

Group Context Learning for Event Recognition

Yimeng Zhang¹, Weina Ge², Ming-Ching Chang², Xiaoming Liu²

¹Cornell University, ²GE Global Research

Presenter:



imagination at work

Event Recognition

Goal: given a video, detect whether an event of interest occurred

Previous works: mainly focused on activities of a single person

Ours: *group-level* events

- Interactions involving a group of people
- Eg. Group forming, group dispersing, group fighting, group flanking, group chasing, etc.

Group-Level Event Recognition



Group Forming



Group Following



Group Forming
Fighting



Related Work

- Action recognition of a single person



[Schuldt. Et al, 04]

- Abnormal movements of a crowd



[Wu et al. 10]

- Events involving a group of people
 - Previous works: logic or rule based methods
Brendel et al. 11, Change et al. 11, Morariu et al. 11
 - Ours: a learning based framework

Approach Overview

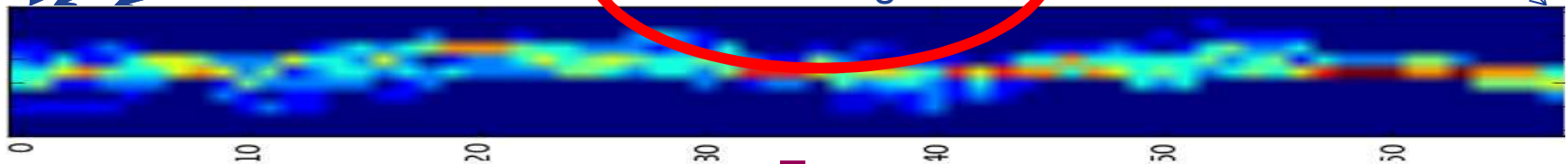
Video



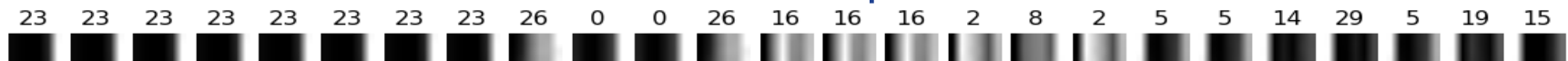
Video Frames



Feature Histogram



Word Sequence



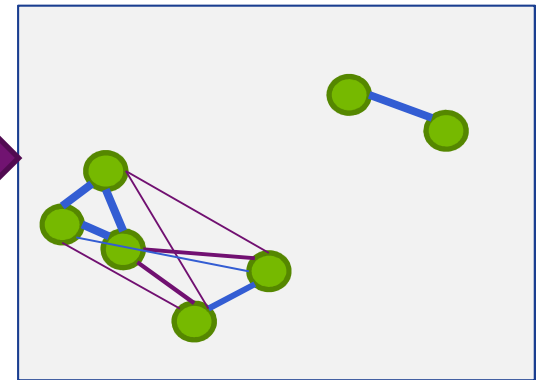
SVM+BOW

Feature Extraction Overview

People Tracking



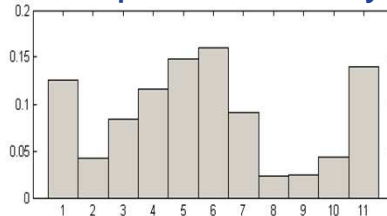
Group Analysis



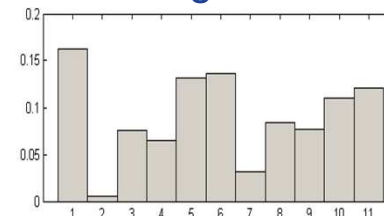
Feature Extraction



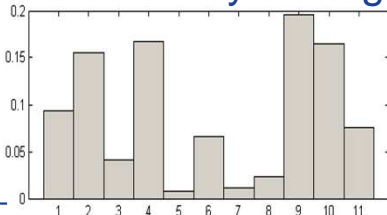
Group Connectivity



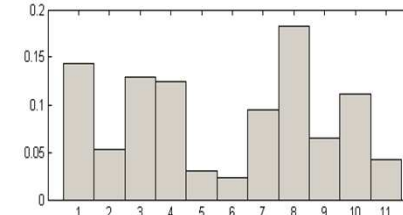
Moving Direction



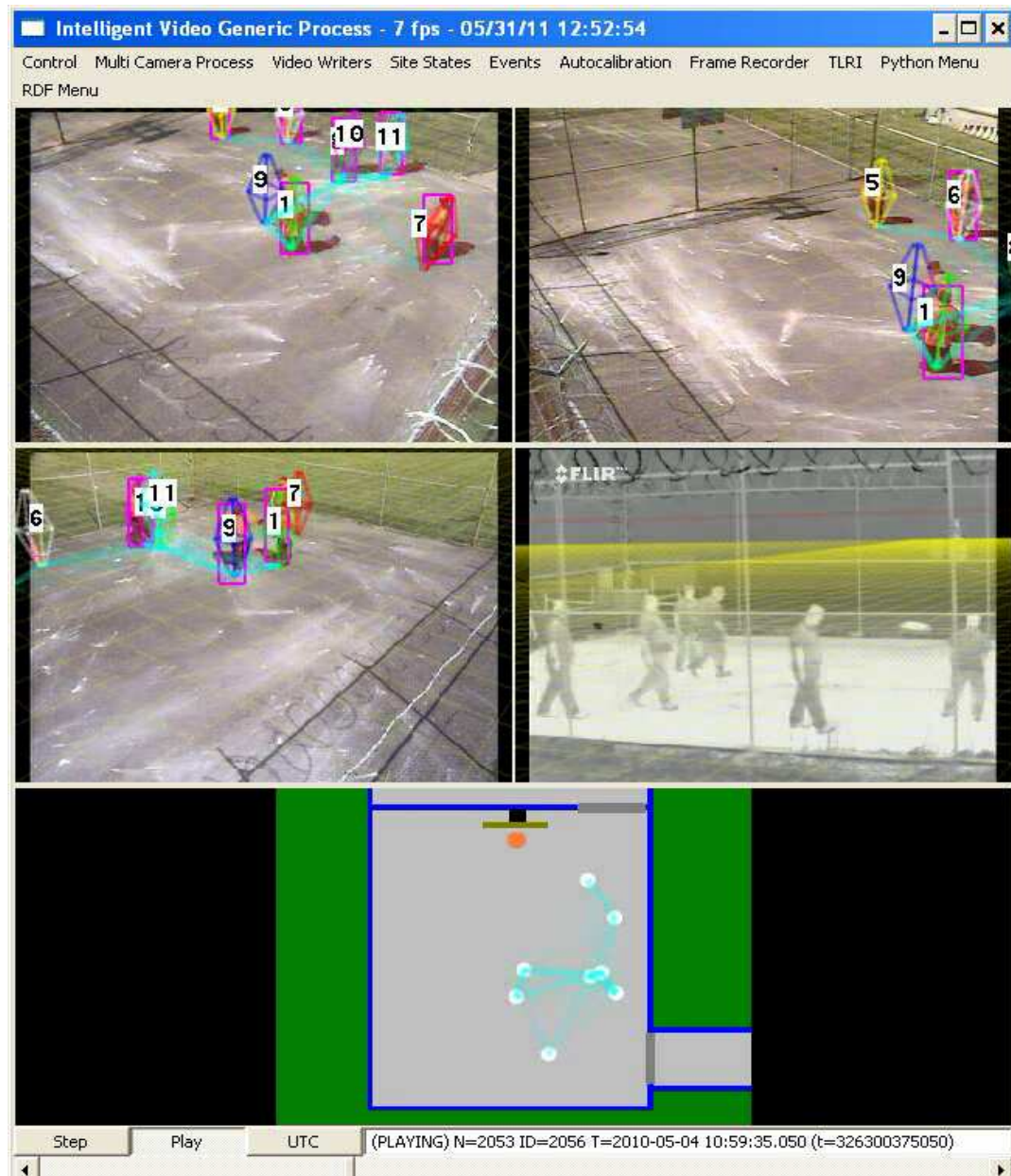
Connectivity Change



Moving Speed



Person Tracking System

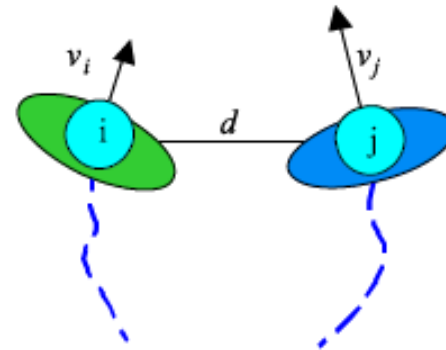
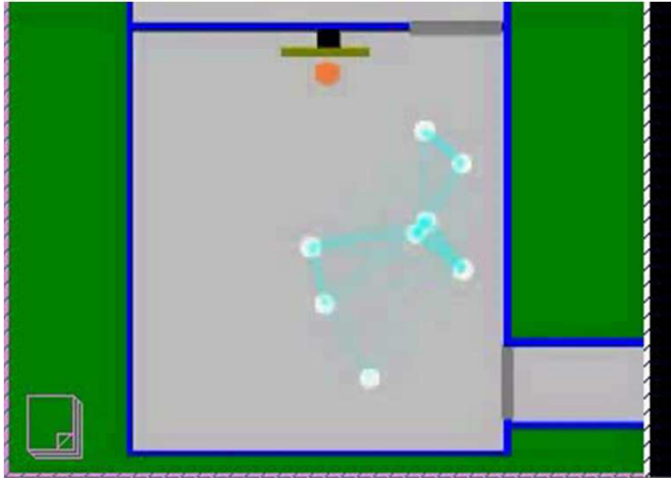


