BAE SYSTEMS

Appearance-based Object Reacquisition for Mobile Manipulation

Matthew R. Walter[1]
Yuli Friedman[2]
Matthew Antone[2]
Seth Teller[1]

Presenter: Mark Keck[2]

[1]MIT CSAIL

[2]BAE Systems AIT

Presentation Outline

- Motivation: task autonomy
- System-level concept
- Reacquisition method
- Experiments and results
- Discussion

Problem Domain

- Developing robotic forklift
  - Capable of autonomous single-pallet outdoor transport
  - Previously required user intervention throughout
- Goal of work is to enable task-level autonomy
  - Higher-level command strings
  - Implies need to reacquire previously seen objects upon revisit

Overview and Goals

- Use sensing to increase autonomy
  - Move from atomic tasks to task sequences
  - Reduce interruptions to user workflow
- Employ visual descriptions of manipulands
  - Learn appearance of each user-selected object
    - Object "model" consists of multiple "views"
    - Each view captures appearance at unique pose
  - Reacquire object locations automatically upon revisit
- Challenges: varying lighting, viewpoint over time

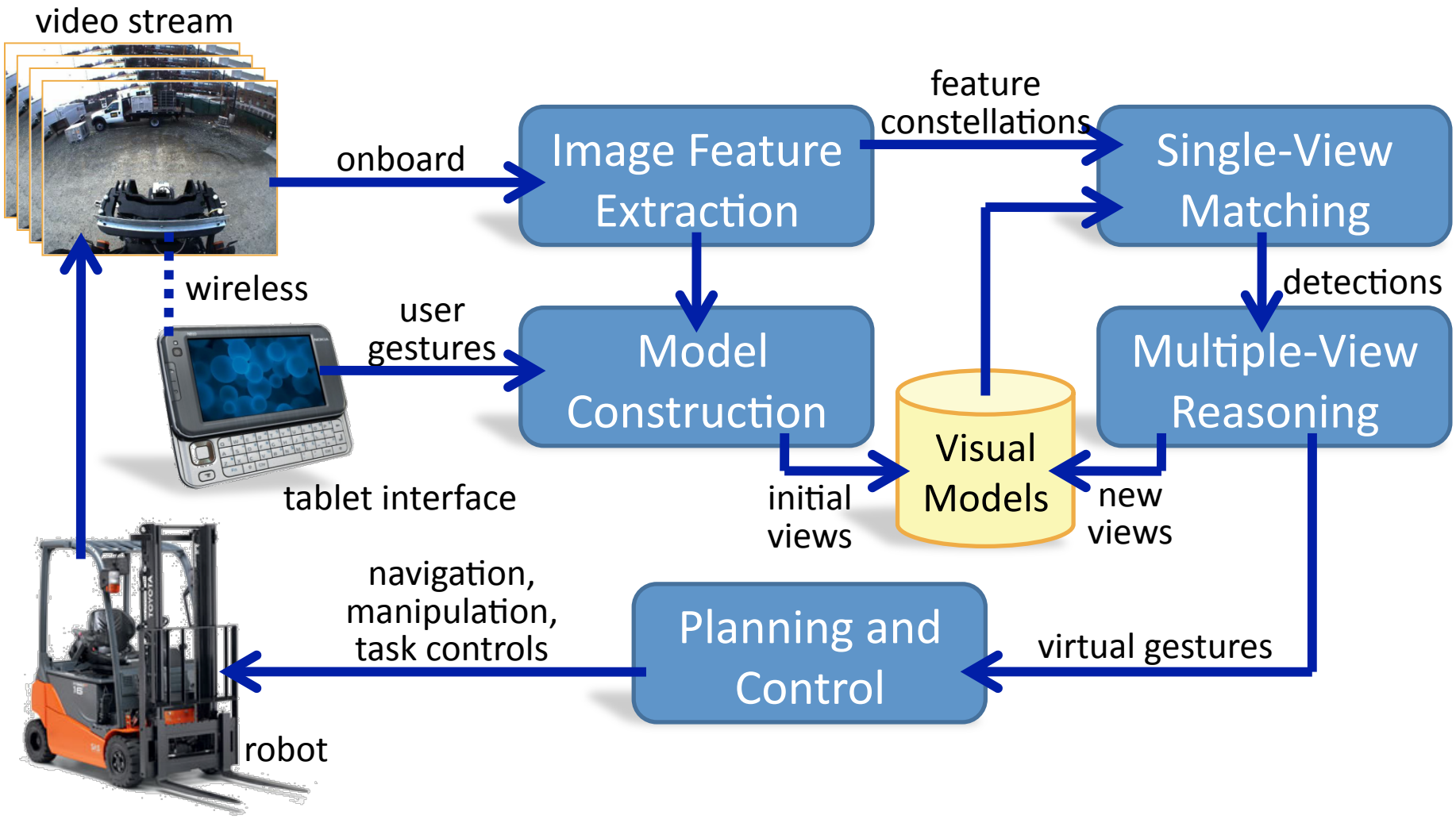
Related Work

- Visual object recognition (category classification)
  - Invariant descriptors and constituent parts
  - Recognize object *categories* rather than *instances*
  - Offline training w/ many samples
  - e.g., Nister06, Hoiem07, Savarese07
- Multiple-view model matching
  - Learn model offline from controlled viewpoints
  - e.g., Lowe01, Gordon06, Collet09
- Vision for gesture-based HRI
  - Person detection and face recognition
  - Body and hand gesture tracking
  - e.g., Nickel07, Haasch05

Task-Level Command Interface

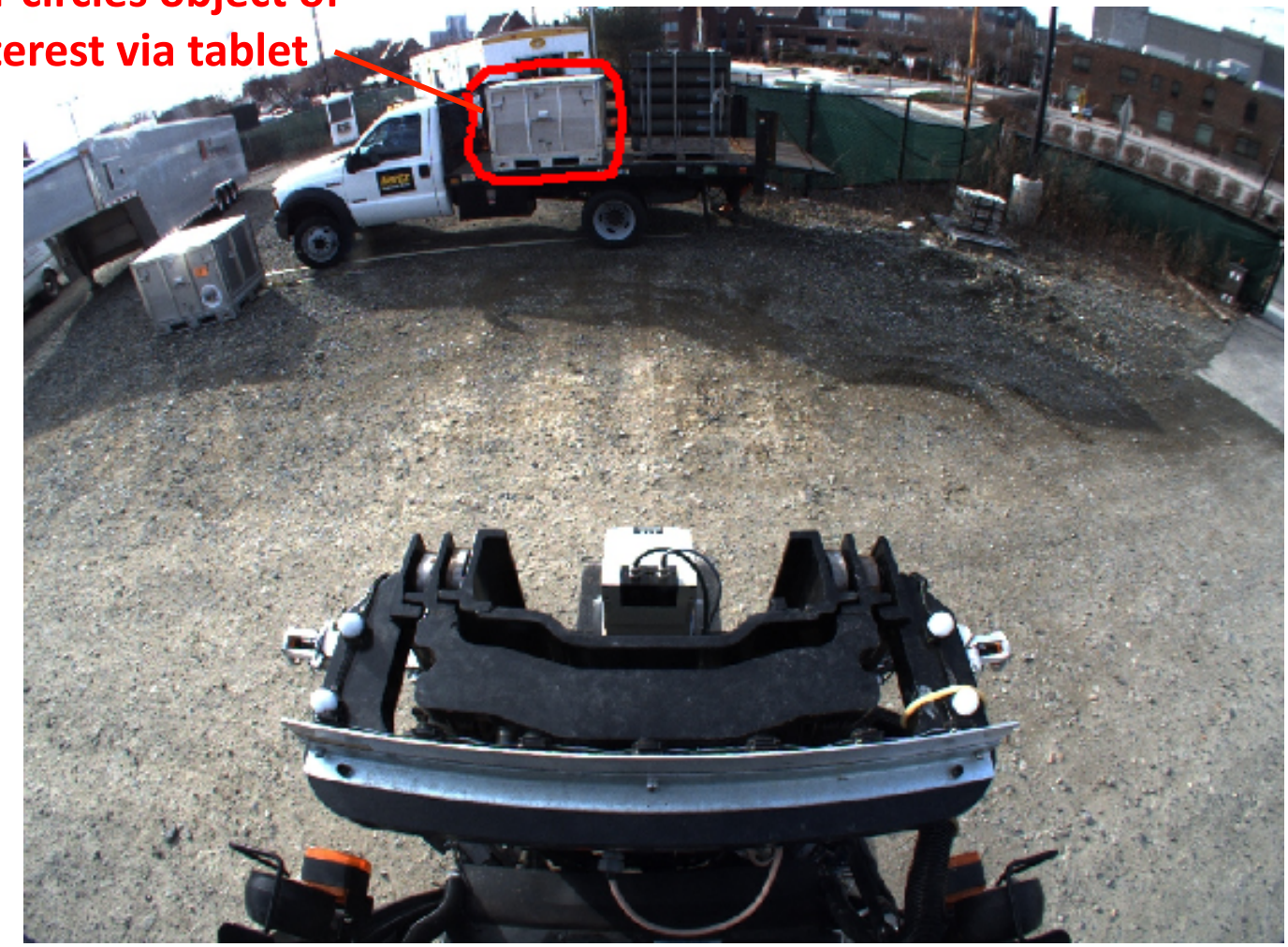
- Human supervisor provides commands
  - Pick up, transport, place palletized cargo
  - Speech and stylus gestures from handheld tablet
  - Previously, supervisor specified one task at a time

