

Video Tracking and Active Learning

Robot Vision Lab

Purdue University

<http://RVL.ecn.purdue.edu>

Avi Kak (kak@purdue.edu)

Noha Elfiky (nelfiky@purdue.edu)

Tracking Passengers and Their Divested Items

1. It is one thing to design a tracking system that would work for a specific pattern of indoor illumination, for a specific checkpoint layout, for a specific floor pattern, etc., and entirely another to design a system that can adapt with minimal human input to different environmental conditions.
2. An additional critical attribute of a tracking system is its resilience to target obscuration.
3. Our experience has shown that good trackers must carry out dynamic foreground and background modeling to maintain a lock on the targets.

Performance Evaluation --- Algorithmic Challenges

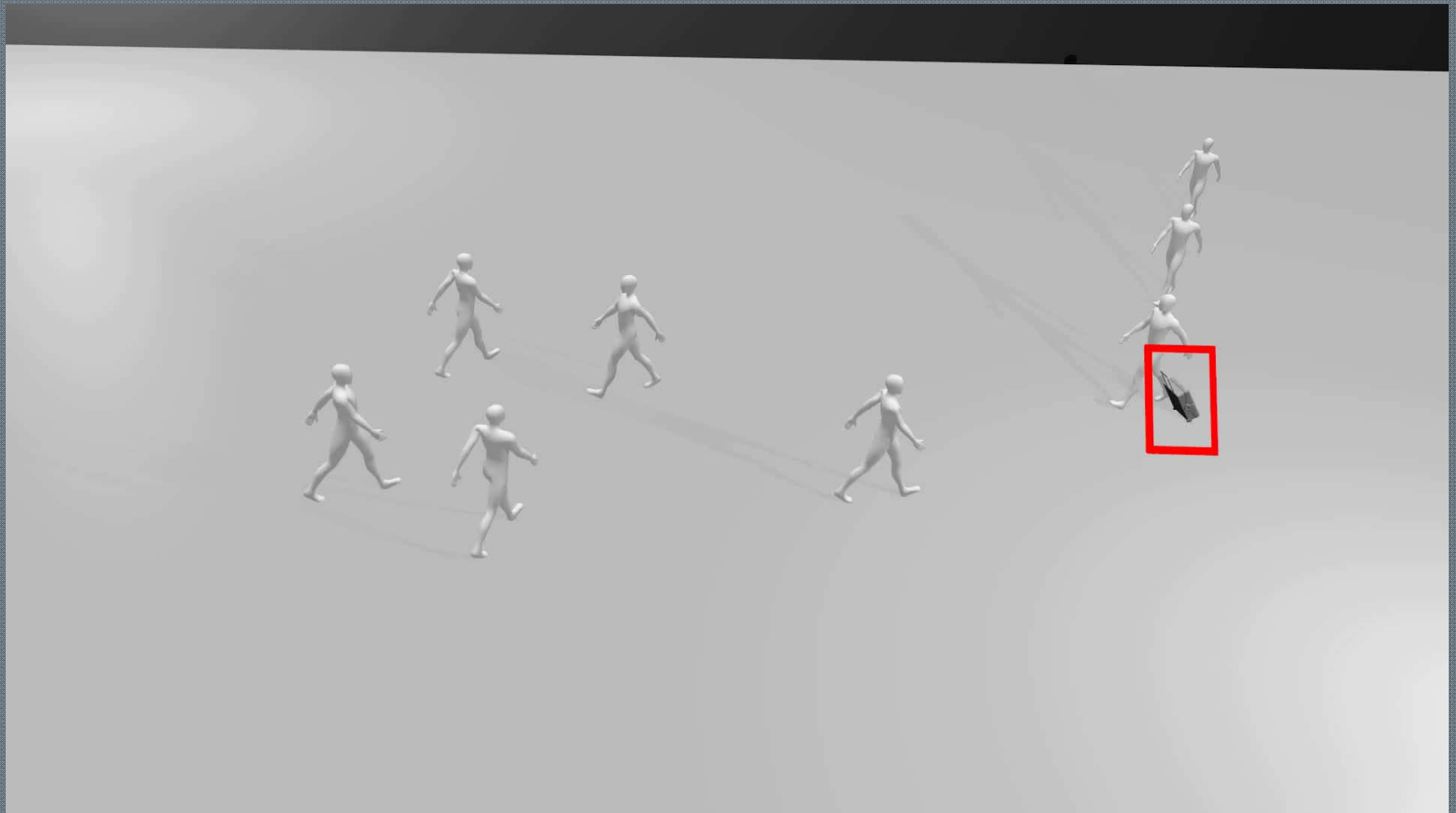
1. Before a tracker can be deployed at a passenger security checkpoint, its performance must be quantified with respect to all of the confounding variables.
2. While real videos make for impressive demonstrations, they do not tell us how a tracker would fare when the appearance parameters for the passengers and their possessions are allowed to vary over a wide range.
3. We can now use modern computer graphics tools to create highly realistic videos with quantified degrees of variability with respect to target appearance, illumination, background, target obscuration, etc. These can subsequently be used to evaluate the tracking algorithms.

