HIDING PRIVACY INFORMATION IN VIDEO SURVEILLANCE SYSTEM

Wei Zhang[1], Sen-ching S. Cheung[1], and Minghua Chen[2]

[1]ECE Department, University of Kentucky,
Lexington, KY 40508
{wzhan4,cheung}@engr.uky.edu

[2]EECS Department, UC Berkeley,
Berkeley, CA 94720
minghua@eecs.berkeley.edu

ABSTRACT

This paper proposes a detailed framework of storing privacy information in surveillance video as a watermark. Authorized personnel is not only removed from the surveillance video as in [1] but also embedded into the video itself, which can only be retrieved with a secrete key. A perceptual-model-based compressed domain video watermarking scheme is proposed to deal with the huge payload problem in the proposed surveillance system. A signature is also embedded into the header of the video as in [2] for authentication. Simulation results have shown that the proposed algorithm can embed all the privacy information into the video without affecting its visual quality. As a result, the proposed video surveillance system can monitor the unauthorized persons in a restricted environment, protect the privacy of the authorized persons but, at the same time, allow the privacy information to be revealed in a secure and reliable way.

1. INTRODUCTION

The video surveillance system has been widely deployed in many circumstances. While it can monitor improper activities in an environment, it also exposes the privacy of innocent people. Although such a concern may not be significant in public spaces such as metro stations, airports or supermarket, employees who work long hours in more confined areas such as office buildings and hospitals may feel their privacy being violated if they are continuously being monitored. In such situations, it is desirable to have a surveillance system that can selectively disregard regular employees and capture only the unauthorized. In [1], Wickramasuriya and Datt proposed such a privacy protecting video surveillance system. It utilized the RFID sensor to identify the authority of the incoming individual, combined this information with an XML-based framework for access control to determine the violations within the space, and finally used certain video masking technique to selectively display the unauthorized people in the video. Although this system in some degree fulfils the privacy-protecting goal, it has a potential security flaw because that it did not keep a record of the privacy information. In the case when authorized personnel were also involved in some maleficent or even criminal behaviors, the surveillance system should have the ability to provide the original surveillance footage when necessary. In this sense the system should provide multiple levels of access to the surveillance video: the general public can only see the unauthorized behavior, while special authorities who possess a secret key, such as the police, can observe all the information happened in the monitored space. The simplest solution to this problem is to store separately a copy of the original surveillance video. Nevertheless, the presence of a separate copy becomes an easy target for illegal tempering and removal, making it very challenging to maintain the security and integrity of the entire system. In this paper, we present a solution to this problem that utilizes watermarking techniques to hide privacy information in the surveillance video and uses digital signature for authentication of the watermarked video. By using a visual perceptual model, the proposed watermarking algorithm can embed all the privacy information in the video with minimal visual distortion. The digital signature is then placed in the ancillary data fields of the compressed video, and thus can detect any kind of changes of the watermarked video. Our proposed solution consummates the original privacy-protecting surveillance concept in [1] and makes the resulting system more reasonable and practical.

This paper is organized as follows: the architecture of a privacy-protecting surveillance system is proposed in Section 2. In Section 3, the challenges of hiding privacy information in the video are discussed and a new video watermarking scheme is proposed. Experimental results on the proposed method are presented in Section 4. Some conclusion remarks are made in Section 5.

2. SYSTEM FRAMEWORK

As discussed above, a desirable video surveillance system should mask the privacy information to the general public, and at the same time be able to provide all the information to the authorities with special security clearances. In this paper, the privacy protecting surveillance system is considered to be installed in a confined area, such as a

0-7803-9134-9/05/$20.00 ©2005 IEEE

hospital or a company building, instead of public spaces such as metro station, supermarket, etc. The detailed definition and requirements of such a surveillance system are discussed in this section.

Suppose all the people that the camera captures in the confined space can be categorized into two groups: the authorized personnel and unauthorized personnel. A privacy protecting surveillance system should meet the following requirements:

1. Record everything in the area.
2. A person with a lower level of security clearance can only see the unauthorized behavior, and all the privacy information must be entirely unperceivable to him/her.
3. Under special conditions, a person with a higher level of security clearance can use a secret key to view the original video, including the privacy information if any.
4. Without the proper secret key, no one can reconstruct the privacy information from the resulting video, even in the extreme case that the exact algorithm of the surveillance system is known to the attacker.
5. The system should provide a reliable way to prove the authenticity and integrity of the resulting video in which the privacy information has been embedded.

