Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data

Zhenfeng Shao, Jiajun Cai, Zhongyuan Wang

Abstract—Video surveillance system has become a critical part in the security and protection system of modern cities, since smart monitoring cameras equipped with intelligent video analytics techniques can monitor and pre-alarm abnormal behaviors or events. However, with the expansion of the surveillance network, massive surveillance video data poses huge challenges to the analytics, storage and retrieval in the Big Data era. This paper presents a novel intelligent processing and utilization solution to big surveillance video data based on the event detection and alarming messages from front-end smart cameras. The method includes three parts: the intelligent pre-alarming for abnormal events, smart storage for surveillance video and rapid retrieval for evidence videos, which fully explores the temporal-spatial association analysis with respect to the abnormal events in different monitoring sites. Experimental results reveal that our proposed approach can reliably pre-alarm security risk events, substantially reduce storage space of recorded video and significantly speed up the evidence video retrieval associated with specific suspects.

Index Terms—Big data, video analysis, surveillance systems, data mining.

1 INTRODUCTION

Modern cities are usually exposed to emergency situations, such as traffic accidents, terrorist attacks and crimes [1]. As a typical example, Paris terror attacks in 2015 left at least 129 people dead. In order to stop criminals and minimize social security dangers, a large number of smart monitoring cameras and surveillance systems have been widely deployed in urban areas. The recorded voice and video data are useful for investigating cases when crimes actually happen. Video surveillance systems have been playing more and more essential roles in crime prevention and forensic.

Driven by video analytics and Big Data techniques in recent years, intelligent video surveillance systems have witnessed rapid development, springing up a large number of representative technical achievements and applications [2], [3], [4]. In the aspect of storage and retrieval for Big Data, Guo et al. [5] designed a method named SVIS to save video data for subsequent analysis of operating platforms. Shavachko et al. [6] created HDFS system to save big video data effectively in order to resolve the problem that storage redundancy and large space consumption are caused by video data. Huang et al. [7] constructed a retrieval systematic method to improve the efficiency and accuracy of retrieval results in the Big Data era. Ding et al. [9] proposed a novel framework for large-scale cross-modality search via collective matrix factorization hashing, which transforms data into binary representation to perform high-speed search with low storage cost. Xiao et al. [10] exploited global redundancy in big surveillance video data of multiple sources in large time span for efficient storage. Besides, with the help of cloud storage, a large number of storage strategies have been developed for big data storage [11], [12], [13].

In the aspect of detecting abnormal behaviors, Kim et al. [14] proposed a novel approach to identify unusual behaviors by using behavior templates. Boiman et al. [15] indicated that if a series of behaviors cannot be reconstructed by continuous clips in video database, they belonged to unknown or anomalous behaviors, and otherwise they belonged to normal behaviors. Chen et al. [16] put forward a novel algorithm based on the acceleration features to detect abnormal crowd behaviors in surveillance video. This method can efficiently alert the suspicious target with fast movement or reverse walking. Cong et al. [17] developed a method which constructs a normal behavior dictionary before performing test samples, and calculates the cost of sparse reconstruction whether it belongs to abnormal behavior. Mehran et al. [18] assumed that abnormal states in crowd could be regarded as anomalous behaviors which happen in scene. Based on this assumption, they delivered social force features to LDA (Latent Dirichlet Allocation) model, which can identify abnormal behaviors in crowd, and simulates the interaction among people in scenes. In some other studies, naïve Bayes algorithm [19], artificial immune system [20], fuzzy clustering with multiple auto-encoders [21] and clustering methods, such as K-NN [22], radius-based [23]
and ant-based ones [24], are used to identify abnormal behaviors. Yuan et al. [25] and Chae et al. [26] proposed an online spatial-temporal analytics and seasonal-trend decomposition for anomaly detection. Cong et al. formulated the abnormal event detection as a matching problem for more robust processing [27] and also used sparse representation to solve this problem [28]. Tran et al. [29] adopted spatiotemporal path search rather than sliding window-based approaches for video event detection. Li et al. [30] proposed a detector which based on a video representation that accounts for both appearance and dynamics, using a set of mixture of dynamic texture models. Liu et al. [31], [32] achieved video segmentation and object tracking based on semantic segmentation, which is helpful to video data analysis. Popoola [33] particularly made a detailed review on abnormal behavior recognition.

Meanwhile, various intelligent surveillance systems have been designed and put into practical uses. Smart monitoring cameras can identify abnormal behaviors of monitoring sites independent of human engagements with the help of the embedded video analytics algorithms. The typical abnormal behaviors include cross-border invasion, entering or leaving area, leaving or taking baggage, wandering, fast movement, gathering and so on. VSAM (Visual Surveillance and Monitoring) system was designed by Carnegie Mellon University [34] to support battlefield awareness. They used automated video understanding technology that enabled a single human operator to monitor activities over a complex area in a distributed network of active video sensors. Zong et al. [35] proposed a system which includes motion detection and target identification. They used adjacent frame differences to mark the motion object and self-mapping neural network to improve the accuracy of detection. Smart Surveillance System [36] by IBM covers the latest computer vision algorithms for automatic event detection in urban surveillance scenarios, specifically designed to handle crowded environments. Knight system created by University of Central Florida [37] is an off-the-shelf surveillance system that detects, categorizes, and tracks moving objects. It also presents a summary in terms of key frames and a textual description of observed activities for final analysis and response decision. V.s-star system [38] completed by Institute of Automation, Chinese Academy of Sciences can automatically interpret human and vehicle motion in surveillance video.

