A NOVEL EYE REGION BASED PRIVACY PROTECTION SCHEME

Dohyoung Lee, Konstantinos N. Plataniotis

Multimedia Lab, The Edward S. Rogers Department of Electrical and Computer Engineering,
University of Toronto, 10 King's College Road, Toronto, Canada

ABSTRACT

This paper introduces a novel eye region scrambling scheme capable of protecting privacy sensitive eye region information present in video contents. The proposed system consists of an automatic eye detection module followed by a privacy enabling JPEG XR encoder module. An object detection method based on a probabilistic model of image generation is used in conjunction with a skin-tone segmentation to accurately locate eye regions in real time. The utilized JPEG XR encoder effectively deteriorate the visual quality of privacy sensitive eye region at low computational cost. Performance of proposed solution is validated using benchmark face recognition algorithms on face image database. Experimental results indicate that the proposed solution is able to conceal identity by preventing successful identification at low computational costs.

Index Terms— privacy protection, eye region, selective scrambling, JPEG XR, eye detection

1. INTRODUCTION

The issue of privacy protection in video contents has become an active research area with the proliferation of surveillance systems and video based human behavioral researches. Especially the advent of video capturing equipments offering high quality recording capability and high computational power has raised the awareness of privacy invasion issue by revealing the identity of face in video data either via human operator or face recognition (FR) software. In literatures, many different solutions are proposed to protect privacy sensitive face regions, including region-based transform-domain scrambling techniques [1, 2]. In these methods, region-of-interests (ROIs) are initially estimated and the corresponding transform coefficients (associated with video encoder) are scrambled by a set of encryption techniques, leading to deterioration of the visual quality in the privacy sensitive regions. Only an authorized user who possesses a valid secret encryption key can revert scrambling process to access meaningful visual data in privacy sensitive regions while an unauthorized user is only able to recognize the background scene.

It has been shown that the application of scrambling techniques allows for an increased level of security, but at the same time, it causes bitstream overhead since changes in transform coefficients may significantly affect the effectiveness of entropy coding. Therefore, existing protection solutions maintain the balance between level of security, coding efficiency, and computational complexity by : i) exploiting a different encryption techniques on different color channel and frequency components, ii) minimizing the region that the protection solution is applied to.

In this paper, a novel eye region based privacy protection scheme is proposed. The eye region based solution is motivated by the fact that this area contains the most discriminative information among three facial regions including nose and mouth regions, in terms of automatic FR algorithm performance [3]. Conversely, protecting eye region should be able to prevent successful identification of face in two perspectives: i) conceal the identify of subject by hiding the most discriminative features, ii) thwart proper initialization of automatic FR algorithms which typically rely on accurate eye detection as an essential preprocessing step. In addition, the eye region based approach reduces computational costs associated with encryption compared to the full facial region based approach as only a subset of facial regions are scrambled.

The proposed system consists of an automatic eye detection module followed by a privacy enabling JPEG XR encoder module employing the subband adaptive scrambling technique [2]. The JPEG XR standard [4] based encryption is considered since it offers a low-complexity solution enabling intra-coding of high-resolution video content. The proposed scheme initially searches for eye regions in input video frame by performing skin-tone segmentation in conjunction with two-staged object detection based on a generative framework [5]. The eye location information is then delivered to privacy enabling JPEG XR encoder that selectively scrambles macroblocks (MB) corresponding to located eye regions.

The rest of this paper is organized as follows. Section 2 presents the proposed eye region privacy protection scheme in detail. Experiment results are reported in Section 3 and conclusion is demonstrated in Section 4.

2. PROPOSED METHODS

The proposed system consists of following main blocks: an automatic eye detection module and a privacy enabling JPEG XR encoder module. (Fig. 1) The input to the system is video data in RGB representation and the output is JPEG XR intra-coded (or Motion JPEG XR) video stream with scrambled eye regions. Each frame is processed independently from previous frames for simplified system architecture.

2.1. Automatic eye detection module

This module is responsible for locating the human eye regions from input video frame via a multi-stage operation. It utilizes both color-based and Haar-like/GentleBoost based object detection methodologies to localize eye coordinates with high accuracy.

