GTC, Santa Jose, 2019

Video-Based Activity Forecasting for Construction Safety Monitoring Use Cases

Speaker: Shuai Tang
University of Illinois at Urbana-Champaign

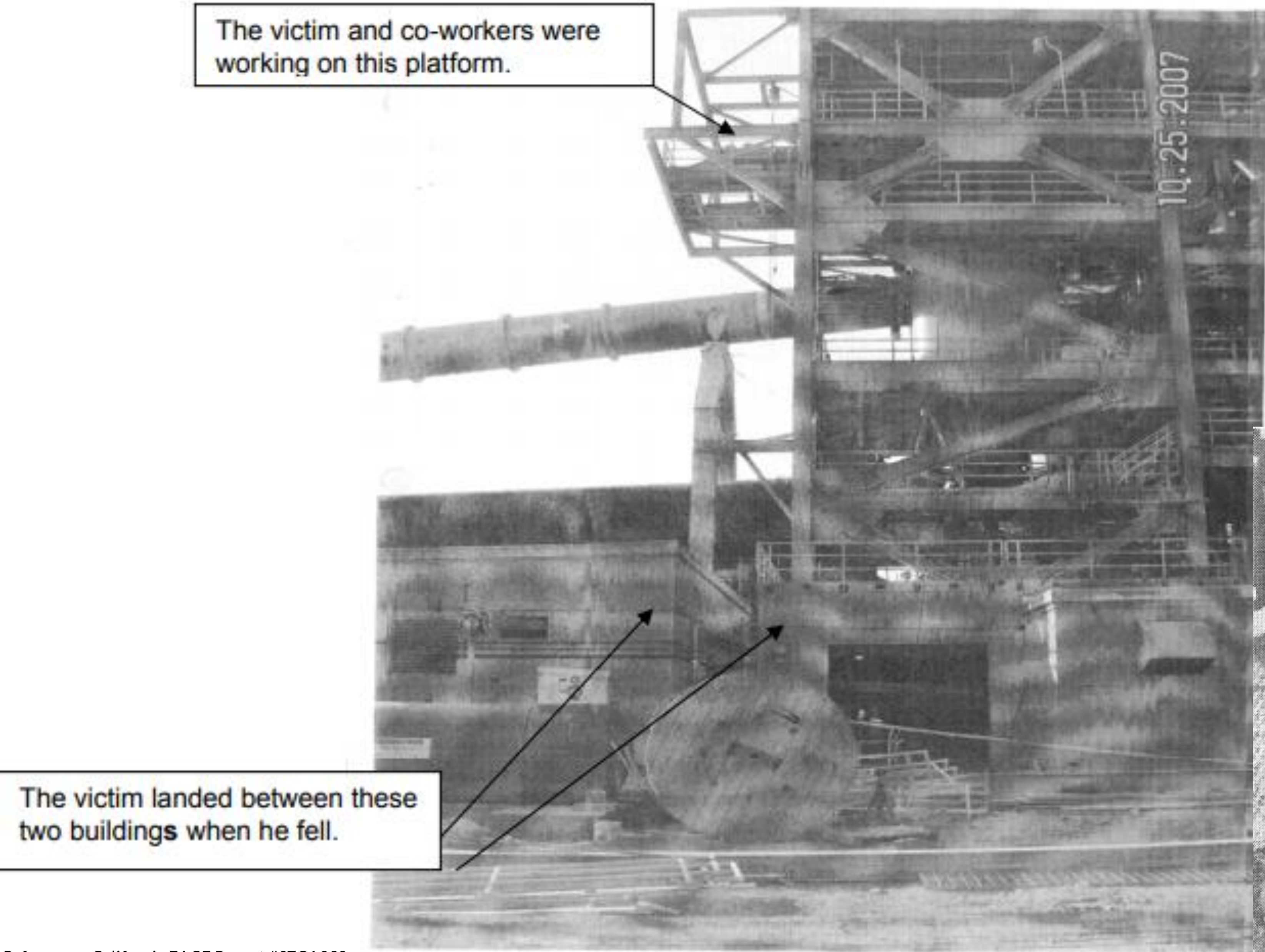
Contributors: Mani Golparvar Fard (University of Illinois)
Milind Naphade (Nvidia), Murali Gopalakrishna (Nvidia), Amit Goel (Nvidia)

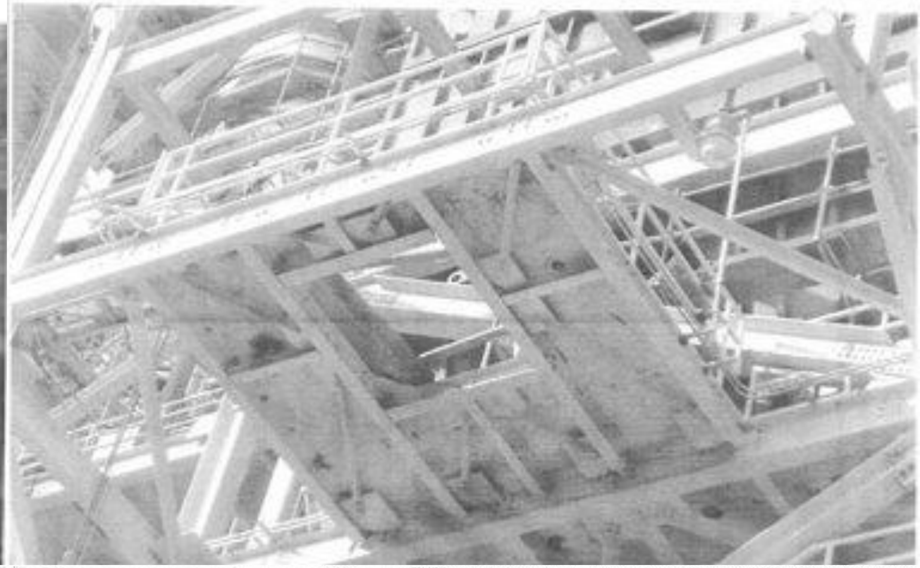
NVIDIA

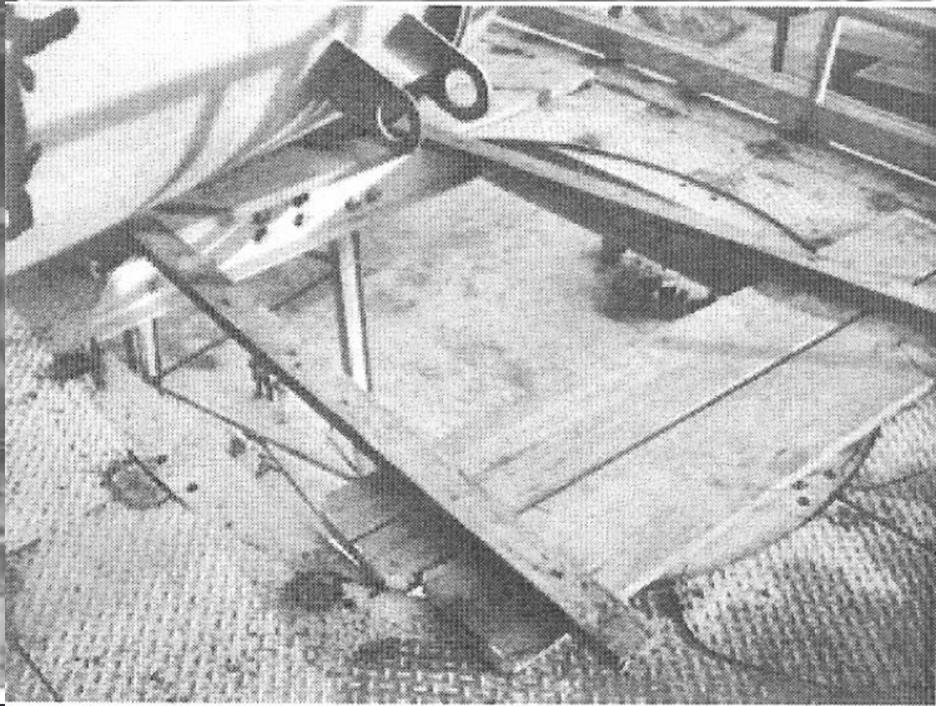
RAAMAC
Real-time and Automated Monitoring and Control

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Worker Dies From Falling 50 Feet







References: California FACE Report #07CA009

14 Worker Deaths Every Day In The US

20.7% of all worker deaths were in construction

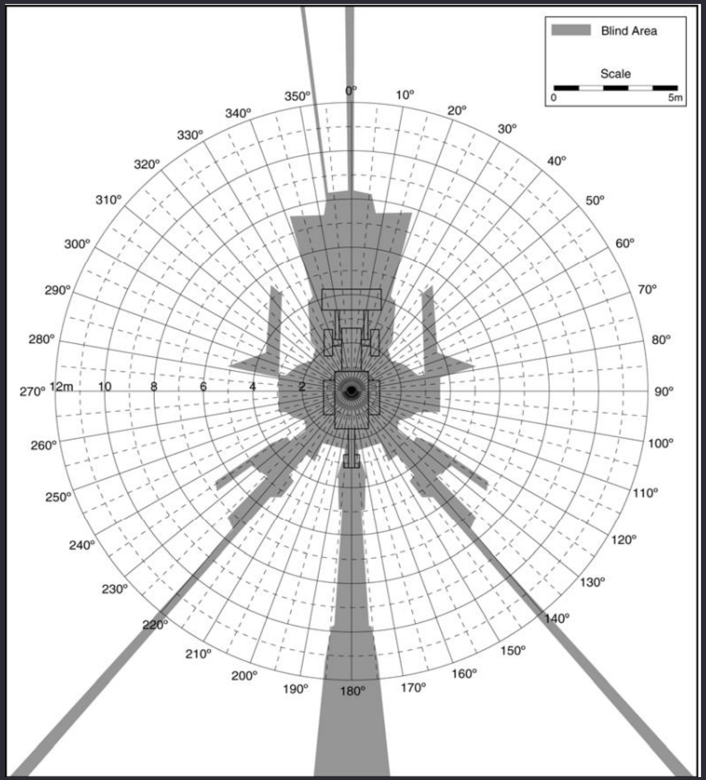
OSHA estimates that eliminating top 4 hazards in construction
save 581 workers' lives

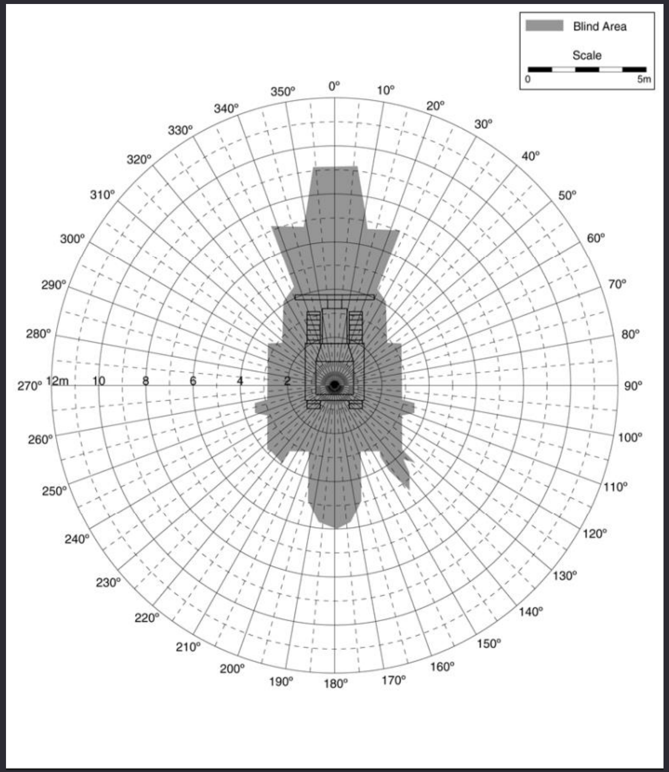
Falls: 381 deaths (39.2%)
Struck By Object: 80 deaths (8.2%)
Electrocution: 71 deaths (7.3%)
Caught-In/Between: 50 deaths (5.1%)

