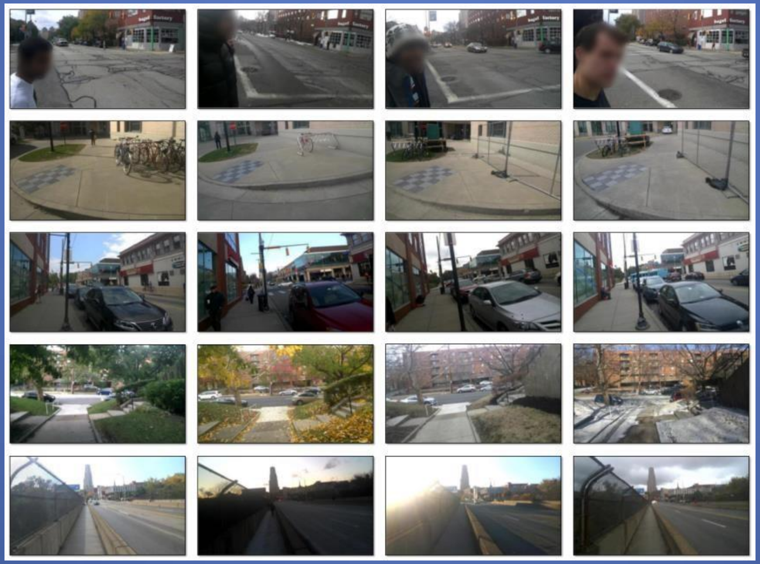
KrishnaCam: Using a Longitudinal, Single-Person, Egocentric Dataset for Scene Understanding Tasks

Krishna Kumar Singh   Kayvon Fatahalian   Alexei A. Efros



Presented By:

Shubham Sharma

Image Credit: Krishna Kumar Singh

Objective

Organize a large egocentric video collection of real-world data from a single individual into a richly annotated database

How much novel visual information does an individual see each day?
Can we predict where the individual might walk next?

Motivation

- *"A baby has brains, but it doesn't know much. Experience is the only thing that brings knowledge, and the longer you are on earth the more experience you are sure to get."* —L. Frank Baum, The Wonderful Wizard of Oz

- The goal is to extract value from life events.

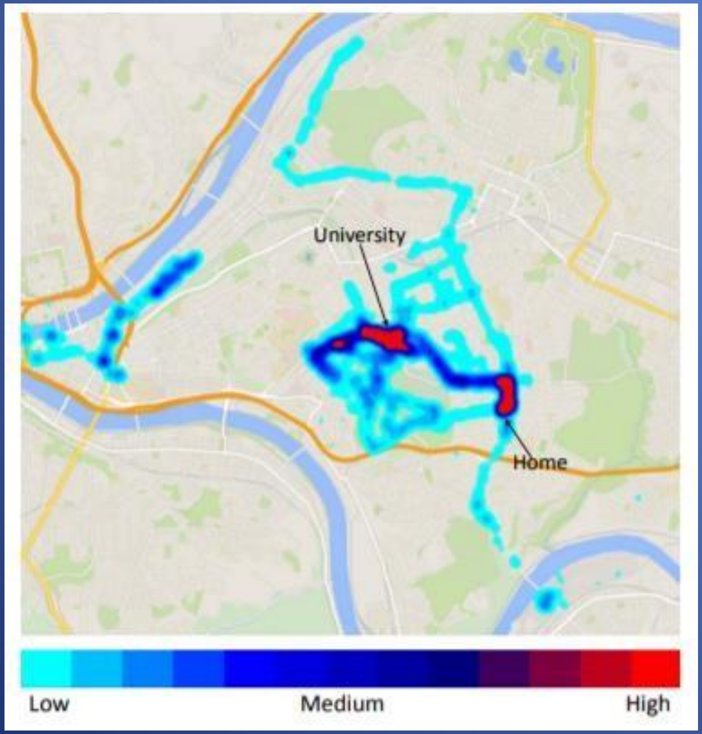
Image credits: Krishna Kumar Singh et al.

Agenda

- Creation of the KrishnaCam new dataset
- Quantification of novel visual data
- Trajectory estimation and motion class prediction
- Experimental evaluation
- Applications
- Strengths and Weaknesses

The KrishnaCam dataset

- Over a period of 9 months, collect and record the events in the life of a graduate student
- Data still being recorded.



Heat map of locations visited

Image Credit: Krishna Kumar Singh et al.

The KrishnaCam dataset

Walking in urban/campus/ residential areas, waiting at intersections and for bus

Shopping, eating

Evening and night recording

Activities in parks, at events

Seasonal change

Socializing with friends

Time-span: 9 months

Duration: 70 hours

Total clips: 460

Device: Google Glass

Data: 720 p, 30 fps Accelerometer, Gyroscope, Orientation, GPS

How much novel visual data is present?

Lot’s of redundant data!

Identify top-5 nearest neighbors of frame in prior recordings.

NN frames constrained to be separated by at least 10 minutes

Novel if the average similarity of its top-5 nearest neighbors is below threshold or if no neighbor.