1) Skin-tone segmentation : Initially, the module employs a skin-tone detector in order to distinguish skin pixels from non-skin pixels in input frame. Skin color is a low-level cue that offers robust face detection performance towards geometrical changes at low computational complexity. However, the use of color-based analysis only is apparently not sufficient to filter out skin-tone like backgrounds from facial region, and thus this block simply eliminates non skin-tone pixels in input frame using simple decision rules so

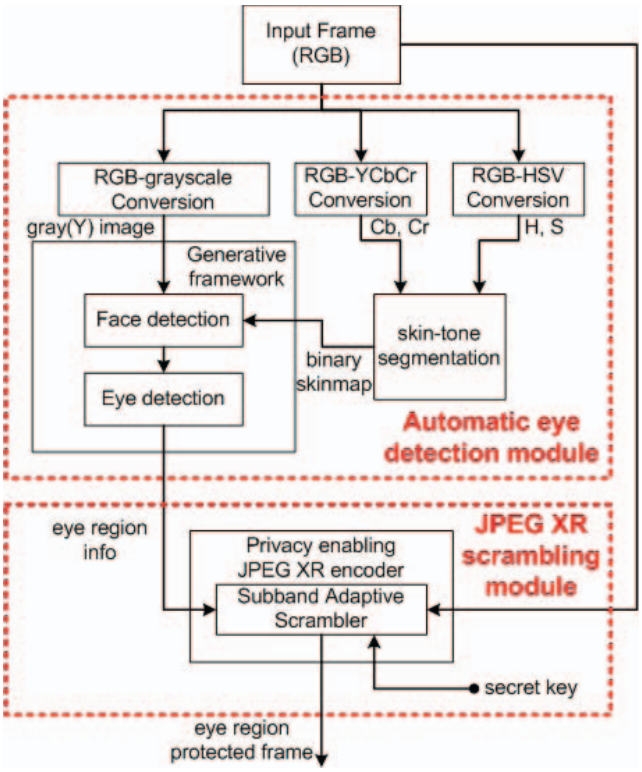

Fig. 1: The proposed eye region scrambling module

that subsequent detection module examines smaller search windows. Thus, this block is expected to achieve low false negative (wrongly classifying a skin pixel as a non-skin) rate.

Despite its dominant usage in digital imaging, RGB color space disallows robust segmentation of skin-tone pixels in varying illumination condition since it doesn't clearly separate the chroma information from the intensity of pixel. Therefore, we transform input RGB data into a color space that tolerates minor variation in the intensity and minimizes the overlap between skin and non-skin distributions. Following a method suggested in [6], we perform detection in multiple color spaces, YCbCr and HSV (Hue,Saturation,Value), and take the union of both detection results in order to compensate the unreliability of the single color space approach. In YCbCr color space, a pixel is classified as a skin if two chrominance (Cb and Cr) values fall within the range specified by thresholds obtained empirically. In HSV color space, only H and S components are used for segmentation as V component of the skin region has a relatively larger variance depending on the light conditions and individual differences. Given input frame x. the binary skinmap x_{Map}, which segments skin/non-skin pixels, can be formulated as follows:

$$x_{Map(i,j)} = \begin{cases} 1(skin) & \text{,if } \begin{cases} \{x_{(i,j)Cb} \in [t_{Cb,l}, t_{Cb,u}], \\ x_{(i,j)Cr} \in [t_{Cr,l}, t_{Cr,u}]\} \text{ or} \\ \{x_{(i,j)H} \in [t_{H,l}, t_{H,u}], \\ x_{(i,j)S} \in [t_{S,l}, t_{S,u}]\} \end{cases} \\ 0(nonskin) & \text{,otherwise} \end{cases} \tag{1}$$

where $t_{k,l}$ and $t_{k,l}$ denote lower and upper threshold values of k color components, respectively.

