資訊科技國際期刊
International Journal of Advanced Information Technologies (IJAIT)

A Retrieval Method Based on Video Fusion for Surveillance Systems

Cheng-Hsiung Hsieh, Hsien-Chou Liao, Chih-En Wu and Yi-Ming Hsieh

College of Informatics,C.Y.U.T
http://www.inf.cyut.edu.tw

Volume 3, Number 1
June, 2009
College of Informatics
Chaoyang University of Technology
ISSN 1996-2568

A Retrieval Method Based on Video Fusion for Surveillance Systems

Cheng-Hsiung Hsieh
Associate Professor
Department of Computer Science and Information Engineering,
Chaoyang University of Technology
e-mail: chhsieh@cyut.edu.tw

Hsien-Chou Liao
Associate Professor
Department of Computer Science and Information Engineering,
Chaoyang University of Technology
e-mail: hcliao@cyut.edu.tw

Chih-En Wu
Graduate Student
Department of Computer Science and Information Engineering,
Chaoyang University of Technology
e-mail: s9527611@cyut.edu.tw

Yi-Ming Hsieh
Graduate Student
Department of Computer Science and Information Engineering,
Chaoyang University of Technology
e-mail: s9527610@cyut.edu.tw

Abstract

Digital surveillance systems have already been used in community security widely. However, one main drawback of current surveillance systems is that the user has to search videos from different cameras to retrieve all the video clips of a suspicious object. It is quite inconvenient and time-consuming. In this paper, we propose a video fusion and retrieval method to solve the problem. The method is mainly based on the temporal information, the spatial relationship of cameras, and the object feature to retrieve the related video clips of a target object. A prototype system is implemented for demonstrating the feasibility of the proposed method. The experimental results show that the average successful fusion rate of the proposed method is as high as 82.7%.

Keywords: *community surveillance systems, video fusion, video query*

1. Introduction

Surveillance system is getting more powerful and popular as related technologies advance. They are used almost everywhere for security purpose. Community is one of the important scenarios. Currently, a surveillance system is only of recording function with moving object detection in general. Recently, several approaches were proposed to improve the performance of surveillance system in different aspects. In [1], an embedded system with smart cameras was presented by integrating video sensing, video processing, and communications into a single device. In [2], a video-based automated event detection framework was designed to perform object segmentation, multiple objects tracking, object trajectory analysis, and relevant events reporting. In [3], an intelligent visual surveillance system for net cage was presented where a function of intruder classification was provided. In [4], a surveillance system was proposed to adapt with different situations from parking areas, commercial districts to housing. In [5], cameras were used as sensors to enhance the surveillance system with situational awareness capability. In [6], a 3-D particle filter based object tracking was reported. It was able to predict and update the probability distribution of moving objects in 3-D domain for multi-camera surveillance system. In [7], the mean-shift tracking algorithm was extended to include the color and shape information such the tracking performance was enhanced. In [8], a semantic based video retrieval method for visual surveillance was proposed and applied in a crowded traffic scenario.

Generally, the approaches described above were to improve the performances of surveillance systems themselves and nothing to do with the user. Consider when searching a suspicious object, in the current surveillance system the user has to retrieve video clips among cameras manually according to time and location information. It is inconvenient and time-consuming task. In order to solve such problem, a video fusion and retrieval method (VFR) is proposed in this paper. In the proposed approach, with the suspicious object queried, all related video clips among different cameras are automatically retrieved and fused. Then the fused video clips are shown in the user’s interface.

The paper is organized as follows. Section 2 describes the proposed VFR method. Then a real world scenario is presented to verify and to evaluate the performance of the VFR method. Finally, conclusion and the future work are given.

2. The Proposed VFR Method

The objective of VFR method is to facilitate the fast retrieval of a specific object from huge number of video clips for the user. It consists of three parts: moving object detection and information extraction, user query, and video fusion. These parts are described in the following sections.

