Object and Activity Recognition in Video Streams

Prof. Ramakant Nevatia
Department of Computer Science and Electrical Engineering
Director, Institute for Robotics and Intelligence Systems (IRIS)
University of Southern California, Los Angeles, CA, USA

Abstract:
Video data is becoming ubiquitous with millions of cameras deployed in a variety of environments. It is not feasible for human users to analyze a significant portion of this data; automated tools are needed to extract meaningful information that may be used to index and browse through large collections of video libraries, or to provide alarms for real time applications

It is apparent that the key elements of the content in a video are the objects in it and the activities that are carried out by the objects (usually humans). First, we need a formal, yet natural, representation for the events and objects that can be operated on by computer programs and can be accessible to human users at the same time. Then, we need methods to actually infer these representations from the raw data streams.

Inference of content from video data faces the traditional challenges of visual analysis as the observed data is dependent on a number of factors including the viewpoint, illumination conditions, object surface properties and varying styles of carrying out the same action. A model-based approach that incorporates the available sources of knowledge is essential to solve these problems.

This talk will describe an Event Description Framework (EDF) which itself is based on an earlier Video Event Representation Language (VERL). Methods for inferring actions- given objects, their properties including trajectories, will be described. The actions may be carried out, in general, by multiple agents acting simultaneously. A brief introduction to methods for detecting and tracking objects will also be provided. Methods used for activity recognition are largely based on statistical graphical models including combinations and modifications of Hidden Markov Models (HMMs).