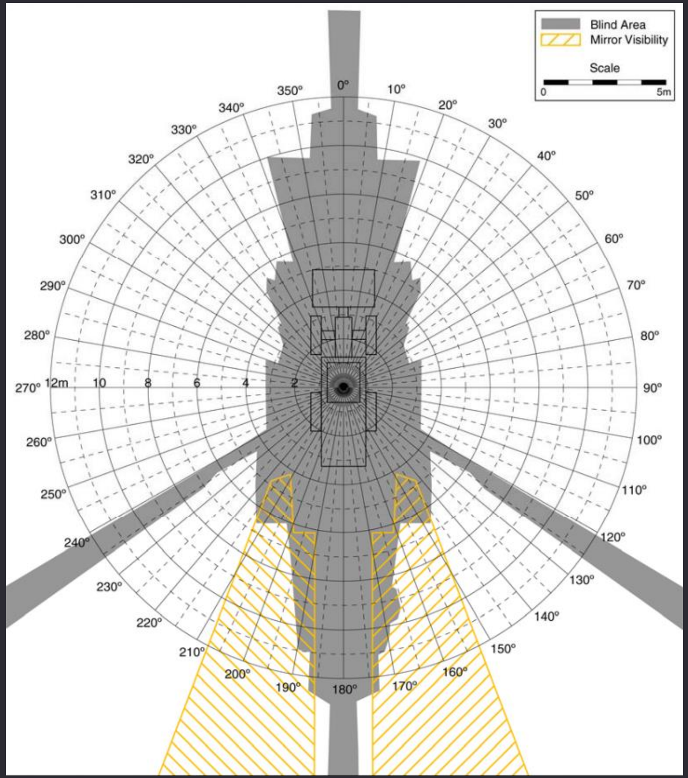
'Careless' foreman crushes woman, 19, with backhoe at Bexar County construction site

By Caleb Downs Updated 7:18 pm CST, Tuesday, January 24, 2017

'Careless' Operator Crushes Worker With Backhoe







References: https://www.cdc.gov/niosh/topics/highwayworkzones/bad/pdfs/catreport2.pdf

Non-fatal Injuries In Construction

RAAMAC
Real-time and Automated Monitoring and Control

- Safety incidents
  - 971 fatal cases
  - 79,810 non-fatal cases involving days away from work
- $1.3 trillion construction expenditure each year
- Financial impact of safety
  - Around $4 million cost per fatal case,
  - Over $42,000 average cost per non-fatal case.

Motivation

Frequency

Safety inspections are taken typically weekly.

Accuracy

50% hazards not recognized by workers

Proactiveness

Safety measurements are often retrospective

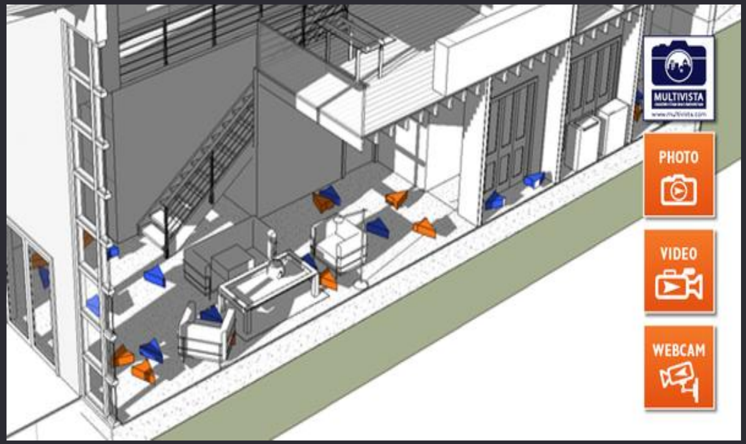
Image sources: Google Image

Overreaching Goal:

