Second International Workshop on Knowledge Discovery and Data Mining

Towards suspicious behavior discovery in video surveillance system

LI Yingjie
[1] Information Engineering Department,
University of Science and Technology Beijiing,
Beijing, China
[2] Information Engineering Department,
Zhejiang Forestry University,
Zhejiang Linan, China
comliy@163.com

YIN Yixin
Information Engineering Department,
University of Science and Technology Beijiing,
Beijing, China
Ustb_yyx@126.com

***Abstract*—Video surveillance systems are becoming common in commercial, industrial, and residential environments. The systems in used are constructed mainly by hard devices with no or very few soft intelligence. It is difficult for human to recognize important events as they happening and to control over unwilling situations by staring at the screens all the time. Soft intelligence to identify human behaviors in the surveillance systems is expected. A system's architecture for this goal is presented in this paper. Bottom-up processing methods and top-down design schemes are integrated in the architecture. The integration may increase the accuracy of relevance algorithms and reduce the computing cost. The feasibility of the system is assured.**

Keywords-video surveillance systems; soft intelligenc; architecture; human behaviors

I. INTRODUCTION

With the decreasing cost of video hardware, video surveillance systems are becoming common in commercial, industrial, and residential environments. A monitor center is established in a care giving zone. Several computers are used to gather and store the video which are captured by cameras distributed in the zone. Monitor screens are used to monitor the environments and to control over unwilling situations. The architecture of the video surveillance system is shown in Fig. 1.

But the increased scale creates difficulties for human trying to recognize important events as they happening and to track people through the monitored space. And the systems in used are constructed mainly by hard devices with no or very few soft intelligence (say CCTV systems). The record videos have been constituted a valuable help in terrorists identification after the crime, but the systems were not capable to give a real time alarm. This limits the efficiency of video surveillance systems. The soft intelligences are called in video surveillance systems. Responding to this, many researches about this have been done. This sensing technology requires scientific and technological knowledge in many different fields, from computer vision and pattern recognition, to behavior analysis, to wireless sensor networks, efficient video streaming, and so on. Some consolidates theories have been transferred in commercial products too. Many commercial surveillance systems aim at a small number of susceptible areas, such as, systems offered by Panasonic [1], Vidient [2], and Vistascape [3]. Although these systems enable many useful functions, including moving direction detection, counting moving target passing through a region, and detection an intruder in a forbidden area, further functions are needed to secure a common zone. International journals together with many international conferences [4] [5] have shown the research activities in new generation video surveillance systems, sensor networks and their integration. Most of the activities in the research laboratories and in the industrial research centers are now centered on human motion capture and human activity monitoring [4] [5]. Functions to identify human behaviors in the intelligent surveillance systems are expected. However, challenges are still crucial for the following facts:

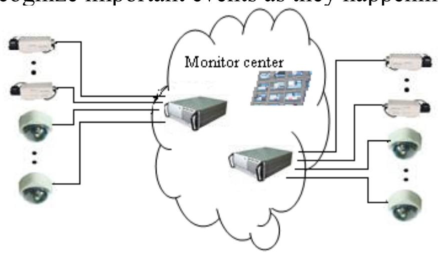

Fig. 1 The architecture of video surveillance system

- The backgrounds of real scene are complex and variation, and relevance researches are done mainly in the laboratory or with criteria.
- It is also a challenging task to identify human motion and their interaction due to the ambiguity caused by body articulation, loose clothing, and mutual occlusion between body parts. Further more; the same action can be normal or abnormal with the different ambient context.

978-0-7695-3543-2/09 $25.00 © 2009 Crown Copyright
DOI 10.1109/WKDD.2009.22

- The video surveillance techniques always lead to tremendous computation. This limits their application in the real time function of surveillance systems.

