Strategic Study
Workshop Series

Algorithm Development for Security Applications

October 2012 Workshop
Final Report

ALERT
AWARENESS AND LOCALIZATION
OF EXPLOSIVES-RELATED THREATS

A Department of Homeland Security Center of Excellence

Northeastern University

Table of Contents

This page intentionally left blank.

1. Executive Summary

A workshop focusing on automated threat recognition (ATR) algorithms for explosion detection systems was held at Northeastern University in Boston on October 24-25, 2012. This workshop was the eighth in a series dealing with algorithm development for security applications[1].

The topic of ATR was chosen for the workshop in order to support the Department of Homeland Security's (DHS) objective of improving the detection performance of existing technologies. Improved detection performance is defined as increased probability of detection, decreased probability of false alarms, lower detected threat mass and an increase in the number of types of explosives detected.

The topics that were addressed at the workshop are as follows:

1. ATR for:
    a. CT-based EDS
    b. Whole body imaging (WBI) & Advanced Imaging Technology (AIT)
    c. Carry-on baggage inspection
    d. Cargo
    e. Trace
    f. Fused systems
2. Risk based screening
3. Behavioral detection
4. Detection explosives implanted in a passenger's body
5. XBS dose
6. Accelerating deployment of third party advances
7. Deterrence

The presentations and discussions concentrated on imaging devices such as CT-based EDS and Advanced Imaging Technology.

The workshop was successful fostering interaction between third parties and vendors, reducing barriers to their working together, now and in the future. It also directly led to increased third party involvement in the development of advanced ATR algorithms. This conclusion is based on anecdotal

---

[1] See www.northeastern.edu/alert/transitioning-technology/strategic-studies/ for additional information on the previous workshops.
See myfiles.neu.edu/groups/ALERT/strategic_studies/ for the final reports for the previous workshops.

evidence of the number of third parties engaging in discussions with vendors during the workshop and the editors' knowledge of third parties consulting for the vendors.

The key findings from the workshop, per the editors of this report, are as follows:

- For an imaging device:
    - ATR should be defined as an operation with images as input and a yes/no decision on the presence of a threat as an output.
    - ATR should include the following steps: segmentation, feature extraction, correction for device imperfections and classification.
- It would be very difficult for a third party to develop, without direct assistance from a vendor, an ATR for a deployed explosive detection device (e.g., an EDS) for the following reasons:
    - Detection requirements are classified.
    - Data from deployed equipment are SSI or classified, and are under export control.
    - There is no publicly available set of images that are representative of challenging ATR problems for explosive detection systems.
    - The business interests of the vendors should be protected.
    - DHS/TSA policies do not allow TSL to test components (e.g., an ATR) separate from a complete scanner.
- Third parties can make advances to ATR by working with data and requirements that are in the public domain. This task could be accomplished through the following steps:
    - Detect a set of benign objects such as peanut butter and rubber sheets.
    - Write detection requirements based on these benign objects.
    - Scan these objects on an equivalent device in a related field. For example, for X-ray based EDS, scan on a medical CT scanner.
    - Provide an environment in which third parties, industry and government can interact.
- The following topics should be considered in detail in the future:
    - Developing and testing ATRs with few training and test samples.
    - Developing metrics for improved performance when the confidence intervals for tests of PD and PFA are large due to small data sets.
    - Funding for researchers from DHS, TSA, government laboratories, and industry.

- Incentives from the TSA for vendors to deploy equipment with improved detection performance. These incentives will lead to the deployment of advanced ATR algorithms.
- Developing ATRs with support for risk-based screening, deterrence and the human in the loop.

2. Disclaimers

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Northeastern University nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation or favoring by the United States government or Northeastern University. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Northeastern University, and shall not be used for advertising or product endorsement purposes.

This document summarizes a workshop at which a number of people participated by discussions and/or presentations. The views in this summary are those of ALERT and do not necessarily reflect the views of all the participants. All errors and omissions are the sole responsibility of ALERT.

This material is based upon work supported by the U.S. Department of Homeland Security under Award Number 2008-ST-061-ED0001.

3. Introduction

The Explosive Division (EXD) of US Department of Homeland Security (DHS) Science & Technology Directorate (S&T), in coordination with the Transportation Security Administration (TSA), has identified detection requirements for future explosive detection scanners that include a larger number of threat categories, lower false alarm rates, lower threat mass and lower total operating costs, all at a constant or increased probability of detection.

One tactic that DHS is pursuing to improve detection performance is to create an environment in which the capabilities and capacities of the established vendors can be augmented or complemented by third-party algorithm development. A third-party developer in this context refers to academia, National Labs and companies other than the incumbent vendors. DHS is particularly interested in adopting the model that has been used by the medical imaging industry, in which university researchers and small commercial companies develop algorithms that are eventually deployed in commercial medical imaging equipment.

A tactic that DHS is using to stimulate academic and industrial third-party algorithm development is to sponsor a series of workshops addressing the research opportunities that may enable the development of next-generation algorithms for homeland security applications. The series of workshops are entitled “Algorithm Development for Security Applications (ADSA).” The workshops are convened by Professor Michael B. Silevitch (NEU) as part of the DHS Center of Excellence (COE) for Awareness and Localization of Explosives-Related Threats (ALERT[2]).

The eighth workshop in the ADSA series was held on October 24-25, 2012, at NEU. The workshop addressed automated threat recognition (ATR) algorithms.

This report discusses what transpired at the workshop and reports a summary of the findings and recommendations.

The workshop was successful fostering interaction between third parties and vendors, reducing barriers to their working together, now and in the future. It also directly led to increased third party involvement in the development of advanced ATR algorithms. This conclusion is based on anecdotal evidence of the number of third parties engaging in discussions with vendors during the workshop and the editors’ knowledge of third parties consulting for the vendors.

---

[2] ALERT in this report refers to the COE at NEU.

4. Discussion

4.1 Objectives

The objective of the workshop was to facilitate the development of improved ATRs for explosive detection equipment. In particular, an objective was to discuss ATRs for the following applications:

1. CT-based EDS
2. Whole body imaging (WBI) & Advanced Imaging Technology (AIT)
3. Carry-on baggage inspection
4. Cargo
5. Trace
6. Fused systems

Most of the workshop dealt with ATRs for imaging devices and in particular for CT-based ATR. Future workshops should address ATRs for non-imaging modalities.

The issues that were addressed centered on the following list of questions:

1. How can and should ATRs be improved?
2. How should the requirement specs for an ATR be established?
3. How should testing of ATRs be modified?
4. How should deterrence and risk-based screening be incorporated into the design of an ATR?
5. How can third parties be involved in the development of improved ATRs?

The purpose of this section is to synthesize the discussion and recommendations in response to these and related questions that surfaced during the discussion.

4.2 ATR Definition

For an imaging device, the following assumptions were made about an ATR:

1. ATR should be defined as a function with images as input and a yes/no decision on the presence of a threat as an output.
2. An ATR should include the following steps: segmentation, feature extraction, correction for device imperfections and classification.

Additional information about these assumptions can be found in the presentation entitled “Review of Automated Threat Detection Algorithms,” which was presented by Carl Crawford at ADSA02.

Some workshop attendees felt that an ATR consists only of the classification step. The editors of this report respectfully disagree because classification for ATR depends strongly on features extracted from the image, and the choice of features requires inclusion of the earlier steps. Some workshop attendees also felt that an ATR should begin with raw data (e.g., projection data for CT-based EDS) as an input. The editors of this report respectfully disagree because reconstruction and ATR require different disciplines and combining these two function would limit the number of third parties who could participate in the development of combined reconstruction/ATR algorithms.

The following topics were briefly discussed in relation to ATRs and need to be addressed by ATR developers in the future:

1. The role of the human in the detection loop (e.g., for OSR).
2. The role of risk based screening and deterrence.

4.3 Barriers for Third Party ATR Development

Ideally, a third-party should have access to training data acquired from a deployed piece of explosive detection equipment and be able to test their ATR at the TSL. It would be very difficult for a third party, without direct assistance from a vendor, to accomplish this goal for the following reasons:

1. Detection requirements are classified.
2. Data from deployed equipment are SSI or classified, and are under export control.
3. There is no publicly available set of images that are representative of challenging ATR problems for explosive detection systems.
4. The business interests of the vendors should be protected.
5. DHS/TSA policies do not allow TSL to test components (e.g., an ATR) separate from a complete scanner.
6. There are privacy concerns with scans on AIT equipment.

4.4 Solution for Third Parties to Develop ATRs

Third parties can make advances to ATR by working with data and requirements that are in the public domain. This task could be accomplished through the following steps:

- Detect a set of benign objects such as peanut butter and rubber sheets.
- Write detection requirements based on these benign objects.
- Scan these objects on an equivalent device in a related field. For example, for CT-based EDS, scan on a medical CT scanner.
- Consider using simulated images.
- Provide ground-truth for the scans or simulated data.
- Provide clarification of requirements based on questions posed by the researchers. Revise the requirements as necessary.
- Test ATRs and provide feedback on results.
- Provide an environment in which third parties, industry and government can interact.
- Allow separate paths (algorithms) for different types of threats and configurations. For example, separate paths could be provided for sheet, bulks, homogeneous objects and textured objects.
- Obtain inputs on this process from industry and TSL.
- Define detection performance based on one of the following standards:
    - Increased PD and for a fixed PFA.
    - Decreased PFA and for a fixed PD.
    - Area under the ROC. Note that a method would have to be developed to be able to generate the ROC curve.

It is recommended that the third parties should first attempt to develop ATRs for CT-based EDS equipment. DHS has provided funding to ALERT to perform this task under a program denoted Task Order 4.

4.5 Other Issues with Third Party ATR Development

The following topics should be considered by researchers and testers in the future:

- Developing and testing ATRs with few samples. This is also known as dealing with statistical significance of training and test samples.
- Preventing, by testers, over-training on test sets.

- What constitutes a permissible set of features in an ATR? For example, can shape be used?
- Developing metrics for improved performance when the confidence intervals for tests of PD and PDF are large due small test sets.
- Developing ATRs with support for risk-based screening, deterrence and the human in the loop.
- Developing ATRs that can be revised in the future to handle emerging threats.
- Developing vendor-neutral ATRs.
- Developing ATRs for fused systems.
- Developing algorithms that eliminate scans without threats to lessen the burden on humans reviewing images.
- Are humans better/worse than ATRs?

4.6 Accelerating Deployment

The following tasks should be performed in order to accelerate the deployment of advanced ATR algorithms, especially those developed by third parties. Many of these tasks are derived from the presentations made by Doug Pearl at ADA07 and ADSA08, and based on the discussion during his presentations.

1. Provide detailed problem statements including:
    a. Short term for vendors and third-party industry
    b. Long term for students
2. Increased incentives from the TSA for vendors to deploy scanners with improved detection performance.
3. Increased incentives for third parties to develop advanced algorithms.
4. Government (DHS/TSA) funding of vendors and third parties.
5. Allowing, if possible, more people access to classified and SSI information or develop non-classified canonical problems capturing ATR challenges.
6. Developing frameworks for protecting:
    a. Intellectual property
    b. Commercial interests of vendors and third parties
7. Reducing transaction costs of working with third parties.
8. Having third parties reduce computational expense of new ATR algorithms. The first of the development of new ATR algorithms should not consider computational expense.

9. Giving third parties access to subject matter expert experts in the field of developing and deploying explosive detection equipment.
10. Fund the science of acceptance criteria (metrics).
11. Modifying acceptance tests (e.g., certification, qualification and CRT) to allow increased involvement of third parties.

4.7 Future ADSA Workshops

1. The following topics should be addressed in future workshops. Note that classification issues may prevent some of these topics from being discussed.
    a. Stand-off detection on personnel and in vehicles
    b. ETD (explosive trace detection)
    c. Chemical sensors
    d. DHS detection problems
    e. Cargo
    f. Special nuclear materials (SNM)
    g. AIT (MMW, XBS) – ATR and reconstruction
    h. Video analytics
    i. Executing grand challenges
    j. New signatures for detecting explosives
    k. Adaptive learning
    l. Combined reconstruction and ATR algorithms
    m. Reducing computational expense of new reconstruction algorithms
2. The following changes should be considered for future ADSA workshops:
    a. More and longer breaks
    b. Presentations
        i. Shorter in number and duration to allow for more discussion.
        ii. Review slides in advance for adherence to presentation methods used at the ADSA workshops. The presentations should not be reviewed for technical content.
        iii. Concentrate on results.
        iv. Obtain permission to release slides in advance.
        v. Provide mentorship to new speakers.
    c. Encourage attendees to stay until the end of the workshop.
    d. Provide abstracts in advance of the workshop to help people decide whether to attend.

5. Acknowledgements

The planning committee would like to thank the following people and organizations for their involvement in the workshop.

- DHS S&T for funding ALERT and sponsoring the workshop.
- Doug Bauer, DHS (retired), and George Zarur, DHS & TSA (retired), for their vision to involve third parties in the development of technologies for security applications.
- Laura Parker, DHS, and Earl Smith, DHS, for coordinating DHS/ALERT activities.
- Northeastern University for hosting the workshop.
- Steve Azevedo, Lawrence Livermore National Laboratory, for taking minutes.

The workshop would not have been a success without the participants, the speakers and the students who presented posters. We extend our heartfelt thanks to them for their contributions.

6. Workshop Planning and Support

The planning committee for the workshop consisted of the following people:

David Castañón, Boston University

Carl Crawford, Csuptwo

Harry Martz, Lawrence Livermore National Laboratory

Michael Silevitch, Northeastern University

The workshop was moderated by:

Carl Crawford, Csuptwo

The final report was assembled and edited by:

Carl Crawford, Csuptwo

The final report was assembled by:

Rachel Parkin, Northeastern University

The final report was reviewed by:

David Castañón, Boston University

Clem Karl, Boston University

Harry Martz, Lawrence Livermore National Laboratory

Logistics for the workshop were led by:

Melanie Smith, Northeastern University

Other logistics, including minute taking and audiovisual assistance, for the workshop were handled by:

Deanna Beirne, Northeastern University

Seda Gokoglu, Northeastern University

Kristin Hicks, Northeastern University

Anne Magrath, Northeastern University

Rachel Parkin, Northeastern University

Can Yegen, Northeastern University

The SSI review was done by:

Horst Wittmann, Northeastern University

7. Appendix: Agenda

7.1 October 24, 2012 - Day 1

TIME	TOPIC	SPEAKER	AFFILIATION
8:30	**Registration/Continental Breakfast**		
9:00	Call to Order	Carl Crawford	Csuptwo
9:05	Welcoming Remarks - ALERT	Michael Silevitch	NEU / ALERT
9:10	Welcoming Remarks - DHS	Laura Parker	DHS
9:15	Workshop Objectives	Carl Crawford	Csuptwo
9:35	ATR for Personnel Screenings	Alex Hudson	Rapiscan
10:00	ATR for Various Modalities	David Perticone	L-3 Communications
10:25	**Break**		
10:50	Open Discussion		
11:35	ATR for Cargo	Sam Song	Telesecurity Sciences
11:40	Feature Extraction in 3D Millimeter-Wave Radar Imaging	Justin Fernandes	Pacific Northwest National Laboratory
12:05	Detection of Liquid and Amorphous Threats in XRD	Sondre Skatter	Morpho Detection
12:30	**Lunch**		
1:15	Threat Detection for Venue Protection	Lisa Sagi-Dolev	Qylur Security Systems
1:40	Computer Aided Detection in Medical Imaging	Robert Nishikawa	University of Chicago
2:05	Clear Bag Concept for Risk Based Screening	Luc Perron	Optosecurity
2:30	Classifier Design for CAXI Project	Jody O'Sullivan	Washington University
2:55	Multi-Stage Decision Systems	Kirill Trapeznikov	Boston University
3:20	X-ray Back Scatter Dose Predictions	Taly Gilat-Schmidt	Marquette University
3:45	Alternative Way for TSA to Acquire Technology	George Zarur	TSA (Retired)
4:10	**ALERT Student Poster Session / Reception Sponsored by Csuptwo**	Students / Carl Crawford	ALERT / Csuptwo
5:10	**Dinner**		
6:00	Effectiveness of Deterrence	Laura Dugan	University of Maryland / START
6:30	Predictive Terrorism Risk for TSA Security Programs	Carter Price	Csuptwo
7:00	**Adjourn**	Carl Crawford	Csuptwo

7.2 October 25, 2012 - Day 2

TIME	TOPIC	SPEAKER	AFFILIATION
07:30	**Continental Breakfast**		
08:00	Day 2 Objectives	Carl Crawford	Csuptwo
08:05	Dynamic ATR	Matthew Merzbacher	Morpho Detection
08:30	ATR - Practical Development Considerations	Richard Bijjani	Robehr Analytics
09:00	EDS Research Problems	Zhengrong Ying	Zomographic
09:20	Aberrant Behavior and Risk Based Screening	Carl Maccario	TSA
09:55	**Break**		
10:30	Discussion: Role of Incentives in Security Imaging	Doug Pearl	Inzight Consulting
11:00	Detection of Implanted Explosives	Steve Azevedo	Lawrence Livermore National Laboratory
11:20	A Math Perspective on Fusion Needs	Ken Jarman	Pacific Northwest National Laboratory
11:45	Fused Sensor System Capabilities and Limitations	Kevin Johnson	Naval Research Laboratory
12:10	Robust Fusion Algorithm for Sensor Failure	Deniz Erdogmus	NEU
12:25	**Lunch**		
12:55	Video Analytics and Anomaly Detection	Venkatesh Saligrama	Boston University
1:15	Imaging Challenges for X-Ray Screening	Brian Tracey, Chris Alvino	Tufts University AS&E
1:35	ECAC Testing	Jean Claude Guilpin	ECAC
2:00	Machine Learning Algorithms for Biomedical Data	Jennifer Dy	NEU
2:25	Low-Rank Analytics for Explosive Detection	Raymond Fu	NEU
2:45	Next Steps	Carl Crawford	Csuptwo
3:40	Closing Remarks - DHS	Laura Parker	DHS
3:50	Closing Remarks – ALERT	Michael Silevitch	NEU / ALERT
4:00	**Adjourn**	Carl Crawford	Csuptwo

Note: The timing in the agenda was only loosely followed due to the amount of discussion that took place during the presentations and to give additional time for participants to network.

8. Appendix: Student Posters

Select posters presented at ADSA08 are available for viewing online at: https://myfiles.neu.edu/groups/ALERT/strategic_studies/ADSA08_posters/

The complete list of student posters presented at ADSA08 is:

STUDENT PRESENTERS	POSTER AUTHORS	PROJECT P.I.'S	POSTER TITLE
Limor Eger / Boston University	Limor Eger / Boston University	W. Clem Karl / Boston University	Classification-aware Methods for Explosives Detection Using Multi-Energy X-ray Computed Tomography
Kirill Trapeznikov / Boston University	Kirill Trapeznikov / Boston University	Venkatesh Saligrama / Boston University David Castanon / Boston University	Multi Stage Classifier Design
Binlong Li / Northeastern University Fei Xiong / Northeastern University	Mustafa Ayazoglu / Northeastern University Caglayan Dicle / Northeastern University Binlong Li / Northeastern University Fei Xiong / Northeastern University	Octavia I. Camps / Northeastern University Mario Sznaier / Northeastern University	Tracking in Large Public Spaces
Caglayan Dicle / Northeastern University Binlong Li / Northeastern University	Mustafa Ayazoglu / Northeastern University Caglayan Dicle / Northeastern University Binlong Li / Northeastern University Necmiye Ozay / Northeastern University	Octavia I. Camps / Northeastern University Mario Sznaier / Northeastern University	Assessment of Complex Threat Scenarios: Behavior Analysis
Borja Gonzalez-Valdes/ Northeastern University Yuri Alvarez / University of Oviedo, Spain	Borja Gonzalez-Valdes/ Northeastern University Yuri Alvarez / University of Oviedo, Spain	Carey Rappaport/ Northeastern University Jose Martinez / Northeastern University	Automatic SAR Processing for Profile Reconstruction and Recognition of Dielectric Objects on the Human Body Surface

STUDENT PRESENTERS	POSTER AUTHORS	PROJECT P.I.'S	POSTER TITLE
Galia Ghazi / Northeastern University	Galia Ghazi / Northeastern University	Carey Rappaport/ Northeastern University Jose Martinez / Northeastern University	Improved Imaging Technique for Automatic Threat Detection
Kang Li / Northeastern University	Kang Li / Northeastern University Jie Hu / Northeastern University	Yun Fu / Northeastern University	Modeling Complex Temporal Composition of Actionlets for Activity Prediction
Ming Shao / Northeastern University	Carlos Castillo / Northeastern University Zhenghong Gu / Northeastern University Ming Shao / Northeastern University	Yun Fu / Northeastern University	Low-Rank Transfer Subspace Learning
Murat Akcakaya / Northeastern University	Murat Akcakaya / Northeastern University Umut Orhan / Northeastern University	Deniz Erdogmus / Northeastern University	Error Dependent Risk Minimization for Detection

9. Appendix: Previous Workshops

Information about the previous seven workshops, including their final reports, can be found at:

www.northeastern.edu/alert/transitioning-technology/strategic-studies

10. Appendix: List of Participants

NAME		AFFILIATION
Richard	Abraham	Machine Vision Analysts
Murat	Akcakaya	Northeastern University
Omar	Al-Kofahi	American Science and Engineering, Inc.
Yuri	Alvarez	Northeastern University
Chris	Alvino	AS&E
Stephen	Azevedo	Lawrence Livermore National Laboratory
Kumar	Babu	Ccuneus Solutions, LLC
Nathaniel	Beagley	Pacific Northwest National Laboratory
John	Beaty	Northeastern University
Moritz	Beckmann	XinRay Systems
Deanna	Beirne	Northeastern University
Richard	Bijjani	Robehr Analytics
Carl	Bosch	SureScan
Emel	Bulat	Northeastern University
John	Bush	Battelle
David	Castañón	Boston University
Terrence	Chen	Siemens
Charles	Choi	General Dynamics AIS
Carl	Crawford	Csuptwo
Andrew	Diamond	Rapiscan Laboratories, Inc.
Caglayan	Dicle	Northeastern University
Synho	Do	Massachusetts General Hospital
Jennifer	Dy	Northeastern University
Limor	Eger	Boston University
Deniz	Erdogmus	Northeastern University
Justin	Fernandes	Pacific Northwest National Laboratory
Yun Raymond	Fu	Northeastern University
Taly	Gilat-Schmidt	Marquette University
Seda	Gokoglu	Northeastern University
Brian	Gonzales	XinRay Systems, LLC

NAME		AFFILIATION
Borja	Gonzalez-Valdes	Northeastern University
David	Goodenough	George Washington University
Jens	Gregor	University of Tennessee
Chris	Gregory	Smiths Detection
Craig	Gruber	Northeastern University
Jean-Claude	Guilpin	European Civil Aviation Conference
William	Hall	Massachusetts Port Authority
Jeffrey	Hamel	IDSS Holdings Inc.
Martin	Hartick	Smiths Heimann
Dominic	Heuscher	University of Utah
Kristin	Hicks	Northeastern University
Matt	Higger	Northeastern University
David	Holden	Rapiscan Systems
Alex	Hudson	Rapiscan Systems
Ken	Jarman	Pacific Northwest National Lab
Olof	Johnson	Photo Detection System, Inc.
Kevin	Johnson	Naval Research Laboratory
Ersel	Karbeyaz	Reveal Imaging Technologies, Inc.
W. Clem	Karl	Boston University
Don	Kim	Transportation Security Administration
Robert	Klueg	Department of Homeland Security
Ronald	Krauss	Department of Homeland Security
Lorena	Kreda	Consultant
Kang	Li	Northeastern University
Binlong	Li	Northeastern University
David	Lieblich	Analogic Corporation
Andrew	Litvin	Analogic Corporation
Chuck	Lloyd	Reveal Imaging Technologies, Inc.
Tony	Macadino	Photo Detection System Inc.
Carl	Maccario	Transportation Security Administration
Michael	Massey	Beth Israel Deaconess Medical Center
Matthew	Merzbacher	Morpho Detection

NAME		AFFILIATION
Christian	Minor	Naval Research Laboratory
Richard	Moore	Massachusetts General Hospital
Robert	Nishikawa	University of Chicago
John	O'Connor	Analogic Corporation
Boris	Oreper	L-3 Communications
Joseph	O'Sullivan	Washington University
Jonathan	Pai	Smiths Detection
Laura	Parker	Department of Homeland Security
Rachel	Parkin	Northeastern University
Tip	Patridge	Netcom
Julia	Pavlovich	Analogic Corporation
Douglas	Pearl	Inzight Consulting
Luc	Perron	Optosecurity
David	Perticone	L-3 Communications
Carter	Price	Rand Corporation
Carey	Rappaport	Northeastern University
Lisa	Sagi-Dolev	Qylur Security Systems, Inc.
Venkatesh	Saligrama	Boston University
David	Schafer	Reveal Imaging Technologies, Inc.
Theodore	Schnackertz	American Science and Engineering, Inc.
Jean-Pierre	Schott	Lawrence Livermore National Laboratory
Anthony	Serino	Raytheon Company
Robert	Shuchatowitz	Reveal Imaging
Michael	Silevitch	Northeastern University
Sergey	Simanovsky	Analogic Corporation
Sondre	Skatter	Morpho Detection
Stephen	Skrzypkowiak	Transportation Security Administration
Adel	Slamani	Quasars
Melanie	Smith	Northeastern University
Steve	Smith	Tek84
Edward	Solomon	Triple Ring Technologies
Serge	Soloviev	Reveal Imaging

NAME		AFFILIATION
Samuel	Song	Telesecurity Sciences
Simon	Streltsov	LongShortWay, Inc.
Greg	Struba	Department of Homeland Security
Ling	Tang	Rapiscan Laboratories, Inc.
Brian	Tracey	Tufts University
Jason	Tracy	Department of Homeland Security
Kirill	Trapeznikov	Boston University
Whitney	Weller	Force 5 Networks, LLC
Dana	Wheeler	Radio Physics Solutions
Alyssa	White	Massachusetts General Hospital
Horst	Wittmann	Northeastern University
Fei	Xiong	Northeastern University
Birsen	Yazici	Rensselaer Polytechnic Institute
Can	Yegen	Northeastern University
Zhengrong	Ying	Zomographic, LLC
George	Zarur	Department of Homeland Security
Francois	Zayek	Unknown
Larry	Zeng	University of Utah
Jun	Zhang	University of Wisconsin-Milwaukee
Margaret	Zhao	Rapiscan Laboratories, Inc.

11. Appendix: Presenter Biographies

Christopher Alvino

American Science and Engineering

Christopher V. Alvino received the B.S. (1998) and M.S. (2001) degrees from Rutgers University and the Ph.D. degree (2005) from Georgia Institute of Technology, all in electrical and computer engineering. During his M.S. degree (1999-2001) he was a signal processing consultant at Sarnoff Corporation in Princeton, NJ where he made important research contributions in microphone arrays, blind source separation, and EEG/MEG signal analysis. Following a postdoctoral fellowship at the University of Pennsylvania from 2005-2006, he joined Siemens Corporate Research (SCR) in Princeton, NJ, as a Research Scientist in medical imaging. In 2011 he joined American Science and Engineering, Billerica, MA as a Senior Scientist in the Image Processing group.

He has made contributions to many areas of both medical imaging and security imaging, including: automated and interactive segmentation, registration, image-based calibration, anomaly detection, and de-noising. His interests in imaging are in both applied optimization methods as well as using machine learning techniques to develop effective computer vision algorithms from large datasets.

Stephen Azevedo

Lawrence Livermore National Laboratory

Dr. Stephen Azevedo is currently Project Engineer for Livermore Explosives Detection Program where he leads R&D efforts in advanced detection systems for aviation security at Lawrence Livermore National Laboratory (LLNL). During his 30+ years at LLNL, he has held a number of technical and leadership positions including Project Leader for National Ignition Facility Shot Data Analysis, Project Leader of the Micropower Impulse Radar (MIR) Project (working on specialized radar systems for various applications including bridge-deck inspection, low-power communications, search-and-rescue, and mine detection) and Deputy Division Leader. His interests have been in the areas of computational signal and image processing research, including computer algorithms, numerical methods, languages, display techniques, and inspection imaging. For eight years, he was Director of the Center for Advanced Signal and Image Sciences (CASIS), and has been on the International Scientific Advisory Committee for the ICALEPCS conference series. He has been a principal investigator for computed tomography research and radar remote sensing, X-ray inspection, nondestructive evaluation and imaging. He has earned four R&D 100 awards for technical excellence.

Dr. Azevedo graduated with his B.S. in Electrical Engineering from U. C. Berkeley in 1977 and received a Masters in E.E. and Biomedical Engineering from Carnegie-Mellon University in 1978. He earned his Ph. D. in 1991 from U. C. Davis (EECS) for his research in model-based tomographic reconstructive imaging. He has been employed at LLNL since 1979.

Richard Bijjani

Robehr Analytics

Dr. Richard Robehr Bijjani has been a thought leader in security technology for over 20 years. He designed and developed many security products including a dozen different Explosive Detection Systems (EDS) utilizing various technologies. The systems he designed managed to successfully exceed the certification requirements of every known EDS detection standard in the world; a unique achievement. In 1990, Richard managed R&D during the development of a dynamic signature verification product at Kumahira Inc., one of the very first biometrics products in the industry. In 1994, he joined InVision Technologies as head of the Algorithm and Machine Vision group where he oversaw the development effort that led to the first successful certification by the FAA, a historic event for the then still nascent industry. He went on to design and certify multiple EDS systems for InVision (now Morpho Detection) and later for Vivid (now L3). In 2002, he co-founded Reveal Imaging (now an SAIC company) where he designed and developed the world’s highest performing automated explosive detection systems to date, which also happen to be the least expensive and the smallest. In January 2012, Richard founded Robehr Analytics where he plans to develop a suite of low cost sensors that he hopes would revolutionize the way people interact with their environment and help enhance and protect their lives. Dr. Bijjani has a Ph.D. in Electrical Engineering from Rensselaer Polytechnic Institute.

Carl Crawford

Csuptwo

Dr. Carl Crawford is president of Csuptwo, LLC, a technology development and consulting company in the fields of medical imaging and homeland security. He has been a technical innovator in the fields of medical and industrial imaging for more than 25 years. Dr. Crawford was the Technical Vice President of Corporate Imaging Systems at Analogic Corporation, Peabody, Massachusetts, where he led the application of signal and image processing techniques for medical and security scanners. He developed the reconstruction and explosive detection algorithms for the Examiner 6000, a computerized tomographic (CT) scanner deployed in airports worldwide. He was also employed at General Electric Medical Systems, Milwaukee, Wisconsin, where he invented the enabling technology for helical (spiral) scanning for medical CT scanners, and at Elscint, where he developed technology for cardiac CT scanners. He also has developed technology for magnetic resonance imaging (MRI), single photon emission tomography (SPECT), positron emission tomography (PET), ultrasound imaging (U/S), and dual energy imaging and automated threat detection algorithms based on computer aided detection (CAD). Dr. Crawford has a doctorate in electrical engineering from Purdue University, is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and an associate editor of IEEE Transactions on Medical Imaging.

Laura Dugan

University of Maryland

Laura Dugan is an Associate Professor in the Department of Criminology and Criminal Justice at the University of Maryland; and is an active member of the National Center for the Study of Terrorism and the Response to Terrorism. Her research examines the consequences of violence and the efficacy of violence prevention/intervention policy and practice. She also designs methodological strategies to overcome data limitations inherent in the social sciences. Dr. Dugan is a co-principal investigator for two important event-based datasets: the Global Terrorism Database (GTD) and the Government Actions in Terrorist Environments (GATE) dataset. The GTD is the most comprehensive source of terrorist incidents, as it records all known attacks across the globe since 1970. The GATE data record government actions related to terrorists and their constituencies for a select set of countries since 1987. Collection on both datasets is on-going. Dr. Dugan's research has been published in top journals in criminology and sociology. She has also published in political science and public policy journals. She received her Ph.D. in Public Policy and Management from Carnegie Mellon University in 1999.

Jennifer Dy

Northeastern University

Jennifer G. Dy is an associate professor at the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, where she first joined the faculty in 2002. She received her M.S. and Ph.D. in 1997 and 2001 respectively from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, and her B.S. degree (Magna Cum Laude) from the Department of Electrical Engineering, University of the Philippines, in 1993.

Her research is in machine learning, data mining and their application to computer vision, health, security, science and engineering, with a particular focus on clustering, multiple clusterings, dimensionality reduction, feature selection and sparse methods, large margin classifiers, learning from the crowds and Bayesian nonparametric models. She received an NSF Career award in 2004. She is an action editor for the journal, Machine Learning since 2007, an editorial board member of the Journal of Machine Learning Research since 2009, organizing/senior/program committee member for ICML, ACM SIGKDD, AAAI, IJCAI, AISTATS and SIAM SDM, and program chair for SIAM SDM 2013.

Deniz Erdogmus

Northeastern University

Deniz Erdogmus received B.S. degrees in EE and Mathematics in 1997, and M.S. in EE in 1999 from the Middle East Technical University, Ankara, Turkey. He received his Ph.D. in ECE from the University of Florida in 2002, where he stayed as a postdoctoral research associate until 2004. He was an Assistant Professor of Biomedical Engineering at the Oregon Health and Science University until 2008. Then he joined Northeastern University, where he is currently an Associate Professor in the Electrical and Computer Engineering Department. His research focuses on statistical signal processing and machine learning with applications to contextual signal, image, and data analysis with applications in cognitive signal processing including brain computer interfaces and technologies that collaboratively improve human performance. He has over 75 journal publications and he has served as an associate editor and program committee member for a number of journals and conferences in these areas, including IEEE Signal Processing Letters, and the following IEEE Transactions: Sig nal Processing, Biomedical Engineering, and Neural Networks.

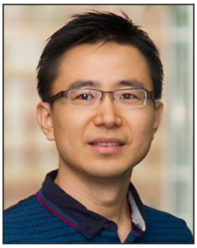
Justin Fernandes

Pacific Northwest National Laboratory

Justin L. Fernandes was born in Denver, Colorado, in 1985. He received a Bachelor of Science degree in Electrical and Computer Engineering and Master of Science degree in Electrical Engineering from Northeastern University in Boston, Massachusetts. From 2010 to present he has worked at Pacific Northwest National Laboratory as a scientist in the Electromagnetics Team under the Applied Physics Group. His research interests include three dimensional synthetic aperture radar, computational electromagnetics, and signal processing.

Yun Raymond Fu

Northeastern University



Dr. Fu is an interdisciplinary faculty member affiliated with College of Engineering and the College of Computer and Information Science at Northeastern University. He received the B.Eng. degree in Information Engineering and the M.Eng. degree in Pattern Recognition and Intelligence Systems from Xi’an Jiaotong University, China, respectively, and the M.S. degree in Statistics and the Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign, respectively. Prior to joining the Northeastern faculty, he was a Scientist working at BBN Technologies, Cambridge, MA, during 2008-2010. He holds a Part-Time Lecturer position in the Department of Computer Science, Tufts University, Medford, MA, in 2009. He was a tenure-track Assistant Professor of the Department of Computer Science and Engineering, State University of New York, Buffalo, during 2010-2012.

Dr. Fu’s research interests are Interdisciplinary research in Machine Learning, Social Media Analytics, Human-Computer Interaction, and Cyber-Physical Systems. He has extensive publications in leading journals, books/book chapters and international conferences/workshops. He serves as associate editor, chairs, PC member and reviewer of many top journals and international conferences/workshops.

Taly Gilat-Schmidt

Marquette University

Taly Gilat Schmidt, Ph. D., is an assistant professor of Biomedical Engineering at Marquette University. Her research interests include medical imaging system design, optimization, and reconstruction. Dr. Schmidt earned an undergraduate degree in Electrical Engineering from the University of Illinois at Urbana Champaign, after which she was employed in the Edison Engineering Program at GE Healthcare. Dr. Schmidt received her M.S. and Ph. D. in Electrical Engineering from Stanford University. She directs the Medical Imaging Systems Laboratory at Marquette University, which is currently conducting research funded by the NIH, DOE, and GE Healthcare.

Jean Claude Guilpin

European Civil Aviation Conference

Mr Jean-Claude Guilpin works since 1997 for the French civil aviation general directorate, in the civil aviation technical department (STAC), which is in charge of the certification of security equipment and canine teams to be used at French airports (www.stac.aviation-civile.gouv.fr).

From 2006 to 2011, he chaired the group of experts of technical aspects of civil aviation security (so called “Technical Task Force”) of the European Civil Aviation Conference (ECAC), and participated in the outcome of the ECAC Common Evaluation Process of security equipment. He also works closely with experts groups of the European Commission and is involved in several coordination activities with others French governmental laboratories working on technical aspects of homeland security.

Alex Hudson

Rapiscan Systems

Alex Hudson is the VP of Global Engineering for Rapiscan Systems Inc. Previously Technical Project Manager on the RTT project for Rapiscan Laboratories Inc. Prior to Rapiscan, Dr. Hudson worked as an R&D Manager in Advanced Development at Varian Inc. Before this he worked as the Supervisor of the Advanced Systems Design Group with Quantum Magnetics (a subsidiary of InVision Technologies, now Morpho Detection). Dr. Hudson has thirteen years of high tech product development experience, with 8 in the field of aviation security, developing technologies and sensors for various applications based on quadrupole resonance (QR), magnetic resonance (MR), computed tomography (CT), line scan X-ray and data fusion. At Varian, Inc. his role was to lead a research group, developing cutting-edge cryogenic RF antenna products and to manage a portfolio of R&D projects created to deliver competitive new magnetic resonance spectroscopy systems. While at Quantum Magnetics, Dr. Hudson was Principal Investigator of a multi-million dollar Quadrupole Resonance (QR) explosive detection grant funded by the Transportation Security Laboratory. As part of this work, he developed a safe test material for QR explosive detection machines, in collaboration with LLNL, which is now commercially available from XM Products. Dr. Hudson holds a BS in Physics from Bristol University, UK and a PhD from Nottingham University, UK in Magnetic Resonance Imaging.

Ken Jarman

Pacific Northwest National Laboratory

Ken Jarman is a Senior Research Scientist at the Pacific Northwest National Laboratory. He holds a Ph.D. (2000) and M.S. (1998) in Applied Mathematics from the University of Colorado. Ken's research focuses on mathematical and statistical techniques for modeling, simulating, and analyzing a variety of threat detection scenarios, including transport of illicit radioactive sources, standoff explosives detection, and techno-social networks of violent non-state actors. Different aspects of this research involve development of models of mathematical physics of novel detection systems, statistical characterization and Monte Carlo simulation of threat/non-threat scenarios, model and sensor data integration, and decision-theoretic analysis. The goal of this research is to quantify and improve the performance of detection systems in the midst of a wide variety of confounding information.

Kevin Johnson

Naval Research Laboratory

Photo
not available

Dr. Kevin Johnson is a staff scientist at the Naval Research Laboratory in Washington D.C. He earned his Ph. D. in Analytical Chemistry from the University of Washington, where his research centered on development of techniques for high-speed gas chromatography coupled with chemometric analysis algorithms. His current research areas are generation and characterization of complex trace vapor mixtures, data fusion algorithms for chemical sensors and instrumentation, and chemometric algorithm development for chemical sensor data.

Carl Maccario

Transportation Security Administration

Photo
not available

Carl is a graduate of Suffolk University in Boston, Massachusetts. He received his Bachelor of Science in 1982. Prior to 9/11/01, Carl's served the Commonwealth of Massachusetts Secretary of State's office as an investigator/auditor for the Securities Division. While employed there he was the coordinator for in-service training. He attended the Massachusetts State Police Academy for Basic Interview Training. He also attended Interview and Interrogation Training at the Essex County House of Corrections, Massachusetts run by the Massachusetts' State Police. As an investigator Carl conducted hundreds of field interviews and audits as well as numerous investigations regarding possible securities fraud. Subsequent to 9/11, Carl left his employment with the State and began a career with Virgin Atlantic Airlines Security as a passenger profiler. He received training in Behavior Pattern Recognition, Document ID Verification, Deception Detection and Eliciting responses from an Israeli security firm hired by Virgin Atlantic.

Shortly after the DHS was created, Carl began his career with the Federal Government utilizing his knowledge and security experience to help design, develop and implement the first Behavior Screening Program for a major international airport which is now being implemented in airports across the United States, and has trained hundreds of security and law enforcement professionals in Suspicious Behavior Detection, detecting deception and eliciting responses. Carl is a certified instructor in Evaluating Truthfulness and Detecting Deception by the Ekman Group, and has attended the Federal Law Enforcement Training Center as a guest of the US Customs and Border Protection to participate in their Detecting Deception and Eliciting responses Training. Carl was recently a guest speaker on detecting deception and is a member of the FBI Behavior Sciences Unit's T.R.A.P. (Terrorist Research and Analysis Project), and is actively working with British counterparts on their Behavior Detection program at London Heathrow Airport and with various other countries interested in behavior detection.

Matthew Merzbacher

Morpho Detection

Dr. Merzbacher is manager of the Machine Vision and Innovation group at Quantum Magnetics (part of the SAFRAN group's Morpho Detection). In addition to managing the group, Dr. Merzbacher works on technical projects, such as break-bulk cargo, DICOS, and the detection algorithms for the MDI family of explosives detection systems. He is chair of the NEMA DICOS Threat Detection Working Group, charged with developing a standard for image interchange in security applications. He joined what was, at the time, InVision Technologies in 2003 as a Research Scientist in the Machine Vision group. Dr. Merzbacher has a Ph.D. in Computer Science from UCLA, specializing in data mining. He has several pending patents on image processing for explosives detection.

Robert Nishikawa

University of Chicago

Robert M. Nishikawa received his B.Sc. in physics in 1981 and his M.Sc. and Ph.D. in Medical Biophysics in 1984 and 1990, respectively, all from the University of Toronto. He is currently an Associate Professor in the Department of Radiology and the Committee on Medical Physics at the University of Chicago. He is director of the Carl J. Vyborny Translational Laboratory for Breast Imaging Research. He is also a fellow of the American Association of Physicists in Medicine (AAPM). His research interests are in computer-aided diagnosis, breast imaging, image quality assessment and evaluation of medical technologies.

Jody O'Sullivan

Washington University

Joseph A. O'Sullivan (F'03) joined the Department of Electrical Engineering at Washington University in 1986, and is now the Samuel C. Sachs Professor of Electrical Engineering. He has joint appointments in the Departments of Radiology and of Biomedical Engineering. He is Dean of the University of Missouri-Saint Louis/Washington University Joint Undergraduate Engineering Program; in this capacity, he sits on the Provost Council at the University of Missouri-Saint Louis. He was Chair of the Faculty Senate Council and Faculty Representative to the Board of Trustees at Washington University 2002–2004. His research interests include information theory, information-theoretic imaging, recognition theory and systems, CT imaging, optical imaging, information hiding, and hyperspectral imaging.

Prof. O'Sullivan was the Publications Editor for the IEEE Transactions on Information Theory, 1992–1995, was the Associate Editor for Detection and Estimation, and was a Guest Associate Editor for the 2000 Special Issue on Information Theoretic Imaging. He was co-chair of the 1999 Information Theory Workshop on Detection, Estimation, Classification, and Imaging. He was local arrangements chair for the IEEE 2003 Statistical Signal Processing Workshop. He was co-chair of the IEEE 2006 International Symposium on Information Theory. He was chair of the Saint Louis Section of the IEEE in 1994. He is a member of Eta Kappa Nu, SPIE, SIAM, AAAS, and ASEE. He was awarded an IEEE Third Millennium Medal.

Laura Parker

Department of Homeland Security

Laura Parker is in the Explosives Division of the Science and Technology Directorate at the Department of Homeland Security (DHS). She works on the Basic Research Program within the Explosives Division to identify critical and enabling science and technology (S&T) to improve S&T customer capabilities to prevent, detect, respond, and mitigate explosives threats. She also has management responsibility for the DHS-sponsored university-based Center of Excellence that addresses explosive threats through fundamental research that is co-lead by Northeastern University and University of Rhode Island.Prior to her present position at DHS, Dr. Parker worked as a contractor providing technical and programmatic support of chemical and biological defense and explosives programs for various Department of Defense (DoD) offices. Dr. Parker has also worked in several DoD laboratories in the field of energetic materials. She obtained her Ph.D. from the Pennsylvania State University in chemistry.

Doug Pearl

Inzight Consulting LLC

Doug Pearl has examined the role of third party involvement and DICOS in the security industry, in party by examining the role of third party involvement and DICOM in the medical industry. He also has extensive experience in the biomedical industry and in the commercial applications of medical diagnostics. He has written on the problem of False Positives in the screening of low risk (low prevalence) populations. He has provided strategy and marketing advice to a variety of biomedical clients, including Fortune 500, public biotechnology and development stage start-up companies. He has extensive experience working with clinicians, scientists and customers to determine key drivers of success in the marketplace, and parallel experience working with senior management, marketing, and R&D to transform this information into relevant actions.

Prior to launching Inzight Consulting LLC (formerly Insight Consulting) in 1993, Doug Pearl was Vice President, Business Development for Matritech, Inc., a then public biotechnology company in Cambridge, MA. Prior to Matritech, he was a consultant at Bain & Company in Boston. Mr. Pearl has a Masters in Management from the Yale School of Management and an undergraduate degree, summa cum laude, from Princeton. He has also worked as a Research Associate at the Harvard School of Public Health.

Luc Perron

Optosecurity

As the Vice-President of Product Management at Optosecurity, Mr. Perron is a directly responsible for Optosecurity's strategic product roadmap. He ensures the liaison between client requirements and product development and often participates in operational trials. Mr. Perron started his career as an Aerospace Engineer in the Canadian Armed Forces and retired with the rank of Major after 20 years of service. During his military career, he occupied several management positions related to the field of software engineering or imaging, including the direction of a Digital Image Processing laboratory for the Military Intelligence in Ottawa and the direction of the Canadian Forces Imaging Test and Evaluation Laboratory, also in Ottawa. In his last military assignment, he was responsible for all software development on board the CP-140 Aurora Maritime Patrol and anti-submarine aircraft. He later became an associate director for DMR Consulting, a Division of Fujitsu, where he lead several high profile IT projects in content management such as the backlog conversion operation for the Quebec Land Titles project.

David Perticone
L-3 Communications

Carter Price
Rand Corporation

Carter C. Price (Ph.D. Applied Mathematics, University of Maryland College Park) is an associate mathematician at the RAND Corporation. While at RAND, Dr. Price has applied modeling, simulation and data mining techniques to a wide variety of problems including both domestic and national security projects. Recent projects include an assessment of the risk models used by the TSA, a study of Unmanned Ground Sensor technology for use by the U.S. Army, and an NIJ study assessing the use of predictive policing. He has also done qualitative work for an assessment of force protection technology for the Army and target tracking technology.

Lisa Sagi-Dolev
Qylur Security Systems

Venkatesh Saligrama

Boston University

Venkatesh Saligrama is a Professor in the Electrical and Computer Engineering Department at Boston University. He holds the Ph.D. degree from MIT. His research interests are in Statistical Signal Processing, Statistical Learning, Video Analysis, Information and Decision theory. He has edited a book on Networked Sensing, Information and Control. He is currently serving as an Associate Editor for IEEE Transactions on Information Theory. He has previously served as an Associate Editor for IEEE Transactions on Signal Processing and has been on the Technical Program Committees of several IEEE conferences. He is the recipient of numerous awards including the Presidential Early Career Award (PECASE), ONR Young Investigator Award, and the NSF Career Award.

Michael Silevitch

Northeastern University

Michael B. Silevitch is currently the Robert D. Black Professor of Electrical and Computer Engineering at Northeastern University in Boston and an elected fellow of the IEEE. His training has encompassed both physics and electrical engineering disciplines. An author/co-author of over 65 journal papers, his research interests include laboratory and space plasma dynamics, nonlinear statistical mechanics, and K-12 science and mathematics curriculum implementation. Of particular interest is the study of the Aurora Borealis, one of nature's most artistic phenomena. Avocations include long distance hiking and the study of 17th Century clocks and watches.

Prof. Silevitch is also the Director of the Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems (Gordon-CenSSIS), a graduated National Science Foundation Engineering Research Center (ERC). Established in September of 2000, the mission of Gordon-CenSSIS is to unify the methodology for finding hidden structures in diverse media such as the underground environment or within the human body. More recently the CenSSIS multidisciplinary enterprise helped lay the foundation for the research and education programs in the Homeland Security Center of Excellence for Awareness and Localization of Explosives Related Threats (ALERT). This Center was funded in 2008 and is co-directed by Prof. Silevitch.

Sondre Skatter

Morpho Detection

Sondre Skatter is Manager of Research and Development in the Newark office of Morpho Detection, Inc. He received the Diploma degree in physics from the Norwegian University of Science and Technology and a Ph.D. from the Norwegian University of Life Sciences. Sondre joined InVision Technologies, Inc., in 1998 to start the adaptation of CTX technology to the wood industry (WoodVision). Sondre led development of data fusion for the systems-of-systems for the QRCT project and the Phoenix XRD program, integrating the CTX 9000 DSi™ with the XRD 3500™. He is currently the principal investigator on the Next Gen XRD program (HSHQDC-11-C-00014) and program manager for MDI's dual energy program.

Sam Song

TeleSecurity Sciences

Samuel M. Song, Ph. D., received the S.B., M.S. and Ph.D. degrees from MIT, UCLA and USC, respectively, all in electrical engineering. From 1983 to 1991, at Hughes Radar Systems Group, as a recipient of Hughes Doctoral Fellowship, he was the lead designer of several radar signal processing algorithms for a number of different radar modes such as search/track and mapping. From 1992 to 1993, at Stanford University, he developed medical image processing algorithms for MRI and CT images. From 1994 to 1995, at UCSF, he led the development of a mini-PACS system for archiving digital radiographic images. From 1995 to 2001, he was a co-director of Communications and Signal Processing Laboratory at Korea University and from 2001 to 2005, he was the director of Visualization Systems Laboratory at Seoul National University. At the two institutions, he advised some thirty graduate students in the field of signal and image.

From 2000 to 2002, he was a technical consultant to AccuImage Diagnostic Corp. responsible for developing a rectangular slab based projection engine for visualizing 3-D medical images. His improvements resulted in over quadrupling the rendering speed which was essential for real-time feedback to the operator. In 2004, during a sabbatical leave, he was a visiting scientist at Rapiscan Security Products, Hawthorne, CA, where he assisted the development of several product lines including the currently deployed multi-view X-ray scanner and X-ray diffraction based threat detection and classification system. Since 2006 he has been the Chief Technology Officer at TeleSecurity Sciences, Inc., Las Vegas, NV, where he has been developing vendor-independent workstation and algorithms for security applications. He is the current Principal Investigator for several development programs at DHS. Dr. Song has authored some hundred peer-reviewed articles, holds seven US Patent has several others pending. He is a member of IEEE, IEEK, KICS, and Eta Kappa Nu. He is also a current voting member of the NEMA-DICOS standards committee.

Brian Tracey

Tufts University

Brian H. Tracey received his Ph.D. in oceanographic engineering (ocean acoustics and signal processing) from MIT/WHOI in 1996. Subsequently he has worked as an acoustical consultant, a member of the technical staff at MIT Lincoln Laboratory (1999-2004), and technical manager for algorithm development at Neurometrix, Inc., a Boston-area medical devices manufacturer (2005-2011). He joined Tufts University as a Research Assistant Professor in February 2011, where he teaches DSP and is currently working on projects including patch-based denoising, image processing for X-ray backscatter systems, and dual-energy computed tomography.

Kirill Trapeznikov

Boston University

Kirill is a PhD candidate in Electrical Engineering working with Prof. Venkatesh Saligrama and Prof. David Castañon in the Information Systems and Sciences Lab at Boston University. He received his BS and an MS in Electrical Engineering from Boston University in 2007 and 2010 and expects to graduate in Spring 2013. Kirill's current research deals with reducing costs in different aspects of machine learning. His work is applied to explosive detection related tasks under ALERT. His other areas of interest are supervised, semi-supervised and unsupervised machine learning: theory and algorithms, statistical signal processing in image reconstruction and inverse problems, and optimization methods. In the past, Kirill has worked on automated alignment and surface characterization in concentrated solar power dish systems at Sandia National Laboratories in Albuquerque, New Mexico.

Zhengrong Ying

Zomographic, LLC

Dr. Ying has been a technical innovator in the areas of CT (Computed Tomography) systems engineering, image reconstruction, image visualization, object detection, and next generations of CT technologies since 2002. Dr. Ying has been performing specialized consulting in medical and security imaging product, technology and business developments at Zomographic LLC since 2009. Dr. Ying obtained his Ph.D. in Electrical Engineering from Boston University in 2002, and MSEE and BSEE from Shanghai Jiaotong University in 1997 and 1994 respectively.

George Zarur

TSA (retired)

12. Appendix: Questionnaire

ADSA attendees were asked to fill out a questionnaire providing feedback on the workshop. The questions are listed below; the answers appear in the next section. Responses are grouped by question and then by person; the first respondent is response A for each question, the second respondent is B, and so on.

1. How can and should Automated Threat Recognitions (ATRs) be improved?
2. How should the requirement specs for an ATR be established?
3. How should testing of ATRs be modified?
4. How should deterrence and risk-based screening be incorporated into the design of an ATR?
5. How can third parties be involved in the development of improved ATRs?

6a. What did you like about this workshop?

6b. What would you like to see changed for future workshops?

7. Do you have recommendations for future workshop formats?
8. What other comments do you have?

13. Appendix: Questionnaire Responses

Question 1: How can and should Automated Threat Recognitions (ATRs) be improved?

A Focus more on the data. MMW is the only modality being used. A diversified set of fundamental techniques needs to be researched at universities (in any reflectance based 3D imaging) in efforts to build resilient feature vectors. 3D range scanners like the Kinect are a great way to start this research.

B
a. multi-modality sensing
b. Fast processing, error minimization
c. Accurate decision making

C The moderation of discussions should be more strict.

D ATR for explosives detection should be improved. How is the question; first, there needs to be improved identification of objects of interest through segmentation, hopefully enabled by better image reconstruction. Second, there needs to be better features selected that separate the threat versus non-threat classes; this includes potentially new signatures (e.g. multispectral CT, X-ray diffraction, others) as well as better analysis of existing signatures. Third, and possibly least important, one needs improvements in robust classifier design, that can extrapolate from limited training data sets and maintain performance on diverse test sets. Finally, one needs ATR designs that are evolving and adaptive as the threat changes, where this adaptation is semi-automated.

E Abstract book available ahead of time.

F ATRs should factor include learning algorithms so that they will improve while in the field.

G I suggest having academic researchers working directly with the industrial manufacturers. The manufacturers have invested tens of millions of dollars in recon, detection and ATR so to expect them to enable third parties to access their data & knowledge with the possible reduction/elimination of revenue is not realistic or feasible. Having one-on-one coordination will enable advances in ATR capabilities to improve security while ensuring fiscal accountability.

H It's already very good.

I No response.

J The main obstacle to the development of better ATR in the US is TSA itself. The current procurement process provides no incentive for equipment manufacturers to develop anything better than the specified minimum standard, it focuses on the individual equipment capabilities rather than the overall screening process or operational efficiency and it leaves no room for innovation (i.e. doing anything different from what is being asked for) or third party involvement.

K - Better organized and more focused sessions.
- Opening with questions (as done) and closing with discussions on the answers to the questions (not done due to time constraints.)
- Availability of presentation material beforehand so that it could be used for note taking and annotation during the presentations.

L I'm quite new to the field, but it seems that fusion is the next evolution of ATR. Threat classes detected from different modalities seem dependent (at least ... in my naive viewpoint). A fusion system could offer the best of all modalities, or at least an active recognition of dependencies could yield better modeling. Of course, fusion methods require a sharing of data which has its own concerns (proprietary and validation).

M ATRs can be improved through several channels. From a technical perspective, ATRs can be improved through the development of and/or fusion of different technologies to improve the overall PD and PFA. From an application perspective, ATRs can be improved through risk-based screening - dynamic screening levels based upon the perceived threat. From an innovation perspective, ATRs can be improved through programs that drive or foster innovation, or more specifically, provide incentive for vendors and third party developers to improve performance.

N Better definition of a long-term plan for what needs to be detected.

O A lot of the comments during the workshop were related to getting more/better data. As someone who performs developmental test and evaluation on these systems I am concerned about over-training of ATR. We need to look closely at what type and amount of data is necessary to train ATR algorithms to make them effective in the task of general screening and not over-trained to any particular data set. After hearing the talks and related discussions, this issue appears to be more of a challenge than many people in the community realize.

P - Set goals
- Set incentives
- Encourage vendors to work with appropriate third parties
- Find public domain problems that third parties can work on, that are analogous to the issues faced by security vendors

Q Since not marking a threat as such will cause it to be missed, greater PD is more important that lower PFA. A tiered approach could perhaps first filter out the truly clean bags and then use a different method to separate what appears to be a threat from what is a threat. This, of course, is more easily said than done.

R I hope we can get more feedback from TSL after submission. If possible, it will be nice for the TSL (IT&E & DT&E) & vendor to work together to define the scopes of the ATR certification requirements and processes.

S No response.

T We are currently observing the status quo solutions within TSA's airports. From a system or systems perspective growing levels of discontent can be heard here and there. The present systems -- in total, are too expensive, too difficult to technically refresh, too manpower intensive, bulbous (ugly even) in appearance, all contributing to growing air traveler irritation. Then, there is the radiation issue. We know that if the topic (e.g. radiation) is even mentioned TSA and the airline industry have lost. More and more travelers are driving or taking the train. We should all never lose sight that the aviation industry has always needed the benefit of government support; more than just guaranteed air mail contracts. So, are our present day non-integrated systems and procedures -- which are working, the equivalent of the bi-plane with fabric covered surfaces? This current paradigm is not going to change within the next 5 years; maybe 10 years (my guess). But what about the time period after that? The ATRs of ten years from now should address solutions to both better aviation security and also air traveler convenience (e.g. less inconvenience). From an algorithm perspective, faster computing will always be a benefit. OK, the present day CONOPS dictates the air traveler literally strip; no shoes, no 'nothing' in our pockets or wrists, etc. What is air traveler convenience? Somehow we do not need to take everything out of our computer bag but the computer. (Mine is really full; and no one ever looks at what's inside; Same with the PCs.) When was the last time your PC was looked at, turned on? The point here is that smaller, incremental improvements have occurred. So, what are the needed small incremental improvements; which are opportunities? The point of this is that the role of ATRs fits into this first generation system – system. How close are we to the true next generation system-system (or total security solution)? The TSA needs better ATRs that progressively improve the status quo; and do so at minimal cost; and let us not forget the DHS / TSA does not have a lot of funds.

U No response.

V Understand where current ATR, in some specific application, does not meet the requirements/expectations of regulators:
- Assess if those expectations are realistic, and, if so, I would suggest assembling a group of industry and select academic investigators with expertise in the specific application, and in more general ATR, for a CLASSIFIED information exchange with the government as intermediary. This meeting, or set of meetings, would be aimed at identifying the critical issues that hinder current technology from achieving what appear to be realistic expectations of ATR.
- A final meeting or set of meetings would be aimed at defining a program and process to best address the critical issues and bring the technology of vendors to the desired level of ATR performance, in each specific application area.

W It should be modularized for the customer, where a regulator could decide what threats to turn on/off without having to go back to the vendor. The vendors would need to be able to quantify the performance of their systems for each sub- category and understand the issue of cross-alarms for detection and FA impact. Third party developers should be allowed to participate on improving the performance on any of the sub-categories. ATR should be improved by moving away from finding explosives based on statistics but rather on finding explosives based on theoretical range of properties.

X By taking into account social science.

Y Better methods to transfer "best practices" to establish the scope and statistical relevance of training data. Most of the presentations included examples where training data was subject to some unintended bias (thin vs. obese subjects, selective threat configurations...). Given the vast possible combinations of threat material morphology, clutter, concealment, environment, etc. it is essential to have a good methodology for establishing a credible, diverse and statistically relevant set of training data.

Z We need to have a better understanding of the physical features of the explosives and develop technologies to detect these unique features. What is the unique signature of the explosives? A pretty picture may not increase the feature detectability.

AA More development data. Correlation of development data with test data. Modification of regulator testing to be statistically valid and representative of stream of commerce.

AB Improved threat definition: more types of threats and configurations and ample access to data.

AC A combination of improved image processing (preprocessing, reconstruction, segmentation), experience with threat signatures, and probabilistic detection software. All data available to the ATR stage should be used to improve performance.

AD 3 fronts: Improved acquisition and improved calibration/reconstruction along with image analysis intelligently partnered with the two preceding processes. The primary challenge is to develop the partnership between all three processes.

AE Need to develop incentives for companies to improve their ATR systems. One way might be to raise the bar for passing a little bit each year in a similar manner to how they raise the required gas mileage for car companies. As described at the meeting, government testing seems to be statistically underpowered because only a small number of scanned images are used. The FDA requires hundreds of images for evaluating the performance of CAD algorithms.

AF Only through a combined effort will ATR's be improved to the level that's acceptable for US and European airline safety. The TSA and ECAC should make sure that the evaluation of the procurement are aligned enough to give vendors incentives. They should also consider "tiering" the acceptances levels (particularly for PFA) to give vendors extra incentive to do well on the tests. Vendors should have the proper avenues open for them to get help from third parties in a efficient and productive way. ALERT and TSA can help this by making sure the appropriate people at vendors have security clearance to understand the explosives and how they appear in the different sensor modalities. National labs are doing a great job and great technical work, but often seem to be working in an open-loop and not in a way that's so connected with vendors. Whether this is the fault of the vendors or is simply a lack of communication is unclear. All in all there were a lot of great ideas in the meeting, but perhaps too many open-ended philosophical discussions. If these were streamlined a little better so that the goal was more to get these questions answered rather than simply open up questions it might have been even more productive.

AG The presentations outlined the industry's progress in ATR. There was a very small representation from the government. It would be good if DHS S&T or TSL be present; first to view the opinion, progress and suggestions from the industry, and second to present their opinion and expectation from the industry for the future.

AH No response.

AI This is a tough question to answer because it's so application-specific. One thing several people brought up (which I agreed with) was the role of computer assist - flagging something for operator review is a more doable problem.

AJ Facilitated evolution without market/purchase barriers that keep incumbents incumbent. Make sterile sets of data with ground truth available.

AK I liked the presentation on dynamic ATRs. We have different threats and customers. We need a way to control false alarms as we increase the threat vectors.

AL Dual energy projection and CT X-ray.

AM No response.

Question 2: How should the requirement specs for an ATR be established?

A TSA does this. I don't see why we need to change them if what TSA is doing is working.

B No response.

C The requirements should be established through mutual discussions.

D This is way out of my depth. Given a context problem domain, and a sufficiently rich representative set of typical data represented in feature space, the potential performance of any classifier is defined by the distribution of the different classes of data over the domain. Unfortunately, there is no simple way of defining the relevant feature space, or characterizing the typical data over the problem domain. Given this, requirements are a difficult thing to guarantee that they are feasible. Either they will be easy to meet, or they will be impossible. I would rather look at best effort type of design, where one tries to do as well as possible. Perhaps the output of such an exercise using limited resources could specify the requirements.

E No response.

F Intel and history should produce a threat set. ATR should be able to handle the most concerning threats that passes through ATR nodes with a reasonable area under the ROC. The level of concern is inversely proportional to the difficulty of obtaining the threat by known adversaries.

G I would recommend a meeting that is at a secret level to be able to discuss the threat scenarios openly and determine exactly what is needed. This can then be directed at a higher [non-sensitive] level to a broader audience.

H No response.

I No response.

J There should be incentive provided to manufacturers to perform better, both from a detection and operational performance point of view, or to introduce new capabilities that can lead to cost savings or better screening experience for travelers.

K - Identify the state of the ATR.
- How to improve and what needs to improve.
- What are the operational constraints?

L This was discussed, but it'd be great if certification wasn't binary. I know it isn't really attractive to complicate the purchasing of devices for airports, but thresholding these designs as "certified" and "not" seems to stifle innovation as there's no economic incentive to do any better than "certified".

M The current process in the US and throughout the world is very reactionary, someone (underwear bomber, shoe bomber, etc.) attempts to or does carry out a terrorist attack and then the regulating bodies worldwide respond by updating the requirements, etc. While this process does work, it can be slow and though it is difficult, it might make more sense to be proactive, using intelligence information coupled with the currently available technologies to drive requirements. This could result in a more nimble and flexible system which potentially yields better security.

N By government-industry partnership, understanding what can be done and at what cost.

O The feature set used for ATR should be based on the physical features of the materials of interest. It is hard to believe that tens to hundreds of features that some ATR algorithms use are truly indicative of the materials of interest. I think using a large number of features reduces the generality of the ATR algorithm and opens the door to over-training and losing insight as to the actual decision path and logic of the algorithm.

P Based on performance goals sought, spec and measure what you want.

Q The government should determine what constitutes acceptable risk such that products that meet the corresponding specs are in the clear if/when a threat makes it through the eye of the needle.

R See response to question 1.

S No response.

T Not sure here, but: the DHS S&T COE has specific mission objectives. One of these is TRL 1 - TRL 3 R&D. In general, the 'real application world' of the production systems does not need to be overly revealed to work the TRL 1 – TRL 3 R&D agenda. The ADSA workshops are non SSI. The ALERT COE can unilaterally develop a set of non-SSI scientific requirements for these non- intrusive imaging systems that have ATR algorithms present. Think as a systems engineer to consider what the needed advancements would functionally look like. Postulate obvious need statements or use cases. Then translate these 'requirements' to more engineering / scientific based detailed requirements. The TSA can have the TRL 4 - TRL 6 or TRL 7 folks vet this open (Non SSI) set of ATR requirements. For example: Distinguish C4 from cheese; or eliminate the need for the 3-1-1 bags (I find this really annoying); achieve checkpoint passenger transaction total time of 10 seconds or less; or achieve maximum passenger queue time of 30 seconds.

U No response.

V The two key metrics for ATR are Pd and Pfa, or their equivalents. Requirements should be, and are, set by government intelligence on known threats. These should also be vetted against what a particular technology and population of products can realistically achieve, or they can be put to the test and determined empirically, irrespective of what a given technology can realistically achieve, and then adjusted, with evaluations of alternatives. These alternatives would be aimed at addressing identified weaknesses that do not appear to be realistically soluble within a specific technology. Pfa requirements should have established baselines by the government, based upon realistic operational requirements and consideration of the Pd requirements, specifically, attendant tradeoffs that may be required to achieve maximum safety while maintaining maximum throughput. Beyond these baselines, airport testing of real operational false alarm rates should provide another target for false alarm rates and these could be captured in desired, but not required ("should" versus "shall"), parts of the specification(s).

W The way it is is fine. Government needs to identify the high risk (more likely to be used) threats and prioritize them.

X No response.

Y Requirements for detection and false alarm rates are required. Requirements for robustness, statistical relevance, region of responsibility should be included but the method of verification and/or certification of these requirements needs to be fully developed.

Z This is the job for physicists and chemists. Auto-detection should be based on the physical and chemical features of the explosive, not the shape or the look of the object. Maybe spectrum recognition is the way to go.

AA By regulators and users, with technology input from industry. e.g. draft spec released for comment. Publication of final spec after review and modification.

AB Tough one.... Don't know.

AC Look at realistic threats and example EDS outputs to find materials that span the feature space (high/low Z, high/low density, etc.). Characterize the EDS machine response to each threat.

AD Need list of threats, at least their composition and general shape.

AE I don't know.

AF Tiering the acceptance levels (particularly on Pfa) is advisable. Also making sure technical evaluation and procurement talk to each other is helpful. Furthermore, they should be established in a way that allows for safety, throughput, and privacy, while paying less attention to where the vendors are and what they are able to achieve today. Vendors will always spend as little money as they can to get the over the acceptance bars unless they have incentives otherwise. It seems clear that most vendors can achieve great things in Pd and Pfa and push the limits of their modalities. That said, what vendors are able to achieve in regards to Pd and Pfa is largely a function of what the incentives are, and what the support is. ATR algorithms are difficult and costly to develop though and TSA needs to understand the NRE development costs that go into them. Partially funding key vendors for ATR development should be a consideration. The biggest costs are data acquisition and NRE algorithm development, and the levels of Pd and Pfa that are achievable are largely a function of these two things (and the incentives of course.)

AG No response.

AH No response.

AI I'm not exactly sure, but it does seem there should be a way for TSA to telegraph their future needs to vendors - something like 'your system passed this year's test, but here are some other tough cases you might want to consider'.

AJ No response.

AK I believe detection performance should be ensured, but after a minimum level of performance, false alarm rate should be an economic factor not a regulation. I am also concerned that the specs and testing do not sufficiently cover all failure modes. I believe a failure mode and effects analysis of the system should be included in the specifications. Requirements that are specific mitigations for failure modes should be made more clear. For example, how are you assured that the correct algorithm is being used in a risk based system?

AL Base requirement for conventional threats. Class requirements for more exotic threats i.e Base + A + D.

AM Through collaboration between regulators and vendors with selected third-party experts who can share best practices in designing development, training, and test datasets. I believe that every one of these constituents has something important to say in how requirements relevant to certification should be set. Assuming critical test information could be protected from vendors as needed. Could a National Academy of Sciences review of the setting of requirements and certification testing be done? Including statisticians (recommend Karen Kafadar who has done this in the past) to address the issue of statistical significance and relevance of the test.

Question 3: How should testing of ATRs be modified?

A I don’t think it should. But there needs to be a method for testing ‘generic’ ATRs in a public venue, this would create a community for contractors to leverage off of. A quick way to start this would be for DHS to fund a graduate school level contest of recognizing objects in 3D scenes. The Microsoft Kinect would be a great tool for this.

B No response.

C No response.

D Don’t know how it is done now. I would suggest that, at a minimum, some feedback on difficult instances must be provided so that failures can be addressed. I would also try to randomize testing to avoid point designs.

E No response.

F Unknown.

G To ensure a level playing field, all the qualification/certification testing must be performed by an appropriate entity [either governmental or private] with sufficient resources & qualifications. This sounds like motherhood and apple pie, but without such an approach, discrepancies are inevitable. Also, following the determination in item 2 of the requirements, clear methods of dissemination of these requirements to the appropriate individuals is a necessity to ensure that the targets are known.

H No response.

I No response.

J It is crucial to provide feedback to manufacturers to help them refine their solutions. Without necessarily divulging test details, there needs to be more feedback than what is currently available to allow for algorithm refinements and the ability to perform proper regression testing. Also, there needs to be more emphasis on operation concerns (i.e. Fa, ConOps, total cost of ownership, etc.)

K - Provide blind tests.
- Use PD and PFA as means to detection.
- Use location of mean PD mean PFA in ROC as a means to comparison between groups.

L As a newcomer, I can't comment.

M It is important that the ATRs actually work from a security and an operational perspective. From a security perspective, things are tested pretty thoroughly at the TSL at the certification level. With the addition of the TSIF in the last 5+ years, testing has improved on the operational side. More live testing in the airport environment would be great, however, there is a ton of political risk and therefore it may not be practical. Barring this, somehow incorporating more "live" testing at the TSIF might make sense.

N It's adequate as it stands today.

O It may be useful to separate the hardware from the software (ATR) testing in some fashion. Simulated image data that is validated to a system platform may be useful for generating unique threat configurations that could test for over-training of ATR. So we may want to "test" the hardware via existing image quality analysis protocols, then probe the ATR with a limited set of real image data combined with simulated image data, then perform a (limited) test of the combined hardware/software system.

P The certification process for EDS (and its cousins) should be statistically valid.

Q Realistic test data should be made available prior to the actual test. Companies should not have to guess where the bar is or how close they were to making it across in case they fail.

R No response.

S No response.

T I think entirely too much time is spent on this topic. Do recall the French representative did note they do not share final test results either. This is not to say the OEMs seeking to build 'production' systems do not have the opportunity to get pre-test results from the TSA staffs; they do; and that information is SSI. It was also noted that each of the checked bag EDS OEMs has their own 'large' set of test or development bags. The theme here is that vendors -- who need to know, know what they need to know! It is also true that the testers know that the OEM will 'game' the test if they think they can prevail. This natural dynamic only reinforces the need to restrict who known what, and when. So the question here is testing of possible TRL 1 -

TRL 3 innovations. It is doubtful an ALERT sponsored TRL 1 - TRL 3 project is going to leap into a TRL 7 fielded system. Such innovations will need to progressively mature through the TRL stages. Actually, the projects of the ALERT ATR work should constitute advances that the commercial OEMs or DHS security leaders find attractive. Typically there is a substantial investment needed to support taking such advancement and incorporating this improvement and / or commercializing the TRL 1 - TRL 3 work into a fielded system. In summary, concentrate less upon the commercialized testing and more upon vetting an ATR advancement to warrant further investment from the OEM or DHS user community. For example: The Next Gen AIT overviewed by PNL. I take as given that their system constituted a 1000x improvement. However, the data processing time was reported to be 6 hours. For this example, commercialization needs to reduce the 360 minutes to what; say 1 minute?

U No response.

V Tests should be unknowable (not able to be gamed) by the test taker, allowing exposure of more information and data on the items that are required to be detected, to facilitate more rapid convergence to the desired capabilities and to reduce cost and time for the government and vendors. This may cost more up front. Testing should be more cooperative between governments to spread the cost and time, to achieve better, more comprehensive and consistent results, and to harmonize as much as possible.

W For safety reasons, the testing should include a few full threat samples for each HME category, but not necessarily scanned in all possible configurations and orientations. In other words, move away from attempting a statistical analysis of performance, but instead do verification by spot checking.

X No response.

Y Testing should be expanded to have adequate statistical relevance. Testing should accommodate continual improvement via ATR updates that sustain detection levels and improve false alarm performance.

Z Should shift emphasis from object shape recognition to spectrum recognition.

AA Apply statistical analysis techniques to data - pre-qualify data sets in test to be representative of variation in the real populations. Require ROC curve submission, rather than one operating point. Give the customer the option to choose the operating point.

AB Somehow streamline it so tests can be run quicker.

AC Tests must be controlled by a third party. I like the idea of having a neutral test director (Carl?) for all participants.

AD No opinion.

AE Larger testing datasets should be used.

AF The testing seems reasonable but TSA (more than ECAC) needs to ensure that the incentive structure is set up correctly. The stages of TSA testing where there is DT&E followed by black box IT&E is a fair and reasonable structure. ECAC has similarly fair policies.

AG No response.

AH No response.

AI Feedback to vendors seems tricky but important.

AJ No response.

AK Testing should be more system level. It should evaluate failure modes outside of the detection and also include conops.

AL Standard validation for base requirements. Additional validation for update class.

AM See answer to question #2.

Question 4: How should deterrence and risk-based screening be incorporated into the design of an ATR?

A Risk based screening could be an improvement. But could also create additional security risk. It gives people more input into the security process. If it were incorporated, I would assume that individuals deemed higher risk due to certain modalities would simply reduce thresholds in other modalities.

B No response.

C I do not know.

D If there are specific threats that you wish to minimize, one could bias the ATR to perform well along those directions.

E No response.

F Successes of ATR should be promoted so that the assumed PD is quite high. This will discourage probing the checkpoints. I am ambivalent about risk-based screening.

G In the U.S., it will be challenging to perform adequate risk-based screening (RBS) due to the concerns with profiling and privacy. However, outside the U.S., alternative lane scenarios are practical. If TSA is able to bring Pre-Check to achieve a sufficient level [say 50%], then the remainder of the population will pass through the "high risk" lane even if they are low risk passengers. The key to all of this is to have a system with a high enough Pd and low enough Pfa to move to alarm only mode [similar to checked bags]. Also, enabling different levels of detection [LAGs, sheet, bulk, etc] for different passengers will enable randomization or switching to higher capabilities based on risk assessment.

H No response.

I No response.

J A risk-based approach that introduces unpredictability for the terrorist and the ability to change detection thresholds based on risk assessment can definitely lead to better overall security. However, to allow this to happen, one must be able to properly assess the impact of changing detection parameters, which requires a different testing approach that allows regulators to push the limits of the detection algorithms and interpret the results beyond a simple threat/no threat indicator.

K Risk-based screening controls adaptive thresholding for increased or decreased detection (PD) and false alarms (PFA) versus threat level - As a result, a user should define an overall threshold that could be adapted as a function of the risk based screening and threat level.

L I don't think the human should ever be taken out of the loop. There is a significant deterrence factor to TSA employees chatting up people in line. They also offer an uncertainty which can't be planned for by potential terrorists. The underwear bomber didn't choose to be the "underwear" bomber for style points, he was reacting to the process at the time and putting a bomb in his underwear was optimal for his goals. Human interactions, however, are more random and less likely to be successfully planned around by terrorists. Additionally, if there is any prior information available about passengers it should be fed into the ATR algorithms. Collect the data and make it available to vendors.

M Some side effects due to physical deterrence typically include operational and financial cost. Placing systems in the checkpoint tends to slow the process and cost money, however, having them there can serve as a deterrent. Having a simple and convenient check-in and/or boarding process might be nice for the passengers, however, it may serve as much of a deterrence. Risk-based screening could easily be incorporated into the ATRs, in fact, many places in the world are altering the security level based upon passenger, destination, or both. The systems within the security environment just need to integrate easily and share the appropriate information.

N As a simple and understandable control.

O Risk-based screening could be conducted via the following protocol: Screen everyone/everything using an ATR-enabled device that provides a probability of an object being a threat. That probability could then be weighted based on a risk-based factor that either increases or decreases the probability of threat status. The initial screening method should have a low false alarm rate. People/Bags that exceed some weighted threat probability are then directed to some secondary screening method that has a higher detection rate than the initial screening method (maybe with a high false alarm as well).

P Rational to adjust ATR based on information exogenous to the scanner (demand higher PD and accept higher PFA for those deemed at higher risk, based on exogenous information).

Q ATR is only a deterrence if the public hears about threats that were caught. Risk-based screening might be achieved for a preset number of risks that the system has been trained to handle.

R No response.

S No response.

T Fundamentally: Have the imaging devices - which feed one or more algorithms need to be built to standardized data control interfaces. DHS S&T has promoted DICOS - a DICOM extension, for the last 5 years. Accordingly, a solution incorporating DICOS has its own merit. That said, this most recent ADSA workshop introduced the comment that there are many algorithms comprising the current SOA. This leads to the notion of an algorithm for controlling / selecting needed algorithms. Better ATR systems will results through incremental improvements to this overall system of systems, Hence, we need to be able to 'parcel' the incremental opportunities for ATR advancement into logical chunks; e.g. AIT systems that are better able to process body folds.

U No response.

V Considering the security and political realities of the regulatory and governmental body(ies) involved.

W Outside the scope of this group. In general an ATR should always be designed, so that the developer cannot have more than 50% certainty that they can defeat it. This has been done by adding some randomness to the decision of detecting difficult corner cases.

X No response.

Y The most obvious implementation would be a "knob" that can be adjusted in advance based on pre-defined risk factors for the passenger or destination, etc. Other factors could be incorporated into risk-based screening such as weight and clutter of a bag for checked or carry-on bags.

Z They are all important. Psychology-based detection and technology-based detection do not have to be combined to “vote”. Use the “winner takes it all” method. Either one of the methods detects something, it should be a red light.

AA Analog input or handle to modify ROC curve operating point depending on risk assessment. Requires ROC curve to be approved by regulator - might only be a smaller range of acceptable points on the ROC.

AB Goes back to regulators, but why not make the ATR’s components of a risk engine, i.e. let them speak the language of probability.

AC Not at all. Let’s see how ATR works on its own, then study deterrence and RBS separately.

AD Based on other factors such as the type of person and place related to the threat modify and refocus the threat detection algorithm.

AE In the medical world, when screening for cancer, radiologists lower their threshold when the patient has risk factors. The same could be done here. The trick is how to get the information to the ATR system.

AF Deterrence is imperative, and should be done on all levels (not just ATR). Deterrence with regard to ATR is more a function of how the ATR is perceived and advertised than how it actually works. For risk-based screening it is much less clear and this probably throws more noise and uncertainty in the process for all involved than it helps. Those trained in applying the risk-based screening must be trained extremely carefully for it to actually work. My impression of this is that it’s not convincing and I would say that this can open up more security holes than anything.

AG No response.

AH No response.

AI This question is way over my pay grade.

AJ Regular inserts to keep the experienced detection rate at 1/300. Multiple levels of inserts.

AK I think an FMEA of the system, not just the ATR, needs to be provided so specific mitigation requirements are understood.

AL Overly optimistic press releases. Fusion of human assessment and machine outputs.

AM Recommend looking at the Morpho DSFP model for incorporating risk-based screening in terms of measures of probability. Needs to be extended to allow for correlated information and possibly options for more flexible mathematical framework such as Dempster-Shafer. Deterrence--recommend looking at where deterrence modeling has been done in the past and who has done it. RAND?

Question 5: How can third parties be involved in the development of improved ATRs?

A DHS should require multiple contractors to develop ATRs for each AIT system. With the current method of ATR development there is no motivation for L3 or Rapiscan to sub out ATR work. Just because L3 is 'passing' the ATR tests, does not mean there are not holes that could easily be filled with some healthy competition. I have seen third party results in the ATR space for AIT, they are impressive and use creative and new techniques that I think would improve results.

B No response.

C Through mutual collaborations.

D Work with vendors, after establishing capabilities on "entrance exam" test suites that should be available to interested third parties.

E No response.

F Toy problems can be widely released along with benchmarks.

G Repeat: third parties should work directly with industrial manufacturers with an NDA and rate of first refusal by the manufacturer to the new algorithms etc. Otherwise, it is unlikely to achieve integration with 3rd parties with industrial groups.

H No response.

I No response.

J If equipment manufacturers are given incentives to do better, they will tend to use more third party help. On the other hand, access to SSI information is a serious limitation at the moment. Programs like the "Great Challenge" are also good ways for third party to get involved and be given the opportunity to be known. But to take full advantage of this, there needs to be a way to turn these R&D results into actual products.

K Need a broader call for proposal and evaluation to include more competing and qualified groups.

L No response.

M Third parties can be involved in several ways: 1. through direct partnerships with new or existing vendors, 2. through programs sponsored by the government or private entities, or 3. through their investment. The barrier for entry within the US tends to be pretty tough - understanding the entire process, getting the right information in a timely manner, being invited to the party so-to-speak - therefore the best options are 1 and 2 above. If vendors drag their feet, the best option is 2, where a program would be put in place and the third party developers receive the support (access, data, etc.) they need.

N By working on focused problems rather than large generalities.

O Grand Challenges are a good start.

P - Set goals.
- Set incentives.
- Allow vendors to seek third parties to help them achieve their goals.
- Reduce "transaction costs" by increasing interaction and networking opportunities for vendors and third parties.

Q With great difficulty. An ATR algorithm requires data and knowledge about the data which in turn may require knowledge of vendor specific system behavior.

R No response.

S No response.

T Look to and understand the mechanisms that drive the world-wide medical imaging industry. The worldwide DICOM / PACS business is characterized by one very large commercial medical imaging company as $8.6 Billion per year. Algorithms -- and the medical industry's version of ATR, are integral to this. The DICOM / PACS market size continues to expand; the large concerns get larger, established third party software concerns get larger, and new small business third parties are added each year; and numerous university SMEs do very well. One last significant point: With technology development, we need to follow the money. It is hard to project where security (e.g. DICOS) or parts inspection (e.g. DICONDE) will ever fund remotely close to the levels past, present, and future -- found within medical imaging. So to be clear, security third parties should want to establish a paradigm built off the established medical paradigm. (Do note: This view may be at odds with established security system OEMs.)

U No response.

V See answer to question #1.

W No response.

X No response.

Y Can evaluate new decision algorithms to improve discrimination of feature data provided by OEMs. Can develop algorithms and statistical approaches to prediction of performance over a Region of Responsibility based on sampled data. Becomes more difficult to involve third parties in upstream algorithms for feature extraction, segmentation, etc. because much more disclosure is required in more sensitive (security) areas.

Z Sub-contracting.

AA TBD.

AB Partnership with industry like the one presented by AS&E and Tufts.

AC Through a (funded) competition.

AD Academics familiar with imaging systems reconstruction and image analysis can greatly add to the expertise and partnership between the processes, as mentioned earlier.

AE I think when the TSA incentivizes companies to improve their ATR, the companies will approach third parties for innovation. Third parties need access to images.

AF Talk to vendors. Most vendors want help improving their ATRs if the issues of trust, clearance, export control, etc. are worked out. There are technical problems to solve. One huge way that third parties can help is to help in the acquisition and evaluation of data in a way that's consistent with a vendor's interests.

AG No response.

AH No response.

AI Data sharing is important; we are a 3rd party who has done work with a vendor, and I didn't appreciate how easy we had it because their images are not SSI.

AJ Open up the data stream via sterilized datasets.

AK By finding a way to provide them with real threat data, not contrived problems.

AL This is hard given competitive / proprietary nature of business. 3rd parties to be driven by requirements specs.

AM Targeted problem areas within the research and development programs of vendors as a start. Rapiscan gave a nice example of how they use multiple external sources for developing classifiers. Might be a good model.

Question 6a: What did you like about the workshop?

A Open atmosphere of dialog.

B No response.

C 1) the choice of talks and discussions 2) the mix of attendees 3) new interesting information.

D Very good discussions among speakers, audience members.

E No response.

F Diverse presentations.

G No response.

H No response.

I What I liked was: 1) The opportunity to learn about the latest developments; 2) The length and pace of the program were just right; 3) The size of the group was good for a gathering of this type; and 4) The menu of topics included was well conceived.

J Good exchanges on what can be done to improve the system rather than simply presenting ATR results. Exposure to what is being done outside of the US or outside of the Air Transportation world.

K - The learning process through the presentations.
- The open discussions.

L It seems data is a big problem in the community and while there are certainly roadblocks to people getting enough of it right now, these won't be going anywhere unless there is a mutual understanding. It was great to see people engage each other on this topic.

M ADSA08 was my first ADSA workshop and I found it to be very worthwhile from the following perspective: helps gain insight into the industry as a whole (not just the technology but the other challenges faced by those within the industry, vendors, regulators, etc.). The problems appear to be very similar from vendor to vendor.

N Breadth of discussion.

O The discussions were very frank and generally productive.

P	Good mix of vendors, third parties.
Q	The open discussions on Day 1.
R	No response.
S	No response.
T	Best technical content yet. Each workshop has progressively learned / improved from the preceding one. This learning has produced gains and challenges. It is genuinely exciting to observe the topic growing.
U	No response.
V	Considering the limitations imposed upon an open discussion of these topics, there was a fairly free exchange of thoughts and ideas.
W	No response.
X	No response.
Y	Diverse participants (academics, industry, medical, security, ...) Good setting for un-interrupted focus on new/different ideas. Excellent logistics.
Z	Short and focused on one topic.
AA	Open discussion. Examples of existing collaboration between vendors and third parties.
AB	Good presentation. Good mix of topics. Good networking opportunity.
AC	Wide range of topics and interests in the ATR issue, including vendors, labs and academia.
AD	Good discourse.
AE	Interaction with different parties. I liked the dinner speakers very much.
AF	Many of the discussions were extremely beneficial. Location was great, food and venue were great. It was run very well except for one technical issue.

AG No response.

AH No response.

AI Some very good talks and lots of opportunity for interaction.

AJ Flexible schedule, high dialog, tough questions asked.

AK Good to see everyone is facing similar problems. Industry speakers were knowledgeable. Video analysis work was very interesting.

AL Networking and Q&A were more enlightening than presentations.

AM I actually like that most talks are not highly technical. I learn a lot more that way about the problem space. Like the interactive format.

Question 6b: What would you like to see changed for future workshops?

A Each workshop should create a challenge for a group of university students. This challenge should be based on the expertise of who are attending the meeting. DHS should fund these students to solve this problem for a year. The students would then report on their results at the next workshop. This would help the meeting focus on actual algorithms, instead of it being simply a method for contractors to complain to DHS.

B No response.

C The moderation of discussions.

D No response.

E No response.

F No response.

G I suggest reducing the number of presentations as it always seems that things become time-crunched.

H No response.

I No changes. Good as is.

J More involvement from TSA to help set long term goals and objectives.

K Less presentations for more discussions or a 3-day workshop for more time.

L I'd love to see breakout sessions with specific guiding questions and a moderator. At times when discussion veered to policy or advanced math it seemed some of the audience couldn't engage.

M Some of the topics were either too compressed (speaker did not have enough time), not well organized (this was more of a case by case basis), or repetitive (same topic covered by different speaker). While I know some speakers will use all the time you give them regardless of how long is provided, it may make sense to reduce the number of speakers and allow for a bit more 'flex' time in the agenda. This time would be scheduled in to compensate to balance the load between speakers who require more and speakers who require less time.

N Wider variety of student papers. How about getting students from other universities? There are a LOT of them out there in many fields.

O Time for each speaker should be longer to allow for enough discussion. Extend the workshop to three days with the same number of speakers.

P More input or information from DHS and TSA would be welcome.

Q More focus on technical aspects. Less focus on policy.

R No response.

S No response.

T The following is meant as constructive criticism.

a. Add another day.

b. For whatever reason, the TSA, TSL, and TSIF participation -- and their SETA support, was down. (Probably the government funding issue) That said, it is not adequate to have just staff from those organizations. A representative -- who is confident to speak within the limits they must abide by, from various DHS organizations is highly desirable.

c. Stop interrupting the presenters. (Title this better time management.) Each presenter works hard on each chart; and are generally not allowed to finish their entire presentation. Having a format for displaying the conclusion chart first (which is liked) is not license to not allow the presenter to go through their prepared remarks. So, technical, on point, interruptions are a part of the format. However, workshop leads should passionately throttle unnecessary kibitzing (read on). It is recognized that presentation skills vary greatly.

d. ALERT PMs should refrain from pandering to the DHS S&T leadership in front of the attendees. There is a time to sell the next or follow-on program, or show worth; but not a continuous dialog at the expense of the presenter's and listener's precious workshop time. ADSA Workshop and ALERT scope creep. Consider this notional division; TRL 1 - TRL3 Work Agenda and ALERT (plus others); TRL 3 - TRL 5 Work Agenda for the TSA Transportation Security Lab (TSL) and TRL 5 - TRL 7 Work Agenda for the TSA Systems Integration Facility

(TSIF). The point here is the ADSA workshop dialogue too-often moves all around these three notional developmental work agenda phases. The work content within each of these areas is significant. Hence, to be a genuine contributor within one area is a great challenge. It is impractical to educate a university researcher on details pertinent to the TSL or TSIF work agendas when what they need to know is the ALERT work agenda set of requirements (and opportunities). With few exceptions, this same observation can be said of the DOE lab SMEs. Clearly, the established OEMs and industry SMEs (as illustrated by a Carl Crawford or Morpho) are knowledgeable regarding the entire developmental spectrum. This knowledge extends to proprietary work their company is performing on within any of the above three areas; work that most often is not to be shared without securing proper controls. Each OEM -- at their discretion, has the opportunity to educate a TRL 1 - TRL 3 researcher or research team regarding what else is needed. Consider a format / tradition followed within the DICOM Standards Committee. A presenter is not permitted to sell; and commercialization -- such as it is, is to be limited to the cover chart. I suspect this style is more broadly followed than just the DICOM technical meetings.

e. The theme of this observation is that if the ADSA Workshop is a technical forum, then non-technical issues need to be minimized.

f. Political agendas. This is a variant of the preceding point. There are always under-currents present; and such communication cannot be prevented. It can be controlled; and it is the role of meeting leadership to manage this. Hence, whenever discussions start to stray, workshop leaders need to redirect from this. Meeting times for such communication can be set for evenings; or otherwise use break-times, etc.

g. Don’t criticize the TSL or TSIF process when the SME leaders covering those domains are not present.

h. The discussion regarding the commercial incentives is interesting. However, I do not see how the ADSA workshops constitute the appropriate forum. For example, NEMA would seem to be a better setting for that conversation; and that conversation needs to include TSA acquisition professionals.

U No response.

V See answer to question #1.

W No response.

X No response.

Y More input from adjacent industries (medical, NDE, ...). There are probably good discrimination algorithms in the NDE (CT or Ultrasound, etc.) industry.

Z Smaller size, focused working groups.

AA No response.

AB Would have been good to move the gov't talks (like the STAC one) up earlier in the conference.

AC Things seemed somewhat rushed... perhaps a third day is needed, with a few breaks for discussion.

AD Describe how partnerships can be created and proposals submitted.

AE No response.

AF The time allocations for talks and discussions was an issue. In the beginning there was considerably more discussion and it went well over schedule. As a result many of the later talks were curbed, or questions were limited. For the most part, the moderator did a great job, but one improvement might be to divert certain discussions to certain times, more with the goal of getting consensus on topics. It seems that the conference generated many questions (which is good) but didn't reach consensus on much.

AG No response.

AH No response.

AI Sometimes the discussions can get a bit philosophical (this was especially true during the fusion ADSA, for example). There is certainly value to that and I think it's an important part of the workshop, but it could be scaled back a bit.

AJ Keep it this way.

AK There were several topics that many people in the room knew much more about, but could not openly discuss for security or economic reasons.

AL More technical detail. This probably means fewer 3rd party vendors.

AM Great talks but the number of them needs to be smaller. Carl already noted this at the opening so no problem.

Question 7: Do you have recommendations for future workshop formats?

A	Focused working groups.
B	No response.
C	Larger size, focused working groups.
D	Fewer speakers.
E	Focused working groups.
F	No response.
G	Smaller size, fewer speakers. To enable more open discussions, I suggest limiting each alternate session to cleared personnel.
H	More speakers.
I	No response.
J	Right size / length.
K	More breaks, fewer speakers, focused working groups.
L	Focused working groups.
M	Fewer speakers. Focused working groups might make sense, however, they would need to be goal driven and really accomplish something...
N	Larger size, more breaks, focused working groups.
O	Larger size, fewer breaks.
P	No response.
Q	Focused working groups.
R	No response.
S	Larger size, more breaks, fewer speakers.
T	Larger size, more breaks, more speakers, focused working groups. Recommendations are:

- a. Work to get the workshops progressively larger; more days.
- b. Breaks provide the opportunity for program management and commercialization discussions.
- c. More and newer speakers. Interestingly, speakers not familiar with the technology details need some background orientation prior to creating their briefings.
- d. Working Groups should be a long term goal; not considered within the next 1 - 3 years.
- e. Venue Changes. Washington D.C., Chicago, California (LLNL?)

U Smaller size, more breaks, fewer speakers, focused working groups.

V Smaller size, focused working groups, classified section(s)

W No response.

X No response.

Y Fewer speakers, focused working groups.

Z Smaller size, focused working groups.

AA More breaks. Missed some follow-up conversations with speakers. Especially towards the end of the symposium. More breaks would have helped here.

AB No response.

AC No response.

AD Focused working groups. Add one day for above.

AE I thought the size was about right. The breaks were about right. Focused groups can be useful, but I think you lose on interactions across different areas of expertise.

AF Fewer speakers, focused working groups.

AG No response.

AH More breaks. Focused working groups.

AI Focused working groups.

AJ More breaks, focused working groups. Continue to react to the real-time group wishes.

AK More breaks, fewer speakers.

AL Focused working groups. Most presentations were very speculative. I would have liked more discussion of implementation rather than theory.

AM Fewer speakers.

Question 8: What other comments do you have?

A No response.

B No response.

C Thank you for great workshops!

D No response.

E No response.

F No response.

G No response.

H No response.

I No response.

J No response.

K I would like to thank the team for the hard work and time they spent preparing for the workshop.

L A great first insight into this community, thanks for the invite.

M While things can be improved, thanks for organizing and keep it up...

N No response.

O No response.

P Thanks for organizing and all the hard work to pull it off!

Q Send a “mark your calendar” email well in advance of the invitation with all the meeting details.

R No response.

S No response.

T One, consider a workshop dedicated to understanding the SOA of medical imaging. The interesting point is that this technology segment continues to experience double-digit growth. Hence, leaders within this market segment are not interested in ‘moving’ or competing within our security enterprise. However, by engaging the correct speakers and crafting a focused technical agenda, a sense of what can be (in that parallel universe) can be gained. Two, repeat the item 1 workshop for the DOD Imaging community.

U No response.

V No response.

W No response.

X No response.

Y No response.

Z Have a clear goal for each workshop. Avoid topics that do not get anywhere, but wasting time.

AA Very enjoyable and useful. Please keep these symposia going!

AB Happy to be invited.

AC No response.

AD No response.

AE Nice workshop. It was well organized and executed.

AF Nice job to the moderator and support staff. Was a quite well-run conference!

AG No response.

AH No response.

AI Nice and helpful workshop. I enjoy these.

AJ No response.

AK No response.

AL Very well run conference, I learned a lot. With more meat three days would be great. Smaller group Q&A with presenters would be nice, maybe three 40 minute sessions.

AM I think this is one of the most effective and successful workshops in which I've participated. Still needs some statisticians.

14. Appendix: Acronyms

TERM	DEFINITION
2D	Two-dimensional
3D	Three-dimensional
ADSA	Algorithm Development for Security Applications (name of workshops at ALERT)
ADSA01	First ADSA workshop held in April 2009 on the check-point application
ADSA02	Second ADSA workshop held in October 2009 on the grand challenge for CT segmentation
ADSA03	Third ADSA workshop held in April 2010 on AIT
ADSA04	Fourth ADSA workshop held in October 2010 on advanced reconstruction algorithms for CT-based scanners.
ADSA05	Fifth ADSA workshop held in May 2011 on fusing orthogonal technologies
ADSA06	Sixth ADSA workshop held in November 2011 on the development of fused explosive detection equipment with specific application to advanced imaging technology
ADSA07	Seventh ADSA workshop held in May 2012 on reconstruction algorithms for CT-based explosive detection equipment
ADSA08	Eighth ADSA workshop to be held in October 2012 on automated target recognition (ATR) algorithms
AIT	Advanced imaging technology. Technology for find objects of interest on passengers. WBI is a deprecated synonym.
ALERT	Awareness and Localization of Explosives-Related Threats, a Department of Homeland Security Center of Excellence at NEU
AT	Advanced technology. Second generation of TRX.
AT2	Second generation of AT.
ATD	Automated threat detection
ATR	Automated threat resolution; a synonym of ATD.
BAA	Broad agency announcement
BDO	Behavioral detection officer. A type of TSO.
BHS	Baggage handling system
BIR	Baggage inspection room
BLS	Bottle Liquids Scanners

TERM	DEFINITION
CAD	Computer aided detection or diagnosis. A term from radiology.
CAT	Credential Authentication Technology
CCL	Connected components labeling
CERT	Certification testing at the TSL
CI	Confidence interval
CNR	Contrast to noise ration
COE	Center of excellence, a DHS designation
CONOP	Concept of operations
COP	Concept of Operation
CRT	Certification readiness testing
CT	Computed tomography
DAS	Data acquisition system
DFT	Direct Fourier Technique
DHS	Department of Homeland Security
DHS S&T	DHS Science & Technology division
DICOM	Digital Imaging and Communications in Medicine; http://medical.nema.org
DICOS	Digital Imaging and Communications in Security. NEMA standard for image format for security; NEMA IIC Industrial Imaging and Communications Technical Committee.
EDS	Explosive detection scanner that passes TSL's CERT.
ETD	Explosive trace detection
EXD	Explosive detection directorate of DHS
FA	False alarm
FAT	Factory acceptance testing
FBI	Federal Bureau of Intelligence
FBP	Filtered back-projection
FDA	Food and Drug Administration
FN	False negative
FP	False positive
GC	Grand challenge
Gordon-CenSSIS	The Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems, a National Science Foundation Engineering Research Center at NEU

TERM	DEFINITION
GT	Ground truth
HME	Homemade explosive
IED	Improvised explosive device
IMS	Ion mobility spectrometry
IP	Intellectual property
IQ	Image quality
IR	Infrared or iterative reconstruction
IRT	Iterative reconstruction
LAC	Linear Attenuation Coefficient
LLNL	Lawrence Livermore National Laboratory
MBIR	Model based iterative reconstruction
MMW	Millimeter wave
MTF	Modulation transfer function
NDA	Non-disclosure agreement
NDE	Non-destructive evaluation
NEMA	National Electrical Manufacturers Association
NEU	Northeastern University
NIST	National Institute of Standards and Technology
NQR	Nuclear Quadrupole Resonance
OCT	Optical Coherence Tomography
OOI	Object of interest
OSARP	On screen alarm resolution protocol/process
OSR	On screen resolution
PD	Probability of detection
PFA	Probability of false alarm
PPV	Positive predictive value
QR	Quadruple resonance
Recon	Reconstruction algorithm
RFI	Request for information
RFP	Request for proposal
ROC	Receiver operator characteristic
ROI	Return on investment or region of interest
SAT	Site acceptance testing

TERM	DEFINITION
SNM	Special nuclear materials
SNR	Signal to noise ratio
SOC	Stream of commerce
SOP	Standard operating procedure
SSI	Sensitive security information
SSP	Slice sensitivity profile
Sv	Sievert. 1 unit of x-ray exposure/dose.
TBD	To be determined
TCO	Total cost of ownership
TIP	Threat image projection
Trace	Synonym of ETD
TRX	TIP-ready X-ray line scanners
TSA	Transportation Security Administration
TSL	Transportation Security Lab, Atlantic City, NJ
TSO	Transportation security officer; scanner operator
WBI	Whole body imaging; a deprecated term for AIT
XBS	X-ray back scatter
XRD	X-ray diffraction
Zeff	Effective atomic number

15. Appendix: Minutes[3]

The ADSA08 minutes were edited for purposes of clarity. All errors in the minutes are due to the editors of this report and not due to the speakers themselves.

15.1 Day 1 Minutes: October 24, 2012

Speaker: Carl Crawford

Carl Crawford: For other modalities it might be different

??: Is the only information you are able to use from the (???) could you have information coming from the outside?

CC: The purpose of these workshops is to bring people together – so if you have suggestions please throw it out.

??: You have to make a decision as to what we are going to do.

CC: This is totally different from what we are hearing. Part of the workshop is to see what these blocks mean, how do we get better blocks, and it seems like we are getting a lot of questions about what this means, and we are going to look into ATR more as we go through these slides. This is acronym soup – there is a list of acronyms in your folder.

Matthew Merzbacher: When you get too specific on your definitions you are just restricting yourself. There's no reason that you can't have a human in your model. It should be able to fold in.

Luc Perron: I agree with Matthew – it's not 1 or 0, it's not always clear. It could be an ATR or it could be an operator. It's still ATR. We should look at this area, because it's still suspicious.

MM: You have to look at characters that are interesting.

Chris Gregory: Are you suggesting that we are limiting this architecture?

MM: We are not limiting all of these in different fields.

Michael B. Silevitch: It will depend on what you are screening. Clearly the screener will have to play an important role in that scenario.

Bill Hall: What the threat is today may not be a threat tomorrow. We have automated tests, and we know what we need.

CC: I assumed that it was all automated, and that was (???)

MM: Video Analytics: It cannot be an issue of putting the machine wherever

---

[3] Inaudible or missing portions of the minutes will be indicated in parentheses as (???).

you want. What if something goes wrong (???) There were some scientists in Italy who were jailed because they didn't know there was going to be an earthquake. Given today's environment, we have to look at what the market is. There are no regulations there.

??: I guess this is a question of where people fit in, and what recognition is, and what algorithms fit in.

John Beaty: A body of ATR that is used in the military for years, and it would be silly for us not to recognize that. If we end up – we have to figure out how to apply it.

Olaf Johnson: There is a major division of what's needed. What be the inputs or outputs. That ambiguity is paralyzing. For industry it's paralyzing to have what's missing.

??: The research has to go on, and we have to be able to do these things the best we can, to change the specifications the best we can. We need to know what these specifications are.

CC: The requirements are what we have to get. Some pictures to show you what we are talking about. Here's the generic system that we had before.

Steve Azevedo: Is throughput an issue?

CC: (Slides are presented)

??: There is nothing in there that says anything about diminishing data and the data that you get to do detection has to be a lot more sophisticated, and that is because it is a government interest.

CC: What does it mean to diminish?

??: Current environment in different countries – that access is greater or lesser in different countries. The more information you have the better, but it takes time and effort and money. There is a disincentive to get there, and it's an interesting research question.

CC: Great question.

Tip Patridge: Are you talking explosives materials?

CC: Every ADSA we talk about a new topic (Slides are presented). How do we deter people using ATR? Is it good to have people's body's searched? There is a deterrence value of strip searching.

??: You have to go without this as an option.

TSA participant: In that diagram, it almost looks like, depending on what the risk based info is, that it's doing something different with the data we have. What is the risk status or level?

CC: We have two other workshops dealing with fused systems.

MBS: Again, remember this whole question is scenario dependent, and if they send something in a package to a FedEx plane, and they blow it up, there is no deterrence. It could be relevant, but it's not one size fits all,

CC: You can download anything on the web today, and what isn't available, I call domain expertise. The reality of how fast someone who knew about my gas furnace could fix it quickly is amazing. I cannot find out how to do that as fast as them.

??: If you go into that example, you can always find the prints to the furnace, and I think that's a problem. They don't have the option to walk up to the machine.

CC: When I got into the field that is what happened; so I think you are right. Richard; I was just agreeing that if you do go on the internet, you can find out what you want, and you can find out how to use explosives. You could get all of this information. What is a corner case: when you push your limit to the edge, and then push it to the edge again? Typically, at least half of the effort of the algorithm is finding those corner cases. Finding an open bomb in the bag is easy, but other ones are more difficult.

??: I understand your analogy, but the guy who fixed your furnace built it, or designed it. What are the really hard problems?

??: I would ask what is more important, but I think everything is important.

??: A couple slides ago you talked about surrogates. But in a real straw man system, where you are working on your little piece of it?

??: Are we going to talk about the scarcity of data? I think that in particular is important; as if you get feedback, it's very ambiguous.

CC: I will sit down in a couple minutes and then other speakers will cover those topics.

??: That could come up later?

CC: We will have a discussion tomorrow about the role of the govt.

JB: From my position, I have other researchers working in areas of segmentation. I try to get the representatives. I want to think about reconstruction. If you want 3rd party people working in any of these areas, they will need a very important step. Every contract that we run, it's the hardest part of what we do. It doesn't seem too reliable.

CC: Someone has to bring them up to speed.

CC: Crowd sourcing is showing up a lot in periodicals as a buzz word. This is why DHS wanted to involve third parties.

??: It's funny how the ADSA works where we're now calling academics "amateurs."

CC: Final reports able to be found online. As a result, NEU has to have a review meeting tonight at 7:00. This is a reconstruction initiative for grand challenges. We have an interesting problem now at ADSA. We have too many speakers and very short slots for our speakers. Sorry to those I couldn't include.

Rule #1 – discussion! Rule #2 – Public domain. #3 - Speaker instructions. Let's follow a specific purpose. The purpose of ADSA is not to talk about math but to have great discussions. These are our objectives. If we don't do a good job framing the problem, by the end we'll be totally lost.

Speaker: Alex Hudson

AH: They're doing a good job looking at images. They're able to find a lot of the threat-type objects we're looking for. Privacy concerns are also a great problem. We have natural problems that we're trying to work against. It's a great concept. The challenge is to have a machine look at humans rather than other humans. We have to get the machine to look at all shapes and sizes of people.

CC: Why are we doing this? What's the point?

AH: TSA decided there was enough risk, so now we are doing it based on what TSA is saying they need.

Conclusions first. ATR is a challenging problem with backscatter images due to low object contrast, pose variations, false alarm mechanisms and data collection limitations. Sometimes certain types of anatomy on the body can look like objects we're looking for, so we need to be aware of this. We have created a framework that permits the problem to be broken up and contributed to by many individual algorithms. Presently, three teams have created 30+ individual algorithms within this framework and we would be open for more collaboration partners. There are different threats we're looking for, so there are different algorithms depending on the different threat. We can separate the problem into many different components and this is great for collaboration. In order to get false alarms as low as we could, we had to develop a lot of specific algorithms. Rapiscan has really mastered this to be able to use different information optimally for the false-alarm [reduction].

MBS: Has Rapiscan developed a technology that can move from back-scatter to other modalities of imagery?

AH: The machine being built that exists (???) the rest is a question of what people are willing to pay for. An idea of how the machine works is as follows:

Raster X-ray pencil beam, vertical sweep and cam, wide area detectors, back scatter and forward scatter image. We want to fill up as much of the space as we can with detectors. The optimal machine is a full-coverage of the body. As you've seen in the airports, the system looks identical in front and back.

Some images: pulled from Rapiscan marketing materials - contrast and appearance of objects in backscatter imaging. Background is black means no backscatter. Subject is white means significant backscatter off body. Proximity, tissue/fat/bone all provides contrast. Objects of interest appear black, grey, or white. You can see belt buckles, keys and buttons on jeans. You can also see clothing segments. You can also greatly see bone, as you can see from the kneecaps that show great contrast.

??: So some types of anatomy are a problem? So another alternative form of screening is a pat-down. Is this also a problem at pat-downs?

AH: I'm not an expert at pat-downs so I can't say. So potentially this will also make the operator of the pat-down needs to do a thorough job.

So the steps for the machine. People stand in a specific way, the images are taken, you apply the image to algorithms and you consolidate and decide what you trust. Then you show the operator and tell them if the person needs a pat-down. There is a set avatar that we have to use according to TSA guidelines, so you have to judge what body part it is congruent to. Security requirements say that you have to pose a specific way and the algorithm is developed to be specific to this pose. Variations in pose need to be handled by the algorithm due to the degrees of freedom in part of the body. There are some segmentation challenges and a lot of room for improvement. There is also a possible performance improvement. It's what I call an upside down algorithm. We need to drive the false alarm rate down and increase performance.

CC: Why did it correlate?

AH: Many reasons including the amount of compromise that each data set has. It's difficult for the regulator and for us to get data sets. I would start applying statistics to the data sample. If they don't correlate then your false-alarm rate will increase.

Carey Rappaport: What about if people hiccup or sneeze or have Parkinson's?

AH: (???)

CR: Huge fusion potential.

Steve Smith: Reduced manpower and a faster decision. Counter example: at

Manchester in UK – had to go to a mm-wave system with ATR, and now they say that false-alarm is now very high. Is there anything in evidence that ATR can reduce manpower?

AH: I'm a firm believer that manpower will be able to do better still, but that's not what the customer is asking for. At the moment a human is perfectly adapted to looking at human form. You're competing head-on at what the human does best. It's difficult to improve on that. That's the difficulty.

Dominic Heuscher: Dose?

AH: Very small.

Robert Nishikawa: How big is dataset?

AH: 500 individuals.

RN: Testing at TSA?

AH: Less than that.

RN: In medical, if we do 0.4 false positives it's unacceptable. Here it's more than that.

AH: (???)

??: Size of object makes it depend?

AH: Not necessarily. I'm not getting all of the object images.

??: It does make a lot of difference for false alarms. Backscatter deployed without ATR is because both vendors said they could do it within 6 months. TSA has no way to enforce promises. Neither of the vendors would give TSA the authority to give 3rd parties the data. If I had my images backscatter, it would be a problem. There was a lot of data, but it couldn't be released, and it stifled the development of the ATR. Once systems were installed, no way to backtrack. It was a political problem.

AH: There was a huge outcry because of this.

??: Well that was after the data collection. There is a report that shows the dollar burden per year for the image analysis.

AH: Quite possible. They don't separate male versus female.

Justin Fernandes: Mismatching performance. Would you say it's the anomalies you're testing on or actual body-type?

AH: Both. You have to take both into consideration. You start with some assumptions. Some are right, some are wrong. You have to fail to be able to convert.

Matthew Merzbacher: Given progress, how far do you think you are from close enough to just small incremental improvements, when is the slope going to slow down?

AH: Well there is also an evolution where the specification is improving so the bar is getting set higher.

Dave Lieblich: You said strong divergence between testing and data set. Seems it's in false-alarm data. Subsets of full population in both sets, but you had 500 people. What population would you need to sample to have a representation of population?

AH: 10,000 would be perfect. The first dataset we collected we used our own peers. Our people are not a representative of the full BMI of the population. You need something that's not a biased dataset. That will also help false-alarm rates.

DL: Has anyone done a study to know statistical significance?

AH: I wish I could comment on that.

??: I don't know if the data set would be significant. The statistical side you can see, though.

Ken Sauer: You have to.

AH: That's why they won't tell you what it is.

KS: As long as it's a statistically relevant sample.

AH: Datasets improve with each iteration.

??: We cannot test with thousands of people. We do our best. We use a number of threats and a good mix of people. We have as many people as we can for research. We hire screeners and train them on male versus female. We assume the genders will have different images. Then we also consider the response of the screener. It's a screener decision about what they see.

Steve Smith: Follow-up on George. Tech 84 can provide images for 3rd party testing.

CC: But vendors work under different regulation.

Richard Bijjani: We need to see images you're concerned with though. As far as sampling size, my concern is really security. We've seen it time and time again; the bad guys know what our problems are. Somehow they get that information. If I were a regulator, I would put a lot more of the real threats. You cannot under-represent those threats.

AH: (???)

RB: But as a vendor you know you need to do that first. So if you have 20% of threats you're being tested on, and your data is only (???); that explains the discrepancy and the number. They won't tell you the test.

AH: We can discuss if that's appropriate.

MM: I think that's great. Maybe it makes sense that slicing based on materials, but for AIT it may make more sense to do on body-type classification.

RB: So it's based on government feedback.

MM: It's the underwear bomber. They're interested if they can find the threat in the underwear, not what kind of threat.

CC: We have a session on this tomorrow.

Doug Bauer: Extraction of data (???)

??: Both machines were being tested with goal of generating ATR. That was 7-8 years ago. Vendors will not do anything until TSA puts money behind it. This is not something new, but now people are buying machines and now the customer requirements are most important.

Jean-Claude Guilpin: In Europe, due to privacy issues, we will not buy specific types of machines.

John Bush: They need to appreciate that we're here. There is a traveling passenger out there who is also looking to be satisfied. Sometimes we lose sight of the fact that the product has to meet security requirements, but the product needs to be fielded with ATR. We can do a great job with security, but eventually the public will push back. They might have ill will. If we look at this from product development POV, we need to know that the public has positive reactions.

CC: Jean-Claude said his dataset was 15. I don't get why you have 500 and still need more.

AH: I'm interested to see who his people were. I don't know how he can get a full scope of the population with that few people.

JC: We have to use a bunch of different types of people, but we don't need you to choose specifically based on the test. We want you to choose based on population. We know it's a limited number of people. It's a reasonable cross-section. We're doing the same thing with baggage. We get a good enough representation, but we know it's not perfect.

CC: Thank the speaker.

Speaker: David Perticone

DP: I want to explain the industrial algorithm development process.

(Summary slide)

I'm going to put it into the perspective of full system development, talk about preliminaries, and then talk about what we do.

(Development slide)

(Solution Space slide)

CC: What's your definition of detection?

DP: PD, whatever that is. It's important to note that ATR projects are only on a 2-3 year scale.

(Goals slide)

DP: PD and PFA are always predefined, but how they are measured changes. A little bit here and there can be the difference between passing and failing. Most of the vendors know the pretest and pre-certification standards; it's all pretty well understood. Now we've got a new system that defines a Rubik's Cube size object in 30 seconds. There was no place to test this so the regulators had to show up. This is the difference between regulatory and pilot testing. On government contracts difficulties may occur.

(Major steps for regulatory approval.)

(Algorithm development sequences.)

DP: You don't always know what the test involves or what the sequence is going to be. The data collection is a very important step – garbage in, garbage out. You need to think about the clean data, but with people, we have to collect volunteers to go in for the body scanners. Simulated data is very useful for physics, it's better for design than the algorithm work. The most important question is will it work, will it meet the spec, but that's hard to do with simulated data.

Steve Azevedo: Are you saying that slowed us down?

DP: It does add cost and it does add time to the schedule. I'd say the simulation is probably part of the due diligence but it won't tell you the answer.

??: Can you speak to your use of experimental design?

DP: We have zero statisticians.

(Segmentation slides)

DP: Creating feature vectors is really like an art form, because there's a lot of things when you make an object that you can decide are interesting.

CC: What do you mean art form?

DP: You'd be surprised at the features people can come up with that really decides if it's a threat or not. It's not obvious. You have to have quite a few features to make good decisions.

Richard Bijjani: If your features don't represent physical qualities that are consistent with explosives, aren't you setting yourself up?

DP: Absolutely.

(Classification/regression)

MBS: It seems to me that the key things you're focusing on are the features themselves.

DP: The features and the models.

MBS: For the academic community where the features of the real threats may not be obvious, can we come up with data sets that will still have value even though they are not classified or SSI, or is the whole thing moot?

DP: I can't speak to the regulatory aspects, but I think you can certainly work with simulants. What's the bane of the algorithm developer? I definitely think it's over-fitting. When you have a simple algorithm, you tend not to get very good predictions. But when it gets too complex, anyone can sit at a desk with a data set and get perfect performance. You can easily fool yourself. And then you go out and test it and you don't pass.

Richard Bijjani and I would go down to the tech center and get the suitcase and put a lead dot so we'd know where the object was. But what happened was that the algorithm actually became a lead dot detector, and when we went to the test there were no lead dots in tests!

RB: If you throw a database of numbers which are features and do tuning without actually knowing what's happening, this can very easily happen. Give anyone a set of data and they'll come up with one hundred percent success. Sometimes the data is not necessarily representative of the real world.

??: The big question in my mind is, how do you optimize the data you're allowed to take?

DP: You get as much high quality data you can and you try to implement the efficiency you collect from the test. Public sets and performance help us dive toward the goal.

??: Throwing in some possibilities of confounding might confound it.

Luc Perron: There's also the other way around. You could introduce artifacts in the test set from the regulatory point of view that will destroy the results. We've come through that a number of times in detection of liquids, for example, that are inside a box.

MM: This discussion is getting way too cynical for my taste! To keep my job, it's important for me to pass the test. To do my job, it's important for me to detect explosives. At least from this vendor's perspective; and I'm sure for everyone here, we are worried about passing the test but our main goal is keeping explosives off planes.

DP: You really need to know what your algorithm is doing and you want to get to the nitty gritty. It's important to recognize the process of getting regulatory approval can be months to years.

Ken Jarman: Is there a big discriminator between the top algorithms?

DP: For us it's getting the correct samples into the correct buckets. It's very empirical – does it work or not?

CC: Coming back to features, it sounds like you have more than the basic four. How can that be?

DP: The algorithm is too simple.

LP: It's all about context as well. In a regular bag vs. a laptop, the results are not going to be the same.

DP: You typically have several algorithms running which might look at a different type of scenario.

??: You have pixels that are arranged with respect to each other spatially.

MM: You have things like variance of the value. You could treat that as a feature, but it's a lot easier to look at common causes and deal with them in a case by case basis. There are realistic passenger bags and unrealistic passenger bags. Your features tend to match to what you might find vs. what's happening at the physics level.

DP: It's either identifying something that's interesting or looking for the anomaly.

JCG: We have access to more threats than you, you can help us by doing some (???) on your technologies. We don't want to defeat a particular technology.

We don't want to defeat the technology because of the way that we design the test. Our wish is to discriminate between good machines and bad machines and not to necessarily defeat a very good machine. I would like you to think detection of explosives rather than just passing the test.

CC: You need to protect your test; vendors need to pass this test. How can we bridge this gap?

JCG: We try to use bags that are representative and threats that are representative of threat categories. We can't test for every variant of dynamite on the market – there are thousands. We assume that our tests are representing overall the possible ranges of detection of the machine. We assume that we are close to the PD of the machine.

CC: Are you better off with simulants?

JCG: In EDS we don't use simulants. We are not talking about commercial (???). We would like to work with simulants but in this case we can't.

David Lieblich: What we do is we find other features from other information that we have to discriminate. We design based on features that are more or less obvious, in what you might consider a direct approach. We use more sophisticated features.

MM: There are two types of testing, white box testing and black box testing. Send someone with no clue about the system and see if they can break it, or send someone who has a huge amount of experience and see where the algorithm might be iffy. I can build a specific bag to defeat our system and I suspect I could to defeat our competitors' systems with some practice. But does that mean that these are bad systems? No. It just means we know where our systems could use improvement.

RB: I really like Dave's chart, but the fact is people do pass the test. Regulators are in the business of trying to detect explosives and they do a good job of that. I don't think we end up where people are just studying for the test. But regulators could make every machine fail if they want; they do enough pretesting to know. But they are very logical and pragmatic people and they know the configurations of interest that they need people to find. And it works in all the regulatory environments I've ever seen.

??: I'm not quite sure what your question is.

CC: My purpose here is to ask open ended questions to stimulate conversation.

??: Kitchen sink is one way and a valid way, but there is also basic machine learning, context, edge variants and you can try to define a priori whether they have some correlation.

DL: Earlier you asked how to close that gap and I think that's one of the most valid questions in this whole process. The communication that we heard from Jean Claude is valuable. It's good to know generalized areas and scenarios where you're weak, not the key to fixing the system, but just in general so there can be quick and proactive feedback and turnaround. The better way to do that is through that feedback and communication. I think today it's done more ad hoc, vendor to vendor and company to company, it's not really a well defined process. I think it's getting better.

??: There's another dimension that goes into the page there and that's how much data you have. As you have more data these things can be easier.

DP: The more data you have the better your life is, but we're never going to have a Google amount of data.

MM: What you said is true but it's on an exponential scale of data. Even Google can't do it.

CC: What if you go to the test center and every stick of dynamite has a piece of lead in it? Do you train on it?

??: If you want to find something that doesn't look exactly like that, then you have a problem.

GZ: You can outsmart yourself that way.

DP: I think the lesson here is that it's not the job of the industry to game the system or vice versa. Industry and regulators need to work together.

Speaker: Sam Song

SS: This is a program we just submitted back in July this year. I think a lot of you know about the CARS program. One was based on the Acura. Basically they have two scanners (built by L3) staggered. The other scanner was built by Rapiscan and that was single energy.

Depending on the PD (or PFA?) that we were getting, we were comparing that to what we got, and we didn't worry about that. (Second slide explanation)

Not just because of the scanner, but by looking at the scanner, we had to fix the PFA by 10 or 20 percent.

Second, the highlights we put in, along with the background, enhance the performance. We convert all of the measurements to steel. We compensate the background and (???) region.

The LP scanner was actually two separate scanners, high and low energy. When you take these images, and do the subtraction, we had to do functional demonstration, to do the visual of it. There was a bump vertically in the mono-pixel.

MBS: What kind of threats were you looking for? The feature space is very broad.

SS: Actually for the liquid problem, the check point, we are talking about water at 7.5. We are trying to distinguish between 6.5 and 7.5. It's a factor of almost 3. It was easier to discriminate. This was actually trying to detect high-material.

I think I mentioned most of these items here – we wanted to devise an algorithm. But finally with all of the features, as mentioned earlier, they wanted either a green light or red light.

In conclusion, we have shown that the original CARS goal was over 90% PD, and less than 3% PFA.

Conceptually, we want to pick cargo here, this would be driven by the scatter lights. Dual energy (???) and we have the detection (Slide 4 explanation).

(Slide 5 explanation)

There are a lot of regions that could be non-visible objects.

(Slide 6 explanation)

We categorized the detection as low-low (LL), other options are low-high (LH), high-low (HL) and high-high (HH). This really helped us tremendously. Low density cargo, has almost perfect detection. No false alarms.

More density and high complexity. We would not want to use the same progress.

(Slide 8 explanation)

As it goes through more material, it gets more photons. When we do that, it turns out that by adding the low and high, the materials (???) to line up here (gestures to middle of graph). Aluminum is at the bottom, with plastic.

CC: Why is one worse than the other?

SS: I don't want to get into that, it's from two different companies.

The scan time was 30 seconds per truck. These are the pixel data, and there is no variation here, but any pixels here.

(Slide 10 explanation)

Because the material is future, it winds up right here – if we do the background compensation it winds up here.

We can classify it as high beam. We are trying to ship all of these points to here. (Slide 11) the cargo classification and high density, and cargo intensity and complexity to compare these high frequency and low coefficients that are LL, LH, and HH. The data we have, we are able to achieve a perfect classification of the cargo type.

For the cargo type of LL, it is almost perfect. LH, it's a little bit, but for H density and High complexity, it's harder to combine these. We have to maximize the PD.

We go through all different combinations to get the maximum feed.

??: Don't you need to have the frequency of the individuals?

SS: It's private here.

The conclusion I went over, but want to end with this slide. We only look inside the cargo region, we had engine blocks, tires, they told us it's okay to share this. I believe that all of the vendors were asked to not (???). And they got the same rules that we had.

??: It proved the point that 3rd parties can't come in and look at other peoples' hardware. Building tables, the program was meant to go through Phase 4. They were very happy with our performance.

??: What testing did they do with CARS?

??: It wasn't cars, it was other stuff.

??: The reason was that systems were not doing well, and the management decided that we were looking to develop something that could be secondary; so they did another iteration on the development.

CC: Cheaper, better equipment is required?

??: There is no question that these systems are state of the art. But when you look at the cost, you can't do it cheaply. That's the problem. There is no balance between cost and what it's worth.

CC: What's the business relationship between Telesecurity?

??: They got all of the data sets from all of the systems.

??: But you also didn't have to build the system. This is the program where we were asked for that simulated data first.

??: Did you use the line scans?

SS: For some reason they don't use that. Might be that it's too big.

??: Then can you tell us where the false alarms were coming from?

SS: A lot of it was from the registration artifacts, because as an example, if we are looking at a pew, it's slightly tilted, it will be way off. We would typically mean we have to get rid of a lot of it. Some of it stays still.

CC: Thank you!

Speaker: Justin Fernandes

(Slide 2 explanation)

JF: Our requirement for whole body imaging is required, so limiting those costs gets the operator out of the room. What we do to generate our images is to scan a linear array. This gives us a whole 3-D image.

PNNL was asked to develop the next generation body scanner. You are not going to see anything on the sensor, and no one is going to see it. We need better resolution and we need tight resolution.

How do we create the correct data set? We need to get the correct sample population.

(Slide 3)

We are here to detect threats. There is a lot of public backlash about mm-wave. For this we get intensity, we also get organization, and different man-made objects. We use this from multiple angles.

(Slide 4 explanation)

Speckle phenomenon – you can see that the false positives are in the system. Its different here (slide picture 1) you can see that the speckle that you get is due to the frequencies we are using.

??: How does the bandwidth affect the speckle?

JF: If you can reduce your range it's (???), you can reduce speckle.

You see again that it picked up the plastic gun.

There are different methods you can use for this. Here are manmade objects, and here we developed a thesis (???) segment. We transformed it into a ring.

Using an artificial network, (Slide 11).

MBS: Many of these features are hidden under clothing

JF: We are still working with clothing. The reason that the old versions failed is that the data was inferior.

(Slide 13)

We subtract a subset of these points, and see here (slide 14) this target has various threats, and a button down oxford, very basic. Here is our reduced subset,

This is developed to create the surface method, and it's an almost water tight surface, that we need to characterize. With Kinect, there is a lot of research being done with 3D data.

(Slide 17)

??: When you say that's significant, that's not significant by a factor of a 1000.

JF: I can't comment on that. Our current system will be much better than the one used for this. I can't comment on the bandwidth. As you can see this is going to significantly change our abilities. The multi-path in the spine is gone. That was an issue. The dielectrics (???) are gone. The conclusion is that with better data, we have better features. Context is everything. With the advent of these comes better data.

(Slide 18)

(Slide 20)

We need data, we are an imaging group, and are just getting into performing, and as we said, there are body mask types, but for us we want to open the scope up to increase the diversity of our data set.

This is a very simple set of targets and scenarios. The total amount of time this will take to do these measurements - 400 hours to do these measurements (slide 22). That is just to give you a real world idea of what it takes to do a data set.

CC: It seems like you don't need an ATR for this.

??: Having the 3D data helps, but there is more data, and more information, the reduction. This is a factor of 10 or a factor of 100.

(Slide 23)

??: Detecting an anomaly can actually be caused by the body itself, so detecting it on the surface, but detecting a watch on the arm is normal, and on the shoulder is not.

MM: It really doesn't matter if you get the world's greatest resolution on the bolt on someone's chest, this won't help you. The 3D info is key. With existing systems out there, you do have info that is not exploited.

JF: It allows for more algorithms to get thrown out.

Luc Perron: It hasn't been explored. There is some information on the system, but not to the full extent that it could be. It mostly works on the 2D image. Without taking into account other information. It's not because the technology couldn't do it, you just have to exploit all of the info that's there.

Speaker: Sondre Skatter

SS: Detection of liquids and amorphous threats in X-ray diffraction (XRD). My overview is that the X-ray diffraction is used by different materials; the spectrum has more information.

(Slide 2)

We are working off of 10 features that are working off of different materials, over 100.

Liquid and amorphous threats - we are working on the technology for Current XRD 3500j; this here is what we are working on. It has a pre-scanner girth, but it is used to target bags.

It was invented for overcoming the weakness which was very slow.

In terms of speed and cost, it was not as good.

Basic imaging concept: When you get the spectrum from crystal material, but the new challenge of liquids, we need to move beyond that.

But this number is a really good job, but there are more features, and even though they don't have a peak, that tells us a lot about the materials,

I will not discuss the individual features, we have hundreds of materials, we have bags, and lab set up, but the representative, were supposed to look like explosives. It gives us a good way to play with our systems lab. There are different materials that are non-threat. We have them in different categories (slide 7).

We have different noise settings, and if we get different results, we will have different settings.

??: Can you say something about the noise?

SS: It's counting.

??: Are the counts listed in the previous slide? Are those totals?

SS: I can't speak to that.

??: Most of these work well with geometry, but they have a focal spot, for a generated beam?

SS: We use an automatic beam. We don't have confounding of the scatter. Multi-scatter, you've made it. If you look at scatter at a certain angle, you can get different scatter energy.

SS: Do you have a picture of how you vocalize that? A bag that you detect: and you need a spectrum for a reason.

SS: They have one way of doing it, but I can't speak to the next generation of how they do it. I don't have a picture.

??: How big would that be?

SS: That's why we are labeled (???). We can actually start talking about images. This feature is compressibility, it can go on (???).

CC: Can you speak to the spectrum for compressibility? This is all invented in the days of physics.

??: What are the red things?

SS: These are washcloths. You have a lot of them outside of the box. Here you see the results (slide 15) and the material here is a high probability here.

(Slide 16)

CC: Go back a slide – I think there are (???) there.

SS: Maybe there is one, does that confuse you?

SS: So now you can see what's false alarming. You have 1.7% false alarm rate.

??: I didn't get it in the beginning but how are these measurements being taken? Single samples?

SS: Yes that's right. The conclusion is that it's working pretty well. It seems very promising, but we have a lot of work to do.

??: What's the typical scan time?

SS: The existing XRD scanners are slow, but this we expect to be 200-255 per hour.

David Castanon: What would be the distribution?

SS: That's something we could look into.

DAC: That's something additional you may want to say.

David Lieblich: You said something about doing better when you get smaller cubes. Could you expand on that?

SS: One thing we see in the current system is that you get a real threat.

??: I saw a bimodal distribution.

SS: Yes, but I split up two groups into one.

Speaker: Lisa Sagi-Dolev

LSD: A bit of space will be taken now from what we were talking about before. Now we will talk about threat detection for public venues.

(Slide 1)

LSD: What can you actually accomplish by having an algorithm? Everything here is driven by that. This is an auto self service security kiosk. You can scan a card to do that: a ticket, boarding pass, anything with a barcode. It locks, turns red, and if things are okay, the person can go. If there's a threat and 100% alarm the systems shut down. If there's an in-between, it alerts a threat and then people are notified to handle it.

(Presentation of how it works)

??: So is it one purse or bag?

LSD: No, it can take multiple items too, but it varies based on the event and location.

So our conclusions are to know what we are looking for. What about you as a customer, and your customers? All of those things need to go inside of a requirements list. Key here is throwing out what you don't need. Then we need to know what you're doing with this information. This is important for segmentation. That brings you to what kind of data set and algorithm you'll be using.

(Slide)

LSD: For aviation, everyone knows the requirements. The language you'll hear is PD/PFA. But in venues it's a different story. Every venue is different. Most important is people flow. That's very different from throughput though. 80,000 people need to get into one game. All people have to go through in a pleasant way. #2 on their list is guest experience. Even if security is a C, if customer experience is an A, that's good enough. Then comes privacy; then conops. The number one driver is the customer experience though; then cost, and then their threat matrix. There's not one unified criteria because event spaces are all different depending on who is attending, the location, or the event. So there's a bit of a difference in thinking for events than for transportation.

(Slide)

LSD: So our approach was dependent on the sensors were (???). At the end of the day this is not so trivial because what I think is happening is that we're trying to make better what we have rather than taking a paper and asking what you really need to go forward. We recreate what we currently have, rather than developing something new. A huge reason to have 3rd party collaboration is having the freedom to not be constrained by the horse trail.

(Slide)

LSD: What are we specifically doing and applying? The first important thing is what kind of data do we have? We take full responsibility for our own destiny so we have our own way of collaborating. We have a guide book that outlines explosives and guns. Though, for our venue, a threat is not just explosives. For instance, a cap of a bottle can be a larger threat because of how people can throw them around. Another venue needs to stop cameras, so yes they want to stop explosives and guns, but they need to find the cameras. So my algorithm needs to find all of these things. So we create bag sets for all of these different items. They're also created dependent on geography. So we have 1000's of these sets. Explosives we do elsewhere. Everything else we do in house. Our testing is on live sets. We have over 10,000 data sets. We have methods to be able to work with our data. We don't give out our data, but people can work with us and access our data through secure ways.

(Slide)

LSD: So where do we live on the ROC curve? Well first we have many ROC curves. We need to let our customers know what they should be aware of and what they should take care of. Something concealed may not be as easy. So it's not just embedded inside that curve. Different venues have different curves because of the needs and types of environments.

??: Are you making your own equipment or are you purchasing and integrated?

LSD: Core technology: we take off of the shelf, but the rest we update for our own purposes. The whole objective is that the customers have the best security there is. So anytime there is an upgrade in technology, we give it to our customers for free. At the end of the day if there is a problem, it's our fault.

??: In the video you showed, you didn't explain if there is an alarm. What happens if there is?

LSD: If there is, there's a door that turns purple. Then, depending on the venue, the response is different. It's all human based though. We can give the okay, but if we don't then give an okay, they need to do a secondary hand search. The human in the loop here, the most important thing they can do some profiling. Yes, profiling is important in this situation. There's no algorithm in the world that can sniff out something being wrong with a person.

John O'Connor: Can you elaborate on core tech?

LSD: Dual energy X-ray, ITM Trace, and nuclear detection, basic gamma detection. Our philosophy is we want some safety. Yes or no, handle it, take care of it. Then we wonder what exactly it is.

RB: Are you worried about radiation leakage?

LSD: No, because there's a locking mechanism that locks the X-ray. We do our data collection on live materials. Contamination is tough. We are also looking at trace. We have a different R&D unit.

MBS: I think the goal of the airport screening is the same thing, yes or no. In public venue the throughput is more. But in a public venue a terrorist would probably use a similar threat. So you must have thought through a tradeoff of the yes-or-no versus the experience.

LSD: Well no, because there is a tradeoff in venues that doesn't exist in an airport. There are still different stories because in airports it's easier than in venues to bring things down. So our life is easier.

Speaker: Robert Nishikawa

RN: Conclusions: Most important factor in developing a large CAD system is high quality large database. In medical world we don't have automated system. One big difference between my field and yours is some disclosures.

(Slide)

RN: I'll talk about mammography because that's my field. For mammography, the false-negative rate is 50%. It still works though because if you miss it one year, you'll find it next year. Cancer grows slow enough that it's okay.

(Slide)

RN: Need for CAD Slide. Nothing in image that's small high contrast. You'll actually need to see zoom image to see contrast. So there are things in image exist that aren't masses. That's what makes detection difficult.

RN: Mammography is most developed system in medicine. Some systems also used for lung and colon cancer. There's another long list too. In terms of development, most important is the database. A case in medical images usually has only one location, one cancer. This is different than other threats in travel that can have many.

??: Do you have same problem in the regular population, that can't really see what's there?

RN: Yes but because cancer is rare, you're probably correct.

So we divide data set into 3 parts, and then when it goes to be evaluated we have a separate database for that. The reason it's important because in what we do we need large numbers. Unless you have enough cases, you cannot select an optimum. We can develop our algorithm, train classifier. We can decide if it's benign or malignant based on numbers the algorithm outputs.

(Slide)

RN: You have to go through the FDA in medicine, so they have to ensure safety and effectiveness. Once you get your algorithm approved, then you can fine tune and improve. Those don't involve same level of evaluation. Systems are used by radiologist. Typically there are 300 cases, 15 radiologists read them, and then it has to be a statistically significant improvement in the curve for PMA.

??: So there is a number associated with that?

RN: All I can show is that it's statistically significant. Not that it's a specific size. There are 7 studies that found 9.3% increase in sensitivity and 12.4 %

increase in recall rate. It's a relative percent increase. Study design is tricky to do evaluation without bias. Even in simple design there is a bias which is tricky.

Some higher level issues are still important. One big problem is indolent cancers. Also benign lesions, a false negative (FN) on aggressive cancer can be fatal. A false positive (FP) adds cost and can affect workflow. There can be some similarities here. Same problem can be on FN and FP between security and medicine.

CC: I think FN can be more expensive in security. Because on a plane one FN kills everyone. In medicine, one FN kills one person.

RN: True. Some differences. Mammography has 2 views of each breast and temporal comparisons. Need to be concerned about radiation dose. False positives can lead to very expensive costs.

MBS: 3D versus 2D is emerging as a technology. Have you looked at algorithm in that domain?

RN: Yes, but with new technology it's tough to get images and do anything from that.

(Slide explanation of mammograms: Observer study). It's not true that if you put a mark on a cancer that a radiologist will recognize that. The reason that is, is because different radiologists have different thresholds for what they'll call cancer.

RN: So problems we have, why this is typical is a radiologist has to believe the computer will help them. One true mark for every 999 false marks is a problem. This is cognitively impenetrable. You can't reduce false negative rate. We don't know if that's true or not. The other thing is there's potentially a big learning curve.

??: How do they ever get past the trust?

RN: Because when we do the studies the prevalence is higher.

This might work in your field. Anything they think is funny, they mark. You can run high sensitivity. There are people who are developing methods to take advantage of data and learn that have complete truth.

Rick Moore: Can you make a comment about basic adoption and reimbursement?

RN: My personal opinion is that there are enough papers in the literature to cast doubt on whether this works or not. A lot of radiologists don't think it works very well except for calcifications. In general I don't think that radiologists think it's that useful. The reason why people buy them is you can get reimbursements- that's my cynical view.

??: Are there any concerns in terms of people?

RN: Originally there were, but I think that's more or less gone now. It's back to what it always was.

David Lieblich: Is this because the criteria are just that you have to have a statistically significant difference?

RN: That's a good question. The problem is that the prevalence is so low; you can't gain experience so that actually works. They'll read literature but don't care until they try and it works for them.

CC: I come back to Matthew at the beginning who said we should put the human back in the loop, but how do we test humans for rare events?

LP: That's the idea of TIP, isn't it?

Speaker: Luc Perron

LP: Airports are struggling to meet their security mandate. Who is the real client? Is it TSA, the airport or us? I'd say the traveler, but the airport comes second. So we need to concentrate on risk-based approach and operations.

We've been known for developing solutions in the detection world (Optosecurity). We've combined 2D imaging with 3D modeling for liquid detection as well as the science X-ray. We implemented the solution and deployed it certified in Europe. For us, an X-ray is an X-ray; we use the raw data out of it and convert it into something that can be used in the field.

(Risk-based Security Screening slide)

LP: Adding unpredictability to the "walls" makes it much more difficult for the terrorist to go through.

CC: But when you move your walls you're also opening up other holes.

LP: The concept is to try to adapt some of the thresholds but make it unpredictable for terrorists so they never know ahead of time what they're dealing with.

(Typical person bag screening process)

LP: "Clear Bag" - Instead of looking for a threat, we look for the absence of a threat. That's what the algorithm does; we look for specific images with very low content and very low probability. It's not just a density finder; it goes a little bit beyond that. You need to do a little bit more contextual analysis. You need the shadow of the side, you need to consider that and process that and recognize some metal. You can now filter that out a bit: Electronics, etc. So you have to fit some intelligence into the algorithm.

LP: This is why we're talking about the risk-based approach.

??: So you're talking about changing the way security is done.

LP: Yes. We're going to do some trials where we compare the results from the algorithms with the results from the screeners, the exact same images, and see what is in the distance. Someone who really wants to go through with a threat will normally try to hide it and not leave it out in the open.

RB: As a private citizen, I want to make sure that 9/11 does not happen again.

LP: There are different opinions related to that, the preventative measure has been to lock the doors for that specific threat. What is the threat? There are a lot of people reconsidering the threats that we should authorize or not authorize. This is more about risk management.

RB: But in aviation security already generated a list of prohibited items. It's not our job to say what is OK, what is not.

Julia Pavlovich: But we can add certain things to the ATR list.

LP: We're talking about saving customers screening and putting money somewhere else, so it's always a question of compromise. It is a paradigm shift. This is something that we are doing now in EU. We're also doing other things like monitoring and linking with other types of detectors and an integrated point of view providing greater awareness.

JOC: It sounds like we potentially need better equipment at the checkpoint to maintain that capability.

LP: Yes but that's a different discussion. Our idea is to automate part of the screening process without worsening the system. When we talk about adjusting the threshold we're talking about some features that we can activate or deactivate so they can select the threshold that meets their own requirement.

RN: With certain things like an iPad, can you do just straight pattern recognition to see if they've been modified?

LP: Yes, but there are multiple tablets. If you recognize some things you can essentially eliminate them, like say workbooks. So you're adding contextual information so you can filter out some common items without a security risk.

RB: You also need to keep updating with new technology iterations.

GZ: The TSA will not accept shape as a criterion. We have considerable problems with people trying to game the system this way.

??: What is the advantage of trying to reduce the screener workload 5-15%?

More in depth screening of some, or reducing screeners?

LP: Both, depending on the airport.

JP: I think there are very specific examples. If there's new tech and you are able to do something you couldn't before.

Speaker: Jody O'Sullivan

JOS: We have a number of classifier designs and that's representative of a large team. We're also looking at different technologies. There are several key components of the CAXSI slide (System Vision). There is a plane with various patterns on it that will allow some to go through and some blocked. We are thinking very carefully about how we modeled the physical signals and how we characterize the signatures. Those signatures will determine the performance of any classification algorithm. We're thinking about the joint design of the aperture and geometry.

There are a lot of limitations; we do use poly-energetic sources which blur the signal. If you're trying to figure out where something is, the nature of this design is that photons could be coming from multiple sources, it's multiplexing. So we have to do a de-blur. Those are the key limitations. So the ideas to combat them: We're directly trying to overcome the blurring effects through the design of the coded aperture. We have a rich, multifaceted design space.

Sondre Skatter: Do these apertures need to block all the X-ray flux, or how much?

JOS: One of the designs we're looking at would have a linear array to effectively measure the attenuation and block the fan. We are generally thinking about completely blocking the primary.

SS: It would have to be pretty thick to do that, right?

JOS: Yes. In general, we just want to make sure we're not getting so much flux.

SS: Is that part of the modeling?

JOS: We're exploring that. We're looking at higher resolution than was discussed earlier.

CC: So you're building an X-ray diffraction with a tomographic regime?

JOS: Rather than only scanning a bag rarely, we would scan every bag. So we try to get the bags coming through at whatever scan ray is needed.

CC: Why are you blocking some of the X-rays?

JOS: We're measuring the scatter because the primary is so much greater than the scatter. There are different motivations that can come from here. There's a measurement space signature, object and abstract/logical, that emerges from this joint design space. We see different opportunities for how you can characterize signatures.

The signal chain that we see is an X-ray source that can be designed in many different ways. There's going to be some scattering and over here there's going to be some sensitivity to the detector. The idea of the types of systems that we're looking at here is that we're going to have new opportunities for optimizing the systems design. For target space we're thinking about the standard momentum transfer. This has been inferred from measurements like the measurement that I had on the last slide. There are going to be limitations determined by the spectral width of the source. A lot of work that I personally do is physical model of the pencil beam/fan-beam.

(Physics-based model slide)

We think about the Poisson model if I do classifications, I use this and I think about using ratio tests and standard statistical methods to quantify the performance. Using the Monte Carlo data and some real data on the pencil beam data, we feel like we're doing pretty well.

The logical space signature, the choice of how we actually design this aperture, we started off with an aperture which was just parallel lines we place in front of the detector. If we applied it in front of a radial from the middle graph, we didn't get as high frequency as with angled. So how much information are we getting through the system?

(Singular Values slide)

For this one choice, there is a whole bunch that leads to much higher values than other choice. For the periodic *x* mass, we get a lot of ambiguity in what any given measurement corresponds to. We should get higher sensitivity. We're also looking at illumination by multiple sources, so you might end up with a source at the top or rotating around the bottom. This type of approach can be taken to think about signatures a little more abstractly. Finally we've also been looking at adaptive sensing strategies where we make some measurements and then adapt to optimize.

GZ: I would like to see a threat and a false alarm object and see how those two spectra distinguish. How is that ability to discriminate?

JOS: We are right now collecting measurements from hundreds of samples at Duke, by themselves or taking a simulant and putting hundreds of objects as clutter. We're going to have a signature analysis review on the project in a couple weeks with DHS. Classification: Performance depends on the problem. You're averaging over different scenarios, but each scenario is and of itself has a performance. We're trying to be able to describe any ambiguities including with clutter.

GZ: When do you think you'll have this data?

JOS: We have some of that data right now, but as a team we have multiple approaches for how we're doing this. We have some available but I don't think we have any published.

JB: With total integration, how long does it take you to get a measurable signal? What is a period of time? How long will a measurement take in real time?

JOS: So what is the MAS that we need to get the answer? Some of our simulations are very reasonable, some unreasonable, we're still working on that.

Speaker: Kirill Trapeznikov

The goal is to minimize measurement cost without sacrificing quality. (Slide 2 read through). We are going to try and generalize that and use it into our framework. We also have this one real live data set and the result of this work is that we can achieve performance by using only a fraction on average.

First I would like to give you two examples (slide) 1st and 2nd sensor. If you have 4 sensors, a low res and high res sensor (???) You can tell that there is zero and 1. MM-wave is a more costly modality, if you have sensors computing a feature. You are not allowed to compute simple features using this system.

We can formulate this as an objective. If someone give you this budget constraint, and use some features for some examples. This is our work and it's sequential architecture.

(Slide explanation)

If you are talking about time, you are going to increase both. I will briefly go into the approach; I can explain any specific questions later.

What is the risk of this entire system? If you make an error, you pay a penalty of 1, but if you want to take the next measurement, you take another penalty. If you know the probability of the distribution, this is an effective program.

Given the measurement, the quantity is given an expectant risk. On the decision space, you are ensuring that you cannot make a decision, you're going to make a standard decision. The tilde would be a half, and if it's less than a half I will classify it has 1.

We don't know the distribution, but we model them. We give an empirical risk of all of the quantities. Usually we are working with classifiers that are making a decision, but here it's with a reject option. We have a couple ways; the problem reduces to a series of learning problems.

We can do this to optimize one stage at a time. We need something to compare this to. One approach is based on a myopic approach, that doesn't take into account the system, but in contrary; our approach takes the system into account.

Before, we have two measurements, and the measurement is 3, an ambiguity region that I only ambiguous in the first region. Both measurements (???).

Ours here, we are going to fix the budgets; so this is kind of proportional of the green area. Only examples of the green region where it requests the green area here, it won't give you additional information. So for the same budget, our approach is lower.

Metrics evaluation, finding error vs. budget: This myopic method changes our findings.

There is a significant difference between each approach. You acquire old measurements from old examples.

Explosives detection data, we turned this into two stage systems, by then we are going to turn it into a classifiable sample, and the data is a 700 by 400 pixel image. We want a simple way to test our algorithm, and we have this one way to do simple pre-processing.

Let's just divide between each region here into 8 numbers. This way we can test if our system works.

We extract overlapping windows, and we compute intensity counts. This is for windows that don't contain the address, and for windows that do.

We can learn a window classifier for every region. This is just to (???) the advantage of the system.

The blue curve is the best you can do, and is the best of all examples, meaning the infrared and the green curve is not the absolute performance.

(Conclusion slide)

Ken Sauer: At some level, this would be similar to the simple analyses.

KT: This problem is solved if you know the distribution. It's hard to estimate.

Taly Gilat-Schmidt: I am telling you my area of research, which is the design of optimization of radiation, so we can optimize the scan so we can reduce the dose to certain organs. But related to security, we are motivated by claims we heard, and some of this chatter that was out there. The radiation that is similar to the flight. The energies that these scanners use are similar to other scans. The radiation dose is distributed differently than to two minutes of the flight. My student was asking how this was deposited in different organs. And we have to understand the organ dose. It is our best guess, and we may not be accurate.

We used our simulation tools to estimate the organ we said something about the results. It's what we expected, to put this into perspective, we have to give you an idea, if we have the X-ray backscatter to give people as an example.

We compared it to this, and it helps put it into perspective. This is in units of micrograms. The X-ray backscatter is lower, but here's where it hit home for me, this study calculated the radiation dose. We will get 2.9 from the (???), it has less dose than that.

The values for the backscatter are very low. Our study has limitations, so we say in our paper; it's only accurate to the order of magnitude. We compared this in the FDA report, and it's consistent. It's 1 or 2 micro-Sieverts.

??: How does this combine?

TGS: It's cumulative.

CC: Is it safe?

TGS: There is a lot of controversy, and I am not qualified to answer that question. Some people think you can recover from low doses, but some people don't.

People on one side of the room will always get more radiation on one side of the room than the other, it is negligible.

??: If you didn't have the bystander, than you couldn't be non-negligible.

??: But I don't have a choice with this.

TGS: We put out a press release so that we could get some press, but I told the reporter that there was negligible risk. That turned bad, as people reasoned on the internet.

When you are looking at the risk of radiation doses, we select organ dose per sensitivity. Then we use Monte Carlo simulations, which model the X-ray between materials. They actually track the photons that they track. Those are the simulation methods that we use. There was one in 2006; there was a single prototype that they use. They took those and leveraged this to look at the organ dose. From that, quantified the effective dose.

Radiation coming off of the scanner vs. on to someone else: We can use these tables in the future.

We thought that there were some limitations, but they have a single unit scanner with shapes, and Monte Carlo software with some annotations.

This is a prototype scanner from what's coming out. They use the FDA conversion factor. They don't look at the organ doses directly.

Of course, in both of these studies, it depends on what is used for this system. There is another study by Peter Rez; and he looked at the images and reverse engineered them, to get the number of photons. He found much higher doses than other studies.

We don't know what processing is used on the imaging.

??: Peter Rez no longer believes in that.

TGS: Okay, I do find that interesting. But the goal of our study is to find the dose of the organ, and to find more flexible Monte Carlo just like the John Hopkins report.

We use the virtual family of phantoms, for 4 different ages and different genders from cadavers. We thought that that would be more realistic. We modeled these tissues in our simulations. What's the dimension of this, in terms of radiation dose, what's the energy of the X-rays? How many X-rays do you have? How long is every area going to be scanned?

Not all specs are available in the public domain. We lay out all of our assumptions in our paper.

Two errors have been pointed out – part of this, they disclosed the distance between the X-ray source and the panel, which we didn't know. The interesting thing is that this is what we wanted. Now we have more information. They are listed on the ANSI standard. We appreciated this, and our original study estimated this much – and we issued a correction from this. There are micro-grades (???) the other tissue.

(Organ dose slide)

The ovaries still get a lot of the dose, and so it is penetrating. But the skin gets the most. The effective dose when we add it all up, this is the TSA and FDA, it's close enough for sources of error. Limitations are based on exposure measurements from the TSA.

We know we have some errors from modeling geometry. But I would not give a statement.How to improve the accuracy? It's hard to measure with the equipment that's out there: More information about the geometry.

My opinion – how to allay public concerns, we can inform that public of quality control and safety, and we need a third party study; Not a govt study.

What's the quality control? How do we know it's working normally, and once you get the better dose estimates, someone needs to quantify the risk.

??: What about scatter radiation, for the officer?

TGS: Operator exposure is unknown.

??: It's comparable to standing next to the baggage scanner. They are the same.

??: Is there info about the variants/ if it puts out 10 watts or something?

TGS: I don't recall seeing those results.

John Beaty: If you get a government study or another study, this looks like it should be trusted.

MBS: Can we converse over this at the reception?

15.2 Day 2 Minutes: October 25, 2012

CC: Welcome to Day 2. A couple of reminders: Please remember to take the questionnaire that is online. There is a link in your folder. We will end by 4:00 today guaranteed. We have George's presentation from yesterday; that we will also add today.

Speaker: George Zarur

GZ: If it isn't broke, don't fix it. But we need to fix it. TSA up until now had the option to do whatever they wanted to do, but they're not going to have that anymore. They're under a lot of scrutiny and there is a problem. What I'm trying to say is the way we were deploying things was not the best way. The concept is the government knows the explosives end of the business and they develop hardware based on image quality and performance. CTs should become standard.

So I think something TSA must do is put some requirements on what is to be delivered by the vendor. In my opinion it should be segmentation, because that is a reality. You do your recon, you do your segmentation. You should be able to say that segmentation is a percent of reality. Then TSA has the option to compete the algorithm. The current way isn't working because TSA has a problem with making the threats. We need to get away from the classical way of doing business. Right now you have your tortuous gauntlet of collecting bags, research, but it needs to be based on image that is based on how good the machine they want to buy is. This is a possibility; they are capable of doing that.

We will decide which threats will be included in the algorithm and at what time. For a long time I used to think that FA rate was most important, but it's even more important than that, because it costs a lot of money to resolve FA. So we have to do something different. Up until today we still don't know if dual energy is useful or not. It either is or isn't. If it is, then why are we buying single-energy systems? All they think is they'll be able to do field upgrades overnight and that cost is minimal.

The driving force is that the delivery system will force vendor to go seek 3rd party, in many way that medical field has. So we want to do the same kind of effort, to force vendors to go abstract the most recent R&D in the medical industry.

??: Why are you only singling out the medical industry?

GZ: Medical people have the best knowledge of CT technology, which is why.

??: True, but DOD may have more information about the type of product you're looking for.

GZ: Yes, but they won't get away from X-ray.

??: But you can apply images from different fields.

GZ: But if I tell the vendor, I can get them the best image, they'll want it.

Pia Dreisetel: Part of the problem is you can't get these systems. How will you do this in terms of hardware?

GZ: With limited success. We wanted to solve this problem over many years ago, but the TSA has said "you do everything". So when DHS said that they have money to solve the problem, I wanted to take it from them and try to do it. I could have gone to Israel for their tech.

??: I'm a 3rd party so I'm happy with the statement, but I'm concerned that you won't combine the other technologies to help with detection. Aren't you blocking the innovation?

GZ: The truth is the CT needs to be improved first. It's a significant problem that's very costly. It's the problem to address today. Coherent X-rays have been around for forever. It hasn't worn out. The Israelis have used them very effectively. I think TSA needs to rethink the way they're doing acquisitions.

TSA needs to know the minimum possible detection rate. It won't be 0. But it needs to be lower. I'm not saying it has to be 8 or 9, but we have to know. That is the target. If indeed it costs so much for resolution of FA rate. In the scheme of things, that's nothing. If you'd improved the hardware, you wouldn't need to spend that. It's a lot of 3rd party work.

??: How do you propose to figure out that number? I don't think it's possible?

GZ: We tried to do a few things that didn't work. I don't know. We ran a few where we could see a decrease in FA rate, so it'll have to be using high quality CT first to see how close we can get to the end-point .

??: You can get a pretty close to the product, by understanding that FA's come from certain sources. It's a function of the resolution

??: You can't figure out what's the best algorithm you can develop. That's impossible.

GZ: I think you can in some way. If you have the best CT algorithms, you can see what your FA rate is going to be.

MBS: Just curious, what if we could create a consortium of vendors, 3rd party, and TSA, and each put some skin in the game, and that consortium could come up with the next generation of CT.

GZ: The TSA was supposed to figure that out for quite a while. This may have some traction. But we need to educate ourselves in the interim. I'm saying the way we could do it right now needs work.

Speaker: Matthew Merzbacher

Justin Fernandes: Yesterday I'm not sure if I mentioned this, but DHS S&T sponsored our work and I needed to note that.

CC: When the paper comes out, please send out the citation.

JF: Sure.

MM: We don't know what we don't know. But surely we can expect to know more tomorrow than we do today. That's reasonable to expect. So dynamic ATR? Why should it be dynamic rather than static? It should change with the situation. It needs to incorporate changes in the environment, changes in technology, and changes in knowledge about the world and circumstances. So environment means new threats. When you search online, the first thing that comes up is Mentos and Coke. Changes in the environment for intelligence, policy and protocol for how we handle threats. If we have protocol that can handle high-res CT, then maybe it's important. Changes in FA. An example that didn't exist 10 years ago that everyone carries today is an electric toothbrush. Changes in technology, new solution, in knowledge. We need to adapt quickly, safely and in a well-understood fashion. We need to justify our decisions. We adjust algorithms based on something.

So the question, is Carl Crawford vetted? If yes, he's a low detection risk. If no, he's a high detection risk. So first we need to know who he is. Validating who the person is, how do you decide, based on my example, it's not so simple as plugging a person into a system. What might the "something" be? Intelligence information, passenger risk, specific threat catalogue, prior data and scans of item? All of this is dynamic ATR. And there's more. Recent similar results: have you seen this person or item recently? Is this something we are used to? So can you fool inductive systems? So then there are practical considerations, randomized elements, and other things. We need a framework for combining knowledge, control and information. It needs to be forward looking.

CC: Are you suggesting that we save images of people?

MM: I think so! But, in our policy is that viable or not? From a tech standpoint, it's a good solution. So the question is, how do we combine the result? Also, how do you control dynamic behavior? What you want is a simple risk-meter dial that a customer can understand. There needs to be just one nice knob so that in the end it's simplified. Also you should understand the dynamic choices. You have to understand what your system is doing at least at some small level.

So as part of a war story, I think voting can be great ways of making decisions, but sometimes there's an erroneous threat correlation. There can be a problem here, though, when voters go as a block. It can be a disaster. The take away I want to leave you with is limiting control improves reporting and robustness, but at the expense of optimization.

CC: Are you going to address how you would test the system?

MM: What should change in ATR is that what we really want is a static ATR with elements that are dynamic and that don't eat into the static base. (Slide on more sophisticated dynamic behavior). What about testing and evaluation with limited resources? They need appropriate testing at both component and system level. When you merge them all together it's all bad. Some techniques we can use are simulations, Monte Carlo, we have to have live testing, black box and white box testing who do and don't know insides of the system, ongoing/evolutionary. False alarms you want to fall off steadily. And there has to be an ongoing way of moving forward. Therefore we should prepare a framework to take advantage of tomorrow's advances whatever they may be. Understandable, controllable, tunable, and testable (???)

??: You also should have regression testing for predictability.

MM: Absolutely, so you don't have to test everything if you can just predict it.

RM: Can you make a comment. It seems to me that the dynamic can eat into static base if they have FP.

MM: If you lower static base low enough, then you don't have the problem, but if the static base is high, then absolutely. To make this field-able, you have to lower the base and add things on.

KS: How do you deal with risk of corner case?

MM: It's similar to what Rick was asking. You have to choose logic carefully about how you do your combination. And if you could use rigid logic then you escape from that.

GZ: Could this migrate to other systems?

MM: I think we'll find that out because we intend to participate in the reconstruction [initiative]. This isn't even a hardware issue it's different. We have 2 main product lines. It's not so hard, it takes some tuning but it only takes a bit of work. The one where we spend 80% of our time is the corner cases. I estimate to make any change is around 2 months. A big architectural change is going to take maybe 6 months.

Speaker: Richard Bijjani

RB: I'm going to take a slightly different path on ATR. First we start with conclusions. So we have an agenda.

What is certification? As many know, there are many different standards. The point is, you need to know what you're supposed to find. We talked about cert yesterday. It all follows the same model. We have different categories of explosives and there's a different detection level for each category. If you fall below the line for each category, you fail. Then you take an average and if you fall below the line, again you fail. All you're doing is adding different categories and having the same idea. Explosives detection system: A lot of people forget about the explosives out. Everyone who wants to participate in this industry needs to learn what an explosive is! There are many different types. What is the common element? For conventional, it's that they're all built in a factory. There's some level of QA. There's some variability but you know about this. The common element is that they all explode and it's what we're trying to prevent.

(Slide)

I did a Google search and this is what I found about an explosive [picture]. Here's the interesting part. A lot of people have always insisted that you have to give me the sample to find explosive. But since day one, we are shown the composition, and the huge variability of the explosive. You have to be able to find the explosive because batch A and B are going to look different. There's a huge variability.