2.1 Moving Object Detection and Information Extraction

The moving object detection provides important information for retrieving related video clips of the moving object of interest. A common temporal differencing technique is applied for the moving object detection. The frames in a video clip are divided into groups. Each group contains 25 frames. The first frame is considered as the reference frame, other frames as compared frames in the group. Then find difference regions between the reference frame and a compared frame. When a difference region reaches some threshold, a moving object is detected whose maximum and minimum coordinates in XY plane are recorded. An example of moving object detection is illustrated in Figure 1 where the moving object is marked by a rectangle. Attributes for the detected moving object are time, left and top coordinates, width, and height. Since the file size to record attributes of moving objects is small, XML files are used, instead of a database. Besides attributes of moving objects, an XML file for a video clip includes the corresponding camera number and filename. An example of XML file for a video clip is shown in Figure 2.

(a) (b)

(c) (d)

Figure 1 An example of moving object detection: (a) reference frame; (b) compared frame; (c) difference frame; (d) detected moving object

```
<Cameras>
    <CameraNumber>03</CameraNumber>
    <Filename>03 (20-02-08 06_40_43).xml</Filename>
    <Time>20080220064043</Time>
    <Object>
        <Time>20080220064047</Time>
        <Left>0</Left>
        <Top>168</Top>
        <Width>63</Width>
        <Height>71</Height>
    </Object>
    <Object>
        <Time>20080220064048</Time>
        <Left>16</Left>
        <Top>104</Top>
        <Width>39</Width>
        <Height>39</Height>
    </Object>
</Cameras>
```

Figure 2 An example of XML file for a video clip

2.2 User Query

The user query in the VFR is included in the user's interface as shown in Figure 3. The purpose of user query is to provide the user an interface to select the object of interest. In the interface, the user is able to select a specific moving object by clicking on it while viewing video clips. The clicked object is then considered as a query and used to retrieve corresponding video clips. By the corresponding XML file, all the related video clips are fused. The video fusion process is described in the following.

2.3 Video Fusion

Given the moving object in the user query, the information recorded in XML file is analyzed and then used to fuse related video clips. Firstly, the moving direction of the object is determined according to the size of recorded rectangles for the moving object. If the object is coming toward the camera, the size of rectangle is getting larger and vise versa. The moving direction and the pre-defined spatial relationship of cameras are used to determine the next possible camera. Then, those video clips of the next possible camera are checked for the object by their XML files. The clip that has the smallest time difference from the current video clip and not exceeding some given threshold is the next video clip. Such process is repeated until all possible clips are checked. When all possible video clips are found, they are fused based on the order of time stamps in the XML files.

3. VFR-Based Surveillance System

In this section, a system based on the VFR method is presented. For the current surveillance system, there is no convenient way for a user, but manually, to search or query some given object. Therefore, it is inconvenient and time-consuming to retrieve a moving object of interest. Consequently, in the VFR-based surveillance system a friendly user interface is provided for query and retrieval of some given object. Moreover, the related video clips for the queried object are fused and shown in the user's interface. The VFR-based system structure is shown in Figure 3. It consists of three parts and each part performs different tasks: (i) preprocessing: detect moving objects in all video clips and record moving objects information in XML files; (ii) user interface: input query information and display fused video clips; (iii) fusion process: determine the video clips related to the user query, link related video clips by time and location information, and fusion related video clips.