For video surveillance system of civilian environment monitoring, such as residential environment, the system goal focuses on discovering suspicious behaviors. Cameras are set in several important sites. Usually, only single camera is set for one site in the civilian monitoring system. The accurate techniques, such as face identification and 3D model building, are not necessary in the system. So, it is feasible to build a more functional system integrating current researches. A system's architecture for this goal is presented in this paper. We organized the paper as follows: first, the system architecture is described in section 2; then, the relative works are viewed in section 3; last, the advantages of the architecture are discussed in section 4.

II. THE SYSTEM ARCHITECTURE

In this section, the problem of classification for imbalanced dataset is defined, and the frame of CRB is illustrated.

To identify human activities accurately in real environment can not meet currently. Instead of identifying all activities in the video sequences, the system function is aiming at real requirements. Patterns are built by requirements of monitored sites and the system will developed only meet the special function. And, more patterns will be accumulated and the identifying ability will increase with the system running. By this meaning, the system architecture is devised as Fig. 2.

The data of the system are divided to three layers. They are initial video sequences, real data and patterns. Abnormity patterns are extracted from requirements of real sites. The abnormity activities are varying with time and scenes. Such as, passing across a line is normal in a street, but it may be abnormal in front of the resident windows; Lingering for a long time is normal in a square, but it may be abnormal in the parking lots. Activity patterns can be built by the abnormity patterns then. Features that must be extracted from video sequences can be ascertained and feature patterns are constructed. All patterns can be regarded as domain knowledge in the system.

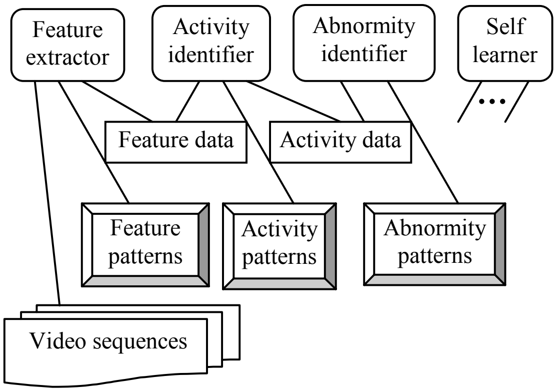

Fig. 2 The system architecture

While running, real feature data are extracted according to feature patterns from video sequences by feature extractor; Activity identifier then identifies the activity of the feature data according to activity patterns; Abnormity identifier identifies whether abnormity is arising according to abnormity patterns. The system sends out an alarm while an abnormity is identified. The alarm means suspicious activity is detected by the system. This may call human's attention in time and further crime may be held back.

Self learner learns new patterns while identifiers being incapability by current patterns. For example, for a new set of feature data, while activity identifier can not identify it according to current activity patterns, it may be sent to self learner; so, new activity patterns may be added or some current patterns may be altered. Interacting with user may be needed in these processes.

III. RELATIVE WORKS

Several layers of techniques should be integrated into the system. They may include foreground segmenting and feature extracting techniques, human recognizing and tracking techniques, and techniques about data analysis, pattern recognition, and fuzzy logic.

Foreground segmentation is usually needed as an initial step in video surveillance applications. Features of foreground are concerned in our system. Background subtraction is typically used to segment moving regions by comparing each new frame to a model of the scene background. But background is varying frequently in real environment. Lipton et al proposed several methods that based on frames difference to adapt the real environment [6] [7]. The methods make difference between adjacent frames to detect foreground region from video sequences. But hollows always rise with these methods. VASM developed a new application to avoid the defaults with combining the methods of background subtraction and the methods of frames difference [8]. Pavel and Korshunov's researches show that there exists a sweet spot in a piece of video sequences [9]. This means that video frames can be reduced without significantly affecting the accuracy of the surveillance tasks. The researches can guide feature extracting with sampling methods to adapt time requirements of multi video streams.

