Privacy Preserving Video Surveillance Using Pedestrian Tracking Mechanism

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

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

ABSTRACT

Video surveillance has become a ubiquitous feature of the modern day life. However, the widespread use of video surveillance has raised concerns about the privacy of people. In this paper, we propose a novel video surveillance with a privacy preserving mechanism. We achieve this by combining the techniques of pedestrian tracking based on a Markov chain with two hidden states, elliptical head contour detection and encryption. The detected pedestrian face/head is obscured by encrypting with a unique key derived from a master key for the privacy preservation purpose. The surveillance video can be viewed with complete privacy or by revoking the privacy of any subset of pedestrians while ensuring complete privacy of the remaining pedestrians. The performance evaluation on many challenging surveillance scenarios shows that the proposed mechanism can effectively and robustly track multiple pedestrians and obscure their faces/head in real time.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy, Intellectual Property Rights*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Object recognition, Tracking*

General Terms

Security, Legal Aspects, Human Factors

Keywords

Video Surveillance, Pedestrian Tracking, Privacy

1. INTRODUCTION

Nowadays, video surveillance systems have been widely deployed in many public places for various applications. It is used in connection with detection and quick resolution of crimes, traffic accidents/violations, gathering of data on the presence and actions of people for various purposes [8]. However, the widespread use of video surveillance has raised concerns about the privacy of people. Thus mechanisms that can satisfy surveillance objectives while ensuring privacy of the people are necessary.

The privacy issues in surveillance systems have been investigated by various authors by looking at different aspects of the problem. Most of the privacy preserving mechanisms are centred around providing privacy by selective modification of the objects in the surveillance scenes. The techniques involve irreversible modifications of the objects such as simple video masking techniques [16], use of black boxes or large pixels [1, 15] to complete object removal, replacing a particular face with a generic face [11, 17], complete object removal followed by inpainting of the background and other foreground objects [14]. All these irreversible techniques aim at the modification of the video without looking into the feasibility of recovering the original video. To securely recover the original video, reversible video modification techniques are applied. The reversible video modification techniques have to be key dependent to prevent misuse of the reversibility property. Some of the reversible video modification techniques are based on encryption techniques [2], transform-domain scrambling of regions of interest [6] and secure coding of arbitrarily shaped visual objects [10].

Whereas the above mechanisms aim at providing and revoking privacy for all the pedestrians collectively, there have been some proposals on providing privacy at an individual level. In [13], a system which conveys sufficient situation information while protecting a specific person's privacy information by identifying the person by face recognition is given. In [18], the authors proposed a framework to store the privacy information in a surveillance video as a watermark and monitor the unauthorized persons in a restricted area while protecting the privacy of the authorized persons. In [5], a system for protecting the privacy of specific individuals in video recordings of medical applications is given. These kinds of privacy preserving mechanisms use people identification mechanisms and perform selective obscuration of the objects before transmission. These mechanisms are application specific and require predefined rules and data for providing individual privacy.

In many security related applications, a surveillance video may be required to be viewed with the privacy of a set of pedestrians revoked while maintaining complete privacy for the remaining set of pedestrians. This can arise when a suspicious activity by one or more pedestrians is detected and their identity has to be quickly found out or the video has to

MiFOR'10, October 29, 2010, Firenze, Italy.

