2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance

An Authentication Mechanism Using Chinese Remainder Theorem for Efficient Surveillance Video Transmission

Tony Thomas*, Sabu Emmanuel, Peng Zhang
School of Computer Engineering
Nanyang Technological University, Singapore
{ttony,asemmanuel,zh0036ng}@ntu.edu.sg

Mohan S. Kankanhalli
School of Computing
National University of Singapore, Singapore
mohan@comp.nus.edu.sg

Abstract

Now-a-days, surveillance cameras have been widely deployed in various security applications. In many surveillance applications, the background changes very slowly and the foreground objects occupy only a relatively small portion of a video frame. In these type of applications, an efficient solution for transmissions over bandwidth-limited networks is to send only the foreground objects for every frame in real time while the background is sent occasionally. At the receiving end of the transmission, the objects and the most recent background can be fused together and the original frame can be reconstructed. However, protecting the authenticity of the video becomes more challenging in this case as a malicious entity can modify/replace/remove the individual foreground objects and background in the video. In this paper, we propose a Chinese remainder theorem based watermarking mechanism for protecting the authenticity of videos transmitted or stored as objects and background. Our mechanism ensures the authenticity between video objects and their associated background.

1. INTRODUCTION

Now-a-days, surveillance cameras have been widely used in public places, work places and homes in connection with detection and quick resolution of crimes, traffic accidents/violations and other incidents. In this type of applications, the images are captured over long periods of time and thus the volume of data generated is enormous. Hence transmission of such videos over bandwidth limited networks and its storage can be a bottleneck in many applications. Object based video segmentation mechanisms can be used to segment a video frame into background and foreground objects [3, 12, 14, 18]. In many surveillance applications, the background changes very slowly, and the ratio of the frame size to the total area occupied by objects in the frame is small. In these types of videos, object segmentation gives an efficient solution for transmission of the video over bandwidth limited networks by sending only the foreground objects frame by frame in real time and the background once over a relatively long time interval.

However, sending a video as background and objects pose new challenges to the security of the video, since the video objects and background can be easily accessed, modified, or replaced by another object or background. Hence there is a need for a secure mechanism for protecting the authenticity of videos transmitted as objects and background as the video could be used as legal testimony in a court later.

In this paper, we propose a Chinese remainder theorem (CRT) [6] based authentication mechanism which can be used for surveillance video transmission with background subtraction. To the best of our knowledge, this problem has not been addressed in the literature yet. The proposed mechanism carries over to the case of efficient storage in a straightforward manner. We establish the authenticity of each frame by embedding a unique joint watermark on each foreground object that is transmitted. Watermark information are computed for each object present in the frame as well as the corresponding background which was filtered out. Then a joint watermark information is computed from them using CRT. Finally a watermark is generated from this joint watermark information and is embedded on each of the object that is transmitted. We use CRT as it has an intrinsic lock property [4] and thus all the foreground objects and background information can be locked as a single watermark. The receiver can unlock the watermark and get information about the objects and the background. The joint watermark can be used to establish the existence of objects in a frame and their association with the background. Thus, our authentication mechanism ensures that the video file has not been falsely manipulated and the foreground objects can be proved to be part of the original video content.

*Thanks to the Agency for Science, Technology and Research (A*STAR), Singapore for supporting this work under the project 'Digital Rights Violation Detection for Digital Asset Management' (Project No: 0721010022).

