Spring 2017 Carnegie Mellon University

Computer Vision

16-385

- **Lecturer:** Kris Kitani
- **TAs:** Prakruti Gogia, Animesh Ramesh, Abhinav Garlapati, Shaurya Shankar, Chen Kong
- **Class:** MW 1:30 to 2:50
- **Room:** DH 1212

today

- staff introduction
- what is computer vision?
- modern applications of computer vision
- administrative stuff (←important)

Prakruti Catherine Gogia

Masters in Computer Vision

pgogia@andrew.cmu.edu

Research interests:

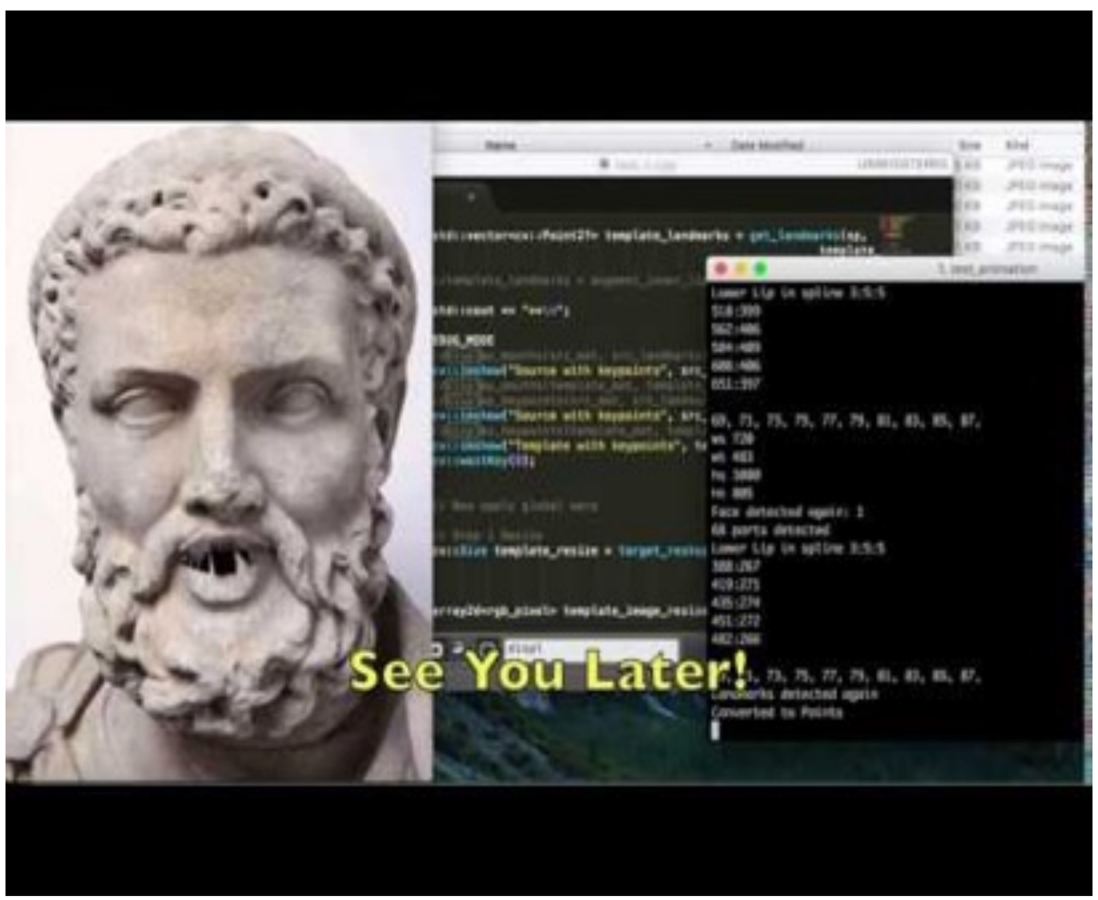
- Semantic segmentation
- Building creative tools using computer vision
- Medical Image Analysis

Office hours:

