Sparse Angle Volumetric Reconstruction for Air Cargo CT

Presented by

David Castañón -- Boston University
François Zayek – Astrophysics, Inc.

August 5, 2020

Acknowledgement

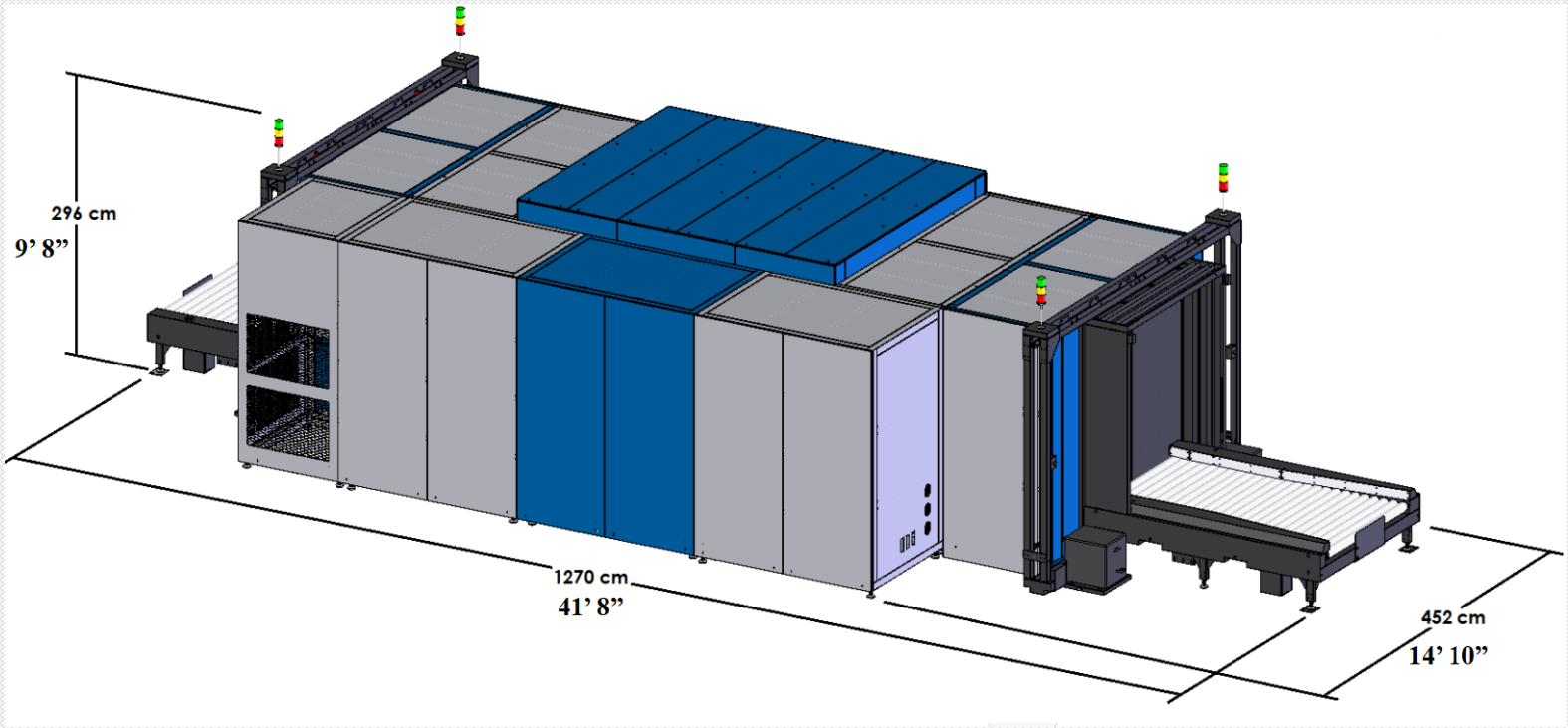
- This work was supported by DHS as part of the ALERT Center of Excellence, and by ASTROPHYSICS, as an ALERT industry partner

- This work was a collaboration between ALERT and ASTROPHYSICS, Inc involving numerous personnel
  - Thanks to all students and engineers involved.