978-0-7695-4264-5/10 $26.00 © 2010 IEEE
DOI 10.1109/AVSS.2010.13

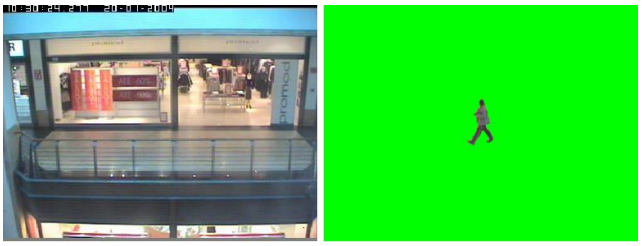

Figure 1. Shopping Center: Background and Object

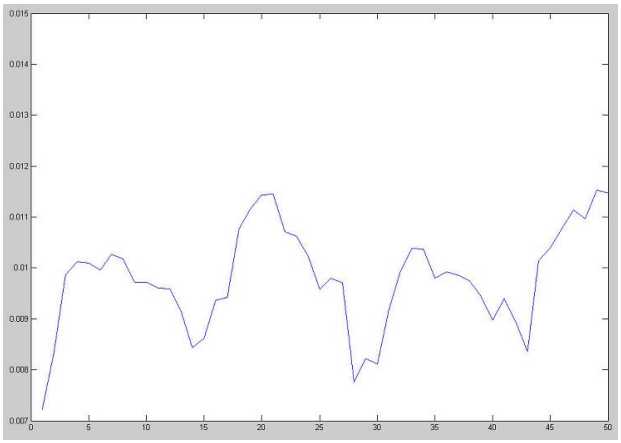

Figure 2. Shopping Center: Ratio of Objects to Frame Size

Figure 3. Street: Background and Objects

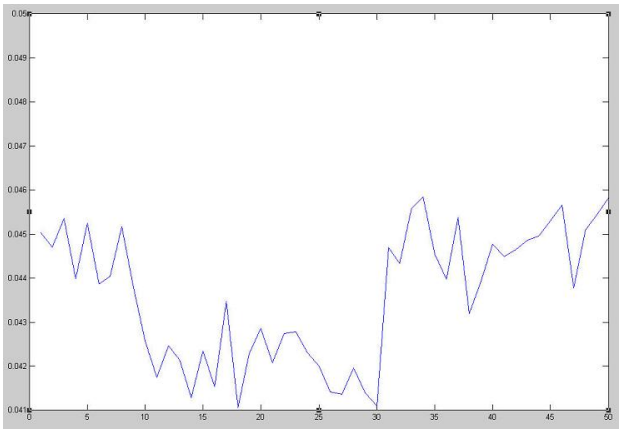

Figure 4. Street: Ratio of Objects to Frame Size

The remaining part of the paper is organized as follows. A discussion on video segmentation and estimates of possible bandwidth savings by transmitting only the foreground objects instead of complete frames are given in Section 2. The proposed authentication mechanism is described in detail in Section 3 and Section 4. The security analysis is carried out in Section 5. Finally, the paper concludes with some observations and future directions in Section 6.

2. Video Segmentation and Analysis

Any frame of a video is a collection of a background and a set of foreground objects. Video segmentation is a process of extraction of the background and the foreground objects from a frame. There exist many algorithms which can efficiently extract out the individual foreground objects from the background in a video frame. These algorithms can be classified into three categories: image segmentation based, motion based and change-detection based [9]. Image segmentation based algorithms first segment a frame into homogeneous regions by spatial similarity and then merge these areas by other criteria [14]. Motion based algorithms treat object as areas with coherent motion and find regions with coherent motion by grouping dense motion vector [18]. Change-detection based algorithms find difference areas temporally that are sensitive to human eyes [3, 19].

In many surveillance video applications, the background remains static or changes slowly or changes with a fixed pattern known a priori. These characteristics of the surveillance videos have been used to simplify, reduce the computational complexity and improve the accuracy of segmentation algorithms [19]. Many of these algorithms use background subtraction technique for video segmentation. In a changing background scenario, a background subtraction method based on adaptive modeling of the background using a proper learning algorithm is used. This method updates the current background according to the real-time change [11].

We now estimate the possible bandwidth saving by transmitting only the foreground objects instead of complete frames. We performed segmentation on two test surveillance videos using the segmentation algorithm given in [12]. The first experiment was performed on the video 'Shopping Center in Portugal' [24] given in Fig. 1. The second experiment was performed on the video 'AVSS PV Medium' [25] given in Fig. 3. The segmentation was performed on 50 consecutive frames of the videos. The ratio of the size of the foreground objects to the total frame size is computed for each frame and are given Fig. 2 and Fig. 4. The x-axis denotes the frame number and the y-axis denotes this ratio. Fig. 2 from the first experiment (one object) shows that the foreground object occupies only around 1% of a frame and thus the bandwidth saving in this case could be close to 99%. Fig. 4 from the second experiment (multi-

ple objects) shows that the foreground objects occupy only around 4.5% of a frame and thus the bandwidth saving in this case could be close to 95.5%.