Methods based on templates are needed to treat the feature data and the activity data to identify human shape and their activities in our frame. Data analyzing techniques are identifying tools in these methods [10] [11]. It is especially challenge to identify the interactions of multiple peoples. Sangho Park and J K Aggarwal presented a method to segment and track multiple body parts in two-person interactions [12]. The method is based on multi-level processing at pixel-level, blob-level and object-level. They label human body part at object level according to domain knowledge. They also presented a method to estimate body parts and interactions based on Bayesian network due to the ambiguity of human body parts and activities in video sequences [13]. Wei Niu et al proposed an efficient representation of human activities that enables recognition of different interaction patterns among a group of people based on simple statistics computed on the tracked trajectories [14].

Data analysis, pattern recognition techniques, and fuzzy logic are also useful in abnormity identifier of the system. The techniques were used successfully in a cardiac surgical patients monitor system [15]. Relevant researches are abundant and further view will not be presented here.

IV. Discussions

In traditional computer vision systems, the flow of information is typically bottom-up. The low level image processing modules take video input, perform early vision tasks such as background subtraction and object detection, and pass this information to the high level to identify further [16]. The system frame follows the processes. But the top-down reasoning also plays roles in the architecture. The high level patterns are used to instruct the low level data extracting process. The low level processes perform forensic analysis of archival video and actively acquire information required to arrive at identity decisions of high level. This may increase the accuracy of relevance algorithms and reduce the computing cost. The difficulties mentioned in section 1 may be weaken or conquered.

References

[1] Panasonic. Security Products. http://www.panasonic.com/ business/ security/

[2] Vidient. http://www.vidient.com/

[3] VistaScape Security System. http://www.vistascape.com/

[4] http://portal.acm.org/citation.cfm?id=1178782

[5] http://www.cvg.cs.rdg.ac.uk/slides/pets.html

[6] Lipton A, Fujiyoshi H, Patil R (1998). “Moving target classification and tracking from real-time video”. In: Proc IEEE Workshop on Applications of Computer Vision, Princeton, NJ, pp.8-14.

[7] Anderson C, Bert P, Vander G (1985). “Change detection and tracking using pyramids transformation techniques”. In: Proc SPIE Conference on Intelligent Robots and Computer Vision, Cambridge, MA, pp.72-78.

[8] Collins R (2000). A system for video surveillance and monitoring: VSAM final report. Carnegie Mellon University: Technical report.

[9] Pavel Korshunov, Wei Tsang Ooi (2005). “Critical video quality for distributed automated video surveillance”. In: Proc of the 13th annual ACM international conference on Multimedia.

[10] Cui Y, Weng J (1997). “Hand segmentation using learning-based prediction and verification for hand sign recognition”. In: Proc of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp.88-93.

[11] Bobick A, Davis J (1996). “Real time recognition of activity using temporal templates”. In: Proc of IEEE workshop on applications of computer vision, Sarasota, Florida, pp.39-42.

[12] Sangho Park and J K Aggarwal (2002). “Segmentation and tracking of interacting human body parts under occlusion and shadowing”. In IEEE Workshop on Motion and Video Computing, Orlando, FL, pp.105-111.

[13] Sangho Park and J K Aggarwal (2006). “Recognition of two-person interactions using a hierarchical Bayesian network”. In: Proc of the 4th ACM international workshop on Video surveillance and sensor networks, Santa Barbara, California, USA.

[14] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang (2004). “Human Activity Detection and Recognition for Video Surveillance”. http://www.cs.ucsb.edu/~yfwang/papers/ icme04_tracking.pdf.

[15] Christian Oberli and Jorge Urzua et al (1999). “An expert system for monitor alarm integration”. Journal of Clinical Monitoring and Computing, 1999, 15, pp.29-35.

[16] Vinay D Shet, David Harwood, and Larry S Davis (2006). “Top-down, bottom-up multivalued default reasoning for identity maintenance”. In: Proc of the 4th ACM international workshop on Video surveillance and sensor networks, Santa Barbara, California, USA.