Results of Novel Visual Data Growth

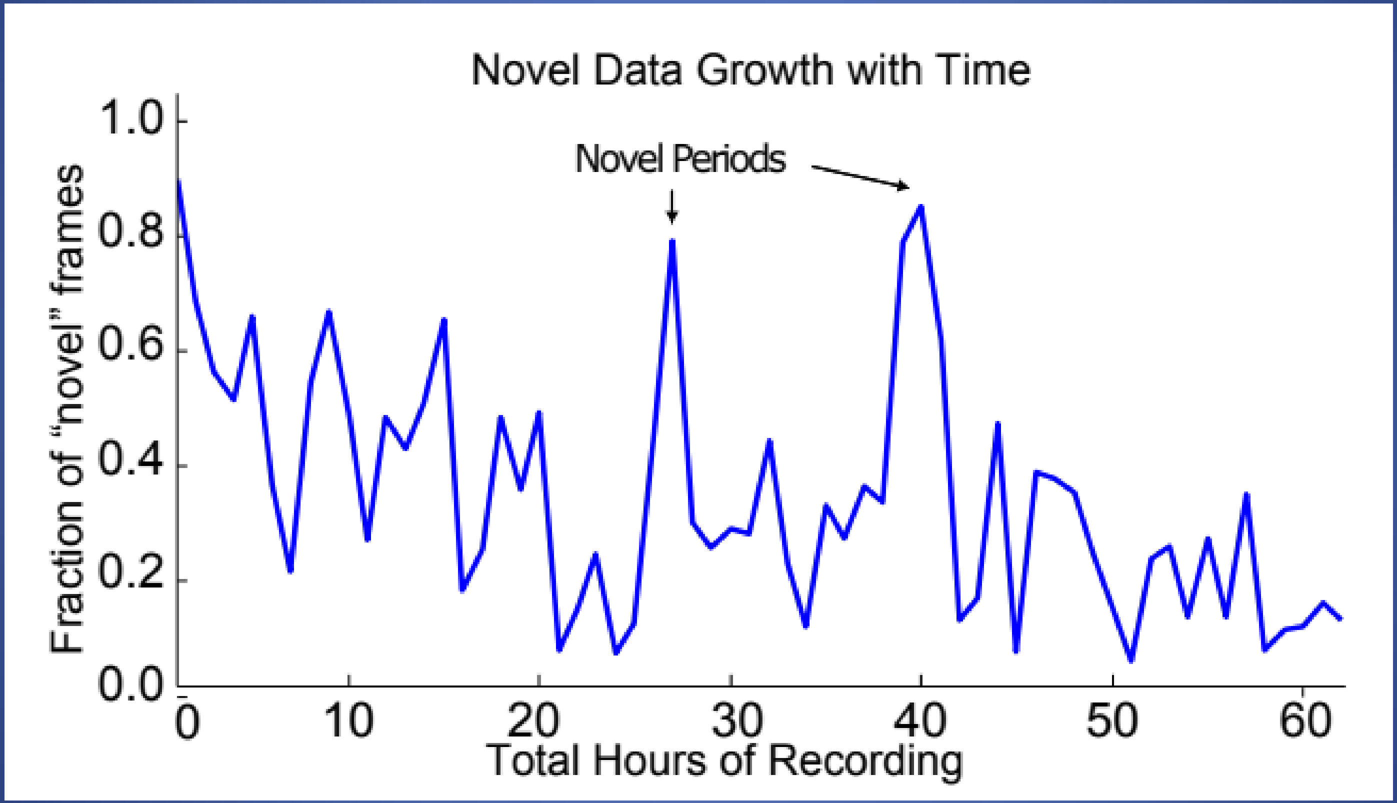


Image Credit: Krishna Kumar Singh et al.

Results of Novel Visual Data Growth

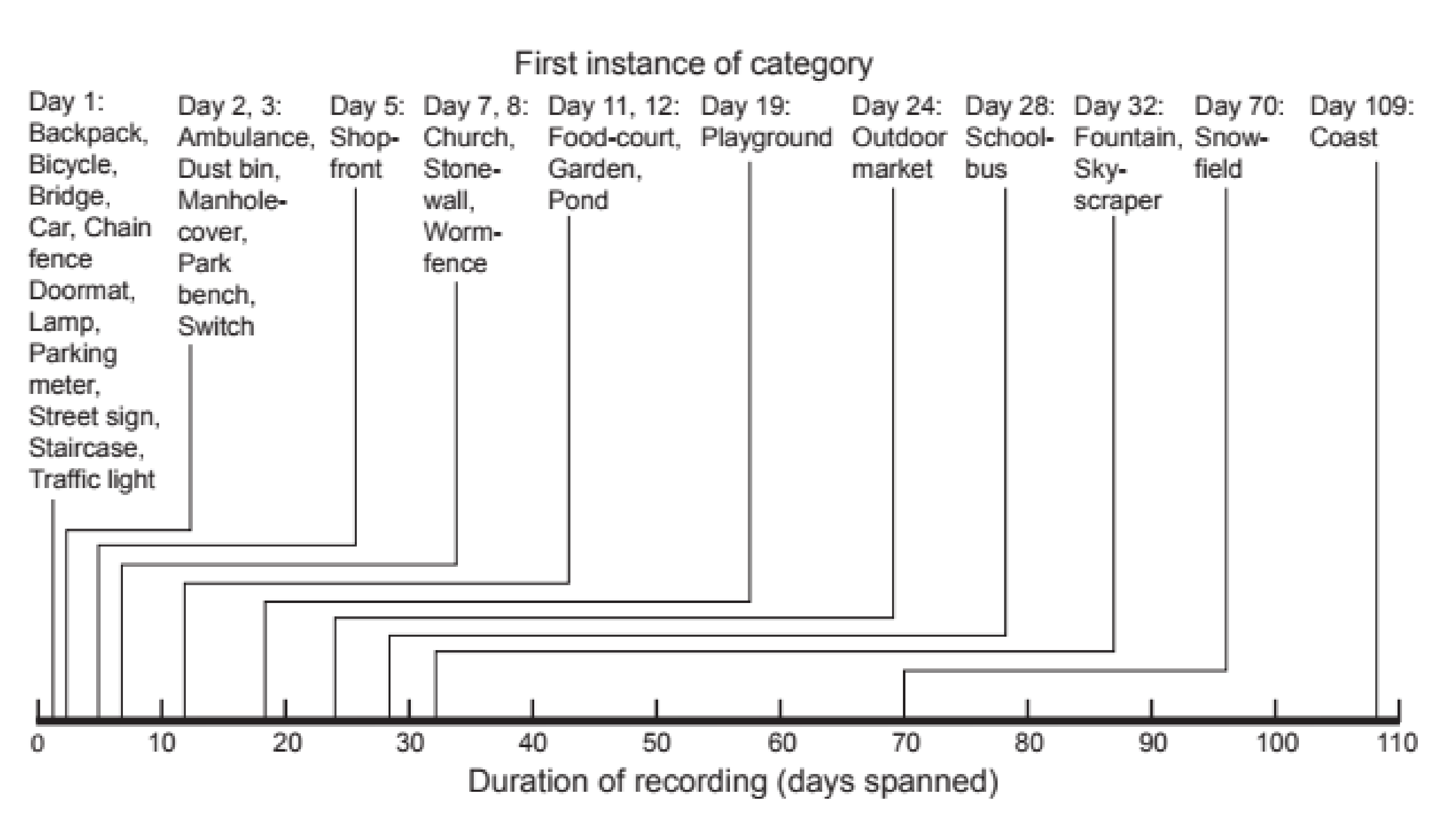


Image Credit: Krishna Kumar Singh et al.

Motion Prediction

- Given a single image, can we predict where the student would walk next in the scene?