Visual-based activity forecasting towards predictive safety monitoring

Opportunity - Growth In Visual Data

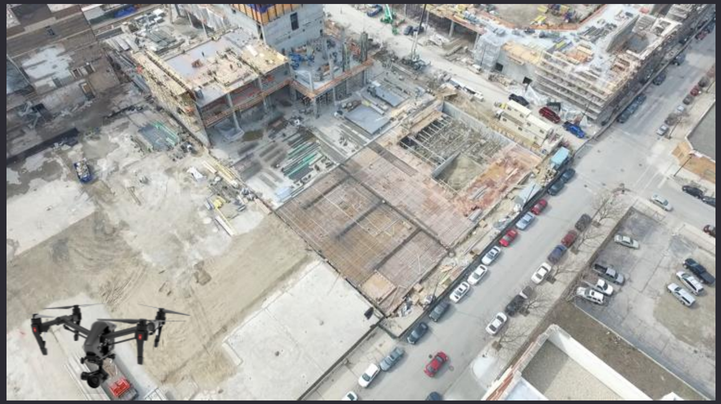
200-1,000 pictures per day



~1,000 pictures per day

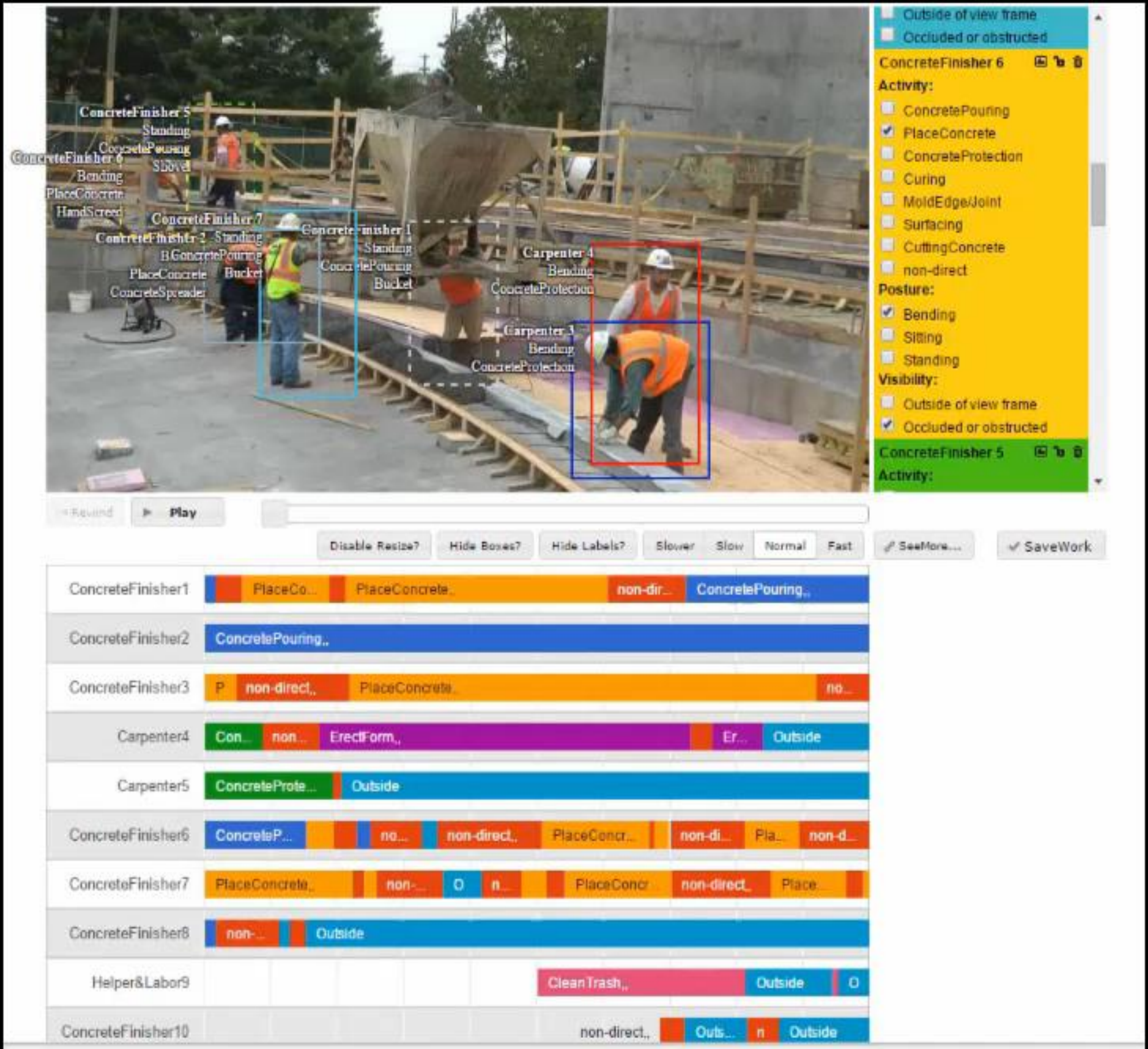
Time-Lapse pictures every 5-15min

1-10 videos per day

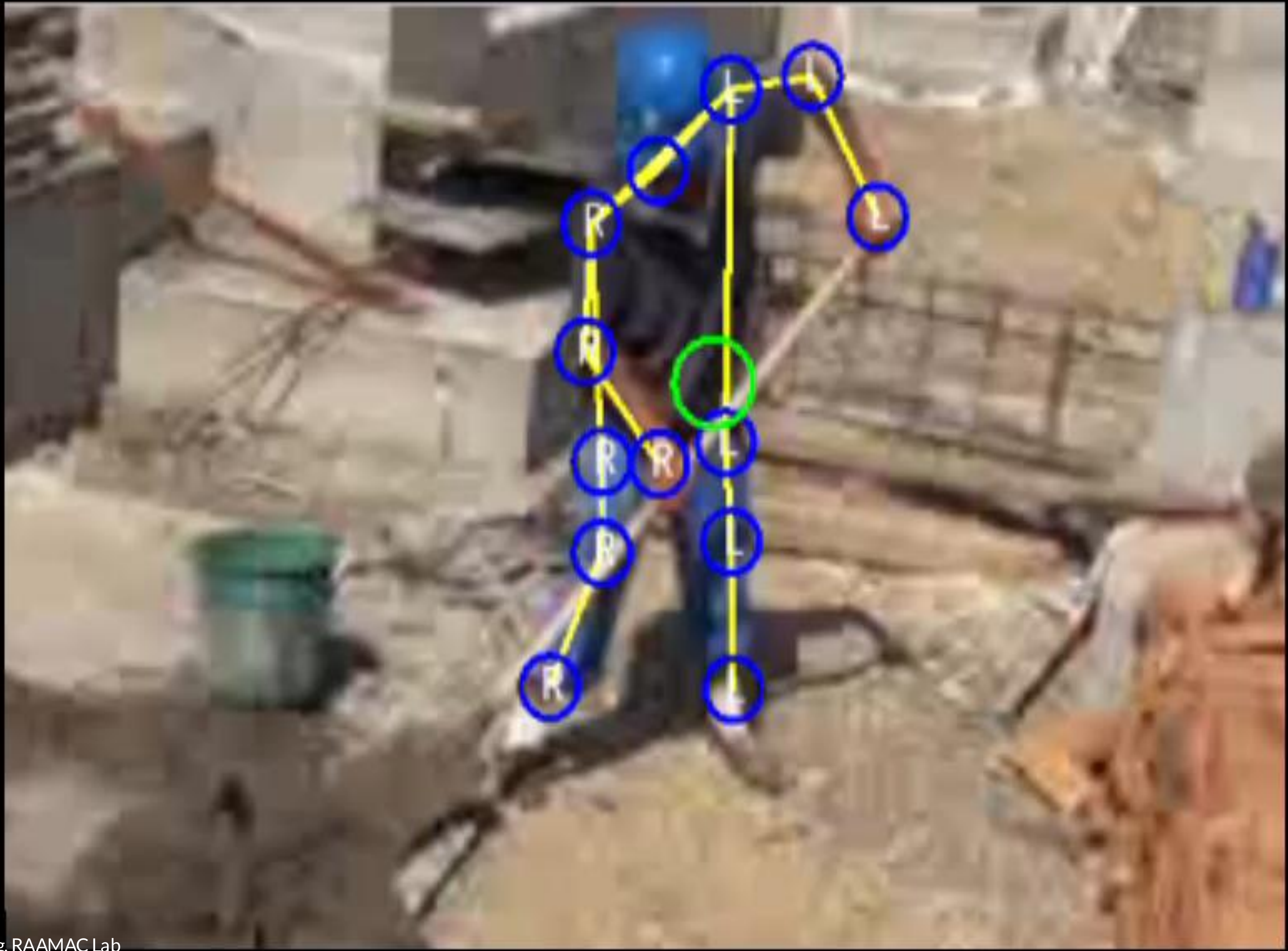


~2,000 images per week

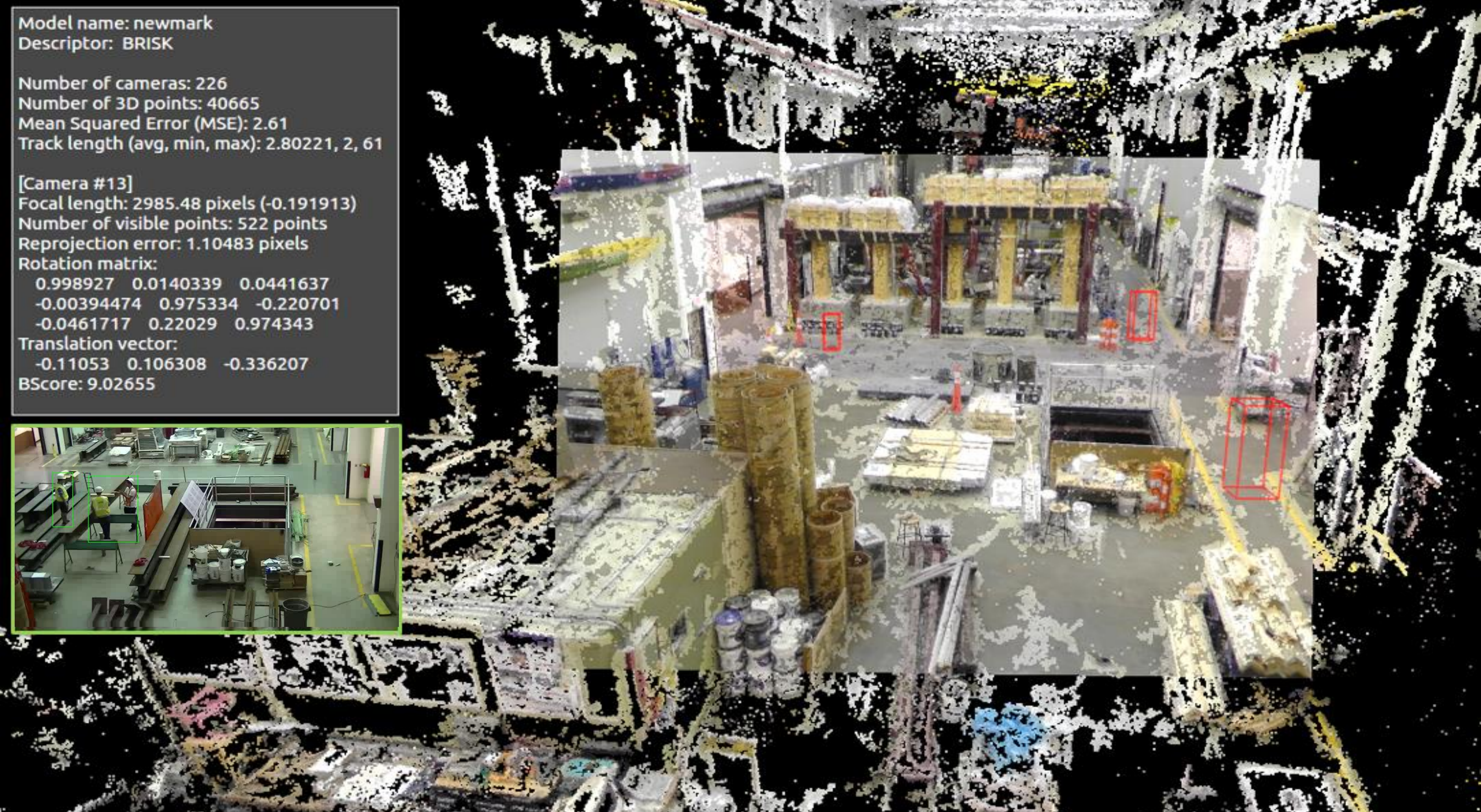
1-5 scans/month



Video sources: RAAMAC Lab



Video sources: Jun Yang, RAAMAC Lab



Video sources: RAAMAC Lab

Documentation, Intervention, Near-Miss Reporting

RAAMAC
Real-time and Automated Monitoring and Control

Right-time Intervention

Near-miss Reporting

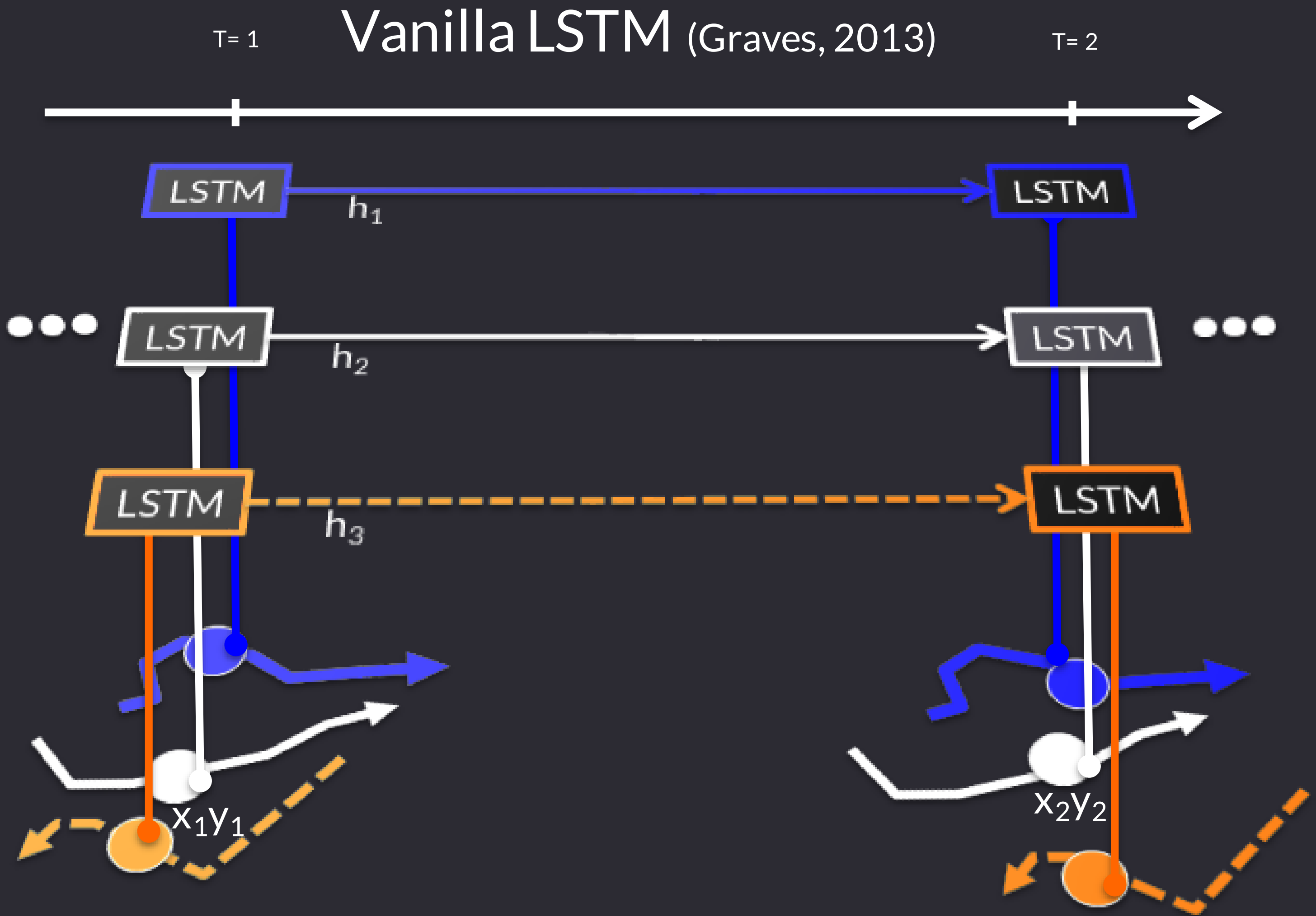
Image sources: Left: http://www.energysafetycanada.com/files/safety-alerts/Safety%20Alert%20-%2010.2018%20-%20Final.pdf Right: http://www.energysafetycanada.com/files/safety-alerts/Safety%20Alert-13%202016.pdf

Big Picture - Computer Vision & Jobsite Cameras

Detect, Track, Model Worker Activities
Understand Work Context
Predict Next Sequence of Activities