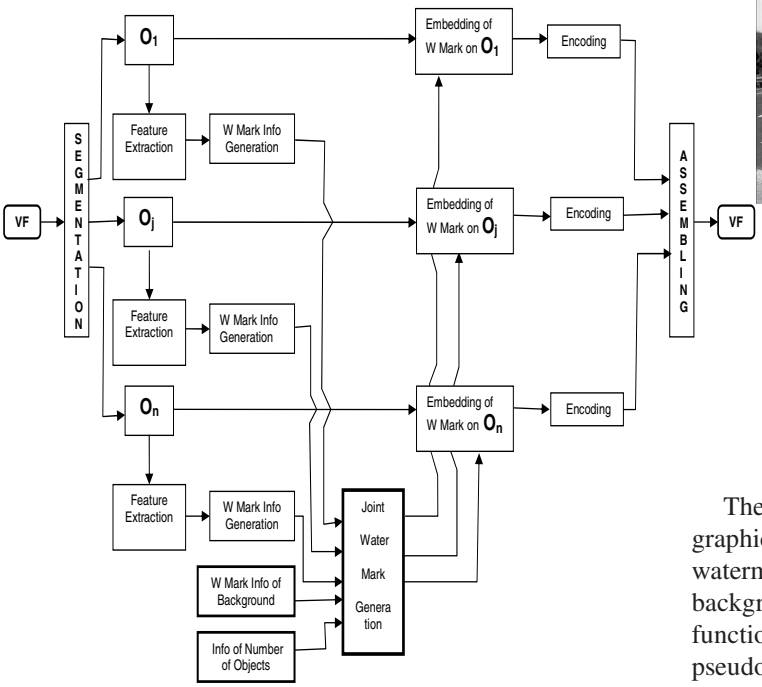

Figure 5. Framework at Sender's Side

3. Proposed Framework at the Sender

In this section, we describe the proposed framework at the sender's side. The sender can be identified with the camera. The framework is illustrated in Fig 5. The details are as follows.

3.1. Assumptions, Initial Setup and Segmentation

We assume that V is a surveillance video such that the background is static or changes very slowly, and the ratio of the total number of pixels occupied by the foreground objects to the total number of pixels in the frame is small. In other words the foreground objects occupy only a relatively small portion of each frame of the video V.

Since the background is either static or changes very slowly, it is updated only at the end of large intervals. So without loss of generality we can assume that the background B is fixed. The background B is initially transmitted without any foreground objects present in it as shown in Fig 6. When the camera starts the transmission, the background B is subtracted from each frame of the video and only the foreground objects are send in real time as shown in Fig 7.

Figure 6. Background without any Objects

Figure 7. Objects without Background

The sender and receiver agree on two secret cryptographic keys K_{mac} and K_{rnd}. K_{mac} is used to generate watermark information of the foreground objects and the background from their features using a cryptographic MAC function $MAC(,)$ [22]. K_{rnd} is used to generate a secure pseudo-random number sequence using a cryptographically secure pseudo-random number generator $PRNG(,)$ [23]. Further, the sender and receiver agree on a set of coprime integers $N_0, N_B, N_1, N_2, \ldots$.

The i-th frame V_i of the video is segmented using a segmentation algorithm with background subtraction and n_i foreground objects $O_i^1, \ldots, O_i^{n_i}$ are extracted out as described in Section 2.

3.2. Chinese Remainder Theorem (CRT)

Let $n_1, \ldots, n_k$ be pairwise coprime positive integers and $r_1, \ldots, r_k$ be any collection of integers. Then the k congruences

$$x \equiv r_i \mod n_i, \text{ for } 1 \leq i \leq k,$$

has a unique solution x such that $0 \leq x < N = n_1 \ldots n_k$. CRT has been used in many security related applications. Some such applications are digital watermarking protocols for the multiparty multilevel DRM architectures [20, 21], watermarking schemes for image authentication [16, 17], a secure broadcast communication scheme [4], a robust t-out-of-n oblivious transfer protocol [2] and a key distribution scheme for conditional access system in digital TV broadcast [10].