Reacquisition System Overview



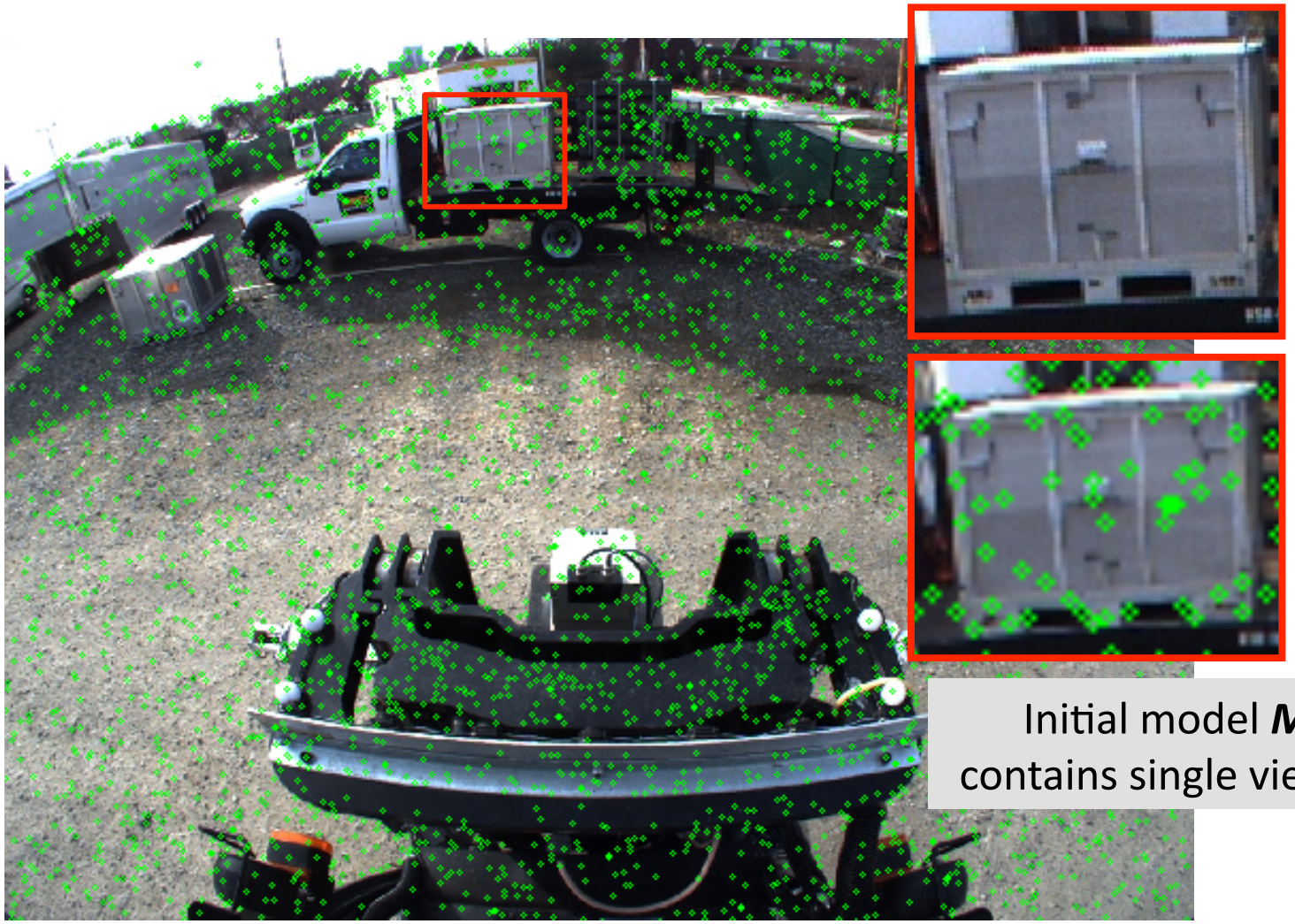
User Initiates Visual Appearance Model

User circles object of interest via tablet



Bot's-eye view of scene

Initial View Consists of Feature Constellation



