An Effective Communication Model for Collaborative Commerce of Web-Based Surveillance Services

Chia-Hui Wang  
Dept. of Information Management,  
Kai-Nan University, Taiwan  
chwang@mail.knu.edu.tw

Ray-I Chang  
Dept. of Information Management,  
National Central University, Taiwan  
rchang@mgt.ncu.edu.tw

Jan-Ming Ho  
Institute of Information Science,  
Academia Sinica, Taiwan  
hoho@iis.sinica.edu.tw

Abstract

In this paper, we design and implement a Web-based information system for providing remote surveillance services. The surveillance data can be treated as a kind of digitized product in E-market. By collaborating different value-added services with the same surveillance data, diverse surveillance applications can be introduced. For example, by collaborating with companies (i.e. virtual enterprises) that supply Web-services such as face-recognition, moving object tracking and abandoned object identification, we can create new and diverse digitized products for supporting security systems, intelligent transportation system (ITS) and biometric system in airports, parking lots and offices. For supporting ubiquitous services, we collaborate with the auto-dialup service of PSTN networks (B2B E-commerce) to send messages to the subscribed user (B2C E-commerce) through his/her phone, pager or fax machine. By collaborating the media streaming service, users can watch live/recorded surveillance videos immediately through wired/wireless Web-devices. To reduce network workload, we apply multicast communication for the surveillance video delivery. Furthermore, an effect and efficient motion detection method is proposed to reduce redundant traffic and storage. Experiments show that our proposed schemes can help the system remit the performance penalty from network congestion.

Keywords: Collaborative Commerce, Web-based information systems, Surveillance, Multicast, Traffic Shaping, Media Streaming, Motion Detection.

1. Introduction

Traditional surveillance systems only provide analog services in hardware. Security guards must stay at the security room and look at arrays of CCTV (Closed Circuit TeleVision) or may play back the recorded videos throughout the video tapes to find out surveillance events. However, humans are poor at keeping alert for a long period of time. The security guards may frequently miss incidents visible on the CCTV screens. Thus such a demanding task is very inefficient.

This research is supported in part by National Science Council, Project No. 91-2416-H-008-008 and Project No. 92-2213-E-424-002, Taiwan.

The pervasive connectivity of the Internet and the powerful architecture of the WWW have created a tremendous opportunity for conducting ubiquitous and automatic video surveillance through the Internet. We introduce the design and implementation of a Web-based surveillance system using Internet multicast communications [1][21]. Users can benefit this efficient and economic solution that conventional autonomous surveillance systems can be integrated together for accessibility and management through the prevalent data networks. It provides not only the substantial security functions of automatic motion detection and random video playback, but also the remote control capabilities. Users can retrieve live or pre-recorded surveillance video through wired/wireless Web-devices or PSTN auto-dialup networks.

Besides, our system architecture can be applied further to collaborate with those virtual enterprises, which have the domain expertise on the delicate video processing such as face recognition, moving object tracking, abandoned object identification and etc., to provide value-added services in potential E-marketplaces on the Internet. The architecture of collaborative commerce of surveillance services is shown in Figure 1: The user who requests the surveillance services sits on top of the pyramidal architecture. Diversified surveillance services are put into the bottom of the pyramid and they are divided into core services and value-added services.

The core services include the live/stored surveillance video streaming service and the proposed motion detection service. To effectively and efficiently deliver
the compressed surveillance videos over Internet by the multicast communication without feedback control, we simply apply the application-level traffic shaping method for the surveillance video streaming service. Then, the motion detection service provides the real-time detection of suspicious motion events in the surveillance video.

Multimedia streams such as the surveillance videos are always sent to the network in a compressed format to save the bandwidth. The prevalent video coding scheme to effectively reduce the redundant video information is the motion compensation [14] by coding the difference between two consecutive pictures. Nevertheless, the noise of varying intensity from daylight or fluorescent lamp may confuse the compressed information that if motions appear in consecutive pictures. Our approach for the effective motion detection service acquires neither sophisticated cameras nor complex video processing algorithms [5-9] to provide automatic surveillance functions by filtering the noise in the variation of compressed video frame sizes.