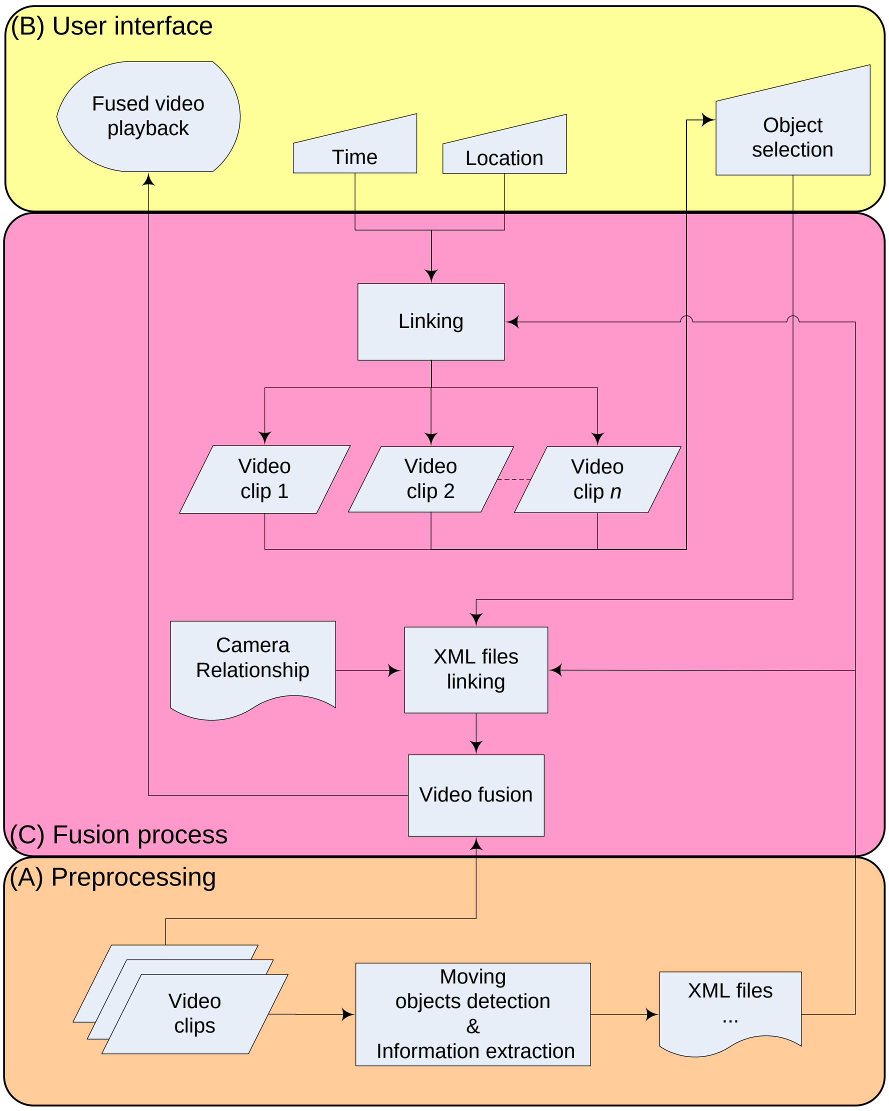

Figure 3 The structure of VFR-based surveillance system

4. Verification and Discussions

In this section, the VRF-based surveillance system described in Section 3 is verified and implemented by Visual Studio.NET. Test video files are collected from some working surveillance system in a community. The video files are from 13 cameras recorded in some day from 6:30 to 8:30 in the morning.

The cameras installed in the community are working on the "alarm" mode. That is, all video clips are not recorded until a moving object is detected. Then, these video clips are analyzed and the information of moving objects in video clips is recorded in XML files. When the user selected a target object in some video clip of a specific camera, related video clips are found and fused by the fusion process described previously. The fused video clips are then shown in the user interface. An illustration of preprocessing, user interface and fusion process is shown in Figure 4.