Initial model ***M*** contains single view ***V***

All SIFT features ***F*** extracted from initial view image

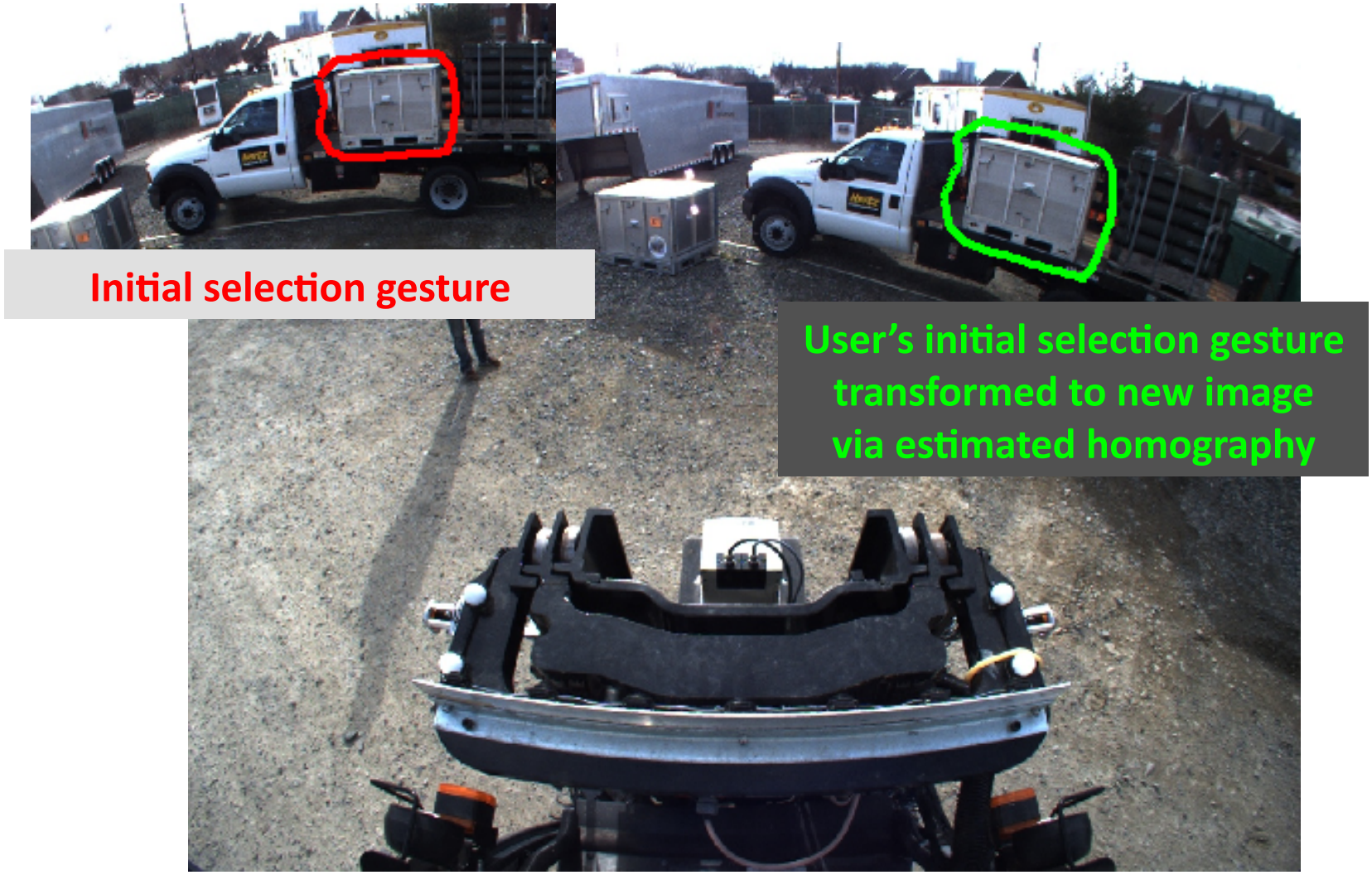
Single-View Matching: Feature Extraction