3.3. Feature Extraction and Watermark Information Generation

Features are extracted from each foreground object O_i^j in V_i as pixel values at random positions as follows. Let

the pixel values of O_i^j are $p_{i,1}^j, p_{i,2}^j, \ldots$ obtained after raster scanning of the object from top to bottom and left to right. Let $r_1, r_2, \ldots$ be a pseudo-random bit sequence (0 and 1) generated using $PRNG()$ with the key K_{rnd}. Let $\mathbf{N}'$ be the subset of the set of positive integers such that $k \in N'$ if and only if $r_k = 0$. Let $Pval_i^j$ denote the concatenation of all pixel values $p_k(O_i^j)$ (including the concatenation RGB components in it) where $k \in \mathbf{N}'$. Let $Cord_i^j$ denotes the concatenation of a set of coordinates to describe the location of the object in the frame. Now, the watermark information W_i^j of the object O_i^j is computed as

$$W_i^j = MAC(Pval_i^j || Cord_i^j, K_{mac}).$$

The watermark information W_B of the background B is computed as

$$W_B = MAC(Pval_B, K_{mac}),$$

where $Pval_B$ is the concatenation of the random pixel values of B computed as in the case of the object O_i^j.

3.4. Joint Watermark Generation and Embedding

Let $N_0, N_B, N_1, N_2, \ldots$ be a collection of relatively prime integers shared between the sender and the receiver. Then the joint watermark information W_i for the frame V_i (containing the background B and n_i foreground objects $O_i^1, \ldots, O_i^{n_i}$) is computed as the solution of the following set of $n_i + 2$ congruences:

$$W_i \equiv n_i \mod N_0; \quad (1)$$

$$W_i \equiv W_B \mod N_B; \quad (2)$$

$$W_i \equiv W_i^j \mod N_i^j, \text{ where } j = 1, \ldots, n_i. \quad (3)$$

The existence and uniqueness of W_i is guaranteed by CRT.

For the simplicity of the discussion, we confine to watermarking in the spatial domain using the spread spectrum watermarking algorithm of Hartung and Girod [8]. The proposed method can be easily modified for watermarking in the transform domain with other watermarking mechanisms. The watermarking on the object O_i^j is carried out as follows.

Let W_i be expressed in the binary form as $W_i = w_i^1 w_i^2 \ldots$ and for $j > 0$, a_j denotes the watermark bits such that

$$a_j = \begin{cases} 1 & \text{if } w_i^j = 1, \\ -1 & \text{if } w_i^j = 0. \end{cases} \quad (4)$$

Now the discrete signal $\{a_j\}_{j>0}$ is spread by the chip-rate cr to obtain the spread sequence

$$b_k = a_j, \text{ where } j.cr \leq k \leq (j+1).cr, \ k \in \mathbf{N}.$$

The spread sequence b_k is amplified with locally adjustable amplitude factor $\alpha_k \geq 0$ and is then modulated by a binary pseudo-noise sequence

$$pn_k, \text{ where } pn_k \in \{-1, 1\}, \ k \in \mathbf{N}.$$

Now the spread spectrum watermark is obtained as

$$wm_k = \alpha_k . b_k . pn_k, \text{ for all } i \in \mathbf{N}.$$

wm_i is then embedded into the object O_i^j using the equations,

$$\widehat{p_{i,k}^j} = p_{i,k}^j + wm_k, \text{ for all } k \in \mathbf{N} \setminus \mathbf{N}'.$$

yielding the watermarked object $\widehat{O}_i^j$. Note that the pixel positions ($\mathbf{N} \setminus \mathbf{N}'$) on which the watermark is embedded is disjoint withe the pixel positions ($\mathbf{N}'$) from which the watermark information was generated. This is to ensure that the receiver can compute the watermark information W_i^j from $\widehat{O}_i^j$.

3.5. Protocol at the Sender's Side

The protocol at the camera is as follows. Initially the background B without any foreground object is captured. The watermark information of the background W_B is generated as in the case of the foreground objects described in Section 3.3. There is no joint watermark generation in this case. The watermark signal is generated from W_B and then embedded into B in a similar manner as in the case of object O_i^j described in Section 3.4. The watermarked background $\widehat{B}$ is then transmitted after encoding. The foreground objects in each frame are then transmitted in real time as follows. We describe the protocol for the i-th frame V_i of the video, where $i \in \mathbf{N}$.

1. V_i is segmented as described in Section 3.1 and n_i foreground objects $O_i^1, \ldots, O_i^{n_i}$ are extracted out [12].

2. For each object O_i^j, the watermark information W_i^j is computed as described in Section 3.3.

3. The joint watermark information W_i is computed and the corresponding watermark is embedded on each object O_i^j as described in Section 3.4.

4. The watermarked objects $\widehat{O}_i^1, \ldots, \widehat{O}_i^{n_i}$ are encoded and assembled as a frame $\widehat{V}_i$ and is transmitted.