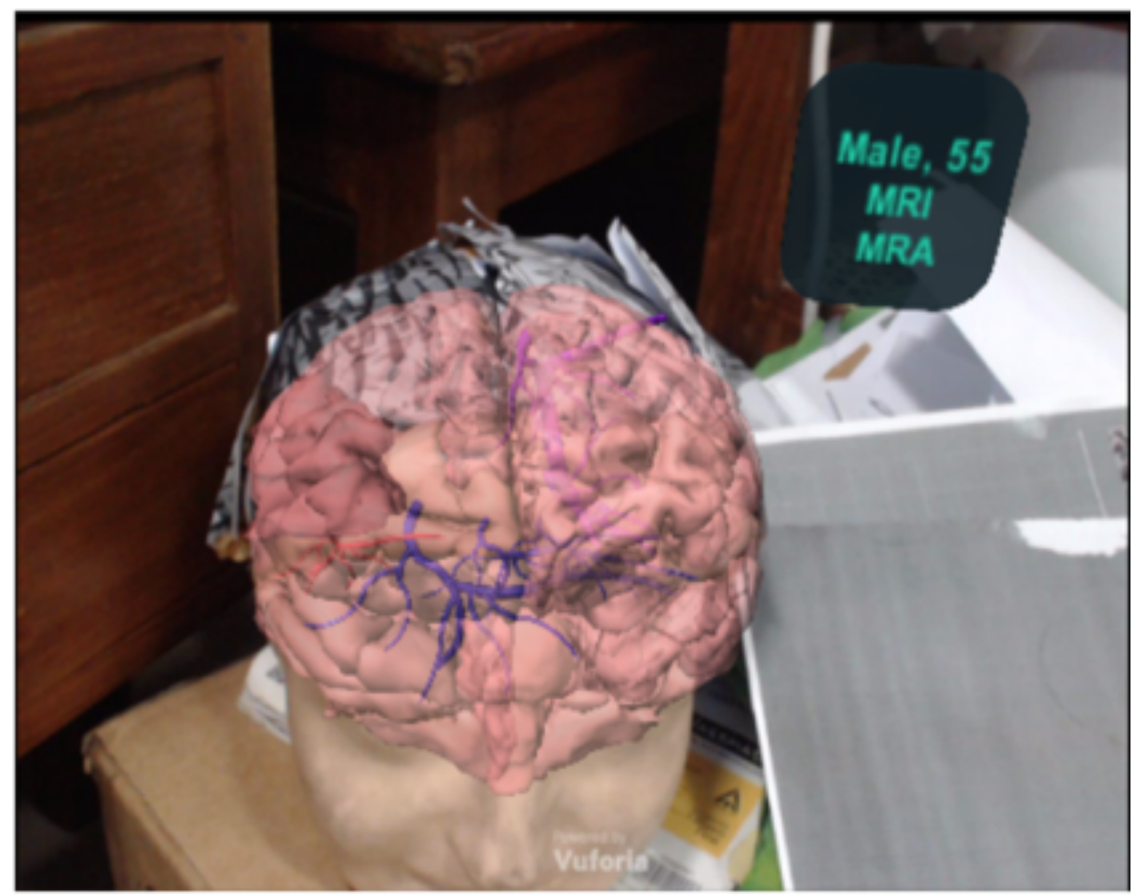
Mondays 6-7pm, EDSH 200

Projects

Snaps that chat! - Animating static images



AR for Surgical Planning



Animesh Ramesh

- 1st Year Master's in Computer Vision, CMU (2016 - 17)
- MSRIT (CS), Bangalore (2012 - 16)
- NUS Research Intern (2015)

Research Interests:

Deep Learning

Autonomous navigation

Machine Learning

Semantic segmentation

Object Recognition

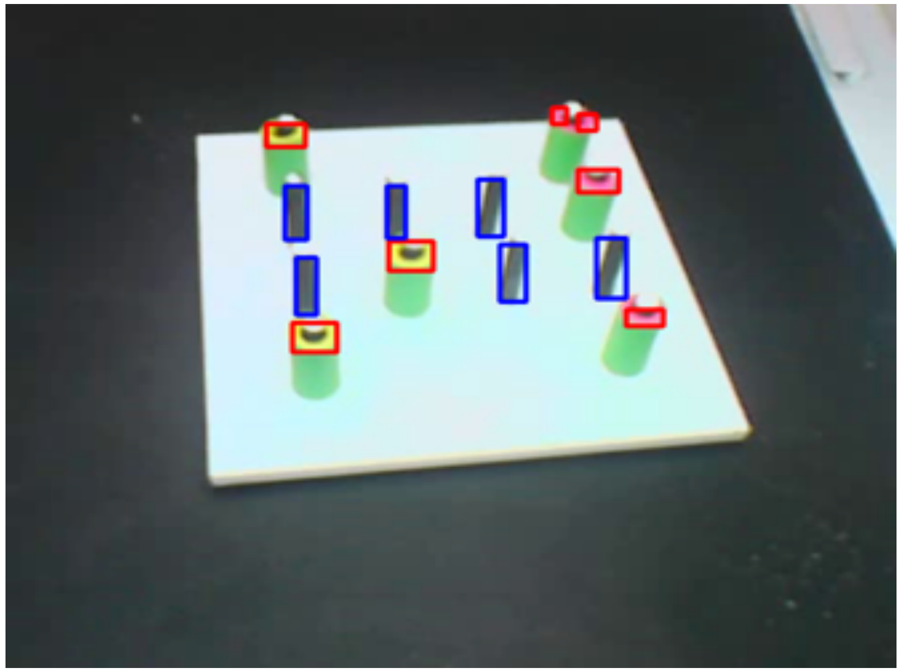
Face Recognition

Office Hours
Wednesdays
4.30-5.30pm
Smith Hall (EDSH) 200

Experience:

- Integrated autonomous navigation to a Robotic Water sensor in Singapore.

- Developed a computer vision system to train medical students for surgeries.



Abhinav Garlapati

Masters in Computer Vision

agarlapa@andrew.cmu.edu

Research Interests:

- Image and Video Understanding
  - Image classification
  - Activity Recognition

Office Hours: Tuesdays 5:00pm-6:00pm EDSH 200

Chen Kong

Third year PhD student
Advisor: Simon Lucey
chenk@cs.cmu.edu

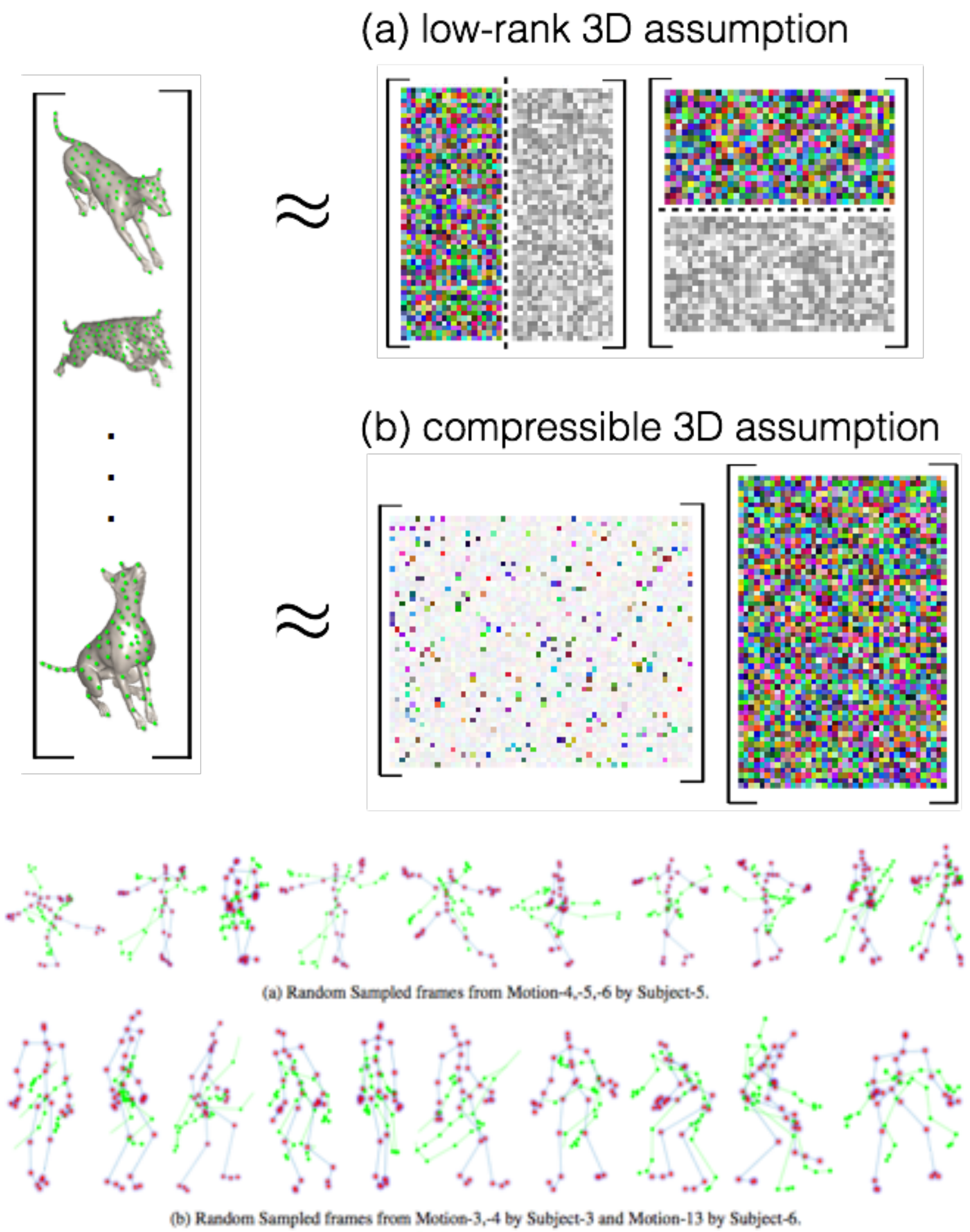
Research Interest:

Non-rigid structure from motion
(Group) sparse dictionary learning
Compressive sensing
Shape estimation from a single image

Office hours:
Friday 3-4pm, EDSH 210

Prior-less Compressible Structure from Motion

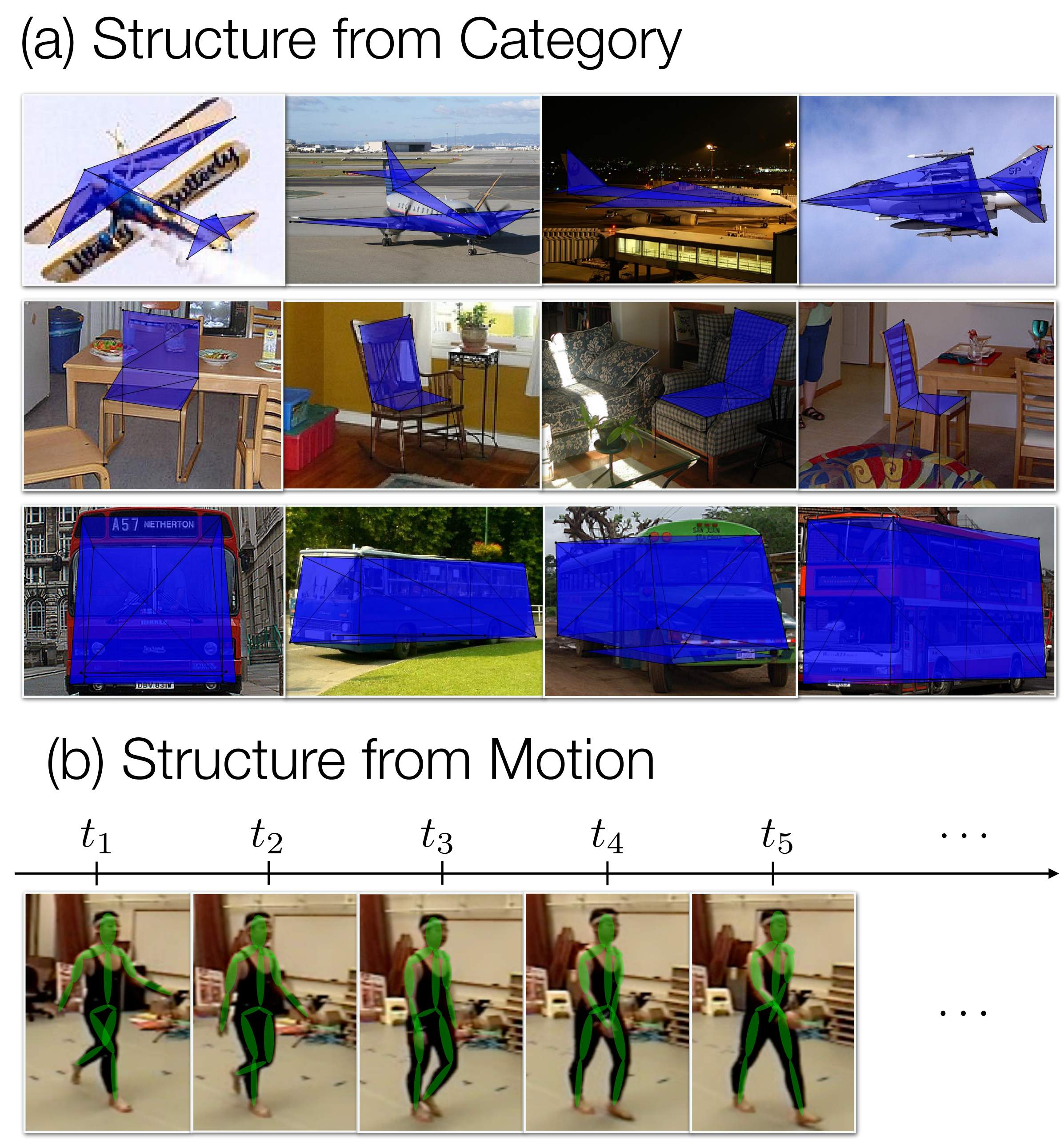
- We demonstrated that a compressible 3D structure under weak perspective projection is 2×3 block-compressible.
- If a 2×3 unique block sparse dictionary learning factorization can be obtained (of the 2D projections), we showed that the compressible 3D structure and camera motion can be recovered solely by the assumption of compressibility.
- The dictionary mutual coherence implies the reconstructibility of the projected 3D structures.



C. Kong and S. Lucey. Prior-less compressible structure from motion. *Computer Vision and Pattern Recognition (CVPR)*, 2016.