SIFT features extracted from new image

Single-View Matching: Homography+RANSAC

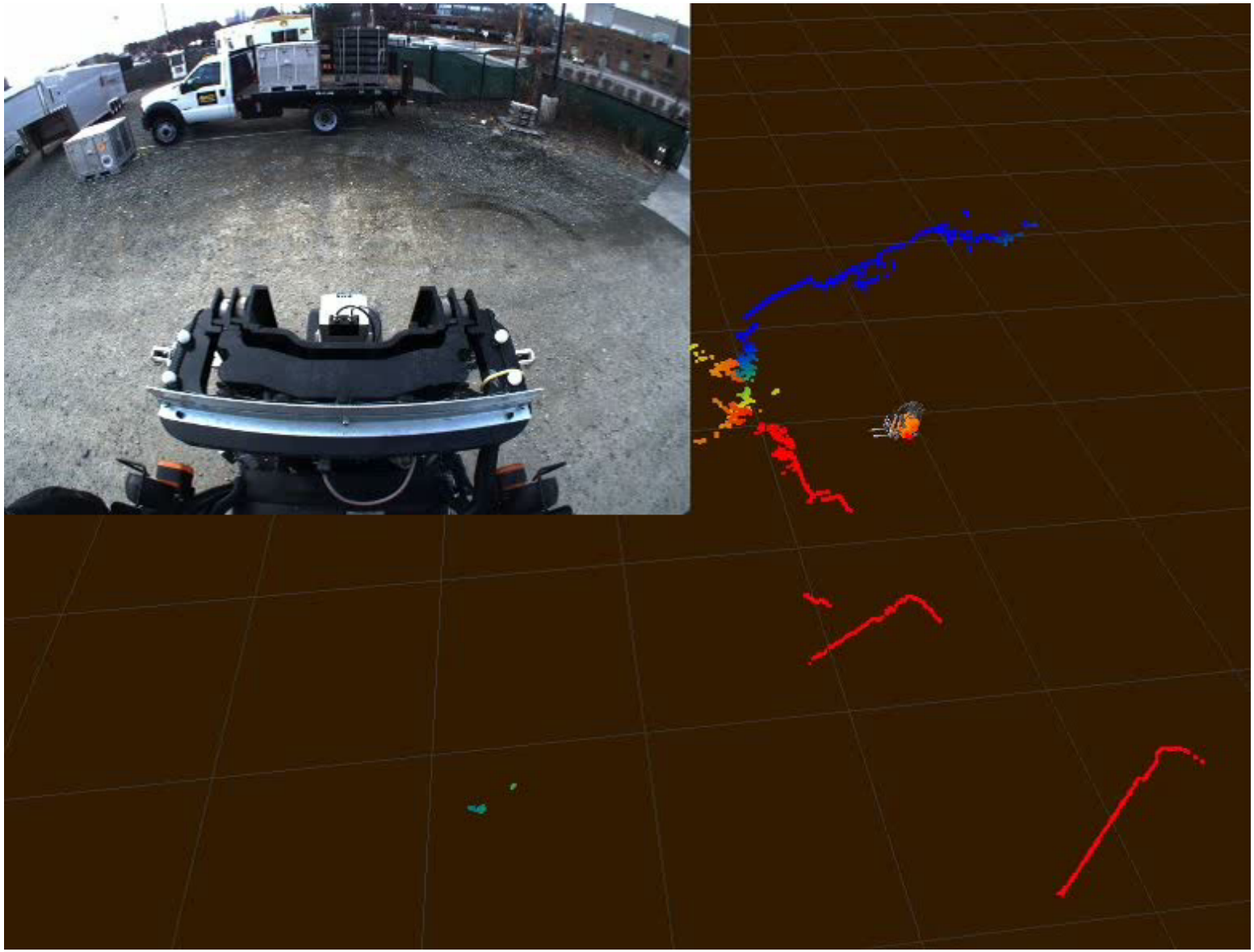
Robustly match to single view after application of geometric constraints