- We want to acknowledge the major contribution of Dr. Fernando Quivira, who was an integral part of developing and implementing the algorithms we will be discussing today

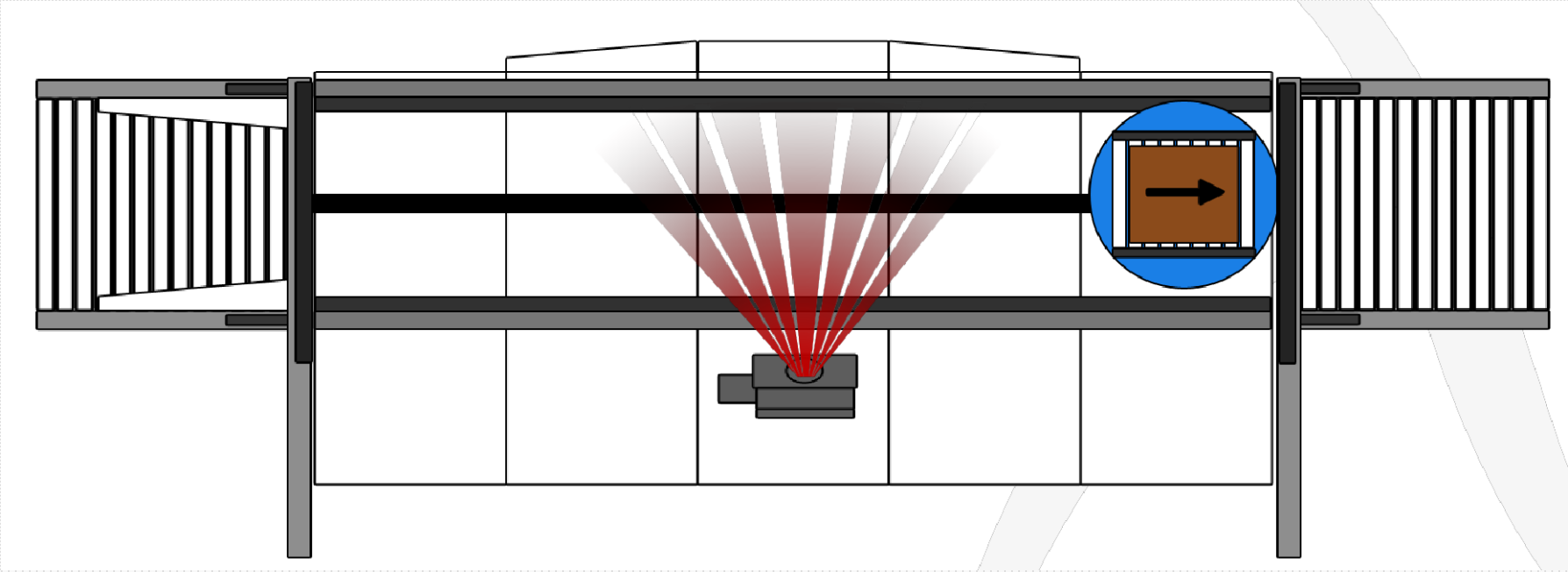
Problem of Interest

- CT inspection of Air Cargo Pallets
  - Detection and identification of materials of interest in containers
- Multi View CT (MVCT) Air Cargo Scanner
  - Dual-energy imager generating multiple radiographic projection images plus volumetric reconstruction