(Slide)

Why is this important? You can find out a lot of FAs. If you're in a specific range, you see that with every 0.01 movement in threshold, you add a FA rate. This is a very expensive region. In other cases it's not much of a problem.

This is where it matters this is where false alarms come from.

(Slide "homemade explosives" explanation)

There are an unlimited number of known oxidizers. These are known; this is chemistry. (Examples of homemade explosives) If you look at the European standard vs. the US standard for explosives: The liquid detection standard in Europe that we don't have here. The rest of the world follows European Civil Aviation Conference (ECAC) or TSA.

Why are the requirements important? Anywhere in here we pass it, you need to be aware of what the requirements are. How do you prepare? Get data, develop algorithms and take test. Go to an airport get some bags, send to Livermore or Israel, and based on that data you get algorithms and when you feel that you're ready, you take the test. CERT management plan of 2010. Carl's difficult questions: It becomes very hard to go over, but we are saying to add more and more explosives, and it keeps your art down.

We are trying to anticipate what the terrorists are doing. And they are looking at what we are doing, and trying to get around it. What is the next step? We all have theories. It depends on what the ingredients are. First, don't look at your images as just images. You have to understand your machine. Know what your resolution is and your noise. Analyze clutter. Every one of the vendors has at least 1000 bags. How much variation is measured. It's not rocket science.

Third, again a hypothetical, Laurence Livermore and TSL generate a cloud, and say I want to find an explosive in this specific range, we don't know what terrorist will use but everything in this range will explode. Here is a sample.

When you take the test, you are going to be tested elsewhere. You start with the cloud, use your clutter and then you get a cloud, this is where this may fall. This black box concentrates on the difficult cases, the normal cases are easy. CEOs don't like that because they want to see progress.

(Algorithm black box slide explanation)

When you are designing your algorithm, you have to make changes, as there is no way you are going to get it right on day 1.

(ROC slide) Your job is to hit that box, start by taking your first guess to a point where you nailed the detection. False alarms change. In Israel it's based on current airports; it depends on the flights and the season. It changes from 12 to 40%. Get your data point, and know where you are.

You know your database, and you don't have to translate, and you get roughly what the false alarm numbers are.

You pass, now what? Your selection is going to go down. PD is at 100%, but we have an obligation to the public to tell them that these machines are not foolproof. At the end of the day, you care about how this system works. There is someone making a decision about each alarm.

??: It's not just a question of missed explosives: But because of the environment of this. What is plastic or glass, if you are not able to investigate this, you are focusing on the wrong thing

RB: Know your tests, know your explosives, you need to know where you need to be.

??: Known and unknown things. We have a lot of things that are built in factories, when you look for ways to cut false alarms,

RB: What people care about mostly, is permissible and non-permissible.

GZ: I know you can predict that based on single or dual, can that be predicted based on values?

RB: I can speak to that based on Reveal; we were given samples, we don't know if its 70 or 40 or 90, but it's very sensitive, and here is 40 grams. What are the possible other things that you can add to this material? What are the expected false alarms? This is what we predict the false alarm rate will be for each of these, and you find the range of properties, and it came very close to the predictions.

Speaker: Zhengrong Ying

ZY: I will share with you my progress that will be useful for the next generation. There are 4 areas that (???) In terms of the geometry, the existing cans (3rd slide explanation).

The medical world - they have some efforts in the community to mimic the CT geometry (???) For the next generation of the stationary CT, is in terms of two features, is the aperture position. Literally generating the position.

??: It is static or dynamic? For the stationary CT (???).

ZY: They are fixed. In the designing stage, you don't have to place them.

In the designing stage, if you align the sources of the images, if you scan 2D, the images for each scanner, you can have improvements of the image even on the second example. If you allow the modulation, then you actually have a

3D cone-beam case. You don't have to have the 360 degrees. It only requires 150 degrees.

Now for the fundamental is that if you look at all the data, they all come up with a conclusion, of thousands of outsources. But you don't need a thousand views. You can still get a very nice image. In terms of the fixed domain, what is the relationship? It's still in the discrete domain.

CC: Are you going to address the numbers needed?

ZY: You don't need that many, there are only 15 sources.

I think that the image for the detection, if you take the finite size of this, the progress is even more interesting.

The reconstruction, is object based, are still are a construction task. There are other categories, I am trying to show that the other generation of the (???).

??: How do you think that thin object reconstruction is different?

ZY: Good question, I came up with a background for that reconstruction. The field is still box or pixel based, and it's not deployed in the field. Both objects need the same for the thin objetcc, and sometimes that continuation is hard to maintain, as the surface of the objects, for the field, the most for this, I have not seen a target in the object, so back to your question; what is the difference between the thin object and the other object?

If you are talking about the thin object if thin and (???)

??: I am not giving out any industry secrets here, but this has been done for 16 or 17 years. What are we supposed to see here?

ZY: I don't know; I guess this has been done in industry for a long time. They have a specific kind of algorithms, so that is my understanding.

??: I can't offer specifics.

ZY: The difficulty is that they cannot offer those specifics.

In the classification field, if you look at the literature, the training process, the test results are not really (???) from the training samples equally. Which for humans to (???).

One of them is feature dependence discovery; I can confirm that the image process is.

For the feature example, the other effects of the CT are the objects. The other state of the object, we have to find teachers.

We have to offer the explanation of why this happens for the correlation such as this.

(Slide 10, explanation)

I would be offering the other way.

??: I need to talk to that point, because when you are taking that test, you have to get the data for that. Otherwise you get no data.

??: One other point is that you have control over both false alarms. You don't have control over par to (???), it's not accessible to you.

ZY: Correct.

This is the process that is seen to be feasible, that is passed in this way. Some of the samples are giving you some ideas is how do you put this up into your algorithms, as long as you maintain certain detection. How can you come up with that kind of framework?

The last area is the multi-energy decomposition you have two modes.

So many of these are in the back and we have another effect; some of things we don't know yet. We have more than two measurements. Because the data are different than the original measurements. In the medical world, trying to maximize some of the contrast.

In this application what do we use to go off of migration. From this application (???)

??: We will see what the target is but we will see.

ZY: that's right, can we actually draw from the cases. Can we come up with more generalized (???)

??: You have 8 different targets?

ZY: Yes I believe so.

I imagine there is a huge false alarm rate.

Speaker: Carl Maccario

Real quick - I think I became really interested in different techniques when people are hiding something and being deceptive. When someone from Virgin Airways hires someone from Israelis (???) Then TSA came along, they started looking at TSA; they started taking over for Logan airport. The gentleman started looking for the person who would undertake from the event.

I want to read you an excerpt: Imagine you are paying for an airline ticket in cash. Let's say you have poor hygiene, no baggage, and when he is interviewed by security officials, he changes his demeanor, the interview takes so long that he misses is flight, and he passes through airport security. He gets on the flight, and tries to ignite an explosive in his shoe.

What sort of behavior should we use to detect these or to find individuals? What should we look to see what these people are planning to do, or what risk they pose? When I came to TSA we tried find a balance between looking at the person through check point and not just what they carry. They were constantly refining the program, and we are improving the program.

It's important to go over what they are wearing; if you read an account from a lady who saw one of the 9/11 people writing down everything he saw.

I am telling you that the interest in this has gone through the roof. There are 13 countries have asked us what we are doing in our program.

The fact that you are engaging people, and training people to look for signs of deception. It's anomalous behavior. Going through checkpoint, 90 percent of this field is just paying attention. Technology is not stopping the bombing. Just like the guy in Times Square who saw a car where there is never a car called the police. You're ability to articulate the suspicious qualifies your observation.

That guy looks really weird, can you talk to him? What does that mean? But if I say this guy over there is hiding around here, and you tell the police everything you see. 90 percent of this is really just paying attention.

(Video)

What were the changes in this video?

21 changes. I have been showing this video to people for years, and only 2 people have been able to find this. Tunnel vision is 1, and 2, the fact that people don't pay attention.

In the behavior detection theory, the fear of discovery, it is managed through observational detection. It's something you can control. It's really a program to identify anomalous behavior based on behavior in a known environment. What you are looking for is deviations from that baseline. You not are solving world hunger, but you are finding behaviors. I don't have time to get into the training of the program. It's the engagement portion. The probability of the detection goes up immensely, if you engage the person and make a risk assessment. They did a study of suicide bomber attacks, part of the results of the study show that suicide behaviors are the same as drug smugglers. Also, the questions that we are piloting now, the questions that are related to the activities now. That (???) was important, that is when it's actively working. In my experience with the undercover community, one woman was held for 12 hours. Why 12 hours? If they suspect that you are going to do something, but can't prove it, if they hold you for 12 hours, you aren't going to meet the people you needed to, to get what you have to do be done.

I called the American Psychological Program; I got pointed to speak to Paul to integrate this into our program. We are working very closely with them to help develop the science of this to find research shown indicators. There is a 100 years worth of research about this. That is what we teach our guys to do.

??: How are programs different? How do you measure the effectiveness?

CM: One of the things we don't do is profiling; demographics get more scrutiny. We don't engage in the length of this that they do. We have two million people flying every day, we can't interview every day. We have Russian security people ask me what we do. I told him we are very high on the science. Other countries are focused on where you are coming from, who you are meeting with. We care about how you react. You will see most people traveling have nothing to hide. Out of a million people in this program, we have 227 people decline to talk to us. We are going through that right now, we have data that tracks the people that we need. How many people do we have? Terrorist are flying around the country. We have caught people who are linked to other things. I can't come out publicly and say we caught a terrorist. We always get the question of how do we test this. If I have a red team member, we are the fear of discovery? You can't create fear that doesn't exist.

??: Where does this interaction take place?

CM: Right now we are piloting this higher level of engagement, and actually we moved around to more than one checkpoint. It's a combination of different officers. It's not stationary. It's putting them up at the ticketing office. We had the secretary of homeland security, and she asked to go through it. We are looking at putting them at the podium. I can ask trip question, and we can do that right now.

??: What are your thoughts on the FAST program? If I get pulled aside, there is nothing telling me that it's not profiling.

CM: I can't stop someone from profiling, I will give you a real life example, someone with an Arabic name gets pulled aside, and the person writes all of these things that that person did that got flagged. But we look at the tape, and they were fine. So they were profiling.

We have machines that are a giant lie detector, and we have some of the best surveillance detection, and the technologies there, but we have to resolve it.

??: Do you have an opinion on automated mapping? Expression?

CM: I am very familiar with that. I am in favor of it, but the real value of this is in a static interview. It's not valuable in a line environment. There are companies out there doing it, and if you can get it working consistently, you pull them into secondary screening. You can see contradictions are cross-cultural. If I open a bag and say, can you put your bag up here and a flash of fear goes on your face, you can see it. There is either a bomb or something embarrassing.

The fear expression - is so easily detected. I gave my daughter the training when she was 13. She was 70 percent accurate.

Speaker: Douglas Pearl

DP: I'm going to ask a lot of questions and hopefully we'll get a lot of feedback. So what are the incentives in place for security vendors? (slide)

CC: Why are you asking this question?

DP: I have a hypothesis that if you improve incentives it will lead to greater third party performance and improved involvement. I am hoping we can get some comments and questions in the table as best we can and I'm also looking for concrete suggestions on how to improve the current state.

So the thing about incentives in a commercial setting (???) Faster upgrades and higher price. There are underlying incentives for hospitals to upgrade too. They compete and are able to charge more.

GZ: Currently there is no DHS incentive. That's the whole problem. There is no mechanism to say that if you do 5 points under this bar (???) we don't even know what is the true cost of operating these boxes; so we don't know the true value of improvement, so it's a pass-fail system more than one that applies for incremental improvements.

DP: Do you as vendors experience financial incentives for improvement?

MM: My experience is that there's no incentive whatsoever, which makes the ROC curve slightly wrong. It's not a box, it's a line. In terms of false alarm reduction, we have ample evidence where we have not seen any incentive whatsoever. Such as simple software upgrades that sit on the shelf for 5 years that we know would improve PFA, but nobody is willing to write a check for any amount of money for it.

DP: Is that because they don't believe you?

MM: I don't know; I'm an engineer. My understanding is that the financial arms of the TSA for procurement vs. operational are different.

??: That little thing that says TSA certification, that little star you put on your system, that's all they care about.

CC: Was the algorithm for distributing procurement ever defined?

David Lieblich: That would have been a way to incentivize within the US but you could look at the market in terms of the rest of the world. If you have a place where the airports do the purchasing, as in many EU countries, you have more of a driver than you do here. Frankly, the airport and the regulators are somewhat at odds. They want different things. The airports are driven by false alarms.

GZ: I think the Europeans are moving ahead simply because the airports are very sensitive to the economics.

DP: If you could improve performance at a higher cost, how does that discussion go?

?? (TSA): Clearly that's a discussion point, ROI. But as George has noticed, in the US it's a bar. Once you clear the bar, you've cleared it. But in other areas the airports are involved, so if you're able to justify a performance increase in x which improves $x + y$ above average, the ROI could be adjusted.

DP: So in Europe they use the cost of total improvements over life and here they don't.

??: Correct.

MM: But the lack of incentive or interest in upgrades is real.

DP: Are they saying that because of a fear about marketing smoke or do you have valid evidence?

MM: A little bit of both because I don't have access to my competitor's numbers, but I do know what my own are and can tell if we did excellent.

GZ: There is hope. We have our best possible opportunity with John Sanders being an administrator. He understands the business and understands the value.

MM: We're in a funny place in terms of the ROC in terms that we're high in terms of a standard. We're up at the top, so it's not clear to me as to the value of the 1% or 2% increase. The standards itself may be off.

LP: There are issues of localization and end users. There's nothing in the procurement system that has you buy one instead of another. Our market is in Europe, where at least once you meet the bar the airport has different incentives. There are different needs for different airports and then you can focus on their specific needs.

JCG: The airports don't know performances (in Europe) they just want to know if it meets the standard or not. They can choose a machine that is more expensive for whatever reason. It's solely for PFA, not PD.

DP: We've got a baseline understanding of some of the issues of incentives; I'd like to put some concrete suggestions on the table.

(Incentive Options slide)

So could we have a system that paid more for better performance after defining goals?

DP: Does any of this make sense?

DL: I think you get certified, and then you're in the real world. Everyone's in the lab before going to the stage. You're not going to be able to address a lot of these issues. I think the vendors that are able to work with the customers on these issues are way ahead of everyone else.

MM: There are various ways to do the incentives, but there are also political considerations here. Who wants to be the representative for the airport that has the low bar?

GZ: That's why it works this way.

??: Is there a good reason why the government certifies false alarm rate and not detection rate?

JCG: If it detects everything but also alarms everything (???) but after that, you could let the market decide. Because false alarms do have a cost.

??: If you bought a car you had to push, you wouldn't buy it.

JCG: And that's what the European Commission began to think about.

MM: So I come in with a very cheap, very high false alarm rate machine that passes detection. So an airport says great, I'll get someone with fast thumbs to go "clear clear clear clear clear" and now we no longer have security.

RB: As Dave was saying, one of the considerations is to pass certification and then you have to do the rest. And that's what the airports care about, that's really a priority. In the past the govt has given money to vendors to improve performance with mixed results. So there is some cynicism from the government side on that front. This was seen as maybe there's an ulterior motive.

DP: You and your current employer said they could improve PD and PFA for free, and it was met with skepticism?

RB: It was rejected. I understand as a vendor that they thought I had an ulterior motive. But with incentives, my biggest fear is – I got into security in the first place because, personally, I care about security. My fear is that once you add incentives you may start gaming the system. Most of your false alarms come from corner cases. So reduce detection in a very small directed way, I have a pretty good idea of what's in the test. I can still pass the test and I'm willfully adding holes in.

DP: I want to ask about the question of goals. Is there sufficient clarity about goals, are they being given to you in a clear way that lets you plan goals for the future? Think MPG for future vehicles, etc. Do you know what parameters are wanted on what time frame?

MM: Quite evidently not, since the goals have changed fairly recently in terms of the roadmap. I – politely – call the TSA the three headed monster. What they say they want, what they think they want, and what they actually need, which one do you build? What they need hoping you can convince them, or try to figure out what they think they want, or what they say they want and then duke it out because you go, you said you wanted this even though you want the other thing?

Don Kim: Well, there really are three heads. Part of the problem is communication, then there's the integrated approach between TSA and other parties. At the end of the day, there will be no winners. I personally think that's a mistake. The other problem is that people want to buy EDSs, but they're expensive, fairly complicated (???) it just isn't fair.

DP: TSA can't raise the bar to the point where everyone might fail.

DK: If we were convinced that there'd be some success relative to that raised bar (???)

DP: Last questions, how about measurements? Can they measure paying better for this?

GZ: Yes, sure.

MM: Can you define a measurement, yes? Does it absolutely apply to every airport? No. But it doesn't matter.

??: You have to use the field data, not the lab PFA.

RB: It works unless something major changes that affects your performance.

MM: Some of the borders are hard borders. It's always seemed a little silly to me that if you're at 81% you're OK and at 79% you're not, and statistically those two are indistinguishable. You're at the mercy of statistical variability which makes it much harder. If you're talking about wholesale changes, that you can measure (???) If you're talking about measuring 5% off, I have my doubts.

GZ: Three-four years ago, we wanted to do exactly that. So we would be able to send to you guys without affecting the throughput of the airport. That can be done.

DP: So is there a sense that if we improve incentives, will performance be improved?

??: I guess it comes down to, improving the capabilities is what we're trying to do all the time. But is the improvement sufficient? Is there an acquisition matrix that actually follows that, so that incentive is driven to a higher market share? If that's not done then it won't improve performance. Also, we have a fiduciary responsibility to our shareholders. It does work but it has to be across the whole board.

Moritz Beckmann: Is it possible to put out an incentive on an ongoing basis rather than focus on new acquisitions, a bonus of sorts?

MM: Yes, that will lead to more third party involvement. Will that lead to improvements in the field? Not necessarily. That's from pure experience but it doesn't mean it can't be different in the future.

Speaker: Steve Azevedo

SA: Detection of Explosives Internal to Humans.

(Summary slide)

Internal = implanted, ingested or inserted.

We're going through the different types of technologies. Extending existing techniques is difficult due to physics constraints.

Going through slides of various imaging techniques

SA: What is the threat quantity of explosives? I don't know, that's what I'm looking for.

DAC: However it matters what is possible to put inside the body, and it depends on what the necessary resolution is.

MBS: So this is essential anomaly solution. With the internal structure of the body being so complicated anyway (???)

CR: And not only is it complicated but it varies from person to person. There are 150 different human models in this room alone.

MBS: It's not clear that microwave will work for this.

SA: And you're right, people are all different. If you could do a differential (???) but it's difficult to get a baseline. (???) The problem of course is that you need to look at skin.

(Mid infrared to visible slide)

SA: CT is extremely invasive. You can argue about type, but that's not something the public will want to put up with. Talk about additional modalities: nuclear resonance, electromagnetic induction, cosmic radiation. (References discussion)

(Summary Slide)

MBS: Basic question: What is the assessment of threat severity? Right now there's no TSA that will hear that there's a spec/concern about threat. I'm just wondering, until there's an acknowledgement of need, will we develop the technology?

SA: Do we know the severity of the threat? Does anyone know?

??: What are you going to do at an airport, when you're telling 1% of the public that you think they have something in their body?

MM: How do you test?

CR: Use turkeys, dogs, pigs.

??: There are at least 3 people making transmission X-rays, about 300 units around world. Customers are relatively happy. For comparison, the dose is okay for screening according to FDA. It is a reasonable tech that is in existence today. Their customer is prisons and diamond mines.

RB: In Colombia, they do have low-dose X-ray scanners for people to see if they swallow drugs. I'm not advocating that, but it does happen.

SA: In prisons or as a secondary scanner, it may be a great solution.

Speaker: Ken Jarman

KJ: This team is sponsored DHS S&T Explosives, for ways to evaluate fusion research. From the perspective of underlying math that goes into fusion and complementary tech. Part of that is to get feedback from discussion groups, and I'd love feedback from you on what is needed and how we evaluate. This is on what's a good research in fusion. I'll talk about one example. Similar topics have already been discussed.

(Conclusions slide)

KJ: We need to study a variety of ATR fusion "models". When do we fuse? That has to consider cost. That means different things. What's feasible? You have to consider if I'm fusing with 2 different vendors, how deep can I go? How much better performance will you really get relative to cost? DHS S&T programmatic strategy is needed to evaluate and prioritize concepts for ATR fusion research investments. They must define the task and define standardized test scenarios and data collections for fusion.

We're not trying to solve a question here. We want to raise questions and get feedback. Scope is checked bags at checkpoints. We have things already out there for detection. This doesn't include many other things that can be included. I want to just focus on MM-wave and metal detection as an example.

They are something different than skin on the surface and different physical properties.

KJ: So correlation is an issue. If you don't account for that, you can over/under predict so account for it when you can. Each of these has some ATR, so now we want to do some fusion.

(View: Carl's diagram from yesterday)

Not all boxes are there for every ATR system. Steps can be there for other imaging tech. Broad data that you put into a spectrum. So all processes get a place, so you see there's already operator fusion. And then some decisions. So at what stage do we do it? And then I have to build new future applications. Greater costs the deeper I go.

CC: One system informs the other one so there is feedback.

KJ: I want to talk about all possible systems; this accounts for all.

MM: You don't have a display fusion, do you?

KJ: I naively didn't think about it. That makes sense though.

Pia Dreisetel: You can have more fusion at different steps to through interception.

??: You mention it'll be more costly.

KJ: Yes, most likely.

??: But if you don't have to incorporate a manufacturer's algorithms, it could be cheaper.

KJ: Greater separation can be talked at another time. We need to be cautious about adding too much criteria though. We can make some simple arguments about how you can get better fusion if you have more information. ROC curve for each sensor (???) ROC curves plus correlation (modeled/estimated). We can get deeper into the info of the sensor. As you go into more information, you think you can get better performance, but you have to be careful. Also, you have greater complexity. So you have to balance. (Explains graph)

KJ: Also look at higher programmatic level to make recommendations. How can you guess what will be successful. So why do you think 2 sensors will give you better performance? Can we come up with some rigorous calculations or statement about why 2 will do better than 1. So on an acquisition from DOD, we're talking about research (???).

??: Have you looked at difference between 2 sensors versus 2 features?

KJ: There could be correlation because they could be physically relevant.

We may not have thought about it in terms of features, but it may just come down to classification. I have thought about it.

??: Well, do you want to look at it in a different way?

Sondre Skatter: Should be best possible from a performance POV.

??: Technically yes, but that may not always be the case.

SS: That presumes a few things.

??: You can do better with that information, but there's still more work to do.

??: Theoretically, you can have maximum chance of separation. If you can manage that there are great ways to deal with that.

KJ: It's not about noise, it's more about the amount of data you're extracting and needing to separate.

Speaker: Kevin Johnson

KJ: Not focused on developing a specific system, but understanding systems. We want to avoid spending and not knowing if the function is working. We want to figure out how to intelligently design systems from the beginning.

KJ: Nothing too controversial in the summary. As people pointed out, the enhancement is by no way guaranteed. Even if you got a benefit, it may not be worth the trouble. A solution that works in one problem domain may not work in another problem. It is possible to estimate best case fused system potential performance against through an understanding of the performance characteristics of component sensor system.

(Target Analytes Slide)

KJ: (Explanation of chemical domain and analysis)

The reason it makes detection complicated, it's an operational definition. There are different groups though so it makes detection through one sensor somewhat difficult. We want to leverage unique sources of information than possible single sensors. What we want to accomplish from the single system is a reduction of false positives. We want to enhance sensitivity. And we want to detect all threats we're interested in.

What do you mean by unique sources of information? Of course, at most basic level, it's sensors that detect different analytes. One step down, we have sensors that detect the same analytes but with differences. Finally we have sensors that weren't designed for explosives specifically but that help through the assessment. (Current landscape slide: explanation through images)

KJ: Wide range of devices. You can group devices into 4 categories: spectrometrics, spectroscopic, chemical adsorption, and chemical reactivity, which is the catch-all. You cans see that there's a wide range of handles that these detectors hang onto, so that implies that we may be able to leverage and get a better solution. Also when you're thinking about combining output to different sensors, you can't always get a usable data. Other techniques give different outputs. In any case the combo generates a measurement space of each component.

(Multisensor System Design)

KJ: What are the requirements? What are the logical concerns, what sensors will be used, how will output design be derived from data? What about cost? Do I want to augment technologies, develop new sensors, use fused data? Also what algorithms will use? As we've illustrated, there's an infinite number of ways you can organize it. This is where I think viewing the system with this approach has a lot of merit. Detection of an analyte is a decision that is made on the basis of measured sensor data. Decision theory provides a framework for an optimal solution. In the framework, the ability of a sensor to detect an analyte rests on the distribution of sensor responses observed when the analyte is present and when it's not. You can visualize adding and removing sensors, within algorithms, all of it is a transformation that will have an impact on performance. Looking at it with this framework lets us evaluate the changes.

(Example slide of overlaps and performance)

KJ: One nice thing about this approach is that there are analogs that are figures of merit (???)

So putting it all together, how are we going to evaluate the system? As stated it can be visualized as a series of measurements made on same sample. System has a characteristic measurement space that contains every possible collection of sensor responses that the system can generate. Using this as design, not as a fieldable solution (???) But it gives you a useful metric benchmark that you can judge other system performance based on.

(Examples of the problem)

KJ: All I know is probability of detection and false alarm. Through this you can see an improvement over using individual system. But say we have 8 such sensors. One might consider is this low performing system giving me a benefit, and it's not.

In other words, adding more sensors doesn't really help, but a fused system is still better than individual sensors. It seems as though it's not worth it to go beyond simple fusion.

(More explanation of graphs and sensors)

KJ: When you turn raw sensor output into decision, you're removing some information but you're keeping the most important. You're always theoretically going to think on raw sensor space. That's not always going to be the case though.

One last example is a common theme is partially selective arrays. In our framework we can get into some of their problems. We can simultaneously detect simultaneous analytes. It's hard to boil that down to a single number.

To develop a space optimal system, you see an interesting phenomenon. The detection problems are very difficult, the ray become more specific, and each sensor detects a new analyte.

??: Is this just a tradeoff?

KJ: It depends. It requires a higher signal of noise, and you gain false positives.

??: Couldn't you have a classifier for each of them?

KJ: Sure, there are infinite permutations. The view I am taking is that we could construct this, and we could characterize this the same way. It's the potential, but I think it's pretty useful itself. It definitely tells you whether or not your system has it.

John Beaty: In chemical systems, all of these modalities will have a different type. If you don't add those in those two elements, it's not practical. I like the construct, but I want you to add dimensions

Speaker: Deniz Erdogmus

DE: I am presenting research that my students and I are working on. Since the second slide is results, we have not heard more of a prediction, and we have been looking into the problem of designing robust fusion, as one of the sensors is broken. What we mean is that the sensor has calibration, and its similar to what is presented as a similar talk, and we are continuing to use this, assuming the sensors are using this, we cannot use techniques from other likes. The best thing we can do equal to other values t stops evaluated at that point. It could be stuck at one particular position. It could be anything. A sensor that is not operating as it should is broken. Can we somehow salvage information from it using other sensors? They are independent. We can replace it or fix it. We can design a robust fusion rule, when some sensors are broken. There are no sensors compared, and we are out of risk for this. We can do it that way. I have more slides at the end about this.

These are three aspects that I mentioned, we tried to learn failed sensitivities online, but after the function was detected, if the sensors correlate decisions, we haven't looked into that case, and what I will show next is robust fusion. How we can get sensor figures. This picture was shown in the other presentation, we cannot detect from a local sensor. We have all these other sensors that we have to be working if that is what they are intended to do. If you look at this graph (Slide 4) (???)

They only care about this sensor, but if it breaks, it depends on the other sensors, so the question is can we use the decision to recalibrate this recycled show that we are remodeling to get additional information.

We have an idea, and moved onto robust fusion. In classical fusions, we will have these two turns. According to minimal decisions here (???) We don't know what these start at. For this part there is failure, there is proper equations, the final decision ruled that robust failures, risk classification, plus correction term in case 1 or more sensors fail. Here are some other results. On the horizontal axis here, you can see that they failed at that rate (Slide 6) each of these curves show if this many numbers of sensors failed out of 5 sensors, then it would cause all of this added risk with respect to robust fusion. If the failure rate is so high, compared to the desire mode, although we can do robust fusion, we grow risk faster.

(Slide 7 explanation) This is the number of sensors; we didn't fail this number of sensors and we took this number from this side here, and you can see the rate.

(Slide 8) What we see here is that if the rate is really low or really high, the gains of robust fusion are actually more. You would like to use robust fusion to be safe in the case that the sensors failed. If a sensor breaks the risk comes from running too fast, and look into taking two sensor profiles than we take a statistical design based on the expected behavior of our model and it has been recognized as a good comparison for how we expect them to behave. From a statistical test, we can subtract a statistic (???). Here what we show is that the sensors are about at 90% presented in this simulation. This assumes that the threat will (???) if the probability of detection changes, then it's really hard to detect if the sensor failed. That is because there are much more innocent travelers passing by.

??: How do you detect how it fails?

DE: If it fails it takes a new detection form that (???), and based on the prior failures, the graphical model has changed.

Speaker: Venkatesh Saligrama

VS: Anomaly detection and video analytics, this is one of our transition tasks we are working with the other transition team; we have a known special detection, we can achieve 100% detection, with an actual replaced detection. On the other side, with similar questions: so Logan Airport gets a lot of data every day, and gets back up, and so the question really arises that if something bad happens in the past, how are you going to manage that. So this motivates how do you try to reduce the amount of data? And yet, still be able to search for this last compression. By the manager from Logan Airport, these are known special detection problems; there are known special detection problems. You are looking for something unusual. They are trying to understand if there are anomalies that there are things going on in forensic searches, in very large areas. Rhode Island researchers had taken back their sensors; that is something that they talked about last time. They talk about how they can get applied for photo sensors.

So I don't have to explain why video analytics are important. So I want to help this guy do a good job. The two main points that I want to mention are, in the forensic context, there is also a storage in dimension; they just don't have enough storage capacity. Why do I say that the kind of techniques that are here are different? It is because most of what we do has a kind of arrangement. Once those features are extracted (???) If there are other instances of these problems, similar types, it applies to different domains. So as I mentioned here as known special factments (???), this is joined project between DHS Cleveland and ALERT. We want to get as good detection as possible, with low false alarm rates. Challenges are obviously ambient noise. Our approach is primarily to calibrate certain features, and they are illustrated here, with blue lines, on each object there are 500 lines.

MBS: This data was taken at Cleveland Airport with real passengers, the passengers went against the flow, in a real airport environment.

VS: As Michael just mentioned, this is a real environment. So some of the performance of the research, I exactly know what they look like, but in this particular context, including the video process, they use both cameras simultaneously. There are 700+ people passing through, and we had a score of 100% detection.

MBS: This problem was driven by the need of the airport. Not an academic problem. Cleveland Airport said that this would save them so much money every year, and time and effort, because they can't take a chance that someone who breaches their exit zone is not a threat.

??: How do you define the detection?

VS: Manually.

??: What kind of guidelines?

VS: This is a counter-flow (???)

MM: For how long?

RM: He has to cross the barrier.

VS: We have to switch topics, this is a different problem; there is a huge data deluge. There are (???) of bytes of data; the second problem is computational overload.

You really do not try to process it. Our data is not precious. Somebody puts in a key word. We are supposed to go in and look through this data. We tried to figure out where the activity takes place.

The number of such events (???) should not be a problem with the amount of data.

The second point is that you want to equal storage. So are the two choices. For instance there is a new turn. The query algorithm has to go in and look at these kinds of things,

What is our operation, we extract some low level features; we break up the video into chunks. We get various kinds of features. That doesn't really help us reduce the data. Then what we have is a hash table, and then take all of these local features, they have multiple instances of times. If hashes these into a half tape. Then a query is generated; it is then in the hash table.

Results: Then there are fancy dynamic algorithms. Time activation of how we are able to do this. This is with the huge storage view. We record a lot of data, and then push the data for that. Here are some of the results as you can see in the index size; you get the experience in many different contexts. The red one is the bad algorithm, which is a fancy thing these days; that is not appropriate, this set up thinks they do much better. There are competing algorithms.

In order to extract the ID, we could extract just a few. You can extract just an hour from several hours of video.

I guess they did mention the stables here. This is another data set; is in video even if you are looking for specific patterns.

MBS: Can you talk about the fluorescent detection?

(Video #2)

VS: What I want to show is this is detection on unknown factors. On the floor detection, but it also detects someone running. It doesn't know what is anomalous. But it observed the running person, but it is only because it doesn't happen in the data usually.

??: The scooter is normal?

VS: Yes, because it happens more often. At URI, Bill has been coming up with different components to come up with bulk imaging techniques. What you see on each column, each row means a different explosive. If you think of these are different sensors, they behave differently. They have variable gains.

Different photo sensors behave differently as well. They have a lot of influence from different sensors. That can be used to detect patterns of explosives. His vision is that you could do this for fewer exposes. Looking for known baggage (???) In that airspace you would be able to detect it. The following problem, the explosives come in different mixtures, they have been talking about unknown factors. Unknown mixtures of composition - You are going to get a combo of these. You're not going to get one of the pure things. Underlined points on stats, other solutions actually, we extended the different technologies, in a situation with unknown positions, so it is an extension with existing literatures. Here is something I want to leave you with – this image, you can think of this mixture of compositions. The second column is a different experiment. This is a different algorithm. We took his data, and got different mixtures. And got (???)

You can certainly develop an anomaly detector, they can look into certain looks, but they start looking at certain bags, that said, they/Rich Radke published a lot of their work.

John Beaty: You can apply a lot of sensors with low distance radar with counter flow. You can look at it from those terms; this is a trigger; which basically say as you can track that figure. We want to track through an extensive area of the airport. So that we in fact have achieved different figures of merit.

Speaker: Chris Alvino

CA: This is about inserting one question on Carl's slide: How to incorporate 3rd parties into ATR. We don't have nearly the false positive rates that Venkatesh does, but we want effective efficient development, essentially, in house we have different problems that are hard, and in general, we give them to a university collaborator; that is the great collaboration between us and Tufts. For the time being, why ATR is hard, Alex convinced us yesterday, here is a list of the problems. I will stay short. The way how Venkatesh is not academically motivated, but industry motivated. But that is how we are trying to come up with the same answers from the same problems. We are deciding how this gets integrated into our algorithm. The proximity *t* helps, but we can schedule a meeting within a day or two.

Speaker: Brian Tracey

BT: Thank you Chris. I had the chance to work with Eric Miller, so the conclusion slide is first. The technical conclusion is that we have done some work doing backscatter trying to get images doing false alarm rates. In particular there are de-noising things that help with this a lot. And there is a careful problem defined from AS&E. Both sides have chipped in time and money, which become more (???)

CC: What was the problem?

BT: They would like us to see if we could reduce the false alarm rates. But the next time, they set out 4 problems where they gave us data, and said can you give us an answer, and there was de-noising as another one.

This is the outline: two different problems. The thing that I want to point out here, is can we get a proof of concept, is that preprocessing is important. Low scale transmission data, but for this person you would see a black and white silhouette. In terms of these lung false alarms, you see these tissues.

What we set up a processing chain, but we kept an edge detector, and we took the tx image, and we won't talk about this, but of course that second part of this, on the one side I expected going in I would have to get segmented lung volumes. Right here, you can see where the images have fully registered. We don't really see it here.

BT: Can estimate BMI from the backscatter only, which helps a little bit but not too much. The transmission really does help you.

Improved de-noising for X-ray backscatter (XBS) (???)

We find a local patch that we're going to choose to characterize the neighborhood of this pixel.

(NLM Improvement slide)

We take a bunch of different patch combination weights and we de-noise using those. In the past people have looked in having a term for this weighting. We said we also want edges to be locally smooth. So the weak edge problem is what we're trying to get at here. This is kind of more directly relevant to the XBS.

CC: What happens with a rare event?

BT: These edge patches here are kind of rare events. There are only a few rare patches.

CA: Worst case it doesn't de-noise it at all, right?

BT: Correct.

(XBS example 1 slide)

MBS: Would this be useful in a CT image?

BT: It's a general de-noising image, it's been used in the general imaging literature so I'm sure it could be used for medical CT. It's a candidate.

SS: Can you comment a little more on what it means to be general enough to avoid IP concerns?

BT: It's kind of AS&E's determination (???) what they will publish.

CA: Also SSI concerns.

DP: Was it discussed in advance whether Tufts owned it or AS&E owned it?

CA: I think Tufts is putting it into the public domain for our applications and we're cherry picking it and can use it as we like. I don't really know the patent attorney agreement but I'm sure there is one.

MBS: The gift is actually a membership agreement within the ALERT consortium. One of the issues is that effectively if a patent is filed, you can file jointly between you, but the work itself is non-proprietary and should be presentable to a general audience. That's the difference between a gift and a contractual agreement. It's a win-win and it works beautifully.

JB: That document is actually quite extensive and talks about IP and a host of ways of return back to the company. You guys are navigating a specific agreement when you worked out. It's great when companies dictate their issues that let us work out.

MM: You don't have SSI problem, but what about export control issues? How do you manage them? One of the things about our data is that it's under export control so it limits collaboration.

CA: We try to pitch problems that are general enough that we don't even have to often give SSI data or data that is deemed export control-worthy.

CC: What are your specific constraints at Morpho?

MM: SSI and export are two different rules. Export is limited by things like nationally and green card status. SSI is limited in different ways. You can't theoretically work with citizens of different nationality without an export license.

CC: Who is putting that restriction on you?

MM: Department of Congress.

MBS: I think art of the issue is what is export control? The image in the machine may not be export control, but the details of how it was generated may be export control.

CA: We're not just giving images out of course; we have some controls they go through before they go out. These are not transmission images in the sense of the dose that goes into medical transmission images.

??: So what do you actually call this? Some call it forward scattering, some call it limited transmission imaging. The distinction is (???) it still doesn't have all the characteristics of transmission. That's why it really does need another name.

Speaker: Jean-Claude Guilpin

JCG: European Civil Aviation Conference (ECAC) ECAC/EU standards are identical for security equipment. ECAC is bigger than the EU, 44 countries as opposed to 27 EU countries. It's a regional organization under the ICAO umbrella. EU regulates civil aviation since 2002; ECAC establishes recommendations since 1955 with invited nations as well like US, Israel, Canada.

We always keep the raw data because we don't want the manufacturers to play with it. If the manufacturers develop new software (???) but we don't change anything on the machine. A type of equipment is evaluated.

MBS: Do you have comparison of whether a manufacturer's device passed a test here and failed it in the EU, or vice versa? Do you have any comparisons as to the difference between the two test protocols?

JCG: No, because the standards are not exactly the same, they are testing on different criteria.

MM: Say you've collected 10K bags during the original test. What happens if the data gets corrupted?

JCG: We can throw some samples out in that eventually, rescan, etc.

??: Can you comment on the type B+? Was that on the table or is that just an experiment? (???)

JCG: (???) There is no standard on this.

??: Do all test centers have the same capabilities, or do some have capabilities the others don't?

JCG: We have 6 test centers, but not all of them are doing all of the testing. For EDS for instance, one center will hold the bags and then dispatch the bags to the other test centers in order to take the same bags through. We try to keep a consistency in the testing insofar as it is possible. The next step is for the technical task force to develop new testing methods. The EU is willing now to look at this type of testing metals to endorse it as the EU level. Also this is not just aviation security; it is a wider homeland security effort. This will certainly go to the direction to establish EU standardization models and body. (I think).

For the manufacturers there is certainly a benefit to us, because nobody else could have had the capability to test so many machines. The process may be a little slow from the manufacturer point of view, but compared to previous methods, it is much improved.

MM: Suppose a member country wants to do selective screening. Is there a method for them to do that within the CEP or CEM, or is that really future?

JCG: If they want special software, it depends.

MM: Are they obligated as members?

JCG: It's EU standards (???) you can say, I want to have additional on this, these are just minimum standards. Maybe the country can pass more.

MM: But they might fail the false alarm rate.

JCG: If we don't know the software, we can't say anything. But the ECAC is not compulsory; it's just a recommendation and an endorsement. The member countries sign an agreement to protect and consider the results, not necessarily endorse them. Your country would test those categories independently.

DL: ECAC tests it and it passes ECAC, and then the country says it has to pass those substances.

Speaker: Jennifer Dy

JD: I work with machine earning and data mining. Most of the time I develop medical imaging (???) The goal here was to segment dermis from epidermis. This is similar to your imaging. I also work with COPD imaging with Brigham and Women's Hospital In particular today I picked the topic of crowd sourcing. It's relevant because you work with different sensors which are like different sources for CS.

JD: So this has broad applicability in collaboration with Siemens in the Machine Learning group.

(Conclusions)

JD: I suspect this is more abstract. Motivation: we started working on this problem because we were looking at medical data. When doctors look at data they don't often agree. They have different diagnosis for similar things. How do you build a learning algorithm that can help the opinion (???) And how do we evaluate this diagnosis. When you collect training data it's very expensive to have labeling of the data done. Now there's software called Amazing Mechanical Turk that can do this. This isn't perfect but it's becoming available now in research. With images, labeling images is also very tedious. So how do you learn from them? It demands new way to learn.

(Challenges slide)

JD: Multiple and unreliable annotators. Another key challenge is that different annotators have different performance. We need to take advantage of that. Also quality of the data, some is clear and some is poor. Our model has to adjust to that.

JD: A standard classifier you have input *x* and output. In many cases, especially in this scenario, you don't know the ground truth. What you have is the inexpensive multiple annotators though to provide those labels. Their accuracy will depend on annotators. This is the challenge and model.

(Math Slide: Explanation of annotators)

JD: Right now, we have a model that can learn from multiple experts, annotators, sources.

JD: So here are the results. If you combine experts by using majority vote, you do better. Our model does work better according to our results.

Because we can detect expertise of annotators, we can do active learning where we only pick the experts that benefit our learning model best. The new paradigm in the crowd sourcing context, we have to intelligently choose which instance to be labeled and decide which annotators to query from. From our results, the ones using most informative is better. Our model will also allow us to learn which are the bad annotators. First we chose adversaries and then we flipped it. Our model shows that the lower the bar the better. The black is our method, we are better with adversaries chosen.

So you can learn from multiple annotators/sources. Our model takes into account the quality of annotation which varies depending on data. We can evaluate reliability of annotators. We also developed an active learning setting approach.

??: Someone who is consistently different than others may be right. Would they be labeled as bad though?

JD: It has to be all consistent. We've built in a classifier.

??: The annotator has a confidence?

JD: Yes we can add that in.

CC: What's the application to explosive detection? Thoughts?

JD: for this it's on the classifier side. So if you have multiple sensors, it has to be where the opinion is always different, then we can use this. If you want to know the ground truth this can work.

??: If you were to build a big fused system and a sensor were to fail, this could fix that problem.

??: Can you label reasons?

JD: Yes you can, but it's tedious.

??: Can you comment on differences using (???) to predict ground truth. Difference?

JD: This one is more general. That model just combines experts, this one builds classifiers too.

Speaker: Raymond Fu

RF: Low end analysis framework, this brings low end analysis framework. This is the intent theory, somehow demonstrates what this means.

Many presenters presented several ways that this happens. As long as we have data, so that I can actually use framework, we will be able to make progress.

How can we use cross-modality data? I want to start at Knowledge Transfer. This is very common, like humans, we can use transfer knowledge to learn English, or French, or any other language, because we understand dictionaries and translation. We have a common assumption, where there is consistency, there is a trinity that in insufficient, cannot always be satisfied. To label the data, and to use data, we only want to pick very useful data models.

Here we look at one direction of low rank analytics. The idea is that if we turn this into a subspace, you throw away a lot of data relation. The learning is very robust. You only need look at the figures. We only assume that some subsets are useful. If you want to do anomaly detection, you can use two different sensors. The data here, red and blue, are target domain, we can separate this data after. Here are the results. We have been trying this low rank transfer subspace learning, for many security reasons. For this classification, and the second, they are both considered manifold learning. Manifold with Noise Effect captures the subtle changes, and subtle uniqueness. How can we mitigate that noise? This is the example, where you can see the noise coming out. The colors come out. If we drew the noise out on these representations, you get these. The idea is that the Robust Manifold Low Rank Recovery is large scale. In this sense, the noise will be ruled out by extra strength. You can get other figures. When I reduce this to 2D, you get this result. If I have noise, and you get this result, but if you concentrate on this low rank consumption, you get a separate result.

The beauty of this framework is that it is sensitive to other changes. But the data structure is very different. The noise is dominant. 2D space in the video, you get these reactions, and they are practical reaction. You can easily get those different actions. We have been running on different detections, we also introduced those other predictions. If someone has the bag in the airport, then it depends on the action and the activity. And we somehow skipped those results as well. I have a couple more slides on this framework, that can motivate people in this field, what can we do? Traditionally we can get data on these learnings; but the competition is verified. Especially because the data goes up significantly (???) Traditional computing cannot be set by another machine.

It's very different to the beginning. The approximation recovery, they will get the representation of the data, when we bypass the global data we get other examples of data sets. We found out that if we cut the manifolding we can achieve even better recovery results. But we speed up significantly. The rest is about matching. Somehow those data are matched with each other. Here is the idea. Even if they have others that have noise, we can only learn the key things that have results. I can take multiple modalities. This can be potentially applied to Tufts.

Speaker: Carl Crawford

CC: Did we achieve our goal for ADSA08? Are we getting better detection? Are we getting better involvement?

My own takeaways (slide 3).

??: For example, the liquid detection in Europe they have a list of threats. At least now we know what threat we need to distinguish. We went in the bag with the laptop or without the laptop.

GZ: We are not totally opposite of what the Europeans are doing.

CC: When I say additional, I don't mean that we can't live with what we have. What is beyond the future. Are there rules?

??: We are going to use physical features, like shape, but to say that it's a square or circle, there are reasonable outlines.

??: You lose your ability to do something if you give too many requirements.

CC: That's a good point.

??: That is part of the TSA spec.

??: If this is a threat, and you always present it as much, then you need some guidance. It's C shaped as opposed to L shaped.

??: Don't disclose it to the public.

CC: What I heard is, how do you get additional information in the US?

??: Sometimes it's not supplied but sometimes you ask. You pick up the phone and you call TSA and you ask. It is just going to explode and they answer you.

CC: Is it an important feature? But the goal of this ADSA is to talk about 3rd party involvement.

That's what I have been hearing.

John Bush: I think that there is a whole body looking at this, and studying what these requirements are. That stuff can't be disclosed in a non-SSI environment. But people who want to know it can get answers.

RB: Unless you get on a vendor list, it's hard to get information.

John Bush: How does the 3rd party source get information?

??: On the US side, we were working with a US partner, but they couldn't pass info to us. Even though we were SSI (???)

??: They couldn't give you the code?

??: No.

??: They have to build a good product, but we have to get something that's saying; how do we get all of the electric toothbrush? Otherwise, we are going to be running around.

As far as I remembered, they are not involved. They are an example.

CC: We now have funding to do T04, what is your advice to ALERT? How do we do it for ATR?

RB: I think you can guide what realistic bags and threats are, and build a database. You may have the vendors come again. What would you like to add?

??: You can't just give a feature.

MBS: If you take the analogy for what we did for segmentation. You define a problem, then we issue a call for proposals to the general 3rd party community, and they told us how they would attack this problem. It may be that that is what we want to do here.

John Beaty: We asked them to help us segment this data, and then tell us what you did. We got 12 or 15 responses.

??: In terms of idea, vendors are going to be competing for this. They will tell you what is hard.

The universities will compete for this.

John Beaty: My concern is that this whole area is highly influenced by (???)

CC: What I just heard you say was this slide here. We should run our own certification test.

DAC: I mean, I think you should collect a data set, and then release it to everyone.

CC: I believe that we have to erase that. Everything is tied together.

??: Include segmentation also.

CC: I think that's too hard.

Luc Perron: When we try new detection solutions, we do need to scan items of our own. We need to find a ground truth, to find limitations of the system. If you have a test set that you provide, you make new images, with specific characteristic, you are in trouble.

CC: The word use up here said what you just said.

LP: This could work for conventional X-rays. If you are starting to extend this to CTs, but there are some very big differences. But now there are some major differences from one system to another. They are very different.

DAC: What you say use a common scanner, we could use a virtual scanner.

CC: That is a goal to do that. We need to develop that.

??: You can get a lot of data from that.

CC: You have to tell me what you need.

??: There was a comment made about how (???) works. All the way back from raw data from machines (???) For AIT, for CT. They are very different from machine to machine. I think that's less useful. Then you are actually able to have a reason to improve the system and a lower PFA. That can be done in the (???) across the board. I think having a common scanner with the lowest common denominator.

CC: A year from now we are going to have people present the results.

Are you going to allow them to show their results?

??: Yes. I would. I can't comment on that or answer this.

CC: If they are SSI they can't.

??: If they are true detect results we can present those.

John Beaty: This is a good idea.

CC: There are different characteristics.

CC: You have an ATR that is trained on threats. There is no way you are going to do that. The results are classified. How do we step around this?

??: The vendors are willing to provide their data. Their results are on real threats. It is SSI.

Laura Parker: I would not be able to run that amount of paperwork.

CC: There are issues behind the scene.

??: The modeling capability that you are building can be used as part of this that is used in the discussion with Tufts that AS&E is defining the problem. They felt comfortable with what they are working with. They have a problem of what is tractable for modeling, and using them with materials with what you can use with your Imatron scanner. You can buy these.

I actually think that would be quite interesting.

CC: The “classified” document for the project, is a whole document. This is part of the learning process. All of the testing will be done virtually. Questions would be answered virtually. We want to solve the problem.

Hearing this over the last two days, this is the only way we can go out and get this done

MBS: We have gotten some good risk, we are going to go back and forth these comments and ideas into a viable plan. We will reach out to the community and get feedback which will go in the March timeframe. I think they will be in time. I wan to thank the audience for participating. This is a very energizing conference for me. We are getting the participation from mainstream, academic, and government.

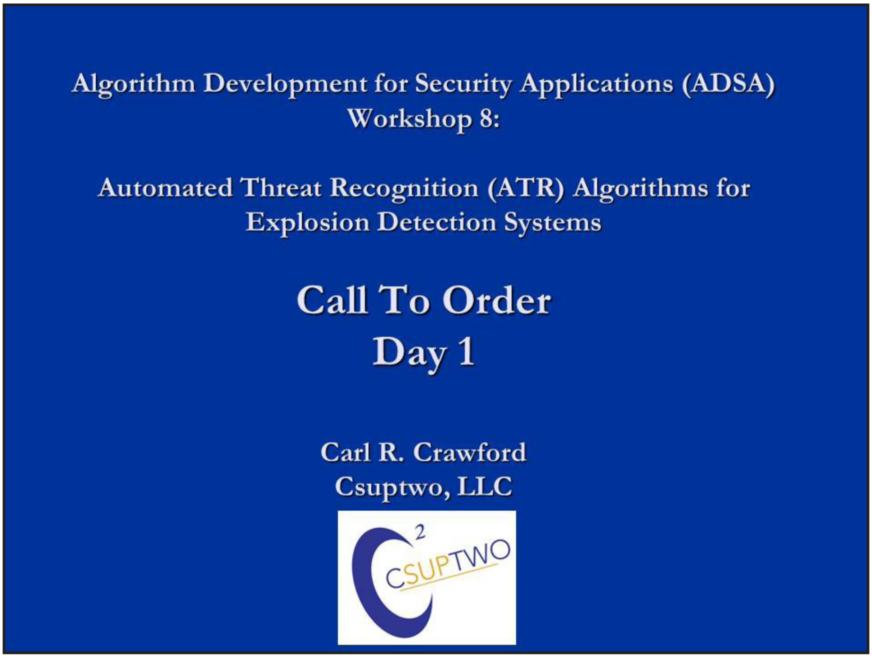
LP: Quickly, thanks to Carl and Michael, and the ALERT team, I can say that ADSA08 is very good. I can see a big difference for everyone.

16. Appendix: Presentations

This section contains the slides presented by speakers at the workshop. The slides appear in the order that talks were given as shown on the agenda. Some of the presentation slides have been redacted to ensure their suitability for public distribution.

PDF versions of selected presentations can be found at the following link:
https://myfiles.neu.edu/groups/ALERT/strategic_studies/ADSA08_Presentations/

16.1 Carl Crawford: Call to Order



Algorithm Development for Security Applications (ADSA)
Workshop 8:

Automated Threat Recognition (ATR) Algorithms for
Explosion Detection Systems

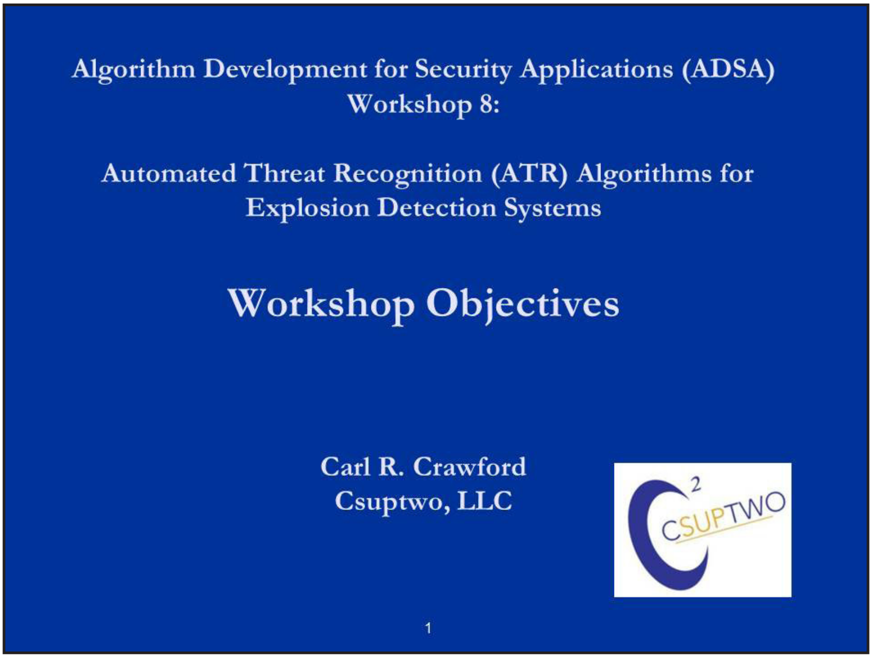
Call To Order
Day 1

Carl R. Crawford
Csuptwo, LLC

Rule #1 – Open Discussions

- This is a workshop
- Conversation and questions expected at all times, especially during presentations
- Moderator responsible for keeping discussions focused
- Not grip-and-grin

16.2 Carl Crawford: Workshop Objectives

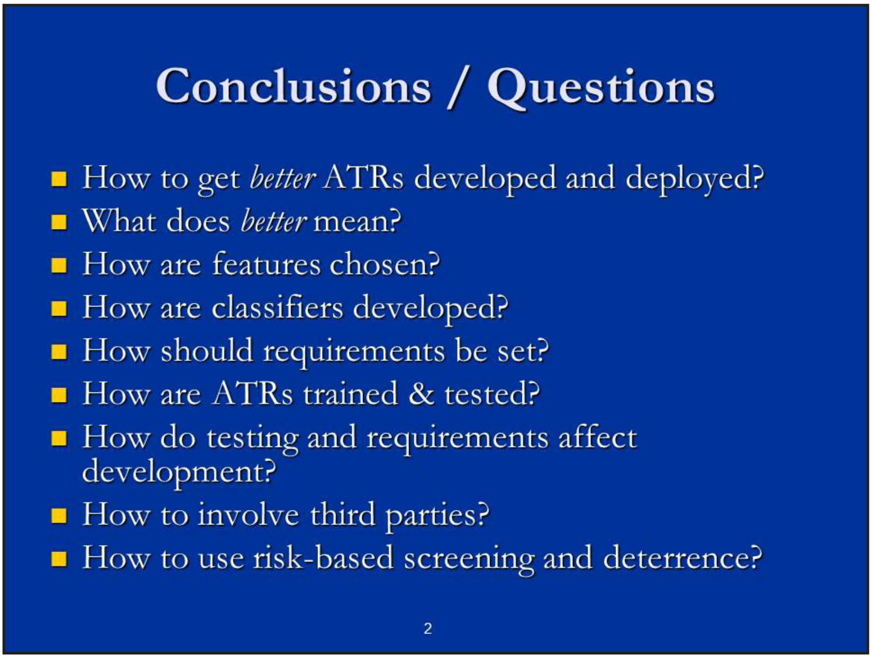


Algorithm Development for Security Applications (ADSA)
Workshop 8:

Automated Threat Recognition (ATR) Algorithms for
Explosion Detection Systems

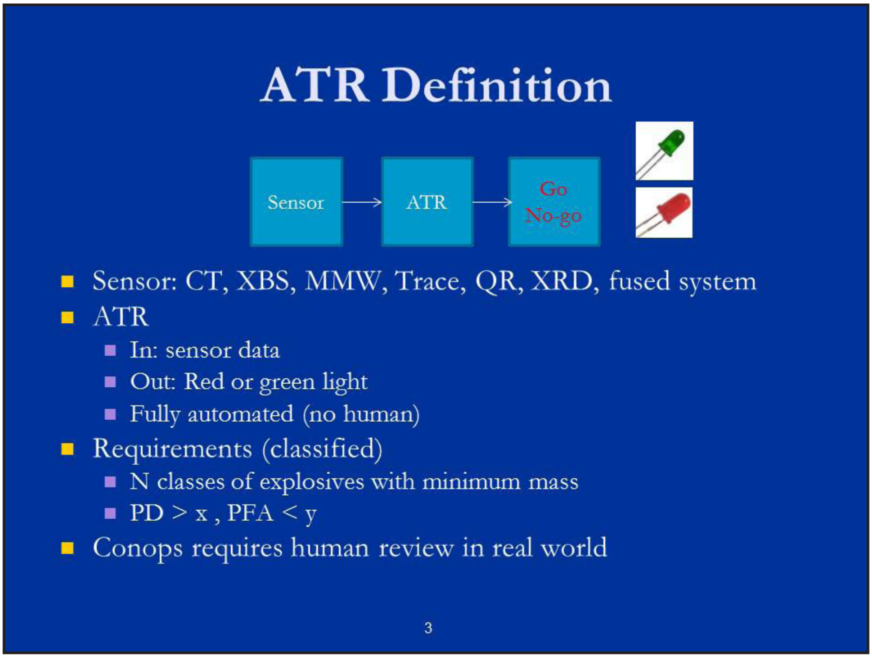
Workshop Objectives

Carl R. Crawford
Csuptwo, LLC



Conclusions / Questions

- How to get *better* ATRs developed and deployed?
- What does *better* mean?
- How are features chosen?
- How are classifiers developed?
- How should requirements be set?
- How are ATRs trained & tested?
- How do testing and requirements affect development?
- How to involve third parties?
- How to use risk-based screening and deterrence?



ATR Definition

Sensor → ATR → Go / No-go

- Sensor: CT, XBS, MMW, Trace, QR, XRD, fused system
- ATR
  - In: sensor data
  - Out: Red or green light
  - Fully automated (no human)
- Requirements (classified)
  - N classes of explosives with minimum mass
  - PD > x , PFA < y
- Conops requires human review in real world

All images from the Internet.
Pictures not to scale.

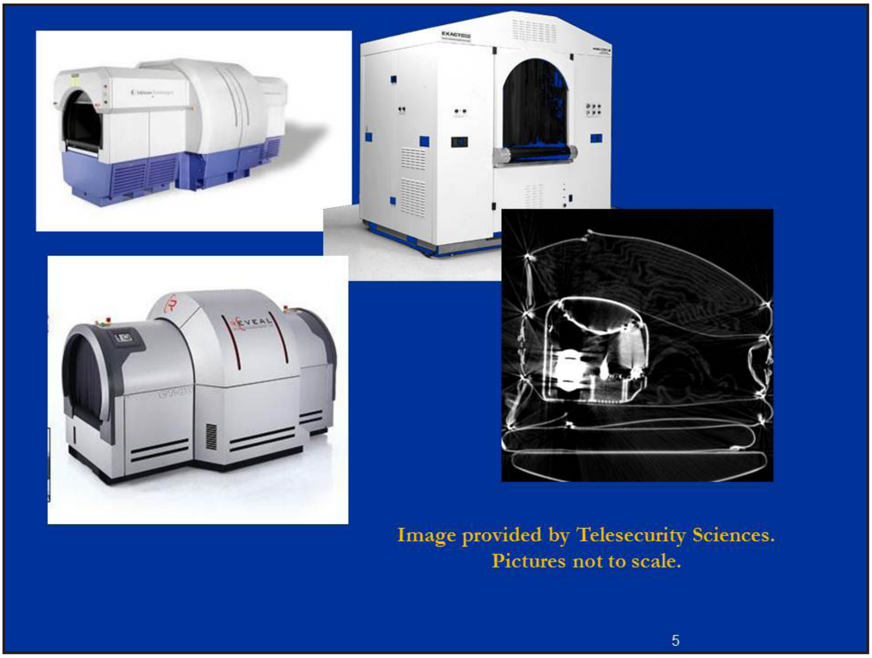
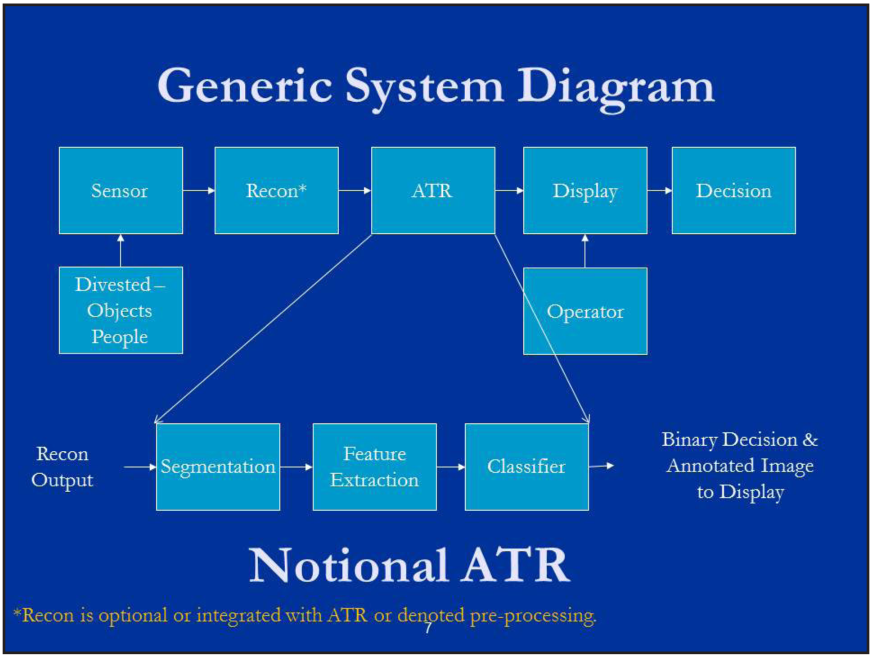


Image provided by Telesecurity Sciences.
Pictures not to scale.

Pictures not to scale.



Generic System Diagram

Sensor → Recon* → ATR → Display → Decision

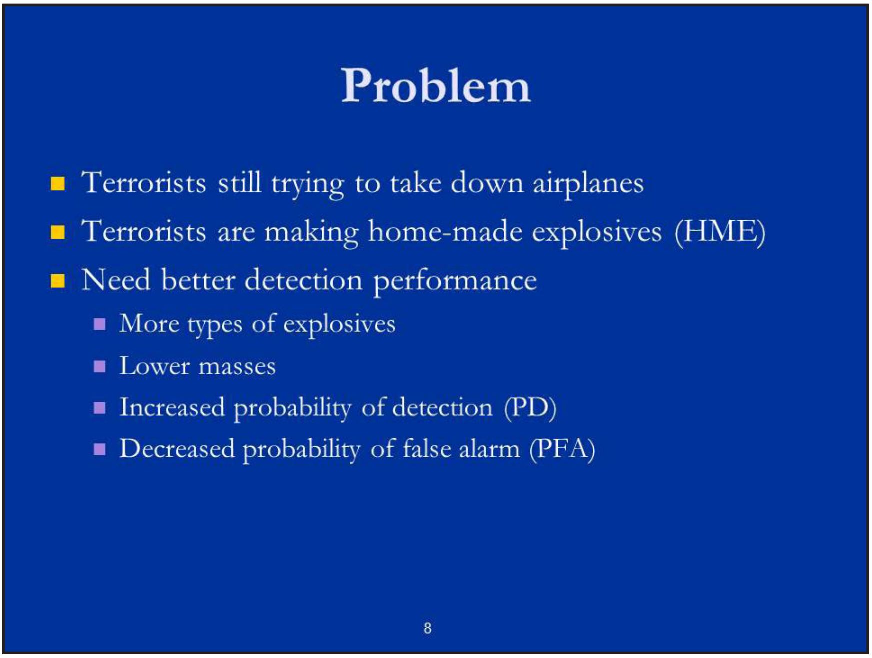
Divested – Objects People → Sensor

Operator → Display

Recon Output → Segmentation → Feature Extraction → Classifier → Binary Decision & Annotated Image to Display

Notional ATR

*Recon is optional or integrated with ATR or denoted pre-processing.



Problem

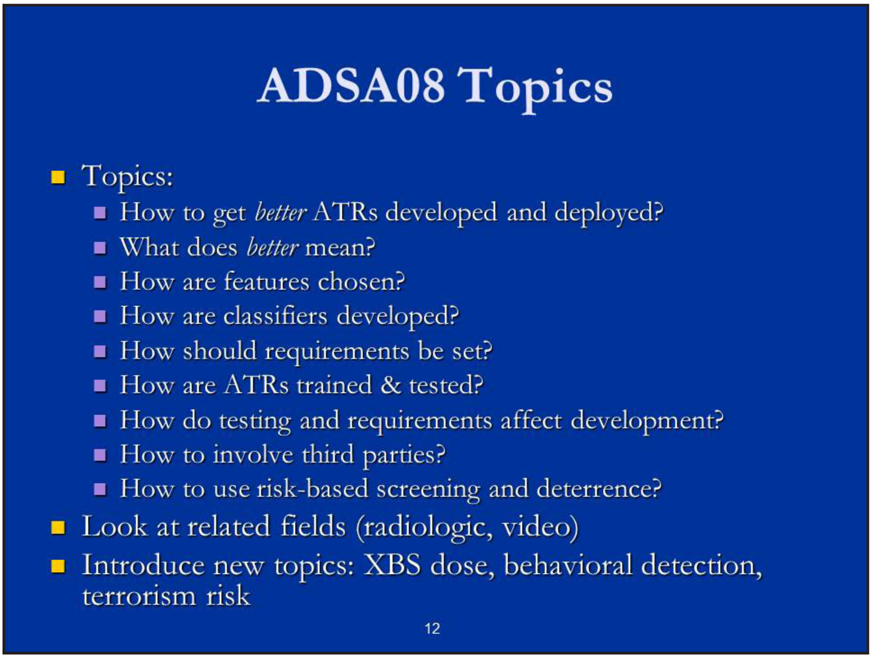
- Terrorists still trying to take down airplanes
- Terrorists are making home-made explosives (HME)
- Need better detection performance
  - More types of explosives
  - Lower masses
  - Increased probability of detection (PD)
  - Decreased probability of false alarm (PFA)

Bin Laden Dead, But …

His Followers Are Still There

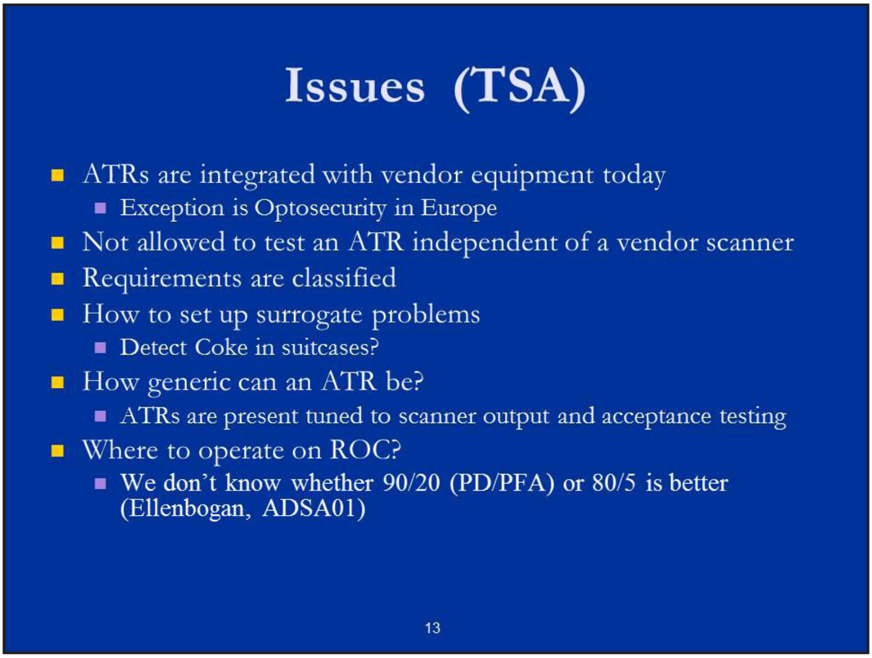
- Plot to bomb airliner foiled, officials say
- A nonmetallic explosive device was recovered that had similarities to the one used in the failed attempt to bomb a Detroit-bound jet in 2009, a U.S. official said.

Man arrested after plotting Federal Reserve bomb, authorities say



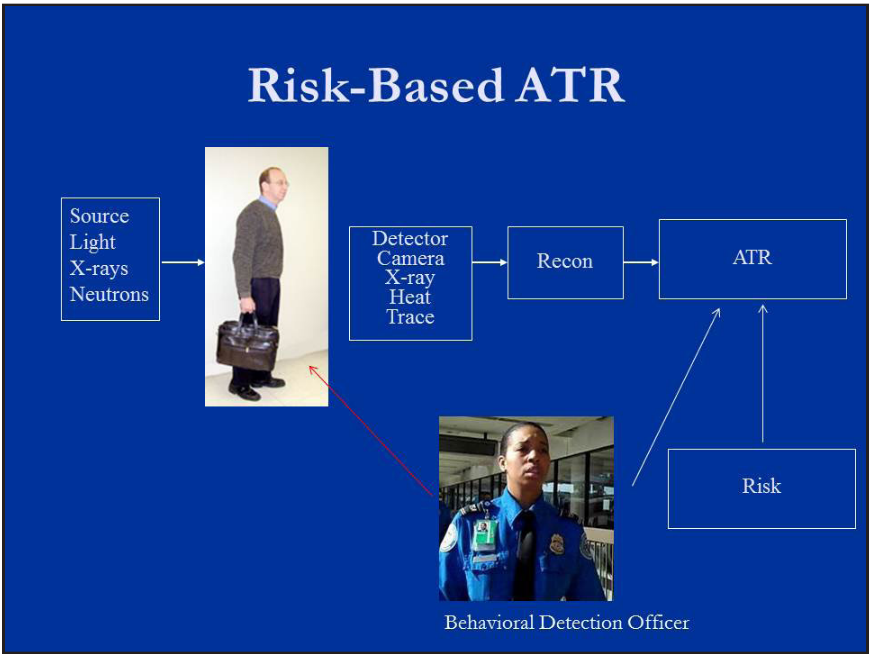
ADSA08 Topics

- Topics:
  - How to get *better* ATRs developed and deployed?
  - What does *better* mean?
  - How are features chosen?
  - How are classifiers developed?
  - How should requirements be set?
  - How are ATRs trained & tested?
  - How do testing and requirements affect development?
  - How to involve third parties?
  - How to use risk-based screening and deterrence?
- Look at related fields (radiologic, video)
- Introduce new topics: XBS dose, behavioral detection, terrorism risk



Issues (TSA)

- ATRs are integrated with vendor equipment today
  - Exception is Optosecurity in Europe
- Not allowed to test an ATR independent of a vendor scanner
- Requirements are classified
- How to set up surrogate problems
  - Detect Coke in suitcases?
- How generic can an ATR be?
  - ATRs are present tuned to scanner output and acceptance testing
- Where to operate on ROC?
  - We don’t know whether 90/20 (PD/PFA) or 80/5 is better (Ellenbogan, ADSA01)



Risk-Based ATR

Source
Light
X-rays
Neutrons

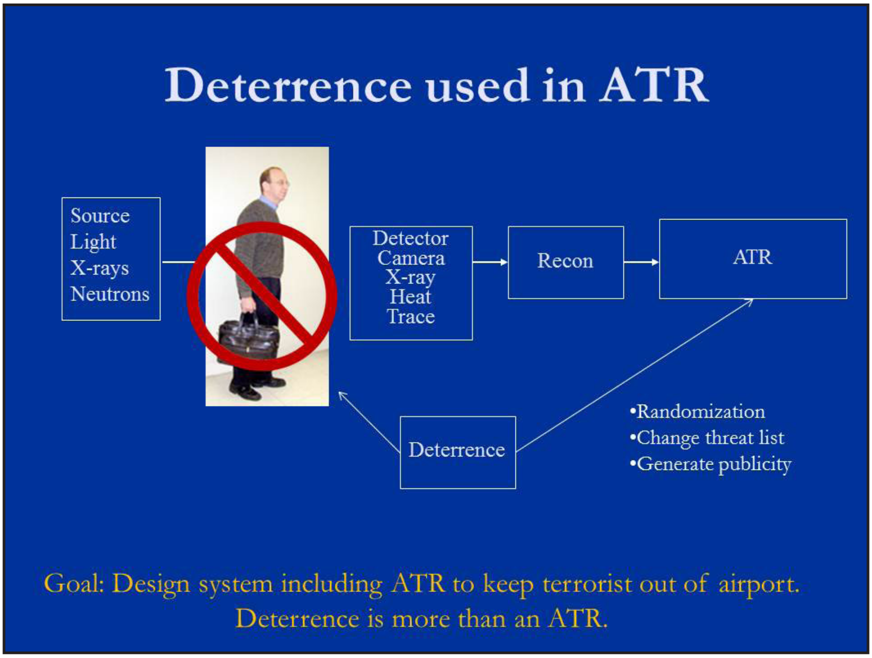
Detector
Camera
X-ray
Heat
Trace

Recon

ATR

Risk

Behavioral Detection Officer



Deterrence used in ATR

Source
Light
X-rays
Neutrons

Detector
Camera
X-ray
Heat
Trace

Recon

ATR

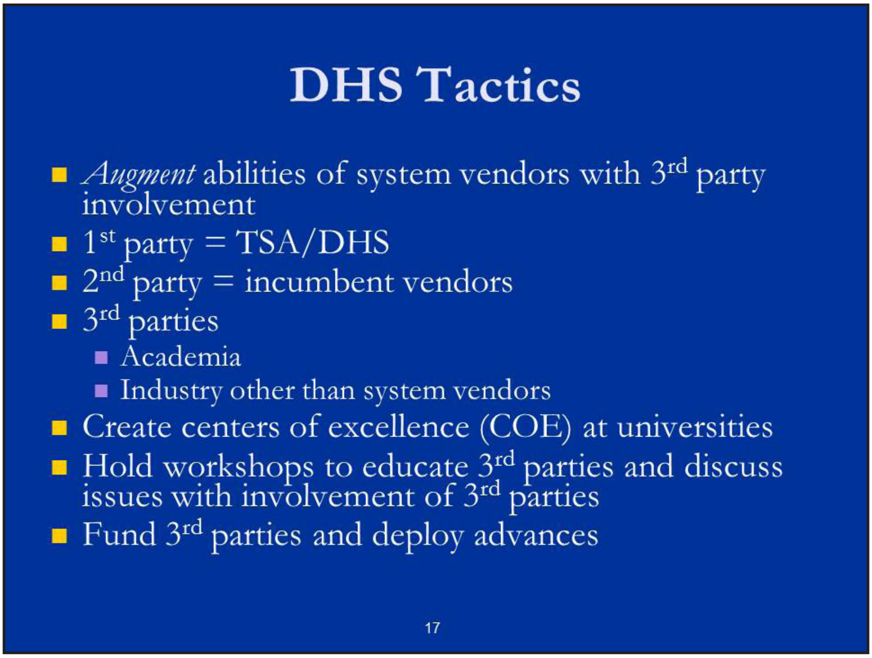
Deterrence

- Randomization
- Change threat list
- Generate publicity

Goal: Design system including ATR to keep terrorist out of airport.
Deterrence is more than an ATR.

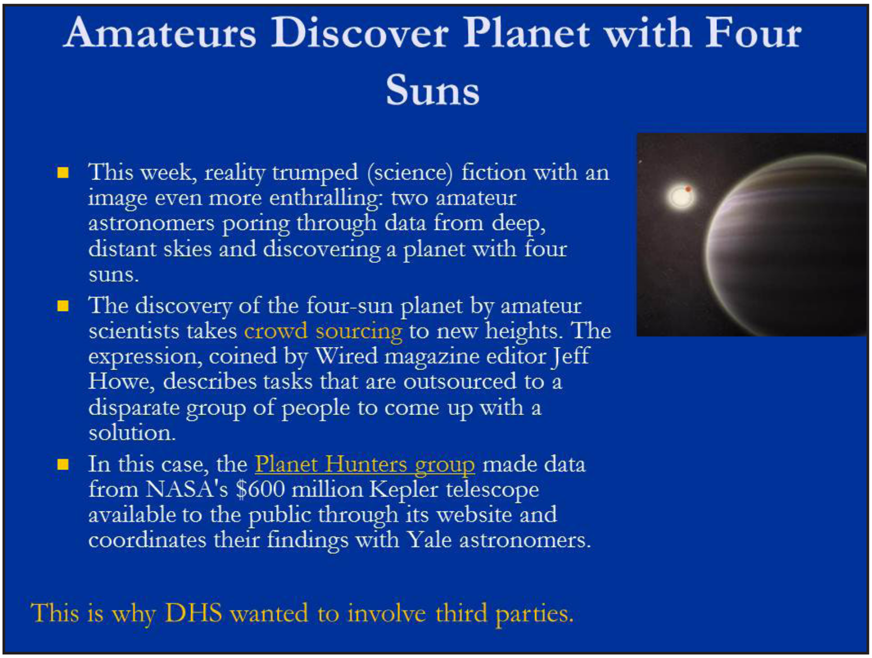
Tools Versus Craft

- Need to fix my gas furnace
  - Can purchase tools and parts from Home Depot
  - Would not do it myself
- ATR
  - Segmentation, classification tools available on web
    - Feature extraction and domain expertise not available on web
  - Requirement specs not on the web
  - ADSA08 mainly interested in how to approach developing an ATR



DHS Tactics

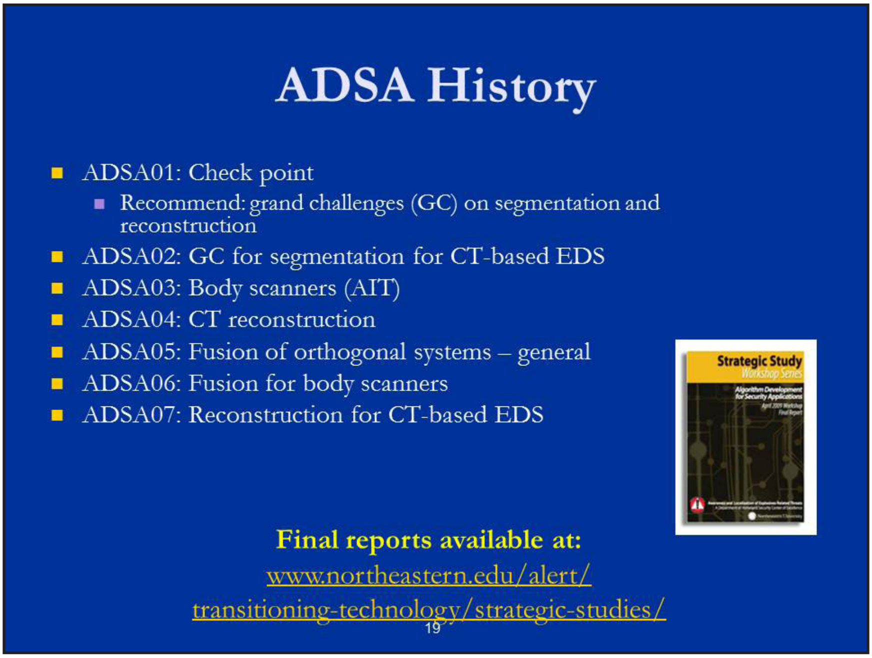
- *Augment* abilities of system vendors with 3rd party involvement
- 1st party = TSA/DHS
- 2nd party = incumbent vendors
- 3rd parties
    - Academia
    - Industry other than system vendors
- Create centers of excellence (COE) at universities
- Hold workshops to educate 3rd parties and discuss issues with involvement of 3rd parties
- Fund 3rd parties and deploy advances



Amateurs Discover Planet with Four Suns

- This week, reality trumped (science) fiction with an image even more enthralling: two amateur astronomers poring through data from deep, distant skies and discovering a planet with four suns.
- The discovery of the four-sun planet by amateur scientists takes crowd sourcing to new heights. The expression, coined by Wired magazine editor Jeff Howe, describes tasks that are outsourced to a disparate group of people to come up with a solution.
- In this case, the Planet Hunters group made data from NASA's $600 million Kepler telescope available to the public through its website and coordinates their findings with Yale astronomers.

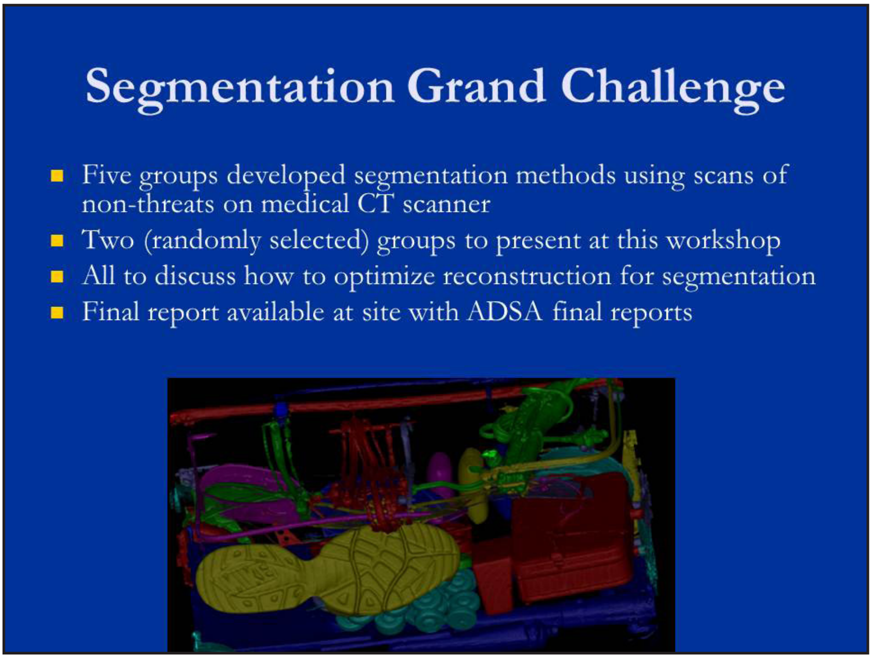
This is why DHS wanted to involve third parties.



ADSA History

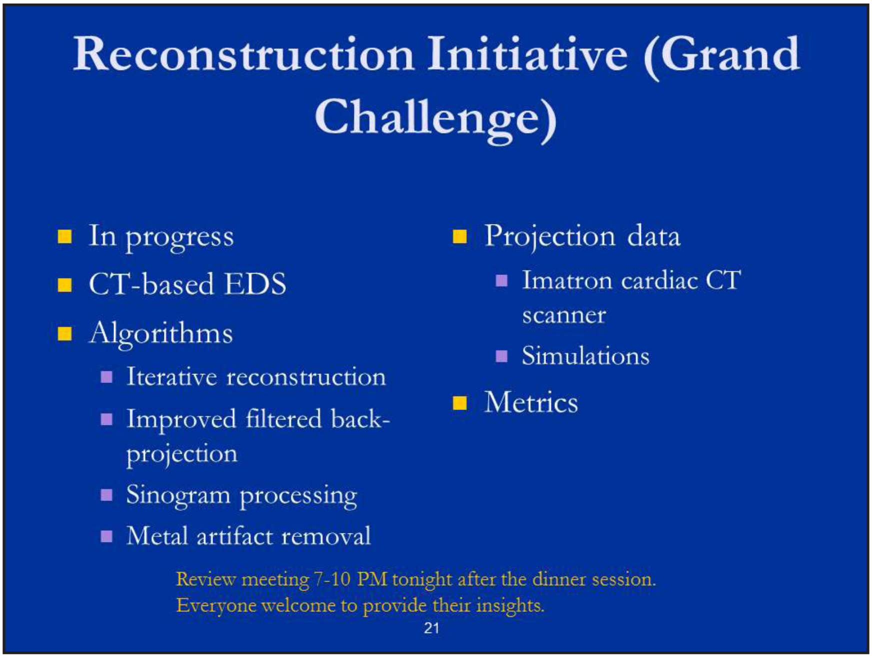
- ADSA01: Check point
  - Recommend: grand challenges (GC) on segmentation and reconstruction
- ADSA02: GC for segmentation for CT-based EDS
- ADSA03: Body scanners (AIT)
- ADSA04: CT reconstruction
- ADSA05: Fusion of orthogonal systems – general
- ADSA06: Fusion for body scanners
- ADSA07: Reconstruction for CT-based EDS

Final reports available at:
www.northeastern.edu/alert/
transitioning-technology/strategic-studies/



Segmentation Grand Challenge

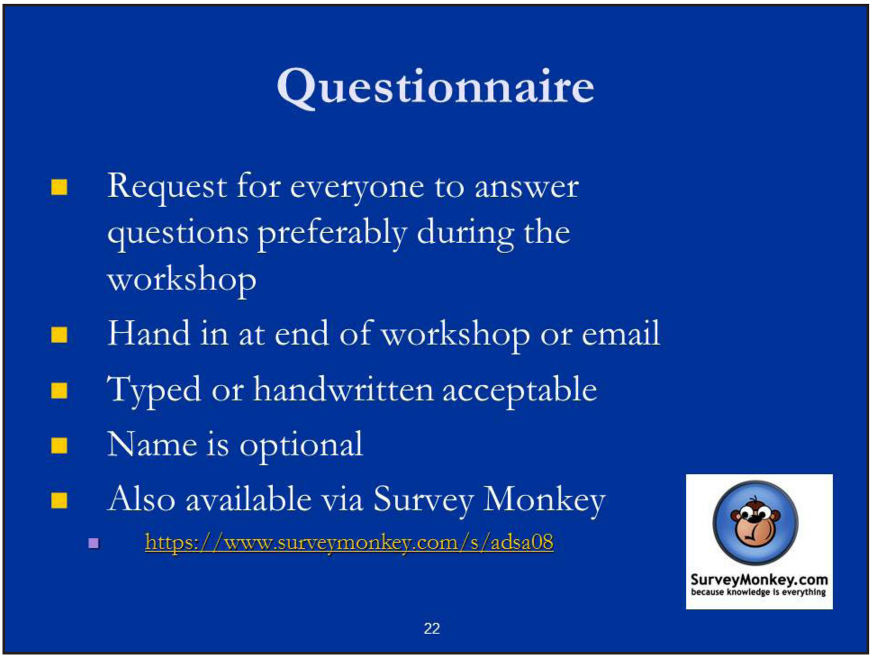
- Five groups developed segmentation methods using scans of non-threats on medical CT scanner
- Two (randomly selected) groups to present at this workshop
- All to discuss how to optimize reconstruction for segmentation
- Final report available at site with ADSA final reports



Reconstruction Initiative (Grand Challenge)

- In progress
- CT-based EDS
- Algorithms
  - Iterative reconstruction
  - Improved filtered back-projection
  - Sinogram processing
  - Metal artifact removal
- Projection data
  - Imatron cardiac CT scanner
  - Simulations
- Metrics

Review meeting 7-10 PM tonight after the dinner session.
Everyone welcome to provide their insights.

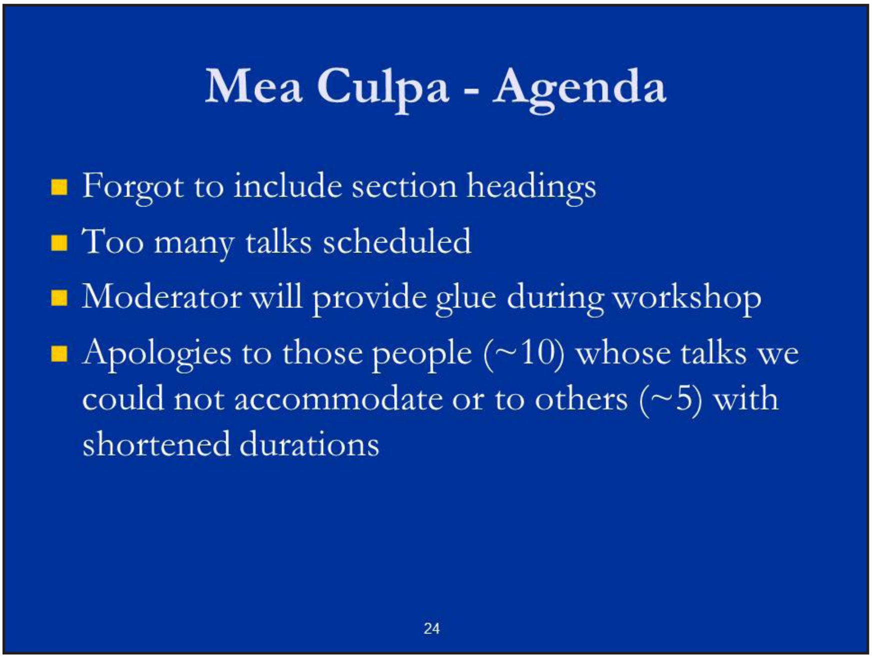


Questionnaire

- Request for everyone to answer questions preferably during the workshop
- Hand in at end of workshop or email
- Typed or handwritten acceptable
- Name is optional
- Also available via Survey Monkey
  - https://www.surveymonkey.com/s/adsa08

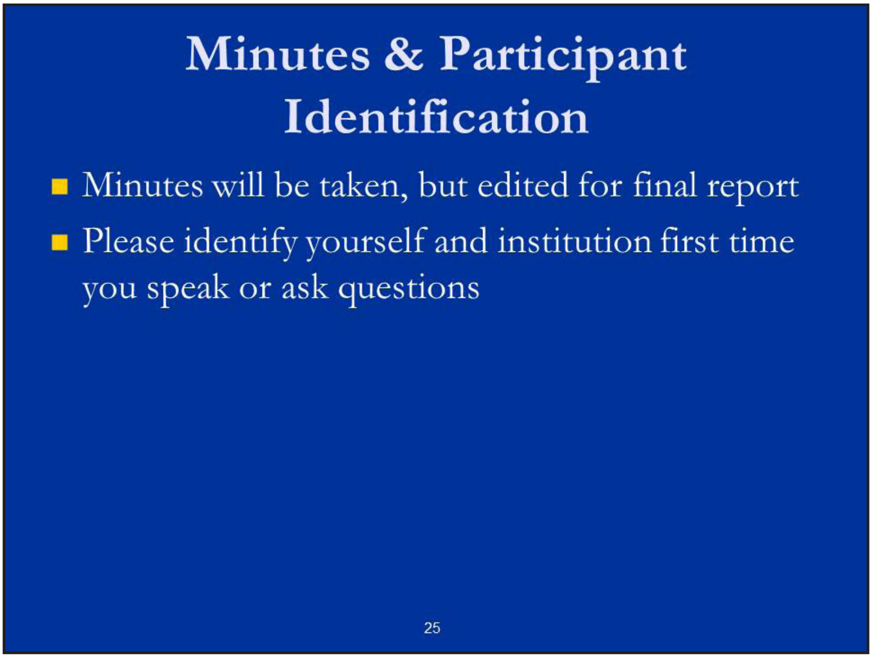
Reception and Dinner

- Reception and dinner tonight part of workshop
- Student poster session during the reception before dinner
- Reconstruction meeting after dinner
- Sorry about conflict with World Series
  - Go Red Sox!



Mea Culpa - Agenda

- Forgot to include section headings
- Too many talks scheduled
- Moderator will provide glue during workshop
- Apologies to those people (~10) whose talks we could not accommodate or to others (~5) with shortened durations

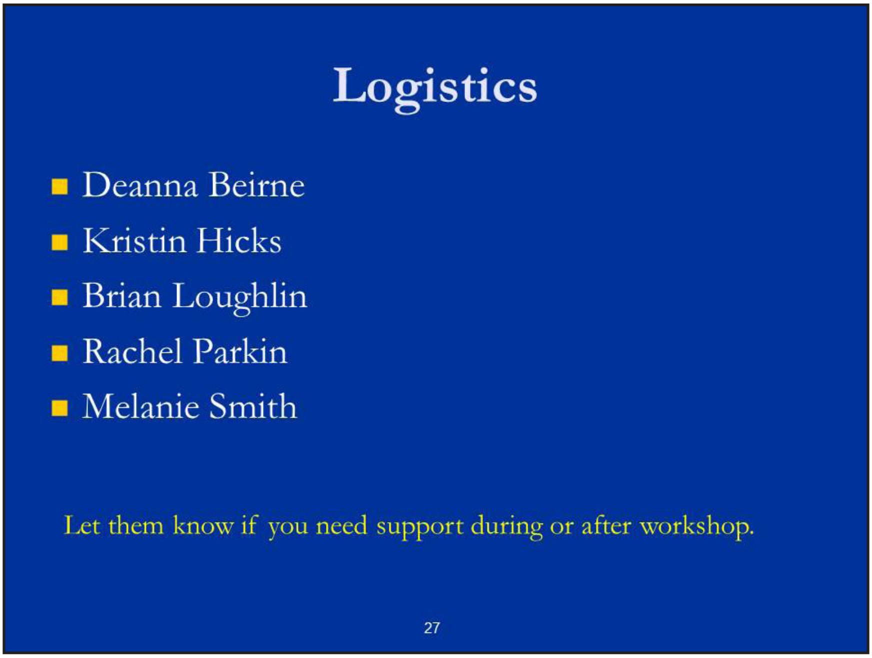


Minutes & Participant Identification

- Minutes will be taken, but edited for final report
- Please identify yourself and institution first time you speak or ask questions

Acknowledgements

- Northeastern University (NEU)
- Awareness and Localization of Explosives-Related Threats (ALERT) Center of Excellence
- Department of Homeland Security (DHS)
- Presenters
- Participants
- Students



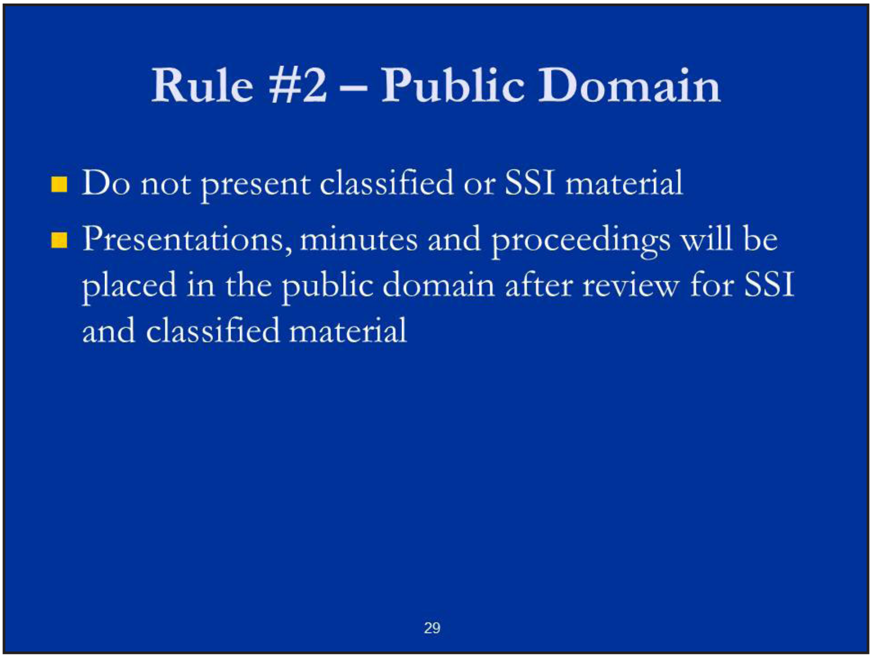
Logistics

- Deanna Beirne
- Kristin Hicks
- Brian Loughlin
- Rachel Parkin
- Melanie Smith

Let them know if you need support during or after workshop.

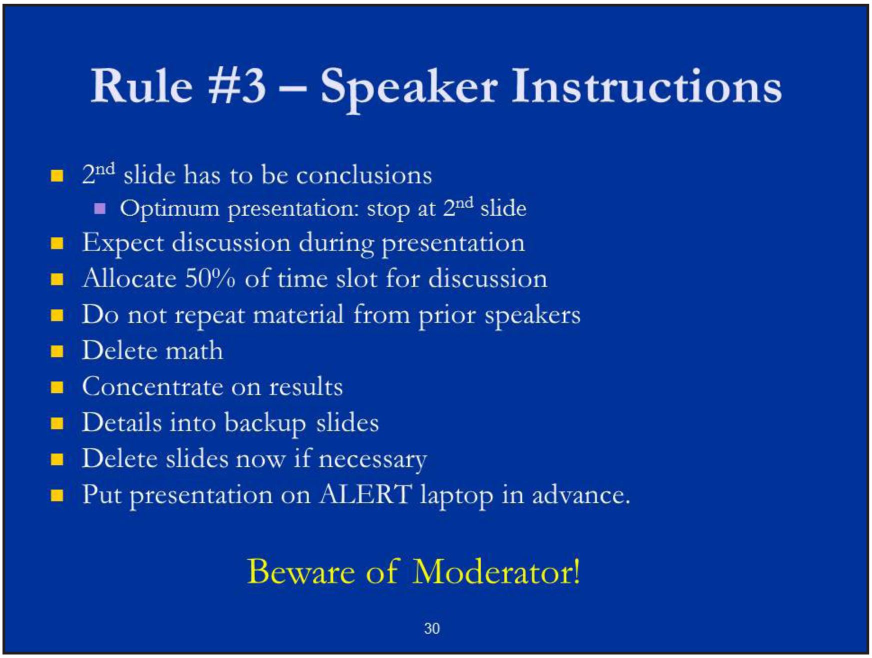
Rule #1 – Open Discussions

- This is a workshop
- Conversation and questions expected at all times, especially during presentations
- Moderator responsible for keeping discussions focused
- Not grip-and-grin



Rule #2 – Public Domain

- Do not present classified or SSI material
- Presentations, minutes and proceedings will be placed in the public domain after review for SSI and classified material



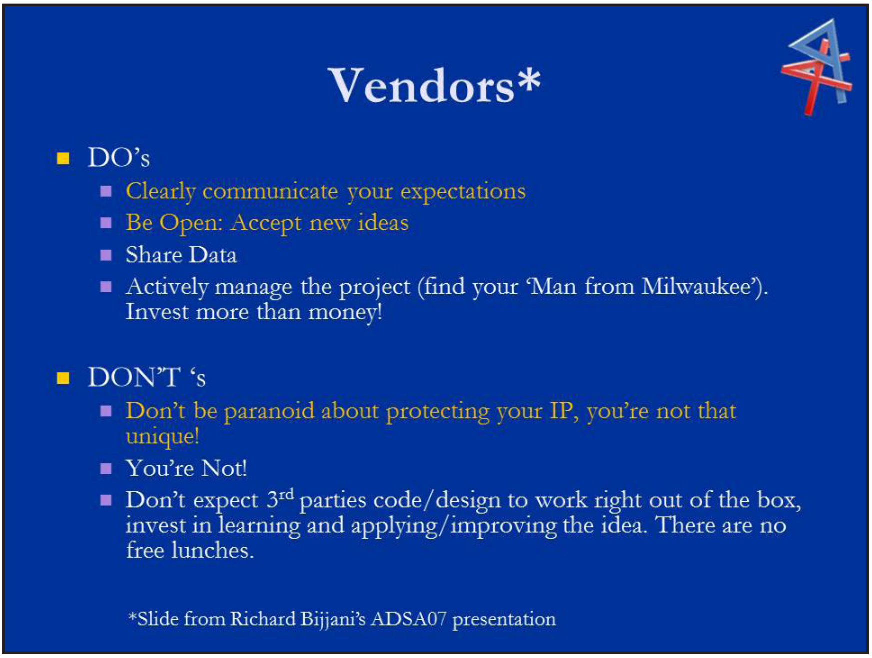
Rule #3 – Speaker Instructions

- 2nd slide has to be conclusions
  - Optimum presentation: stop at 2nd slide
- Expect discussion during presentation
- Allocate 50% of time slot for discussion
- Do not repeat material from prior speakers
- Delete math
- Concentrate on results
- Details into backup slides
- Delete slides now if necessary
- Put presentation on ALERT laptop in advance.

Beware of Moderator!

Logitech R800

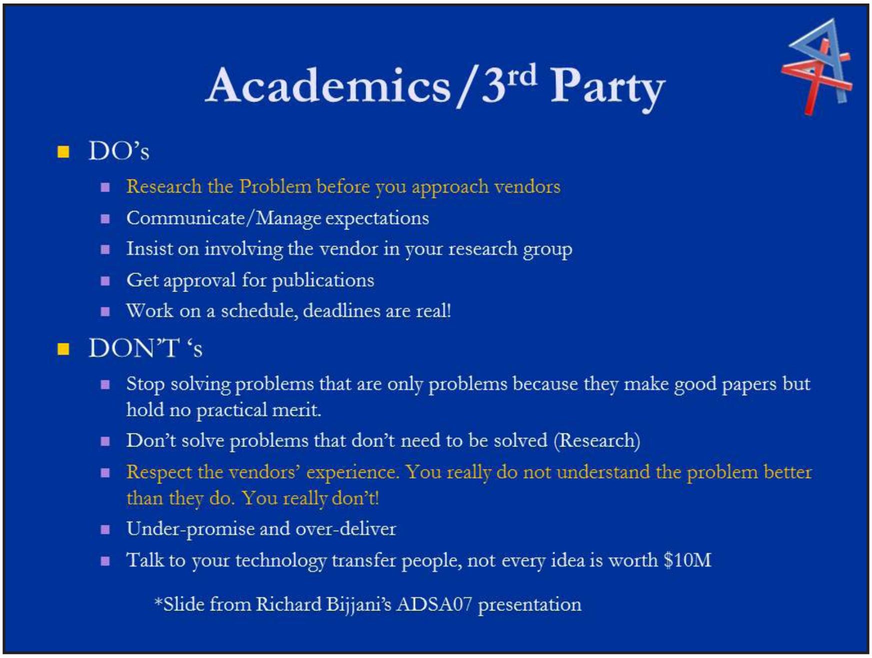
- Slide advancer
- Laser pointer
- Count-down timer
    - Vibrate at 5, 2, 0 minutes left
- Explodes if 1 minute late!!



Vendors*

- DO's
    - Clearly communicate your expectations
    - Be Open: Accept new ideas
    - Share Data
    - Actively manage the project (find your 'Man from Milwaukee'). Invest more than money!
- DON'T 's
    - Don't be paranoid about protecting your IP, you're not that unique!
    - You're Not!
    - Don't expect 3rd parties code/design to work right out of the box, invest in learning and applying/improving the idea. There are no free lunches.

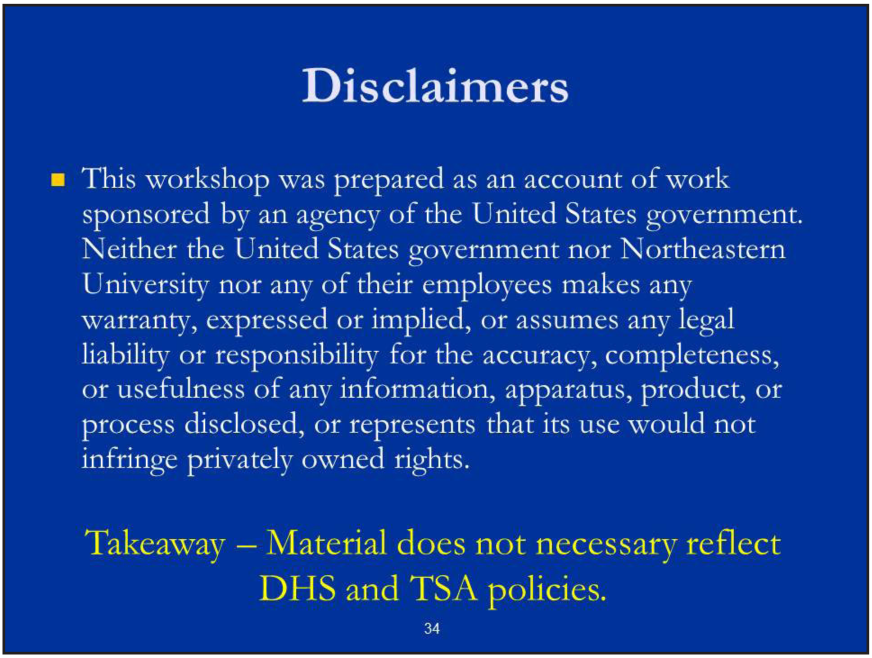
*Slide from Richard Bijjani's ADSA07 presentation



Academics/3rd Party

- DO's
  - Research the Problem before you approach vendors
  - Communicate/Manage expectations
  - Insist on involving the vendor in your research group
  - Get approval for publications
  - Work on a schedule, deadlines are real!
- DON'T 's
  - Stop solving problems that are only problems because they make good papers but hold no practical merit.
  - Don't solve problems that don't need to be solved (Research)
  - Respect the vendors' experience. You really do not understand the problem better than they do. You really don't!
  - Under-promise and over-deliver
  - Talk to your technology transfer people, not every idea is worth $10M

*Slide from Richard Bijjani's ADSA07 presentation



Disclaimers

- This workshop was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Northeastern University nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Takeaway – Material does not necessary reflect DHS and TSA policies.

Questions & Answers

- "Knowledge in general and self-knowledge in particular are gained not only from discovering logical answers but also from formulating logical, even though unanswerable, questions."*
- Framing problem is good start; need to do more over time.

*Soloveitchik, Joseph B. (2009-07-01). The Lonely Man of Faith (Kindle Locations 184-185). Random House, Inc.. Kindle Edition.

Final Remarks

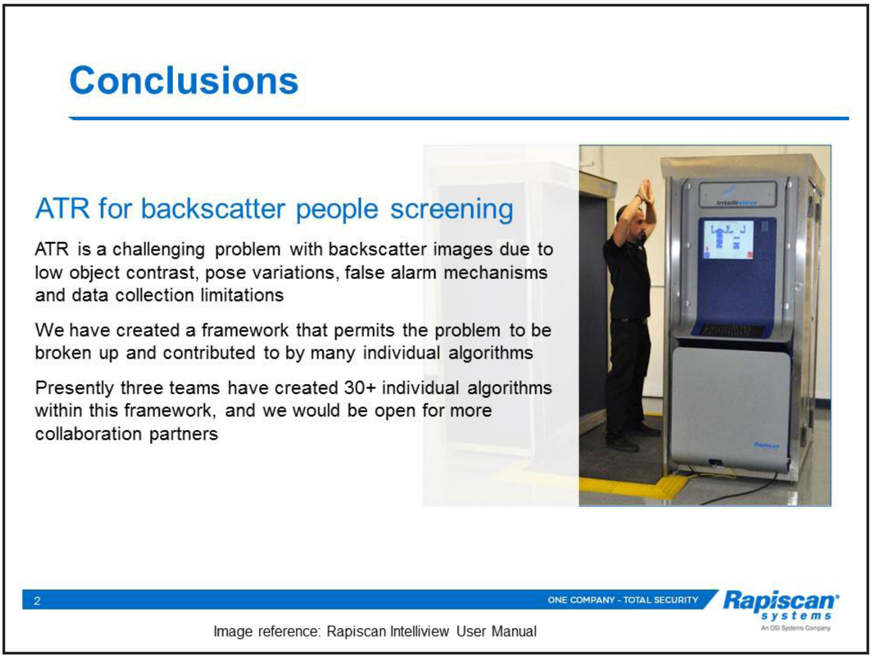
- "Terrorism causes a loss of life and a loss of quality of life," Lisa Dolev, Qylur
- Need improved technology
- Thank you for participating

16.3 Alex Hudson: ATR for Personnel Screenings

ONE COMPANY - TOTAL SECURITY

Rapiscan AIT ATR
Alex Hudson

www.rapiscansystems.com



Conclusions

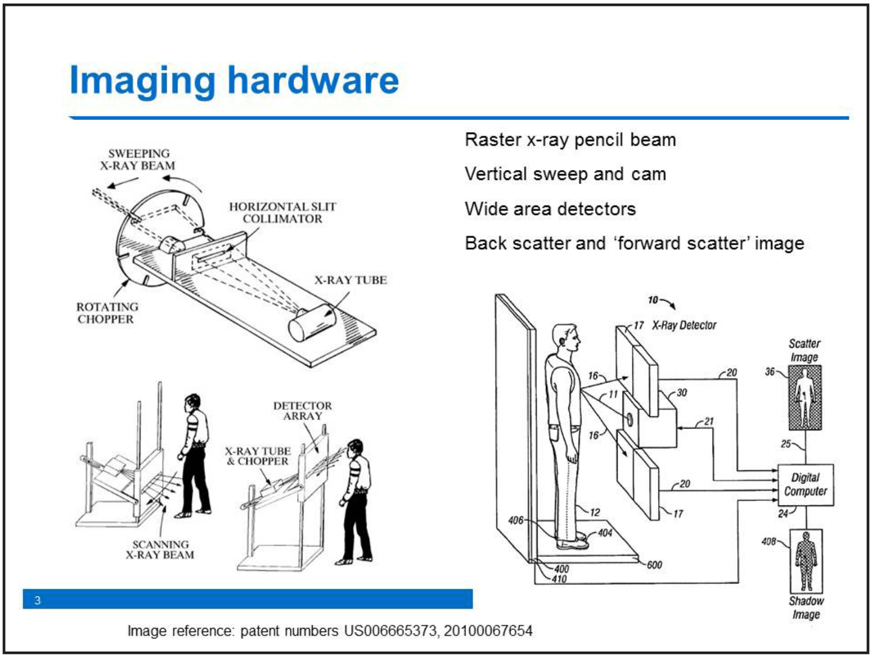
ATR for backscatter people screening

ATR is a challenging problem with backscatter images due to low object contrast, pose variations, false alarm mechanisms and data collection limitations

We have created a framework that permits the problem to be broken up and contributed to by many individual algorithms

Presently three teams have created 30+ individual algorithms within this framework, and we would be open for more collaboration partners

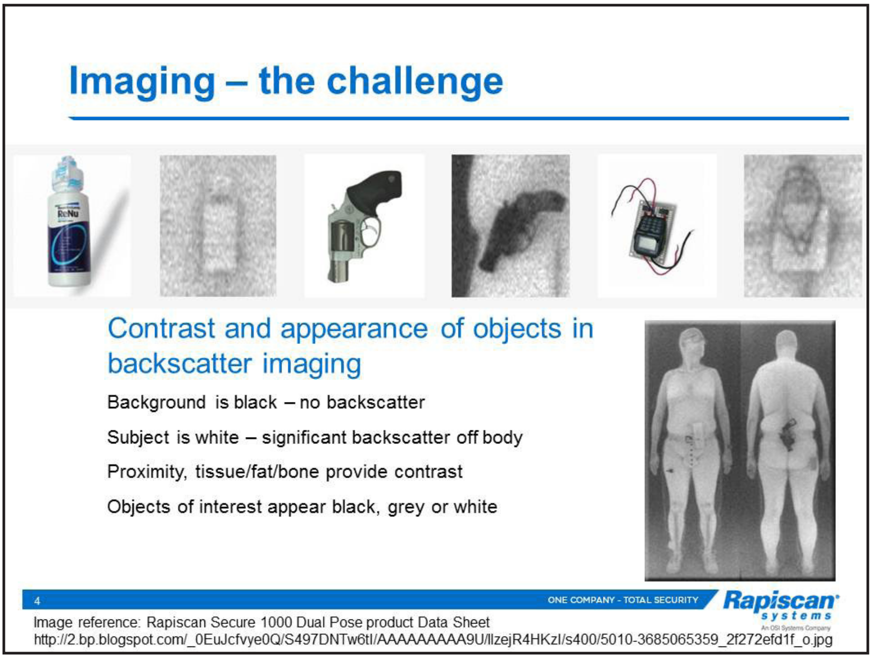
Image reference: Rapiscan Intelliview User Manual



Imaging hardware

- Raster x-ray pencil beam
- Vertical sweep and cam
- Wide area detectors
- Back scatter and ‘forward scatter’ image

Image reference: patent numbers US006665373, 20100067654



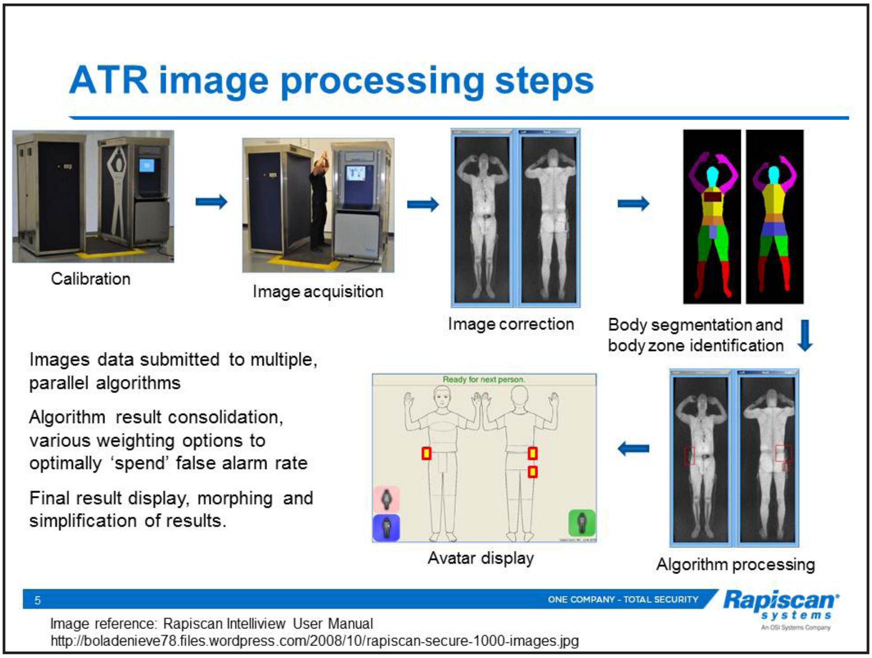
Imaging – the challenge

Contrast and appearance of objects in backscatter imaging

- Background is black – no backscatter
- Subject is white – significant backscatter off body
- Proximity, tissue/fat/bone provide contrast
- Objects of interest appear black, grey or white

ONE COMPANY - TOTAL SECURITY Rapiscan systems An OSI Systems Company

Image reference: Rapiscan Secure 1000 Dual Pose product Data Sheet
http://2.bp.blogspot.com/_0EuJcfvye0Q/S497DNTw6tI/AAAAAAAAA9U/llzejR4HKzI/s400/5010-3685065359_2f272efd1f_o.jpg



ATR image processing steps

Calibration → Image acquisition → Image correction → Body segmentation and body zone identification ↓

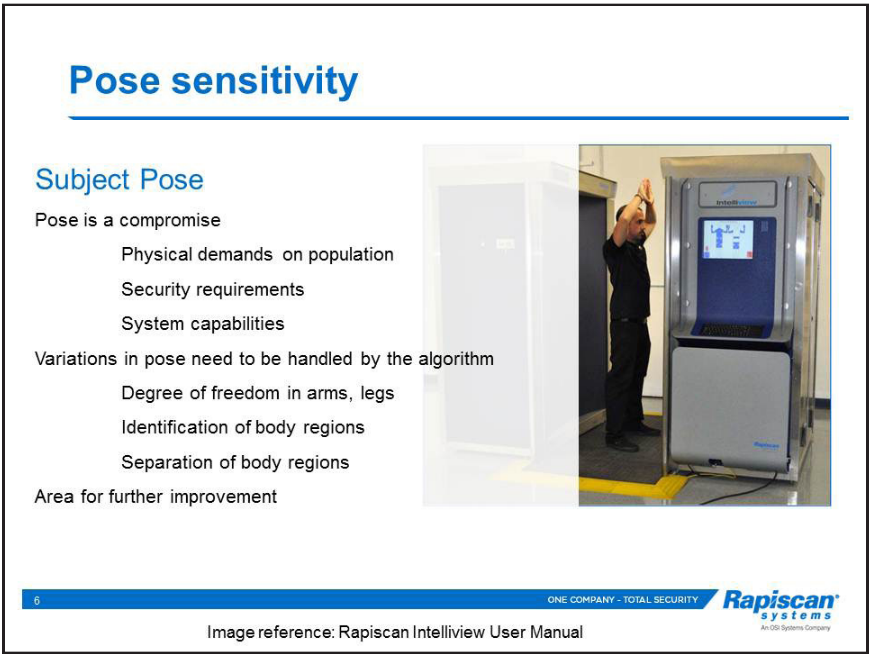
Algorithm processing → Avatar display

Images data submitted to multiple, parallel algorithms

Algorithm result consolidation, various weighting options to optimally 'spend' false alarm rate

Final result display, morphing and simplification of results.

Image reference: Rapiscan Intelliview User Manual
http://boladenieve78.files.wordpress.com/2008/10/rapiscan-secure-1000-images.jpg



Pose sensitivity

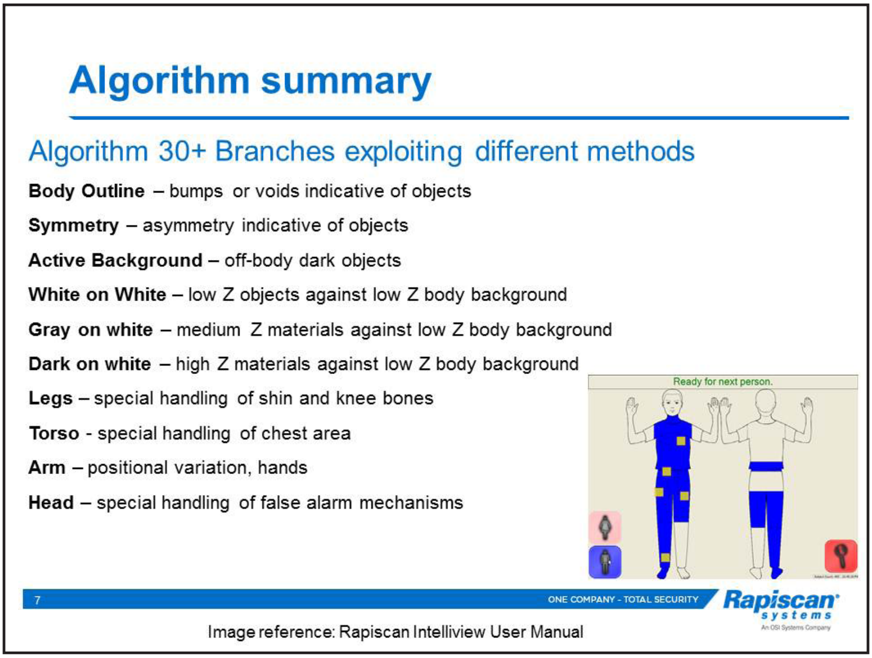
Subject Pose

Pose is a compromise
- Physical demands on population
- Security requirements
- System capabilities

Variations in pose need to be handled by the algorithm
- Degree of freedom in arms, legs
- Identification of body regions
- Separation of body regions

Area for further improvement

Image reference: Rapiscan Intelliview User Manual



Algorithm summary

Algorithm 30+ Branches exploiting different methods

Body Outline – bumps or voids indicative of objects

Symmetry – asymmetry indicative of objects

Active Background – off-body dark objects

White on White – low Z objects against low Z body background

Gray on white – medium Z materials against low Z body background

Dark on white – high Z materials against low Z body background

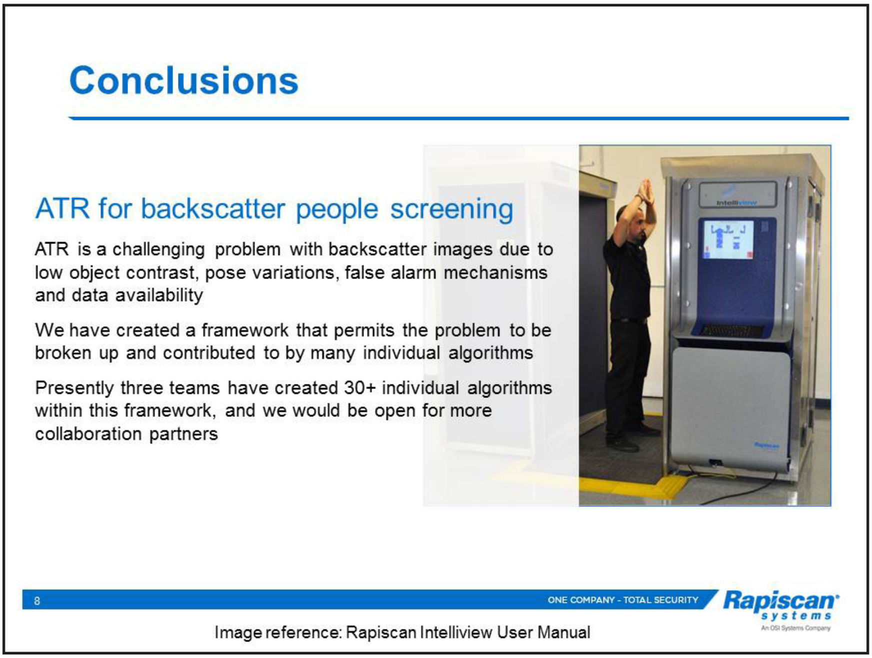
Legs – special handling of shin and knee bones

Torso - special handling of chest area

Arm – positional variation, hands

Head – special handling of false alarm mechanisms

Image reference: Rapiscan Intelliview User Manual



Conclusions

ATR for backscatter people screening

ATR is a challenging problem with backscatter images due to low object contrast, pose variations, false alarm mechanisms and data availability

We have created a framework that permits the problem to be broken up and contributed to by many individual algorithms

Presently three teams have created 30+ individual algorithms within this framework, and we would be open for more collaboration partners

Image reference: Rapiscan Intelliview User Manual

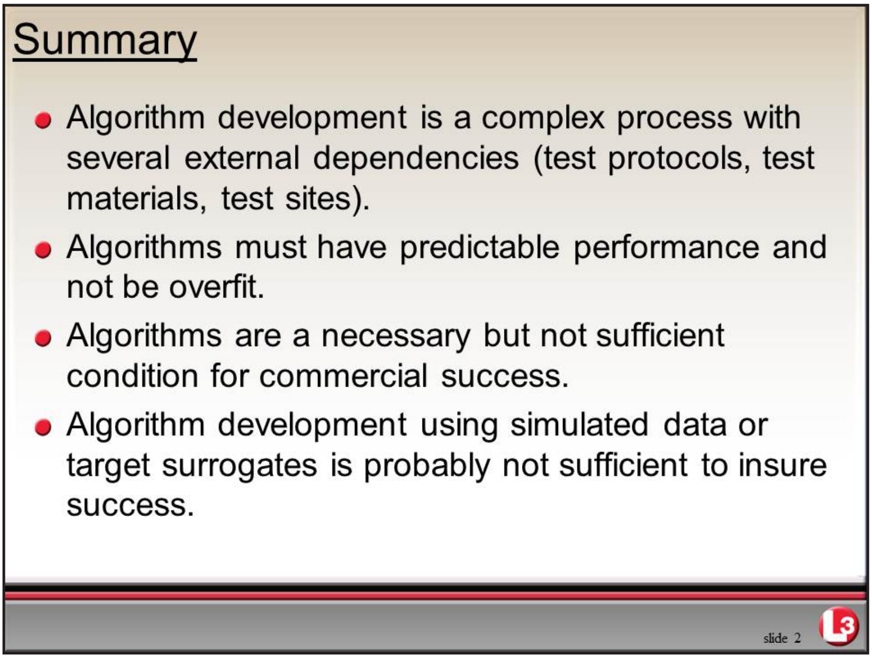
16.4 David Perticone: ATR for various modalities

L-3 Communications Security & Detection Systems

ATR for various modalities

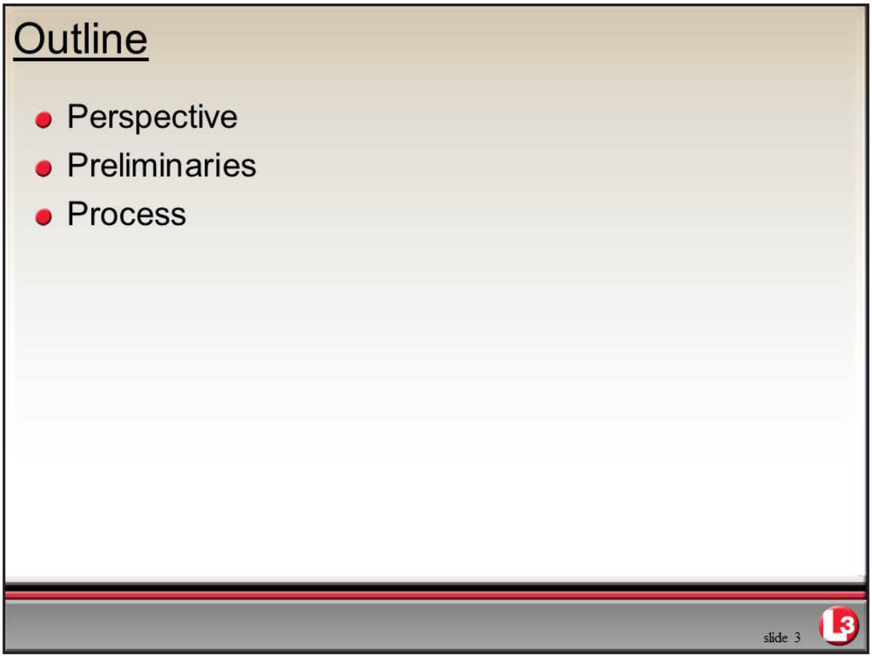
Dr. David Perticone
Engineering Fellow

ADSA08 10/24/2012



Summary

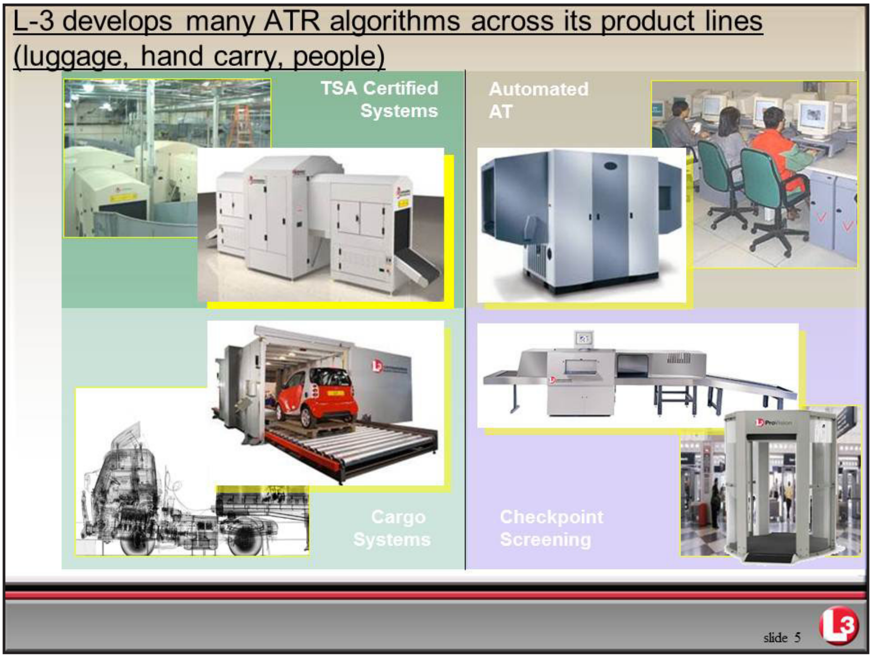
- Algorithm development is a complex process with several external dependencies (test protocols, test materials, test sites).
- Algorithms must have predictable performance and not be overfit.
- Algorithms are a necessary but not sufficient condition for commercial success.
- Algorithm development using simulated data or target surrogates is probably not sufficient to insure success.



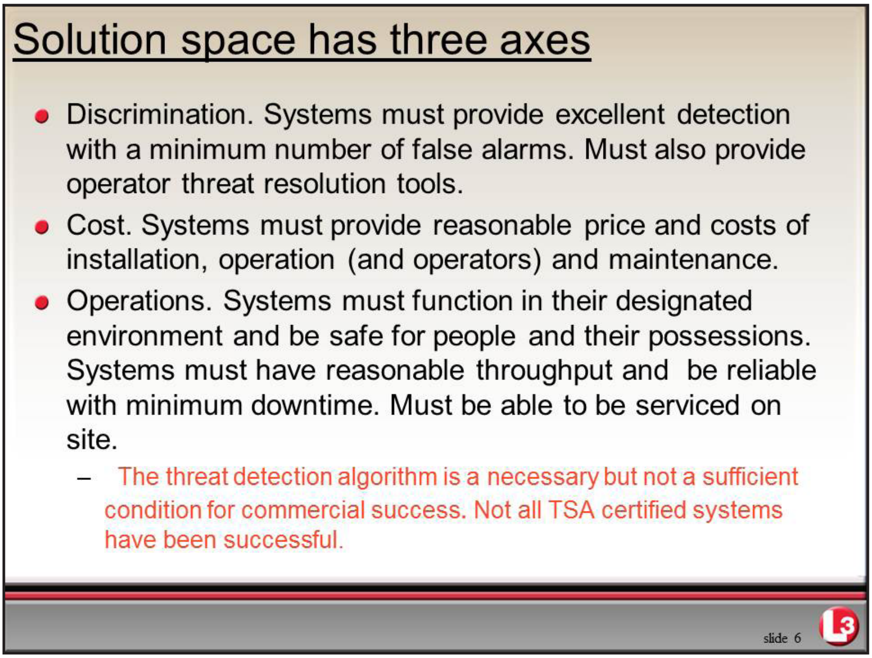
Outline

- Perspective
- Preliminaries
- Process

Perspective

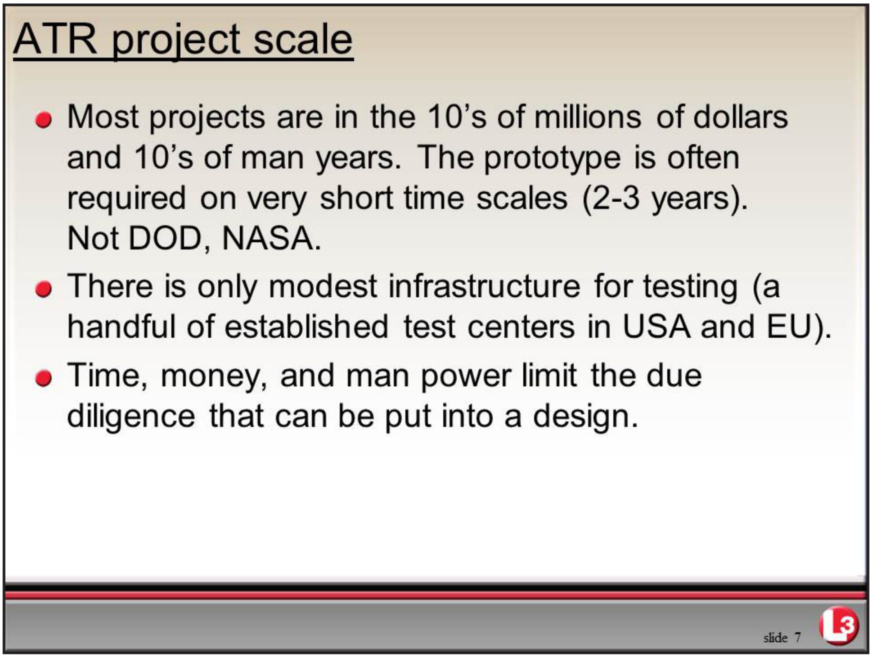


L-3 develops many ATR algorithms across its product lines (luggage, hand carry, people)



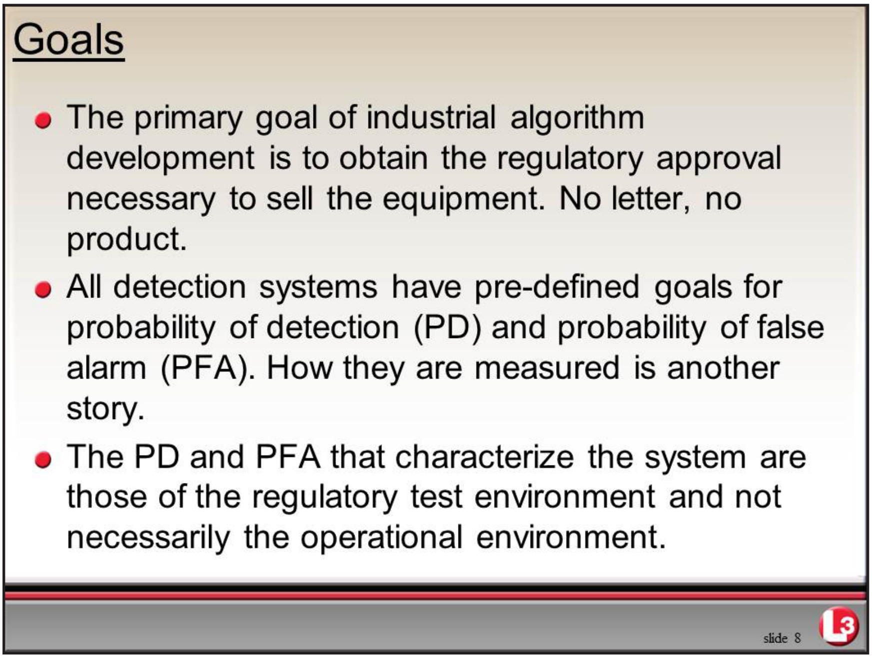
Solution space has three axes

- Discrimination. Systems must provide excellent detection with a minimum number of false alarms. Must also provide operator threat resolution tools.
- Cost. Systems must provide reasonable price and costs of installation, operation (and operators) and maintenance.
- Operations. Systems must function in their designated environment and be safe for people and their possessions. Systems must have reasonable throughput and be reliable with minimum downtime. Must be able to be serviced on site.
  - The threat detection algorithm is a necessary but not a sufficient condition for commercial success. Not all TSA certified systems have been successful.



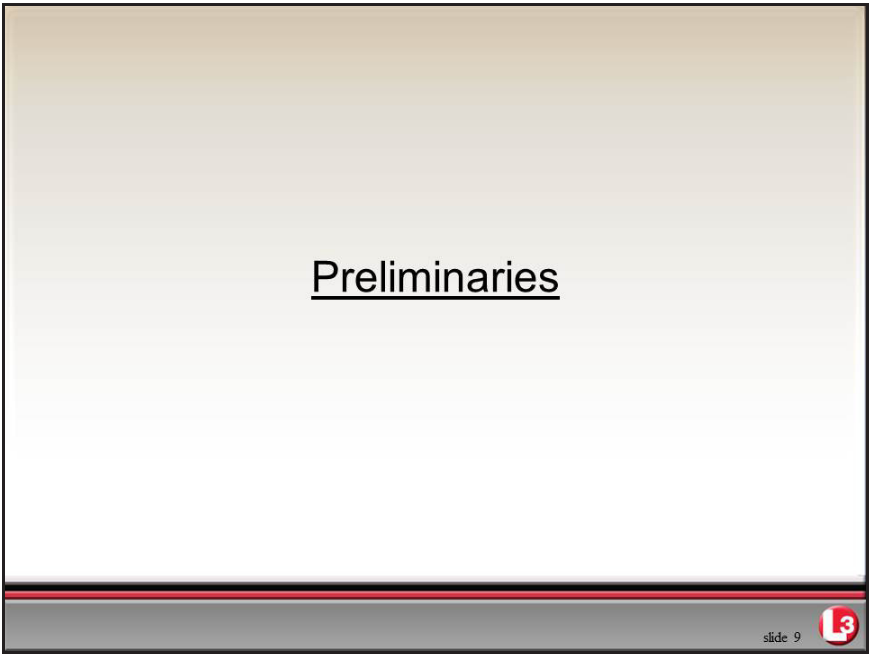
ATR project scale

- Most projects are in the 10's of millions of dollars and 10's of man years. The prototype is often required on very short time scales (2-3 years). Not DOD, NASA.
- There is only modest infrastructure for testing (a handful of established test centers in USA and EU).
- Time, money, and man power limit the due diligence that can be put into a design.

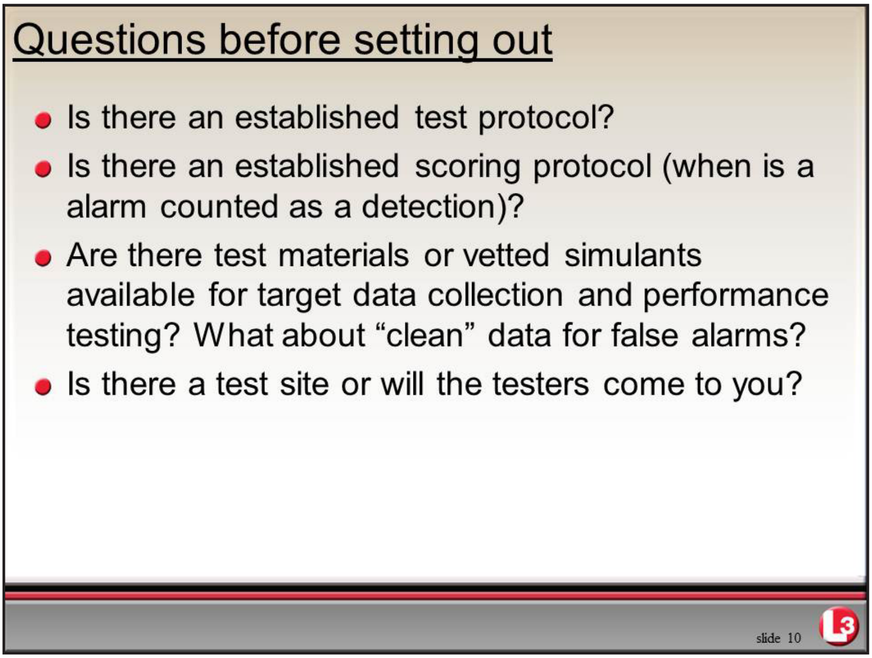


Goals

- The primary goal of industrial algorithm development is to obtain the regulatory approval necessary to sell the equipment. No letter, no product.
- All detection systems have pre-defined goals for probability of detection (PD) and probability of false alarm (PFA). How they are measured is another story.
- The PD and PFA that characterize the system are those of the regulatory test environment and not necessarily the operational environment.

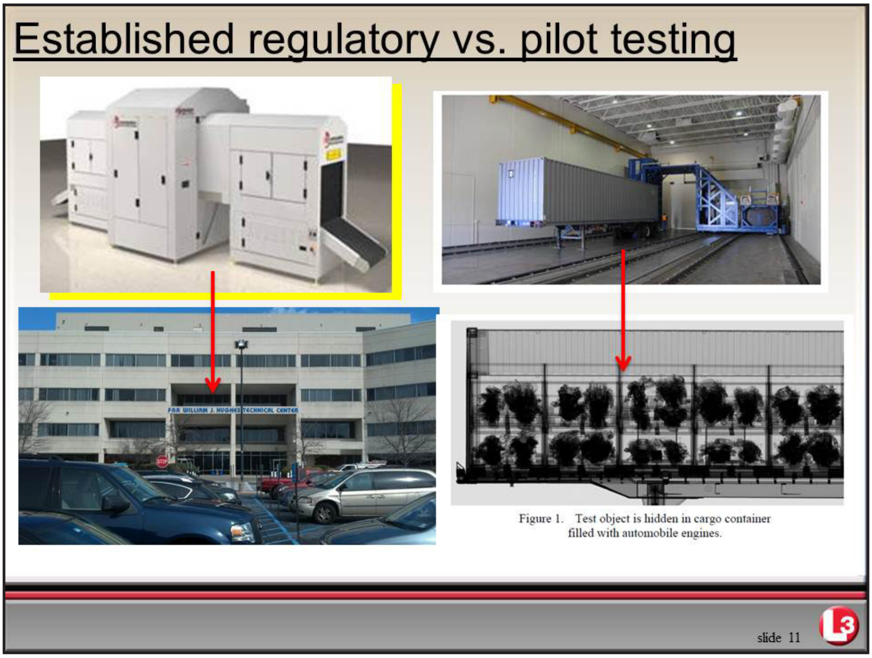


Preliminaries



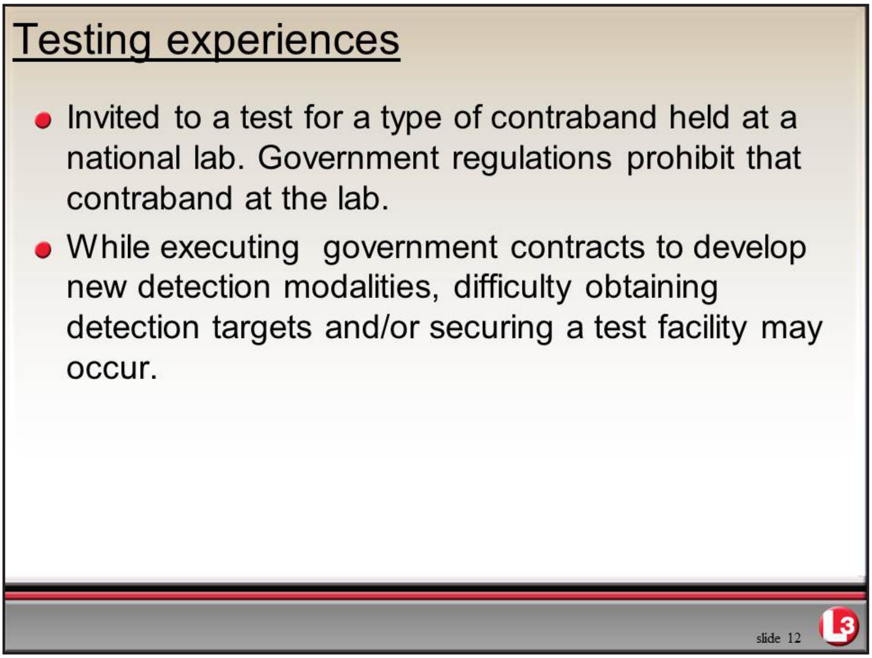
Questions before setting out

- Is there an established test protocol?
- Is there an established scoring protocol (when is a alarm counted as a detection)?
- Are there test materials or vetted simulants available for target data collection and performance testing? What about "clean" data for false alarms?
- Is there a test site or will the testers come to you?



Established regulatory vs. pilot testing

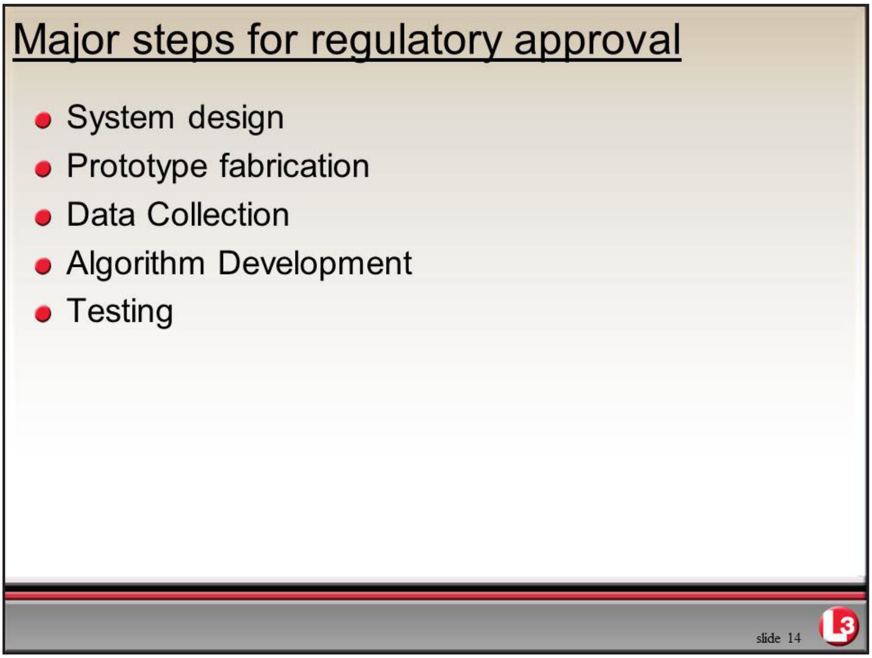
Figure 1. Test object is hidden in cargo container filled with automobile engines.



Testing experiences

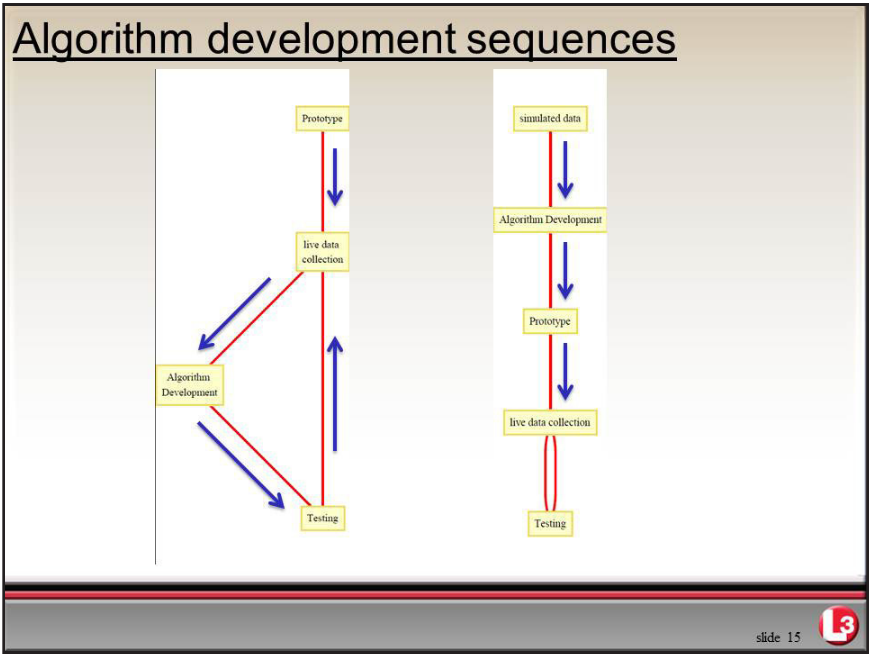
- Invited to a test for a type of contraband held at a national lab. Government regulations prohibit that contraband at the lab.
- While executing government contracts to develop new detection modalities, difficulty obtaining detection targets and/or securing a test facility may occur.

Process



Major steps for regulatory approval

- System design
- Prototype fabrication
- Data Collection
- Algorithm Development
- Testing



Algorithm development sequences

Prototype

live data collection

Algorithm Development

Testing

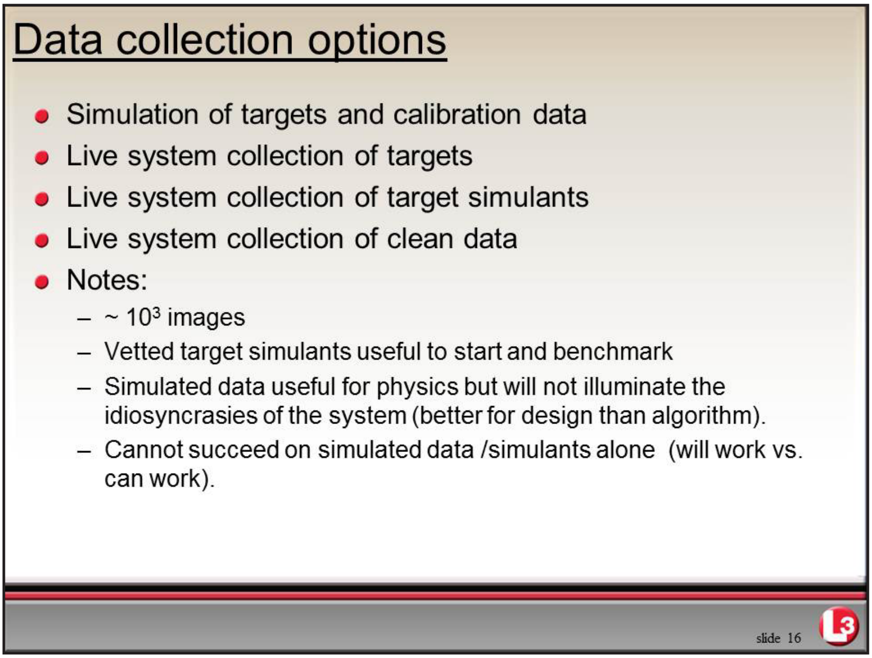
simulated data

Algorithm Development

Prototype

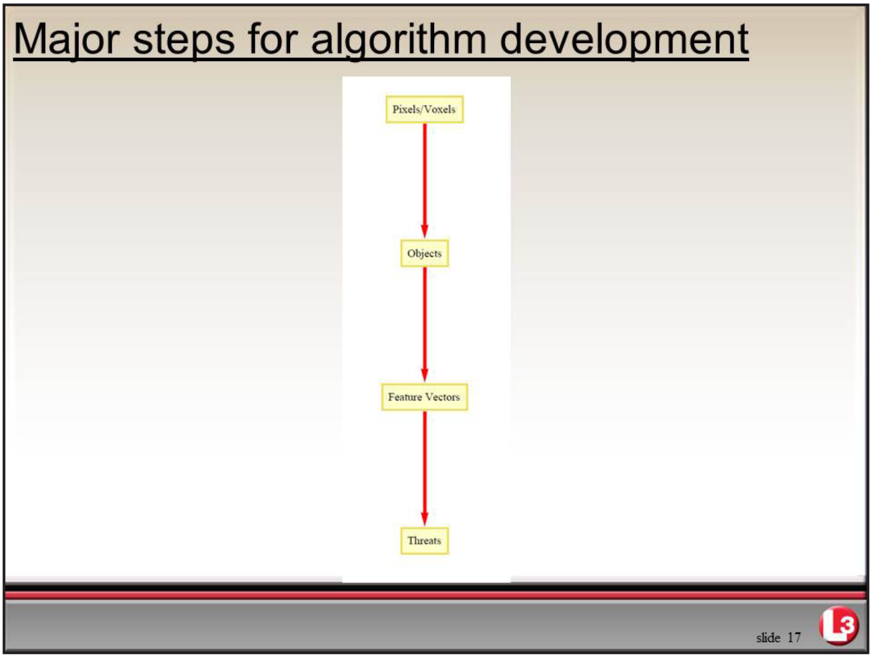
live data collection

Testing



Data collection options

- Simulation of targets and calibration data
- Live system collection of targets
- Live system collection of target simulants
- Live system collection of clean data
- Notes:
  - ~ 10^3 images
  - Vetted target simulants useful to start and benchmark
  - Simulated data useful for physics but will not illuminate the idiosyncrasies of the system (better for design than algorithm).
  - Cannot succeed on simulated data /simulants alone (will work vs. can work).



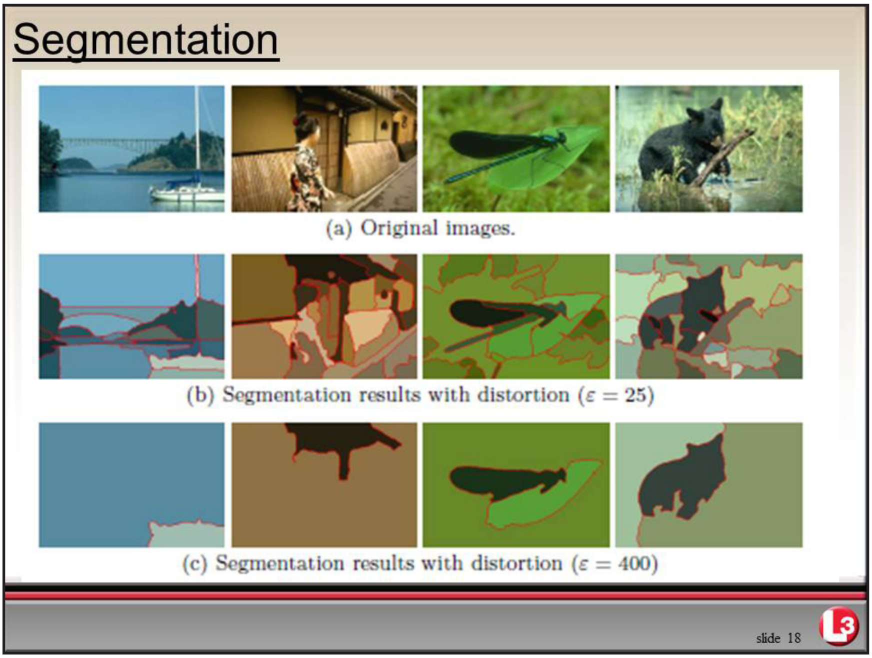
Major steps for algorithm development

Pixels/Voxels

Objects

Feature Vectors

Threats

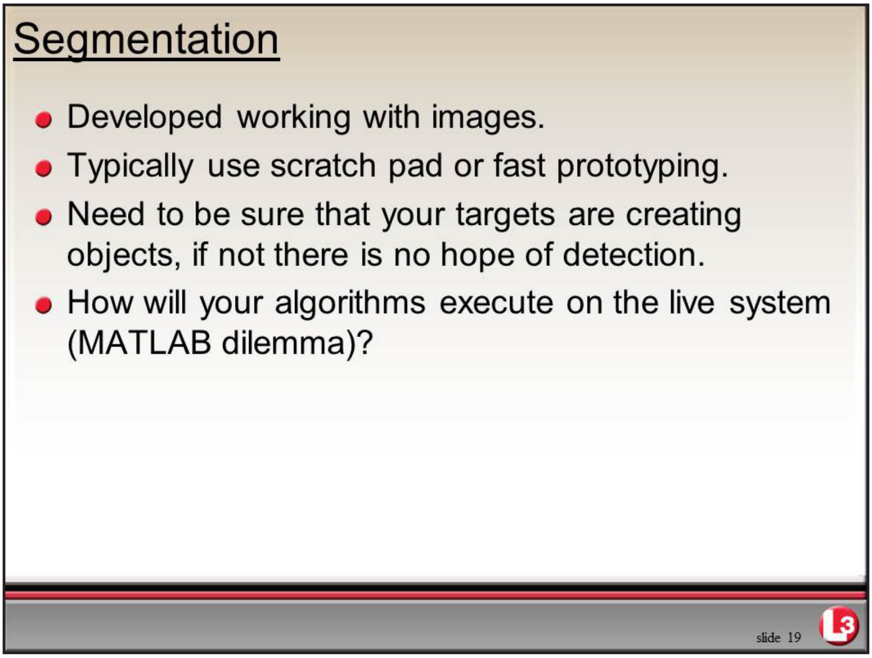


Segmentation

(a) Original images.

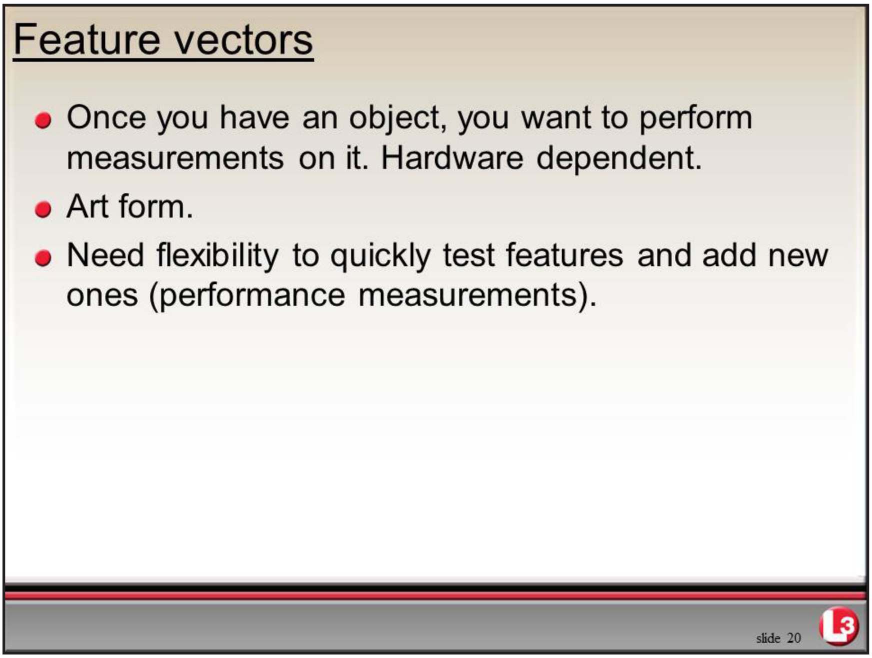
(b) Segmentation results with distortion ($\varepsilon = 25$)

(c) Segmentation results with distortion ($\varepsilon = 400$)



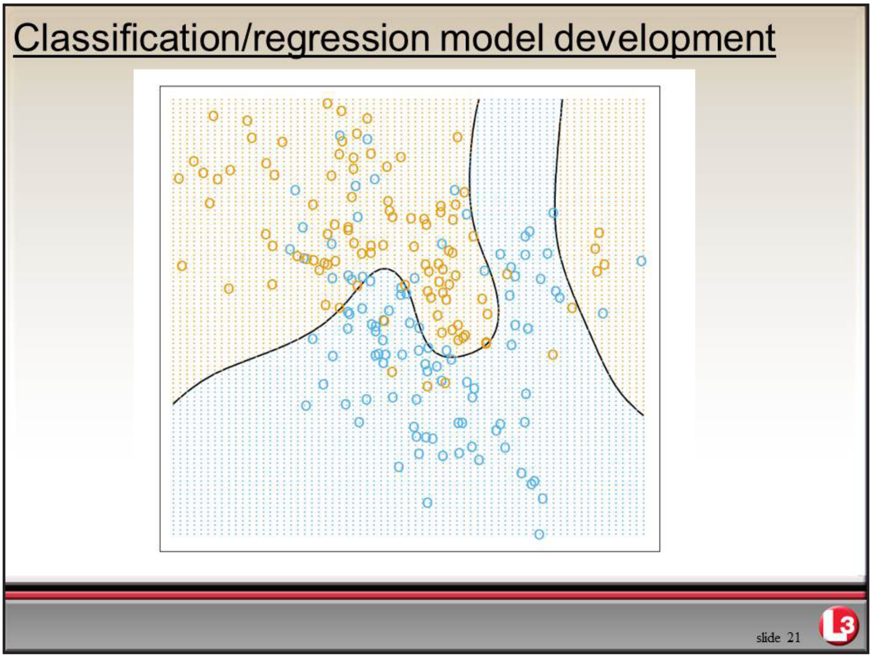
Segmentation

- Developed working with images.
- Typically use scratch pad or fast prototyping.
- Need to be sure that your targets are creating objects, if not there is no hope of detection.
- How will your algorithms execute on the live system (MATLAB dilemma)?

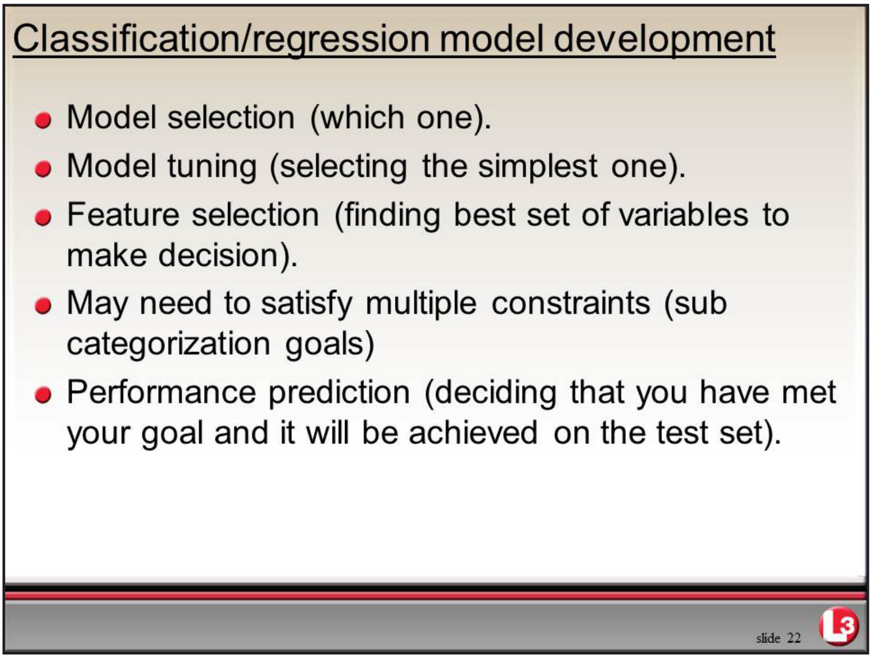


Feature vectors

- Once you have an object, you want to perform measurements on it. Hardware dependent.
- Art form.
- Need flexibility to quickly test features and add new ones (performance measurements).

