

## Temporal Segmentation of Egocentric Videos

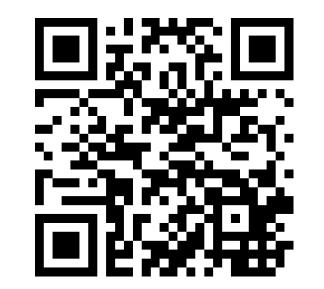
Yair Poleg<sup>1</sup>

Chetan Arora<sup>1</sup> <sup>2</sup>

Shmuel Peleg<sup>1</sup>

<sup>1</sup>The Hebrew University of Jerusalem

<sup>2</sup> IIIT Delhi

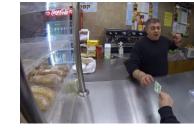


## Egocentric Videos

Egocentric Video = First person point of view

















Used by:











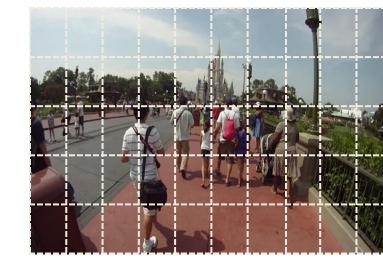
## Life Logging Egocentric Videos

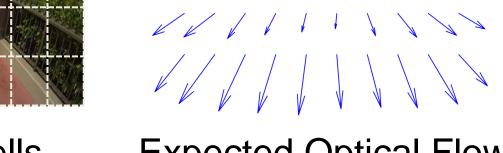
- Always-on, infinite, unstructured video
- Video in the wild. Head is always moving, and head motion is always dominant.
- Very long and boring

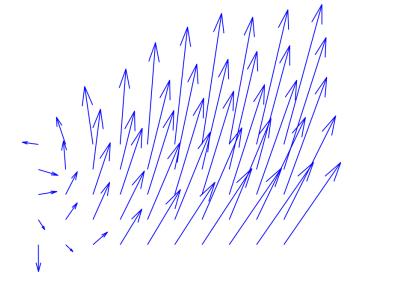
## Objective: Divide a long egocentric video into chapters Sitting Riding Driving Standing Input Video Vearer is at one place. May have natural head motion or even taking a few steps if standin hanging location inside a vehicle with partial view of surroundings Stationary Faster motion compared to walking. Less head motion. E.g. bicycle. Forward motion with sideways view. Riding in the train should also be considered this class.

# Our Approach Demo Video

(i) Estimate optical flow at fixed cells in all frames. No feature tracking.





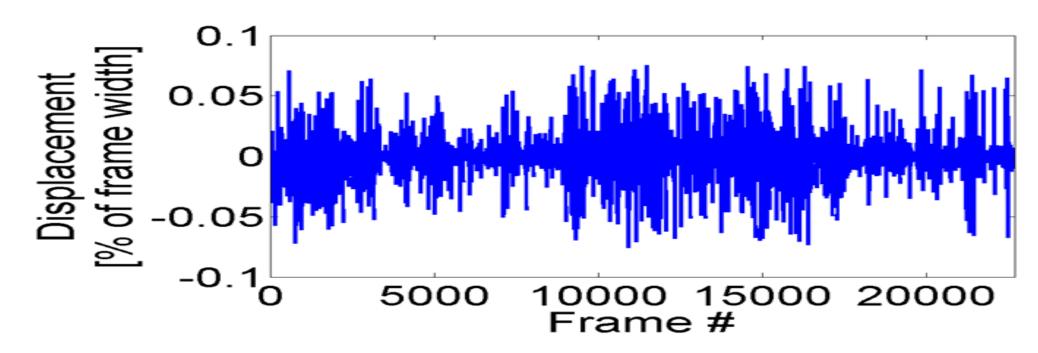


Grid of 10x5 cells

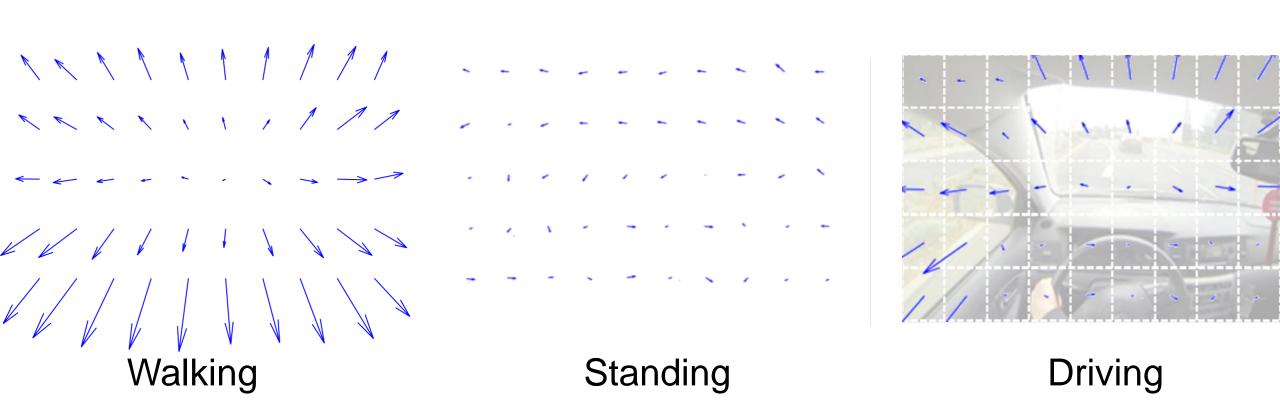
Expected Optical Flow Measured Optical Flow