Activity Forecasting – Computer Vision

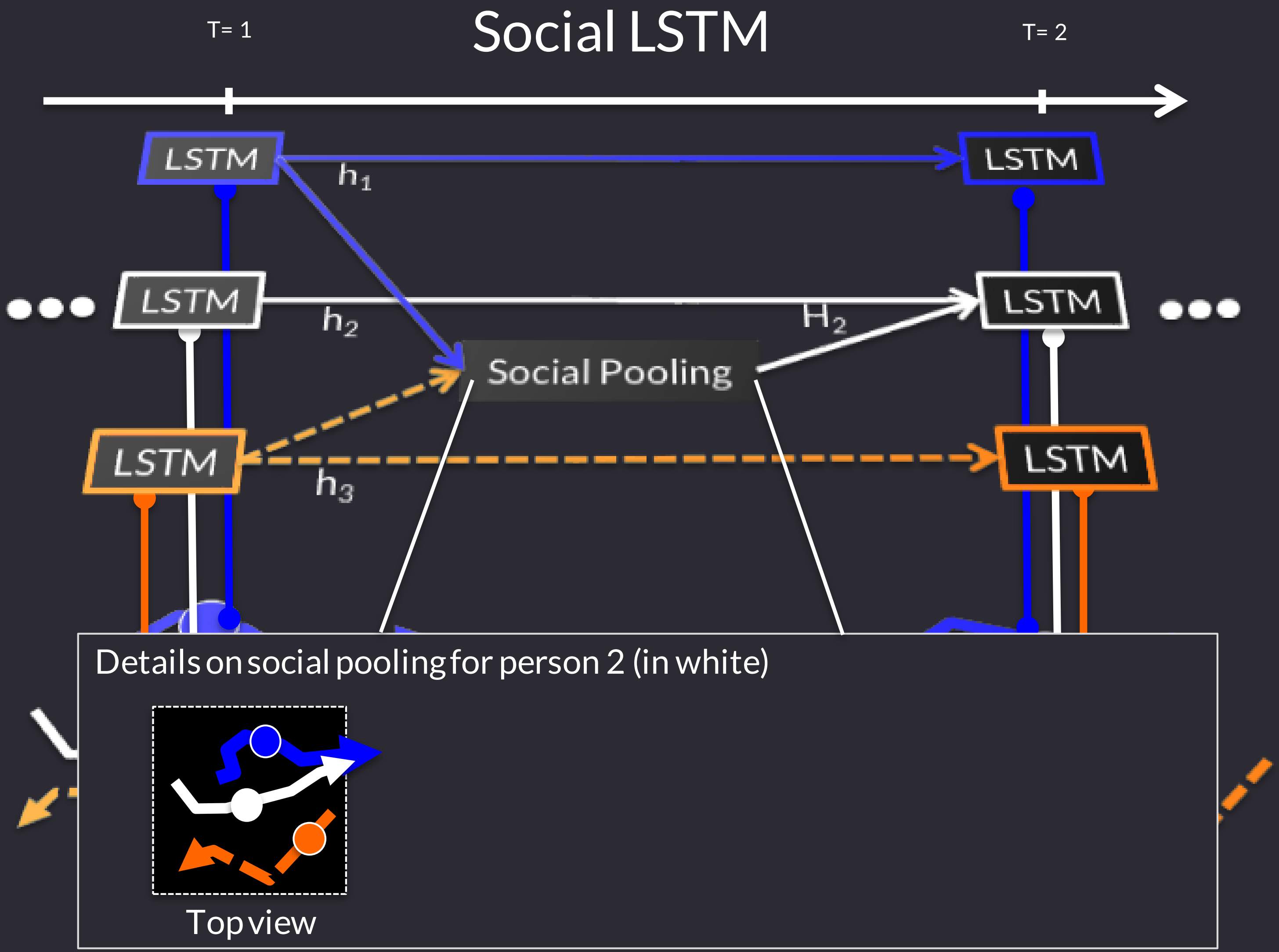
Social LSTM (Alahi *et al.* 2016)



Activity Forecasting – Computer Vision

RAAMAC
Real-time and Automated Monitoring and Control

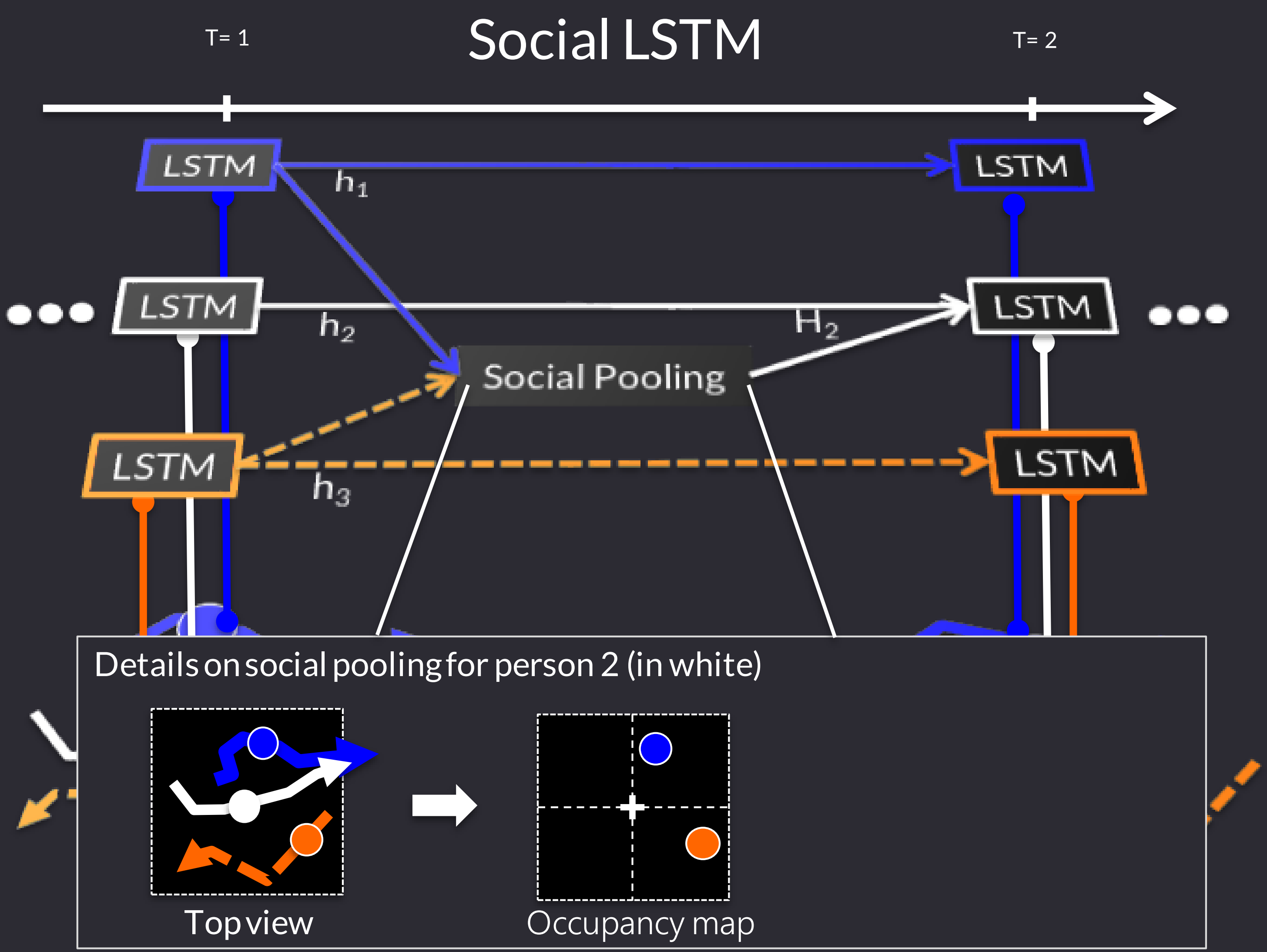
Social LSTM (Alahi *et al.* 2016)



Slide: Alexandre Alahi

Activity Forecasting – Computer Vision

Social LSTM (Alahi *et al.* 2016)

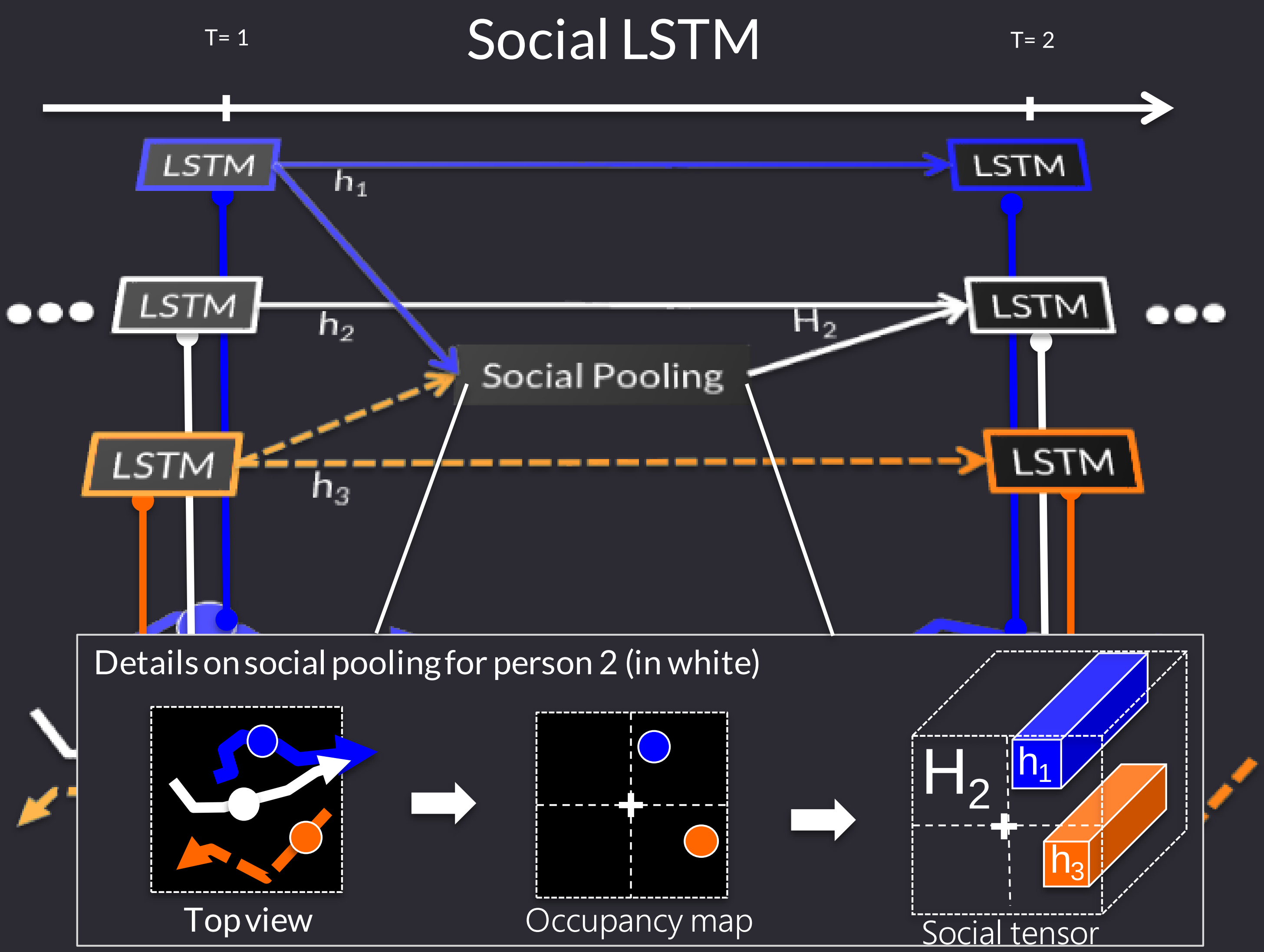


Slide: Alexandre Alahi

Activity Forecasting – Computer Vision

RAAMAC
Real-time and Automated Monitoring and Control

Social LSTM (Alahi *et al.* 2016)

