Intelligent Video Surveillance

Rogerio Feris, IBM T. J. Watson Research Center

Surveillance video is generally used in two key modes, watching for known threats in real-time and searching for events of interest after the fact. Typically, real-time alerting is a localized function, e.g. an airport security center receives and reacts to a "perimeter breach alert", while investigations often tend to encompass a large number of geographically distributed cameras like the London bombing, or Washington sniper incidents. Enabling effective search of surveillance video for investigation and preemption involves indexing the video along multiple dimensions. As the number of surveillance cameras increases, the need for intelligent systems based on video analytics that automatically parse video content and extract events of interest is imperative. In this talk, I will discuss the emerging field of "intelligent video surveillance", its applications, challenges, and limitations.

After a brief overview of video content extraction techniques and applications, I will introduce the IBM Smart Vision Suite (SVS), which is a cutting-edge smart surveillance system with automatic event detection and search capabilities. In particular, I will focus on a specific problem which is quite relevant to criminal investigation processes: how to automatically search for people in surveillance videos. Current research on this topic focuses on approaches based on face recognition, where the goal is to establish the identity of a person given an image of a face. However, face recognition is still a very challenging problem, especially in low resolution images with variations in pose and lighting, which is often the case in surveillance data. I will describe a new approach which addresses this problem in an alternative way by finding people based on semantic attributes, such as facial hair, eyewear, clothing color, etc. These attributes can be extracted using detectors learned from large amounts of training data. At the interface, the user can specify a set of personal characteristics, and the system then retrieves events that match the provided description. For example, a possible query is "show me the bald people who entered a given building last Saturday wearing a red shirt and sunglasses." This capability is useful in several applications, such as finding suspects or missing people.

I will conclude the talk discussing limitations and emerging research problems in the field of intelligent video surveillance. In particular, I will briefly cover my view on new directions for research, including the shift from blob-centric to object-centric video analytics, the development of non-usual computational cameras, and learning from large amounts of data.