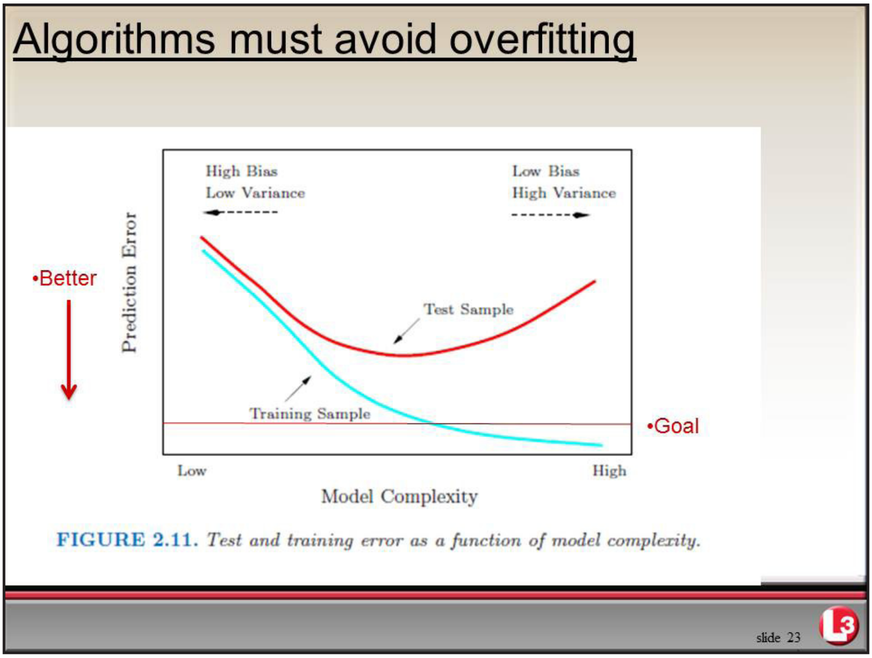


Classification/regression model development



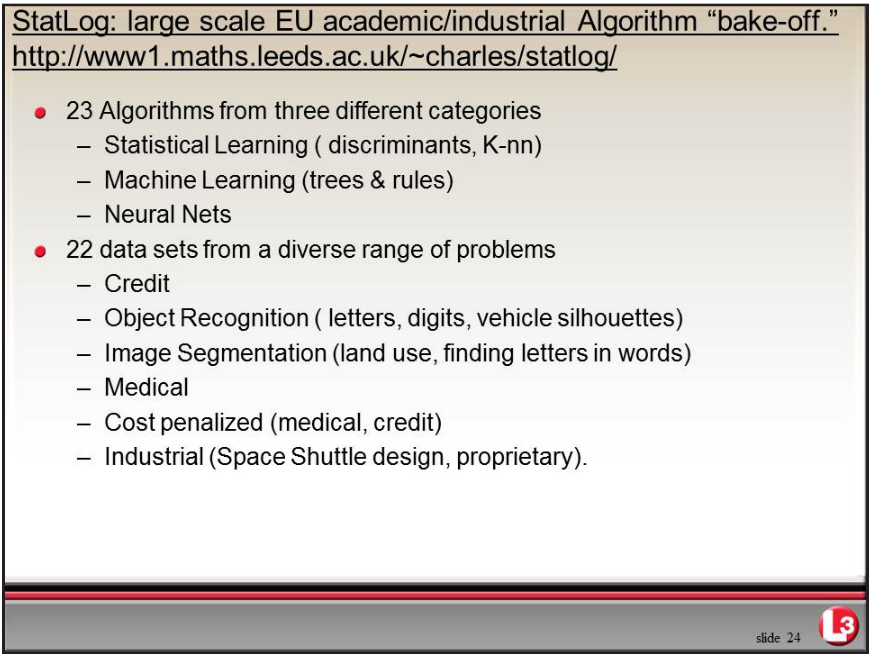
Classification/regression model development

- Model selection (which one).
- Model tuning (selecting the simplest one).
- Feature selection (finding best set of variables to make decision).
- May need to satisfy multiple constraints (sub categorization goals)
- Performance prediction (deciding that you have met your goal and it will be achieved on the test set).



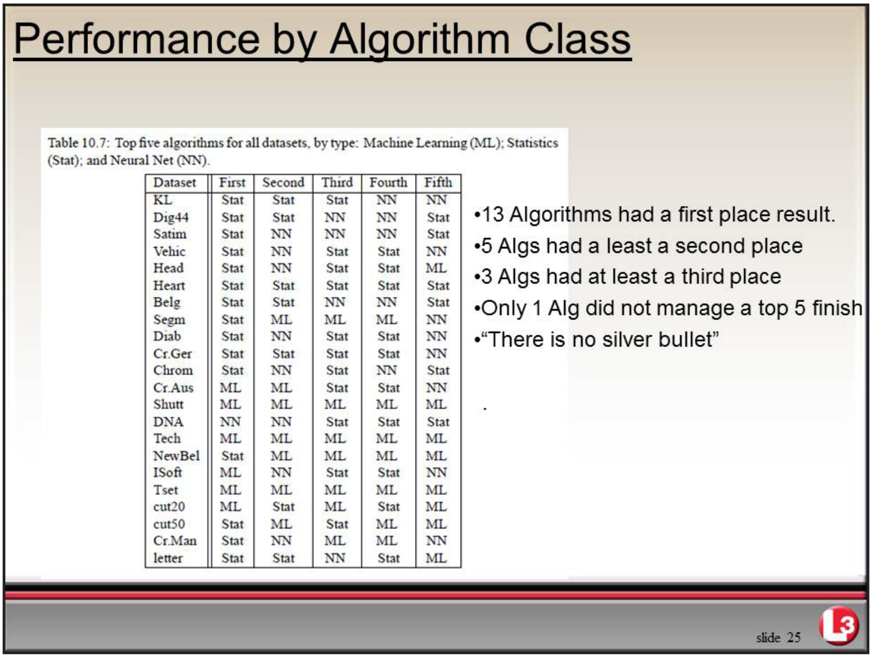
Algorithms must avoid overfitting

FIGURE 2.11. *Test and training error as a function of model complexity.*



StatLog: large scale EU academic/industrial Algorithm "bake-off." http://www1.maths.leeds.ac.uk/~charles/statlog/

- 23 Algorithms from three different categories
  - Statistical Learning (discriminants, K-nn)
  - Machine Learning (trees & rules)
  - Neural Nets
- 22 data sets from a diverse range of problems
  - Credit
  - Object Recognition (letters, digits, vehicle silhouettes)
  - Image Segmentation (land use, finding letters in words)
  - Medical
  - Cost penalized (medical, credit)
  - Industrial (Space Shuttle design, proprietary).



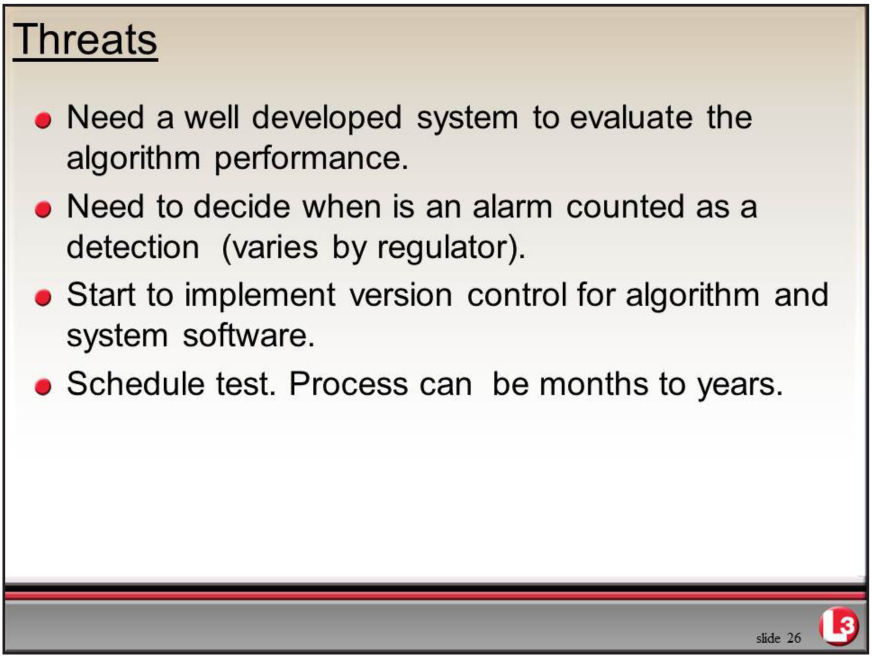
Performance by Algorithm Class

Table 10.7: Top five algorithms for all datasets, by type: Machine Learning (ML); Statistics (Stat); and Neural Net (NN).

Dataset	First	Second	Third	Fourth	Fifth
KL	Stat	Stat	Stat	NN	NN
Dig44	Stat	Stat	NN	NN	Stat
Satim	Stat	NN	NN	NN	Stat
Vehic	Stat	NN	Stat	Stat	NN
Head	Stat	NN	Stat	Stat	ML
Heart	Stat	Stat	Stat	Stat	Stat
Belg	Stat	Stat	NN	NN	Stat
Segm	Stat	ML	ML	ML	NN
Diab	Stat	NN	Stat	Stat	NN
Cr.Ger	Stat	Stat	Stat	Stat	NN
Chrom	Stat	NN	Stat	NN	Stat
Cr.Aus	ML	ML	Stat	Stat	NN
Shutt	ML	ML	ML	ML	ML
DNA	NN	NN	Stat	Stat	Stat
Tech	ML	ML	ML	ML	ML
NewBel	Stat	ML	ML	ML	ML
ISoft	ML	NN	Stat	Stat	NN
Tset	ML	ML	ML	ML	ML
cut20	ML	Stat	ML	Stat	ML
cut50	Stat	ML	Stat	ML	ML
Cr.Man	Stat	NN	ML	ML	NN
letter	Stat	Stat	NN	Stat	ML

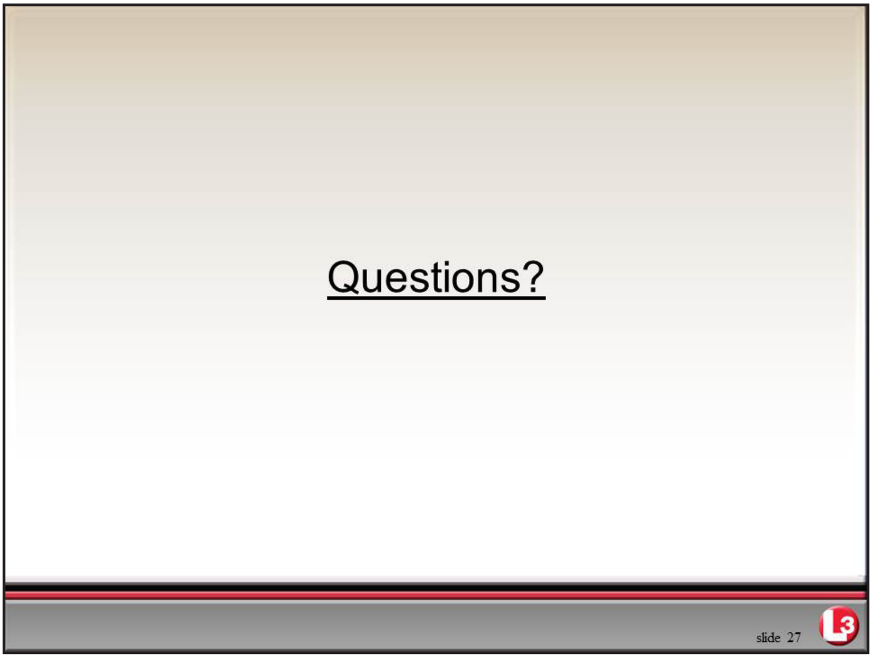
- •13 Algorithms had a first place result.
- •5 Algs had a least a second place
- •3 Algs had at least a third place
- •Only 1 Alg did not manage a top 5 finish
- •“There is no silver bullet”

.



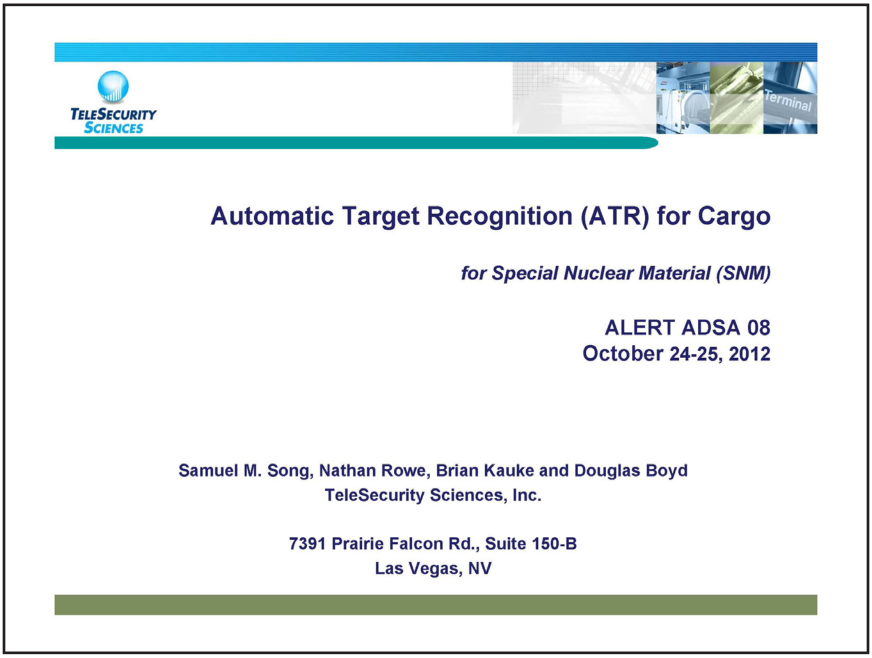
Threats

- Need a well developed system to evaluate the algorithm performance.
- Need to decide when is an alarm counted as a detection (varies by regulator).
- Start to implement version control for algorithm and system software.
- Schedule test. Process can be months to years.



Questions?

16.5 Sam Song: ATR for Cargo



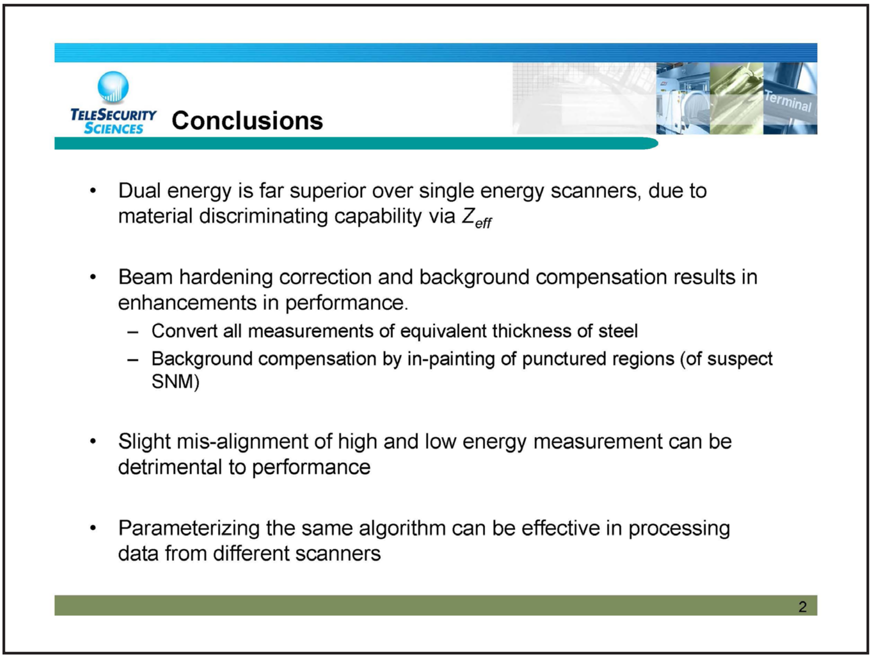
Automatic Target Recognition (ATR) for Cargo

for Special Nuclear Material (SNM)

ALERT ADSA 08
October 24-25, 2012

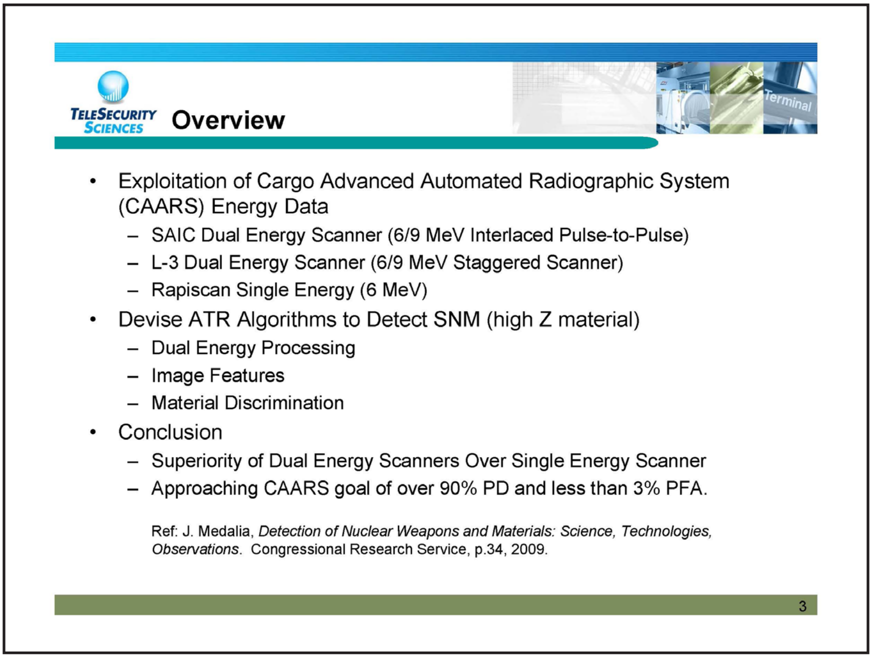
Samuel M. Song, Nathan Rowe, Brian Kauke and Douglas Boyd
TeleSecurity Sciences, Inc.

7391 Prairie Falcon Rd., Suite 150-B
Las Vegas, NV



Conclusions

- Dual energy is far superior over single energy scanners, due to material discriminating capability via Z_{eff}
- Beam hardening correction and background compensation results in enhancements in performance.
  - Convert all measurements of equivalent thickness of steel
  - Background compensation by in-painting of punctured regions (of suspect SNM)
- Slight mis-alignment of high and low energy measurement can be detrimental to performance
- Parameterizing the same algorithm can be effective in processing data from different scanners

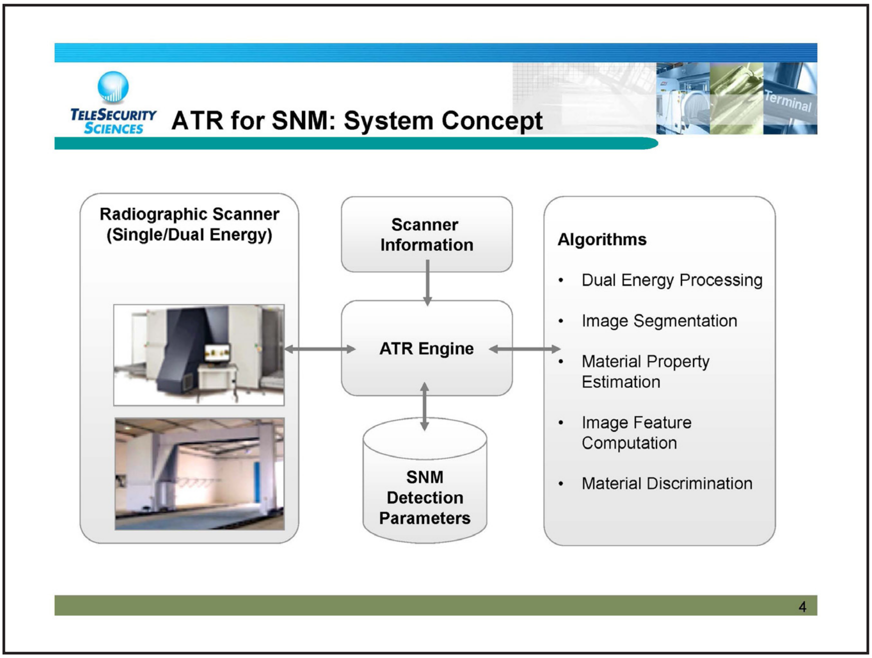


TELESECURITY SCIENCES

Overview

- Exploitation of Cargo Advanced Automated Radiographic System (CAARS) Energy Data
  - SAIC Dual Energy Scanner (6/9 MeV Interlaced Pulse-to-Pulse)
  - L-3 Dual Energy Scanner (6/9 MeV Staggered Scanner)
  - Rapiscan Single Energy (6 MeV)
- Devise ATR Algorithms to Detect SNM (high Z material)
  - Dual Energy Processing
  - Image Features
  - Material Discrimination
- Conclusion
  - Superiority of Dual Energy Scanners Over Single Energy Scanner
  - Approaching CAARS goal of over 90% PD and less than 3% PFA.

Ref: J. Medalia, *Detection of Nuclear Weapons and Materials: Science, Technologies, Observations*. Congressional Research Service, p.34, 2009.



TELESECURITY SCIENCES

ATR for SNM: System Concept

Radiographic Scanner (Single/Dual Energy)

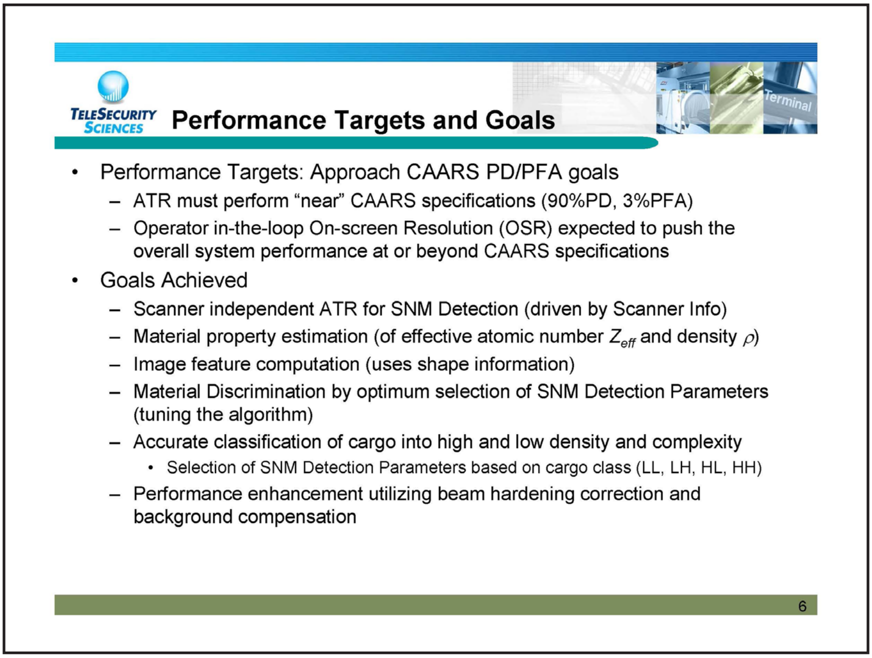
Scanner Information

ATR Engine

SNM Detection Parameters

Algorithms

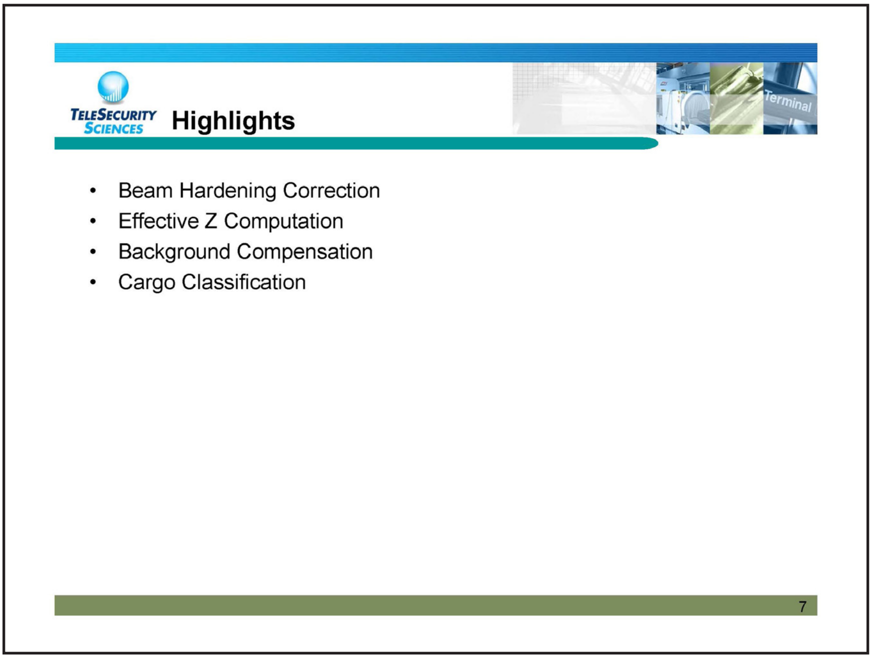
- Dual Energy Processing
- Image Segmentation
- Material Property Estimation
- Image Feature Computation
- Material Discrimination



TeleSecurity Sciences

Performance Targets and Goals

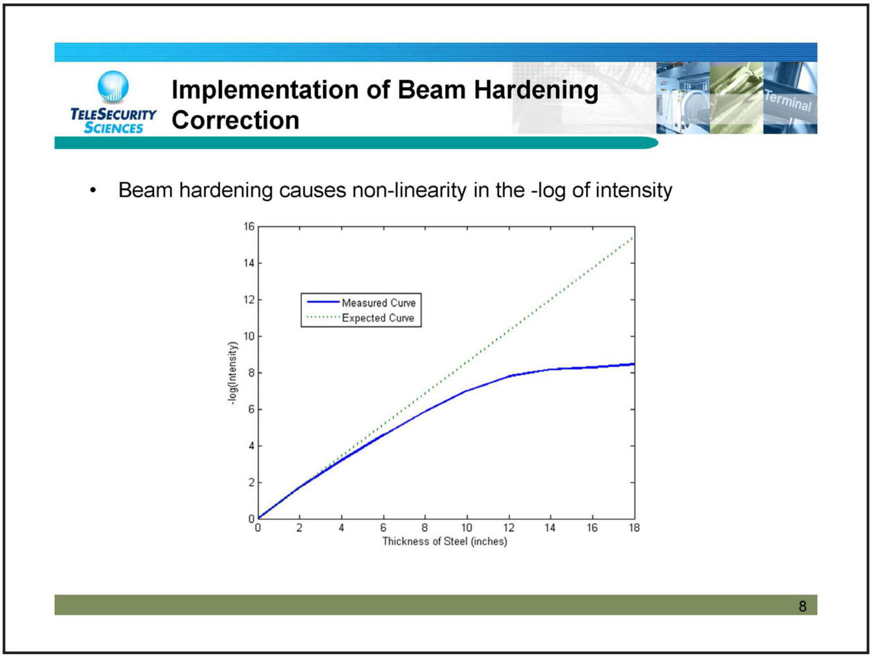
- Performance Targets: Approach CAARS PD/PFA goals
  - ATR must perform "near" CAARS specifications (90%PD, 3%PFA)
  - Operator in-the-loop On-screen Resolution (OSR) expected to push the overall system performance at or beyond CAARS specifications
- Goals Achieved
  - Scanner independent ATR for SNM Detection (driven by Scanner Info)
  - Material property estimation (of effective atomic number Z_{eff} and density ρ)
  - Image feature computation (uses shape information)
  - Material Discrimination by optimum selection of SNM Detection Parameters (tuning the algorithm)
  - Accurate classification of cargo into high and low density and complexity
    - Selection of SNM Detection Parameters based on cargo class (LL, LH, HL, HH)
  - Performance enhancement utilizing beam hardening correction and background compensation



TeleSecurity Sciences

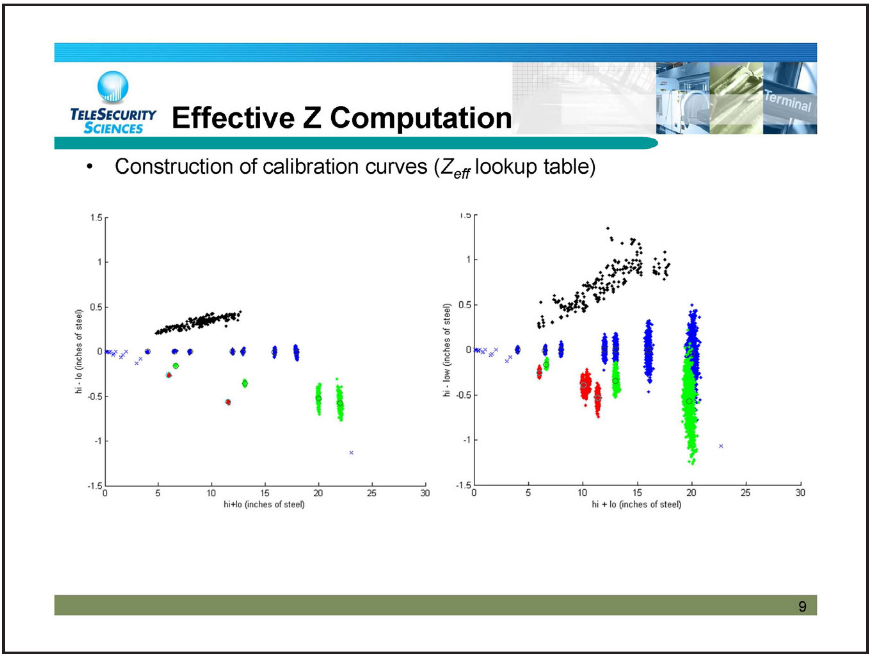
Highlights

- Beam Hardening Correction
- Effective Z Computation
- Background Compensation
- Cargo Classification



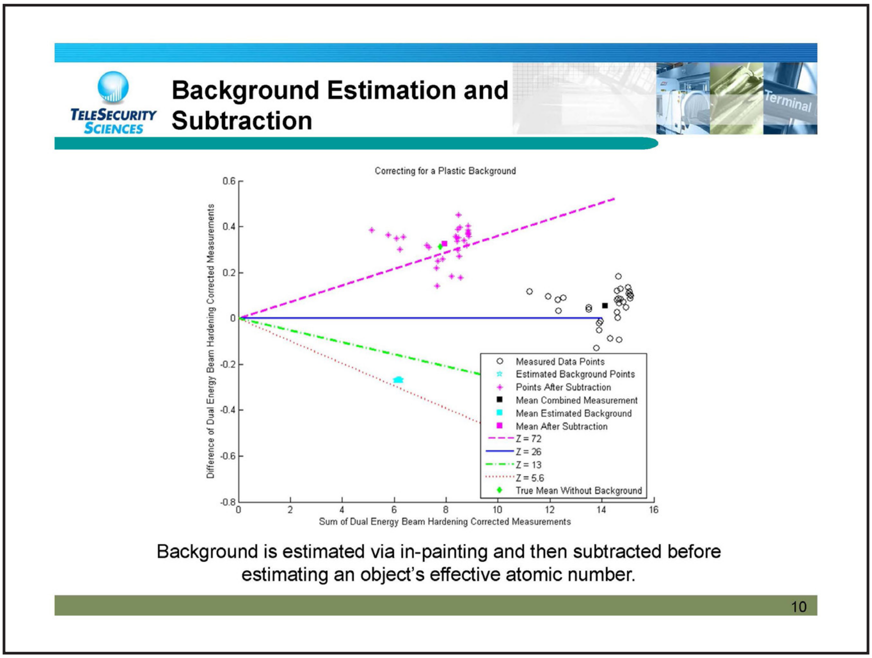
Implementation of Beam Hardening Correction

- Beam hardening causes non-linearity in the -log of intensity



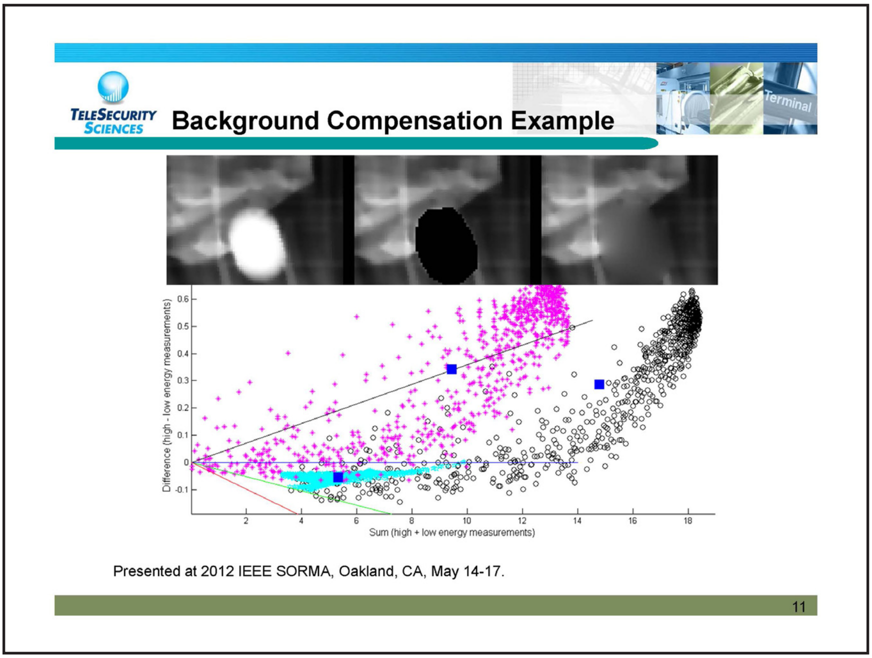
Effective Z Computation

- Construction of calibration curves (Z_{eff} lookup table)



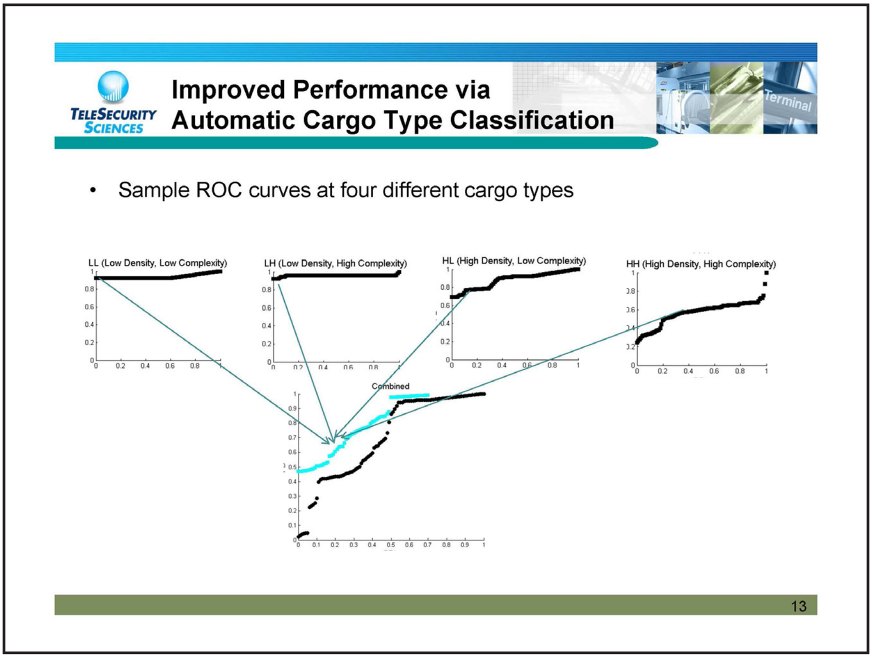
Background Estimation and Subtraction

Background is estimated via in-painting and then subtracted before estimating an object's effective atomic number.



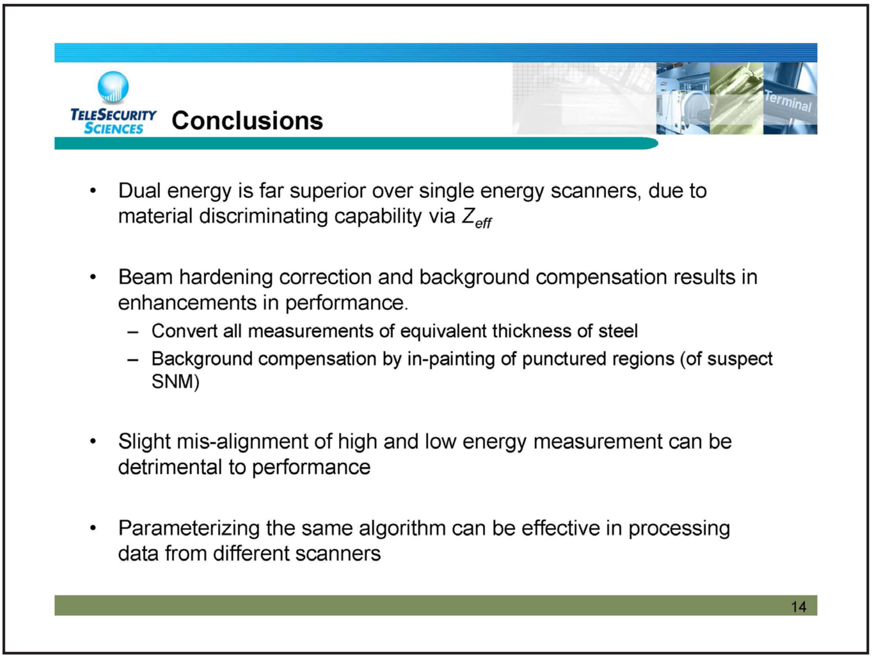
Background Compensation Example

Presented at 2012 IEEE SORMA, Oakland, CA, May 14-17.



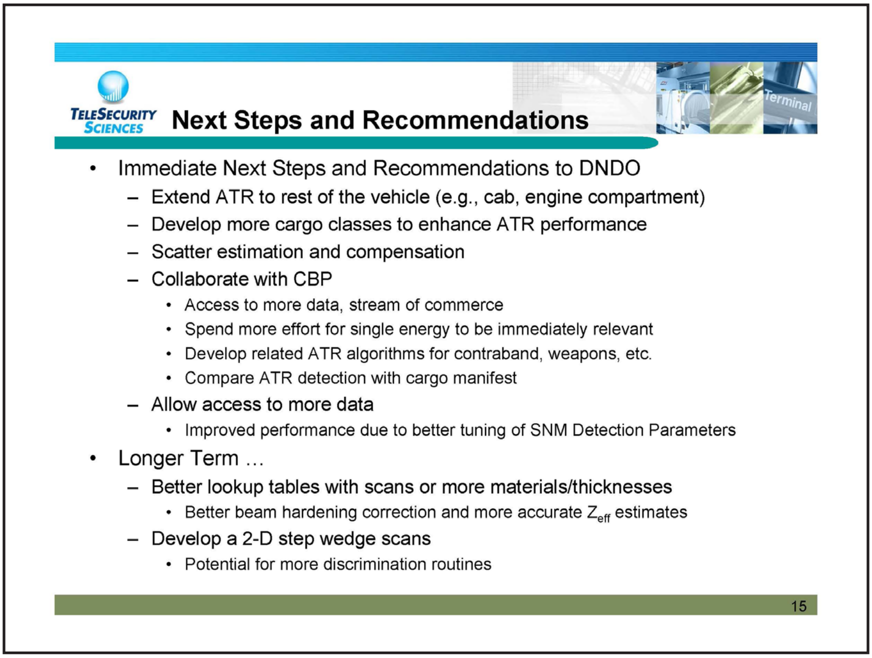
Improved Performance via Automatic Cargo Type Classification

- Sample ROC curves at four different cargo types



Conclusions

- Dual energy is far superior over single energy scanners, due to material discriminating capability via Z_{eff}
- Beam hardening correction and background compensation results in enhancements in performance.
  - Convert all measurements of equivalent thickness of steel
  - Background compensation by in-painting of punctured regions (of suspect SNM)
- Slight mis-alignment of high and low energy measurement can be detrimental to performance
- Parameterizing the same algorithm can be effective in processing data from different scanners

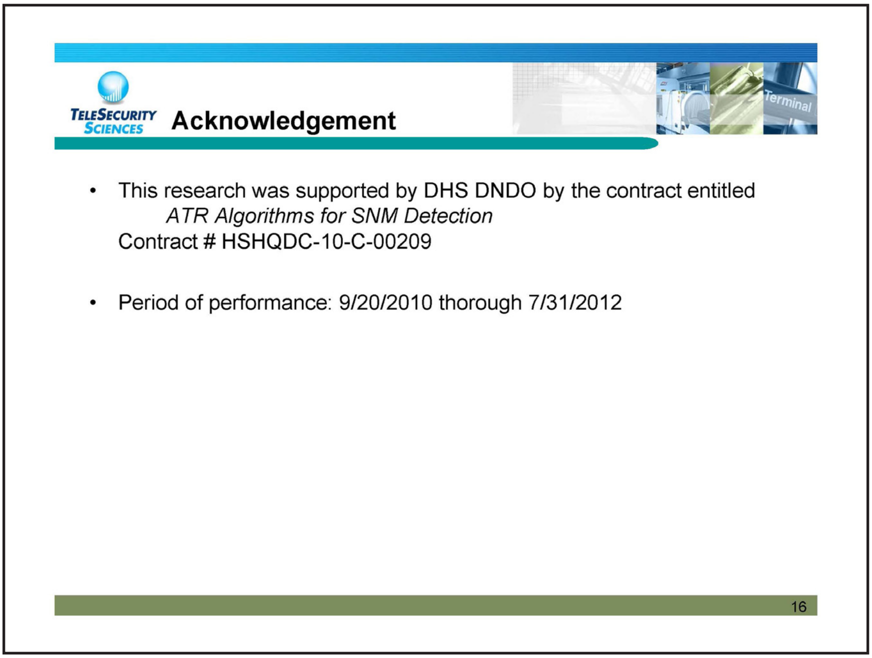


TeleSecurity Sciences

Next Steps and Recommendations

- Immediate Next Steps and Recommendations to DNDO
  - Extend ATR to rest of the vehicle (e.g., cab, engine compartment)
  - Develop more cargo classes to enhance ATR performance
  - Scatter estimation and compensation
  - Collaborate with CBP
    - Access to more data, stream of commerce
    - Spend more effort for single energy to be immediately relevant
    - Develop related ATR algorithms for contraband, weapons, etc.
    - Compare ATR detection with cargo manifest
  - Allow access to more data
    - Improved performance due to better tuning of SNM Detection Parameters
- Longer Term …
  - Better lookup tables with scans or more materials/thicknesses
    - Better beam hardening correction and more accurate Z_{eff} estimates
  - Develop a 2-D step wedge scans
    - Potential for more discrimination routines

15



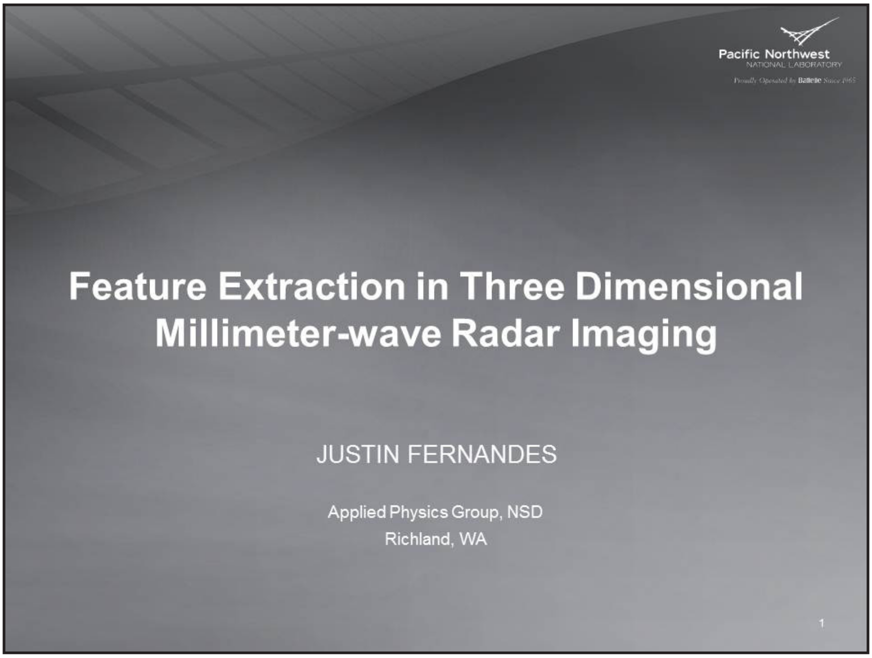
TeleSecurity Sciences

Acknowledgement

- This research was supported by DHS DNDO by the contract entitled
  ATR Algorithms for SNM Detection
  Contract # HSHQDC-10-C-00209
- Period of performance: 9/20/2010 thorough 7/31/2012

16

16.6 Justin Fernandes: Feature Extraction in 3D Millimeter-Wave Radio Imaging

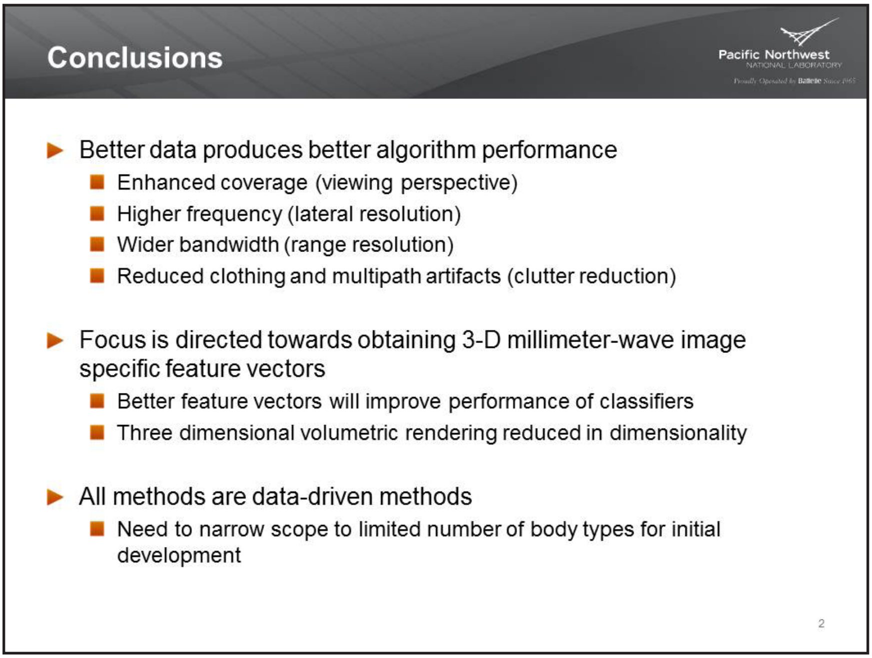


Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Feature Extraction in Three Dimensional Millimeter-wave Radar Imaging

JUSTIN FERNANDES

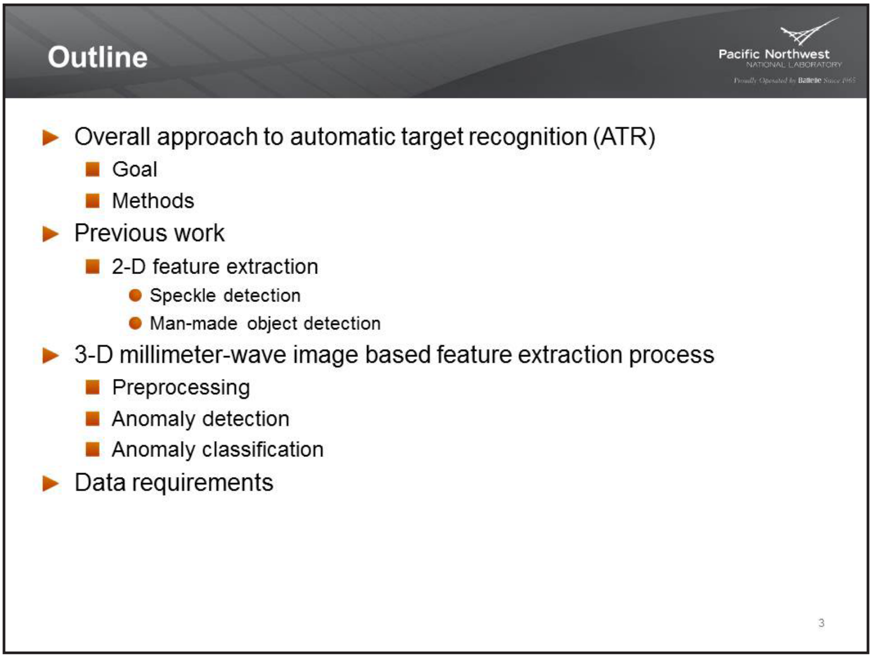
Applied Physics Group, NSD
Richland, WA



Conclusions

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

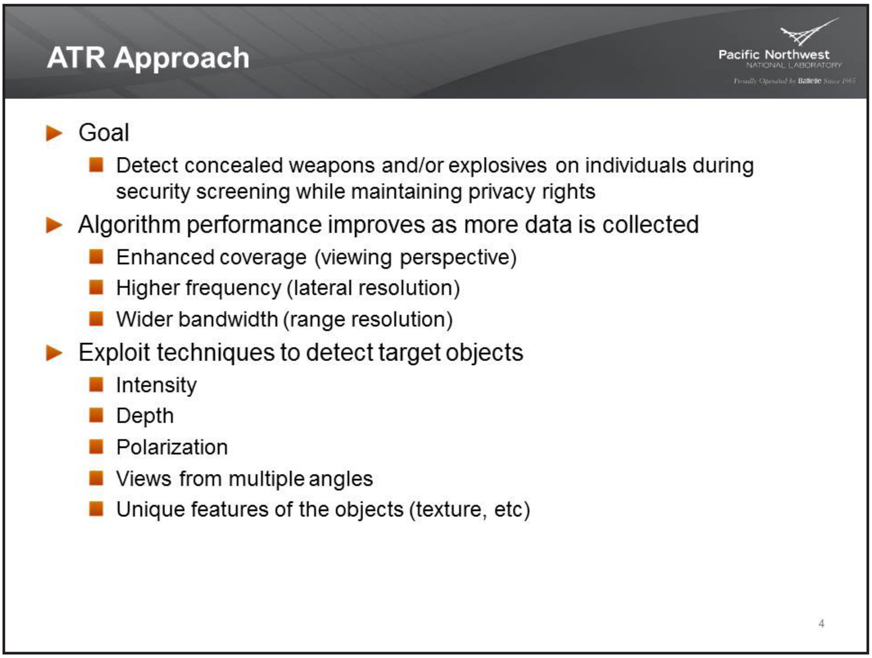
- Better data produces better algorithm performance
  - Enhanced coverage (viewing perspective)
  - Higher frequency (lateral resolution)
  - Wider bandwidth (range resolution)
  - Reduced clothing and multipath artifacts (clutter reduction)
- Focus is directed towards obtaining 3-D millimeter-wave image specific feature vectors
  - Better feature vectors will improve performance of classifiers
  - Three dimensional volumetric rendering reduced in dimensionality
- All methods are data-driven methods
  - Need to narrow scope to limited number of body types for initial development



Outline

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

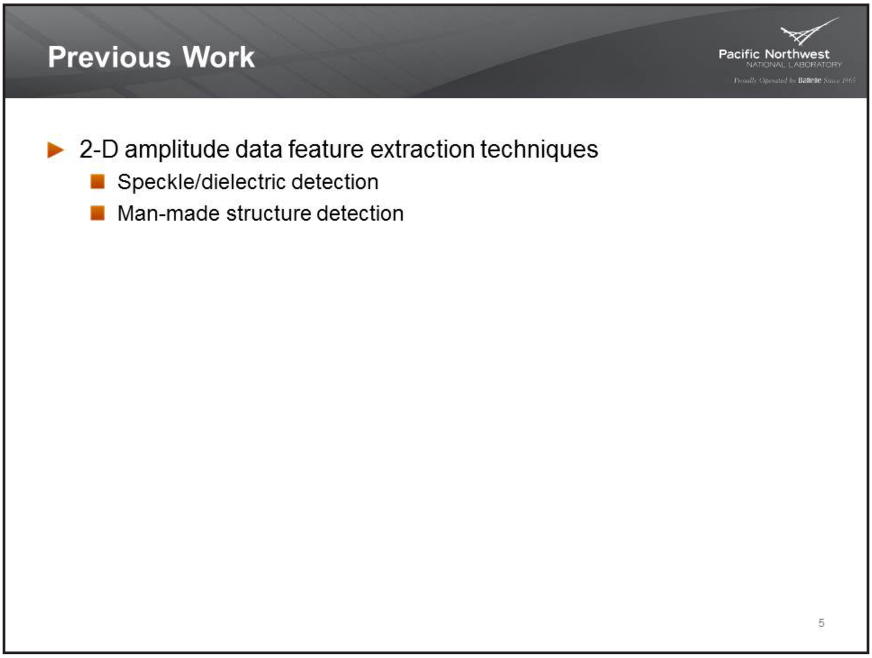
- Overall approach to automatic target recognition (ATR)
  - Goal
  - Methods
- Previous work
  - 2-D feature extraction
    - Speckle detection
    - Man-made object detection
- 3-D millimeter-wave image based feature extraction process
  - Preprocessing
  - Anomaly detection
  - Anomaly classification
- Data requirements



ATR Approach

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

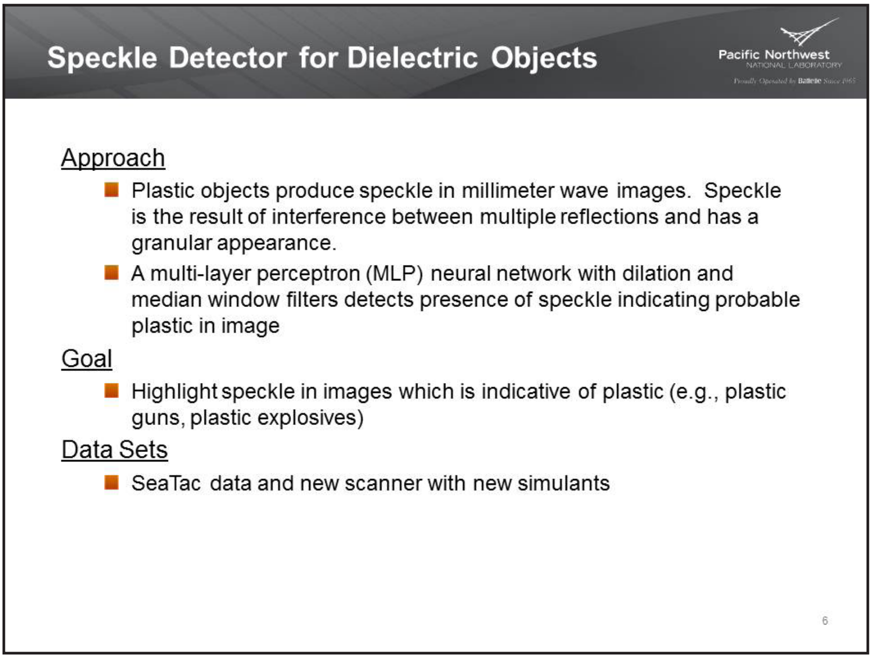
- Goal
  - Detect concealed weapons and/or explosives on individuals during security screening while maintaining privacy rights
- Algorithm performance improves as more data is collected
  - Enhanced coverage (viewing perspective)
  - Higher frequency (lateral resolution)
  - Wider bandwidth (range resolution)
- Exploit techniques to detect target objects
  - Intensity
  - Depth
  - Polarization
  - Views from multiple angles
  - Unique features of the objects (texture, etc)



Previous Work

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- 2-D amplitude data feature extraction techniques
  - Speckle/dielectric detection
  - Man-made structure detection



Speckle Detector for Dielectric Objects

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Approach

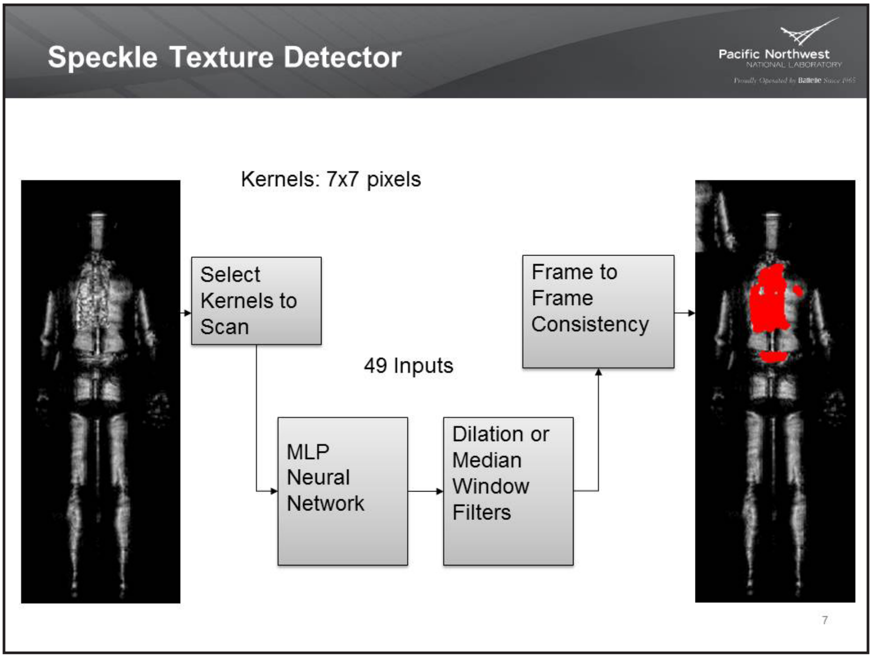
- Plastic objects produce speckle in millimeter wave images. Speckle is the result of interference between multiple reflections and has a granular appearance.
- A multi-layer perceptron (MLP) neural network with dilation and median window filters detects presence of speckle indicating probable plastic in image

Goal

- Highlight speckle in images which is indicative of plastic (e.g., plastic guns, plastic explosives)

Data Sets

- SeaTac data and new scanner with new simulants



Speckle Texture Detector

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Kernels: 7x7 pixels

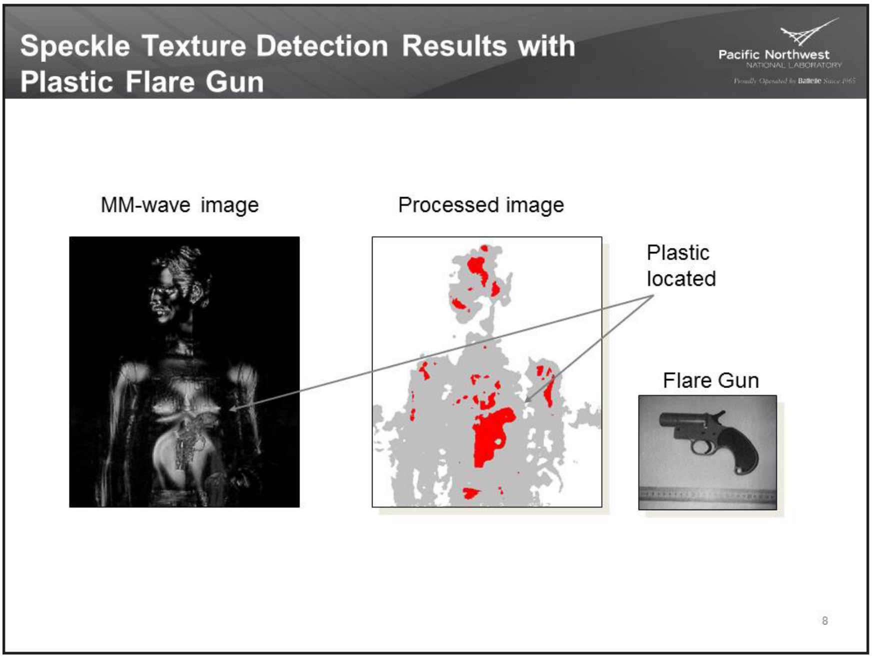
Select Kernels to Scan

49 Inputs

MLP Neural Network

Dilation or Median Window Filters

Frame to Frame Consistency



Speckle Texture Detection Results with Plastic Flare Gun

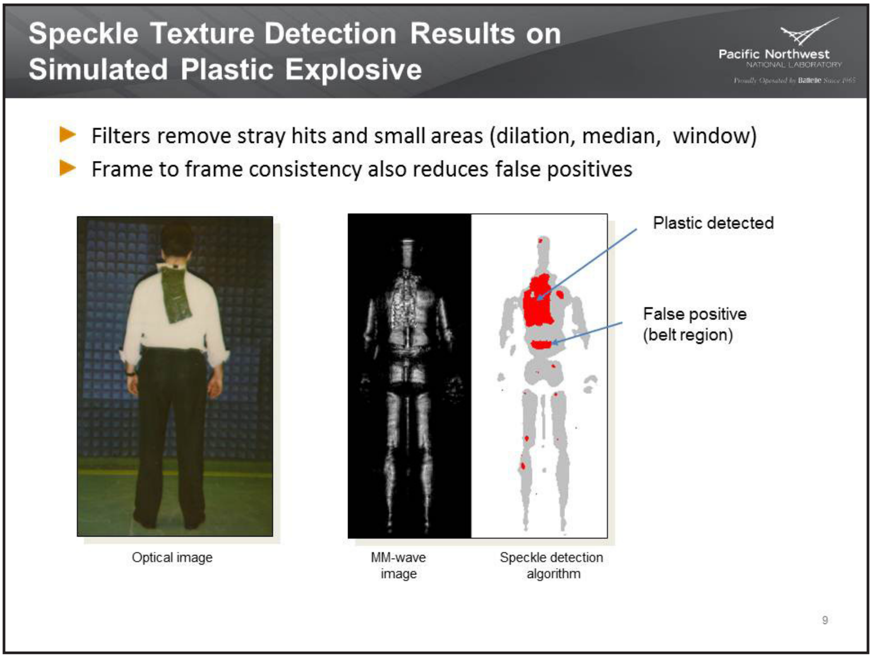
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

MM-wave image

Processed image

Plastic located

Flare Gun



Speckle Texture Detection Results on Simulated Plastic Explosive

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Filters remove stray hits and small areas (dilation, median, window)
- Frame to frame consistency also reduces false positives

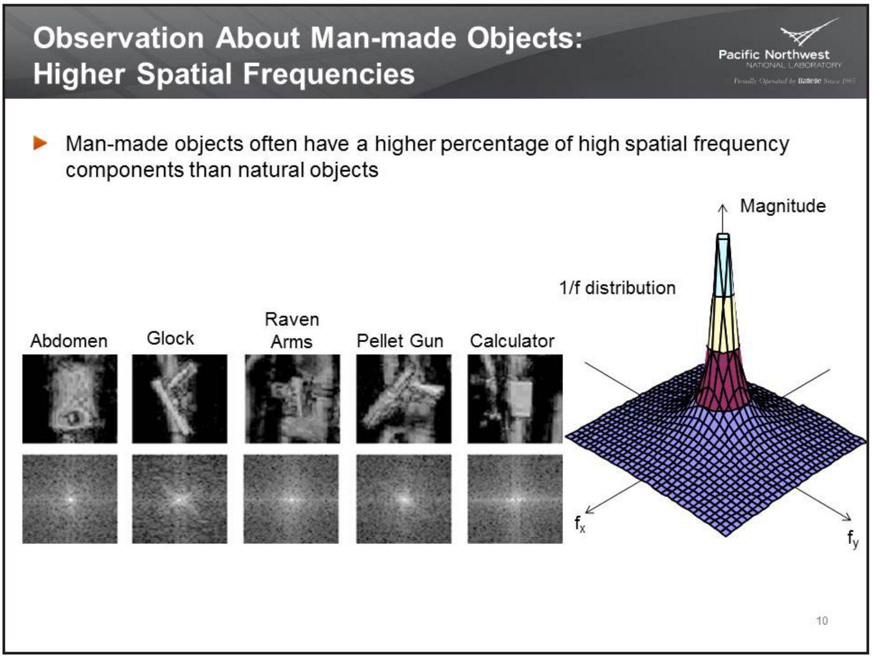
Plastic detected

False positive
(belt region)

Optical image

MM-wave
image

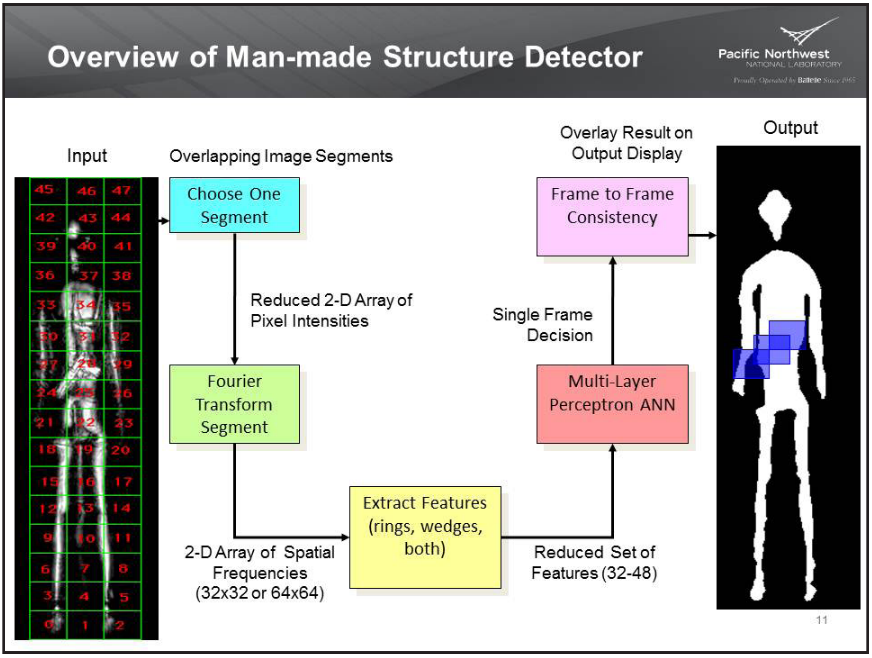
Speckle detection
algorithm



Observation About Man-made Objects: Higher Spatial Frequencies

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Man-made objects often have a higher percentage of high spatial frequency components than natural objects



Overview of Man-made Structure Detector

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Input

Overlapping Image Segments

Choose One Segment

Reduced 2-D Array of Pixel Intensities

Fourier Transform Segment

2-D Array of Spatial Frequencies (32x32 or 64x64)

Extract Features (rings, wedges, both)

Reduced Set of Features (32-48)

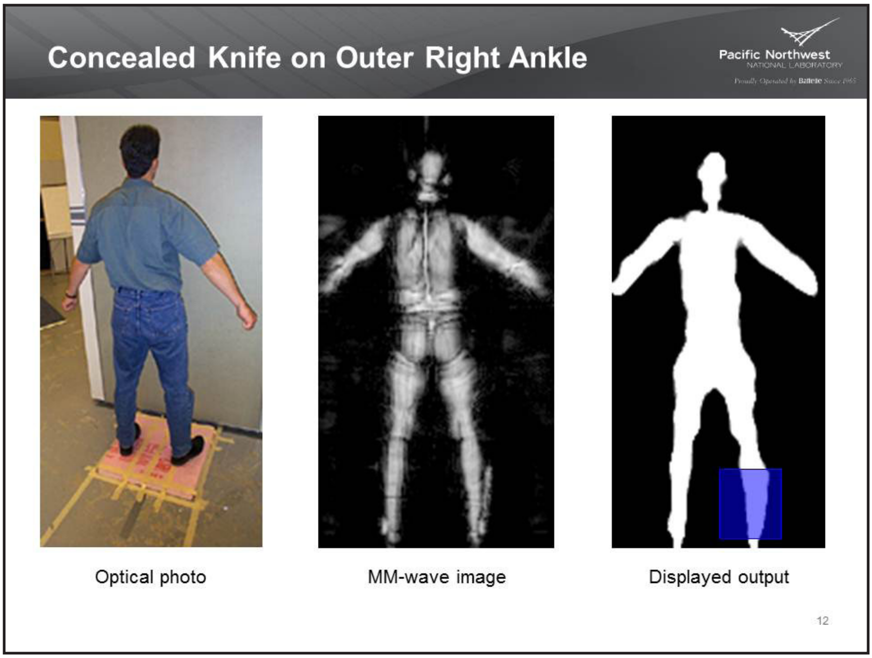
Multi-Layer Perceptron ANN

Single Frame Decision

Frame to Frame Consistency

Overlay Result on Output Display

Output



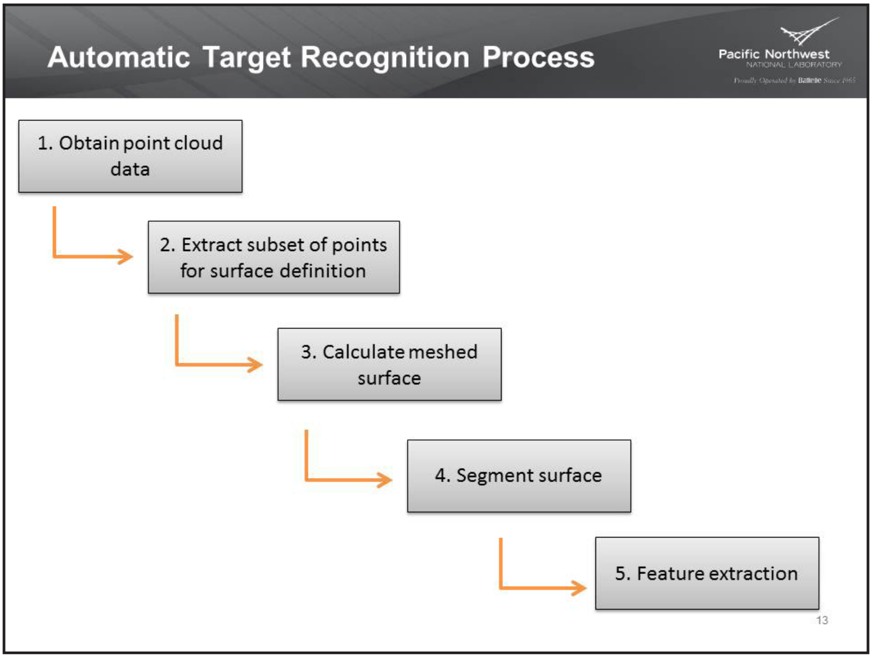
Concealed Knife on Outer Right Ankle

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Optical photo

MM-wave image

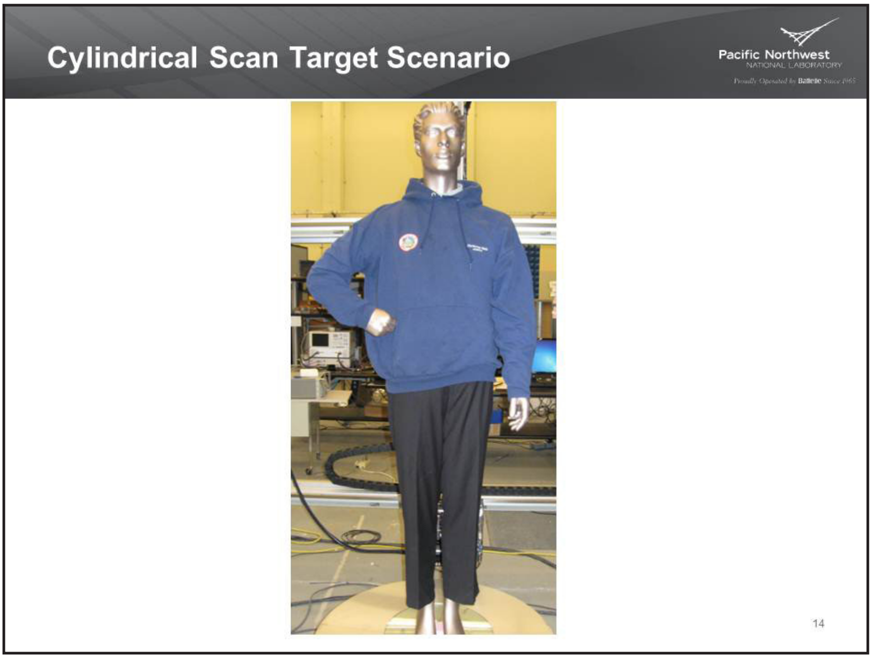
Displayed output



Automatic Target Recognition Process

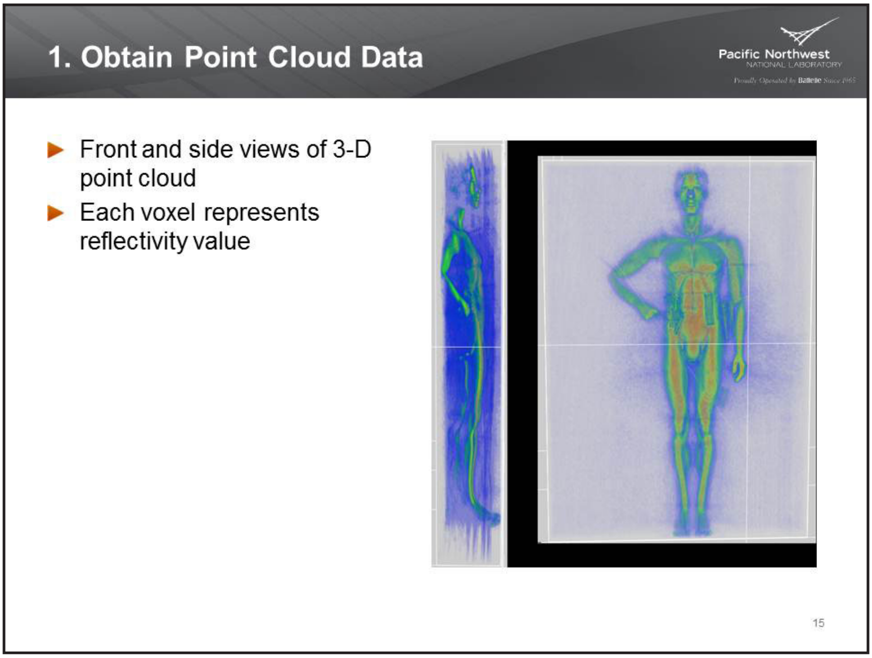
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

1. Obtain point cloud data
2. Extract subset of points for surface definition
3. Calculate meshed surface
4. Segment surface
5. Feature extraction



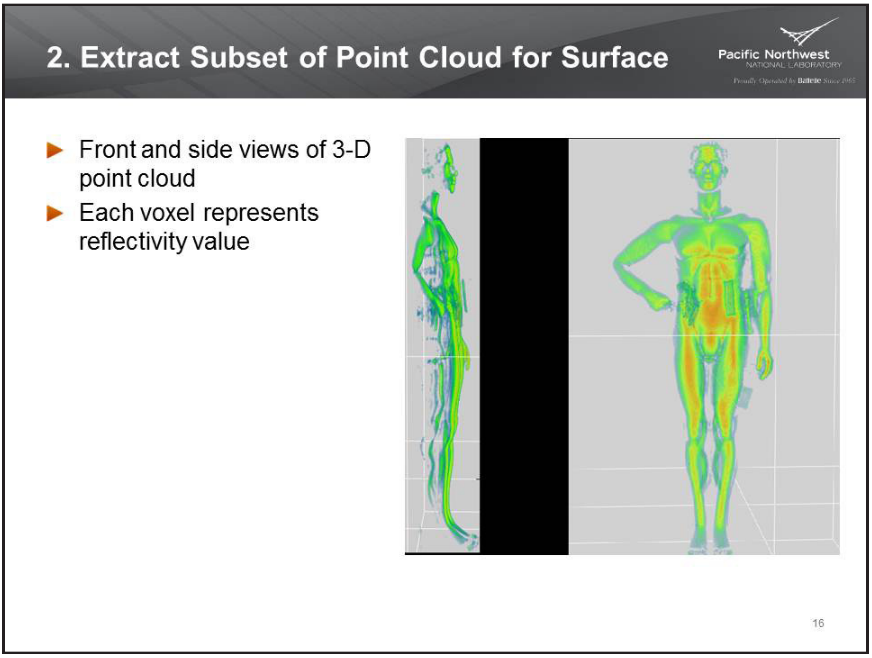
Cylindrical Scan Target Scenario

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



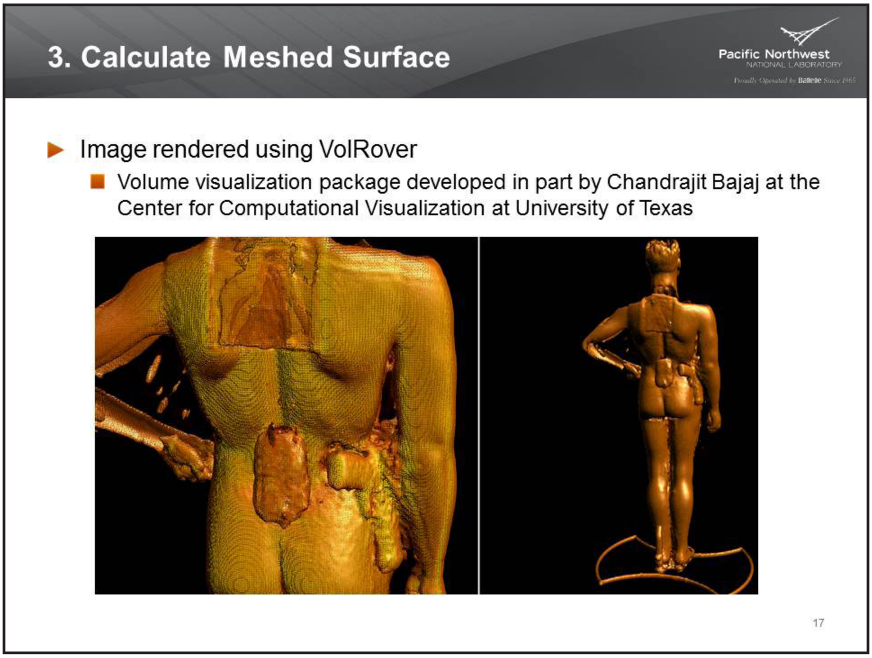
1. Obtain Point Cloud Data

- Front and side views of 3-D point cloud
- Each voxel represents reflectivity value



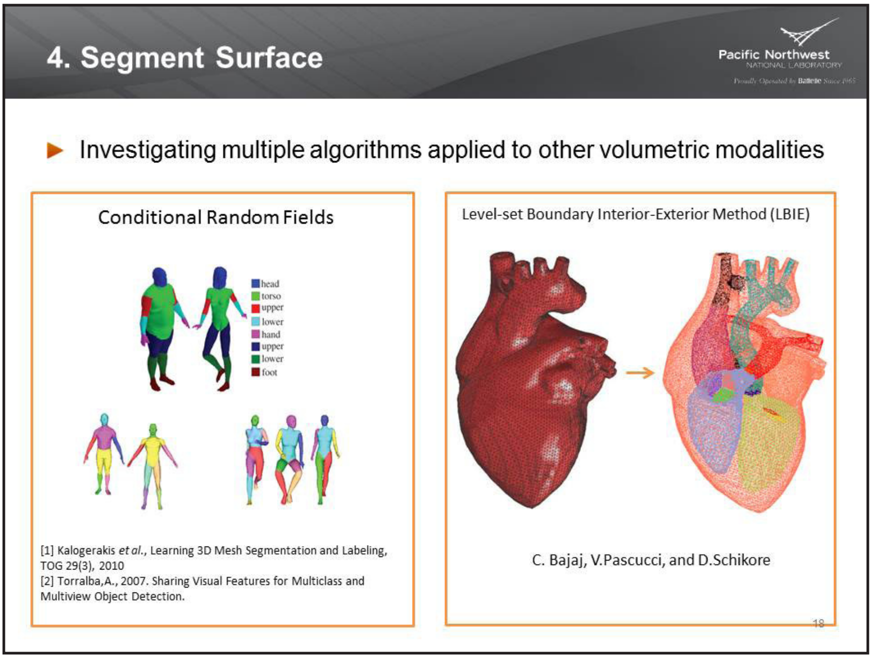
2. Extract Subset of Point Cloud for Surface

- Front and side views of 3-D point cloud
- Each voxel represents reflectivity value



3. Calculate Meshed Surface

- Image rendered using VolRover
  - Volume visualization package developed in part by Chandrajit Bajaj at the Center for Computational Visualization at University of Texas



4. Segment Surface

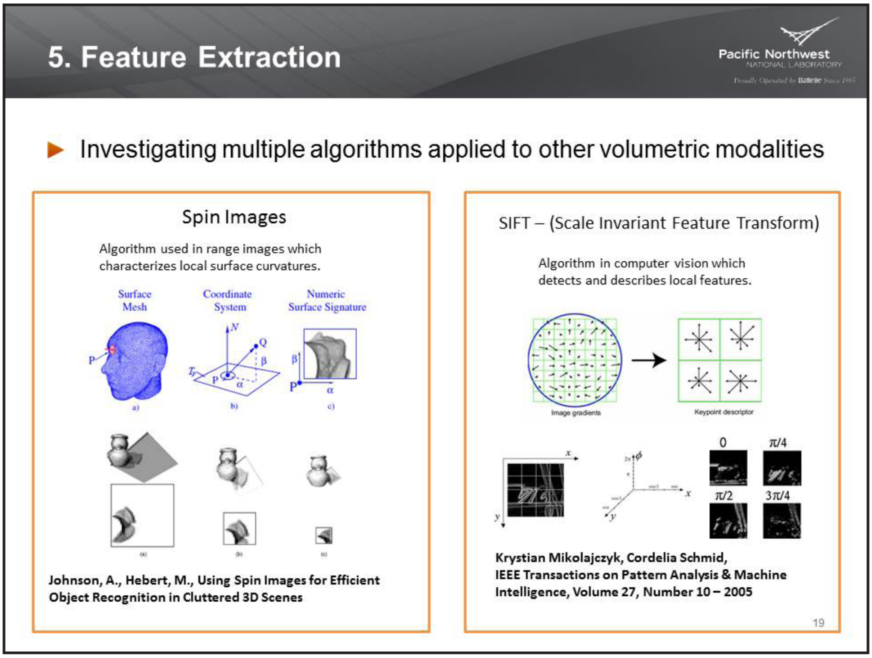
- Investigating multiple algorithms applied to other volumetric modalities

Conditional Random Fields

[1] Kalogerakis *et al.*, Learning 3D Mesh Segmentation and Labeling, TOG 29(3), 2010
[2] Torralba,A., 2007. Sharing Visual Features for Multiclass and Multiview Object Detection.

Level-set Boundary Interior-Exterior Method (LBIE)

C. Bajaj, V.Pascucci, and D.Schikore



5. Feature Extraction

Pacific Northwest NATIONAL LABORATORY

- Investigating multiple algorithms applied to other volumetric modalities

Spin Images

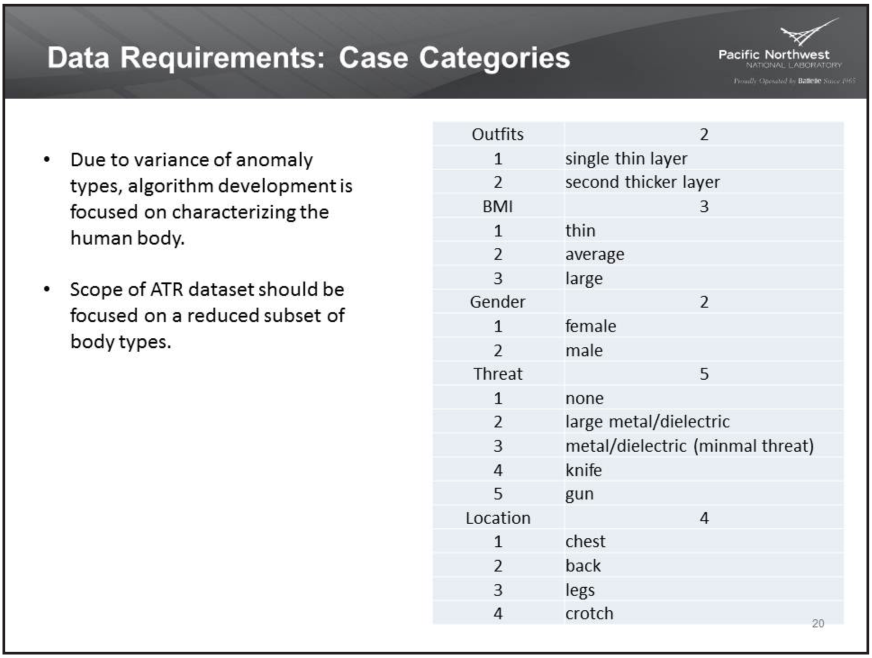
Algorithm used in range images which characterizes local surface curvatures.

Johnson, A., Hebert, M., Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes

SIFT – (Scale Invariant Feature Transform)

Algorithm in computer vision which detects and describes local features.

Krystian Mikolajczyk, Cordelia Schmid, IEEE Transactions on Pattern Analysis & Machine Intelligence, Volume 27, Number 10 – 2005

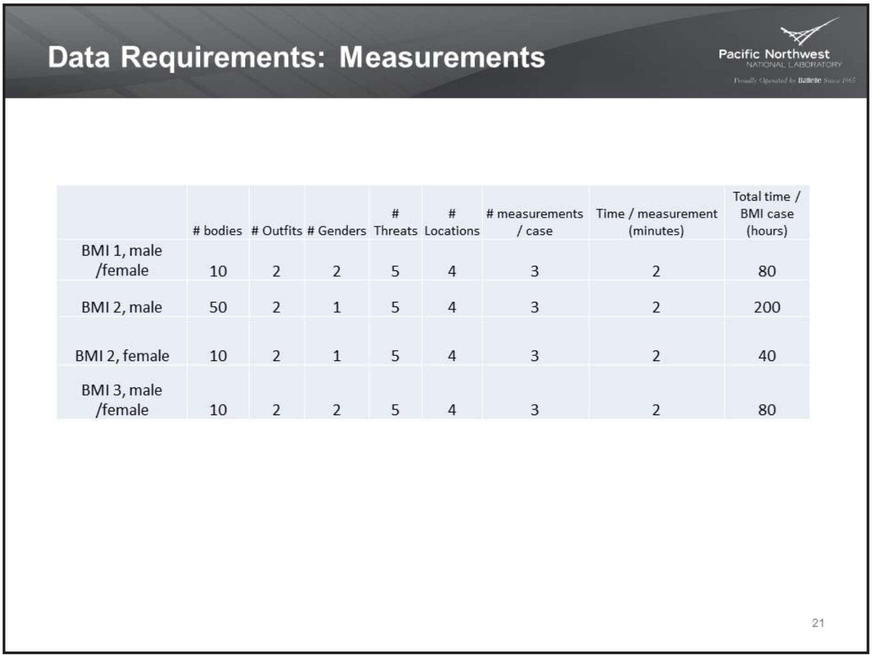


Data Requirements: Case Categories

Pacific Northwest NATIONAL LABORATORY

- Due to variance of anomaly types, algorithm development is focused on characterizing the human body.
- Scope of ATR dataset should be focused on a reduced subset of body types.

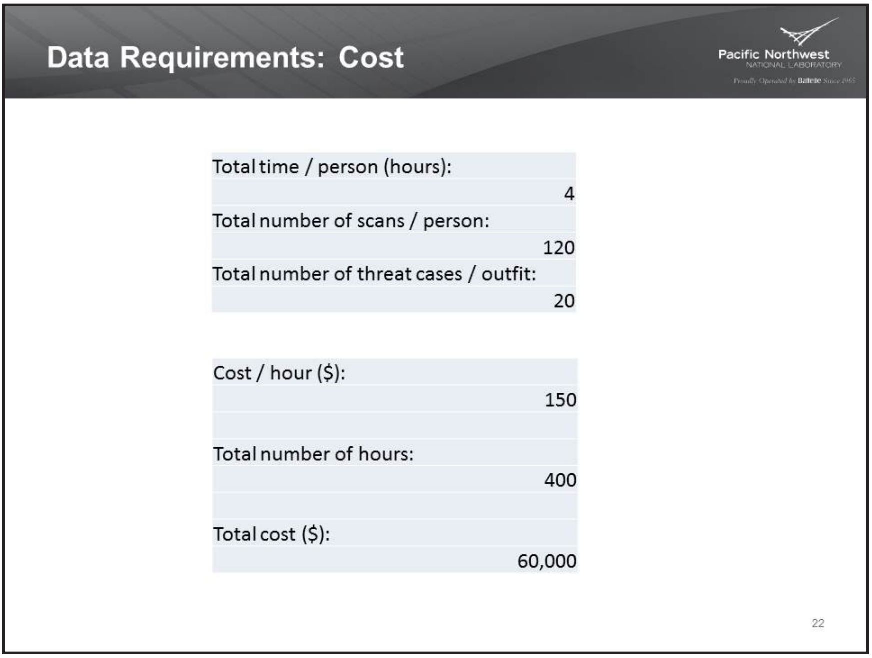
Outfits	2
1	single thin layer
2	second thicker layer
BMI	3
1	thin
2	average
3	large
Gender	2
1	female
2	male
Threat	5
1	none
2	large metal/dielectric
3	metal/dielectric (minmal threat)
4	knife
5	gun
Location	4
1	chest
2	back
3	legs
4	crotch



Data Requirements: Measurements

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

	# bodies	# Outfits	# Genders	# Threats	# Locations	# measurements / case	Time / measurement (minutes)	Total time / BMI case (hours)
BMI 1, male /female	10	2	2	5	4	3	2	80
BMI 2, male	50	2	1	5	4	3	2	200
BMI 2, female	10	2	1	5	4	3	2	40
BMI 3, male /female	10	2	2	5	4	3	2	80

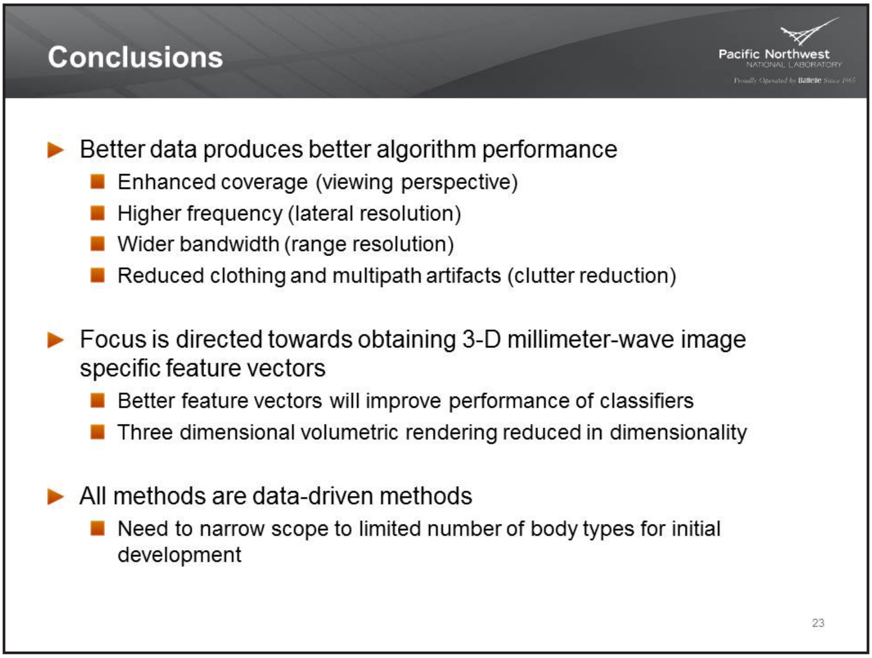


Data Requirements: Cost

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Total time / person (hours):	4
Total number of scans / person:	120
Total number of threat cases / outfit:	20

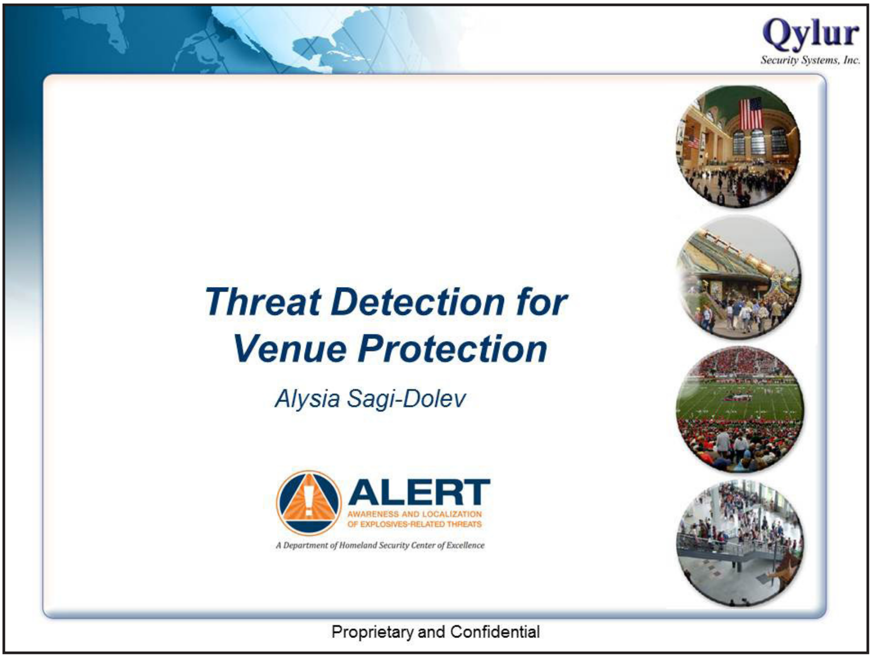
Cost / hour ($):	150
Total number of hours:	400
Total cost ($):	60,000



Conclusions

- Better data produces better algorithm performance
  - Enhanced coverage (viewing perspective)
  - Higher frequency (lateral resolution)
  - Wider bandwidth (range resolution)
  - Reduced clothing and multipath artifacts (clutter reduction)
- Focus is directed towards obtaining 3-D millimeter-wave image specific feature vectors
  - Better feature vectors will improve performance of classifiers
  - Three dimensional volumetric rendering reduced in dimensionality
- All methods are data-driven methods
  - Need to narrow scope to limited number of body types for initial development

16.7 Lisa Sagi-Dolev: Threat Detection for Venue Protection



Qylur
Security Systems, Inc.

Threat Detection for Venue Protection

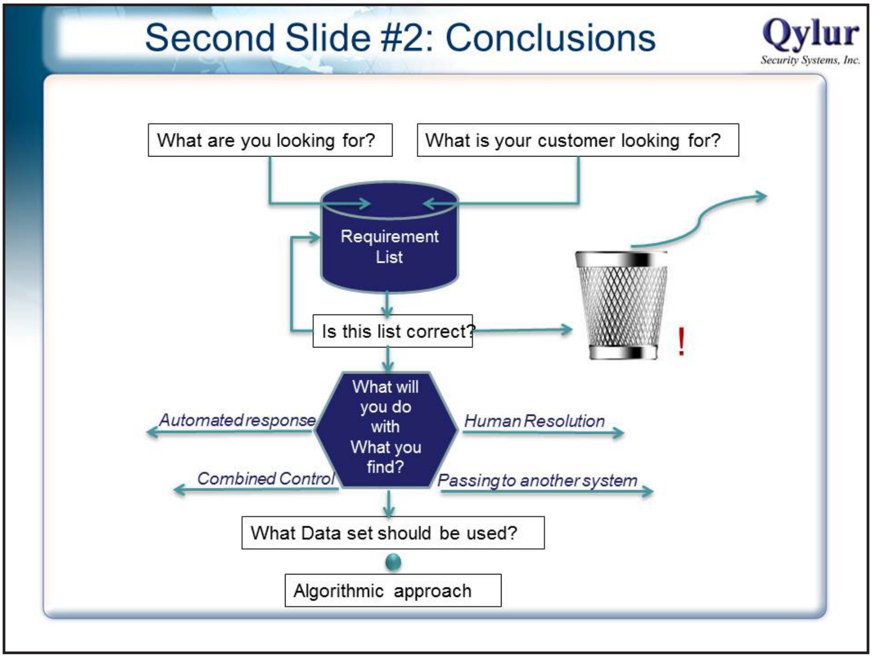
Alysia Sagi-Dolev

ALERT
AWARENESS AND LOCALIZATION OF EXPLOSIVES-RELATED THREATS
A Department of Homeland Security Center of Excellence

Proprietary and Confidential

Qylur
Security Systems, Inc.

Automated Self Service Security Checkpoint Powered by QyFuse
An *Automated Fused Threat Detection Algorithm*



Second Slide #2: Conclusions

Qylur
Security Systems, Inc.

What are you looking for?

What is your customer looking for?

Requirement List

Is this list correct?

What will you do with What you find?

Automated response

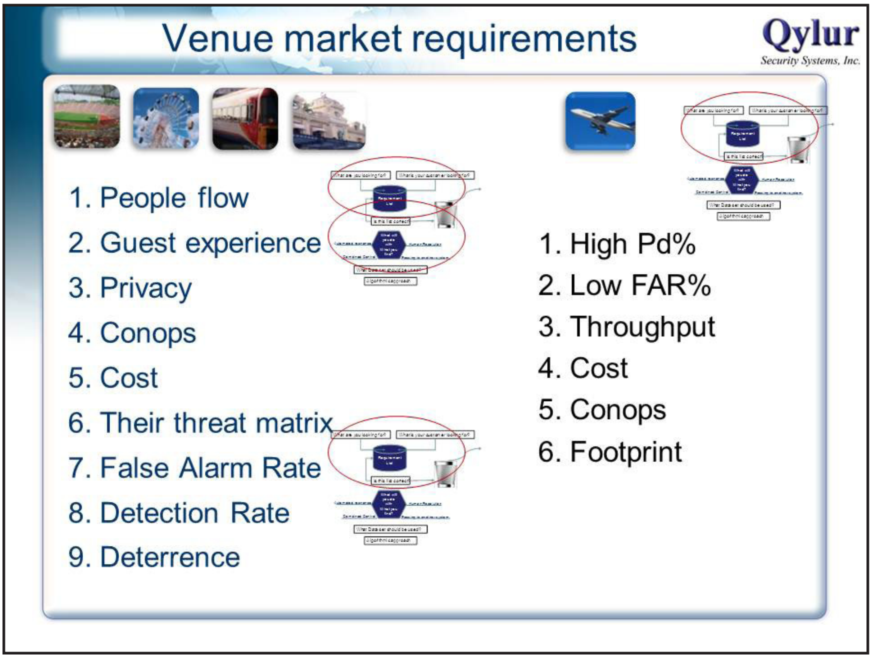
Human Resolution

Combined Control

Passing to another system

What Data set should be used?

Algorithmic approach

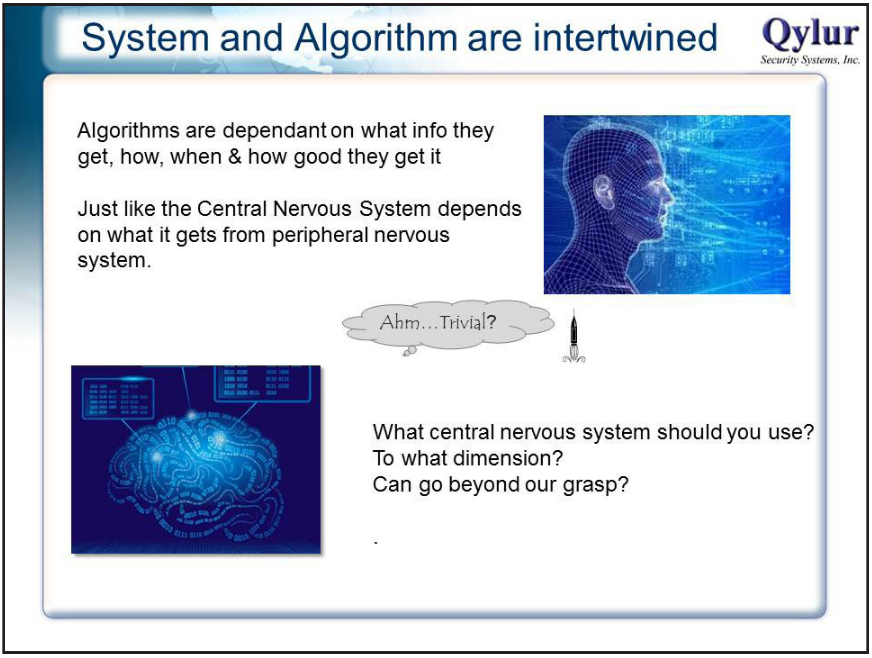


Venue market requirements

Qylur
Security Systems, Inc.

1. People flow
2. Guest experience
3. Privacy
4. Conops
5. Cost
6. Their threat matrix
7. False Alarm Rate
8. Detection Rate
9. Deterrence

1. High Pd%
2. Low FAR%
3. Throughput
4. Cost
5. Conops
6. Footprint



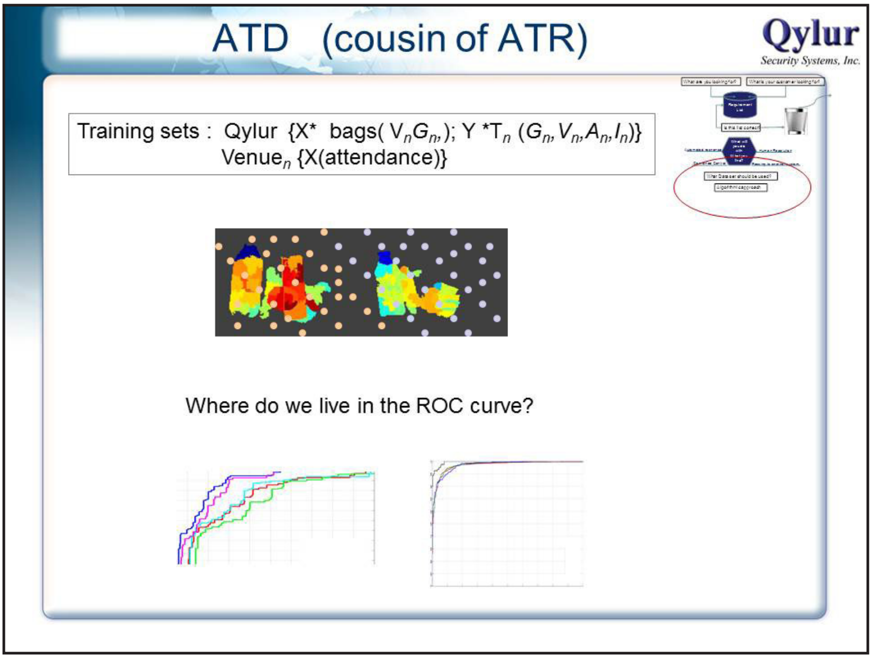
System and Algorithm are intertwined

Qylur
Security Systems, Inc.

Algorithms are dependant on what info they get, how, when & how good they get it

Just like the Central Nervous System depends on what it gets from peripheral nervous system.

What central nervous system should you use?
To what dimension?
Can go beyond our grasp?

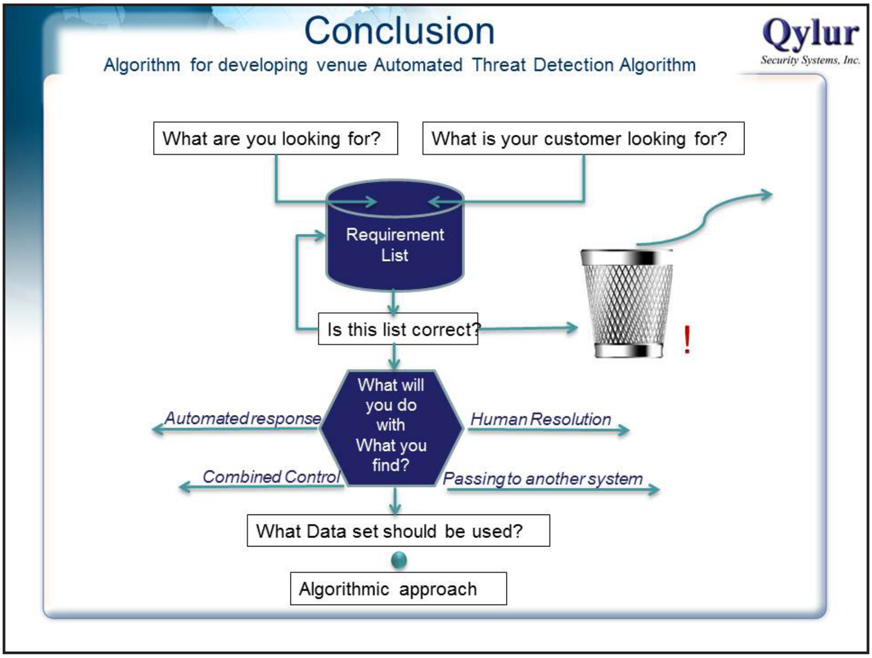


ATD (cousin of ATR)

Qylur
Security Systems, Inc.

Training sets : Qylur {X* bags($V_n G_n$,); Y *T_n (G_n, V_n, A_n, I_n)}
Venue$_n$ {X(attendance)}

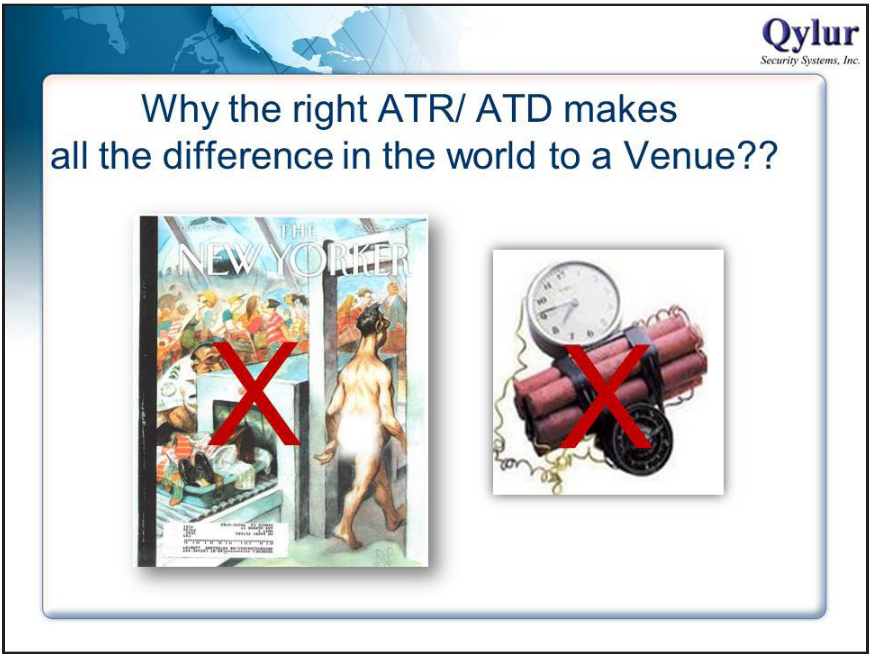
Where do we live in the ROC curve?



Conclusion

Algorithm for developing venue Automated Threat Detection Algorithm

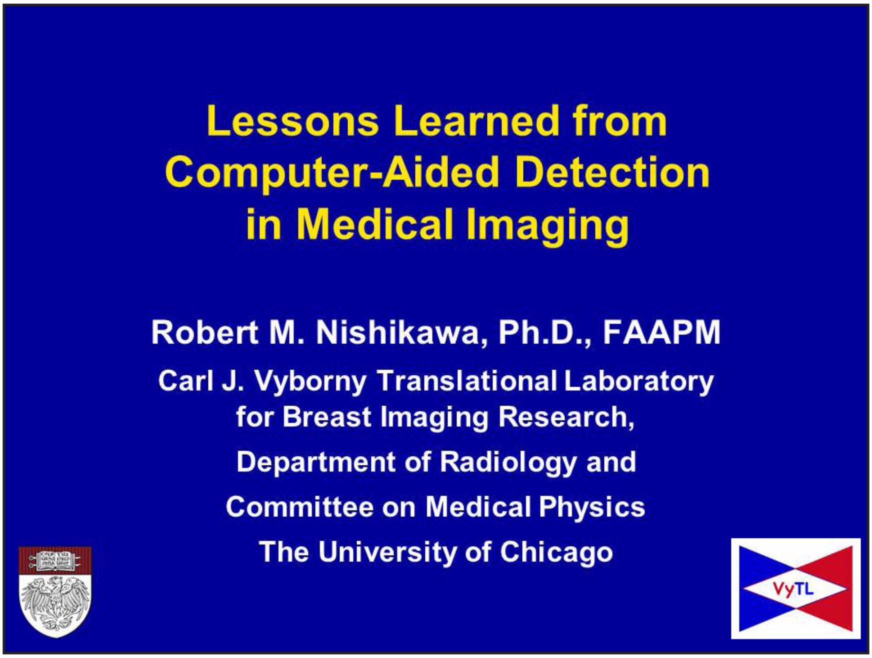
Qylur
Security Systems, Inc.



Qylur
Security Systems, Inc.

Why the right ATR/ ATD makes all the difference in the world to a Venue??

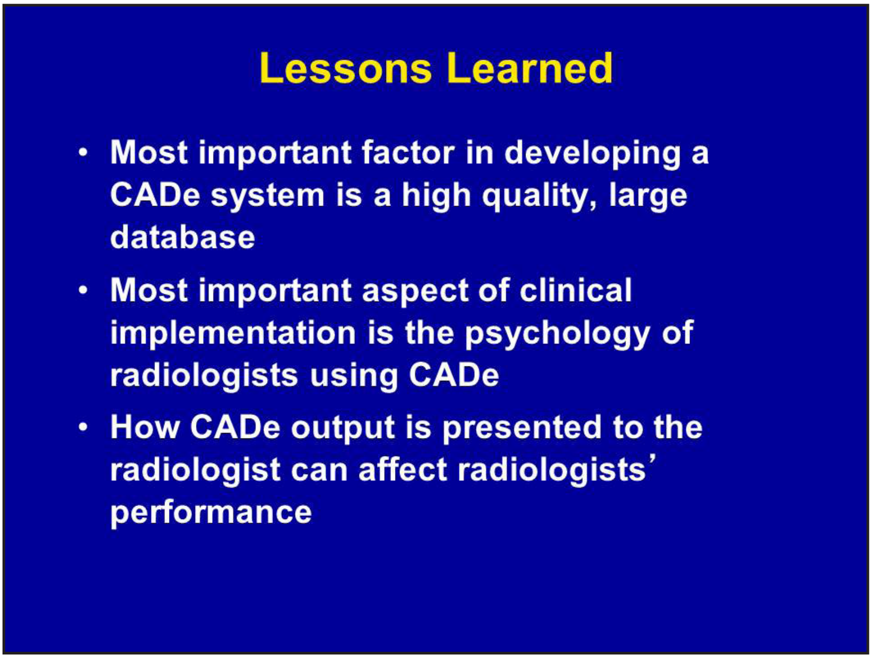
16.8 Robert Nishikawa: Computer Aided Detection in Medical Imaging



Lessons Learned from Computer-Aided Detection in Medical Imaging

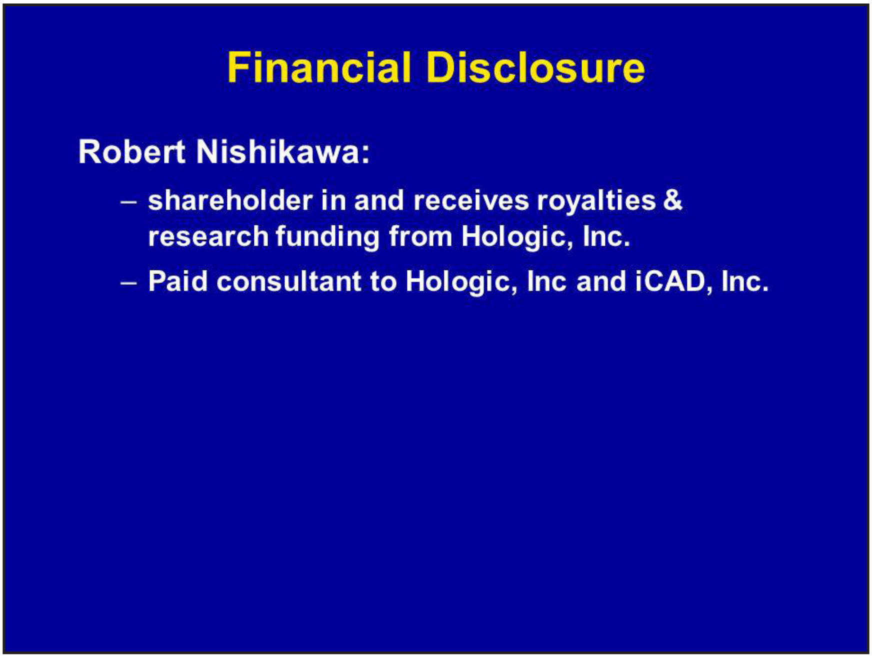
Robert M. Nishikawa, Ph.D., FAAPM

Carl J. Vyborny Translational Laboratory
for Breast Imaging Research,
Department of Radiology and
Committee on Medical Physics
The University of Chicago



Lessons Learned

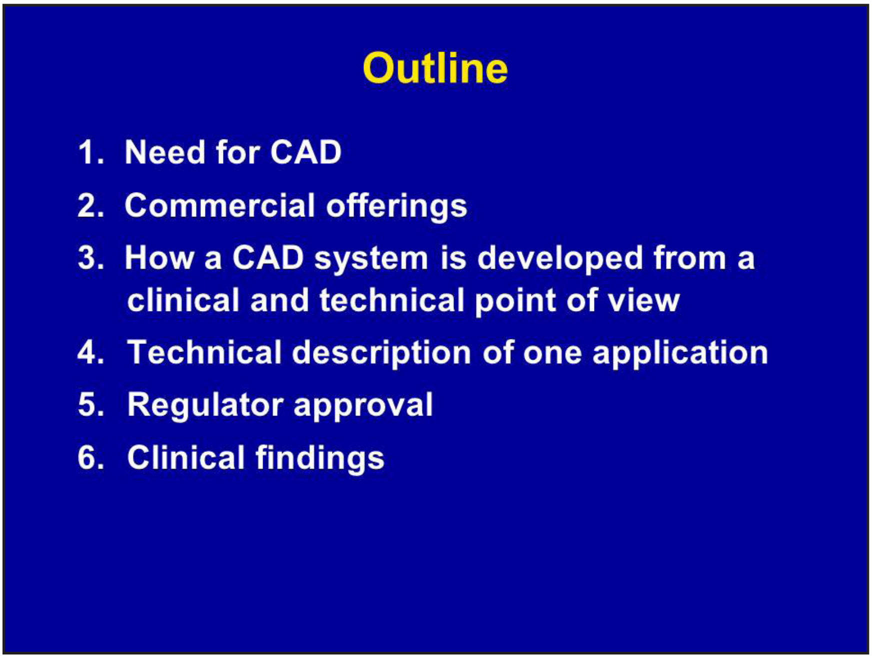
- Most important factor in developing a CADe system is a high quality, large database
- Most important aspect of clinical implementation is the psychology of radiologists using CADe
- How CADe output is presented to the radiologist can affect radiologists' performance



Financial Disclosure

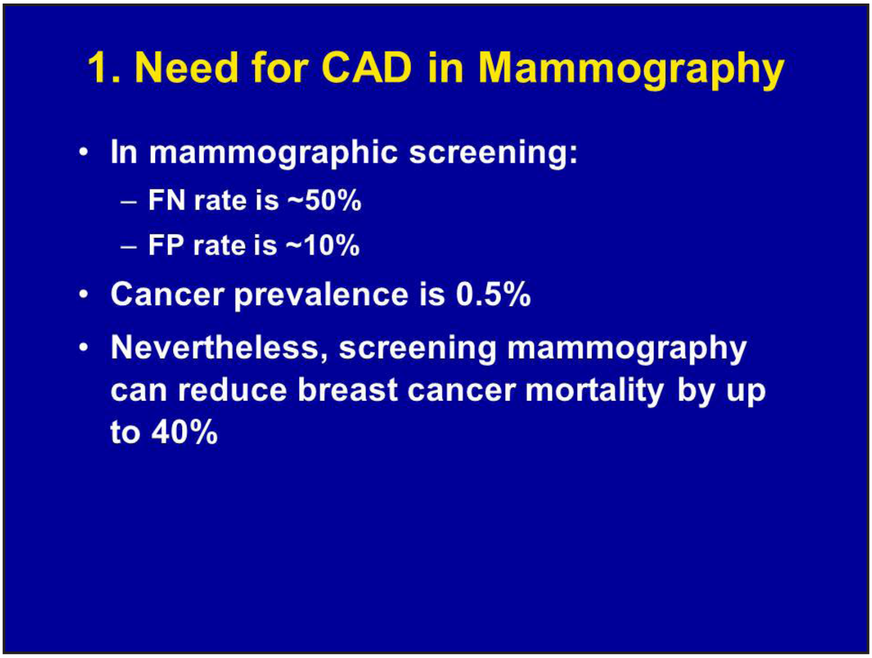
Robert Nishikawa:

- shareholder in and receives royalties & research funding from Hologic, Inc.
- Paid consultant to Hologic, Inc and iCAD, Inc.



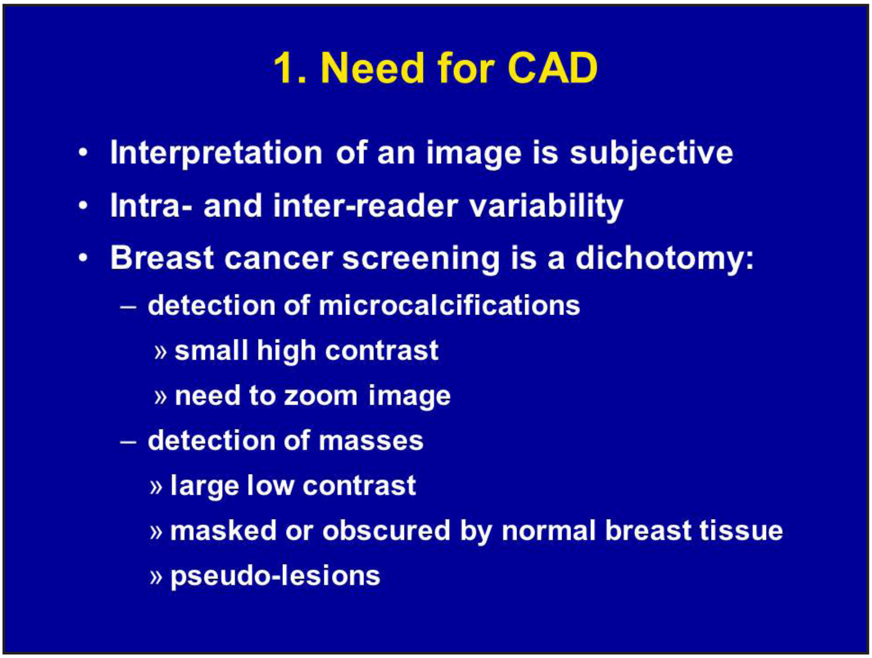
Outline

1. Need for CAD
2. Commercial offerings
3. How a CAD system is developed from a clinical and technical point of view
4. Technical description of one application
5. Regulator approval
6. Clinical findings



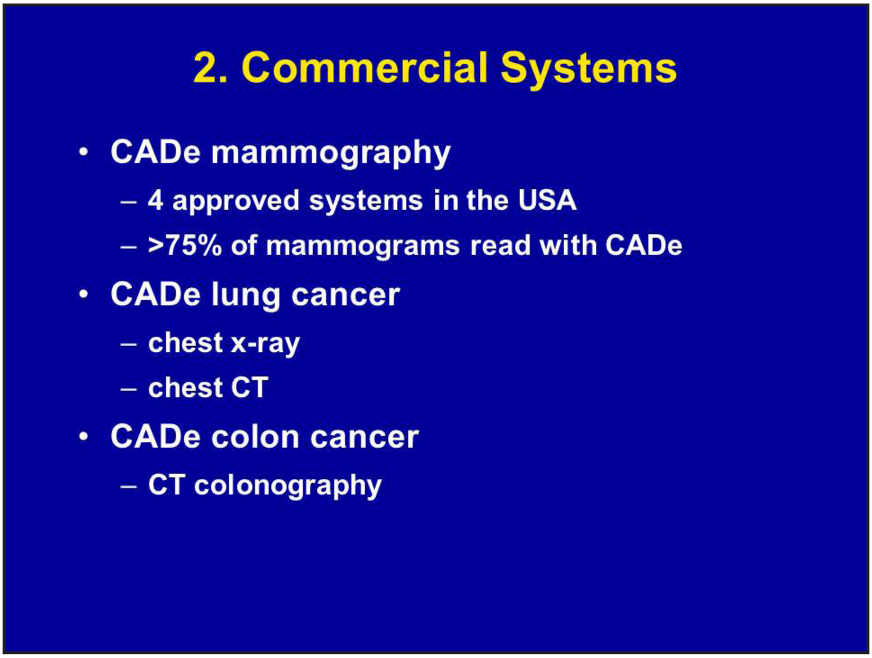
1. Need for CAD in Mammography

- In mammographic screening:
  - FN rate is ~50%
  - FP rate is ~10%
- Cancer prevalence is 0.5%
- Nevertheless, screening mammography can reduce breast cancer mortality by up to 40%



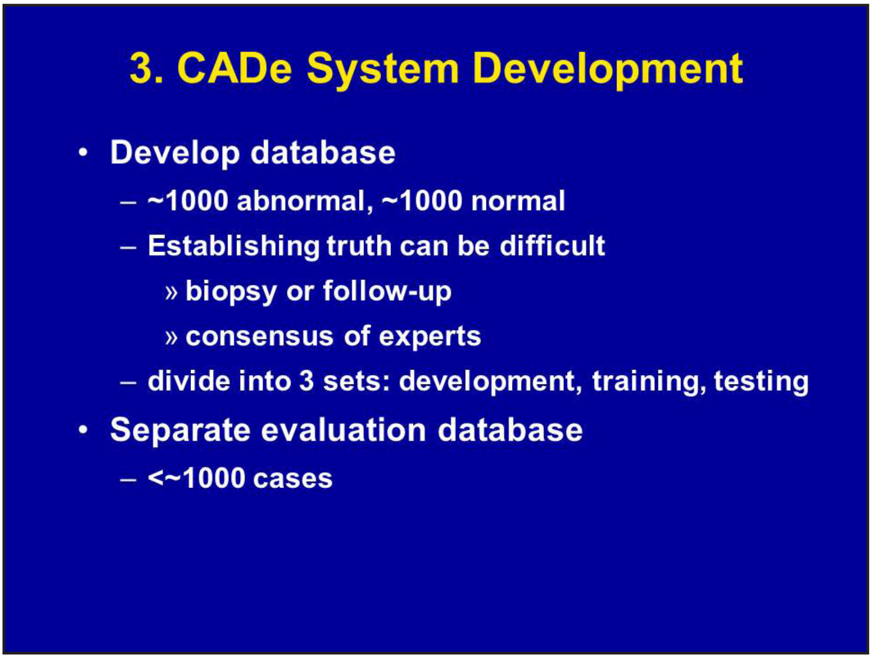
1. Need for CAD

- Interpretation of an image is subjective
- Intra- and inter-reader variability
- Breast cancer screening is a dichotomy:
  - detection of microcalcifications
    - » small high contrast
    - » need to zoom image
  - detection of masses
    - » large low contrast
    - » masked or obscured by normal breast tissue
    - » pseudo-lesions



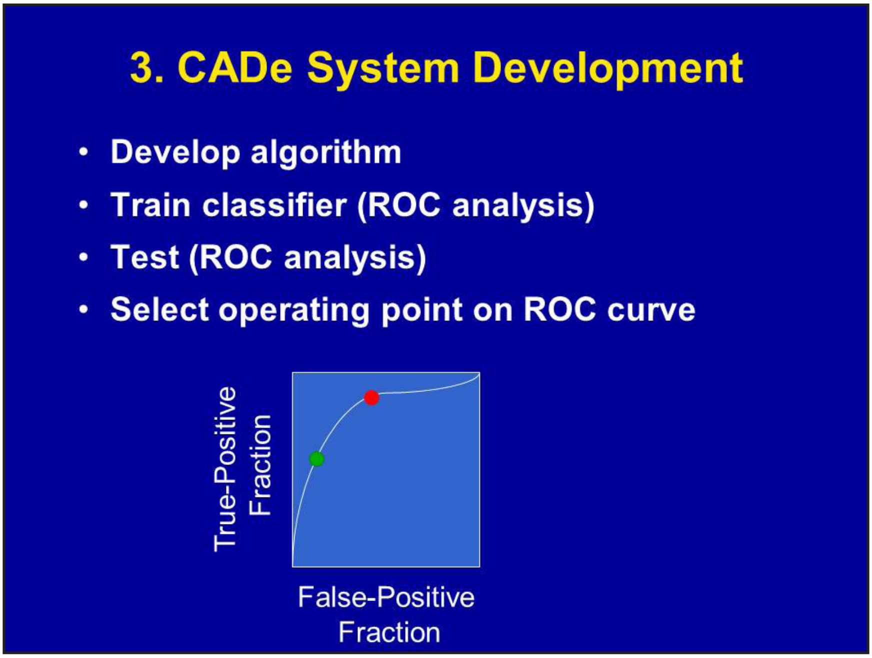
2. Commercial Systems

- CADe mammography
  - 4 approved systems in the USA
  - >75% of mammograms read with CADe
- CADe lung cancer
  - chest x-ray
  - chest CT
- CADe colon cancer
  - CT colonography



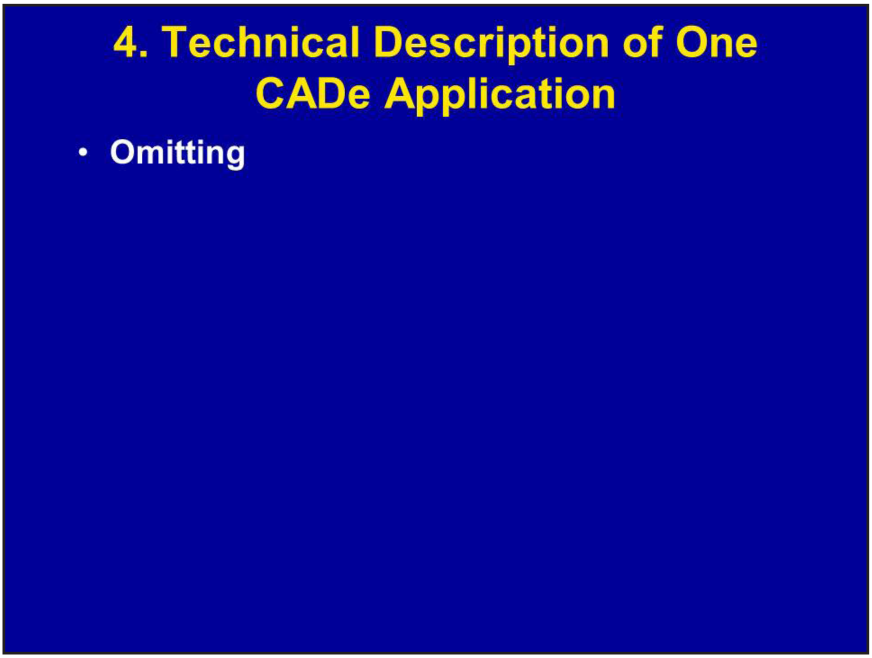
3. CADe System Development

- Develop database
  - ~1000 abnormal, ~1000 normal
  - Establishing truth can be difficult
    - » biopsy or follow-up
    - » consensus of experts
  - divide into 3 sets: development, training, testing
- Separate evaluation database
  - <~1000 cases



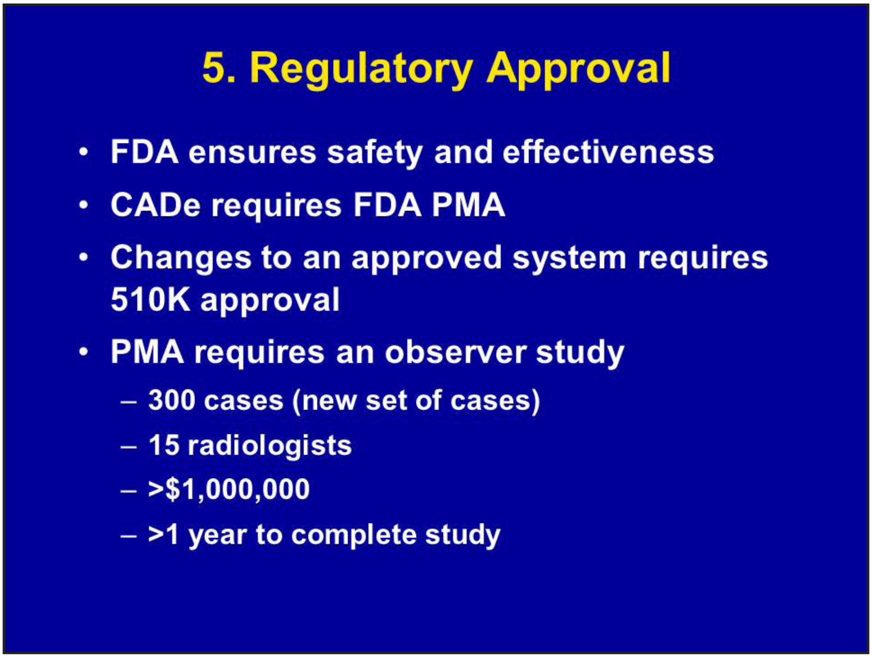
3. CADe System Development

- Develop algorithm
- Train classifier (ROC analysis)
- Test (ROC analysis)
- Select operating point on ROC curve



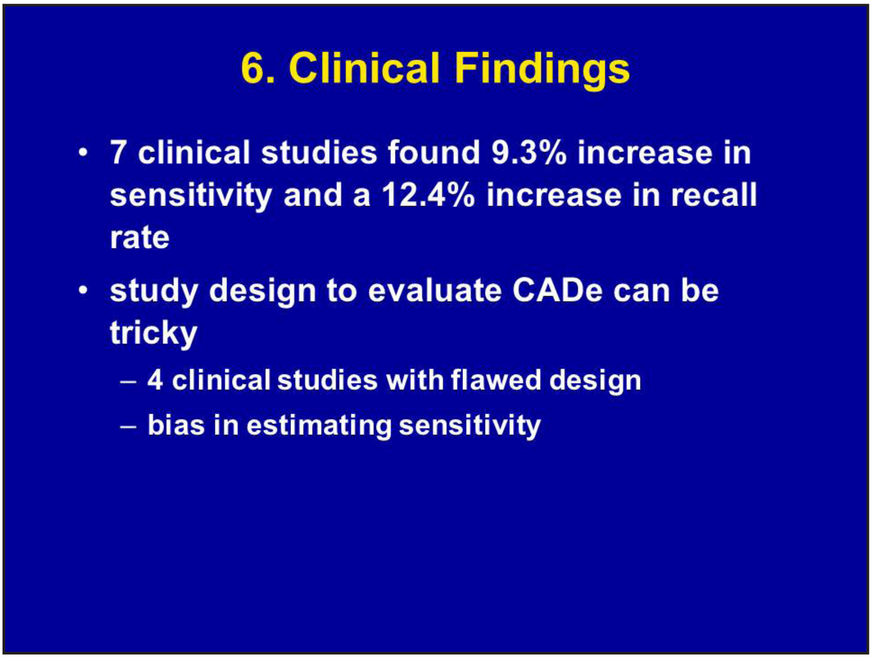
4. Technical Description of One CADe Application

- Omitting



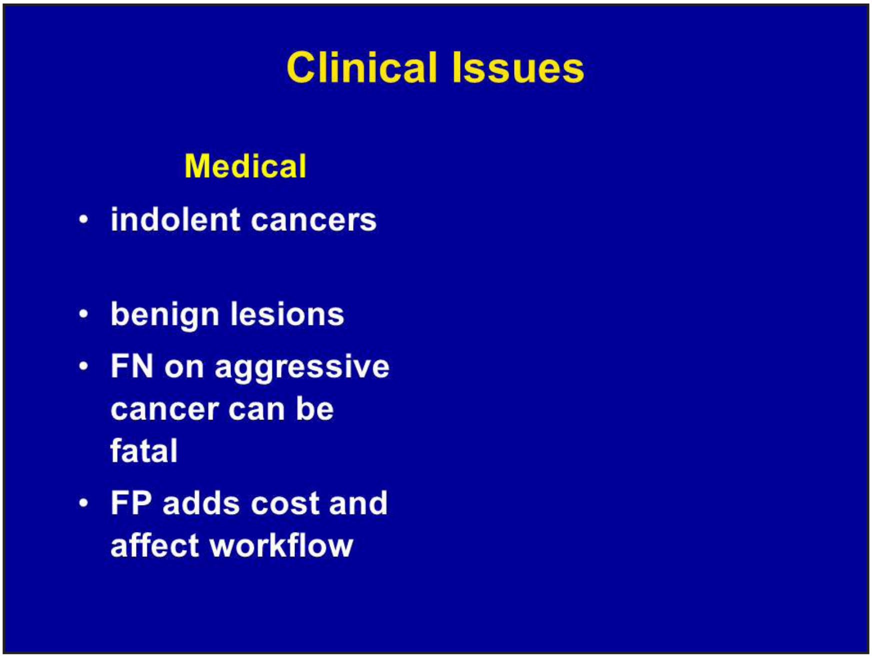
5. Regulatory Approval

- FDA ensures safety and effectiveness
- CADe requires FDA PMA
- Changes to an approved system requires 510K approval
- PMA requires an observer study
  - 300 cases (new set of cases)
  - 15 radiologists
  - >$1,000,000
  - >1 year to complete study



6. Clinical Findings

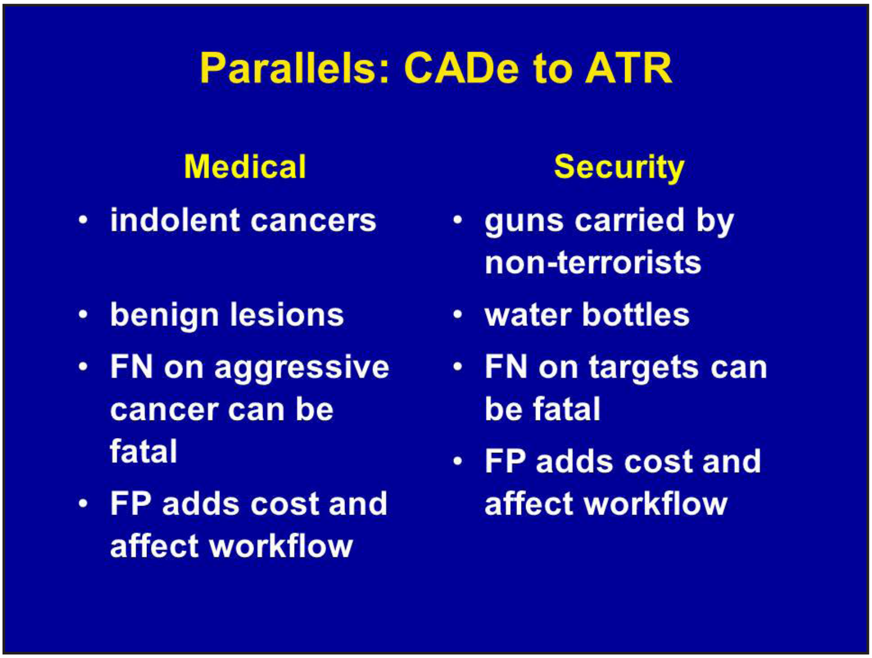
- 7 clinical studies found 9.3% increase in sensitivity and a 12.4% increase in recall rate
- study design to evaluate CADe can be tricky
  - 4 clinical studies with flawed design
  - bias in estimating sensitivity



Clinical Issues

Medical

- indolent cancers
- benign lesions
- FN on aggressive cancer can be fatal
- FP adds cost and affect workflow



Parallels: CADe to ATR

Medical	Security
• indolent cancers	• guns carried by non-terrorists
• benign lesions	• water bottles
• FN on aggressive cancer can be fatal	• FN on targets can be fatal
• FP adds cost and affect workflow	• FP adds cost and affect workflow

Differences

- Mammography has 2 views of each breast and temporal comparisons
- Need to be concerned about radiation dose
  - retakes for ambiguous findings are not done

CADe as a Second Reader

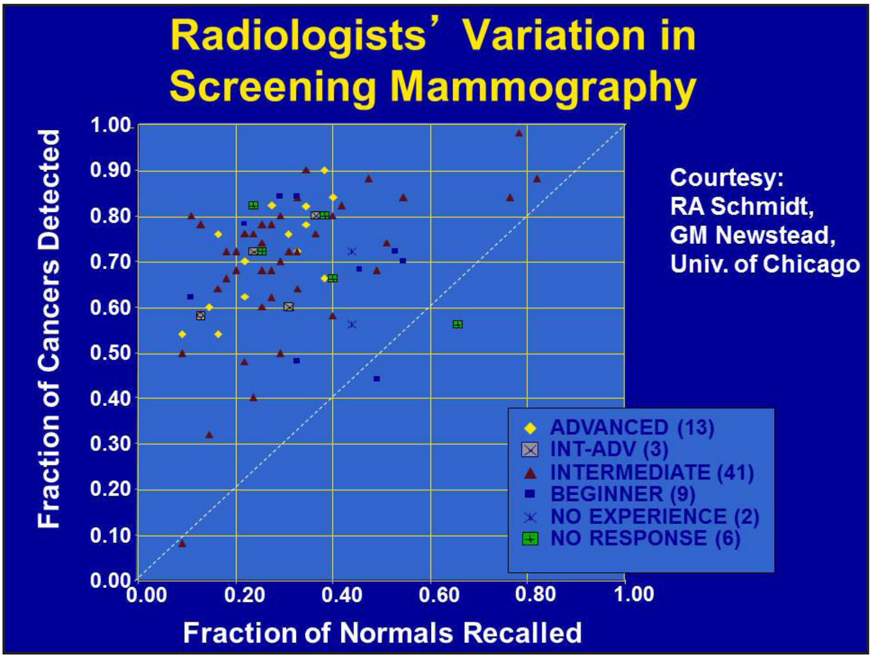
0 radiologists detected without CADe

3 radiologists detected with CADe

5 radiologist ignored the correct CADe mark (lower asterisk)

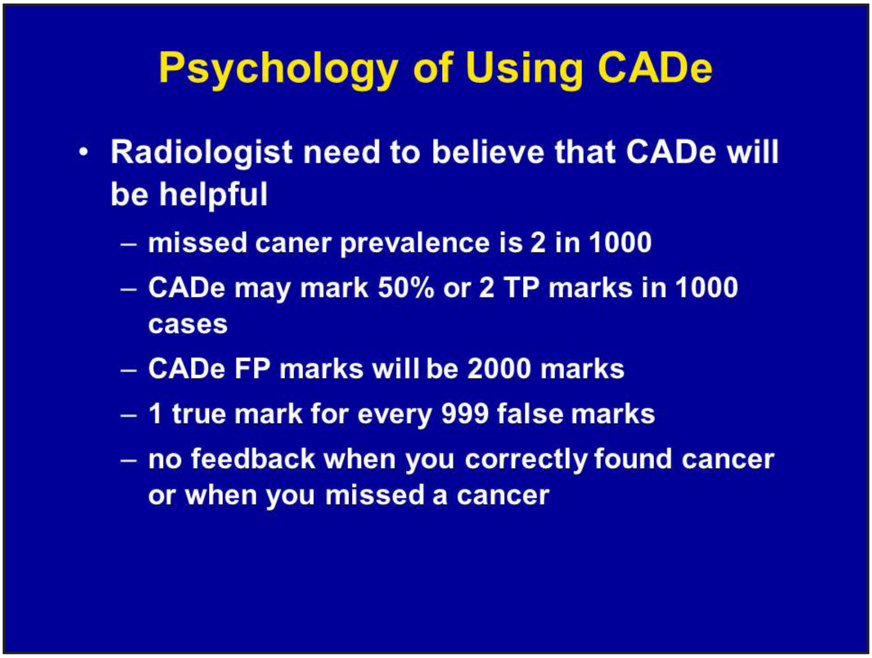
Observer Study

- 8 radiologists reading 300 screening exams
- 69 cancers (all missed clinically)
- reading without CADe sensitivity = 0.549
- reading with CADe sensitivity = 0.603
- 9.9% in sensitivity (12.4% increase in recall rate)
- radiologists ignored 70% of TP marks



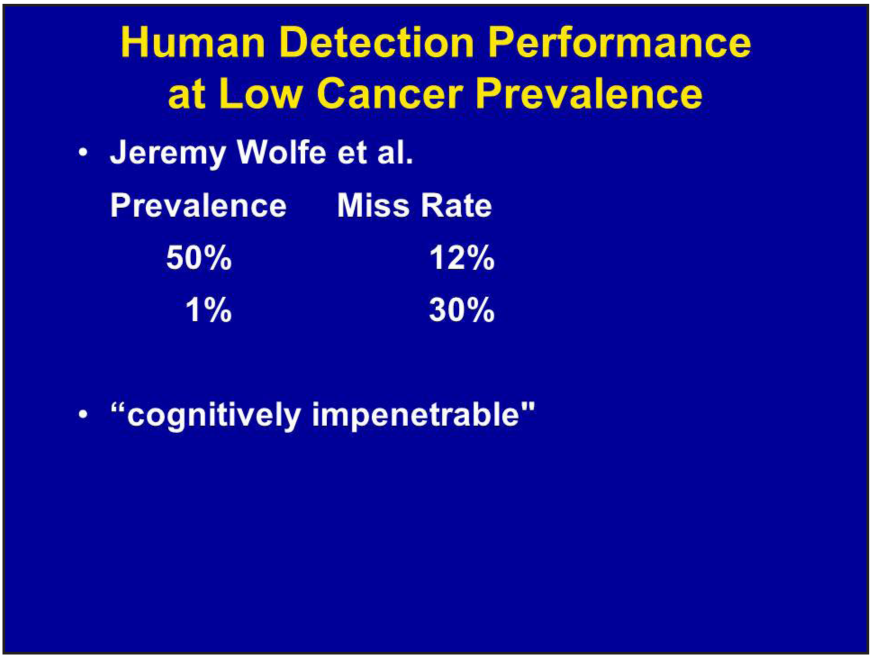
Radiologists' Variation in Screening Mammography

Courtesy:
RA Schmidt,
GM Newstead,
Univ. of Chicago



Psychology of Using CADe

- Radiologist need to believe that CADe will be helpful
  - – missed caner prevalence is 2 in 1000
  - – CADe may mark 50% or 2 TP marks in 1000 cases
  - – CADe FP marks will be 2000 marks
  - – 1 true mark for every 999 false marks
  - – no feedback when you correctly found cancer or when you missed a cancer

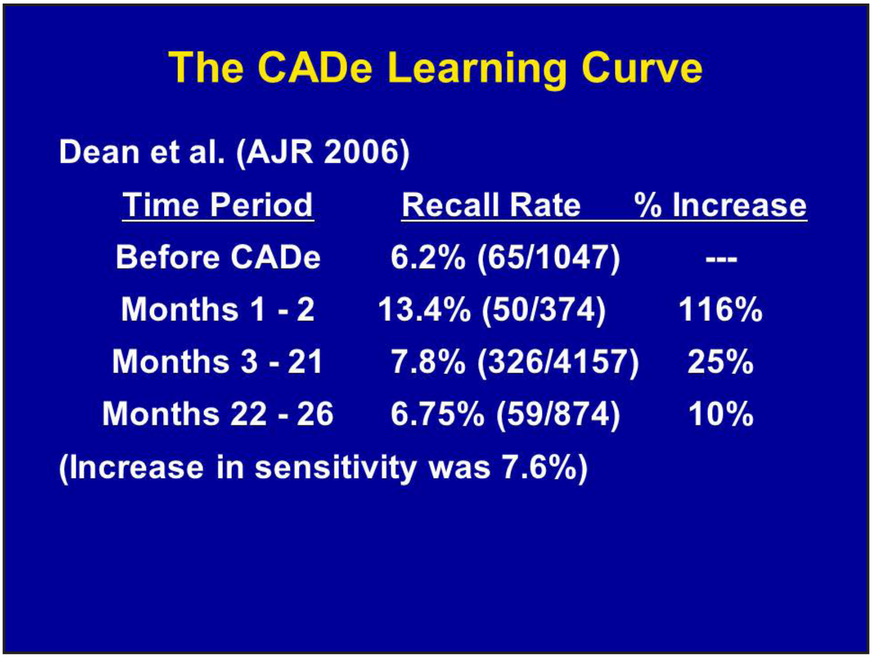


Human Detection Performance at Low Cancer Prevalence

- Jeremy Wolfe et al.

Prevalence	Miss Rate
50%	12%
1%	30%

- “cognitively impenetrable"

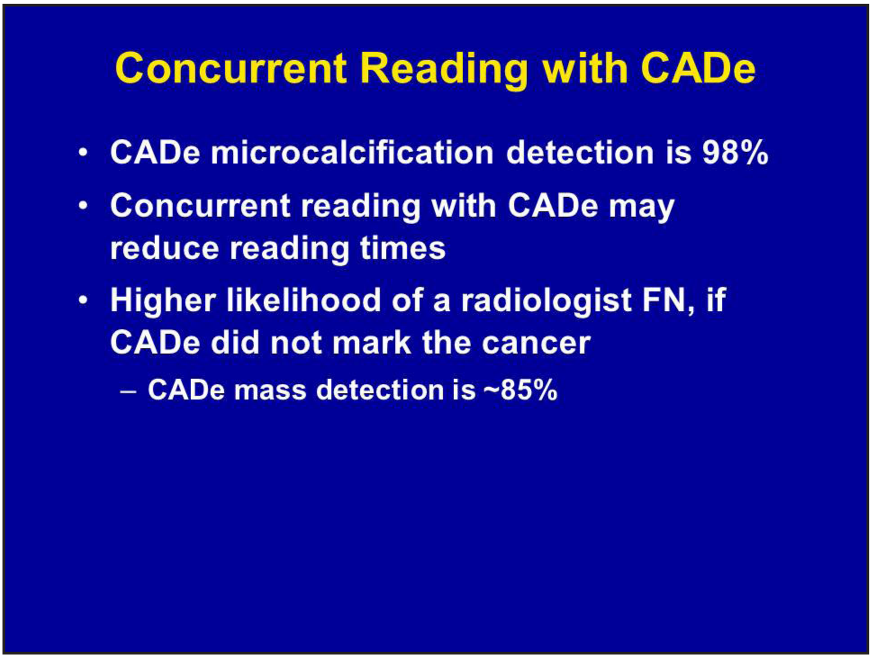


The CADe Learning Curve

Dean et al. (AJR 2006)

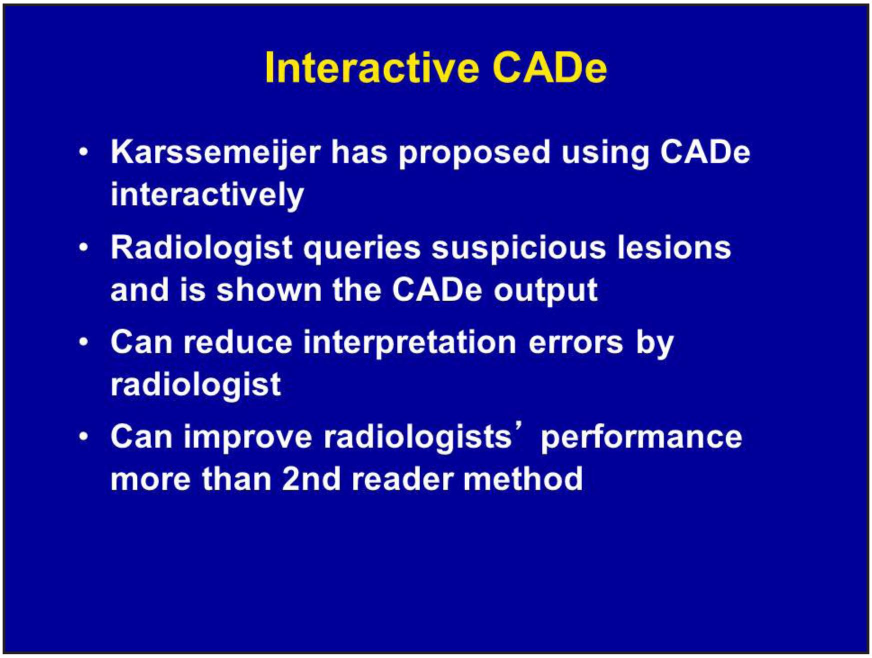
Time Period	Recall Rate	% Increase
Before CADe	6.2% (65/1047)	---
Months 1 - 2	13.4% (50/374)	116%
Months 3 - 21	7.8% (326/4157)	25%
Months 22 - 26	6.75% (59/874)	10%

(Increase in sensitivity was 7.6%)



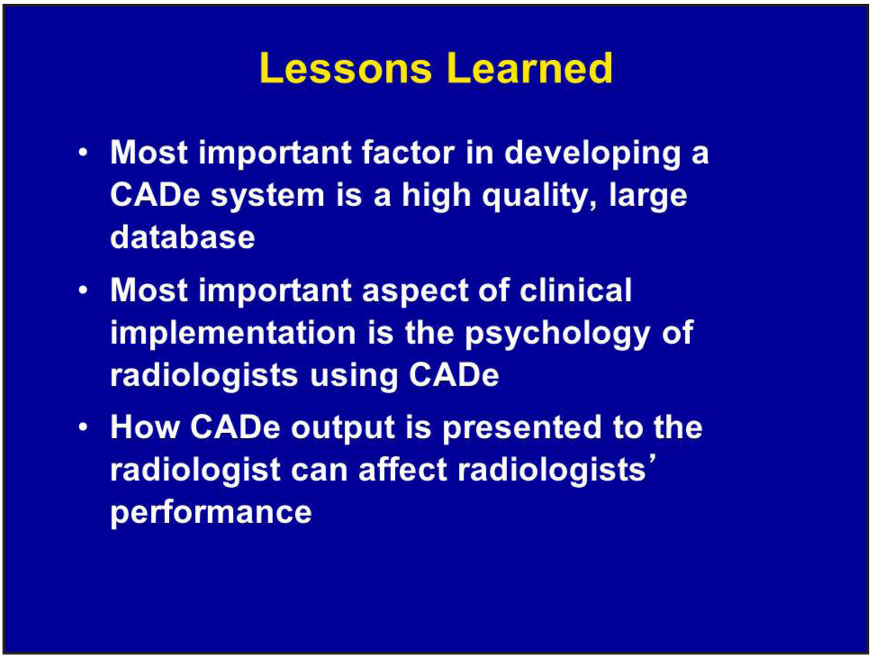
Concurrent Reading with CADe

- CADe microcalcification detection is 98%
- Concurrent reading with CADe may reduce reading times
- Higher likelihood of a radiologist FN, if CADe did not mark the cancer
  - CADe mass detection is ~85%



Interactive CADe

- Karssemeijer has proposed using CADe interactively
- Radiologist queries suspicious lesions and is shown the CADe output
- Can reduce interpretation errors by radiologist
- Can improve radiologists' performance more than 2nd reader method



Lessons Learned

- Most important factor in developing a CADe system is a high quality, large database
- Most important aspect of clinical implementation is the psychology of radiologists using CADe
- How CADe output is presented to the radiologist can affect radiologists' performance

16.9 Luc Perron: Clear Bag Concept for Risk Based Screening

OPTOSECURITY

eVelocity
Integrated Security Screening Software

"Clear Bag": A New Risk-Based Screening Approach

ADSA08, Boston, October 24th, 2012

Proprietary

Conclusions

1. Airports are struggling to meet their security mandate
   - Need greater focus on Operational requirements
2. A Risk-Based Approach to Security Screening can provide some relief
   - Adding unpredictability through ATR is an excellent way to minimize the risk
3. "Clear Bag" can reduce screener workload by 5% to 15%

OPTOSECURITY Proprietary 2

Unique Liquid Explosive Detection Solution

100% Full

60% Full

30% Full

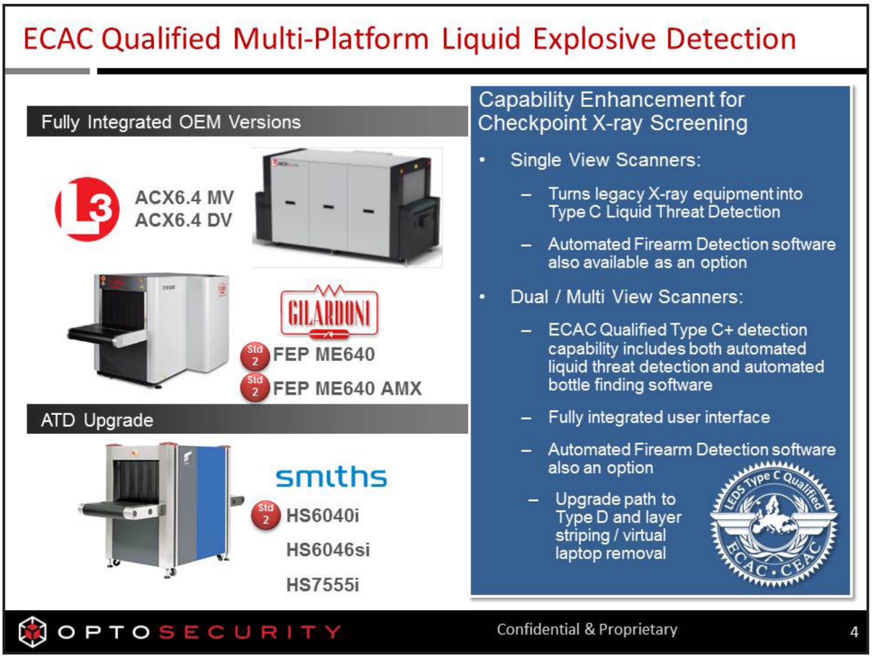
Automatic Meniscus Finding for Partially Filled Bottles

Multiple patents issued in Canada, pending in US & Europe

OPTOSECURITY

Proprietary

3



ECAC Qualified Multi-Platform Liquid Explosive Detection

Fully Integrated OEM Versions

ACX6.4 MV
ACX6.4 DV

GILARDONI

Std 2 **FEP ME640**
Std 2 **FEP ME640 AMX**

ATD Upgrade

smiths

Std 2 **HS6040i**
HS6046si
HS7555i

Capability Enhancement for Checkpoint X-ray Screening

- Single View Scanners:
  - Turns legacy X-ray equipment into Type C Liquid Threat Detection
  - Automated Firearm Detection software also available as an option
- Dual / Multi View Scanners:
  - ECAC Qualified Type C+ detection capability includes both automated liquid threat detection and automated bottle finding software
  - Fully integrated user interface
  - Automated Firearm Detection software also an option
  - Upgrade path to Type D and layer striping / virtual laptop removal

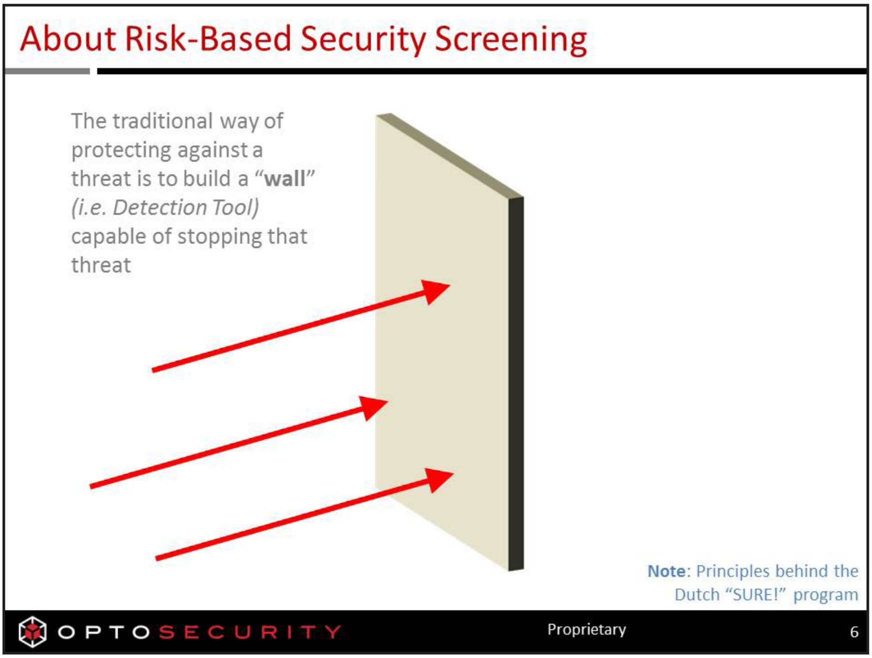
OPTOSECURITY

Confidential & Proprietary

4

ABOUT RISK-BASED SECURITY SCREENING

OPTOSECURITY Confidential & Proprietary 5

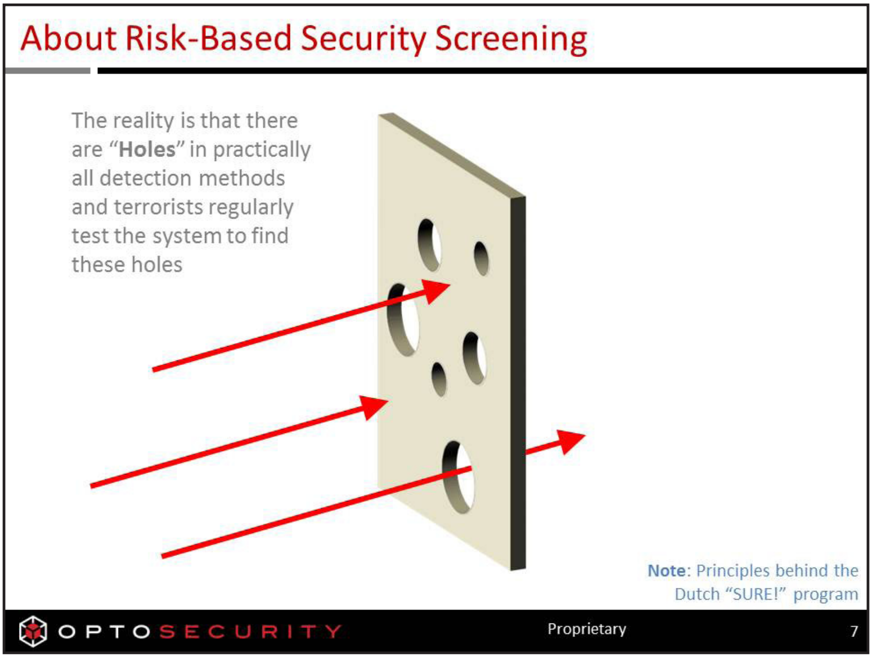


About Risk-Based Security Screening

The traditional way of protecting against a threat is to build a "**wall**" *(i.e. Detection Tool)* capable of stopping that threat

Note: Principles behind the Dutch "SURE!" program

OPTOSECURITY Proprietary 6

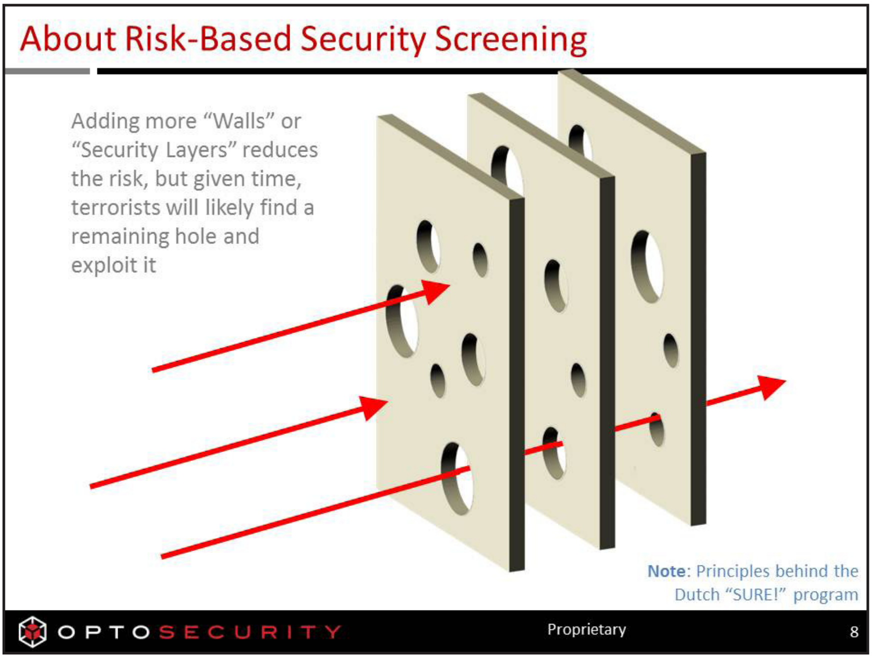


About Risk-Based Security Screening

The reality is that there are "**Holes**" in practically all detection methods and terrorists regularly test the system to find these holes

Note: Principles behind the Dutch "SURE!" program

OPTOSECURITY Proprietary 7

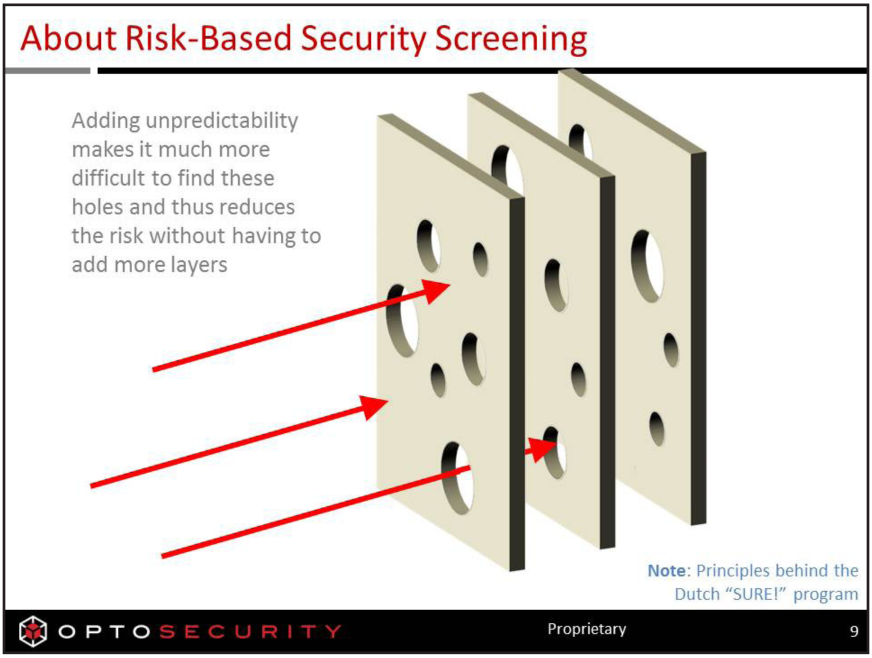


About Risk-Based Security Screening

Adding more "Walls" or "Security Layers" reduces the risk, but given time, terrorists will likely find a remaining hole and exploit it

Note: Principles behind the Dutch "SURE!" program

OPTOSECURITY Proprietary 8

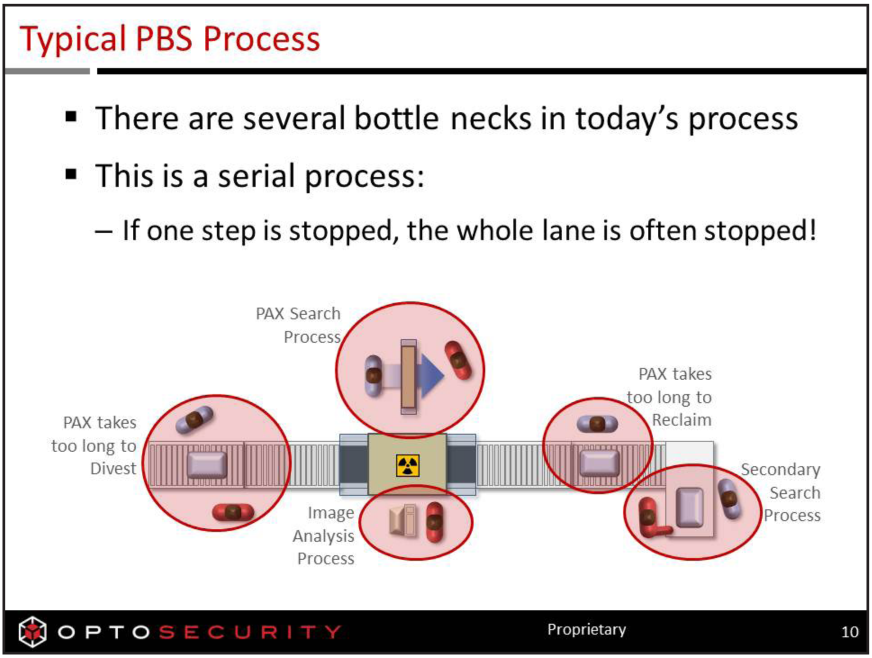


About Risk-Based Security Screening

Adding unpredictability makes it much more difficult to find these holes and thus reduces the risk without having to add more layers

Note: Principles behind the Dutch "SURE!" program

OPTOSECURITY Proprietary 9



Typical PBS Process

- There are several bottle necks in today's process
- This is a serial process:
  - If one step is stopped, the whole lane is often stopped!

