Chapter XV
Event Detection, Query, and Retrieval for Video Surveillance

Ying-li Tian
IBM T. J. Watson Research Center, USA

Max Lu
IBM T. J. Watson Research Center, USA

Arun Hampapur
IBM T. J. Watson Research Center, USA

Andrew Senior
IBM T. J. Watson Research Center, USA

Lisa Brown
IBM T. J. Watson Research Center, USA

Chiao-fe Shu
IBM T. J. Watson Research Center, USA

Rogerio Feris
IBM T. J. Watson Research Center, USA

Yun Zhai
IBM T. J. Watson Research Center, USA

ABSTRACT

Video surveillance automation is 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, for example, 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 event detection, query and retrieval of surveillance video for preemption, and investigation, involves indexing the video along multiple dimensions. This chapter presents a framework for event detection and surveillance search that includes: video parsing, indexing, query and retrieval mechanisms. It explores video parsing techniques that automatically extract index data from video indexing, which stores data in relational tables; retrieval which uses SQL queries to retrieve events of interest and the software architecture that integrates these technologies.

1. INTRODUCTION

Video analysis and video surveillance are active areas of research. The key challenges are video-based event detection and large-scale data management and retrieval. While detecting and tracking objects is a critical capability for smart surveillance, the most critical challenge in video-based surveillance (from the perspective of a human intelligence analyst) is retrieval of the analysis output to detect events of interest and identify trends. In this chapter, we describe a specific system, the IBM Smart Surveillance Solution, in order to detail an open and extensible framework for extracting events in video which can be used for real-time alerting, searching during investigations with unpredictable characteristics, or exploring normative (or anomalous) behaviors.

Current systems have begun to look into automatic event detection. These are often point solutions for detecting license plate numbers, abandoned objects, or motion in restricted locations. However, the area of context-based interpretation of the events in a monitored space is still in its infancy. Challenges here include: using knowledge of time and deployment conditions to improve video analysis, using geometric models of the environment and other object and activity models to interpret events, and using learning techniques to improve system performance and detect unusual events. The first hurdle that must be overcome is to provide extensible search capabilities based on the broadest possible set of meaningful event metadata which can be provided by state-of-the-art point solutions.

This chapter explores these issues using as an example the IBM Smart Surveillance Solution. Its architecture is outlined as an example of a system which addresses the problems of indexing event metadata and providing extensible search. Its components provide examples of video parsing, indexing and retrieval methods which are deployed by the system. Lastly, its interface shows many examples of how an end-user may search for specific information regarding a real-world investigation.

2. BACKGROUND

Video surveillance systems which run 24/7 (24 hours a day and seven days a week) create a large amount of data including videos, extracted features, alerts, statistics etc. Designing systems to manage this extensive data and make it easily accessible for query and search is a very challenging and potentially rewarding problem. However, the vast majority of research in video indexing has taken place in the field of multimedia, in particular for authored or produced video such as news or movies, and spontaneous but broadcast video such as sporting events. Efforts to apply video indexing to completely spontaneous video such as surveillance data are still emerging.

The work in video indexing of broadcast video has focused on such tasks as shot boundary detection, story segmentation and high level semantic concept extraction. The latter is based on the classification of video, audio, and text into a small (10-20) but increasing number of semantically interesting categories such as outdoor, people, building, road, vegetation, and vehicle. For broadcast video, the goal is to find a high level indexing scheme to facilitate retrieval. The task objectives are very different for surveillance video. For surveillance video, the primary interest is to learn higher level behavior patterns. In both broadcast and surveillance video, there exists a semantic gap between the feasible low level feature set and the high level semantics or ontology desired by the system users.

Because of its practical nature, surveillance video analysis has been extensively explored. However, compared to the vast amount of research in broadcast video search, such as (Hauptmann, 2006; Naphade,

Related Content

Analysis of Text Identification Techniques Using Scene Text and Optical Character Recognition

Monica Gupta, Alka Choudharyand Jyotsna Parmar (2021). *International Journal of Computer Vision and Image Processing (pp. 39-62).*

www.irma-international.org/article/analysis-of-text-identification-techniques-using-scene-text-and-optical-character-recognition/288385

A New Image Distortion Measure Based on Natural Scene Statistics Modeling

Abdelkaher Ait Abdelouahad, Mohammed El Hassouni, Hocine Cherifiand Driss Aboutajdine (2012). *International Journal of Computer Vision and Image Processing (pp. 1-15).*

www.irma-international.org/article/new-image-distortion-measure-based/68001

Securing Digital Image with Authentication Code

Siva Charan Murahariraoand Manik Lal Das (2014). *Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies (pp. 203-215).*

www.irma-international.org/chapter/securing-digital-image-with-authentication-code/108417

A Compilation of Methods and Datasets for Group and Crowd Action Recognition

Luis Felipe Borja, Jorge Azorin-Lopezand Marcelo Saval-Calvo (2017). *International Journal of Computer Vision and Image Processing (pp. 40-53).*

www.irma-international.org/article/a-compilation-of-methods-and-datasets-for-group-and-crowd-action-recognition/188760

2D and 3D Visual Attention for Computer Vision: Concepts, Measurement, and Modeling

Vincent Ricordel, Junle Wang, Matthieu Perreira Da Silvaand Patrick Le Callet (2016). *Innovative Research in Attention Modeling and Computer Vision Applications (pp. 1-44).*

www.irma-international.org/chapter/2d-and-3d-visual-attention-for-computer-vision/139000