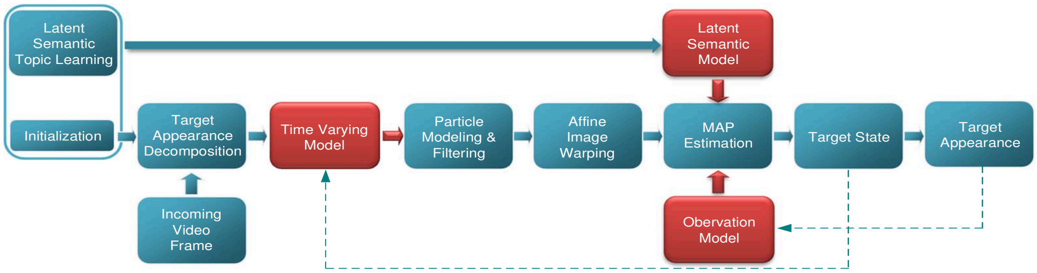


Figure 1: Pipeline of Two State Markov Chain Tracking Algorithm

be shown to an audience with the identity of only such pedestrians exposed. In this paper, we propose a novel mechanism in which a surveillance video can be viewed with complete privacy for all the pedestrians or by revoking the privacy of any set of pedestrians while ensuring complete privacy of the remaining pedestrians. With the proposed mechanism, an authorized person can selectively remove the privacy protection of any set of pedestrians using a combination of keys (while ensuring complete privacy of the remaining set of pedestrians). This mechanism ensures that the privacy of innocent pedestrians are always protected even in situations where that of a set of suspicious ones are to be revoked. We achieve this by combining the mechanisms of pedestrian tracking, head contour detection and encryption. Privacy is provided to the pedestrians by obscuring their face/head using an encryption algorithm. The video is normally viewed with complete privacy to the pedestrians. The goals of the video surveillance are still met as face/head obscuration of the pedestrians allows monitoring of their activities without knowing their identities. To make the discussions simple, in this paper we do not address the issue whether face/head obscuration alone can guarantee complete privacy to the pedestrians. The experimental evaluations with many challenging surveillance scenes show that the proposed mechanism can effectively and robustly track multiple pedestrians and obscure their faces/head in real time.

The remaining part of the paper is organized as follows. The proposed mechanism is described in detail in Section 2. The performance of the proposed system is demonstrated through experimental results in Section 3. Finally, the paper concludes with observations, remarks and future directions for research in Section 4.

2. PRIVACY PRESERVING VIDEO TRANSMISSION AND VIEWING MECHANISM

The proposed system has been built by combining the techniques of pedestrian tracking, head contour detection and encryption. We first use a novel, robust and efficient pedestrian tracking algorithm based on a Markov chain with two hidden states to locate/track the pedestrians and then an elliptical head contour detection algorithm is applied to detect the location of the face/head of each tracked pedestrian. The face/head of each pedestrian is then obscured by encrypting in the spatial domain with a unique key derived from a master key. We perform all the above operations on the raw video data. Finally, the output video stream is subjected to the standard encoding process. The resultant video can be viewed with complete/selective privacy to the pedestrians. We now describe each of these steps in detail.

2.1 Identification and Tracking of Pedestrians

We model the pedestrian tracking problem as an inference task in a temporal Markov chain with two hidden state variables. The major steps involved are represented schematically in Fig. 1. The proposed model is given in Fig. 2. The details are given below.

For a pedestrian P, let X_t describe the affine motion parameters (and thereby the location) of P at time t and $\mathbf{I}_t = \{d_t^1, d_t^2, \ldots, d_t^n\}$ denotes a collection of the estimated image patches of P at time t, where n is a predefined number of sample estimates. Let $\mathbf{z} = \{z_1, z_2, \ldots, z_k\}$ be a collection of k latent semantic topics associated with the pedestrians and $\mathbf{w} = \{w_1, w_2, \ldots, w_m\}$ is a collection of m codewords . We consider the temporal Markov chain with hidden states X_t and $\mathbf{z}$. It is assumed that the hidden state $\mathbf{X}_t$ is independent of the latent semantic topics $\mathbf{z}$ and the codewords $\mathbf{w}$ [7]. Let $\mathcal{I}_t = \{\mathbf{I}_1, \ldots, \mathbf{I}_{t-1}, \mathbf{I}_t\}$. Now, from the Fig 2:(a) the whole tracking process can be accomplished by maximizing the probability $p(\mathbf{X}_t|\mathcal{I}_t, \mathbf{z}, \mathbf{w}) \cdot p(\mathcal{I}_t|\mathbf{z}, \mathbf{w}) \cdot p(\mathbf{z}|\mathbf{w})$ for each time stage t. We have the following relations,

$$p(\mathbf{X}_t|\mathcal{I}_t, \mathbf{z}, \mathbf{w}) \cdot p(\mathcal{I}_t|\mathbf{z}, \mathbf{w}) \cdot p(\mathbf{z}|\mathbf{w}) \propto p(\mathbf{X}_t|\mathcal{I}_t) \cdot p(\mathcal{I}_t|\mathbf{z}, \mathbf{w}),$$

$$p(\mathbf{X}_t|\mathcal{I}_t) \propto p(\mathbf{I}_t|\mathbf{X}_t) \int p(\mathbf{X}_t|\mathbf{X}_{t-1}) p(\mathbf{X}_{t-1}|\mathcal{I}_{t-1}) d\mathbf{X}_{t-1} = \mathbf{Y}_t.$$