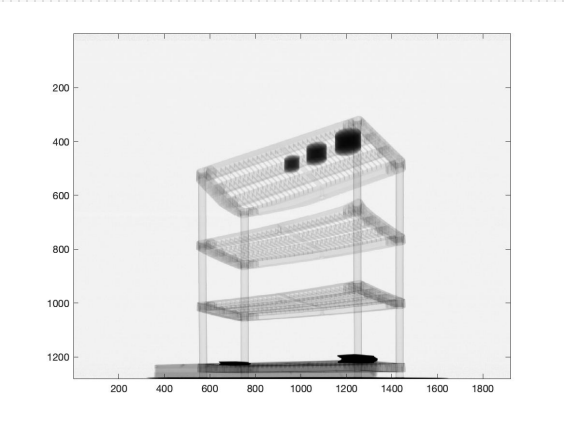
How MVCT Operates

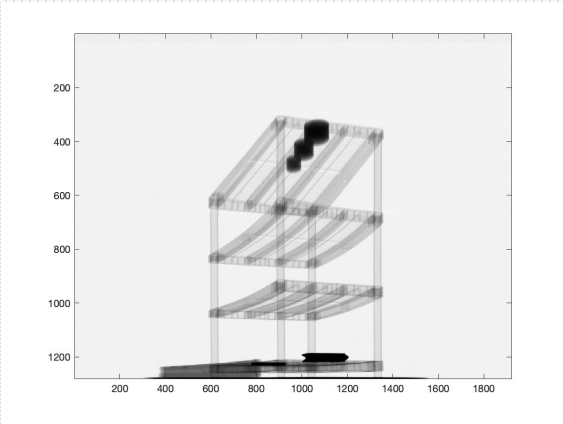
- Cargo loaded onto belt, scanned using sparse beam array on each pass
- Rotated after a pass, scanned getting different projections on subsequent passes

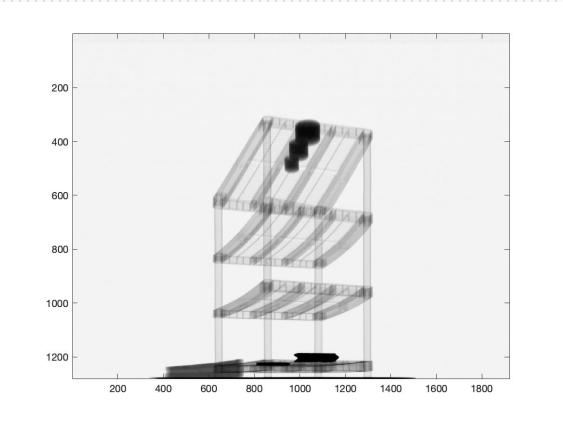


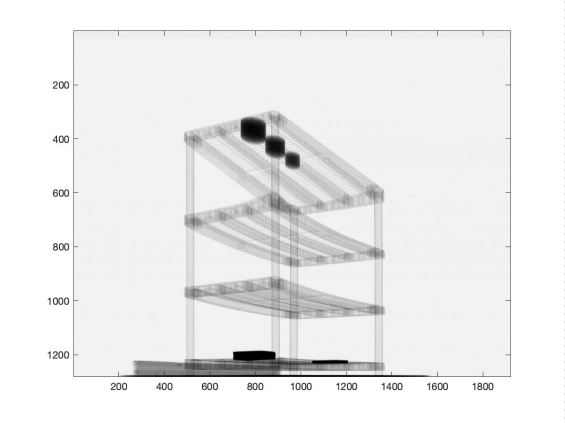
Sample projection data

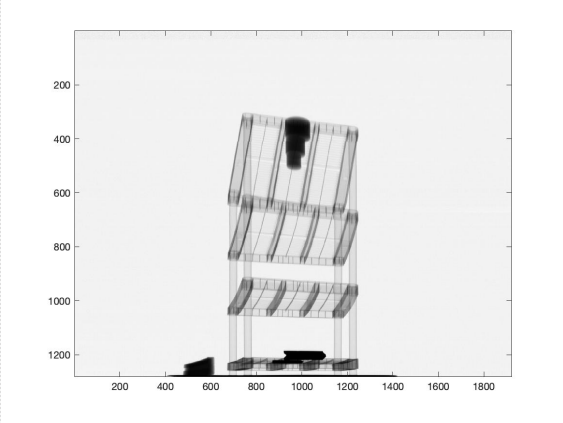
- Raw input to reconstruction – dual energy projections

