Research in human action recognition has advanced along multiple fronts in recent years to address various types of actions including simple, isolated actions in staged data (e.g., KTH dataset), complex actions (e.g., Hollywood dataset) and naturally occurring actions in surveillance videos (e.g., VIRAT dataset). Several techniques including those based on gradient, flow and interest-points have been developed for their recognition. Most perform very well in standard action recognition datasets, but fail to produce similar results in more complex, large-scale datasets. Action recognition on large categories of unconstrained videos taken from the web is a very challenging problem compared to datasets like KTH (six actions), IXMAS (thirteen actions), and Weizmann (ten actions). Challenges like camera motion, different viewpoints, huge inter-class variations, cluttered background, occlusions, bad illumination conditions, and poor quality of web videos cause the majority of the state-of-the-art action recognition approaches to fail. Also, an increased number of categories and the inclusion of actions with high confusion, add to the challenges.

In this thesis, we analyze the reasons for this less than successful generalization by considering a state-of-the-art technique based histogram of oriented gradients in spatio-temporal volumes as an example. This analysis may prove useful in developing robust and effective techniques for action recognition.

Secondly, we propose using the scene context information obtained from moving and stationary pixels in the key frames, in conjunction with motion features, to solve the action recognition problem on a large (50 actions) dataset with videos from the web. We perform a combination of early and late fusion on multiple features to handle the huge number of categories. We demonstrate that scene context is a very important feature to perform action recognition on huge datasets.

Finally, we focus on solving practical problems in representing actions by bag of spatio-temporal features (i.e. cuboids), which have proven valuable for action recognition in recent literature. We propose a framework to index large scale motion features using Sphere/Rectangle-tree (SR-tree) for incremental action detection and recognition.

Major: Computer Vision

Educational Career:
Bachelor's of Electronics and Communication, BS, 2002, Jawaharlal Nehru Technological University, Hyderabad, A.P, India
Master's of Electronic Systems and Engineering Management, MS, 2006, Fachhochschule SÃ¼dwestfalen, Soest, Germany

Committee in Charge:
Dr. Mubarak Shah, Chair, Electrical Engineering and Computer Science
Dr. Gita Reese Sukthankar, Electrical Engineering and Computer Science
Dr. Lei Wei, Electrical Engineering and Computer Science
Dr. Brian E. Moore, Department of Mathematics

Approved for distribution by Dr. Mubarak Shah, Committee Chair, on May 31, 2012.

The public is welcome to attend.