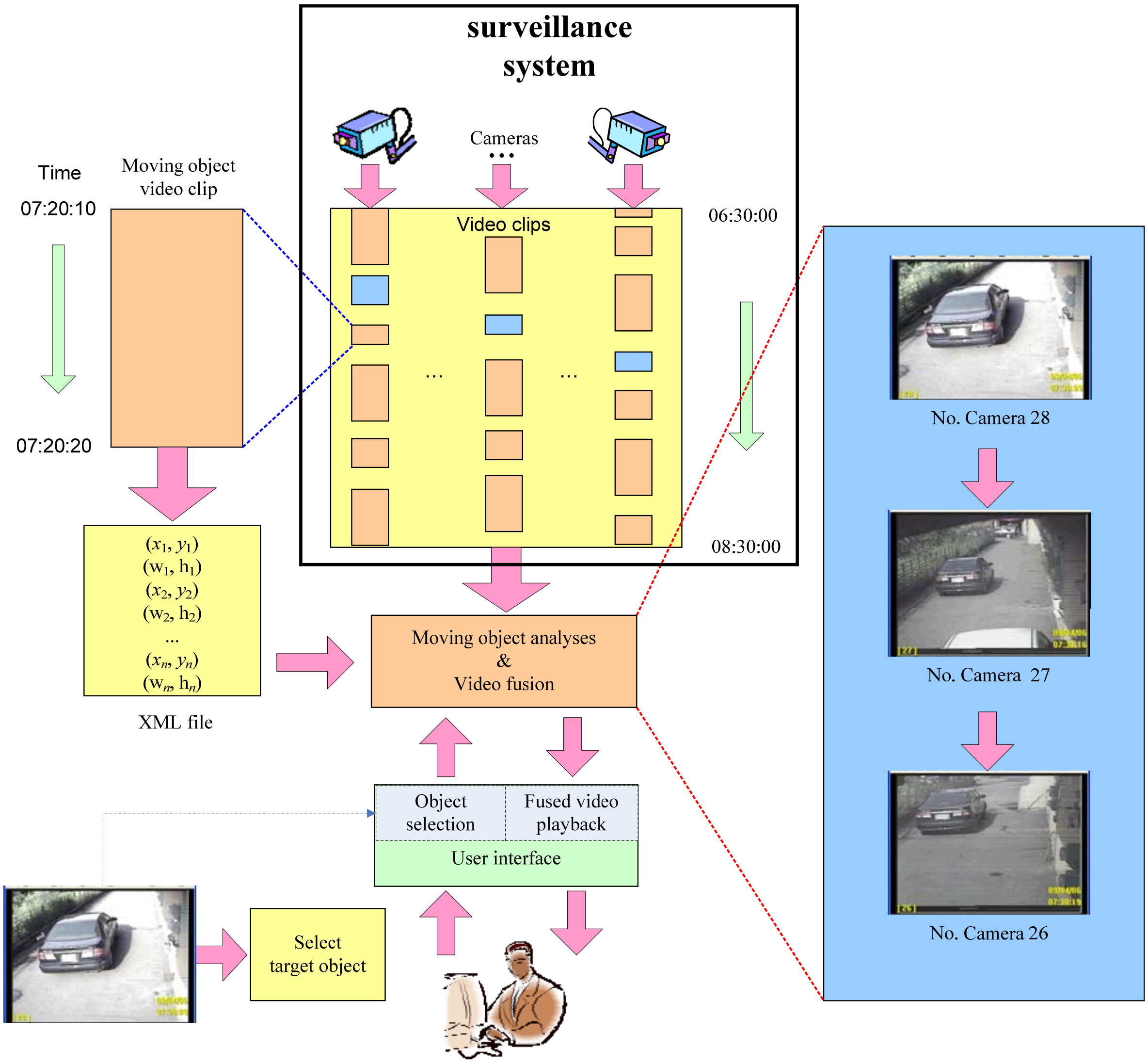

Figure 4 An illustration of preprocessing, user interface and fusion process

A screen shot of the interface for VFR-based surveillance system is shown in Figure 5. In the interface, a user is able to select a specific camera for browsing video clips. The video clips of the camera are displayed on the right-hand side. If an object of interest is found, the user can set the time period for video fusion. When the “Search” button is pressed, all possible video clips are checked and those related video clips are listed in the “Play List”. When playing the video clips, the user can adjust the playback speed to provide further convenience in video inspection. Moreover, the trajectory of the moving object is shown in the community map in the lower right hand side in Figure 5.

To evaluate the performance of VFR-based system, several experiments have been done. First, the total numbers of moving objects are recorded. For moving objects of all the video clips, the fusion process is performed. The number of related video clips that should be fused and the actual video clips that are fused are recorded, respectively. Then, a successful rate is computed. The experimental result is listed in Table 1. Some notes on Table 1 are given as follows. The number of cameras means the number of cameras that a moving object has passed by. This also means the number of related video clips for the moving object. The number of moving objects is listed under the “Number of objects” column. The column “Number of fusions” indicates the number of fusion performed which is equal to the product of “Number of cameras” and “Number of objects”. The “Number of video clips” means the number of video clips should be fused correctly. It is resulted from (“Number of fusions”)×(“Number of cameras”-1). For example, in the case of “Number of cameras” is 6, 5 video clips should be fused correctly for a single fusion operation. Thus it comes to the total number is 510 = 102×(6-1) video clips. In this case, the actual number of video clips fused successfully is 429. Consequently, the successful rate is 429/510, i.e., 84.1%.