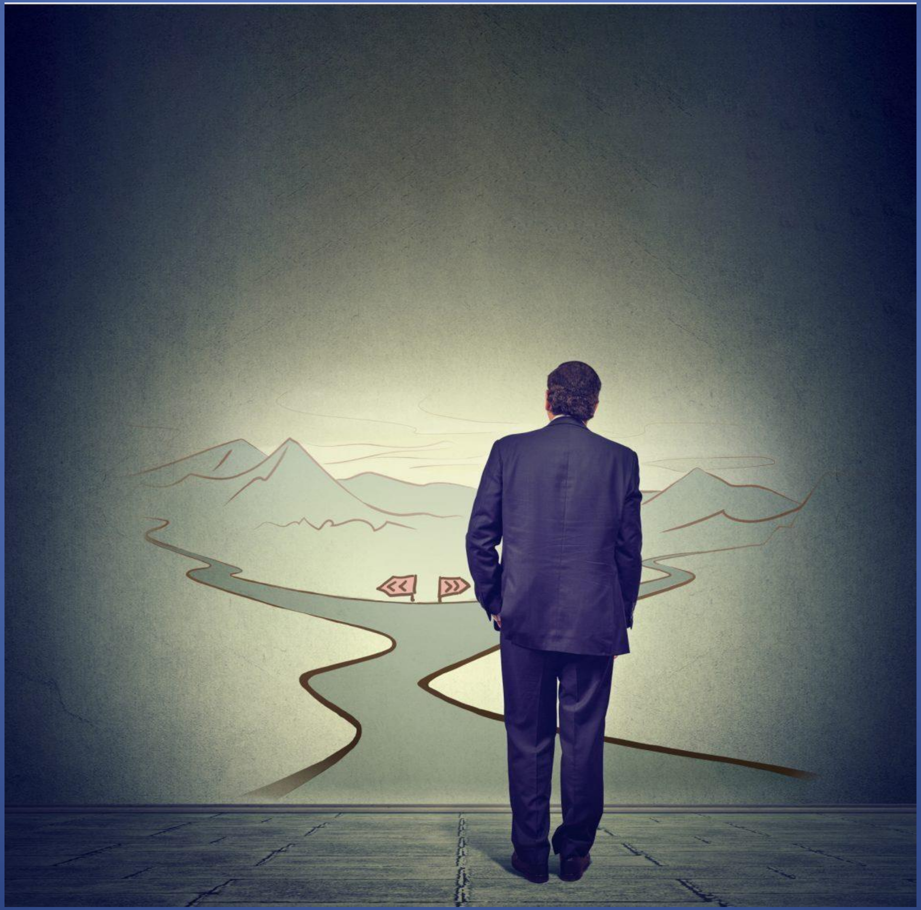


Image Credit: http://paragonroad.com/krishna-pendyala-legacy-by-design-not-by-default/

Motion Prediction: Ground-Truth data

How do we get ground-truth trajectories in this huge dataset? Manual annotation?

I am not labeling that!

Image Credit: https://beinspiredchannel.com/frustrated-frustration/

Motion Prediction: Ground-Truth

- Estimating ground-truth motion trajectories: GPS is inaccurate for location prediction.

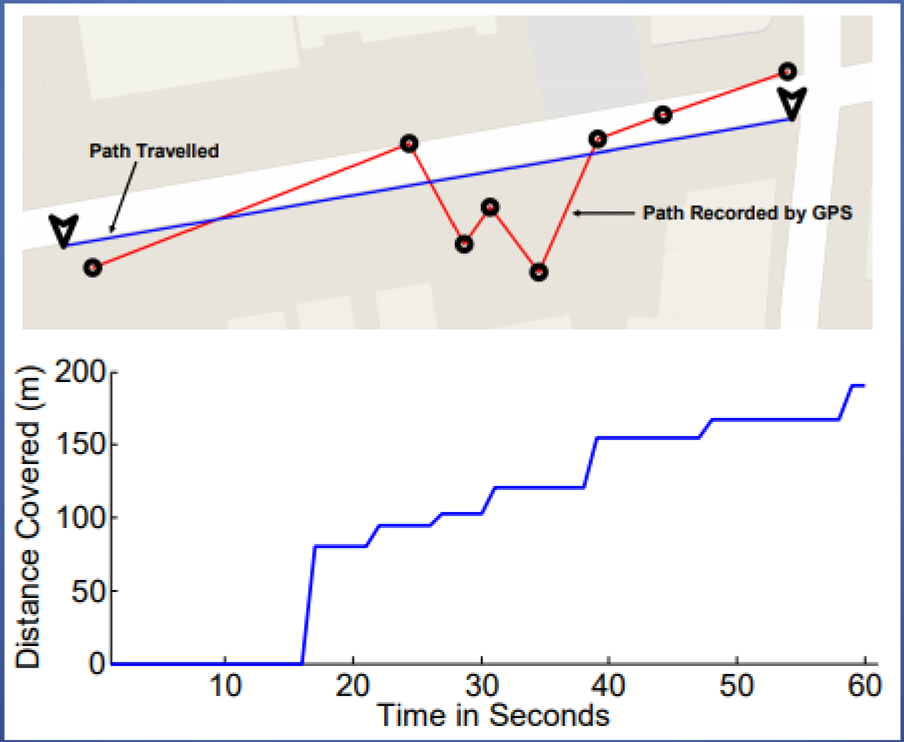
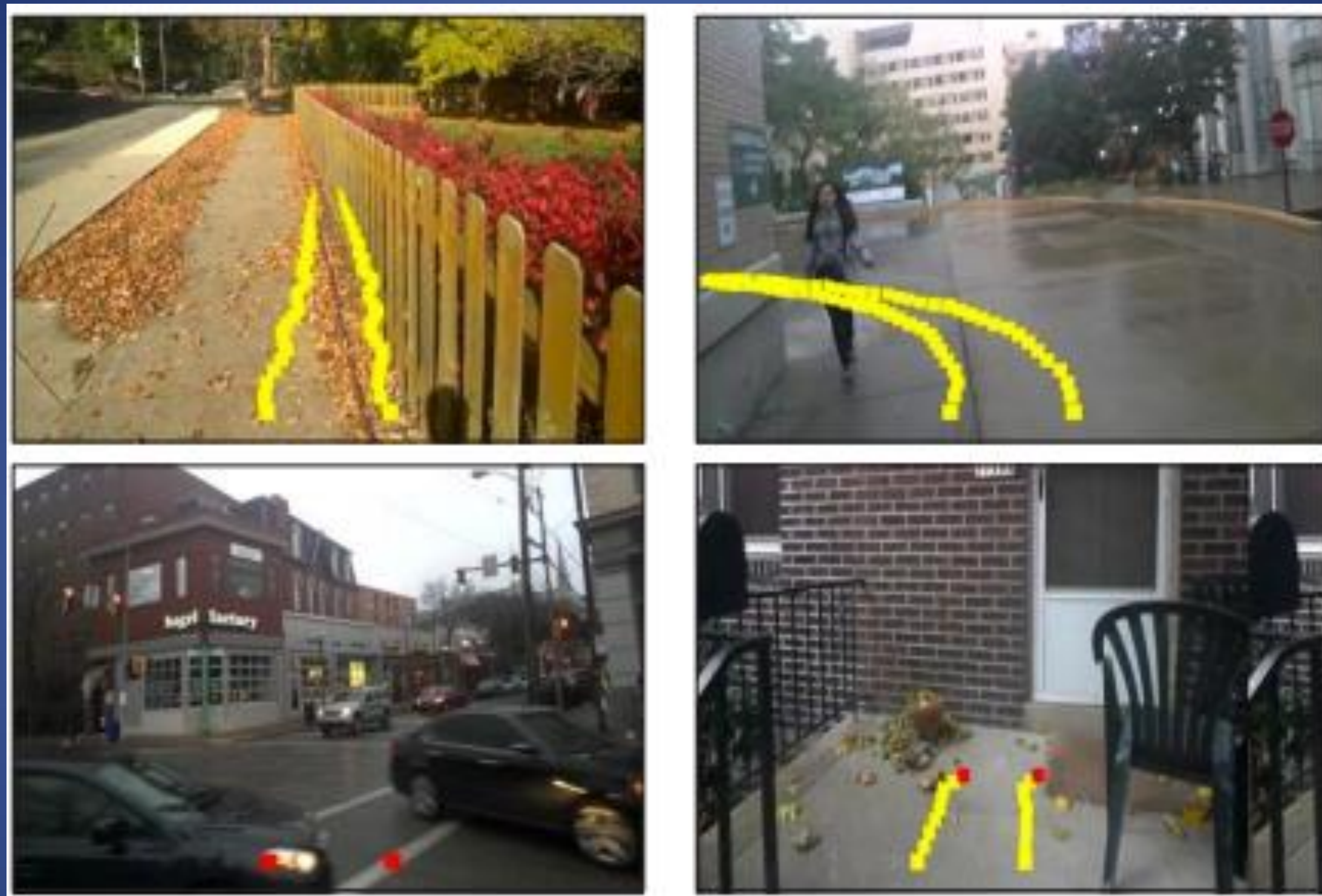