Single-View Matching: Final Object Detection



Detection of object location in new image

Single-View Matching: Example Video



Multiple-View Reasoning

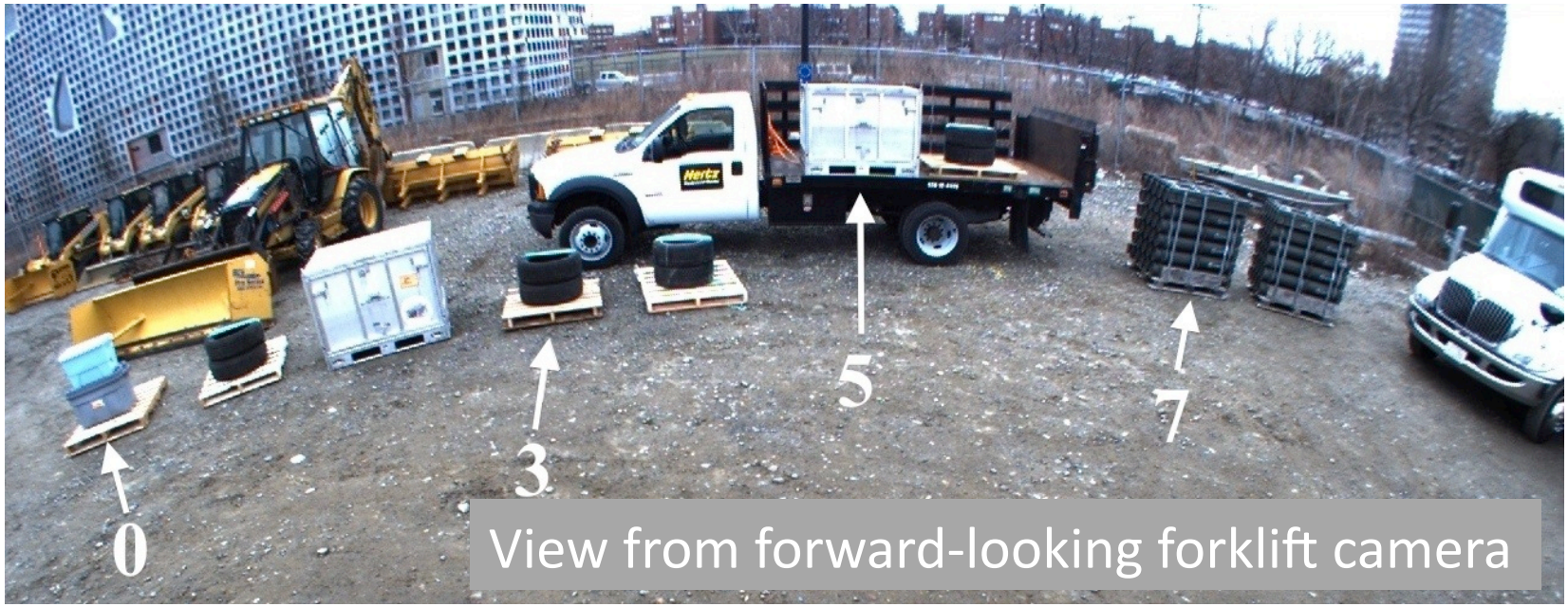
- Each object appearance model has many views
  - Multiple views capture different aspects, distances
  - Increase robustness to pose and lighting variation
- Matching produces multiple hypotheses
  - Several matches per image and per model
  - Want to produce at most one coherent hypothesis
- Reasoning component resolves ambiguity
  - Examines hypotheses and likelihoods
  - Voting produces at most one detection per model