Since the latent semantic analysis process is not a temporal inference, the maximum probability for each time stage t can be obtained as follows:

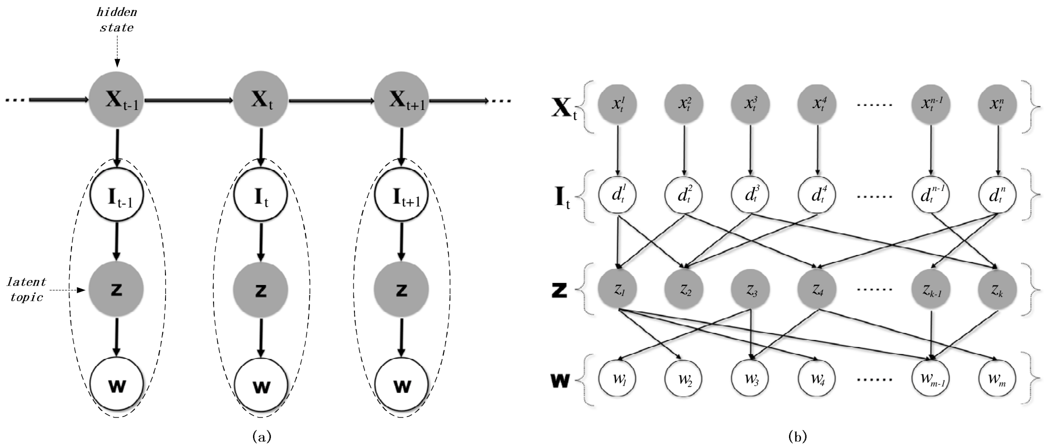
$$\max\Big\{p(\mathbf{X}_t|\mathcal{I}_t, \mathbf{z}, \mathbf{w}) \cdot p(\mathcal{I}_t|\mathbf{z}, \mathbf{w})\Big\} = \max\Big\{\mathbf{Y}_t \cdot p(\mathbf{I}_t|\mathbf{z}, \mathbf{w})\Big\}$$
$$= \max\Big\{\mathbf{Y}_t \cdot p(d_t^1|\mathbf{z}, \mathbf{w}), \ldots, \mathbf{Y}_t \cdot p(d_t^n|\mathbf{z}, \mathbf{w})\Big\}.$$

Thus, tracking at each time stage t is achieved by maximizing the following quantity for each $i \in [1, n]$,

$$p(d_t^i|\mathbf{z}, \mathbf{w}) p(\mathbf{I}_t|\mathbf{X}_t) \int p(\mathbf{X}_t|\mathbf{X}_{t-1}) p(\mathbf{X}_{t-1}|\mathcal{I}_{t-1}) d\mathbf{X}_{t-1}. \quad (1)$$

We estimate the three probabilities $p(\mathbf{X}_t|\mathbf{X}_{t-1})$, $p(\mathbf{I}_t|\mathbf{X}_t)$ and $p(d_t^i|\mathbf{z}, \mathbf{w})$ in the Expression 1 above by adopting the following three probabilistic models.

1. **Time-Varying Motion Model**: The state transition probability $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ is estimated using this model. For the time-varying motion model, each pedestrian to be tracked is characterized by an affine image warp,


hidden state
latent topic
X_{t-1}
X_t
X_{t+1}
I_{t-1}
I_t
I_{t+1}
z
w
X_t
x_t^1 x_t^2 x_t^3 x_t^4 ...... x_t^{n-1} x_t^n
I_t
d_t^1 d_t^2 d_t^3 d_t^4 ...... d_t^{n-1} d_t^n
z
z_1 z_2 z_3 z_4 ...... z_{k-1} z_k
w
w_1 w_2 w_3 w_4 ...... w_{m-1} w_m

(a)

(b)

Figure 2: (a) Two Hidden State Markov Chain Model, (b) Probability Map of Time t

which is represented by the hidden state $\mathbf{X}_t$ composed of 6 parameters: $\mathbf{X}_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \varphi_t)$, where $x_t, y_t, \theta_t, s_t, \alpha_t, \varphi_t$ denote x, y translation, rotation angle, scale, aspect ratio and skew direction at time t respectively. As in [9], the distribution of each parameter of $\mathbf{X}_t$ is assumed to be Gaussian centered around $\mathbf{X}_{t-1}$ and the corresponding diagonal covariance matrix Ψ of the Gaussian distribution is made up of 6 parameters denoting the variance of the affine parameters, σ_x^2, σ_y^2, σ_θ^2, σ_s^2, σ_α^2 and σ_φ^2. Thus,

$$p(\mathbf{X}_t|\mathbf{X}_{t-1}) = \mathcal{N}(\mathbf{X}_t; \mathbf{X}_{t-1}, \Psi).$$