Structure from Object Category

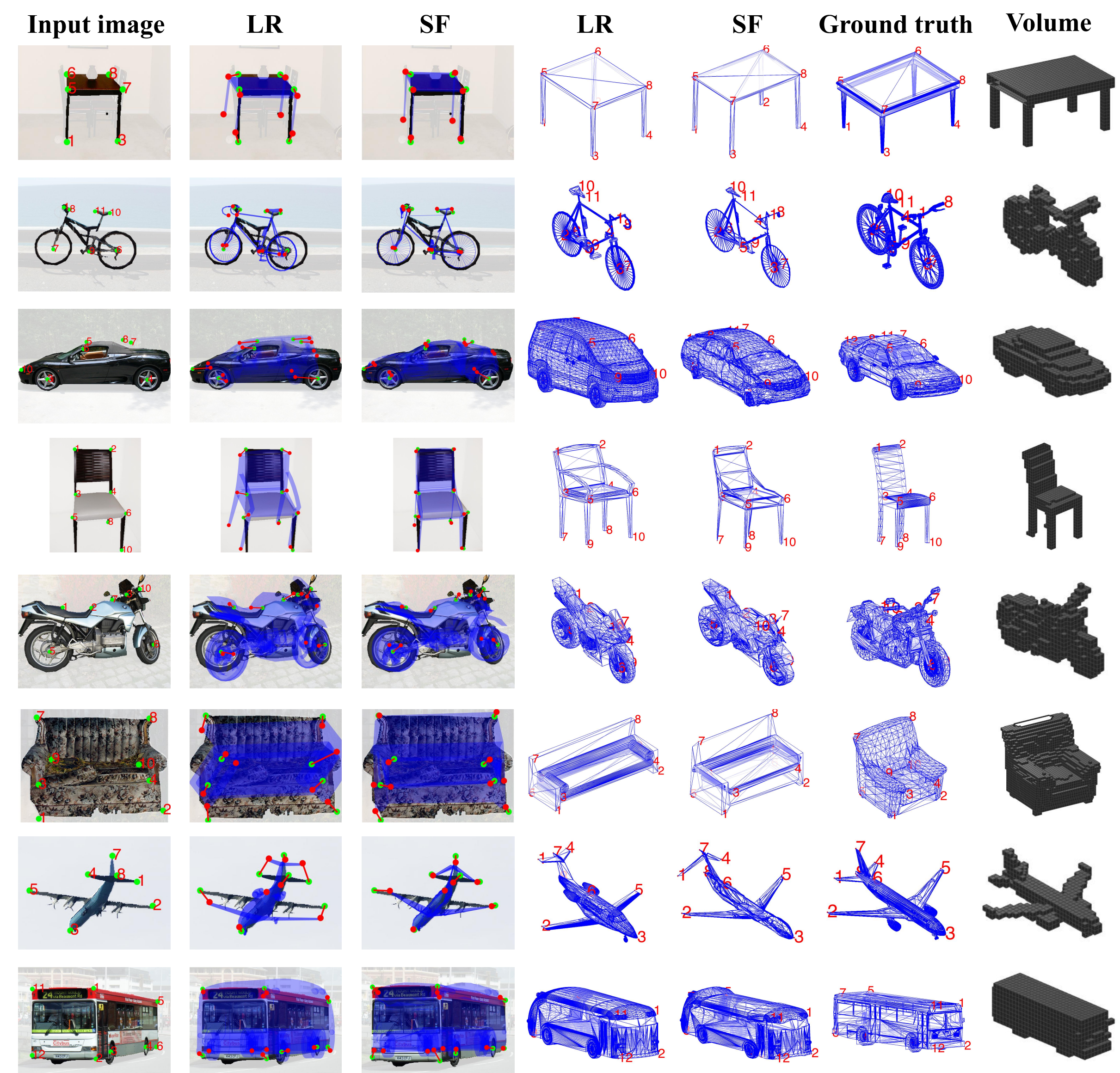
- We introduced the concept of Structure from Category to reconstruct 3D shapes of generic object categories from a sequence of images.
- Unlike most existing NRS*f* M methods, our approach requires no additional constraint on the shape or camera motion. Instead, all shapes and camera motion parameters (including shape bases) are jointly estimated through an augmented sparse shape-space model.
- Our framework can be applied for large scale 3D reconstruction.

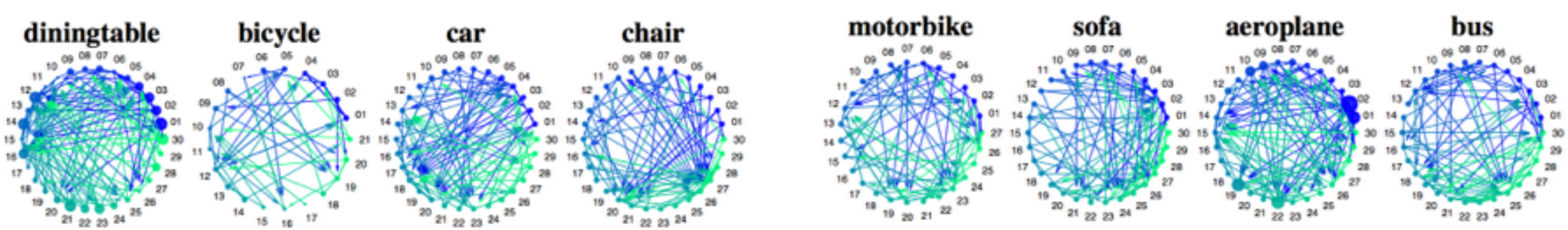


C. Kong, R. Zhu, H. Kiani, and S. Lucey. Structure from category: a generic and prior-less approach. *International Conference on 3D Vision (3DV)*, 2016.

Dense 3D Reconstruction from a Single Image

- We proposed a novel graph embedding demonstrating that a deformable, dense 3D model can be inferred only from local dense correspondence, eschewing the need for global correspondence.
- We proposed a two-step coarse-to-fine strategy using 2D landmarks and silhouette to reconstruct a deformable dense model from a single image.
- Impressive results were shown on both synthetic and real-world natural images





Kumar Shaurya Shankar

3rd Year PhD Student

kumarsha@cs.cmu.edu

Office Hours: Thurs 12-1 PM NSH 2201

Flying Through The Forests of Endor

https://www.youtube.com/watch?v=hNsP6-K3Hn4A

Odometry In The Real World

Conventional digital cameras have limited dynamic range

https://www.youtube.com/watch?v=rvp17MZdbis

Conventional 6DoF LK Tracking

What parameterized warp best minimizes a measure of dissimilarity between a reference image and a candidate image?