As for applications, Garcia et al. [39] introduced their work into public transit by highway, which aims to improve the safety and accessibility of people and monitors abnormal behaviors. Huang et al. [38] showed the usage of their system in subway and important affairs which can monitor the scene intelligently including indoor and outdoor in 24 hours. Arroyo et al. [40] used their system to detect suspicious behaviors in shopping malls. Jiang et al. [41] introduced image super-resolution technique into video surveillance application to enhance the quality of video big data for obtaining pronounced face recognition effects. Hancke et al. [42] declared that by monitoring people’s actions, it was likely to determine a violent action and discern the potential criminals, which was thus able to maintain the public security.

To satisfy the actual demands, a lot of places have been installed with the smart cameras. Most of those products can provide pre-alarming functions for these special occasions, like banks. However, there are some limitations in using these smart cameras for detecting abnormal events:

(1) The existing intelligent surveillance systems can only detect and alarm single abnormal event yet without bridging the spatial and temporal association among multiple unusual events. However, it is quite not convincing to judge suspicious behavior by a single monitoring. As the case of wandering in the front of a bank, the occasional wander outside the bank may be a usual behavior for awaiting others. It only makes sense to treat the wandering as suspicion when it happens repeatedly or takes a long time.

(2) The huge amount of video acquired by the city-scale monitoring network results in the rapid increasing of storage costs. IDC (International Data Corporation) calculated that surveillance video data accounts for 65 percent of whole data, which was far more than any other data like transaction data, medical data, entertainment and social media data. Since the massive surveillance video needs to be stored for several months or years, it leads to a large storage cost.

(3) The amount of false alarming resulted from the data explosion is beyond the limitations of manual processing. Traditional methods for obtaining evidences highly depend on the surveillance video within or near the accident site. However, when the incident passes through a wide range of space and time, it is hard to find any valuable evidences on the criminals from massive surveillance video, which hampers the efficiency of resolving cases.

The recent emerging smart monitoring cameras are able to automatically identify abnormal behaviors through the built-in intelligent algorithms, greatly boosting the performance of the surveillance system. However, the above mentioned three major challenges have not been fundamentally resolved. The essential reason is that the existing system only individually accepts alarm information from each front-end camera and makes a limited range of alarming, without performing collaborative analysis among geospatially interrelated camera network. Besides, the detection results on unusual behaviors are not fully exploited in terms of deep utilization, paying little attention to storage and retrieval on massive video but for event alarming.

Some new observations on criminal activities and security operations can help us optimize intelligent processing of big surveillance video data. First, criminals often inspect various places in different time before committing crimes, which are captured by cameras located in different sites. Through temporally and spatially associative validation within camera network, false alarming can thus be ruled out. Second, video storage is mainly used for post-investigation, and so the video without unusual behaviors does not have to be preserved for long term. Third, the actual occurrence of a case is almost accompanied by harbingers of unusual behaviors. Therefore, when
the real case occurs, in favor of surveillance videos catching abnormal behaviors, we can reduce video search scale within valid space-time range in investigating criminal cues.

Following this idea, this paper advocates an intelligent processing approach to big surveillance video data driven by smart front-end cameras. In our approach, we do not natively and passively receive and process the alarming information from smart front-end cameras, but make full use of spatial and temporal attributes of multi-site monitoring cameras to perform collaborative association analysis. This approach will disclose the intrinsic relationships and reveal hidden patterns among a number of seemingly separate abnormal behaviors. In methodology, it is applied to three major procedures of intelligent video surveillance system: danger alarming before an event, high-efficiency storage during an event, and rapid evidence retrieval after an event. This way, we can improve the alarming accuracy of the abnormal behaviors with inherent association, the efficiency of the video preservation associated with the abnormal behaviors, and the discovery efficiency of the case clues under the abnormal behavior constraint.

The remainder of the paper is organized as follows. Section 2 particularly presents our solution to association analytics, smart storage and rapid retrieval. Issues on system implementation are discussed in Section 3. In Section 4, our approach is verified with several practical examples. Section 5 concludes the paper.

2 The Proposed Approach

To make better use of surveillance video data, we present a methodology to maximize the role of smart cameras in surveillance system. First of all, an abnormal behavior database is established for smart cameras, in order to store and manage the warnings in emergency. Secondly, surveillance videos are stored selectively according to the warning information, which substantially reduces storage space. At last, when video evidences are needed in some cases, those associated with abnormal behaviors are traced and accessed preferentially.

