Advances in Computer Science Research, (ACSR), volume 52
2016 International Conference on Computer Engineering and Information Systems (CEIS-16)

New Approach of Abnormal Events Judgement and Early-warning Based on Combination of Audio Analysis and Video Analysis

Jun Hu, Xue-Ming Shu, Shi-Yang Tang, Shi-Fei Shen
Department of Engineering Physics, Institute of Public Safety Research, Tsinghua University, Beijing 100084, China
E-mail: hujun18810654098@126.com, shuxm@tsinghua.edu.cn, 1007939330@qq.com, shensf@tsinghua.edu.cn

***Abstract*-The monitoring and early-warning of abnormal events in public places is an important part of the public security research. But nowadays, the monitoring way is basically video surveillance, and the judgement and analysis of abnormal events is based on the video content. This single way has low accuracy and analytic algorithms have some limitations, so the single audio analysis method can't offer timely and effective early-warning. Taking a sudden car fire event for example, this article combines the audio analysis and video analysis to judge abnormal events and also designs an new surveillance system, offering an effective way to early-warning.**

Keywords-abnormal events; audio analysis; video analysis

I. INTRODUCTION

We live in an age of surveillance, and surveillance guarantees the security. Nowadays, there are lots of surveillance cameras in public places. According to IMS Health Inc., a research company mainly focus on market research, in the year of 2010, there were more than 10 million additional surveillance cameras in China, and the number continued to increase at nearly 20% yearly. Although there are plenty of monitoring information, the main analysis method in practice is manual analysis. So many study focus on the intelligent video surveillance, namely analyzing the video surveillance automatically.

So far, there are many research production in the field of intelligent video surveillance. Many abnormal events can be detected automatically. For example, J. Kim et al. [1] uses the Markov Random Field (MRF) model and the maximum a posteriori (MAP) to estimate the abnormal degree, Y. Zhang [2] et al. combines motion and appearance cues for anomaly detection based on Support Vector Data Description (SVDD). Totally, the abnormal events, including vehicle abnormal behavior [3, 4], restricted-area access [5], group fighting [6] and carrying cases [7] can be detected.

However, the detections are not so perfect. First, the definition of abnormal events is changeable in different video content, it is hard to label an abnormal event for videos. Second, the analytic algorithms is not suitable for all different video content, the background of video has significant influence on the effectiveness of analytic algorithms. The analytic algorithms in video analysis have some limitations. Nearly all observed data contain noise, while the analytic algorithms is sensitive to noise. And on the other hand, the video surveillance can't contain enough information to analyze.

As a man has eyes and ears, the surveillance system need both video and audio. With audio analysis, we can *hear* more information, and then the early-warning will be more timely and effective. The study of audio analysis begins on Bell Labs in the early 1950s, and now have great progress [8]. Actually, in the monitoring field, there are another kind of sound besides normal environmental sound, which is called abnormal sound. Abnormal sound can regard as interpretation and presentation of abnormal state and accidents. The typical abnormal sound in monitoring field includes gunfire, blasts, glass smash and screeching. We can infer abnormal events by analyzing abnormal sound. There are some research of audio analysis, for example, Regunathan R [9] design an audio classifying system to extract the Mel Frequency Cepstrum Coefficient (MFCC) of different sound, and then classify and recognize different sound, such as alarm, hitting the wall and opening and locking doors based on Gaussian Mixture Model(GMM). Using the same method, Clavel C [10] recognizes some kinds of gunfire under the background of noise.

This article combines the audio analysis and video analysis to analyze a sudden car fire event. Detecting abnormal sounds, and then analyze video information to justify the judgment of abnormal event. Both audio analysis and video analysis are based on the Matlab software. The data sources are from a car fire event happened in a town, which is recorded by a mobile phone. The data contains video and audio information and is stored in MP4 format. The total length of time is 47 seconds. The event is shown in Fig. 1.

Figure 1. A car fire event happened in a city.