PAX Search Process

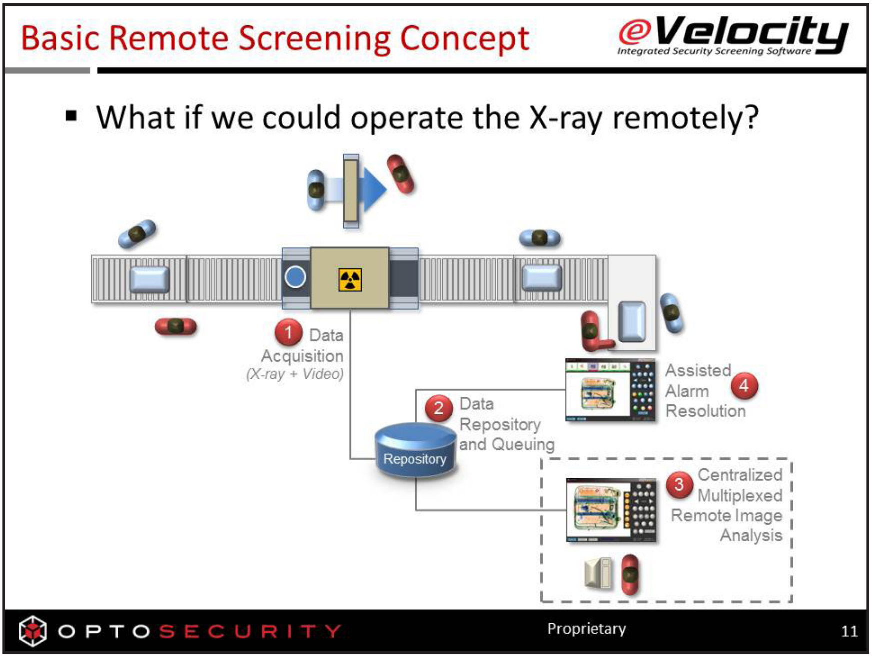
PAX takes too long to Divest

Image Analysis Process

PAX takes too long to Reclaim

Secondary Search Process

OPTOSECURITY Proprietary 10



Basic Remote Screening Concept

eVelocity
Integrated Security Screening Software

- What if we could operate the X-ray remotely?

① Data Acquisition *(X-ray + Video)*

② Data Repository and Queuing

Repository

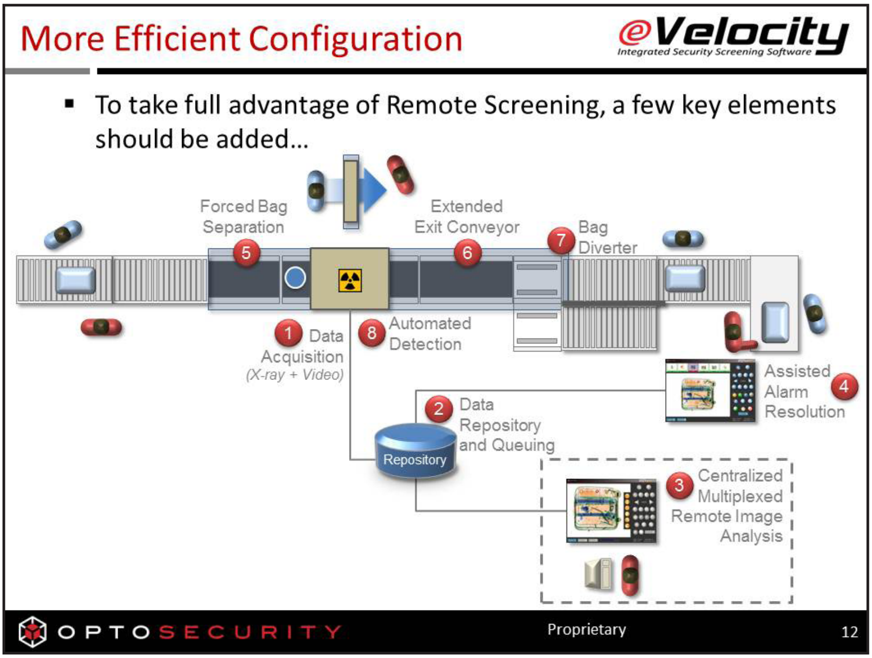
③ Centralized Multiplexed Remote Image Analysis

④ Assisted Alarm Resolution

OPTOSECURITY

Proprietary

11



More Efficient Configuration

eVelocity
Integrated Security Screening Software

- To take full advantage of Remote Screening, a few key elements should be added…

⑤ Forced Bag Separation

⑥ Extended Exit Conveyor

⑦ Bag Diverter

① Data Acquisition *(X-ray + Video)*

⑧ Automated Detection

② Data Repository and Queuing

Repository

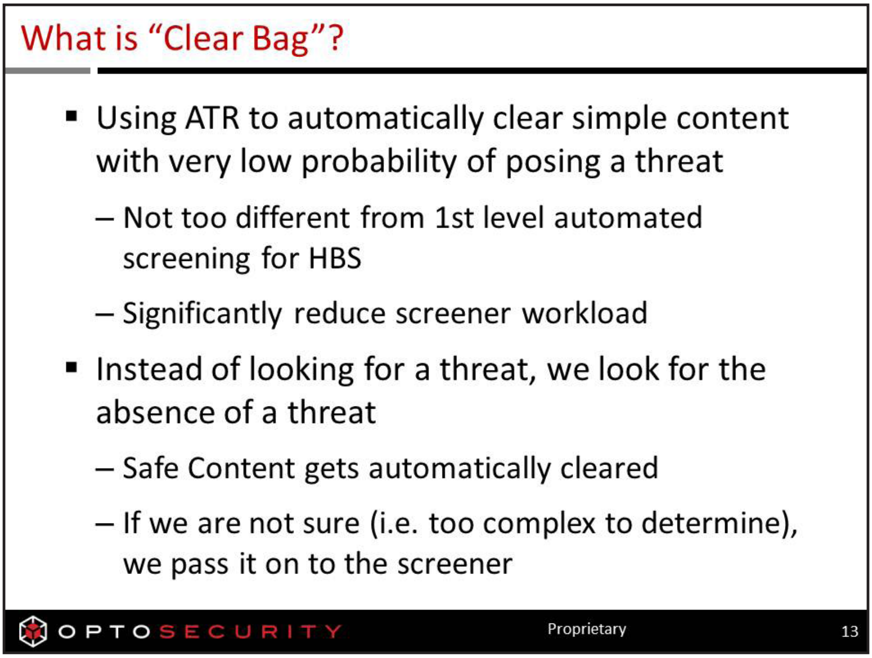
③ Centralized Multiplexed Remote Image Analysis

④ Assisted Alarm Resolution

OPTOSECURITY

Proprietary

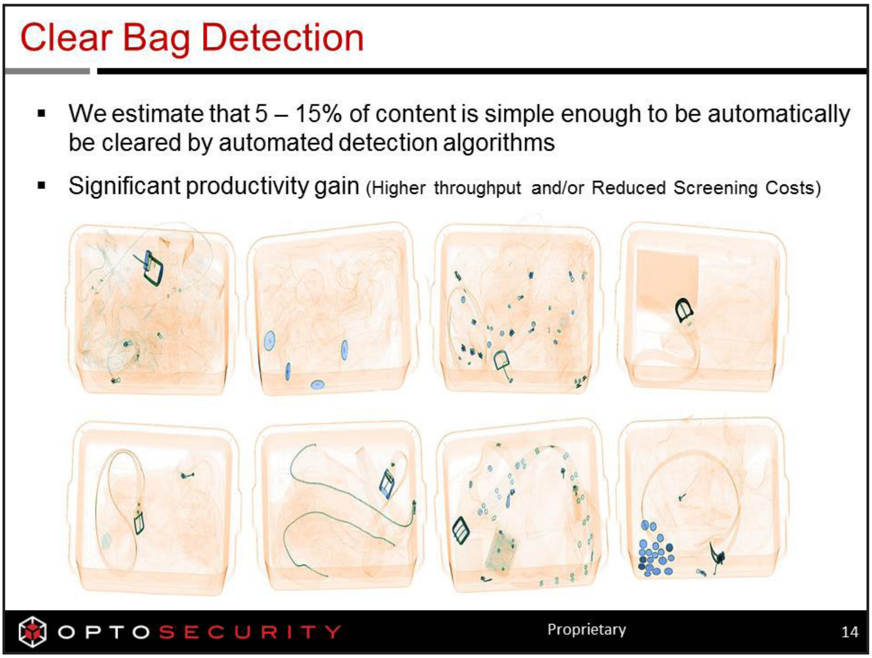
12



What is "Clear Bag"?

- Using ATR to automatically clear simple content with very low probability of posing a threat
  - Not too different from 1st level automated screening for HBS
  - Significantly reduce screener workload
- Instead of looking for a threat, we look for the absence of a threat
  - Safe Content gets automatically cleared
  - If we are not sure (i.e. too complex to determine), we pass it on to the screener

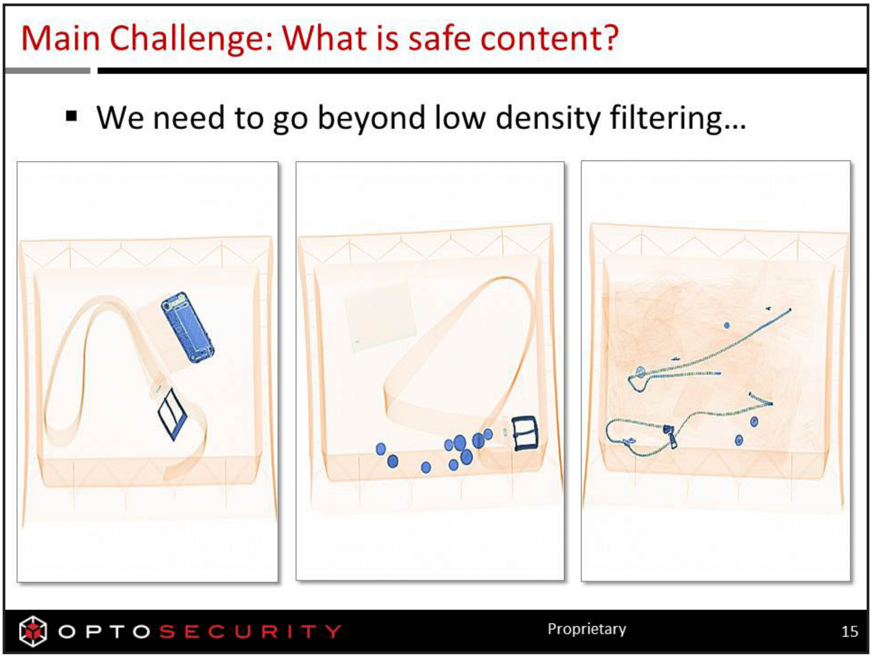
OPTOSECURITY Proprietary 13



Clear Bag Detection

- We estimate that 5 – 15% of content is simple enough to be automatically be cleared by automated detection algorithms
- Significant productivity gain (Higher throughput and/or Reduced Screening Costs)

OPTOSECURITY Proprietary 14



Main Challenge: What is safe content?

- We need to go beyond low density filtering…

Paradigm Shift in Checkpoint Security Screening

- Basic "Clear Bag" module has already been integrated into eVelocity suite
- Off-Line Testing with real images from Schiphol has shown an average of 7% auto-clear
  - Based on approximately 20000 images from multiple checkpoints
- Qualification testing currently in progress at TNO
  - Formal Test Methodology developed by Dutch Regulators
- First operational deployment in Staff Checkpoint being planned for December

What else can we do?

Real-Time KPI

Demand Management

- Metal Detector Data Acquisition
- AIT Data Acquisition
- Trace Detector Data Acquisition
- RT Monitoring & Reporting
- Flight Data Acquisition
- Real Time (RT) Queue Management
- RT Equipment Demand
- RT Staffing Demand
- RT Service & Maintenance

Integrated View leads to Situation Awareness

OPTOSECURITY Proprietary 17

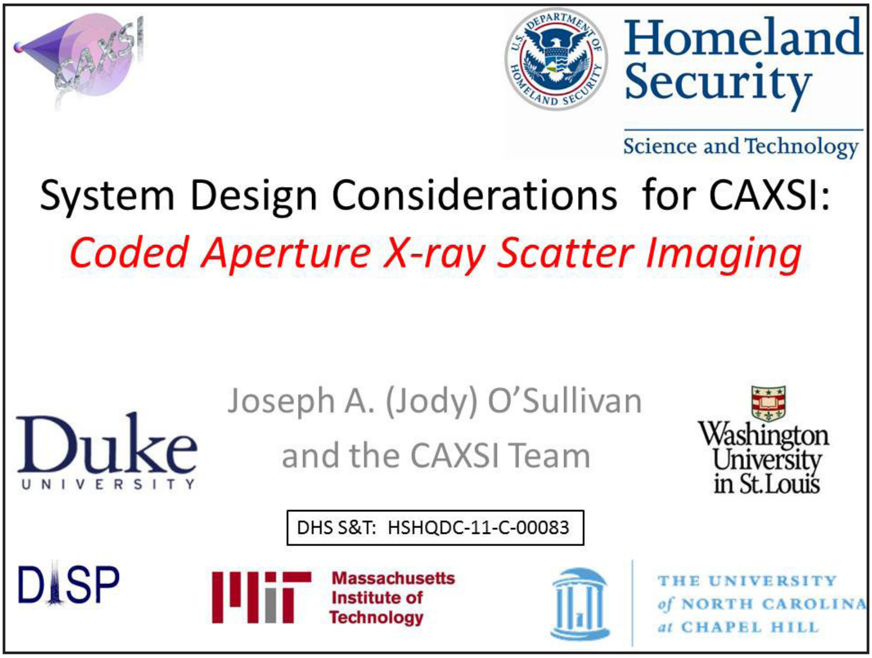
Conclusions

1. Airports are struggling to meet their security mandate
   - Need greater focus on Operational requirements
2. A Risk-Based Approach to Security Screening can provide some relief
   - Adding unpredictability through ATR is an excellent way to minimize the risk
3. “Clear Bag” can reduce screener workload by 5% to 15%

OPTOSECURITY Proprietary 18


29
29
SORTIE
EXIT
OPTOSECURITY

16.10 Jody O'Sullivan: Classifier Design for CAXI Project



Homeland Security
Science and Technology

System Design Considerations for CAXSI:
Coded Aperture X-ray Scatter Imaging

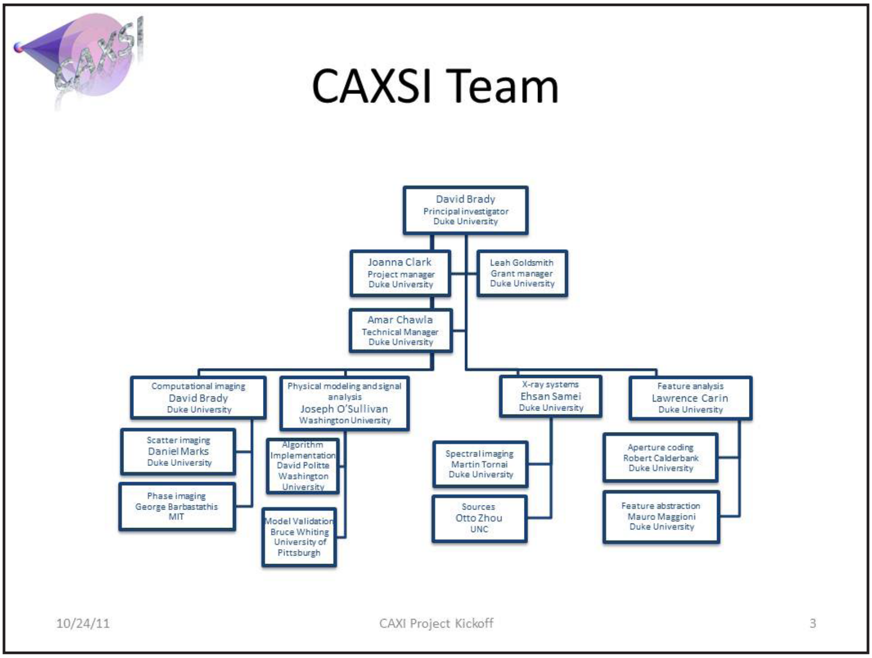
Joseph A. (Jody) O'Sullivan
and the CAXSI Team

DHS S&T: HSHQDC-11-C-00083

Duke University
Washington University in St. Louis
DISP
Massachusetts Institute of Technology
The University of North Carolina at Chapel Hill

CAXSI Team

- David Brady, Larry Carin, Robert Calderbank, Amarpreet Chawla, Anuj Kapadia, Kalyani Krishnamurthy, Andrew Holmgren, Pooyan Bagherzadeh, Ehsan Samei, Martin Tornai, Mauro Maggioni, Randy McKinley, Scott Wolter, Duke University
- Jody O'Sullivan, David Politte, Ikenna Odinaka, Washington University
- Bruce Whiting, University of Pittsburgh
- Otto Zhou, Kenneth MacCabe, UNC
- George Barbastathis, Jon Petrocelli, Lei Tian, MIT



CAXSI Team

David Brady
Principal investigator
Duke University

Joanna Clark
Project manager
Duke University

Leah Goldsmith
Grant manager
Duke University

Amar Chawla
Technical Manager
Duke University

Computational imaging
David Brady
Duke University

Scatter imaging
Daniel Marks
Duke University

Phase imaging
George Barbastathis
MIT

Physical modeling and signal analysis
Joseph O'Sullivan
Washington University

Algorithm Implementation
David Politte
Washington University

Model Validation
Bruce Whiting
University of Pittsburgh

X-ray systems
Ehsan Samei
Duke University

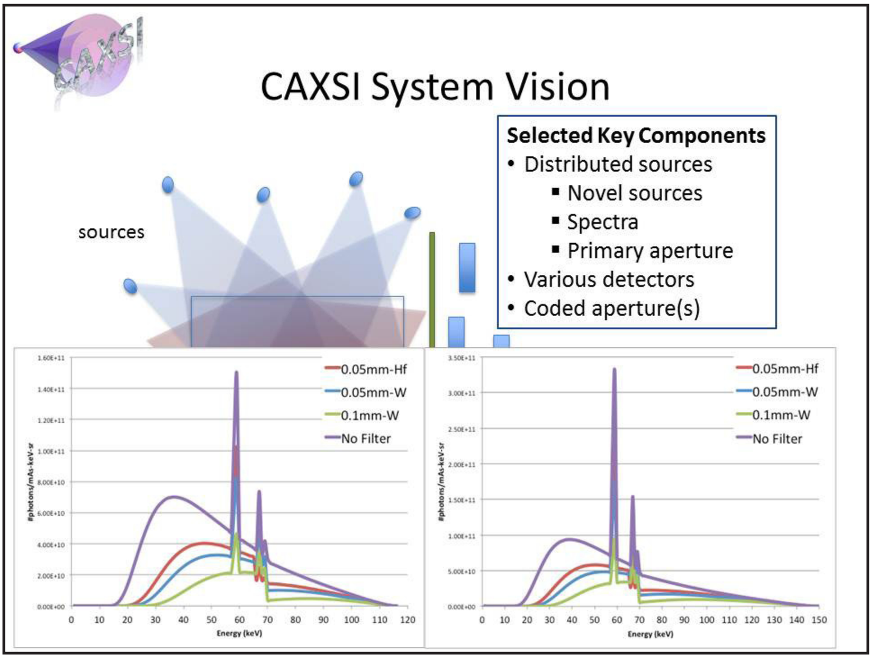
Spectral imaging
Martin Tornai
Duke University

Sources
Otto Zhou
UNC

Feature analysis
Lawrence Carin
Duke University

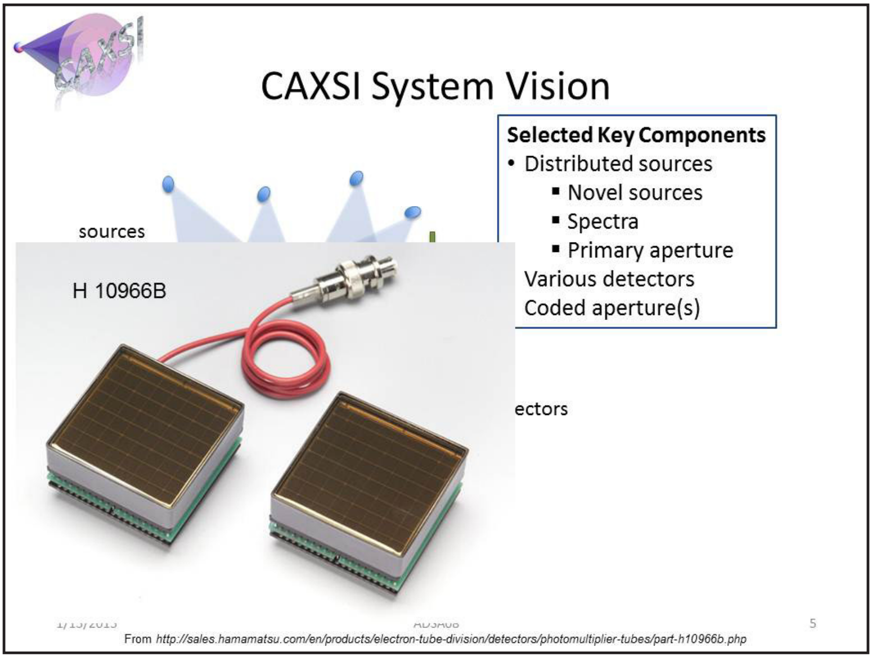
Aperture coding
Robert Calderbank
Duke University

Feature abstraction
Mauro Maggioni
Duke University



CAXSI System Vision

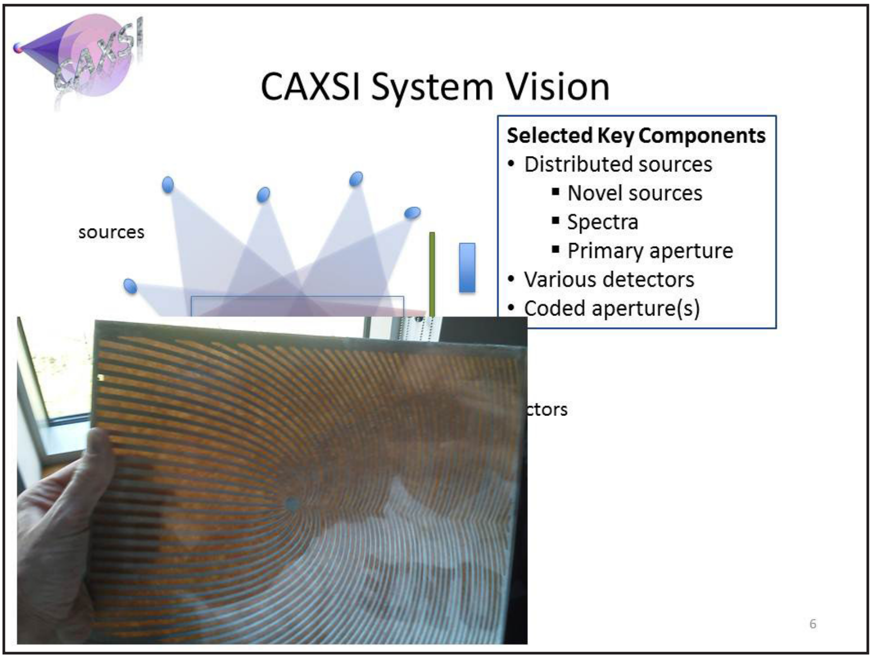
Selected Key Components
- Distributed sources
    - Novel sources
    - Spectra
    - Primary aperture
- Various detectors
- Coded aperture(s)



CAXSI System Vision

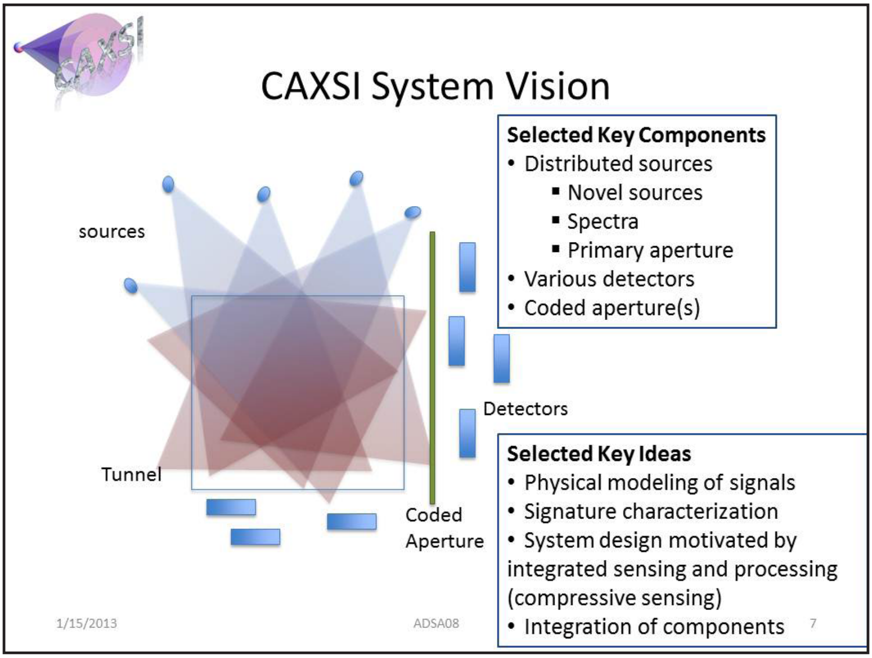
Selected Key Components
- Distributed sources
  - Novel sources
  - Spectra
  - Primary aperture
- Various detectors
- Coded aperture(s)

From *http://sales.hamamatsu.com/en/products/electron-tube-division/detectors/photomultiplier-tubes/part-h10966b.php*



CAXSI System Vision

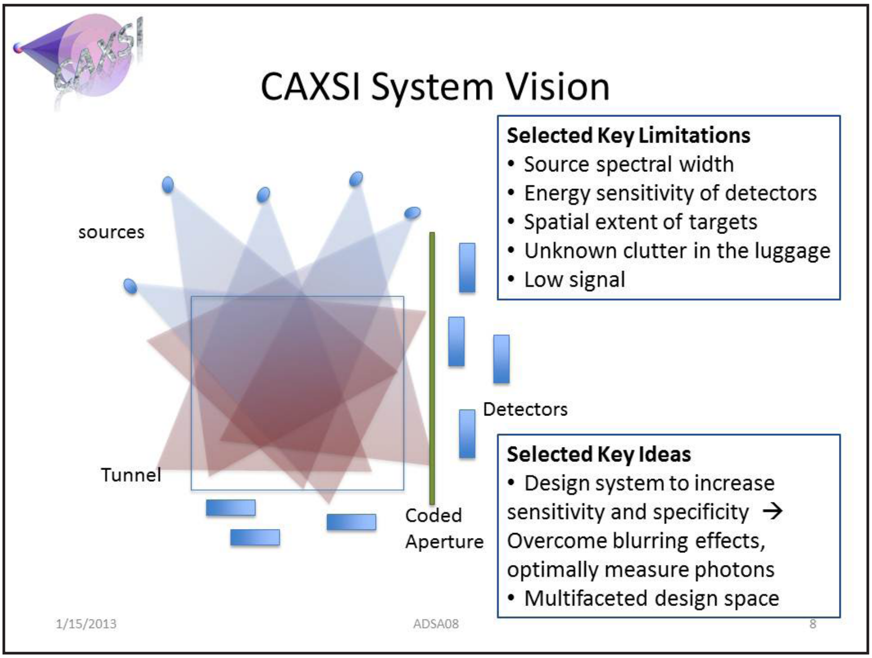
Selected Key Components
- Distributed sources
  - Novel sources
  - Spectra
  - Primary aperture
- Various detectors
- Coded aperture(s)



CAXSI System Vision

Selected Key Components
- Distributed sources
  - Novel sources
  - Spectra
  - Primary aperture
- Various detectors
- Coded aperture(s)

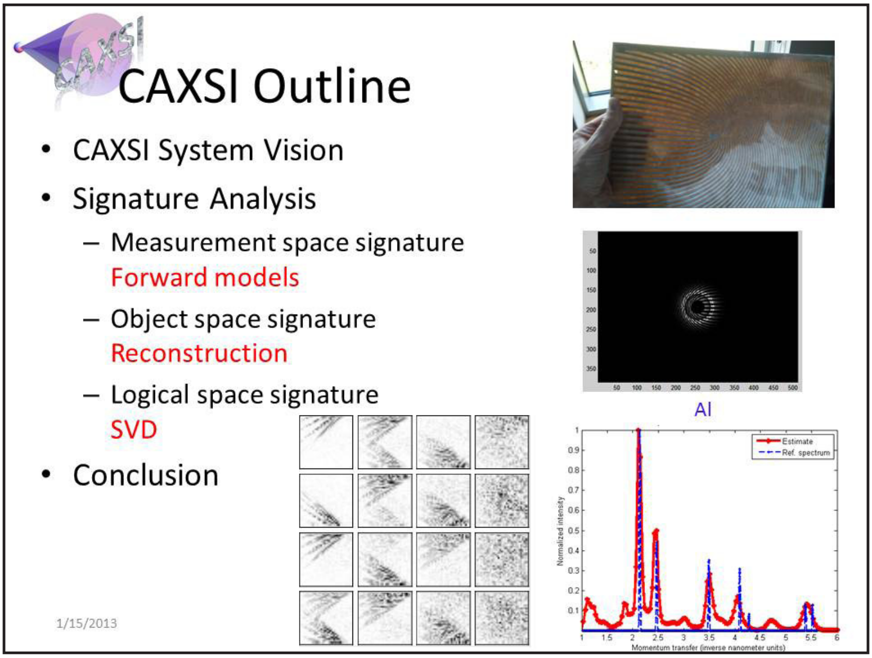
Selected Key Ideas
- Physical modeling of signals
- Signature characterization
- System design motivated by integrated sensing and processing (compressive sensing)
- Integration of components



CAXSI System Vision

Selected Key Limitations
- Source spectral width
- Energy sensitivity of detectors
- Spatial extent of targets
- Unknown clutter in the luggage
- Low signal

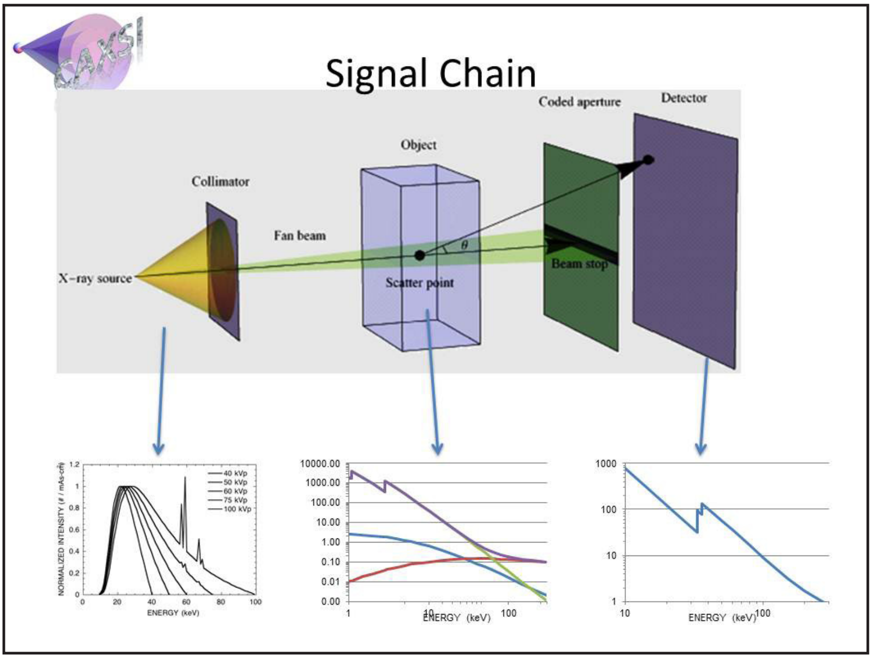
Selected Key Ideas
- Design system to increase sensitivity and specificity → Overcome blurring effects, optimally measure photons
- Multifaceted design space



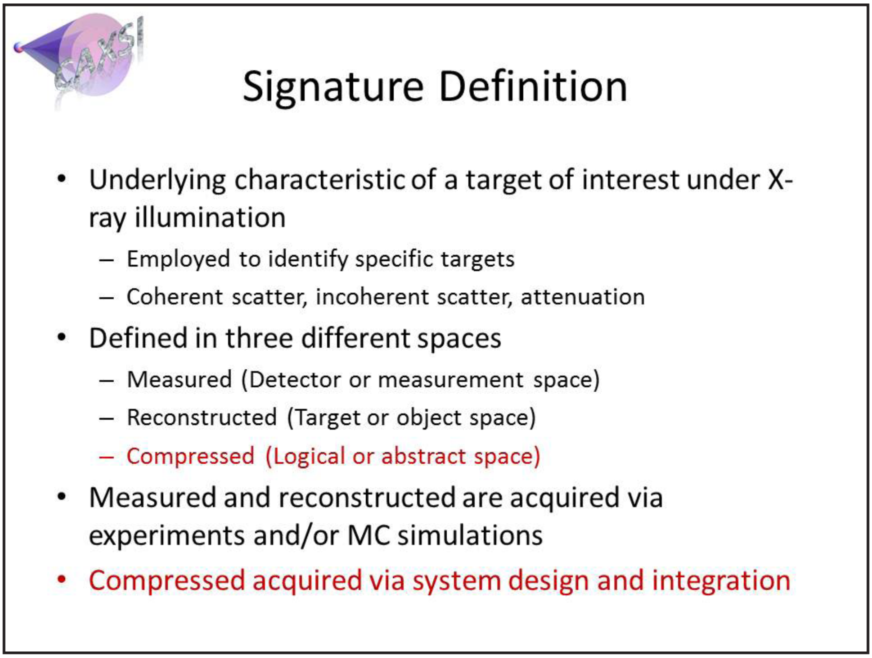
CAXSI Outline

- CAXSI System Vision
- Signature Analysis
  - Measurement space signature
    Forward models
  - Object space signature
    Reconstruction
  - Logical space signature
    SVD
- Conclusion

1/15/2013

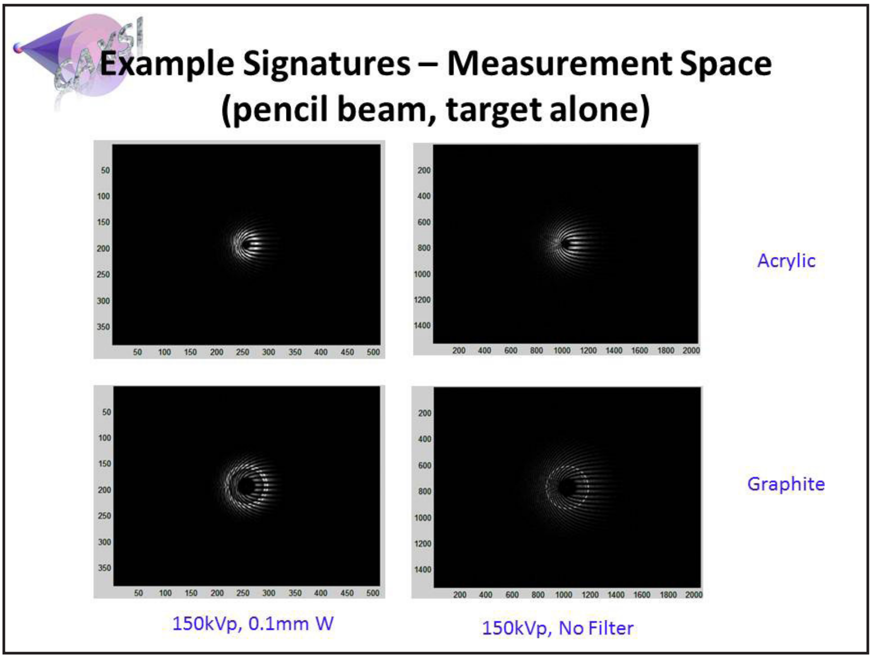


Signal Chain



Signature Definition

- Underlying characteristic of a target of interest under X-ray illumination
  - Employed to identify specific targets
  - Coherent scatter, incoherent scatter, attenuation
- Defined in three different spaces
  - Measured (Detector or measurement space)
  - Reconstructed (Target or object space)
  - Compressed (Logical or abstract space)
- Measured and reconstructed are acquired via experiments and/or MC simulations
- Compressed acquired via system design and integration



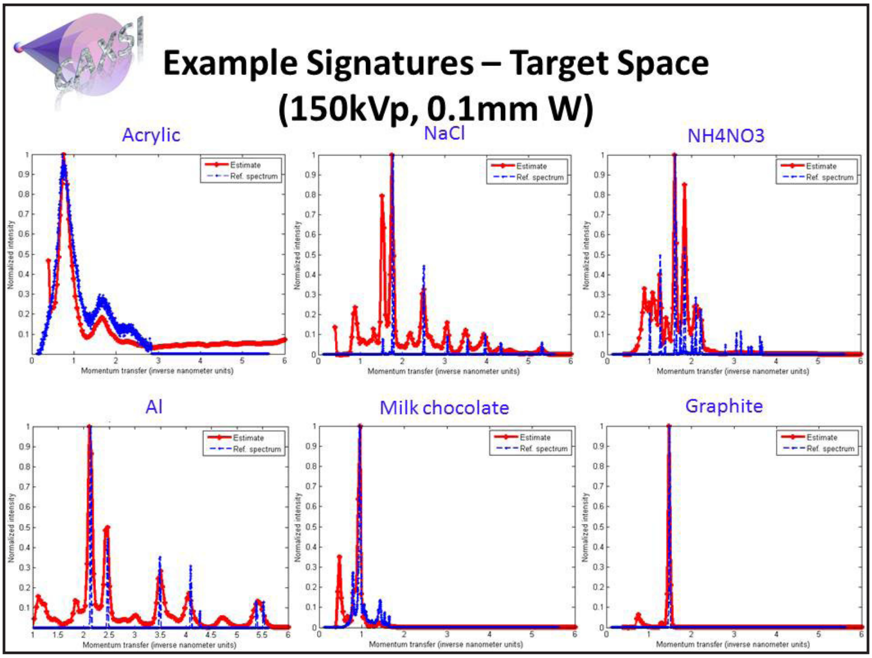
Example Signatures – Measurement Space (pencil beam, target alone)

Acrylic

Graphite

150kVp, 0.1mm W

150kVp, No Filter



Example Signatures – Target Space (150kVp, 0.1mm W)

Acrylic

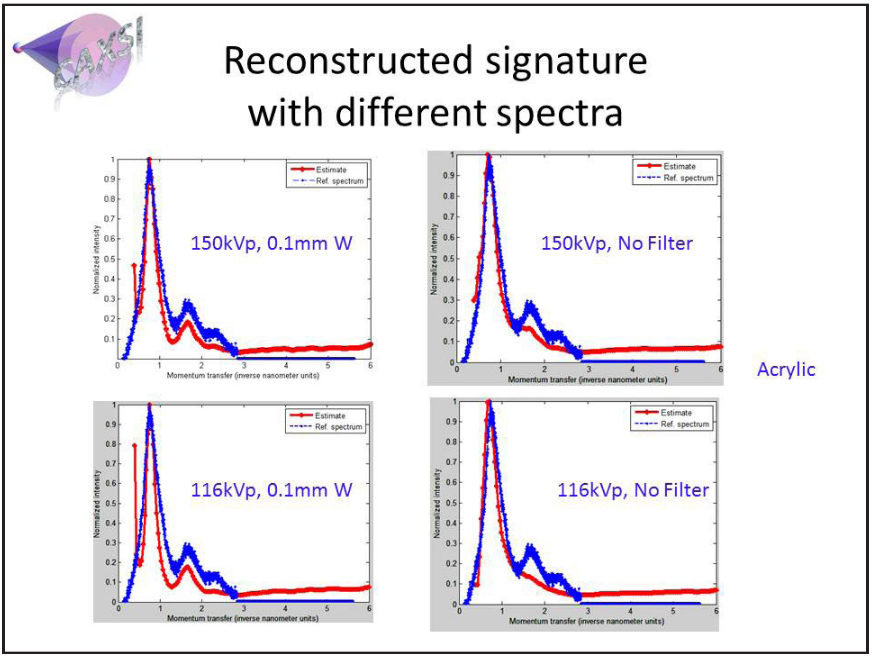
NaCl

NH4NO3

Al

Milk chocolate

Graphite



Reconstructed signature with different spectra

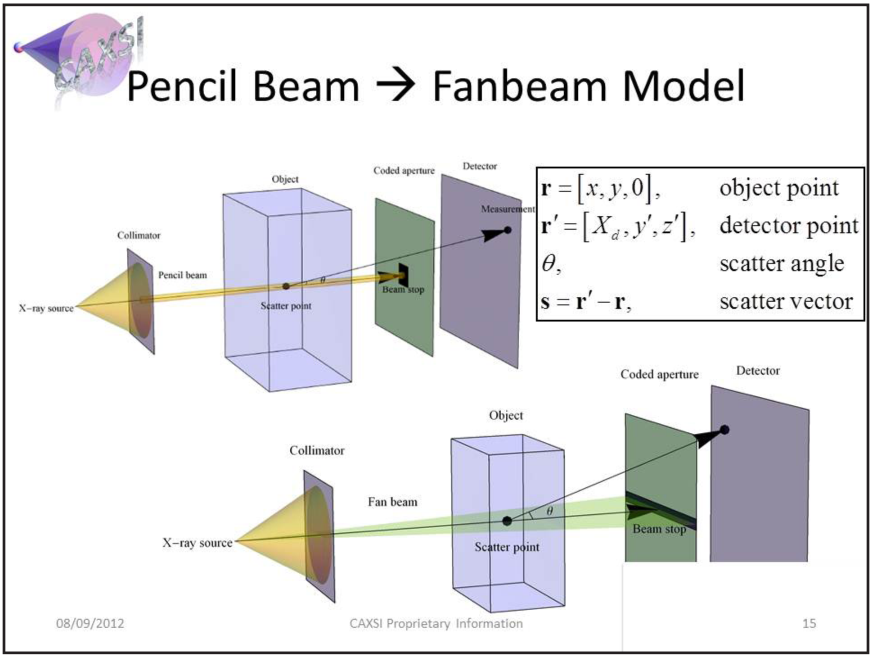
150kVp, 0.1mm W

150kVp, No Filter

116kVp, 0.1mm W

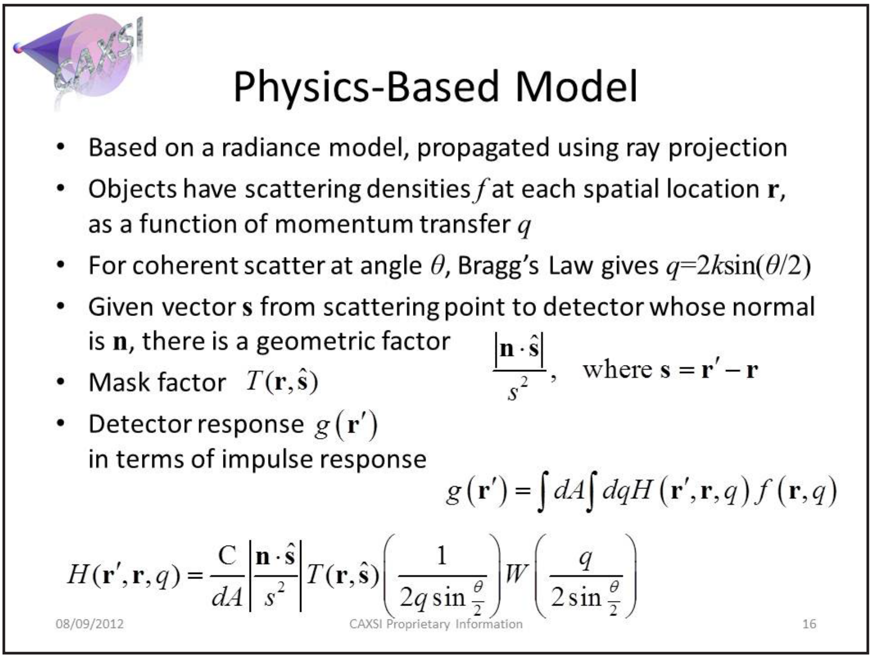
116kVp, No Filter

Acrylic



Pencil Beam → Fanbeam Model

$\mathbf{r} = [x, y, 0]$,	object point
$\mathbf{r}' = [X_d, y', z']$,	detector point
θ,	scatter angle
$\mathbf{s} = \mathbf{r}' - \mathbf{r}$,	scatter vector

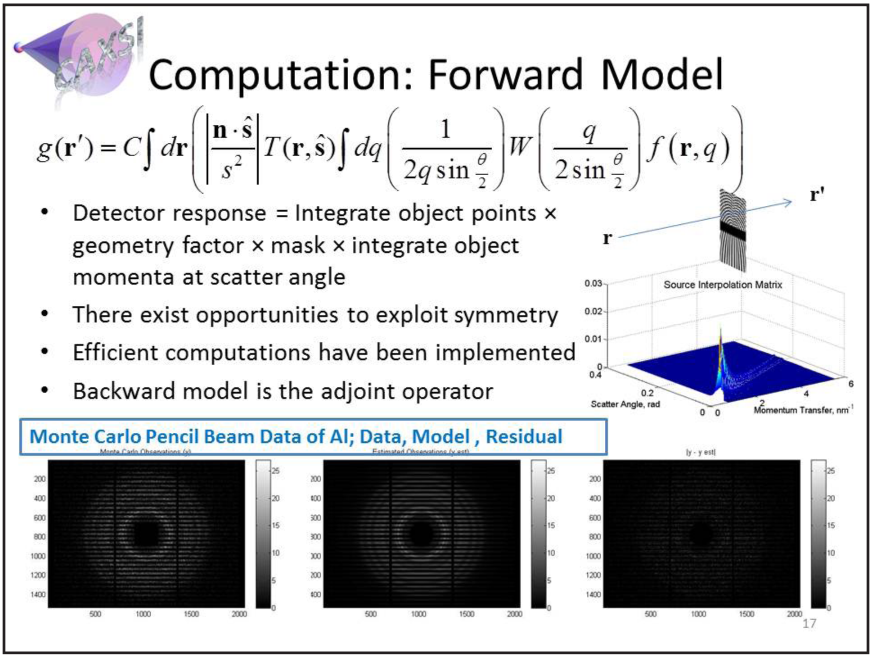


Physics-Based Model

- Based on a radiance model, propagated using ray projection
- Objects have scattering densities f at each spatial location $\mathbf{r}$, as a function of momentum transfer q
- For coherent scatter at angle θ, Bragg's Law gives $q=2k\sin(\theta/2)$
- Given vector $\mathbf{s}$ from scattering point to detector whose normal is $\mathbf{n}$, there is a geometric factor $\frac{|\mathbf{n}\cdot\hat{\mathbf{s}}|}{s^2}$, where $\mathbf{s} = \mathbf{r}' - \mathbf{r}$
- Mask factor $T(\mathbf{r}, \hat{\mathbf{s}})$
- Detector response $g(\mathbf{r}')$ in terms of impulse response

$$g(\mathbf{r}') = \int dA \int dq H(\mathbf{r}', \mathbf{r}, q) f(\mathbf{r}, q)$$

$$H(\mathbf{r}', \mathbf{r}, q) = \frac{C}{dA}\left|\frac{\mathbf{n}\cdot\hat{\mathbf{s}}}{s^2}\right| T(\mathbf{r}, \hat{\mathbf{s}}) \left(\frac{1}{2q\sin\frac{\theta}{2}}\right) W\left(\frac{q}{2\sin\frac{\theta}{2}}\right)$$

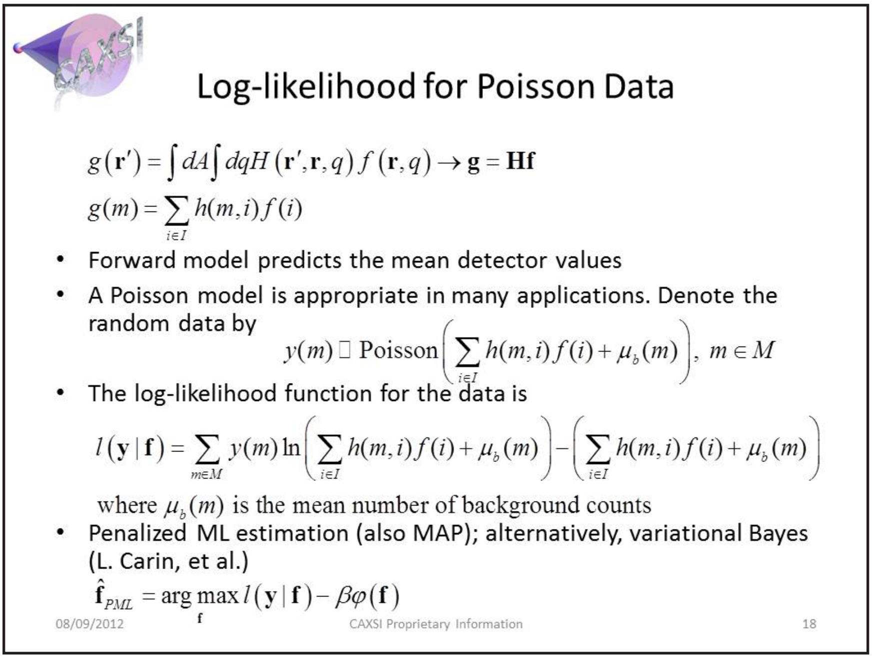


Computation: Forward Model

$$g(\mathbf{r}') = C\int d\mathbf{r}\left(\left|\frac{\mathbf{n}\cdot\hat{\mathbf{s}}}{s^2}\right| T(\mathbf{r},\hat{\mathbf{s}})\int dq\left(\frac{1}{2q\sin\frac{\theta}{2}}\right)W\left(\frac{q}{2\sin\frac{\theta}{2}}\right)f(\mathbf{r},q)\right)$$

- Detector response = Integrate object points × geometry factor × mask × integrate object momenta at scatter angle
- There exist opportunities to exploit symmetry
- Efficient computations have been implemented
- Backward model is the adjoint operator

Monte Carlo Pencil Beam Data of Al; Data, Model , Residual



Log-likelihood for Poisson Data

$$g(\mathbf{r}') = \int dA\int dq H(\mathbf{r}',\mathbf{r},q)\, f(\mathbf{r},q) \rightarrow \mathbf{g} = \mathbf{H}\mathbf{f}$$

$$g(m) = \sum_{i\in I} h(m,i) f(i)$$

- Forward model predicts the mean detector values
- A Poisson model is appropriate in many applications. Denote the random data by

$$y(m) \sim \text{Poisson}\left(\sum_{i\in I} h(m,i) f(i) + \mu_b(m)\right),\ m\in M$$

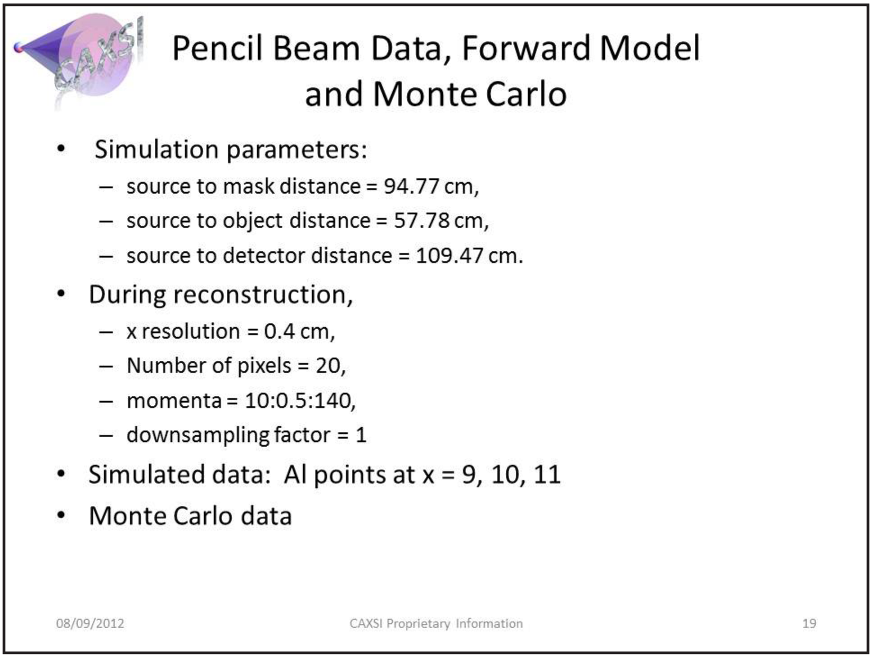
- The log-likelihood function for the data is

$$l(\mathbf{y}\,|\,\mathbf{f}) = \sum_{m\in M} y(m)\ln\left(\sum_{i\in I} h(m,i) f(i) + \mu_b(m)\right) - \left(\sum_{i\in I} h(m,i) f(i) + \mu_b(m)\right)$$

where $\mu_b(m)$ is the mean number of background counts

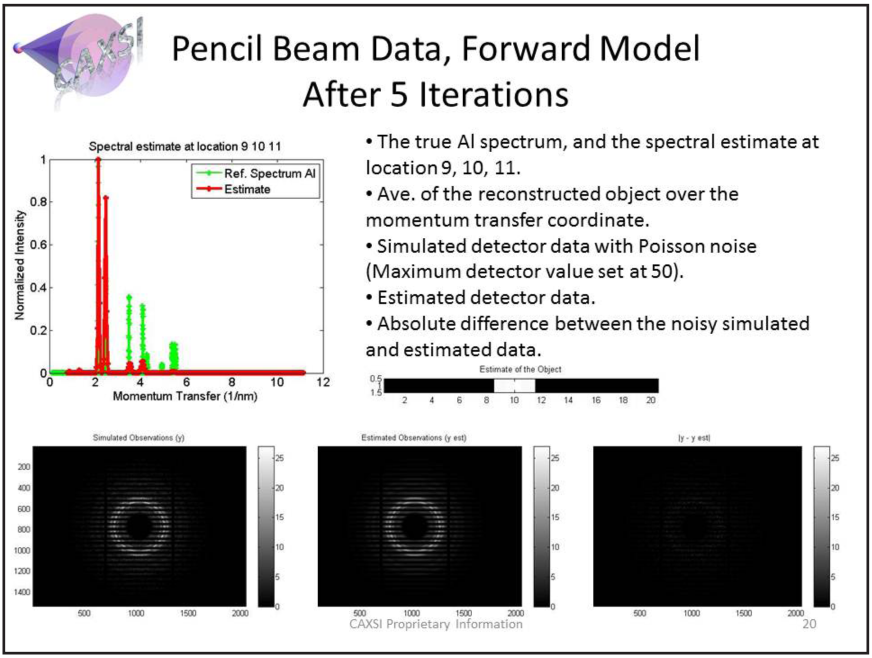
- Penalized ML estimation (also MAP); alternatively, variational Bayes (L. Carin, et al.)

$$\hat{\mathbf{f}}_{PML} = \arg\max_{\mathbf{f}} l(\mathbf{y}\,|\,\mathbf{f}) - \beta\varphi(\mathbf{f})$$



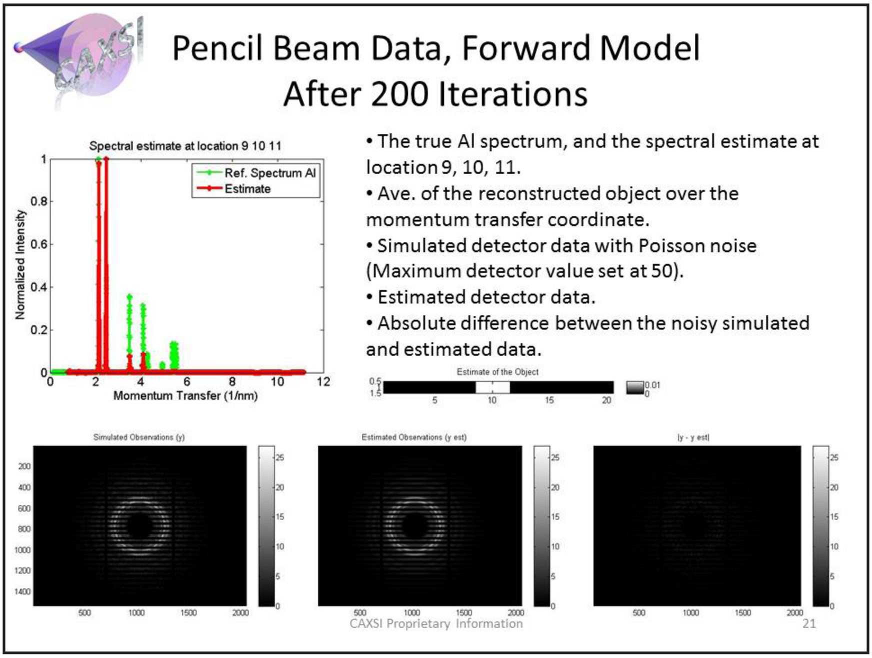
Pencil Beam Data, Forward Model and Monte Carlo

- Simulation parameters:
  - source to mask distance = 94.77 cm,
  - source to object distance = 57.78 cm,
  - source to detector distance = 109.47 cm.
- During reconstruction,
  - x resolution = 0.4 cm,
  - Number of pixels = 20,
  - momenta = 10:0.5:140,
  - downsampling factor = 1
- Simulated data: Al points at x = 9, 10, 11
- Monte Carlo data



Pencil Beam Data, Forward Model After 5 Iterations

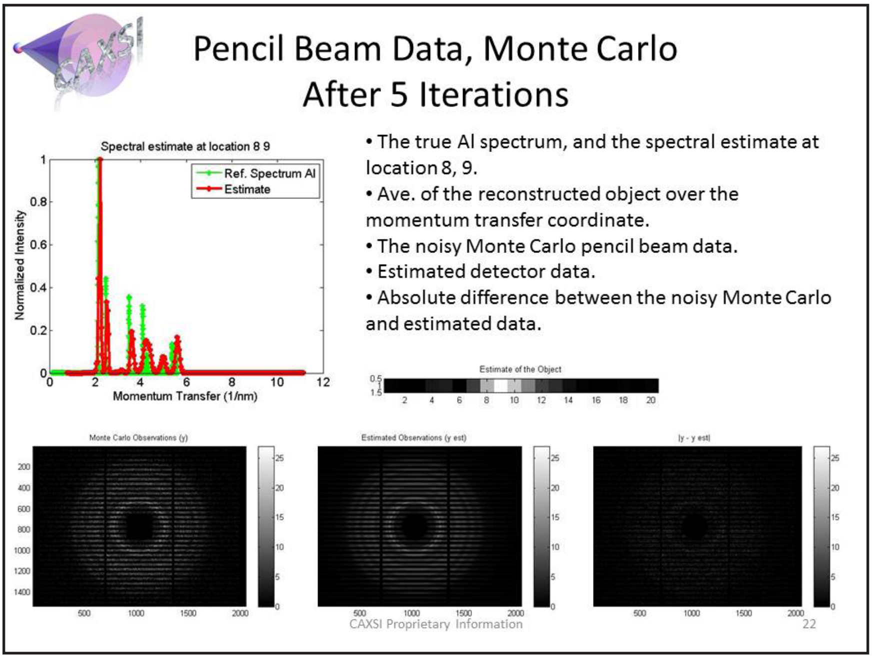
- The true Al spectrum, and the spectral estimate at location 9, 10, 11.
- Ave. of the reconstructed object over the momentum transfer coordinate.
- Simulated detector data with Poisson noise (Maximum detector value set at 50).
- Estimated detector data.
- Absolute difference between the noisy simulated and estimated data.



Pencil Beam Data, Forward Model
After 200 Iterations

- The true Al spectrum, and the spectral estimate at location 9, 10, 11.
- Ave. of the reconstructed object over the momentum transfer coordinate.
- Simulated detector data with Poisson noise (Maximum detector value set at 50).
- Estimated detector data.
- Absolute difference between the noisy simulated and estimated data.

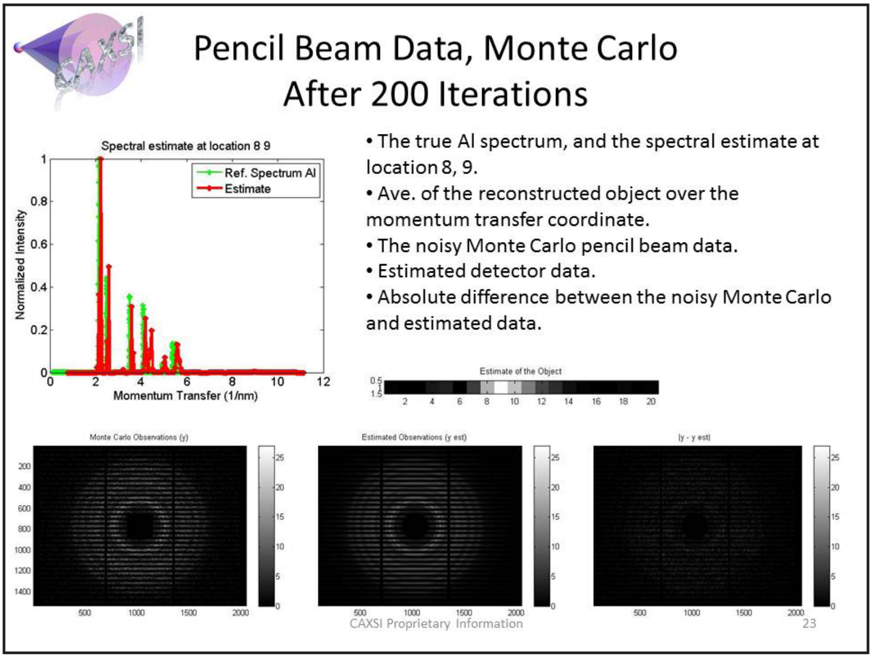
CAXSI Proprietary Information



Pencil Beam Data, Monte Carlo
After 5 Iterations

- The true Al spectrum, and the spectral estimate at location 8, 9.
- Ave. of the reconstructed object over the momentum transfer coordinate.
- The noisy Monte Carlo pencil beam data.
- Estimated detector data.
- Absolute difference between the noisy Monte Carlo and estimated data.

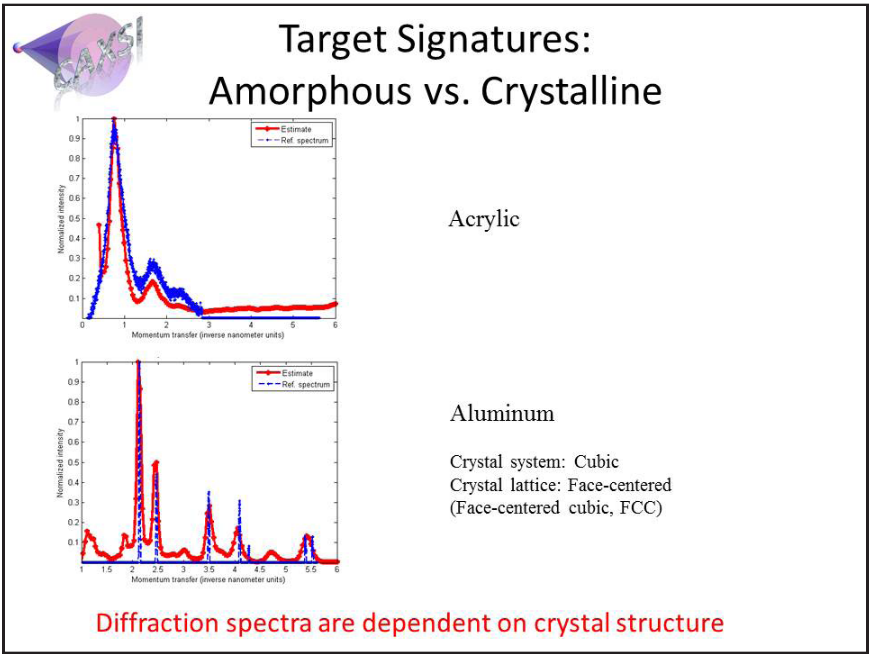
CAXSI Proprietary Information



Pencil Beam Data, Monte Carlo After 200 Iterations

- The true Al spectrum, and the spectral estimate at location 8, 9.
- Ave. of the reconstructed object over the momentum transfer coordinate.
- The noisy Monte Carlo pencil beam data.
- Estimated detector data.
- Absolute difference between the noisy Monte Carlo and estimated data.

CAXSI Proprietary Information



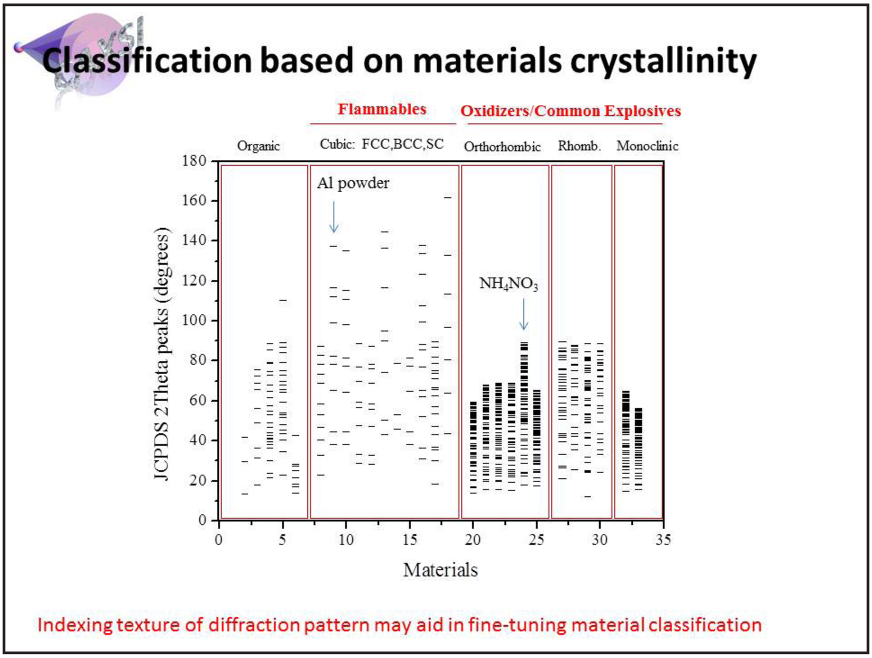
Target Signatures: Amorphous vs. Crystalline

Acrylic

Aluminum

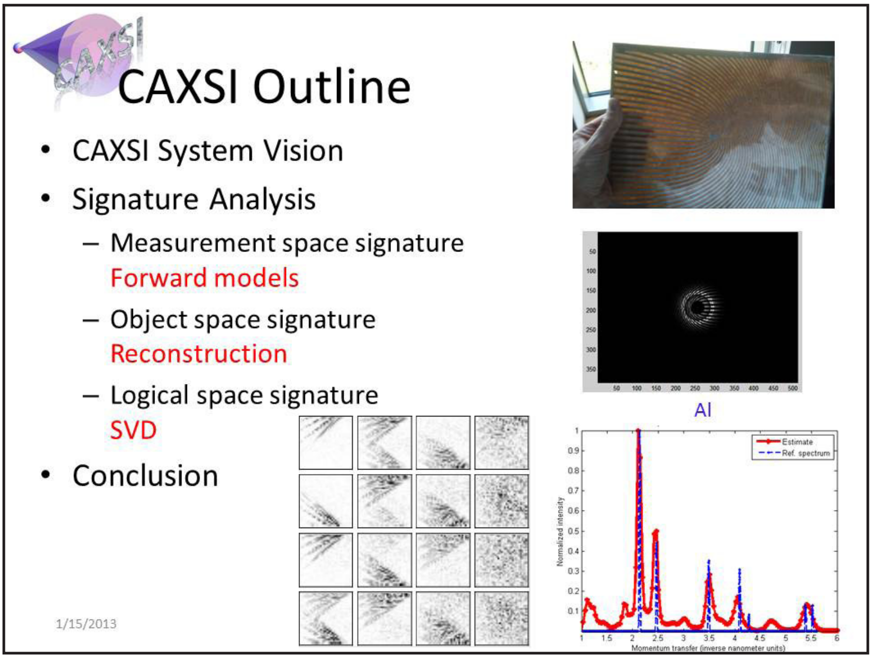
Crystal system: Cubic
Crystal lattice: Face-centered
(Face-centered cubic, FCC)

Diffraction spectra are dependent on crystal structure



Classification based on materials crystallinity

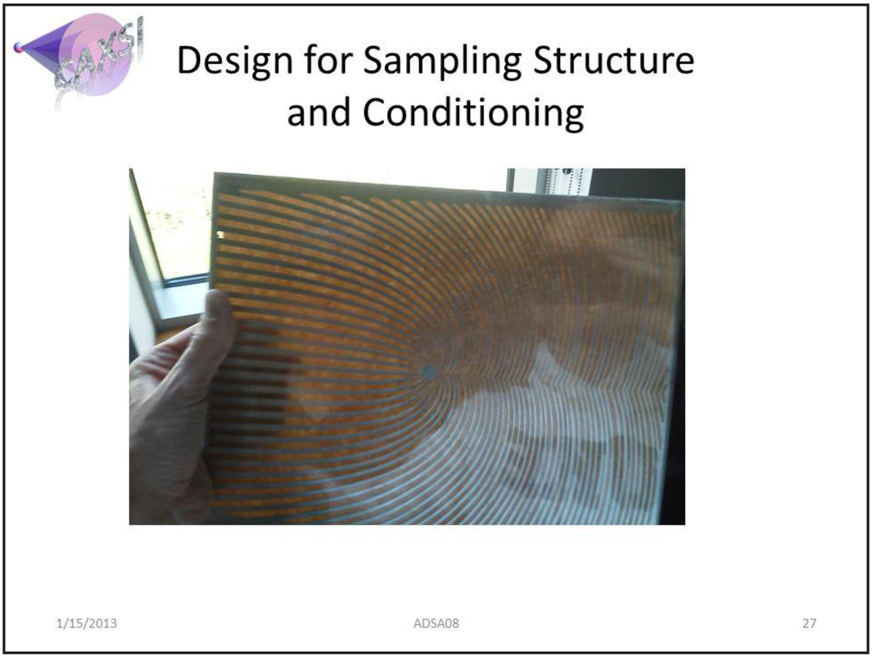
Indexing texture of diffraction pattern may aid in fine-tuning material classification



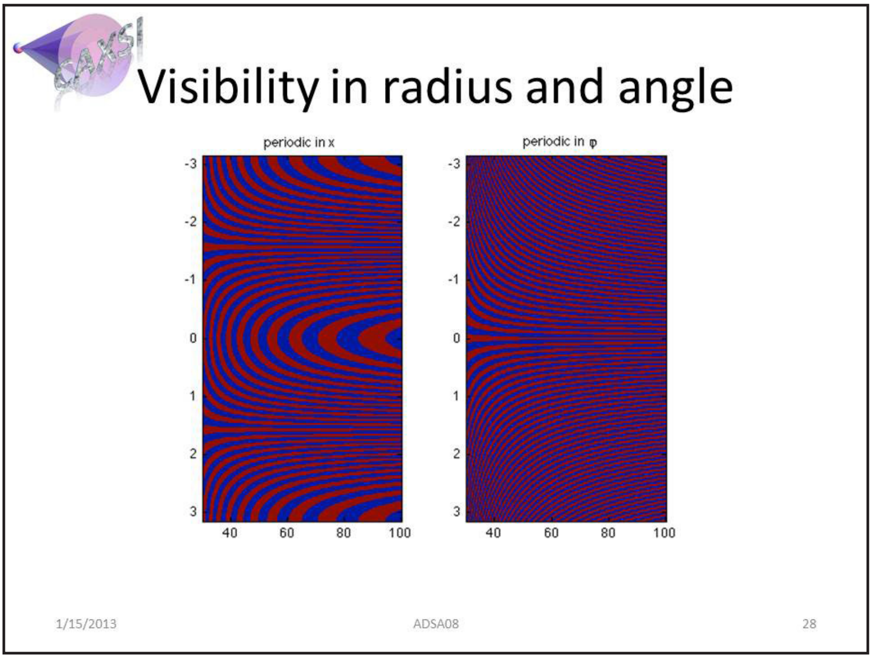
CAXSI Outline

- CAXSI System Vision
- Signature Analysis
  - Measurement space signature
    Forward models
  - Object space signature
    Reconstruction
  - Logical space signature
    SVD
- Conclusion

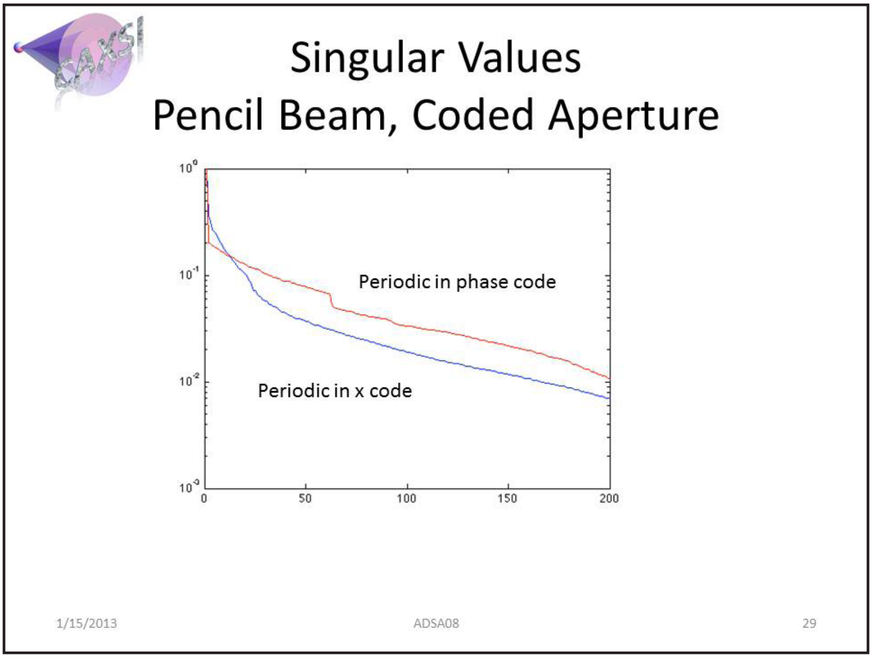
1/15/2013



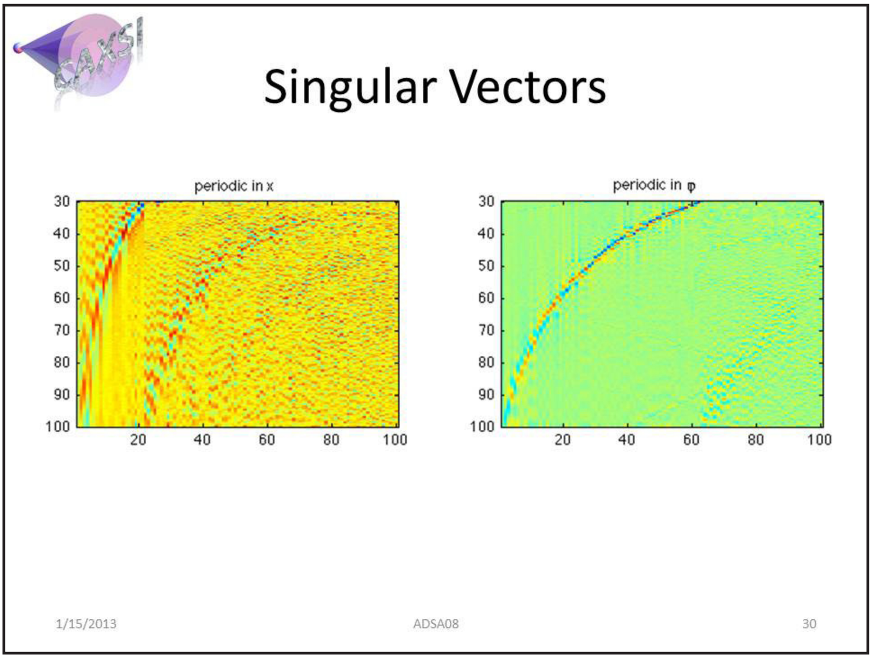
Design for Sampling Structure and Conditioning



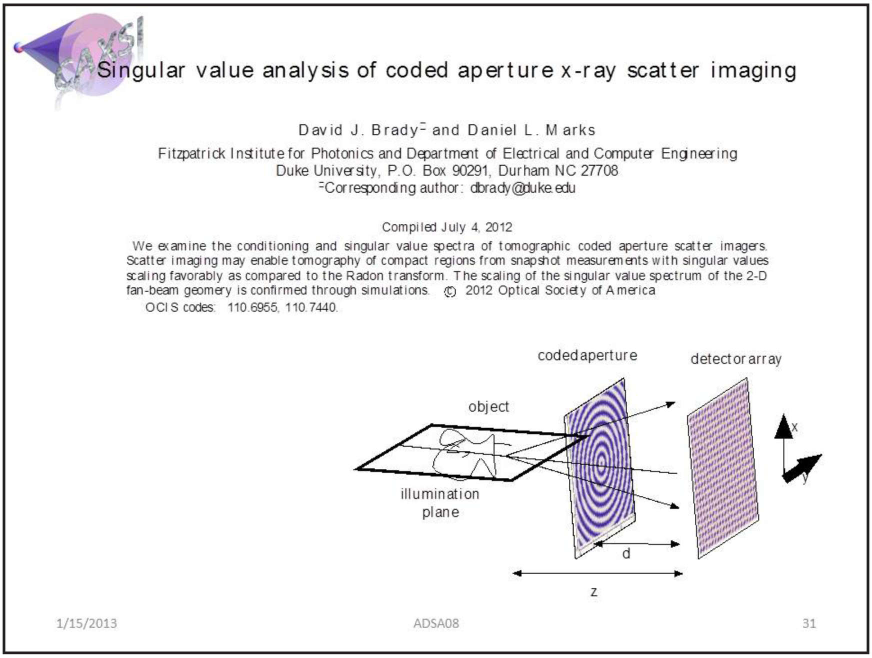
Visibility in radius and angle



Singular Values
Pencil Beam, Coded Aperture



Singular Vectors



Singular value analysis of coded aperture x-ray scatter imaging

David J. Brady* and Daniel L. Marks
Fitzpatrick Institute for Photonics and Department of Electrical and Computer Engineering
Duke University, P.O. Box 90291, Durham NC 27708
*Corresponding author: dbrady@duke.edu

Compiled July 4, 2012

We examine the conditioning and singular value spectra of tomographic coded aperture scatter imagers. Scatter imaging may enable tomography of compact regions from snapshot measurements with singular values scaling favorably as compared to the Radon transform. The scaling of the singular value spectrum of the 2-D fan-beam geomery is confirmed through simulations. © 2012 Optical Society of America

OCIS codes: 110.6955, 110.7440.

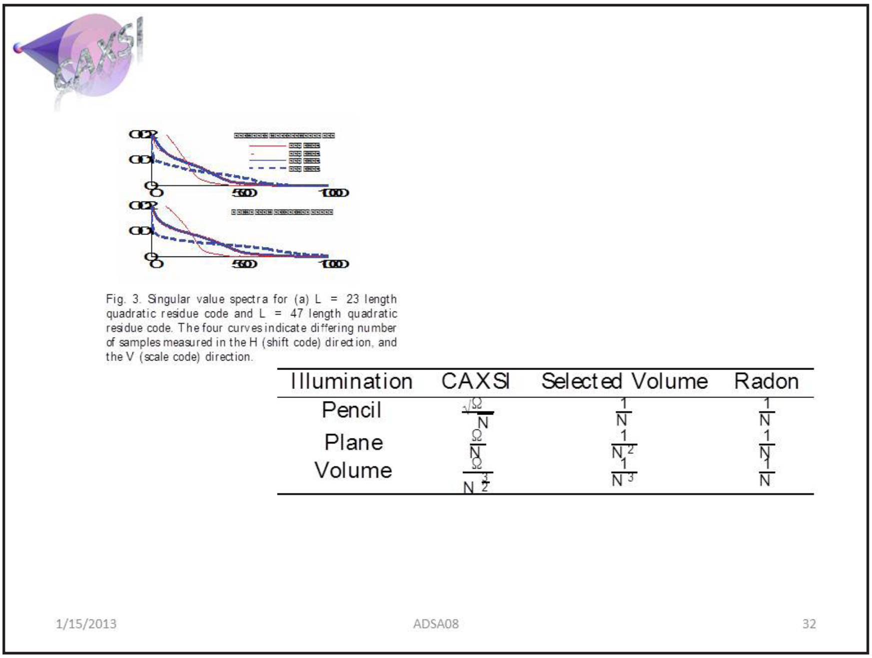
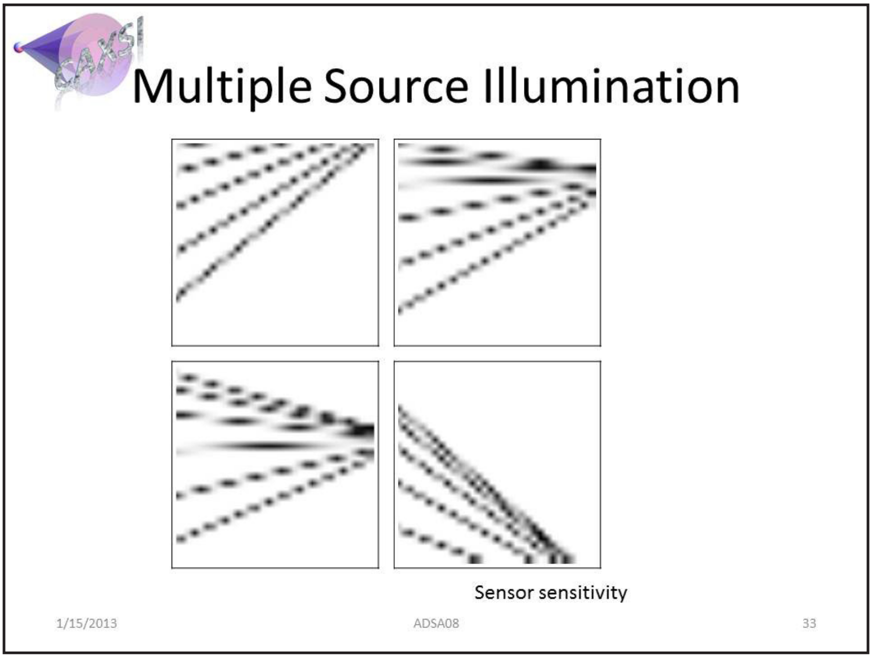


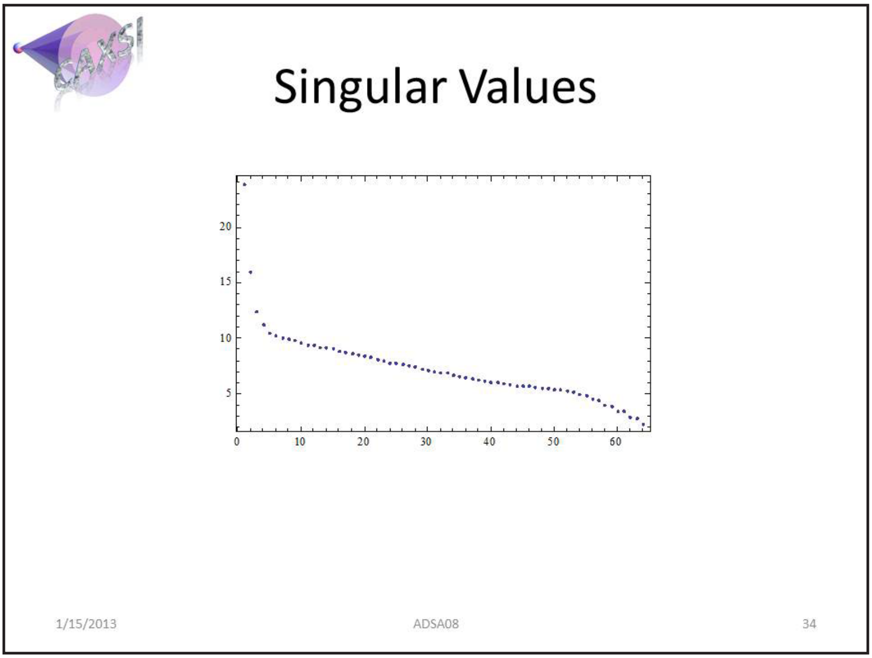
Fig. 3. Singular value spectra for (a) L = 23 length quadratic residue code and L = 47 length quadratic residue code. The four curves indicate differing number of samples measured in the H (shift code) direction, and the V (scale code) direction.

Illumination	CAXSI	Selected Volume	Radon
Pencil	$\frac{\sqrt{\Omega}}{N}$	$\frac{1}{N}$	$\frac{1}{N}$
Plane	$\frac{\Omega}{N}$	$\frac{1}{N^2}$	$\frac{1}{N}$
Volume	$\frac{\Omega}{N^{\frac{3}{2}}}$	$\frac{1}{N^3}$	$\frac{1}{N}$

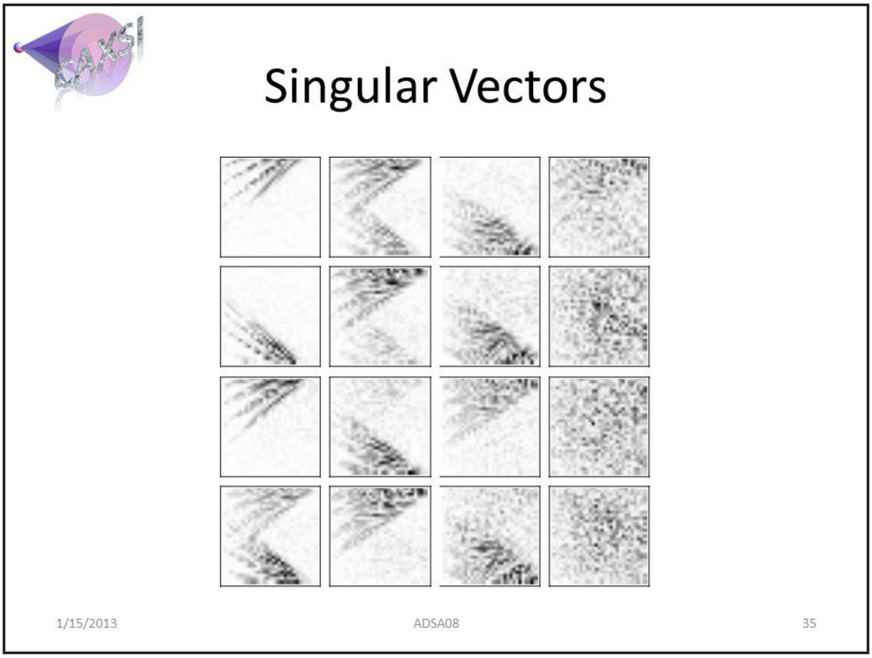


Multiple Source Illumination

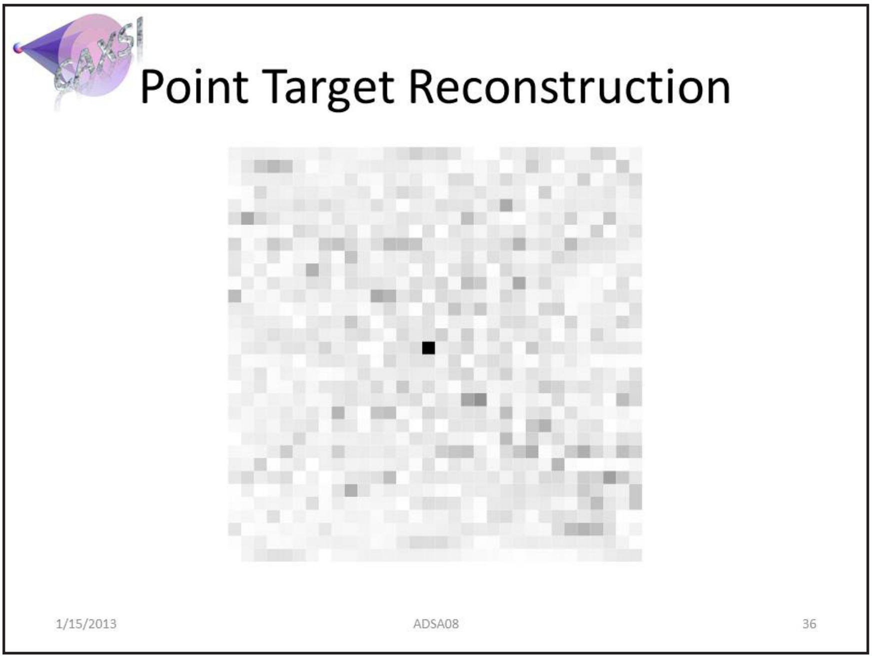
Sensor sensitivity



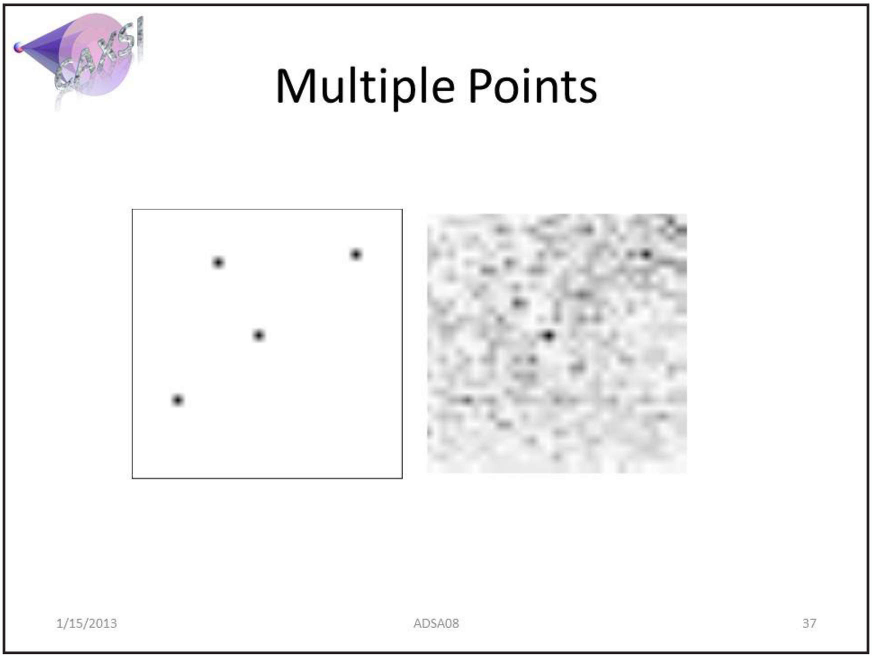
Singular Values



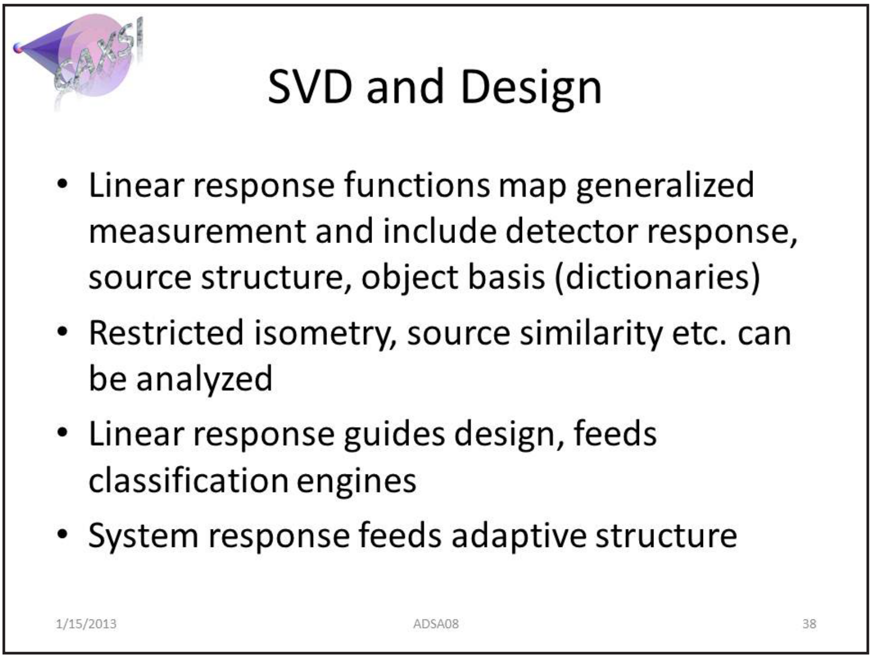
Singular Vectors



Point Target Reconstruction

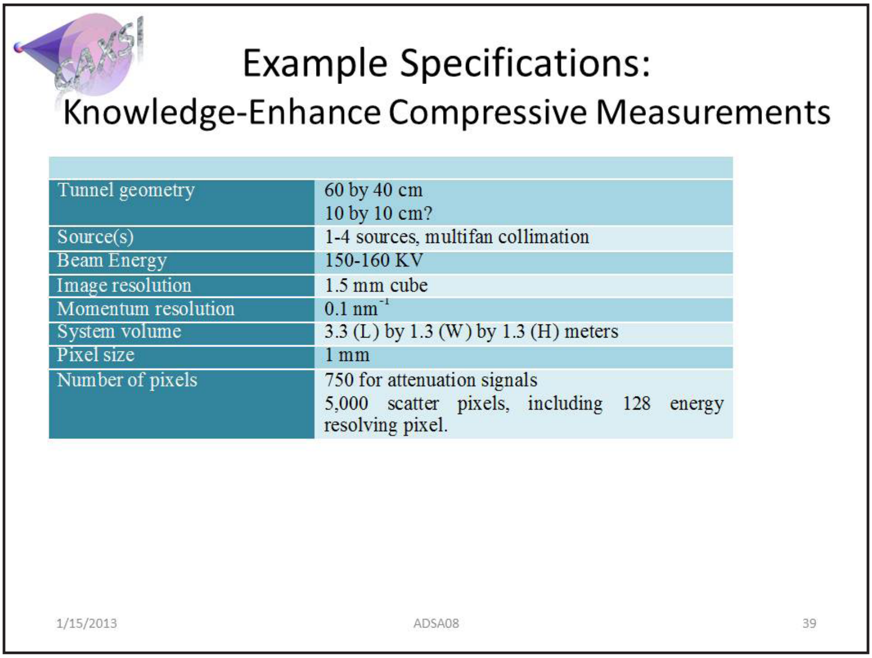


Multiple Points



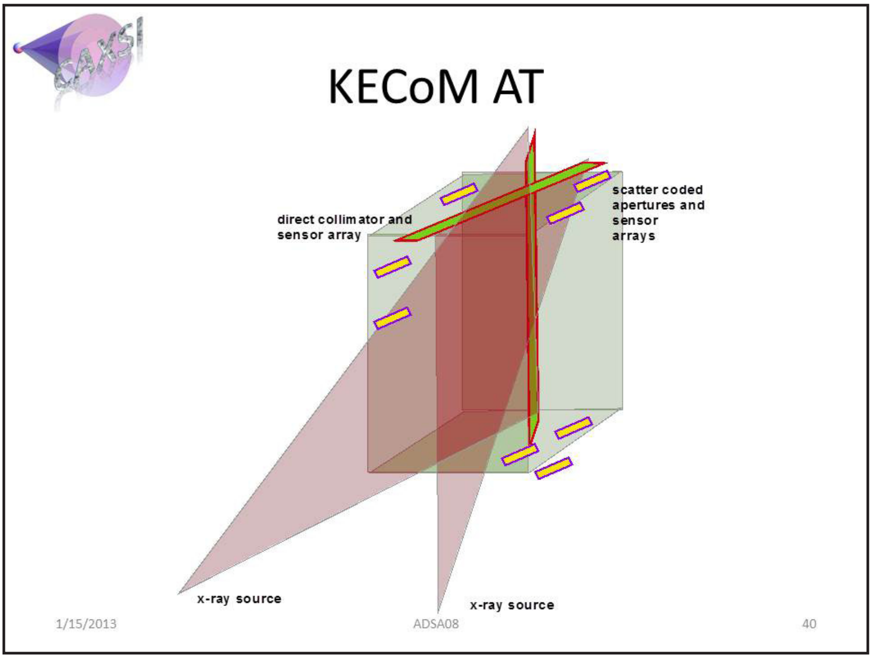
SVD and Design

- Linear response functions map generalized measurement and include detector response, source structure, object basis (dictionaries)
- Restricted isometry, source similarity etc. can be analyzed
- Linear response guides design, feeds classification engines
- System response feeds adaptive structure



Example Specifications:
Knowledge-Enhance Compressive Measurements

Tunnel geometry	60 by 40 cm 10 by 10 cm?
Source(s)	1-4 sources, multifan collimation
Beam Energy	150-160 KV
Image resolution	1.5 mm cube
Momentum resolution	0.1 nm^{-1}
System volume	3.3 (L) by 1.3 (W) by 1.3 (H) meters
Pixel size	1 mm
Number of pixels	750 for attenuation signals 5,000 scatter pixels, including 128 energy resolving pixel.



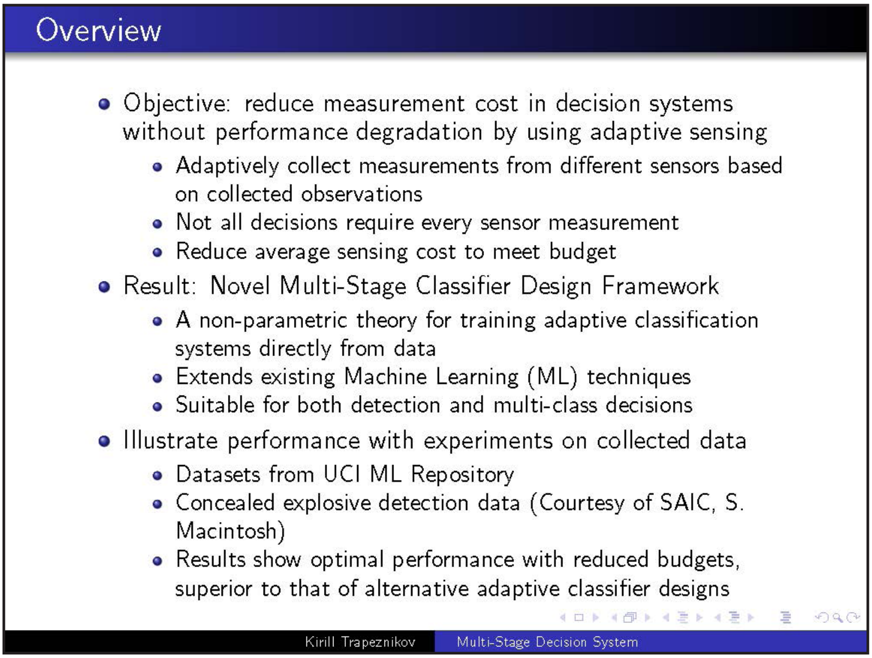
KECoM AT

16.11 Kirill Trapeznikov: Multi-Stage Decision Systems

Multi-Stage Decision System

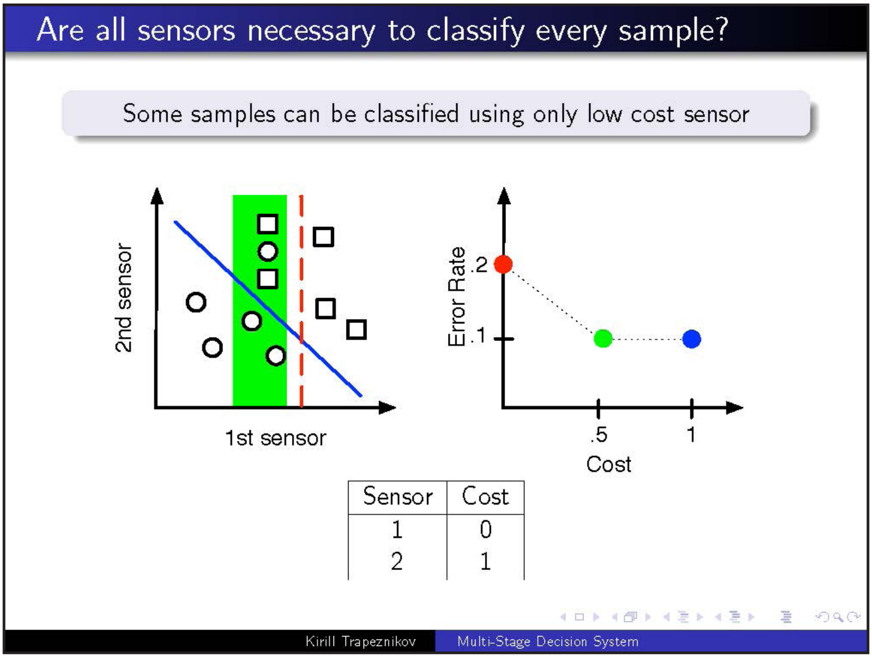
Kirill Trapeznikov
(Venkatesh Saligrama and David Castañón, Boston University)

October 24th, 2012



Overview

- Objective: reduce measurement cost in decision systems without performance degradation by using adaptive sensing
  - Adaptively collect measurements from different sensors based on collected observations
  - Not all decisions require every sensor measurement
  - Reduce average sensing cost to meet budget
- Result: Novel Multi-Stage Classifier Design Framework
  - A non-parametric theory for training adaptive classification systems directly from data
  - Extends existing Machine Learning (ML) techniques
  - Suitable for both detection and multi-class decisions
- Illustrate performance with experiments on collected data
  - Datasets from UCI ML Repository
  - Concealed explosive detection data (Courtesy of SAIC, S. Macintosh)
  - Results show optimal performance with reduced budgets, superior to that of alternative adaptive classifier designs

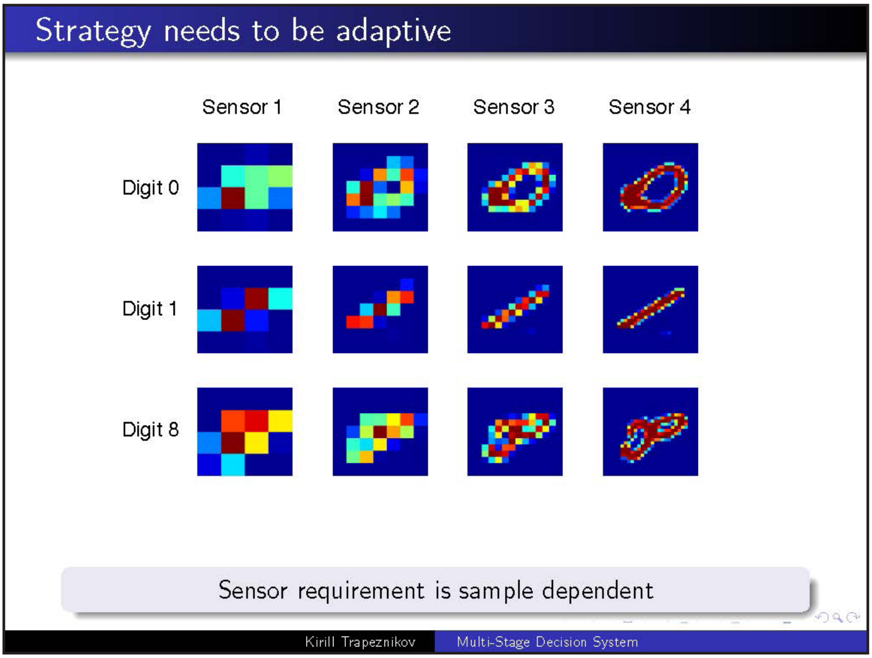


Are all sensors necessary to classify every sample?

Some samples can be classified using only low cost sensor

Sensor	Cost
1	0
2	1

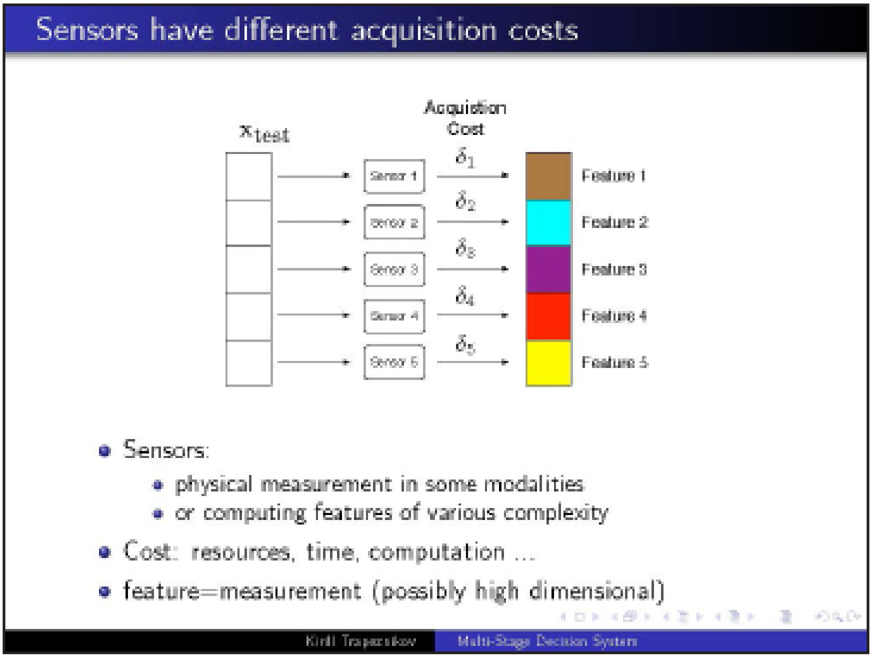
Kirill Trapeznikov   Multi-Stage Decision System



Strategy needs to be adaptive

Sensor requirement is sample dependent

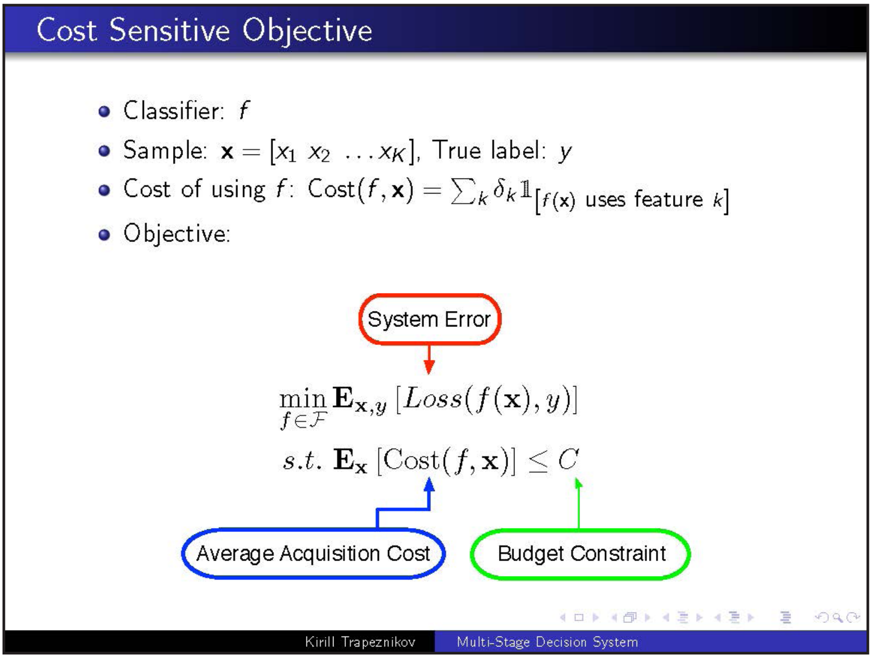
Kirill Trapeznikov   Multi-Stage Decision System



Sensors have different acquisition costs

- Sensors:
  - physical measurement in some modalities
  - or computing features of various complexity
- Cost: resources, time, computation ...
- feature=measurement (possibly high dimensional)

Kirill Trapeznikov Multi-Stage Decision System



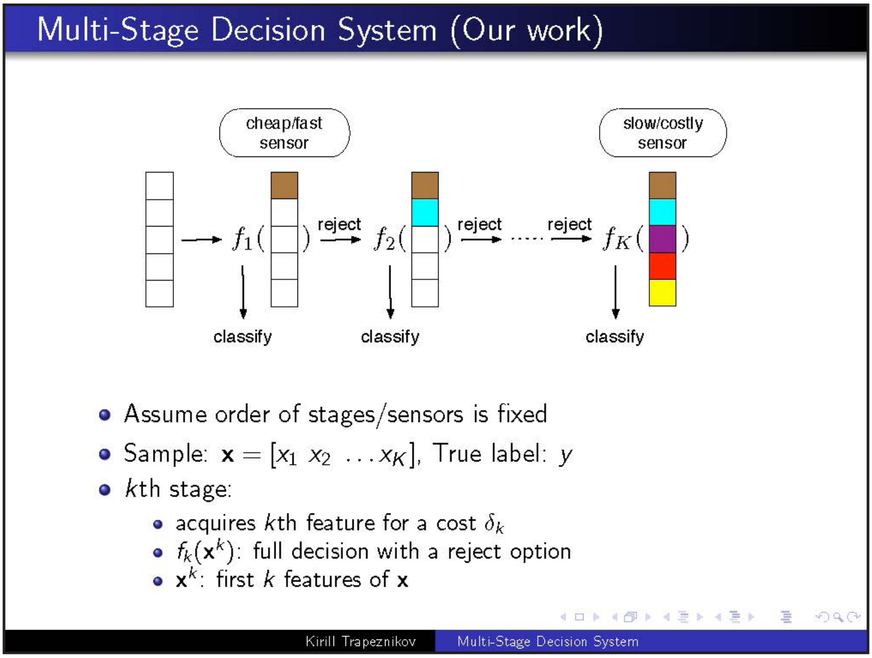
Cost Sensitive Objective

- Classifier: f
- Sample: $\mathbf{x} = [x_1\ x_2\ \dots x_K]$, True label: y
- Cost of using f: $\text{Cost}(f, \mathbf{x}) = \sum_k \delta_k \mathbb{1}_{[f(\mathbf{x}) \text{ uses feature } k]}$
- Objective:

$$\min_{f \in \mathcal{F}} \mathbf{E}_{\mathbf{x},y} \left[Loss(f(\mathbf{x}), y) \right]$$

$$s.t.\ \mathbf{E}_{\mathbf{x}} \left[\text{Cost}(f, \mathbf{x}) \right] \leq C$$

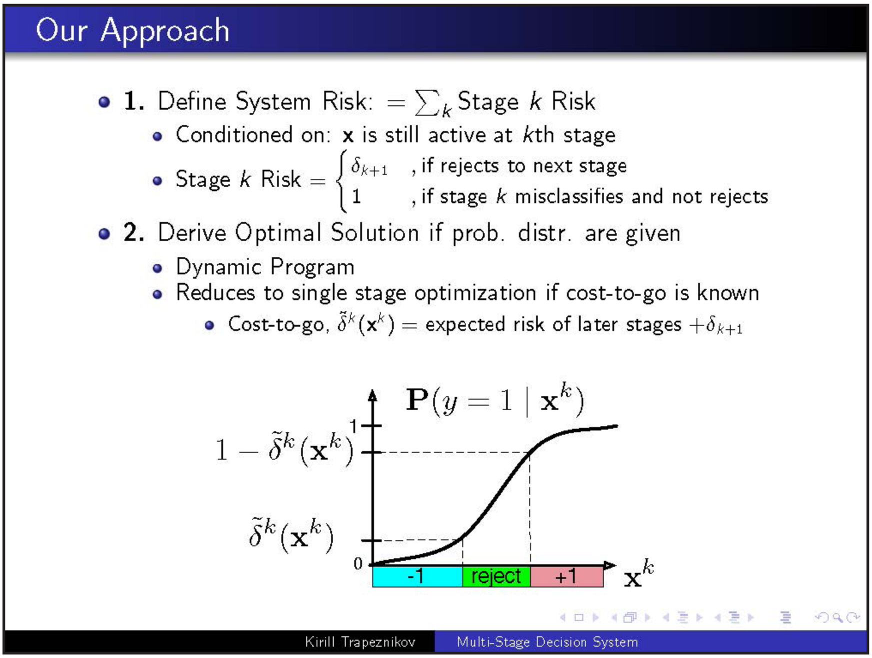
Kirill Trapeznikov Multi-Stage Decision System



Multi-Stage Decision System (Our work)

- Assume order of stages/sensors is fixed
- Sample: $\mathbf{x} = [x_1 \; x_2 \; \dots x_K]$, True label: y
- kth stage:
  - acquires kth feature for a cost δ_k
  - $f_k(\mathbf{x}^k)$: full decision with a reject option
  - $\mathbf{x}^k$: first k features of $\mathbf{x}$

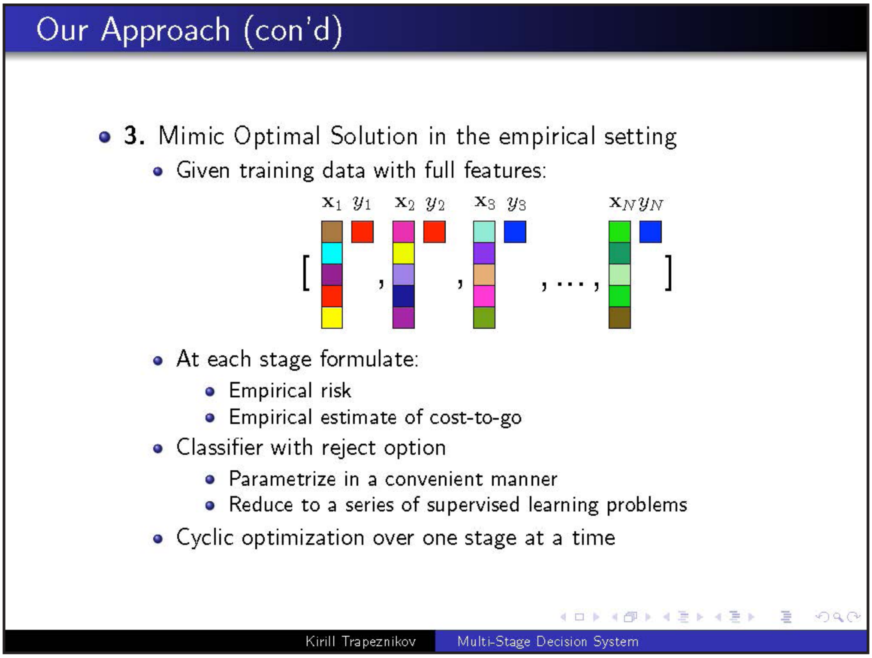
Kirill Trapeznikov | Multi-Stage Decision System



Our Approach

- **1.** Define System Risk: $= \sum_k$ Stage k Risk
  - Conditioned on: $\mathbf{x}$ is still active at kth stage
  - Stage k Risk $= \begin{cases} \delta_{k+1} & \text{, if rejects to next stage} \\ 1 & \text{, if stage } k \text{ misclassifies and not rejects} \end{cases}$
- **2.** Derive Optimal Solution if prob. distr. are given
  - Dynamic Program
  - Reduces to single stage optimization if cost-to-go is known
    - Cost-to-go, $\tilde{\delta}^k(\mathbf{x}^k) =$ expected risk of later stages $+\delta_{k+1}$

Kirill Trapeznikov | Multi-Stage Decision System



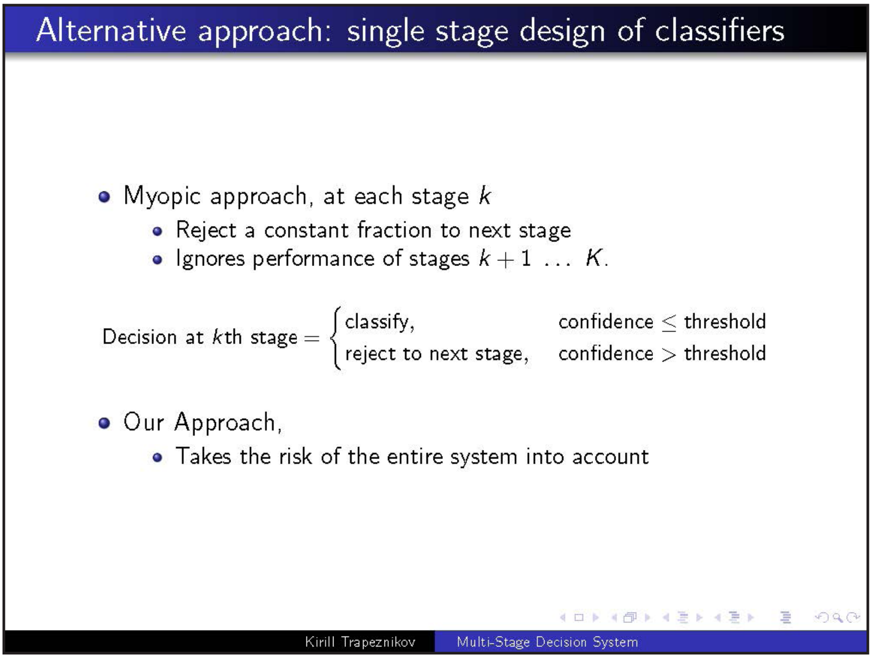
Our Approach (con'd)

- **3.** Mimic Optimal Solution in the empirical setting
  - Given training data with full features:

$$\left[\ \underset{\mathbf{x}_1\ y_1}{}\ ,\ \underset{\mathbf{x}_2\ y_2}{}\ ,\ \underset{\mathbf{x}_3\ y_3}{}\ ,\dots,\ \underset{\mathbf{x}_N y_N}{}\ \right]$$

  - At each stage formulate:
    - Empirical risk
    - Empirical estimate of cost-to-go
  - Classifier with reject option
    - Parametrize in a convenient manner
    - Reduce to a series of supervised learning problems
  - Cyclic optimization over one stage at a time

Kirill Trapeznikov | Multi-Stage Decision System



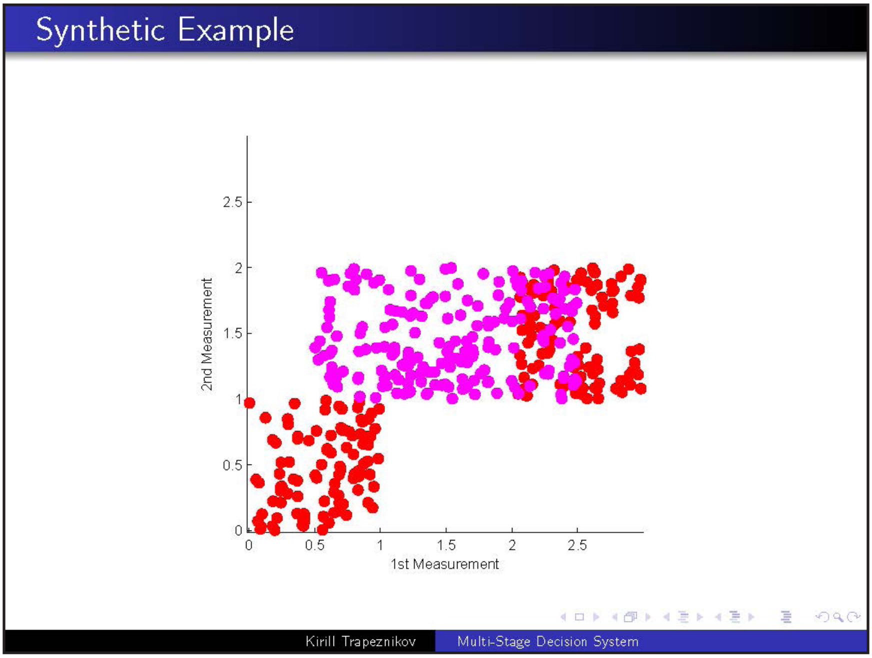
Alternative approach: single stage design of classifiers

- Myopic approach, at each stage k
  - Reject a constant fraction to next stage
  - Ignores performance of stages $k+1\ \dots\ K$.

$$\text{Decision at } k\text{th stage} = \begin{cases} \text{classify,} & \text{confidence} \leq \text{threshold} \\ \text{reject to next stage,} & \text{confidence} > \text{threshold} \end{cases}$$

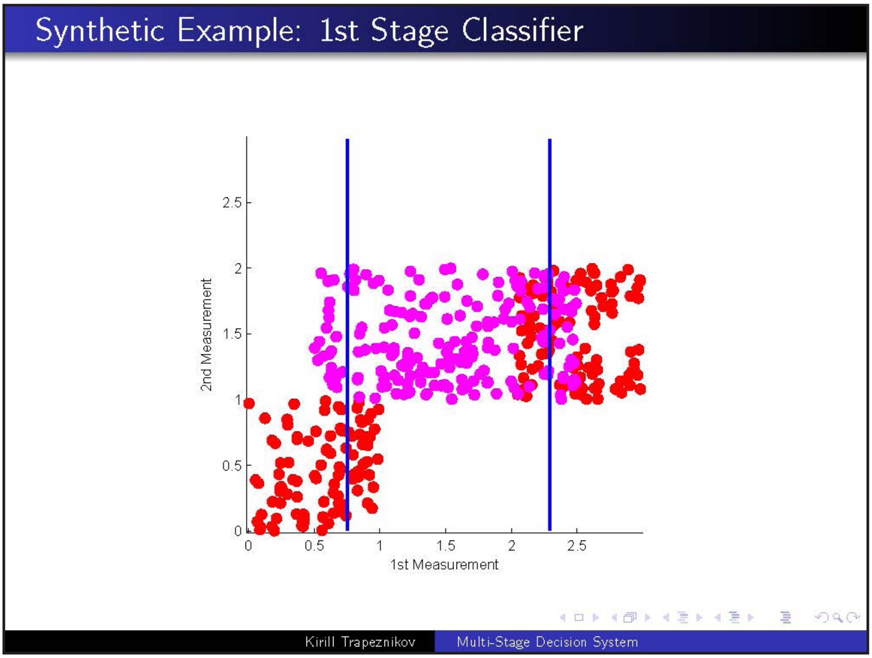
- Our Approach,
  - Takes the risk of the entire system into account

Kirill Trapeznikov | Multi-Stage Decision System



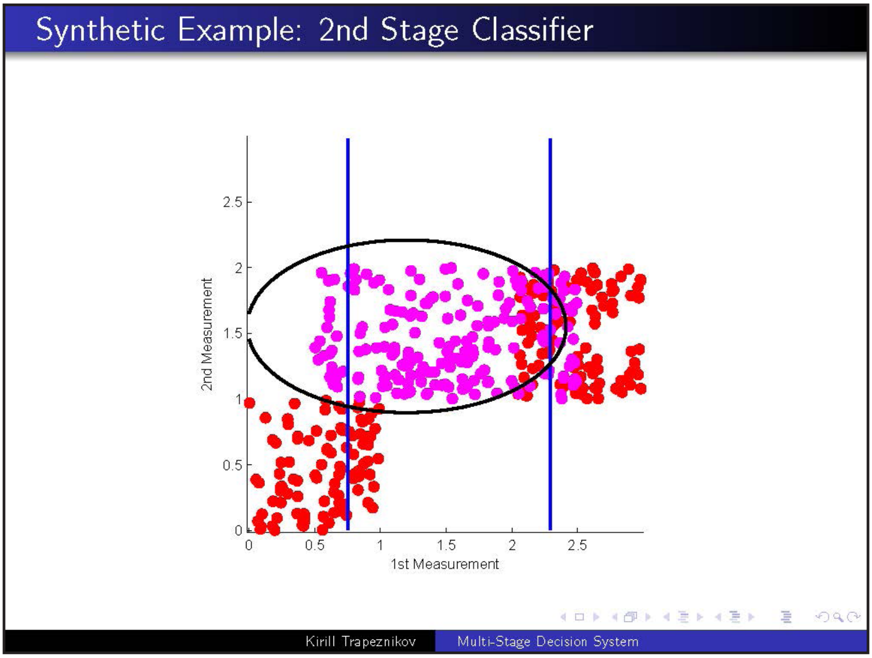
Synthetic Example

Kirill Trapeznikov Multi-Stage Decision System



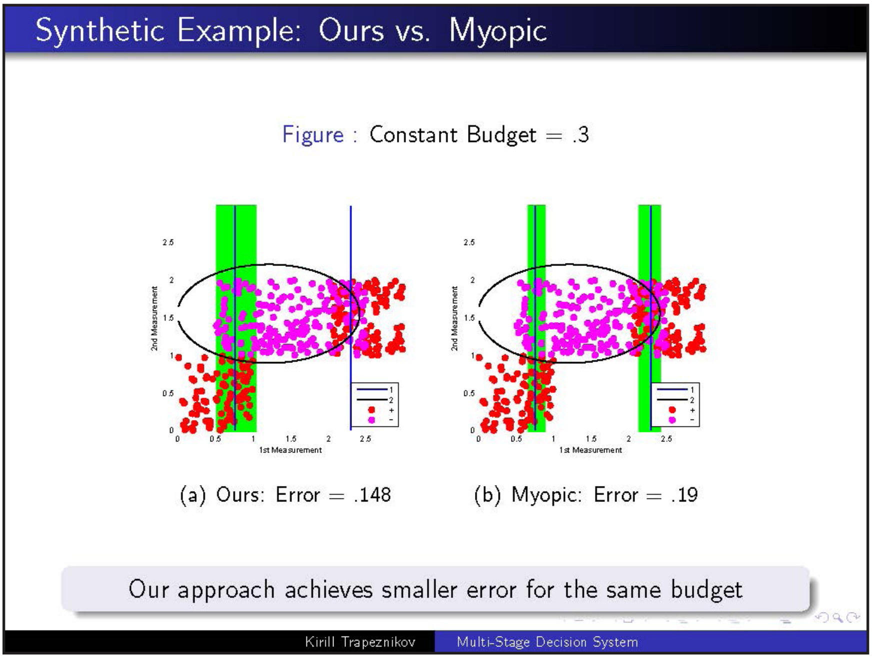
Synthetic Example: 1st Stage Classifier

Kirill Trapeznikov Multi-Stage Decision System



Synthetic Example: 2nd Stage Classifier

Kirill Trapeznikov | Multi-Stage Decision System



Synthetic Example: Ours vs. Myopic

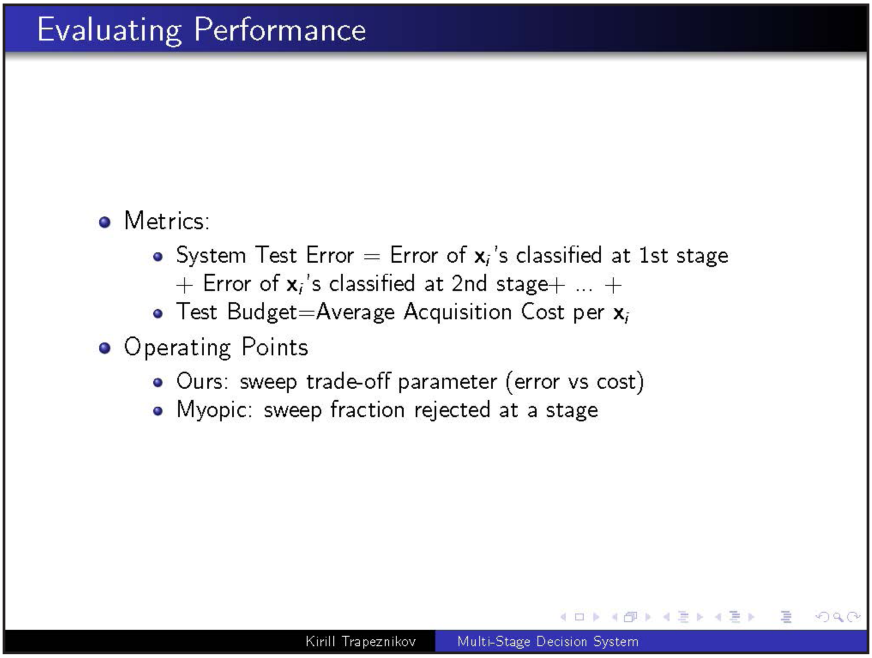
Figure : Constant Budget = .3

(a) Ours: Error = .148

(b) Myopic: Error = .19

Our approach achieves smaller error for the same budget

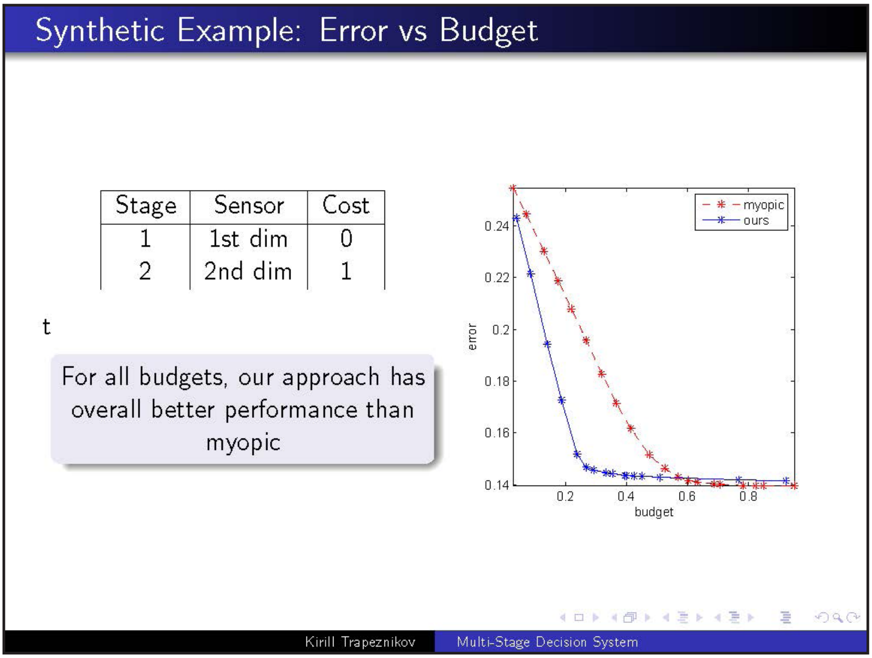
Kirill Trapeznikov | Multi-Stage Decision System



Evaluating Performance

- Metrics:
  - System Test Error = Error of $\mathbf{x}_i$'s classified at 1st stage + Error of $\mathbf{x}_i$'s classified at 2nd stage+ ... +
  - Test Budget=Average Acquisition Cost per $\mathbf{x}_i$
- Operating Points
  - Ours: sweep trade-off parameter (error vs cost)
  - Myopic: sweep fraction rejected at a stage

Kirill Trapeznikov | Multi-Stage Decision System



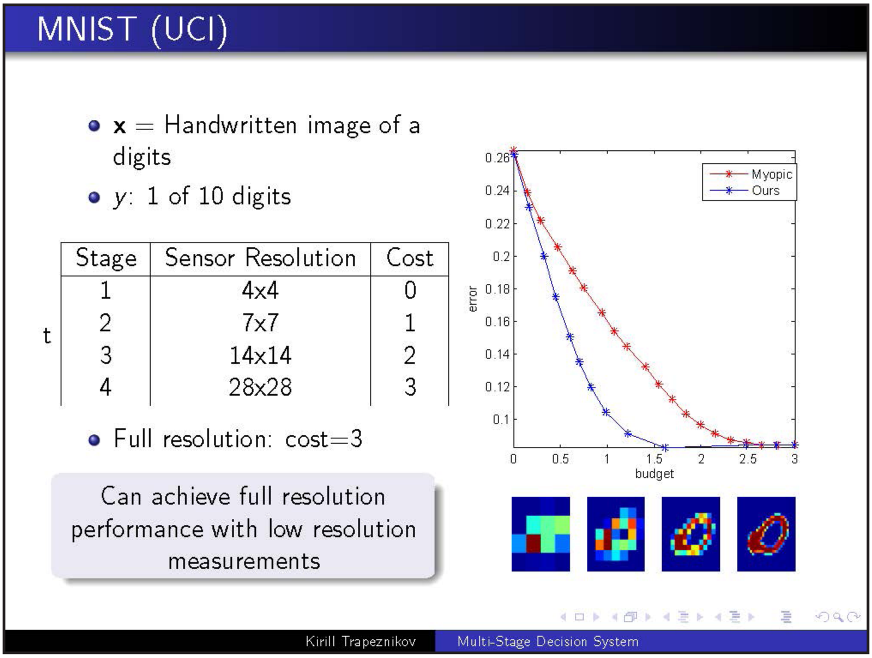
Synthetic Example: Error vs Budget

Stage	Sensor	Cost
1	1st dim	0
2	2nd dim	1

t

For all budgets, our approach has overall better performance than myopic

Kirill Trapeznikov | Multi-Stage Decision System



MNIST (UCI)

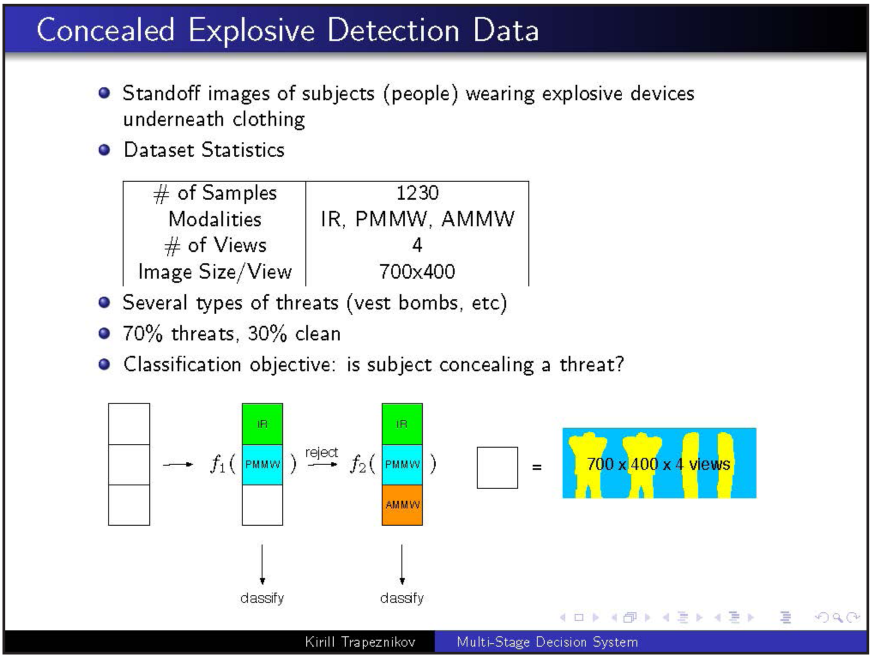
- **x** = Handwritten image of a digits
- y: 1 of 10 digits

	Stage	Sensor Resolution	Cost
t	1	4x4	0
	2	7x7	1
	3	14x14	2
	4	28x28	3

- Full resolution: cost=3

Can achieve full resolution performance with low resolution measurements

Kirill Trapeznikov | Multi-Stage Decision System



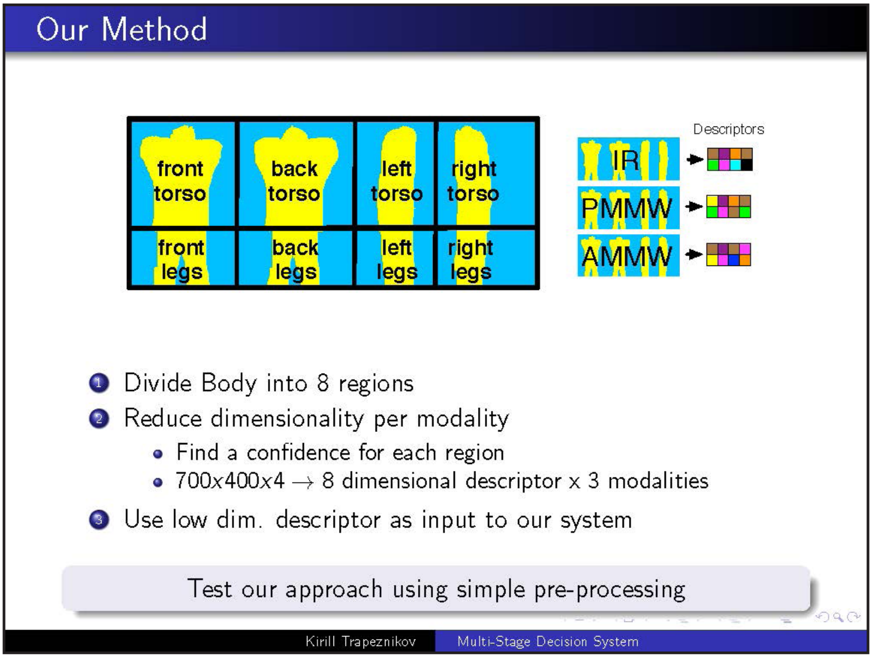
Concealed Explosive Detection Data

- Standoff images of subjects (people) wearing explosive devices underneath clothing
- Dataset Statistics

# of Samples	1230
Modalities	IR, PMMW, AMMW
# of Views	4
Image Size/View	700x400

- Several types of threats (vest bombs, etc)
- 70% threats, 30% clean
- Classification objective: is subject concealing a threat?

Kirill Trapeznikov | Multi-Stage Decision System

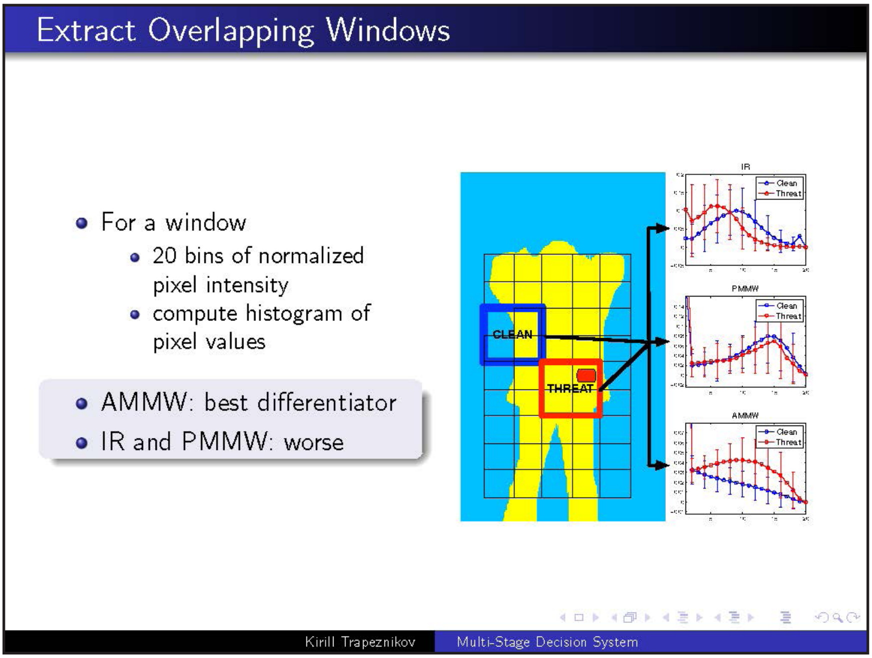


Our Method

1. Divide Body into 8 regions
2. Reduce dimensionality per modality
   - Find a confidence for each region
   - 700x400x4 → 8 dimensional descriptor x 3 modalities
3. Use low dim. descriptor as input to our system

Test our approach using simple pre-processing

Kirill Trapeznikov Multi-Stage Decision System

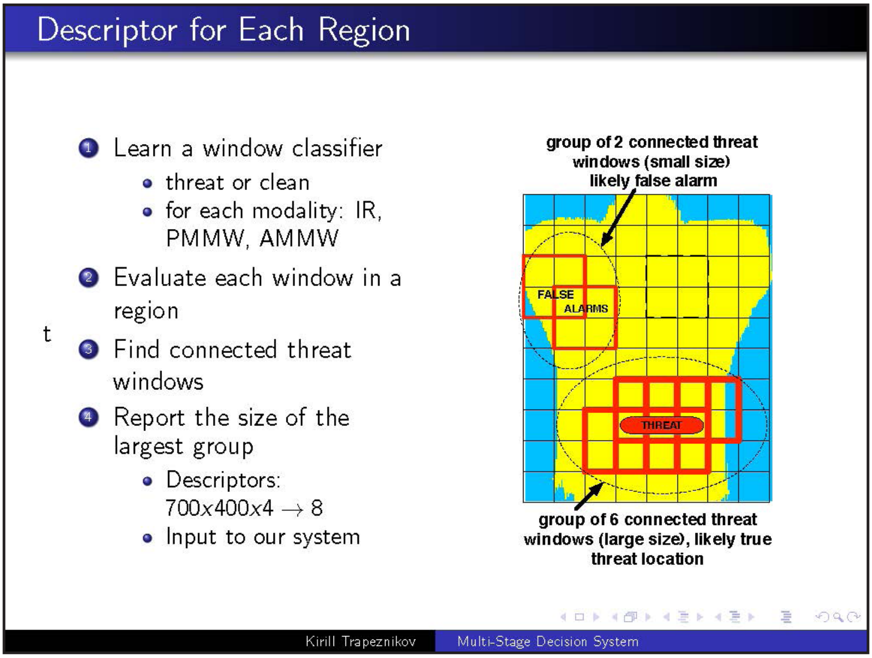


Extract Overlapping Windows

- For a window
  - 20 bins of normalized pixel intensity
  - compute histogram of pixel values

- AMMW: best differentiator
- IR and PMMW: worse

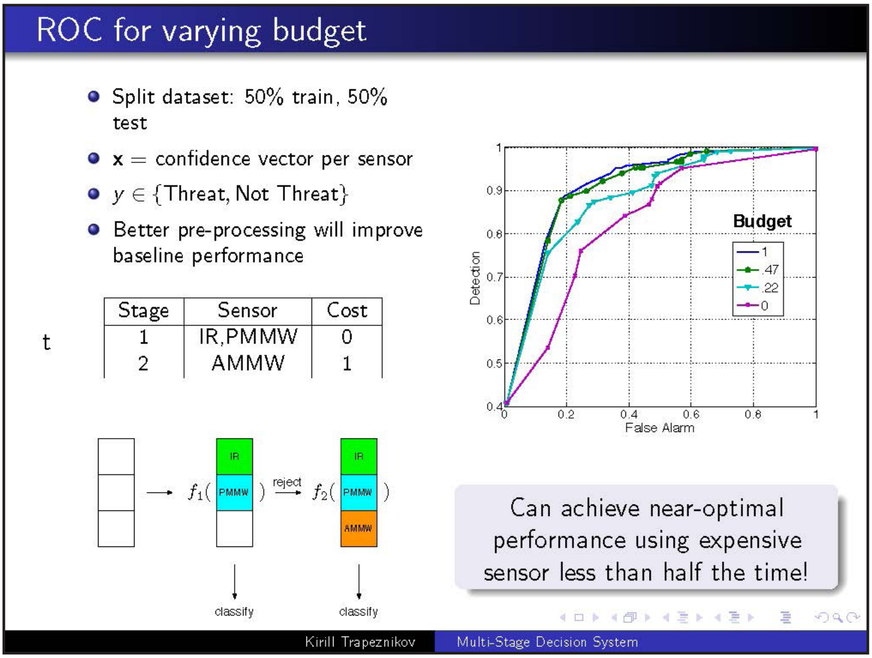
Kirill Trapeznikov Multi-Stage Decision System



Descriptor for Each Region

1. Learn a window classifier
    - threat or clean
    - for each modality: IR, PMMW, AMMW
2. Evaluate each window in a region
3. Find connected threat windows
4. Report the size of the largest group
    - Descriptors: 700x400x4 → 8
    - Input to our system

t



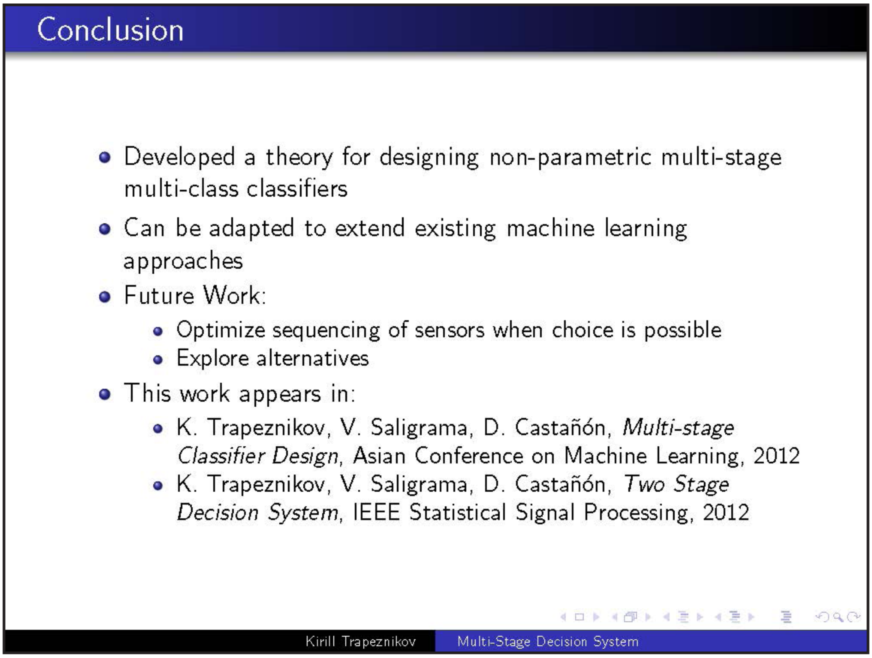
ROC for varying budget

- Split dataset: 50% train, 50% test
- $\mathbf{x}$ = confidence vector per sensor
- $y \in \{\text{Threat, Not Threat}\}$
- Better pre-processing will improve baseline performance

Stage	Sensor	Cost
1	IR,PMMW	0
2	AMMW	1

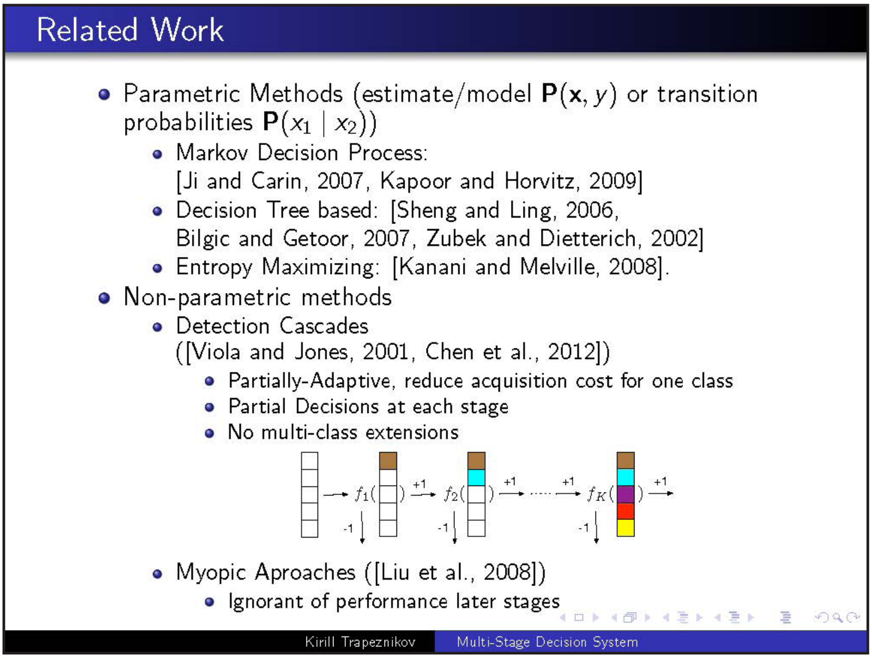
t

Can achieve near-optimal performance using expensive sensor less than half the time!



Conclusion

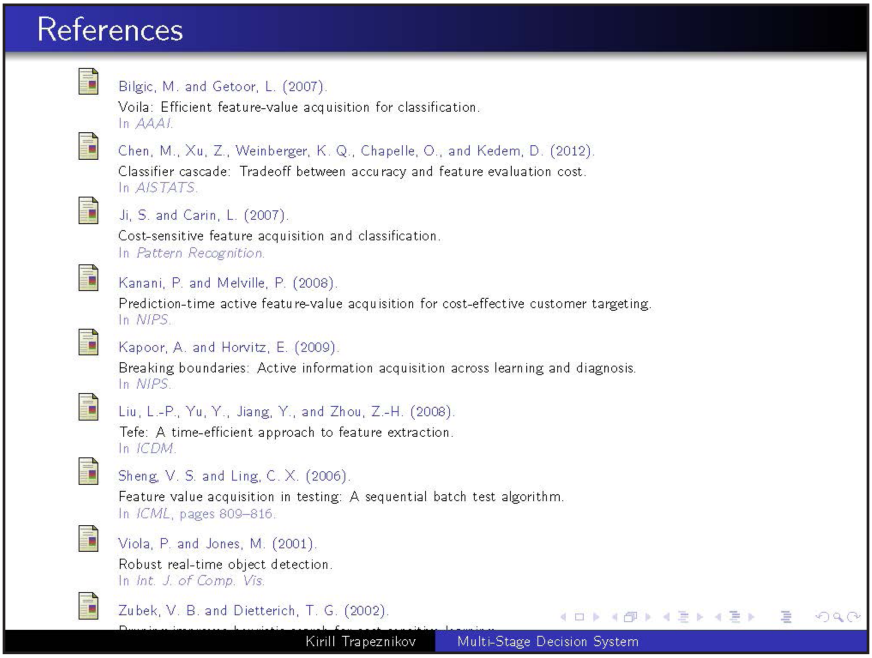
- Developed a theory for designing non-parametric multi-stage multi-class classifiers
- Can be adapted to extend existing machine learning approaches
- Future Work:
  - Optimize sequencing of sensors when choice is possible
  - Explore alternatives
- This work appears in:
  - K. Trapeznikov, V. Saligrama, D. Castañón, *Multi-stage Classifier Design*, Asian Conference on Machine Learning, 2012
  - K. Trapeznikov, V. Saligrama, D. Castañón, *Two Stage Decision System*, IEEE Statistical Signal Processing, 2012



Related Work

- Parametric Methods (estimate/model $\mathbf{P}(\mathbf{x}, y)$ or transition probabilities $\mathbf{P}(x_1 \mid x_2)$)
  - Markov Decision Process:
    [Ji and Carin, 2007, Kapoor and Horvitz, 2009]
  - Decision Tree based: [Sheng and Ling, 2006, Bilgic and Getoor, 2007, Zubek and Dietterich, 2002]
  - Entropy Maximizing: [Kanani and Melville, 2008].
- Non-parametric methods
  - Detection Cascades ([Viola and Jones, 2001, Chen et al., 2012])
    - Partially-Adaptive, reduce acquisition cost for one class
    - Partial Decisions at each stage
    - No multi-class extensions

  - Myopic Aproaches ([Liu et al., 2008])
    - Ignorant of performance later stages



References

Bilgic, M. and Getoor, L. (2007).
Voila: Efficient feature-value acquisition for classification.
In *AAAI*.

Chen, M., Xu, Z., Weinberger, K. Q., Chapelle, O., and Kedem, D. (2012).
Classifier cascade: Tradeoff between accuracy and feature evaluation cost.
In *AISTATS*.

Ji, S. and Carin, L. (2007).
Cost-sensitive feature acquisition and classification.
In *Pattern Recognition*.

Kanani, P. and Melville, P. (2008).
Prediction-time active feature-value acquisition for cost-effective customer targeting.
In *NIPS*.

Kapoor, A. and Horvitz, E. (2009).
Breaking boundaries: Active information acquisition across learning and diagnosis.
In *NIPS*.

Liu, L.-P., Yu, Y., Jiang, Y., and Zhou, Z.-H. (2008).
Tefe: A time-efficient approach to feature extraction.
In *ICDM*.

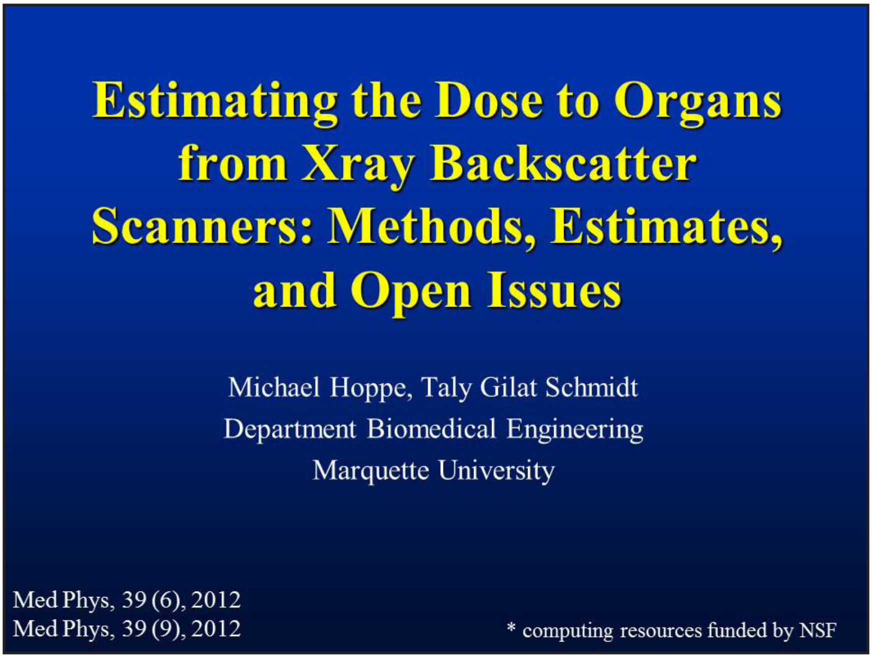
Sheng, V. S. and Ling, C. X. (2006).
Feature value acquisition in testing: A sequential batch test algorithm.
In *ICML*, pages 809–816.

Viola, P. and Jones, M. (2001).
Robust real-time object detection.
In *Int. J. of Comp. Vis.*

Zubek, V. B. and Dietterich, T. G. (2002).