Therefore, while collaborating with the valued-added services of E-Services such as face recognition, moving object tracking, abandoned object detection and etc., the proposed motion detection service can save more bandwidth and storage for the collaborative commerce of surveillance services. Experimental results show that the core services can remit the performance penalty from congested networks and can effectively detect the surveillance events in real-time.

This paper is organized as follows: The system architecture of the collaborative Web-based surveillance services is described in Section II. In Section III, we describe the applied schemes to furnish the core services for supporting the communications of collaborative surveillance services. Experiments and performance results for the applied schemes are presented in Section IV. Section V includes conclusion and future work.

Compressed Video Pumping Server (CVPS): The analog signal of monitoring camera is sent to the input of video capture card at CVPS. CVPS will invoke the installed video codec to compress the input data into coded video frames. Then, CVPS will apply proposed traffic shaping scheme to effectively deliver the coded video frames to the network by multicast connection.

Web Browser Plug-in Client (WBPC): At client site, WBPC is a plug-in software module running with the Web browser. To receive the selected multicasting video streams and display them.

Video Recording and Querying Server (VRQS): VRQS will receive the coded video data from CVPS of the surveillance service. It performs following jobs:
- Save the video frames into multimedia files for the query of playback.
- Apply the proposed motion detection scheme based on received frame sizes regardless of the applied codec to detect the suspicious events.
- Acknowledge security guard’s pager or phone by the autodialing device or other mobile data gateway [10] right after an alarm is set.
- May apply further the value-added services from other video processing techniques (e.g. [9]) by the way as shown in Figure 3 to identify the motion objects while an alarm is set to reduce the redundant processing overhead.

Web Server CGI Scripts (WSCS): WSCS basically provides two following functions:
- Receive interactive commands such as selection of surveillance spots or query of stored videos playback from WBPC user.
- Reply WBPC users the corresponding attributes in WSCS database upon the command requests.

Figure 3 shows the sequence of surveillance video frames processed by VRQS and forwards to value-added services.

2. System Architecture of Collaborative Surveillance Services

Basically, the collaborative surveillance system can be divided into four subsystems as shown in Figure 2:

Core Services

Fig. 2. An overview of the components of the collaborative surveillance services.
Logs of remote user access information such as access location (i.e., IP address), access time or period and selected surveillance spot. While they collaborating with other value-added service, XML or UDDI [13][12] can be used to provide interfaces for data exchanges between organizations or companies.

3. Applied Schemes In Core Surveillance Services

3.1. Surveillance video streaming

The non-discriminated sharing of network resources in Internet makes no guarantee of timely delivery between senders and receivers. While deploying Internet applications of multimedia communications, the performance metrics such as delay, delay jitter, and loss rate of packets should be considered. In the proposed collaborative commerce of surveillance services, the incomplete video pictures may not be able to identify the suspicious moving object or the intruder due to surveillance data packets lost in the network. Therefore, packet loss should be considered more important than the other metrics.

In this paper, the proposed scheme to alleviate the impact of packet loss is based on a codec-level macro-block updating scheme proposed in [15] (they increase codec robustness for video transmission over the Internet) to revise and integrate the previous work of application-level traffic shaping mechanism [20][21]. To reduce the buffering burstiness on the data path from CVPS to the network for delivering video packets, we not only divide the compressed video frames into small PDUs (Protocol Data Unit, i.e. packet), but also revise the previous shaping method since the applied codec using motion estimation may introduce large variations of the packets numbers in different video frames.

3.2. Motion detection in surveillance video

The composite color signal of a video image consists of a luminance component and two chrominance components. These components are much highly compressed by coding the current picture with respected to previous picture. Thus, the degree of motion definitely affects the size of inter-frame that contains different information with previous frame.

In a sequence of video frames producing with frame rate $R$, a motion $I_j$ with time of duration $d_j$ can be defined by a sequence set of inter-frame sizes. The size of the set $I_j$ is $d_j$, called motion length. The extra coding length paid for the motion is $\Delta m$. Thus, a motion can be ideally defined by the following equations:

$$I = \{i_1, i_{i+1}, \ldots, i_{j-1}, i_j, \ldots, i_{\infty}\}, \forall i \in I_j$$

$$d_j = d_i \times R > 0$$

$$\Delta m = i_j - i_{j-1} > 0$$

In Equation (1), without considering the length of header information in coded video frame, the non-motion inter-frame size $i_{j-1}$ is zero because of no extra coding data given for motion. Theoretically, if the motion length $d_i$ is greater than zero, then a motion can be promptly detected to set an alarm.