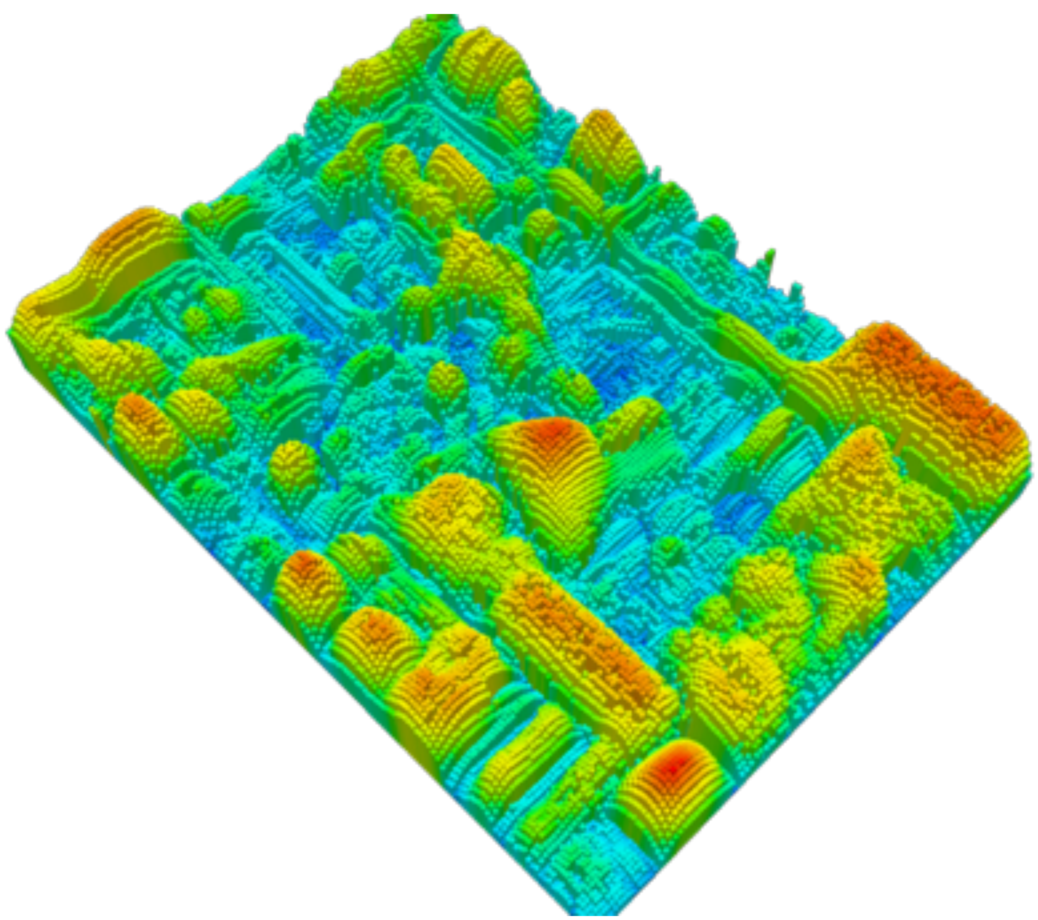
Brightness Constancy Assumption!

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{\mathbf{x} \in \mathcal{I}_l} \boxed{\left\| \mathcal{I}_r(\mathbf{w}(\mathbf{x}; \boldsymbol{\theta})) - \mathcal{I}_l(\mathbf{x}) \right\|^2}$$

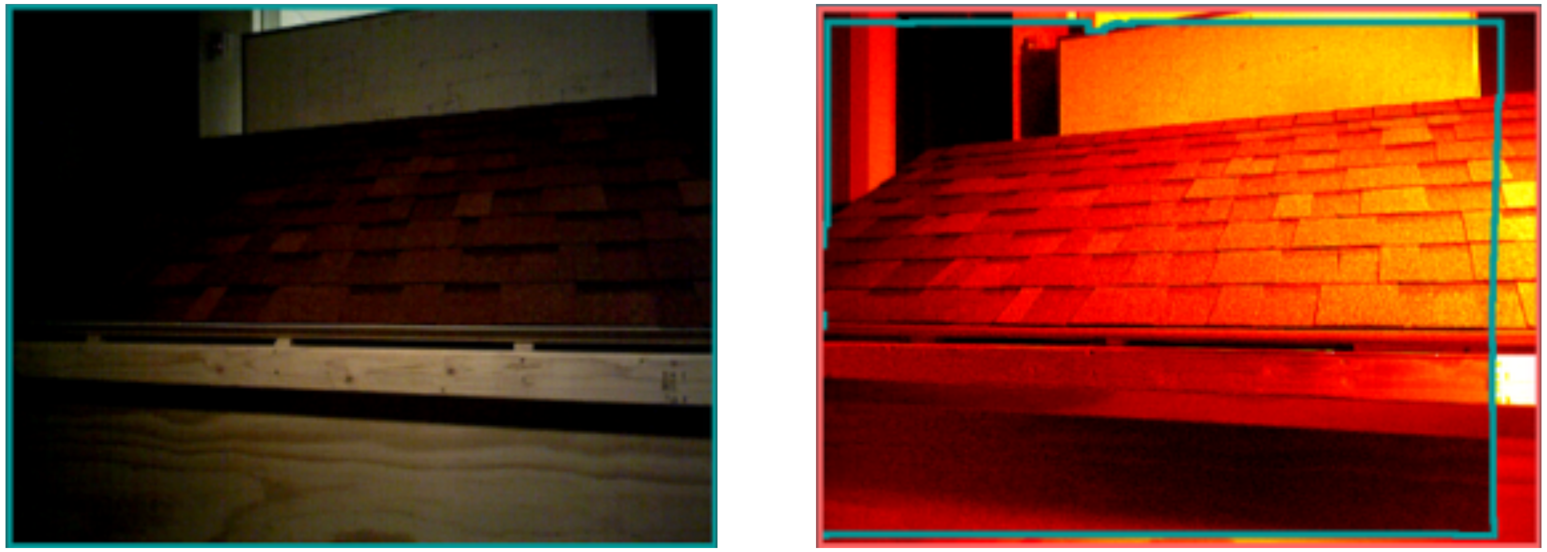
This is fundamentally violated in dynamic conditions!