← ← ← ← · · · · · → →

(ii) Optical flow is dominated by head motion.

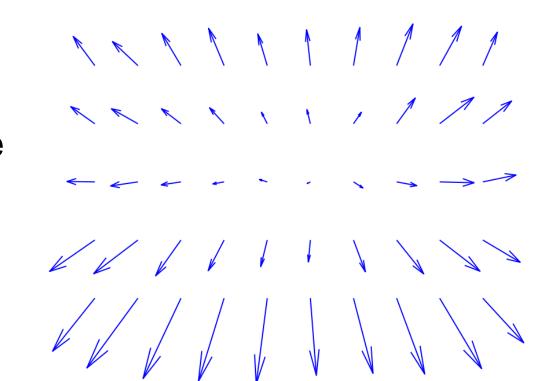


(iii) Remove effects of head motion by temporal smoothing of optical flow



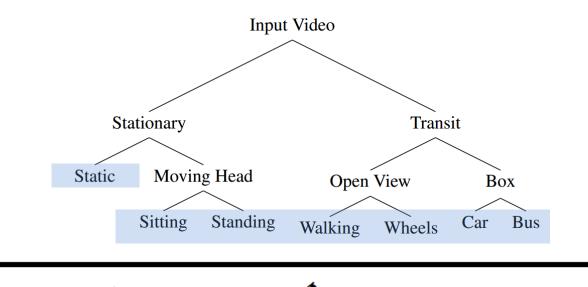
#### Feature Vector

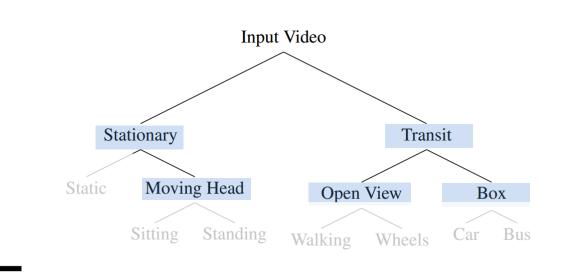
- Radial Projection Response
  - Find FOE, compute expansion template
- Count motion vectors satisfying template
- Motion vectors magnitude clusters
- Statistical information:
- Number of blocks with valid optical flow
- Average & variance of flow magnitudes



### Results

- Dataset:
  - Training set:140 sequences, 65hrs, ~3.5M frames
- All labeled into 7 classes by students
- Training:
  - Randomly pick a sequences until we cover 12K samples per class
- Training sequences are excluded from the test
- Test Set: Unseen sequences





|          | Walking     | á   | Standing          | <b>1</b> 5  | Wheels      | Sitting     | Static      |
|----------|-------------|-----|-------------------|-------------|-------------|-------------|-------------|
|          | 40          | Car | Sta               | Bus         | 4,          | Silv        | Skir        |
| Walking  | <b>83</b> % | 0%  | 6%                | 6%          | 4%          | 1%          | 0%          |
| Car      | 1%          | 74% | 3%                | 15%         | 0%          | 3%          | 4%          |
| Standing | 14%         | 2%  | $oldsymbol{47}\%$ | 4%          | 0%          | 31%         | 2%          |
| Bus      | 3%          | 19% | 27%               | <b>43</b> % | 0%          | 7%          | 1%          |
| Wheels   | 9%          | 0%  | 0%                | 6%          | <b>86</b> % | 0%          | 0%          |
| Sitting  | 3%          | 1%  | 22%               | 1%          | 0%          | <b>62</b> % | 10%         |
| Static   | 0%          | 1%  | 1%                | 0%          | 0%          | 1%          | <b>97</b> % |

| Class Label      | Accuracy | # Samples |
|------------------|----------|-----------|
| Static-Moving    | 91%      | 1083115   |
| Sitting-Standing | 82%      | 1036217   |
| Box-Open         | 87%      | 1197623   |
| Car-Bus          | 76%      | 228108    |
| Walking-Wheels   | 82%      | 969515    |