Organization of the Rest of the Presentation

1. I'll first show a couple of synthetically generated rudimentary videos just to point to the sort of problems we need to address in tracking passengers and their divested items at airport security checkpoints.
2. I'll then go over the current state of the art with regard to what can be done today with trackers that are based on real-time dynamic modeling of foreground/background pixels as targets are being tracked.
3. I'll then say a couple of words about our ongoing work on recognizing faces by combining evidence from multiple viewpoints.
4. Subsequently, I'll go over our Active Learning Framework meant to reduce the burden on humans when it comes to generating the ground-truth data for training ML algorithms.
5. Finally, I'll present "What Can Purdue RVL Bring to the DHS/TSA Table?"

Tracking a Roller Bag and Detecting its Hand-Off

Cheaper way to generate data and can include exceptions/events easily



We Must Combine Evidence from Multiple Cameras

Combining Evidence For Faces is Best Done on Manifolds

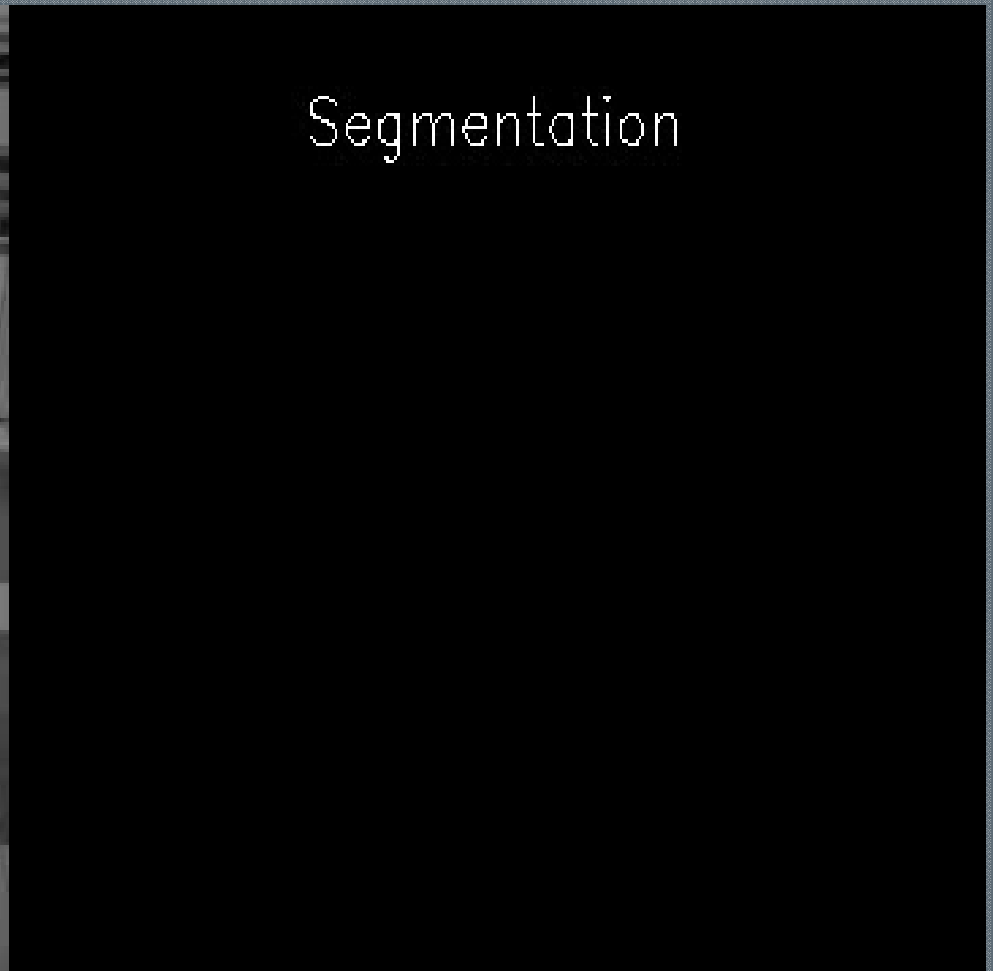


Where Is the State-of-the-Art in Multi-Camera Tracking of Objects and People?

1. Results based on a new approach to tracking: Combined segmentation and tracking (CVPR2010, ECCV2010)
2. Tracking multiple vehicles simultaneously in wide-area aerial video
3. Tracking vehicles through shadows and occlusions in wide-area aerial video
4. Tracking multiple humans with a network of cameras
5. Tracking rigid objects in real time (with pose estimation)

Combined Tracking and Segmentation

A New Approach to Tracking



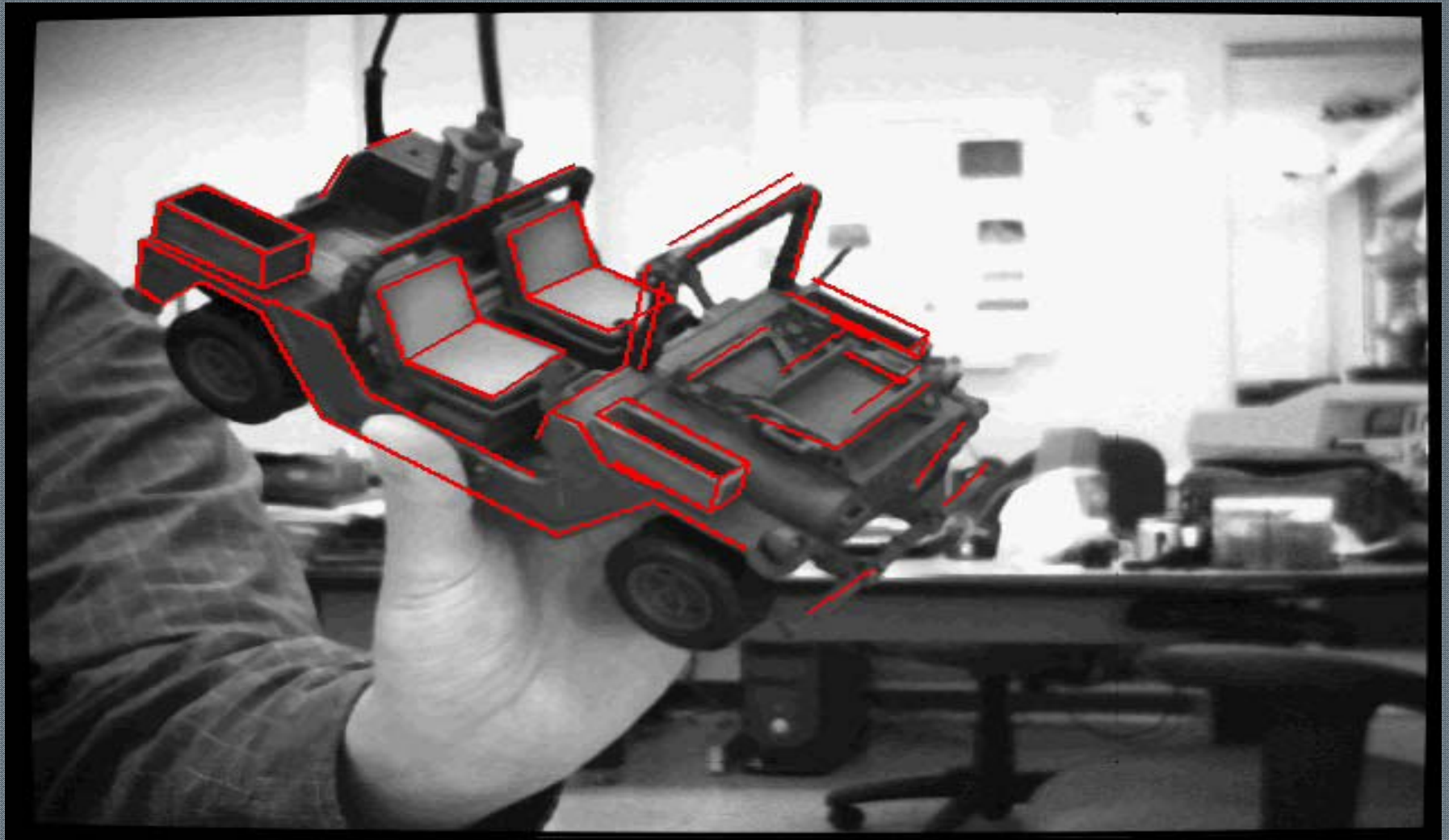
State of the Art: Tracking Vehicles in Wide-Angle Aerial Imagery