2. **Observation Model**: The relationship $p(\mathbf{I}_t|\mathbf{X}_t)$ between the observations $\mathbf{I}_t$ and the hidden states $\mathbf{X}_t$ is estimated using this model. We use $\mathbf{I}_t$ to denote a collection of the estimated image patches from the hidden state $\mathbf{X}_t$. Suppose that the sample $\mathbf{I}_t$ is drawn from a subspace spanned by U and centered at $\boldsymbol{\mu}$. Let Σ denote the matrix of singular values corresponding to the columns of U, $\mathbf{I}$ denote the identity matrix and $\varepsilon\mathbf{I}$ denotes the additive Gaussian noise in the observation process. Then as in [12], the probability of a sample drawn from the subspace is estimated as:

$$p(\mathbf{I}_t|\mathbf{X}_t) \quad = \quad \mathcal{N}(\mathbf{I}_t; \boldsymbol{\mu}, UU^\top + \varepsilon\mathbf{I}) \cdot \mathcal{N}(\mathbf{I}_t; \boldsymbol{\mu}, U\Sigma^{-2}U^\top).$$

3. **Probabilistic Latent Semantic Analysis (*pLSA*) Model**: The *pLSA* model [7] is employed in the testing phase to find the maximal pedestrian likelihood probability $p(d_t^i|\mathbf{z}, \mathbf{w})$, for $1 \leq i \leq n$. For the *pLSA* model in [7], the variable d_t^i denotes a document, while in our case it denotes an estimated image patch which is the observation. As in [7] the likelihood of each estimated sample at time t to be a pedestrian is obtained as,

$$p(d_t^i|\mathbf{z}, \mathbf{w}) \propto p(d_t^i|\mathbf{z}) \propto \sum_{w \in \mathbf{w}} p(\mathbf{z}|w)p(w|d_t^i).$$

Since latent semantic analysis process is not a temporal inference process, by using all the above three models in the Expression 1, the tracking of the pedestrian can be performed.

2.2 Pedestrian Head Contour Detection

The tracking algorithm from the previous section yields a set of sub-images, corresponding to the bounding boxes of the pedestrians in each video frame. The next task is to locate the head contour of each pedestrian in the bounding boxes. Here, we employ the head contour detection method proposed by Zou *et. al.* [19] based on quadrant arcs, which is efficient and robust with respect to arbitrary head pose and wide distance range. We briefly describe the main steps below.

For each bounding box in a frame, we first perform the Canny operator on it to obtain an edge map of single pixel width. Then all the pixels in the edge map are scanned to obtain the connected edges having single pixel width and length over a pre-defined threshold. Each connected edge is put into one of the four sets of arcs: 1st quadrant arcs; 2nd quadrant arcs; 3rd quadrant arcs; 4th quadrant arcs (see Fig. 3). One arc is then selected from each quadrant arc set according to some specified rules. All the pixels on the selected arcs constitute the candidate fitted points set. The ellipse with minimum errors from the candidate fitted points is then fitted out using the least square method as a quadratic polynomial $x^2 + k_0xy + k_1y^2 + k_2x + k_3y + k_4 = 0$. Finally, based to human head's characteristics, the most likely ellipse is determined as the elliptical head contour.

2.3 Face Obscuration Using Encryption

Let K denote a master key, $\mathbf{MAC}(.)$ denote a cryptographic MAC function and $\mathbf{Stream}(.)$ denote a secure stream cipher which generates integer random values in the range $[0, 255]$. We now describe the pedestrian face obscuration using the encryption mechanism below.

1. **Encryption Key Derivation**: Let a pedestrian P is detected by the tracking algorithm for the first time in the i_0-th frame and the centre of the head contour ellipse of P be (x_P, y_P) in this frame. Then the unique key for encryption corresponding to P is obtained as

$$K_P = \mathbf{MAC}(K, i_0||x_P||y_P),$$

where $||$ is the concatenation operator.