Image Credit: Krishna Kumar Singh et al.

Motion Prediction: Ground Truth

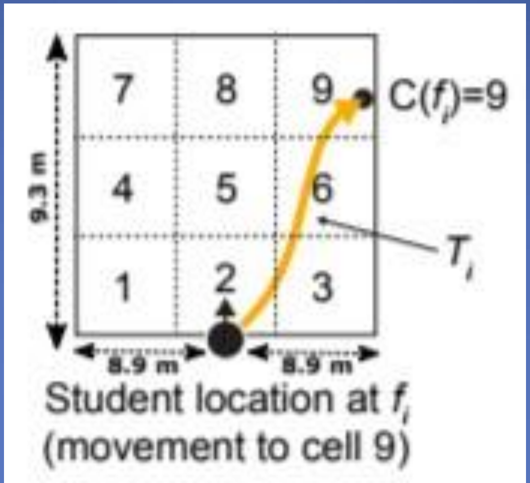
- A multi-class SVM is trained with accelerometer and orientation sensor readings.
- 4 classes of velocity: stationary, slow, regular and fast.
- Using this velocity and orientation, find 7 second trajectories.



Ground truth 7-second motion trajectories obtained from accelerometer and orientation measurements. The red dots represent stationary behavior.

Image credits: Krishna Kumar Singh et al.

Motion Class prediction

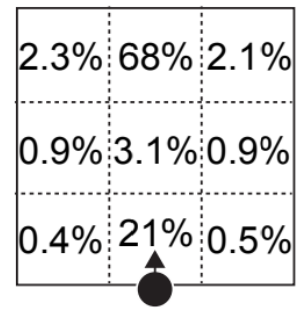


- Ground partitioning. C(fi) is the final position
- To learn C(fi), modify the final softmax layer of the MIT Places-Hybrid Network to predict nine motion classes
- Training: 38 hours (681,565 frames, September 18 to March 2)
- Testing: 252,209 frames (collected between 38 and 52 hours)

Image Credit: Krishna Kumar Singh et al.

Results

Motion Class Prediction Accuracy (Unweighted)	Unvisited (%)	Visited (%)	Overall (%)
Fine Tuned	58.4	81.2	73.4
NN	54.9	81.4	72.2
Chance	43.2	51.3	48.5



Distribution of movements in training set

- The dataset is heavily biased towards instances of walking straight.
- To remove bias, for each training frame, scale the gradient used for back-propagation by the size of the frame's motion category

Image Credit: Krishna Kumar Singh et al.

Results: weighted model

Class	Unvisited (%)	Visited (%)	Overall (%)
1 (Hard-Left)	1.2	11.2	7.9
2 (Stop)	26.2	56.2	41.8
3 (Hard-Right)	4.8	29.3	22.9
4 (Med-Left)	9.6	27.7	23.6
5 (Med-Straight)	25.6	27.3	26.6
6 (Med-Right)	7.1	22.6	18.9
7 (Soft-Left)	20.4	48.4	40.3
8 (Straight)	35.4	57.4	51.3
9 (Soft-Right)	16.8	38.9	31.6
Overall	16.4	35.4	29.4

Per-class motion prediction accuracy

Image Credits: Krishna Kumar Singh et al.