3.6. Efficiency of Transmission

The major computations involved at the sender's side are computations of $MAC(,)$ in Section 3.3. The $MAC(,)$ for each object can be computed in parallel. MAC is used in IPSec and SSL protocols and there are now very efficient hardware implementations of MAC/HAMC functions

with throughput of atleast 1600 Mbps [15]. The other operations are computation of one CRT and watermark generation and embedding using Hartung's spread spectrum algorithm. CRT can be computed very efficiently [20]. Hartung's watermarking algorithm is efficient and can be used in real-time video transmission. Thus, we do not expect any substantial slow down in real-time video transmission due to the proposed authentication mechanism.

The transmission can be made more efficient by watermarking only a selected set of frames of the video.

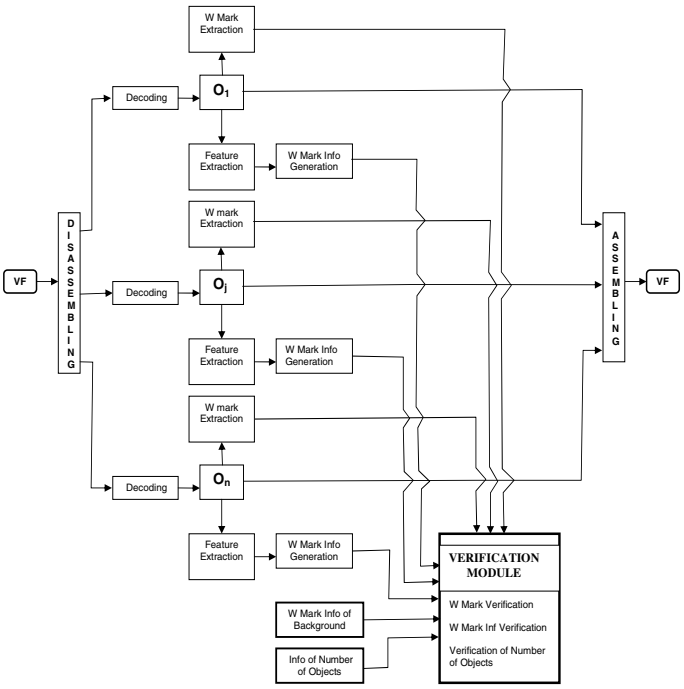

Figure 8. Framework at Receiver's Side

4. Proposed Framework at the Receiver

The framework of the system at the receiver's side is given in Fig 8. We describe the steps performed by the receiver for the i-th frame $\widehat{V}_i$ of the video. The receiver first extracts out the watermarked objects $\widehat{O}_i^1, \ldots, \widehat{O}_i^{n_i}$ from the frame $\widehat{V}_i$. The other steps involved are described in detail below.

4.1. Watermark and Watermark Information Extraction

The watermark is retrieved as described in [8] and the embedded information bits $\{a_j\}_{j>0}$ are obtained. The joint watermark information W_i is then computed from $\{a_j\}_{j>0}$ using Equation 4.

The number of objects n_i, watermark information of the background W_B and watermark information of the objects O_i^j are computed using the same set of congruences, 1 to 3 given in Section 3.4.

4.2. Authenticity Verification

The receiver first verifies that the number of objects in the received frame is equal to n_i computed from Eqn. 1. It then computes the object watermark information from the watermarked object $\widehat{O}_i^j$ using the same way as that by the sender given in Section 3.3. This is possible because the set of pixels from which watermark information are computed and the set of pixels on which the joint watermark is embedded are disjoint. The receiver checks whether this object watermark information computed from the pixels matches with W_i^j computed from Eqn 3. It also checks whether the watermark information of the background computed from the pixels matches with W_B computed from Eqn. 2. These verifications are repeated for all the objects and if all the verifications are successful $\widehat{V}_i$ is accepted as authentic.

4.3. Protocol at the Receiver's Side

The protocol at the receiver's side is as follows. Initially the background $\widehat{B}$ without any foreground object is received. The watermark signal W_B is extracted from B in a similar manner as described in Section 4.1. The watermark information of the background is then generated from B as in the case of foreground objects described in Section 4.2. The receiver then compares the extracted watermark signal with the generated one. If they match, the background $\widehat{B}$ is accepted as authentic and is stored. The foreground objects in each frame are received in real time and is processed as follows. We describe the protocol for the i-th frame $\widehat{V}_i$ of the video, where $i \in \mathbf{N}$.