2.1 Multi-point Association Analysis Based Pre-alarming

In order to identify the events which have significant differences from occasional normal behaviors, characteristics of criminal behaviors can be disclosed with analysis of experience and historical data. Besides, anomalous behaviors can be judged in terms of several dimensions, which include occurrence frequency, duration, identity of target and the surrounding range. To facilitate analysis, single-site period analysis and multi-site association analysis are engaged.

2.1.1 Association Analysis Model

For multi-point association analysis, it is assumed that $T$ smart cameras track abnormal behaviors with $m_i$ denoting the number of abnormal behaviors confirmed on $k$ moment. Based on that, we have a formally tractable problem:

$$F \triangleq \{f_i\}_i = 0,1, \ldots, m_i; \ t = 0,1, \ldots, T.$$  \hspace{1cm} (1)

where $F$ is the basic information matrix of joint events, which is constituted by the probability statistics distance of abnormal behavior and smart cameras, from smart cameras $t$ and abnormal behavior $i$. $f_i$ is the statistical distance between abnormal behavior $i$ and smart camera $t$, which refers to as the probability that a certain abnormal behavior corresponds to a camera. In actual surveillance occasions, whether a certain abnormal behavior is confidently captured by a given camera depends on a variety of factors, such as the view field of the camera, the distance and orientation of the target, and even the lighting conditions. For example, if a target is far from a stationary camera and thus beyond its observation scope, the unusual behavior incurred by this target is difficult to detect. Under a scenario of relay monitoring of multiple chained cameras, the possibility of a specific abnormal behavior corresponding to these cameras is further governed by the coincidence degree between the moving target trajectory and the monitoring sites on the path. Therefore, the statistical distance is an inherent indicator more subjected to the geographical environment than the camera’s capability (i.e., false or missed detection).

Further, we use $\Theta$ to specify the generalized joint event set under the condition that matrix $F$ satisfies the new feasibility rules. $\Theta$ is the generalized subset of events which satisfies the condition that each abnormal behavior can be captured by smart cameras (one or more, including the case of uncaptured). $\Theta$ is the generalized subset of events that each smart camera has a source of abnormal behaviors (one or more, including the case of no abnormal behavior). Combining $\Theta_i$ and $\Theta_i$, we show the abnormal behaviors that smart monitoring cameras can observe as follows,

$$\Theta = \{\Theta_i, \Theta_i\}. \hspace{1cm} (2)$$

When applied to the actual situation, apart from satisfying the possible relationship between the smart cameras and abnormal behaviors, a combination of temporal and spatial conditions for association analysis is also required.

Association analysis is a kind of analysis that seeks the temporal and spatial correlation of abnormal behaviors, which can figure out more sensitive behaviors for public security rather than those isolated behaviors happening by chance.

This paper adopts temporal autocorrelation function to measure temporal correlation characteristics. The temporal autocorrelation of abnormal behaviors refers to the frequency or the attribute value of abnormal behaviors which exhibit correlation characteristics in time. Assume that the abnormal behaviors are sequentially captured by smart cameras in time, let $z(t)$ represent the frequency or an attribute value of abnormal behaviors happened at the time of $t$, the temporal autocorrelation function with the delay of $k$ is defined as:

$$\rho_k = \frac{Cov(z(t), z(t+k))}{\sigma_{z(t)} \sigma_{z(t+k)}} = \frac{\text{E}[z(t)z(t+k) - \bar{z}(t)\bar{z}(t+k)]}{\sqrt{\text{E}[z(t)z(t) - \bar{z}(t)^2] \text{E}[z(t+k)z(t+k) - \bar{z}(t+k)^2]}} \hspace{1cm} (3)$$
Fig. 1. Abnormal behavior association analysis based pre-alarming

\[
\rho_i \text{ not only describes the correlation of abnormal behaviors detected by smart cameras in arbitrary intervals but also reveals its alteration trend with delay } k.
\]

The spatial autocorrelation of abnormal behaviors puts emphasis on the distribution of massive abnormal behaviors in whole surveillance network. It is used to determine whether the overall spatial distribution of events presents a clustering feature. The common spatial autocorrelation index is Moron’s I, namely,

\[
I = \frac{1}{\sum n \sum j} \times \frac{\sum i=1 \sum j=1 \omega_j(z_i - \bar{z})(z_j - \bar{z})}{\sum i=1 \sum j=1 (z_i - \bar{z})^2} (4)
\]

where \(z_i\) and \(z_j\) are spatial variables, representing the frequency or an attribute value of abnormal behaviors occurring within \(n\) different monitoring sites with \(\bar{z} = \frac{\sum i=1 \sum j=1 z_i}{n}\). Weight coefficient \(\omega_j\) indicates the adjacency relationship between the site \(i\) and the site \(j\) where abnormal behaviors really happened.