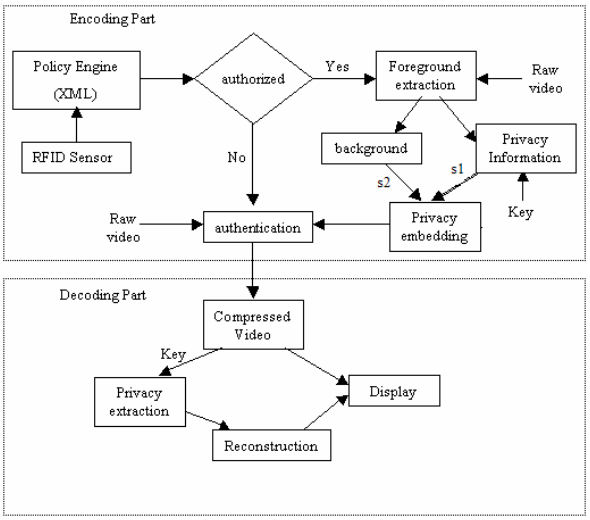

Fig 1. Framework of the Privacy-protecting surveillance system

A surveillance system that meets the first and second requirements has been reported in [1]. In this paper, we describe a system that meets all the above five requirements. The diagram of the proposed system, which is an extension to the one described in [1], is shown in Figure 1.

By using an identity sensor (e.g. RFID), the policy engine first determines the authority of the incoming person. If an authorized person is detected, the system will first extract the corresponding foreground object in each frame, and then compress and encrypt this privacy information as a regular video bitstream, denoted as s1. The void in the video left behind the foreground objects is filled in by the background model. The resulting video is then compressed into another bitstream s2, with s1 being embedded as a watermark by a privacy-embedding scheme. To authenticate s2 against any form of tampering, we borrow the idea from [2] to compute a signature of the watermarked video and embed it into the header of the video. In this way, any changes can be detected by our system, which meets the fifth requirement described earlier. The decoder is designed to combine the privacy information with the video only after proper authentication and the presence of the secret key.

3. HIDING PRIVACY INFORMATION

The major challenge of the above proposed system is that how to imperceivably hide the large amount of privacy information in the video. It is natural to think of using watermarking techniques in this context. However, most current video watermarking techniques are designed for authentication or copyright protection instead of hiding information. Thus, most current video watermarking system can only embed a few bytes of a company logo or other types of identification information. On the other hand, the size of the privacy information in our application is far beyond several bytes even after compression (usually more than 3000 bits per 352x288 frame). Obviously, the perceptual quality after embedding such a large payload is of great concern in this situation. In [3], a similar research has been performed on still images. The authors proposed to hide a large volume of information into the nonzero DCT terms after quantization. Although the method is based on the human visual system (HVS), it did not formally use any perceptual model and thus could not guarantee a minimal perceptual distortion. In addition, surveillance videos do not typically have much foreground activities. Thus, the compressed bitstream is likely to contain many zeros in the quantized DCT domain, making the above approach difficult to accommodate such a huge payload. In the sequel, we will propose two new high-capacity video watermarking algorithms that are based on a DCT-based perceptual model.

Due to the popularity of Discrete Cosine Transform (DCT) in many video compression standards, we adopt the DCT perceptual model described in [4]. Consider the luminance and contrast masking of HVS as described in [4], the final perceptual mask $s[i, j, k]$ that indicates the maximum permissible alternation to the $(i, j)^{th}$ coefficient of the k^{th} 8×8 DCT block of an image can be calculated as:

$$s[i,j,k] = \max\{t_L(i,j,k), |C[i,j,k]|^w \cdot t_L(i,j,k)^{1-w}\} \quad (1)$$

$$\text{with } t_L(i,j,k) = t(i,j)\left(\frac{C[0,0,k]}{C_{0,0}}\right)^{\alpha_T}, \; i,j \in \{0,1,\cdots,7\} \quad (2)$$

where $t(i,j)$ is the frequency sensitivity threshold, $w = 0.7$ is a constant, $C[0,0,k]$ is the DC term of block k, $\alpha_T = 0.649$ is a constant, and $C_{0,0}$ is the average luminance of the image. With this perceptual model, we can compute a perceptual mask value for each DCT coefficient in the current frame.