Model Augmentation

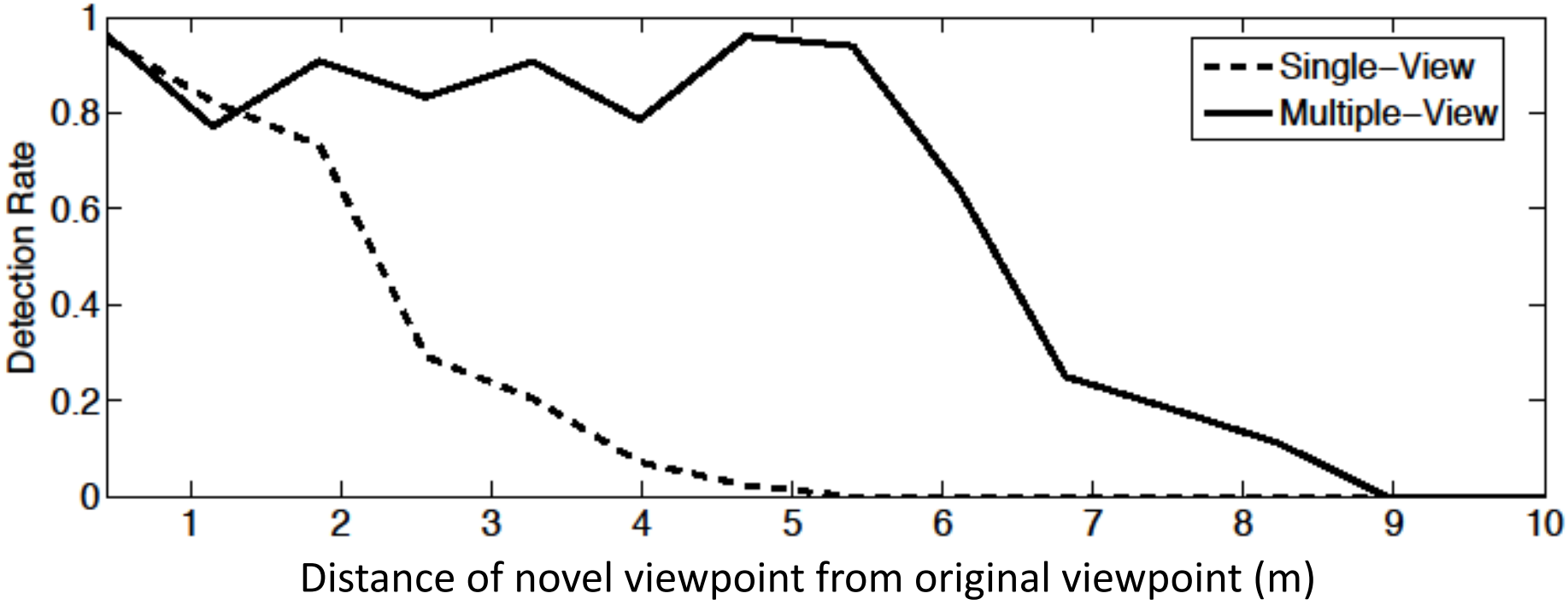
- New views opportunistically added to model
  - Addition requires high-confidence single-view match
  - Views added only when object aspect/distance is distinct
- More robust to viewpoint, illumination variation

Experimental Setup

- Preliminary experiments devised to test concept
  - Variety of loaded pallets placed within camera view
  - Loads have both similar and differing appearances
  - User selected four pallets for pickup and transport
- Counted false detections and missed detections
  - Ground truth generated via manual annotation