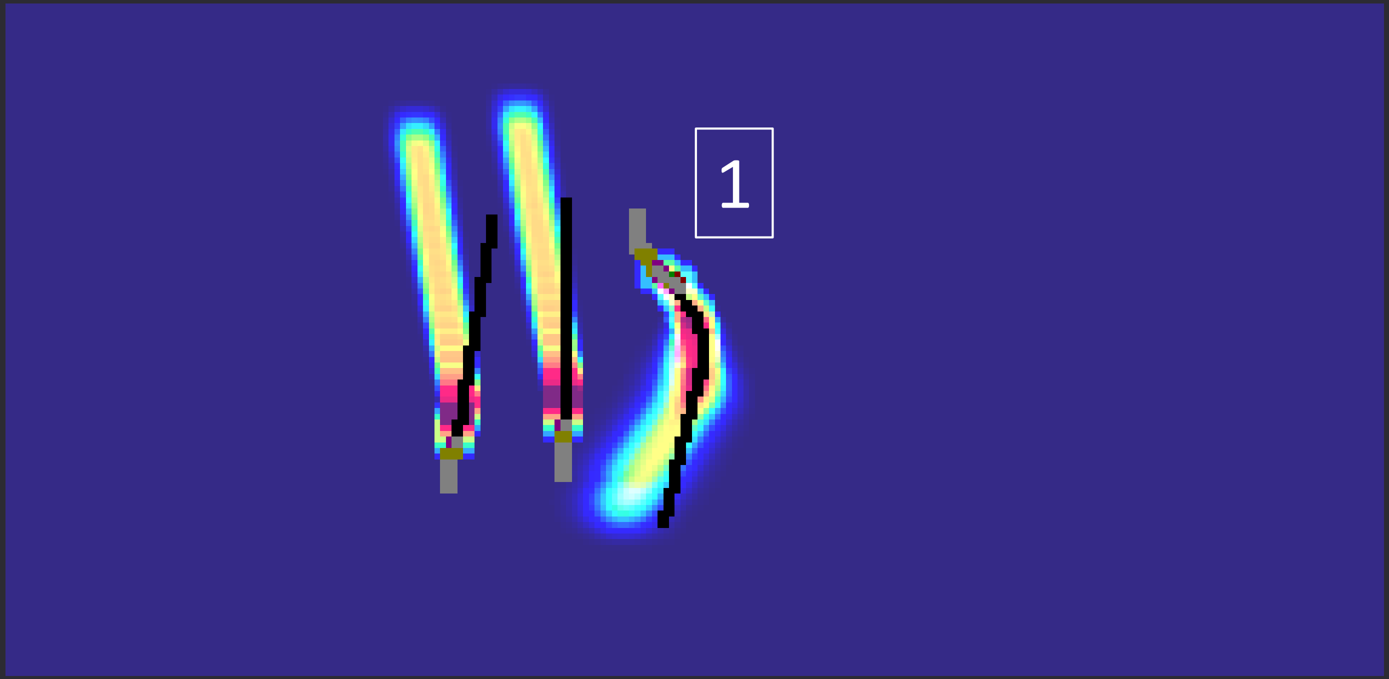


Slide: Alexandre Alahi

Activity Forecasting – Computer Vision

Social LSTM (Alahi *et al.* 2016)

- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution



Social LSTM learned to turn around a group

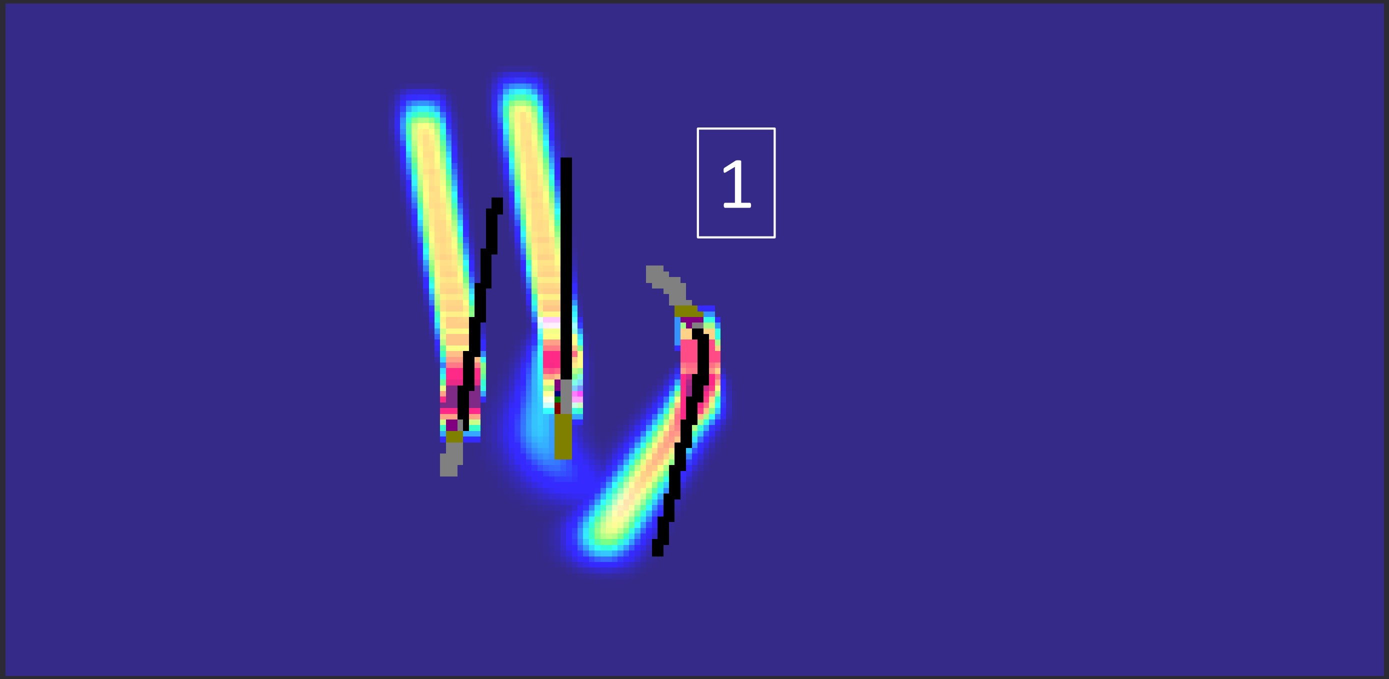
Slide: Alexandre Alahi

Activity Forecasting – Computer Vision

RAAMAC
Real-time and Automated Monitoring and Control

Social LSTM (Alahi *et al.* 2016)

- Black line is the ground truth trajectory
- Gray line is the past
- Heatmap is the predicted distribution



Social LSTM learned to turn around a group

Slide: Alexandre Alahi

From Crowd Scenes To Construction Sites

Crowd scenes from UCY and ETH dataset

Example construction sites, Google Image

From Crowd Scenes To Construction Sites

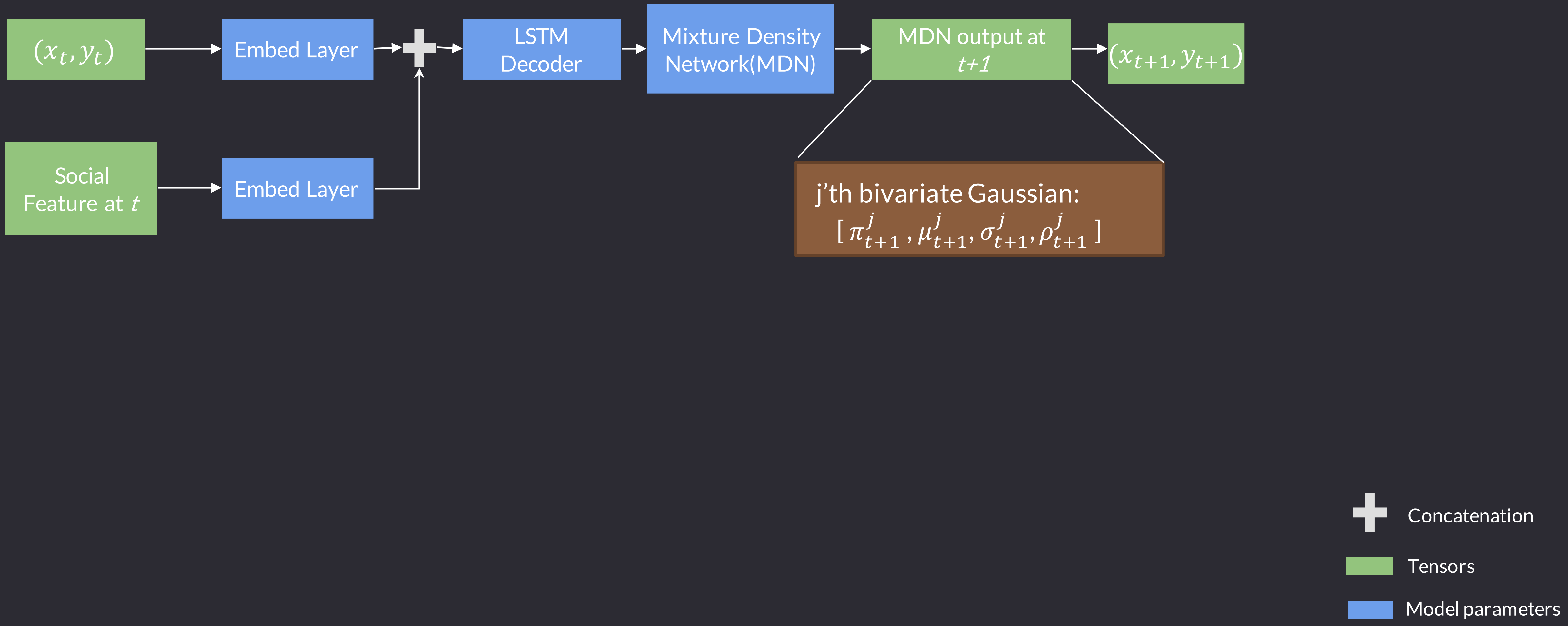
RAAMAC
Real-time and Automated Monitoring and Control