2. **Pedestrian Face Obscuration**: Let the output of the stream cipher be $\mathbf{Stream}(K_P) = r_1, r_2, \ldots$. The

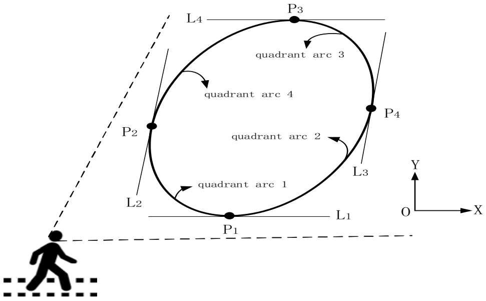


Figure 3: Pedestrian Head Contour

elliptical head contour of the pedestrian P in the j-th frame ($j \geq i$) is raster scanned from left to right and top to bottom and the interior pixel point Pix_i, where $i > 0$ is encrypted as,

$$\widehat{Pix_i} = Pix_i + r_i \mod 256.$$

Note that Pix_i can be a 3-tuple corresponding to the RGB components.

2.4 The Final Video Bit Stream

The output from the previous stage is subjected to any standard video coding algorithm. In addition to the video bit stream, the parameters of the head contour of each pedestrian in each frame $< k_0, k_1, k_2, k_3, k_4 >$ is added as a header after encryption with a key K_{head}. The head contour information corresponds to an additional data of 10 bytes per pedestrian per frame of the video which is negligible in most of the applications. The head contour information of each pedestrian is provided along with the video data to enable an authorized person to efficiently and accurately locate the head contour of the pedestrians and decrypt the face.

2.5 Privacy Preserving Video Viewing

The surveillance video after reception is decoded and is made available for viewing. The viewers will be able to watch the pedestrians with their obscured face only. Thus their privacy is protected. There is a trusted super-user U who has access to the master key K and the key K_{head}. The encrypted head contour data is stored securely by the super user U. If a need arises to expose the identity of a specific pedestrian P to a set of viewers such as crime investigating agencies or court or public, the super-user U performs the operations described below.

1. **Decryption Key Derivation**: The decryption key K_P is the same as the encryption key and is computed using the head contour information in exactly the same way as in Section 2.3.

2. **Pedestrian Face Un-obscuring**: Let the output of the stream cipher be $\textbf{Stream}(K_P) = r_1, r_2, \ldots$. The elliptical head contour parameters of the pedestrian P in the j-th frame ($j \geq i$) is obtained from the stored head contour data after decryption with the key K_{head}. The head contour is then located in the video frame. It is then raster scanned from left to right and top to bottom and the interior pixel point $\widehat{Pix_i}$, where $i > 0$ is decrypted as,

$$Pix_i = \widehat{Pix_i} - r_i \mod 256.$$

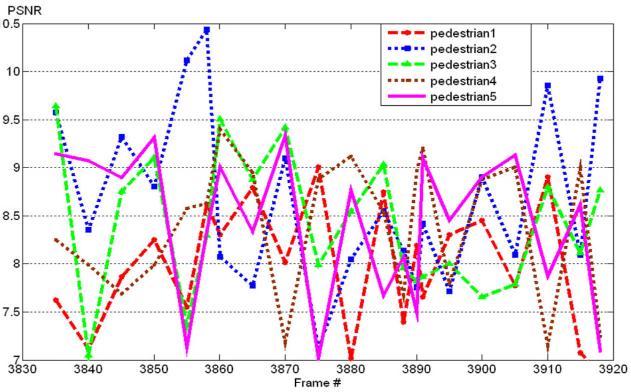


Figure 4: PSNR Values of Faces After Obscuration

2.6 Security Analysis

In this section, we carry out the analysis of the privacy protection guaranteed by the proposed mechanism. We examine the following:

1. whether the encryption used for obscuring the faces is secure and provides sufficient privacy protection?