2.1.2 Association Analysis Procedure

In the light of the theoretical models discussed before, we can readily conduct multi-site association analytics by exploiting the plentiful alarming information from front-end smart cameras, which roughly involves three major steps.

(1) Abnormal behavior database is constructed firstly by recording the descriptive information of abnormal behaviors from front-end smart cameras. The description metadata at least consists of the types of monitoring target and abnormal behavior, name of monitoring site, ID of smart camera, time of event’s beginning and ending, and so on. Other useful information which is beneficial to abnormal event detection could also be recorded if needed. Meanwhile, to facilitate fast video investigation, snapshots or video clips associated with a specific behavior are extracted and stored together with metadata. The database is updated in real time, which is continuously supplemented with the latest behaviors on-the-fly whenever they are detected by front-end smart cameras. The abnormal behavior database provides the metadata for the subsequent analysis. Moreover, since not all the unusual behaviors benefit from multi-point correlation analysis, for instance, an occasional traffic accident or an event happening in a completely isolated monitoring site, we need to figure out the ones suitable for further analysis from the abnormal behavior database. Particularly, Eqs. (1) and (2) are used to generate the candidates, with the geographical map of cameras and the recoded abnormal behaviors in the database as input.

(2) Upon construction of the behavior database, historical data in a single monitoring site is then examined collectively with its temporal autocorrelation in Eq. (3) based on such attributes as lasting duration or occurrence frequency of recoded behaviors in the database. The behavior with temporal autocorrelation larger than preset empirical threshold will be claimed as a dangerous one. When an abnormal behavior warning is triggered by a certain smart camera, the intelligent surveillance system will immediately start the temporal correlation analysis with respect to historical records in the database. According to comprehensive analysis, judgment of whether it is merely an isolated action will be drawn.

(3) Similarly, we need to carry out multi-site association analysis formulated in Eq. (4), due to the action of criminal suspects’ anti-investigation. Usually, the target of the suspect is not limited to one, but may focus on multiple objectives and choose the most vulnerable one to attack. Therefore, we extend association analysis from single site to multiple sites based on historical data so as to reveal the regular patterns of abnormal behavior in time and space. This will help system trigger more scientific and reasonable pre-alarming. Note that, when computing the spatial autocorrelation index in Eq. (4), the weight coefficients account for the geographical topology of the camera network.

This process flowchart is shown in Fig. 1, in which the assistive techniques refer to as those serving for specific purposes (but beyond the scope of this paper). For instance, people who wander for many times in one occasion should be identified by technology of person re-identification [43].

In the typical case of wandering, if there appear logics listed below, it will be declared as a high-risk event.

(1) Wandering happening in different sites simultaneously: when a smart monitoring camera discovers the pedestrian wandering, according to spatial relevance, we
quickly find the alarm records from the surrounding monitoring sites. Meanwhile, we count the frequency within a certain interval and hence calculate its temporal autocorrelation so that we can obtain the relevance of the behavior over a certain period of time. If a couple of wandering events are detected by the cameras in different locations at same time, the wandering behavior will be claimed as a high-risk one.

(2) Frequent occurrences of wandering in short term and several sites: we summarize the alarm information of the wandering events in the region by the time scale of day, week and month. If the hovering frequency in the analysis period is highly correlated and much higher than the normal historical data at the time of the absence of the case, the behavior will be determined as a suspicious one. The relevant monitoring sites are also remarked as high-risk protection ones. This analysis can be started in a definite time or while wandering is detected by smart cameras.

(3) Same person appearing in different sites: we identify whether the same person is present in the hovering behaviors that are highly relevant in terms of time and space. The higher the frequency of the same person is, the more suspicious. If it is found that wandering behavior is due to the same person, it is highly suspicious abnormal behavior. How to identify pedestrians is achieved by a technique so called person re-identification, which is also mentioned in Fig. 1, seen as an assistive technology.

2.2 Abnormal Behavior-aware Smart Storage of Surveillance Video

The demands for abnormal behaviors detection vary remarkably from distinct places. For example, the important facilities should be paid more attention to the person who is entering the area, while the public places should be more concerned about the social security events like fighting or brawl. Therefore, this paper classifies the monitoring sites into different kinds in order to conduct the risk assessment and construct appropriate abnormal behavior set. As to choose the loosest site to commit crimes, the unusual behavior information of the same type of monitoring site within a certain range around itself is still the important basis for the risk assessment of the monitoring site. Finally, there is close relationship between premonitory abnormal behavior and security events, but it does not mean the causal relationship. Therefore, historical alarms and records of incidents are used to excavate abnormal behaviors associated with the incidents that actually occurred. Alternatively, they are also used to optimally correct risk assessment model and strengthen the role of the key factors in dependent variables of the model.

To facilitate the establishment of a security risk assessment model, based on the essence of the monitoring sites, the sites are roughly divided into three categories in this paper:

- Important facilities: power plants, chemical plants, bridges, wharf, etc.;
- Financial institutions: banks, financial branches, ATM machines, etc.;
- Public places: airports, railway stations, parks, schools, museums, stadiums, scenic spots, streets, shopping malls, entertainments, etc.