Kirill Trapeznikov Multi-Stage Decision System

16.12 Taly Gilat-Schmidt: X-Ray Backscatter Dose Predictions

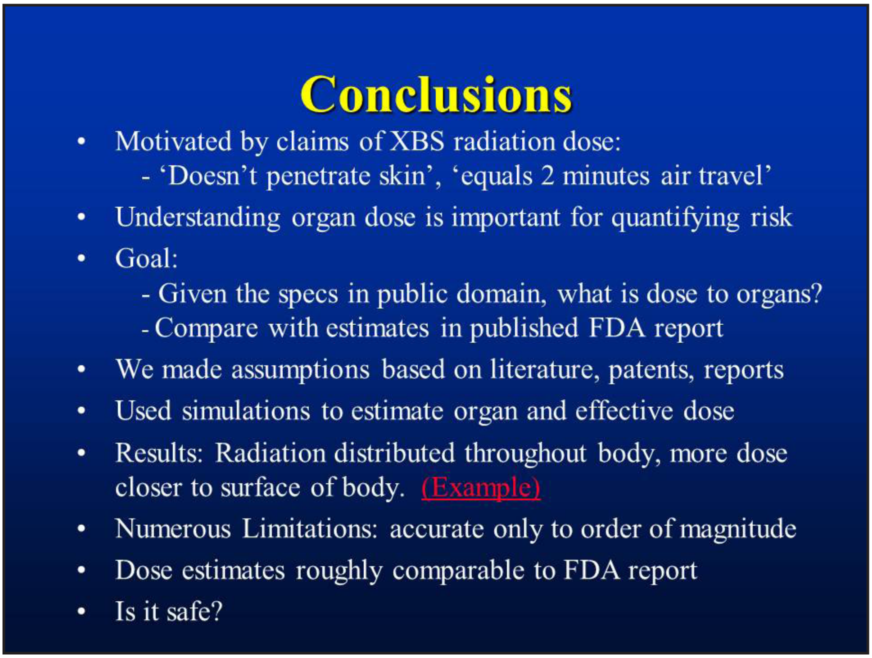


Estimating the Dose to Organs from Xray Backscatter Scanners: Methods, Estimates, and Open Issues

Michael Hoppe, Taly Gilat Schmidt
Department Biomedical Engineering
Marquette University

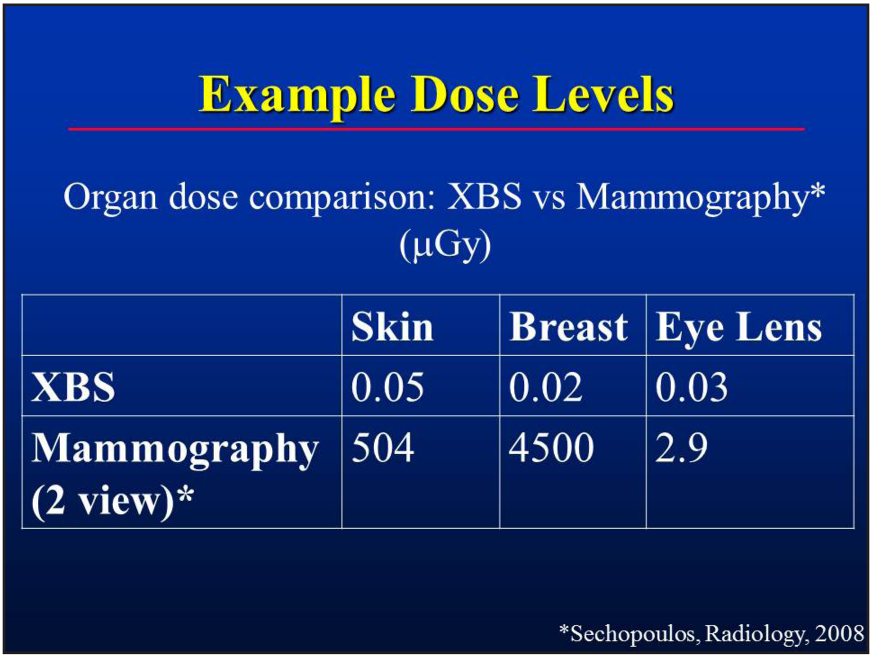
Med Phys, 39 (6), 2012
Med Phys, 39 (9), 2012

* computing resources funded by NSF



Conclusions

- Motivated by claims of XBS radiation dose:
  - 'Doesn't penetrate skin', 'equals 2 minutes air travel'
- Understanding organ dose is important for quantifying risk
- Goal:
  - Given the specs in public domain, what is dose to organs?
  - Compare with estimates in published FDA report
- We made assumptions based on literature, patents, reports
- Used simulations to estimate organ and effective dose
- Results: Radiation distributed throughout body, more dose closer to surface of body. (Example)
- Numerous Limitations: accurate only to order of magnitude
- Dose estimates roughly comparable to FDA report
- Is it safe?

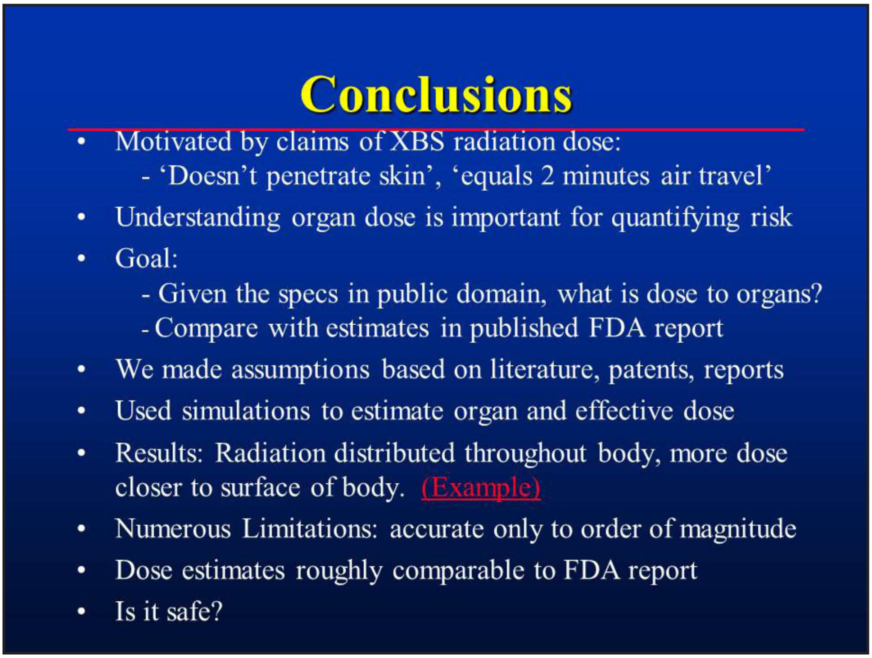


Example Dose Levels

Organ dose comparison: XBS vs Mammography*
(μGy)

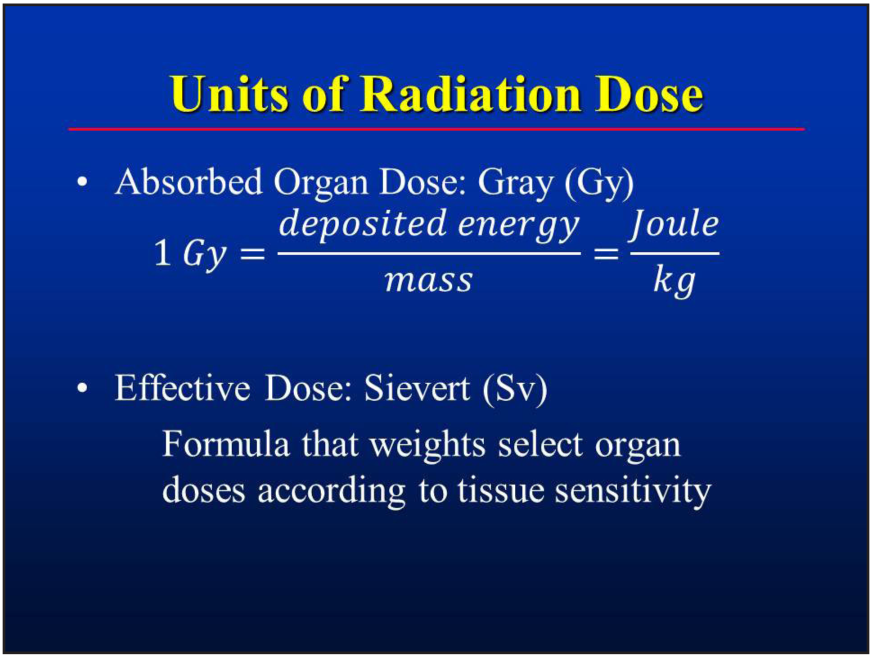
	Skin	Breast	Eye Lens
XBS	0.05	0.02	0.03
Mammography (2 view)*	504	4500	2.9

*Sechopoulos, Radiology, 2008



Conclusions

- Motivated by claims of XBS radiation dose:
  - ‘Doesn’t penetrate skin’, ‘equals 2 minutes air travel’
- Understanding organ dose is important for quantifying risk
- Goal:
  - Given the specs in public domain, what is dose to organs?
  - Compare with estimates in published FDA report
- We made assumptions based on literature, patents, reports
- Used simulations to estimate organ and effective dose
- Results: Radiation distributed throughout body, more dose closer to surface of body. (Example)
- Numerous Limitations: accurate only to order of magnitude
- Dose estimates roughly comparable to FDA report
- Is it safe?



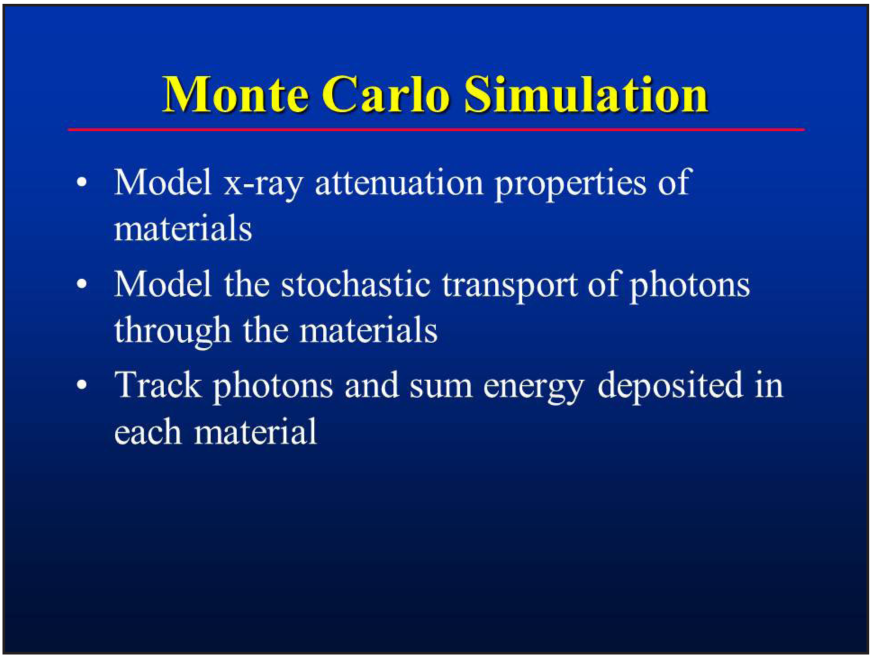
Units of Radiation Dose

- Absorbed Organ Dose: Gray (Gy)

$$1\ Gy = \frac{deposited\ energy}{mass} = \frac{Joule}{kg}$$

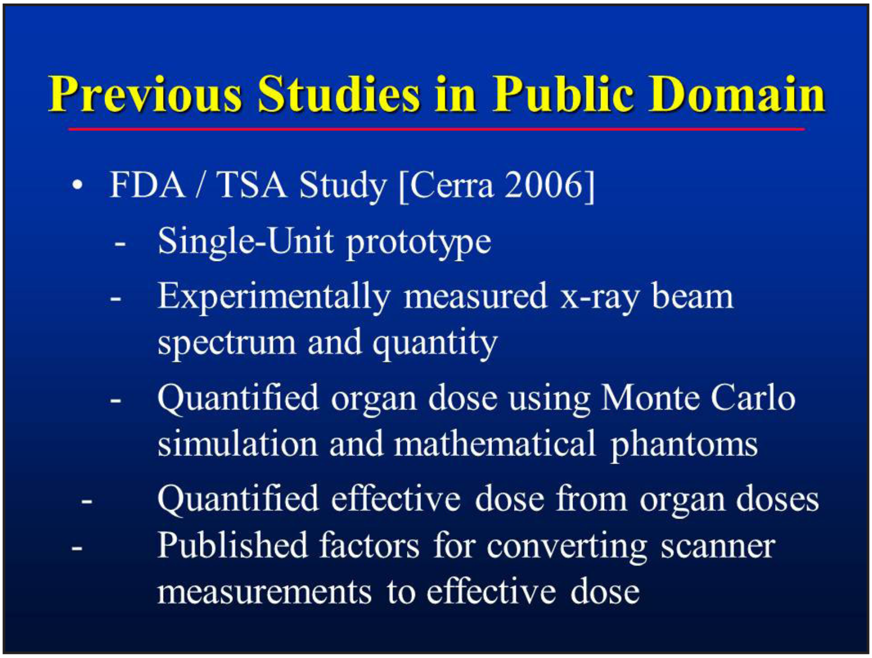
- Effective Dose: Sievert (Sv)

  Formula that weights select organ doses according to tissue sensitivity



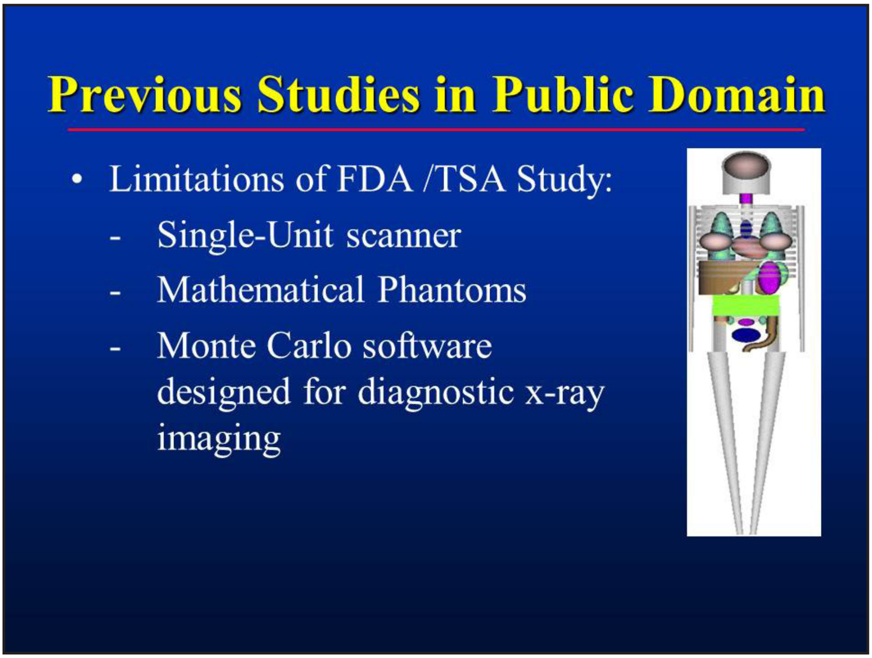
Monte Carlo Simulation

- Model x-ray attenuation properties of materials
- Model the stochastic transport of photons through the materials
- Track photons and sum energy deposited in each material



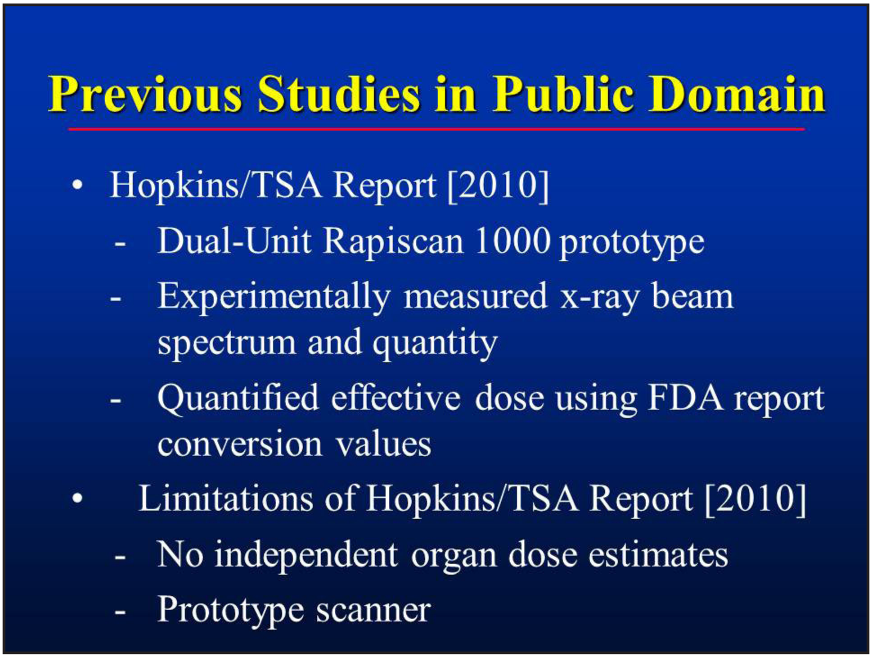
Previous Studies in Public Domain

- FDA / TSA Study [Cerra 2006]
  - Single-Unit prototype
  - Experimentally measured x-ray beam spectrum and quantity
  - Quantified organ dose using Monte Carlo simulation and mathematical phantoms
  - Quantified effective dose from organ doses
  - Published factors for converting scanner measurements to effective dose



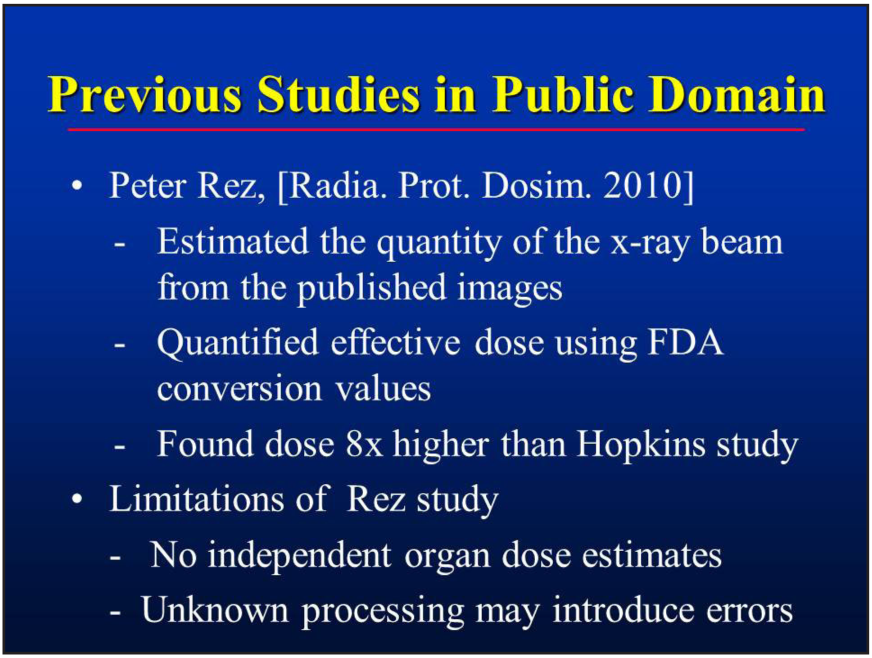
Previous Studies in Public Domain

- Limitations of FDA /TSA Study:
  - Single-Unit scanner
  - Mathematical Phantoms
  - Monte Carlo software designed for diagnostic x-ray imaging



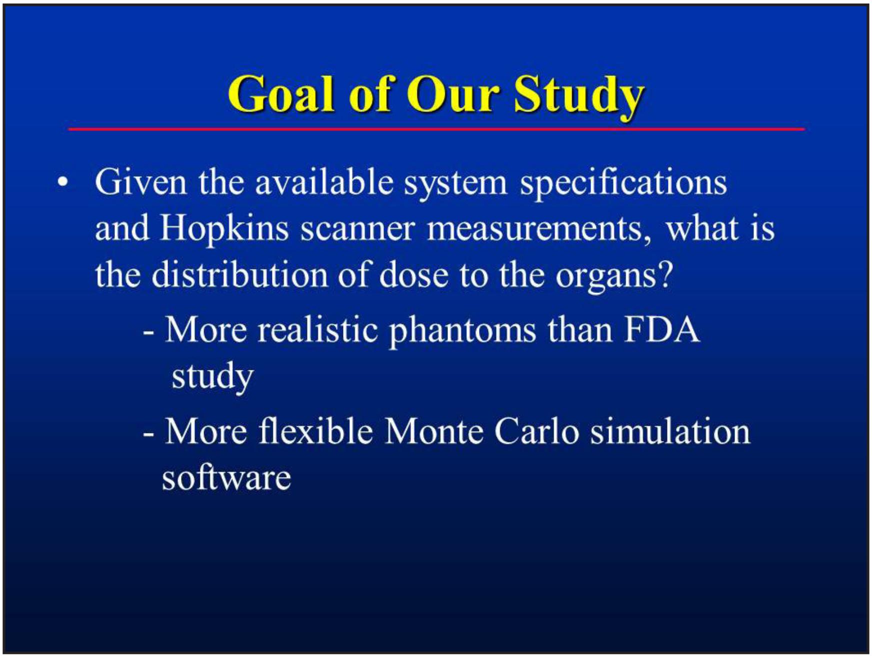
Previous Studies in Public Domain

- Hopkins/TSA Report [2010]
  - Dual-Unit Rapiscan 1000 prototype
  - Experimentally measured x-ray beam spectrum and quantity
  - Quantified effective dose using FDA report conversion values
- Limitations of Hopkins/TSA Report [2010]
  - No independent organ dose estimates
  - Prototype scanner



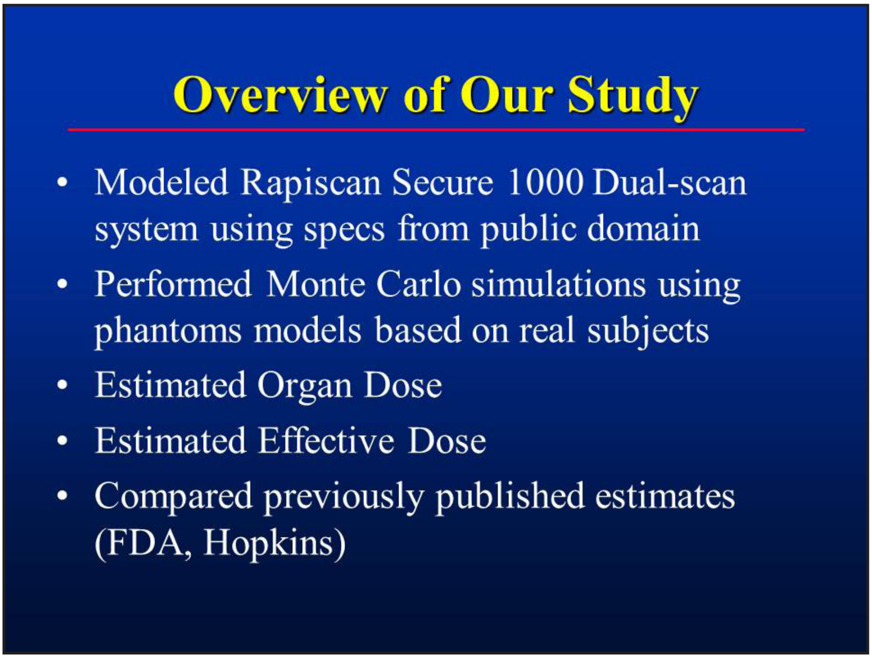
Previous Studies in Public Domain

- Peter Rez, [Radia. Prot. Dosim. 2010]
  - Estimated the quantity of the x-ray beam from the published images
  - Quantified effective dose using FDA conversion values
  - Found dose 8x higher than Hopkins study
- Limitations of Rez study
  - No independent organ dose estimates
  - Unknown processing may introduce errors



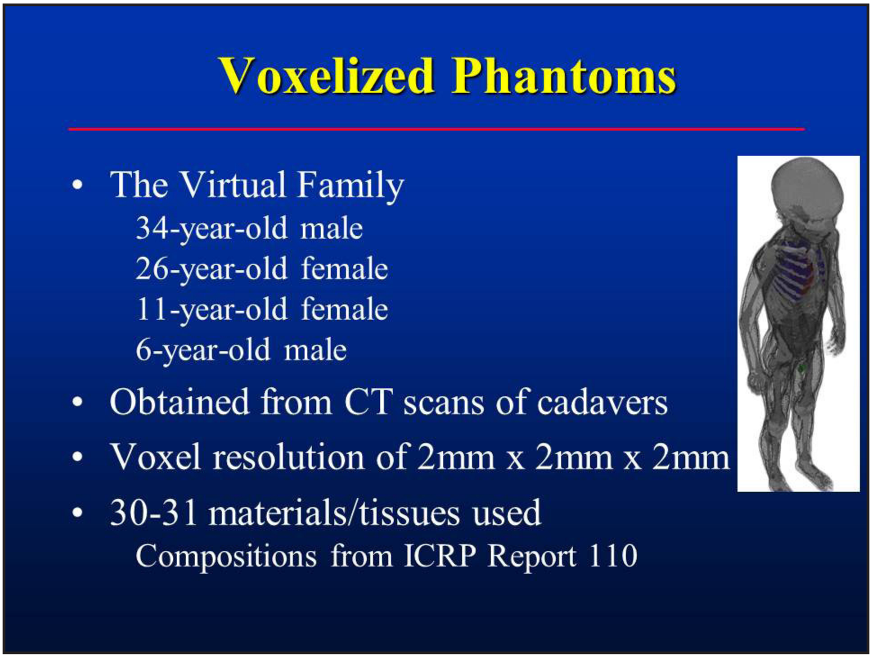
Goal of Our Study

- Given the available system specifications and Hopkins scanner measurements, what is the distribution of dose to the organs?
  - More realistic phantoms than FDA study
  - More flexible Monte Carlo simulation software



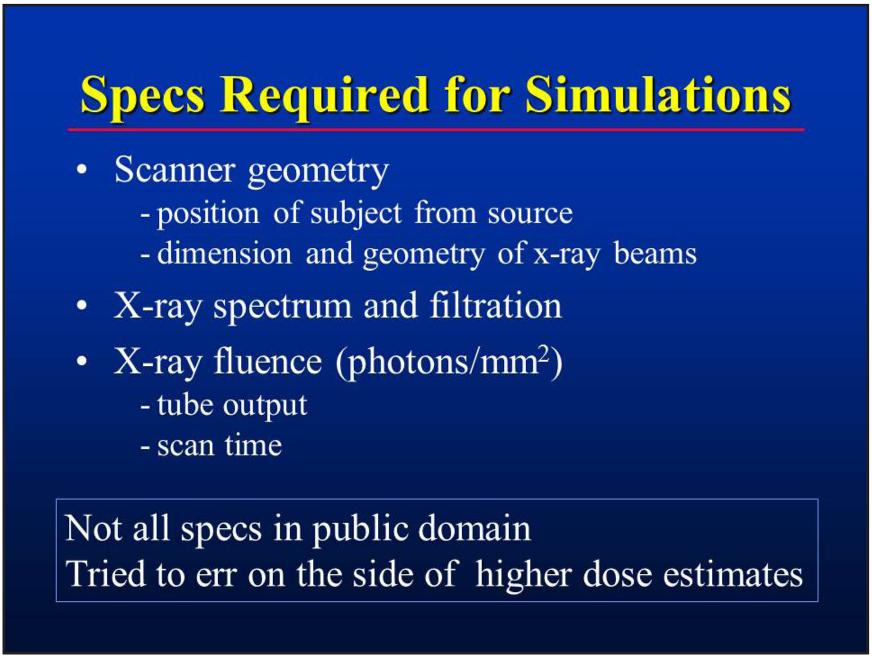
Overview of Our Study

- Modeled Rapiscan Secure 1000 Dual-scan system using specs from public domain
- Performed Monte Carlo simulations using phantoms models based on real subjects
- Estimated Organ Dose
- Estimated Effective Dose
- Compared previously published estimates (FDA, Hopkins)



Voxelized Phantoms

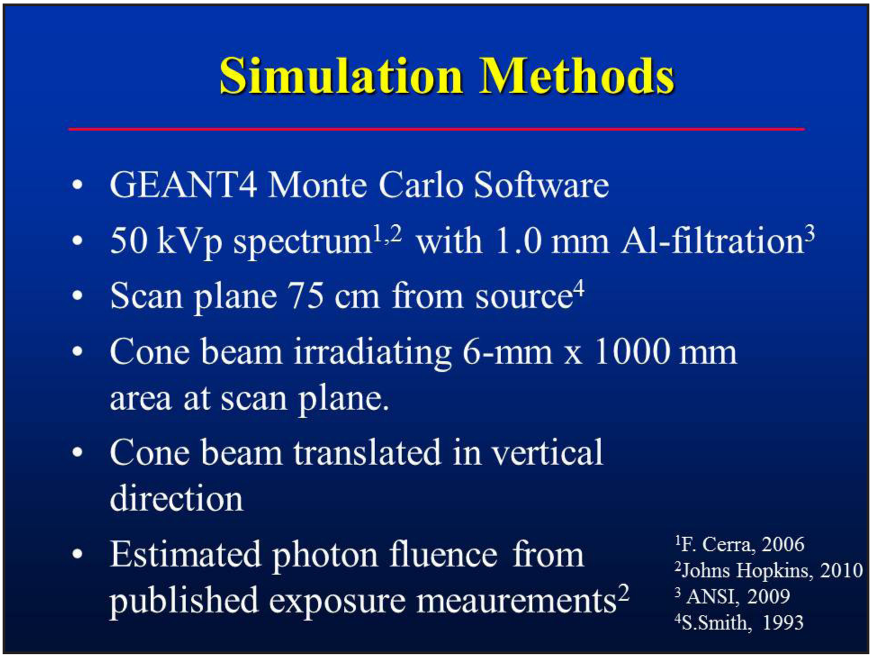
- The Virtual Family
  - 34-year-old male
  - 26-year-old female
  - 11-year-old female
  - 6-year-old male
- Obtained from CT scans of cadavers
- Voxel resolution of 2mm x 2mm x 2mm
- 30-31 materials/tissues used
  - Compositions from ICRP Report 110



Specs Required for Simulations

- Scanner geometry
  - position of subject from source
  - dimension and geometry of x-ray beams
- X-ray spectrum and filtration
- X-ray fluence (photons/mm^2)
  - tube output
  - scan time

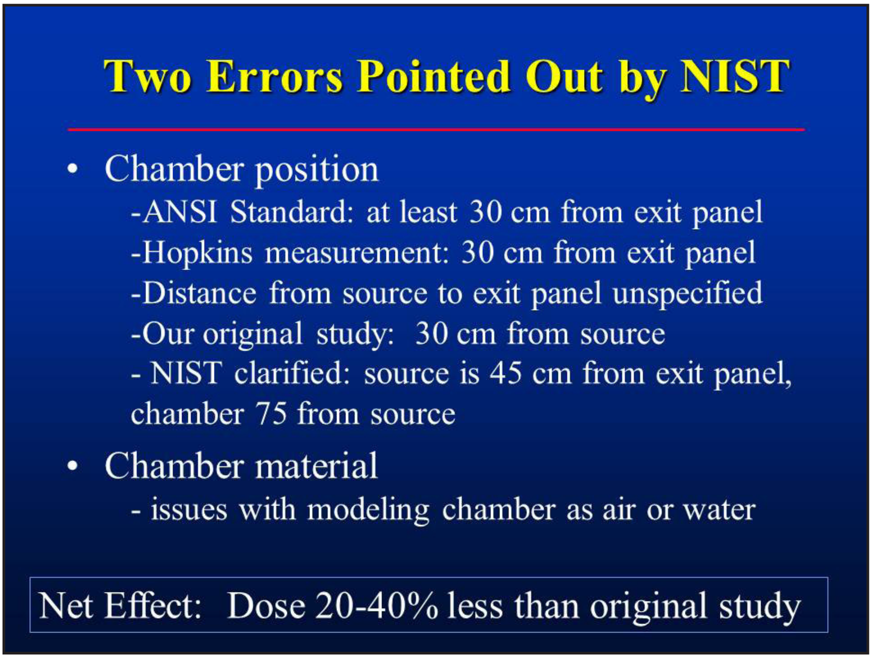
Not all specs in public domain
Tried to err on the side of higher dose estimates



Simulation Methods

- GEANT4 Monte Carlo Software
- 50 kVp spectrum[1,2] with 1.0 mm Al-filtration[3]
- Scan plane 75 cm from source[4]
- Cone beam irradiating 6-mm x 1000 mm area at scan plane.
- Cone beam translated in vertical direction
- Estimated photon fluence from published exposure meaurements[2]

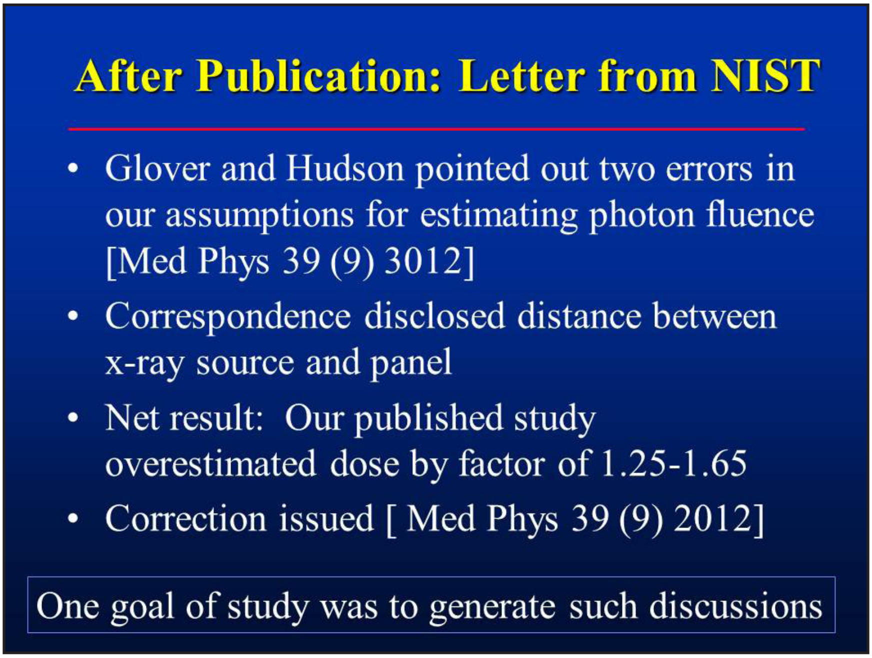
[1]F. Cerra, 2006
[2]Johns Hopkins, 2010
[3] ANSI, 2009
[4]S.Smith, 1993



Two Errors Pointed Out by NIST

- Chamber position
  - -ANSI Standard: at least 30 cm from exit panel
  - -Hopkins measurement: 30 cm from exit panel
  - -Distance from source to exit panel unspecified
  - -Our original study: 30 cm from source
  - - NIST clarified: source is 45 cm from exit panel, chamber 75 from source
- Chamber material
  - - issues with modeling chamber as air or water

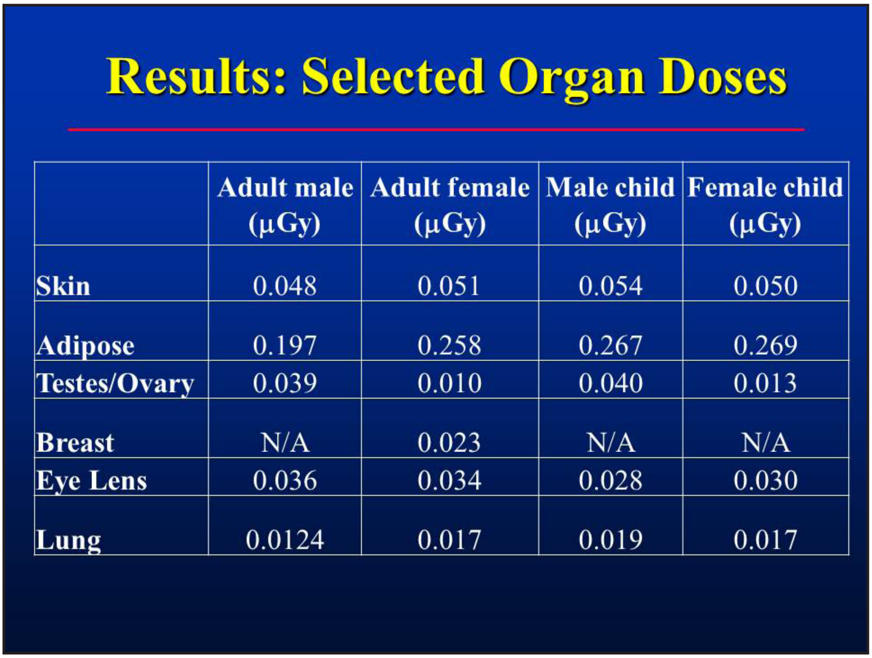
Net Effect: Dose 20-40% less than original study



After Publication: Letter from NIST

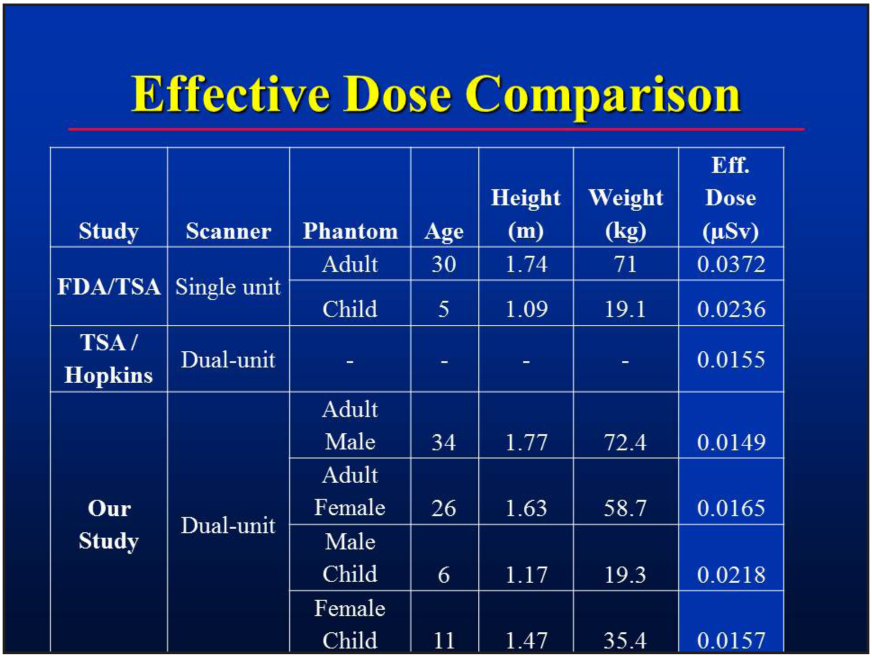
- Glover and Hudson pointed out two errors in our assumptions for estimating photon fluence [Med Phys 39 (9) 3012]
- Correspondence disclosed distance between x-ray source and panel
- Net result: Our published study overestimated dose by factor of 1.25-1.65
- Correction issued [Med Phys 39 (9) 2012]

One goal of study was to generate such discussions



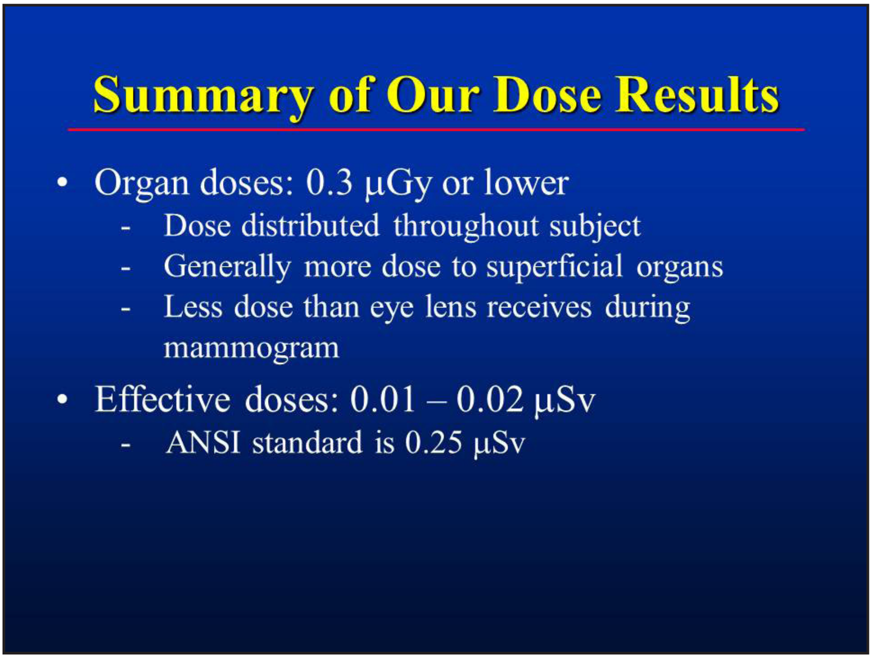
Results: Selected Organ Doses

	Adult male (μGy)	Adult female (μGy)	Male child (μGy)	Female child (μGy)
Skin	0.048	0.051	0.054	0.050
Adipose	0.197	0.258	0.267	0.269
Testes/Ovary	0.039	0.010	0.040	0.013
Breast	N/A	0.023	N/A	N/A
Eye Lens	0.036	0.034	0.028	0.030
Lung	0.0124	0.017	0.019	0.017



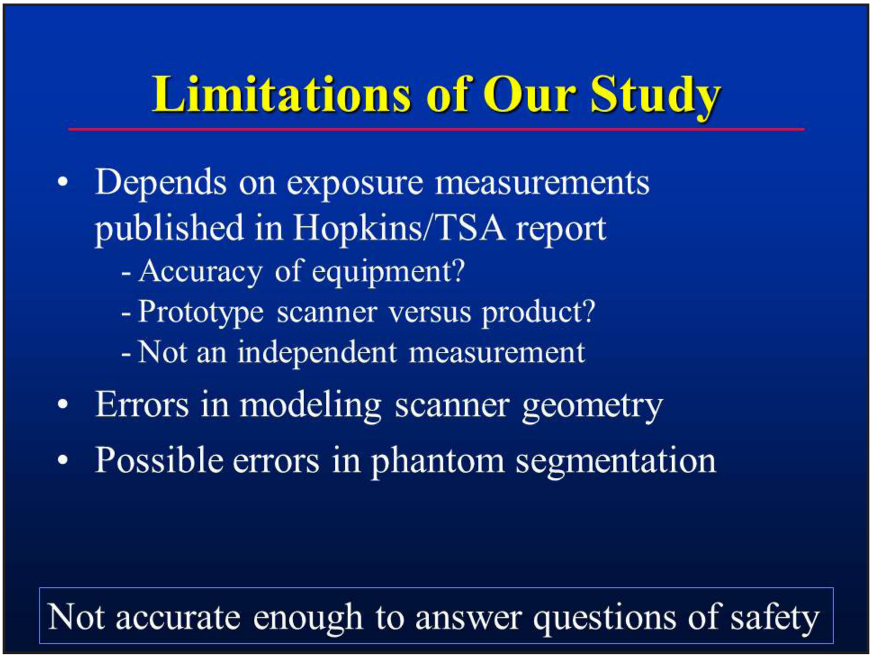
Effective Dose Comparison

Study	Scanner	Phantom	Age	Height (m)	Weight (kg)	Eff. Dose (µSv)
FDA/TSA	Single unit	Adult	30	1.74	71	0.0372
		Child	5	1.09	19.1	0.0236
TSA / Hopkins	Dual-unit	-	-	-	-	0.0155
Our Study	Dual-unit	Adult Male	34	1.77	72.4	0.0149
		Adult Female	26	1.63	58.7	0.0165
		Male Child	6	1.17	19.3	0.0218
		Female Child	11	1.47	35.4	0.0157



Summary of Our Dose Results

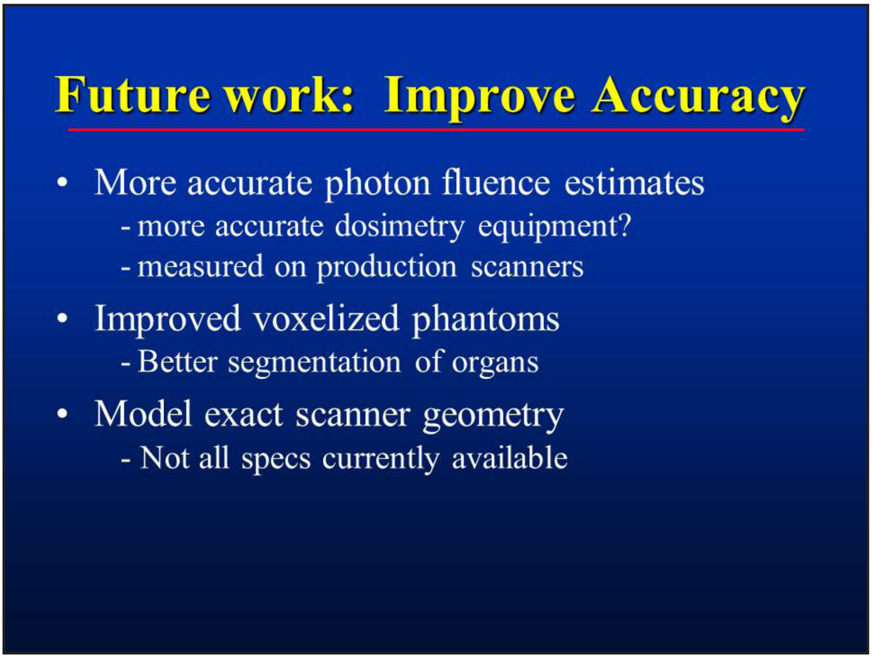
- Organ doses: 0.3 µGy or lower
  - Dose distributed throughout subject
  - Generally more dose to superficial organs
  - Less dose than eye lens receives during mammogram
- Effective doses: 0.01 – 0.02 µSv
  - ANSI standard is 0.25 µSv



Limitations of Our Study

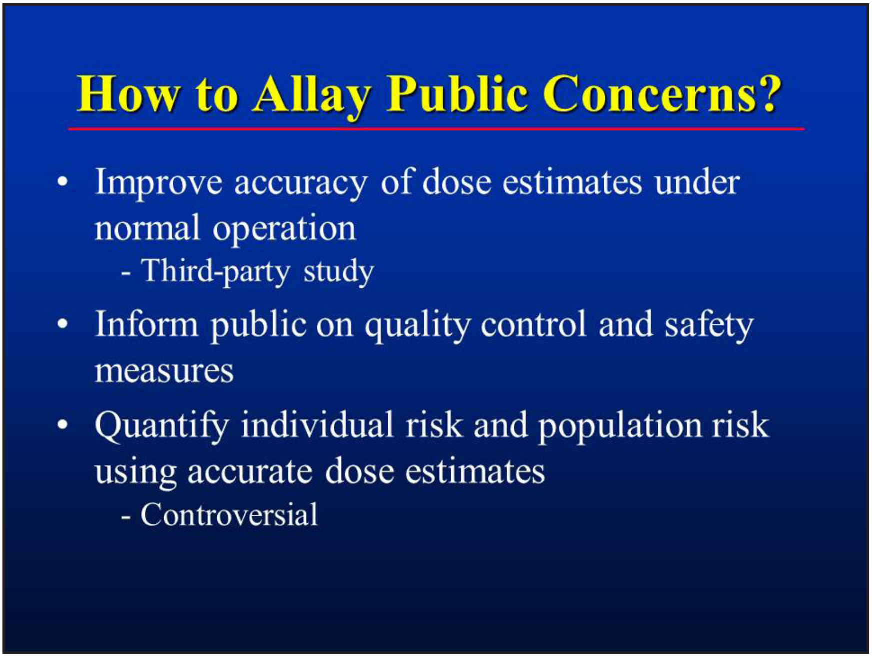
- Depends on exposure measurements published in Hopkins/TSA report
  - Accuracy of equipment?
  - Prototype scanner versus product?
  - Not an independent measurement
- Errors in modeling scanner geometry
- Possible errors in phantom segmentation

Not accurate enough to answer questions of safety



Future work: Improve Accuracy

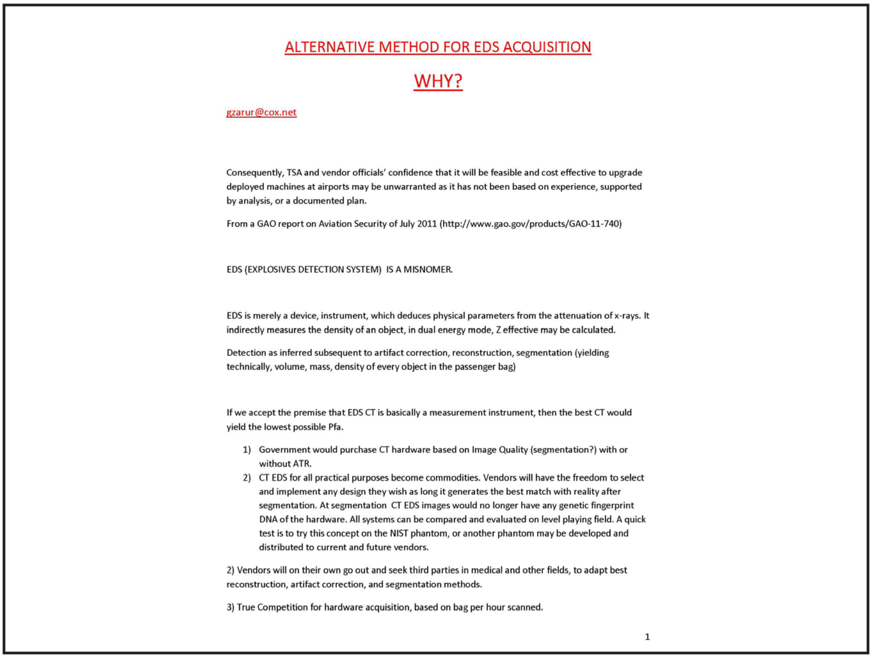
- More accurate photon fluence estimates
  - more accurate dosimetry equipment?
  - measured on production scanners
- Improved voxelized phantoms
  - Better segmentation of organs
- Model exact scanner geometry
  - Not all specs currently available



How to Allay Public Concerns?

- Improve accuracy of dose estimates under normal operation
  - Third-party study
- Inform public on quality control and safety measures
- Quantify individual risk and population risk using accurate dose estimates
  - Controversial

16.13 George Zarur: Alternative Way for TSA to Acquire Technology



ALTERNATIVE METHOD FOR EDS ACQUISITION

WHY?

gzarur@cox.net

Consequently, TSA and vendor officials' confidence that it will be feasible and cost effective to upgrade deployed machines at airports may be unwarranted as it has not been based on experience, supported by analysis, or a documented plan.

From a GAO report on Aviation Security of July 2011 (http://www.gao.gov/products/GAO-11-740)

EDS (EXPLOSIVES DETECTION SYSTEM) IS A MISNOMER.

EDS is merely a device, instrument, which deduces physical parameters from the attenuation of x-rays. It indirectly measures the density of an object, in dual energy mode, Z effective may be calculated.

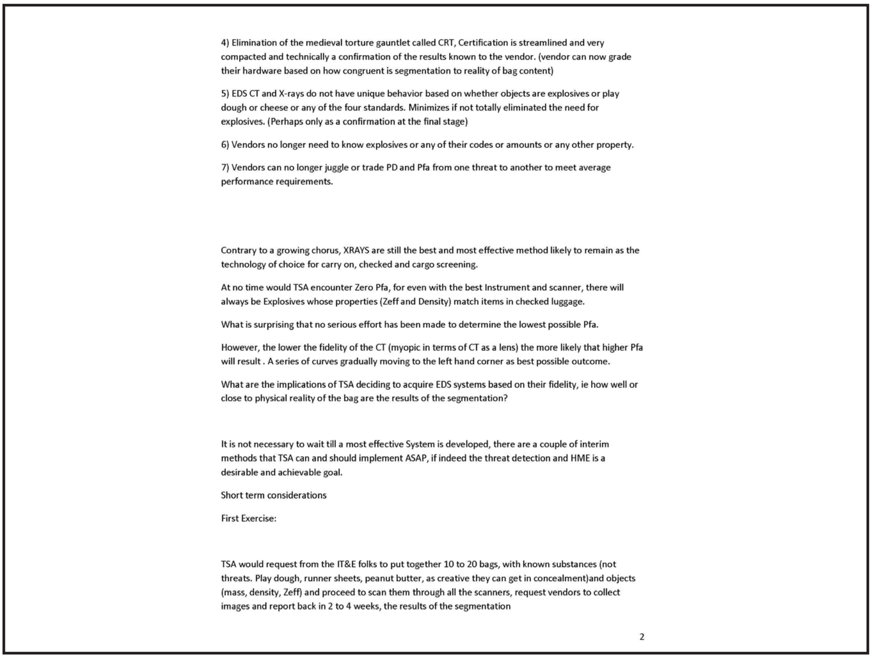
Detection as inferred subsequent to artifact correction, reconstruction, segmentation (yielding technically, volume, mass, density of every object in the passenger bag)

If we accept the premise that EDS CT is basically a measurement instrument, then the best CT would yield the lowest possible Pfa.

1) Government would purchase CT hardware based on Image Quality (segmentation?) with or without ATR.
2) CT EDS for all practical purposes become commodities. Vendors will have the freedom to select and implement any design they wish as long it generates the best match with reality after segmentation. At segmentation CT EDS images would no longer have any genetic fingerprint DNA of the hardware. All systems can be compared and evaluated on level playing field. A quick test is to try this concept on the NIST phantom, or another phantom may be developed and distributed to current and future vendors.

2) Vendors will on their own go out and seek third parties in medical and other fields, to adapt best reconstruction, artifact correction, and segmentation methods.

3) True Competition for hardware acquisition, based on bag per hour scanned.



4) Elimination of the medieval torture gauntlet called CRT, Certification is streamlined and very compacted and technically a confirmation of the results known to the vendor. (vendor can now grade their hardware based on how congruent is segmentation to reality of bag content)

5) EDS CT and X-rays do not have unique behavior based on whether objects are explosives or play dough or cheese or any of the four standards. Minimizes if not totally eliminated the need for explosives. (Perhaps only as a confirmation at the final stage)

6) Vendors no longer need to know explosives or any of their codes or amounts or any other property.

7) Vendors can no longer juggle or trade PD and Pfa from one threat to another to meet average performance requirements.

Contrary to a growing chorus, XRAYS are still the best and most effective method likely to remain as the technology of choice for carry on, checked and cargo screening.

At no time would TSA encounter Zero Pfa, for even with the best Instrument and scanner, there will always be Explosives whose properties (Zeff and Density) match items in checked luggage.

What is surprising that no serious effort has been made to determine the lowest possible Pfa.

However, the lower the fidelity of the CT (myopic in terms of CT as a lens) the more likely that higher Pfa will result . A series of curves gradually moving to the left hand corner as best possible outcome.

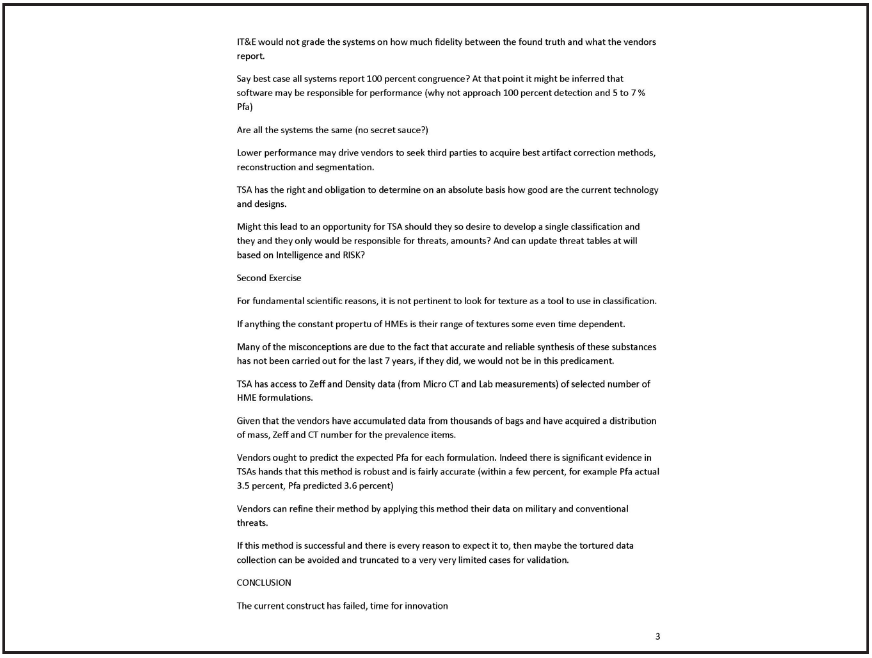
What are the implications of TSA deciding to acquire EDS systems based on their fidelity, ie how well or close to physical reality of the bag are the results of the segmentation?

It is not necessary to wait till a most effective System is developed, there are a couple of interim methods that TSA can and should implement ASAP, if indeed the threat detection and HME is a desirable and achievable goal.

Short term considerations

First Exercise:

TSA would request from the IT&E folks to put together 10 to 20 bags, with known substances (not threats. Play dough, runner sheets, peanut butter, as creative they can get in concealment)and objects (mass, density, Zeff) and proceed to scan them through all the scanners, request vendors to collect images and report back in 2 to 4 weeks, the results of the segmentation



IT&E would not grade the systems on how much fidelity between the found truth and what the vendors report.

Say best case all systems report 100 percent congruence? At that point it might be inferred that software may be responsible for performance (why not approach 100 percent detection and 5 to 7 % Pfa)

Are all the systems the same (no secret sauce?)

Lower performance may drive vendors to seek third parties to acquire best artifact correction methods, reconstruction and segmentation.

TSA has the right and obligation to determine on an absolute basis how good are the current technology and designs.

Might this lead to an opportunity for TSA should they so desire to develop a single classification and they and they only would be responsible for threats, amounts? And can update threat tables at will based on Intelligence and RISK?

Second Exercise

For fundamental scientific reasons, it is not pertinent to look for texture as a tool to use in classification.

If anything the constant propertu of HMEs is their range of textures some even time dependent.

Many of the misconceptions are due to the fact that accurate and reliable synthesis of these substances has not been carried out for the last 7 years, if they did, we would not be in this predicament.

TSA has access to Zeff and Density data (from Micro CT and Lab measurements) of selected number of HME formulations.

Given that the vendors have accumulated data from thousands of bags and have acquired a distribution of mass, Zeff and CT number for the prevalence items.

Vendors ought to predict the expected Pfa for each formulation. Indeed there is significant evidence in TSAs hands that this method is robust and is fairly accurate (within a few percent, for example Pfa actual 3.5 percent, Pfa predicted 3.6 percent)

Vendors can refine their method by applying this method their data on military and conventional threats.

If this method is successful and there is every reason to expect it to, then maybe the tortured data collection can be avoided and truncated to a very very limited cases for validation.

CONCLUSION

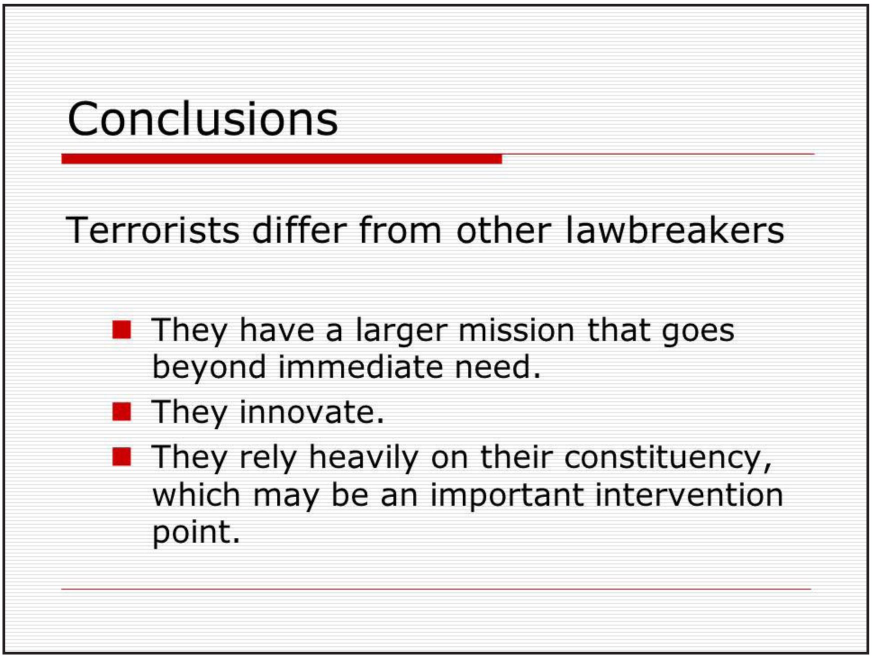
The current construct has failed, time for innovation

16.14 Laura Dugan: Effectiveness of Deterrence

Deterring Terrorism: Lessons Learned in Social Science Research

Laura Dugan
University of Maryland
Department of Criminology and Criminal Justice &
The National Consortium for the Study of Terrorism and Response to Terrorism (START)

Some of this research was supported by the U.S. Department of Homeland Security (DHS) Science and Technology Directorate's Office of University Programs through START. Any opinions, findings, conclusions or recommendations presented here are solely the authors' and are not representative of DHS or the United States Government.



Conclusions

Terrorists differ from other lawbreakers

- They have a larger mission that goes beyond immediate need.
- They innovate.
- They rely heavily on their constituency, which may be an important intervention point.

Deterrence: Rooted in Rational Choice Theory

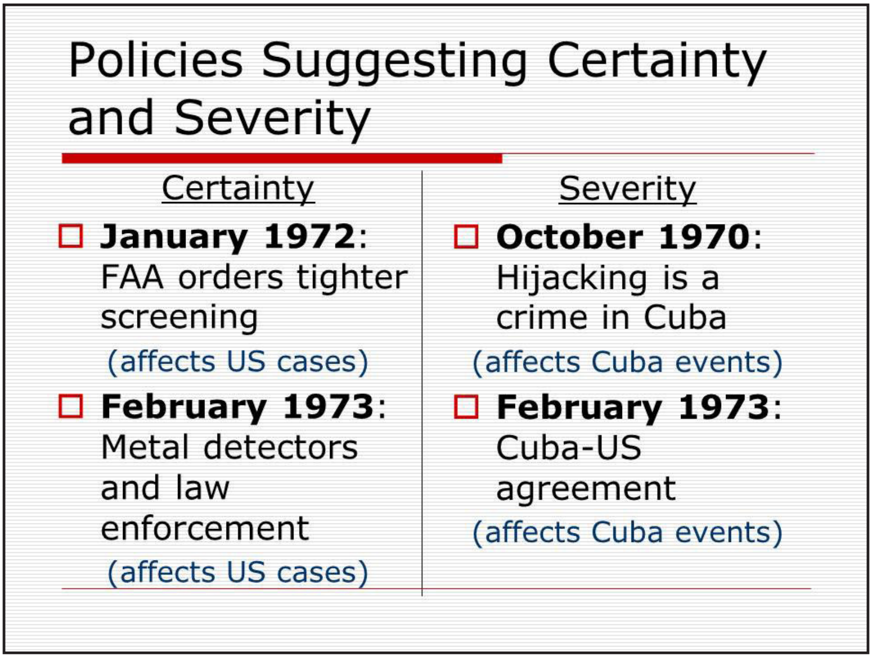
$$E(u_{terror}) = p\, U(y-F) + (1-p)\, U(y)$$

where p = perceived probability of punishment
y = anticipated benefits of perpetration; and
F = perceived penalty of the act

Lesson: Raise the costs of perpetration through increased certainty and severity so that the utility of perpetration falls below the benefit of the act. In other words: DETERRENCE

Testing a Rational Choice Model of Airline Hijacking

Dugan, Laura, Gary LaFree, and Alex Piquero, *Criminology* 2005



Policies Suggesting Certainty and Severity

Certainty	Severity
☐ **January 1972**: FAA orders tighter screening (affects US cases)	☐ **October 1970**: Hijacking is a crime in Cuba (affects Cuba events)
☐ **February 1973**: Metal detectors and law enforcement (affects US cases)	☐ **February 1973**: Cuba-US agreement (affects Cuba events)

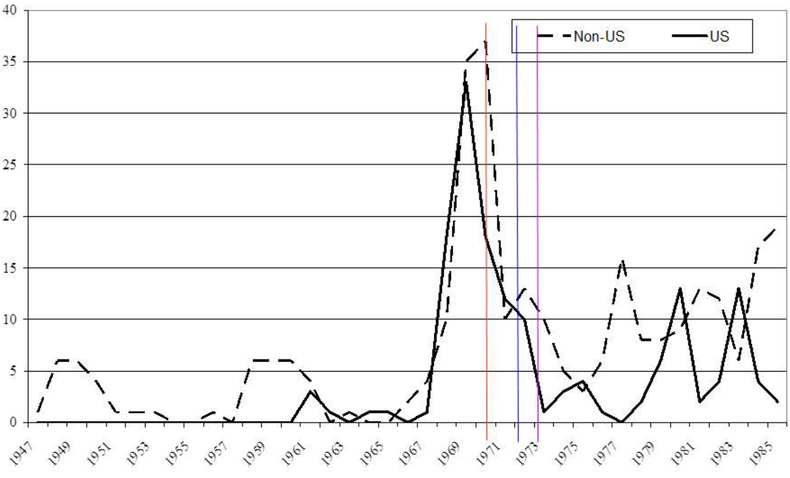
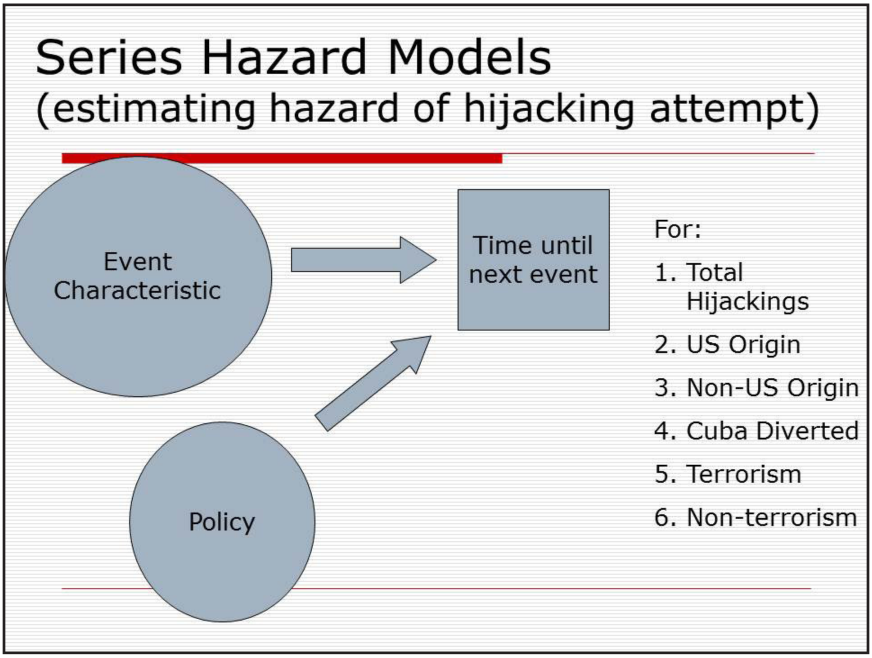
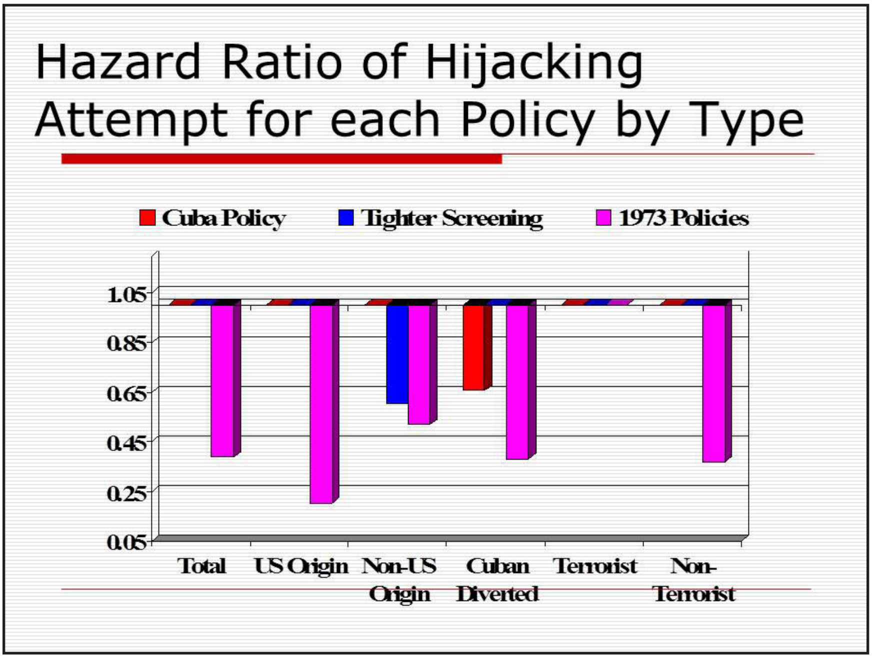


Figure 2. US and Non-US Successful Hijackings, 1946 to 1985



Series Hazard Models
(estimating hazard of hijacking attempt)

For:
1. Total Hijackings
2. US Origin
3. Non-US Origin
4. Cuba Diverted
5. Terrorism
6. Non-terrorism



Hazard Ratio of Hijacking Attempt for each Policy by Type

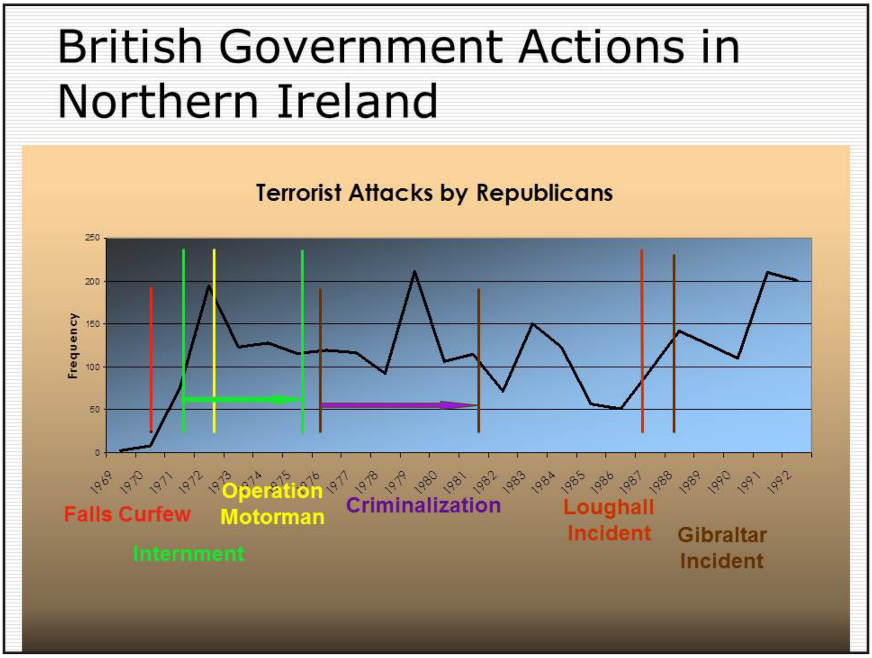
Conclusions on Deterrence

- A new hijacking attempt is less likely when the certainty of apprehension is increased.
- Compared to those who hijack for other reasons, hijacking attempts by terrorists will be less affected by counter terrorism measures that raise the severity or certainty of punishment.

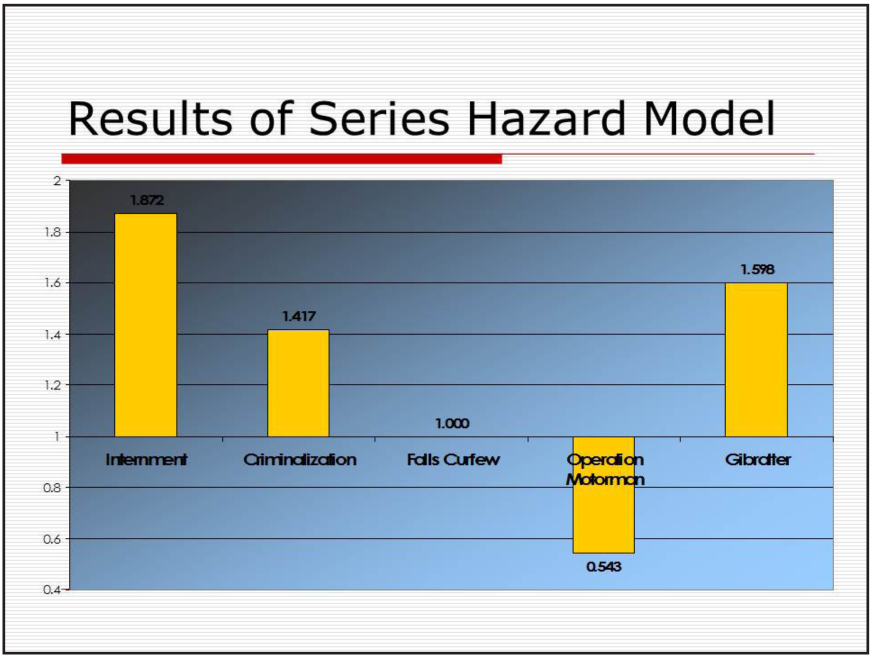
The Impact of British Counterterrorism Strategies on Political Violence in Northern Ireland:

Comparing Deterrence and Backlash Models

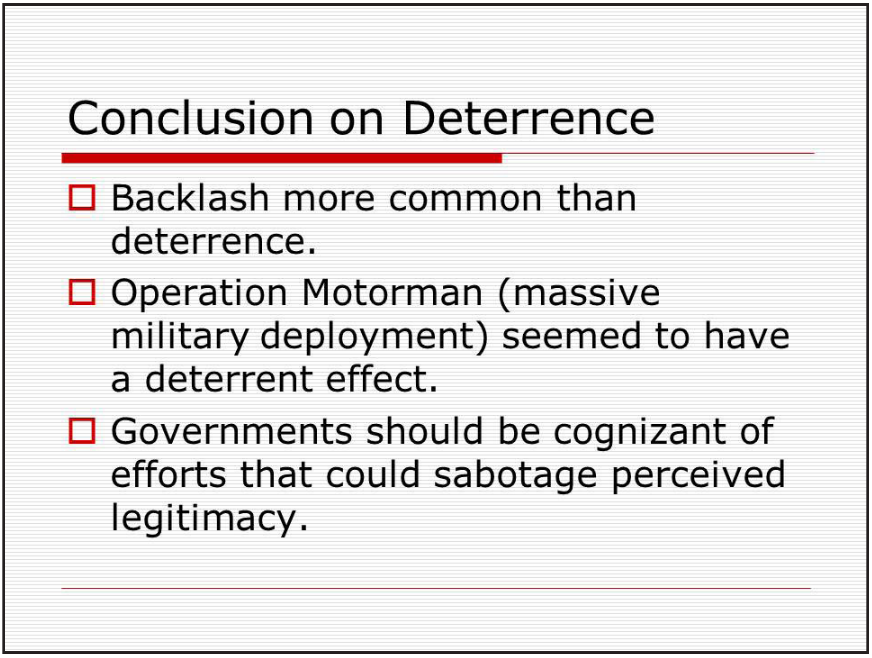
LaFree, Gary, Laura Dugan and Raven Korte, *Criminology* 2009



British Government Actions in Northern Ireland

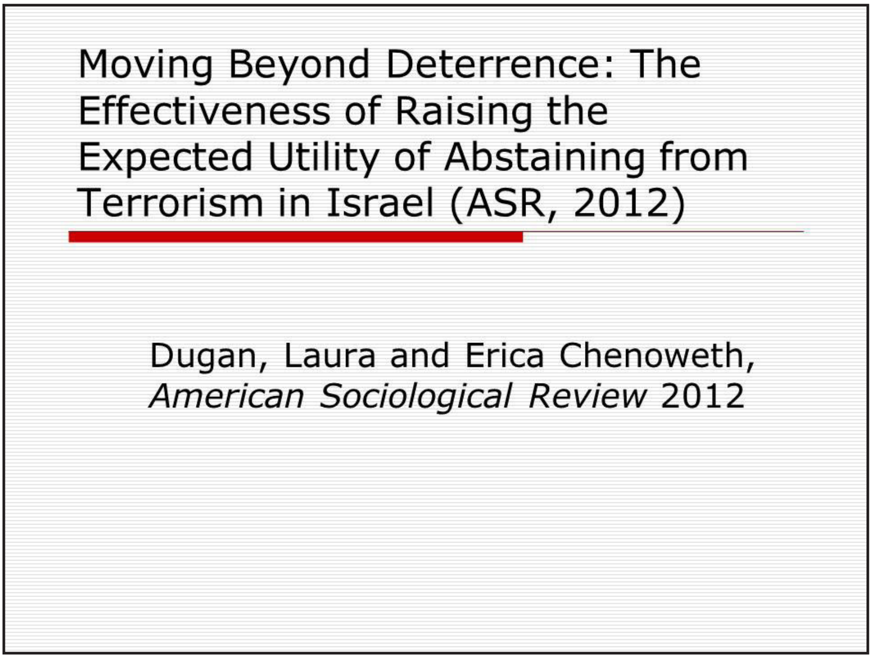


Results of Series Hazard Model



Conclusion on Deterrence

- Backlash more common than deterrence.
- Operation Motorman (massive military deployment) seemed to have a deterrent effect.
- Governments should be cognizant of efforts that could sabotage perceived legitimacy.



Moving Beyond Deterrence: The Effectiveness of Raising the Expected Utility of Abstaining from Terrorism in Israel (ASR, 2012)

Dugan, Laura and Erica Chenoweth,
American Sociological Review 2012

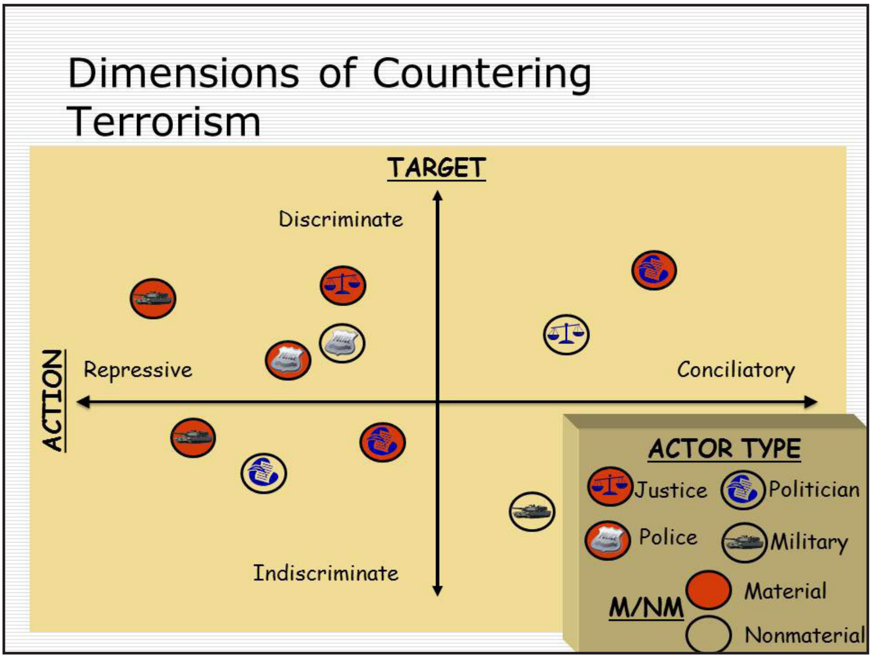
An Underutilized Component of Rational Choice: Raising the Benefits of Abstaining from Terrorism

$$E(u_{\mathrm{nonterror}}) = q\, U(x+G) + (1\text{-}q)\, U(x)$$

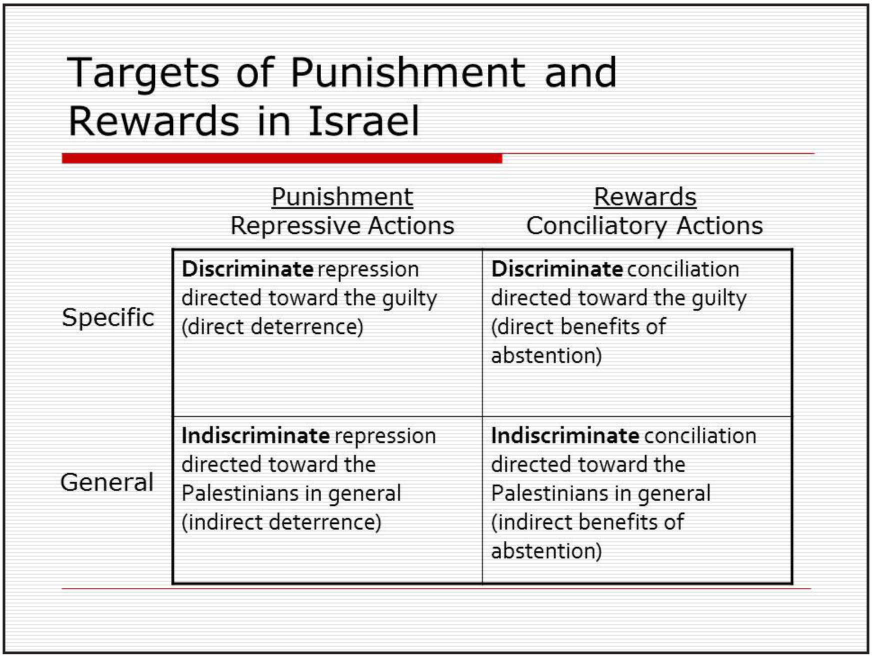
where q = perceived probability of rewards from abstention

x = value of current situation; and

G = anticipated rewards of abstention



Dimensions of Countering Terrorism



Targets of Punishment and Rewards in Israel

	Punishment Repressive Actions	Rewards Conciliatory Actions
Specific	**Discriminate** repression directed toward the guilty (direct deterrence)	**Discriminate** conciliation directed toward the guilty (direct benefits of abstention)
General	**Indiscriminate** repression directed toward the Palestinians in general (indirect deterrence)	**Indiscriminate** conciliation directed toward the Palestinians in general (indirect benefits of abstention)

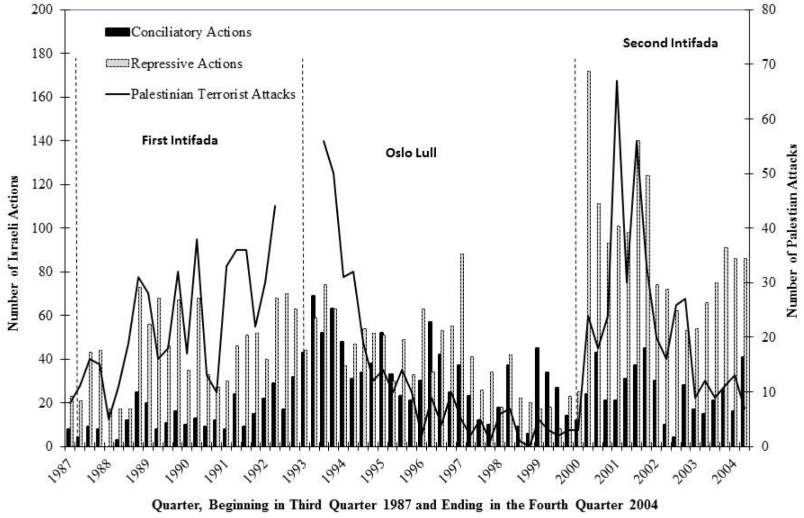
Tactical Regimes of the Israeli-Palestinian Conflict

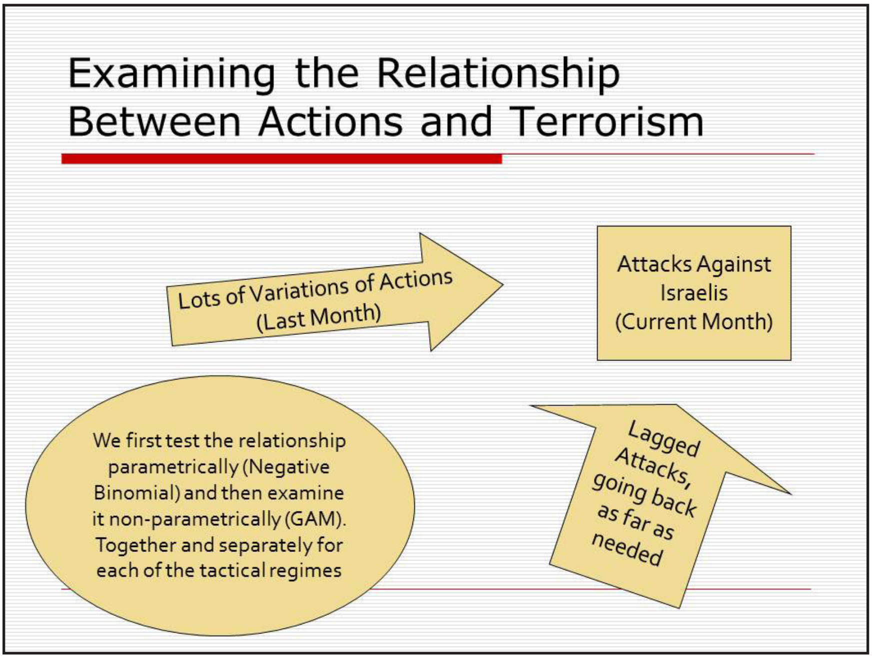
The First Intifada (1987-1993)
- Started as nonviolent
- Dominated by secular nationalists
- Hamas became active near the end

The Oslo Lull (1993-2000)
- Negotiators established Palestinian Authority
- Palestinians recognized 1967 borders
- Neither side held to agreement

The Second Intifada (2000-2004)
- Violent from the beginning
- Dominated by religious groups
- Known for deadly suicide attacks

Figure 1. Quarterly Repressive and Conciliatory Actions by Israel and Palestinian Terrorist Attacks





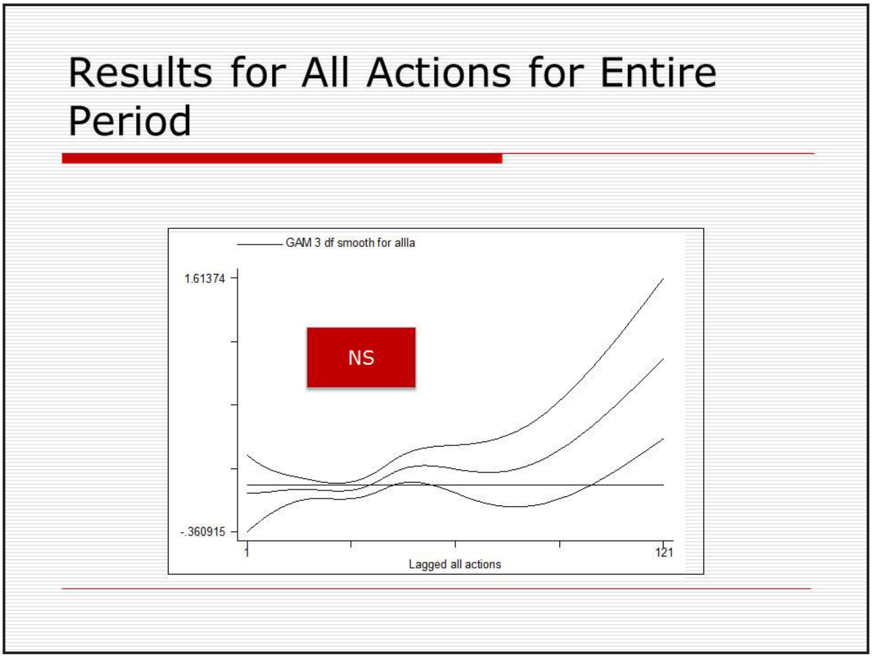
Examining the Relationship Between Actions and Terrorism

Lots of Variations of Actions (Last Month)

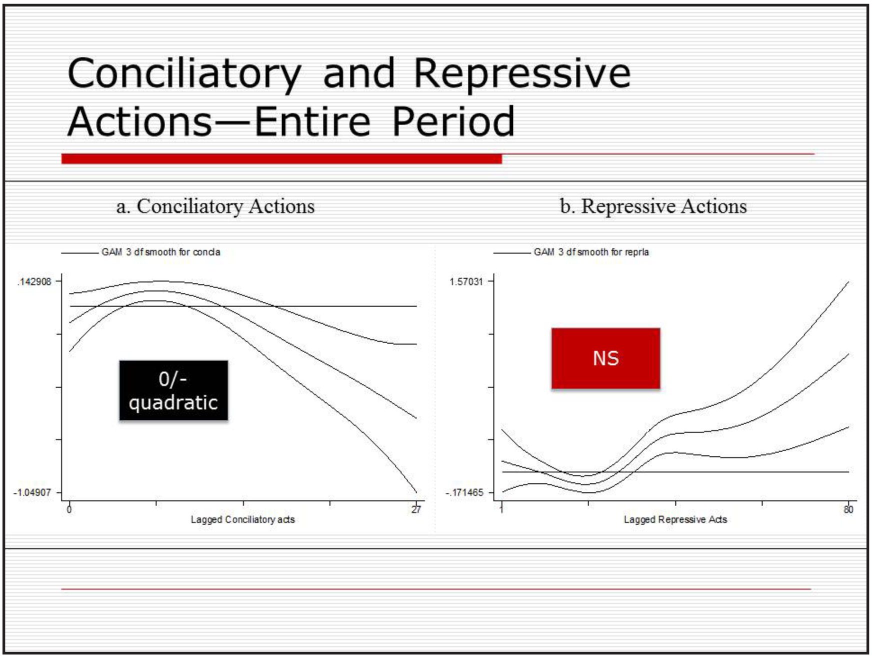
Attacks Against Israelis (Current Month)

We first test the relationship parametrically (Negative Binomial) and then examine it non-parametrically (GAM). Together and separately for each of the tactical regimes

Lagged Attacks, going back as far as needed



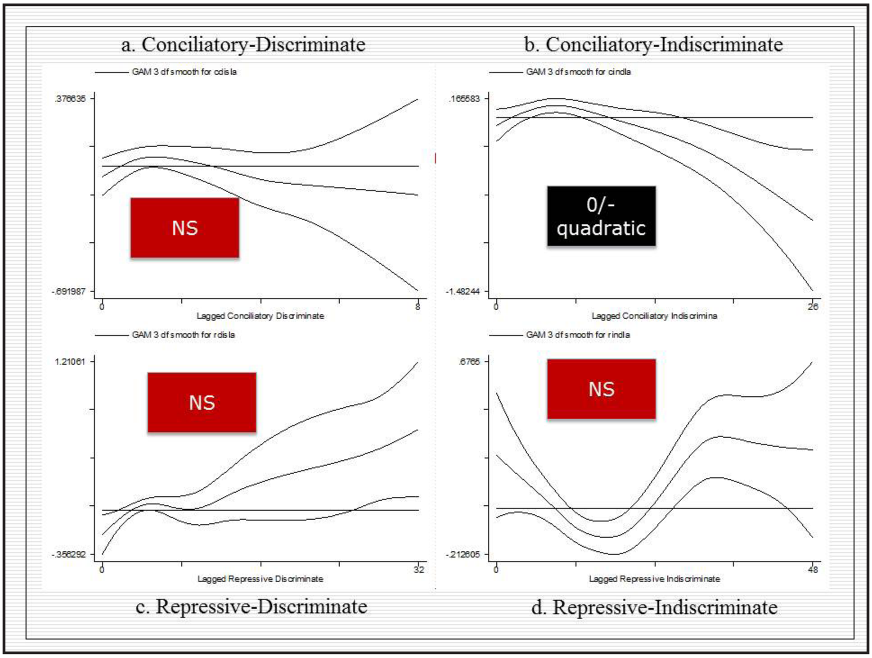
Results for All Actions for Entire Period



Conciliatory and Repressive Actions—Entire Period

a. Conciliatory Actions

b. Repressive Actions

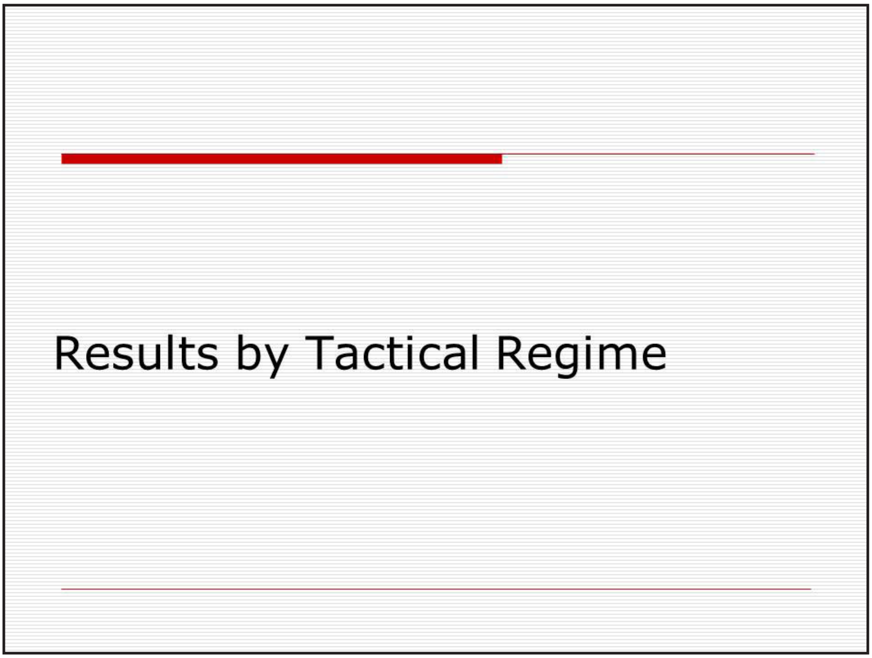


a. Conciliatory-Discriminate

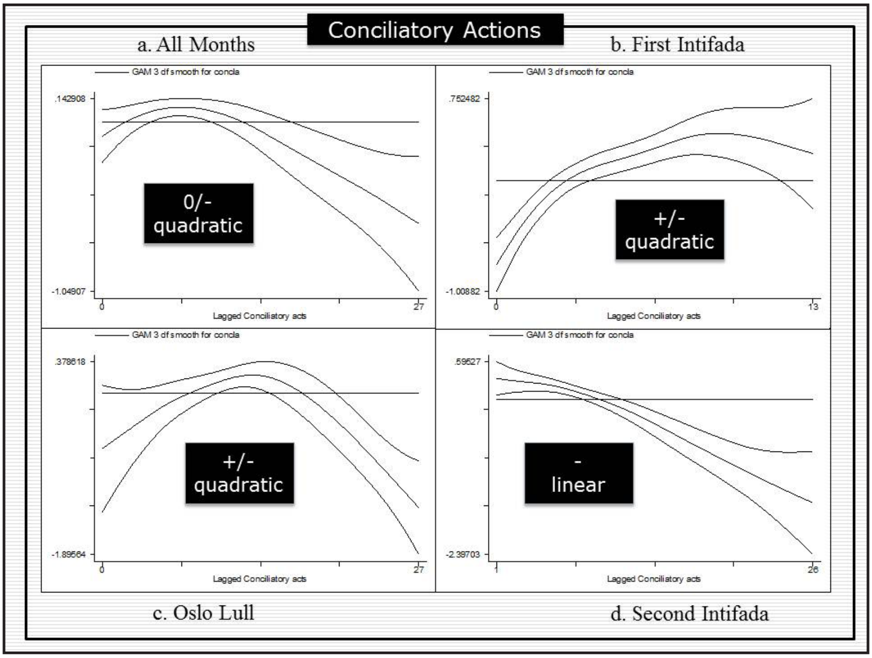
b. Conciliatory-Indiscriminate

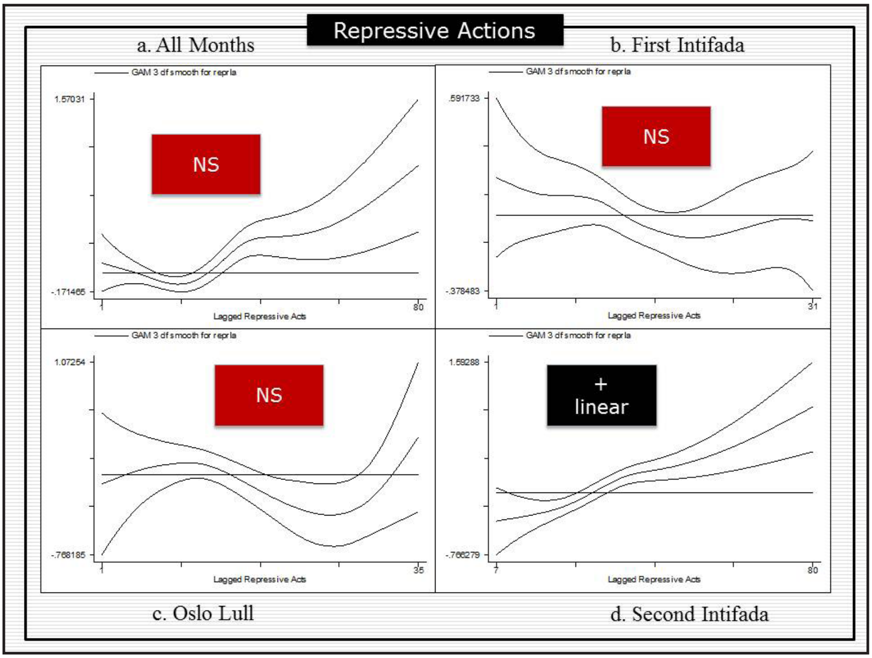
c. Repressive-Discriminate

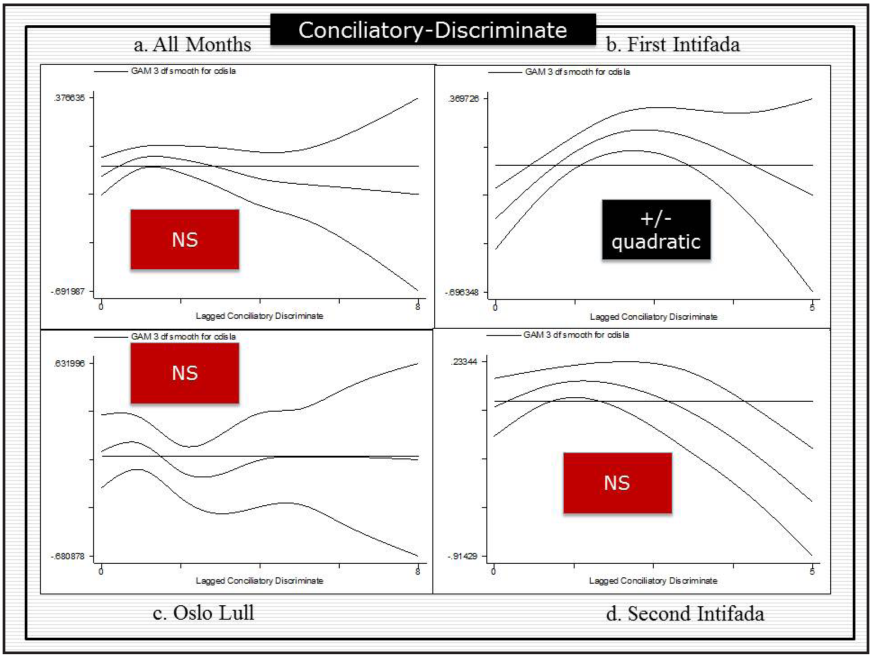
d. Repressive-Indiscriminate

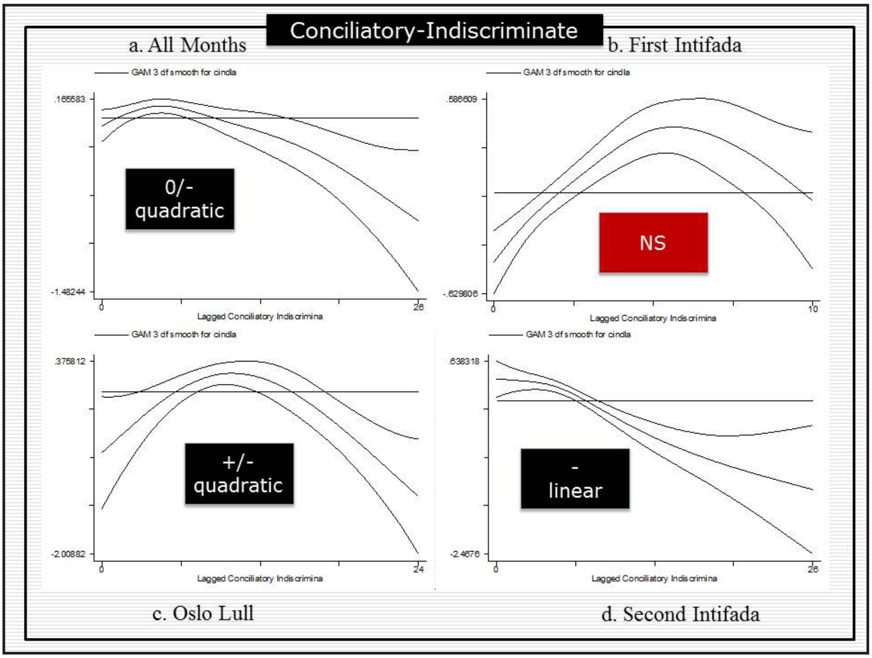


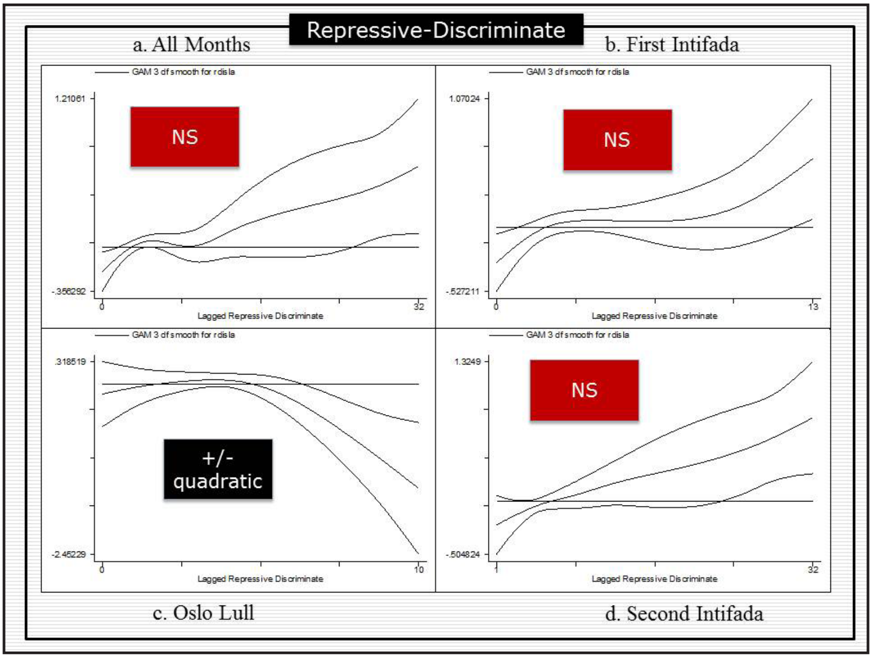
Results by Tactical Regime

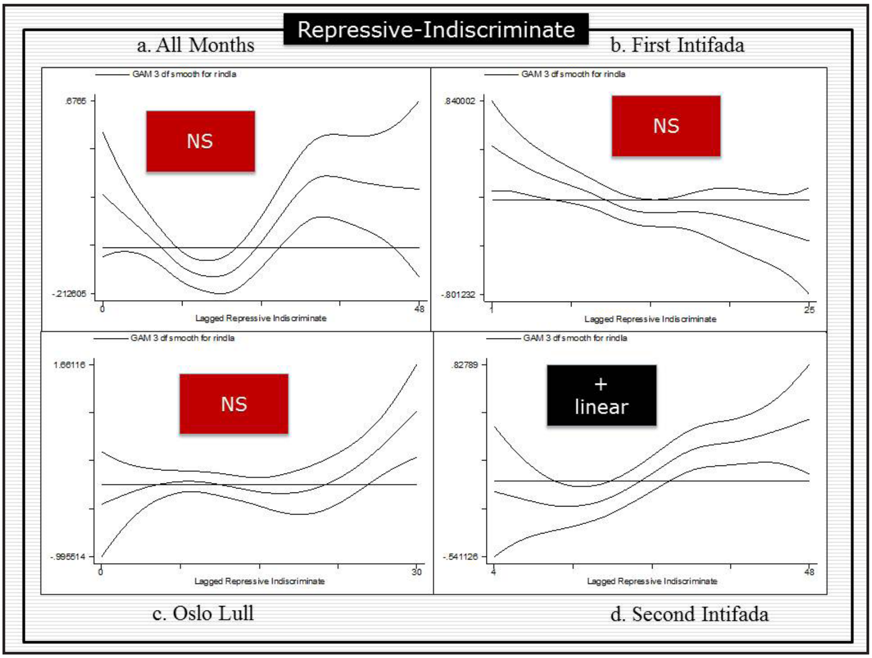

Conciliatory Actions
a. All Months
b. First Intifada
GAM 3 df smooth for concla
.142908
-1.04907
0
27
Lagged Conciliatory acts
0/-
quadratic
GAM 3 df smooth for concla
.752482
-1.00882
0
13
Lagged Conciliatory acts
+/-
quadratic
GAM 3 df smooth for concla
.378618
-1.89564
0
27
Lagged Conciliatory acts
+/-
quadratic
GAM 3 df smooth for concla
.59527
-2.39703
1
26
Lagged Conciliatory acts
-
linear
c. Oslo Lull
d. Second Intifada

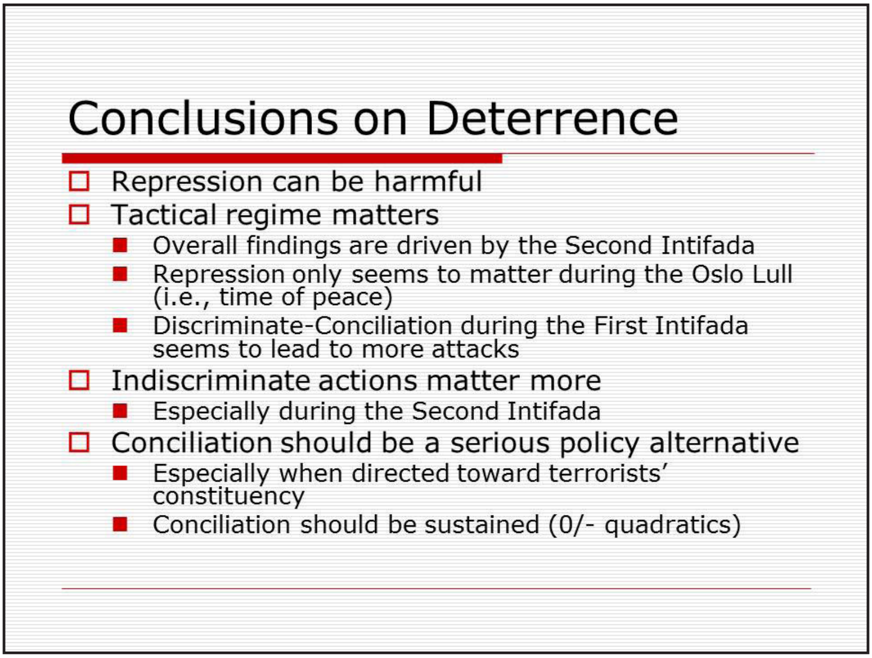

Repressive Actions
a. All Months
b. First Intifada
GAM 3 df smooth for repria
1.57031
-.171465
1
80
Lagged Repressive Acts
NS
GAM 3 df smooth for repria
.591733
-.378483
1
31
Lagged Repressive Acts
NS
GAM 3 df smooth for repria
1.07254
-.768185
1
35
Lagged Repressive Acts
NS
GAM 3 df smooth for repria
1.59288
-.766279
7
80
Lagged Repressive Acts
+
linear
c. Oslo Lull
d. Second Intifada


Conciliatory-Discriminate
a. All Months
GAM 3 df smooth for cdisla
.376635
-.691987
0
8
Lagged Conciliatory Discriminate
NS
b. First Intifada
GAM 3 df smooth for cdisla
.369726
-.696348
0
5
Lagged Conciliatory Discriminate
+/-
quadratic
GAM 3 df smooth for cdisla
.631996
-.680878
0
8
Lagged Conciliatory Discriminate
NS
GAM 3 df smooth for cdisla
.23344
-.91429
0
5
Lagged Conciliatory Discriminate
NS
c. Oslo Lull
d. Second Intifada


Conciliatory-Indiscriminate
a. All Months
GAM 3 df smooth for cindla
.165583
-1.48244
0
26
Lagged Conciliatory Indiscrimina
0/-
quadratic
b. First Intifada
GAM 3 df smooth for cindla
.586609
-.629806
0
10
Lagged Conciliatory Indiscrimina
NS
GAM 3 df smooth for cindla
.375812
-2.00882
0
24
Lagged Conciliatory Indiscrimina
+/-
quadratic
GAM 3 df smooth for cindla
.638318
-2.4676
0
26
Lagged Conciliatory Indiscrimina
-
linear
c. Oslo Lull
d. Second Intifada

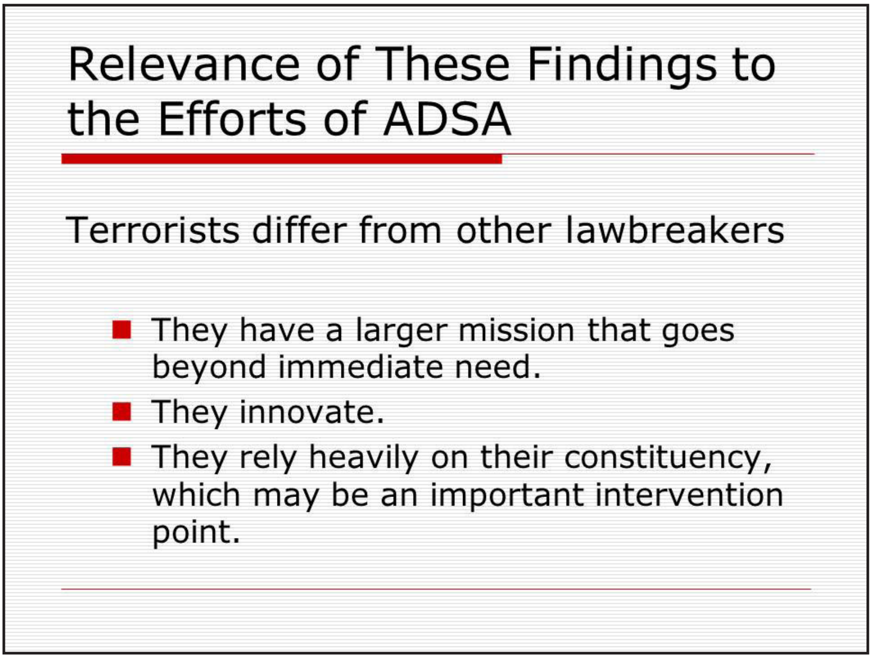

Repressive-Discriminate
a. All Months
b. First Intifada
c. Oslo Lull
d. Second Intifada
GAM 3 df smooth for rdisla
1.21061
-.356292
0
32
Lagged Repressive Discriminate
NS
1.07024
-.527211
0
13
NS
.318519
-2.45229
0
10
+/-
quadratic
1.3249
-.504824
1
32
NS


Repressive-Indiscriminate
a. All Months
b. First Intifada
c. Oslo Lull
d. Second Intifada
GAM 3 df smooth for rindla
.6765
-.212605
0
48
Lagged Repressive Indiscriminate
NS
.840002
-.801232
1
25
NS
1.66116
-.995514
0
30
NS
.82789
-.541126
4
48
+
linear



Conclusions on Deterrence

- Repression can be harmful
- Tactical regime matters
  - Overall findings are driven by the Second Intifada
  - Repression only seems to matter during the Oslo Lull (i.e., time of peace)
  - Discriminate-Conciliation during the First Intifada seems to lead to more attacks
- Indiscriminate actions matter more
  - Especially during the Second Intifada
- Conciliation should be a serious policy alternative
  - Especially when directed toward terrorists' constituency
  - Conciliation should be sustained (0/- quadratics)

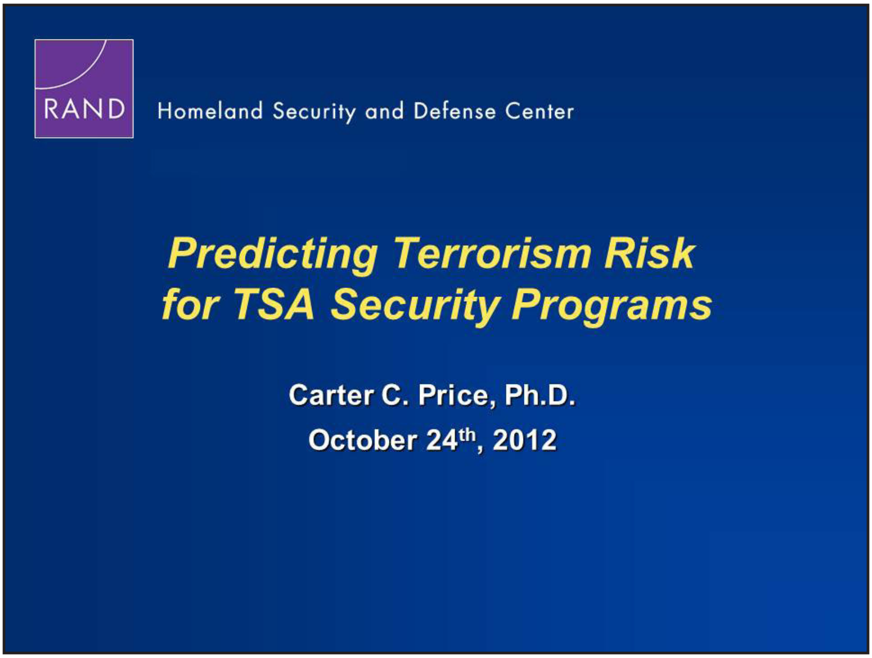


Relevance of These Findings to the Efforts of ADSA

Terrorists differ from other lawbreakers

- They have a larger mission that goes beyond immediate need.
- They innovate.
- They rely heavily on their constituency, which may be an important intervention point.

16.15 Carter Price: Predictive Terrorism Risk for TSA Security Programs



RAND Homeland Security and Defense Center

Predicting Terrorism Risk for TSA Security Programs

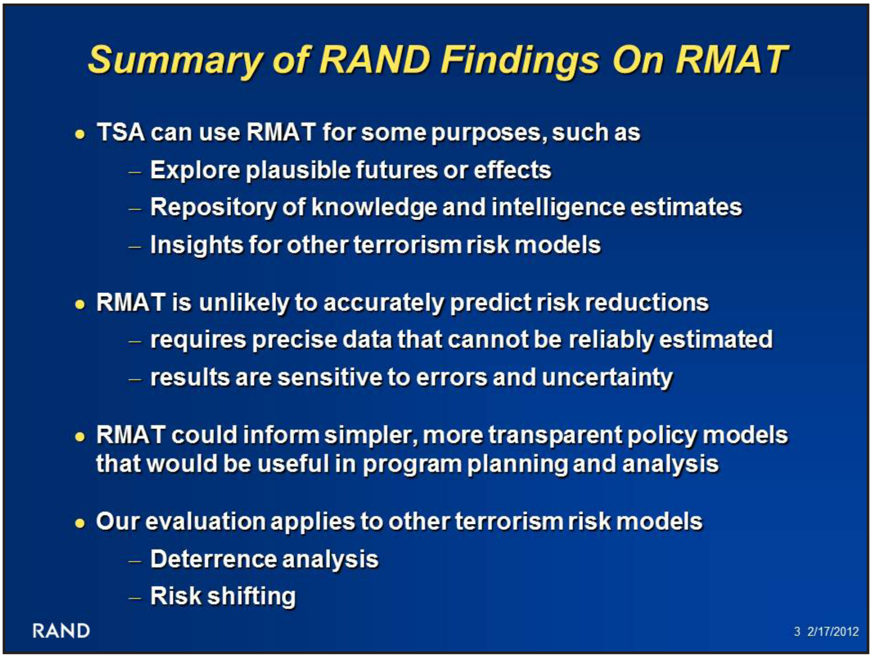
Carter C. Price, Ph.D.
October 24th, 2012

RMAT Is One of Several TSA Risk Tools

- Risk Management Analysis Tool (RMAT) simulates terrorist attacks on aviation system
    - Developed by The Boeing Company in conjunction with TSA and other industry and agency stakeholders
- RMAT designed to estimate risk reductions attributable to new programs, accounting for
    - Terrorist intentions, targeting preferences, and tactics
    - Effectiveness of existing layers of security
    - Likely damage from 60 kinds of attack
- TSA asked RAND to independently validate RMAT

RAND

2 2/17/2012

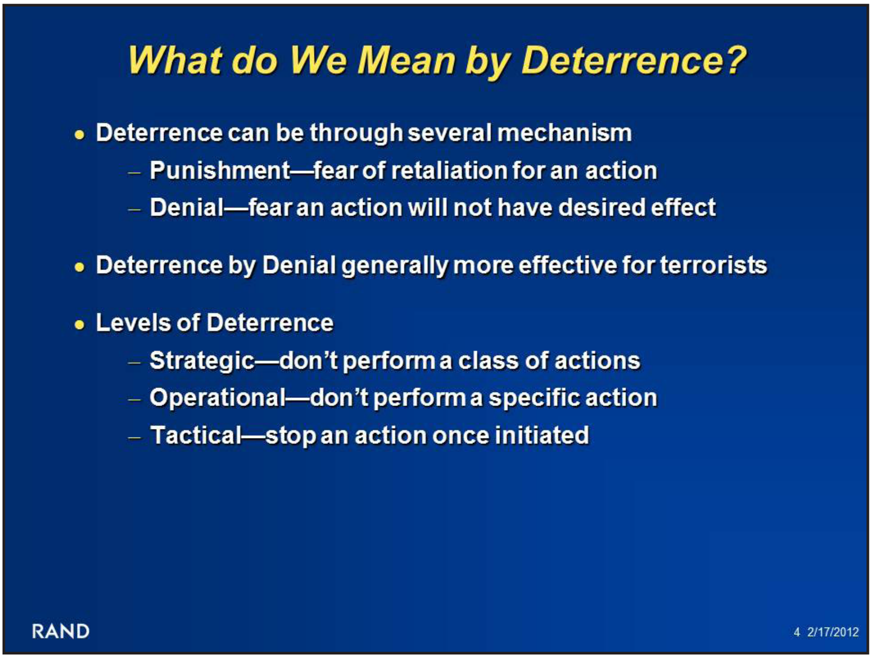


Summary of RAND Findings On RMAT

- TSA can use RMAT for some purposes, such as
  - Explore plausible futures or effects
  - Repository of knowledge and intelligence estimates
  - Insights for other terrorism risk models
- RMAT is unlikely to accurately predict risk reductions
  - requires precise data that cannot be reliably estimated
  - results are sensitive to errors and uncertainty
- RMAT could inform simpler, more transparent policy models that would be useful in program planning and analysis
- Our evaluation applies to other terrorism risk models
  - Deterrence analysis
  - Risk shifting

RAND

3 2/17/2012

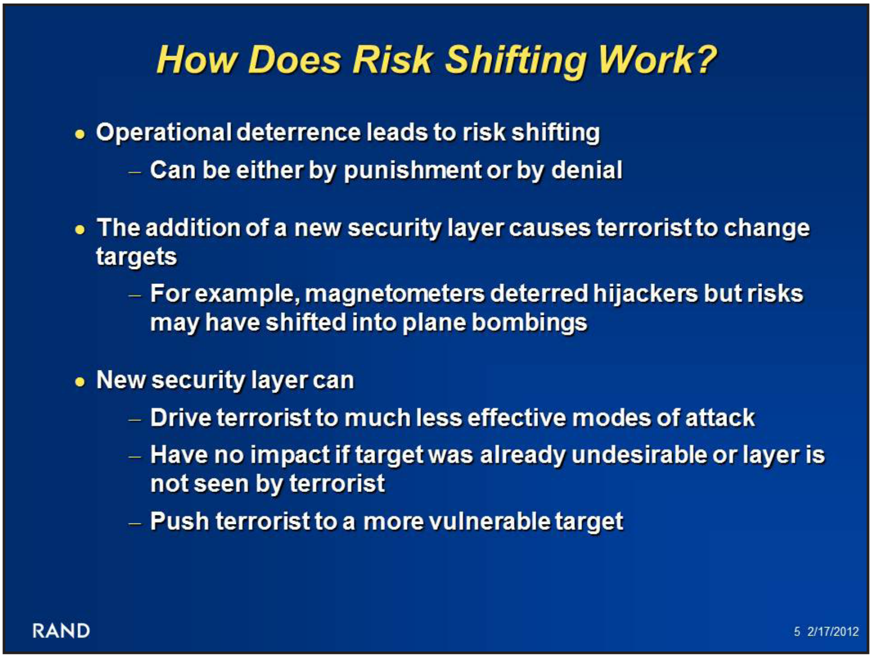


What do We Mean by Deterrence?

- Deterrence can be through several mechanism
  - Punishment—fear of retaliation for an action
  - Denial—fear an action will not have desired effect
- Deterrence by Denial generally more effective for terrorists
- Levels of Deterrence
  - Strategic—don't perform a class of actions
  - Operational—don't perform a specific action
  - Tactical—stop an action once initiated

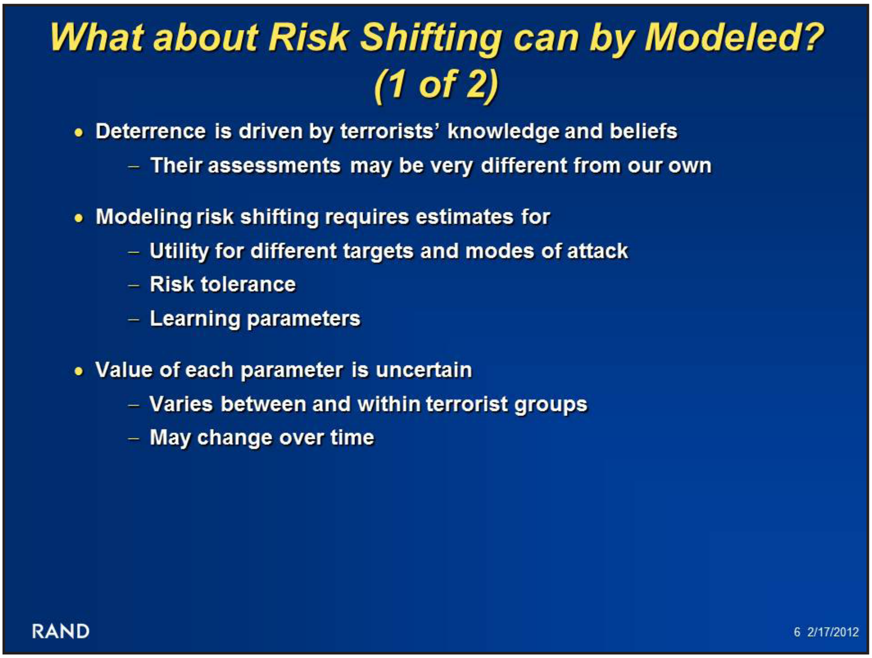
RAND

4 2/17/2012



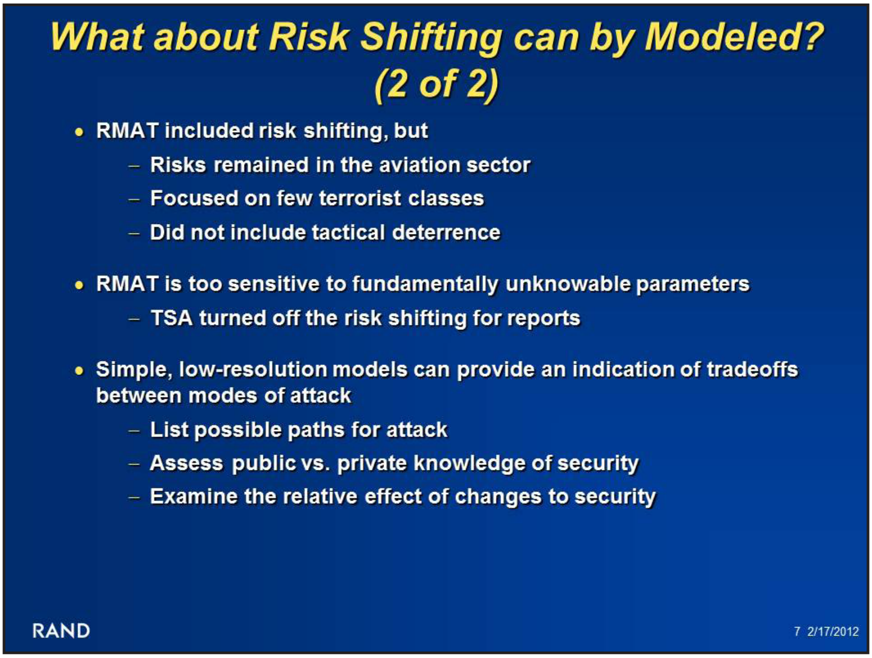
How Does Risk Shifting Work?

- Operational deterrence leads to risk shifting
  - Can be either by punishment or by denial
- The addition of a new security layer causes terrorist to change targets
  - For example, magnetometers deterred hijackers but risks may have shifted into plane bombings
- New security layer can
  - Drive terrorist to much less effective modes of attack
  - Have no impact if target was already undesirable or layer is not seen by terrorist
  - Push terrorist to a more vulnerable target



What about Risk Shifting can by Modeled? (1 of 2)

- Deterrence is driven by terrorists' knowledge and beliefs
  - Their assessments may be very different from our own
- Modeling risk shifting requires estimates for
  - Utility for different targets and modes of attack
  - Risk tolerance
  - Learning parameters
- Value of each parameter is uncertain
  - Varies between and within terrorist groups
  - May change over time



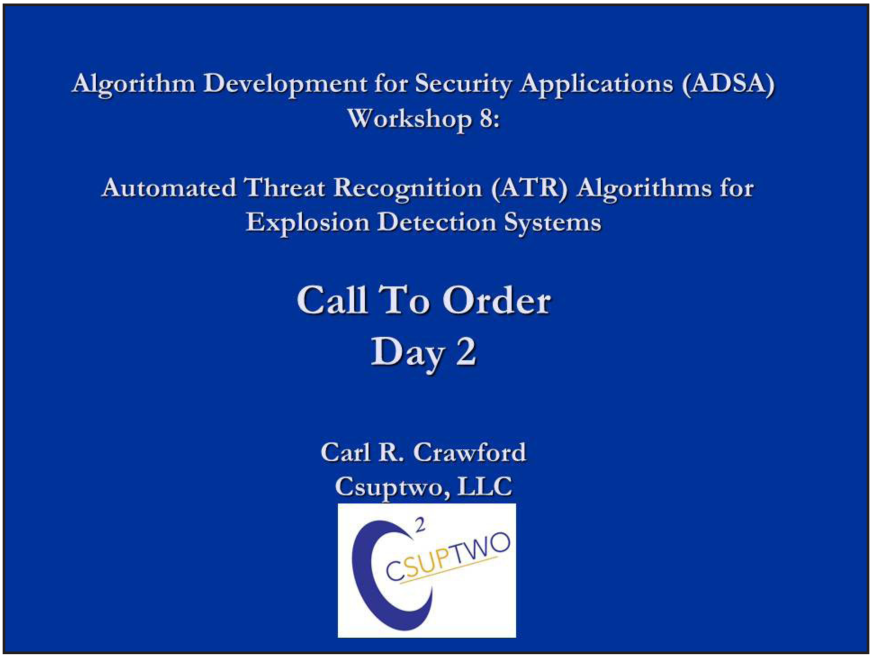
What about Risk Shifting can by Modeled? (2 of 2)

- RMAT included risk shifting, but
  - Risks remained in the aviation sector
  - Focused on few terrorist classes
  - Did not include tactical deterrence
- RMAT is too sensitive to fundamentally unknowable parameters
  - TSA turned off the risk shifting for reports
- Simple, low-resolution models can provide an indication of tradeoffs between modes of attack
  - List possible paths for attack
  - Assess public vs. private knowledge of security
  - Examine the relative effect of changes to security

RAND Homeland Security and Defense Center

A MULTI-UNIT RESEARCH CENTER

16.16 Carl Crawford: Day 2 Objectives

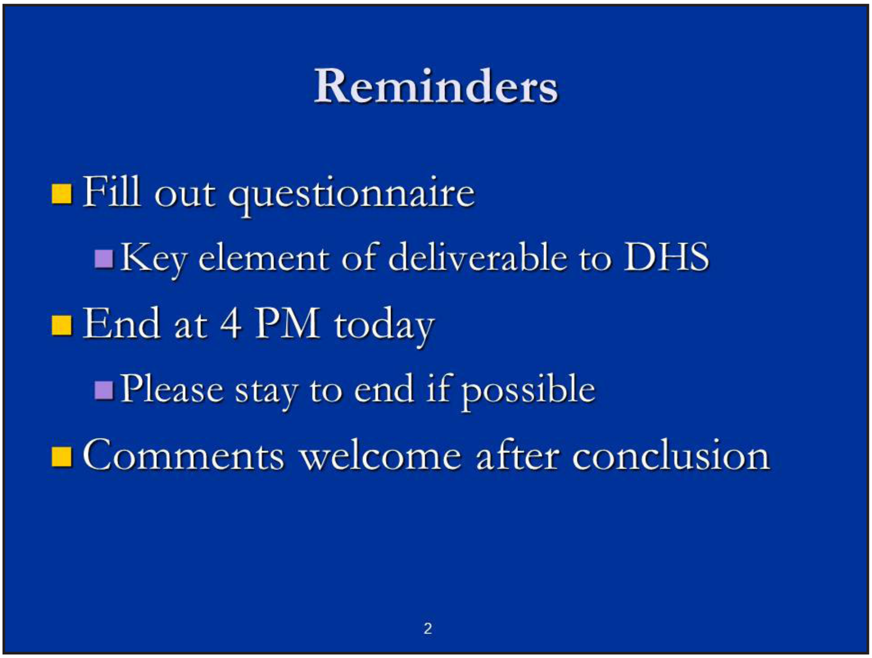


Algorithm Development for Security Applications (ADSA)
Workshop 8:

Automated Threat Recognition (ATR) Algorithms for
Explosion Detection Systems

Call To Order
Day 2

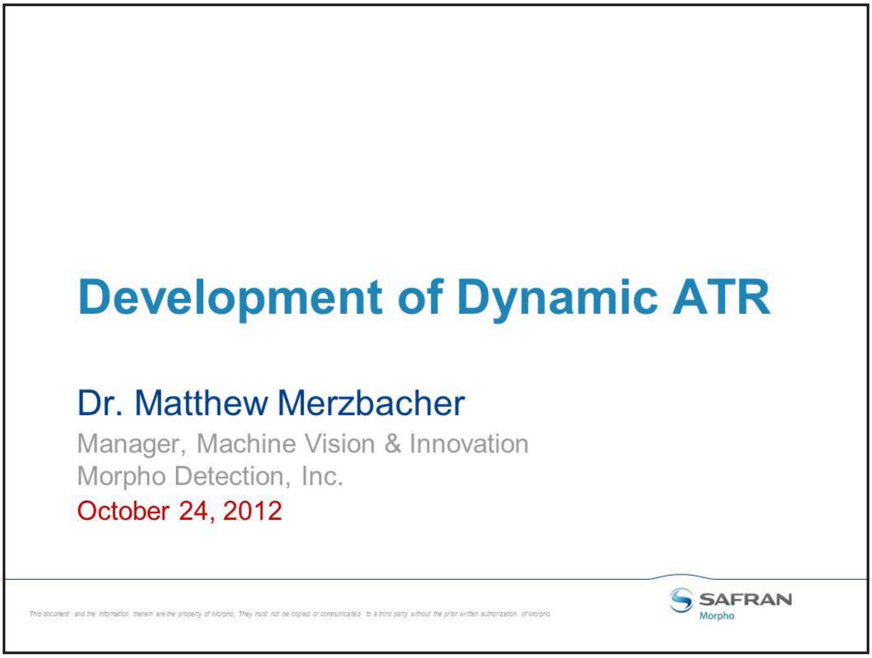
Carl R. Crawford
Csuptwo, LLC



Reminders

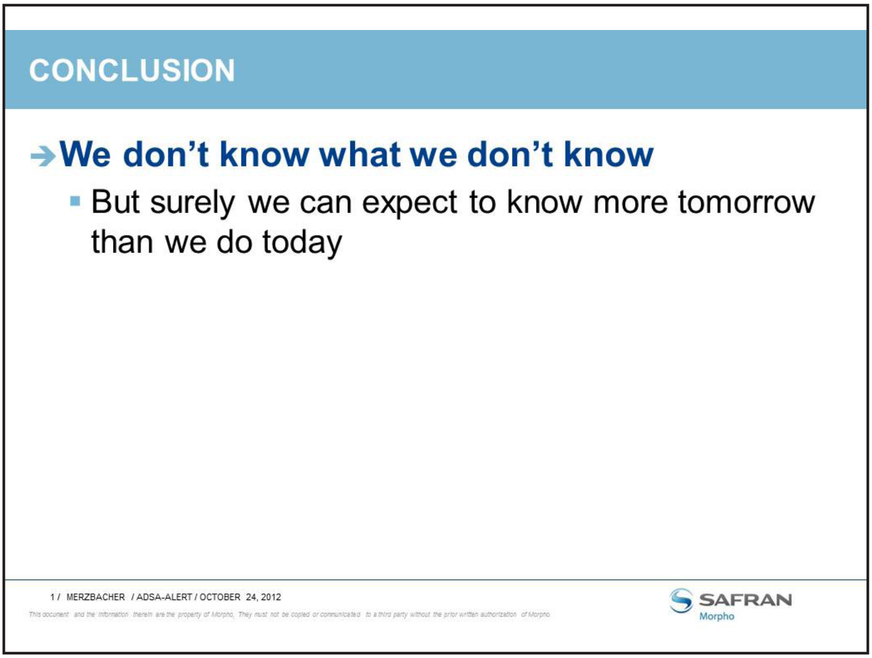
- Fill out questionnaire
  - Key element of deliverable to DHS
- End at 4 PM today
  - Please stay to end if possible
- Comments welcome after conclusion

16.17 Matthew Merzbacher: Dynamic ATR



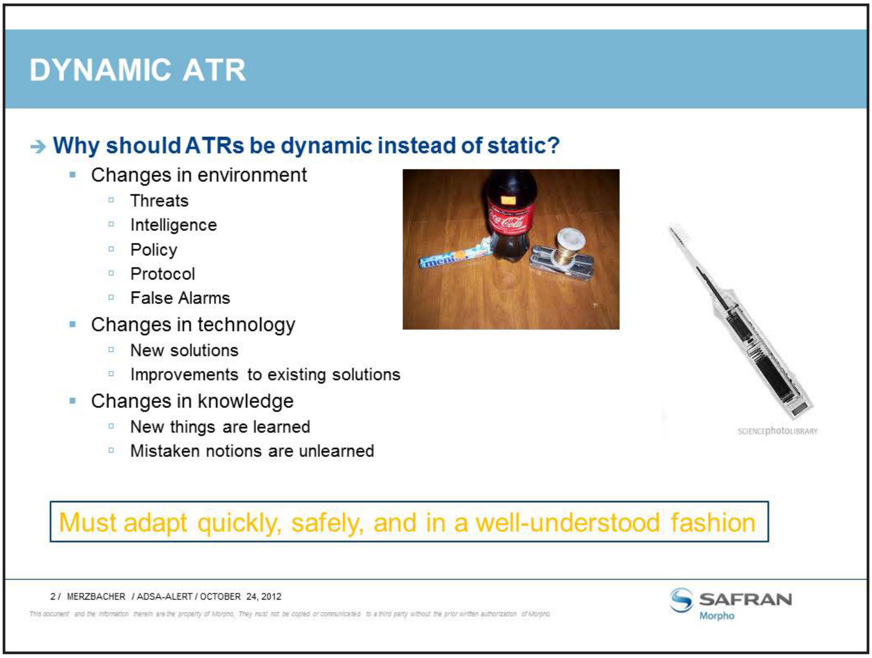
Development of Dynamic ATR

Dr. Matthew Merzbacher
Manager, Machine Vision & Innovation
Morpho Detection, Inc.
October 24, 2012



CONCLUSION

→ **We don't know what we don't know**
  - But surely we can expect to know more tomorrow than we do today



DYNAMIC ATR

→ **Why should ATRs be dynamic instead of static?**

- Changes in environment
    - Threats
    - Intelligence
    - Policy
    - Protocol
    - False Alarms
- Changes in technology
    - New solutions
    - Improvements to existing solutions
- Changes in knowledge
    - New things are learned
    - Mistaken notions are unlearned

SCIENCEphotoLIBRARY

Must adapt quickly, safely, and in a well-understood fashion

SAFRAN
Morpho

LEARNING TO CRAWL

North Carolina physics professor in Argentine jail on drug charges

By Joshua Rhett Miller / Published March 20, 2012 / FoxNews.com

Paul Frampton, UNC Physics Professor, Asks For Double His Salary From Argentine Prison

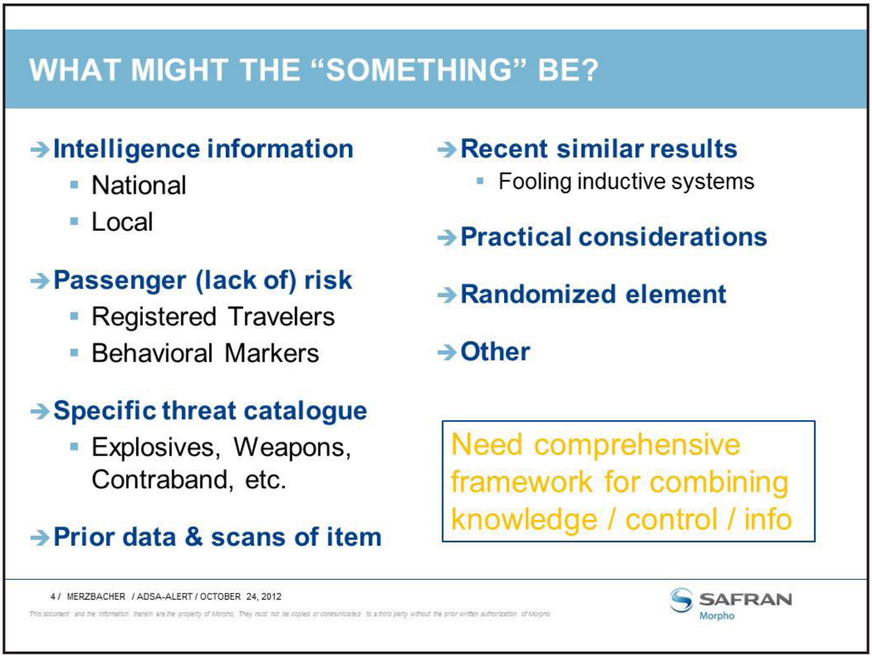
Posted: 10/24/2012 1:14 pm EDT Updated: 10/24/2012 1:15 pm EDT

Distinguished British scientist faces 16 years in Argentine jail after being caught with suitcase of cocaine

A distinguished British scientist is languishing in an Argentine jail on suspicion of drugs smuggling.

'caught with 2kg of cocaine' held

SAFRAN
Morpho

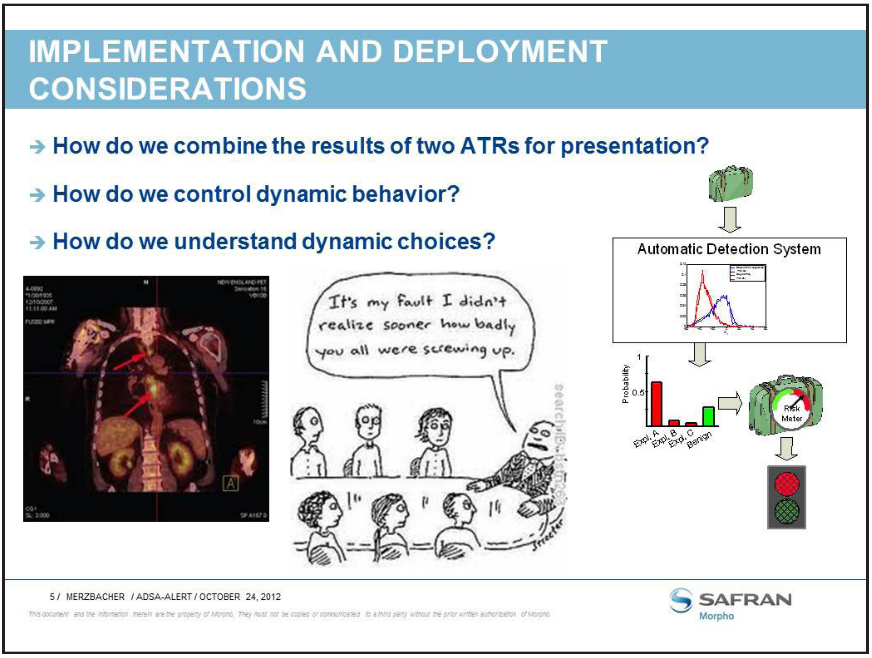


WHAT MIGHT THE “SOMETHING” BE?

- **Intelligence information**
  - National
  - Local
- **Passenger (lack of) risk**
  - Registered Travelers
  - Behavioral Markers
- **Specific threat catalogue**
  - Explosives, Weapons, Contraband, etc.
- **Prior data & scans of item**
- **Recent similar results**
  - Fooling inductive systems
- **Practical considerations**
- **Randomized element**
- **Other**

Need comprehensive framework for combining knowledge / control / info

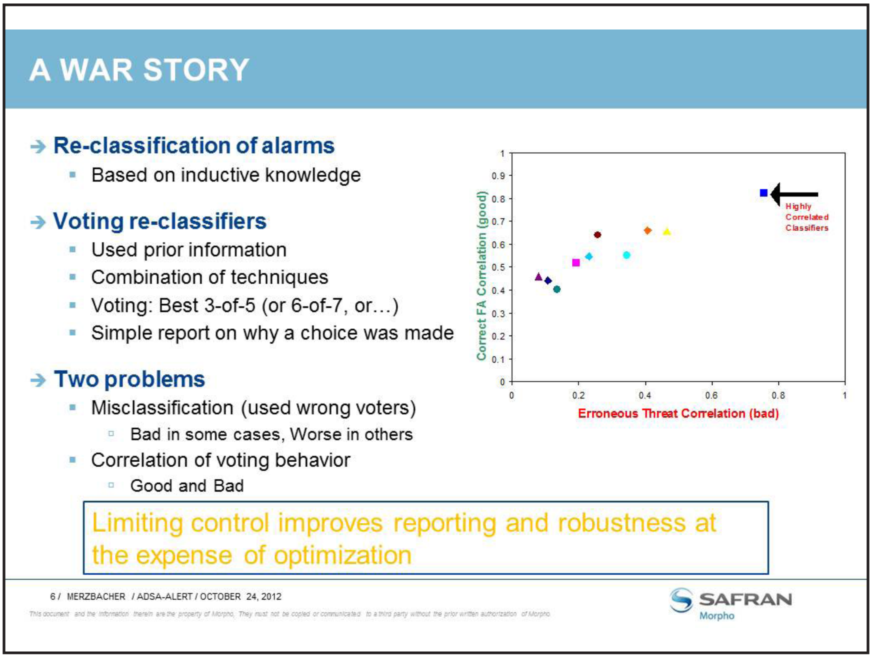
4 / MERZBACHER / ADSA-ALERT / OCTOBER 24, 2012



IMPLEMENTATION AND DEPLOYMENT CONSIDERATIONS

- **How do we combine the results of two ATRs for presentation?**
- **How do we control dynamic behavior?**
- **How do we understand dynamic choices?**

5 / MERZBACHER / ADSA-ALERT / OCTOBER 24, 2012

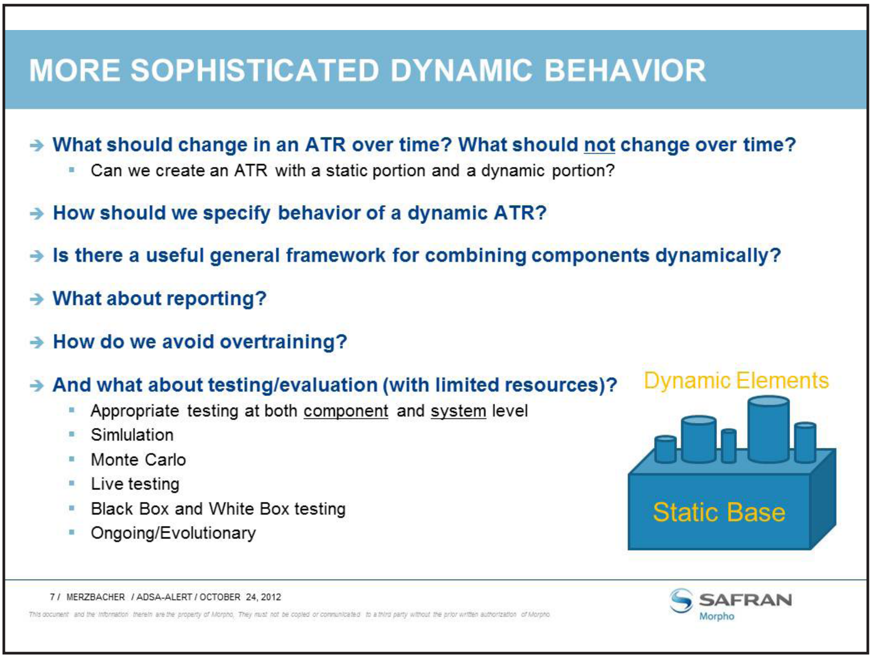


A WAR STORY

- **Re-classification of alarms**
  - Based on inductive knowledge
- **Voting re-classifiers**
  - Used prior information
  - Combination of techniques
  - Voting: Best 3-of-5 (or 6-of-7, or…)
  - Simple report on why a choice was made
- **Two problems**
  - Misclassification (used wrong voters)
    - Bad in some cases, Worse in others
  - Correlation of voting behavior
    - Good and Bad

Limiting control improves reporting and robustness at the expense of optimization

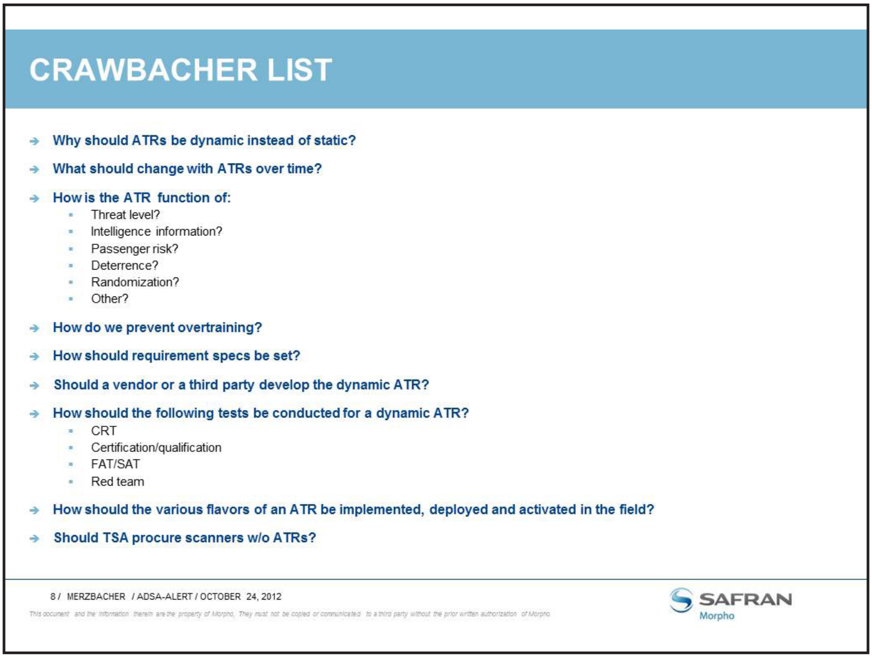
SAFRAN
Morpho



MORE SOPHISTICATED DYNAMIC BEHAVIOR

- **What should change in an ATR over time? What should not change over time?**
  - Can we create an ATR with a static portion and a dynamic portion?
- **How should we specify behavior of a dynamic ATR?**
- **Is there a useful general framework for combining components dynamically?**
- **What about reporting?**
- **How do we avoid overtraining?**
- **And what about testing/evaluation (with limited resources)?**
  - Appropriate testing at both component and system level
  - Simlulation
  - Monte Carlo
  - Live testing
  - Black Box and White Box testing
  - Ongoing/Evolutionary

SAFRAN
Morpho

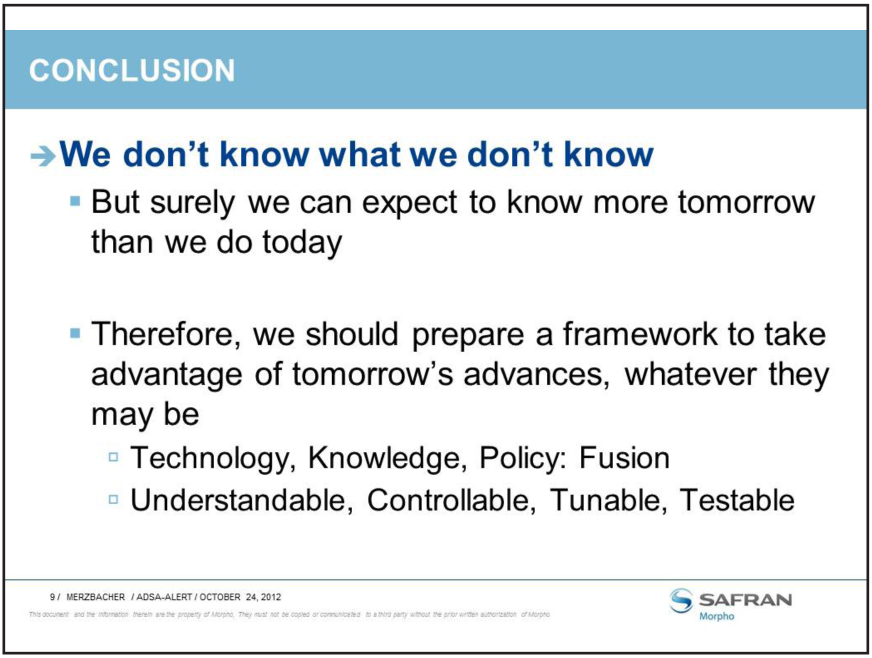


CRAWBACHER LIST

- Why should ATRs be dynamic instead of static?
- What should change with ATRs over time?
- How is the ATR function of:
  - Threat level?
  - Intelligence information?
  - Passenger risk?
  - Deterrence?
  - Randomization?
  - Other?
- How do we prevent overtraining?
- How should requirement specs be set?
- Should a vendor or a third party develop the dynamic ATR?
- How should the following tests be conducted for a dynamic ATR?
  - CRT
  - Certification/qualification
  - FAT/SAT
  - Red team
- How should the various flavors of an ATR be implemented, deployed and activated in the field?
- Should TSA procure scanners w/o ATRs?

8 / MERZBACHER / ADSA-ALERT / OCTOBER 24, 2012

SAFRAN
Morpho



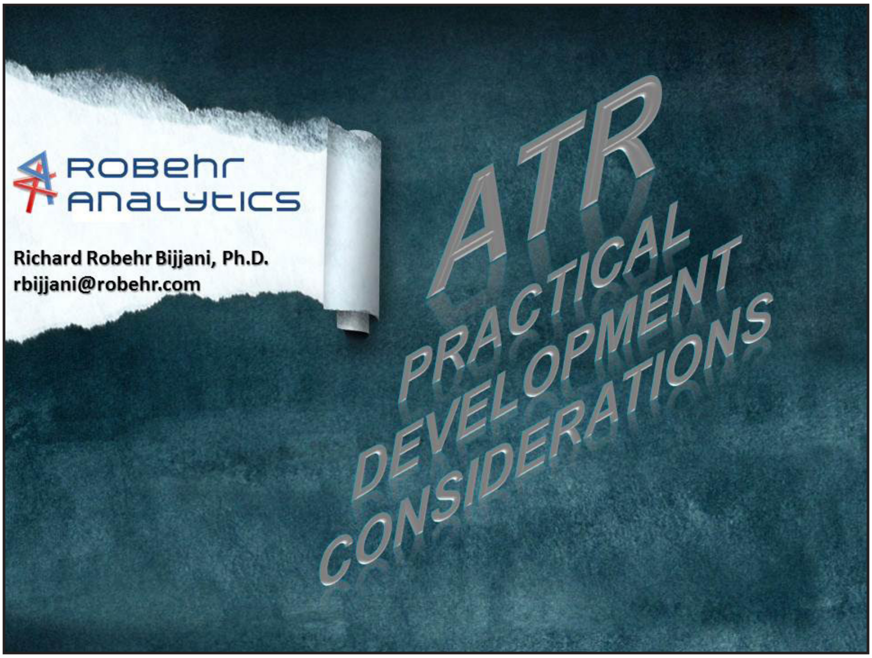
CONCLUSION

- **We don't know what we don't know**
  - But surely we can expect to know more tomorrow than we do today
  - Therefore, we should prepare a framework to take advantage of tomorrow's advances, whatever they may be
    - Technology, Knowledge, Policy: Fusion
    - Understandable, Controllable, Tunable, Testable

9 / MERZBACHER / ADSA-ALERT / OCTOBER 24, 2012

SAFRAN
Morpho

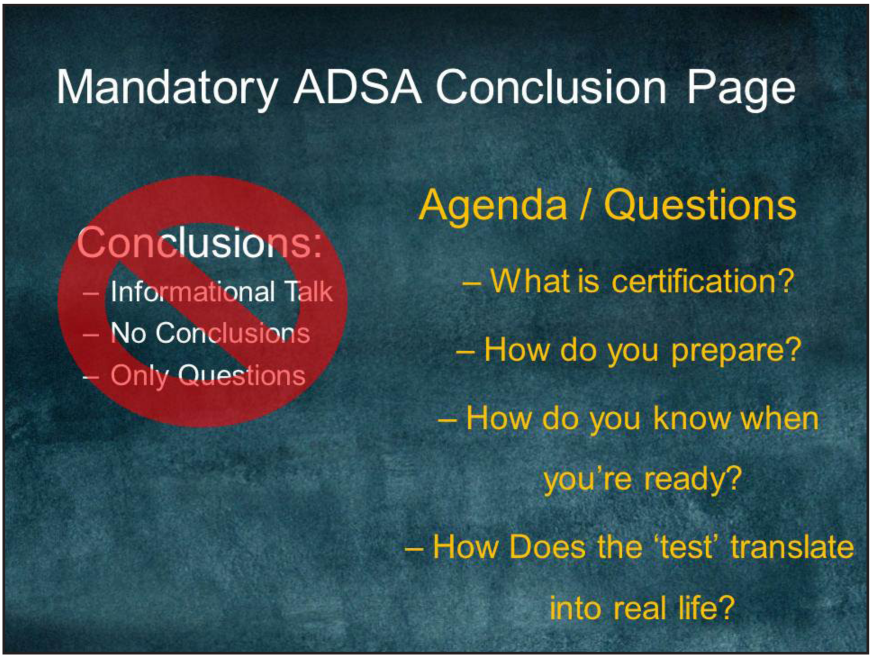
16.18 Richard Bijjani: ATR — Practical Development Considerations



ROBEHR ANALYTICS

Richard Robehr Bijjani, Ph.D.
rbijjani@robehr.com

ATR
PRACTICAL
DEVELOPMENT
CONSIDERATIONS



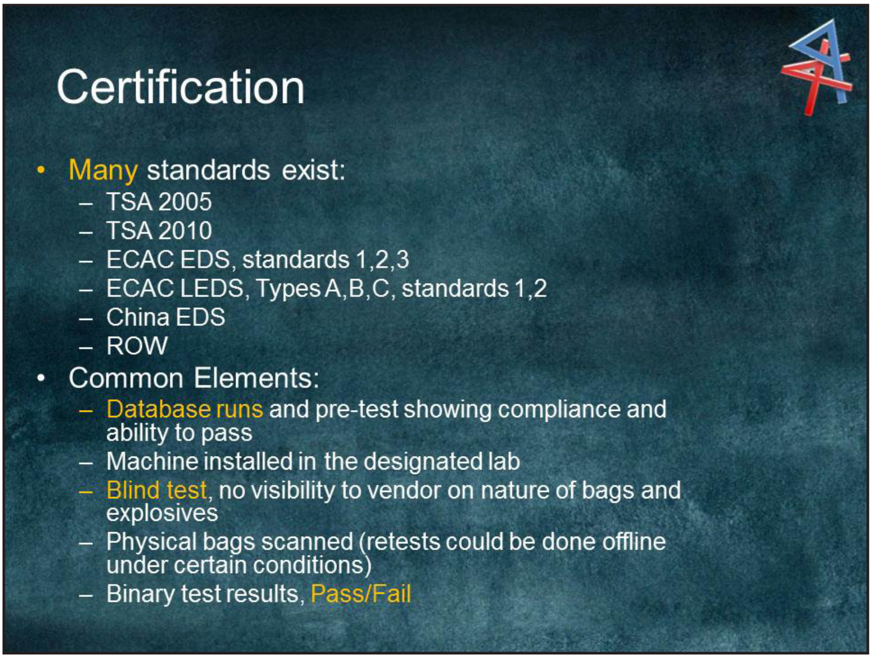
Mandatory ADSA Conclusion Page

Conclusions:
- Informational Talk
- No Conclusions
- Only Questions

Agenda / Questions
- What is certification?
- How do you prepare?
- How do you know when you're ready?
- How Does the 'test' translate into real life?

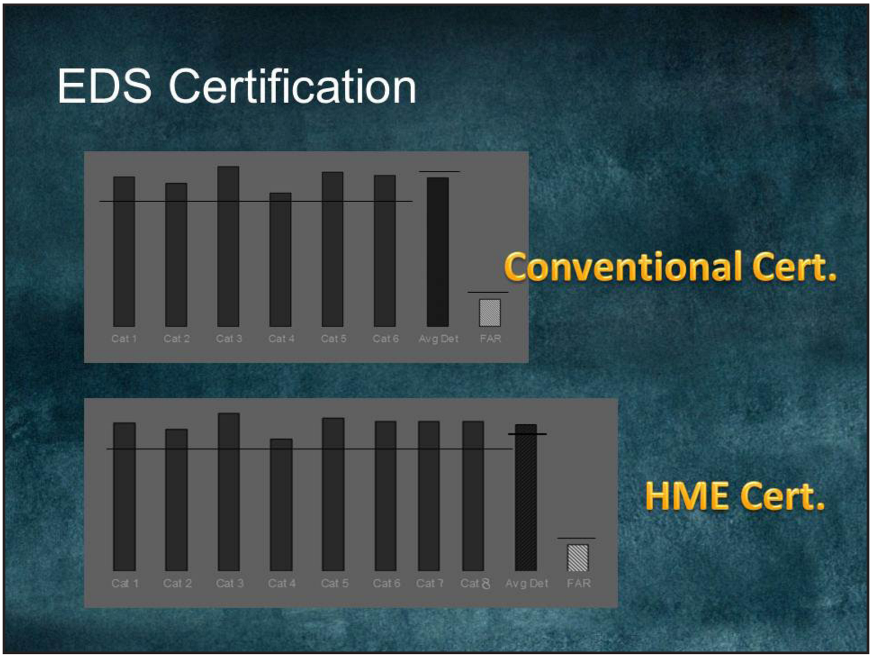
TSA, ECAC, ISA, China, ROW

WHAT IS CERTIFICATION?



Certification

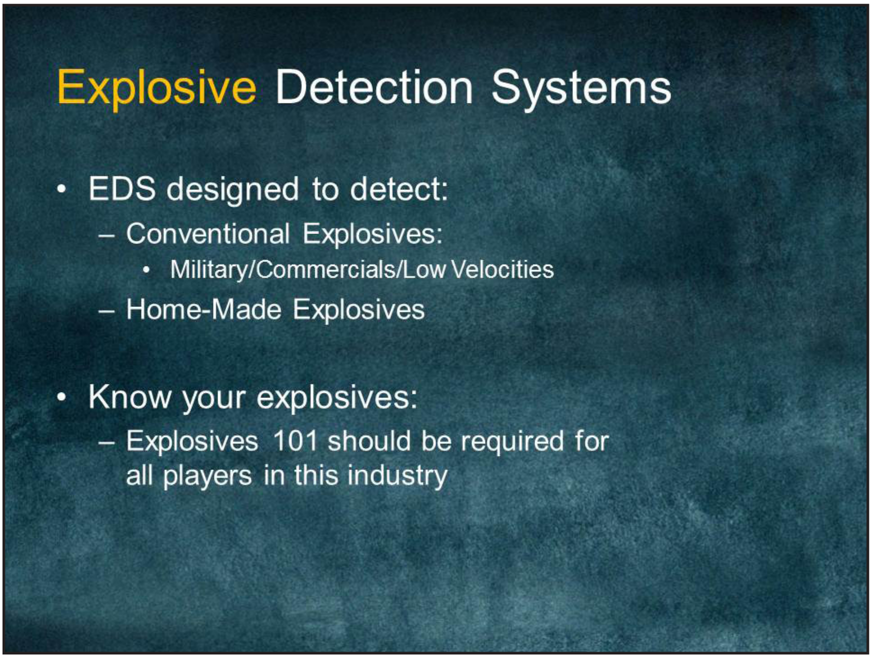
- Many standards exist:
  - TSA 2005
  - TSA 2010
  - ECAC EDS, standards 1,2,3
  - ECAC LEDS, Types A,B,C, standards 1,2
  - China EDS
  - ROW
- Common Elements:
  - Database runs and pre-test showing compliance and ability to pass
  - Machine installed in the designated lab
  - Blind test, no visibility to vendor on nature of bags and explosives
  - Physical bags scanned (retests could be done offline under certain conditions)
  - Binary test results, Pass/Fail



EDS Certification

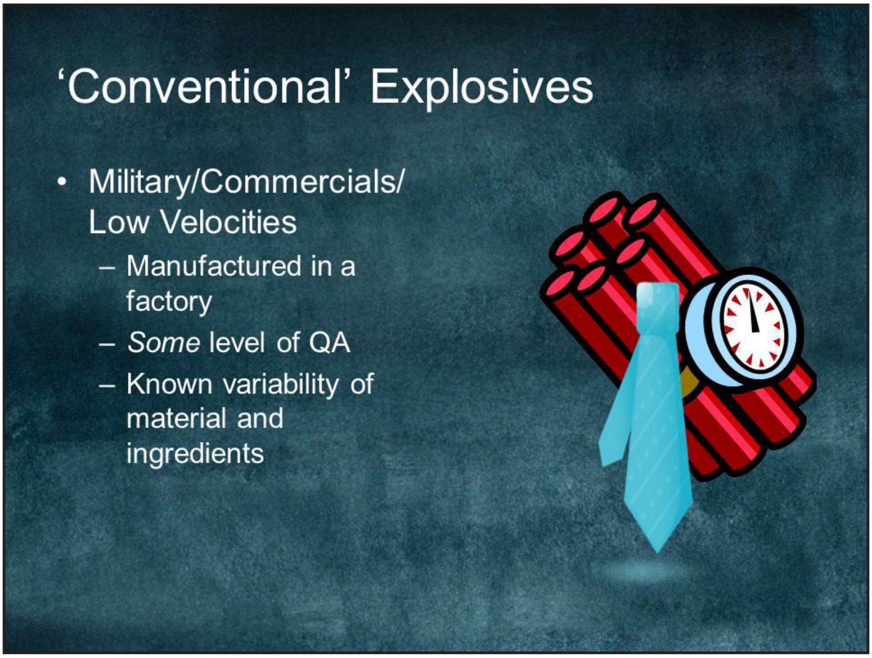
Conventional Cert.