State of the Art: Tracking Vehicles Through Shadows and Obscurations



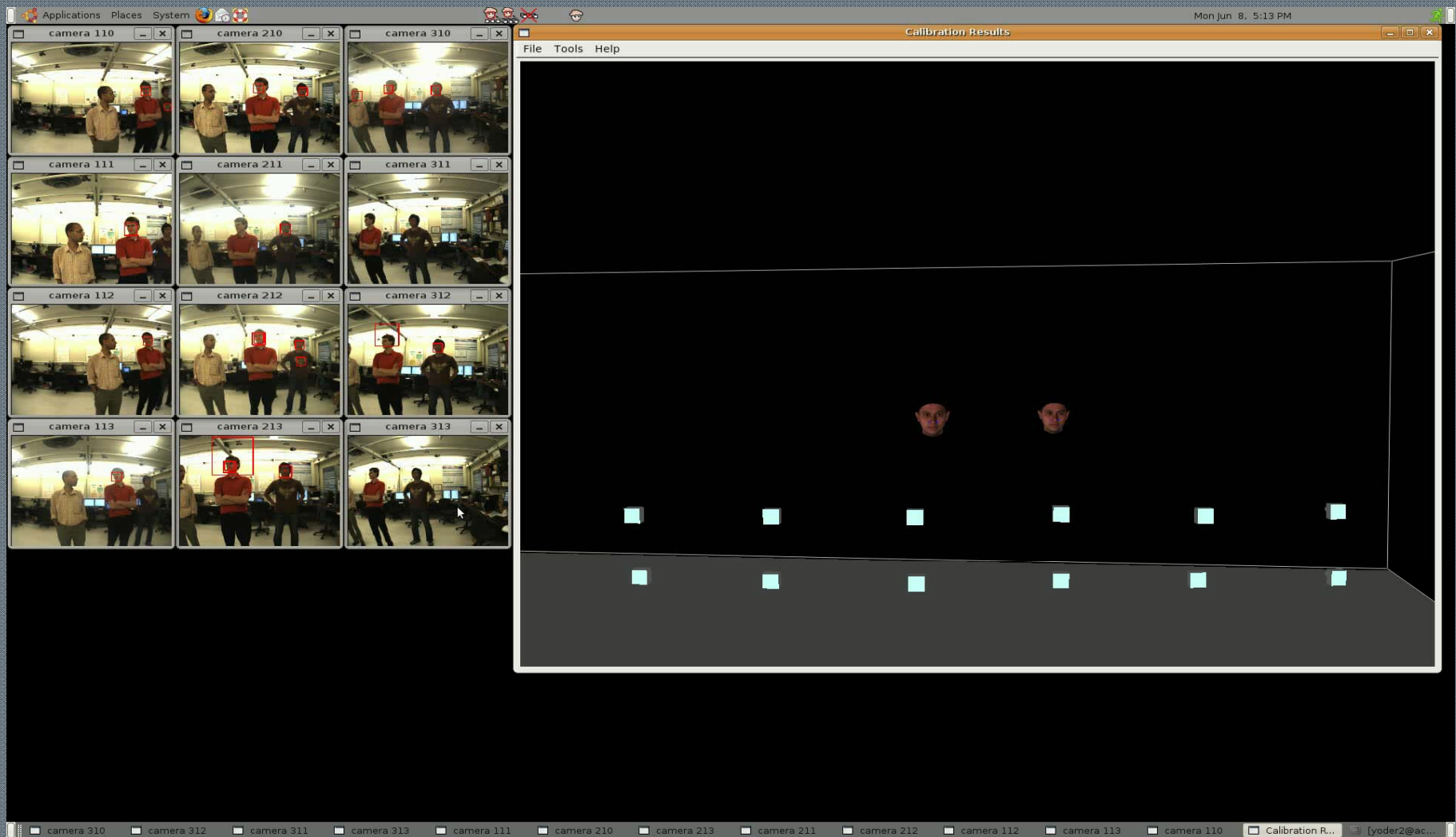
Tracking 3D Objects and Estimating Their Poses in Real Time



Real-Time Tracking of Faces