Meanwhile, this paper also categorizes the most typical anomalous behaviors involved in each type of monitoring targets:

- Important facilities: invasion of the region, cross-border invasion, entering/leaving area, high-density flow and so on;
- Financial institutions: wandering, suspicious face, suspicious license plates, and other fast movement;
- Public places: leaving or taking baggage, fast movement, illegal parking, gathering, fighting, chasing and so on.

2.2.1 Construction for Risk Assessment Model

According to the established abnormal behavior database, risk assessment model is constructed based on the potential security risk consequences of anomalous behaviors. Specifically, a risk weight is given for each type of behavior to form a risk-weighted table. The abnormal behavior of high potential security risk will be assigned to a large
Fig. 3. Fast retrieval procedure driven by abnormal behaviors

weight value, and vice versa. The occurrence frequency of unusual behavior is counted, which is combined with the risk weight so as to obtain the risk value. The risk value is further mapped into five risk levels.

The determination of risk weights not only depends on the potential hazards of different unusual behaviors (for instance, fighting is more serious than fast movement), but also accounts for the geographical environment of monitoring sites (for instance, financial institutions are more important than ordinary public places). According to historical criminal cases and the opinions of experts, monitoring target types (corresponding to geographical environment) typically include important facilities (e.g., military facilities, power plants, urban infrastructures), financial institutions and public places. In contrast, unusual behaviors are much more diversified, for example, typically, running, fighting, wandering, gathering, and so on.

Furthermore, it is acknowledged that abnormal behavior poses a risk to the place of its occurrence, and meanwhile it may also be dangerous to the surrounding targets as the criminal space is not stationary. Therefore, the risk weights of unusual behaviors should be examined from two perspectives: local risk weights and surrounding risk weights, which indicate the degree of hazard to in-place and around targets, respectively. For example, gathering has almost the same effect on the current locations and the surroundings, but leaving/taking baggage has little impact on the places other than the local one.

On the basis of the above discussion, the initial settings of the risk-weighted table are shown in Table 1.

Risk value $R$ is then calculated by:

$$R = \sum_{i=1}^{N} w_i n_i$$

where $N$ is the total warned number of abnormal behaviors, $w_i$ is the risk weight corresponding to abnormal behavior $i$, and $n_i$ is the frequency of abnormal behavior $i$.

The risk value $R$ is mapped into five levels according to Eq. (6):

$$L = \begin{cases} A, & R \geq 50 \\ B, & 50 > R \geq 30 \\ C, & 30 > R \geq 10 \\ D, & 10 > R \geq 5 \\ E, & 0 \leq R < 5 \end{cases}$$

where $A$ represents extremely high, $B$ represents high, $C$ represents medium, $D$ represents general, and $E$ represents no risk, respectively.

Depending on the different security requirements of monitoring sites, a set of high-risk abnormal behaviors are constructed for each type of monitoring site. The abnormal behaviors in the set are regarded as potential threats to this type of monitoring site. The definition of high-risk abnormal behaviors set is listed as follows:

$$S_f = \{\text{Invasion of the region, Cross-border invasion, Gathering}\}$$
$$S_p = \{\text{Suspicious face, Fast movement}\}$$
$$S_F = \{\text{Gathering, Fighting}\}$$

where $S_f$ represents important facilities, $S_p$ represents financial institutions, and $S_F$ represents public places.

2.2.2 Smart Storage Strategies for Surveillance Video

It should be verified that whether an actual security event occurs at local monitoring site or similar monitoring sites within a certain range. If a security event has occurred, all the surveillance videos are stored in accordance with the common requirements. Otherwise, according to risk assessment results, three different surveillance video storage strategies can be carried out: delete, not delete and partially delete.

With the growth of accidents or unusual events, the risk assessment model will be updated and corrected. By using large historical data of anomalous behaviors and records of security incidents, the unusual behaviors lead-
ing to actual security incidents are defined as premonitory ones. Then, based on abnormal behavior database, the frequency and types of the premonitory abnormal behavior in a week, which causes the security event, are counted and sorted from more to less.

If several abnormal behaviors have same frequency, the one which has larger weight is highlighted. We select the first five abnormal behaviors to update the high-risk behavior set of corresponding monitoring targets and double their respective risk weights in the risk-weighted table. If the unusual behavior is not in the risk weight table, we add an entry to the table and set its risk weight value to the maximum value in the table. The risk assessment model should be regularly revised in this way. The whole process of smart storage is shown in Fig. 2.

2.3 Rapid Evidence Video Retrieval Driven by Abnormal Events

Abnormal behaviors, which have a strong correlation with specific events, not only refer to behaviors that influence the process of the event, but also represent those unusual actions which do happen after the incident. These features are an important basis for subsequent abnormal event retrieval.