Group Analysis

Construct a group connectivity graph

Node: detected individuals

Edge: the probability of two individuals being in the same group



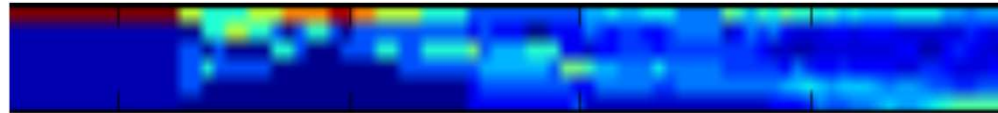
Group connectivity:
distance, motion, trajectory

[Chang et al. ICCV 11]

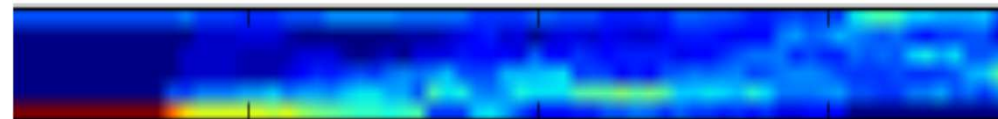
Group Connectivity Feature

Histogram of the edge weights of the group connectivity graph

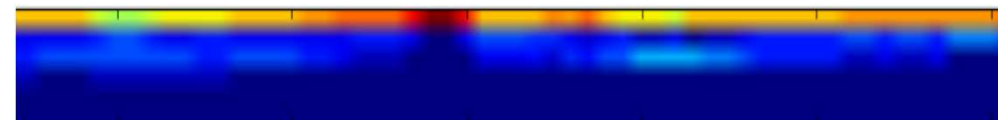
Group Dispersing



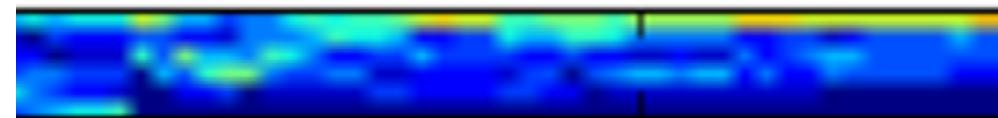
Group Forming



Group Following



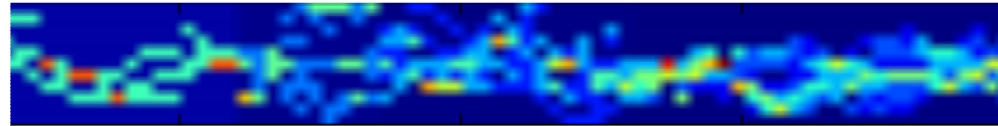
Group Fighting



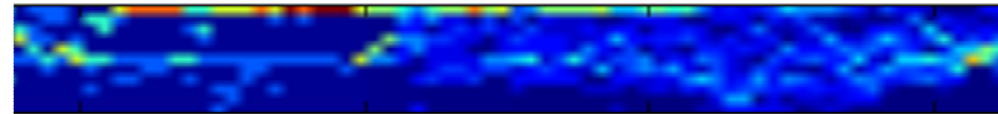
Connectivity Change Feature

Histogram of the connectivity changes of the same individual pairs between the current frame and the t^{th} previous frame