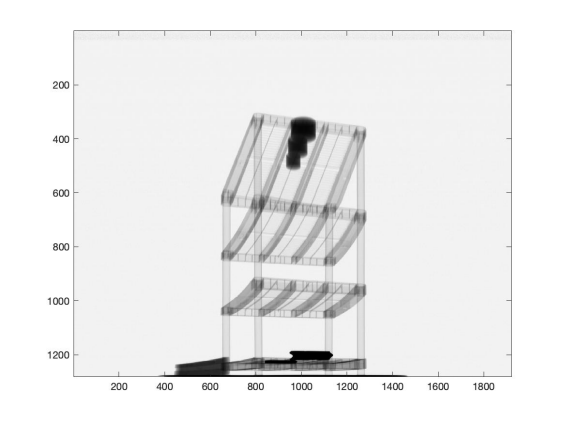






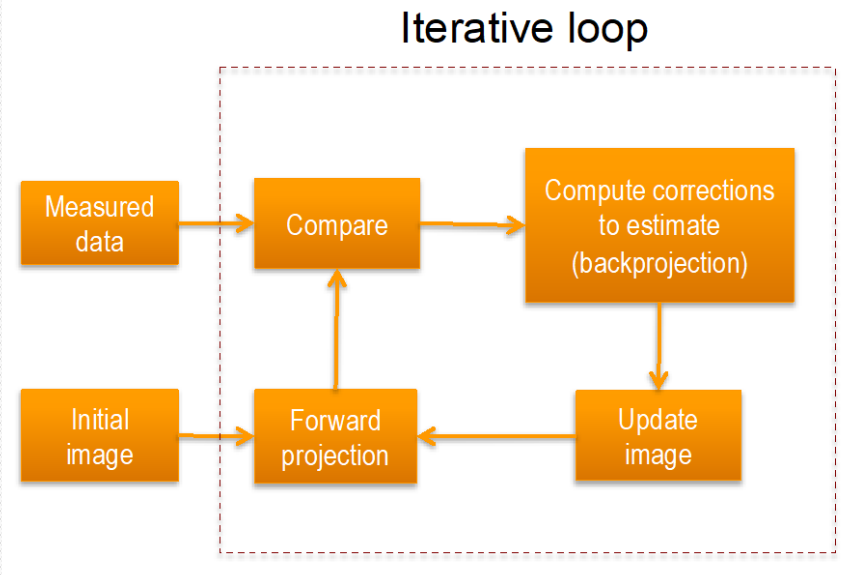






Challenges for Reconstruction

- Sparse projections requires volumetric reconstruction, with ~10^8 unknowns
  - Hard to partition reconstruction into regions
- Irregular sampling makes direct reconstruction (e.g. filtered backprojection) hard to apply
- Iterative reconstruction:
  - Algebraic techniques
  - Exploit models of sensor-source locations
  - MAP reconstruction using optimization



$$\arg\max_{\text{img}} \log P(\text{meas}|\text{img}) + f(\text{img})$$

Algorithm Model

- Statistical measurement model in photon counts (Beer-Lambert)

$$Y_i \sim \text{Poisson}\{b_i e^{-A_i \mu} + r_i\}$$

  - b_i is source flux, r_i is background flux model
  - μ is attenuation volumetric field we wish to image
  - A_i is line integral for detector measurement I, depending on source position, detector position, object rotation and translation on belt