II. AUDIO ANALYSIS

To analyze the audio information, we should extract the sound from the MP4 format data firstly. Using some software tools, the sound information is stored in the wave format.

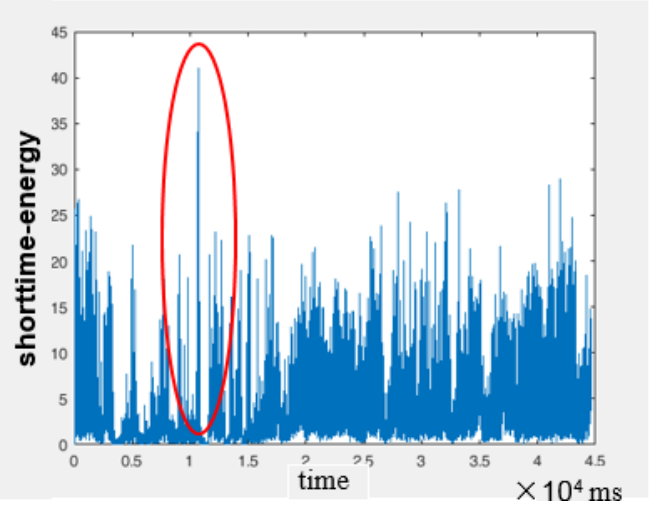

Figure 2 The shorttime-energy of audio information.

The analysis from the domain of time is shown in Fig. 2. It is revealed that the transient peak change dramatically in the time of 11s. And the result shows that on this time points, the sound change rapidly, maybe there are abnormal event occurring.

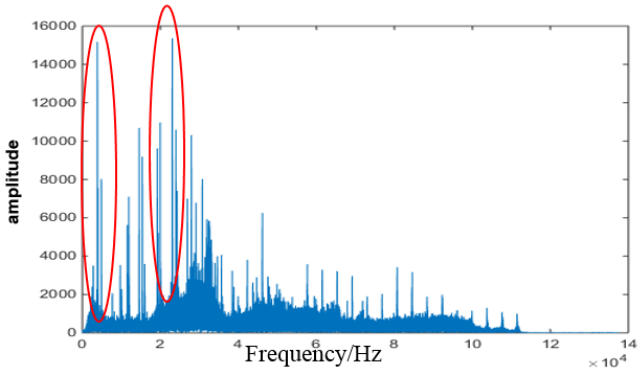

Figure 3. Spectrum of audio information.

The spectrum of the sound is shown in Fig. 3. It is easy to find out there are two peak in the frequency domain. The other part can be explained as the background noise while the two peaks means the abnormal sound (such as explosion, alarm and scream) when abnormal event occurs.

Further study can recognize the scream or explosion by analyzing the Mel Frequency Cepstrum Coefficient (MFCC) of different sound. In this way, we can recognize the event type. But as the article mainly focus on the judgement of abnormal events, the MFCC analysis is not involved.

III. VIDEO ANALYSIS

To represent the abnormity, a common characteristic value in video analysis is the similarity between two pixel matrixes. As the example is fire, when the fire breaks out, the flare and plume come together. Comparing the two adjacent frames, many pixel points will become more bright (flare) or dark (plume). We define the similarity as:

$$\mathrm{sim}(A_i, A_{i-1}) = \frac{|A_i - A_{i-1}|}{m \times n} \quad (1)$$

where A_i and A_{i-1} represent the pixel matrixes of frame i and frame $i-1$. A_i and A_{i-1} have the same size, they are both m rows and n columns, and the element of pixel matrix is the RGB value of each pixel point. To simplify analysis, we convert the image to the binary one. Fig. 4 shows the original image and binary image for a certain frame of the video, with the method of binary conversion, all the binary value of pixel point is available.