Nevertheless, the varying intensity of light (e.g. daylight or fluorescent lamp) in the video in which even human vision may not tell the difference also affects the size of video frame. Considering the coding length variation introducing from the zero-mean white noise $X_k$ of unfixed light intensity, the obtained frame size $i_k$ in inter-frame sequence can be redefined as follows:

$$i_k = i_0 + X_k, k > 0$$

Consequently, considering the coding lengths of inter-frames without motion, they (i.e. $i_k = |X_k|$, without header information) will vary due to the white noise. They are hard to be identified as non-motion inter-frames by Equation (1) to detect the coming motion. Therefore, our motion detection scheme proposes to learn a threshold $T$ for the maximum size of non-motion inter-frame in an initial period. It is feasible to force no motion shown in the surveillance video for a short period of the system initialization. Since the white noise of background light may change from time to time, the threshold $T$ will be adjusted by a modified moving average method [17] called selective moving average method to predict the maximum size of non-motion inter-frame.

To reduce the false alarms further, another threshold $D$ for minimum motion length is proposed. $D$ is a threshold for properly prolonging the detection time to prevent false alarms from the interference of noise. For example, two consecutive non-motion inter-frames may set a false alarm if occasionally $|X_k|_i$ is larger than a learned $T$ and less than $|X_k|$. Therefore, as long as the frame rate $R$ is high enough, we can extend the detection time to a minimum threshold $D$ of motion length without incurring miss alarms and degrading the response time for setting an alarm. The value of $D$ can be obtained by estimating minimum motion duration for any possible motion in the video scene of surveillance. Thus, we can summarily define a motion $I_j$ in our motion detection scheme as follows:

$$I = \{i_1, i_{i+1}, \ldots, i_{j-1}, i_j, \ldots, i_{\infty}\}, \forall i \in I_j$$

$$d_j > D$$

Note that, the two-tier thresholds of $T$ and $D$ play key roles for the precision in the proposed motion detection algorithm. $T$ is an adaptive threshold to predict the maximum size of non-motion inter-frame to help to pick up candidate inter-frames that may have motion. On the other hand, $D$ is a minimum threshold for defining a motion to defend the precision of motion detection. It
helps to prevent false alarms from the possibly over-optimistic value of $T$. Therefore, our motion detection algorithm can provide an effect and efficient mechanism for the automatic surveillance system to achieve relatively low false alarm without complex video processing overheads. Evaluation results are shown in subsection B of next section.

---

**4. Experiments and Performance Results**

A series of experiments are conducted on the test-bed as shown in Figure 4 to investigate the performance of the proposed surveillance service. The captured video format is CIF (352x288 pixels) H.261 in our experiments. In the test-bed, we elaborate the experiments for narrowband Internet across an MBone router (maximum bandwidth of the tunnel is up to 128kbps). In the following subsections, we will examine representative experimental results from the traffic shaping scheme and the motion detection method, respectively.

### 4.1. Reducing frame loss by traffic shaping

In our experiments, we propose two testing strategies to validate the performance of enhanced traffic shaping method. In strategy A, each segmented video packet is of 512 bytes long. In strategy B, each video packet is of 128 bytes long. Because of the system capability in the experiments, each CVPS machine will maintain its output rate to 5 frames per second ($i.e. R = 5$) for a H.261 compressed stream. In the traffic shaping method, the inter-burst interval is adaptive ($i.e. 1/(pR)$) rather than a constant and its value will depends on the processing time of segmentation overhead and the system function call for sending out the packet in Windows system.

In the experiments, each surveillance spot will contribute more than 32 Kbits/s traffic. If the number of total surveillance spots is larger than 4, the MBone tunnel with a bandwidth limitation of 128 Kbps would not be able to sustain such overloaded traffic. Thus, the transmitted packets will be lost in the network. Therefore, we simply evaluate the performance of proposed strategies by deploying this case.

As shown in Figure 5, after we apply strategies A and B in the traffic shaping method, the packet losses are all decreased without sacrificing the frame rate. The application-level traffic shaping method controls the time distance within the sending packets to reduce the burst arrival in the buffers along the path to the receivers. Then, it effectively decreases the chance of buffer flooding to reduce packet loss without affecting the video frame rate.

