Linking Video Information Across Cameras

Vassilios Morellas
Department of Computer Science & Engineering
University of Minnesota

Abstract:

Video camera systems currently abound. They can be deployed in as diverse domains as security and surveillance, transportation, robotics, medical, multimedia, inspection, etc. The bulk of research in the last years has concentrated on single camera video analytics. Despite their significant advances high rates of false alarms limits their appeal in the market place. In addition, there is a reasonable argument between developers and users about the degree of automation that multi camera systems can provide. Our research has targeted the development of video analytics in single cameras and incremental linking of video information across a camera network. Our work spans development of video modules (behaviors) involving a single person (e.g., tracking, thrown object detection, camera tampering, etc.) or multiple people—group behaviors, (e.g., people counting, abnormal crowd behaviors, etc.)

Background/foreground separation is a fundamental capability existent in every modern camera system. A background/foreground detection technique will be presented that works under operating conditions not compatible with the deployment of the dominant Gaussian mixture model methodology. Our method is based on nonparametric modeling of image regions (layers) and the detection of foreground via a layer propagation process that uses the ``A-contrario'' framework for automatic thresholding. Region modeling allows for robust operation under nominal motions of the camera, in dynamic backgrounds and in cases of stop/go object motions.

Tracking objects across cameras is viewed as matching the image regions that comprise the object to be tracked. Foreground objects are modeled using both appearance information and structural information encoding adjacency relationships of the constituent to the object regions. Graph modeling is use to encode structural info. A vertex in the graph represents an “image region” in the foreground, while the topological relationship between two segments is encoded by an edge between the corresponding vertices. This provides a simple yet powerful and generic representation of the foreground corresponding to various classes of objects. The effectiveness of this model is discussed in the context of an “example based query” type of application for people matching in videos even in cases of occlusions.

When tracking people appearances across multiple cameras, problems like illumination variations, pose changes, etc. need to be taken into account. We have been exploring the use of Region Covariance Descriptors (RCD). A typical RCD is the covariance matrix formed from the feature vectors of an image region usually found after background subtraction. RCD has been found to be quite effective in single camera tracking. However, these descriptors perform poorly when the pose of the subject changes, which is a problem that needs to be considered when tracking people across multiple cameras. We have been developing a framework for incremental exploration of an appearance space and the ultimate learning of people appearances. This is to say that due to the lack of a robust appearance model, a procedure is set forth to explore people appearance distributions and eventually clusters corresponding to specific people of interest. This approach facilitates tracking across camera networks and takes advantage of the wealth of data readily available from tracking in a single camera.