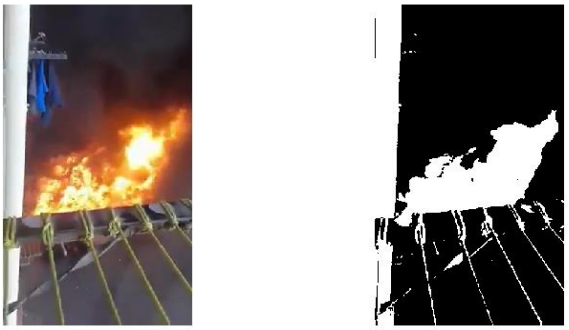

Figure 4 Original image (left) and binary image(right).

From the audio analysis, we get some abnormal time points. Now we will analysis the video information around those time points, calculating the similarity around different time point. To normalize the similarity, the rate R is introduced:

$$\mathrm{R} = \frac{N_{same}}{N_{total}} \quad (2)$$

where N_{total} means the number of total pixel points, and N_{same} means the number of pixel points whose difference of binary value between A_i and A_{i-1} in Eq.1 is equal to zero.

The lower R represent the lower similarity, while the higher abnormity. Setting a threshold value of R, we can define the abnormal event. The result is shown in Table 1.

TABLE I. TABLE R OF DIFFERENT TIME POINTS

Time[second]	R[%]
5	70.88
7	76.34
10	70.59
11	67.61
15	52.47

From Table 1, it can be found that the value of R is lower at the time of 11s and 15s, which means the image has big change. At the time of 11s, the fire breaks out and the flare occurs, which leads to the binary value change of a part of region in the image; while at the time of 15s, the plume begins, the black smoke also change the binary value of a larger part of region in the image.

Combined with the audio analysis, the video analysis offer the information that the fire occurs at 11s, the car begins to smoke, and also at 15s, people begin to scream. With the method of audio analysis and video analysis, we can conclude that there are abnormal event on the 11s. However, there are less obvious image change at 5s and 7s. Actually, there are just sound of the automobile horn.

This example is fire accident, the abnormal sound (scream) and abnormal image (flare) happen nearly simultaneously. For many social security accidents, the abnormal sound (scream) is much earlier than abnormal behavior (fighting). In those cases, the audio is a very important information resource. And for other abnormal events, there are many other different features to represent the abnormity in different ways. To represent the abnormity fully, an abnormal event analysis database is necessary, which need further specific study and is not discussed in this article.

IV. DESIGN OF SURVEILLANCE SYSTEM

According to the analyses above, a new surveillance system is proposed. The surveillance system topology graph is shown as Fig. 4. The system mainly contains the following modules:

Video capture module. The cameras collect the video signals from the real world, and transfer the information to the processors. This module is a sampling and quantization module.

Audio capture and alarm module. The capture part is the same as video capture module, but this module has another important part, namely alarm part, which can release early-warning alarm.

Analysis module. This module can be divided into two sub-modules, video analysis sub-module and audio analysis sub-module. When many abnormal event happens, the detection of abnormal sound is early than abnormal behaviors, the abnormal audio signal occurs firstly, and then the video analysis sub-module further confirm the abnormal information. But it is noted that the two sub-module are independent, as some abnormal events are silent (such as carrying cases).

Storage module. The function of this module is storing data.

Management module. This module is a human-computer interaction module, users can watch and listen to the information, change surveillance fields and manage the data.

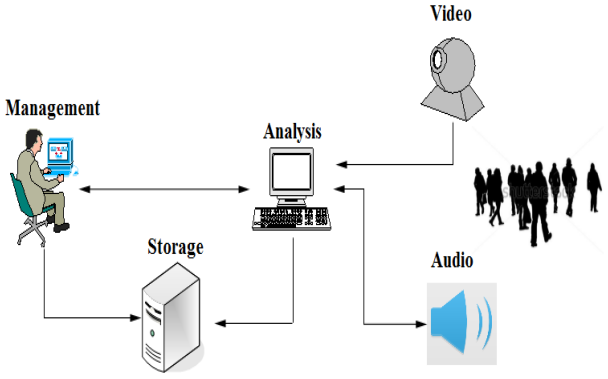

Figure 5. Surveillance system topology graph.

The flow chart is shown as Fig. 5. The system can judge the abnormal events and alarm automatically.