1. $\widehat{V}_i$ is dissembled, decoded and the foreground objects $\widehat{O}_i^1, \ldots, \widehat{O}_i^{n_i}$ are obtained.
2. The watermark and watermark information are extracted and computed as described in Section 4.1.
3. Authenticity verification is performed as described in Section 4.2.
4. The objects $\widehat{O}_i^1, \ldots, \widehat{O}_i^{n_i}$ are placed on the background $\widehat{B}$ using the location coordinates and the frame is reconstructed.

5. Security Analysis

In this section, we carry out the security analysis of the proposed mechanism. A malicious entity can carry out one or more of the following three attacks (in the decreasing order of importance).

1. adding or removal or modification or replacement of some or all of the objects/background;

2. tampering with the watermarks on objects/background;

3. extraction of watermark from objects.

The security of the proposed scheme against the above attacks is based on the following three primitives (in the decreasing order of importance) used:

1. strength and security of $MAC(,)$ and K_{mac};

2. robustness of the watermarking algorithm;

3. strength and security of $PRNG(,)$ and K_{rnd}.

If the MAC function $MAC(,)$ and the key K_{mac} are secure an attacker will not be able to generate any valid watermark information W_i^j or W_B and hence a valid joint watermark.

If the pseudo-random number generator $PRNG(,)$ and the key K_{rnd} are secure an attacker will neither be able to identify the pixels positions in an object/background used to generate the watermark information W_i^j or W_B nor the pixel positions where the joint watermark W_i is embedded. Hence the attacker will neither be able to generate any valid watermark information W_i^j or W_B (hence a valid joint watermark) nor able to extract the joint watermark from any object or able to embed a valid/invalid joint watermark on any object.

If the watermarking algorithm used is robust an attacker will not be able to tamper with the watermark on any object.

We now discuss the security against the attacks mentioned in the beginning. Throughout this section by *Step 3* we mean the Step 3 of the protocol at the receiver's side given in Section 4.3.

If an attacker has added (inserted) an extra object O_i^* into the frame $\widehat{V}_i$, the receiver can detect it in *Step 3* since the number of objects in the received frame will not match with the number n_i (computed from the joint watermark W_i using Eqn. 2). Further, the new object will be identified in the watermark extraction process in *Step 3*, since the watermark extracted from O_i^* will differ from that of the original n_i objects (unless the same watermark has been embedded on it) and the watermark information of O_i^* computed from its pixels will not match with any watermark information W_i^j derived from the joint watermark using Eqn. 4. Similarly, if an object has been removed from the frame or an object/background has been replaced/modified it will be detected in *Step 3*.

An attacker may tamper with the watermarks on the objects/background. This attack can be avoided by using a robust watermark embedding mechanism.

An attacker may try to extract the watermark from a valid object and try to insert it on a new object. This attack can be easily avoided by making the watermarking pixel positions random using a secure $PRNG(,)$ and K_{rnd}.

6. CONCLUSION

In this paper, we have presented a novel mechanism using CRT to check the authenticity of a surveillance video transmitted as foreground objects and backgrounds. The experiments in Section 2, demonstrated that in many surveillance video applications, one may be able to save considerable bandwidth by only transmitting the foreground objects in real time and the backgrounds once in a while. The proposed authentication mechanism takes care of the security concerns related to the manipulation of a video frame.

For the simplicity of the discussion, in this paper we built our authentication mechanism using the spatial spread spectrum watermarking algorithm of Hartung and Girod [8] in the uncompressed domain. We will carry out the extension of this work to the compressed domain with more rigorous robustness analysis and performance evaluations as future work.

References

[1] L. Atzori, D.D. Giusto, C. Perra, "A Novel Block-based Video Segmentation Algorithm", IEEE International Conference on Multimedia and Expo (ICME), Tokyo, 2001.

[2] C. C. Chang, J. S. Lee, "Robust t-out-of-n Oblivious Transfer Mechanism Based on CRT", Journal of Network and Computer Applications, 32, pp.226-235, 2009.

[3] S.Y. Chien, Y. W. Huang, B.Y. Hsieh, S.Y. Ma, L.G. Chen, "Fast Video Segmentation Algorithm with Shadow Cancellation, Global Motion Compensation and Adaptive Threshold Techniques", IEEE Transactions on Multimedia, Vol. 6, No. 5, pp. 732-748, 2004.