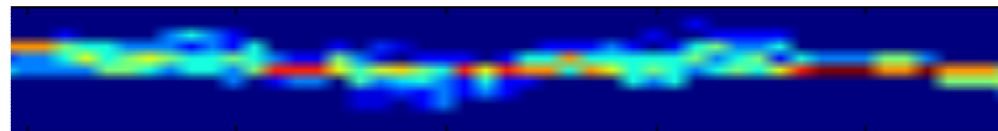
Group Dispersing



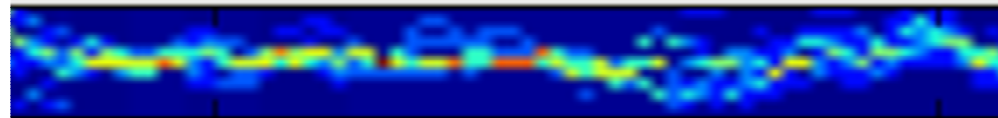
Group Forming



Group Following



Group Fighting

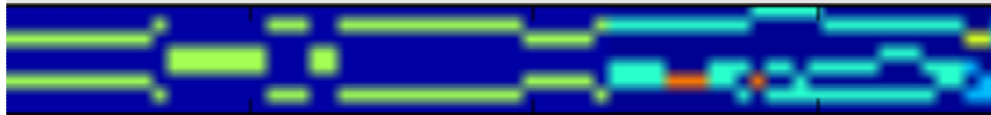


Motion Direction Feature

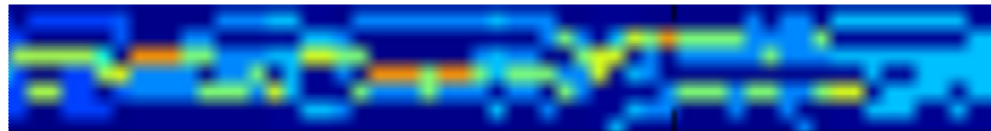
Histogram of the motion directions of people