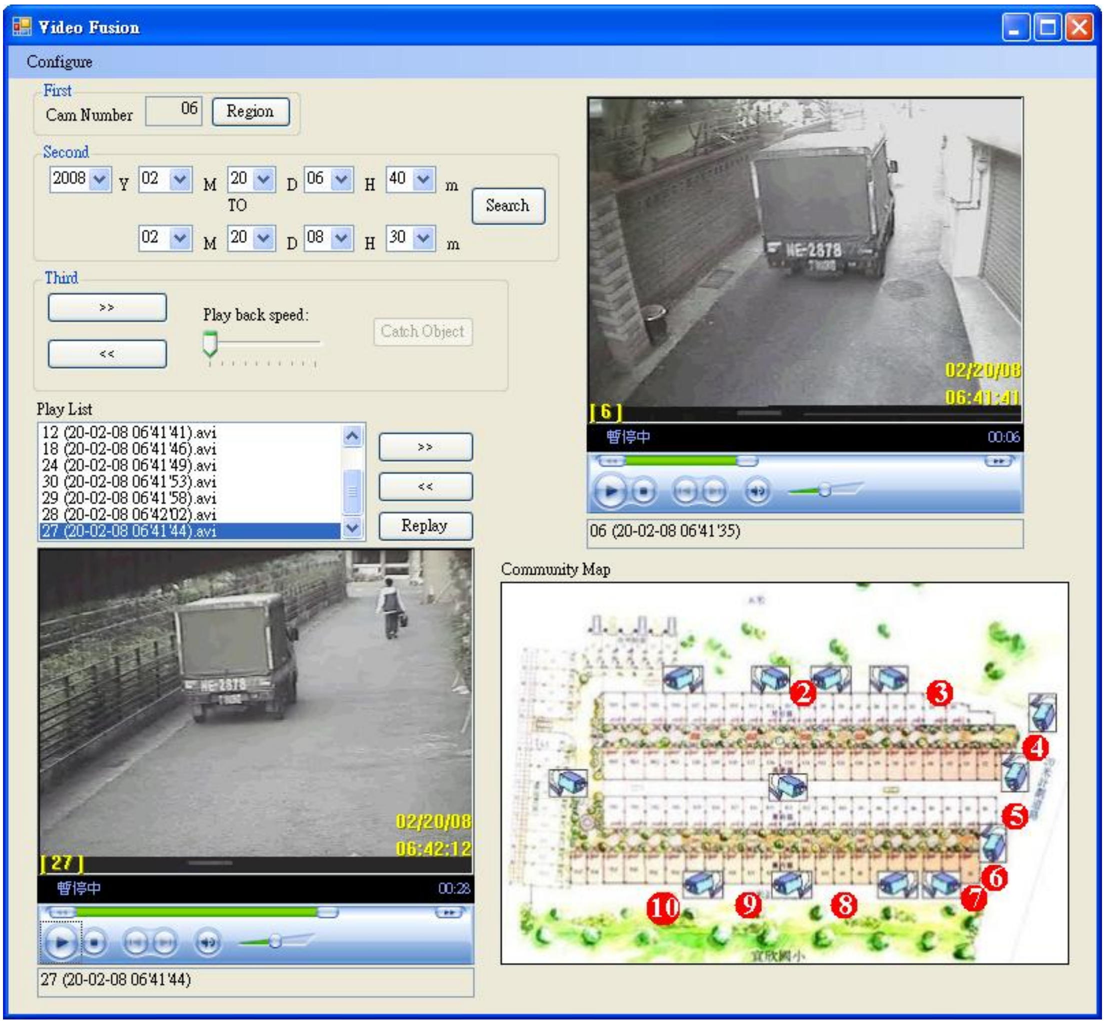

Figure 5 A screen shot of the interface for VFR-based surveillance system

In Table 1, the performance of VFR method achieves 82.7% overall successful rate. It is acceptable but there are still remaining some problems to be solved. Some of failure cases are analyzed and possible reasons are given. They are listed as follows:

(A) The time period of a moving object shown in a video clip is not long enough. Therefore, the information may not be recorded in the XML file and the fusion fails.