Construction sites often change drastically

Approach – data-driven, context rich, and sequence-to-sequence models

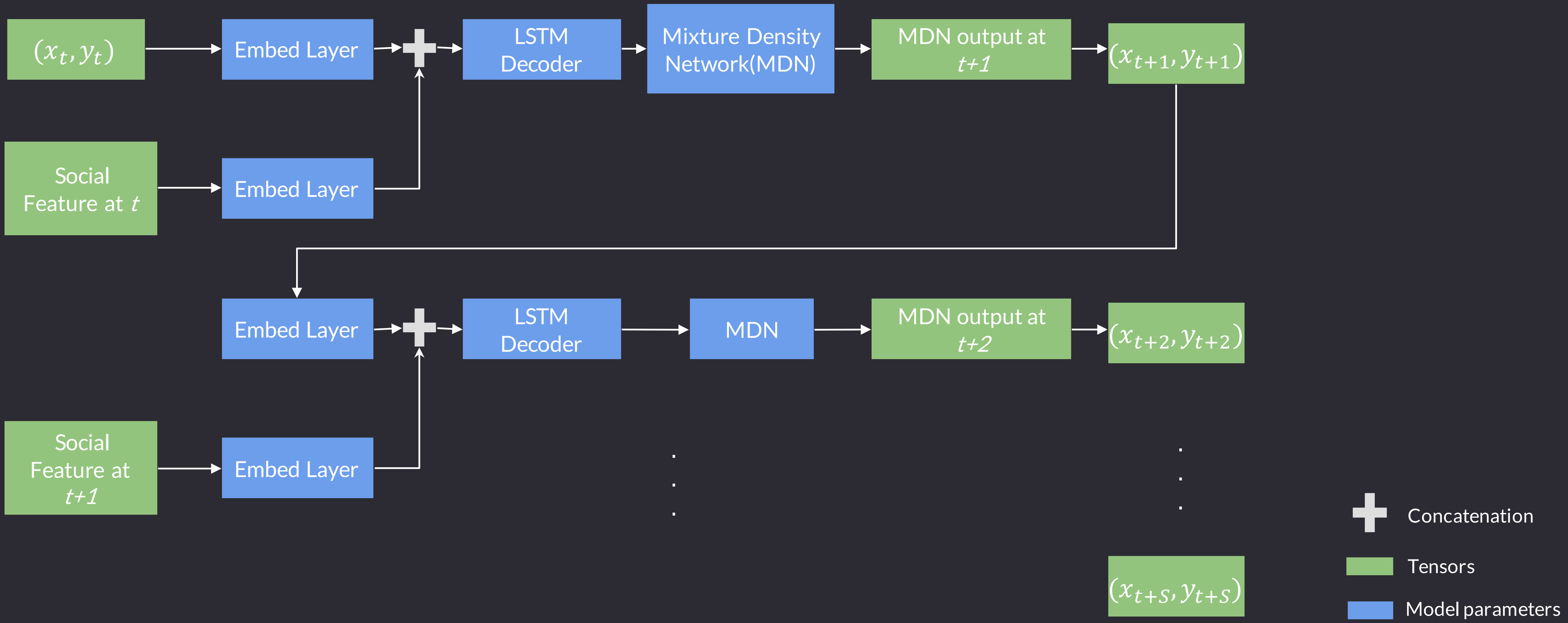
Model Architecture (Social LSTM)

For *i* ‘th trajectory at time *t*… predict *i* ‘s location at *t+1*



Model Architecture (Social LSTM)

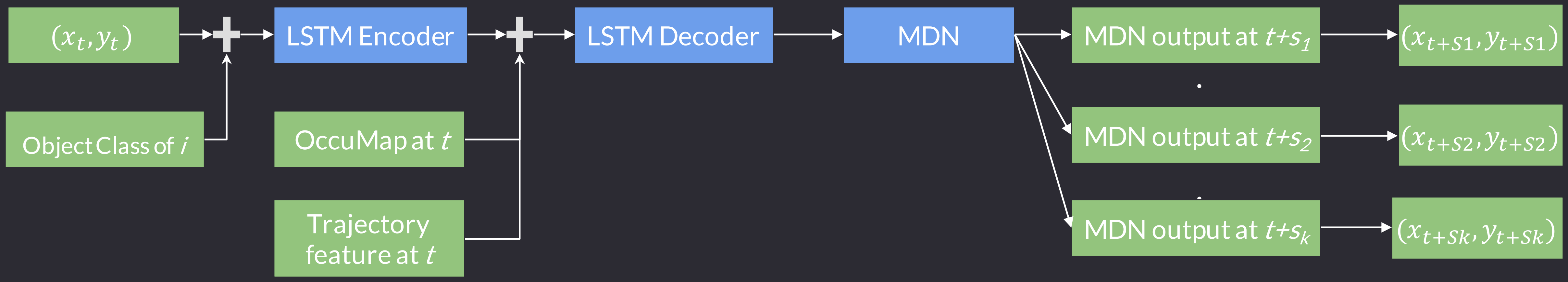
For *i* ‘th trajectory at time *t*… predict *i* ‘s location at *t+1*



Model Architecture (Ours)

RAAMAC
Real-time and Automated Monitoring and Control

For i 'th trajectory at time t… predict i 's location at $\{t+s_1, t+s_2, \ldots, t+s_k\}$



Concatenation

Tensors

Model parameters

Model Architecture (Ours) - Occupancy Map

Model Architecture (Ours)

RAAMAC
Real-time and Automated Monitoring and Control

Trajectory Features From Common Trajectories

Color Code and Movement
Blue South West to North East
Lime: North East to South West
Red: East to West
Yellow: North to South

Length: Average length of all trajectories belonging to the cluster

Thickness: Cluster size (number of Trajectories in the cluster)

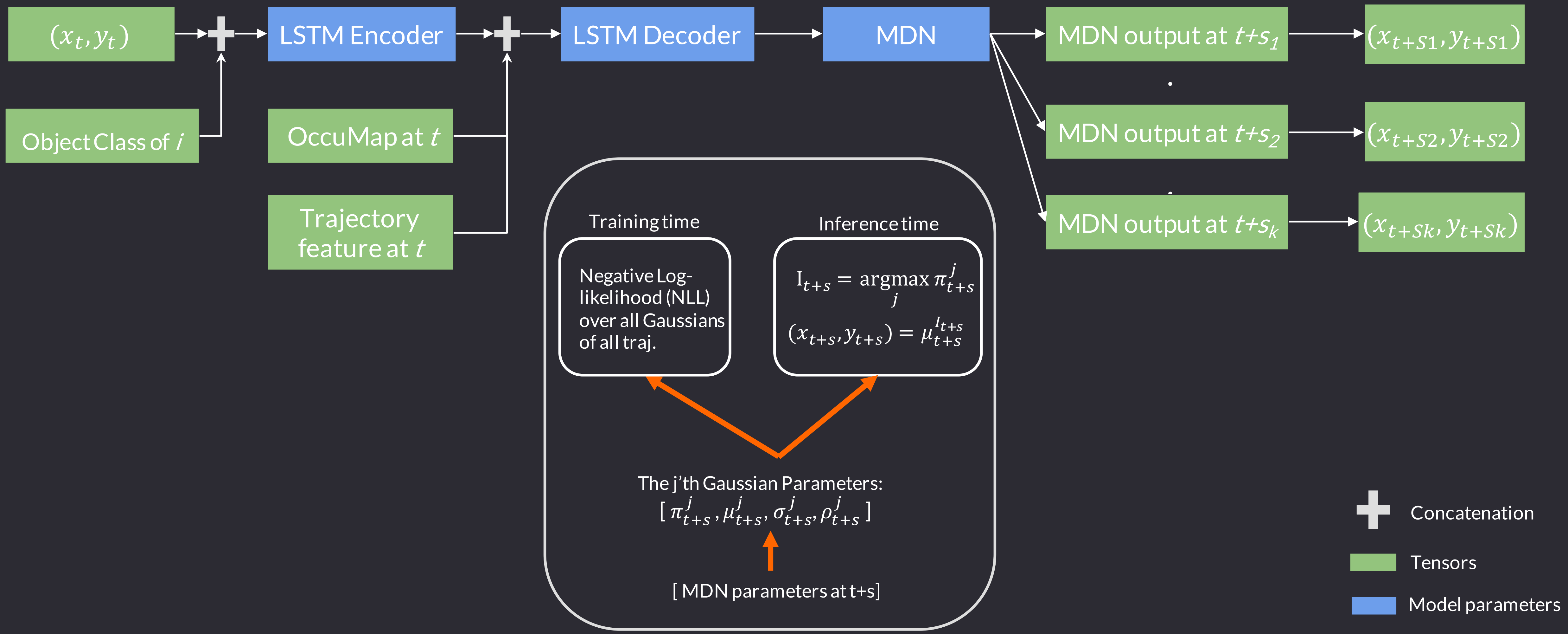
Model Architecture (Ours)

RAAMAC
Real-time and Automated Monitoring and Control

Iteratively Use Predicted Locations As Inputs Lead to Large Deviations

Model Architecture (Ours)

For i ‘th trajectory at time t… predict i ‘s location at $\{t+s_1, t+s_2, \dots, t+s_k\}$



Case Study At Nvidia Voyager Site

RAAMAC
Real-time and Automated Monitoring and Control

Experiment Setup

- Voyager dataset:
  - 1,464 mins (24.4 hrs) of 1080p videos
  - Trainval set (from 76 clips): person 1630, vehicle 1752
  - Test set (from 29 clips): person 143, vehicle 161
  - Traj. duration : [30, 2000] steps , endpts dist. > 50 pixels

- TrajNet dataset:
  - 58 scenes from UCY, ETH and SSD dataset
  - 11,448 pedestrian traj.
  - 20 steps each traj., world coordinates in meter.

Implementation Details

- Running on one RTX 2080 Ti GPU with Nvidia docker image

- Optimization tricks:
  - gradient clipping to 50% gradient norm
  - Adam optimizer, lr = 0.005, lr decay to 50%

- Dynamic length batches
- Pre-computed features for accelerating training speed.
- Training time:
  - Voyager: 1 hr for 1000 epochs with 3 MDN output heads
  - Trajnet: ~30 mins for 1700 epochs with 12 MDN output heads

Experimental Results – Voyager dataset

RAAMAC
Real-time and Automated Monitoring and Control

Experiment results and ablation study (error in pixels)

Group	ID	Method	RMSE@10	RMSE@20	RMSE@40
Baselines	1	Linear Reg ($p = 1$)	62.47	68.59	82.51
	2	VAR ($p = 5$)	46.85	90.27	163.02
	3	MLP + Reg	14.17	27.08	50.16
	4	LSTM+Reg	8.67	14.65	27.39
Ours	5	LSTM+MDN	7.42	13.26	25.25
	6	LSTM+MDN (single output)	7.51 (0.22)*	13.30 (0.34)	25.20 (0.45)
	7	LSTM+MDN+OccuMap	7.24 (0.02)	**12.70 (0.008)**	24.30 (0.01)
	8	LSTM+MDN+Attribute	**7.22 (0.0003)**	12.95 (0.01)	24.74 (0.02)
	9	LSTM+Traj. Feature	7.39 (0.03)	12.89 (0.05)	24.45 (0.03)
	10	LSTM+MDN+OccuMap +Attribute	7.30 (0.09)	12.71 (0.005)	**24.22 (0.004)**
	11	LSTM+MDN+OccuMap +Attribute + Traj. Feature	7.36 (0.04)	13.06 (0.03)	24.54 (0.008)

* p-values against method 5 (LSTM+MDN), $p < 0.05$ means two results are different with statistical significance