Directions normalized by subtracting the mean of all directions

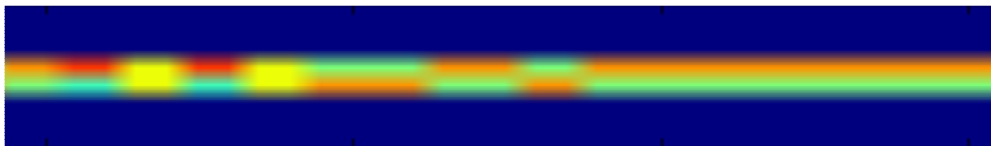
Group Dispersing



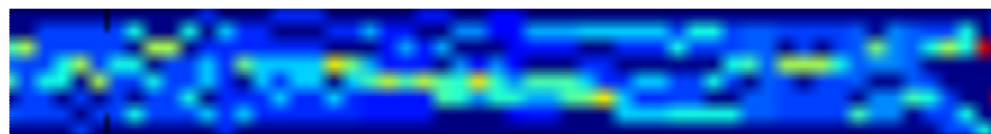
Group Forming



Group Following



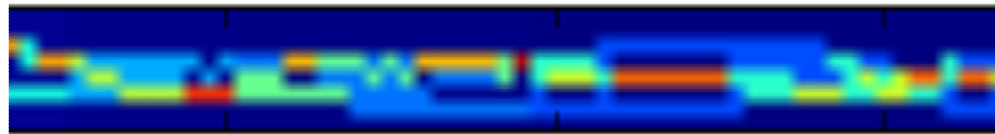
Group Fighting



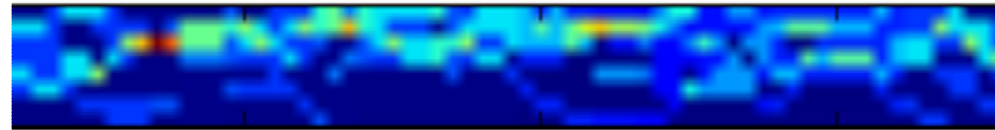
Motion Speed Feature

Histogram of the motion speed of people

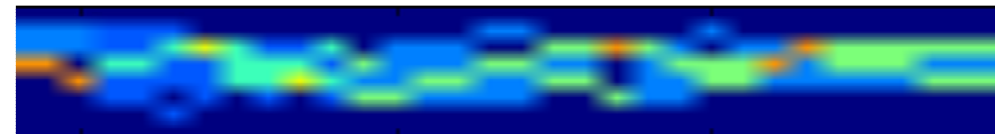
Group Dispersing



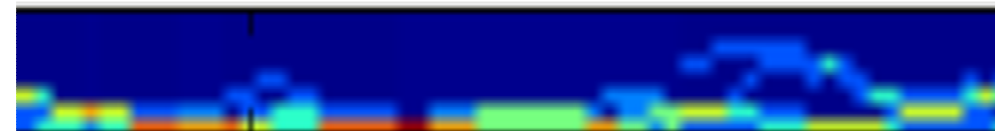
Group Forming



Group Following



Group Fighting

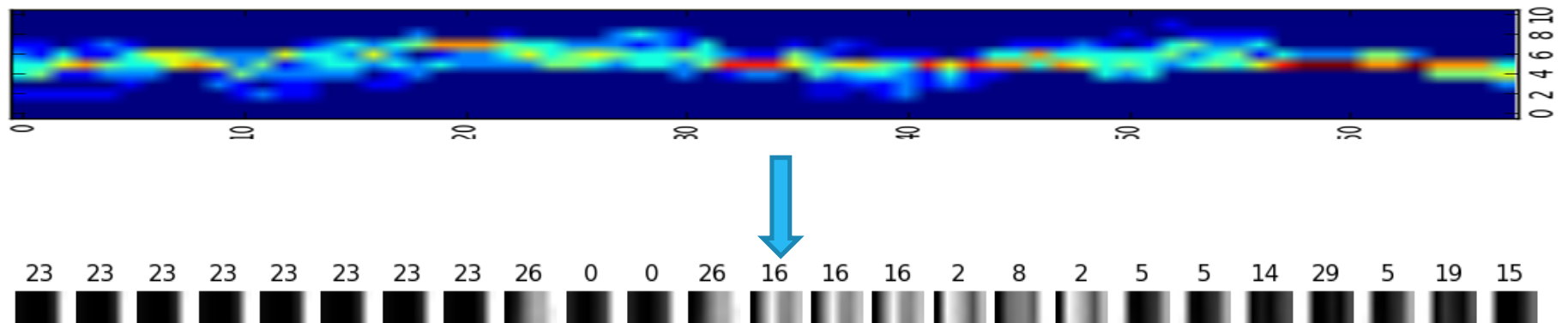


Group Context Words

The histograms of consecutive 4 frames are clustered.

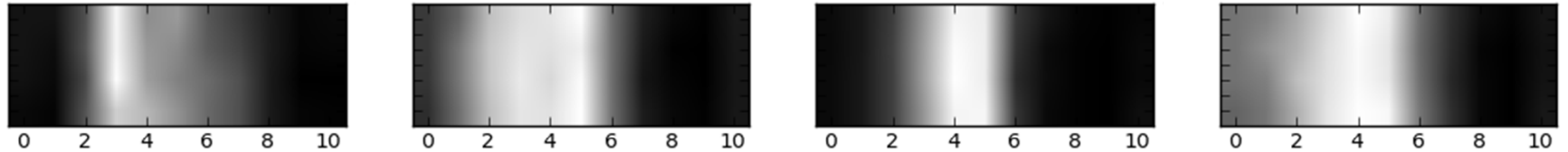
Create words for each feature type

Concatenate BOW of each feature type as input to the SVM

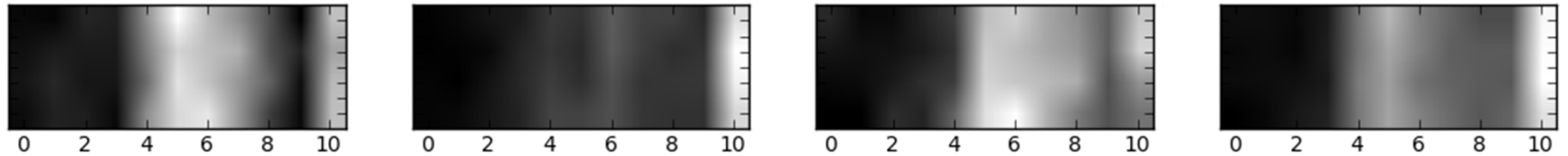


Connectivity Change Words

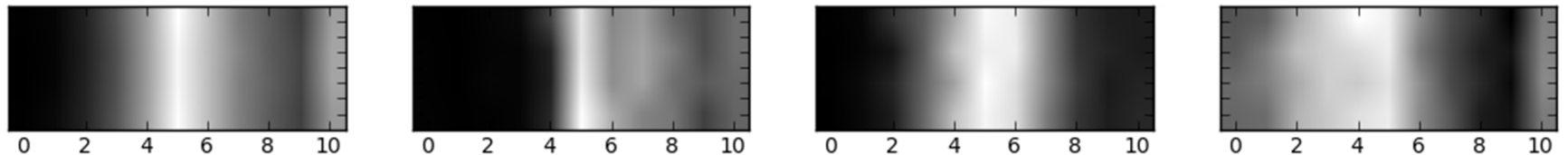
Group Dispersing



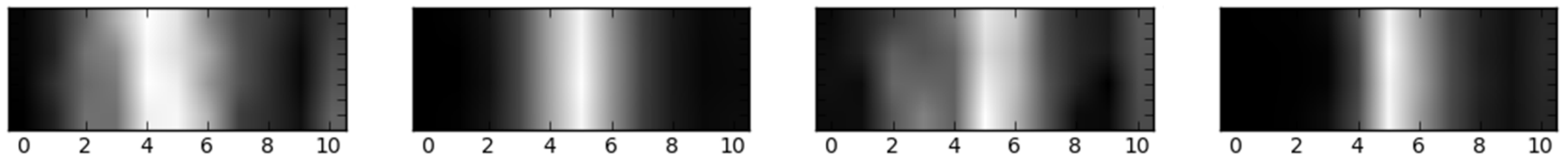
Group Forming



Group Fighting



Group Following



Experiments

Mock Prison Riot(MPR) Dataset

19 surveillance video sequences of 3 to 6 minute length taken in a prison

6 events: group forming, group dispersing, group following, group chasing, group flanking, and group fighting



...



...



Event Recognition

First experiment: classify the segmented video clips into one of the event categories