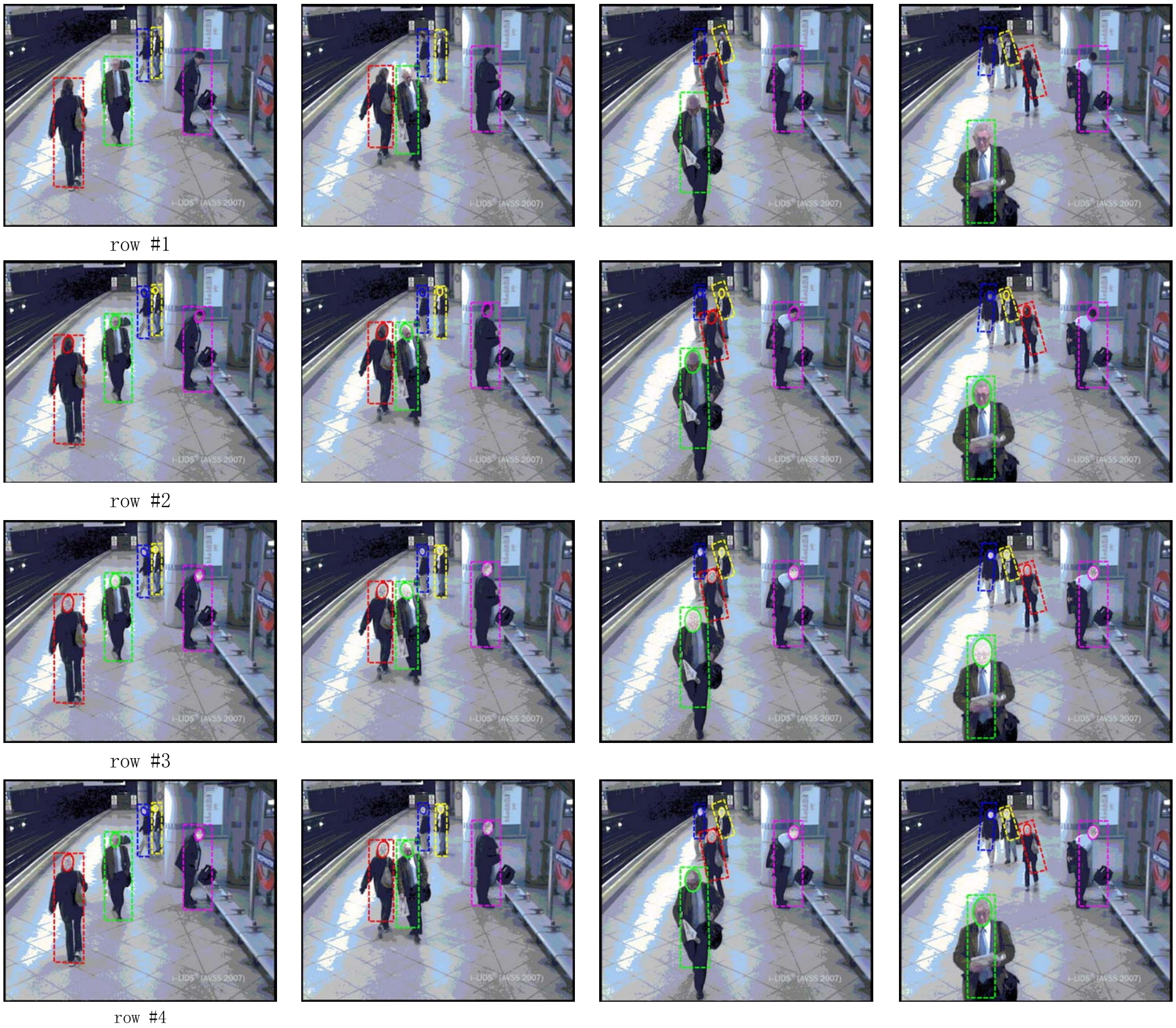
2. whether revoking the privacy of a set pedestrians compromises the privacy of the remaining pedestrians?

The encryption scheme we used for obscuring the face of the pedestrians is the same as the homomorphic symmetric key encryption scheme proposed in [4] for data aggregation in wireless networks. This has been proved to be secure in [4].

Theorem 2.1. *The encryption scheme used for face obscuration is perfectly secure.*

The strength and security of this encryption scheme depends on the underlying stream cipher $\textbf{Stream}(,)$ used. We suggest the use of a strong stream cipher such as RC4. Further, the face obscuration by this encryption algorithm provides sufficient privacy protection as demonstrated in Fig. 4. The PSNR values of the faces of the 5 pedestrians after obscuration in 80 frames of the video "AVSS_AB_Medium" (shown in Fig. 5) from *AVSS2007* dataset is plotted in Fig. 4. Here color of the curve corresponds to the color of the bounding box of the pedestrian in Fig 5. It can be observed from the graph that the PSNR values of the faces after obscuration lies mostly in the range 7-9 dB. Thus, this obscuration mechanism adds sufficient noise into the face of the pedestrians to provide an acceptable level of privacy protection to the pedestrians.