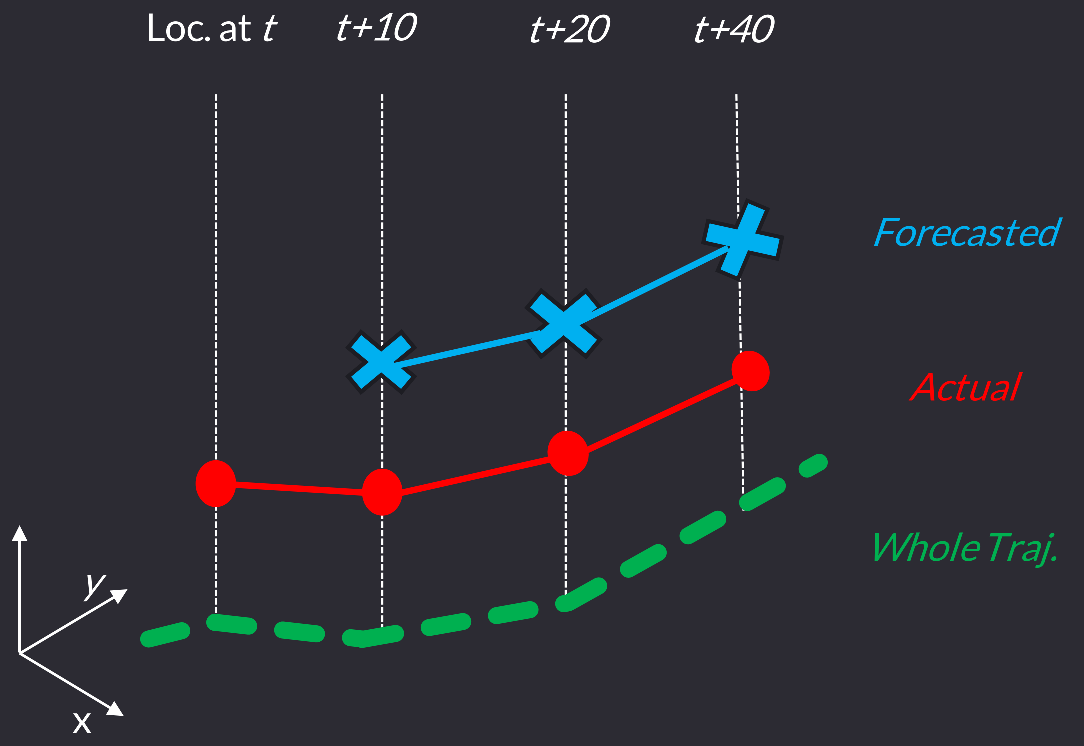
Experimental Results – TrajNet dataset

Tentative comparison between Social LSTM and Ours (error in meters)

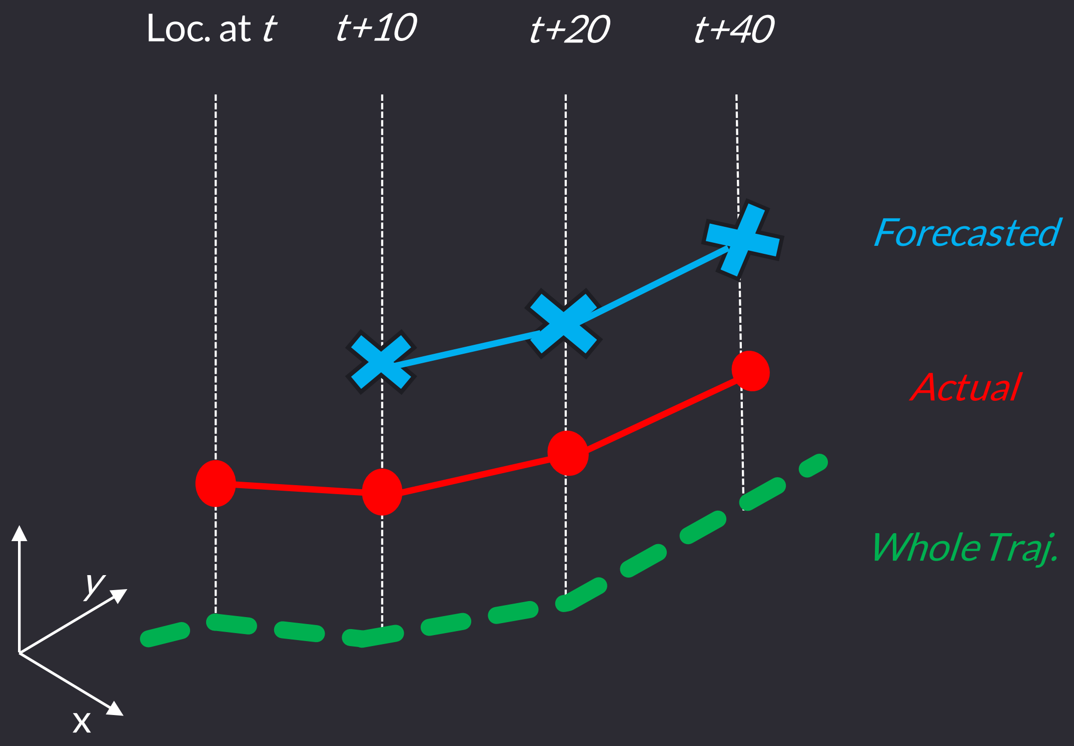
Group	ID	Method	Average error	Final error	Mean error
Social LSTM*	9	Occupancy LSTM	2.1105	3.12	1.101
	10	Social LSTM	1.3865	2.098	**0.675**
Ours**	4	LSTM+Reg	1.039	1.382	0.696
	5	LSTM+MDN	1.036	1.377	0.694
	7	LSTM+MDN+OccuMap	**1.028**	**1.370**	0.686

*Unofficial Implementation from https://github.com/quancore/social-lstm
**cross validation result on train set because evaluation server not available