---

**4.2. Detecting motion with low false alarms**

In the experiment, the video frame rates range from 2 to 5 frames per second ($i.e. 2 < R <= 5$) due to different system capabilities in the installed machines. The maximum threshold $T$ of inter-frame size is learned from a silent period $S_{1,32}$ ($i.e. N = 32$). So the learning period $N$ will roughly take a short time in 6 to 16 seconds. Then, $T$ will be adjusted by moving average method with a small window size 8 ($i.e. W = 8$) to keep track of the latest average size of non-motion inter-frames.

In the experiments, the value of $D$ is set to a small number of 2. It indicates that the surveillance system can detect the suspicious event in which the motion lasts more than 0.4 to 1 second according to the frame rates in different surveillance videos. The video length of each surveillance spots in the experiments is two hours. Then, we review the video files to find out the true alarms to examine the performance. The experimental results of the proposed adaptive motion detection algorithm are shown in Table 1. They show low false alarm rate in the proposed automatic surveillance system.

<table>
<thead>
<tr>
<th>Camera Location</th>
<th>Set Alarms</th>
<th>False Alarms</th>
<th>Test Time Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 elevator</td>
<td>10</td>
<td>5</td>
<td>1700:1859</td>
</tr>
<tr>
<td>B1 garage</td>
<td>13</td>
<td>0</td>
<td>1700:1859</td>
</tr>
<tr>
<td>Eastern entrance</td>
<td>33</td>
<td>0</td>
<td>2100:2259</td>
</tr>
<tr>
<td>Lobby</td>
<td>32</td>
<td>0</td>
<td>2100:2259</td>
</tr>
<tr>
<td>West side entrance</td>
<td>0</td>
<td>0</td>
<td>1700:1859</td>
</tr>
<tr>
<td>IF cargo elevator</td>
<td>1</td>
<td>0</td>
<td>1700:1859</td>
</tr>
</tbody>
</table>

Though one surveillance spot shows high false alarm ratio, it is because the background light is too dim to
have high resolution in the surveillance video. Such a surveillance video with low resolution degrades the degree of accuracy to predict the threshold $T$. Then, it is hard to prevent false alarms for small $D$. However, if the resolution of this surveillance video can be enhanced (i.e. install illuminant source such as inexpensive infrared equipment), the false alarms can be reduced further because the motion will affect much larger increment of inter-frame size than white noise does. Therefore, according to the experimental results for the motion detection scheme in our surveillance system, we have well-predicted $T$ with small $D$ by the selective moving average method to achieve much low false alarm ratio without introducing miss alarms.

5. Conclusions and Future Work

In this paper, we described an effective communication model for collaborative commerce of Web-based surveillance services. We summarize the advantages of this proposed model as follows:

- **Efficient and economic solution model to integrate conventional autonomous surveillance systems for accessibility and management through prevalent data networks.**
- **Ubiquitous E-Services for users such as selecting surveillance videos among different surveillance spots through wire/wireless Web devices (B2C) and automatic notification of suspicious events via auto-dialing system (B2B).**
- **Deploying one-to-many multicast transmission model for Web-based surveillance systems presents the benefit to save workload of network and surveillance video servers.**
- **Without affecting the video frame rate, proposed application-level traffic shaping mechanism smooth the burst of sending packets to network to improve the quality of service in surveillance video streaming due to the packet loss is reduced.**
- **Low-complexity motion detection algorithm not only efficiently detects suspicious motion events in the surveillance videos using the prevalent compression scheme of motion estimation, but also economically reduces the storage and network bandwidth in this collaborative commerce model.**

The experiments on H.261 video frames show good results of low false alarm ratio for the proposed motion detection method. However, the ideal goal of motion detection is to achieve virtually zero false alarm for all kinds of surveillance video scenes. To explore and elaborate the optimal noise eliminating function with the low complexity and virtually zero false alarms without miss ratio is our technical goal in the near future.

Besides, some surveillance videos are sensitive to the privacy. We will need to collaborate with other value-added services such as watermarking or encryptions to effectively protect the surveillance videos for supporting collaborative commerce of surveillance services.

Acknowledgements

The authors would like to thank Prof. Jie-Yong Juang and Prof. Shun-Chin Hsu for their valuable comments.

References