In the era of big data, typical behaviors can be summarized through data mining from massive historical cases, with the assistance of specific expert knowledge. Here are some typical associations between unusual behaviors and security events that have been verified:

- Wandering outside or around the spot before bank robbery;
- Running nearby after robbery;
- Entering (or leaving) the community before (or after) burglary;
- Gathering before group incidents such as riot;
- Gathering before affray;
- Moving against crowds after the violence;
- Illegal parking in violence.

Therefore, establishing mapping relationship between typical events and abnormal behaviors that could be detected by smart monitoring cameras is indispensable. After occurrence of a case, by searching abnormal behavior database, all of unusual behaviors that have a strong correlation with the case according to event-abnormal behavior association table can be identified, including their occurrence time and spots. The meticulous analysis can be further performed with related snapshot images and videos captured by smart cameras. In this way, there is a significant decrease in data-analysis scale during the process of video investigation, and efficiency in finding evidences can be raised accordingly. The outline of this part is shown in Fig. 3.

2.3.1 Event-abnormal Behavior Correlation Model

In the light of common sense and expert knowledge, several typical correlation principals of event-abnormal behavior are chosen as the initial entries, which are shown in Table 2.

We then update the correlation model of event-abnormal behavior table as Algorithm 1.

### Algorithm 1. Event-abnormal Behavior Correlation Model Table Updating Algorithm

1. Definition of event set:
   \[ E = \{ \text{Bank robbery, Robbery, Riot} \} \]
2. Definition of abnormal behavior set:
   \[ A = \{ \text{Regional invasion, Entering/Leaving the area, Wandering, Fast movement, Fighting} \} \]
3. With the increasing number of different types of abnormal behaviors detected by smart monitoring cameras, set \( A \) shall be enlarged correspondently.
4. With the expansion of abnormal behavior set \( A \), cases or events which are closely related to newly added abnormal behaviors will be analyzed and added to the events set \( E \).
5. For an given entry \( E_i \) in event set \( E \), according to the historical case data, the frequency of associated abnormal behaviors will be counted and sorted as set \( S_1 \).
6. For an given entry \( E_i \) in event set \( E \), according to experts' knowledge, all the relevant unusual behaviors are enumerated by priority and labeled as set \( S_2 \).
7. Get intersection elements
   \[ S = S_1 \cap S_2 \]
   Select the very first two behaviors in \( S \), and record them into the correlation model table as the most likely associated unusual behaviors with a particular event.

2.3.2 The Use of the Abnormal Behavior Database

Retrieval, statistics, analysis and other operations can be performed with respect to the abnormal behavior database, which thus can provide sufficient data for video evidence collection, social security situation census and security risk assessment.

Here are major functions of the abnormal behavior database.

- Retrieval: abnormal behavioral constrained video forensics. Retrieve under types of behavior, site or time, and show retrieval image outcome along with time and site of the event.
- Statistics: survey of social security situation. By time statistics, the curve shows the number of abnormal behaviors per month in a year. By location statistics, the curve shows the number of abnormal behaviors occurred in a year with respect to site. The two-dimensional statistics in terms of time and places, with a two-dimensional surface showing the number of abnormal behaviors occurred a year.
- Analytics: the assessment of security risk. Cluster the spatial and temporal attributes of abnormal behavior, draw a security surface map, and visualize levels of risk by different colors with geographic information. Try to figure out spatial and temporal attributes of abnormal behaviors and predict their trends, find key nodes of security, and conclude information in favor of decision-making. For example, thieves are willing to steal at the end of a year or robberies are more likely to happen in a secluded place. All these information can
provide scientific reference for police assignment in the future.

The most appealing application is to obtain video evidences associated with specific abnormal behaviors. According to the existing cases or events, the correlation model table of event-abnormal behavior will be looked up to obtain the type of the behavior. Then the abnormal behavior database will be retrieved by type to get its time and site. After that, all the snapshot images and videos that are relevant to this kind of unusual behavior are checked, and all of them in different sites are summarized into a complete summary video. Routes of the targets’ movement can also be tracked by doing this. At last, according to actual demands, index field of original videos in the database is used to locate the complete surveillance videos.

3 SOFTWARE DESIGN AND IMPLEMENTATION

Although smart cameras have gained popularity in recent years, there is no enough system software that can exert its features completely. Currently, there exist three major problems for using smart cameras: easy to produce false alarm, requirement for large storage and huge manual retrieval workload. Integrating smart cameras with the method mentioned in Section 2, we have designed the software to address the problems of pre-alarming, storage and retrieval. The framework of software is shown in Fig. 4, which consists of abnormal behavior database and three functional modules.

The alarm information of unusual behaviors, which is captured by front-end smart cameras, is used to construct the abnormal behavior database. The abnormal behavior database will provide the metadata for the subsequent analysis. The descriptive information provided by metadata includes the type of monitoring site and unusual behavior, name of monitoring site, ID of smart camera, time of abnormal behavior occurring, time of abnormal behavior ending, index of surveillance video, snapshot or video clip of associated behavior.