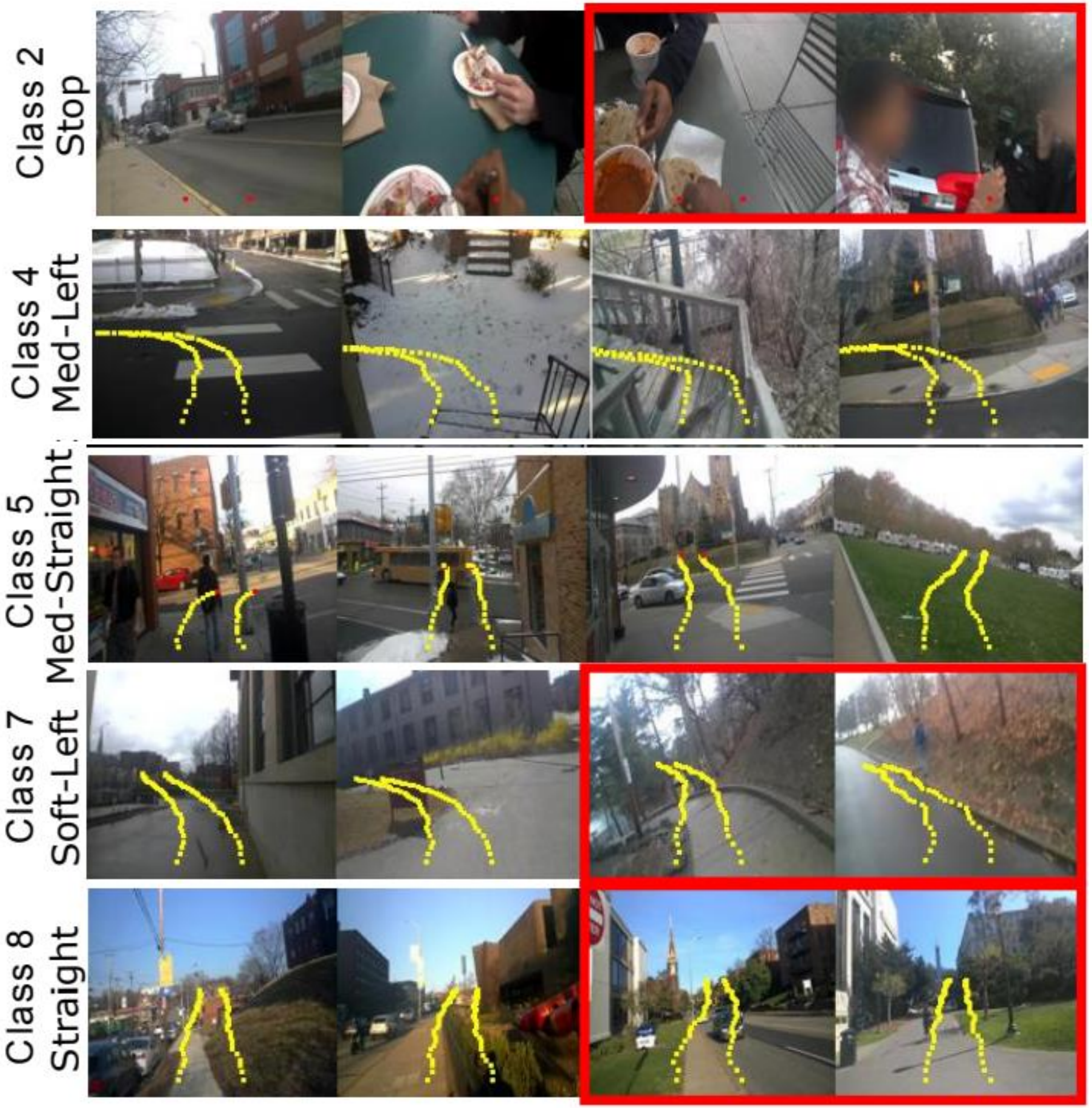


Image credits: Krishna Kumar Singh et al.

Predicting Trajectories

- Future trajectory as average of the frame trajectories of top-10 nearest neighbors separated by 10 minutes.
- Training: First 38 hours of recording (681,565 frames after temporal subsampling)
- Testing: 40,000 test frames (20,000 unvisited, 20,000 visited) randomly chosen from 38 and 52 hours.

RESULTS

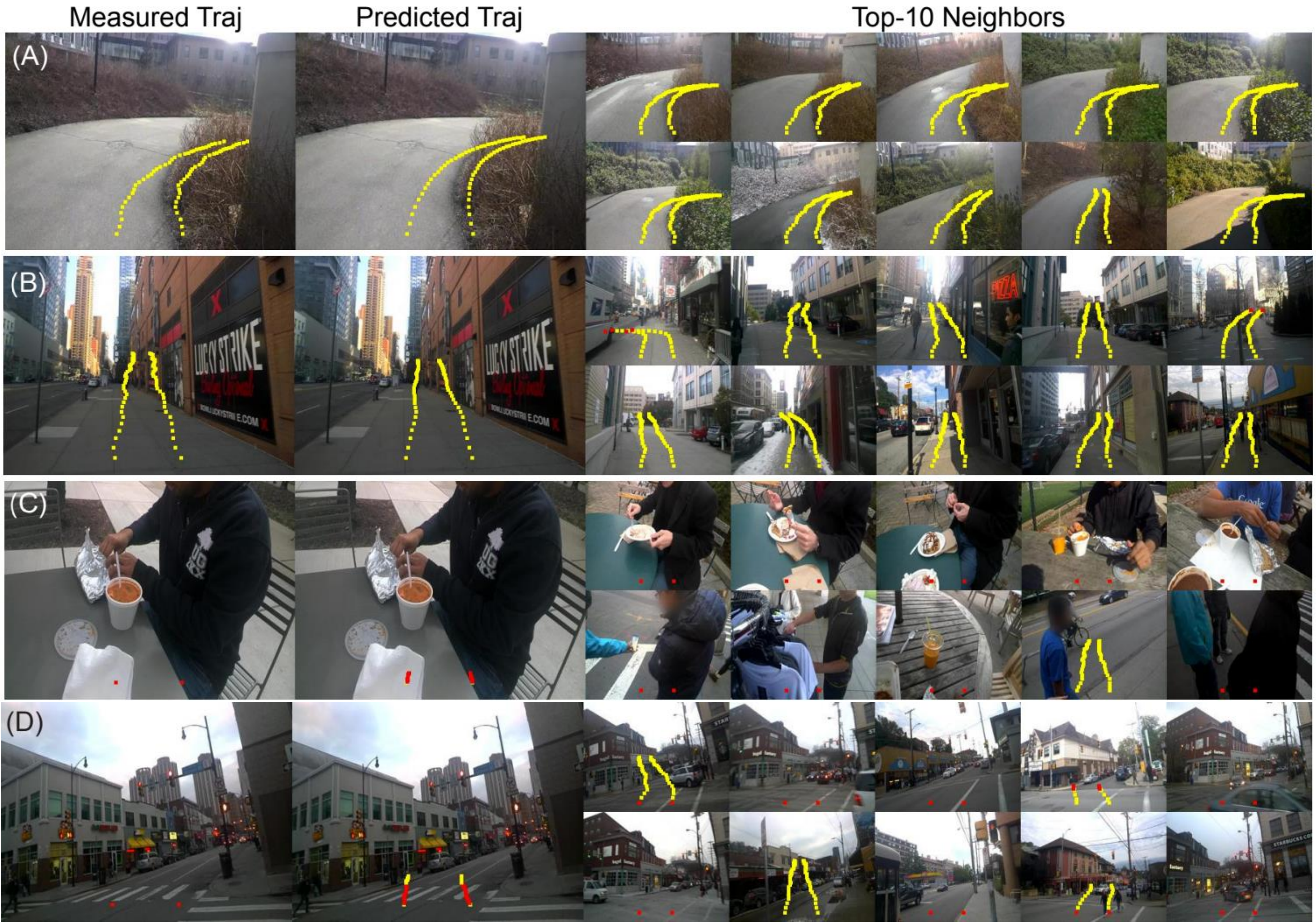
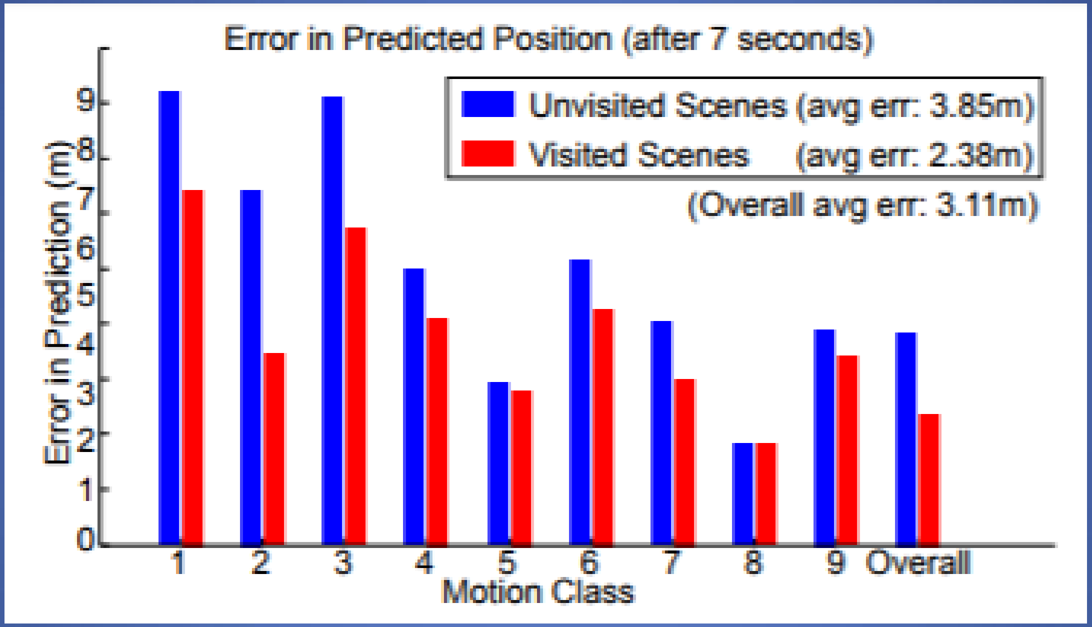


