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

Krishna Kumar Singh, Kayvon Fatahalian, Alexei Efros

Objective

Organize a large egocentric video collection (of real-world data from a single individual) into a richly annotated database that facilitates rapid analysis and enables exploration of new visual data understanding applications.

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

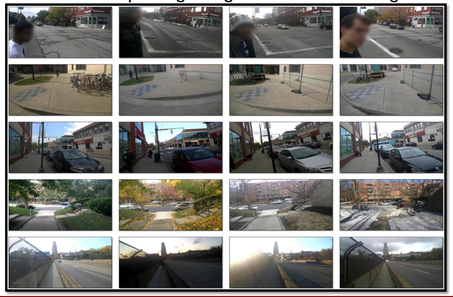
Total clips: 460

Device: Google Glass

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

Virtual Webcams

Webcams Capturing Long-term Scene Changes



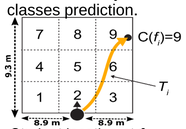
- Although the egocentric camera is not stationary, long-term recording captures changes in a scene over time.
- Each row in figure captures (from top to bottom) changes in companion (person), movement of bicycle stand, parked cars, season and lighting

Trajectory Estimation

- Motion trajectory is estimated by accelerometer and orientation sensor in the smart phone.
- Trajectory represents motion in next 7 seconds.
- Yellow color indicates motion, red means stop.
- GPS was not accurate for short term trajectories.

Motion Class Prediction

- MIT Places Hybrid-CNN is fine-tuned for 9 motion classes prediction.



Student location at f_i (movement to cell 9)

2.3%	68%	2.1%
0.9%	3.1%	0.9%
0.4%	21%	0.5%

Distribution of movements in training set

Most Confident Predictions (Red Box: Novel Places)

Motion Class Prediction Accuracy

	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

Trajectory Prediction

- Trajectory is predicted by averaging trajectory of top 10 nearest neighbors in deep feature space (Pool-5 MIT Places Hybrid-CNN).

Predict Turn

Predict straight at novel place

Predict stop while eating

Predict Stop at intersection

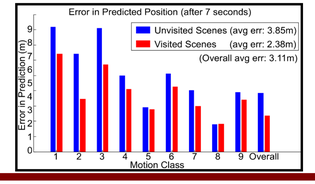
Predict stop in front of car at novel place

Predict turn

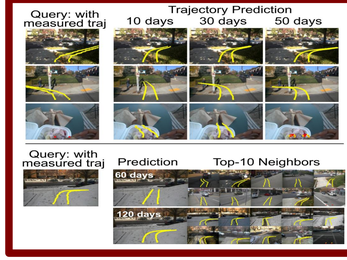
Predict stop at novel bus stop

Trajectory Prediction on SUN Dataset

Prediction Error



Value of Extended Recording

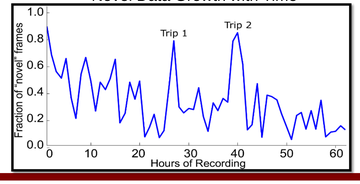


- Longer recording is needed to adequately sample rare events:
  - 10 days for common turn.
  - 30 days for less common turn.
  - 50 days for rare eating event.
- Not until four months of recording had occurred did snowing days begin to appear in the dataset, making prediction robust to seasonal change.
- Using only 50% and 25% of the training data, decreases motion class prediction accuracy relatively by 29% and 51% respectively.

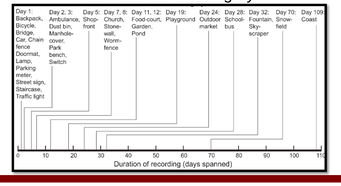
Novel Data Growth

- Quantify amount of novel visual data observed by the camera each day.
- Novel visual frames and semantic classes decreases with time.

Novel Data Growth with Time

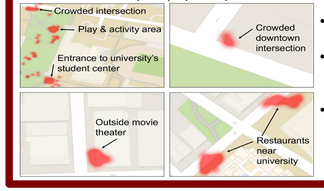


First Instance of Category



Popular Places

Heatmap of popular places



- 17% of the dataset contains at least one human.
- Popular places are found correlating pedestrian detections with GPS measurements.
- Red regions indicate locations where (on average) more than four people are present in images.