- Resulting optimization for iterative reconstruction

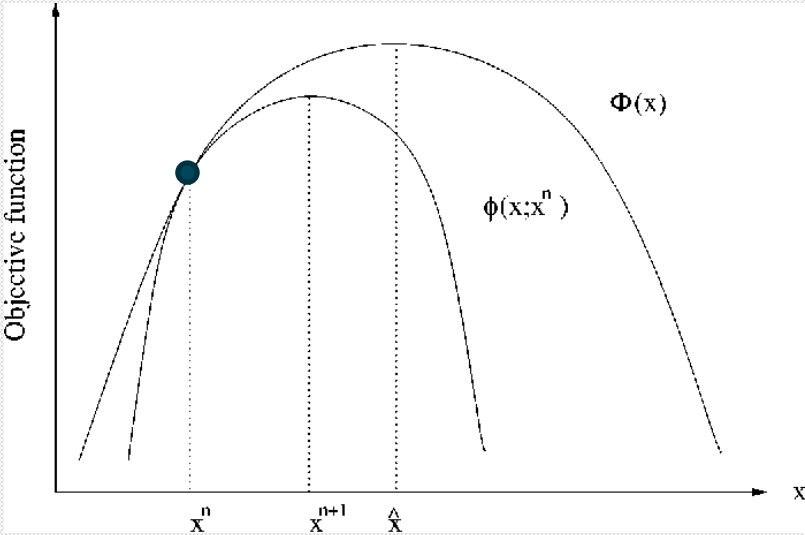
$$\min_{\mu \geq 0} \sum_{i=1}^{N} b_i e^{-\ell_i} + r_i - Y_i \log(b_i e^{-\ell_i} + r_i) + \frac{\beta}{2} \sum_{j=1}^{P} \sum_{k \in N_j} w_{jk} \psi(\mu_j - \mu_k)$$

$$\ell_i = A_i \mu, \;\; A_i \text{ tomographic line integral}; \quad N_j \text{ neighborhood of voxel } j$$

$$\psi(\cdot) = \text{ regularization function}, \quad w_{jk} \text{ directional weights}$$

Algorithm Choice

- Desire high degree of parallelism for executing iterations
  - Number of measurements N, voxels M is $O(10^8)$
- Would also like to pipeline data processing with data collection to avoid reconstruction delays
- Algorithm choice: Separable Paraboloidal surrogates (Erdogan-Fessler '99)
  - Given existing solution guess μ^{old}, compute separable paraboloidal upper bound of objective function
  - Surrogate in 10^8 dimensions
  - Update solution by optimizing separable bound: can be done voxel by voxel because of separability
  - Voxel-by-voxel optimization update in closed form (quadratic optimization, no line search)
  - Guarantees of convergence



Computations (simplified)

- For each measurement Y_i, forward project current field estimate, compute predicted measurement error

$$\hat{\mu} = \mu^{\text{old}};\ \ell_i = A_i\hat{\mu};\ \hat{y}_i = b_i e^{-\ell_i},\ i = 1, \ldots, N$$

  - Data parallel prediction…forward line integral projection coefficients computed on the fly for each thread (cheaper than storage)
  - Data retrieval of relevant field values into each parallel thread is slow point
- Optimization update: For each voxel j, update field estimate using predicted measurement errors

$$num(j) = \sum_i a_{ij}(\hat{y}_i - Y_i); \quad den(j) = \sum_i a_{ij}\ell_i\hat{y}_i$$

$$\mu_j^{\text{new}} = \left[\mu^{\text{old}}(1 + \frac{num(j)}{den(j)})\right]_+$$

  - Data parallel, but voxels must access measurements with line integrals intersecting voxels → irregular data movements make it slower