We now prove that revoking the privacy of a set of pedestrians does not compromise the privacy of the remaining set of pedestrians in the video. The privacy of each pedestrian can be associated with the key K_P used to generate the key stream used for obscuring its face. We will prove that even if the keys corresponding to a set of pedestrians are available to an adversary, he cannot compute the key corresponding to a different pedestrian. This is because, we computed



row #1

row #2

row #3

row #4

Figure 5: Effect of the Proposed Mechanism on Surveillance Video

each key K_P as an output of a secure MAC function. To be considered secure, a MAC function must resist existential forgery under chosen-plaintext attacks. That is even if an adversary has access to an oracle which can generate MACs for messages of the adversary's choosing, the adversary cannot guess the MAC of another message without performing infeasible amounts of computation.

Theorem 2.2. *If* $\mathbf{MAC}(.)$ *is a MAC function which is resistant to existential forgery under chosen-plaintext attacks, then it is infeasible to revoke the privacy of any pedestrian using the privacy information of any other set of pedestrians.*

Proof. Let $P_1, \ldots, P_n$ be a set of pedestrians whose privacy has been revoked and the keys $K_{P_1}, \ldots, K_{P_n}$ respectively associated with them are available to an adversary. We will prove that using this information, the privacy of any other pedestrian P_0 cannot be revoked. In other words an adversary cannot derive the key K_{P_0} using the keys $K_{P_1}, \ldots, K_{P_n}$. Let

$$K_{P_j} = \mathbf{MAC}(K, i_j || x_{P_j} || y_{P_j}), \text{ for } 0 \leq j \leq n.$$

Now, computing K_{P_0} from $K_{P_1}, \ldots, K_{P_n}$ corresponds to an existential forgery of the MAC function $\mathbf{MAC}(.)$ under the chosen plain texts $\{i_j || x_{P_j} || y_{P_j} : 1 \leq j \leq n\}$. By assumption, this is infeasible. Hence, the privacy of any pedestrian cannot be revoked using the privacy information of any other set of pedestrians.

3. EXPERIMENTAL RESULTS

To verify the performance of the proposed mechanism, we choose the test videos from the popular surveillance datasets, *AVSS2007*, *CAVIAR PETS2006* and *PETS2007*.

Fig. 5 shows the results on the video "AVSS_AB_Medium" from *AVSS2007* dataset. The first row shows the results of tracking of multiple pedestrians by employing the proposed tracking mechanism and it shows that the proposed tracking can robustly track multiple pedestrian even if occlusion occurs. The second row shows the results of the head contour detection and the third row shows the effect of obscuring the interior of elliptical head contours of pedestrians. The last row presents the effect of viewing with the privacy of

Table 1: Performance of the Proposed System with Different Number of Estimation Particles (EP)

Scenes & Number of Pedestrians	EP = 25 Avg. fps	EP = 50 Avg. fps	EP = 75 Avg. fps	EP = 100 Avg. fps	EP =125 Avg. fps
S3-T7-A_Scene4 (6 pedestrians)	6.677	6.305	5.905	5.604	5.274
EnterExitCrossingPaths2cor (1 pedestrian)	7.605	7.479	7.341	7.185	7.134
EnterExitCrossingPaths1cor (3 pedestrians)	7.405	7.226	6.911	6.667	6.432
S01-GENERAL_LOITERING_1_Scene4 (4 pedestrians)	6.543	6.440	6.020	5.792	5.545
AVSS_AB_Medium (5 pedestrians)	7.094	6.806	6.467	6.226	6.025
S07-ABANDONED_BAG_1_Scene 2 (5 pedestrians)	6.829	6.460	6.052	5.837	5.519

one pedestrian revoked. We leave out the other visual performance results on the datasets due to space limitations.

Table 1 shows the overall performance of the proposed system. Different columns show the effect of varying the number of estimation particles used by the tracking algorithm. Increasing the number of particles increases the accuracy of tracking but slows the performance. We performed the experiments in the environment: INTEL Core2 Duo 6600, 2GB RAM, WINDOWS XP and MATLAB 2009a. We can see that the average performance is around 6.5 fps. The proposed mechanism can fulfil the real-time requirement with C/C++ implementation by optimizing many loops.