HME Cert.



Explosive Detection Systems

- EDS designed to detect:
  - Conventional Explosives:
    - Military/Commercials/Low Velocities
  - Home-Made Explosives
- Know your explosives:
  - Explosives 101 should be required for all players in this industry



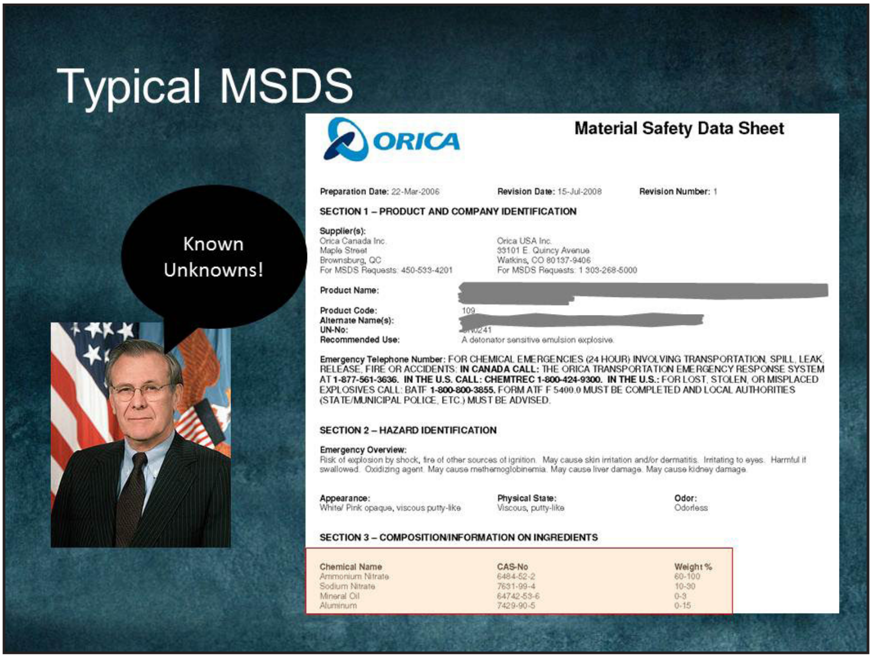
‘Conventional’ Explosives

- Military/Commercials/ Low Velocities
  - Manufactured in a factory
  - *Some* level of QA
  - Known variability of material and ingredients

Categories of Explosives (courtesy of TSL)

Dark Knight Movie Scene

Typical MSDS



ORICA

Material Safety Data Sheet

Preparation Date: 22-Mar-2006 **Revision Date:** 15-Jul-2008 **Revision Number:** 1

SECTION 1 – PRODUCT AND COMPANY IDENTIFICATION

Supplier(s):

Orica Canada Inc.
Maple Street
Brownsburg, QC
For MSDS Requests: 450-533-4201

Orica USA Inc.
33101 E. Quincy Avenue
Watkins, CO 80137-9406
For MSDS Requests: 1 303-268-5000

Product Name: [redacted]

Product Code: 109
Alternate Name(s): [redacted]
UN-No: 0241
Recommended Use: A detonator sensitive emulsion explosive.

Emergency Telephone Number: FOR CHEMICAL EMERGENCIES (24 HOUR) INVOLVING TRANSPORTATION, SPILL, LEAK, RELEASE, FIRE OR ACCIDENTS: **IN CANADA CALL:** THE ORICA TRANSPORTATION EMERGENCY RESPONSE SYSTEM AT **1-877-561-3636. IN THE U.S. CALL: CHEMTREC 1-800-424-9300. IN THE U.S.:** FOR LOST, STOLEN, OR MISPLACED EXPLOSIVES CALL: BATF **1-800-800-3855.** FORM ATF F 5400.0 MUST BE COMPLETED AND LOCAL AUTHORITIES (STATE/MUNICIPAL POLICE, ETC.) MUST BE ADVISED.

SECTION 2 – HAZARD IDENTIFICATION

Emergency Overview:
Risk of explosion by shock, fire of other sources of ignition. May cause skin irritation and/or dermatitis. Irritating to eyes. Harmful if swallowed. Oxidizing agent. May cause methemoglobinemia. May cause liver damage. May cause kidney damage.

Appearance:	**Physical State:**	**Odor:**
White/ Pink opaque, viscous putty-like	Viscous, putty-like	Odorless

SECTION 3 – COMPOSITION/INFORMATION ON INGREDIENTS

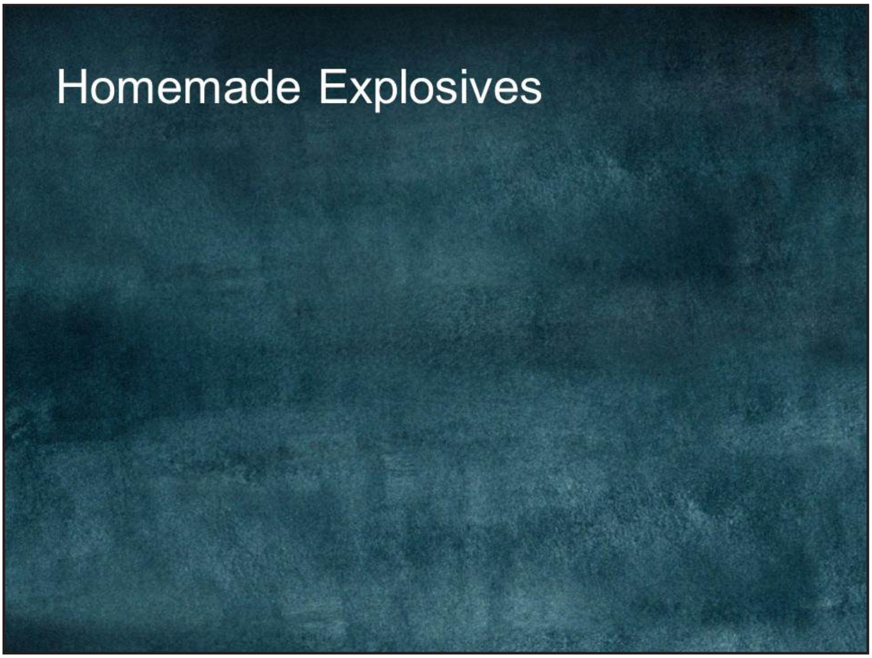
Chemical Name	CAS-No	Weight %
Ammonium Nitrate	6484-52-2	60-100
Sodium Nitrate	7631-99-4	10-30
Mineral Oil	64742-53-6	0-3
Aluminum	7429-90-5	0-15

Sensitivity to Ingredients

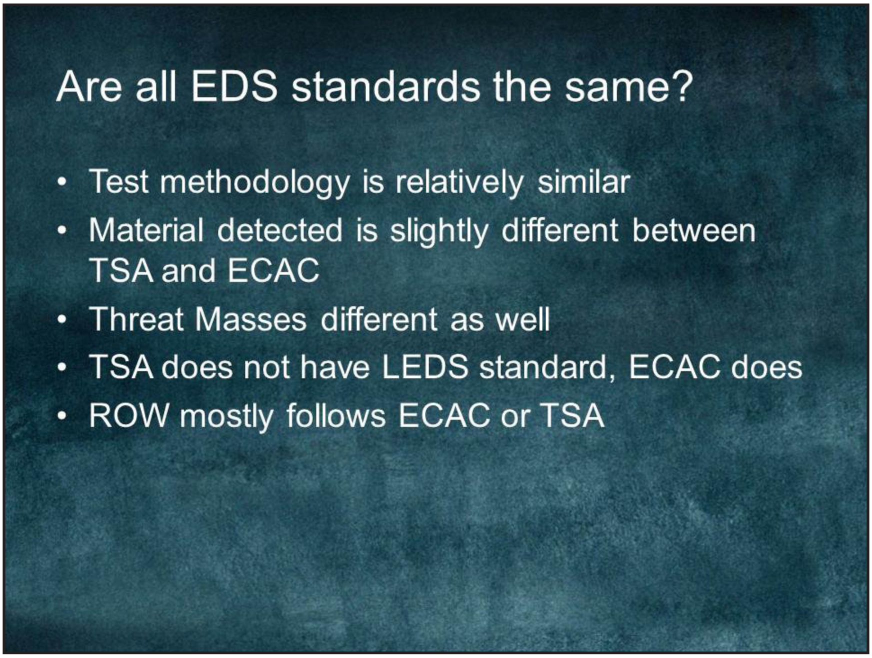
Homemade Explosives

Unknown Unknowns!!

- What are they?
  - Liquids or solids/powders
  - Unknown formulations
  - Unknown ingredients
  - Unknown Contaminants
  - Unknown tapping pressure
  - No QA
- What do we know?
  - Explode
  - *Limited* list of known oxidizers
  - Unlimited list of fuels
  - Certain requirements for oxidizer/fuel ratios

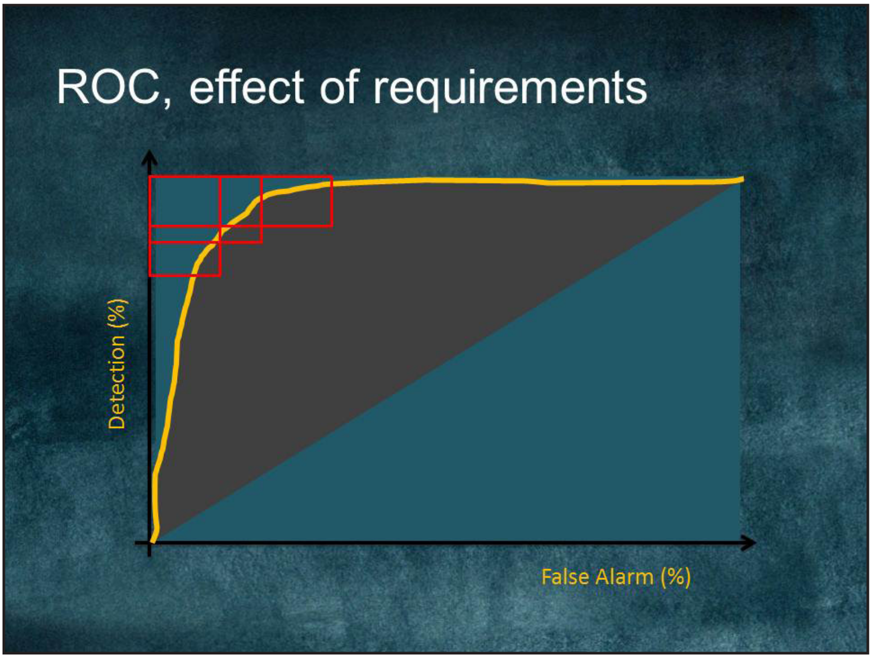


Homemade Explosives

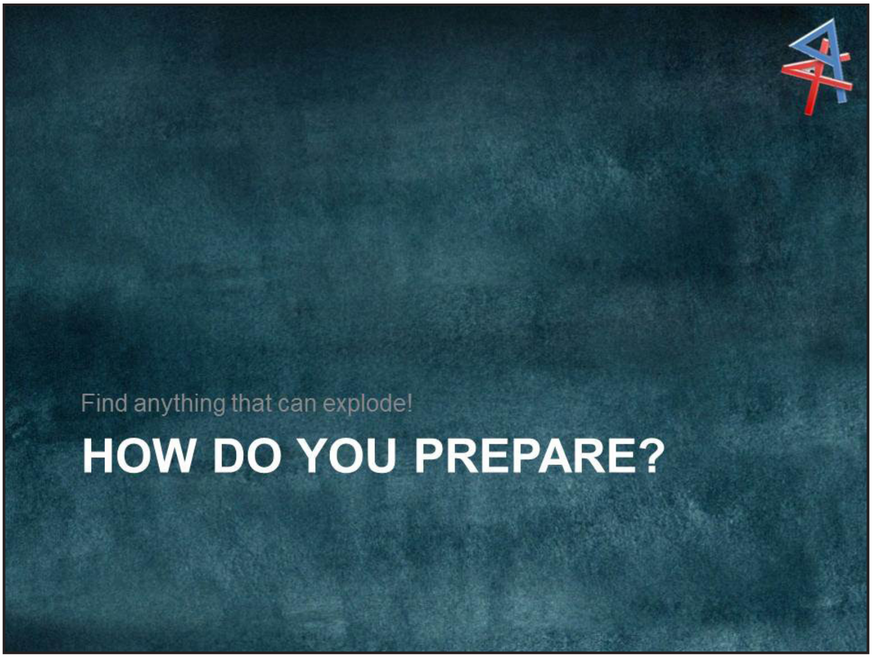


Are all EDS standards the same?

- Test methodology is relatively similar
- Material detected is slightly different between TSA and ECAC
- Threat Masses different as well
- TSA does not have LEDS standard, ECAC does
- ROW mostly follows ECAC or TSA



ROC, effect of requirements



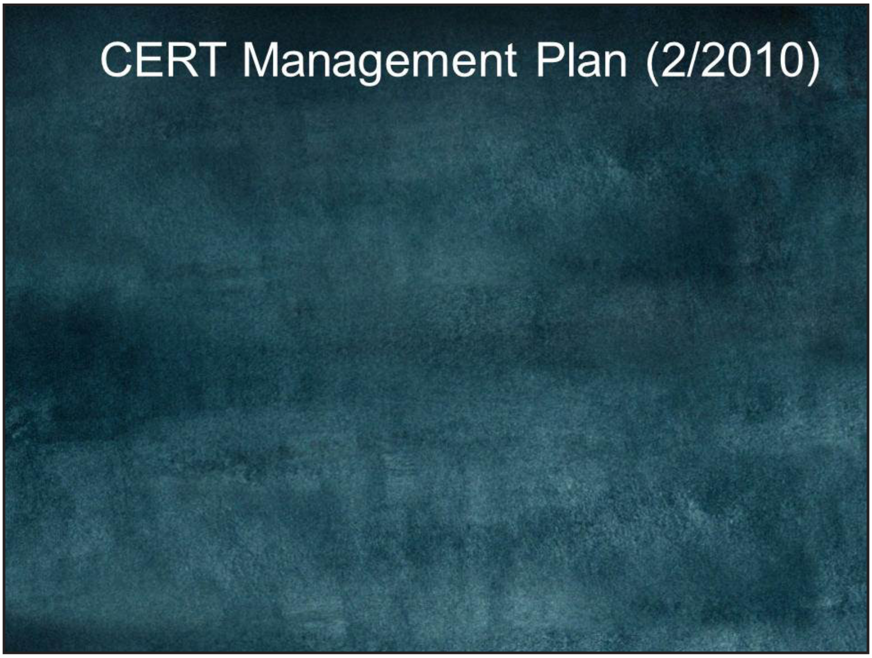
Find anything that can explode!

HOW DO YOU PREPARE?

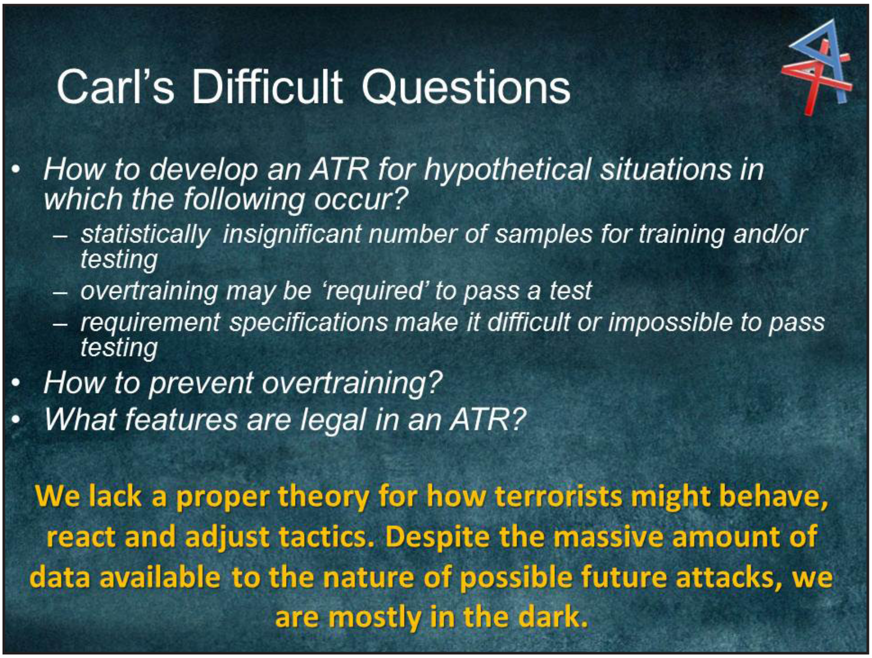
Preparation (the long winding road to Certification)

- Get Data
- Develop Algorithms
- Take Test

Rinse and repeat



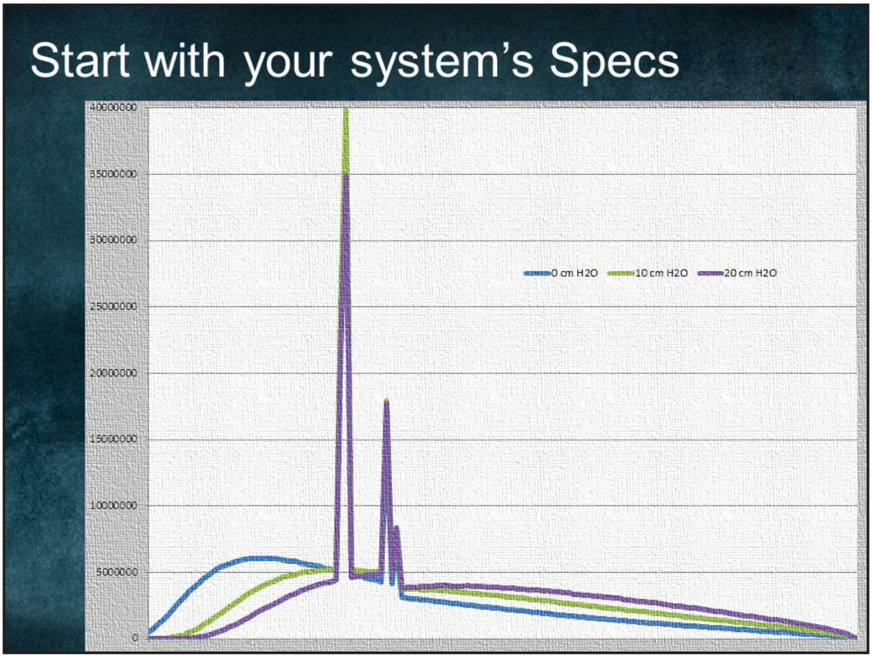
CERT Management Plan (2/2010)



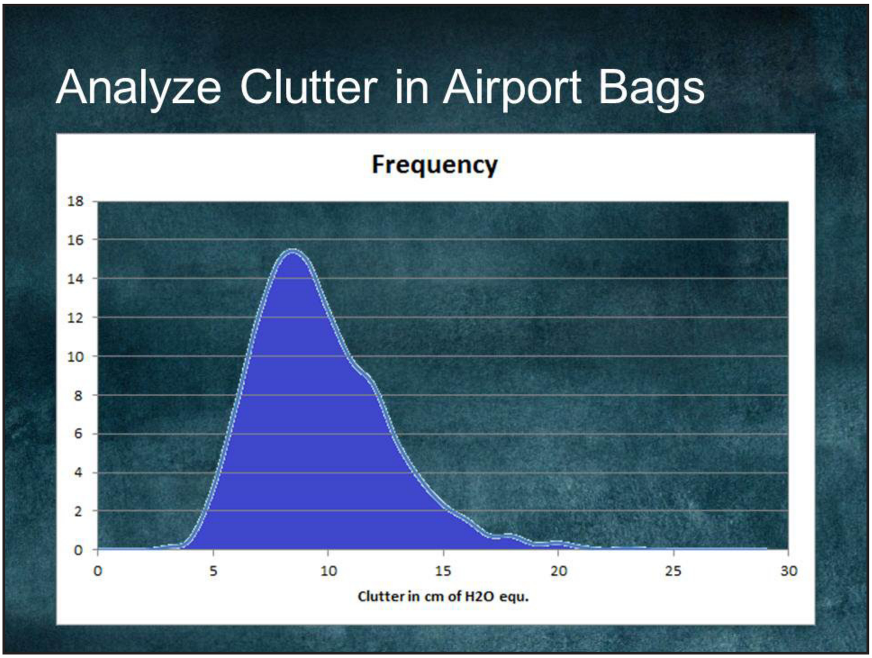
Carl’s Difficult Questions

- *How to develop an ATR for hypothetical situations in which the following occur?*
  - *statistically insignificant number of samples for training and/or testing*
  - *overtraining may be ‘required’ to pass a test*
  - *requirement specifications make it difficult or impossible to pass testing*
- *How to prevent overtraining?*
- *What features are legal in an ATR?*

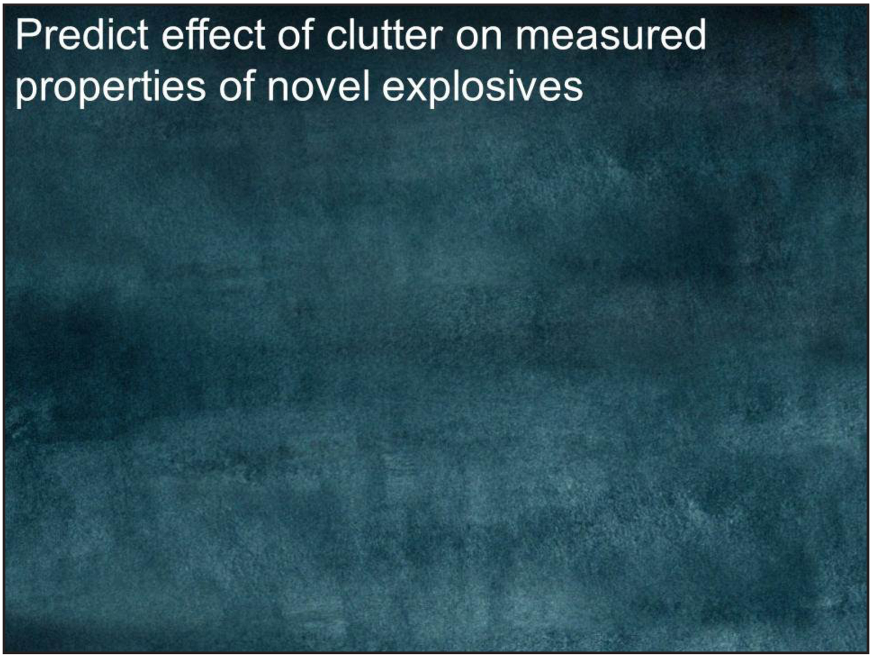
We lack a proper theory for how terrorists might behave, react and adjust tactics. Despite the massive amount of data available to the nature of possible future attacks, we are mostly in the dark.



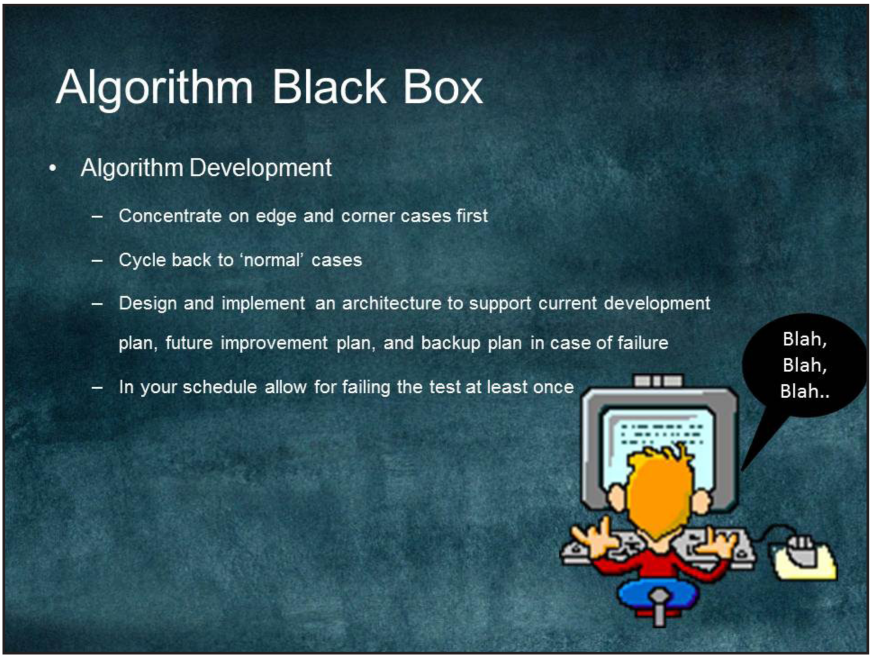
Start with your system's Specs



Analyze Clutter in Airport Bags

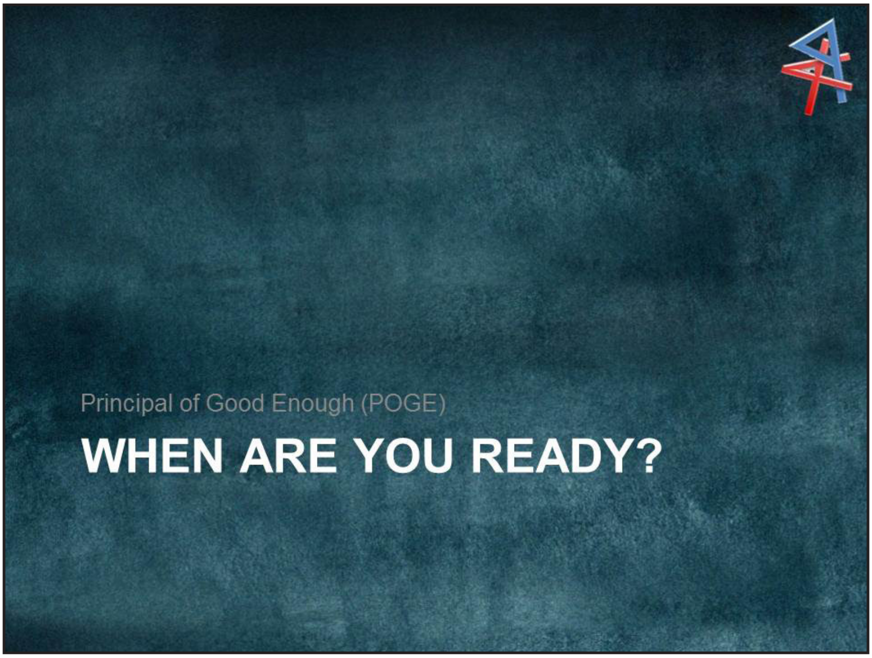


Predict effect of clutter on measured properties of novel explosives



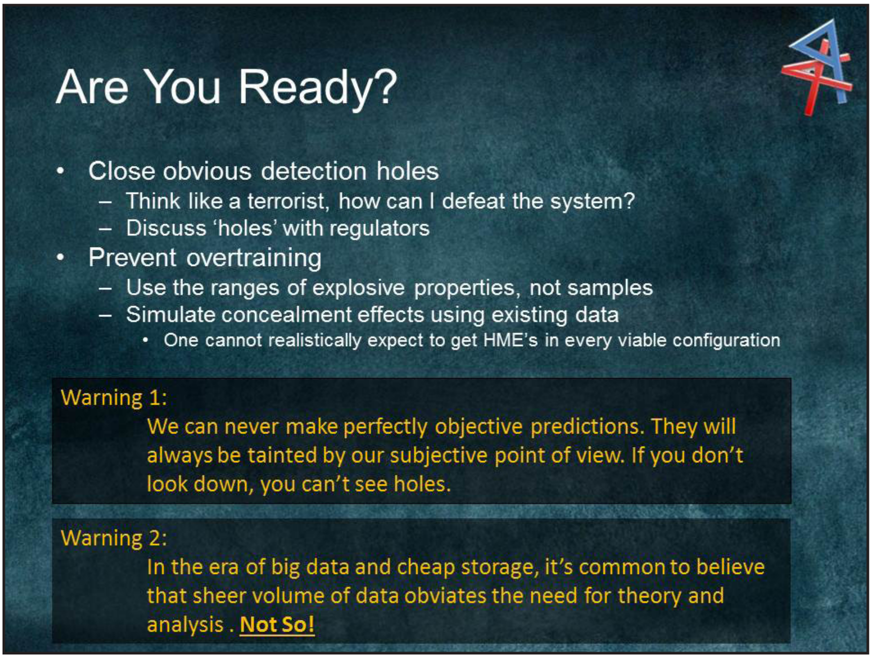
Algorithm Black Box

- Algorithm Development
  - Concentrate on edge and corner cases first
  - Cycle back to 'normal' cases
  - Design and implement an architecture to support current development plan, future improvement plan, and backup plan in case of failure
  - In your schedule allow for failing the test at least once



Principal of Good Enough (POGE)

WHEN ARE YOU READY?

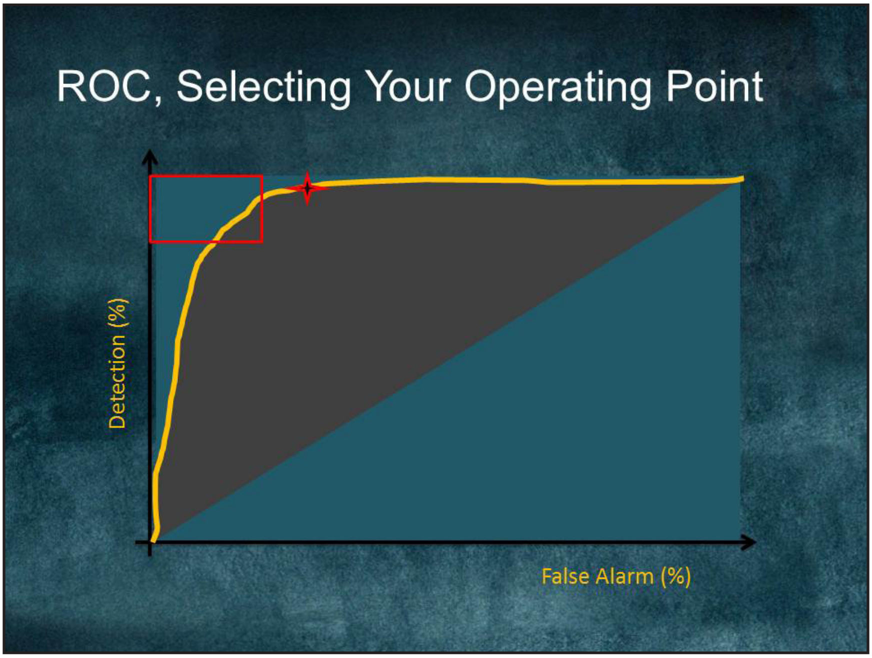


Are You Ready?

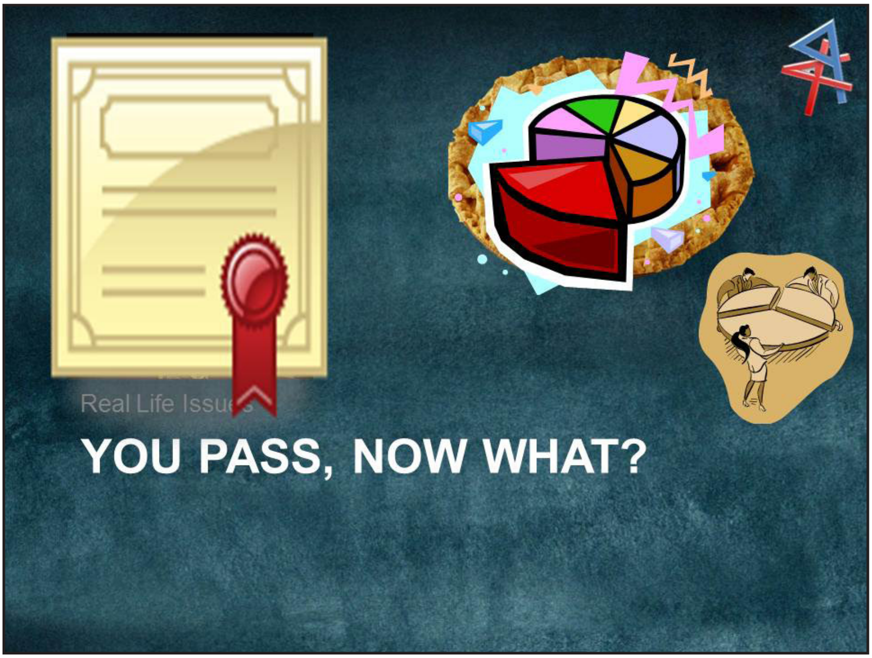
- Close obvious detection holes
  - Think like a terrorist, how can I defeat the system?
  - Discuss ‘holes’ with regulators
- Prevent overtraining
  - Use the ranges of explosive properties, not samples
  - Simulate concealment effects using existing data
    - One cannot realistically expect to get HME’s in every viable configuration

Warning 1:
We can never make perfectly objective predictions. They will always be tainted by our subjective point of view. If you don’t look down, you can’t see holes.

Warning 2:
In the era of big data and cheap storage, it’s common to believe that sheer volume of data obviates the need for theory and analysis . **Not So!**

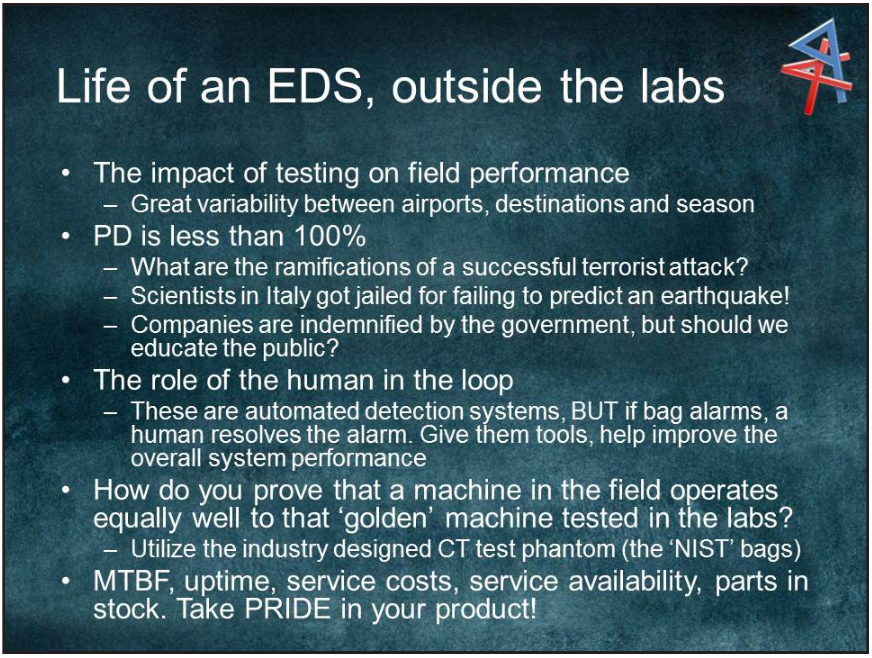


ROC, Selecting Your Operating Point



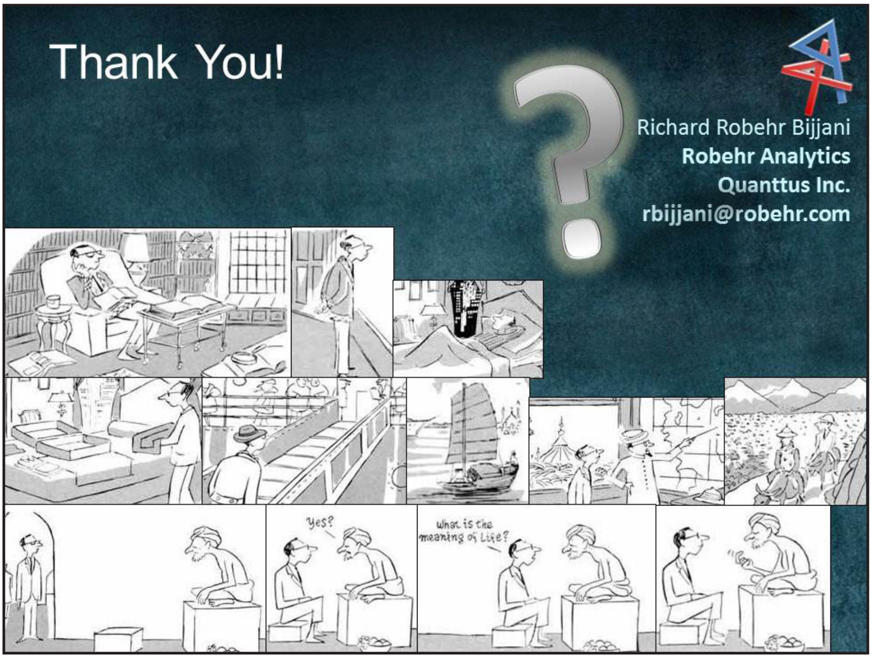
Real Life Issues

YOU PASS, NOW WHAT?



Life of an EDS, outside the labs

- The impact of testing on field performance
  - Great variability between airports, destinations and season
- PD is less than 100%
  - What are the ramifications of a successful terrorist attack?
  - Scientists in Italy got jailed for failing to predict an earthquake!
  - Companies are indemnified by the government, but should we educate the public?
- The role of the human in the loop
  - These are automated detection systems, BUT if bag alarms, a human resolves the alarm. Give them tools, help improve the overall system performance
- How do you prove that a machine in the field operates equally well to that 'golden' machine tested in the labs?
  - Utilize the industry designed CT test phantom (the 'NIST' bags)
- MTBF, uptime, service costs, service availability, parts in stock. Take PRIDE in your product!

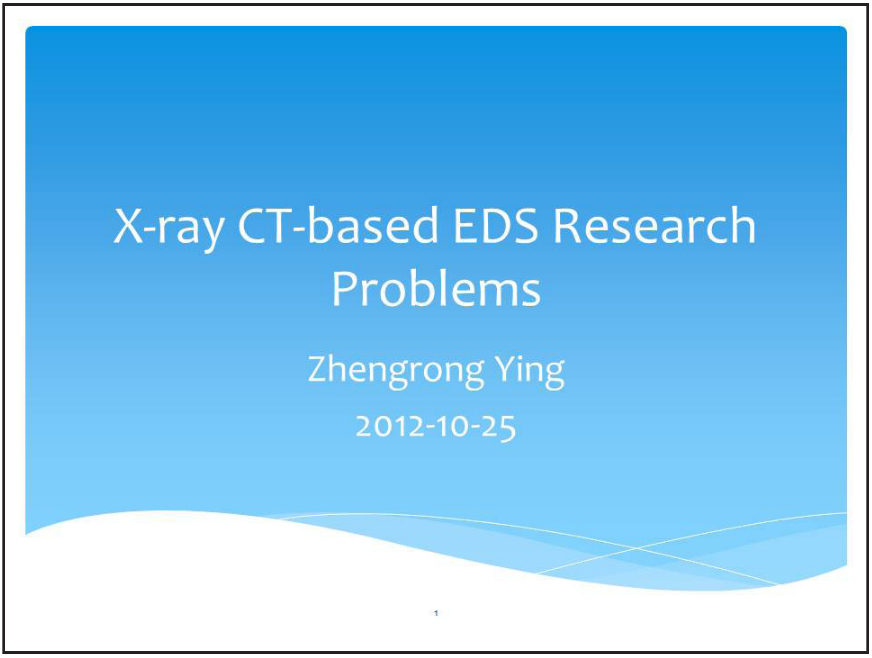


Thank You!

Richard Robehr Bijjani
Robehr Analytics
Quanttus Inc.
rbijjani@robehr.com


Listen, I've got my own problems!
dumas

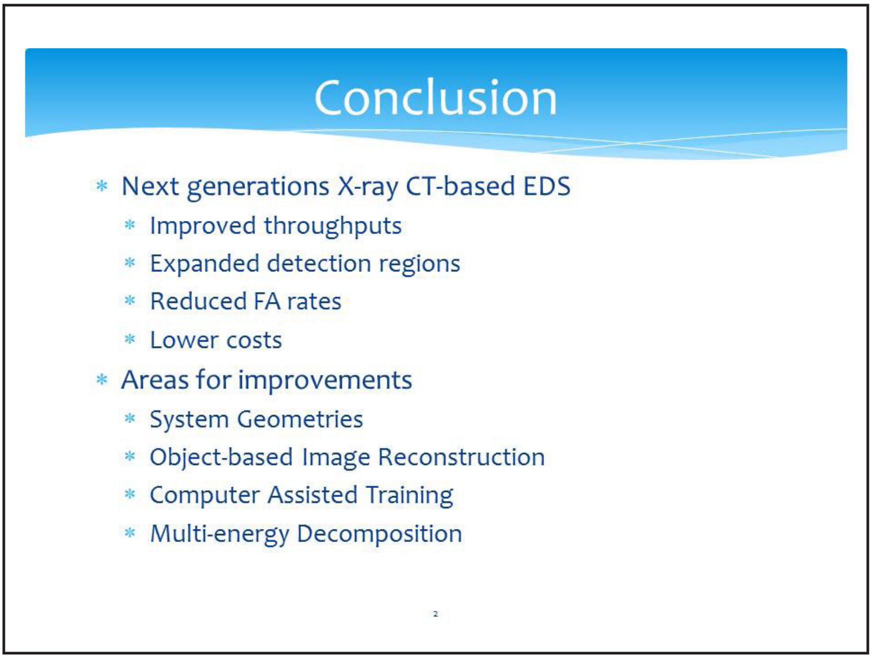
16.19 Zhengrong Ying: EDS Research Problems



X-ray CT-based EDS Research Problems

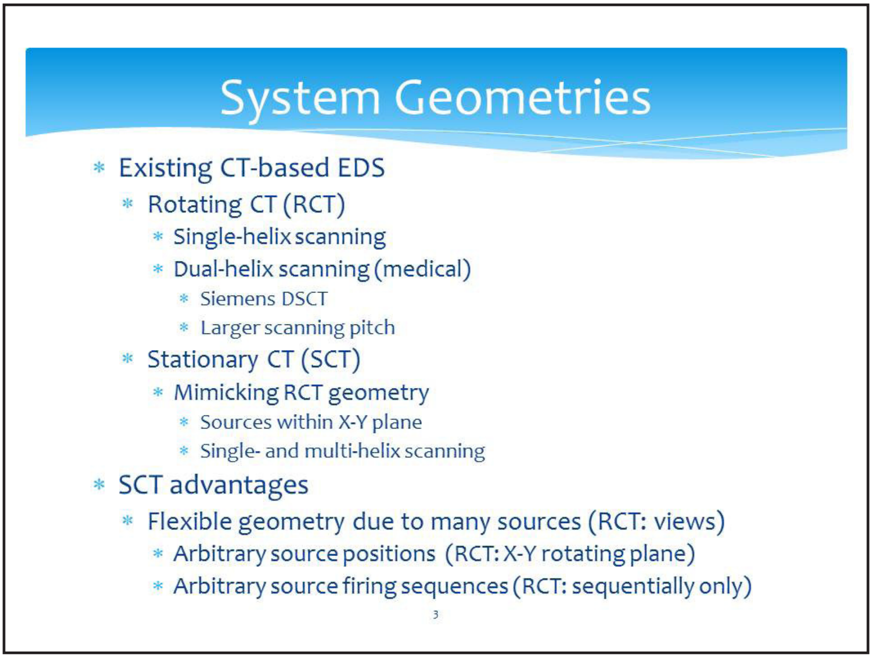
Zhengrong Ying

2012-10-25



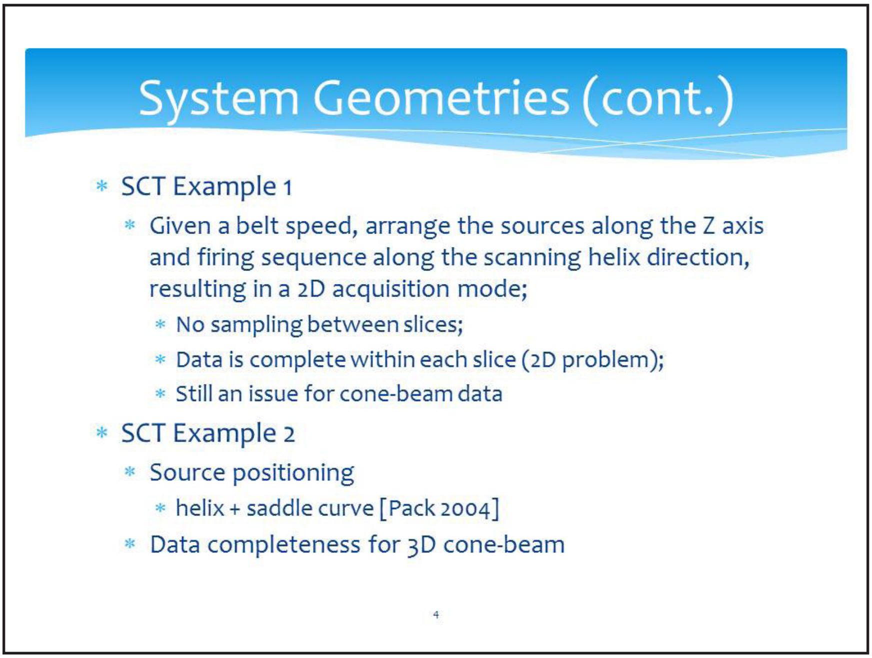
Conclusion

- Next generations X-ray CT-based EDS
  - Improved throughputs
  - Expanded detection regions
  - Reduced FA rates
  - Lower costs
- Areas for improvements
  - System Geometries
  - Object-based Image Reconstruction
  - Computer Assisted Training
  - Multi-energy Decomposition



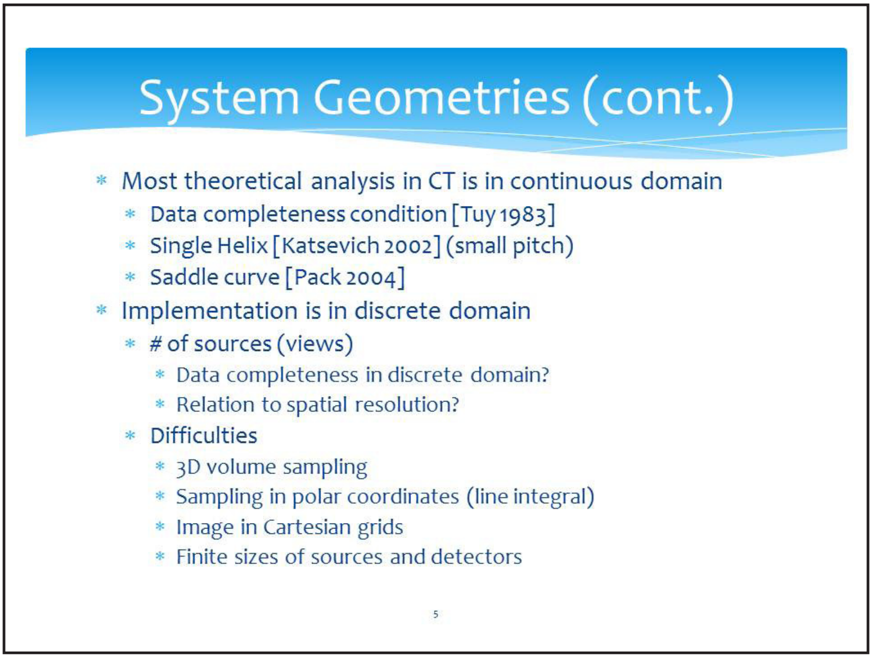
System Geometries

- Existing CT-based EDS
  - Rotating CT (RCT)
    - Single-helix scanning
    - Dual-helix scanning (medical)
      - Siemens DSCT
      - Larger scanning pitch
  - Stationary CT (SCT)
    - Mimicking RCT geometry
      - Sources within X-Y plane
      - Single- and multi-helix scanning
- SCT advantages
  - Flexible geometry due to many sources (RCT: views)
    - Arbitrary source positions (RCT: X-Y rotating plane)
    - Arbitrary source firing sequences (RCT: sequentially only)



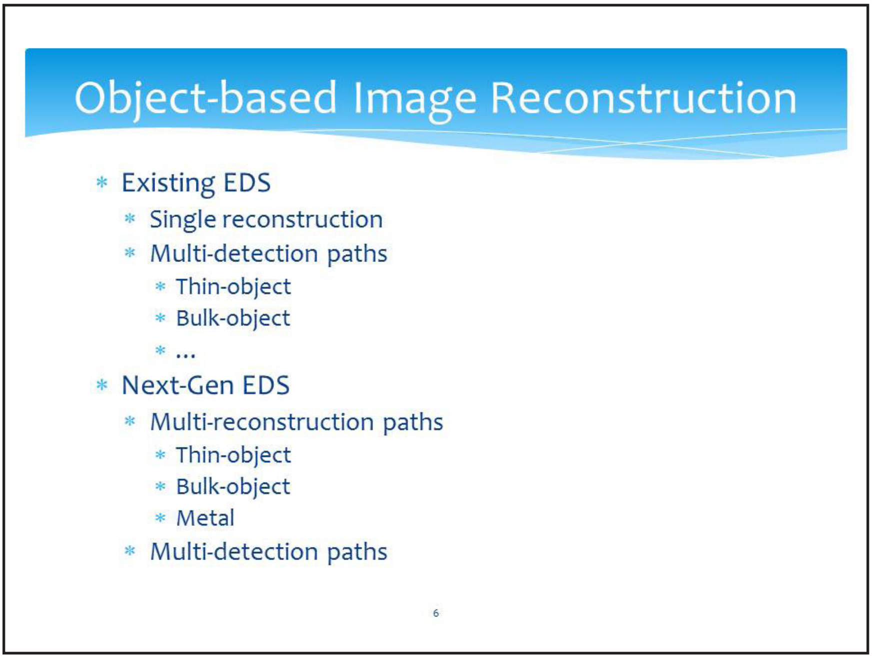
System Geometries (cont.)

- SCT Example 1
  - Given a belt speed, arrange the sources along the Z axis and firing sequence along the scanning helix direction, resulting in a 2D acquisition mode;
    - No sampling between slices;
    - Data is complete within each slice (2D problem);
    - Still an issue for cone-beam data
- SCT Example 2
  - Source positioning
    - helix + saddle curve [Pack 2004]
  - Data completeness for 3D cone-beam



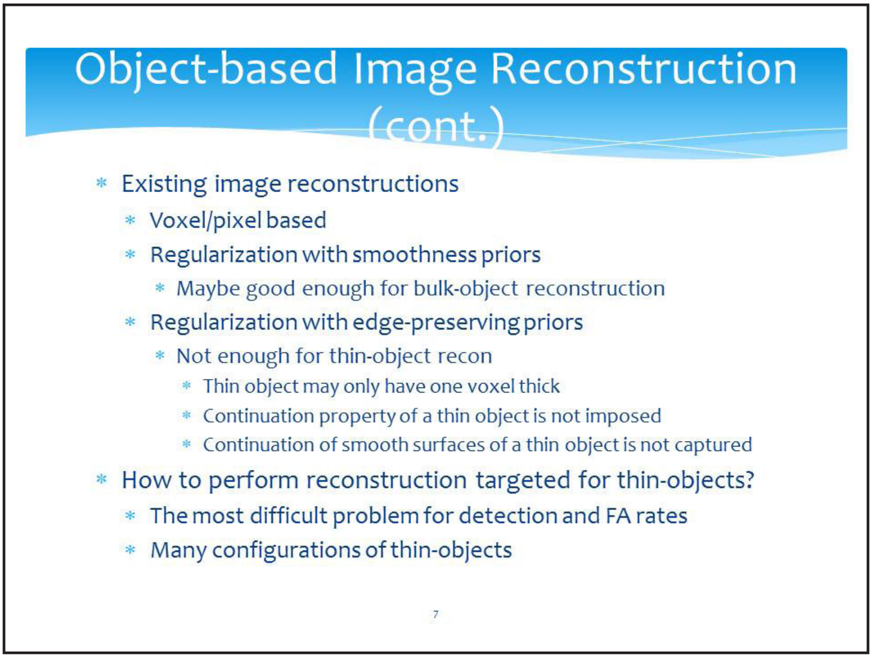
System Geometries (cont.)

- Most theoretical analysis in CT is in continuous domain
  - Data completeness condition [Tuy 1983]
  - Single Helix [Katsevich 2002] (small pitch)
  - Saddle curve [Pack 2004]
- Implementation is in discrete domain
  - # of sources (views)
    - Data completeness in discrete domain?
    - Relation to spatial resolution?
  - Difficulties
    - 3D volume sampling
    - Sampling in polar coordinates (line integral)
    - Image in Cartesian grids
    - Finite sizes of sources and detectors



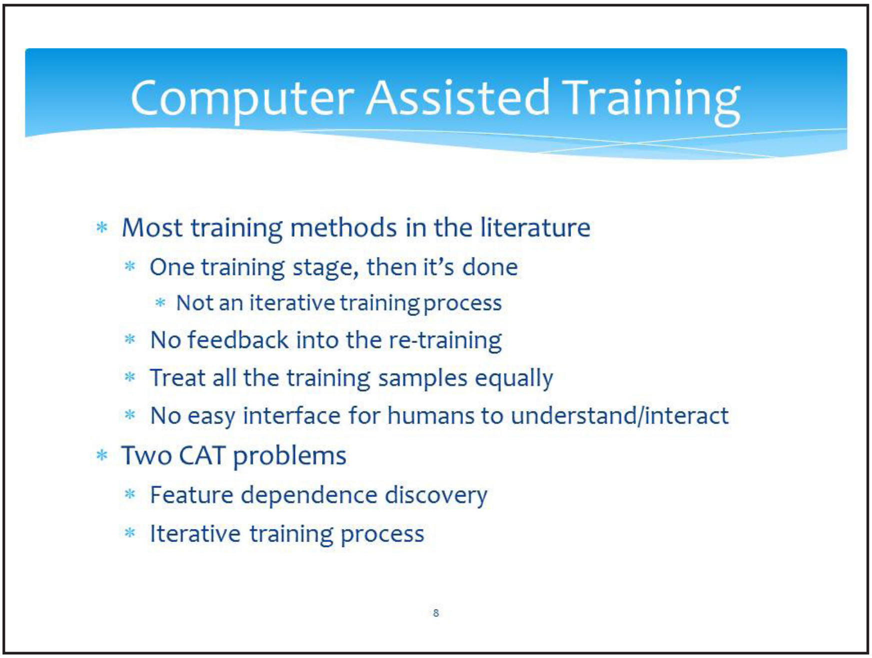
Object-based Image Reconstruction

- Existing EDS
  - Single reconstruction
  - Multi-detection paths
    - Thin-object
    - Bulk-object
    - …
- Next-Gen EDS
  - Multi-reconstruction paths
    - Thin-object
    - Bulk-object
    - Metal
  - Multi-detection paths



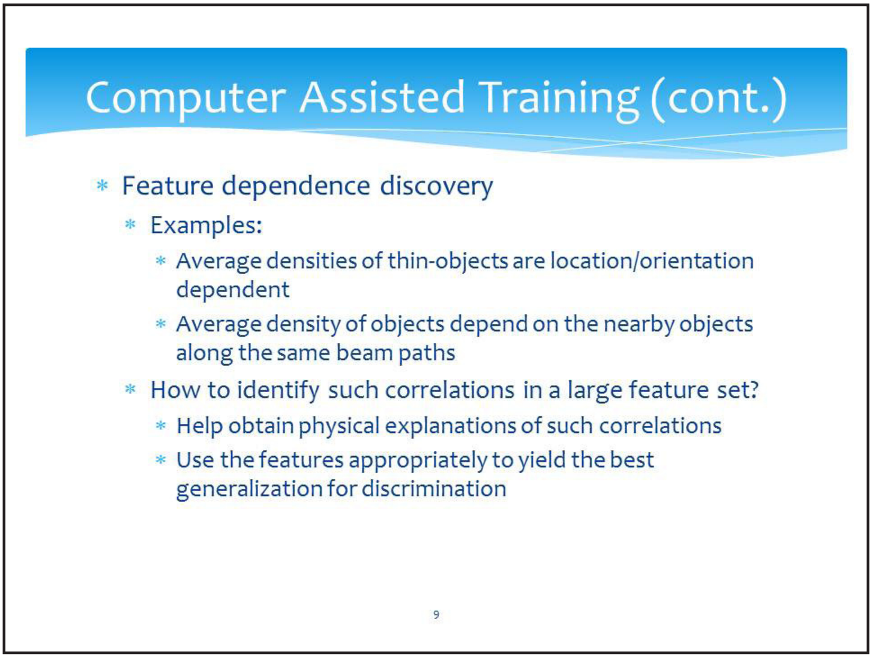
Object-based Image Reconstruction (cont.)

- Existing image reconstructions
  - Voxel/pixel based
  - Regularization with smoothness priors
    - Maybe good enough for bulk-object reconstruction
  - Regularization with edge-preserving priors
    - Not enough for thin-object recon
      - Thin object may only have one voxel thick
      - Continuation property of a thin object is not imposed
      - Continuation of smooth surfaces of a thin object is not captured
- How to perform reconstruction targeted for thin-objects?
  - The most difficult problem for detection and FA rates
  - Many configurations of thin-objects



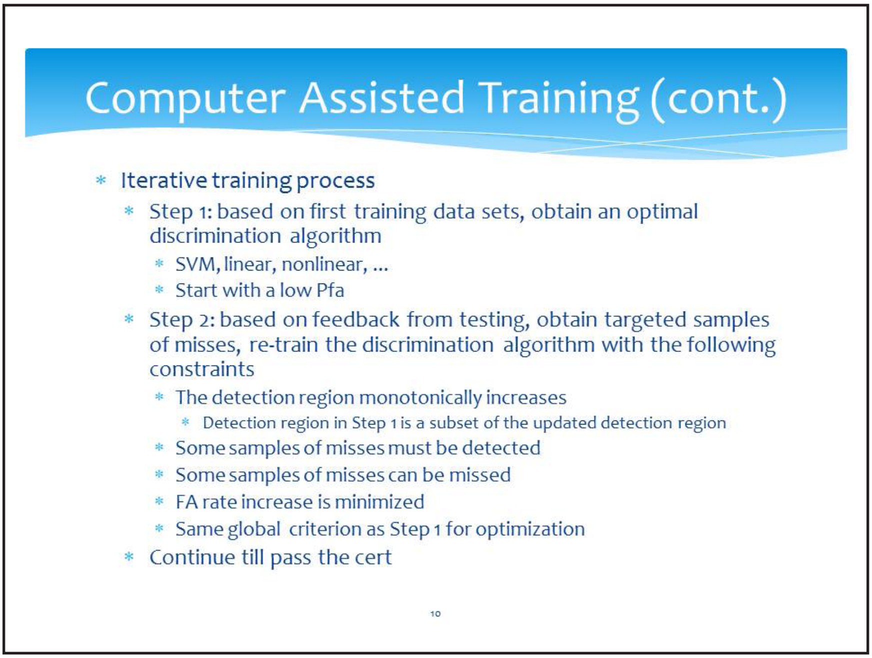
Computer Assisted Training

- Most training methods in the literature
  - One training stage, then it's done
    - Not an iterative training process
  - No feedback into the re-training
  - Treat all the training samples equally
  - No easy interface for humans to understand/interact
- Two CAT problems
  - Feature dependence discovery
  - Iterative training process



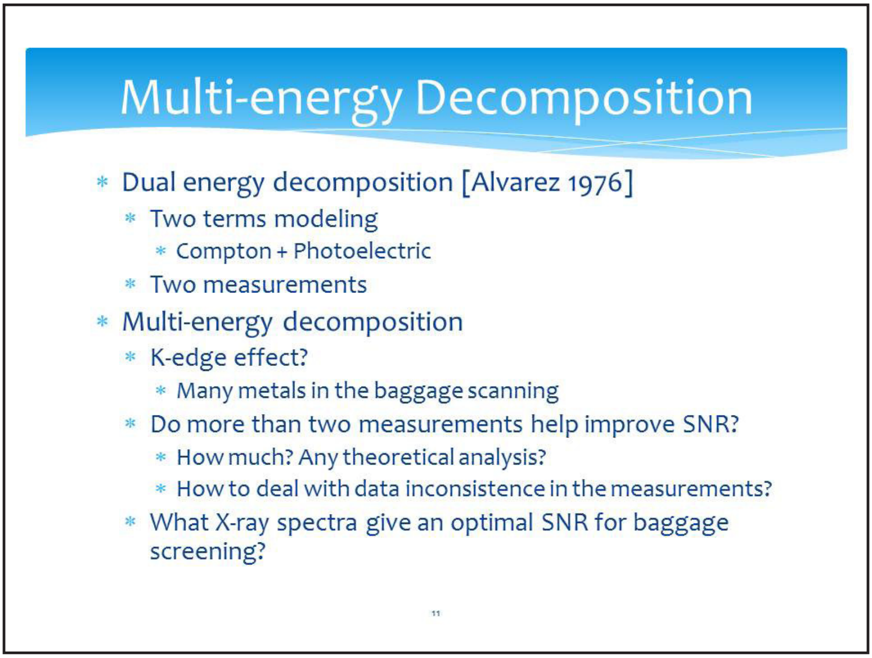
Computer Assisted Training (cont.)

- Feature dependence discovery
  - Examples:
    - Average densities of thin-objects are location/orientation dependent
    - Average density of objects depend on the nearby objects along the same beam paths
  - How to identify such correlations in a large feature set?
    - Help obtain physical explanations of such correlations
    - Use the features appropriately to yield the best generalization for discrimination



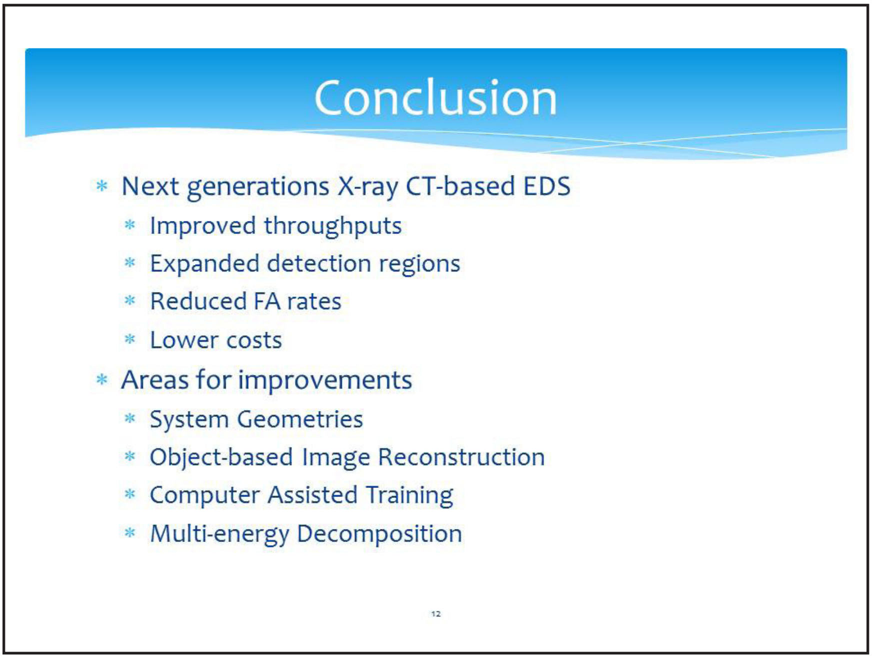
Computer Assisted Training (cont.)

- Iterative training process
  - Step 1: based on first training data sets, obtain an optimal discrimination algorithm
    - SVM, linear, nonlinear, …
    - Start with a low Pfa
  - Step 2: based on feedback from testing, obtain targeted samples of misses, re-train the discrimination algorithm with the following constraints
    - The detection region monotonically increases
      - Detection region in Step 1 is a subset of the updated detection region
    - Some samples of misses must be detected
    - Some samples of misses can be missed
    - FA rate increase is minimized
    - Same global criterion as Step 1 for optimization
  - Continue till pass the cert



Multi-energy Decomposition

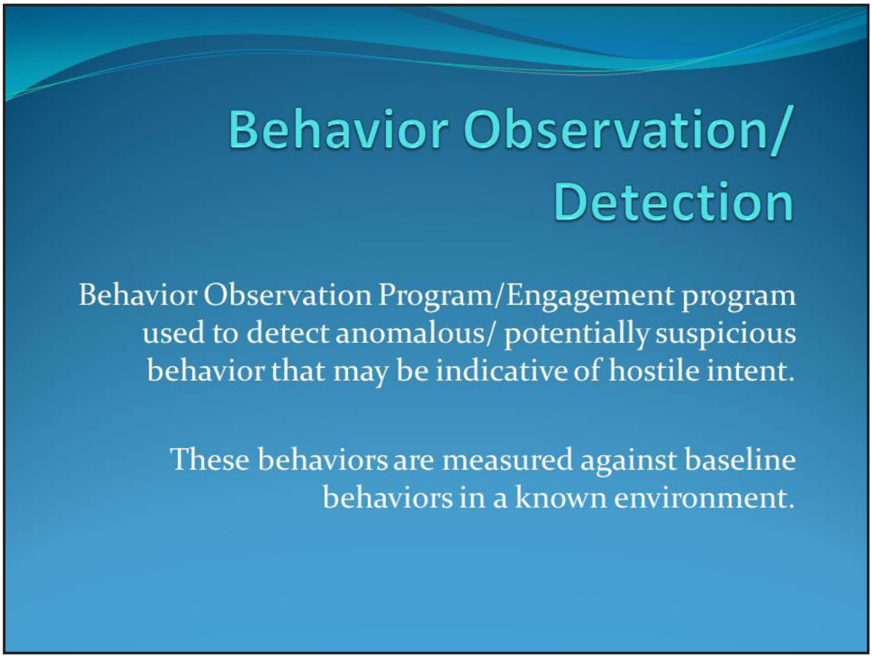
- Dual energy decomposition [Alvarez 1976]
  - Two terms modeling
    - Compton + Photoelectric
  - Two measurements
- Multi-energy decomposition
  - K-edge effect?
    - Many metals in the baggage scanning
  - Do more than two measurements help improve SNR?
    - How much? Any theoretical analysis?
    - How to deal with data inconsistence in the measurements?
  - What X-ray spectra give an optimal SNR for baggage screening?



Conclusion

- Next generations X-ray CT-based EDS
  - Improved throughputs
  - Expanded detection regions
  - Reduced FA rates
  - Lower costs
- Areas for improvements
  - System Geometries
  - Object-based Image Reconstruction
  - Computer Assisted Training
  - Multi-energy Decomposition

16.20 Carl Maccario: Aberrant Behavior and Risk Based Screening



Behavior Observation/ Detection

Behavior Observation Program/Engagement program used to detect anomalous/ potentially suspicious behavior that may be indicative of hostile intent.

These behaviors are measured against baseline behaviors in a known environment.

Behavior Observation/ Detection Concept

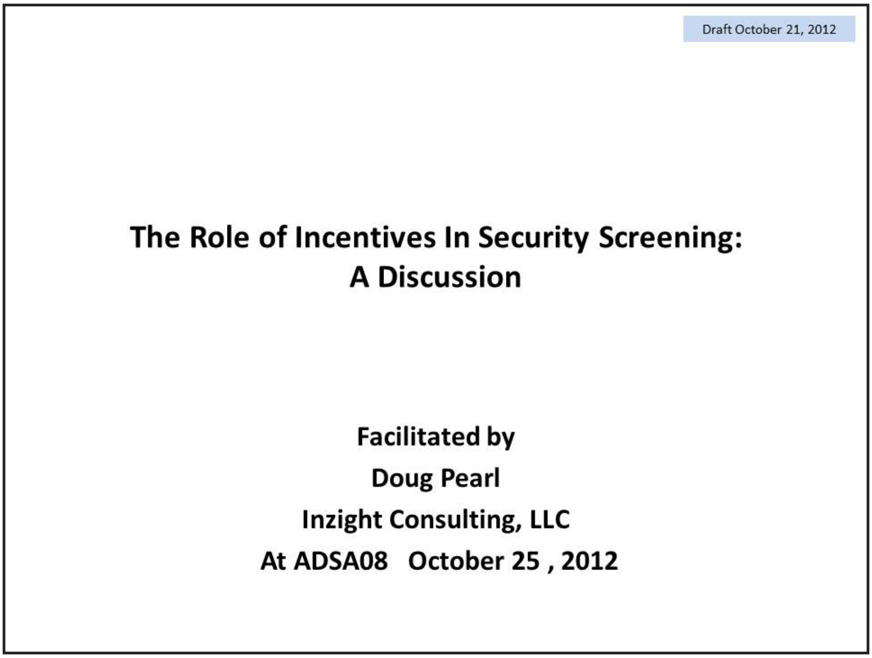
Involves fear of discovery

Person who fears being discovered will suffer mental stress, fear, or anxiety

Manifested through involuntary physical and physiological reactions

For the terrorist, fear of discovery not related to loss of life but fear of failing to complete the mission

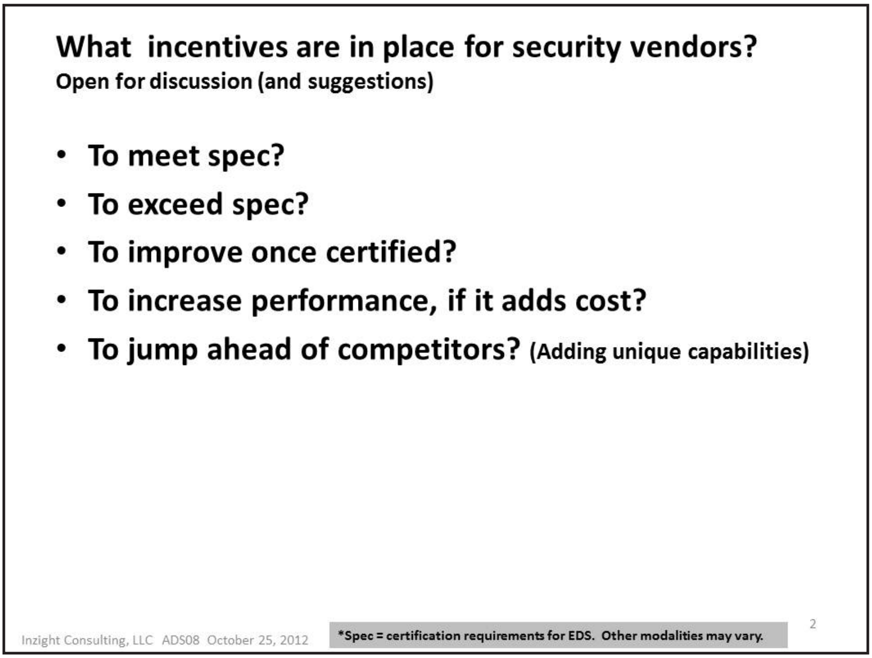
16.21 Doug Pearl: Discussion: Role of Incentives in Security Imaging



Draft October 21, 2012

**The Role of Incentives In Security Screening:
A Discussion**

**Facilitated by
Doug Pearl
Inzight Consulting, LLC
At ADSA08 October 25 , 2012**



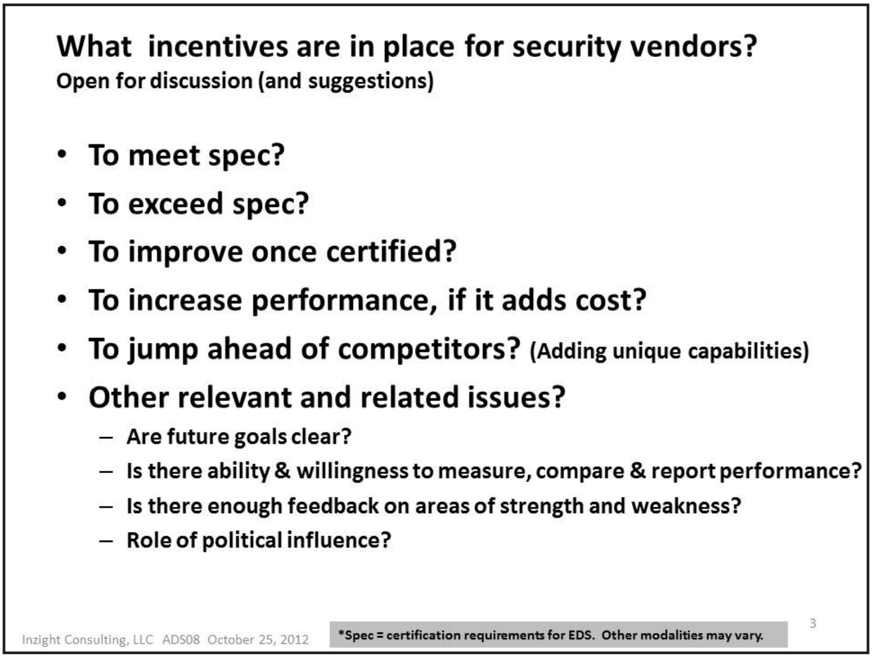
What incentives are in place for security vendors?
Open for discussion (and suggestions)

- **To meet spec?**
- **To exceed spec?**
- **To improve once certified?**
- **To increase performance, if it adds cost?**
- **To jump ahead of competitors? (Adding unique capabilities)**

Inzight Consulting, LLC ADS08 October 25, 2012

***Spec = certification requirements for EDS. Other modalities may vary.**

2

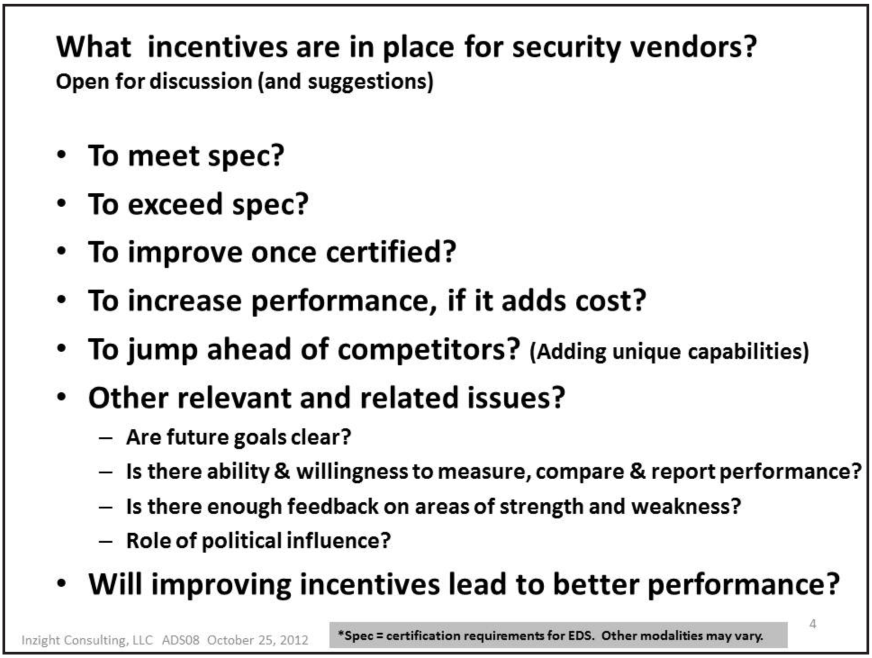


What incentives are in place for security vendors?

Open for discussion (and suggestions)

- To meet spec?
- To exceed spec?
- To improve once certified?
- To increase performance, if it adds cost?
- To jump ahead of competitors? (Adding unique capabilities)
- Other relevant and related issues?
  - Are future goals clear?
  - Is there ability & willingness to measure, compare & report performance?
  - Is there enough feedback on areas of strength and weakness?
  - Role of political influence?

*Spec = certification requirements for EDS. Other modalities may vary.

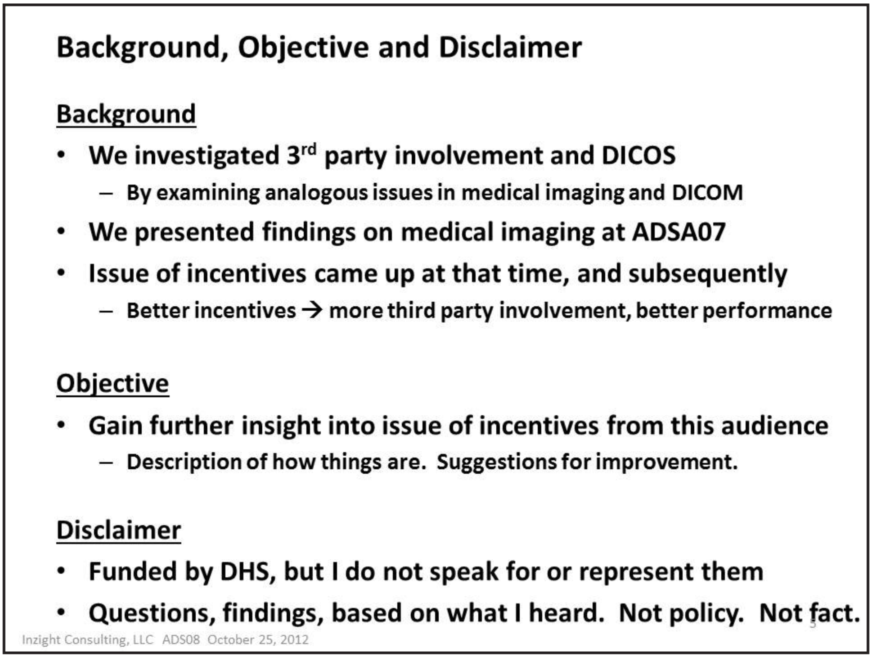


What incentives are in place for security vendors?

Open for discussion (and suggestions)

- To meet spec?
- To exceed spec?
- To improve once certified?
- To increase performance, if it adds cost?
- To jump ahead of competitors? (Adding unique capabilities)
- Other relevant and related issues?
  - Are future goals clear?
  - Is there ability & willingness to measure, compare & report performance?
  - Is there enough feedback on areas of strength and weakness?
  - Role of political influence?
- Will improving incentives lead to better performance?

*Spec = certification requirements for EDS. Other modalities may vary.



Background, Objective and Disclaimer

Background

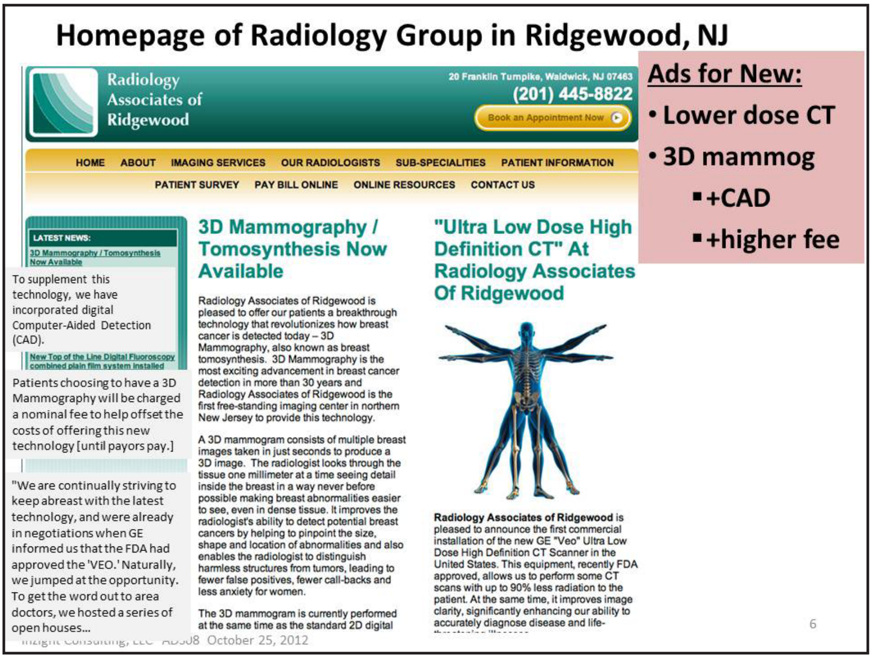
- We investigated 3rd party involvement and DICOS
  - By examining analogous issues in medical imaging and DICOM
- We presented findings on medical imaging at ADSA07
- Issue of incentives came up at that time, and subsequently
  - Better incentives → more third party involvement, better performance

Objective

- Gain further insight into issue of incentives from this audience
  - Description of how things are. Suggestions for improvement.

Disclaimer

- Funded by DHS, but I do not speak for or represent them
- Questions, findings, based on what I heard. Not policy. Not fact.



Homepage of Radiology Group in Ridgewood, NJ

To supplement this technology, we have incorporated digital Computer-Aided Detection (CAD).

Patients choosing to have a 3D Mammography will be charged a nominal fee to help offset the costs of offering this new technology [until payors pay.]

"We are continually striving to keep abreast with the latest technology, and were already in negotiations when GE informed us that the FDA had approved the 'VEO.' Naturally, we jumped at the opportunity. To get the word out to area doctors, we hosted a series of open houses...

3D Mammography / Tomosynthesis Now Available

Radiology Associates of Ridgewood is pleased to offer our patients a breakthrough technology that revolutionizes how breast cancer is detected today – 3D Mammography, also known as breast tomosynthesis. 3D Mammography is the most exciting advancement in breast cancer detection in more than 30 years and Radiology Associates of Ridgewood is the first free-standing imaging center in northern New Jersey to provide this technology.

A 3D mammogram consists of multiple breast images taken in just seconds to produce a 3D image. The radiologist looks through the tissue one millimeter at a time seeing detail inside the breast in a way never before possible making breast abnormalities easier to see, even in dense tissue. It improves the radiologist's ability to detect potential breast cancers by helping to pinpoint the size, shape and location of abnormalities and also enables the radiologist to distinguish harmless structures from tumors, leading to fewer false positives, fewer call-backs and less anxiety for women.

The 3D mammogram is currently performed at the same time as the standard 2D digital

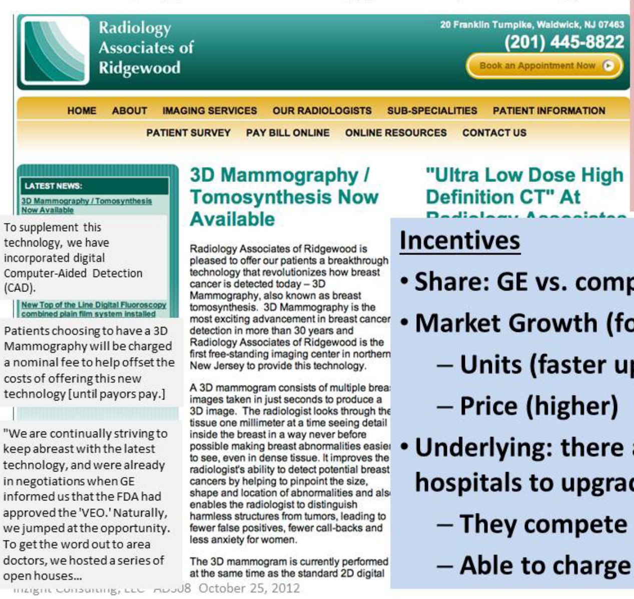
"Ultra Low Dose High Definition CT" At Radiology Associates Of Ridgewood

Radiology Associates of Ridgewood is pleased to announce the first commercial installation of the new GE "Veo" Ultra Low Dose High Definition CT Scanner in the United States. This equipment, recently FDA approved, allows us to perform some CT scans with up to 90% less radiation to the patient. At the same time, it improves image clarity, significantly enhancing our ability to accurately diagnose disease and life-

Ads for New:

- Lower dose CT
- 3D mammog
  - +CAD
  - +higher fee

Homepage of Radiology Group in Ridgewood, NJ



Radiology Associates of Ridgewood

20 Franklin Turnpike, Waldwick, NJ 07463

(201) 445-8822

3D Mammography / Tomosynthesis Now Available

Radiology Associates of Ridgewood is pleased to offer our patients a breakthrough technology that revolutionizes how breast cancer is detected today – 3D Mammography, also known as breast tomosynthesis. 3D Mammography is the most exciting advancement in breast cancer detection in more than 30 years and Radiology Associates of Ridgewood is the first free-standing imaging center in northern New Jersey to provide this technology.

A 3D mammogram consists of multiple brea images taken in just seconds to produce a 3D image. The radiologist looks through the tissue one millimeter at a time seeing detail inside the breast in a way never before possible making breast abnormalities easie to see, even in dense tissue. It improves the radiologist's ability to detect potential breast cancers by helping to pinpoint the size, shape and location of abnormalities and als enables the radiologist to distinguish harmless structures from tumors, leading to fewer false positives, fewer call-backs and less anxiety for women.

The 3D mammogram is currently performed at the same time as the standard 2D digital

"Ultra Low Dose High Definition CT" At

To supplement this technology, we have incorporated digital Computer-Aided Detection (CAD).

Patients choosing to have a 3D Mammography will be charged a nominal fee to help offset the costs of offering this new technology [until payors pay.]

"We are continually striving to keep abreast with the latest technology, and were already in negotiations when GE informed us that the FDA had approved the 'VEO.' Naturally, we jumped at the opportunity. To get the word out to area doctors, we hosted a series of open houses...

Ads for New:

- Lower dose CT
- 3D mammog
  - +CAD
  - +higher fee

Incentives

- Share: GE vs. competitors
- Market Growth (for all vendors)
  - Units (faster upgrades)
  - Price (higher)
- Underlying: there are incentives for hospitals to upgrade
  - They compete
  - Able to charge more

Inzight Consulting, LLC ADS08 October 25, 2012

What incentives are in place for security vendors?

Open for discussion (and suggestions)

- To meet spec?
- To exceed spec?
- To improve once certified?
- To increase performance, if it adds cost?
- To jump ahead of competitors? (Adding unique capabilities)

Inzight Consulting, LLC ADS08 October 25, 2012

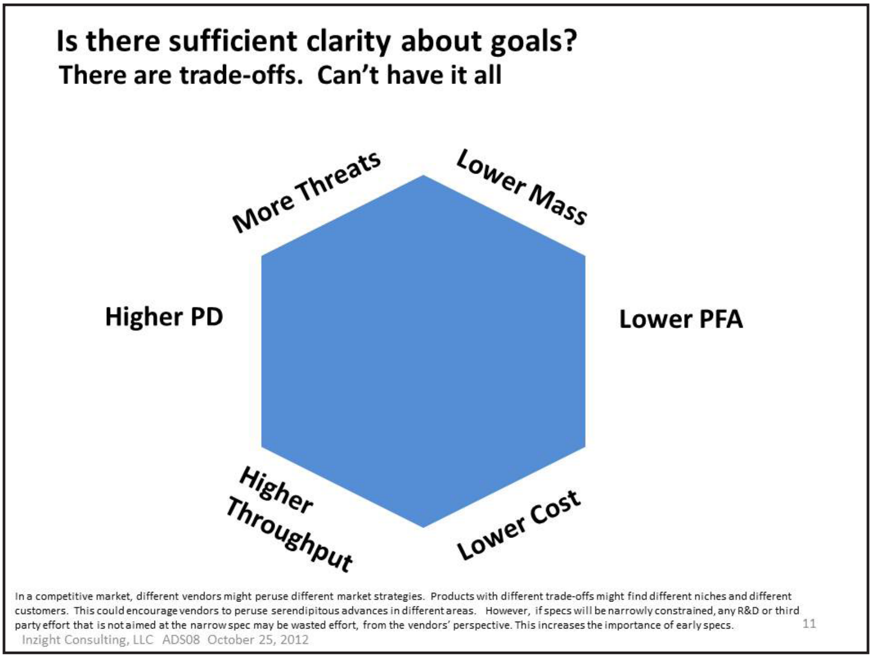
*Spec = certification requirements for EDS. Other modalities may vary.

8

Options to Increase Incentives: To Be Discussed

Plausible? Desirable? Barriers to change?

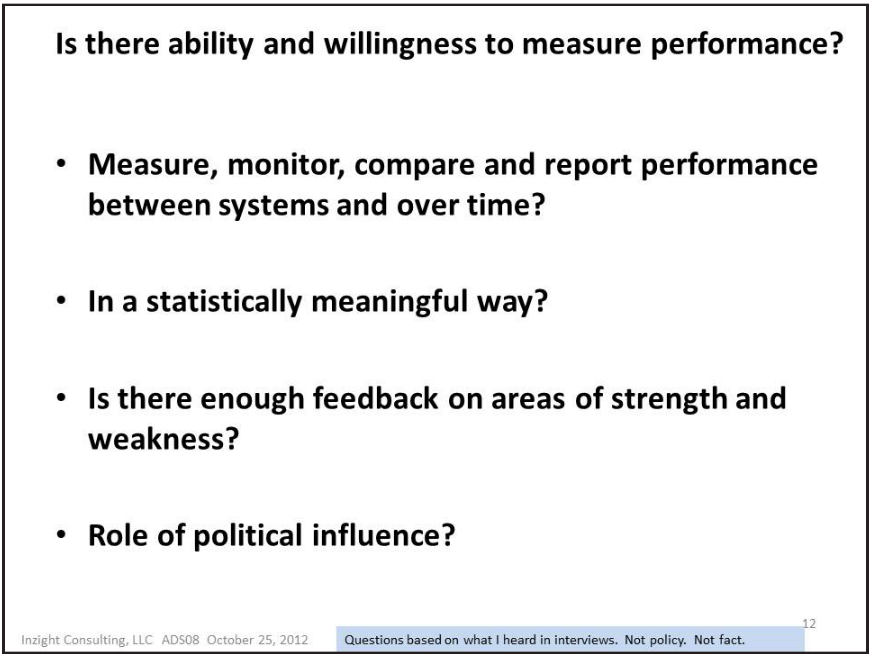
- Pay more for better performance (after defining goals)
  - Higher price or market share. (Consider using "non-minimum shalls" ?)
- Increase focus on TCO (total cost of ownership)
- Create market for upgrades
  - That improve performance or decrease TCO
  - Monitor PFA in the field. Reward improvement in PFA, at constant PD.
- Reduce requirement for multiple-source
  - Increase the incentive for vendors to go beyond competitors' capabilities
- Make contingent offers to buy
- Other?



Is there sufficient clarity about goals?

There are trade-offs. Can't have it all

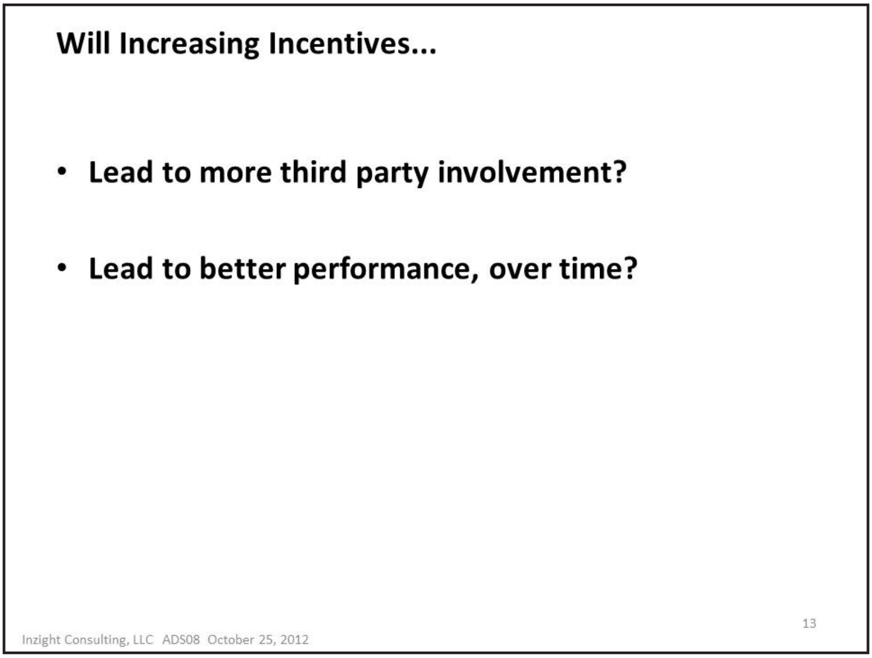
In a competitive market, different vendors might peruse different market strategies. Products with different trade-offs might find different niches and different customers. This could encourage vendors to peruse serendipitous advances in different areas. However, if specs will be narrowly constrained, any R&D or third party effort that is not aimed at the narrow spec may be wasted effort, from the vendors' perspective. This increases the importance of early specs.



Is there ability and willingness to measure performance?

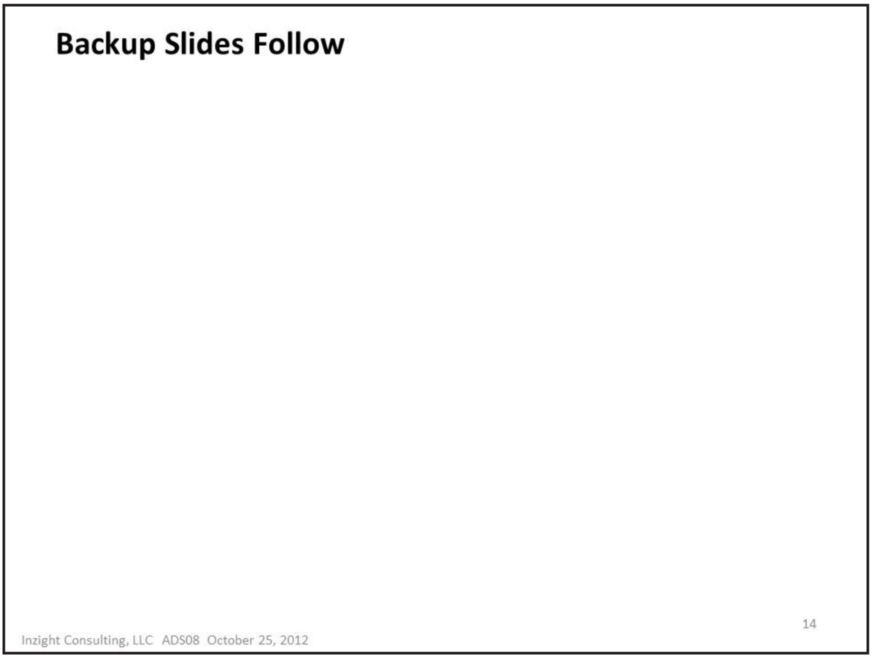
- Measure, monitor, compare and report performance between systems and over time?
- In a statistically meaningful way?
- Is there enough feedback on areas of strength and weakness?
- Role of political influence?

Questions based on what I heard in interviews. Not policy. Not fact.

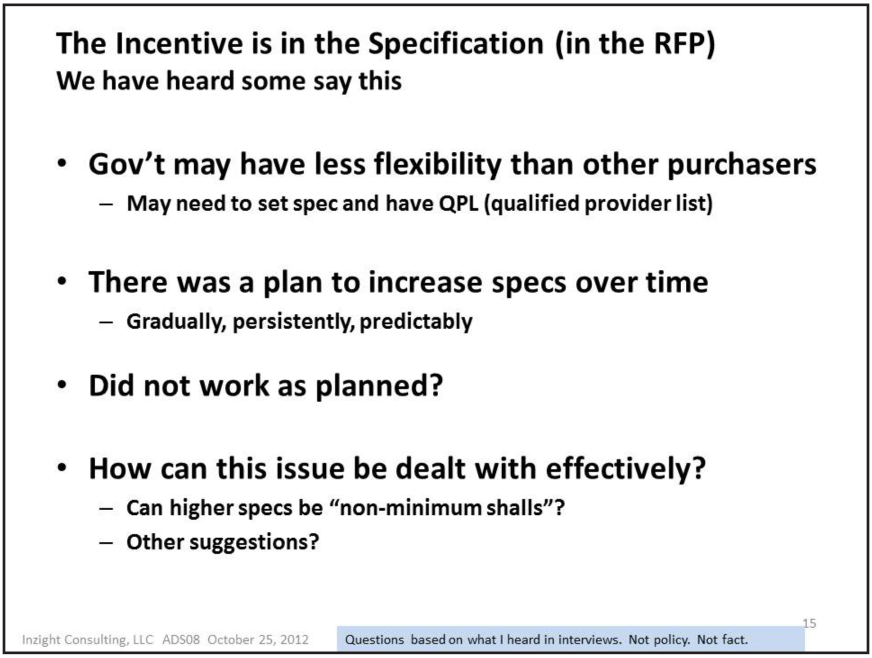


Will Increasing Incentives...

- Lead to more third party involvement?
- Lead to better performance, over time?



Backup Slides Follow



The Incentive is in the Specification (in the RFP)

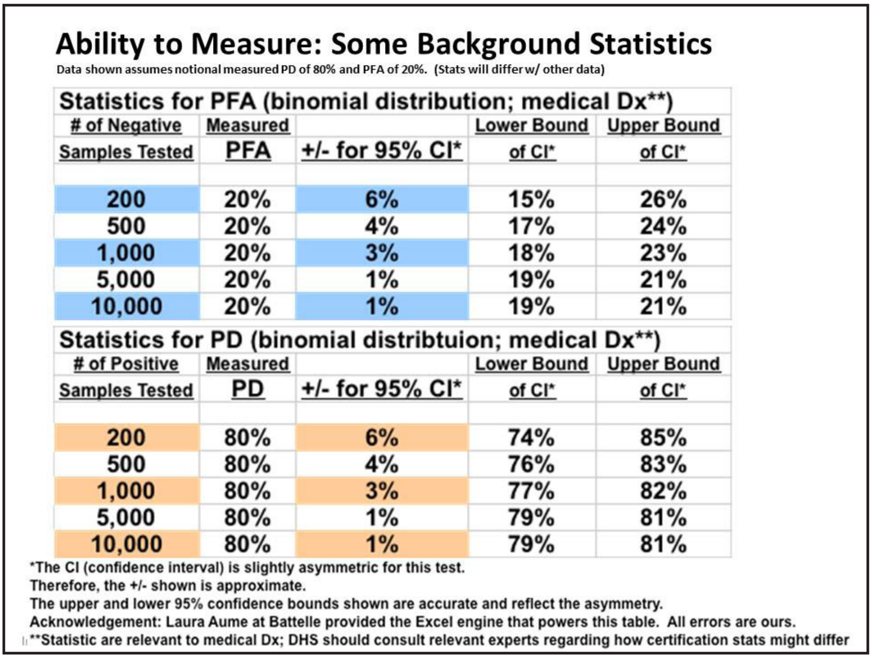
We have heard some say this

- Gov't may have less flexibility than other purchasers
  - May need to set spec and have QPL (qualified provider list)
- There was a plan to increase specs over time
  - Gradually, persistently, predictably
- Did not work as planned?
- How can this issue be dealt with effectively?
  - Can higher specs be "non-minimum shalls"?
  - Other suggestions?

Inzight Consulting, LLC ADS08 October 25, 2012

Questions based on what I heard in interviews. Not policy. Not fact.

15



Ability to Measure: Some Background Statistics

Data shown assumes notional measured PD of 80% and PFA of 20%. (Stats will differ w/ other data)

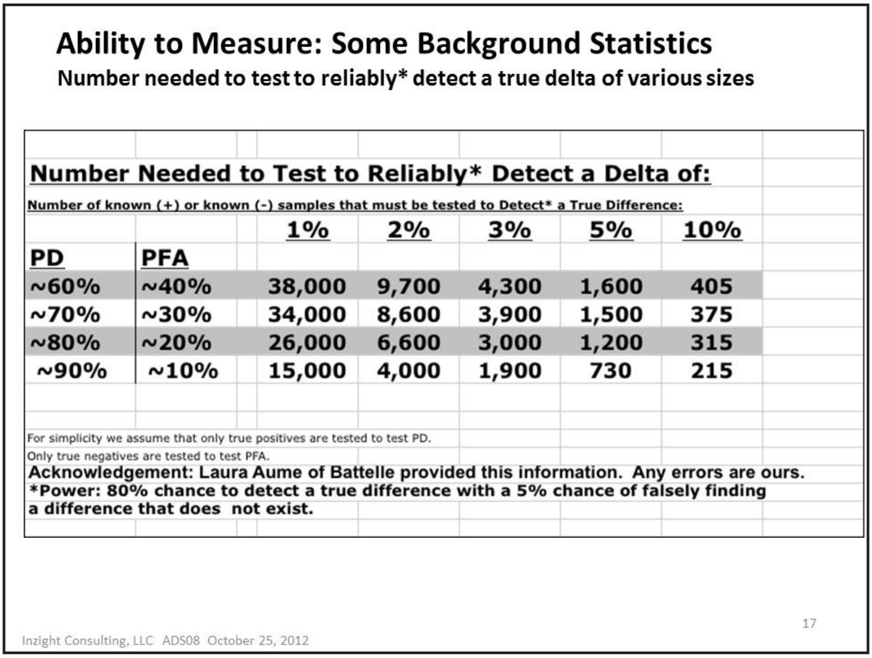
Statistics for PFA (binomial distribution; medical Dx)**

# of Negative Samples Tested	Measured PFA	+/- for 95% CI*	Lower Bound of CI*	Upper Bound of CI*
200	20%	6%	15%	26%
500	20%	4%	17%	24%
1,000	20%	3%	18%	23%
5,000	20%	1%	19%	21%
10,000	20%	1%	19%	21%

Statistics for PD (binomial distribtuion; medical Dx)**

# of Positive Samples Tested	Measured PD	+/- for 95% CI*	Lower Bound of CI*	Upper Bound of CI*
200	80%	6%	74%	85%
500	80%	4%	76%	83%
1,000	80%	3%	77%	82%
5,000	80%	1%	79%	81%
10,000	80%	1%	79%	81%

*The CI (confidence interval) is slightly asymmetric for this test.
Therefore, the +/- shown is approximate.
The upper and lower 95% confidence bounds shown are accurate and reflect the asymmetry.
Acknowledgement: Laura Aume at Battelle provided the Excel engine that powers this table. All errors are ours.
**Statistic are relevant to medical Dx; DHS should consult relevant experts regarding how certification stats might differ



Ability to Measure: Some Background Statistics

Number needed to test to reliably* detect a true delta of various sizes

Number Needed to Test to Reliably* Detect a Delta of:

Number of known (+) or known (-) samples that must be tested to Detect* a True Difference:

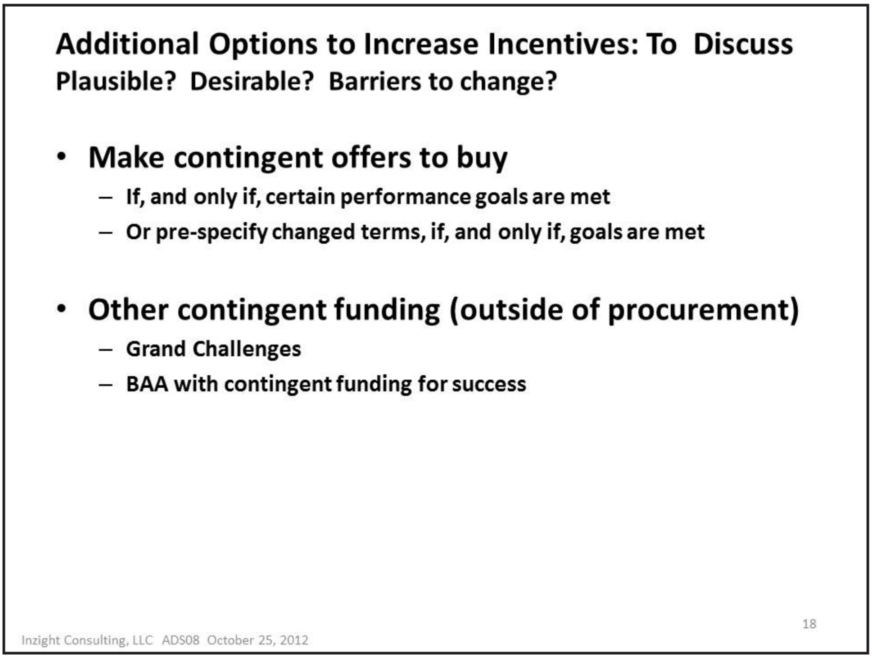
PD	PFA	1%	2%	3%	5%	10%
~60%	~40%	38,000	9,700	4,300	1,600	405
~70%	~30%	34,000	8,600	3,900	1,500	375
~80%	~20%	26,000	6,600	3,000	1,200	315
~90%	~10%	15,000	4,000	1,900	730	215

For simplicity we assume that only true positives are tested to test PD.

Only true negatives are tested to test PFA.

Acknowledgement: Laura Aume of Battelle provided this information. Any errors are ours.

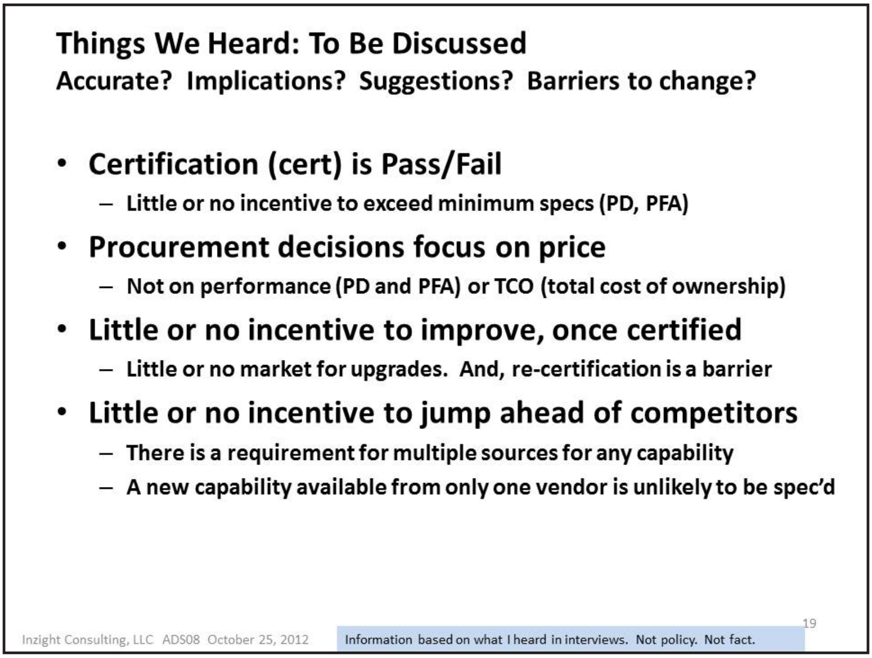
***Power: 80% chance to detect a true difference with a 5% chance of falsely finding a difference that does not exist.**



Additional Options to Increase Incentives: To Discuss

Plausible? Desirable? Barriers to change?

- **Make contingent offers to buy**
  - If, and only if, certain performance goals are met
  - Or pre-specify changed terms, if, and only if, goals are met
- **Other contingent funding (outside of procurement)**
  - Grand Challenges
  - BAA with contingent funding for success



Things We Heard: To Be Discussed

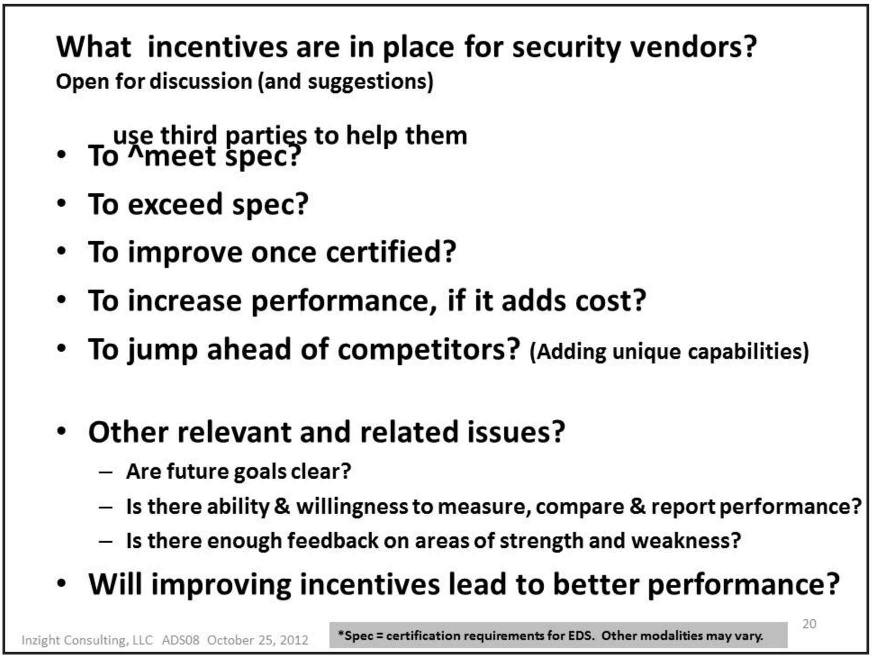
Accurate? Implications? Suggestions? Barriers to change?

- Certification (cert) is Pass/Fail
  - Little or no incentive to exceed minimum specs (PD, PFA)
- Procurement decisions focus on price
  - Not on performance (PD and PFA) or TCO (total cost of ownership)
- Little or no incentive to improve, once certified
  - Little or no market for upgrades. And, re-certification is a barrier
- Little or no incentive to jump ahead of competitors
  - There is a requirement for multiple sources for any capability
  - A new capability available from only one vendor is unlikely to be spec'd

Inzight Consulting, LLC ADS08 October 25, 2012

Information based on what I heard in interviews. Not policy. Not fact.

19



What incentives are in place for security vendors?

Open for discussion (and suggestions)

use third parties to help them
- To ^meet spec?
- To exceed spec?
- To improve once certified?
- To increase performance, if it adds cost?
- To jump ahead of competitors? (Adding unique capabilities)

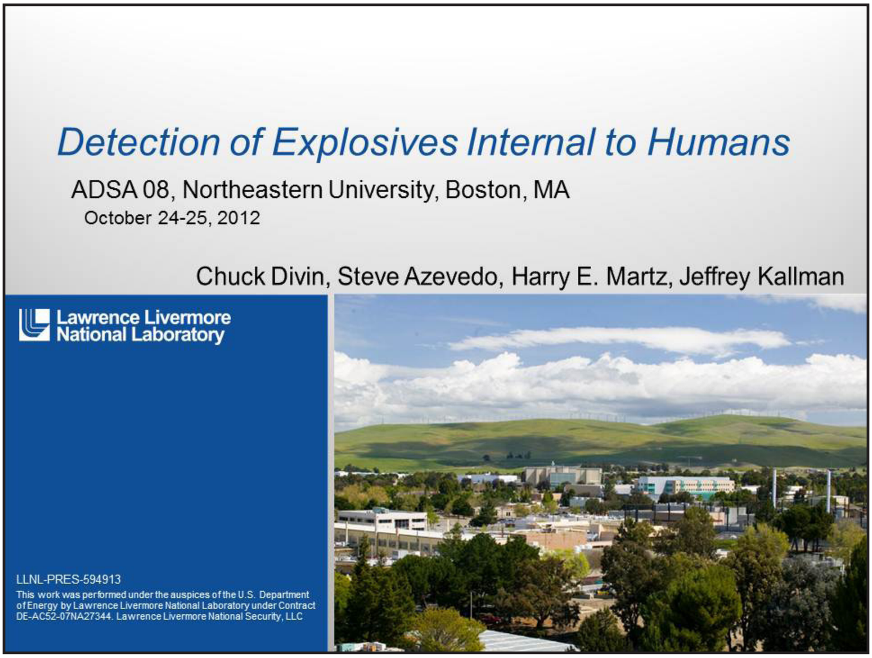
- Other relevant and related issues?
  - Are future goals clear?
  - Is there ability & willingness to measure, compare & report performance?
  - Is there enough feedback on areas of strength and weakness?
- Will improving incentives lead to better performance?

Inzight Consulting, LLC ADS08 October 25, 2012

*Spec = certification requirements for EDS. Other modalities may vary.

20

16.22 Steve Azevedo: Detection of Implanted Explosives



Detection of Explosives Internal to Humans

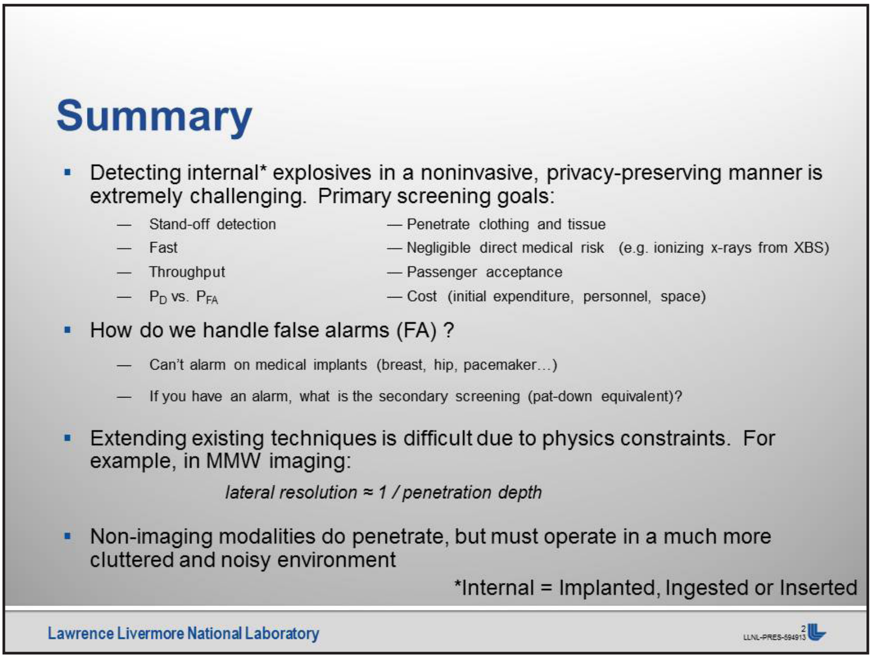
ADSA 08, Northeastern University, Boston, MA
October 24-25, 2012

Chuck Divin, Steve Azevedo, Harry E. Martz, Jeffrey Kallman

Lawrence Livermore National Laboratory

LLNL-PRES-594913

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC



Summary

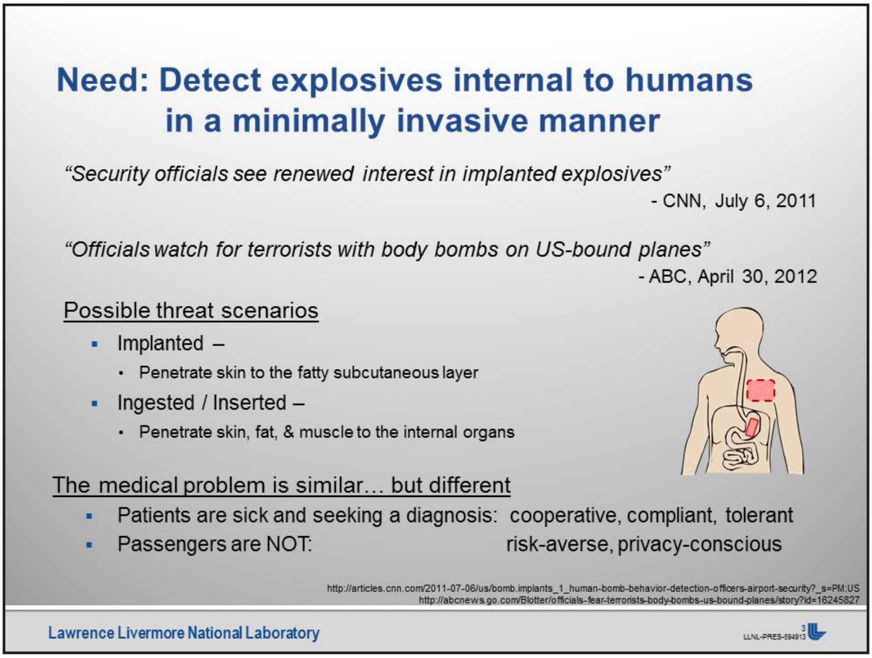
- Detecting internal* explosives in a noninvasive, privacy-preserving manner is extremely challenging. Primary screening goals:
  - — Stand-off detection
  - — Fast
  - — Throughput
  - — P_D vs. P_{FA}
  - — Penetrate clothing and tissue
  - — Negligible direct medical risk (e.g. ionizing x-rays from XBS)
  - — Passenger acceptance
  - — Cost (initial expenditure, personnel, space)
- How do we handle false alarms (FA) ?
  - — Can't alarm on medical implants (breast, hip, pacemaker…)
  - — If you have an alarm, what is the secondary screening (pat-down equivalent)?
- Extending existing techniques is difficult due to physics constraints. For example, in MMW imaging:

  lateral resolution ≈ 1 / penetration depth
- Non-imaging modalities do penetrate, but must operate in a much more cluttered and noisy environment

*Internal = Implanted, Ingested or Inserted

Lawrence Livermore National Laboratory

LLNL-PRES-594913 2



Need: Detect explosives internal to humans in a minimally invasive manner

"Security officials see renewed interest in implanted explosives"
- CNN, July 6, 2011

"Officials watch for terrorists with body bombs on US-bound planes"
- ABC, April 30, 2012

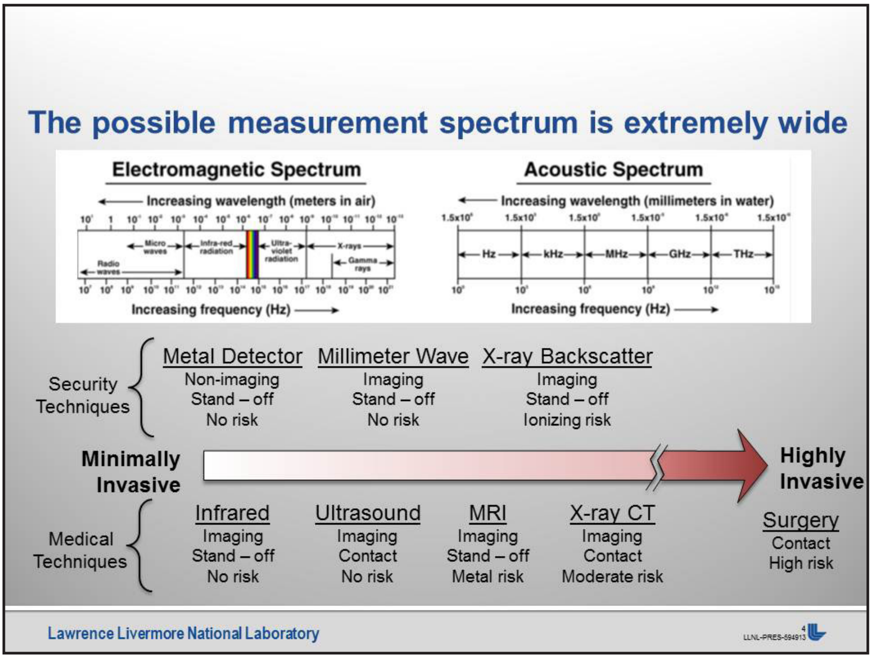
Possible threat scenarios

- Implanted –
  - Penetrate skin to the fatty subcutaneous layer
- Ingested / Inserted –
  - Penetrate skin, fat, & muscle to the internal organs

The medical problem is similar… but different

- Patients are sick and seeking a diagnosis: cooperative, compliant, tolerant
- Passengers are NOT: risk-averse, privacy-conscious

http://articles.cnn.com/2011-07-06/us/bomb.implants_1_human-bomb-behavior-detection-officers-airport-security?_s=PM:US
http://abcnews.go.com/Blotter/officials-fear-terrorists-body-bombs-us-bound-planes/story?id=16245827

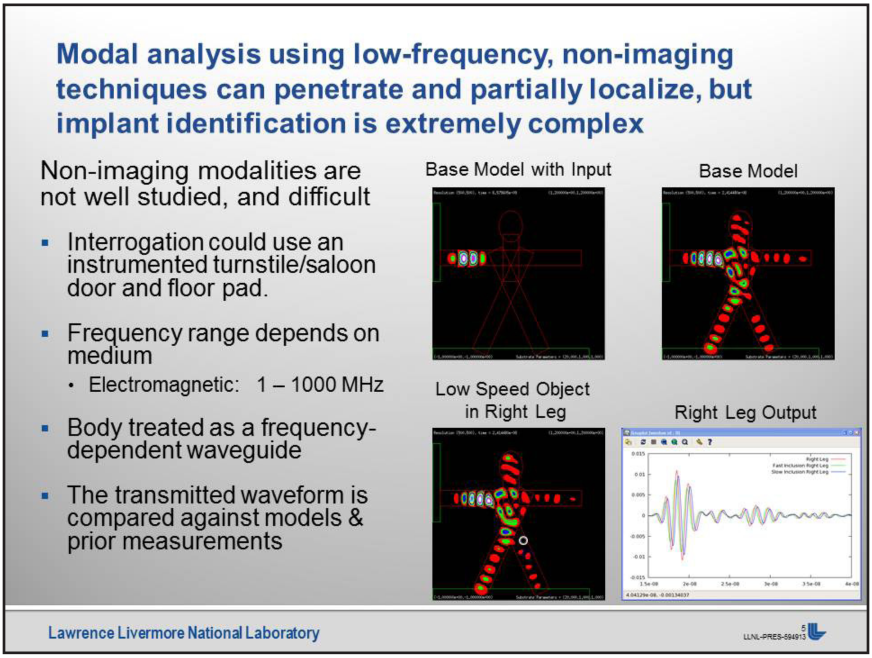


The possible measurement spectrum is extremely wide

	Metal Detector	Millimeter Wave	X-ray Backscatter
Security Techniques	Non-imaging Stand – off No risk	Imaging Stand – off No risk	Imaging Stand – off Ionizing risk

Minimally Invasive → **Highly Invasive**

	Infrared	Ultrasound	MRI	X-ray CT	Surgery
Medical Techniques	Imaging Stand – off No risk	Imaging Contact No risk	Imaging Stand – off Metal risk	Imaging Contact Moderate risk	Contact High risk



Modal analysis using low-frequency, non-imaging techniques can penetrate and partially localize, but implant identification is extremely complex

Non-imaging modalities are not well studied, and difficult

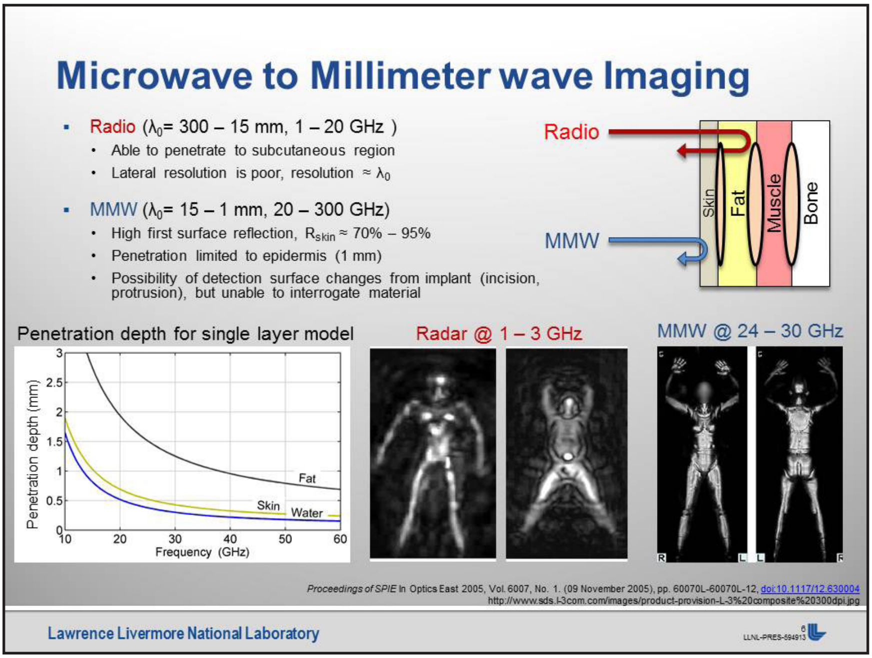
- Interrogation could use an instrumented turnstile/saloon door and floor pad.
- Frequency range depends on medium
  - Electromagnetic: 1 – 1000 MHz
- Body treated as a frequency-dependent waveguide
- The transmitted waveform is compared against models & prior measurements

Base Model with Input

Base Model

Low Speed Object in Right Leg

Right Leg Output



Microwave to Millimeter wave Imaging

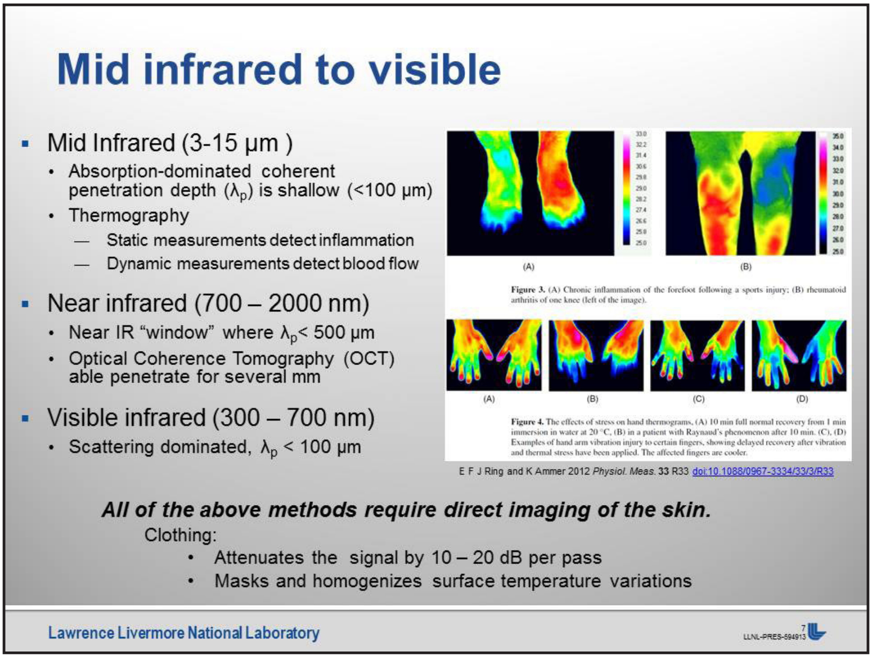
- Radio (λ_0= 300 – 15 mm, 1 – 20 GHz)
  - Able to penetrate to subcutaneous region
  - Lateral resolution is poor, resolution ≈ λ_0
- MMW (λ_0= 15 – 1 mm, 20 – 300 GHz)
  - High first surface reflection, R_{skin} ≈ 70% – 95%
  - Penetration limited to epidermis (1 mm)
  - Possibility of detection surface changes from implant (incision, protrusion), but unable to interrogate material

Penetration depth for single layer model

Radar @ 1 – 3 GHz

MMW @ 24 – 30 GHz

Proceedings of SPIE In Optics East 2005, Vol. 6007, No. 1. (09 November 2005), pp. 60070L-60070L-12, doi:10.1117/12.630004
http://www.sds.l-3com.com/images/product-provision-L-3%20composite%20300dpi.jpg



Mid infrared to visible

- Mid Infrared (3-15 µm)
  - Absorption-dominated coherent penetration depth (λ_p) is shallow (<100 µm)
  - Thermography
    - — Static measurements detect inflammation
    - — Dynamic measurements detect blood flow
- Near infrared (700 – 2000 nm)
  - Near IR "window" where λ_p< 500 µm
  - Optical Coherence Tomography (OCT) able penetrate for several mm
- Visible infrared (300 – 700 nm)
  - Scattering dominated, λ_p < 100 µm

Figure 3. (A) Chronic inflammation of the forefoot following a sports injury; (B) rheumatoid arthritis of one knee (left of the image).

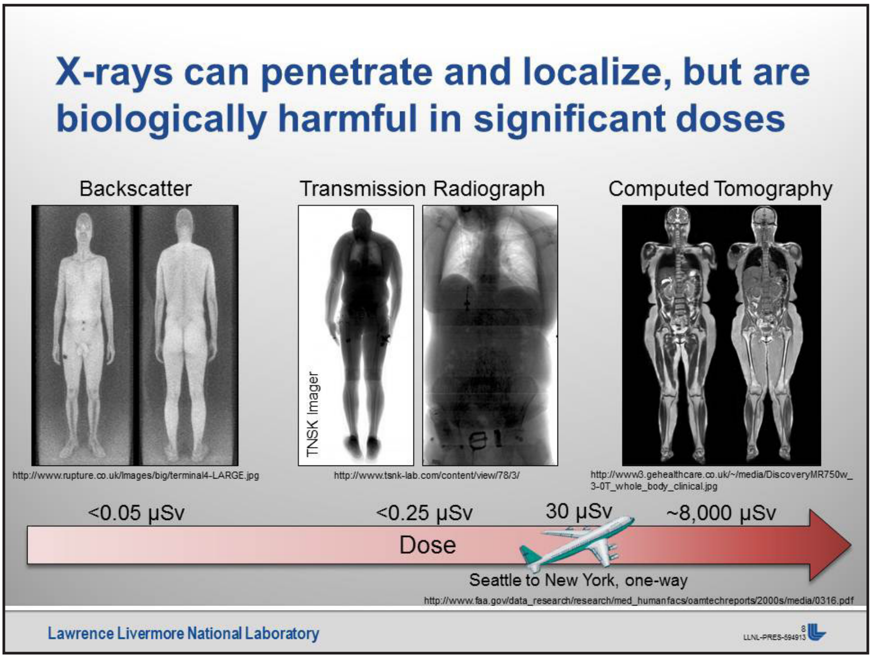
Figure 4. The effects of stress on hand thermograms. (A) 10 min full normal recovery from 1 min immersion in water at 20 °C, (B) in a patient with Raynaud's phenomenon after 10 min. (C), (D) Examples of hand arm vibration injury to certain fingers, showing delayed recovery after vibration and thermal stress have been applied. The affected fingers are cooler.

E F J Ring and K Ammer 2012 *Physiol. Meas.* 33 R33 doi:10.1088/0967-3334/33/3/R33

All of the above methods require direct imaging of the skin.

Clothing:

- Attenuates the signal by 10 – 20 dB per pass
- Masks and homogenizes surface temperature variations



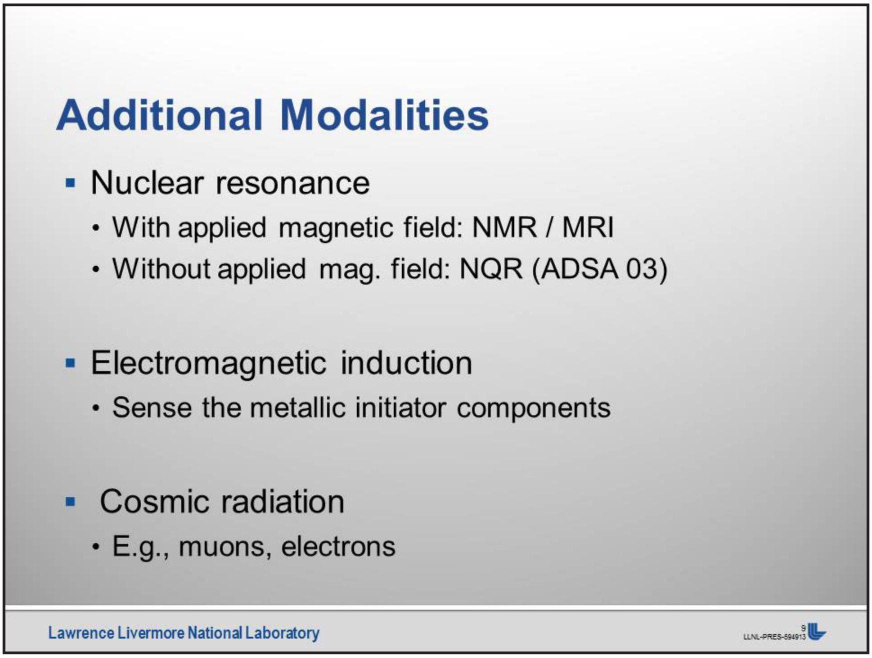
X-rays can penetrate and localize, but are biologically harmful in significant doses

http://www.rupture.co.uk/Images/big/terminal4-LARGE.jpg

http://www.tsnk-lab.com/content/view/78/3/

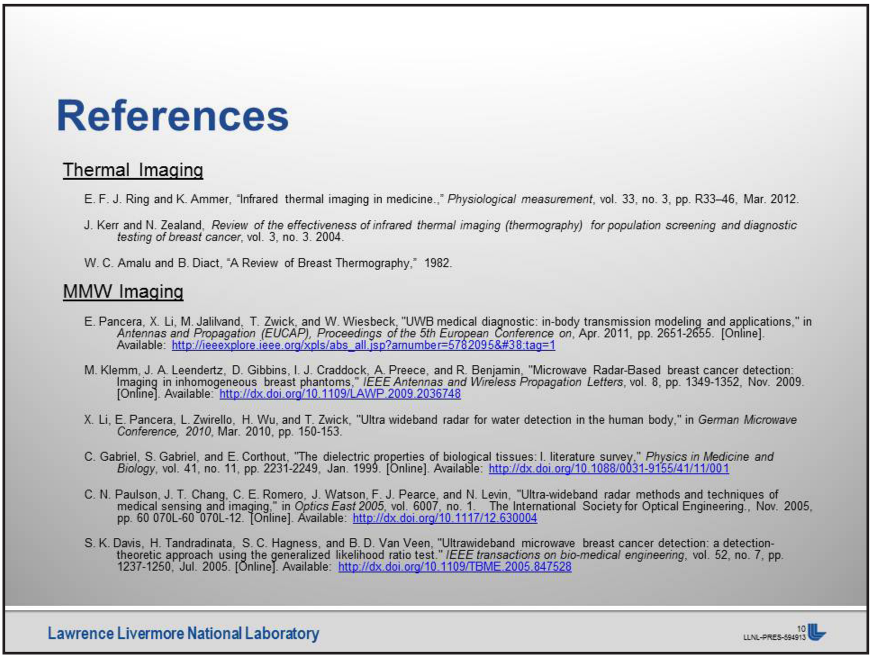
http://www3.gehealthcare.co.uk/~/media/DiscoveryMR750w_3-0T_whole_body_clinical.jpg

http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0316.pdf



Additional Modalities

- Nuclear resonance
  - With applied magnetic field: NMR / MRI
  - Without applied mag. field: NQR (ADSA 03)
- Electromagnetic induction
  - Sense the metallic initiator components
- Cosmic radiation
  - E.g., muons, electrons



References

Thermal Imaging

E. F. J. Ring and K. Ammer, "Infrared thermal imaging in medicine.," *Physiological measurement*, vol. 33, no. 3, pp. R33–46, Mar. 2012.

J. Kerr and N. Zealand, *Review of the effectiveness of infrared thermal imaging (thermography) for population screening and diagnostic testing of breast cancer*, vol. 3, no. 3. 2004.

W. C. Amalu and B. Diact, "A Review of Breast Thermography," 1982.

MMW Imaging

E. Pancera, X. Li, M. Jalilvand, T. Zwick, and W. Wiesbeck, "UWB medical diagnostic: in-body transmission modeling and applications," in *Antennas and Propagation (EUCAP), Proceedings of the 5th European Conference on*, Apr. 2011, pp. 2651-2655. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5782095&tag=1

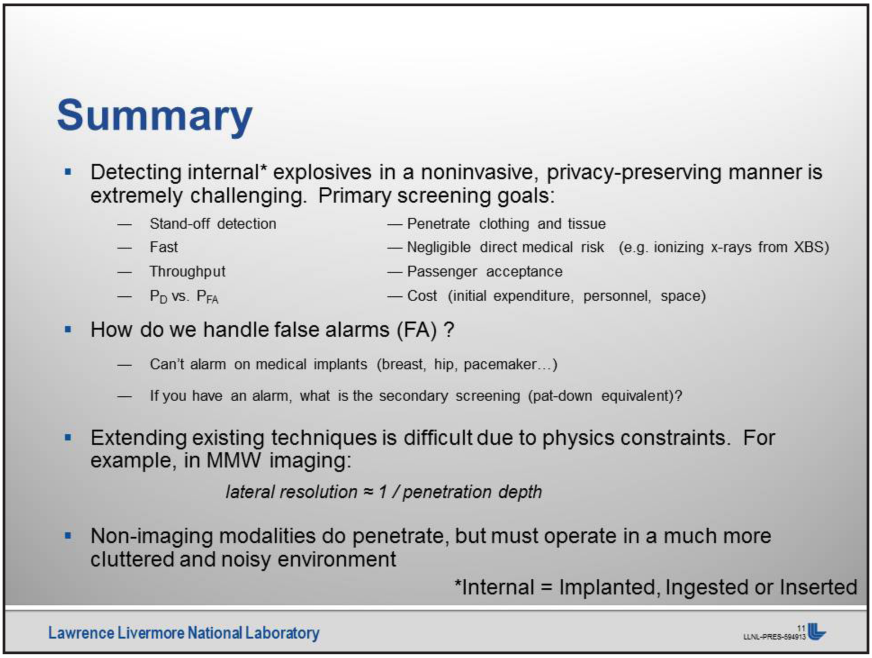
M. Klemm, J. A. Leendertz, D. Gibbins, I. J. Craddock, A. Preece, and R. Benjamin, "Microwave Radar-Based breast cancer detection: Imaging in inhomogeneous breast phantoms," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 1349-1352, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1109/LAWP.2009.2036748

X. Li, E. Pancera, L. Zwirello, H. Wu, and T. Zwick, "Ultra wideband radar for water detection in the human body," in *German Microwave Conference, 2010*, Mar. 2010, pp. 150-153.

C. Gabriel, S. Gabriel, and E. Corthout, "The dielectric properties of biological tissues: I. literature survey," *Physics in Medicine and Biology*, vol. 41, no. 11, pp. 2231-2249, Jan. 1999. [Online]. Available: http://dx.doi.org/10.1088/0031-9155/41/11/001

C. N. Paulson, J. T. Chang, C. E. Romero, J. Watson, F. J. Pearce, and N. Levin, "Ultra-wideband radar methods and techniques of medical sensing and imaging," in *Optics East 2005*, vol. 6007, no. 1. The International Society for Optical Engineering., Nov. 2005, pp. 60 070L-60 070L-12. [Online]. Available: http://dx.doi.org/10.1117/12.630004

S. K. Davis, H. Tandradinata, S. C. Hagness, and B. D. Van Veen, "Ultrawideband microwave breast cancer detection: a detection-theoretic approach using the generalized likelihood ratio test." *IEEE transactions on bio-medical engineering*, vol. 52, no. 7, pp. 1237-1250, Jul. 2005. [Online]. Available: http://dx.doi.org/10.1109/TBME.2005.847528



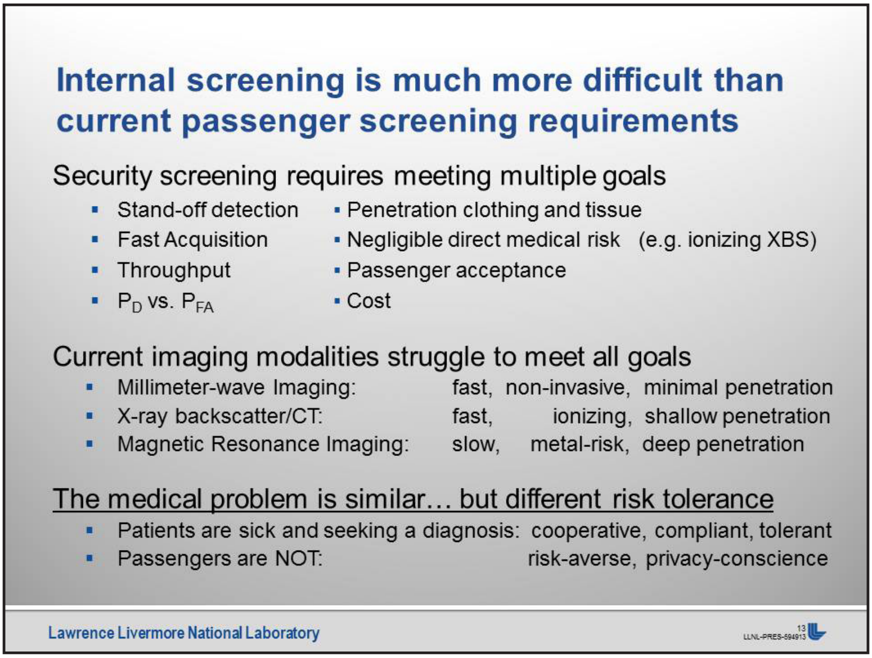
Summary

- Detecting internal* explosives in a noninvasive, privacy-preserving manner is extremely challenging. Primary screening goals:
  - Stand-off detection
  - Fast
  - Throughput
  - P_D vs. P_{FA}
  - Penetrate clothing and tissue
  - Negligible direct medical risk (e.g. ionizing x-rays from XBS)
  - Passenger acceptance
  - Cost (initial expenditure, personnel, space)
- How do we handle false alarms (FA) ?
  - Can't alarm on medical implants (breast, hip, pacemaker…)
  - If you have an alarm, what is the secondary screening (pat-down equivalent)?
- Extending existing techniques is difficult due to physics constraints. For example, in MMW imaging:

  lateral resolution ≈ 1 / penetration depth
- Non-imaging modalities do penetrate, but must operate in a much more cluttered and noisy environment

*Internal = Implanted, Ingested or Inserted

Lawrence Livermore
National Laboratory



Internal screening is much more difficult than current passenger screening requirements

Security screening requires meeting multiple goals

- Stand-off detection
- Fast Acquisition
- Throughput
- P_D vs. P_{FA}

- Penetration clothing and tissue
- Negligible direct medical risk (e.g. ionizing XBS)
- Passenger acceptance
- Cost

Current imaging modalities struggle to meet all goals

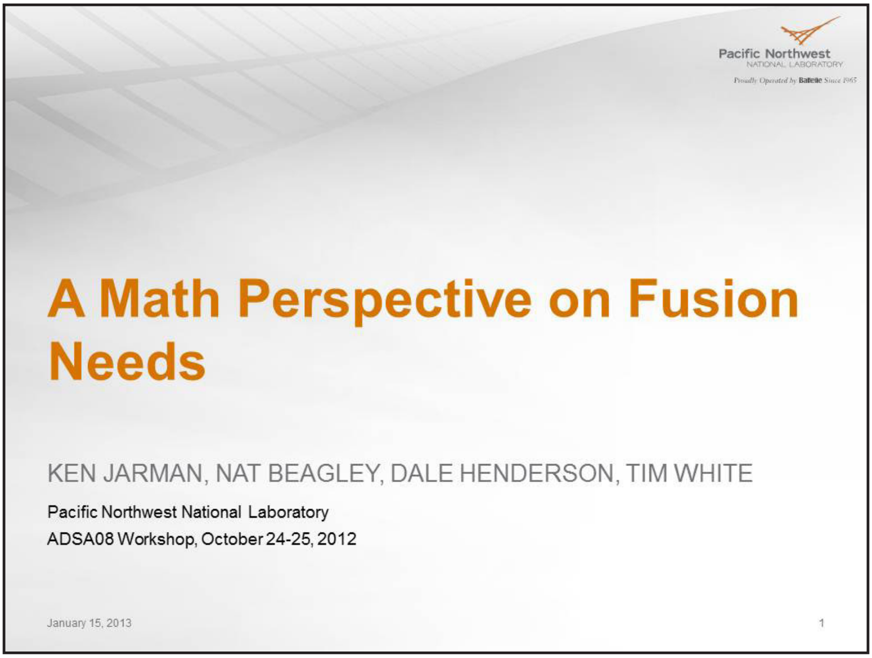
- Millimeter-wave Imaging: fast, non-invasive, minimal penetration
- X-ray backscatter/CT: fast, ionizing, shallow penetration
- Magnetic Resonance Imaging: slow, metal-risk, deep penetration

The medical problem is similar… but different risk tolerance

- Patients are sick and seeking a diagnosis: cooperative, compliant, tolerant
- Passengers are NOT: risk-averse, privacy-conscience

Lawrence Livermore National Laboratory

16.23 Ken Jarman: A Math Perspective on Fusion Needs

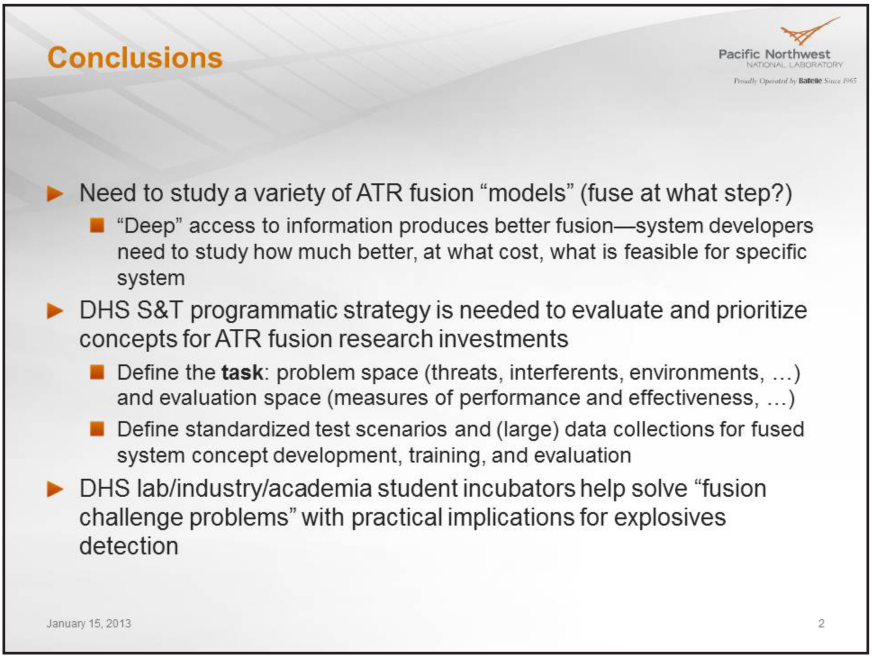


Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

A Math Perspective on Fusion Needs

KEN JARMAN, NAT BEAGLEY, DALE HENDERSON, TIM WHITE

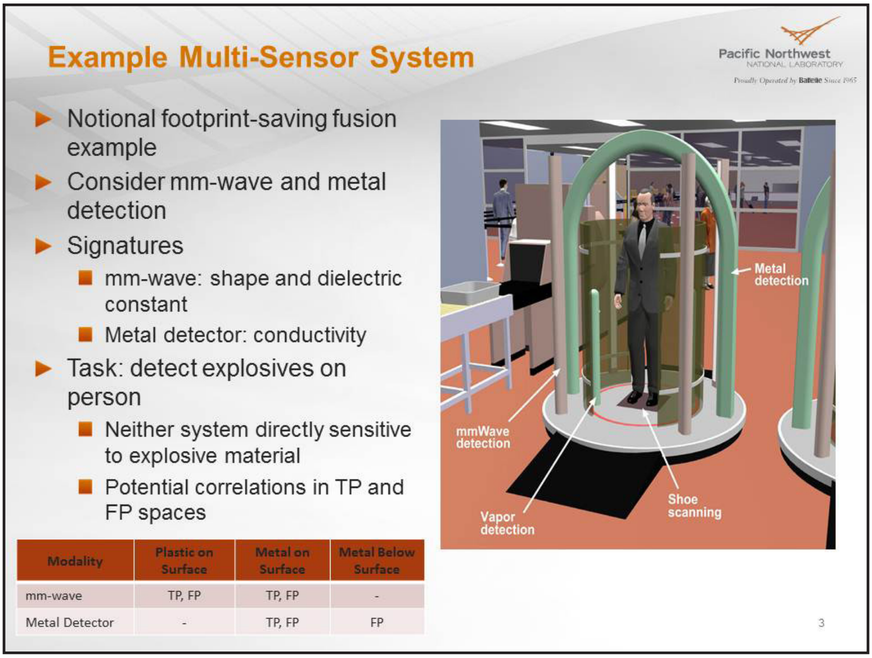
Pacific Northwest National Laboratory
ADSA08 Workshop, October 24-25, 2012



Conclusions

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Need to study a variety of ATR fusion "models" (fuse at what step?)
  - "Deep" access to information produces better fusion—system developers need to study how much better, at what cost, what is feasible for specific system
- DHS S&T programmatic strategy is needed to evaluate and prioritize concepts for ATR fusion research investments
  - Define the **task**: problem space (threats, interferents, environments, …) and evaluation space (measures of performance and effectiveness, …)
  - Define standardized test scenarios and (large) data collections for fused system concept development, training, and evaluation
- DHS lab/industry/academia student incubators help solve "fusion challenge problems" with practical implications for explosives detection

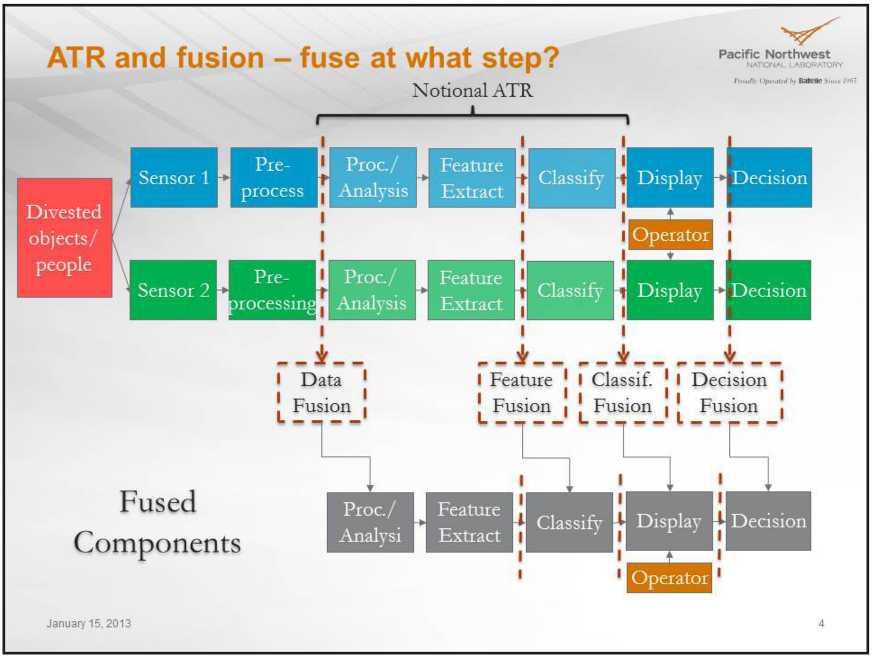


Example Multi-Sensor System

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Notional footprint-saving fusion example
- Consider mm-wave and metal detection
- Signatures
  - mm-wave: shape and dielectric constant
  - Metal detector: conductivity
- Task: detect explosives on person
  - Neither system directly sensitive to explosive material
  - Potential correlations in TP and FP spaces

Modality	Plastic on Surface	Metal on Surface	Metal Below Surface
mm-wave	TP, FP	TP, FP	-
Metal Detector	-	TP, FP	FP



ATR and fusion – fuse at what step?

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Notional ATR

Divested objects/ people

Sensor 1 → Pre- process → Proc./ Analysis → Feature Extract → Classify → Display → Decision

Operator

Sensor 2 → Pre- processing → Proc./ Analysis → Feature Extract → Classify → Display → Decision

Data Fusion

Feature Fusion

Classif. Fusion

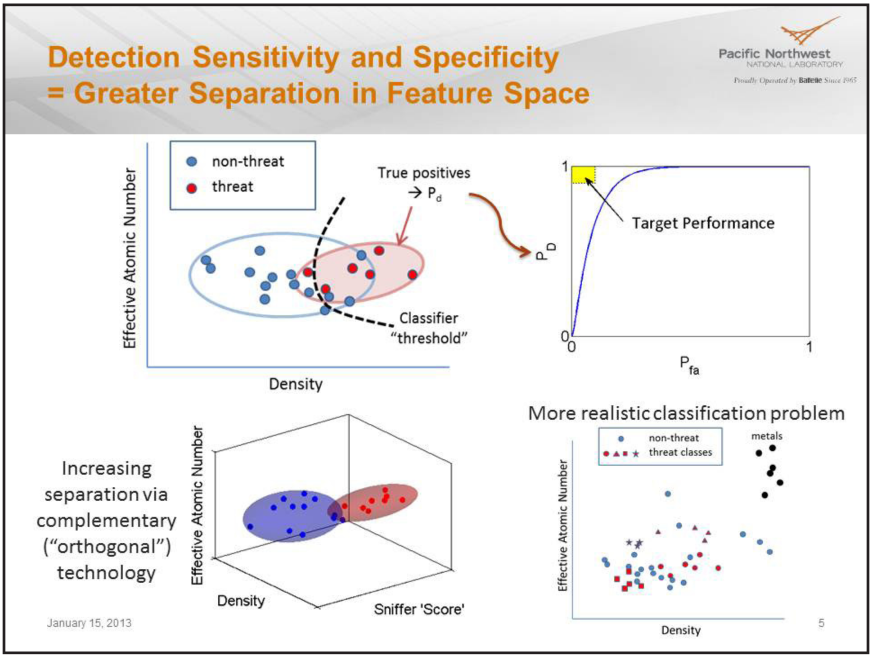
Decision Fusion

Fused Components

Proc./ Analysi → Feature Extract → Classify → Display → Decision

Operator

January 15, 2013



Detection Sensitivity and Specificity = Greater Separation in Feature Space

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

True positives → P_d

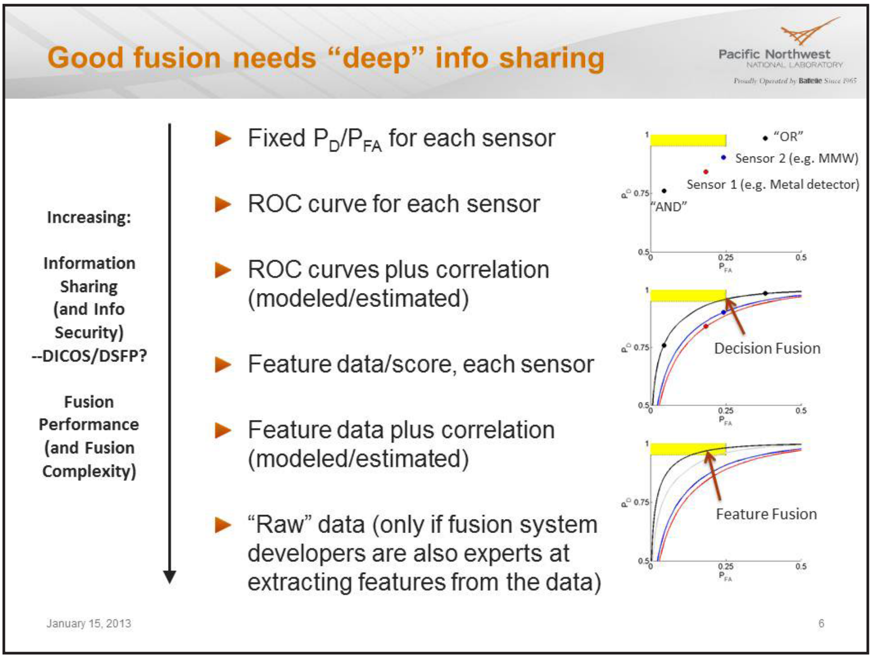
Classifier "threshold"

Target Performance

Increasing separation via complementary ("orthogonal") technology

More realistic classification problem

January 15, 2013



Good fusion needs "deep" info sharing

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Increasing:

Information Sharing (and Info Security) --DICOS/DSFP?

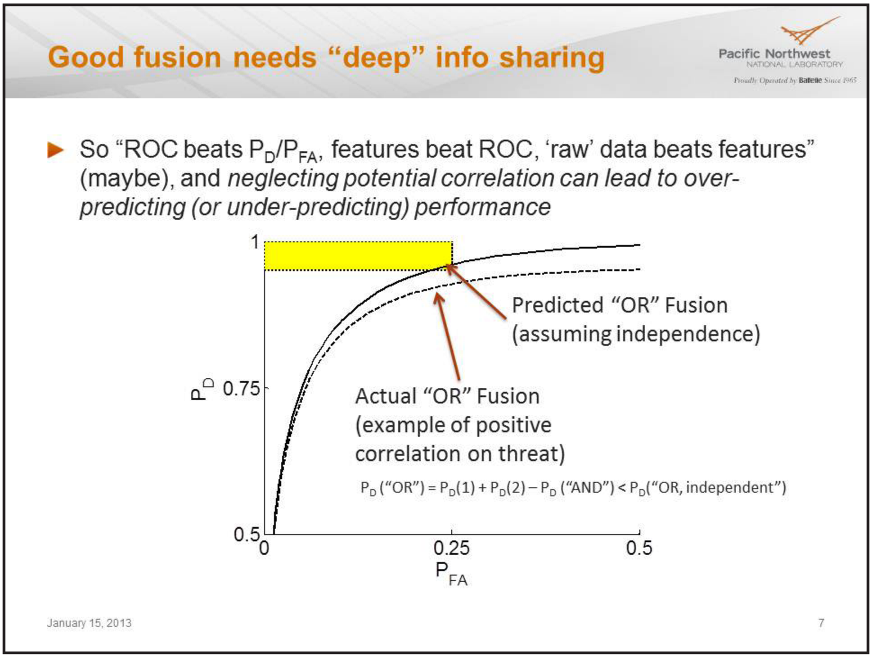
Fusion Performance (and Fusion Complexity)

- Fixed P_D/P_{FA} for each sensor
- ROC curve for each sensor
- ROC curves plus correlation (modeled/estimated)
- Feature data/score, each sensor
- Feature data plus correlation (modeled/estimated)
- "Raw" data (only if fusion system developers are also experts at extracting features from the data)

Decision Fusion

Feature Fusion

January 15, 2013



Good fusion needs "deep" info sharing

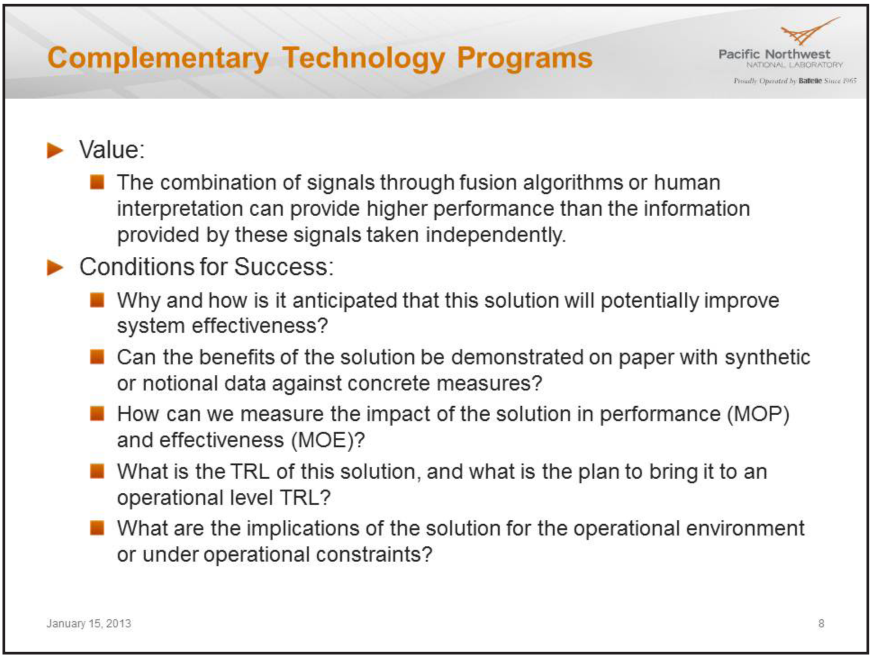
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- So "ROC beats P_D/P_{FA}, features beat ROC, 'raw' data beats features" (maybe), and *neglecting potential correlation can lead to over-predicting (or under-predicting) performance*

Predicted "OR" Fusion (assuming independence)

Actual "OR" Fusion (example of positive correlation on threat)

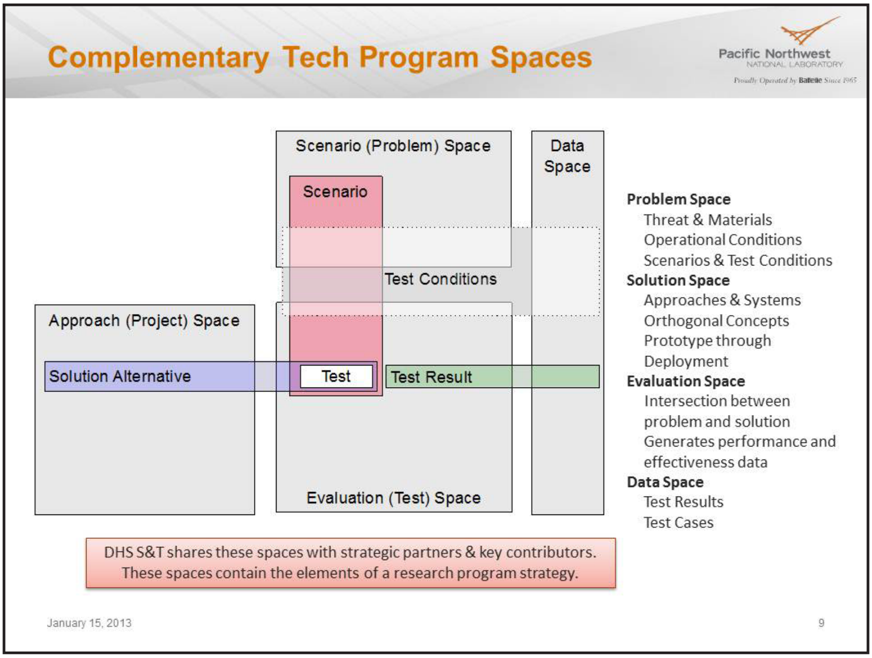
$P_D(\text{"OR"}) = P_D(1) + P_D(2) - P_D(\text{"AND"}) < P_D(\text{"OR, independent"})$



Complementary Technology Programs

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Value:
  - The combination of signals through fusion algorithms or human interpretation can provide higher performance than the information provided by these signals taken independently.
- Conditions for Success:
  - Why and how is it anticipated that this solution will potentially improve system effectiveness?
  - Can the benefits of the solution be demonstrated on paper with synthetic or notional data against concrete measures?
  - How can we measure the impact of the solution in performance (MOP) and effectiveness (MOE)?
  - What is the TRL of this solution, and what is the plan to bring it to an operational level TRL?
  - What are the implications of the solution for the operational environment or under operational constraints?