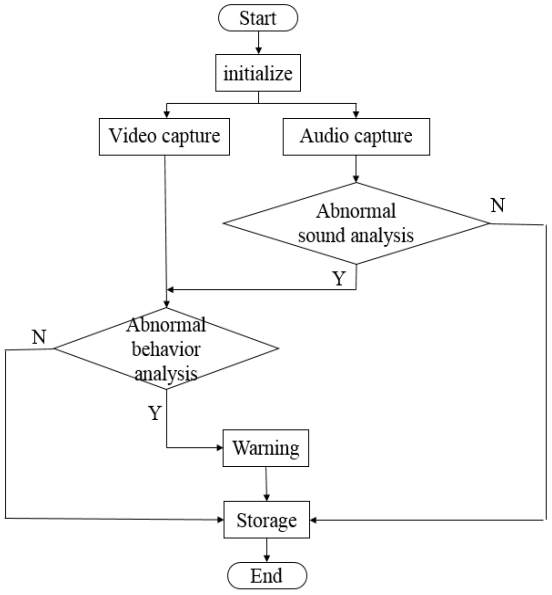

Figure 6. Flow chart of surveillance system.

Comparing with the single video surveillance, the system can collect both image information and sound information. Both types of information corroborate each other to improve the accurate of judgement. And on the other hand, the audio output device can release early-warning alarm to avoid more losses.

V. SUMMARY

This article takes a car fire event for example, showing the combination of audio analysis and video analysis can effectively judge abnormal events, and then designs a new surveillance system, contains video surveillance and audio surveillance. This system can not only collect more information, but also warn early by alarming from audio output device. However, the information in audio and video

are abundant. In addition to judgement of abnormal events and then warning timely, we can also identify people by recognizing the face information in video and voice information in audio, which need further study.

ACKNOWLEDGEMENT

Supported by the National Science & Technology Pillar Program during the 12th Five-year Plan Period (No. 2015BAK12B03), Fire Department of Ministry of Public Security Research Programs (No.2014XFCX10), and the Collaborative Innovation Center of Public Safety.

REFERENCES

[1] J. Kim, K. Grauman, Observe locally, infer globally: a space–time mrf for detecting abnormal activities with incremental updates, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, IEEE Computer Society,2009,pp.2921–2928.

[2] Y. Zhang, H. Lu, L. Zhang. Combining motion and appearance cues for anomaly detection, in: Pattern Recognition 51(2016):pp.443–452.

[3] G. Zen, E. Ricci, Earth mover's prototypes: a convex learning approach for discovering activity patterns in dynamic scenes, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp.3225–3232.

[4] Z. Fu, W. Hu, T. Tan, Similarity based vehicle trajectory clustering and anomaly detection, in: Proceedings of IEEE International Conference on Image Processing, vol.2, 2005, pp.602–605.

[5] J. Konrad, Motion detection and estimation, in: Handbook of Image and Video Processing, vol.207, 2000, p.225.

[6] X. Cui, Q. Liu, M. Gao, D.N. Metaxas, Abnormal detection using interaction energy potentials, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp.3161–3167.

[7] I. Haritaoglu, D. Harwood, L.S. Davis, W 4: real-time surveillance of people and their activities, IEEE Trans. Pattern Anal. Mach. Intell. vol.22, 2000, pp.809–830.

[8] Y. Lee, Han D.K., Hanseok Ko. Acoustic signal based abnormal event detection in indoor environment using multiclass adaboost. Consumer Electronics, 2013, 615-622.

[9] Visalakshi R, Dhanalakshmi P, Palanivel S. Analysis of Throat Microphone Using MFCC Features for Speaker Recognition. COMPUTATIONAL INTELLIGENCE, CYBER SECURITY AND COMPUTATIONAL MODELS, 2015(56):11-14.

[10] C. Clavel, T. Ehrette, G. Richard. Events detection for an audio-based surveillance system. IEEE International Conference on Multimedia and Expo, 2005.