(B) When multiple moving objects are in a video clip, objects may be occluded and thus the size of corresponding rectangles recorded in the XML file are incorrect. It causes error on determining the direction of a moving object. Consequently, the fusion fails.

(C) The user doesn't select an object appropriately in a video clip. An example is shown in Figure 6. In Figure 6(a), two objects are overlapped. If the user clicks an object to be fused at this moment, it fails the video fusion since attributes of the object are mixed with the other object. If the user selects an object at the case as shown in Figure 6(b), the fusion process will be successful.

(D) The auto exposure mode of a camera may cause error in the moving object detection.

An example is show in Figure 7. Note that the brightness of backgrounds in Figures 7(a) and 7(b) are different because of the auto exposure of the camera. It causes the detection error based on the background subtraction technique.

Table 1 The experimental result

Number of cameras	Number of objects	Number of Fusions	Number of video clips	Number of video clips fused successfully	Successful rate (%)
2	38	76	76	63	82.9
3	19	57	114	94	82.5
4	20	80	240	201	83.8
5	5	25	100	85	85
6	17	102	510	429	84.1
7	18	126	756	582	77
8	7	56	392	322	82.1
10	5	50	450	385	85. 6
11	1	11	110	110	100
Overall	130	583	2,746	2,271	82.7

(a) **(b)**

Figure 6 The occlusion of moving objects

(a) **(b)**

Figure 7 An example of problem caused by auto exposure mode of a camera

5. Conclusion and Future Work

In this paper, a video fusion based retrieval (VFR-based) method is presented to provide the user of surveillance systems a convenient way to query and track an object of interest in video clips. By this doing, the inconvenience of users for the current surveillance systems, whose function is simply recording with moving object detection in generally, may be relieved when searching an object of interest among video clips. A prototype VFR-based surveillance system is implemented by Visual Studio.NET. With a real-world video clips obtained from a given community surveillance system, the proposed VFR-based method is verified. In the experiment, the overall successful fusion rate 82.7% is achieved by VFR-based method. Though it is acceptable, the VFR-based should be improved for better performance. Several failure cases in the VFR-based approach with possible reasons are listed. The information provides a way to improve the performance of VFR-based method in the future work.

References

[1] Bramberger, M., Doblander, A., Maier, A., Rinner, B. and Schwabach, H., “Distributed Embedded Smart Cameras for Surveillance Applications,” ***IEEE Computer***, Vol. 39, Issue 2, pp. 68- 75, 2006.

[2] Cheng, H. Y. and Hwang, J. N., “Automated Event Detection for Video Surveillance via Hidden Markov Models,” ***The 20th Computer Vision, Graphics and Image Processing Conference***, pp. 165-172, 2007.

[3] W.-C. Hu., C.-Y. Su., C.-C. Chuang, and, C.-W. Huang, “Intelligent Visual Surveillance in Net Cage,” **The 20th *Computer Vision, Graphics and Image Processing Conference***, pp. 158-164, 2007.

[4] Thiel, G., “Automatic CCTV Surveillance Towards the VIRTUAL GUARD,” ***IEEE Aerospace and Electronic Systems Magazine***, Vol. 15, Issue 7, pp. 3-9, 2000.

[5] Trivedi, M. M., Gandhi, T. L. and Huang, K. S., “Distributed Interactive Video Arrays for Event Capture and Enhanced Situational Awareness,” ***IEEE Intelligent Systems***, Vol. 20, Issue 5, pp. 58 -66, 2005.

[6] Wang, Y. C. and Wang, S. J., “Moving Objects Tracking and Correspondence over

Multi-Camera Surveillance System," ***Workshop on Consumer Electronics and Signal Processing***, pp. 47-52, 2007.

[7] J. Wang and Y. Yagi, "Integrating Color and Shape-Texture Features for Adaptive Real-Time Object Tracking," ***IEEE Transactions on Image Processing***, Volume 17, Issue 2, Feb. 2008.

[8] W. Hu; D. Xie, Fu Z., Zeng W., and M. Steve, "Semantic-Based Surveillance Video Retrieval," ***IEEE Transactions on Image Processing***, Volume 16, Issue 4, Apr. 2007.