Pipeline of Computations

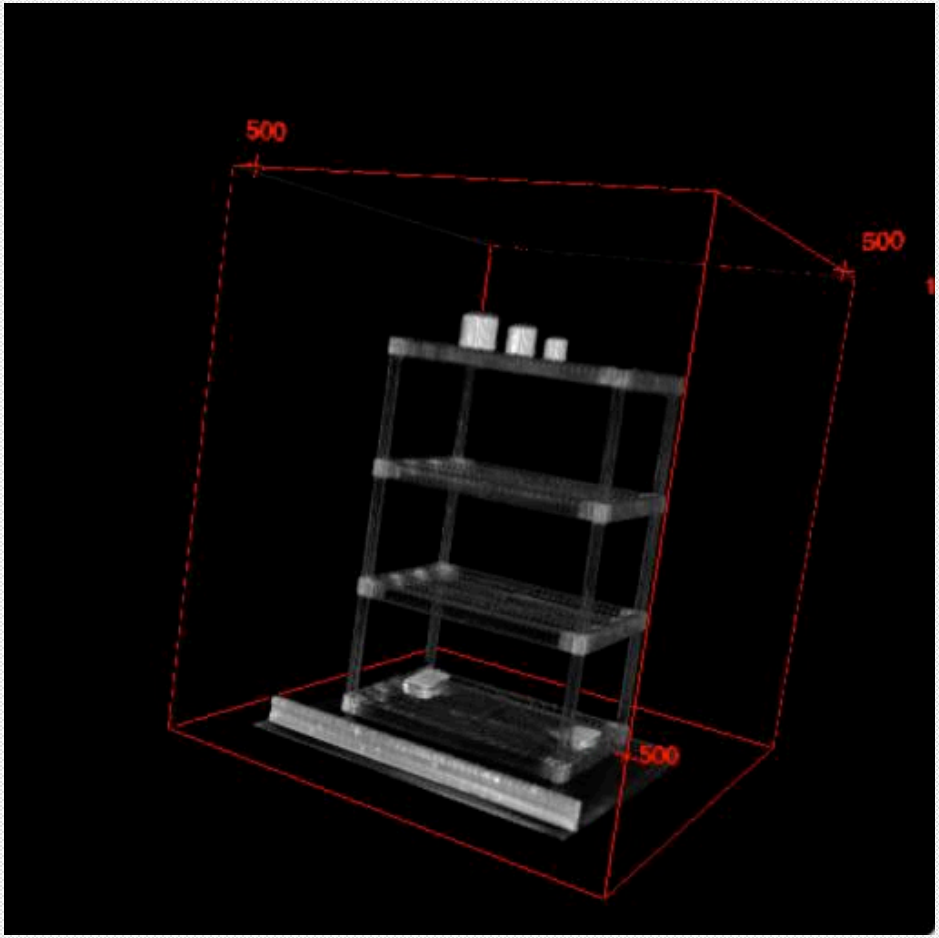
- Goal: perform useful computations during measurements
- Approach: ordered subsets
  - Begin iterations using subsets of measurements from each pass
  - Image reconstructed from interim measurements, refined using measurements from most recent passes
  - Use last pass for image refinement
- Advantage: simpler forward projection technique, less measurements used in each pass implies less data transfer to GPU memory
- Design how much processing needs to be done, order in which passes get processed

Implementation in GPUs

- Target GPU: NVIDIA gaming GeForce family
  - Current version: NVIDIA GeForce RTX 2080 Ti
    - 11GB of GDDR6 VRAM, 4,352 CUDA cores and a boost clock of 1,635MHz.
    - Number of CUDA cores indicates number of parallel threads we can compute
- Implementation in CUDA language
  - Forward projection: Divide measurements into blocks of threads for CUDA execution.
  - Each measurement computes the projection row A_i and accesses reconstructed field values to predict measurements
  - Back projection and field update: Divide voxels into blocks of threads. Threads recompute column elements a_{iij} , loop compute numerator and denominator from measurements
  - Threads update individual voxels in parallel

Results to date:

- Full reconstruction of $5x10^7$ voxels, dual energy, in 5.5 seconds using 2 GPUs
  - With pipelining, can reduce further so delay is minimal
  - Can also increase resolution and maintain throughput requirements
- Provides volumetric view to operators as complement to projections



Extensions

- Improved mapping onto GPU structures
  - Slow part: accessing the field values for projection and backprojection
  - Exploring use of texture memory, other parts of GPUs
- Anisotropic, edge-preserving regularization coupling high/low reconstruction
  - Demonstrated improvement in contrast
  - Need to design efficient parallel implementation on GPUs
- Basis decomposition techniques for reconstruction of coefficient images
  - Useful for beam-hardening corrections
  - Have designed fully parallel decomposition algorithms, small overhead over current version

Conclusion

- Translating university research into commercial products is a challenging task
  - Experimental data from implemented systems often deviates from assumptions
  - Application requires system engineering tradeoffs in speed, accuracy
- This work presented an opportunity to transition new concepts in multi-spectral sparse CT imaging into an important application: CT inspection of air cargo
- The collaboration between ALERT and ASTROPHYSICS continues, as new directions (requirements for algorithms) emerge from interactions with customers