Image credits: Krishna Kumar Singh et al.

Unexpected prediction: (G) staring at road at angle is indicative of waiting at bus stop. (Note: nearest neighbors are from bus stops in a different part of town.)

Image credits: Krishna Kumar Singh et al.

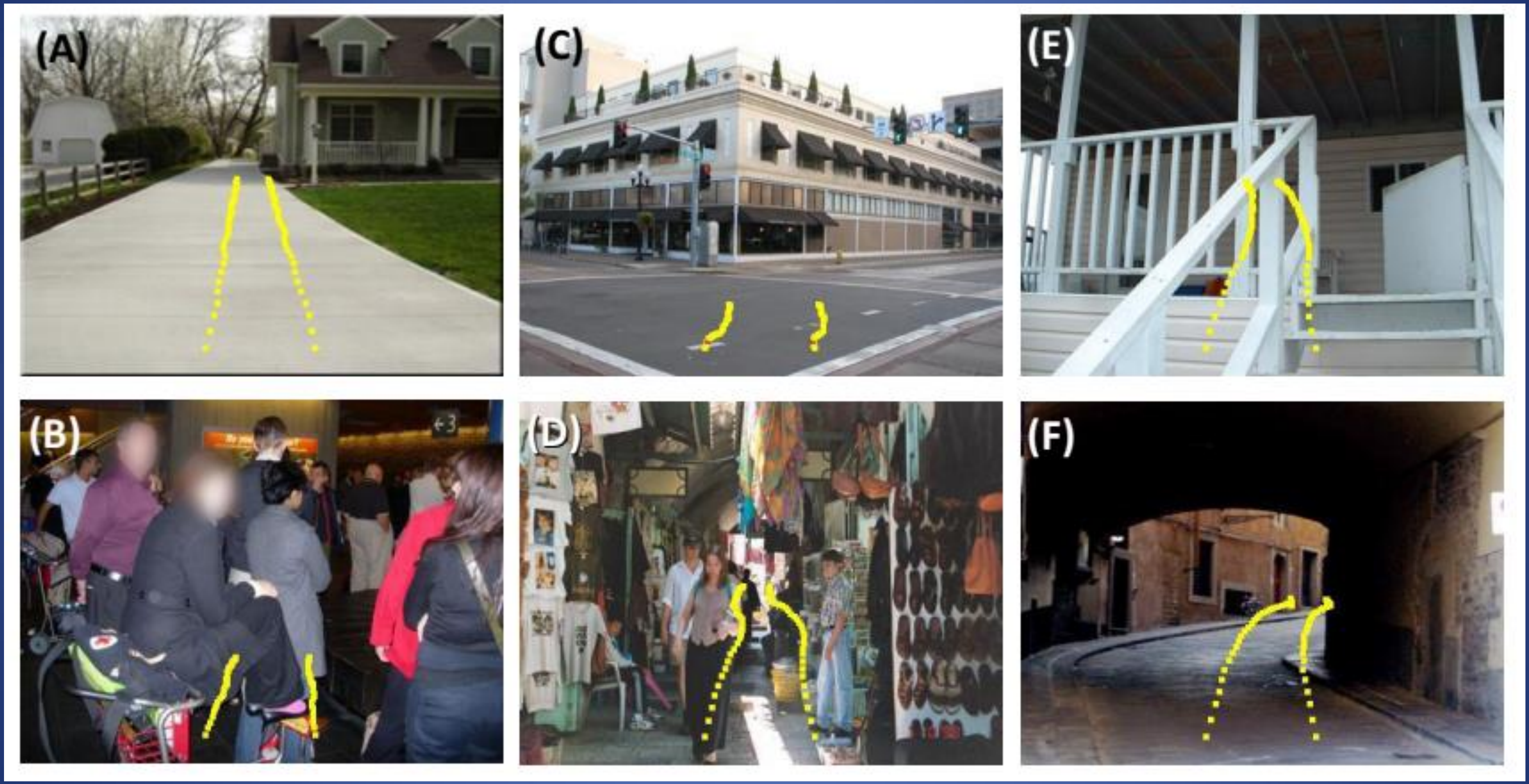
RESULTS



Error measure: Distance (in meters) between the predicted position and the measured position seven seconds into the future.