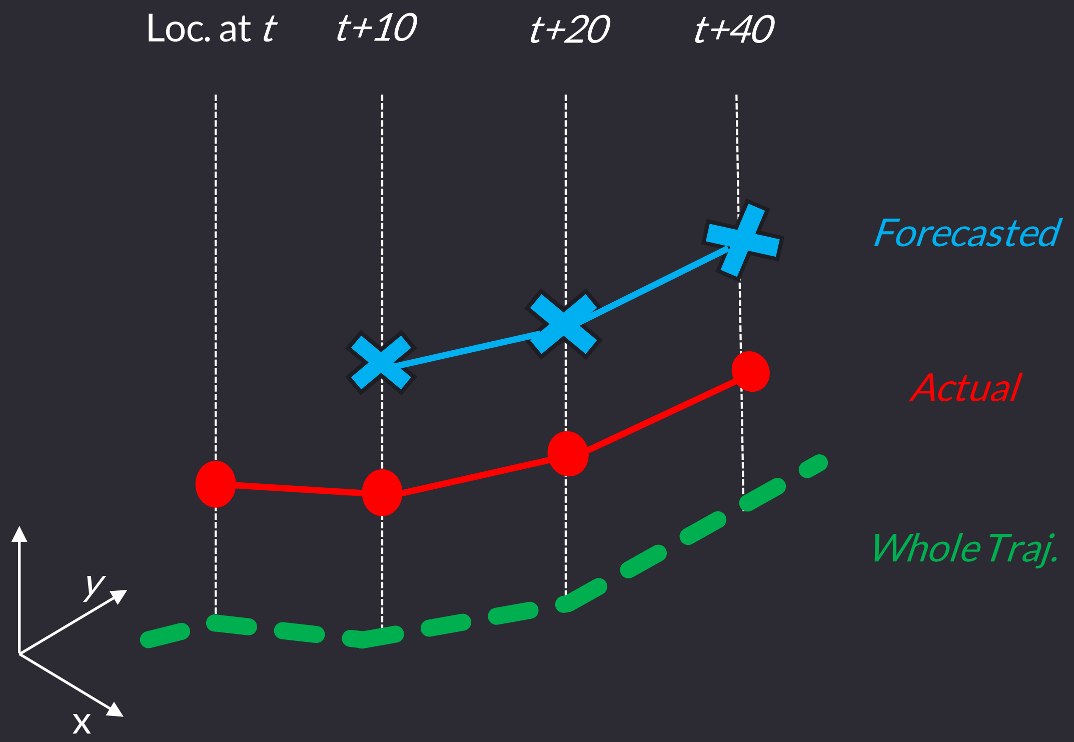
Qualitative Results – Easy Example



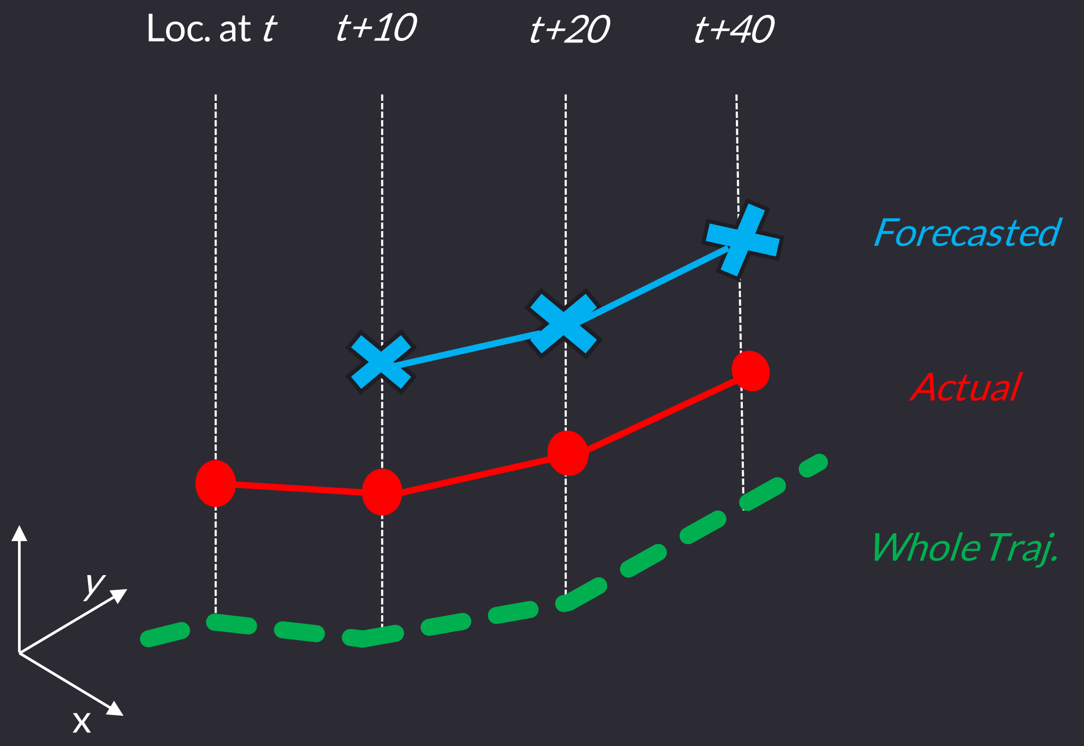
Qualitative Results – Easy Example



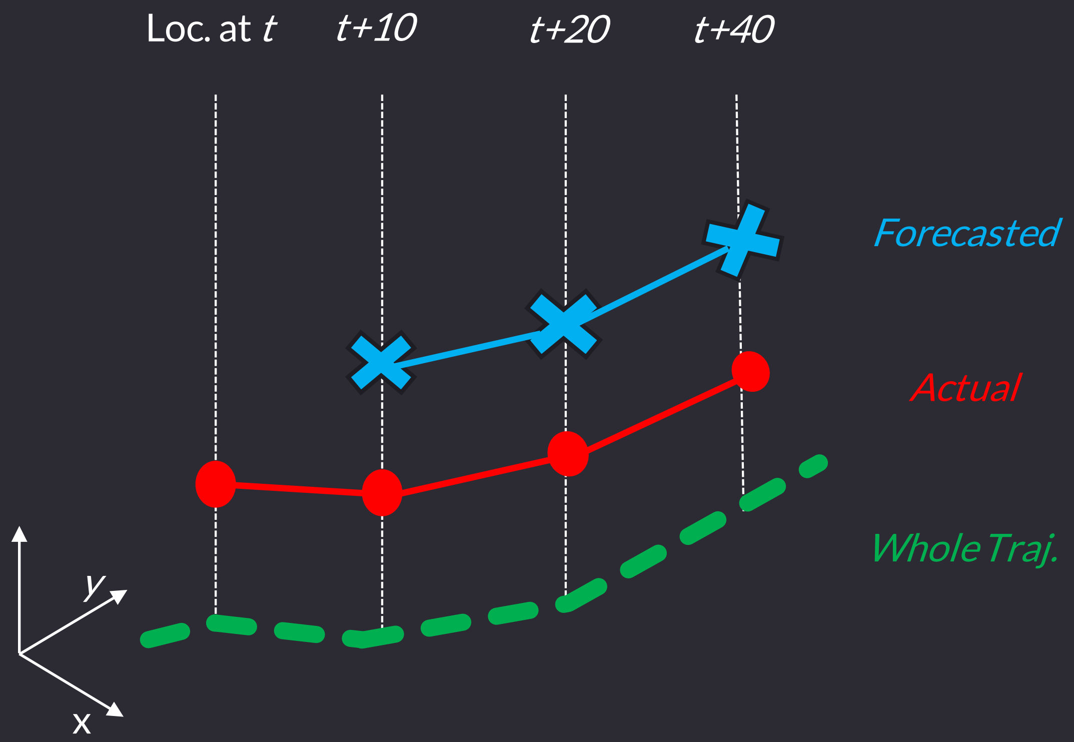
Qualitative Results - Intermediate Difficulty



Qualitative Results - Intermediate Difficulty



Qualitative Results – Hard Example



A Safety Application Prototype

Input video

Object Detection
+
Object Tracking

Task: Forecast Entering Excavation Zone Events

1. Using trajectory forecasting model to predict person/vehicle's future locations in 0.5/1.0/2.0 seconds
2. Matching predictions to human-defined excavation zones.

Object detection + tracking:

- Mask RCNN (Resnet-101 backbone, Caffe2 Model zoo) for Person & Vehicle
- SORT for tracking Person & Vehicle objects

A Safety Application Prototype

Admin panel to modify regions of interest

Figure 1

Keyboard-binding instructions :

-press '1': Insert mode off and edit mode on;
-press '2': Insert mode on and edit mode off;
-press '3': export current polygons to visdom display;
-press 'r': refresh frame

In insert mode:
click points and finish drawing by click back the first point
-hold 'shift' and left mouse click to move the whole polygon;
-hold left mouse click to move one vertex;
-press 'esc' to start a new polygon

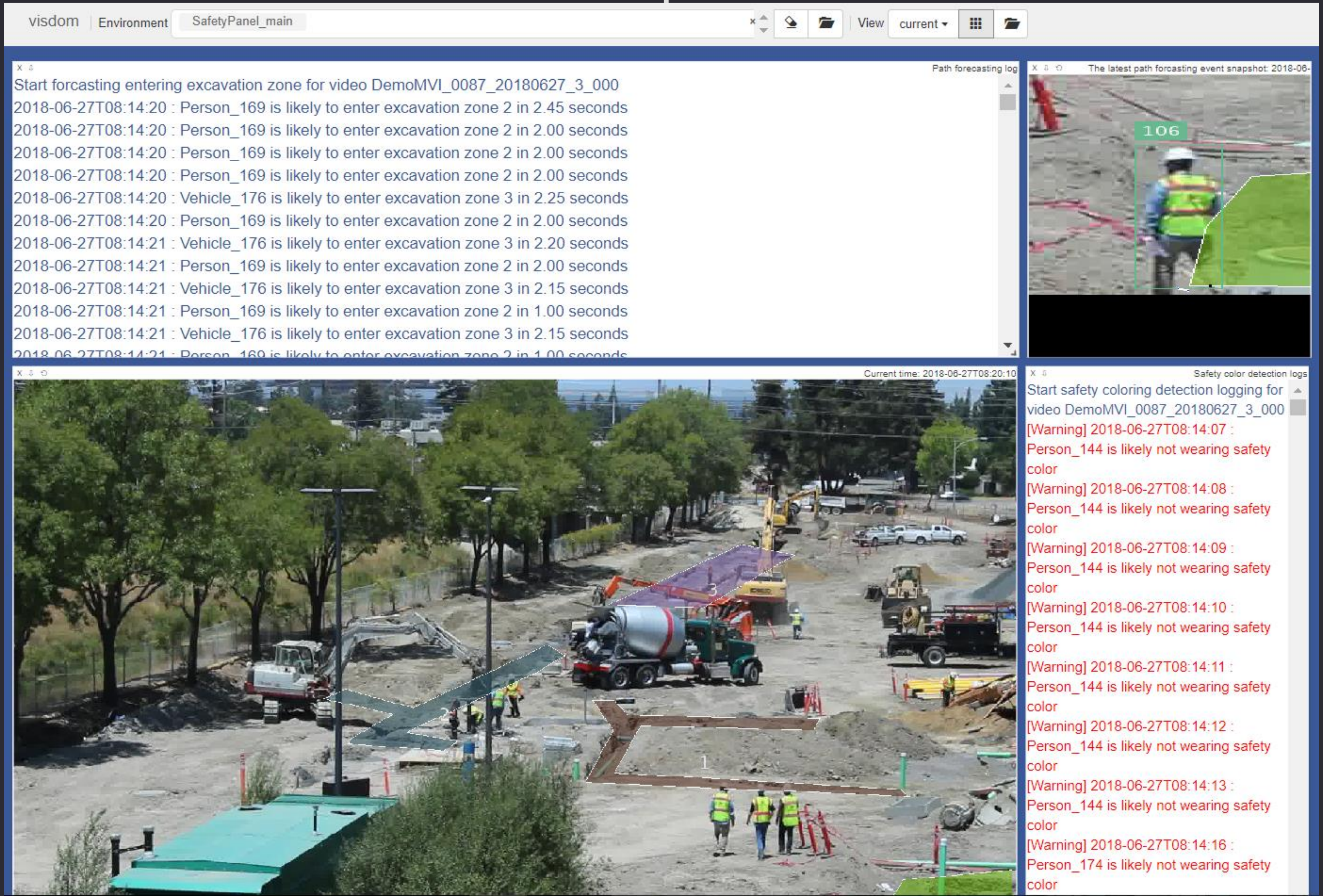
In edit mode:
-press '4': delete the polygon containing the point
-press 't': toggle vertex markers on and off.
-press 'd': delete the vertex under point when markers on
-press 'i': insert a vertex at point when markers on,
within 10 pixels of the line connecting two existing vertices.

Ploy at time: 701, Insert new polygon mode: off

A Safety Application Prototype

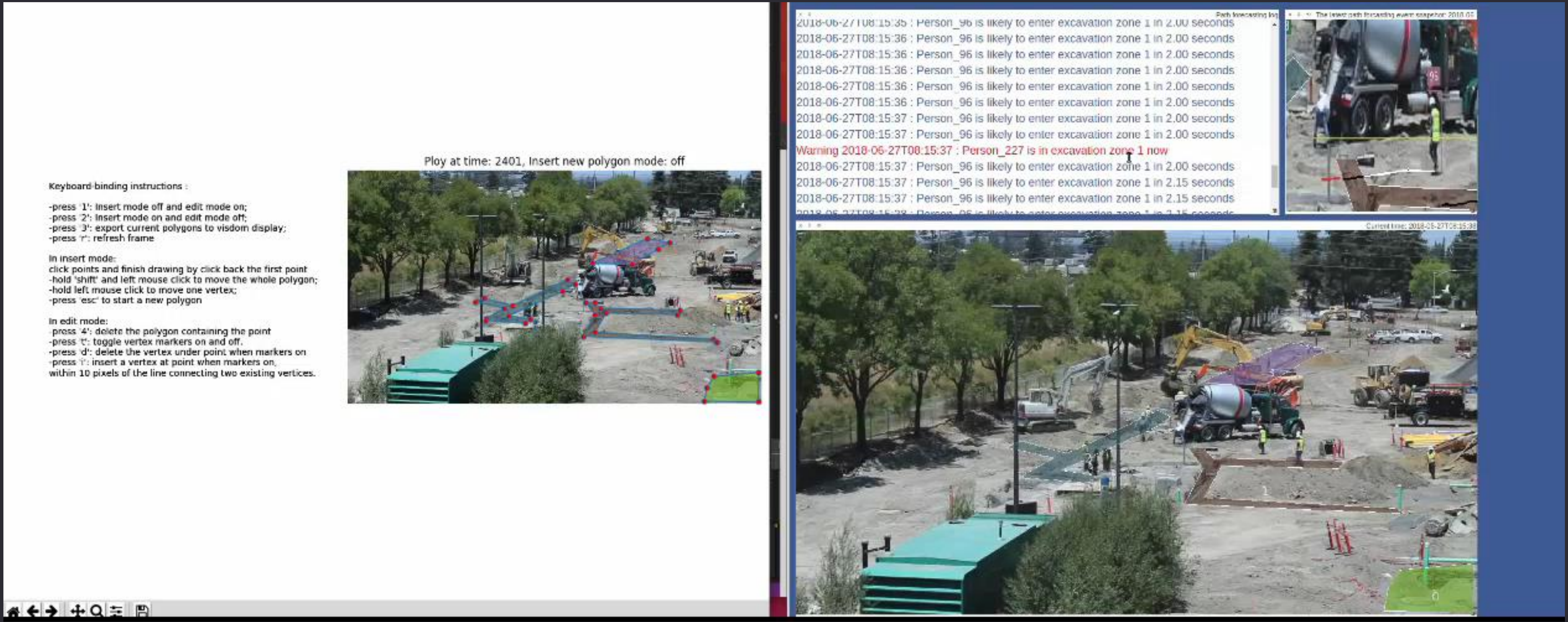
RAAMAC
Real-time and Automated Monitoring and Control

Viewer panel



A Safety Application Prototype

Demo Video



Summary

- Improving construction safety requires more frequent, accurate and proactive inspections.

- We show detection, tracking, and trajectory forecasting models are promising ways to improve predictive construction safety management.

GTC, Santa Jose, 2019

Video-Based Activity Forecasting for Construction Safety Monitoring Use Cases

Shuai Tang

stang30@illinois.edu

NVIDIA

RAAMAC

Real-time and Automated Monitoring and Control

ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN