In this dissertation, we address the problem of discovery and representation of group activity of humans and objects in a variety of scenarios, commonly encountered in vision applications. The overarching goal is to devise a discriminative representation of human motion in social settings that captures a wide variety of human activities observable in video sequences. Such motion emerges from the collective behavior of individuals and their interactions and is a significant source of information typically employed for applications such as event detection, behavior recognition, and activity recognition. We present new representations of human group motion for static cameras, and propose algorithms for their application to a variety of problems.

We first propose a novel method to model and learn the scene activity of a crowd using Social Force Model for the first time in the computer vision community. We present a method to densely estimate the interaction forces between people in a crowd, observed by a static camera. The patterns of activities of the objects in the scene are modeled in the form of volumes of interaction forces.

Second, we propose a method based on the Lagrangian framework for fluid dynamics, by introducing a streakline representation of flow. We propose a method to distinguish different group behaviors such as divergent/convergent motion and lanes using this framework. Finally, we introduce flow potentials as a discriminative feature to recognize crowd behaviors in a scene. Results of extensive experiments are presented for multiple real life crowd sequences involving pedestrian and vehicular traffic.

The proposed method exploits optical flow as the low level feature and performs integration and clustering to obtain coherent group motion patterns. However, we observe that in crowd video sequences, as well as in a variety of other vision applications, the co-occurrence and inter-relation of motion patterns are the main characteristics of group behaviors. In other words, the group behavior of objects is a mixture of individual actions or behaviors in specific geometrical layout and temporal order.

We, therefore, propose a new representation for group behaviors of humans using the inter-relation of motion patterns in a scene. The representation is based on bag of visual phrases of spatio-temporal visual words. We present a method to match the high-order spatial layout of visual words that preserve the geometry of the visual words under similarity transformations.

Major: Electrical Engineering

Educational Career:
Bachelor's of Electrical Engineering, BS, 2003, K.N.Toosi University of Technology, Tehran, Iran
Master's of Electrical Engineering, MS, 2006, K.N.Toosi University of Technology, Tehran, Iran

Committee in Charge:
Dr. Mubarak Shah, Chair, EECS
Dr. Gita R. Sukthankar, EECS
Dr. Marshall Tappen, EECS
Dr. Aman Behal, EECS
Dr. Brian E. Moore, MATH

Approved for distribution by Dr. Mubarak Shah, Committee Chair, on October 7, 2011.

The public is welcome to attend.