Human Activity Recognition using Deep Neural Network with Contextual Information

University of Houston, Computer Science Department, Quantitative Imaging Laboratory
Li Wei and Shishir K. Shah

Introduction

Recognize the human activity in the video.

The problem is challenge because:

- Tremendous intra-class variation and inter-class similarity of human activities
- The visual appearance differences, subject motion variabilities, and viewpoint changes.

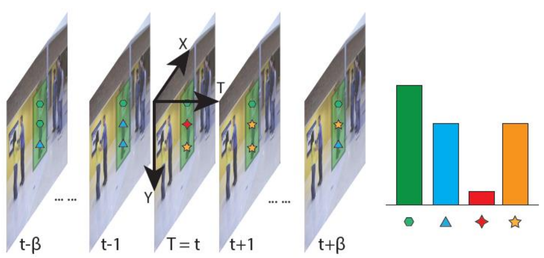
Motivation and Contribution

- Motion features based approaches not able to capture the interaction between persons.

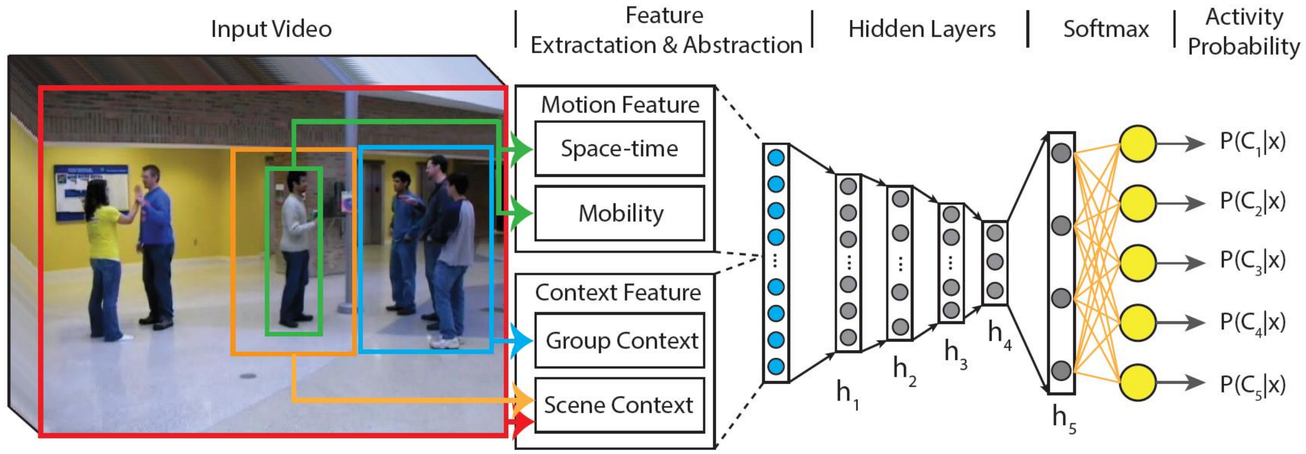
- Introduce the scene context features that describe the environment of the subject at global and local levels.
- Propose group context feature that encodes the group interaction and group structures.
- A deep neural network model to recognize the human activity

Motion Feature

- **Space-time features**: Space-time interest points (STIP)
- **Mobility feature**: The velocity of person in 3D space



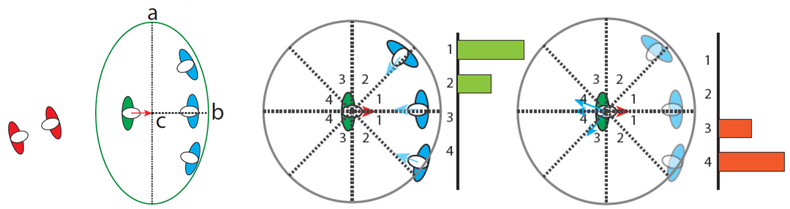
Algorithm Overview and Network Structure



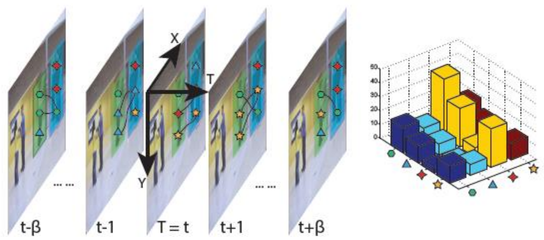
Context Features

Group Based Context

- **Structure**: relative positions and directions of people within interaction regions.

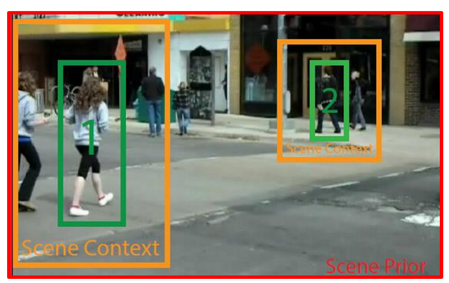


- **Interaction**: motion interactions between group members.



Scene Based Context – Recognize using Place-CNN [2]

- **Global Scene**: the global scene attributes of the video
- **Local Scene**: the scene around the person locally.



Global Scene
crosswalk: 0.54 gas_station: 0.30

Local Scene #1
crosswalk: 0.70 parking_lot: 0.07

Local Scene #2
Phone_booth: 0.54 lobby: 0.30

Results and Comparison

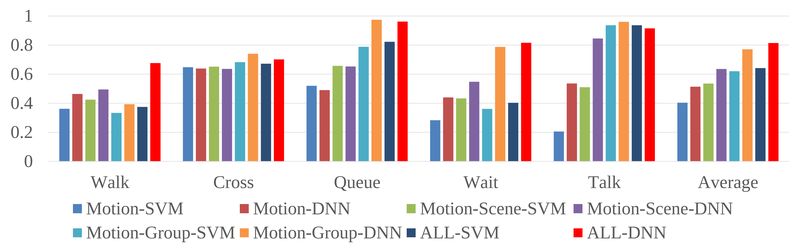
Evaluate our approach in Collective Activity Dataset [2].

- Activities: Crossing, Waiting, Queueing, Walking, Talking
- 70 videos, 2 mins average length. 4-folds cross validation

Confusion matrix of recognition results

	Cross	Wait	Queue	Walk	Talk
Cross	0.70	0.10	0.00	0.20	0.00
Wait	0.09	0.82	0.00	0.08	0.01
Queue	0.00	0.00	0.96	0.03	0.01
Walk	0.19	0.03	0.07	0.68	0.04
Talk	0.02	0.01	0.03	0.03	0.91

Compare with Appearance Based Approach and SVM Model



References

[1] Li Wei, Shishir. K. Shah, **Human Activity Recognition using Deep Neural Network with Contextual Information**, ECCV 2016 (In submission)

[2] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. **Learning deep features for scene recognition using places database.** In Advances in Neural Information Processing Systems

[3] W. Choi, K. Shahid, and S. Savarese. What are they doing? : **Collective activity classification using spatiotemporal relationship among people**. ICCV 2009 Workshop

PhDPosters.com
Your Research Poster, Printed For Less