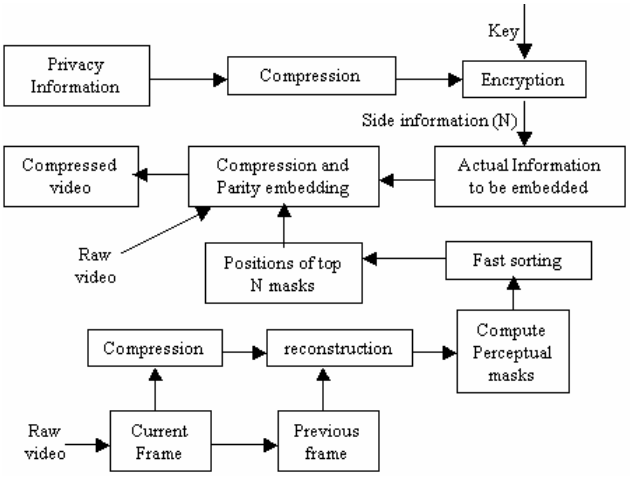

Fig 2: Embedding Method A

We assume the number of bits N of the privacy information for each frame is less than the total number of DCT coefficients. If we can embed at most one bit in each coefficient, the best places to hide the information according to the above perceptual model are the DCT terms with the highest mask values according to (1). Based on this idea, we develop our first embedding algorithm, denoted as Method A, which is illustrated in Fig. 2. The privacy information is first compressed and encrypted. The raw video with the privacy information removed is also compressed and the previous reconstructed frame is then used to estimate the perceptual mask of every DCT coefficient in the current frame based on (1). The watermark is then stored, using parity embedding, in the DCT coefficients of the *residual frame* that correspond to the N coefficients that are the least sensitive according to the computed perceptual masks. For this embedding method, a blind decoder needs to know the value of N before decoding. We can insert it into the header of the video or embed it also as a watermark in some fixed positions.

Method A minimizes the perceptual distortion by spreading the watermark bits across the entire frame. As surveillance video is typically quite static, many DCT blocks do not have any non-zero coefficients. Embedding watermark into these zero blocks may significantly increase the bitrate due to the long runlength codewords used for these rare DCT patterns. An easy way to solve this problem is to use as few blocks as possible. However, the fewer blocks used for embedding, the more visible the watermarks will likely to be. In order to explore the tradeoff between the bitrate and the perceptual quality, we consider an alternative approach, Method B, in which the embedding positions are chosen block-wise rather than coefficient-wise in the quantized DCT domain. In Method B, we embed the same number of watermark bits in the blocks that have the highest DC coefficients based on luminance masking. Let '*1/ratio*' be the percentage of blocks being used for embedding. Then a bigger *ratio* indicates fewer blocks are used for embedding, and thus more bits are embedded in those blocks. Watermark bits within the same DCT block are embedded in a zig-zag fashion starting from the highest frequency coefficients.

5. EXPERIMENTS

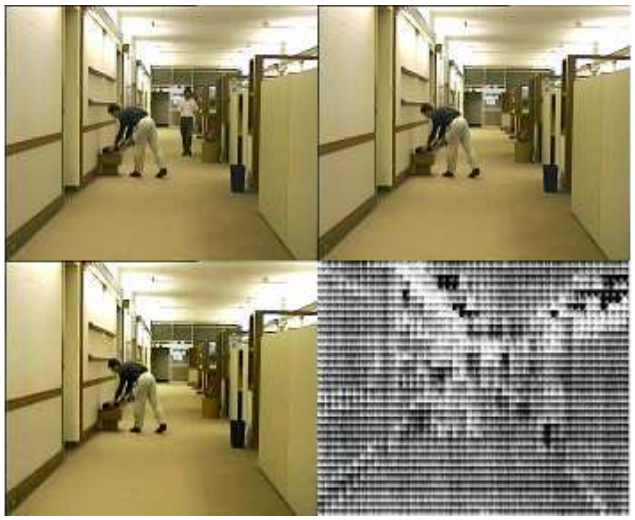

Fig. 3 Experimental results. Top-left: Original video; Top-right: Uncompressed mask video; Bottom left: Mask video, compressed w/ watermark. Privacy information size: 112 Kbytes (3060 bits per 352x288 frame). QP is 10 and fixed. Bottom right: Perceptual masks of this frame (enhanced for display).

Since this paper is mainly focused on hiding privacy information in the video surveillance system, we will only present experimental results relevant to this part of the system. H.263 version 2 is used to compress all the video sequences in our experiments. We have tested our algorithm on the "hall monitor" sequence that contains two persons in an office setting. The sequence has 300 frames and the frame size is 352x288.

To extract moving foreground objects, we compare each frame with the first frame of the sequence, which has no moving objects. Pixels that are substantially different are classified as foreground. Foreground objects are

created based on connected-component grouping on the foreground pixels, expanded to form bounding boxes, and finally tracked based on the overlap of objects from frame to frame. Suppose the person on the right in the top left image of Fig. 3 is classified as authorized personnel and thus needs to be masked out. To provide the appropriate masking background that is visually compatible with the surrounding, we construct it using an adaptive background model based on a Kalman filter [5], paste it over the extracted foreground, and apply a H.263-style de-blocking filter to smooth the transition between the fill-in background and the rest of the image. The resulting image is shown in the top right corner of Fig. 3. The extracted foreground frames are then treated as another video, compressed, encrypted, and embedded into the reconstructed video using Method A. The bottom left of Fig.3 shows the same frame reconstructed from the compressed bitstream with watermark. The Quantization Parameter (QP) is set at 10 and around 3060 bits of watermark is embedded in this frame. As shown in the perceptual masks located at bottom right of Fig. 3, the watermark is embedded to the bright or visually complex places, such as the bright area on the ceiling and wall-wall joints, where the watermarks can be successfully masked.