Image Credit: Krishna Kumar Singh et al.

RESULTS



Results on the SUN database

Image Credit: Krishna Kumar Singh et al.

Value of longer recordings

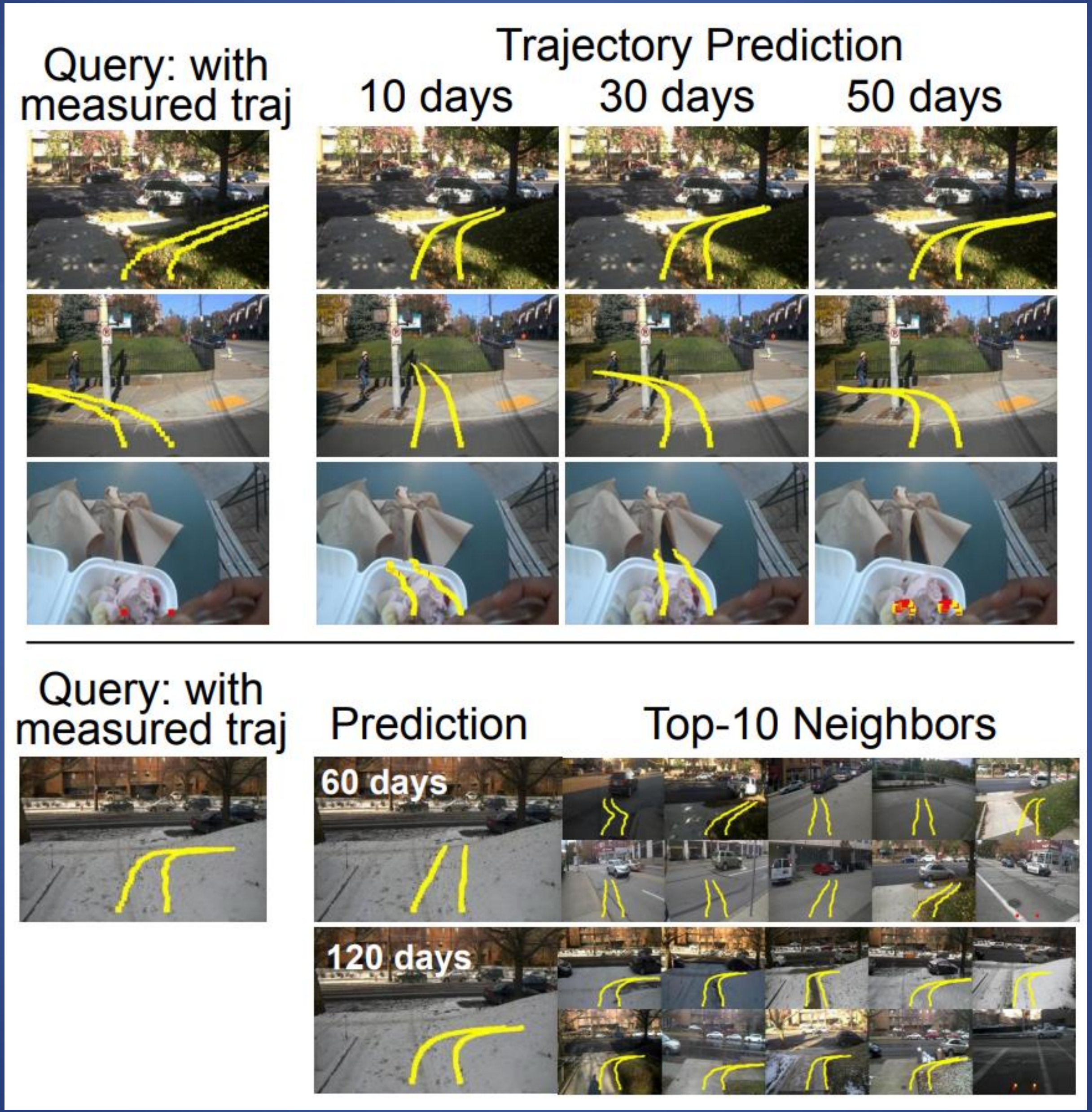


Image Credit: Krishna Kumar Singh et al.

