Action recognition represents one of the most challenging problems in computer vision given that it embodies the combination of several uncertain attributes, such as the subtle variability associated with individual human behavior, the challenges that come with viewpoint variations, scale changes, and different temporal extents. Nevertheless, action recognition solutions are critical in a host of important application domains, such as video indexing, surveillance, human-computer interface design, analysis of sports videos, and the development of intelligent environments.

In this dissertation, we investigate template-based action recognition algorithms that can incorporate the information contained in a set of training examples, and we explore how these algorithms perform in action recognition and video summarization.

First, we introduce a template-based method for recognizing human actions called Action MACH. Our approach is based on a Maximum Average Correlation Height (MACH) filter. MACH is capable of capturing intra-class variability by synthesizing a single Action MACH filter for a given action class. We generalize the traditional MACH filter to video (3D spatiotemporal volume), and vector valued data. By analyzing the response of the filter in the frequency domain, we avoid the high computational cost commonly incurred in template-based approaches. Vector valued data is analyzed using the Clifford Fourier transform, a generalization of the Fourier transform intended for both scalar and vector-valued data.

Next, we address three seldom explored challenges in template-based action recognition. The first is the recognition and localization of human actions in aerial videos obtained from unmanned aerial vehicles, a new medium which presents unique challenges due to the small number of pixels per human, pose, and moving camera. The second issue we address is the incorporation of multiple positive and negative examples of a target action class when generating an action template. We address this issue by employing the Fukunaga-Koontz Transform as a means of generating a single quadratic template which, unlike traditional temporal templates (which rely on positive examples alone), effectively captures the variability associated with an action class by including both positive and negative examples in the template training process.

A third challenge we address is the role of action recognition in video summarization. To solve this problem we introduce action-specific video summaries in an effort to provide a means of generating a compact video representation based on a set of activities of interest, while preserving the scene dynamics of the original video. In our approach, a user specifies which activities interest him and the video is automatically condensed to a short clip which captures the most relevant events based on the user's preference. We follow the output summary video format of non-chronological video synopsis approaches, in which different events which occur at different times may be displayed concurrently, even though they never occur simultaneously in the original video. However, instead of assuming that all moving objects are interesting, priority is given to specific activities of interest which pertain to a user's query. This provides an efficient means of browsing through large collections of video for events of interest.
Dr. Fernando Gomez, EECS
Dr. Raymond Surette, Criminal Justice and Legal Studies

Approved for distribution by Dr. Mubarak Shah, Committee Chair, on June 22, 2010.

The public is welcome to attend.