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

Krishna Kumar Singh, Prof. Kayvon Fatahalian, Prof. Alexei A. Efros

Carnegie Mellon University
University of California, Berkeley
University of California, Davis
Opportunity: ubiquitous visual sensing will soon provide the opportunity to record a large fraction of life’s events
Time Span: 9 months

Duration: 70 hours

Total Clips: 460 (10-20 minutes)

Locations: Various neighborhoods in Pittsburgh

Device: Google Glass

Data: 720 p, 30 fps
Accelerometer, Gyroscope, Orientation, GPS
Dataset contains diverse life experiences

Walking in different neighborhoods

Visiting parks

Talking to people

Shopping and eating outdoor

Different time of day

Different seasons
Prediction Task:
Where Will Krishna Move Next?
Motion trajectories

• For a given frame, trajectory represents motion in the next 7 seconds.

• Trajectories are generated through sensor data and used as ground-truth.

• Yellow lines indicate movement.

• Red dots indicate stopping.

• Our goal is to predict these trajectories.
Prediction Problem

Prediction Algorithm

[Diagram showing the process of prediction problem and prediction algorithm with images of pedestrians and street scenes]
Trajectory prediction using nearest neighbors

• Goal: Predict future trajectory for a given frame.

Ground Truth
Trajectory

Top 10 nearest neighbors using fifth layer deep features and cosine similarity as distance kernel

Predicted trajectory is average of trajectories of top 10 nearest neighbors
Predicting common human behaviors

People walk on sidewalk
Predicting common human behaviors

People remain stationary while eating

People stop soon after approaching a traffic button
Prediction of Krishna specific behaviors

Turning left at particular intersection

Turning right outside my house
Predicting behavior due to transient objects

Predicting stop, if car is in front
Is big data actually necessary for this task? (how much data do you really need?)
Prediction of rare events require more training data
Prediction of rare events require more training data

Ground Truth

Prediction using 2 months of training data

Top 10 Nearest Neighbors

Prediction using 4 months of training data

Top 10 Nearest Neighbors
Prediction failure cases: bifurcations

Junction where both left and right turn possible

Stopping at an intersection, waiting at a traffic light, or continuing to walk
Density Estimate

Left-Right Bifurcation

Stop-Go Bifurcation

Density estimate of where camera wearer will be in 7 sec
Amount of novel data decreases with time
Virtual Webcam (capture changes)

- Person
- Bicycle stand
- Car
- Season
- Time of day
Crowded locations in dataset

- Crowded Intersection
- Play and Activity Area
- Outside University Center
- Crowded Intersection at Downtown
- Outside Movie Theater
- Restaurants Near University
Summary

• Simple nearest neighbor predictions are effective due to heavy redundancy in dataset.
• Simple nearest neighbor is able to generalize for the novel places for which we have seen diverse set of examples.
• Prediction of rare events require long term training data.
Thank You