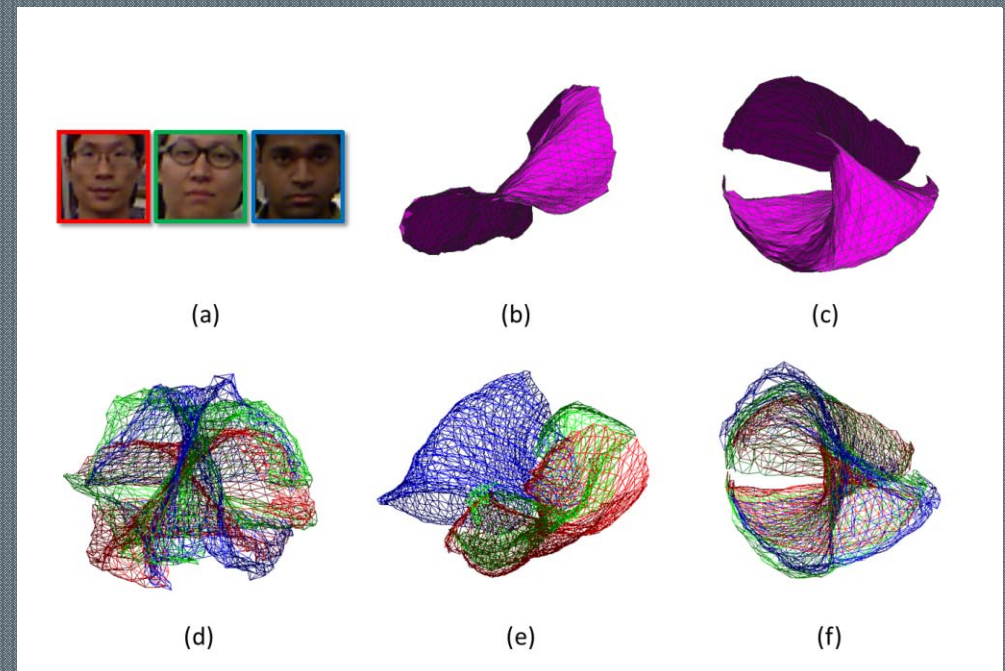
The Intelligent Shelf Project

Tracking Faces with Multiple Cameras



You Need to Cluster Camera Data on a Manifold When Combining Face Evidence from Multiple Viewpoints