As a kind of concrete realization, the abnormal behavior database has its own construction and methods for accessing to information, which is shown as below:

1. The recorded type of abnormal behavior is exactly kept same as the type detected and labeled by smart cameras.
2. The name of monitoring site and the ID of smart camera show the specific geographic location.
3. Recording the time of abnormal behavior when it occurs and ends according to actual alarming.
4. Snapshot or video clip of associated behavior is regarded as a compressed result, which has already removed irrelevant frames from original video.
5. There are two methods to obtain snapshot or video clip: if smart camera has the function of delivering snapshot, you can save it directly. Otherwise, it needs to be done by back-end platform.
6. The index of surveillance video is created for locating videos quickly when case happens.
7. The remark includes some personalized information of behaviors, like the number of people who are wandering and direction of running.

Besides, it contains the description of actual cases relevant to unusual behaviors, with generally recorded by manually.

The functional modules of software include pre-alarming for abnormal events, smart storage for surveillance video and retrieval based on association analysis. We will have an adequate description about our experiments in Section 4. By using our software, the pre-alarming accuracy will be improved and it will be more scientific to manage the storage spaces. What’s more, the data volume will be decreased when we perform analysis on the surveillance video data which is relevant to specific case, and it will greatly enhance the efficiency of the discovery of valuable clues at the same time.

4 PRACTICAL EXAMPLES

4.1 Security Risk Pre-alarming

In the major bank robbery cases happened previously, criminals usually cautiously observed environment around the spot for many times before robberies, so as to decide best chances for committing crimes and routes for escaping as well. The longest observation time even approached 3 hours in a crime committed in 2012 in China. The behavior features are very distinguished from normal activities, such as wandering in the target area seemingly to be aimless, repeatedly entering or leaving the place without doing any business. Fig. 5 shows the location of committing crimes.

According to the method proposed in this article, in terms of criminal’s wandering, we adopt single-site period analysis and person re-identification technology to judge whether there are people staying for too long, and fire alarming if the answer is “yes”. In the upcoming cases, by using multi-spot association analysis, we also observe that the criminal not only wandered for a long time in the spot but also behaved similarly in many other places, which thus produces alarming as soon as possible due to its high spatial correlation. The interface of pre-alarming module is shown in Fig. 6.
Fig. 5. Location of crime scene.

The blue point in the map indicates the long wandering detected by smart cameras in the spot. The red points indicate the similar behavior which happened near the spot. If red point was double clicked, it would turn into green and display detailed information about this point. Note that, the existing person re-identification method [43] is used to identify the same person wandering in different places.

We used these crime videos described above as experimental data, with the outcome listed in Table 3.

As shown in Table 3, it is concluded that the criminal is willing to check out the location as carefully as possible before committing crime, which provide us with great support for detecting criminal and warning. In this situation, the existing surveillance system did not generate any alarming for the potential criminal. On the contrary, when the suspect behaves abnormally at the first time, police is able to alert the possible criminal and prevent the crime. As for the upcoming crimes, both single-site analysis and multi-site association analysis are used to achieve pre-alarming ahead more efficiently.

Fig. 6. Interface of pre-alarming module.

4.2 Smart Storage of Surveillance Video

Three ATM machines in Wuhan University are selected as experimental subjects. The location map is shown as Fig. 7.

Storage strategy proposed in this paper can greatly save storage space. When monitoring financial institutions, unusual behaviors can be discerned due to local and surrounding risk values calculated by initial risk-weighted table, including wandering, suspicious face, leaving or taking baggage, and fast movement. The risk values could be used to decide whether the video should be stored or not and their suitable preservation term as well. If the surveillance video contains high-risk abnormal behaviors, like suspicious faces and fast movement around financial institutions especially, its preservation period should be extended. If a crime is committed in or near the spot and recorded by smart cameras, this video should be preserved forever in case of being checked and recorded in the future.

Fig. 7. Distribution of experimental targets.

Monitoring videos in three days of these three smart cameras are randomly chosen as experimental data. The simulation outcome is shown in Table 4.

Table.4 shows the comparison results of the proposed strategy with the traditional one. Traditionally, the videos collected by monitoring cameras are preserved for at least 3 months. Undoubtedly, this requirement causes a huge unnecessary waste for storage space, since most of videos are useless. In our strategy, most of videos do not have to be stored for long term and part of them can even be deleted immediately. Moreover, the bold number in Table.4 indicates that financial institutions of high-risk abnormal behaviors enjoy an extended storage period.

4.3 Association Analysis and Retrieval

In the case mentioned in section 4.1, police established a special investigation team for collecting evidences. 1500 policemen were assembled to check all videos in the city by a totally manual way, identifying 329 video clips containing criminal clues ultimately, which thus caused a huge waste in labor resources.