4. CONCLUSIONS

In this paper, we have proposed a novel privacy preserving surveillance video transmission mechanism in which the received video can be viewed with complete privacy or by revoking the privacy of any set of pedestrians while ensuring privacy of the remaining pedestrians. The performance evaluation with many challenging surveillance scenes shows that the proposed mechanism can perform in real time.

For the simplicity of the discussion, in this paper we have not addressed the issue whether face obscuration alone can guarantee privacy of the pedestrians. We will extend this work towards full body obscuration in the future.

5. ACKNOWLEDGEMENT

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).

6. REFERENCES

[1] A. M. Berger, "Privacy Mode for Acquisition Cameras and Camcorders", US Patent 6067399, Sony Corporation, May 2000.

[2] T. E. Boult, "Pico: Privacy through Invertible Cryptographic Obscuration", Proceedings of the Computer Vision for Interactive and Intelligent Environments, pp. 27-38, 2005.

[3] P. Carrillo, H. Kalva, S. Magliveras,"Compression Independent Object Encryption for Ensuring Privacy in Video Surveillance" , ICME 2008, Hanover, Germany.

[4] C. Castelluccia, E. Mykletun, G. Tsudik, "Efficient Aggregation of Encrypted Data in Wireless Sensor Networks', International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp. 109 - 117, 2005.

[5] D. Chen, Y. Chang, R. Yan, J. Yang, "Tools for Protecting the Privacy of Specific Individuals in Video", EURASIP Journal on Applied Signal Processing, Issue 1, pp.107- 107, 2007.

[6] F. Dufaux, T. Ebrahimi, "Scrambling for Video Surveillance with Privacy", 2006 Conference on Computer Vision and Pattern Recognition Workshop, NY, 2006.

[7] T. Hofmann, "Probabilistic Latent Semantic Indexing, Proc. ACM SIGIR, pp. 50-57, 1999.

[8] K. S. Huang, M. M. Trivedi, "Integrated Detection, Tracking, and Recognition of Faces with Omnivideo Array in Intelligent Environments", EURASIP Journal on Image and Video Processing, vol. 2008, 2008.

[9] M. Isard, A. Blake, "Condensation Conditional Density Propagation for Visual Tracking", International Journal of Computer Vision, 29(1):5-28, 1998.

[10] K. Martin, K. N. Plataniotis, "Privacy Protected Surveillance Using Secure Visual Object Coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, pp. 1152-1162, 2008.

[11] E. M. Newton, L. Sweeney, B. Malin, "Preserving Privacy by De-identifying Face Images", IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232-243, 2005.

[12] D.A. Ross, J. Lim, R.S. Lin, M.H. Yang. "Incremental Learning for Robust Visual Tracking", International Journal of Computer Vision, 77(1-3):125-141, 2008.

[13] S. Tansuriyavong, S.-I.Hanaki, "Privacy Protection by Concealing Persons in Circumstantial Video Image", Proceedings of the Workshop on Perceptive User Interfaces, pp. 1-4, 2001.

[14] M. V. Venkatesh, S.-C. Cheung, J. Zhao, "Efficient Object Based Video Inpainting", Pattern Recognition Letters, vol. 30, no. 2, pp. 168-179, 2009.

[15] J. Wada, K. Kaiyama, K. Ikoma, H. Kogane, "Monitor Camera System and Method of Displaying Picture from Monitor Camera Thereof", European patent, EP 1081955 A2, Matsushita, April 2001.

[16] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacy Protecting Data Collection in Media Spaces", ACM International Conference on Multimedia, pp. 48-55, 2004.

[17] X. Yu, N. Babaguchi, "Privacy Preserving: Hiding a Face in a Face", Proceedings of ACCV, LNCS, vol. 4844, pp. 651-661, 2007.

[18] W. Zhang, S.C. Cheung, M. Chen, "Hiding Privacy Information in Video Surveillance System", Proceedings of ICIP, vol. 3, pp. 868-871, 2005.

[19] W. Zou, Y. Li, K. Yuan, D. Xu, "Real-time Elliptical Head Contour Detection under Arbitrary Pose and Wide Distance Range, Journal of Visual Communication and Image Representation, Vol. 20, pp. 217-228, 2009.