APPLICATIONS OF THE DATASET

VIRTUAL WEBCAM

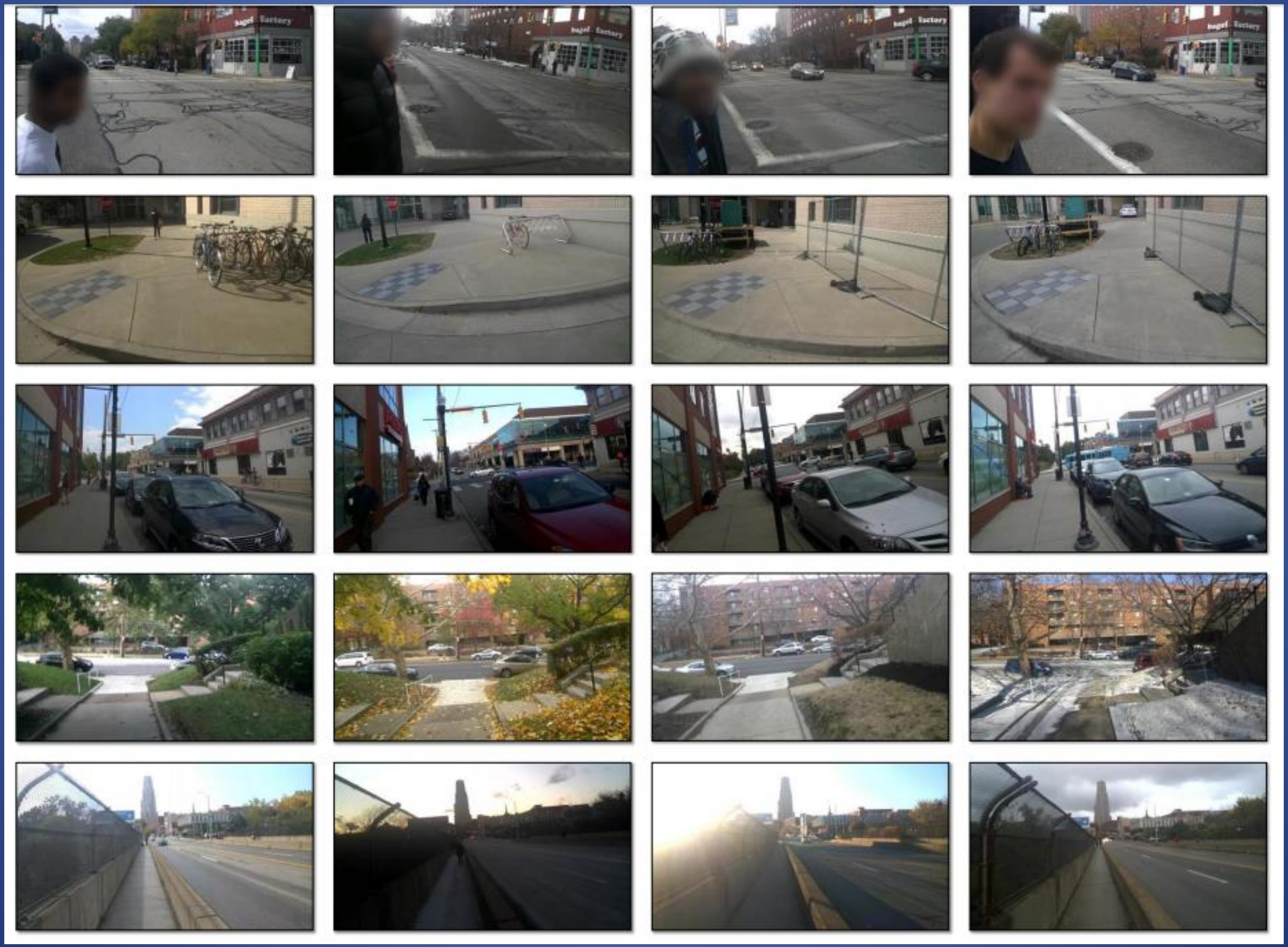


Image Credit: Krishna Kumar Singh et al.

APPLICATIONS OF THE DATASET

- Finding popular places: Correlate pedestrian detection with GPS location.

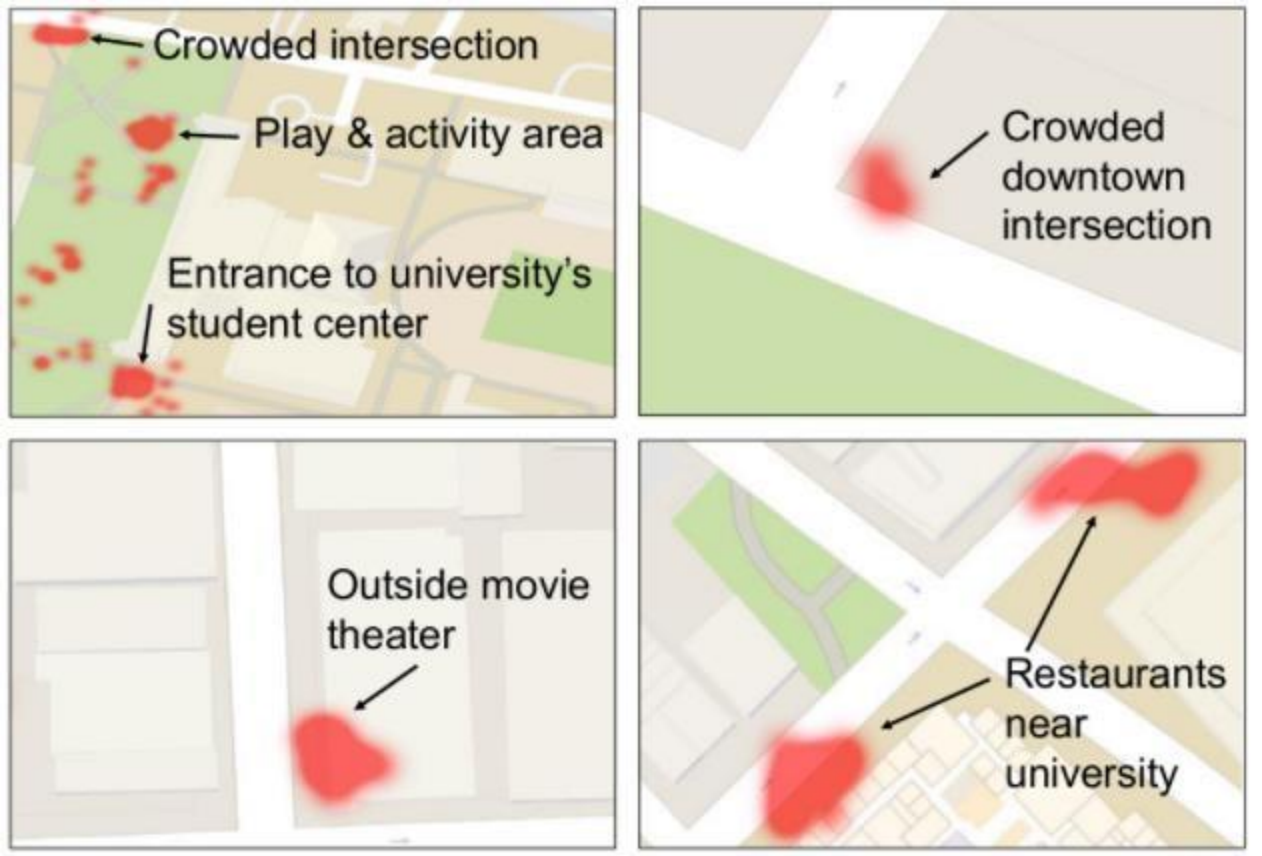


Image Credit: Krishna Kumar Singh et al.

STRENGTHS

- Creation of a huge egocentric dataset
- Using simple methods like NN
- New analyses that shed light on the nature of an individual's daily visual environment
- No manual annotations required

WEAKNESSES

- Single person only!
- Failure in trajectory prediction in fast movement.
- Low prediction accuracy in per- class motion prediction.
- No novel algorithms created

OPEN ISSUE: IS SUCH A DATASET USEFUL FOR MANY APPLICATIONS, AS IT IS EXTREMELY BIASED TO THE LIFE OF A PARTICULAR INDIVIDUAL?

POSSIBLE EXTENSIONS/FUTURE WORK

Motion prediction based on recent video history.

Using advanced techniques to enhance accuracy.

Application of dataset: giving good trajectory predictions to intoxicated individuals.

Analyzing motion of other individuals.

SUMMARY

Collected a large-scale, motion annotated, egocentric video stream

Solve scene understanding tasks

Opinion: Great dataset, huge scope for improvement in algorithms