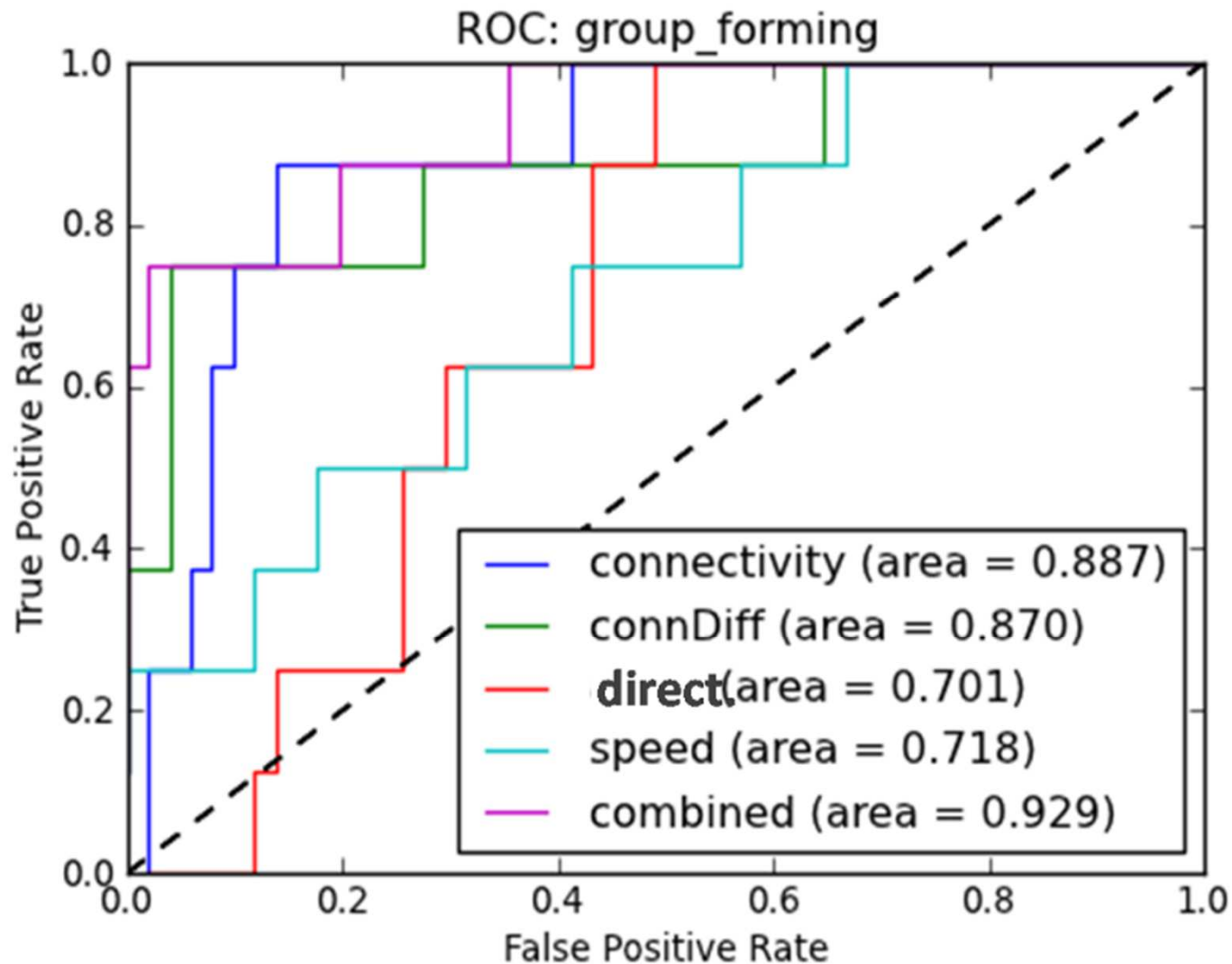
177 video clips of 2 to 30 seconds

60% for training, 40 % for testing

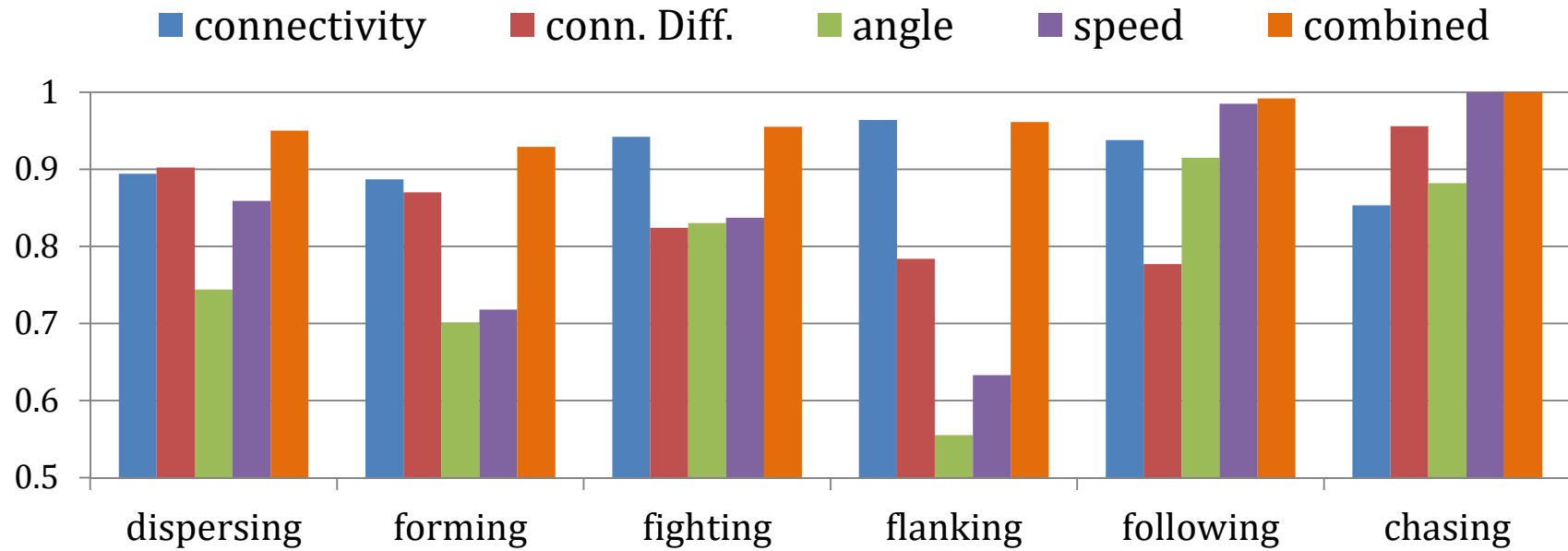


Group dispersing

ROC Curve for Group Forming



AUC scores



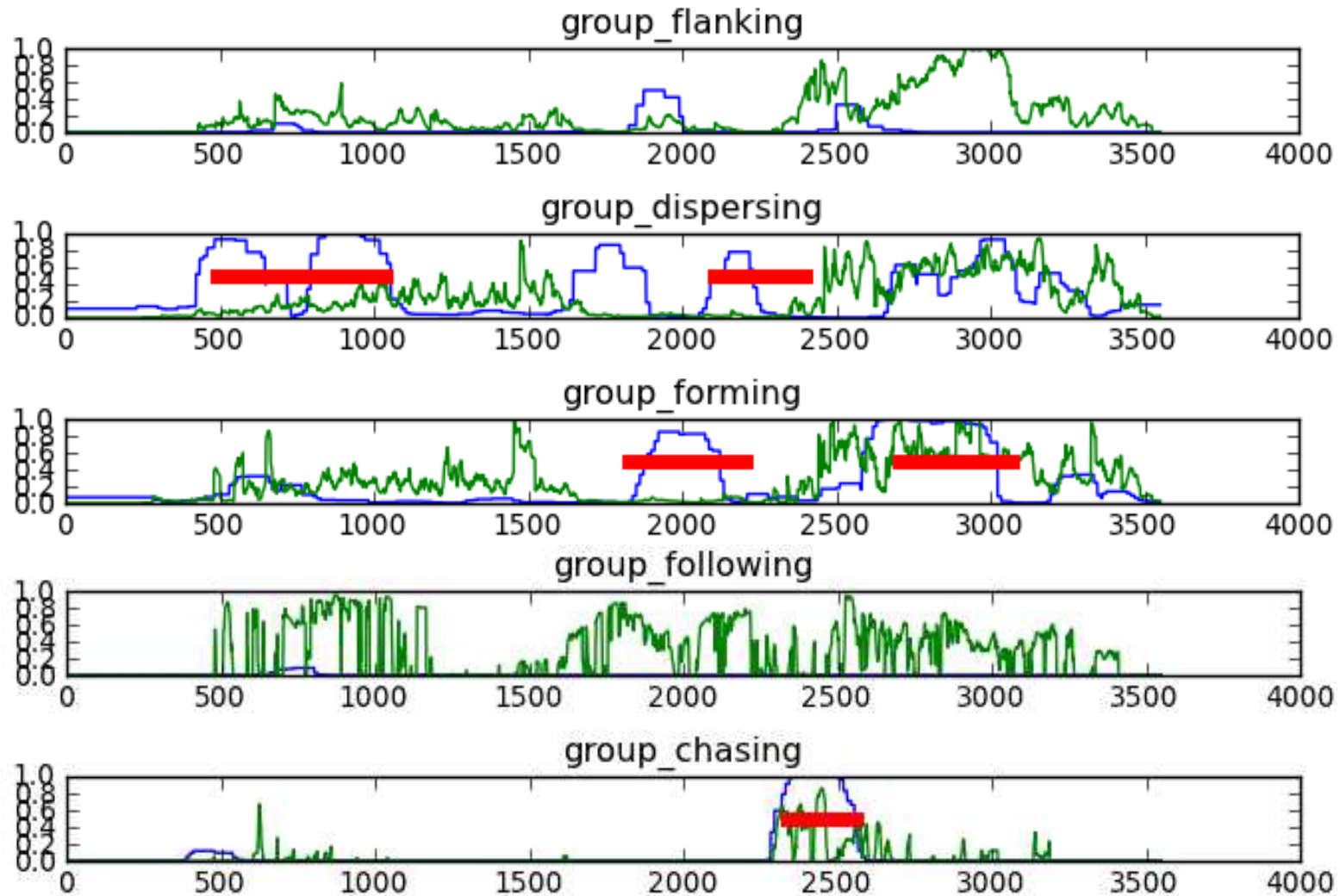
Event Detection

Second experiment: online event detection from continuous videos

Make prediction for each frame using the previous 4 seconds

60% of the 19 videos for training, 40% for testing

Predicted probabilities



Ground truth Learning Rule [Chang, et al. 11]

AUC Score Comparison

	Dispersing	Forming	Flanking	Following	Chasing	Fighting
rule-based	0.592	0.658	0.921	0.667	0.981	N/A
ours	0.926	0.811	0.959	0.827	1.000	0.834

Running Time: 0.02s per frame including person detection and tracking

Conclusion

A novel learning based framework for group-level event recognition

Robust features that capture the group context of individuals in a video

A system that detects events in real time