By using the method of association analysis and retrieval introduced in this paper, all of videos associated with criminals are identified fast and accurately. In this case, the criminals robbed near a bank. According to the mapping of abnormal behavior association model table, abnormal behaviors that are most likely to happen are wandering and fast movement. Then the unusual behavior database is retrieved by keywords “wandering” and “fast movement”, with all of relevant smart cameras located this way. Afterwards, valuable clues are easily screened out in surveillance videos from these smart cameras.
### Table 1

**Risk-weighted Table**

<table>
<thead>
<tr>
<th>Monitoring target types</th>
<th>Abnormal behaviors</th>
<th>Local risk weights</th>
<th>Surrounding risk weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important facilities</td>
<td>Invasion of the region</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Cross-border invasion</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Gathering</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Financial institutions</td>
<td>Wandering</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Suspicious face</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Leaving/taking baggage</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Public places</td>
<td>Leaving/taking baggage</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Fast movement</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Illegal parking</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Gathering</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Fighting</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>

### Table 2

**Event-abnormal behavior association table**

<table>
<thead>
<tr>
<th>Type of events</th>
<th>Type of abnormal behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank robbery</td>
<td>Wandering</td>
</tr>
<tr>
<td>Disturbance</td>
<td>Gathering</td>
</tr>
<tr>
<td>Robbery</td>
<td>Fast movement</td>
</tr>
<tr>
<td>House breaking</td>
<td>Entering or leaving area</td>
</tr>
<tr>
<td>Affray</td>
<td>Gathering</td>
</tr>
<tr>
<td>Terrorist crime</td>
<td>Inverse waking, illegal parking</td>
</tr>
<tr>
<td>Group incidents</td>
<td>Gathering</td>
</tr>
</tbody>
</table>

### Table 3

**Results of pre-alarming experiment**

<table>
<thead>
<tr>
<th>Monitoring sites</th>
<th>Single-site period analysis</th>
<th>Multi-site collaborative analysis</th>
<th>Pre-alarming: existing surveillance (Yes/No)</th>
<th>Pre-alarming: proposed surveillance (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Long-time wandering</td>
<td>Frequent wandering</td>
<td>Multiple places</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>√</td>
<td>×</td>
<td>×</td>
<td>No</td>
</tr>
<tr>
<td>B</td>
<td>×</td>
<td>√</td>
<td>√</td>
<td>No</td>
</tr>
<tr>
<td>C</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>No</td>
</tr>
<tr>
<td>D</td>
<td>√</td>
<td>×</td>
<td>√</td>
<td>No</td>
</tr>
</tbody>
</table>

### Table 4

**Results of storage experiment**

<table>
<thead>
<tr>
<th>Monitoring sites</th>
<th>No.</th>
<th>Abnormal behavior</th>
<th>Risk values</th>
<th>Strategy of storage (traditional)</th>
<th>Strategy of storage (proposed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wandering</td>
<td>Suspicious face</td>
<td>Leaving/Taking baggage</td>
<td>Fast movement</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
We apply our method to another sensational bank robbery that happened in Wuhan in 2011. The abnormal behavior database is built by historical data of smart cameras near the crime scene, and association analysis is then conducted to find the video clues related to the case. The results are tabulated in Table 5. As shown, relevant surveillance videos are located faster with the help of association analytics. The proposed approach spends near 73 seconds on finding the key surveillance video, while the police almost took two days manually to find it. The bold item in the table is the exact surveillance video that police found, which helps them identify the criminal.

In summary, by a combination of snapshot images, original surveillance videos and unusual events, valuable clues can be found out much easier, which thus helps the police boost their investigation efficiency.

5 CONCLUSION

In contrast to the traditional video surveillance system, the proposed solution contributes to make full use of detected and alarmed events by smart monitoring cameras, which thus effectively improves the performance of intelligent surveillance system, promotes the ability to danger pre-alarming, and greatly saves the storage space for surveillance video data. Meanwhile, the surveillance video data relevant to specific cases will be scaled down, which will greatly improve the efficiency for discovering valuable investigation clues. Several practical cases demonstrate that our approach outperforms the existing solutions.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2016YFB0502603, 2016YFB0100901), National Natural Science Foundation of China (61671332, U1404618), the Fundamental Research Funds for the Central Universities (2042016kf0179, 2042016kf0191), Guangzhou Science and Technology Project (201604020070), Wuhan Chen Guang Project (2016070204010114), Special Task of Technical Innovation in Hubei Province (2016AAA018).

REFERENCES


Zhenfeng Shao received his Ph.D. degree from Wuhan University, China, in 2004. He is now a Professor of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. His research interests are smart city and data mining.
Jiajun Cai received the BEng degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2016. He is currently working toward the MEng degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include image processing and smart city.

Zhongyuan Wang received his Ph.D. degree in Communication and Information System in Wuhan University and is currently working as a professor with Computer School at Wuhan University. His research interests include video compression, image processing and multimedia big data analytics as well. He directed three projects from NSFC and has published more than 80 papers in distinguished international conferences and journals.