[4] G.H. Chiou, W.T. Chen, "Secure Broadcasting Using the Secure Lock", IEEE Transaction on Software Engineering, Vol. 15, No. 8, pp. 929-934, 1989.

[5] P. Correia, F. Pereira, "Objective Evaluation of Relative Segmentation Quality", IEEE International Conference on Image Processing (ICIP), pp. 308-311, 2000.

[6] C. Ding, D. Pei, A. Salomaa, "Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography", World Scientific, June 1999.

[7] C. E. Erdem, B. Sankur, A. M.Tekalp, "Performance Measures for Video Object Segmentation and Tracking", IEEE Trans. Image Processing, vol. 13, no. 7, pp. 937-951, 2004.

[8] F. Hartung, B. Girod, “Watermarking of Uncompressed and Compressed Video”, Signal Processing, Vol. 66, No. 3, pp. 283-301, 1998.

[9] W. Hu, T. Tan, L. Wang, S. Maybank, “Survey on Visual Surveillance of Object Motion and Behaviors”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Appl. and Reviews, Vol. 34, No. 3, pp. 334-352, 2004.

[10] B. Hu, W. Ye, Sui-Li Feng, Xiao-Liang Wang, X. Xie, “Key Distribution Scheme Based on Two Cryptosystems for Hierarchical Access Control”, ICACT 2006, pp. 1723-1728.

[11] O. Javed, Shafique, K., Shah, M., “A Hierarchical Approach to Robust Background Subtraction Using Color and Gradient Information”, WMVC, Orlando, 2002.

[12] L. Li, W. Huang, I.Y.H. Gu, Q. Tian, “Foreground Object Detection from Videos Containing Complex Background”, ACM International Conference on Multimedia, pp. 2-10, CA, USA, 2003.

[13] V. Y. Mariano, J. Min, J. H. Park, R. Kasturi, D. Mihalcik, H. Li, D. Doermann, T. Drayer, “Performance Evaluation of Object Detection Algorithms”, ICPR02, Quebec City, Canada, 2002.

[14] T. Meier, K.N. Ngan, “Video Segmentation for Content-based Coding”, IEEE TCSVT, Vol. 9, No. 8, pp. 1190-1203, Dec. 1999.

[15] H. E. Michail, A. P. Kakarountas, A. Milidonis, C. E. Goutis, “Efficient Implementation of the Keyed-hash Message Authentication Code (HMAC) Using the SHA-1 Hash Function”, ICECS 2004, pp. 567 - 570.

[16] J. C. Patra, A. Karthik, P. K. Meher, C. Bornand, “Robust CRT-based Watermarking Technique for Authentication of Image and Document”, IEEE International Conference on Systems, Man and Cybernetics, SMC 2008, pp.3250 - 3255, 2008.

[17] J. C. Patra, J. E. Phuab, C. Bornand, “A Novel DCT Domain CRT-based Watermarking Scheme for Image Authentication Surviving JPEG Compression”, Digital Signal Processing, Article in Press.

[18] A. Shamim, J.A. Robinson, “Object-based Video Coding by Global-to-local Motion Segmentation”, IEEE TCSVT, vol. 12, no. 12, pp. 1106-1116, Dec. 2002.

[19] H. Sun, T. Feng, T. Tan, “Spatio-temporal Segmentation for Video Surveillance”, ICPR 00, Spain.

[20] T. Thomas, S. Emmanuel, A. Das, M. Kankanhalli, “A CRT Based Watermark for Multiparty Multilevel DRM Architecture”, ICME 2009, New York, USA.

[21] T. Thomas, S. Emmanuel, A. V. Subramanyam, M. Kankanhalli, “Joint Watermarking Scheme for Multiparty Multilevel DRM Architecture”, IEEE TIFS, Vol. 4, No. 4, pp.758-767, Dec. 2009.

[22] NIST FIPS 198, http://csrc.nist.gov/publications/fips/fips198/fips-198a.pdf.

[23] NIST SP 800-90, http://csrc.nist.gov/ publications/nistpubs/800-90/SP800-90revised_March2007.pdf.

[24] CAVIAR Test Case Scenarios, http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/.

[25] AVSS 2007, Datasets, http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007_d.html.