Human Activity Recognition

J.K. Aggarwal

Department of Electrical and Computer Engineering,
The University of Texas at Austin,
Austin, Texas, 78712
aggarwaljk@mail.utexas.edu

Motion is an important cue for the human visual system Mobiles have fascinated children, Zeno (circa 500 B.C.) studied moving arrows to pose a paradox and Zeke is investigating the human brain devoted to the understanding of motion. In computer vision research, motion has played an important role for the past thirty years. An important application of motion research has emerged in the past decade. This area is devoted to the study of people – facial expression recognition, gesture recognition, whole-body tracking and human activity recognition. The broad area is at times called "looking at people." In addition, study of human motion is of interest to number of disciplines including psychology, kinesiology, choreography, computer graphics and human-computer interaction.

Recent interest in certain applications like surveillance, sports-video analysis, monitoring a car driver has heightened our interest in human activity recognition. A major goal of current computer vision research is to recognize and understand human motion, activities and continuous activity. Initially, we focused on tracking a single person; today we focus on tracking, recognizing and understanding interactions among several people, for example at an airport or at a subway station. Interpreting such a scene is complex, because similar configurations may have different contexts and meanings. In addition, occlusion and correspondence of body parts in an interaction present serious difficulties to understanding the activity.

Prof. Aggarwal's interest in motion started with the study of motion of rigid planar objects and it gradually progressed to the study of human motion. The current work includes the study of interactions at the gross (blob) level and at the detailed (head, torso, arms and legs) level. The two levels present different problems in terms of observation and analysis. For blob level analysis, we use a modified Hough transform called the Temporal Spatio-Velocity transform to isolate pixels with similar velocity profiles. For the detailed-level analysis, we employ a multi-target, multi-assignment strategy to track blobs in consecutive frames. An event hierarchy consisting of pose, gesture, action and interaction is used to describe human-human interaction. A methodology is developed to describe the interaction at the semantic level.

Professor Aggarwal will present analysis and results, and discuss the applications of the research.

S.K. Pal et al. (Eds.): PReMI 2005, LNCS 3776, p. 39, 2005.