Complementary Tech Program Spaces

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Problem Space
- Threat & Materials
- Operational Conditions
- Scenarios & Test Conditions

Solution Space
- Approaches & Systems
- Orthogonal Concepts
- Prototype through Deployment

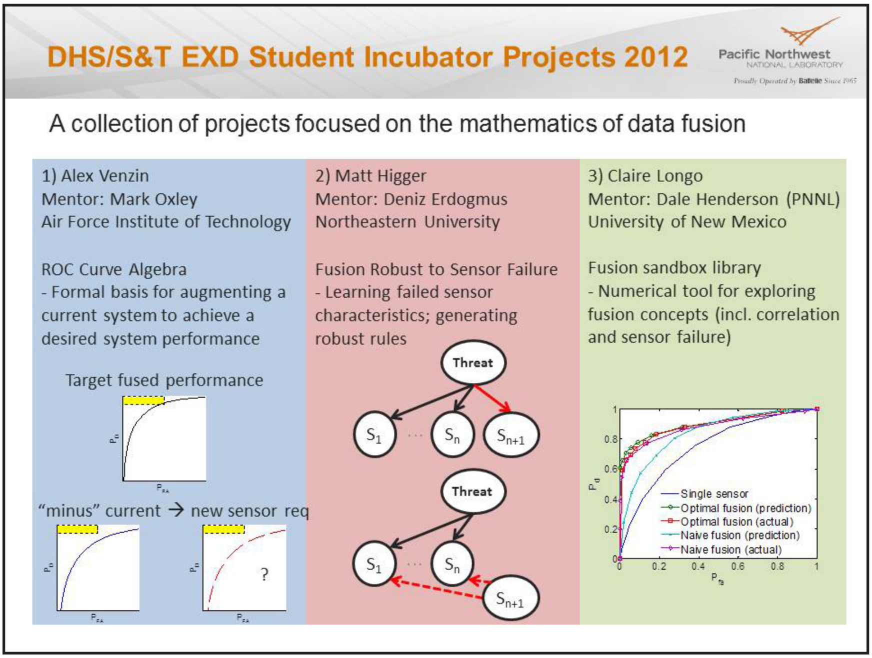
Evaluation Space
- Intersection between problem and solution
- Generates performance and effectiveness data

Data Space
- Test Results
- Test Cases

DHS S&T shares these spaces with strategic partners & key contributors.
These spaces contain the elements of a research program strategy.

January 15, 2013

9



DHS/S&T EXD Student Incubator Projects 2012

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

A collection of projects focused on the mathematics of data fusion

1) Alex Venzin
Mentor: Mark Oxley
Air Force Institute of Technology

ROC Curve Algebra
- Formal basis for augmenting a current system to achieve a desired system performance

Target fused performance

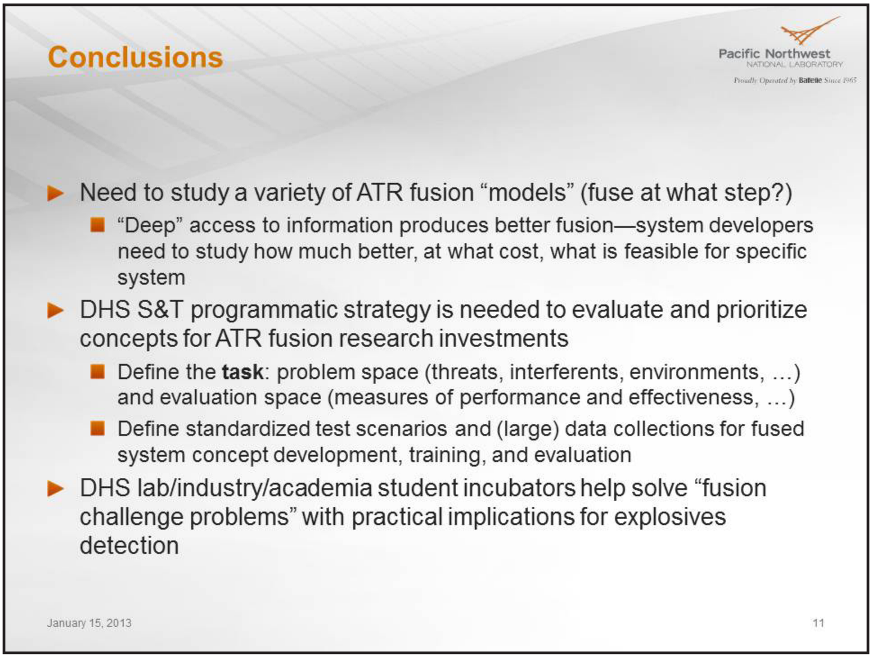
"minus" current → new sensor req

2) Matt Higger
Mentor: Deniz Erdogmus
Northeastern University

Fusion Robust to Sensor Failure
- Learning failed sensor characteristics; generating robust rules

3) Claire Longo
Mentor: Dale Henderson (PNNL)
University of New Mexico

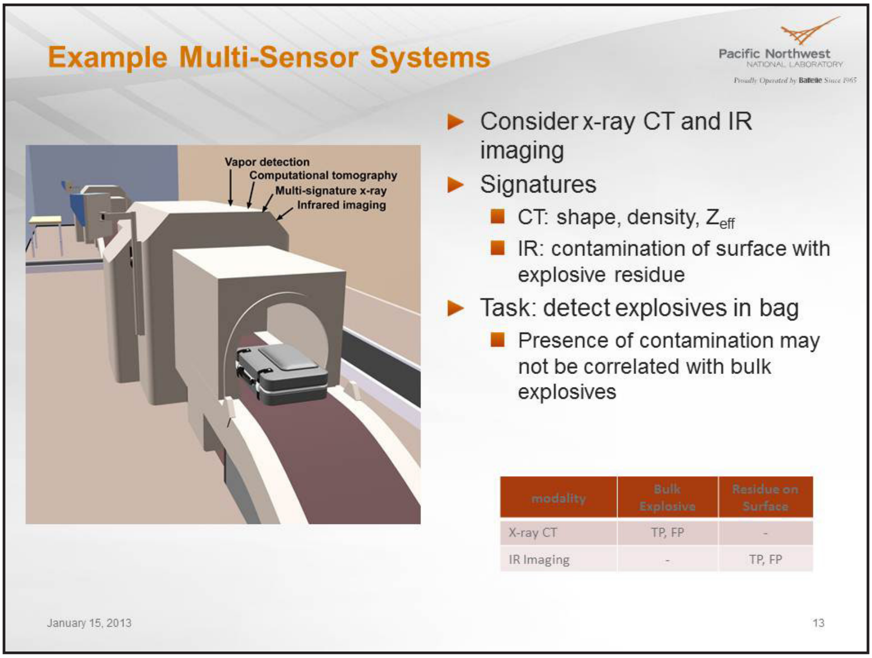
Fusion sandbox library
- Numerical tool for exploring fusion concepts (incl. correlation and sensor failure)



Conclusions

- Need to study a variety of ATR fusion “models” (fuse at what step?)
    - “Deep” access to information produces better fusion—system developers need to study how much better, at what cost, what is feasible for specific system
- DHS S&T programmatic strategy is needed to evaluate and prioritize concepts for ATR fusion research investments
    - Define the **task**: problem space (threats, interferents, environments, …) and evaluation space (measures of performance and effectiveness, …)
    - Define standardized test scenarios and (large) data collections for fused system concept development, training, and evaluation
- DHS lab/industry/academia student incubators help solve “fusion challenge problems” with practical implications for explosives detection

Additional Slides: OT Strategy

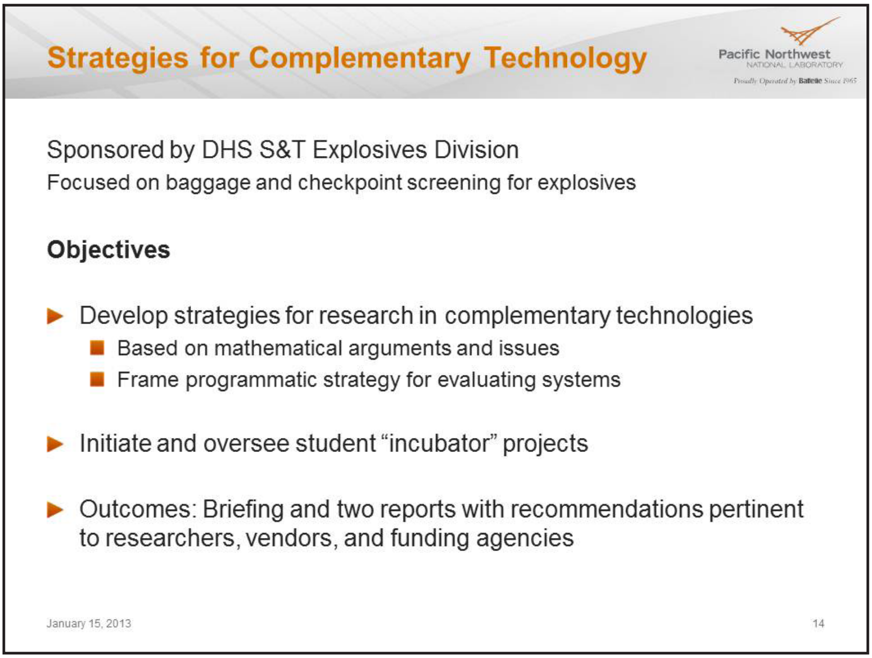


Example Multi-Sensor Systems

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Consider x-ray CT and IR imaging
- Signatures
  - CT: shape, density, Z_{eff}
  - IR: contamination of surface with explosive residue
- Task: detect explosives in bag
  - Presence of contamination may not be correlated with bulk explosives

modality	Bulk Explosive	Residue on Surface
X-ray CT	TP, FP	-
IR Imaging	-	TP, FP



Strategies for Complementary Technology

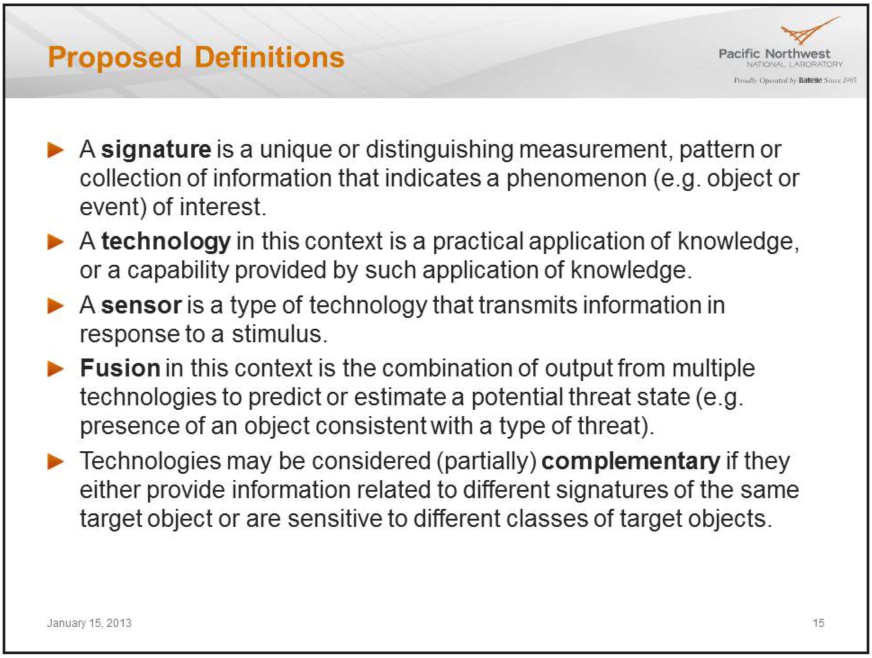
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Sponsored by DHS S&T Explosives Division

Focused on baggage and checkpoint screening for explosives

Objectives

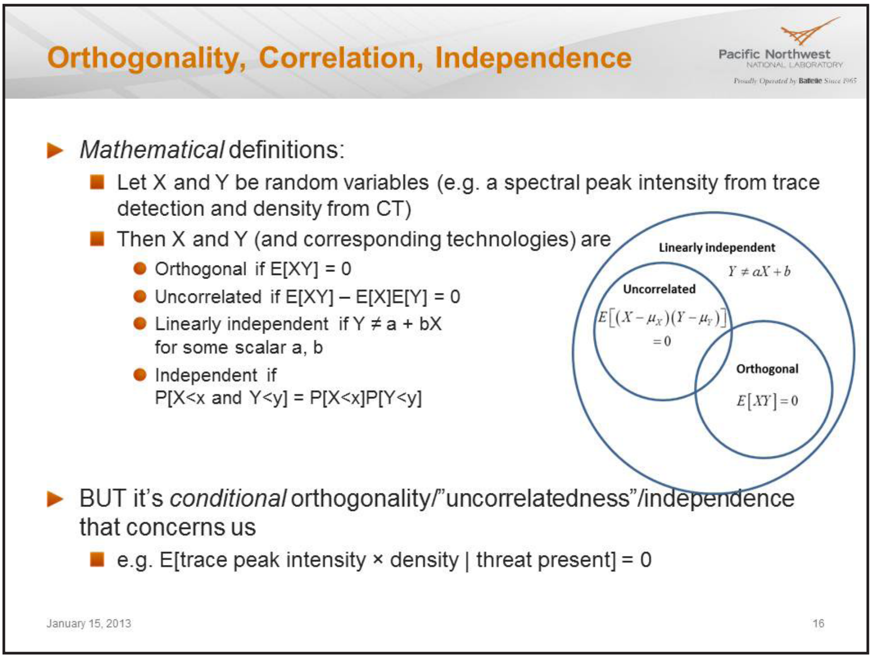
- Develop strategies for research in complementary technologies
  - Based on mathematical arguments and issues
  - Frame programmatic strategy for evaluating systems
- Initiate and oversee student "incubator" projects
- Outcomes: Briefing and two reports with recommendations pertinent to researchers, vendors, and funding agencies



Proposed Definitions

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- A **signature** is a unique or distinguishing measurement, pattern or collection of information that indicates a phenomenon (e.g. object or event) of interest.
- A **technology** in this context is a practical application of knowledge, or a capability provided by such application of knowledge.
- A **sensor** is a type of technology that transmits information in response to a stimulus.
- **Fusion** in this context is the combination of output from multiple technologies to predict or estimate a potential threat state (e.g. presence of an object consistent with a type of threat).
- Technologies may be considered (partially) **complementary** if they either provide information related to different signatures of the same target object or are sensitive to different classes of target objects.

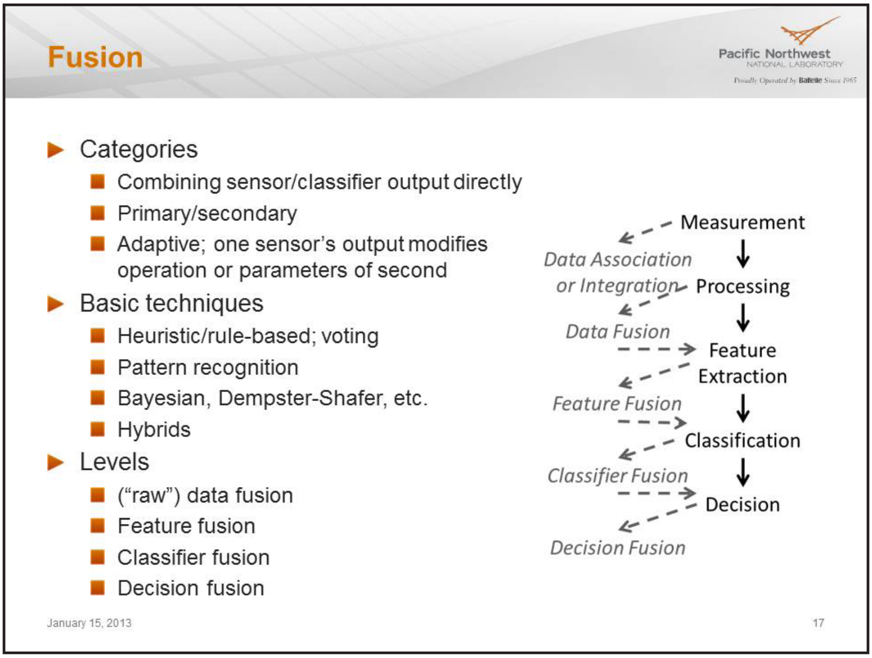


Orthogonality, Correlation, Independence

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- *Mathematical* definitions:
  - Let X and Y be random variables (e.g. a spectral peak intensity from trace detection and density from CT)
  - Then X and Y (and corresponding technologies) are
    - Orthogonal if E[XY] = 0
    - Uncorrelated if E[XY] – E[X]E[Y] = 0
    - Linearly independent if Y ≠ a + bX for some scalar a, b
    - Independent if P[X<x and Y<y] = P[X<x]P[Y<y]

- BUT it's *conditional* orthogonality/"uncorrelatedness"/independence that concerns us
  - e.g. E[trace peak intensity × density | threat present] = 0



Fusion

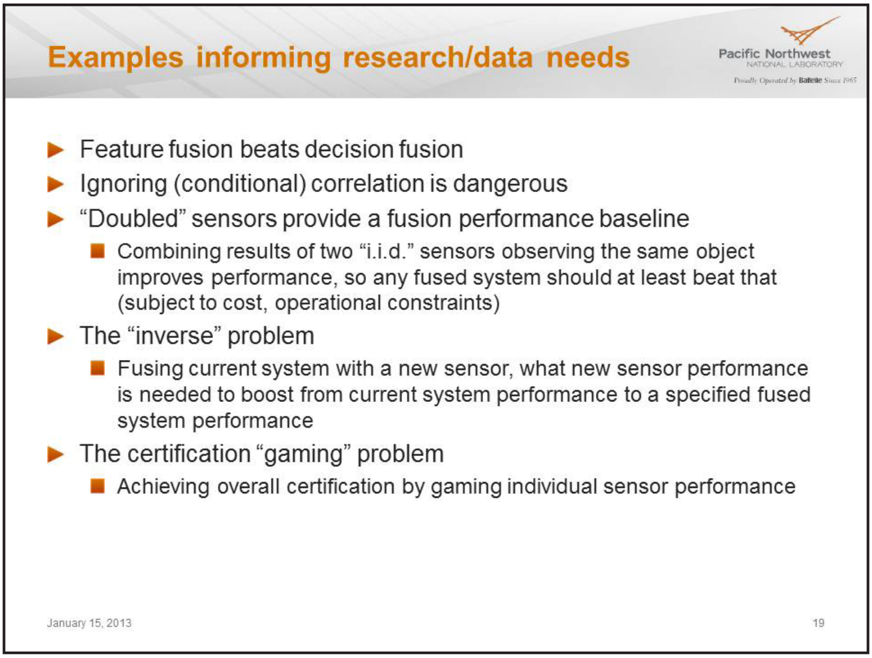
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Categories
  - Combining sensor/classifier output directly
  - Primary/secondary
  - Adaptive; one sensor's output modifies operation or parameters of second
- Basic techniques
  - Heuristic/rule-based; voting
  - Pattern recognition
  - Bayesian, Dempster-Shafer, etc.
  - Hybrids
- Levels
  - ("raw") data fusion
  - Feature fusion
  - Classifier fusion
  - Decision fusion

Fusion Research and Data Needs

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- DoD Wisdom
  - Fusion framework elements
  - Fusion methodology
  - Categorized "pitfalls"



Examples informing research/data needs

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Feature fusion beats decision fusion
- Ignoring (conditional) correlation is dangerous
- “Doubled” sensors provide a fusion performance baseline
  - Combining results of two “i.i.d.” sensors observing the same object improves performance, so any fused system should at least beat that (subject to cost, operational constraints)
- The “inverse” problem
  - Fusing current system with a new sensor, what new sensor performance is needed to boost from current system performance to a specified fused system performance
- The certification “gaming” problem
  - Achieving overall certification by gaming individual sensor performance

January 15, 2013

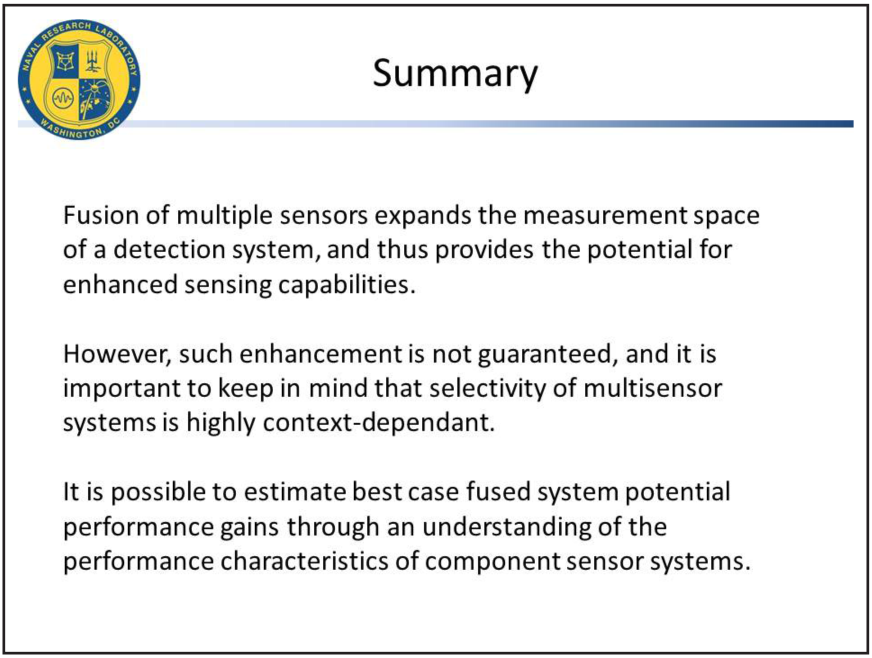
16.24 Kevin Johnson: Fused Sensor System Capabilities and Limitations

Understanding fused sensor system capabilities and limitations

Kevin Johnson
Naval Research Laboratory
Washington, DC 20375

Christian Minor
Nova Research, Inc.
Alexandria, VA 22308

This research was funded and supported by
Department of Homeland Security
Science and Technology Directorate
Contract HSHQDC-11-X-00561

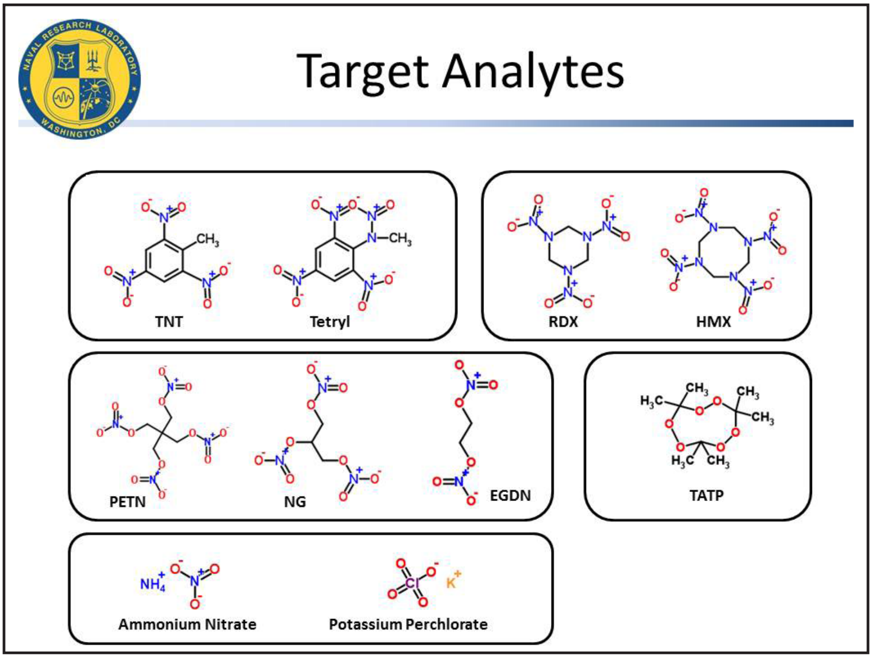


Summary

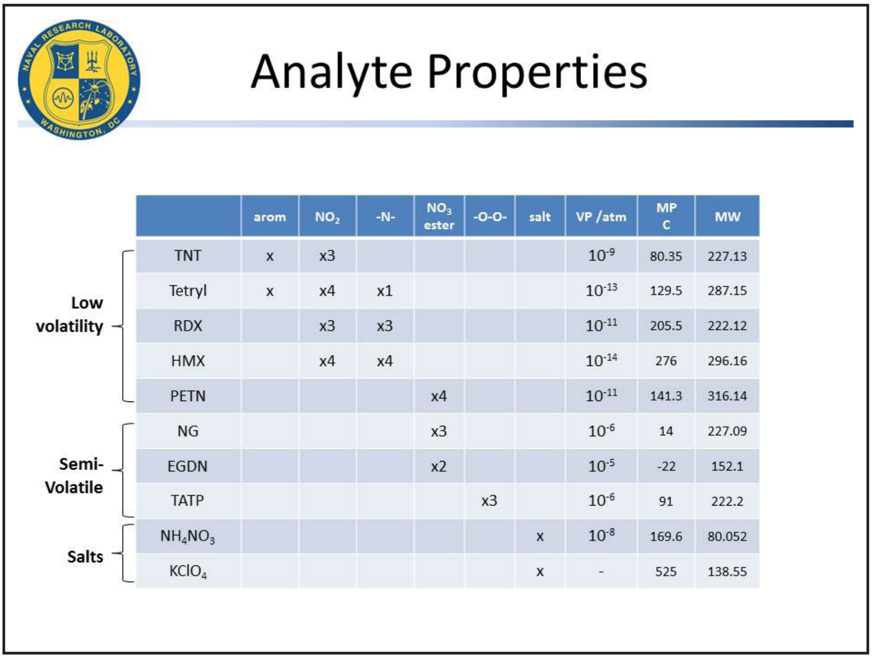
Fusion of multiple sensors expands the measurement space of a detection system, and thus provides the potential for enhanced sensing capabilities.

However, such enhancement is not guaranteed, and it is important to keep in mind that selectivity of multisensor systems is highly context-dependant.

It is possible to estimate best case fused system potential performance gains through an understanding of the performance characteristics of component sensor systems.

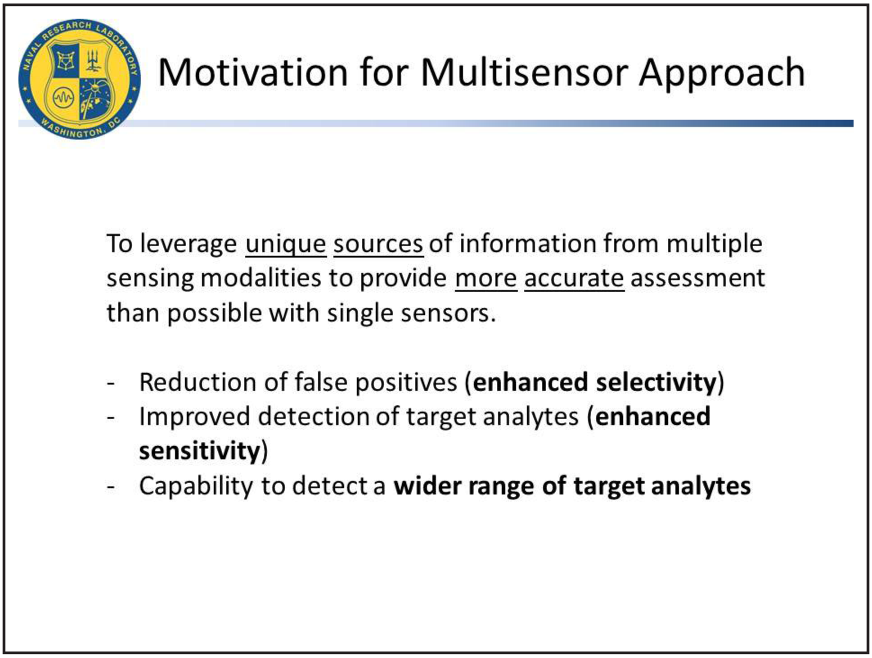


Target Analytes



Analyte Properties

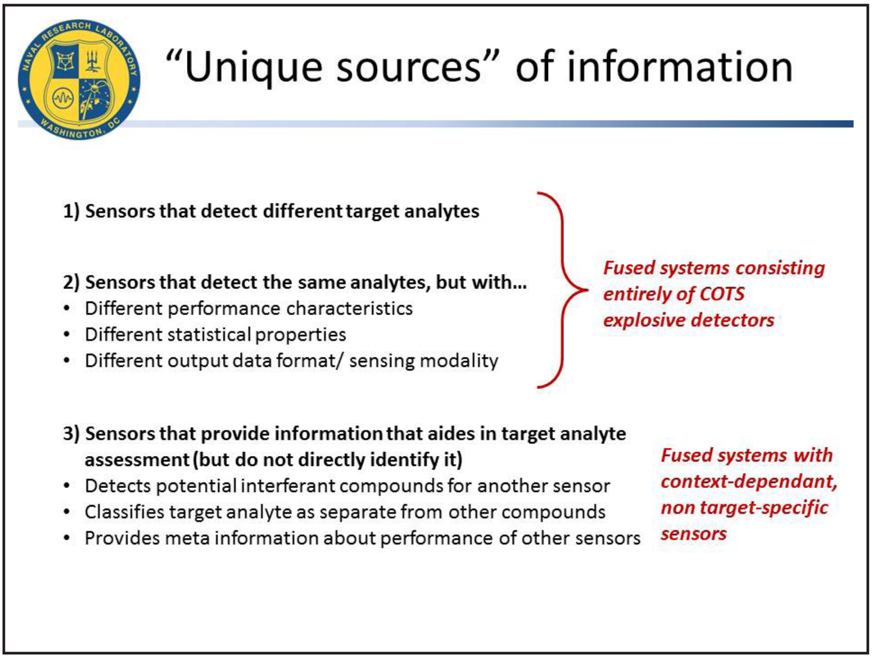
		arom	NO_2	-N-	NO_3 ester	-O-O-	salt	VP /atm	MP C	MW
Low volatility	TNT	x	x3					10^{-9}	80.35	227.13
	Tetryl	x	x4	x1				10^{-13}	129.5	287.15
	RDX		x3	x3				10^{-11}	205.5	222.12
	HMX		x4	x4				10^{-14}	276	296.16
	PETN				x4			10^{-11}	141.3	316.14
Semi-Volatile	NG				x3			10^{-6}	14	227.09
	EGDN				x2			10^{-5}	-22	152.1
	TATP					x3		10^{-6}	91	222.2
Salts	NH_4NO_3						x	10^{-8}	169.6	80.052
	$KClO_4$						x	-	525	138.55



Motivation for Multisensor Approach

To leverage unique sources of information from multiple sensing modalities to provide more accurate assessment than possible with single sensors.

- Reduction of false positives (**enhanced selectivity**)
- Improved detection of target analytes (**enhanced sensitivity**)
- Capability to detect a **wider range of target analytes**



“Unique sources” of information

1) Sensors that detect different target analytes

2) Sensors that detect the same analytes, but with…
- Different performance characteristics
- Different statistical properties
- Different output data format/ sensing modality

Fused systems consisting entirely of COTS explosive detectors

3) Sensors that provide information that aides in target analyte assessment (but do not directly identify it)
- Detects potential interferant compounds for another sensor
- Classifies target analyte as separate from other compounds
- Provides meta information about performance of other sensors

Fused systems with context-dependant, non target-specific sensors

Current Landscape

FirstDefender RM
(Thermo Scientific)

Hazmat ID Ranger
(Smiths Detection)

Quantum Sniffer QS-H150
(Implant Sciences Corp.)

TrueDefender FT/FTG
(Thermo Scientific)

Fido Verdict (ICx Technologies)

MobileTrace
(Morpho Detection)

Multi-Mode Threat Detector
(Smiths Detection)

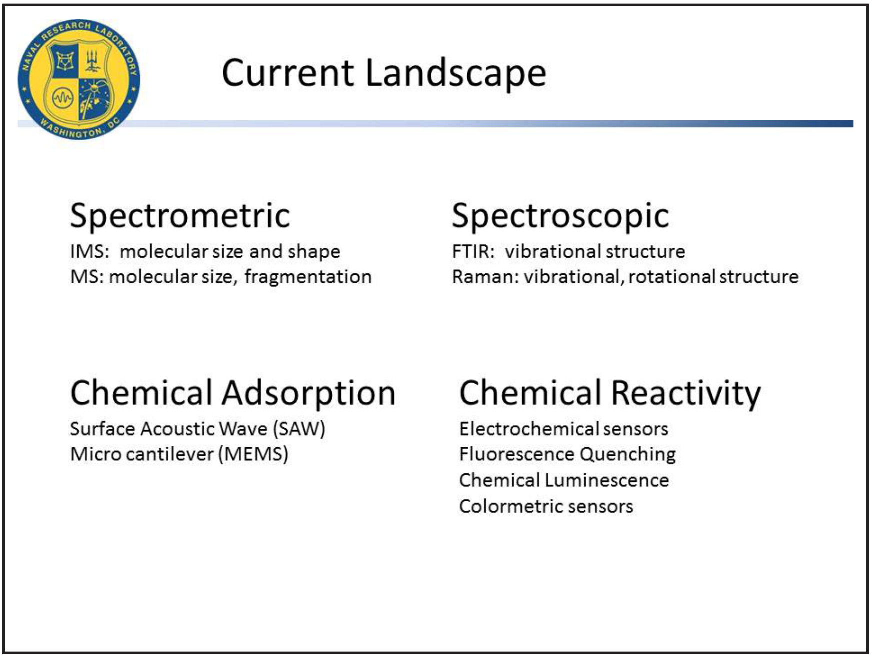
EVD 3000+
(Scintrex)

E3500 Chemilux
(Scintrex)

NevadaNano Self Sensing Array
(Nevada Nanotech Systems, Inc.)

Fido PaxPoint (ICx Technologies)

zNose Model 4500
(Electric Sensor Technology)



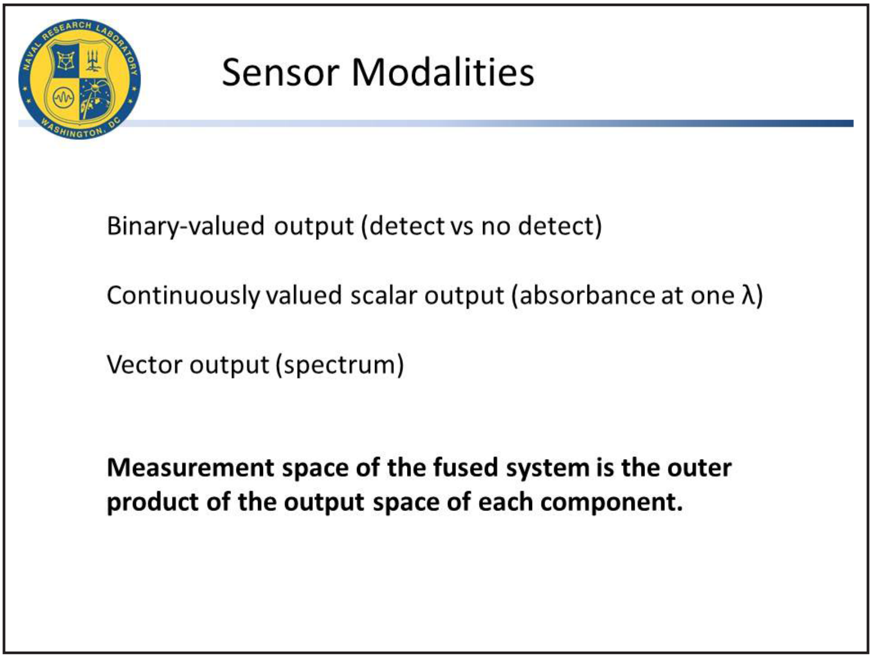
Current Landscape

Spectrometric
IMS: molecular size and shape
MS: molecular size, fragmentation

Spectroscopic
FTIR: vibrational structure
Raman: vibrational, rotational structure

Chemical Adsorption
Surface Acoustic Wave (SAW)
Micro cantilever (MEMS)

Chemical Reactivity
Electrochemical sensors
Fluorescence Quenching
Chemical Luminescence
Colormetric sensors



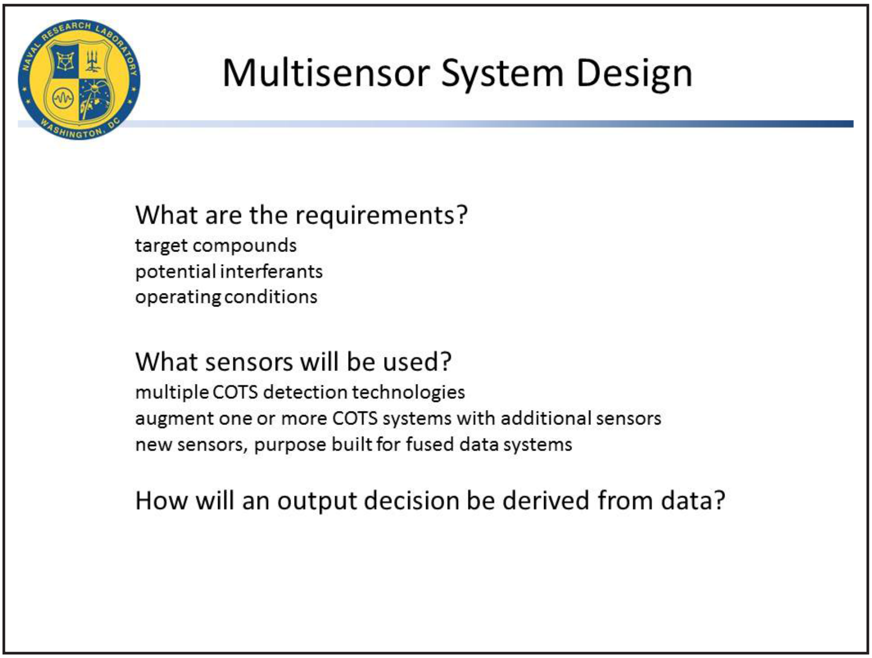
Sensor Modalities

Binary-valued output (detect vs no detect)

Continuously valued scalar output (absorbance at one λ)

Vector output (spectrum)

Measurement space of the fused system is the outer product of the output space of each component.



Multisensor System Design

What are the requirements?
target compounds
potential interferants
operating conditions

What sensors will be used?
multiple COTS detection technologies
augment one or more COTS systems with additional sensors
new sensors, purpose built for fused data systems

How will an output decision be derived from data?

Detection and Decision Theory

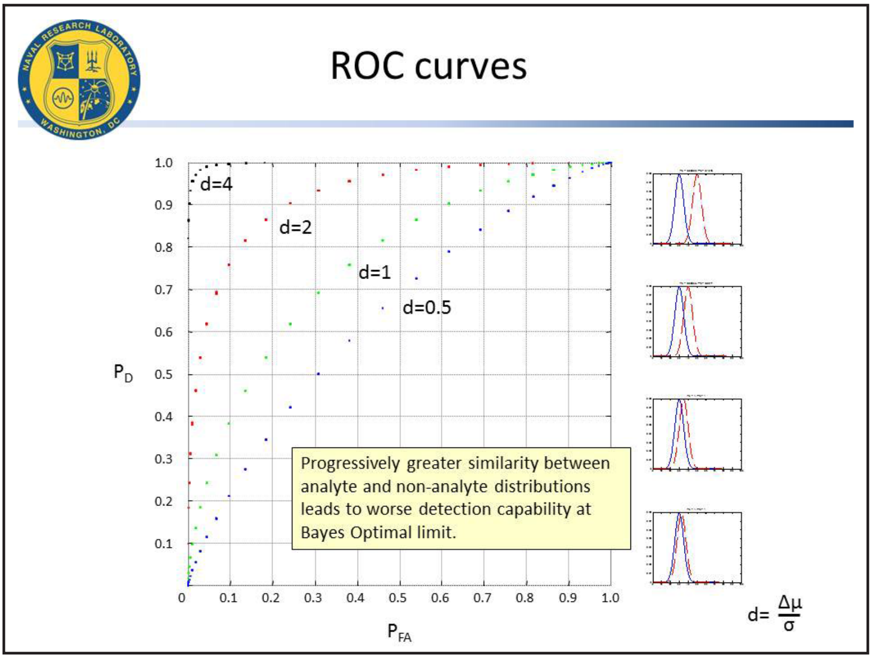
Detection of an analyte is a decision that is made on the basis of measured sensor data.

Decision theory: arrive at an optimal decision, given observable evidence and knowledge of the statistical distributions.

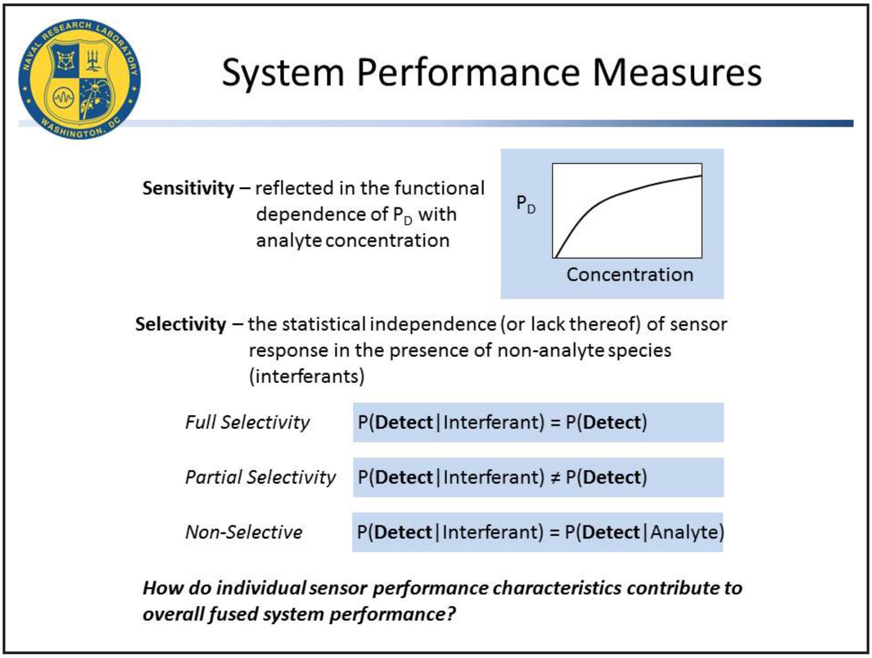
Ability of a sensor (or multisensor system) to detect an analyte rests on the distribution of sensor responses observed when the analyte is present and when it is not.

Adding or removing sensors, incorporating other compounds and interferences, and varying the amount of each present will alter these distributions, leading to changes in the system's detection ability.

ROC curves



Progressively greater similarity between analyte and non-analyte distributions leads to worse detection capability at Bayes Optimal limit.



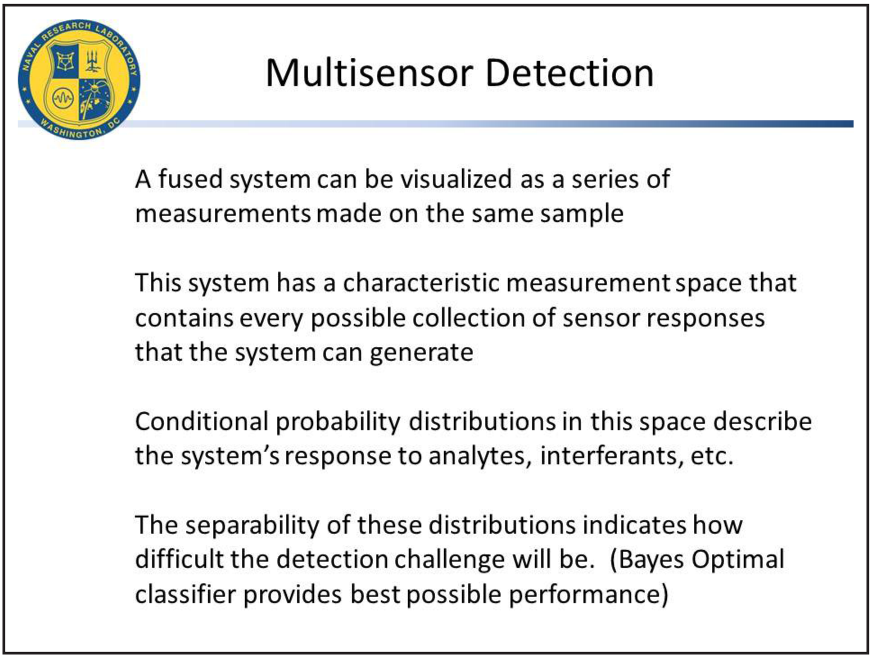
System Performance Measures

Sensitivity – reflected in the functional dependence of P_D with analyte concentration

Selectivity – the statistical independence (or lack thereof) of sensor response in the presence of non-analyte species (interferants)

Full Selectivity	P(**Detect**|Interferant) = P(**Detect**)
Partial Selectivity	P(**Detect**|Interferant) ≠ P(**Detect**)
Non-Selective	P(**Detect**|Interferant) = P(**Detect**|Analyte)

How do individual sensor performance characteristics contribute to overall fused system performance?



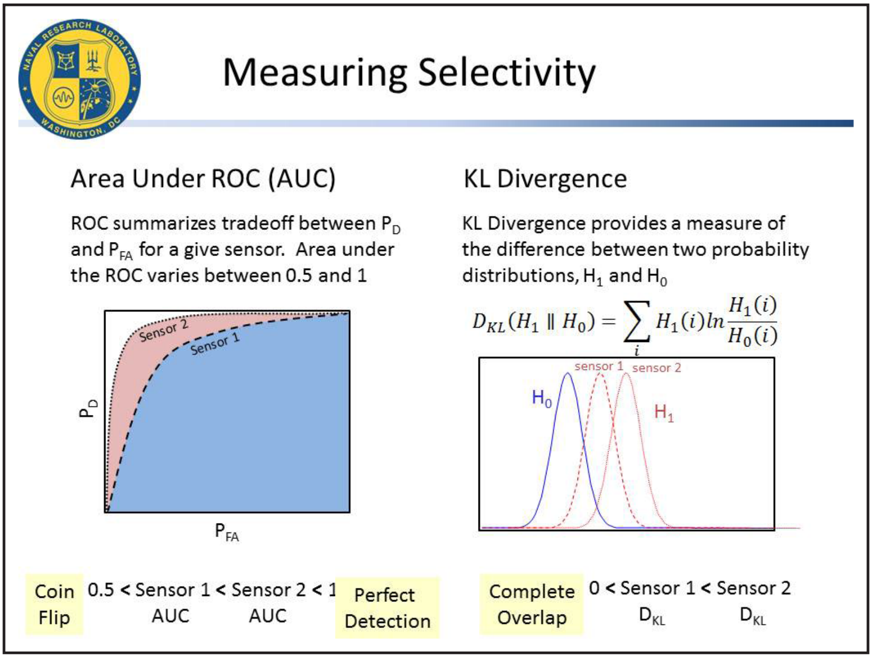
Multisensor Detection

A fused system can be visualized as a series of measurements made on the same sample

This system has a characteristic measurement space that contains every possible collection of sensor responses that the system can generate

Conditional probability distributions in this space describe the system's response to analytes, interferants, etc.

The separability of these distributions indicates how difficult the detection challenge will be. (Bayes Optimal classifier provides best possible performance)



Measuring Selectivity

Area Under ROC (AUC)

ROC summarizes tradeoff between P_D and P_{FA} for a give sensor. Area under the ROC varies between 0.5 and 1

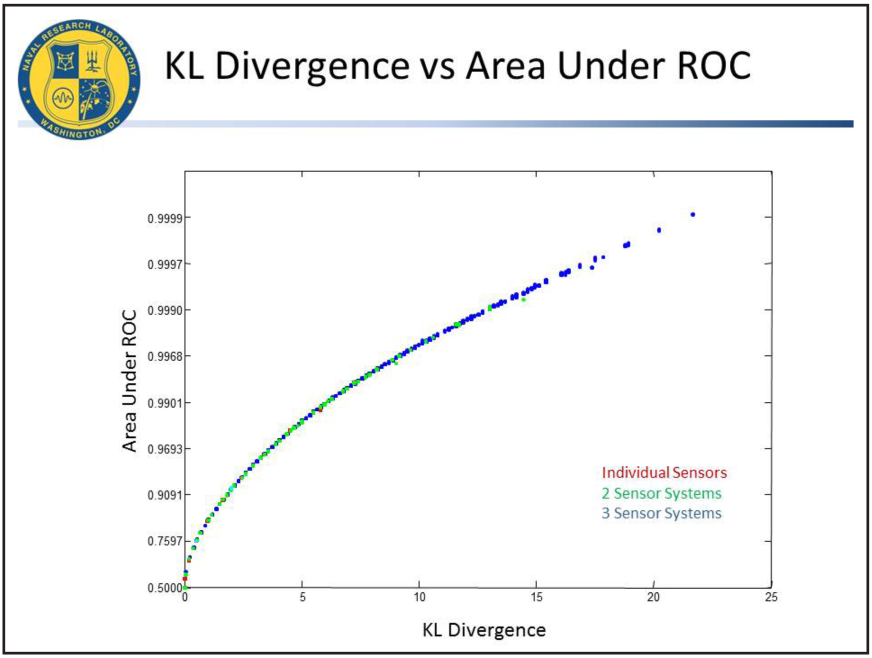
Coin Flip | 0.5 < Sensor 1 AUC < Sensor 2 AUC < 1 | Perfect Detection

KL Divergence

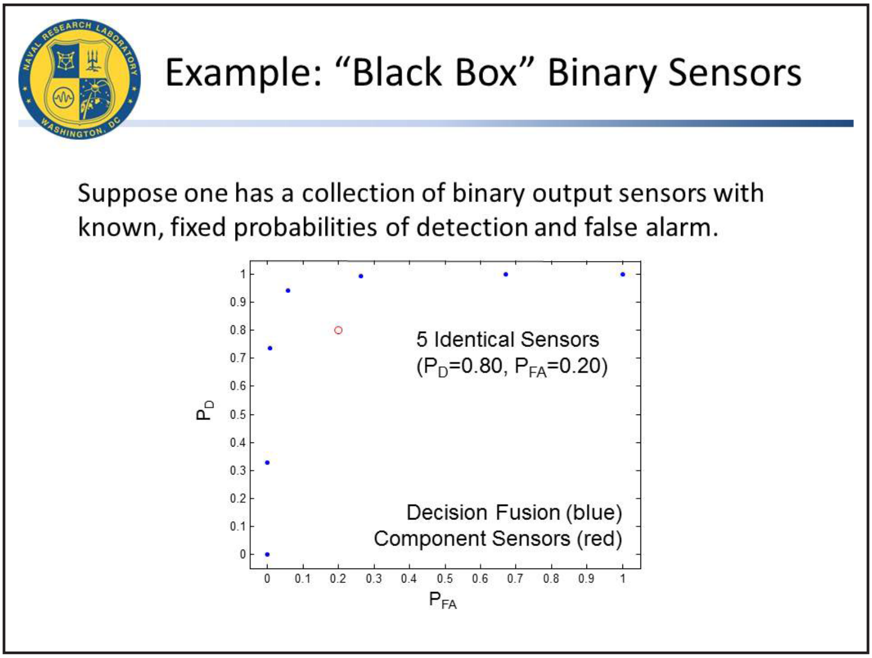
KL Divergence provides a measure of the difference between two probability distributions, H_1 and H_0

$$D_{KL}(H_1 \parallel H_0) = \sum_i H_1(i) ln \frac{H_1(i)}{H_0(i)}$$

Complete Overlap | 0 < Sensor 1 D_{KL} < Sensor 2 D_{KL}

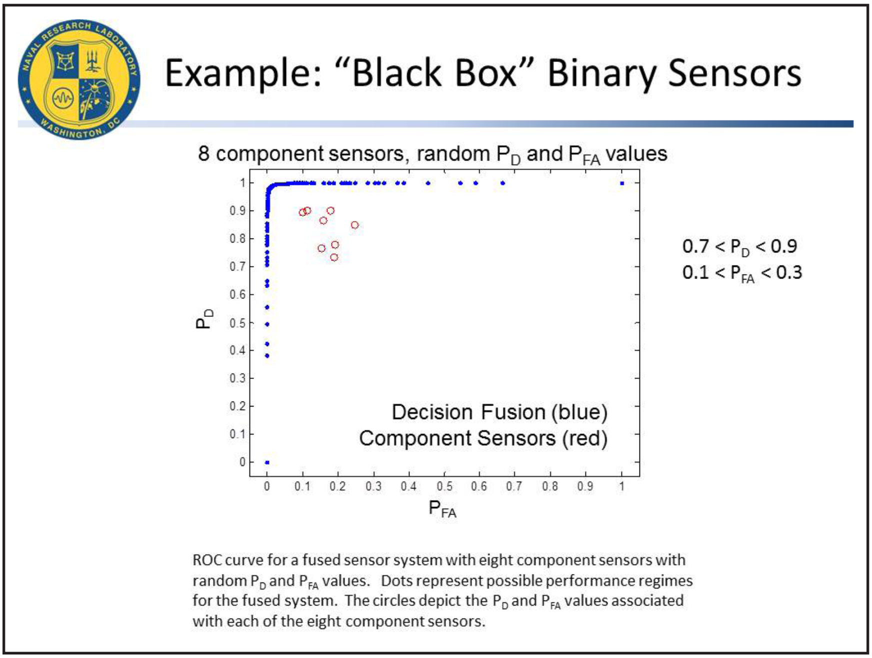


KL Divergence vs Area Under ROC



Example: "Black Box" Binary Sensors

Suppose one has a collection of binary output sensors with known, fixed probabilities of detection and false alarm.

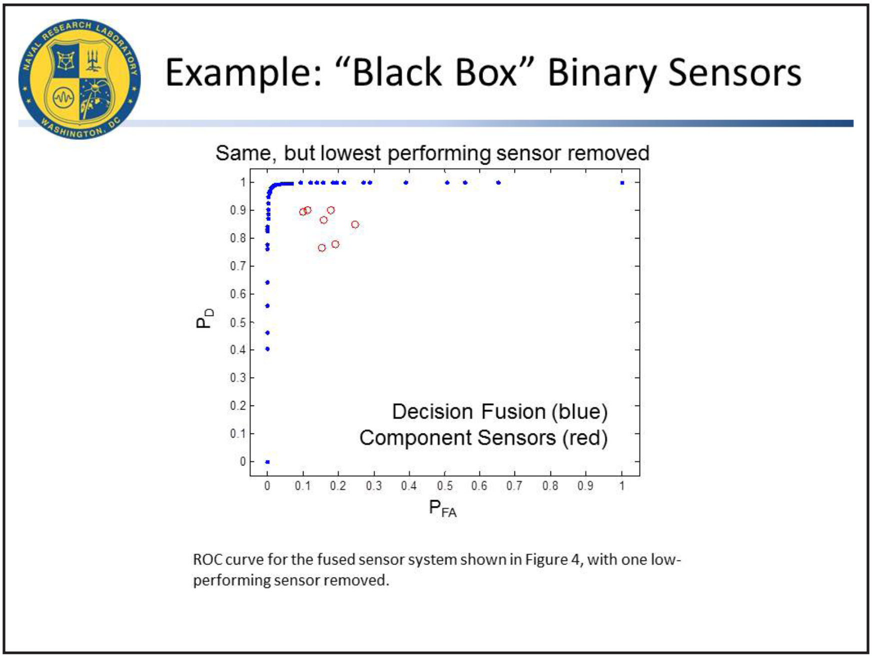


Example: "Black Box" Binary Sensors

8 component sensors, random P_D and P_{FA} values

$0.7 < P_D < 0.9$
$0.1 < P_{FA} < 0.3$

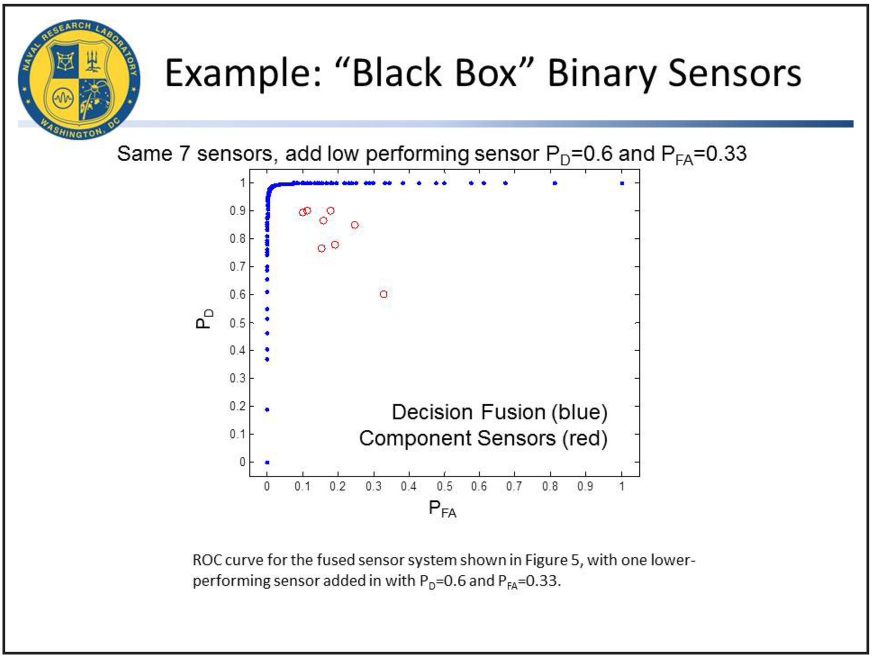
ROC curve for a fused sensor system with eight component sensors with random P_D and P_{FA} values. Dots represent possible performance regimes for the fused system. The circles depict the P_D and P_{FA} values associated with each of the eight component sensors.



Example: "Black Box" Binary Sensors

Same, but lowest performing sensor removed

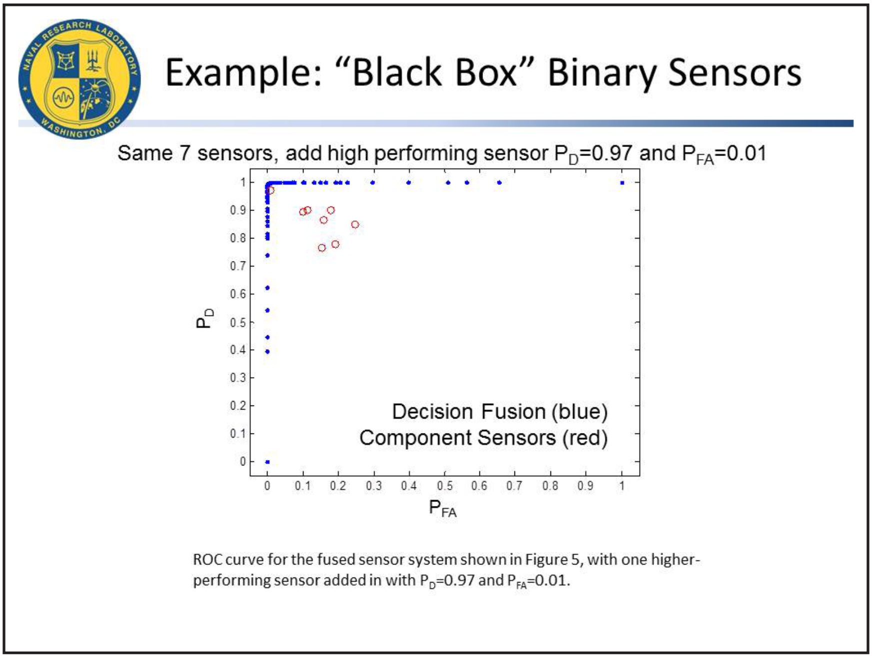
ROC curve for the fused sensor system shown in Figure 4, with one low-performing sensor removed.



Example: "Black Box" Binary Sensors

Same 7 sensors, add low performing sensor P_D=0.6 and P_{FA}=0.33

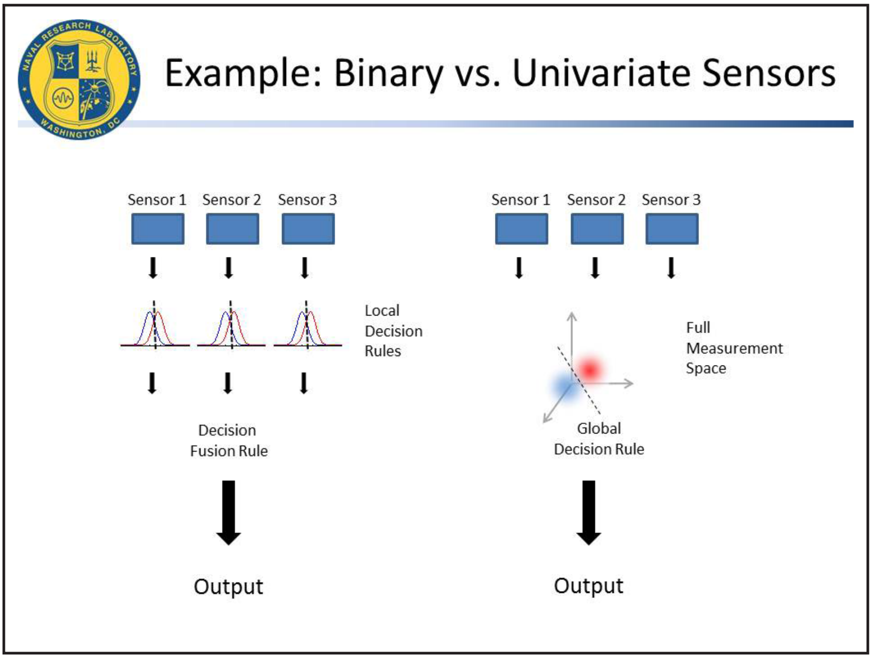
ROC curve for the fused sensor system shown in Figure 5, with one lower-performing sensor added in with P_D=0.6 and P_{FA}=0.33.



Example: "Black Box" Binary Sensors

Same 7 sensors, add high performing sensor P_D=0.97 and P_{FA}=0.01

ROC curve for the fused sensor system shown in Figure 5, with one higher-performing sensor added in with P_D=0.97 and P_{FA}=0.01.



Example: Binary vs. Univariate Sensors

Sensor 1 Sensor 2 Sensor 3

Local
Decision
Rules

Decision
Fusion Rule

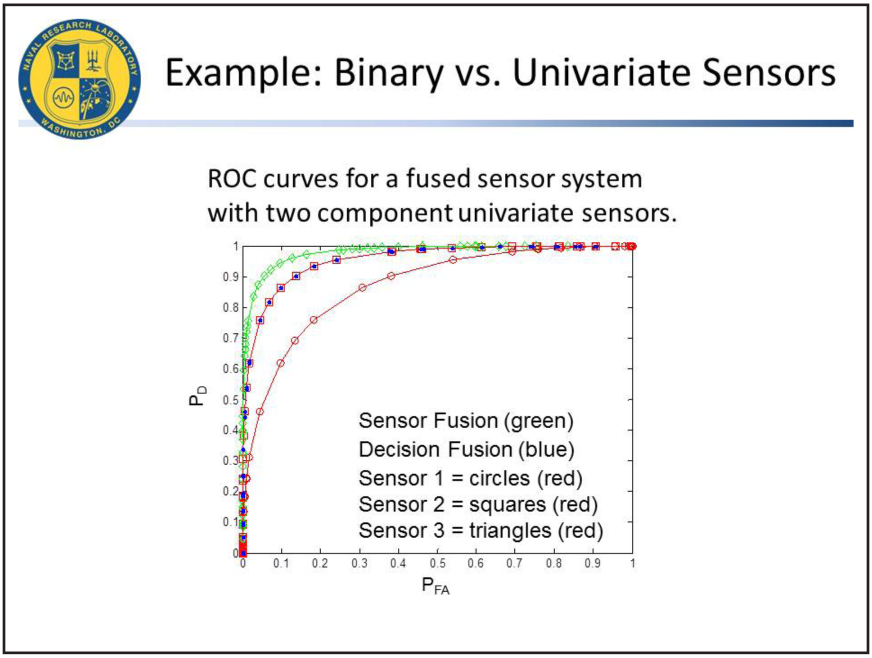
Output

Sensor 1 Sensor 2 Sensor 3

Full
Measurement
Space

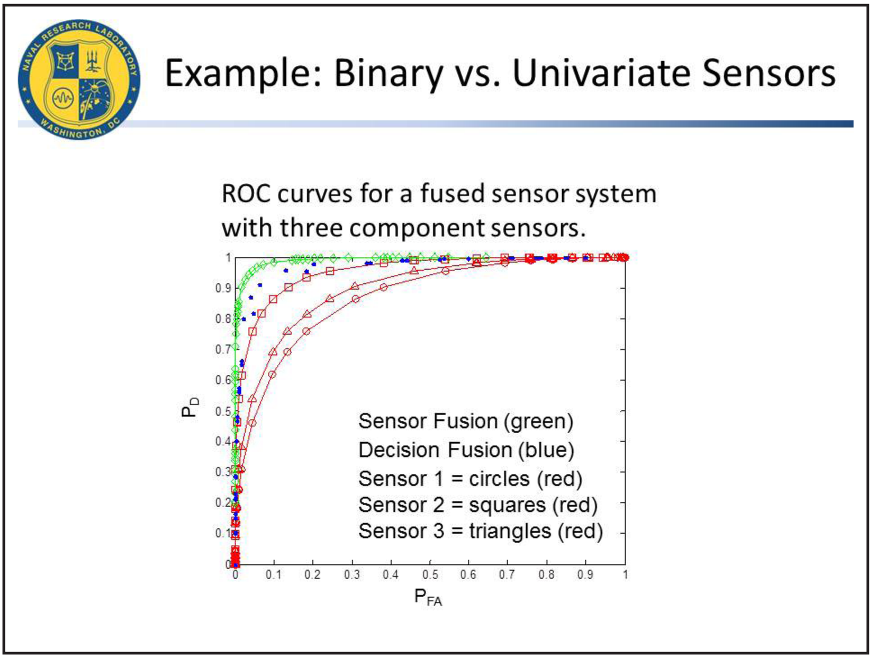
Global
Decision Rule

Output



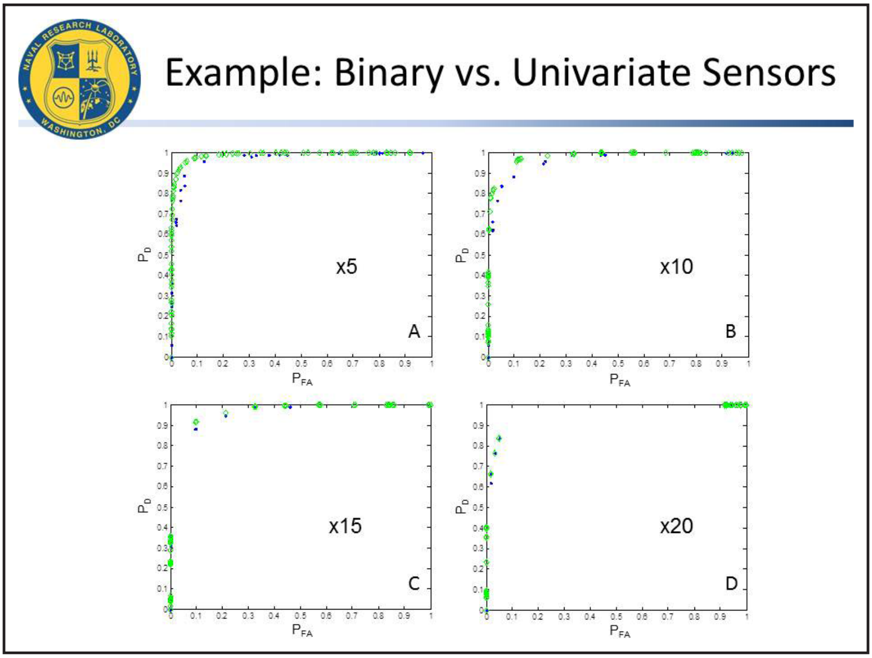
Example: Binary vs. Univariate Sensors

ROC curves for a fused sensor system
with two component univariate sensors.

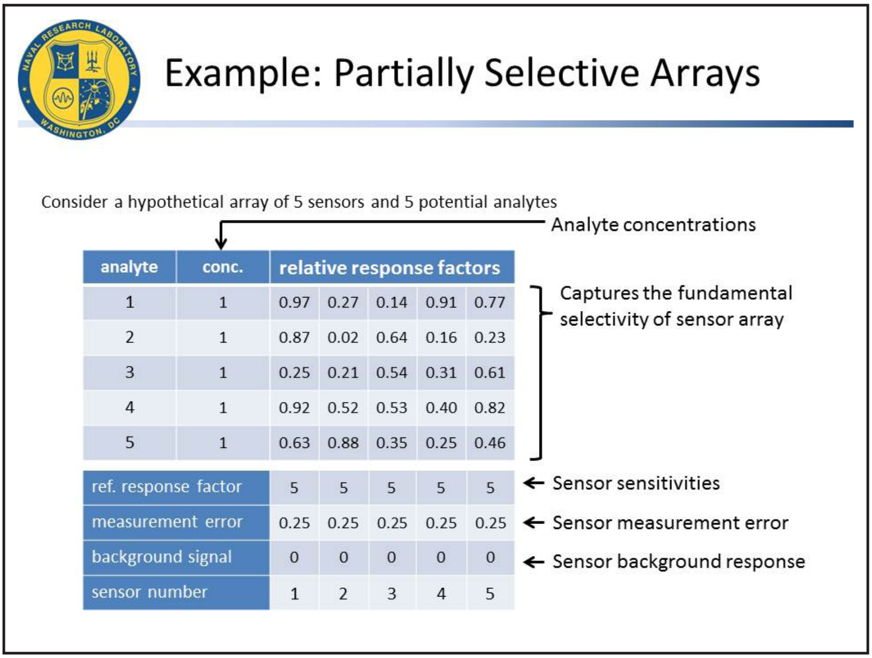


Example: Binary vs. Univariate Sensors

ROC curves for a fused sensor system
with three component sensors.



Example: Binary vs. Univariate Sensors



Example: Partially Selective Arrays

Consider a hypothetical array of 5 sensors and 5 potential analytes

Analyte concentrations

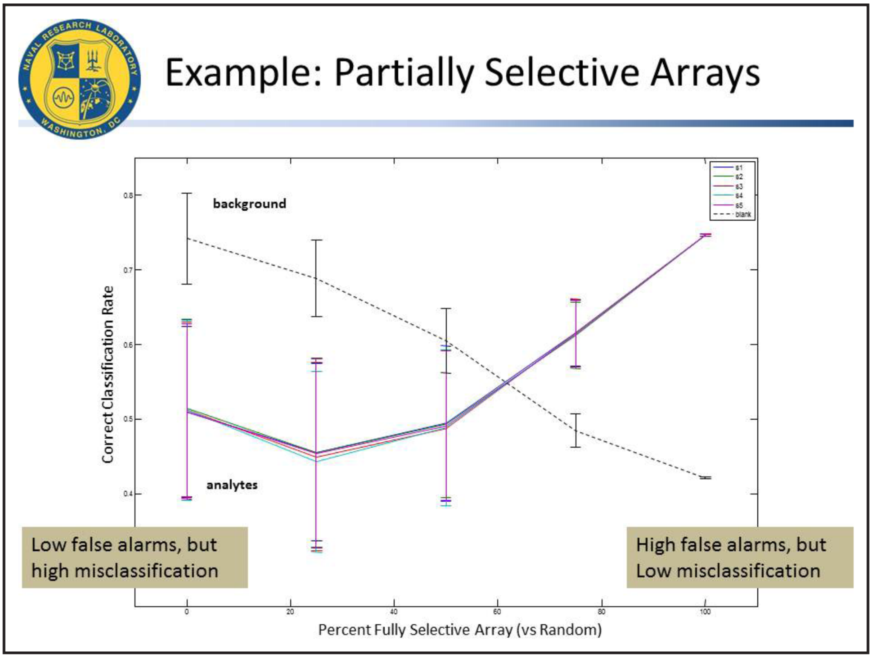
analyte	conc.	relative response factors				
1	1	0.97	0.27	0.14	0.91	0.77
2	1	0.87	0.02	0.64	0.16	0.23
3	1	0.25	0.21	0.54	0.31	0.61
4	1	0.92	0.52	0.53	0.40	0.82
5	1	0.63	0.88	0.35	0.25	0.46
ref. response factor		5	5	5	5	5
measurement error		0.25	0.25	0.25	0.25	0.25
background signal		0	0	0	0	0
sensor number		1	2	3	4	5

Captures the fundamental selectivity of sensor array

← Sensor sensitivities

← Sensor measurement error

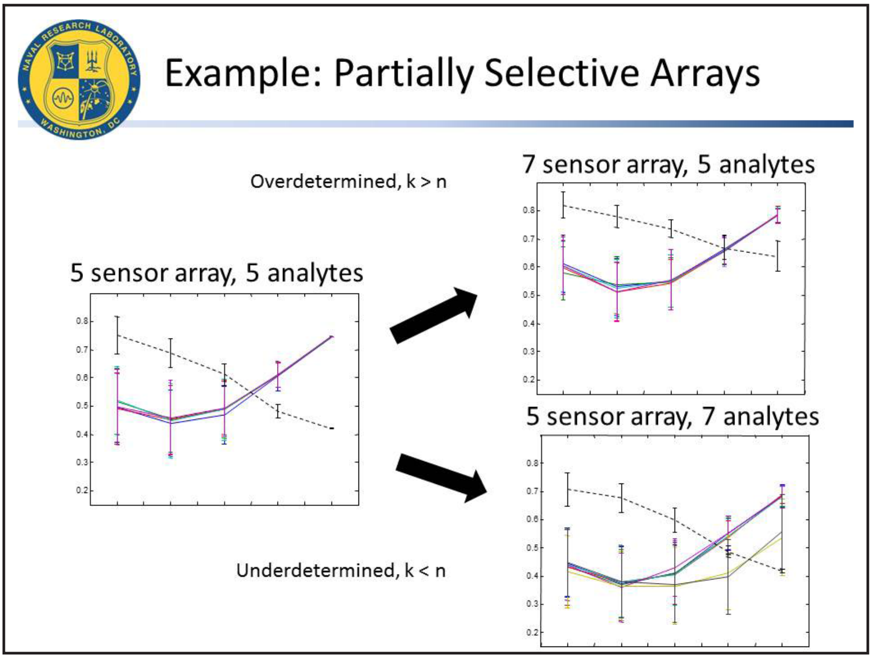
← Sensor background response



Example: Partially Selective Arrays

Low false alarms, but
high misclassification

High false alarms, but
Low misclassification



Example: Partially Selective Arrays

Overdetermined, k > n

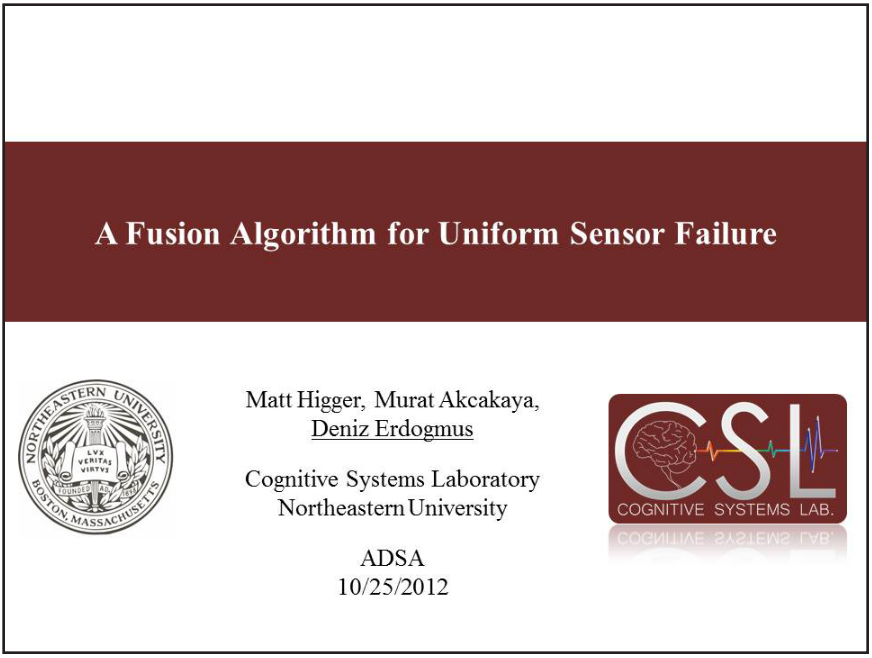
5 sensor array, 5 analytes

7 sensor array, 5 analytes

5 sensor array, 7 analytes

Underdetermined, k < n

16.25 Deniz Erdogmus: Robust Fusion Algorithm for Sensor Failure

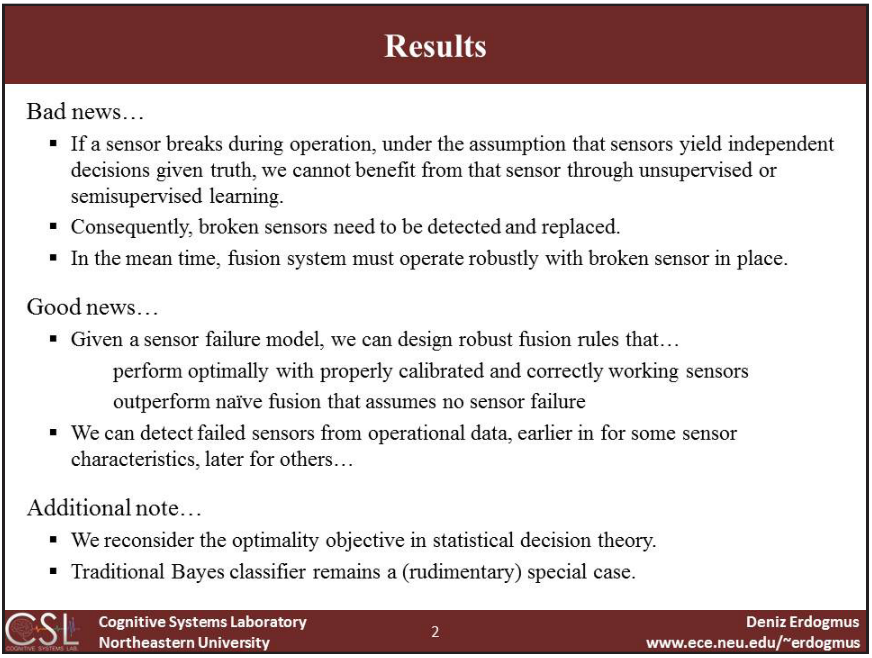


A Fusion Algorithm for Uniform Sensor Failure

Matt Higger, Murat Akcakaya,
Deniz Erdogmus

Cognitive Systems Laboratory
Northeastern University

ADSA
10/25/2012



Results

Bad news…

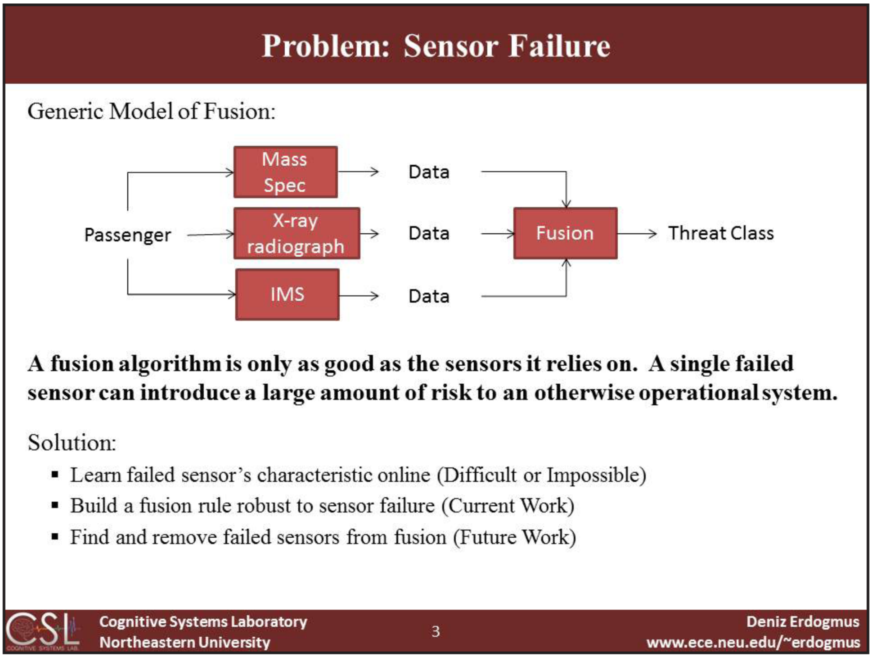
- If a sensor breaks during operation, under the assumption that sensors yield independent decisions given truth, we cannot benefit from that sensor through unsupervised or semisupervised learning.
- Consequently, broken sensors need to be detected and replaced.
- In the mean time, fusion system must operate robustly with broken sensor in place.

Good news…

- Given a sensor failure model, we can design robust fusion rules that…
  perform optimally with properly calibrated and correctly working sensors
  outperform naïve fusion that assumes no sensor failure
- We can detect failed sensors from operational data, earlier in for some sensor characteristics, later for others…

Additional note…

- We reconsider the optimality objective in statistical decision theory.
- Traditional Bayes classifier remains a (rudimentary) special case.



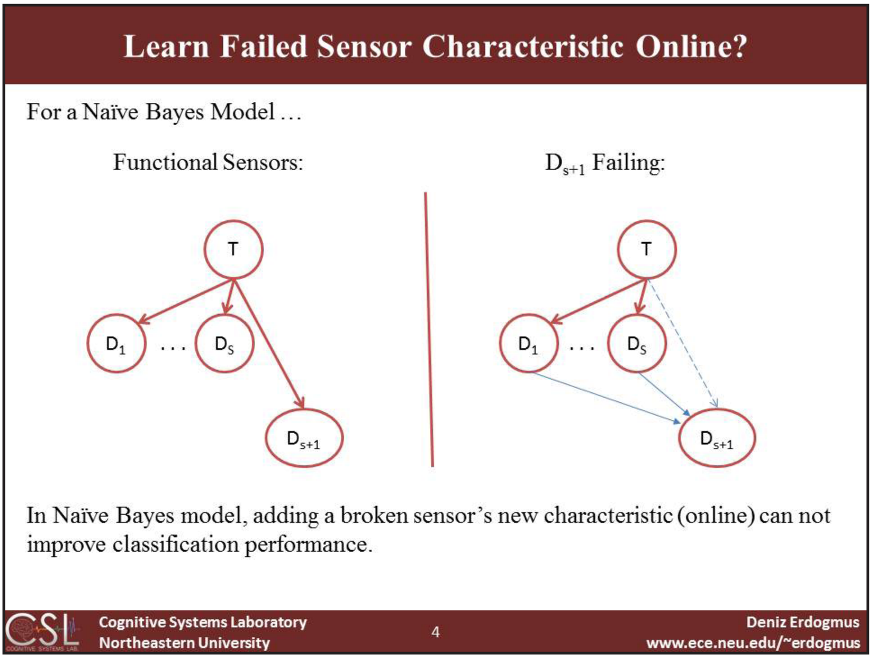
Problem: Sensor Failure

Generic Model of Fusion:

A fusion algorithm is only as good as the sensors it relies on. A single failed sensor can introduce a large amount of risk to an otherwise operational system.

Solution:

- Learn failed sensor's characteristic online (Difficult or Impossible)
- Build a fusion rule robust to sensor failure (Current Work)
- Find and remove failed sensors from fusion (Future Work)



Learn Failed Sensor Characteristic Online?

For a Naïve Bayes Model …

Functional Sensors:

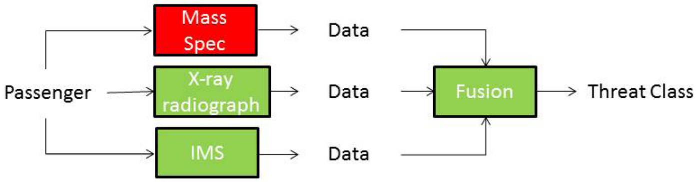
D_{s+1} Failing:

In Naïve Bayes model, adding a broken sensor's new characteristic (online) can not improve classification performance.

Build a Fusion Rule Robust to Sensor Failure

Given a uniform model of sensor failure, we can minimize risk by deciding:

$$
\begin{aligned}
F(d_{1:S}) &= 1 \quad \text{when} \quad \log\frac{r_{1|0}P_T(1)}{r_{0|1}P_T(0)} + \sum_{D_i} log\frac{P_{D_i|T}(d_i|1)}{P_{D_i|T}(d_i|0)} + log\frac{\sum_{r=0}^{r=S} C_1(d_{1:S},r)}{\sum_{r=0}^{r=S} C_0(d_{1:S},r)} < 0 \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$



Passenger

Mass Spec

X-ray radiograph

IMS

Data

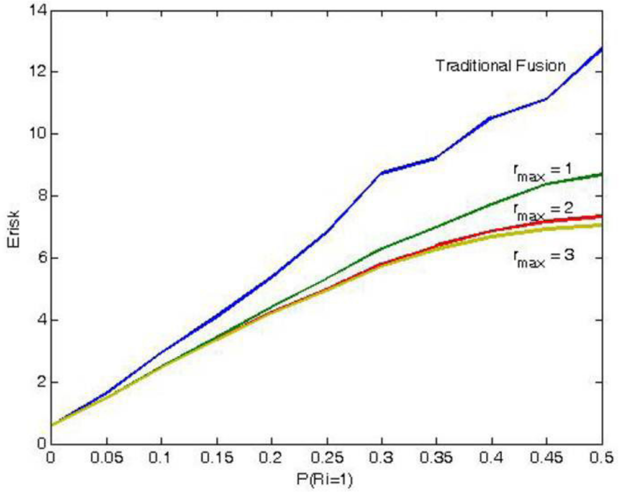
Data

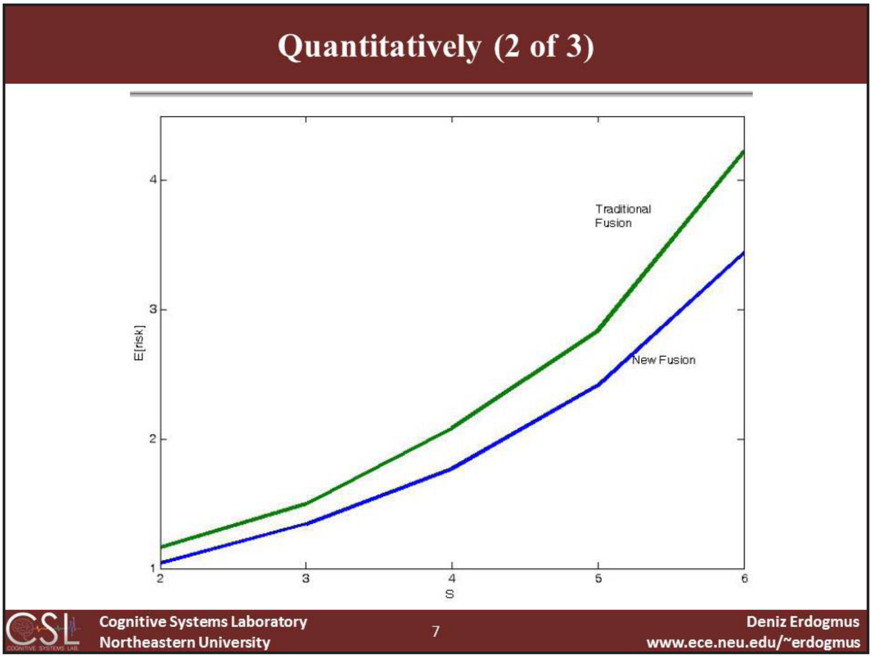
Data

Fusion

Threat Class

Quantitatively (1 of 3)

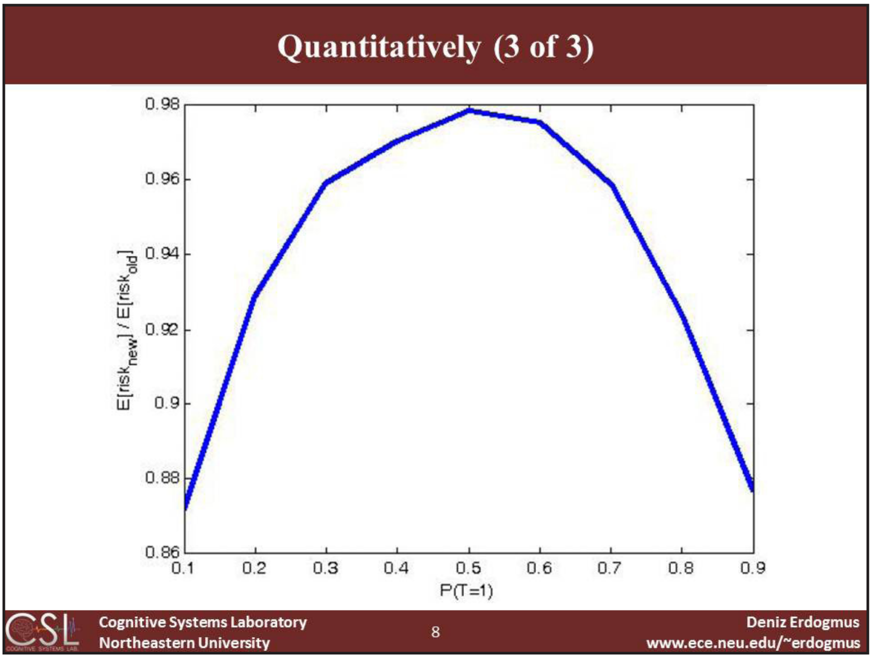




Quantitatively (2 of 3)

Cognitive Systems Laboratory
Northeastern University

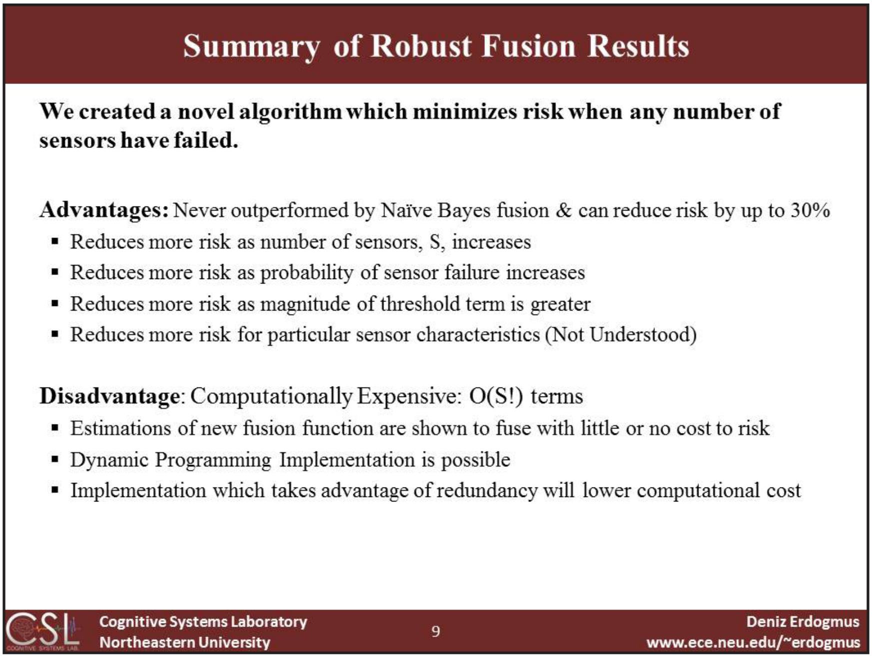
Deniz Erdogmus
www.ece.neu.edu/~erdogmus



Quantitatively (3 of 3)

Cognitive Systems Laboratory
Northeastern University

Deniz Erdogmus
www.ece.neu.edu/~erdogmus



Summary of Robust Fusion Results

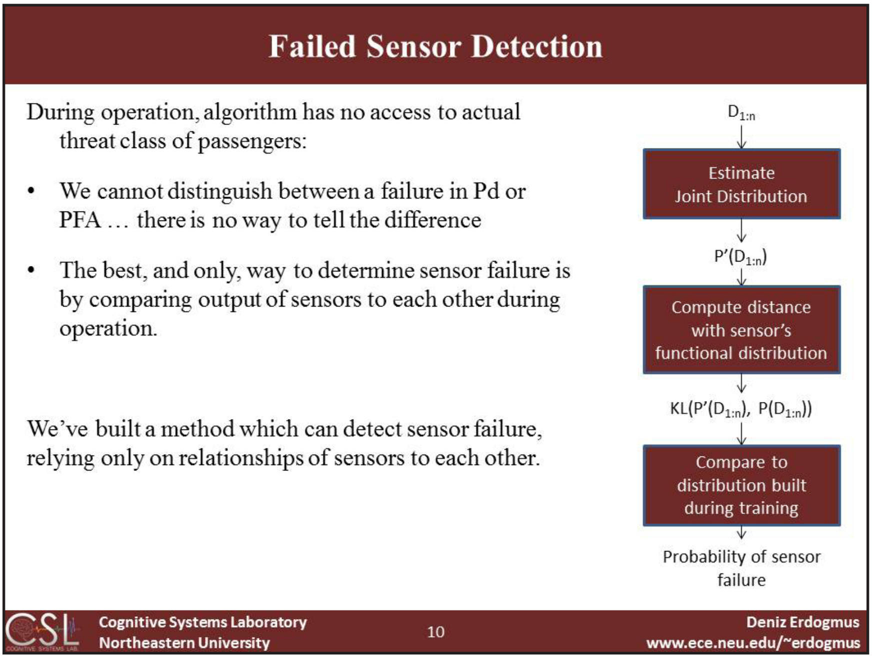
We created a novel algorithm which minimizes risk when any number of sensors have failed.

Advantages: Never outperformed by Naïve Bayes fusion & can reduce risk by up to 30%

- Reduces more risk as number of sensors, S, increases
- Reduces more risk as probability of sensor failure increases
- Reduces more risk as magnitude of threshold term is greater
- Reduces more risk for particular sensor characteristics (Not Understood)

Disadvantage: Computationally Expensive: O(S!) terms

- Estimations of new fusion function are shown to fuse with little or no cost to risk
- Dynamic Programming Implementation is possible
- Implementation which takes advantage of redundancy will lower computational cost



Failed Sensor Detection

During operation, algorithm has no access to actual threat class of passengers:

- We cannot distinguish between a failure in Pd or PFA … there is no way to tell the difference
- The best, and only, way to determine sensor failure is by comparing output of sensors to each other during operation.

We've built a method which can detect sensor failure, relying only on relationships of sensors to each other.

$D_{1:n}$

Estimate Joint Distribution

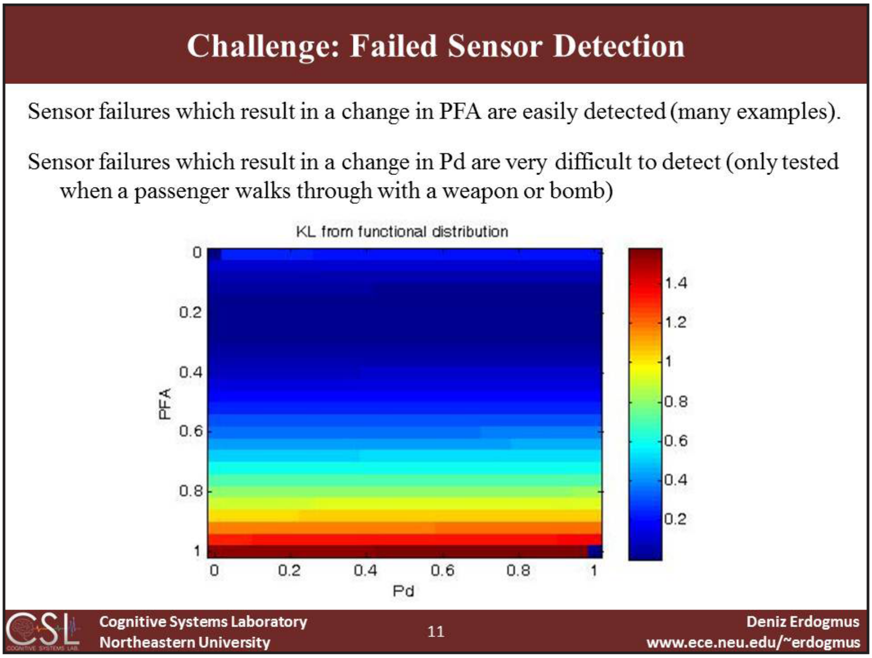
$P'(D_{1:n})$

Compute distance with sensor's functional distribution

$KL(P'(D_{1:n}), P(D_{1:n}))$

Compare to distribution built during training

Probability of sensor failure

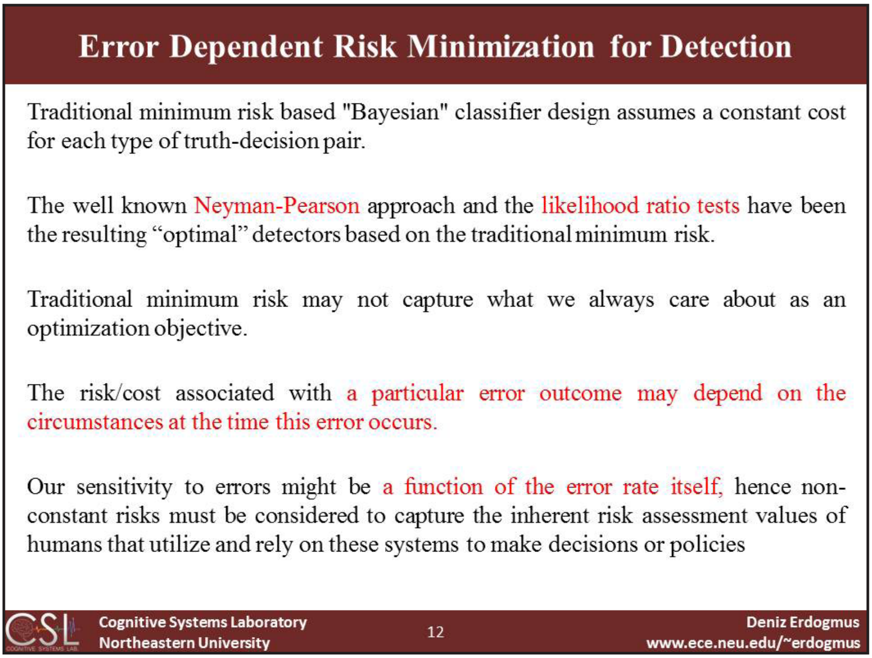


Challenge: Failed Sensor Detection

Sensor failures which result in a change in PFA are easily detected (many examples).

Sensor failures which result in a change in Pd are very difficult to detect (only tested when a passenger walks through with a weapon or bomb)

KL from functional distribution



Error Dependent Risk Minimization for Detection

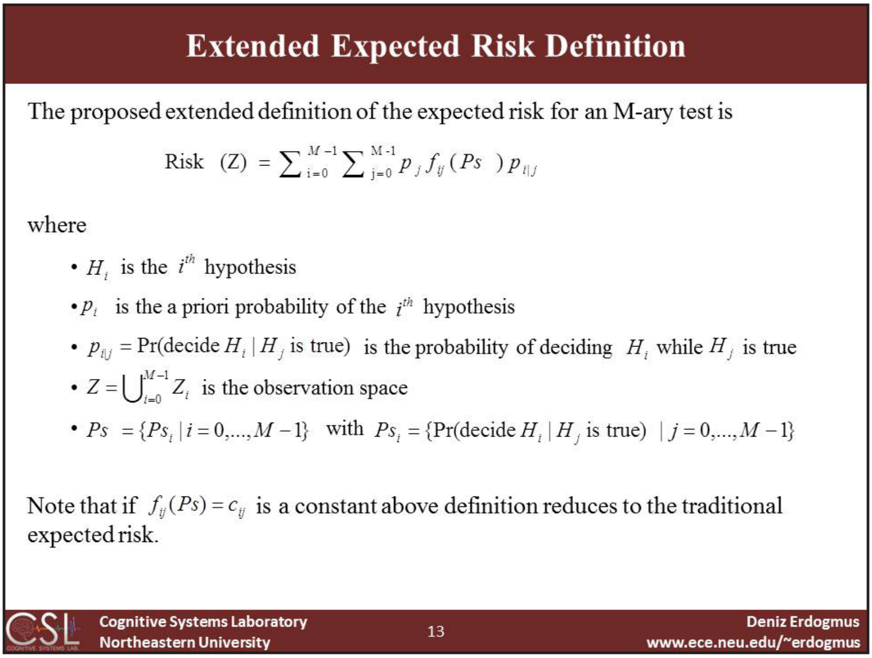
Traditional minimum risk based "Bayesian" classifier design assumes a constant cost for each type of truth-decision pair.

The well known Neyman-Pearson approach and the likelihood ratio tests have been the resulting “optimal” detectors based on the traditional minimum risk.

Traditional minimum risk may not capture what we always care about as an optimization objective.

The risk/cost associated with a particular error outcome may depend on the circumstances at the time this error occurs.

Our sensitivity to errors might be a function of the error rate itself, hence non-constant risks must be considered to capture the inherent risk assessment values of humans that utilize and rely on these systems to make decisions or policies



Extended Expected Risk Definition

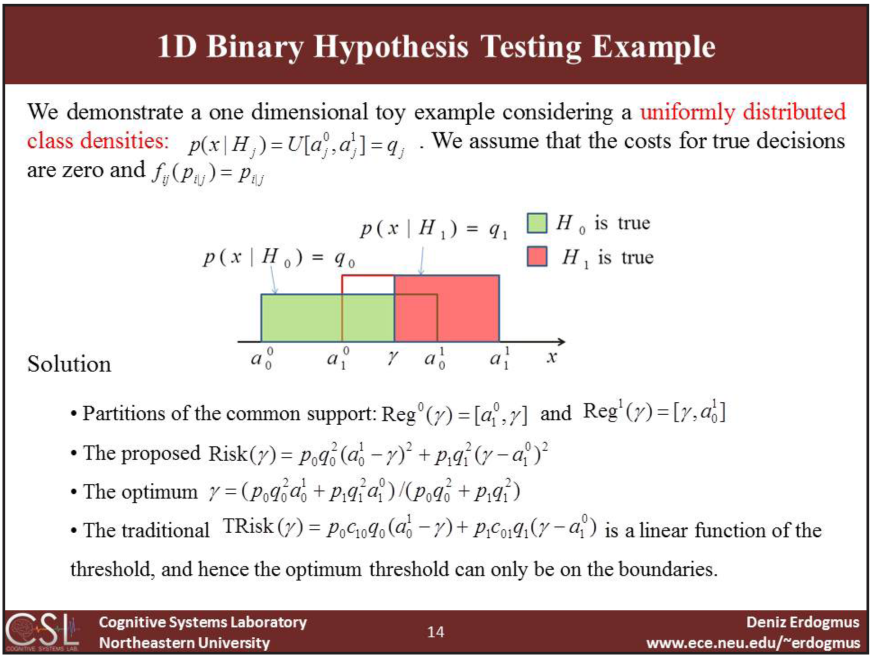
The proposed extended definition of the expected risk for an M-ary test is

$$\text{Risk}\ (Z) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} p_j f_{ij}(Ps\ ) p_{i|j}$$

where

- H_i is the i^{th} hypothesis
- p_i is the a priori probability of the i^{th} hypothesis
- $p_{i|j} = \Pr(\text{decide } H_i \mid H_j \text{ is true})$ is the probability of deciding H_i while H_j is true
- $Z = \bigcup_{i=0}^{M-1} Z_i$ is the observation space
- $Ps\ = \{Ps_i \mid i = 0,\ldots,M-1\}$ with $Ps_i = \{\Pr(\text{decide } H_i \mid H_j \text{ is true}) \mid j = 0,\ldots,M-1\}$

Note that if $f_{ij}(Ps) = c_{ij}$ is a constant above definition reduces to the traditional expected risk.



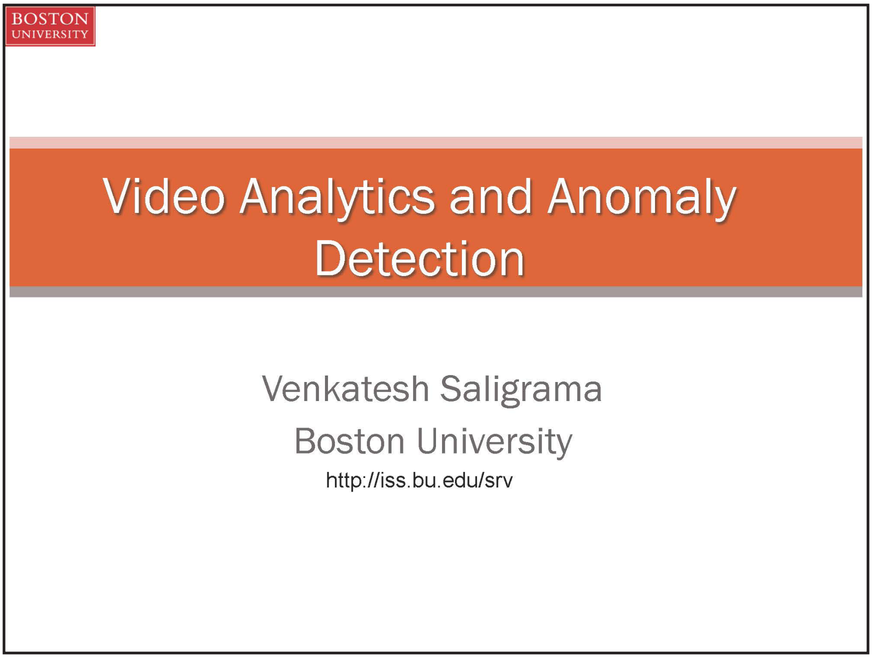
1D Binary Hypothesis Testing Example

We demonstrate a one dimensional toy example considering a uniformly distributed class densities: $p(x \mid H_j) = U[a_j^0, a_j^1] = q_j$. We assume that the costs for true decisions are zero and $f_{ij}(p_{i|j}) = p_{i|j}$

Solution

- Partitions of the common support: $\text{Reg}^0(\gamma) = [a_1^0, \gamma]$ and $\text{Reg}^1(\gamma) = [\gamma, a_0^1]$
- The proposed $\text{Risk}(\gamma) = p_0 q_0^2 (a_0^1 - \gamma)^2 + p_1 q_1^2 (\gamma - a_1^0)^2$
- The optimum $\gamma = (p_0 q_0^2 a_0^1 + p_1 q_1^2 a_1^0)/(p_0 q_0^2 + p_1 q_1^2)$
- The traditional $\text{TRisk}(\gamma) = p_0 c_{10} q_0 (a_0^1 - \gamma) + p_1 c_{01} q_1 (\gamma - a_1^0)$ is a linear function of the threshold, and hence the optimum threshold can only be on the boundaries.

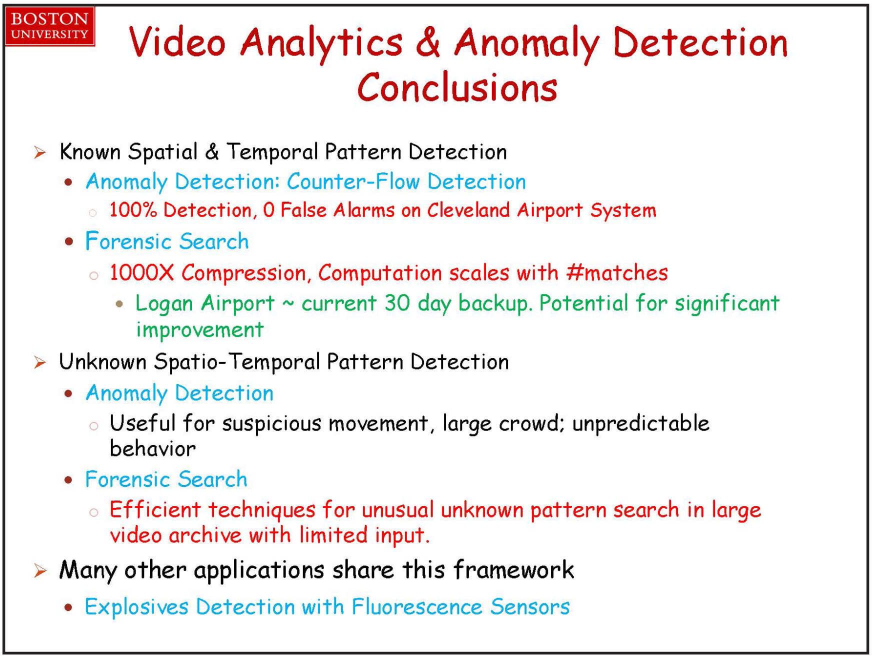
16.26 Venkatesh Saligrama: Video Analytics and Anomaly Detection



BOSTON UNIVERSITY

Video Analytics and Anomaly Detection

Venkatesh Saligrama
Boston University
http://iss.bu.edu/srv



BOSTON UNIVERSITY

Video Analytics & Anomaly Detection Conclusions

- Known Spatial & Temporal Pattern Detection
  - Anomaly Detection: Counter-Flow Detection
    - 100% Detection, 0 False Alarms on Cleveland Airport System
  - Forensic Search
    - 1000X Compression, Computation scales with #matches
      - Logan Airport ~ current 30 day backup. Potential for significant improvement
- Unknown Spatio-Temporal Pattern Detection
  - Anomaly Detection
    - Useful for suspicious movement, large crowd; unpredictable behavior
  - Forensic Search
    - Efficient techniques for unusual unknown pattern search in large video archive with limited input.
- Many other applications share this framework
  - Explosives Detection with Fluorescence Sensors

BOSTON UNIVERSITY

Autonomous Video Surveillance

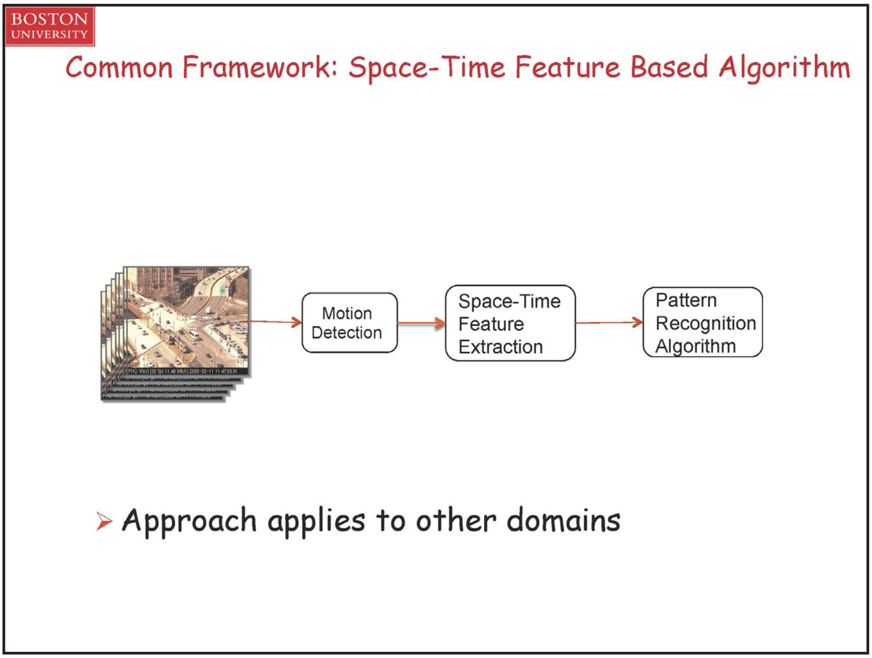
Networked Cameras are Everywhere

Information Overload

How to help this guy?

Computation Overload:
Most video footage stored but rarely analyzed.

Storage Overload
At Logan Airport: about 30 days of footage stored

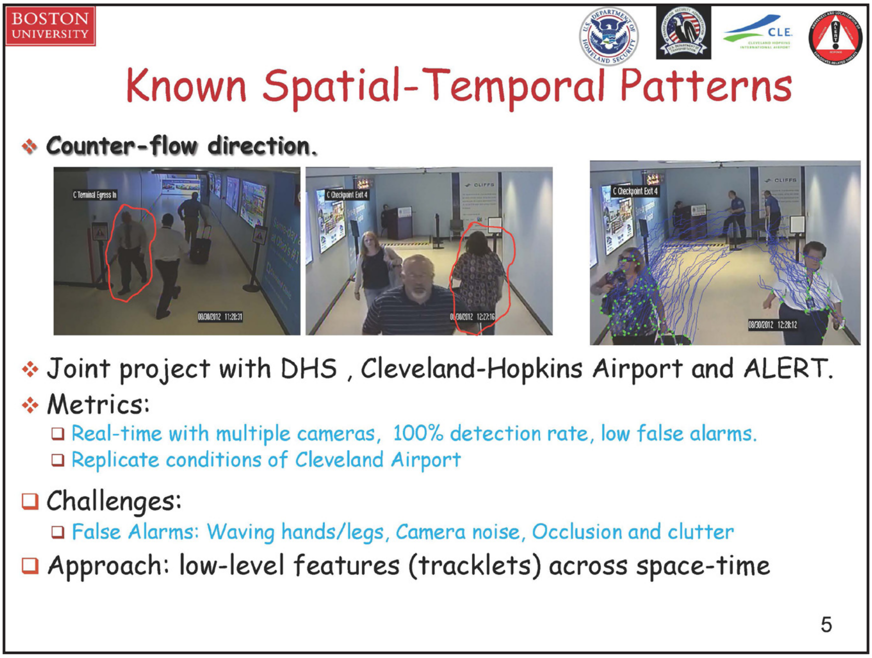


BOSTON UNIVERSITY

Common Framework: Space-Time Feature Based Algorithm

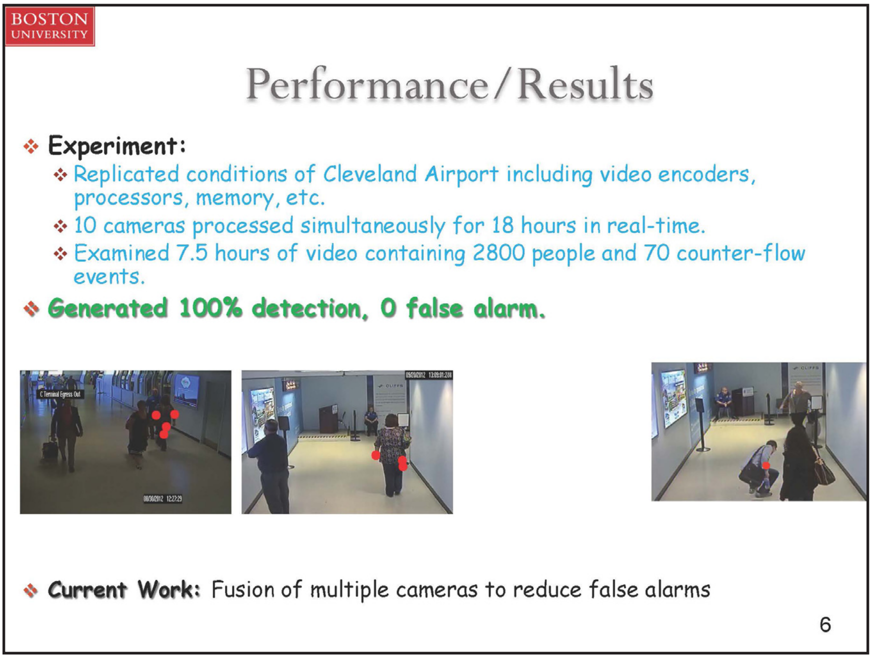
Motion Detection → Space-Time Feature Extraction → Pattern Recognition Algorithm

- Approach applies to other domains



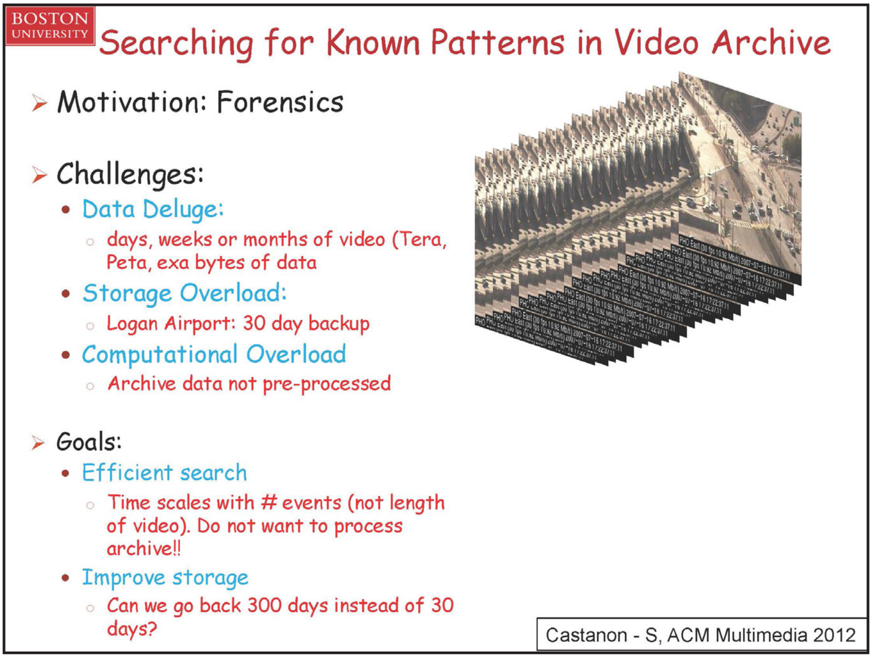
Known Spatial-Temporal Patterns

- **Counter-flow direction.**
- Joint project with DHS , Cleveland-Hopkins Airport and ALERT.
- Metrics:
  - ☐ Real-time with multiple cameras, 100% detection rate, low false alarms.
  - ☐ Replicate conditions of Cleveland Airport
- ☐ Challenges:
  - ☐ False Alarms: Waving hands/legs, Camera noise, Occlusion and clutter
- ☐ Approach: low-level features (tracklets) across space-time



Performance/Results

- **Experiment:**
  - Replicated conditions of Cleveland Airport including video encoders, processors, memory, etc.
  - 10 cameras processed simultaneously for 18 hours in real-time.
  - Examined 7.5 hours of video containing 2800 people and 70 counter-flow events.
- **Generated 100% detection, 0 false alarm.**
- **Current Work:** Fusion of multiple cameras to reduce false alarms

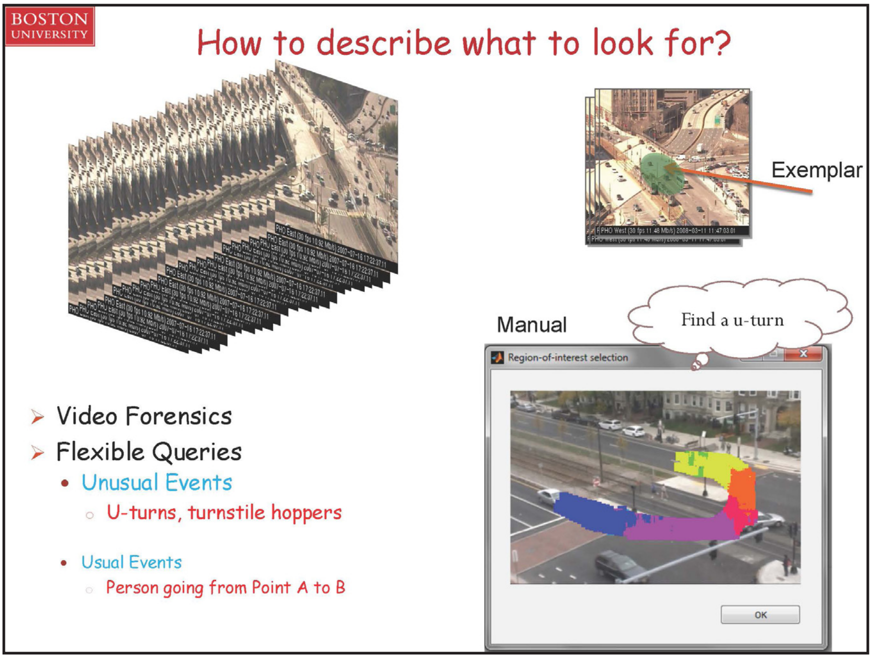


BOSTON UNIVERSITY

Searching for Known Patterns in Video Archive

- Motivation: Forensics
- Challenges:
  - Data Deluge:
    - days, weeks or months of video (Tera, Peta, exa bytes of data
  - Storage Overload:
    - Logan Airport: 30 day backup
  - Computational Overload
    - Archive data not pre-processed
- Goals:
  - Efficient search
    - Time scales with # events (not length of video). Do not want to process archive!!
  - Improve storage
    - Can we go back 300 days instead of 30 days?

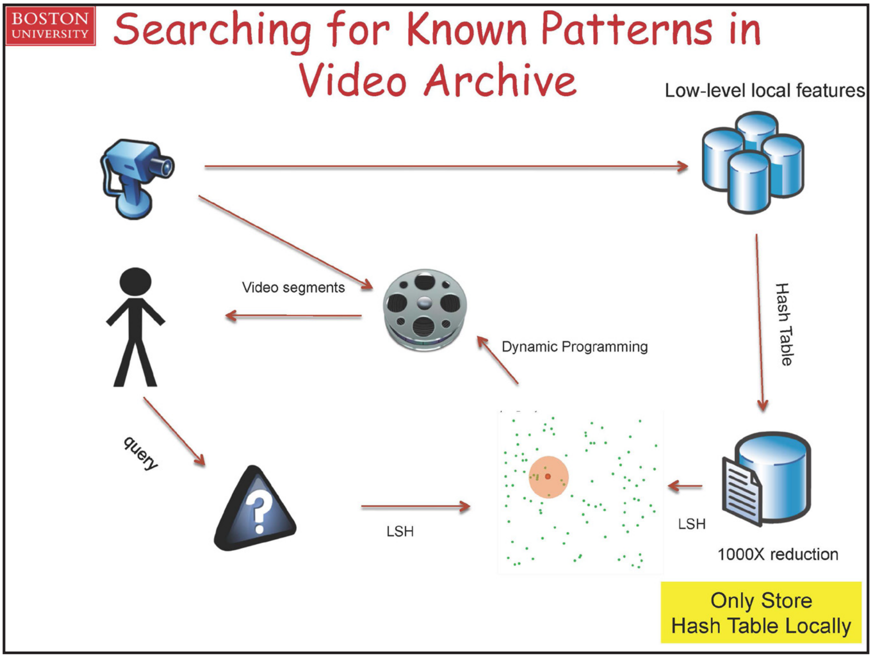
Castanon - S, ACM Multimedia 2012



BOSTON UNIVERSITY

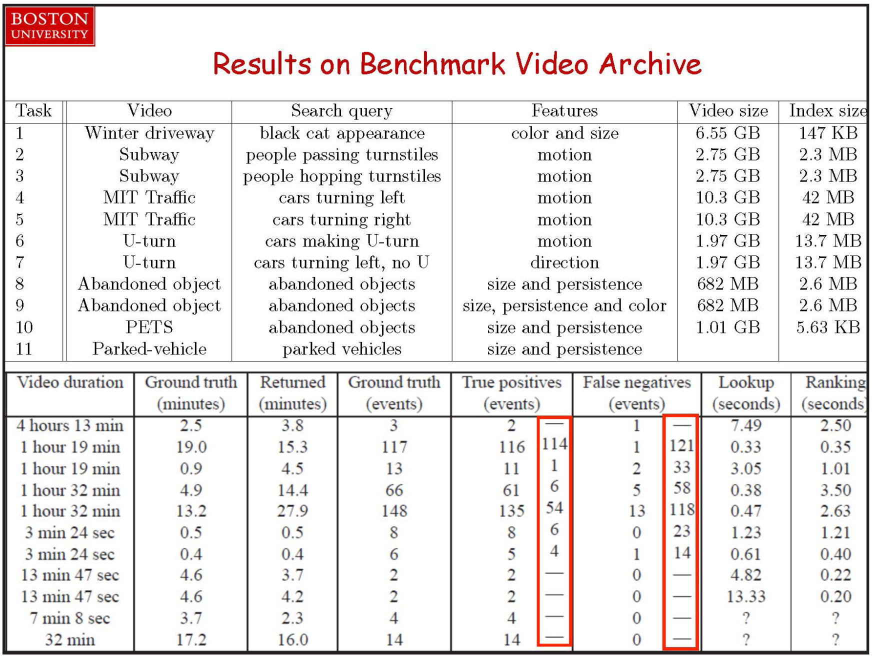
How to describe what to look for?

- Video Forensics
- Flexible Queries
  - Unusual Events
    - U-turns, turnstile hoppers
  - Usual Events
    - Person going from Point A to B



Searching for Known Patterns in Video Archive

Only Store
Hash Table Locally



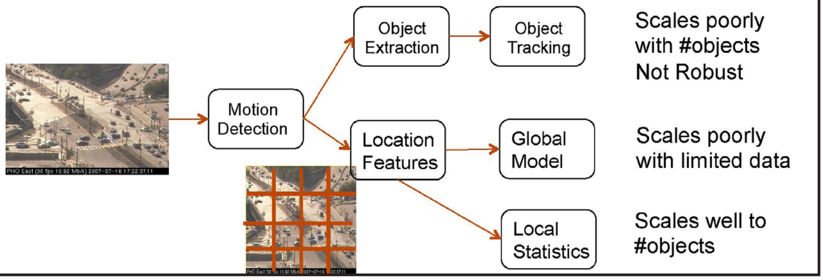
Results on Benchmark Video Archive

Task	Video	Search query	Features	Video size	Index size
1	Winter driveway	black cat appearance	color and size	6.55 GB	147 KB
2	Subway	people passing turnstiles	motion	2.75 GB	2.3 MB
3	Subway	people hopping turnstiles	motion	2.75 GB	2.3 MB
4	MIT Traffic	cars turning left	motion	10.3 GB	42 MB
5	MIT Traffic	cars turning right	motion	10.3 GB	42 MB
6	U-turn	cars making U-turn	motion	1.97 GB	13.7 MB
7	U-turn	cars turning left, no U	direction	1.97 GB	13.7 MB
8	Abandoned object	abandoned objects	size and persistence	682 MB	2.6 MB
9	Abandoned object	abandoned objects	size, persistence and color	682 MB	2.6 MB
10	PETS	abandoned objects	size and persistence	1.01 GB	5.63 KB
11	Parked-vehicle	parked vehicles	size and persistence		

Video duration	Ground truth (minutes)	Returned (minutes)	Ground truth (events)	True positives (events)		False negatives (events)		Lookup (seconds)	Ranking (seconds
4 hours 13 min	2.5	3.8	3	2	—	1	—	7.49	2.50
1 hour 19 min	19.0	15.3	117	116	114	1	121	0.33	0.35
1 hour 19 min	0.9	4.5	13	11	1	2	33	3.05	1.01
1 hour 32 min	4.9	14.4	66	61	6	5	58	0.38	3.50
1 hour 32 min	13.2	27.9	148	135	54	13	118	0.47	2.63
3 min 24 sec	0.5	0.5	8	8	6	0	23	1.23	1.21
3 min 24 sec	0.4	0.4	6	5	4	1	14	0.61	0.40
13 min 47 sec	4.6	3.7	2	2	—	0	—	4.82	0.22
13 min 47 sec	4.6	4.2	2	2	—	0	—	13.33	0.20
7 min 8 sec	3.7	2.3	4	4	—	0	—	?	?
32 min	17.2	16.0	14	14	—	0	—	?	?

Detecting Unknown Unusual Patterns

- Location-based attributes
  - highly correlated in space-time-feature space
- Learn Global Joint Space-time-Feature Model
  - Topic Modeling, MRFs, Mixture models, Sparse Dictionaries (Wang'10,Kim'09,Mahadevan'11, Cong'11, …)
- Main Drawback with Global Models
  - Nominal behavior is too complicated



Motion Detection

Object Extraction

Object Tracking

Scales poorly with #objects
Not Robust

Location Features

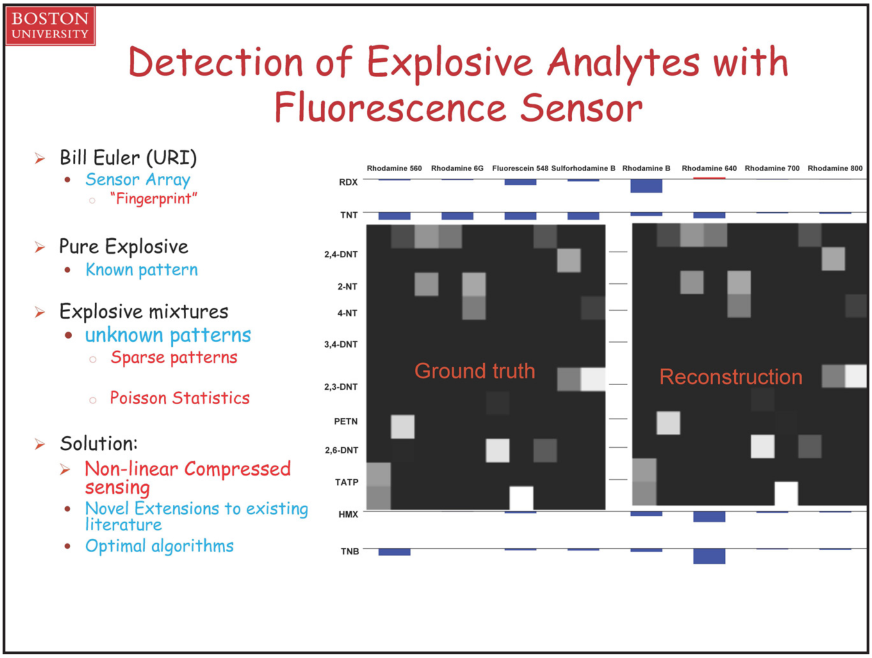
Global Model

Scales poorly with limited data

Local Statistics

Scales well to #objects

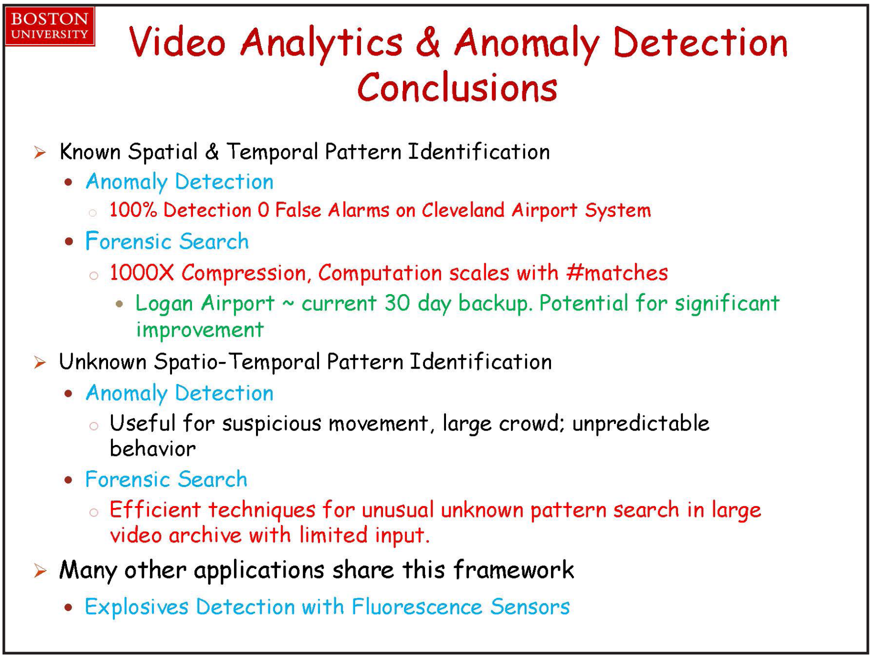
Video Anomaly Detection
Based on Local Statistical Aggregates



BOSTON UNIVERSITY

Detection of Explosive Analytes with Fluorescence Sensor

- Bill Euler (URI)
  - Sensor Array
    - "Fingerprint"
- Pure Explosive
  - Known pattern
- Explosive mixtures
  - unknown patterns
    - Sparse patterns
    - Poisson Statistics
- Solution:
  - Non-linear Compressed sensing
  - Novel Extensions to existing literature
  - Optimal algorithms



BOSTON UNIVERSITY

Video Analytics & Anomaly Detection Conclusions

- Known Spatial & Temporal Pattern Identification
  - Anomaly Detection
    - 100% Detection 0 False Alarms on Cleveland Airport System
  - Forensic Search
    - 1000X Compression, Computation scales with #matches
      - Logan Airport ~ current 30 day backup. Potential for significant improvement
- Unknown Spatio-Temporal Pattern Identification
  - Anomaly Detection
    - Useful for suspicious movement, large crowd; unpredictable behavior
  - Forensic Search
    - Efficient techniques for unusual unknown pattern search in large video archive with limited input.
- Many other applications share this framework
  - Explosives Detection with Fluorescence Sensors

16.27 Brian Tracey: Imaging Challenges for X-Ray Screening

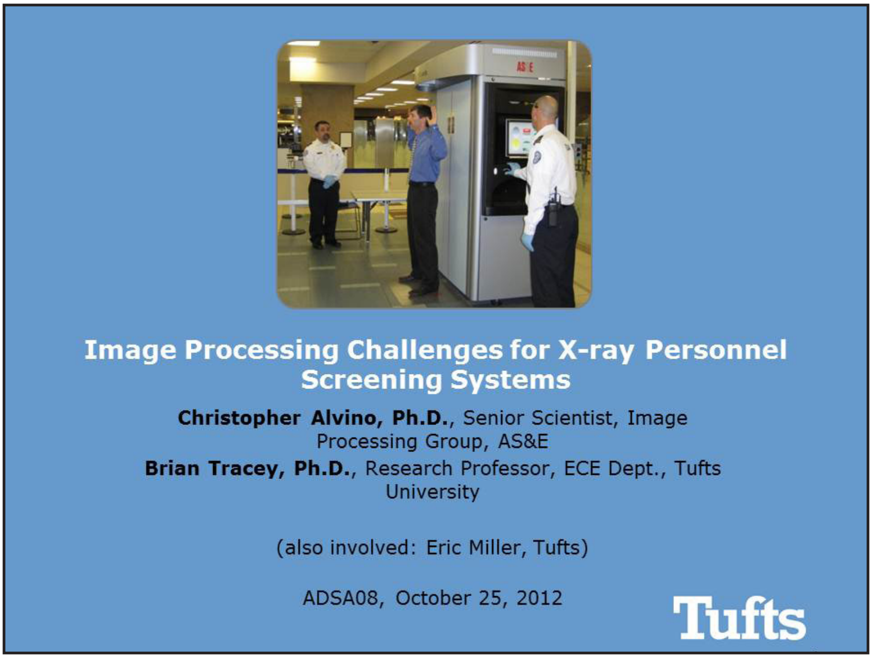


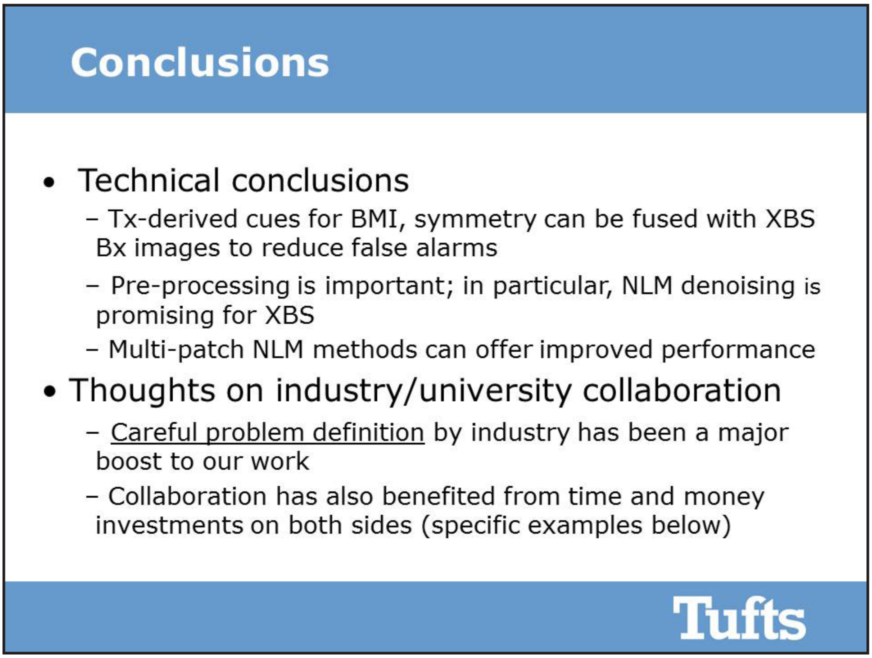
Image Processing Challenges for X-ray Personnel Screening Systems

Christopher Alvino, Ph.D., Senior Scientist, Image Processing Group, AS&E

Brian Tracey, Ph.D., Research Professor, ECE Dept., Tufts University

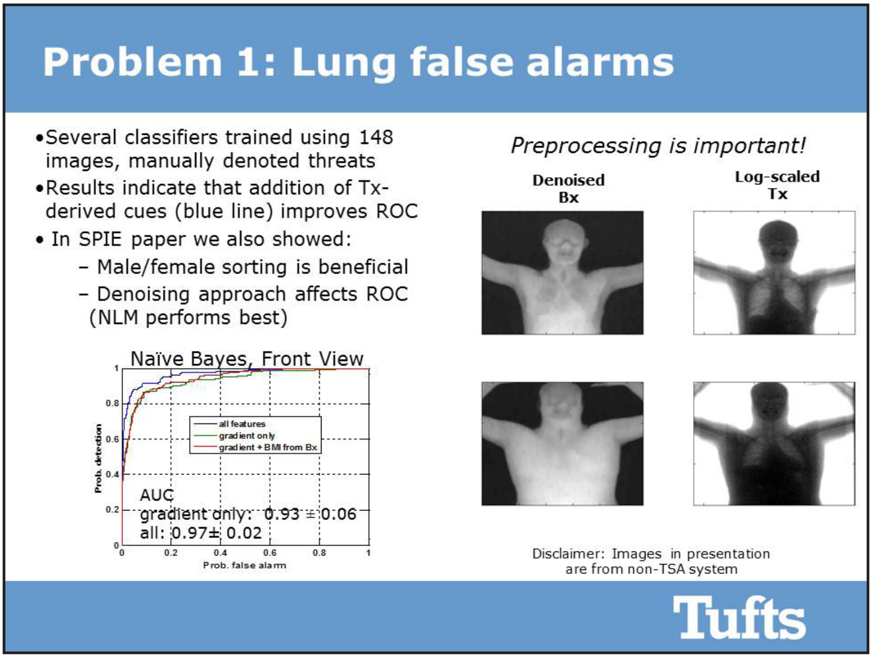
(also involved: Eric Miller, Tufts)

ADSA08, October 25, 2012



Conclusions

- Technical conclusions
  - Tx-derived cues for BMI, symmetry can be fused with XBS Bx images to reduce false alarms
  - Pre-processing is important; in particular, NLM denoising is promising for XBS
  - Multi-patch NLM methods can offer improved performance
- Thoughts on industry/university collaboration
  - Careful problem definition by industry has been a major boost to our work
  - Collaboration has also benefited from time and money investments on both sides (specific examples below)



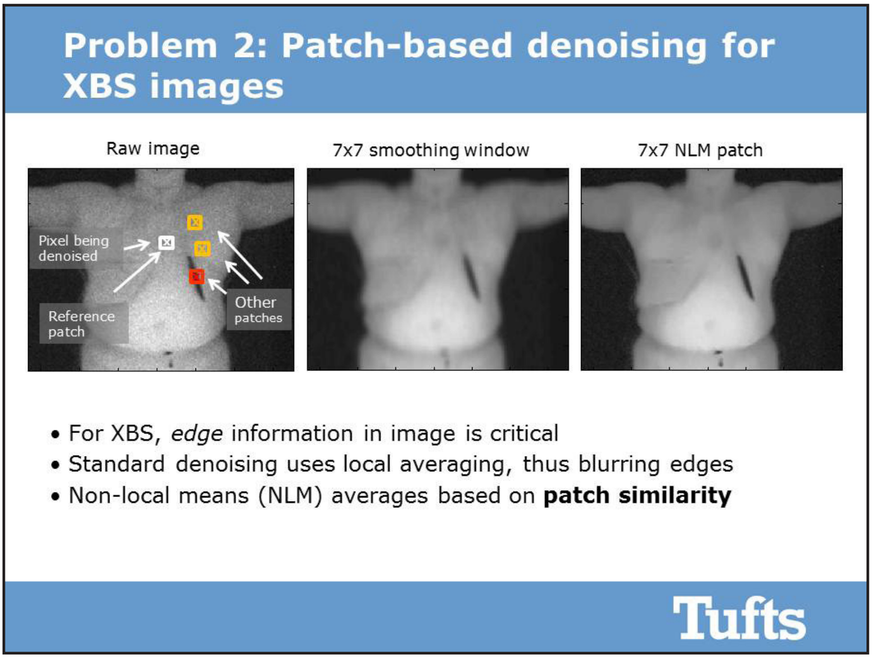
Problem 1: Lung false alarms

- Several classifiers trained using 148 images, manually denoted threats
- Results indicate that addition of Tx-derived cues (blue line) improves ROC
- In SPIE paper we also showed:
  - Male/female sorting is beneficial
  - Denoising approach affects ROC (NLM performs best)

Preprocessing is important!

Disclaimer: Images in presentation are from non-TSA system

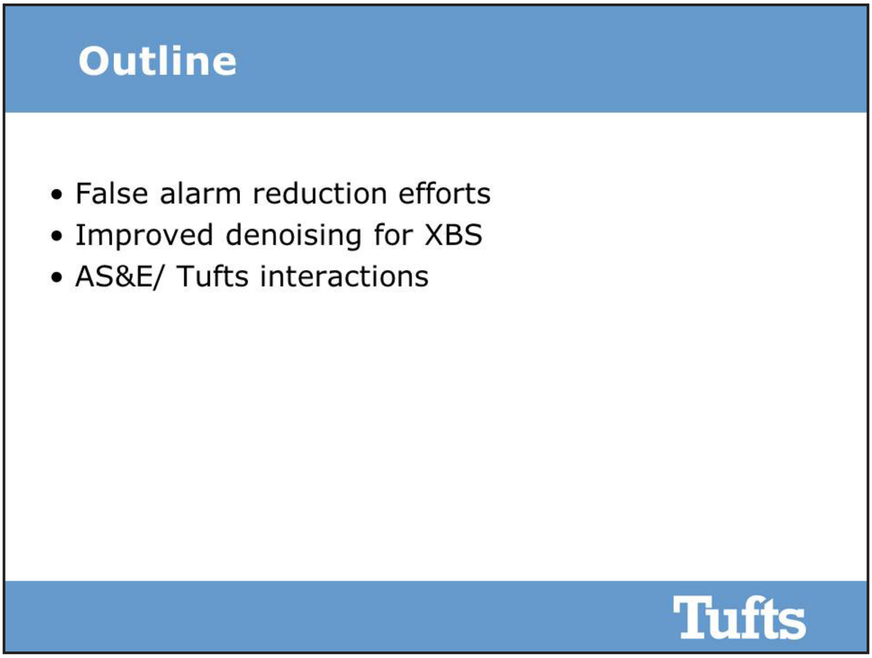
Tufts



Problem 2: Patch-based denoising for XBS images

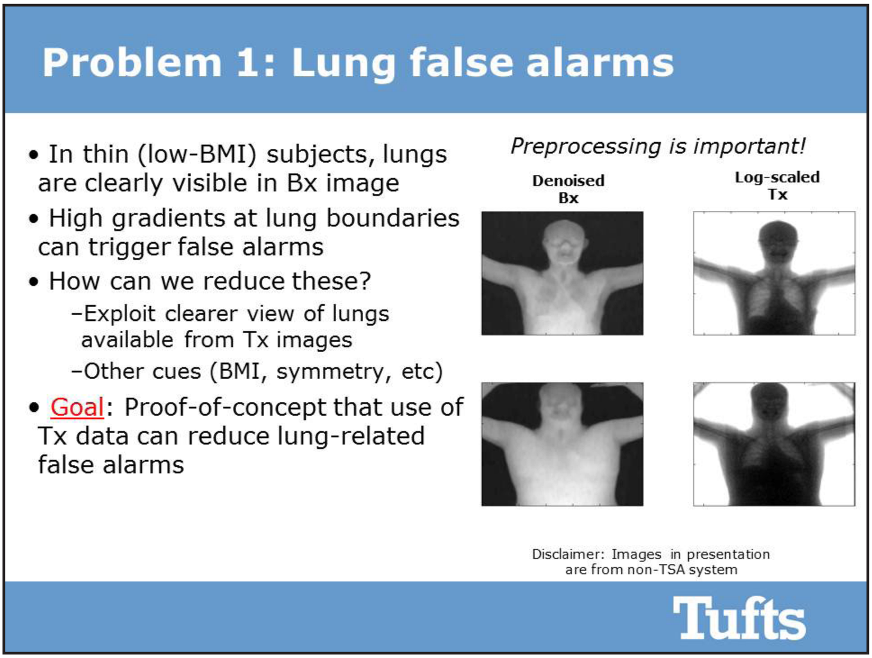
- For XBS, *edge* information in image is critical
- Standard denoising uses local averaging, thus blurring edges
- Non-local means (NLM) averages based on **patch similarity**

Tufts



Outline

- False alarm reduction efforts
- Improved denoising for XBS
- AS&E/ Tufts interactions



Problem 1: Lung false alarms

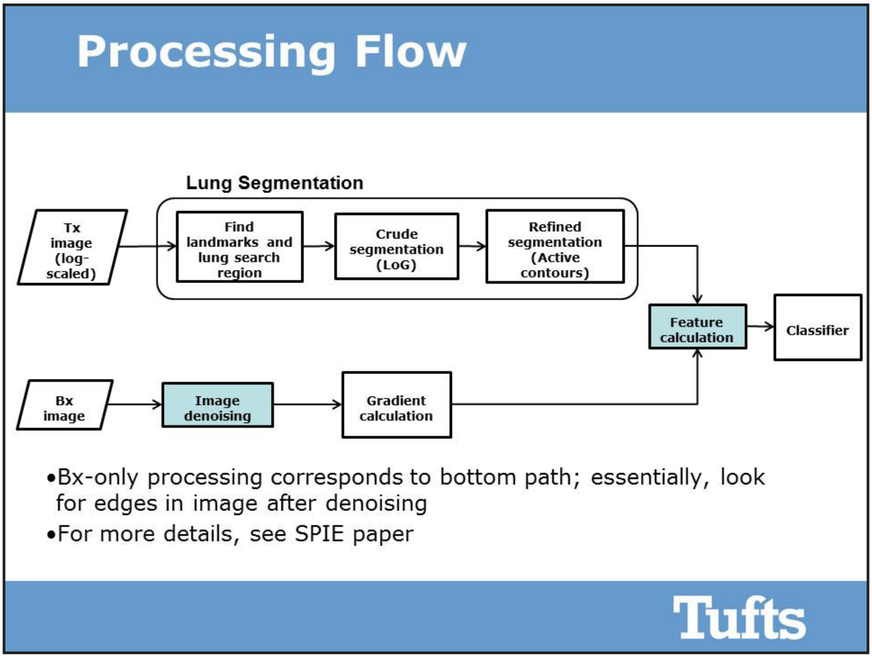
- In thin (low-BMI) subjects, lungs are clearly visible in Bx image
- High gradients at lung boundaries can trigger false alarms
- How can we reduce these?
  - –Exploit clearer view of lungs available from Tx images
  - –Other cues (BMI, symmetry, etc)
- Goal: Proof-of-concept that use of Tx data can reduce lung-related false alarms

Preprocessing is important!