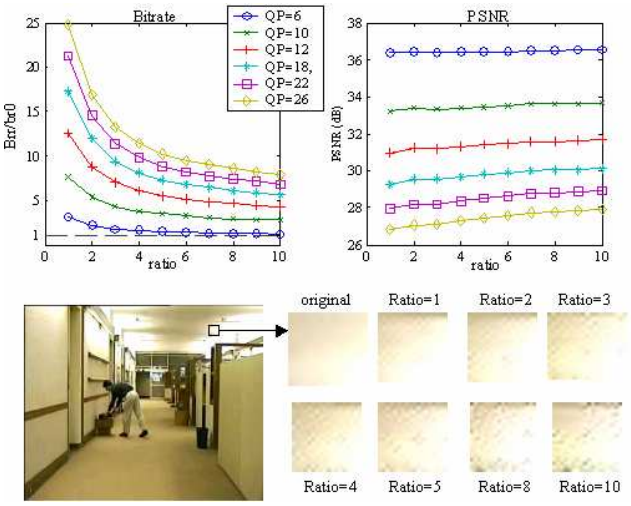

Fig.4: Tradeoff between bitrate and perceptual quality. Top left: brr/br0 vs. ratio for different QP's; top right: PSNR vs. ratio for different QP's; bottom: visibility of watermarks at different ratios, where the QP is fixed at 10.

While Method A produces excellent reconstructed video quality, the bitrate increases from 119 kbps of the unwatermarked case to 922 kbps after inserting the watermark. To study the tradeoff between bitrate and reconstructed video quality, we run Method B using different ratio values, where the reciprocal of the ratio determines the percentage of DCT blocks used for embedding. For QP = 10, we observe that a ratio of 3 can reduce the bitrate to 628 kbps while keeping a similar visual quality as Method A. A higher ratio can further reduce the bitrate, but produce obvious artifacts as shown in the bottom part of Fig. 4. Let br0 and brr be the average bitrates before and after watermarking. In the top-left corner of Fig. 4, we plot the relative increase in bitrate, brr/br0, as a function of the ratio values. For high QP, there is a dramatic drop in both brr/br0 and visual quality as we increase the ratio values. This coincides with what we anticipate when reducing the number of blocks used for embedding. When we lower the QP, the effect becomes less dramatic as there are fewer zero blocks. The PSNR plots at the top right corner of Fig. 4 does not show any dependency on the ratio values as we modify as most one bit per coefficient. Thus, PSNR is inappropriate to measure visual quality in our application. With the above relationships in mind, we can use either Method A to get minimal perceptual distortion if storage is not the major concern or Method B with an appropriate ratio to obtain a good compromise between bitrate and perceptual quality.

6. CONCLUSION

In this paper, we have presented a privacy-preserving surveillance video system where the privacy information can be retrieved in a secure and reliable manner. An effective information hiding method has been proposed to hide all the privacy information into the host video with minimal perceptual distortion. A digital signature is inserted into the resulting video for authentication. Future research will be focused on minimizing the bitrate given the amount of privacy information and perceptual distortion requirement.

7. REFERENCES

[1] J. Wickramasuriya, M. Datt, S. Mehrotra and N. Venkatasubramanian, "Privacy Protecting Data Collection in Media Spaces," *ACM International Conference on Multimedia*, New York, NY, Oct. 2004

[2] M. Pramateftakis, T.Oelbaum, K.Diepold, "Authentication of MPEG-4 based Surveillance Video," *International Conference on Image Processing (ICIP)*, Singapore, 2004.

[3] K. Solanki, N. Jacobsen, C. Chandrasekaran, U. Madhow and B. S. Manjunath, "High-Volume Data Hiding in Images: Introducing Perceptual Criteria into Quantization Based Embedding," *Proc. 2002 IEEE Military Communications Conference (Milcom 2002)*, Anaheim, CA, October 2002.

[4] Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom, "Digital watermarking", *Morgan Kaufmann Publishers*, 2002

[5] K.-P. Karmann and A. Brandt, "Moving object recognition using and adaptive background memory,"in *Time-Varying Image Processing and Moving Object Recognition*, V. Cappellini, ed., 2, pp. 289-307, Elsevier Science Publishers B.V., 1990.