1. A universally accepted tenet in face recognition community is that when you look at the same face from different viewpoints, the data you get resides on a manifold.
2. At Purdue, we have developed powerful algorithms for clustering information on manifolds.
3. A manifold, while being locally Euclidean, is non-Euclidean in the original measurement space.
4. So popular approaches like PCA and LDA for dimensionality reduction do not work



Active Learning Framework for Designing Event Detectors for Videos

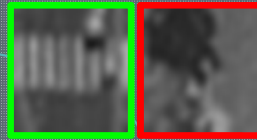
A big lesson learned from our ongoing IARPRA FINDER project:

- It can be extremely fatiguing for humans to generate the ground-truth data for training machine learning algorithms for complex applications.
- After marking a few samples as positive or negative, the human mind becomes a poor judge of whether a new sample is similar to those already marked or has anything new to contribute to the training dataset. As a result, the human-generated training datasets tend to be highly redundant.
- As a part of our FINDER project, we have developed an Active Learning framework in which, at the beginning, the human only supplies strongly positive and strongly negative samples to the system.
- Subsequently, the system asks for help only on an as-needed basis.

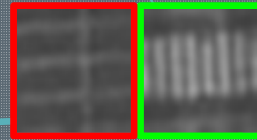
Active Learning: An Introduction



Find and label initial set of samples

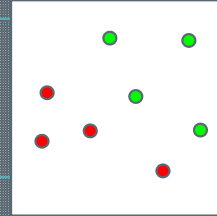


Label uncertain samples

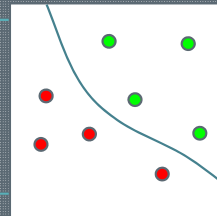


Train detector:

Extract Features

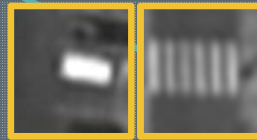
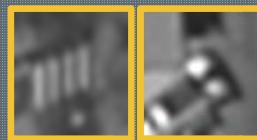
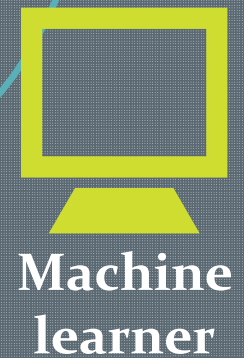
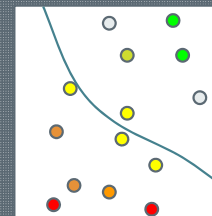


Train Classifier



Apply detector to new regions

Select samples with uncertain estimated labels



Active Learning for Detecting Roller-Bag Hand-Off

A Human shows strongly positive and negative examples to the system

Active Learning System
Creates Initial Decision
Surface from the Flagged
Video Segments

When it Encounters Video
Segments That are Too Close to
the Decision Surface, it Elicits
help from a human

Subsequently, it improves the
Decision Surface Based on the
feedback received from the human

What Can Purdue RVL Bring to the DHS/TSA Table?

1. We have over 20 years of experience in creating computer-vision based trackers for our industrial and government sponsors.
2. We will use computer graphics to create evaluation datasets for security checkpoint tracking of passengers and their divested items.
3. We will use our dynamic foreground/background estimation based framework to design trackers for passenger checkpoint applications.
4. We will use video event detection algorithms to detect changes in the states of the items divested by passengers.
5. Given the visual complexity of a security check point, most ML algorithms will require large amounts of human-annotated training data. We will use our Active Learning framework to mitigate the human burden involved.