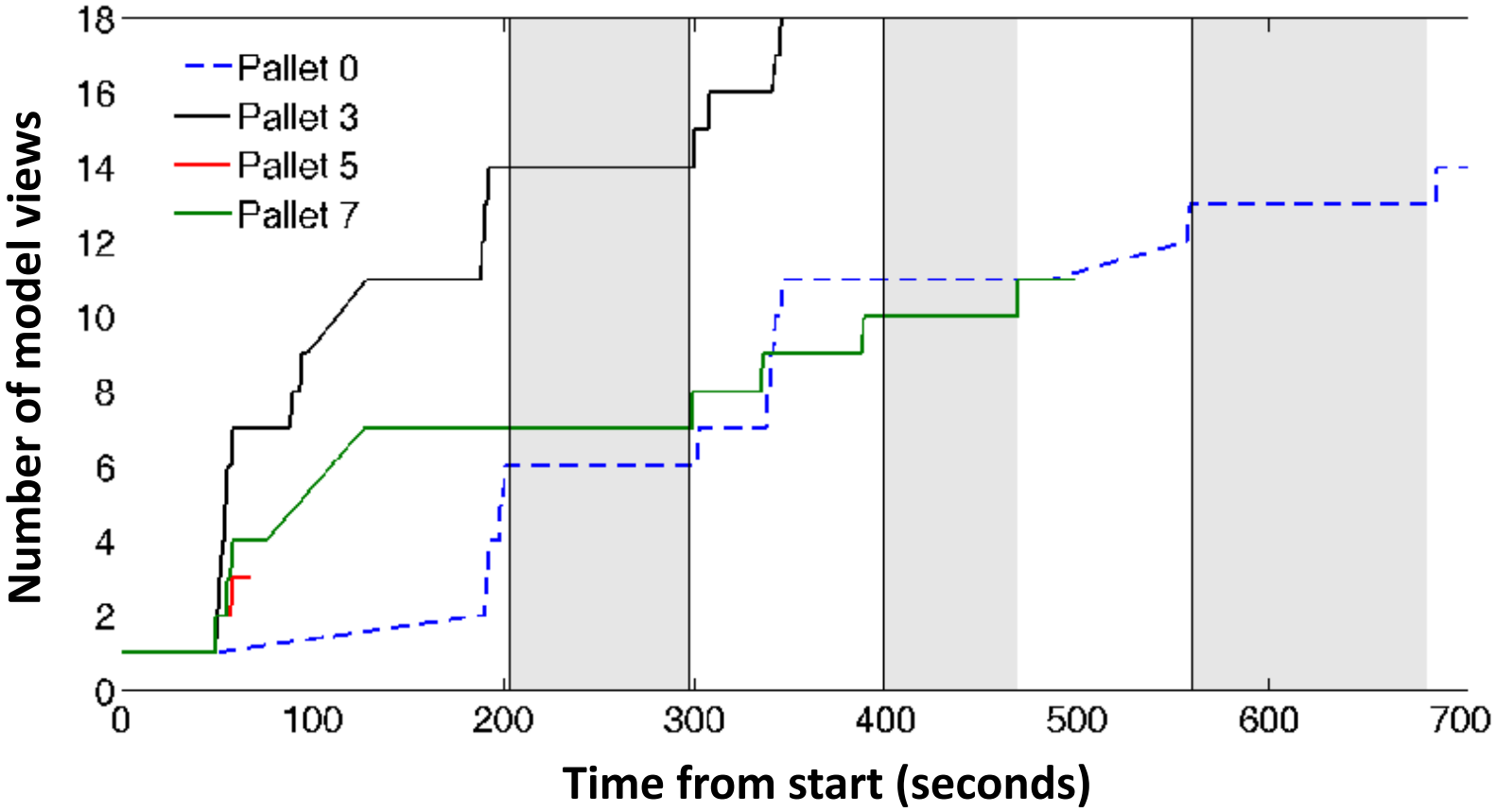


Single vs. Multiple View Detection



(False detection rate was *very low*, in XX trials -- **verify**)

Model Augmentation Over Time



Conclusions and Discussion

- Vision-based matching method
  - Motivated by real-world robotics task
  - Matches objects and their visual context
  - Multiple views used to increase persistence
- Limitations:
  - Scalability: exhaustive matching; models persist
  - Homography assumes dominant planar surface
  - Matching fails for very distant novel viewpoints

Future Work

- Possible extensions to improve robustness
  - Incorporate LIDAR
    - Object/context segmentation
    - Metrical distance of SIFT keypoints
  - Relax planar object assumption
    - Constellation graph matching
    - Bag-of-words recognition
  - Explicitly incorporate visual context
- Currently extending this approach
  - More accurate matching using quasi-3D descriptors
    - Multiple frames more closely coupled
    - Create local scene maps in vicinity of manipuland

Thank You. Questions?