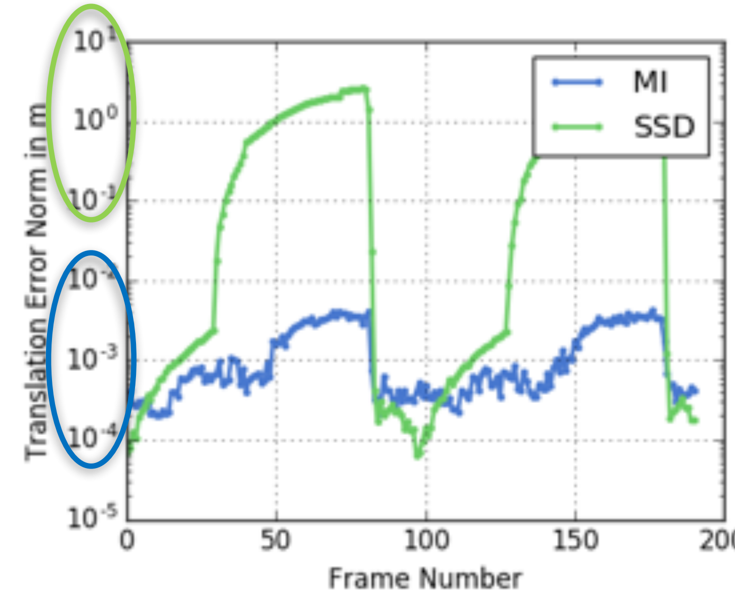
Mutual Information for Registration

- Images are a joint distribution of spatial locations and intensity
- Mutual Information is an established measure of divergence for distributions
- Focus on *relative* comparisons as opposed to absolute measures



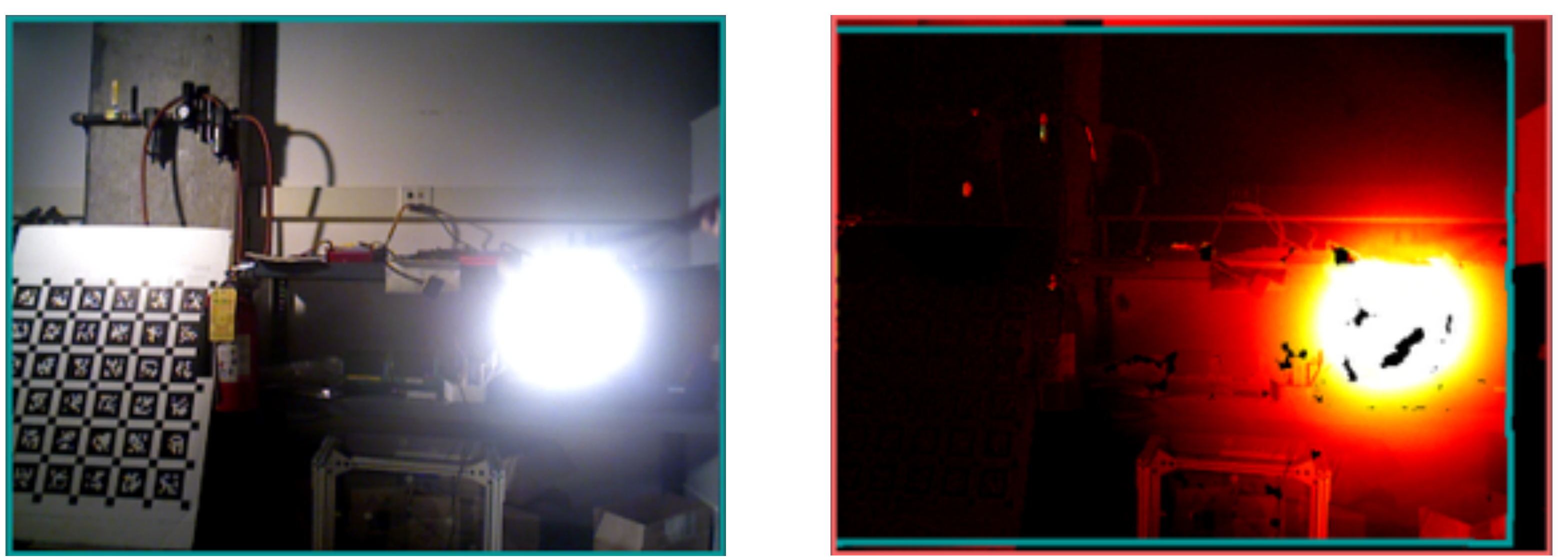
Comparison under Dynamic Lighting

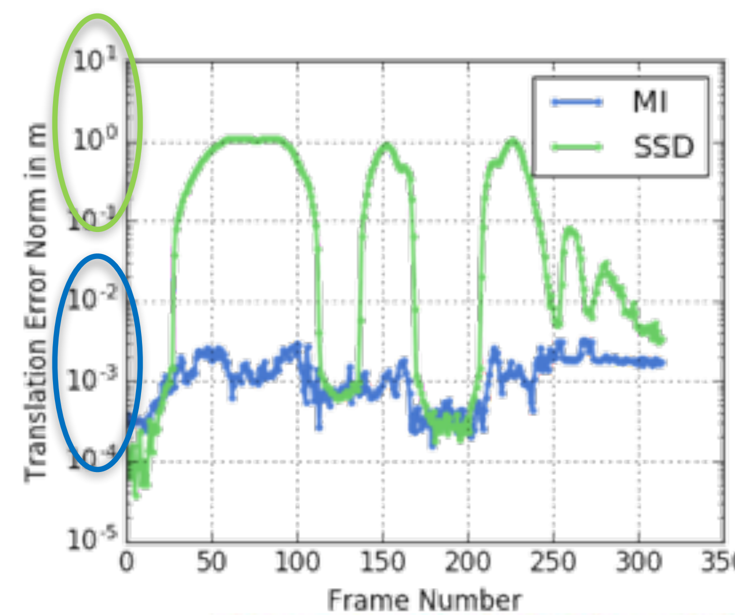




Varying Global Illumination

Three orders of magnitude smaller per frame mean error! (10^{-3} vs 10^{0} m)



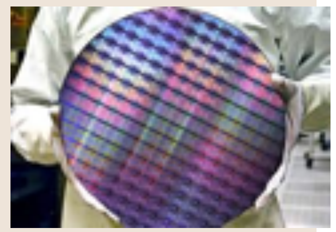


Varying Local Illumination

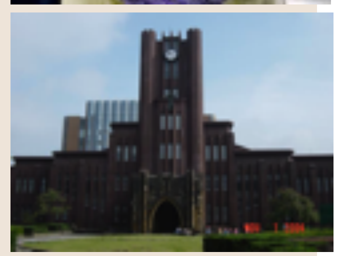
Related Publication:
K. S. Shankar and N. Michael, “Robust Direct Visual Odometry using Mutual Information”, International Symposium on Safety, Security and Rescue Robotics [Best Student Paper Award]

