Team: UCSB_UCR_VCG
TRECVID 2012: Surveillance Event Detection

* University of California, Riverside, CA-92521
**University of California, Santa Barbara, CA-93106

Introduction

- Seven Activities: (1) CellToEar, (2) Embrace, (3) ObjectPut, (4) PeopleMeet, (5) PeopleSplitUp, (6) PersonRuns, and (7) Pointing.
- Challenges: background noise, clutter, difference of viewpoints, large crowd, illumination variation, occlusion, etc.

Approaches

- **Spatio-temporal cuboid based approach:** activities like CellToEar, Embrace, ObjectPut, and Pointing are the results of articulated motion of human parts. For these activities, we exploit spatio-temporal sliding cuboid based approach.
- **Track based approach:** In the activities like PeopleMeet, PeopleSplitUp and PersonRuns, the characteristics of trajectories of the persons of interest in the activities are discriminative. For these activities, we exploit track based approach.

Spatio-temporal Cuboid Based Approach: Feature extraction

- Event video clips are segmented from the video corpus and spatial extent of the activity regions are drawn.
- STIP features are generated and collected those, belong to the activity regions.
- STIP features are clustered into visual words using k-mean (400) algorithm.
- Video clips are represented using histograms of visual words.
- Discriminative classifiers are trained for each camera-activity pair using SVM.

Spatio-temporal Cuboid Based Approach: Evaluation

- Activities tend to occur more in some parts of the video frame, which are distinct for different cameras and activities.
- We utilize this prior information from the training videos in the evaluation phase in order to reduce the number of false alarms.

Experiments and Results

- We keep five frame temporal and twenty pixel spatial distance between two overlapping cuboids.
- For PeopleMeet and PeopleSplitUp, the current system uses training instances from VIRAT Dataset release 1.
- Tracks with 5% highest velocity are classified as PersonRuns.