Denoised Bx

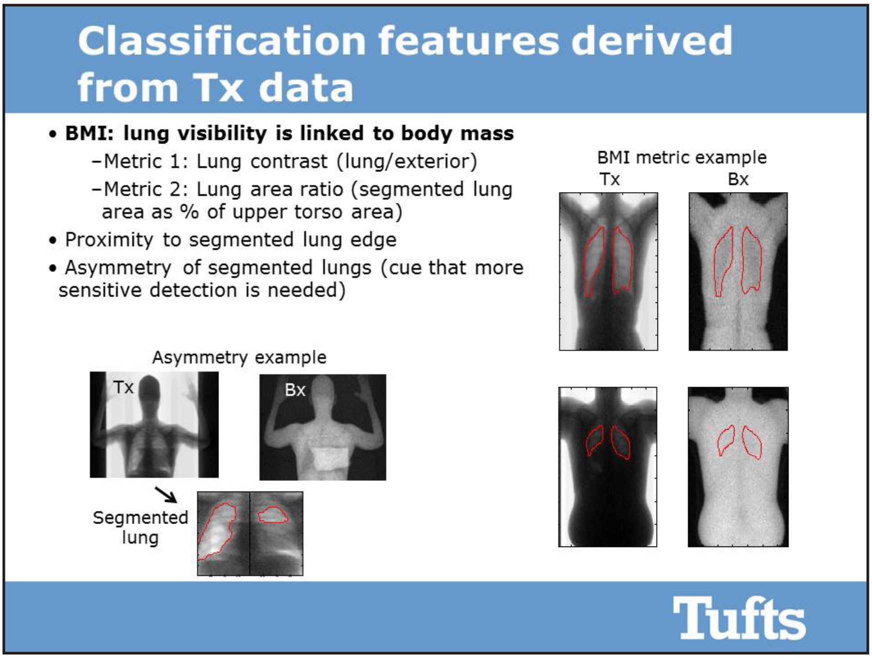
Log-scaled Tx

Disclaimer: Images in presentation are from non-TSA system



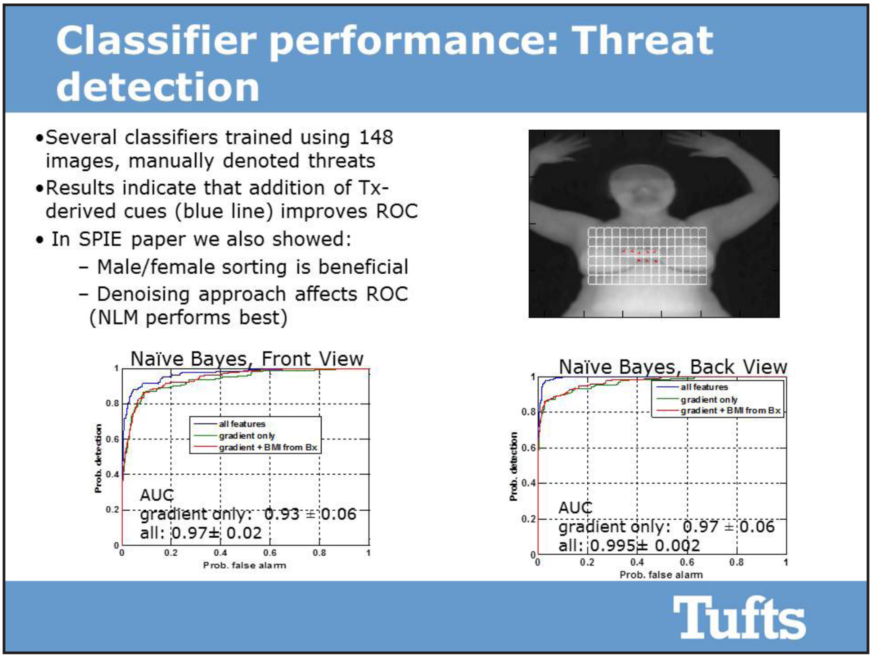
Processing Flow

- Bx-only processing corresponds to bottom path; essentially, look for edges in image after denoising
- For more details, see SPIE paper



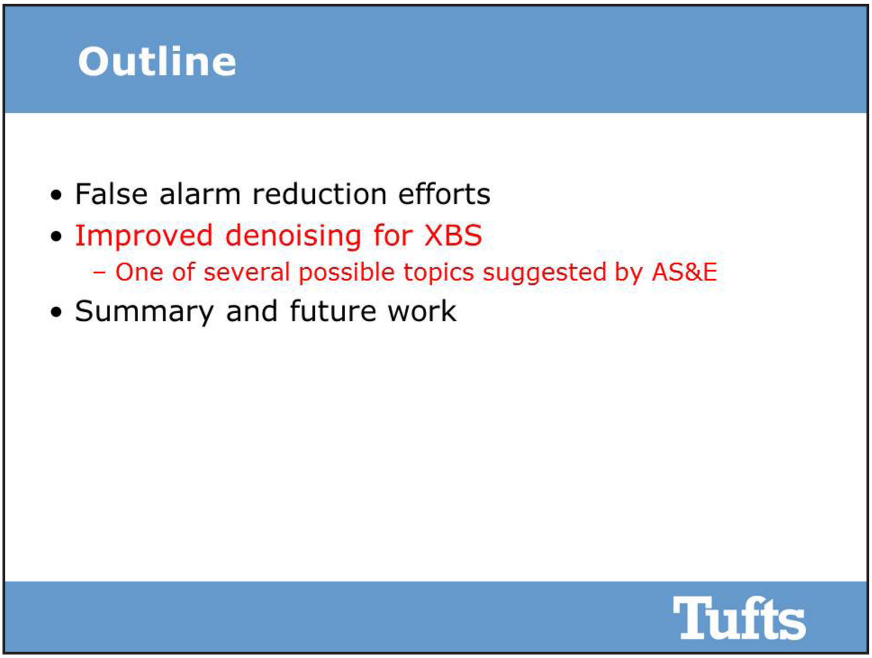
Classification features derived from Tx data

- **BMI: lung visibility is linked to body mass**
  - –Metric 1: Lung contrast (lung/exterior)
  - –Metric 2: Lung area ratio (segmented lung area as % of upper torso area)
- Proximity to segmented lung edge
- Asymmetry of segmented lungs (cue that more sensitive detection is needed)



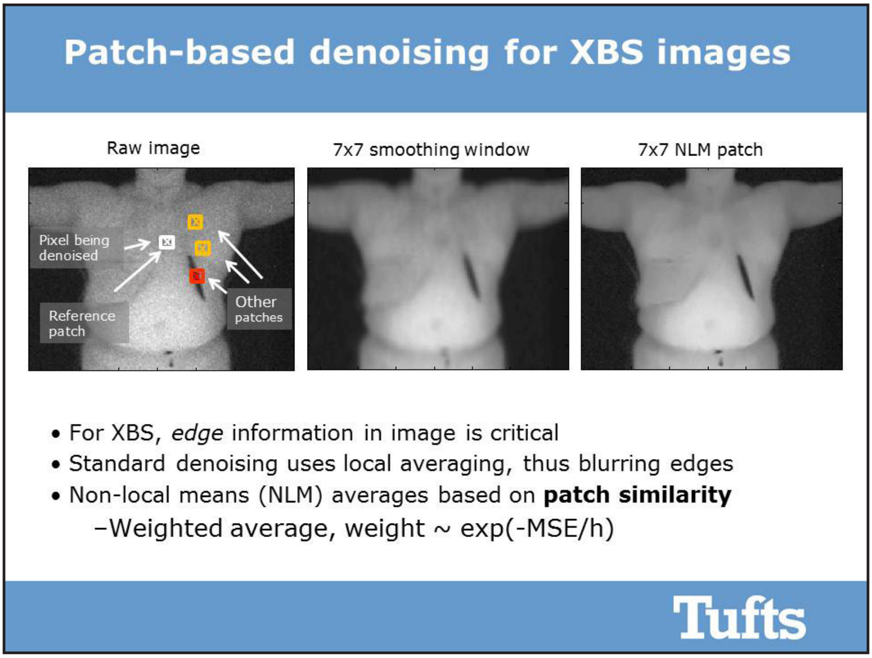
Classifier performance: Threat detection

- Several classifiers trained using 148 images, manually denoted threats
- Results indicate that addition of Tx-derived cues (blue line) improves ROC
- In SPIE paper we also showed:
  - Male/female sorting is beneficial
  - Denoising approach affects ROC (NLM performs best)



Outline

- False alarm reduction efforts
- Improved denoising for XBS
  - One of several possible topics suggested by AS&E
- Summary and future work



Patch-based denoising for XBS images

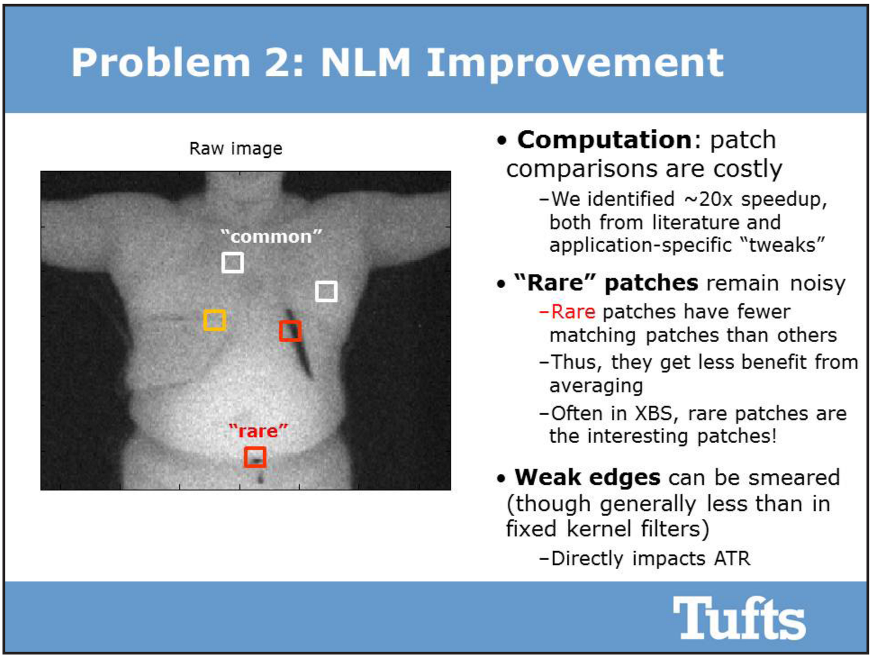
Raw image

7x7 smoothing window

7x7 NLM patch

- For XBS, *edge* information in image is critical
- Standard denoising uses local averaging, thus blurring edges
- Non-local means (NLM) averages based on **patch similarity**
  - –Weighted average, weight ~ exp(-MSE/h)

Tufts

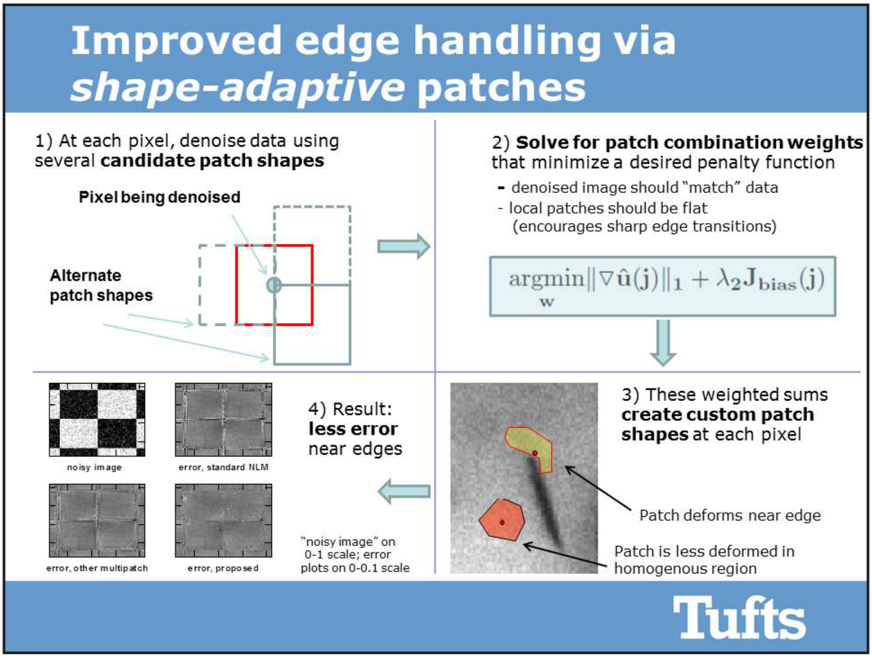


Problem 2: NLM Improvement

Raw image

- **Computation**: patch comparisons are costly
  - –We identified ~20x speedup, both from literature and application-specific "tweaks"
- **"Rare" patches** remain noisy
  - –Rare patches have fewer matching patches than others
  - –Thus, they get less benefit from averaging
  - –Often in XBS, rare patches are the interesting patches!
- **Weak edges** can be smeared (though generally less than in fixed kernel filters)
  - –Directly impacts ATR

Tufts



Improved edge handling via *shape-adaptive* patches

1) At each pixel, denoise data using several **candidate patch shapes**

2) **Solve for patch combination weights** that minimize a desired penalty function

- denoised image should "match" data
- local patches should be flat (encourages sharp edge transitions)

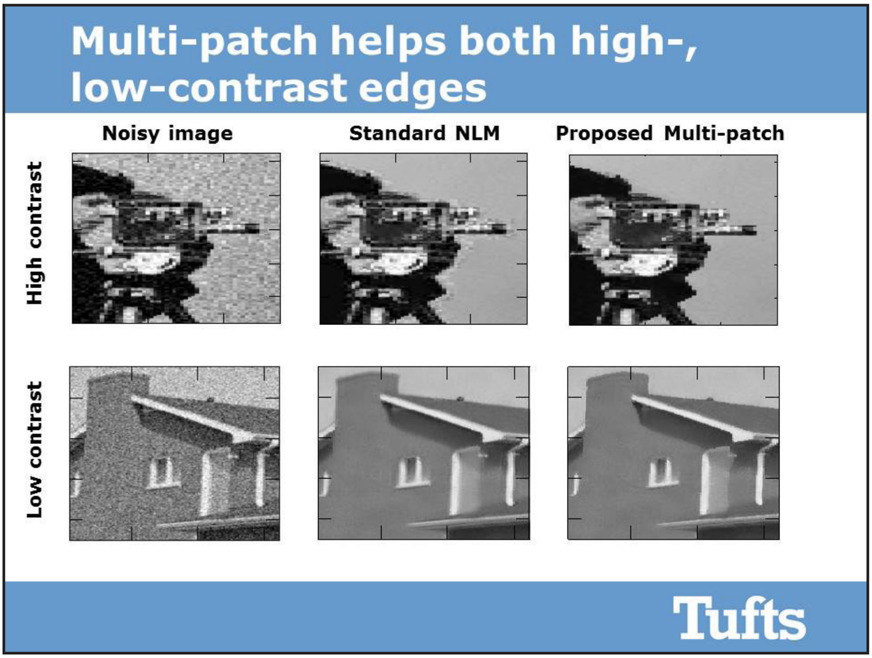
$$\underset{w}{\operatorname{argmin}} \|\nabla \hat{\mathbf{u}}(\mathbf{j})\|_1 + \lambda_2 \mathbf{J}_{\mathbf{bias}}(\mathbf{j})$$

3) These weighted sums **create custom patch shapes** at each pixel

4) Result: **less error** near edges

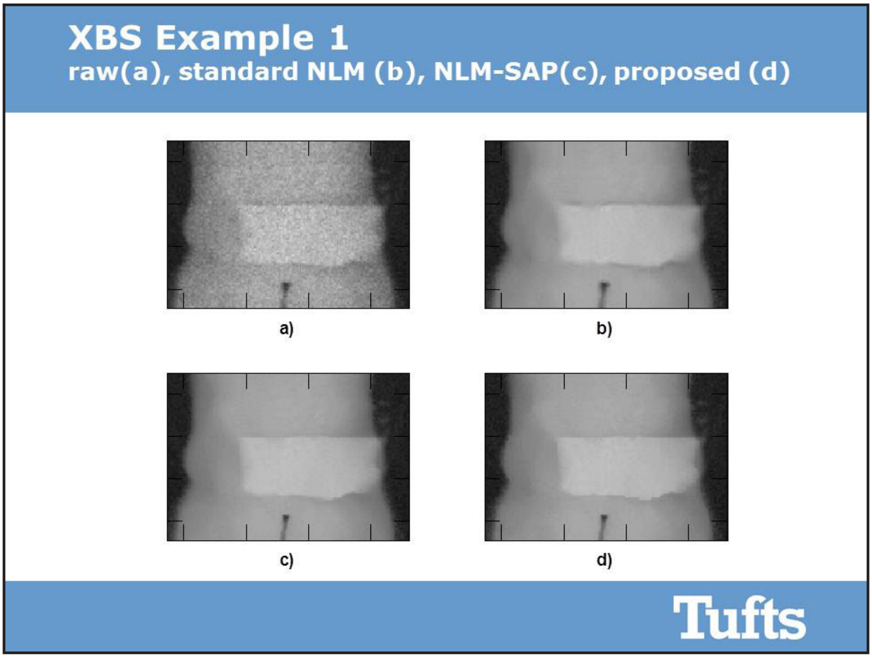
"noisy image" on 0-1 scale; error plots on 0-0.1 scale

Tufts



Multi-patch helps both high-, low-contrast edges

Tufts



XBS Example 1

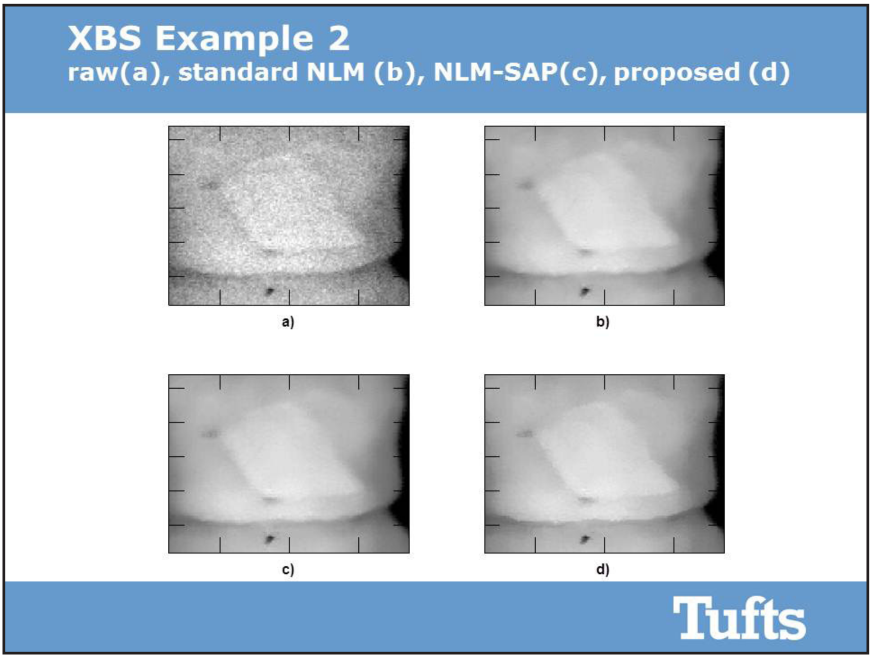
raw(a), standard NLM (b), NLM-SAP(c), proposed (d)

a)

b)

c)

d)



XBS Example 2

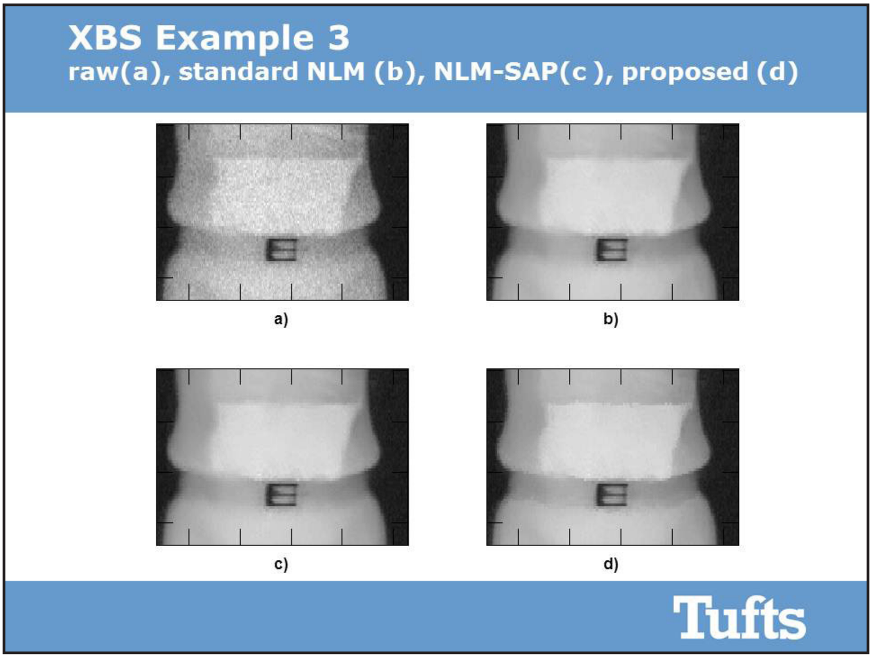
raw(a), standard NLM (b), NLM-SAP(c), proposed (d)

a)

b)

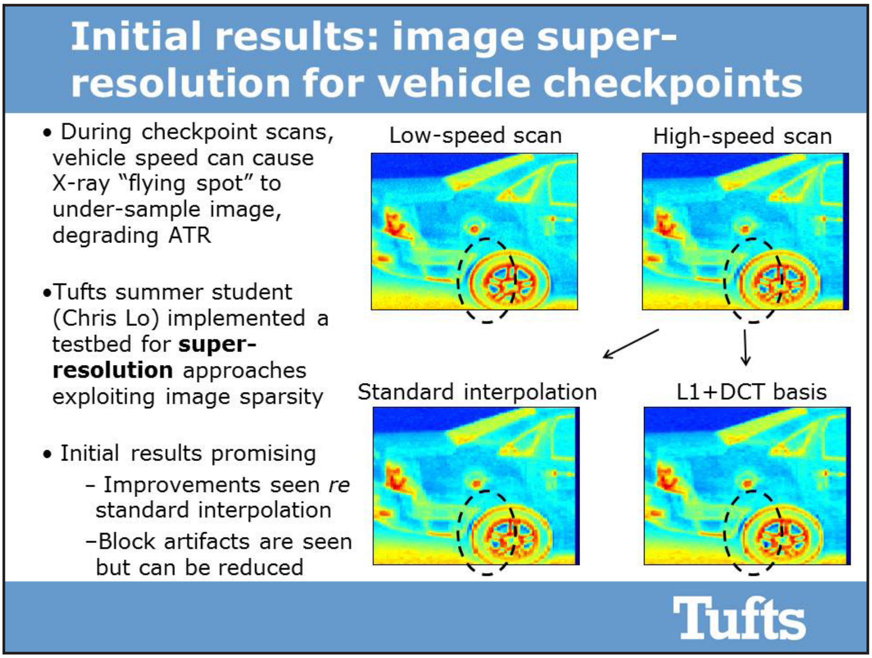
c)

d)



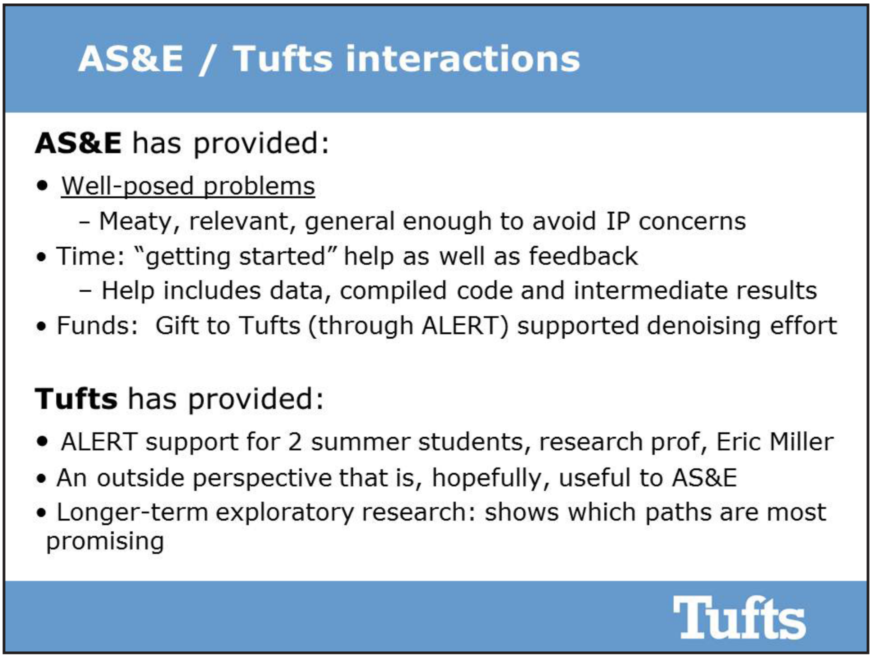
XBS Example 3

raw(a), standard NLM (b), NLM-SAP(c), proposed (d)



Initial results: image super-resolution for vehicle checkpoints

- During checkpoint scans, vehicle speed can cause X-ray "flying spot" to under-sample image, degrading ATR
- Tufts summer student (Chris Lo) implemented a testbed for **super-resolution** approaches exploiting image sparsity
- Initial results promising
  - Improvements seen *re* standard interpolation
  - Block artifacts are seen but can be reduced



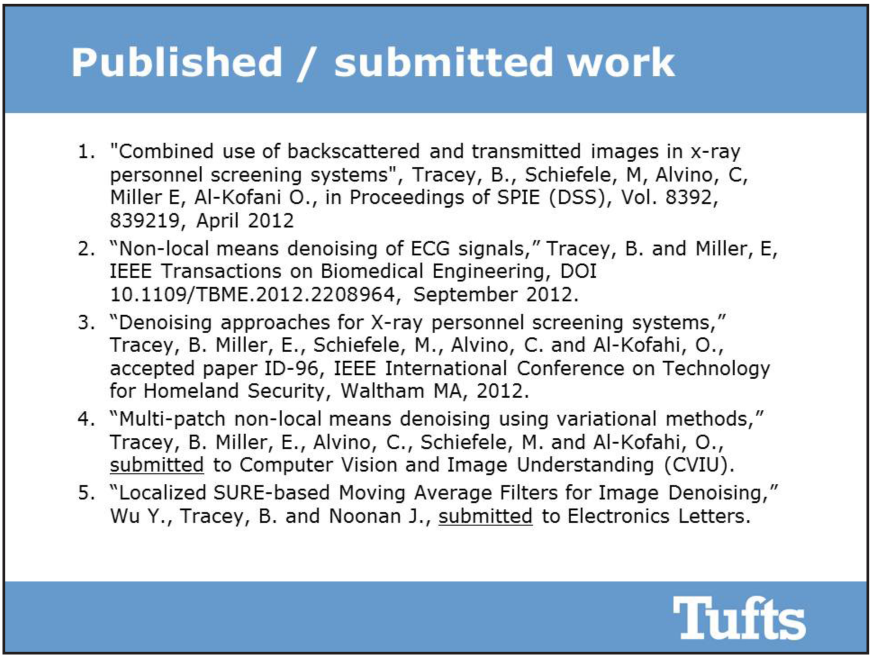
AS&E / Tufts interactions

AS&E has provided:

- Well-posed problems
  - Meaty, relevant, general enough to avoid IP concerns
- Time: "getting started" help as well as feedback
  - Help includes data, compiled code and intermediate results
- Funds: Gift to Tufts (through ALERT) supported denoising effort

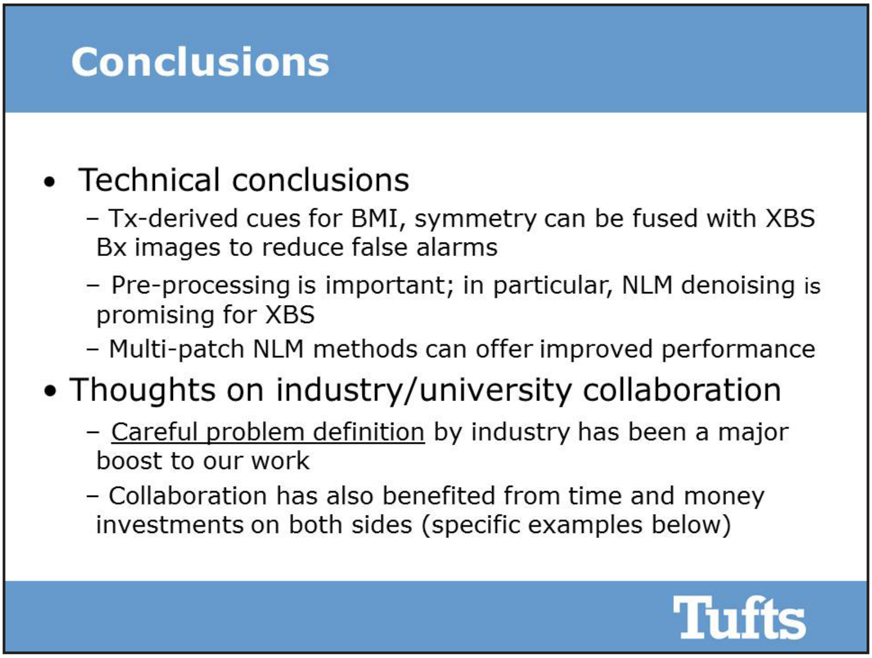
Tufts has provided:

- ALERT support for 2 summer students, research prof, Eric Miller
- An outside perspective that is, hopefully, useful to AS&E
- Longer-term exploratory research: shows which paths are most promising



Published / submitted work

1. "Combined use of backscattered and transmitted images in x-ray personnel screening systems", Tracey, B., Schiefele, M, Alvino, C, Miller E, Al-Kofani O., in Proceedings of SPIE (DSS), Vol. 8392, 839219, April 2012
2. "Non-local means denoising of ECG signals," Tracey, B. and Miller, E, IEEE Transactions on Biomedical Engineering, DOI 10.1109/TBME.2012.2208964, September 2012.
3. "Denoising approaches for X-ray personnel screening systems," Tracey, B. Miller, E., Schiefele, M., Alvino, C. and Al-Kofahi, O., accepted paper ID-96, IEEE International Conference on Technology for Homeland Security, Waltham MA, 2012.
4. "Multi-patch non-local means denoising using variational methods," Tracey, B. Miller, E., Alvino, C., Schiefele, M. and Al-Kofahi, O., submitted to Computer Vision and Image Understanding (CVIU).
5. "Localized SURE-based Moving Average Filters for Image Denoising," Wu Y., Tracey, B. and Noonan J., submitted to Electronics Letters.



Conclusions

- Technical conclusions
  - – Tx-derived cues for BMI, symmetry can be fused with XBS Bx images to reduce false alarms
  - – Pre-processing is important; in particular, NLM denoising is promising for XBS
  - – Multi-patch NLM methods can offer improved performance
- Thoughts on industry/university collaboration
  - – Careful problem definition by industry has been a major boost to our work
  - – Collaboration has also benefited from time and money investments on both sides (specific examples below)

Tufts

16.28 Chris Alvino: Imaging Challenges for X-Ray Screening

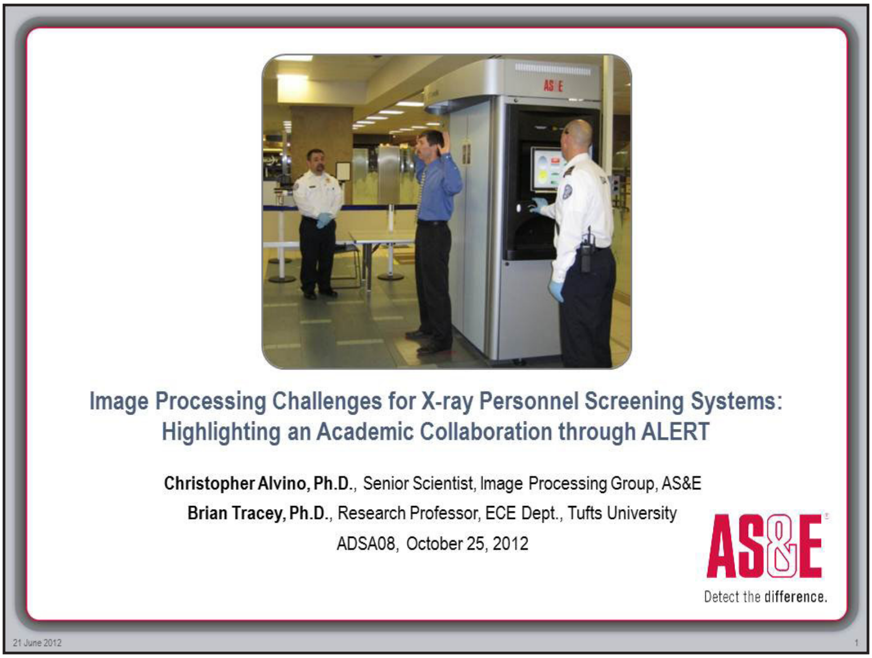


Image Processing Challenges for X-ray Personnel Screening Systems: Highlighting an Academic Collaboration through ALERT

Christopher Alvino, Ph.D., Senior Scientist, Image Processing Group, AS&E
Brian Tracey, Ph.D., Research Professor, ECE Dept., Tufts University
ADSA08, October 25, 2012

AS&E
Detect the difference.

Conclusion: Industrial-Academic Collaboration

AS&E

ATR for Personnel Screening (AIT) a Challenging Problem

Some problems very application specific

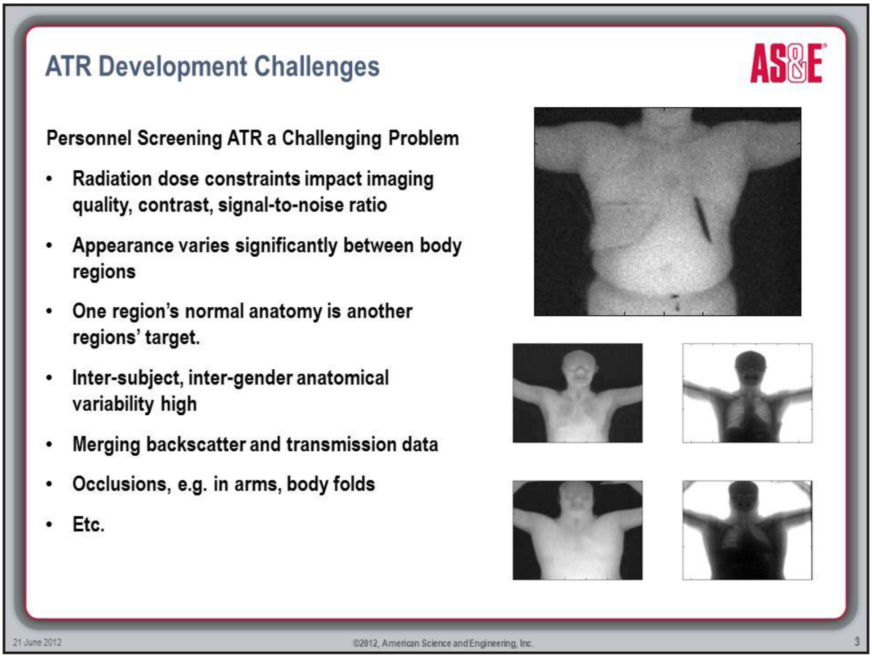
- Domain knowledge required
- Machine knowledge required

Many challenging problems are more general enough to be transferred

Want Effective, Efficient ATR Development

University Collaboration (ALERT) Key

- We benefit from sharing ideas, state-of-the-art academic knowledge and advanced prototypes
- University researchers benefit from motivating practical problems

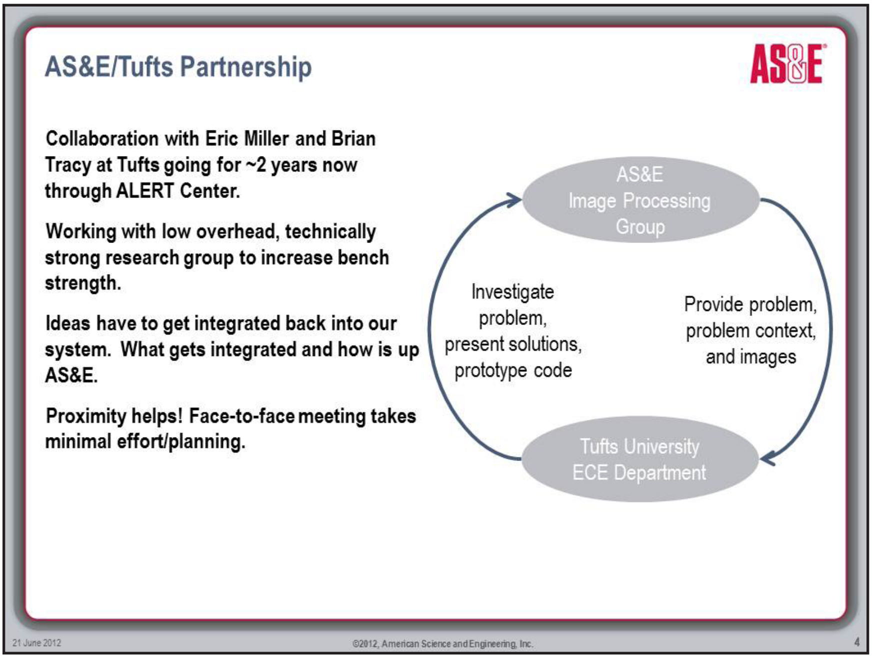


ATR Development Challenges

Personnel Screening ATR a Challenging Problem

- Radiation dose constraints impact imaging quality, contrast, signal-to-noise ratio
- Appearance varies significantly between body regions
- One region's normal anatomy is another regions' target.
- Inter-subject, inter-gender anatomical variability high
- Merging backscatter and transmission data
- Occlusions, e.g. in arms, body folds
- Etc.

21 June 2012



AS&E/Tufts Partnership

Collaboration with Eric Miller and Brian Tracy at Tufts going for ~2 years now through ALERT Center.

Working with low overhead, technically strong research group to increase bench strength.

Ideas have to get integrated back into our system. What gets integrated and how is up AS&E.

Proximity helps! Face-to-face meeting takes minimal effort/planning.

AS&E
Image Processing
Group

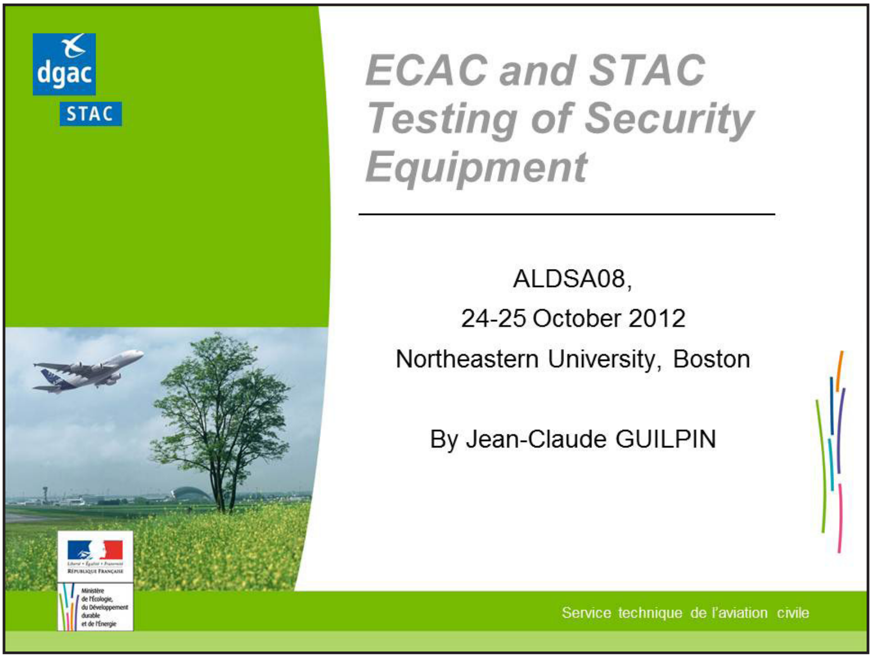
Provide problem,
problem context,
and images

Tufts University
ECE Department

Investigate
problem,
present solutions,
prototype code

21 June 2012

16.29 Jean Claude Guilpin: ECAC Testing



dgac
STAC

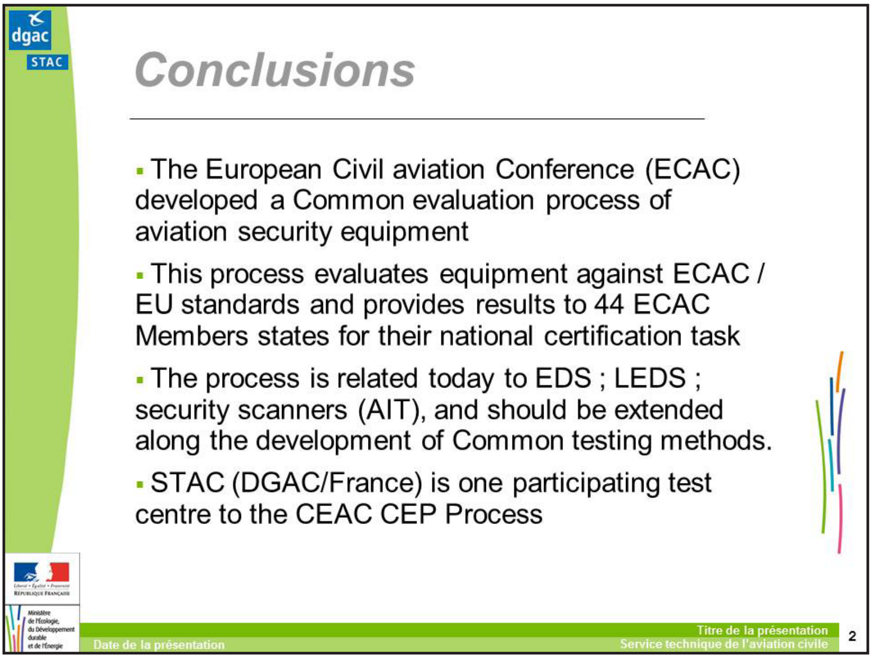
ECAC and STAC Testing of Security Equipment

ALDSA08,
24-25 October 2012
Northeastern University, Boston

By Jean-Claude GUILPIN

Ministère de l'Écologie, du Développement durable et de l'Énergie

Service technique de l'aviation civile



dgac
STAC

Conclusions

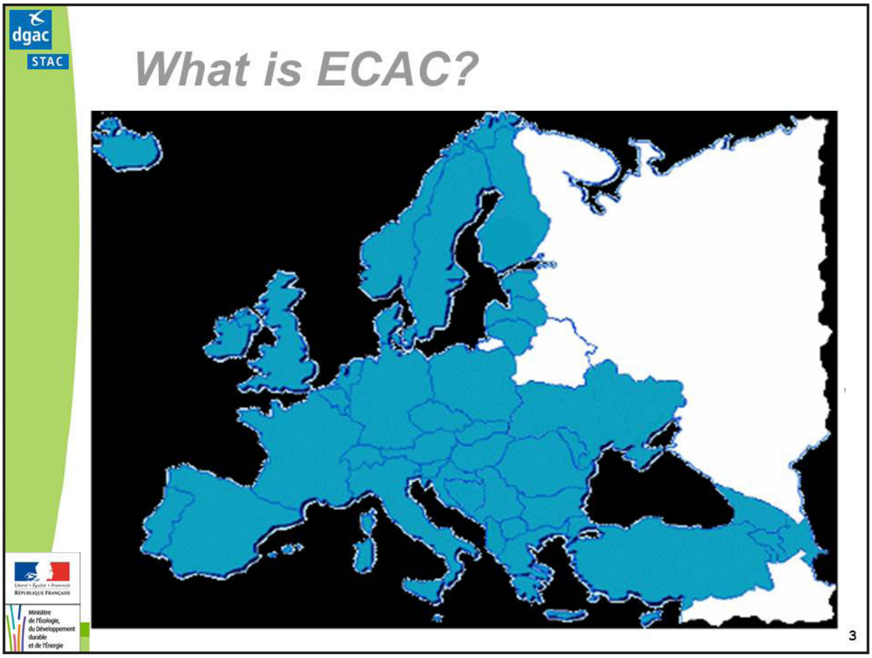
- The European Civil aviation Conference (ECAC) developed a Common evaluation process of aviation security equipment
- This process evaluates equipment against ECAC / EU standards and provides results to 44 ECAC Members states for their national certification task
- The process is related today to EDS ; LEDS ; security scanners (AIT), and should be extended along the development of Common testing methods.
- STAC (DGAC/France) is one participating test centre to the CEAC CEP Process

Ministère de l'Écologie, du Développement durable et de l'Énergie

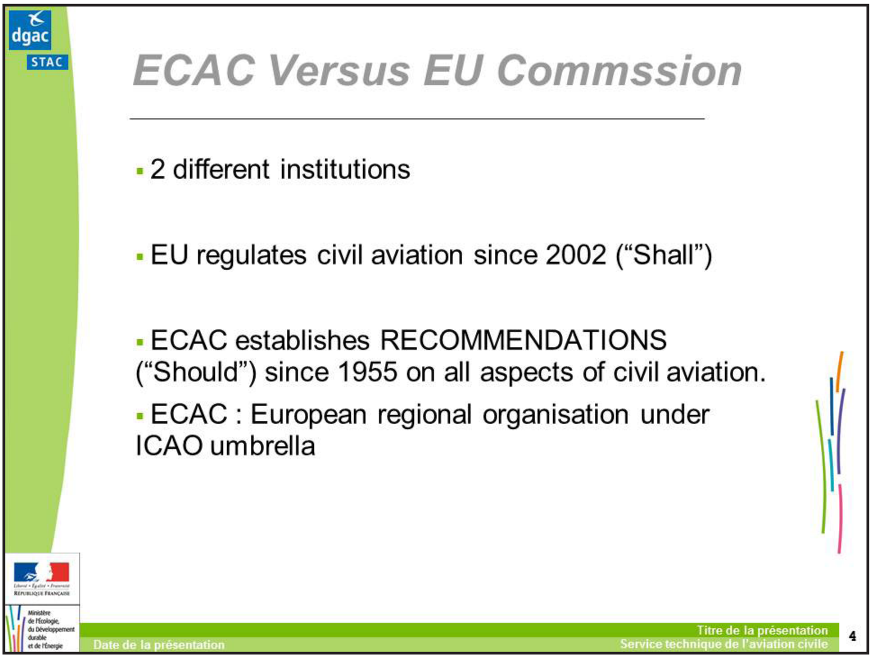
Date de la présentation

Titre de la présentation
Service technique de l'aviation civile

2

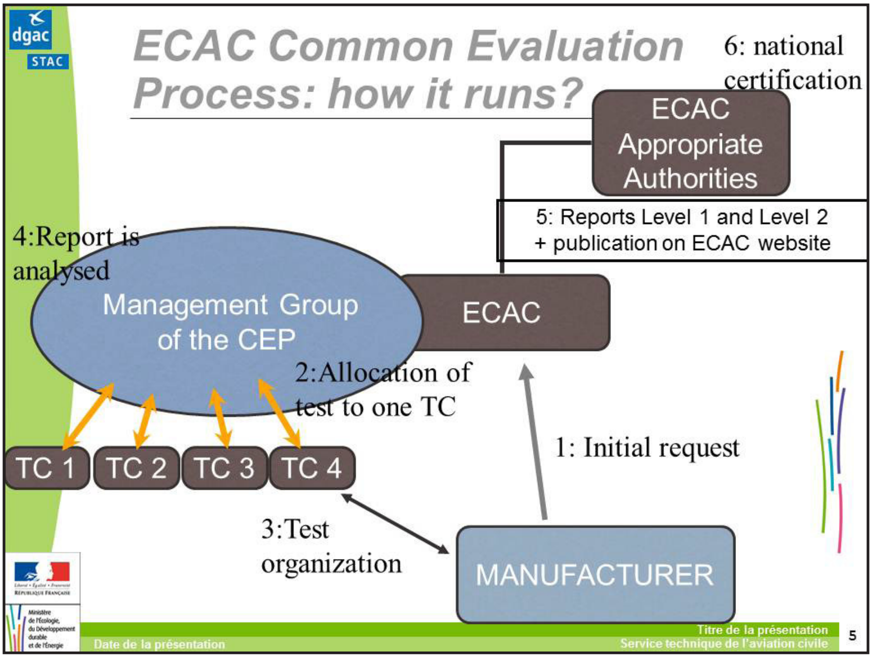


What is ECAC?



ECAC Versus EU Commssion

- 2 different institutions
- EU regulates civil aviation since 2002 (“Shall”)
- ECAC establishes RECOMMENDATIONS (“Should”) since 1955 on all aspects of civil aviation.
- ECAC : European regional organisation under ICAO umbrella



ECAC Common Evaluation Process: how it runs?

6: national certification

ECAC Appropriate Authorities

5: Reports Level 1 and Level 2
+ publication on ECAC website

4:Report is analysed

Management Group of the CEP

ECAC

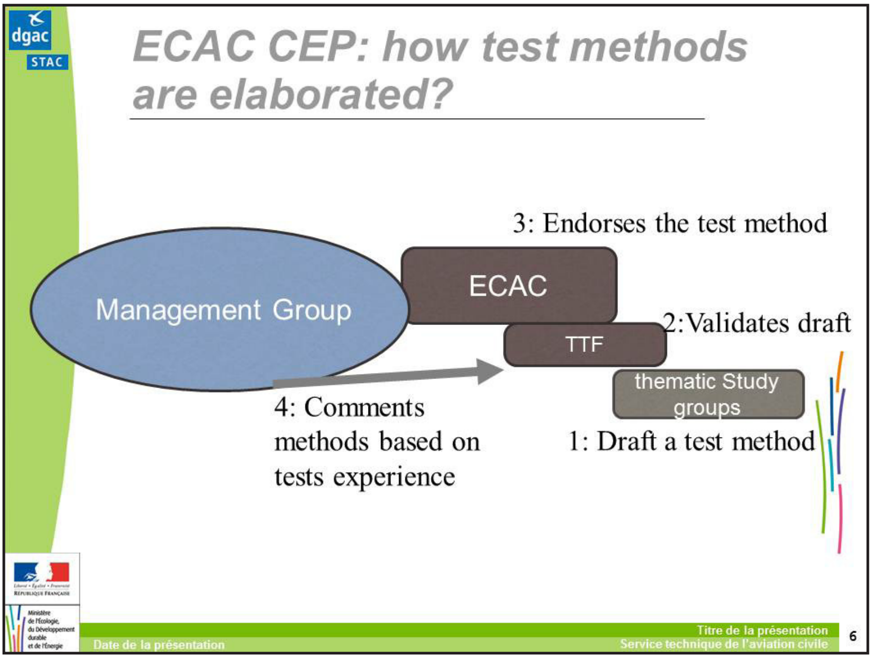
2:Allocation of test to one TC

TC 1 TC 2 TC 3 TC 4

1: Initial request

3:Test organization

MANUFACTURER



ECAC CEP: how test methods are elaborated?

3: Endorses the test method

Management Group

ECAC

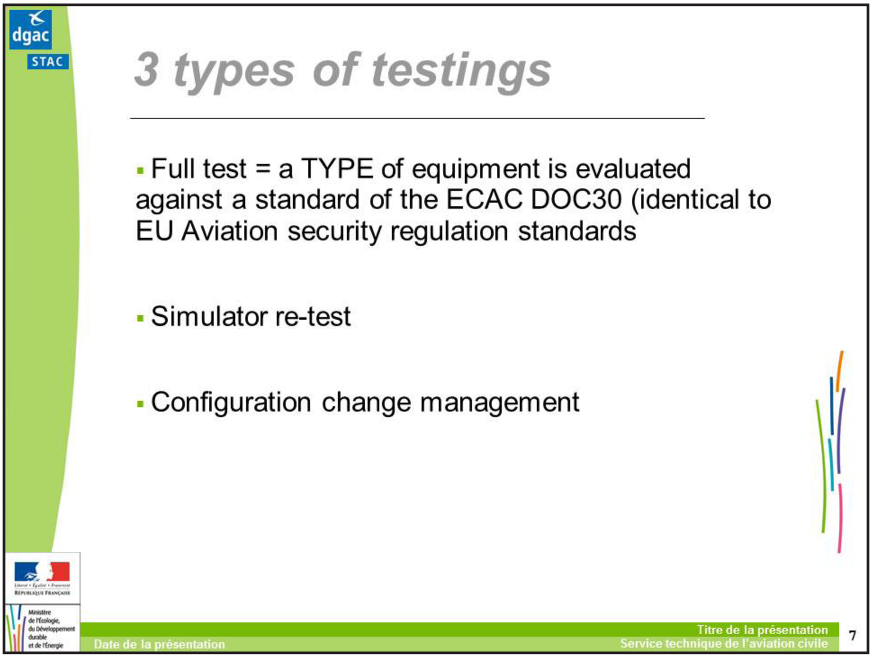
TTF

2:Validates draft

thematic Study groups

1: Draft a test method

4: Comments methods based on tests experience



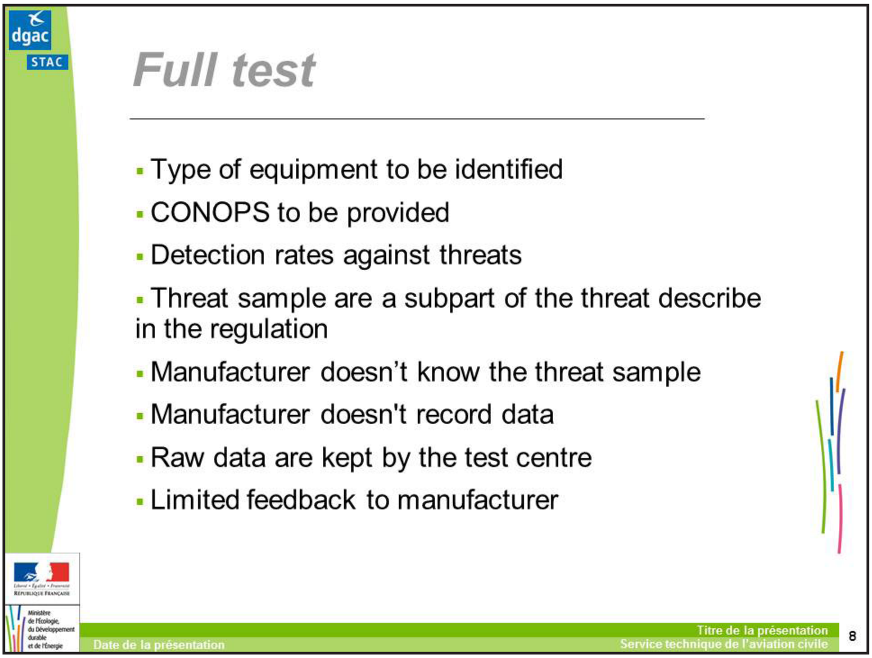
3 types of testings

- Full test = a TYPE of equipment is evaluated against a standard of the ECAC DOC30 (identical to EU Aviation security regulation standards
- Simulator re-test
- Configuration change management

Date de la présentation

Titre de la présentation
Service technique de l'aviation civile

7



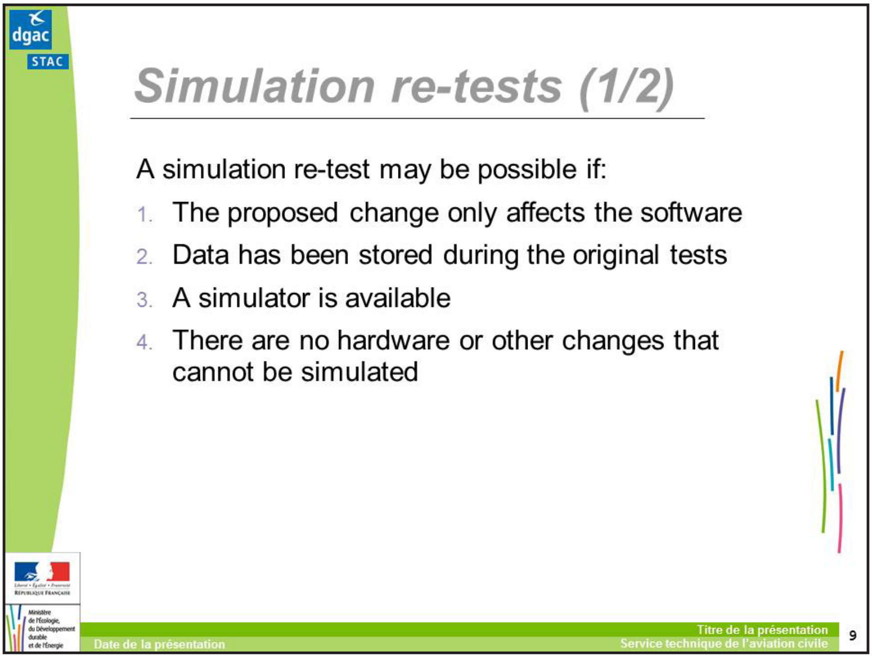
Full test

- Type of equipment to be identified
- CONOPS to be provided
- Detection rates against threats
- Threat sample are a subpart of the threat describe in the regulation
- Manufacturer doesn't know the threat sample
- Manufacturer doesn't record data
- Raw data are kept by the test centre
- Limited feedback to manufacturer

Date de la présentation

Titre de la présentation
Service technique de l'aviation civile

8



dgac
STAC

Simulation re-tests (1/2)

A simulation re-test may be possible if:

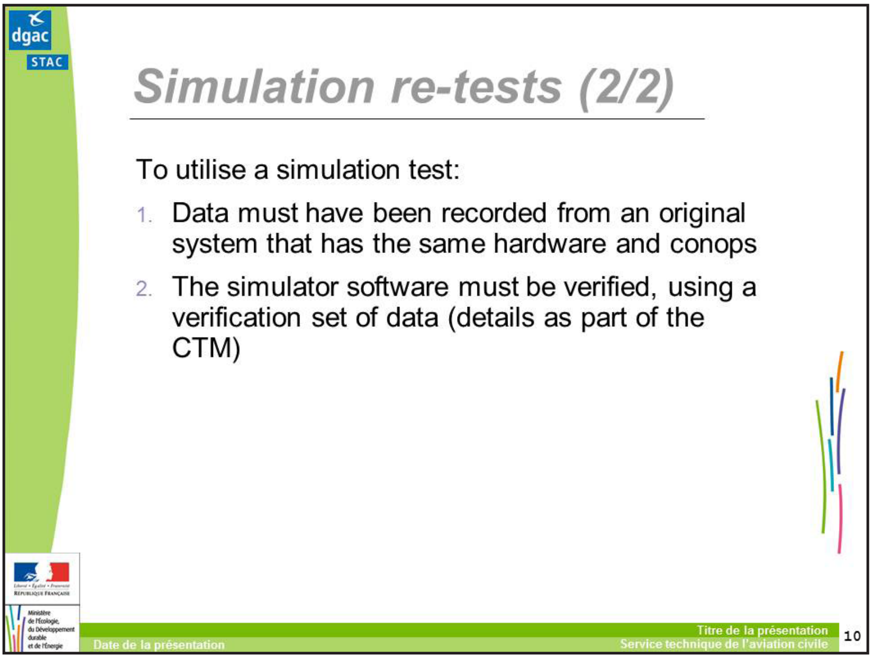
1. The proposed change only affects the software
2. Data has been stored during the original tests
3. A simulator is available
4. There are no hardware or other changes that cannot be simulated

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère
de l'Écologie,
du Développement
durable
et de l'Énergie

Date de la présentation

Titre de la présentation
Service technique de l'aviation civile



dgac
STAC

Simulation re-tests (2/2)

To utilise a simulation test:

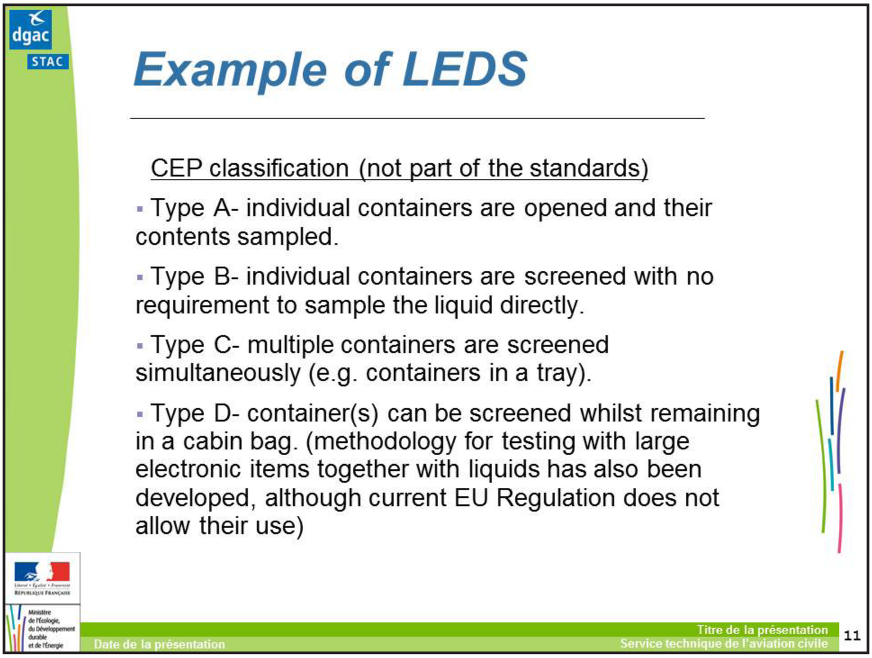
1. Data must have been recorded from an original system that has the same hardware and conops
2. The simulator software must be verified, using a verification set of data (details as part of the CTM)

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère
de l'Écologie,
du Développement
durable
et de l'Énergie

Date de la présentation

Titre de la présentation
Service technique de l'aviation civile



dgac
STAC

Example of LEDS

CEP classification (not part of the standards)

- Type A- individual containers are opened and their contents sampled.
- Type B- individual containers are screened with no requirement to sample the liquid directly.
- Type C- multiple containers are screened simultaneously (e.g. containers in a tray).
- Type D- container(s) can be screened whilst remaining in a cabin bag. (methodology for testing with large electronic items together with liquids has also been developed, although current EU Regulation does not allow their use)

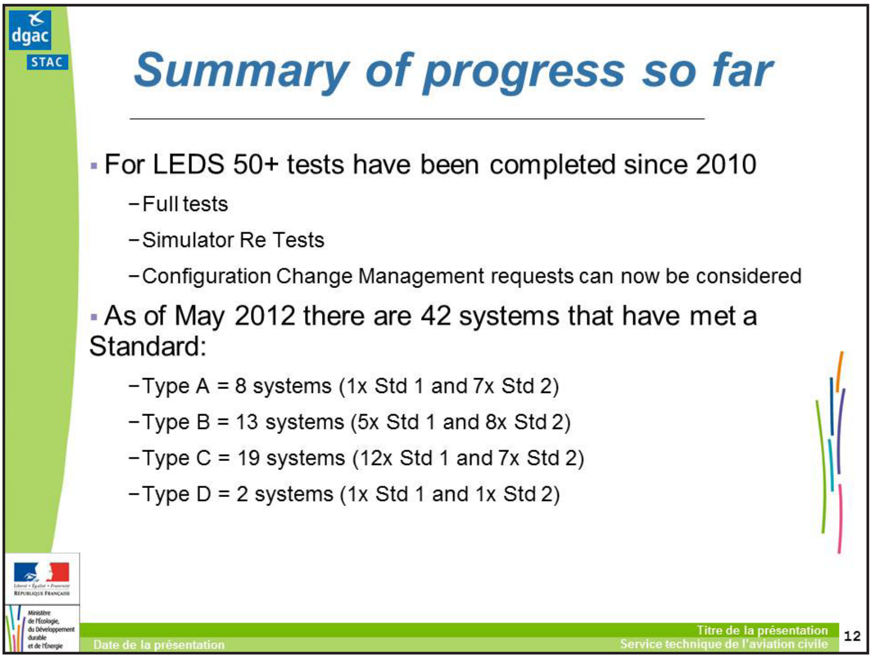
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère
de l'Écologie,
du Développement
durable
et de l'Énergie

Date de la présentation

Titre de la présentation
Service technique de l'aviation civile

11



dgac
STAC

Summary of progress so far

- For LEDS 50+ tests have been completed since 2010
  - Full tests
  - Simulator Re Tests
  - Configuration Change Management requests can now be considered
- As of May 2012 there are 42 systems that have met a Standard:
  - Type A = 8 systems (1x Std 1 and 7x Std 2)
  - Type B = 13 systems (5x Std 1 and 8x Std 2)
  - Type C = 19 systems (12x Std 1 and 7x Std 2)
  - Type D = 2 systems (1x Std 1 and 1x Std 2)

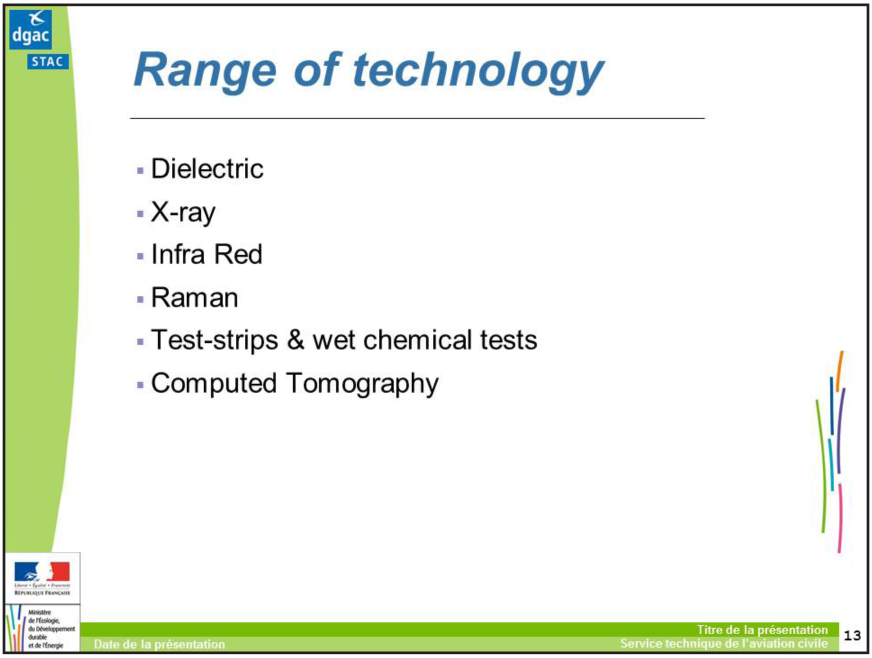
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère
de l'Écologie,
du Développement
durable
et de l'Énergie

Date de la présentation

Titre de la présentation
Service technique de l'aviation civile

12



Range of technology

- Dielectric
- X-ray
- Infra Red
- Raman
- Test-strips & wet chemical tests
- Computed Tomography

Ministère
de l'Écologie,
du Développement
durable
et de l'Énergie

Date de la présentation

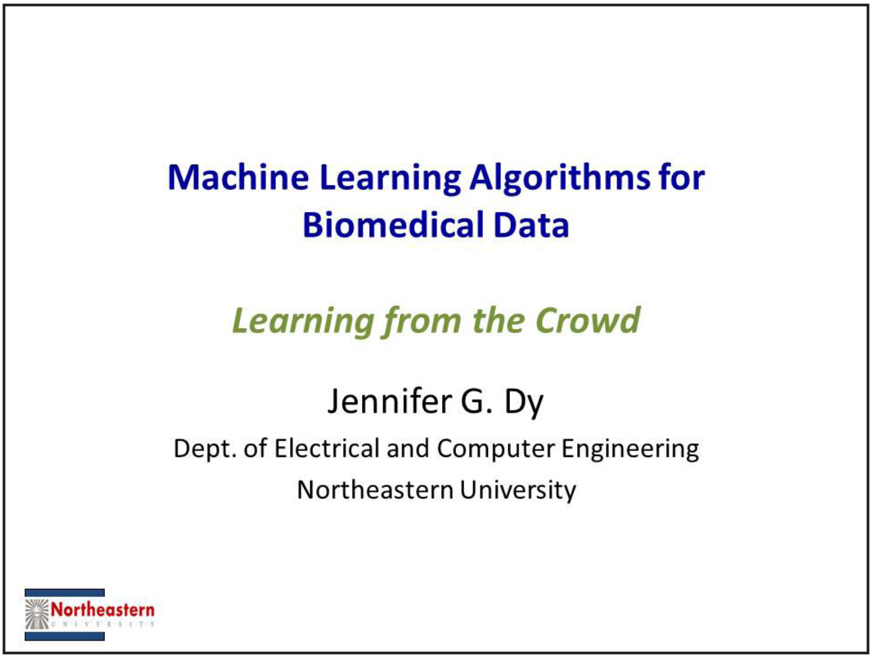
Titre de la présentation
Service technique de l'aviation civile

13

Thanks for attention

Jean-Claude GUILPIN
STAC/SE/Security division
jean-claude.guilpin@aviation-civile.gouv.fr
+ 33 1 49 56 81 34

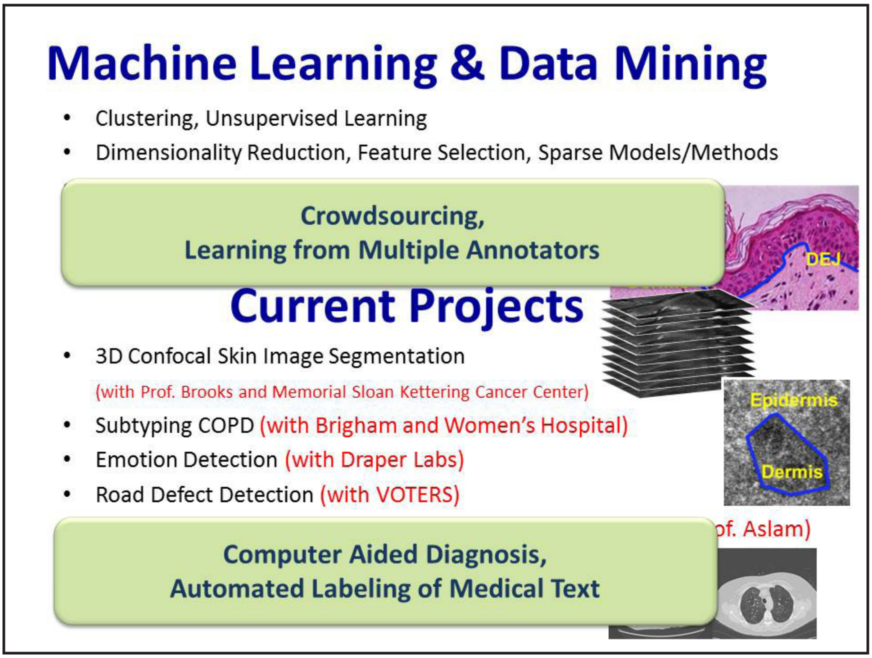
16.30 Jennifer Dy: Machine Learning Algorithms for Biomedical Data



Machine Learning Algorithms for
Biomedical Data

Learning from the Crowd

Jennifer G. Dy
Dept. of Electrical and Computer Engineering
Northeastern University



Machine Learning & Data Mining

- Clustering, Unsupervised Learning
- Dimensionality Reduction, Feature Selection, Sparse Models/Methods

Crowdsourcing,
Learning from Multiple Annotators

Current Projects

- 3D Confocal Skin Image Segmentation
  (with Prof. Brooks and Memorial Sloan Kettering Cancer Center)
- Subtyping COPD (with Brigham and Women's Hospital)
- Emotion Detection (with Draper Labs)
- Road Defect Detection (with VOTERS)

of. Aslam)

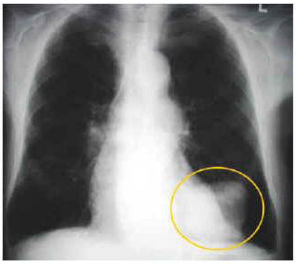
Computer Aided Diagnosis,
Automated Labeling of Medical Text

Conclusions

- We provided a probabilistic model that allows learning from multiple annotators/crowd whose annotations may be noisy;
- Our model takes into account that the quality of annotation may vary with data;
- This model can deal with missing annotators/data;
- Our model can also be utilized to evaluate annotators even when ground truth is not available; and
- We can also utilize our model to select the most trustworthy/accurate annotator for each new instance labeling
- We've developed an approach that can intelligently select samples to label and the associated annotators to query (active learning from multiple annotators).

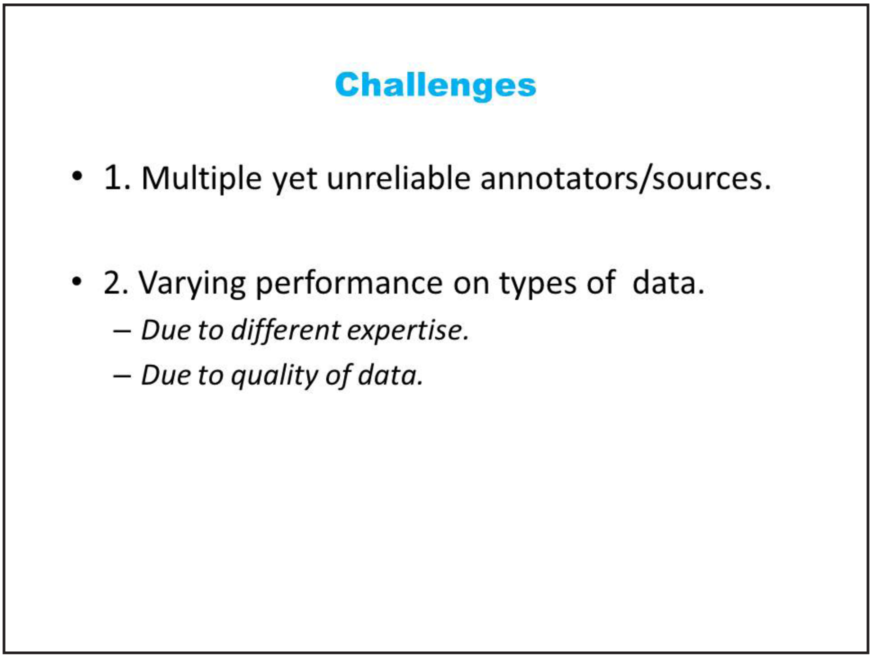
Motivation

- ***Multiple Expert Diagnoses***



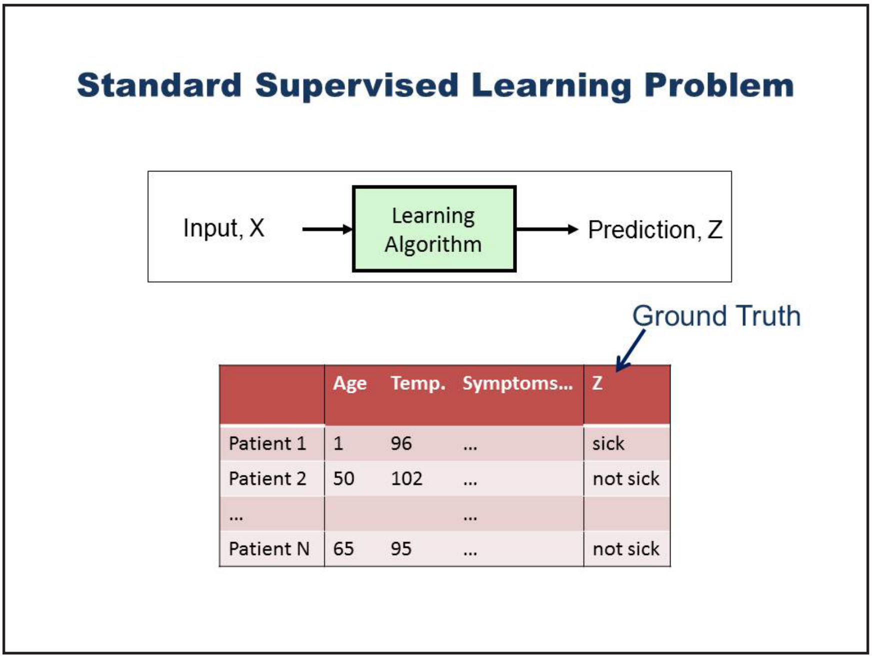
- ***Amazon Mechanical Turk***

1. *How should the patients be diagnosed when doctors disagree?*
2. *How do we evaluate the doctors' diagnoses?*



Challenges

- 1. Multiple yet unreliable annotators/sources.
- 2. Varying performance on types of data.
  - *Due to different expertise.*
  - *Due to quality of data.*



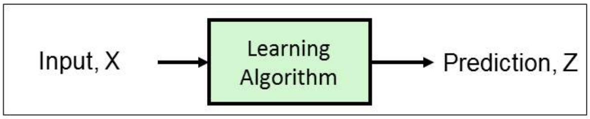
Standard Supervised Learning Problem

Input, X → Learning Algorithm → Prediction, Z

Ground Truth

	Age	Temp.	Symptoms…	Z
Patient 1	1	96	…	sick
Patient 2	50	102	…	not sick
…			…	
Patient N	65	95	…	not sick

Multiple Annotator Learning Problem


Input, X

Learning Algorithm

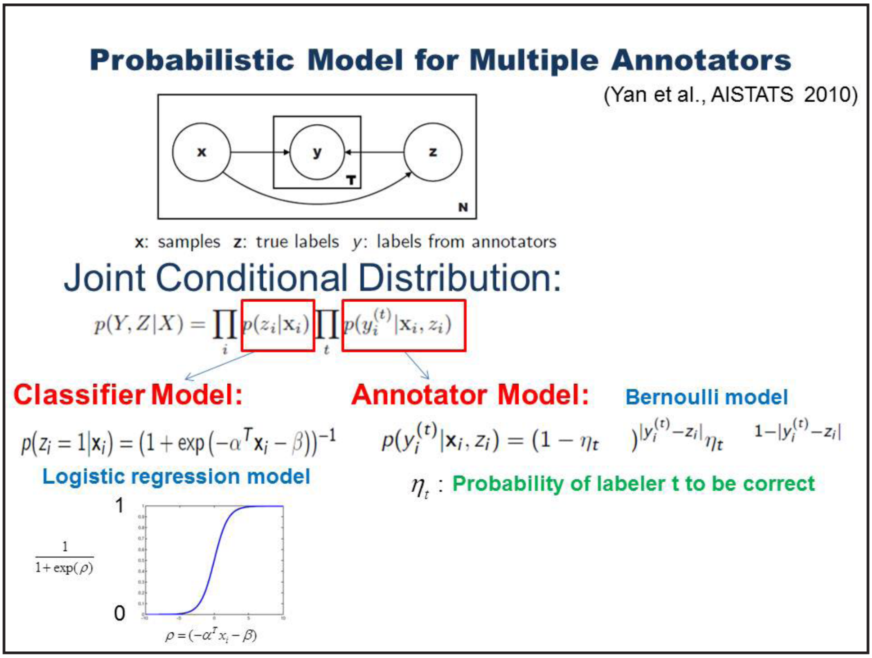
Prediction, Z

	Age	Temp.	Symptoms…	Ann. Y_1	Ann. Y_2	Ann. …	Ann. Y_T
Patient 1	1	96	…	not sick	sick	…	sick
Patient 2	50	102	…	sick	sick	…	sick
…			…				
Patient N	65	95	…	not sick	not sick	…	sick

- No objective ground truth.
- Multiple inexpensive annotators may be available, but will depend on annotator idiosyncrasies.

Challenges

- 1. Multiple yet unreliable annotators/sources.
- 2. Varying performance on types of data.
  - *Due to different expertise.*
  - *Due to quality of data.*



Probabilistic Model for Multiple Annotators

(Yan et al., AISTATS 2010)

x: samples **z**: true labels *y*: labels from annotators

Joint Conditional Distribution:

$$p(Y,Z|X)=\prod_i \boxed{p(z_i|\mathrm{x}_i)}\prod_t \boxed{p(y_i^{(t)}|\mathrm{x}_i,z_i)}$$

Classifier Model:

$$p(z_i=1|\mathrm{x}_i)=(1+\exp(-\alpha^T\mathrm{x}_i-\beta))^{-1}$$

Logistic regression model

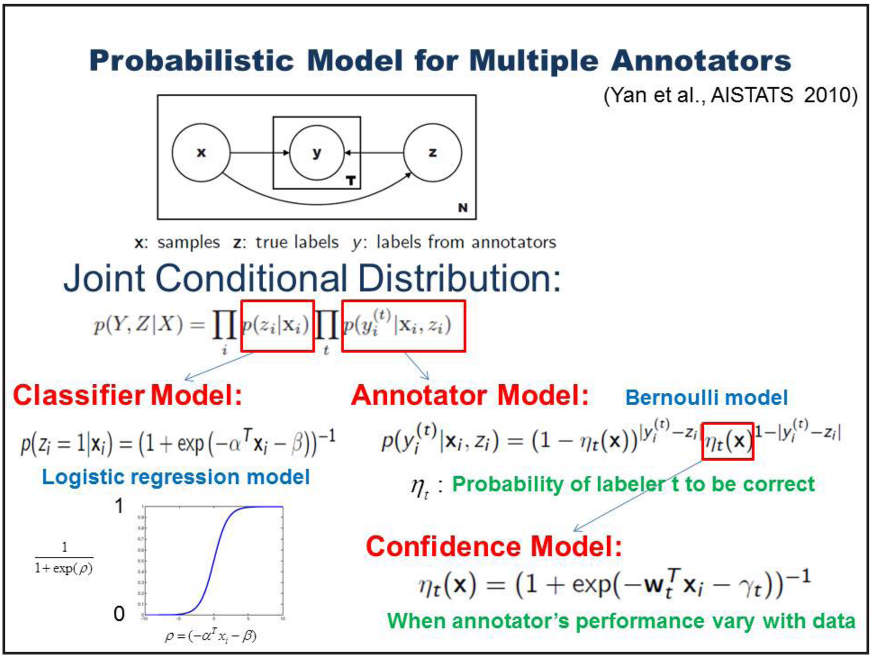
$$\frac{1}{1+\exp(\rho)}$$

$$\rho=(-\alpha^T x_i-\beta)$$

Annotator Model: Bernoulli model

$$p(y_i^{(t)}|\mathrm{x}_i,z_i)=(1-\eta_t)^{|y_i^{(t)}-z_i|}\eta_t^{1-|y_i^{(t)}-z_i|}$$

η_t : Probability of labeler t to be correct



Probabilistic Model for Multiple Annotators

(Yan et al., AISTATS 2010)

x: samples **z**: true labels *y*: labels from annotators

Joint Conditional Distribution:

$$p(Y,Z|X)=\prod_i \boxed{p(z_i|\mathrm{x}_i)}\prod_t \boxed{p(y_i^{(t)}|\mathrm{x}_i,z_i)}$$

Classifier Model:

$$p(z_i=1|\mathrm{x}_i)=(1+\exp(-\alpha^T\mathrm{x}_i-\beta))^{-1}$$

Logistic regression model

$$\frac{1}{1+\exp(\rho)}$$

$$\rho=(-\alpha^T x_i-\beta)$$

Annotator Model: Bernoulli model

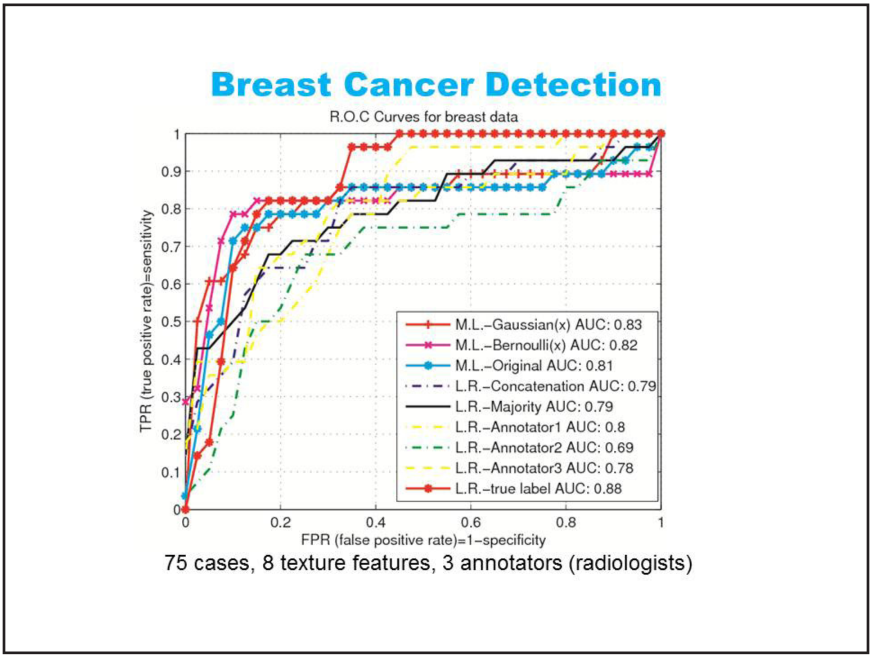
$$p(y_i^{(t)}|\mathrm{x}_i,z_i)=(1-\eta_t(\mathbf{x}))^{|y_i^{(t)}-z_i|}\boxed{\eta_t(\mathbf{x})}^{1-|y_i^{(t)}-z_i|}$$

η_t : Probability of labeler t to be correct

Confidence Model:

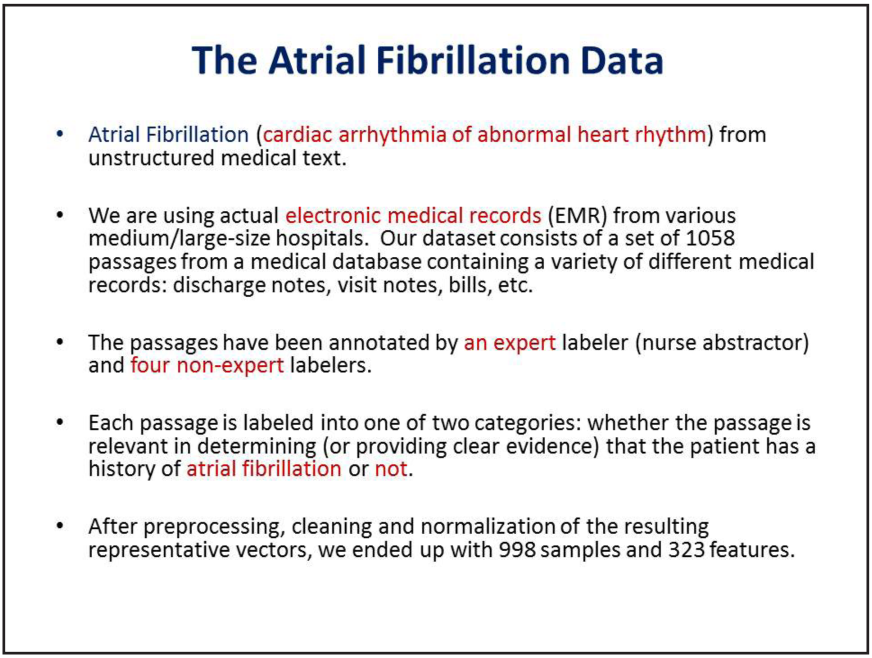
$$\eta_t(\mathbf{x})=(1+\exp(-\mathbf{w}_t^T\mathbf{x}_i-\gamma_t))^{-1}$$

When annotator's performance vary with data



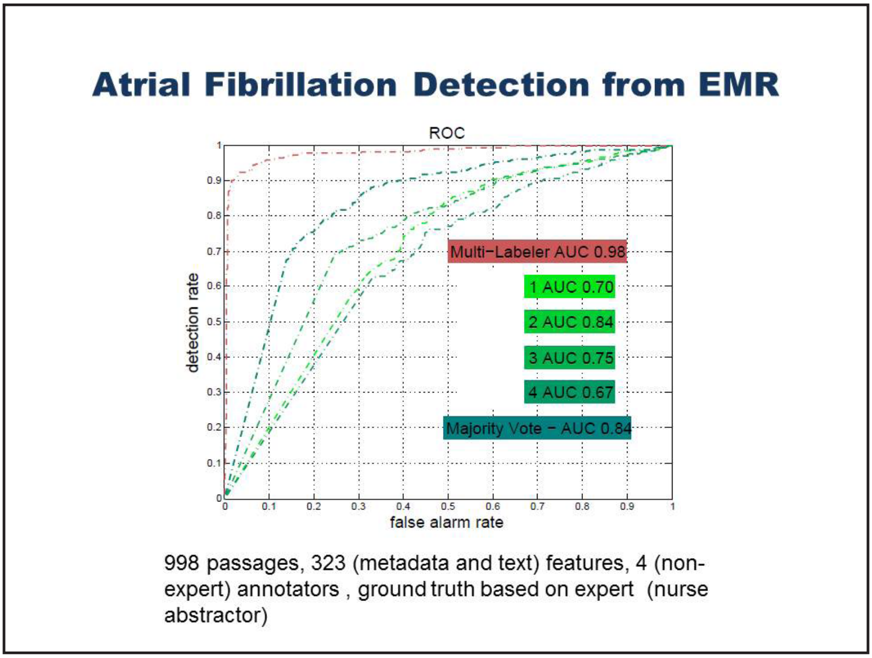
Breast Cancer Detection

75 cases, 8 texture features, 3 annotators (radiologists)



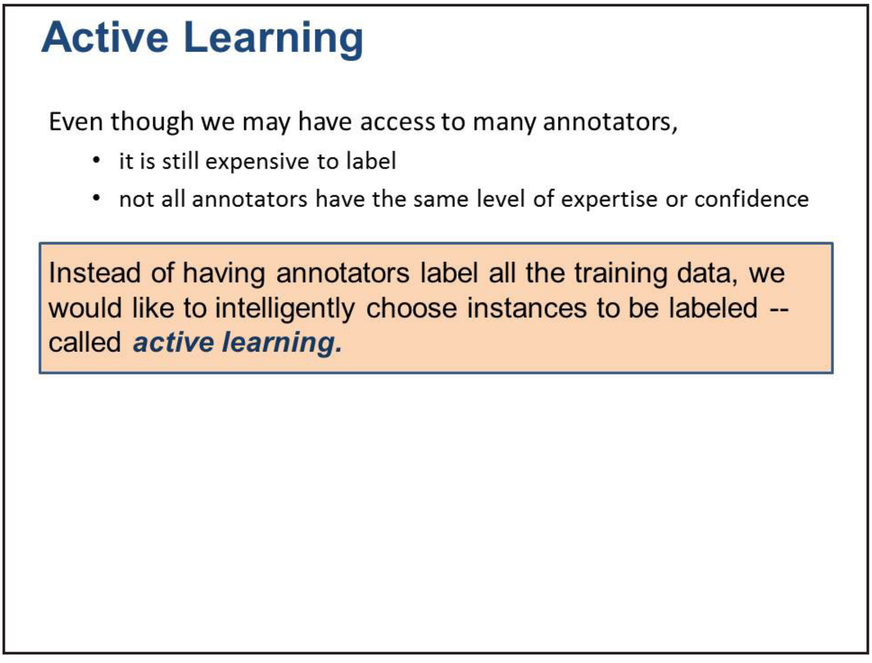
The Atrial Fibrillation Data

- Atrial Fibrillation (cardiac arrhythmia of abnormal heart rhythm) from unstructured medical text.
- We are using actual electronic medical records (EMR) from various medium/large-size hospitals. Our dataset consists of a set of 1058 passages from a medical database containing a variety of different medical records: discharge notes, visit notes, bills, etc.
- The passages have been annotated by an expert labeler (nurse abstractor) and four non-expert labelers.
- Each passage is labeled into one of two categories: whether the passage is relevant in determining (or providing clear evidence) that the patient has a history of atrial fibrillation or not.
- After preprocessing, cleaning and normalization of the resulting representative vectors, we ended up with 998 samples and 323 features.



Atrial Fibrillation Detection from EMR

998 passages, 323 (metadata and text) features, 4 (non-expert) annotators , ground truth based on expert (nurse abstractor)



Active Learning

Even though we may have access to many annotators,

- it is still expensive to label
- not all annotators have the same level of expertise or confidence

Instead of having annotators label all the training data, we would like to intelligently choose instances to be labeled -- called ***active learning.***

New Paradigm:
Active Learning from Multiple Annotators

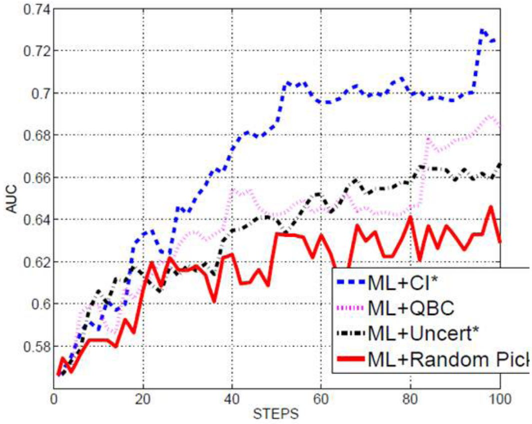
New Challenges:

- Intelligently choose instances to be labeled.
- Intelligently decide which annotator(s) to query from.

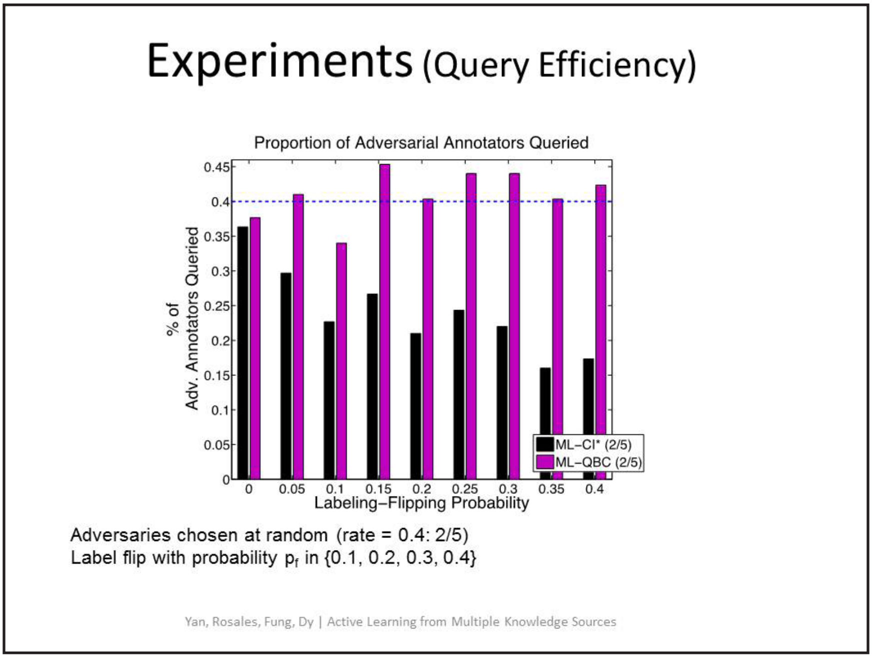
Two Strategies:

1. Uncertainty Sampling (ICML 2011)
2. Most Informative Sample and Annotator (AISTATS 2012)

Atrial Fibrillation Detection from EMR

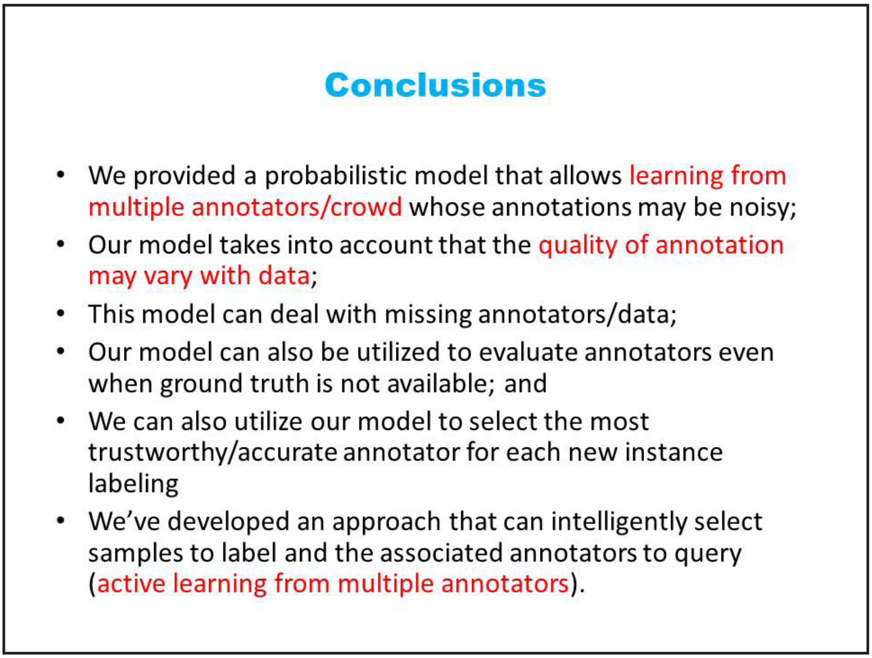


998 passages, 323 (metadata and text) features, 4 (non-expert) annotators , ground truth based on expert (nurse abstractor), 30 random initial training, 300 active pool, the rest as test



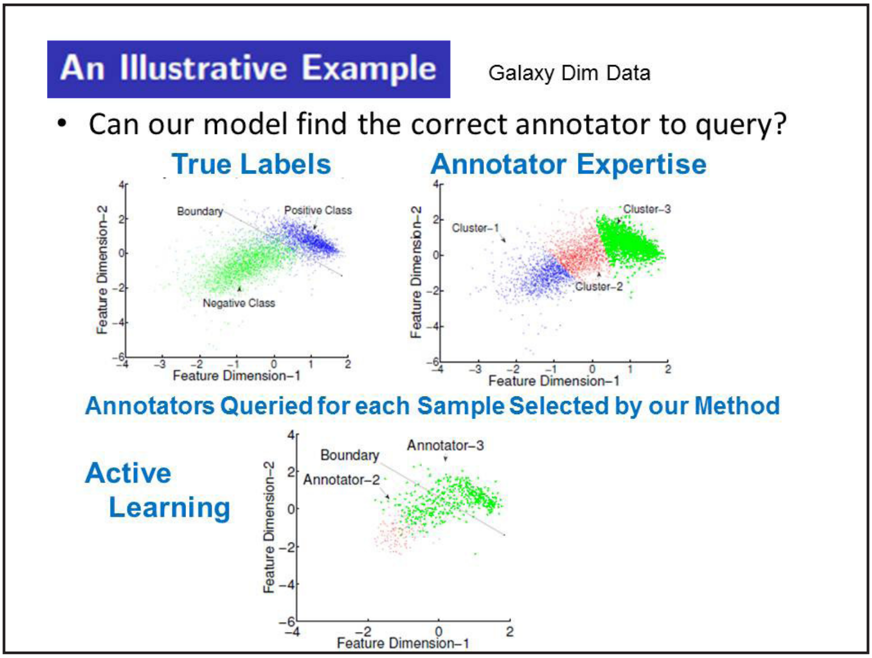
Experiments (Query Efficiency)

Adversaries chosen at random (rate = 0.4: 2/5)
Label flip with probability p_f in {0.1, 0.2, 0.3, 0.4}



Conclusions

- We provided a probabilistic model that allows learning from multiple annotators/crowd whose annotations may be noisy;
- Our model takes into account that the quality of annotation may vary with data;
- This model can deal with missing annotators/data;
- Our model can also be utilized to evaluate annotators even when ground truth is not available; and
- We can also utilize our model to select the most trustworthy/accurate annotator for each new instance labeling
- We've developed an approach that can intelligently select samples to label and the associated annotators to query (active learning from multiple annotators).



An Illustrative Example

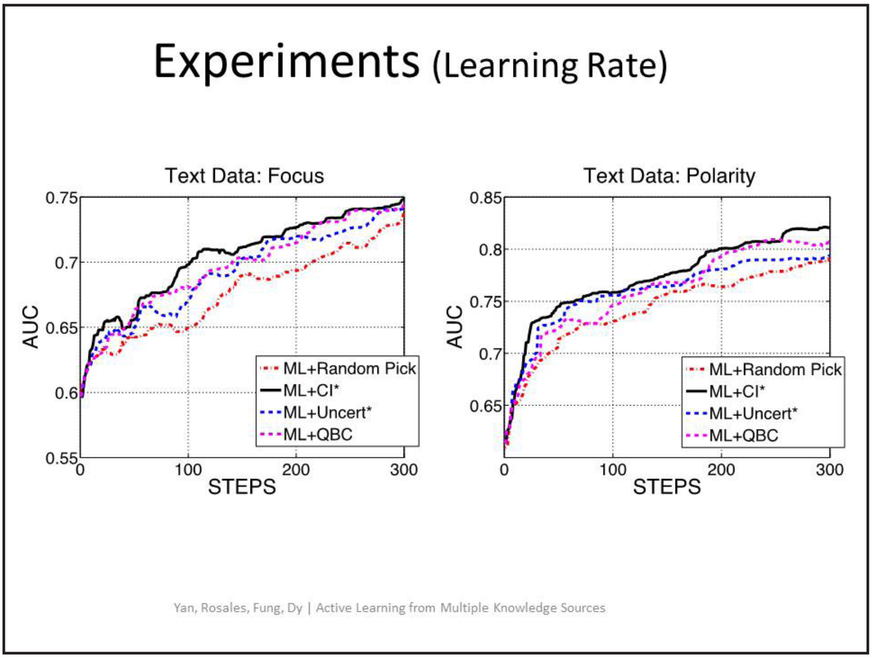
Galaxy Dim Data

- Can our model find the correct annotator to query?

True Labels **Annotator Expertise**

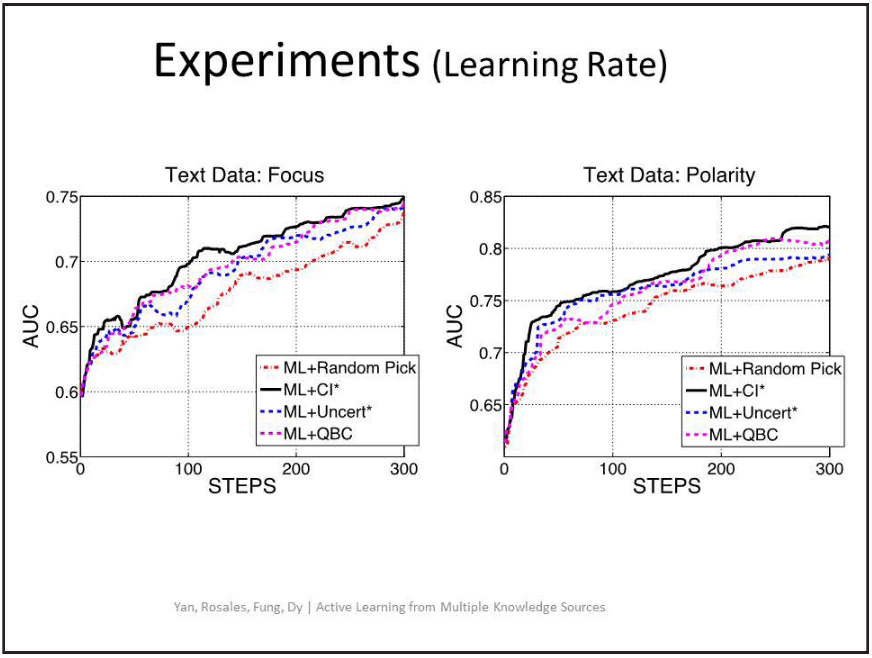
Annotators Queried for each Sample Selected by our Method

Active Learning



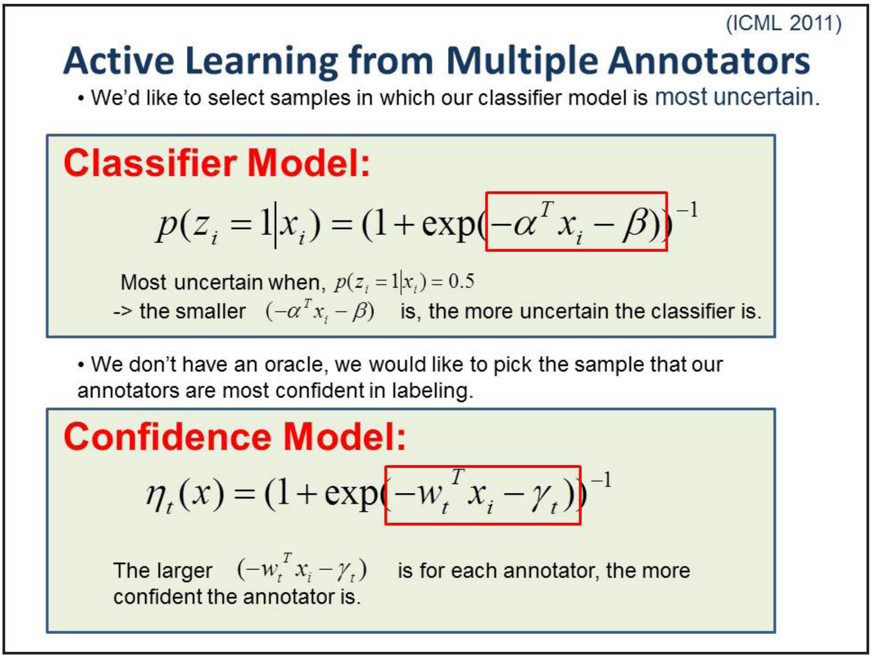
Experiments (Learning Rate)

Yan, Rosales, Fung, Dy | Active Learning from Multiple Knowledge Sources



Experiments (Learning Rate)

Yan, Rosales, Fung, Dy | Active Learning from Multiple Knowledge Sources



(ICML 2011)

Active Learning from Multiple Annotators

- We'd like to select samples in which our classifier model is most uncertain.

Classifier Model:

$$p(z_i = 1 | x_i) = (1 + \exp(\boxed{-\alpha^T x_i - \beta}))^{-1}$$

Most uncertain when, $p(z_i = 1 | x_i) = 0.5$

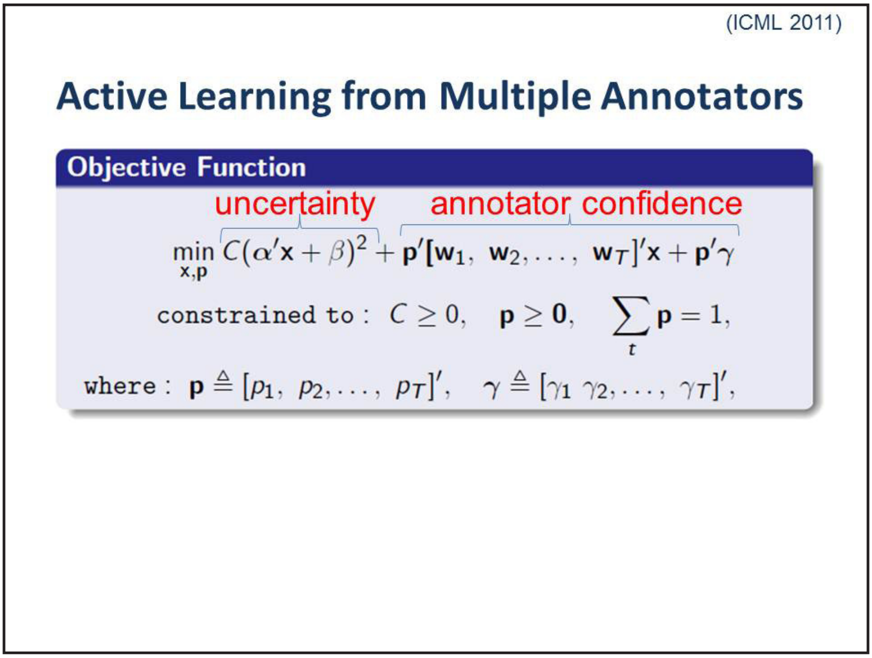
-> the smaller $(-\alpha^T x_i - \beta)$ is, the more uncertain the classifier is.

- We don't have an oracle, we would like to pick the sample that our annotators are most confident in labeling.

Confidence Model:

$$\eta_t(x) = (1 + \exp(\boxed{-w_t^T x_i - \gamma_t}))^{-1}$$

The larger $(-w_t^T x_i - \gamma_t)$ is for each annotator, the more confident the annotator is.



(ICML 2011)

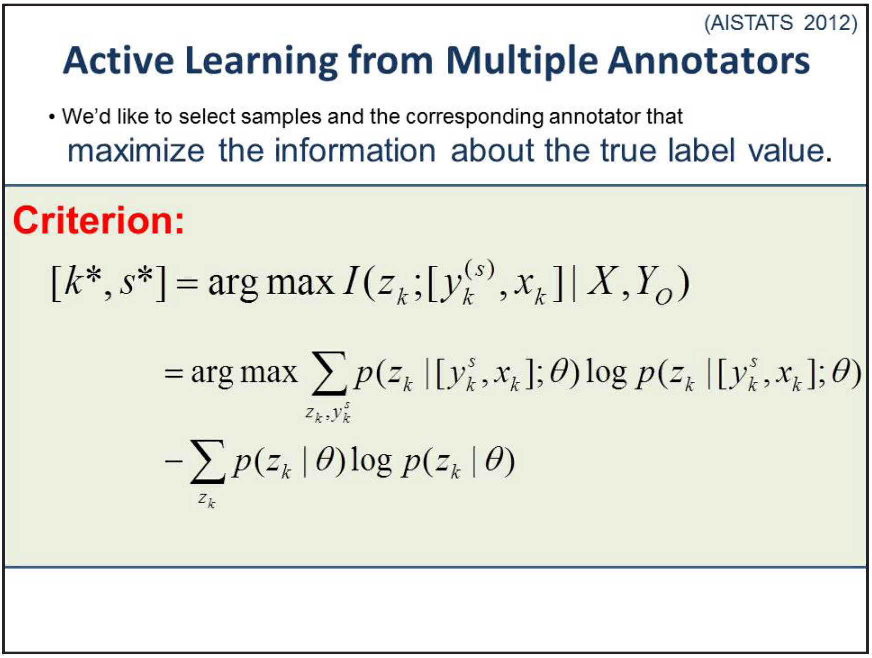
Active Learning from Multiple Annotators

Objective Function

$$\min_{\mathbf{x},\mathbf{p}} \overbrace{C(\alpha'\mathbf{x}+\beta)^2}^{\text{uncertainty}} + \overbrace{\mathbf{p}'[\mathbf{w}_1,\ \mathbf{w}_2,\ldots,\ \mathbf{w}_T]'\mathbf{x}+\mathbf{p}'\gamma}^{\text{annotator confidence}}$$

constrained to : $C \geq 0,\quad \mathbf{p} \geq \mathbf{0},\quad \sum_t \mathbf{p} = 1,$

where : $\mathbf{p} \triangleq [p_1,\ p_2,\ldots,\ p_T]',\quad \gamma \triangleq [\gamma_1\ \gamma_2,\ldots,\ \gamma_T]',$



(AISTATS 2012)

Active Learning from Multiple Annotators

- We'd like to select samples and the corresponding annotator that maximize the information about the true label value.

Criterion:

$$[k^*, s^*] = \arg\max I(z_k;[y_k^{(s)}, x_k] \mid X, Y_O)$$

$$= \arg\max \sum_{z_k, y_k^s} p(z_k \mid [y_k^s, x_k];\theta) \log p(z_k \mid [y_k^s, x_k];\theta)$$

$$-\sum_{z_k} p(z_k \mid \theta) \log p(z_k \mid \theta)$$

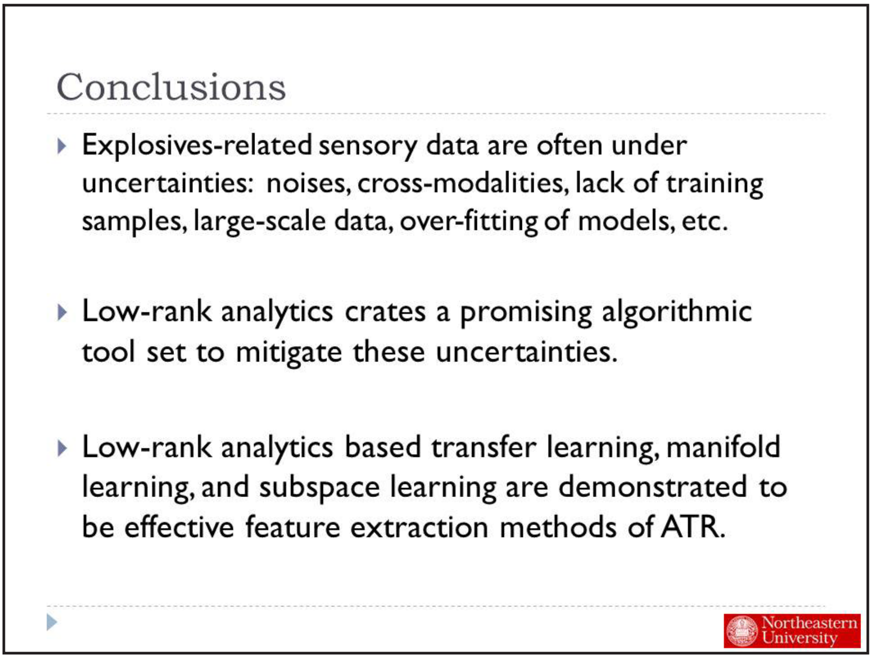
16.31 Raymond Fu: Low-Rank Analytics for Explosive Detection

Northeastern University

ALERT
AWARENESS AND LOCALIZATION
OF EXPLOSIVES-RELATED THREATS

Low-Rank Analytics
for Explosive Detection

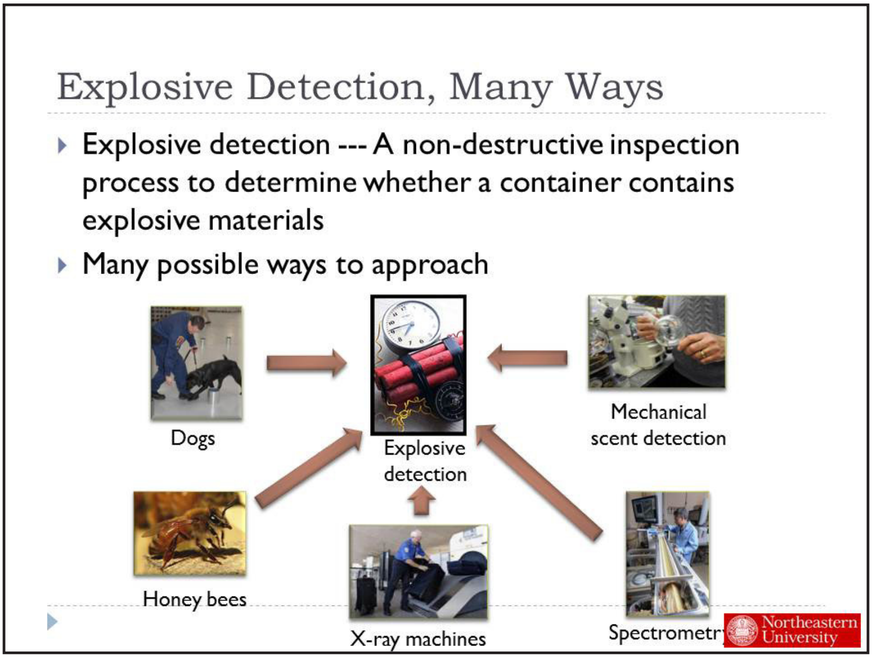
Yun Raymond Fu, Assistant Professor
ECE, Northeastern University



Conclusions

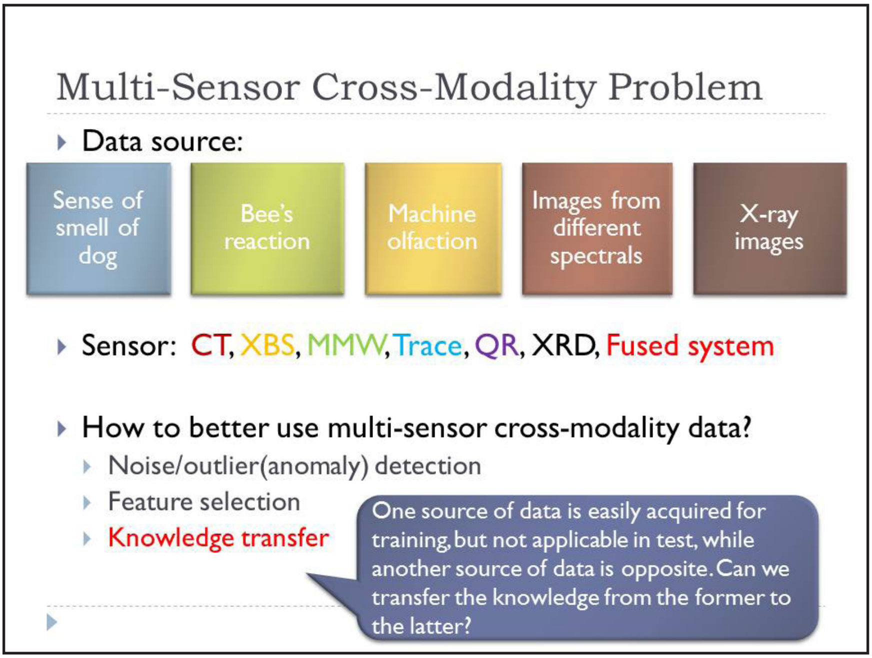
- Explosives-related sensory data are often under uncertainties: noises, cross-modalities, lack of training samples, large-scale data, over-fitting of models, etc.
- Low-rank analytics crates a promising algorithmic tool set to mitigate these uncertainties.
- Low-rank analytics based transfer learning, manifold learning, and subspace learning are demonstrated to be effective feature extraction methods of ATR.

Northeastern University



Explosive Detection, Many Ways

- Explosive detection --- A non-destructive inspection process to determine whether a container contains explosive materials
- Many possible ways to approach



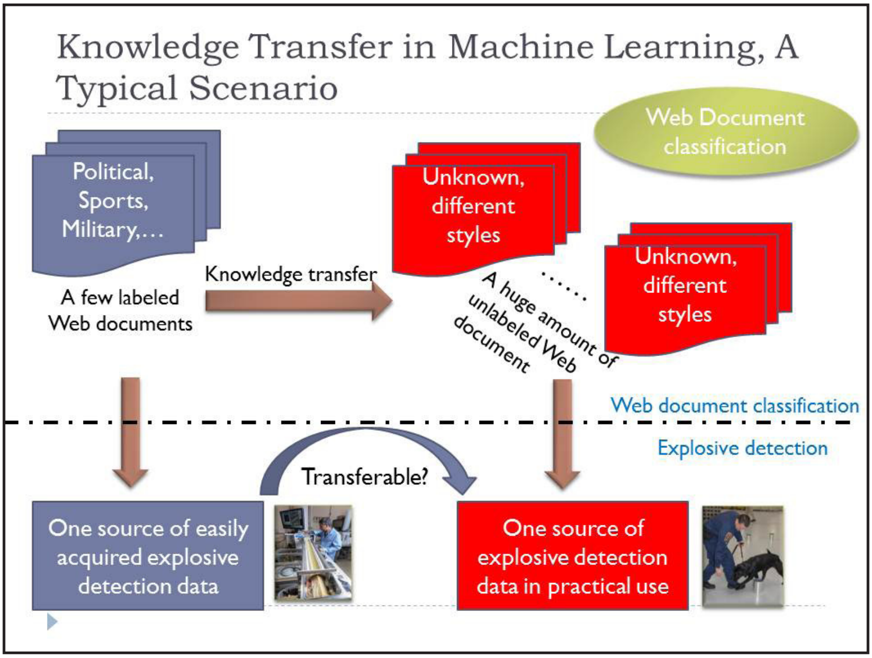
Multi-Sensor Cross-Modality Problem

- Data source:

Sense of smell of dog	Bee's reaction	Machine olfaction	Images from different spectrals	X-ray images

- Sensor: CT, XBS, MMW, Trace, QR, XRD, Fused system
- How to better use multi-sensor cross-modality data?
  - Noise/outlier(anomaly) detection
  - Feature selection
  - Knowledge transfer

One source of data is easily acquired for training, but not applicable in test, while another source of data is opposite. Can we transfer the knowledge from the former to the latter?



Knowledge Transfer in Machine Learning, A Typical Scenario

Web Document classification

Political, Sports, Military,…

A few labeled Web documents

Knowledge transfer

Unknown, different styles

A huge amount of unlabeled Web document

……

Unknown, different styles

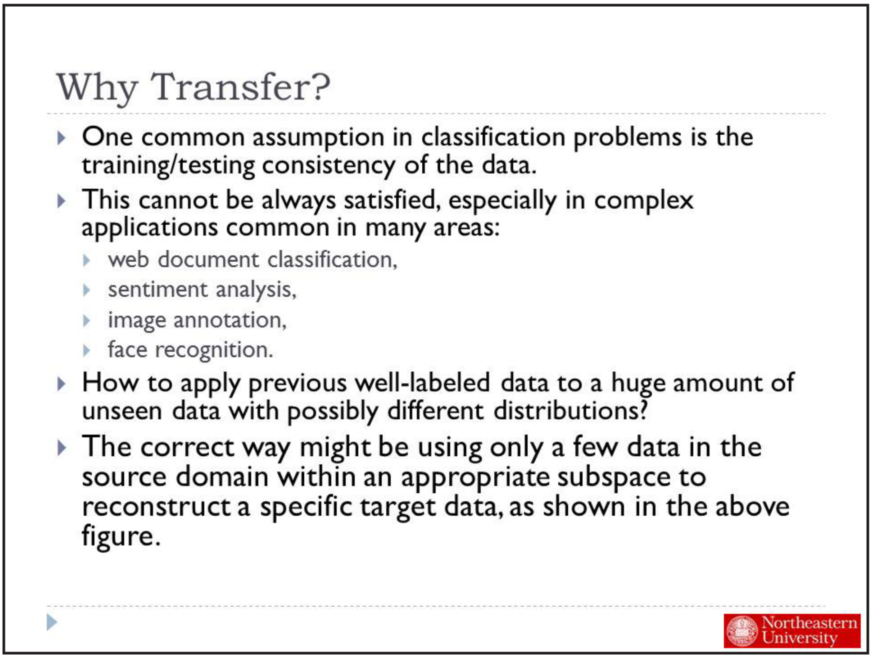
Web document classification

Explosive detection

Transferable?

One source of easily acquired explosive detection data

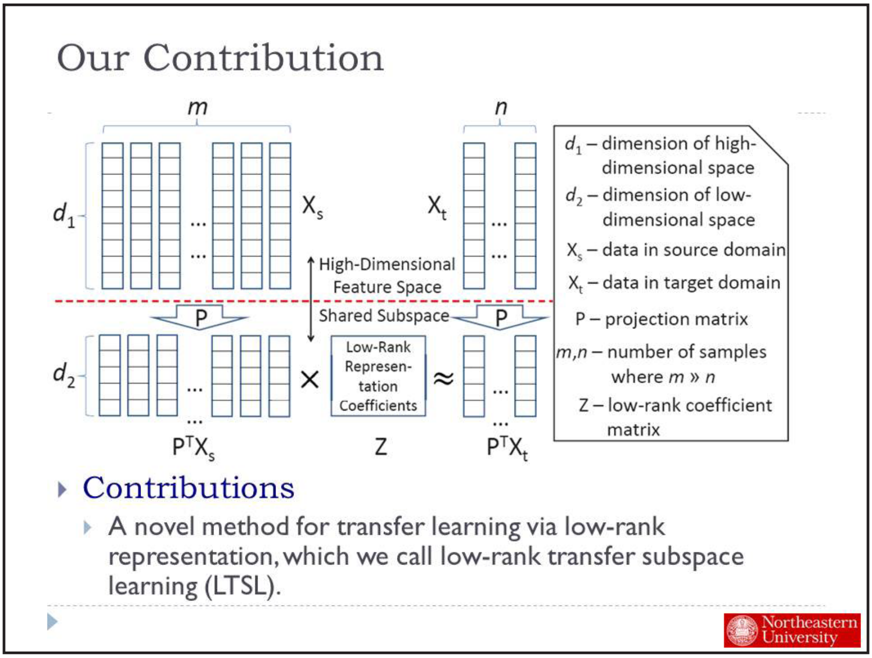
One source of explosive detection data in practical use



Why Transfer?

- One common assumption in classification problems is the training/testing consistency of the data.
- This cannot be always satisfied, especially in complex applications common in many areas:
    - web document classification,
    - sentiment analysis,
    - image annotation,
    - face recognition.
- How to apply previous well-labeled data to a huge amount of unseen data with possibly different distributions?
- The correct way might be using only a few data in the source domain within an appropriate subspace to reconstruct a specific target data, as shown in the above figure.

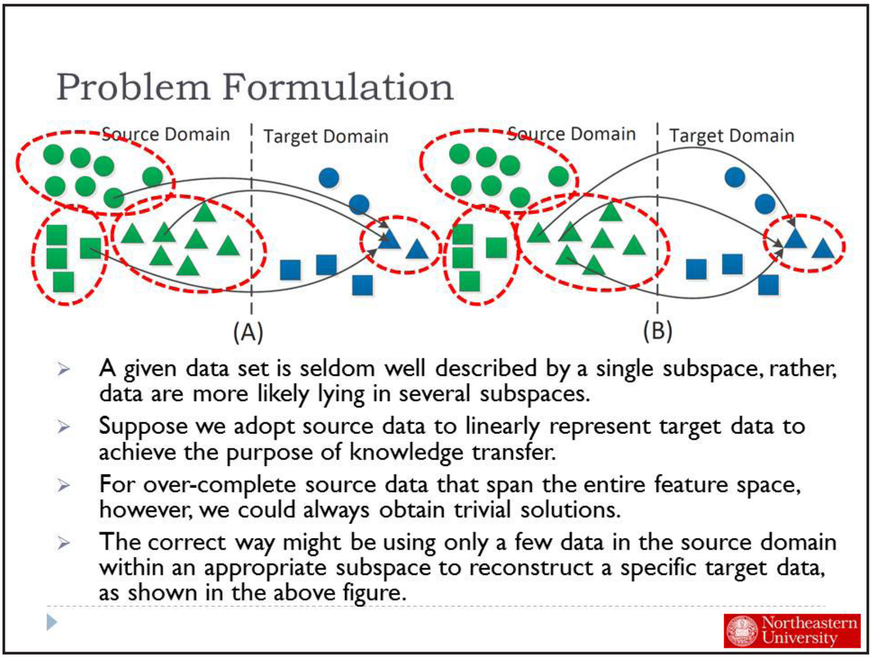
Northeastern University



Our Contribution

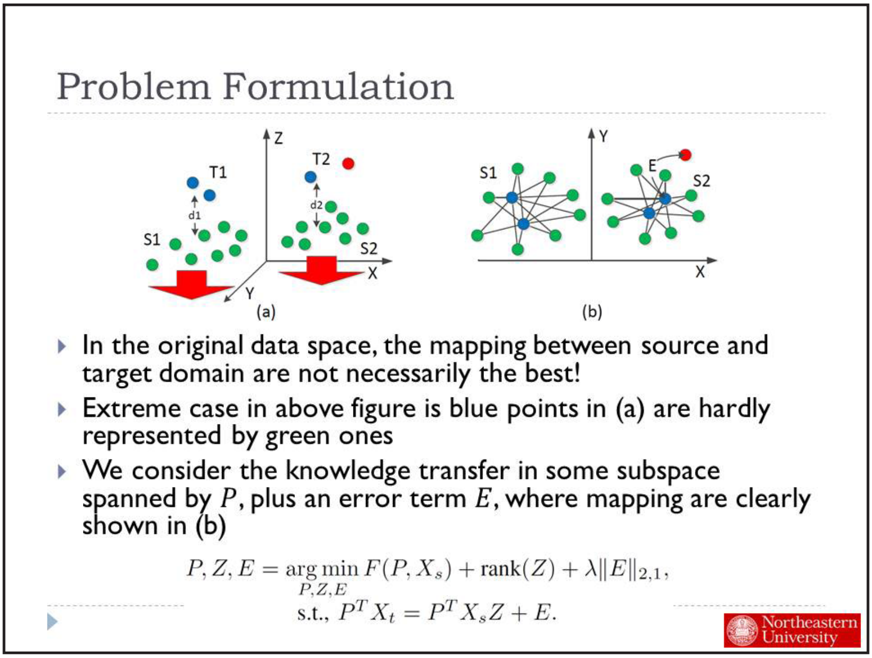
d_1 – dimension of high-dimensional space
d_2 – dimension of low-dimensional space
X_s – data in source domain
X_t – data in target domain
P – projection matrix
m,n – number of samples where $m \gg n$
Z – low-rank coefficient matrix

- Contributions
  - A novel method for transfer learning via low-rank representation, which we call low-rank transfer subspace learning (LTSL).



Problem Formulation

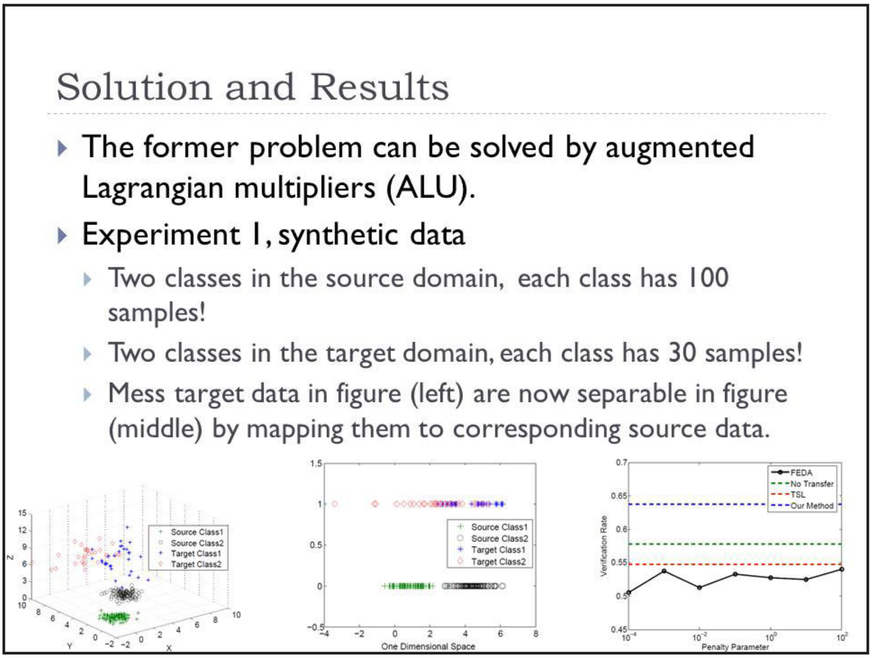
- A given data set is seldom well described by a single subspace, rather, data are more likely lying in several subspaces.
- Suppose we adopt source data to linearly represent target data to achieve the purpose of knowledge transfer.
- For over-complete source data that span the entire feature space, however, we could always obtain trivial solutions.
- The correct way might be using only a few data in the source domain within an appropriate subspace to reconstruct a specific target data, as shown in the above figure.



Problem Formulation

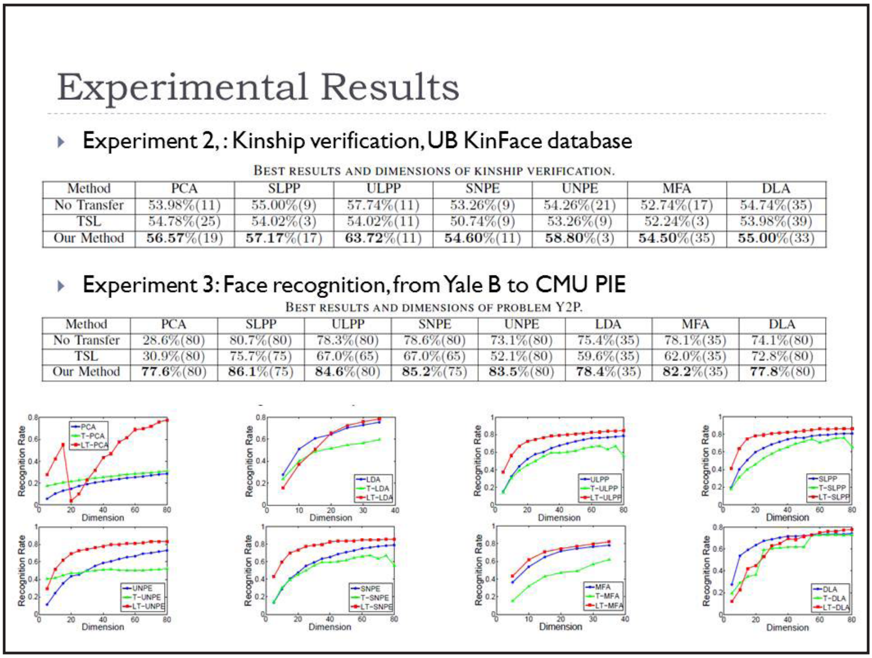
- In the original data space, the mapping between source and target domain are not necessarily the best!
- Extreme case in above figure is blue points in (a) are hardly represented by green ones
- We consider the knowledge transfer in some subspace spanned by P, plus an error term E, where mapping are clearly shown in (b)

$$P, Z, E = \underset{P,Z,E}{\arg\min}\, F(P, X_s) + \text{rank}(Z) + \lambda\|E\|_{2,1},$$
$$\text{s.t.,}\ \ P^T X_t = P^T X_s Z + E.$$



Solution and Results

- The former problem can be solved by augmented Lagrangian multipliers (ALU).
- Experiment 1, synthetic data
  - Two classes in the source domain, each class has 100 samples!
  - Two classes in the target domain, each class has 30 samples!
  - Mess target data in figure (left) are now separable in figure (middle) by mapping them to corresponding source data.



Experimental Results

- Experiment 2, : Kinship verification, UB KinFace database

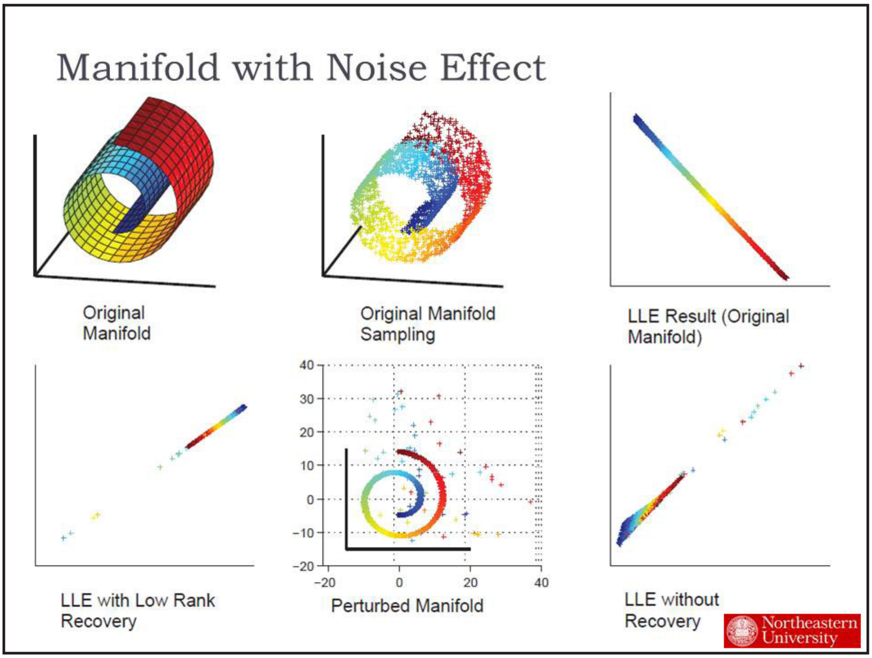
BEST RESULTS AND DIMENSIONS OF KINSHIP VERIFICATION.

Method	PCA	SLPP	ULPP	SNPE	UNPE	MFA	DLA
No Transfer	53.98%(11)	55.00%(9)	57.74%(11)	53.26%(9)	54.26%(21)	52.74%(17)	54.74%(35)
TSL	54.78%(25)	54.02%(3)	54.02%(11)	50.74%(9)	53.26%(9)	52.24%(3)	53.98%(39)
Our Method	**56.57**%(19)	**57.17**%(17)	**63.72**%(11)	**54.60**%(11)	**58.80**%(3)	**54.50**%(35)	**55.00**%(33)

- Experiment 3: Face recognition, from Yale B to CMU PIE

BEST RESULTS AND DIMENSIONS OF PROBLEM Y2P.

Method	PCA	SLPP	ULPP	SNPE	UNPE	LDA	MFA	DLA
No Transfer	28.6%(80)	80.7%(80)	78.3%(80)	78.6%(80)	73.1%(80)	75.4%(35)	78.1%(35)	74.1%(80)
TSL	30.9%(80)	75.7%(75)	67.0%(65)	67.0%(65)	52.1%(80)	59.6%(35)	62.0%(35)	72.8%(80)
Our Method	**77.6**%(80)	**86.1**%(75)	**84.6**%(80)	**85.2**%(75)	**83.5**%(80)	**78.4**%(35)	**82.2**%(35)	**77.8**%(80)



Manifold with Noise Effect

Original Manifold

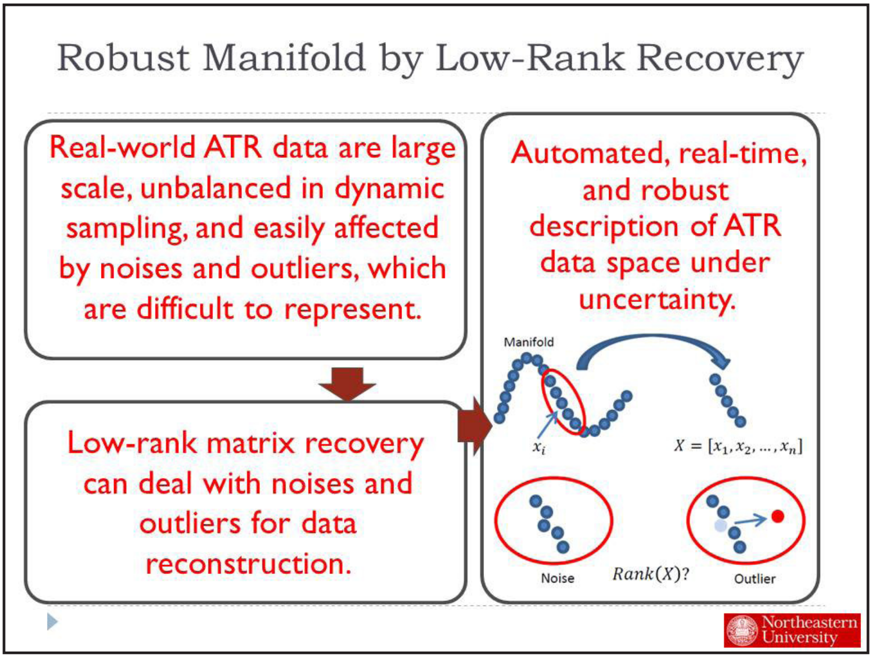
Original Manifold Sampling

LLE Result (Original Manifold)

LLE with Low Rank Recovery

Perturbed Manifold

LLE without Recovery



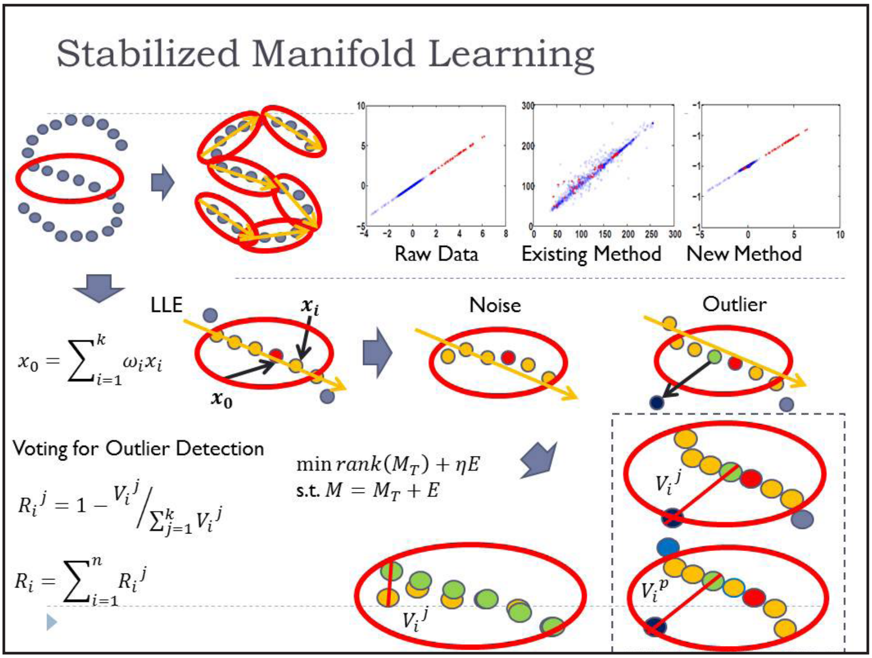
Robust Manifold by Low-Rank Recovery

Real-world ATR data are large scale, unbalanced in dynamic sampling, and easily affected by noises and outliers, which are difficult to represent.

Low-rank matrix recovery can deal with noises and outliers for data reconstruction.

Automated, real-time, and robust description of ATR data space under uncertainty.

Northeastern University



Stabilized Manifold Learning

$$x_0 = \sum_{i=1}^{k} \omega_i x_i$$

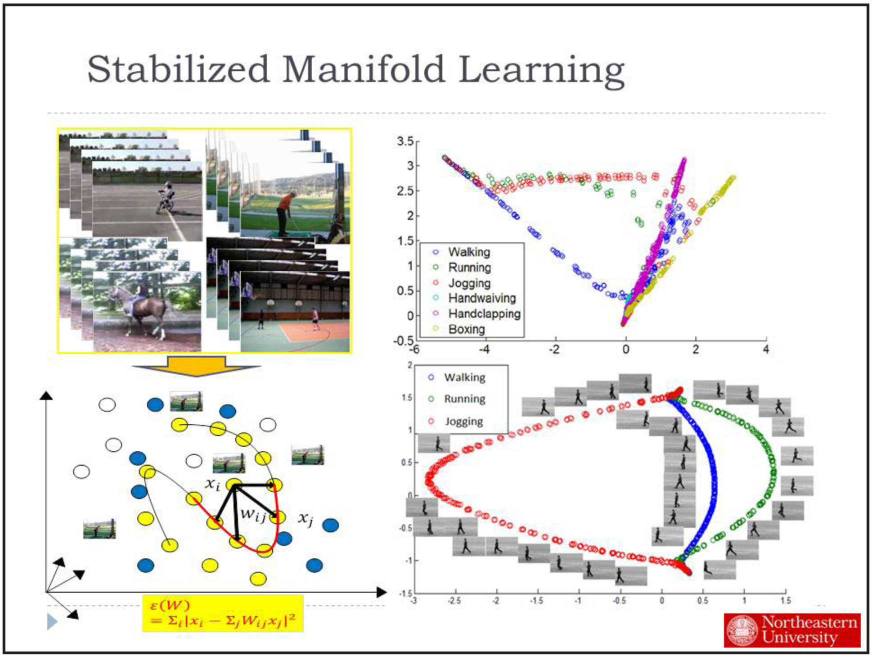
Voting for Outlier Detection

$$R_i^{\,j} = 1 - {V_i^{\,j}} \Big/ {\textstyle\sum_{j=1}^{k}} V_i^{\,j}$$

$$R_i = \sum_{i=1}^{n} R_i^{\,j}$$

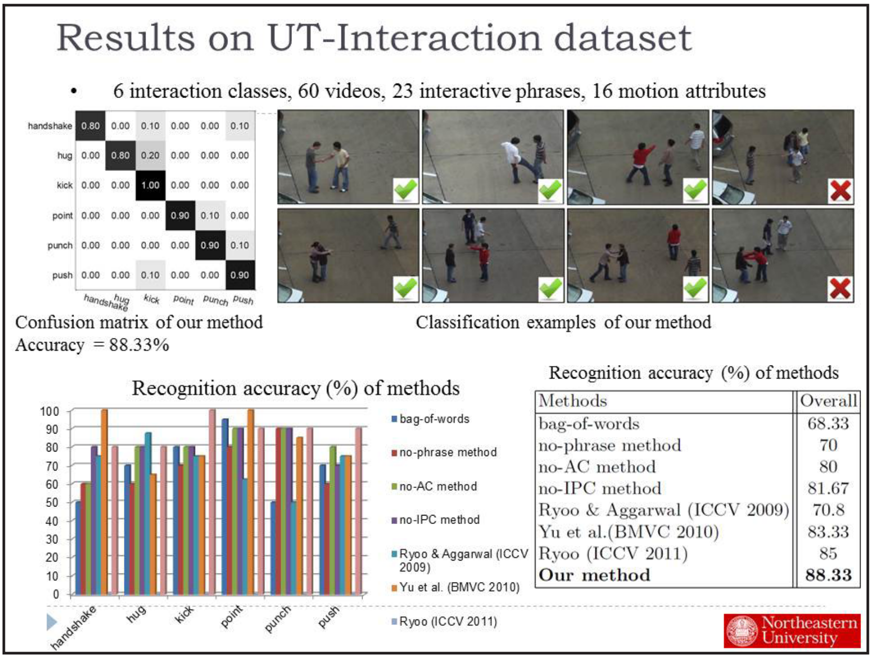
$$\min rank(M_T) + \eta E$$

s.t. $M = M_T + E$



Stabilized Manifold Learning

$\varepsilon(W) = \Sigma_i |x_i - \Sigma_j W_{ij} x_j|^2$



Results on UT-Interaction dataset

- 6 interaction classes, 60 videos, 23 interactive phrases, 16 motion attributes

	handshake	hug	kick	point	punch	push
handshake	0.80	0.00	0.10	0.00	0.00	0.10
hug	0.00	0.80	0.20	0.00	0.00	0.00
kick	0.00	0.00	1.00	0.00	0.00	0.00
point	0.00	0.00	0.00	0.90	0.10	0.00
punch	0.00	0.00	0.00	0.00	0.90	0.10
push	0.00	0.00	0.10	0.00	0.00	0.90

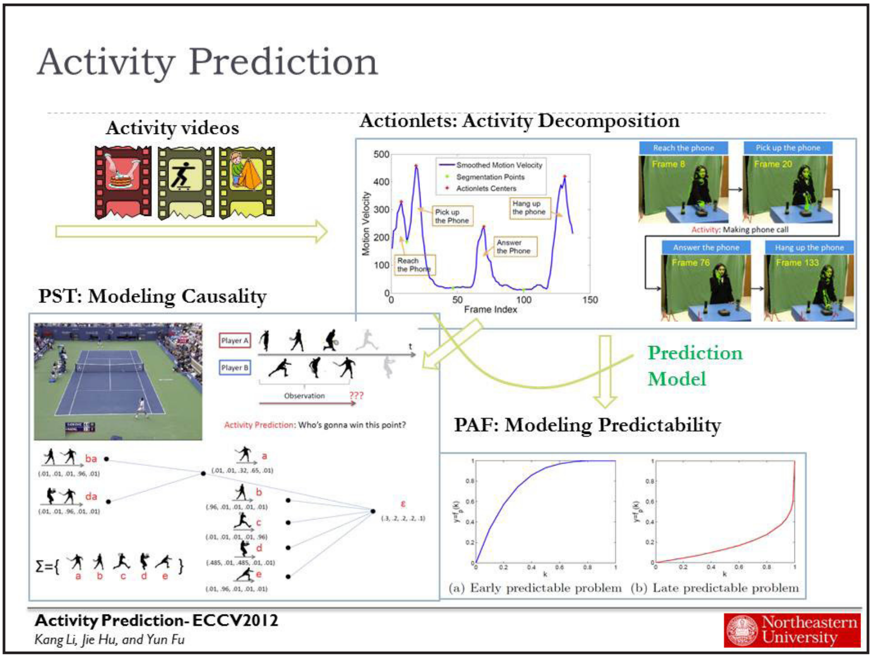
Confusion matrix of our method
Accuracy = 88.33%

Classification examples of our method

Recognition accuracy (%) of methods

Recognition accuracy (%) of methods

Methods	Overall
bag-of-words	68.33
no-phrase method	70
no-AC method	80
no-IPC method	81.67
Ryoo & Aggarwal (ICCV 2009)	70.8
Yu et al.(BMVC 2010)	83.33
Ryoo (ICCV 2011)	85
Our method	**88.33**



Activity Prediction

Activity videos

Actionlets: Activity Decomposition

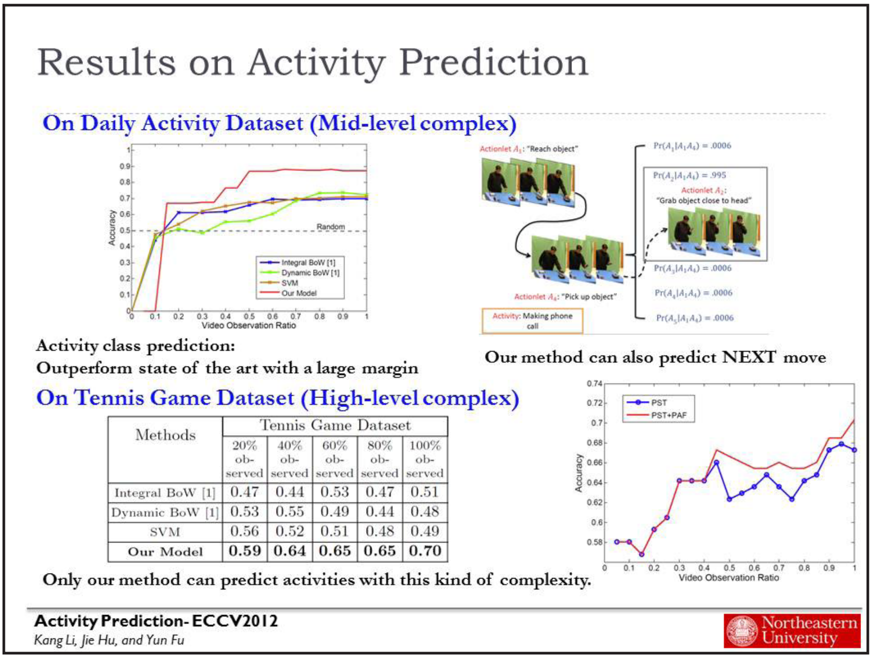
PST: Modeling Causality

Prediction Model

PAF: Modeling Predictability

Activity Prediction- ECCV2012
Kang Li, Jie Hu, and Yun Fu

Northeastern University



Results on Activity Prediction

On Daily Activity Dataset (Mid-level complex)

Activity class prediction:
Outperform state of the art with a large margin

Our method can also predict NEXT move

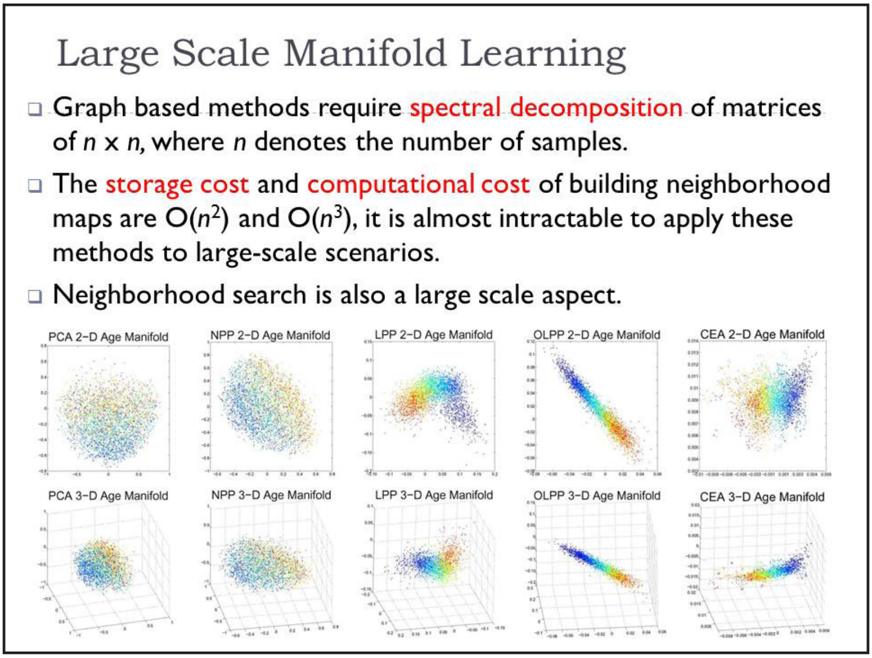
On Tennis Game Dataset (High-level complex)

Methods	Tennis Game Dataset				
	20% observed	40% observed	60% observed	80% observed	100% observed
Integral BoW [1]	0.47	0.44	0.53	0.47	0.51
Dynamic BoW [1]	0.53	0.55	0.49	0.44	0.48
SVM	0.56	0.52	0.51	0.48	0.49
Our Model	**0.59**	**0.64**	**0.65**	**0.65**	**0.70**

Only our method can predict activities with this kind of complexity.

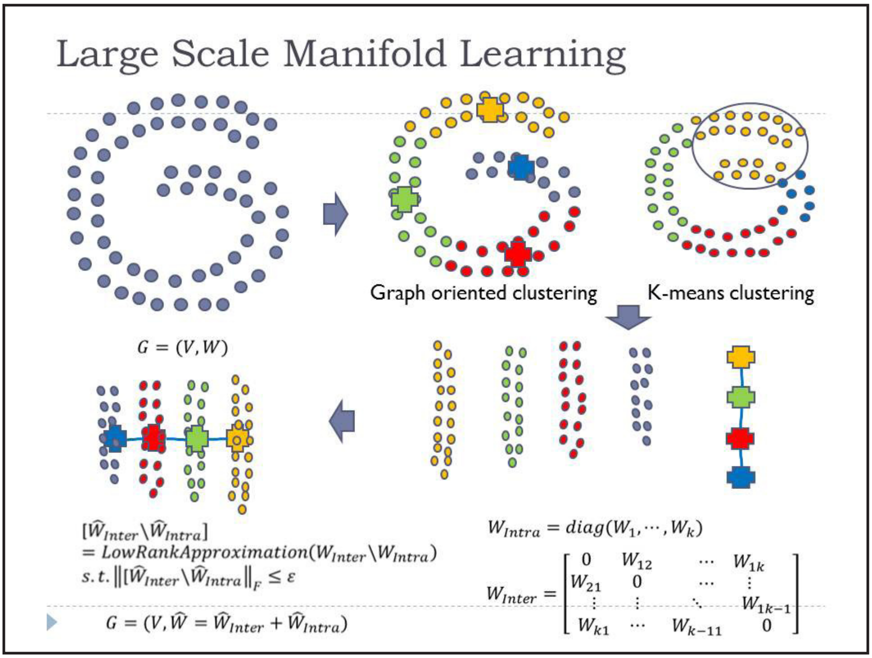
Activity Prediction- ECCV2012
Kang Li, Jie Hu, and Yun Fu

Northeastern University



Large Scale Manifold Learning

- Graph based methods require spectral decomposition of matrices of *n* x *n*, where *n* denotes the number of samples.
- The storage cost and computational cost of building neighborhood maps are O(n^2) and O(n^3), it is almost intractable to apply these methods to large-scale scenarios.
- Neighborhood search is also a large scale aspect.



Large Scale Manifold Learning

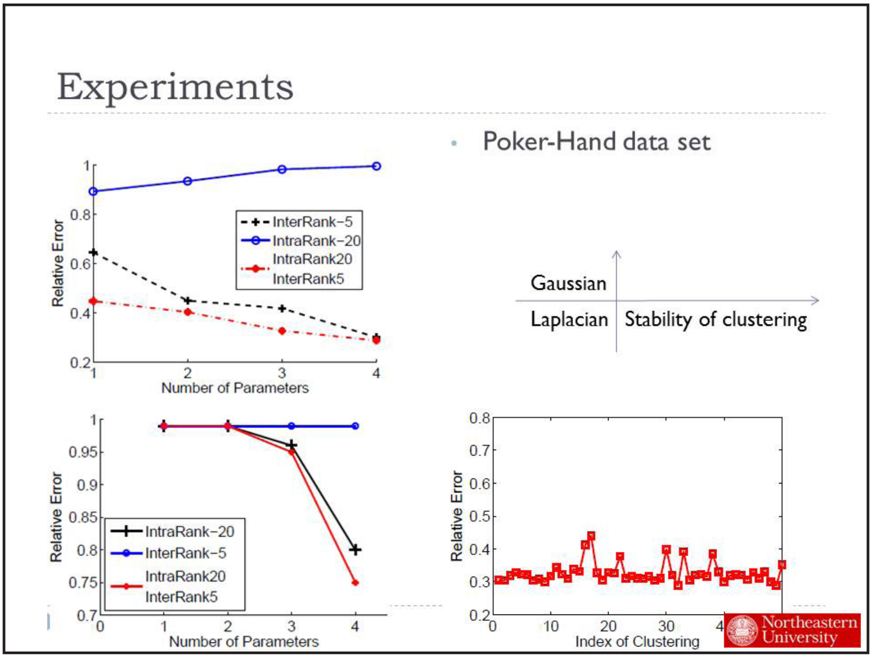
$G = (V, W)$

$$[\widehat{W}_{Inter} \backslash \widehat{W}_{Intra}] = LowRankApproximation(W_{Inter} \backslash W_{Intra})$$
$$s.t. \left\| [\widehat{W}_{Inter} \backslash \widehat{W}_{Intra} \right\|_F \leq \varepsilon$$

$$W_{Intra} = diag(W_1, \cdots, W_k)$$

$$W_{Inter} = \begin{bmatrix} 0 & W_{12} & \cdots & W_{1k} \\ W_{21} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & W_{1k-1} \\ W_{k1} & \cdots & W_{k-11} & 0 \end{bmatrix}$$

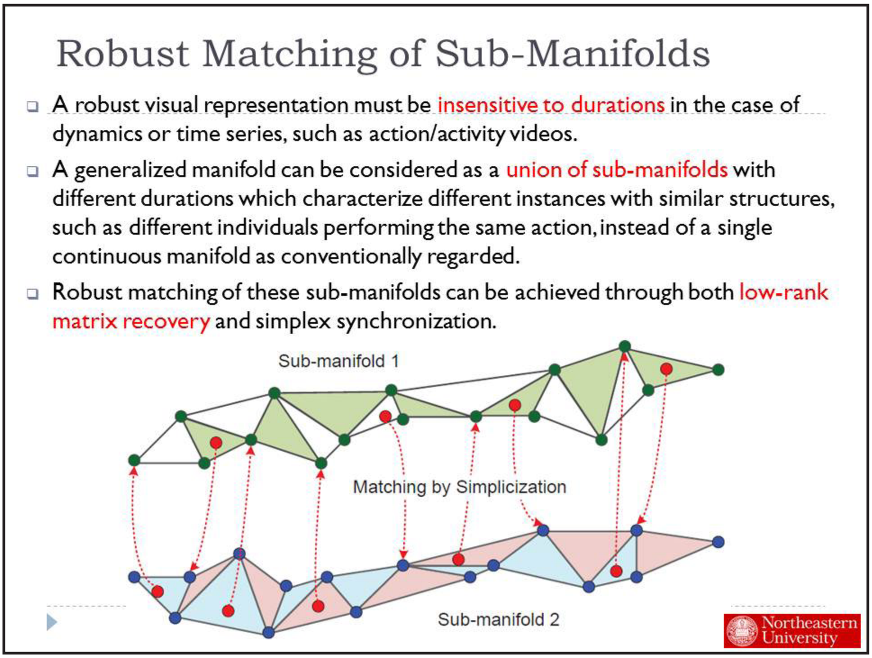
- $G = (V, \widehat{W} = \widehat{W}_{Inter} + \widehat{W}_{Intra})$



Experiments

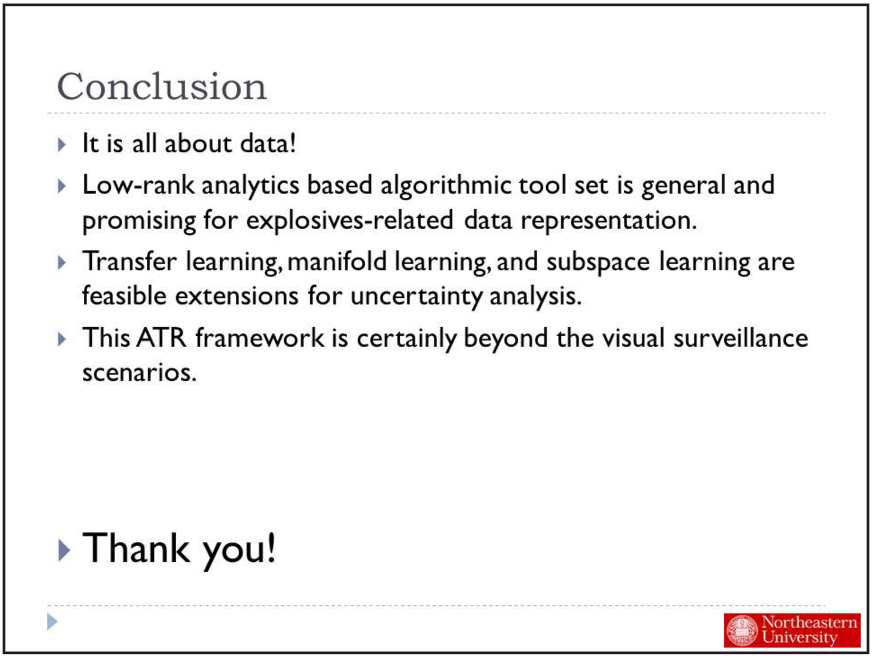
- Poker-Hand data set

Gaussian
Laplacian
Stability of clustering



Robust Matching of Sub-Manifolds

- A robust visual representation must be insensitive to durations in the case of dynamics or time series, such as action/activity videos.
- A generalized manifold can be considered as a union of sub-manifolds with different durations which characterize different instances with similar structures, such as different individuals performing the same action, instead of a single continuous manifold as conventionally regarded.
- Robust matching of these sub-manifolds can be achieved through both low-rank matrix recovery and simplex synchronization.



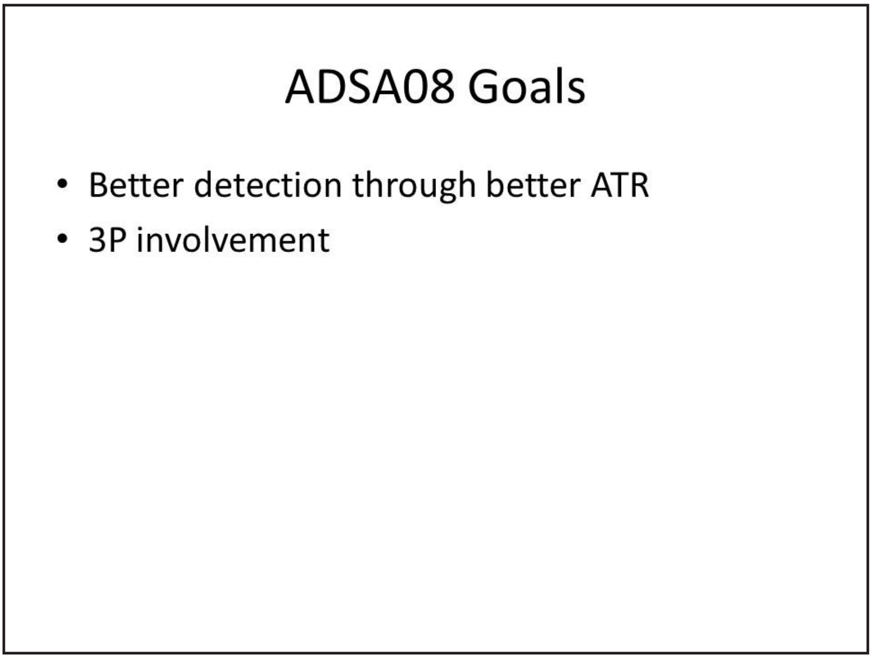
Conclusion

- It is all about data!
- Low-rank analytics based algorithmic tool set is general and promising for explosives-related data representation.
- Transfer learning, manifold learning, and subspace learning are feasible extensions for uncertainty analysis.
- This ATR framework is certainly beyond the visual surveillance scenarios.

- Thank you!

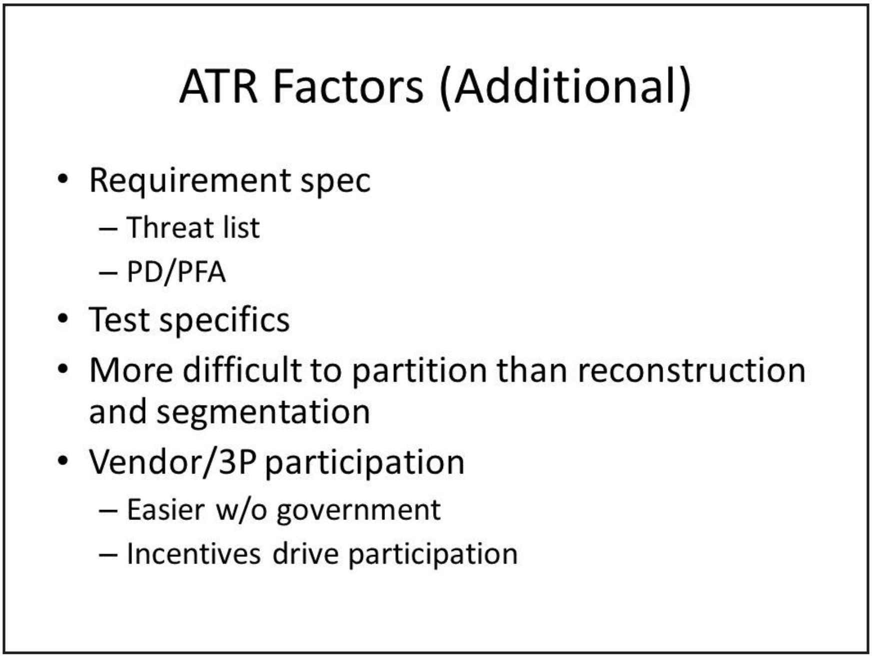
16.32 Carl Crawford: Next Steps

Third Party ATR Development



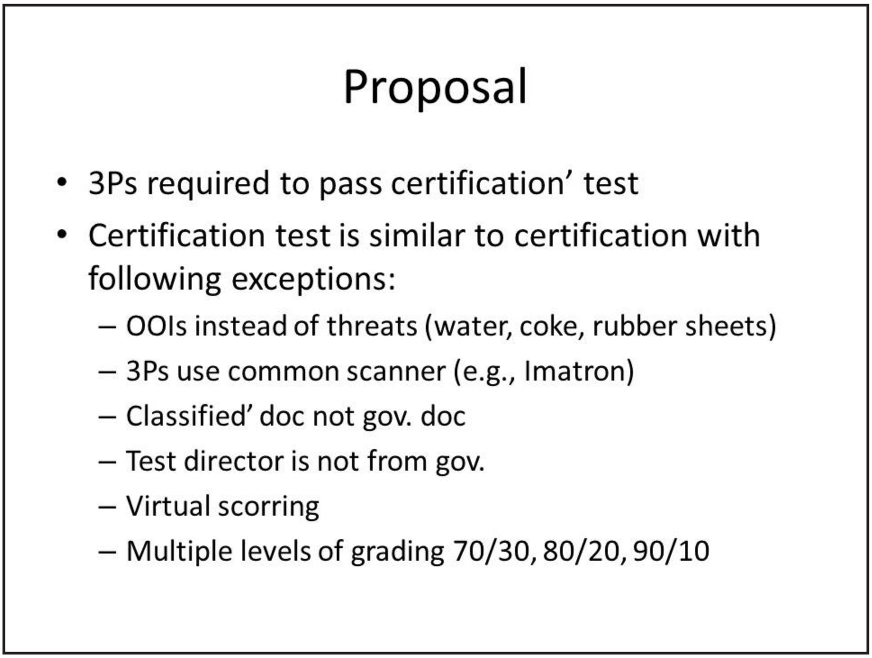
ADSA08 Goals

- Better detection through better ATR
- 3P involvement



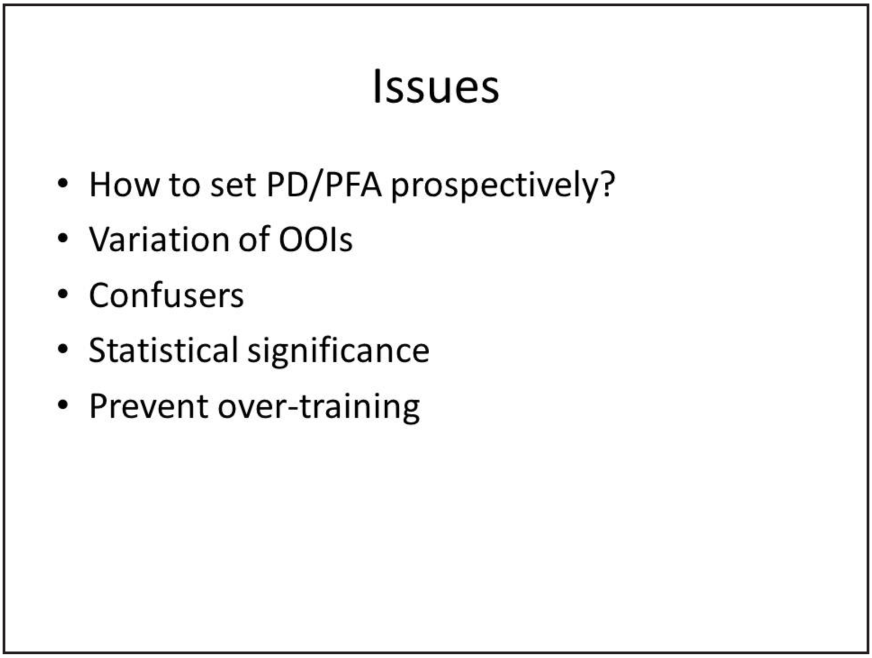
ATR Factors (Additional)

- Requirement spec
  - Threat list
  - PD/PFA
- Test specifics
- More difficult to partition than reconstruction and segmentation
- Vendor/3P participation
  - Easier w/o government
  - Incentives drive participation



Proposal

- 3Ps required to pass certification' test
- Certification test is similar to certification with following exceptions:
  - OOIs instead of threats (water, coke, rubber sheets)
  - 3Ps use common scanner (e.g., Imatron)
  - Classified' doc not gov. doc
  - Test director is not from gov.
  - Virtual scorring
  - Multiple levels of grading 70/30, 80/20, 90/10



Issues

- How to set PD/PFA prospectively?
- Variation of OOIs
- Confusers
- Statistical significance
- Prevent over-training

ALERT
AWARENESS AND LOCALIZATION
OF EXPLOSIVES-RELATED THREATS

Awareness and Localization of Explosives-Related Threats

Northeastern University — 360 Huntington Avenue — Boston MA 02115
phone: 617.373.4673 — fax: 617.373.8627 — web: www.neu.edu/alert

This material is based upon work supported by the U.S. Department of Homeland Security under Award Number 2008-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.