Kris Kitani

University of Southern California (1995-1999)



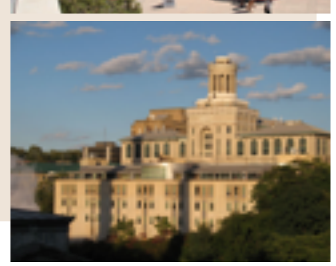
KLA-Tencor Japan (2000-2003)



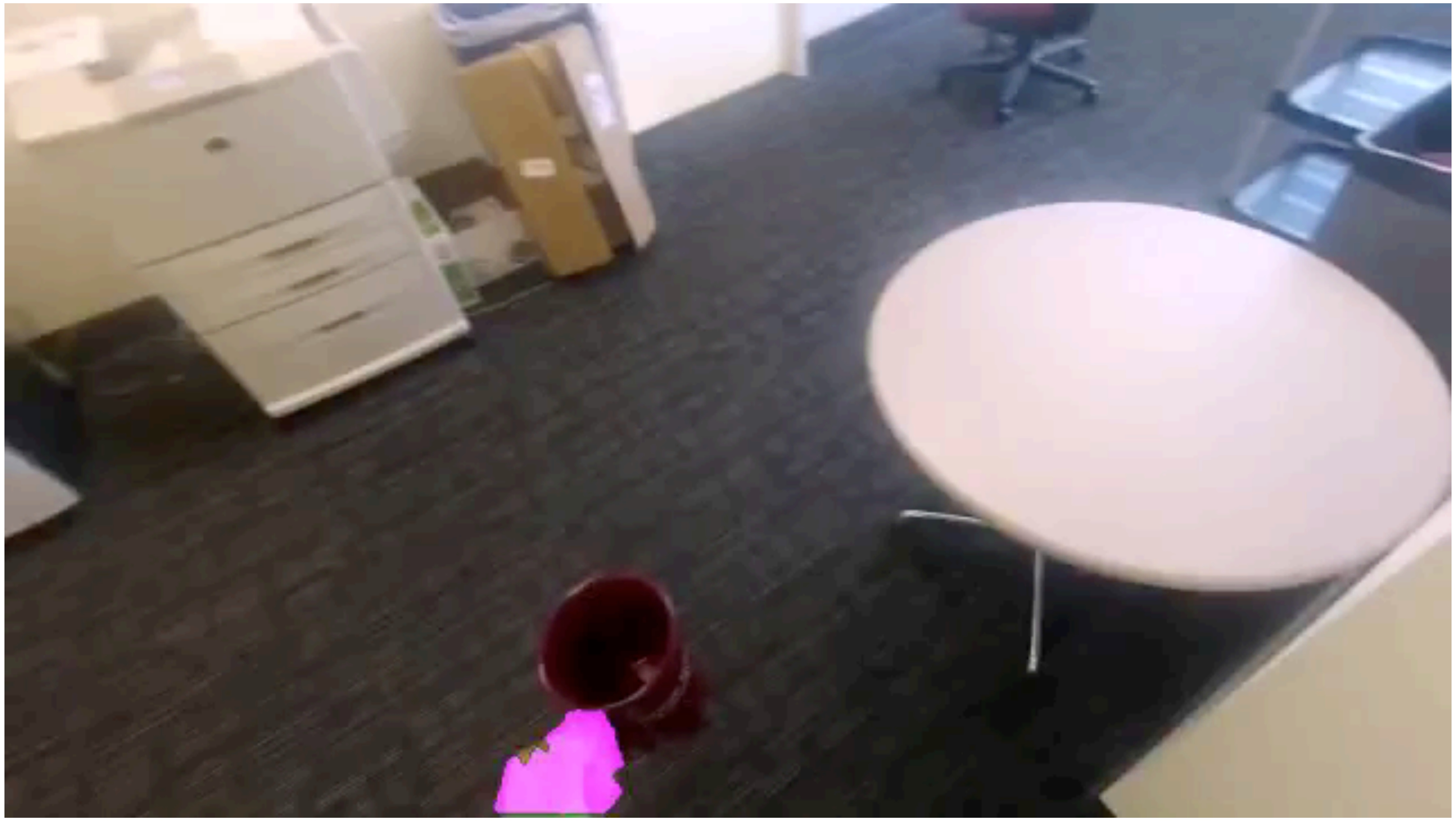
University of Tokyo (2003-2008)

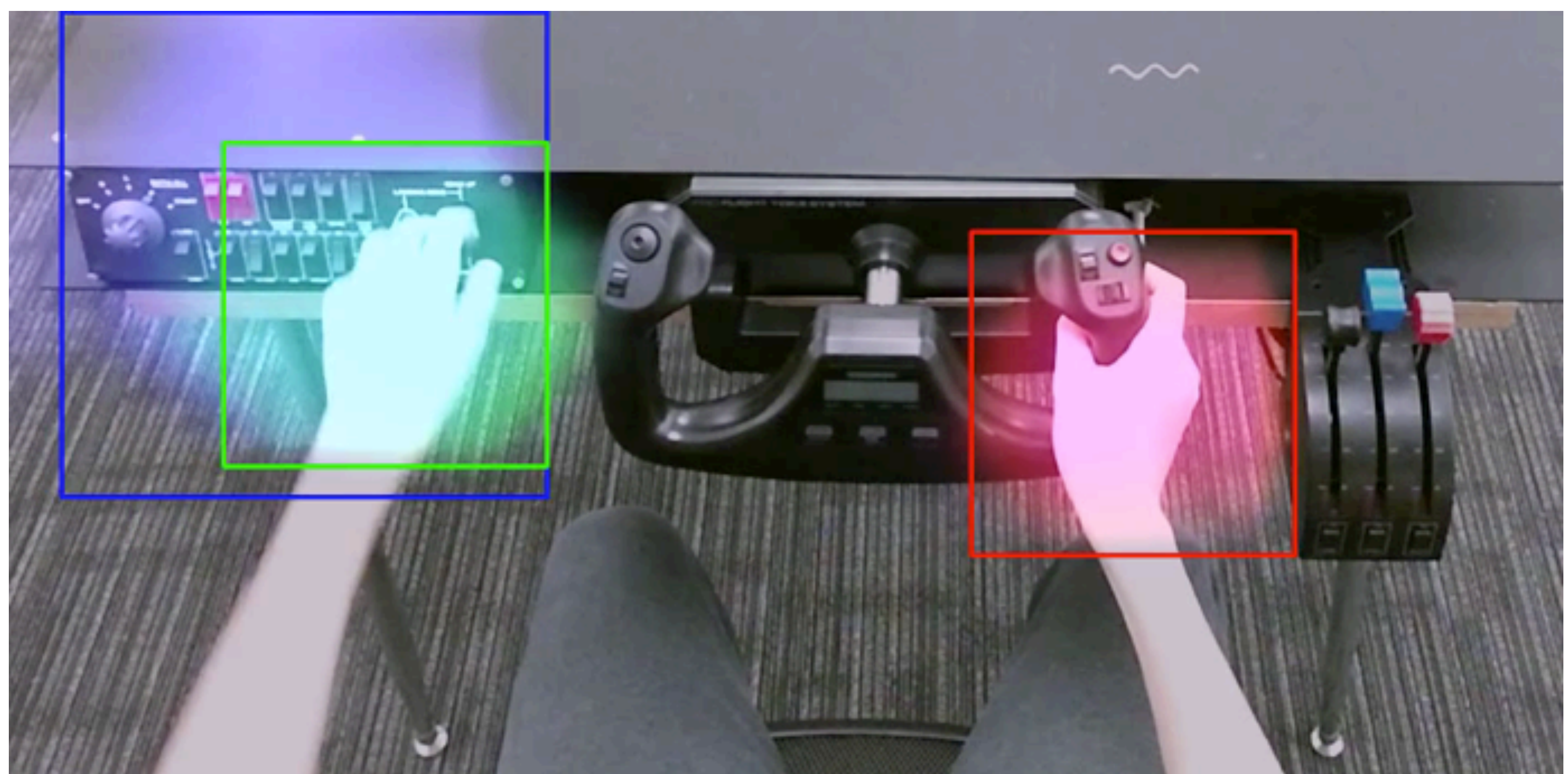
University of Electro-Communications (2008-2011)

University of California, San Diego (2010)

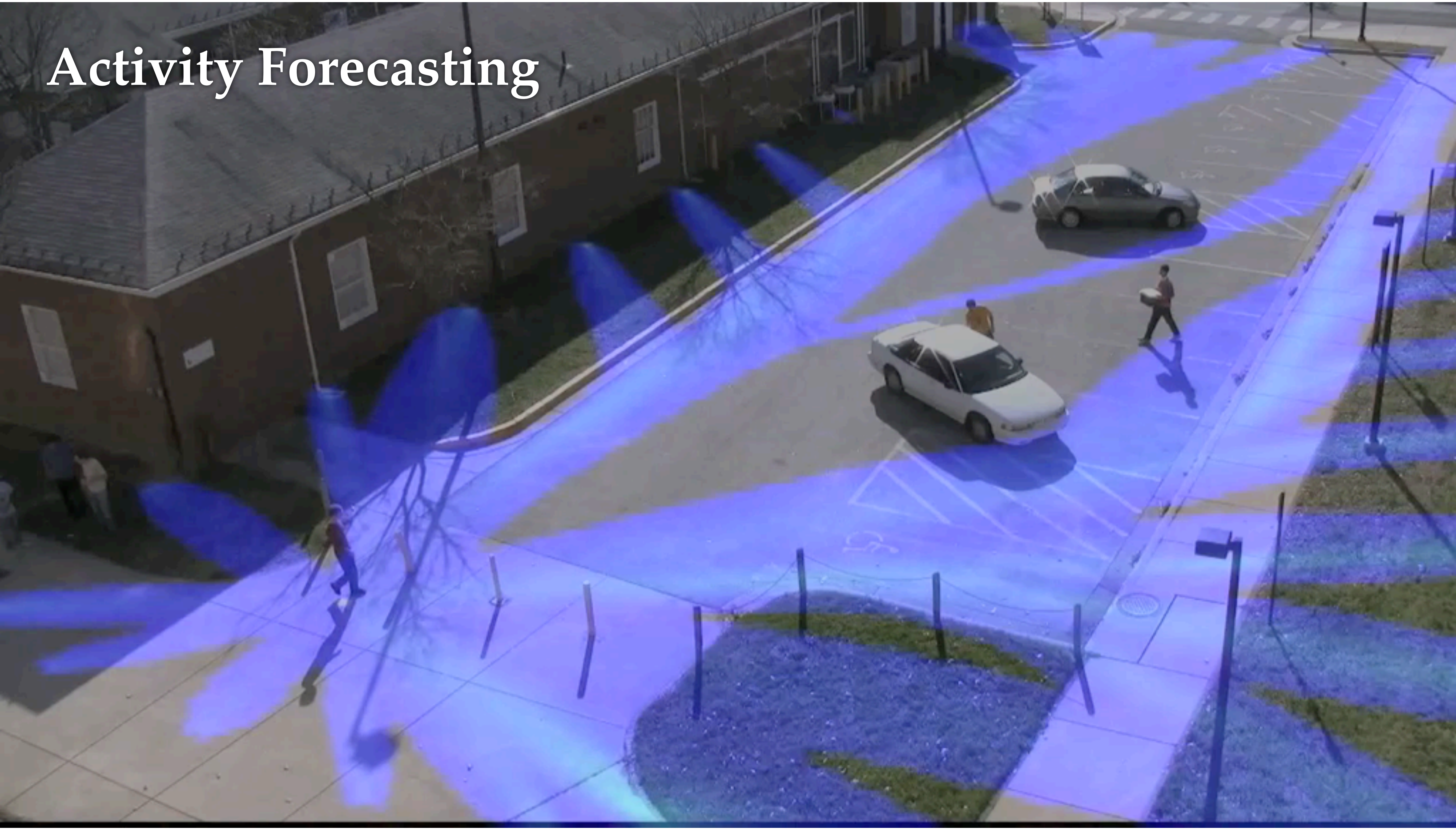


Carnegie Mellon University (2011-present)





Activity Forecasting



Given an occluded interaction video

extrapolate the missing image sequence

4. Experimental Evaluation

Ours　　DPM

Town Center Dataset

NavCog