2) Generative framework based eye detection : In this work, we adopt a generative framework based object detection scheme [5] to determine the location of eyes in real time via a two stage process where the first stage is specialized on locating facial regions from general background while the second stage is responsible for finding eye from the located facial region. This scheme requires development of separate likelihood detection models for face versus non-face and eye versus non-eye, which are learned by a GentleBoost method. During training the GentleBoost selects a set of the most discriminative Haar-like features at multiple positions and scales. This eye detection methodology is adopted since it fulfills following requirements: i) computational efficiency due to use of simple Haar-like features, allowing real time operation, ii) robust performance against complex background and variable illumination conditions.

In the first stage, the input frame is scanned at different pixel locations with different scales to obtain face versus non-face likelihood ratio. To enhance the accuracy of detection performance, we propose two modifications: i) the binary skinmap generated from earlier segmentation stage is used to reduce search regions for face, and ii) a novel RGB to grayscale conversion [7] is incorporated instead of standard NTSC conversion to generate the grayscale image input to this module. The conversion formula is given as follows:

$$x_{(i,j)Y} = \alpha \times x_{(i,j)R} + \beta \times x_{(i,j)G} + \gamma \times x_{(i,j)B} \tag{2}$$

where (α, β, γ), the weights corresponding to color channels, R,G, and B are defined as $(0.85, 0.10, 0.05)$, respectively. This conversion allows for an optimal face detection performance than the NTSC conversion by increasing contribution of red channel in grayscale image estimation, which takes the significant proportion in the skin-tone signal. Once face is detected, similar multi-scale iterative search is performed to locate eye regions within a facial region by evaluating eye versus non-eye likelihood ratio.

Overall, the automatic eye detection module combines results from both skin-tone detector and generative framework based detector to provide robust detection performance against geometric, illumination, and background variations.

2.2. Privacy enabling JPEG XR encoder module

The output of eye detection module contains pixel location information of both eyes in given input frame. In this work, we employ the privacy enabling JPEG XR encoder [2] to scramble macroblocks (MB: consists of 16x16 pixels) of input frame where eye region is located. This encryption module offers a cost-effective solution for de-identification, enabling intra-coding of high-resolution video content. Due to its simplified architecture and low memory footprint, it is able to facilitate real time operation in various environments. In addition, it produces video stream compatible with a worldwide standard, enabling potential large-scale adoption.

In order to achieve protection of privacy sensitive content, the privacy enabling JPEG XR encoder utilizes three encryption techniques to the transform coefficients of frequency subbands in MB basis. (summarized in Table 1). This subband adaptive technique provides an optimized balance between the level of security, the coding efficiency, and the computational complexity of scrambling tools.

3. EXPERIMENTAL RESULTS

The prototype of the proposed system is built by integrating a MATLAB implementation of the generative model framework, available

Subband	Scrambling technique
DC	**RLS** (Random Level Shift) : Shift DC coefficient value by a random integer number X, where $X \in [-2^{L-1}, 2^{L-1}]$
LP	**RP** (Random Permutation) : Rearrange LP coefficients within a macroblock in random order
HP	**RSI** (Random Sign Inversion) : Flip the sign bit of HP coefficients randomly

Table 1: Overview of JPEG XR subband adaptive scrambling

in the Machine Perception ToolBox (MPT) from the UCSD Machine Perception Laboratory [5] into the privacy enabling JPEG XR encoder module. Within MPT eye detection tool, the proposed skin tone segmentation module and the novel grayscale conversion module are embedded. The threshold parameters for skin tone segmentation are determined by inheriting the settings previously used in [8,9] as $(t_{Cb,l}, t_{Cb,u}, t_{Cr,l}, t_{Cr,u}) = (77, 127, 133, 173)$ given $Cb, Cr \in [0, 255]$, and $(t_{H,l}, t_{H,u}, t_{S,l}, t_{S,u}) = (0, 0.14, 0.2, 0.68)$ given $H, S \in [0, 1]$. Considering a tradeoff between visual security and bitstream overhead, the JPEG XR encoder module is operated in follow configuration: i) scrambling only applied to luma (Y) channel, ii) the RLS parameter L for DC coefficient is set to 8 (refer to Table 1), iii) the quantization parameter(QP) value of 15 is used for all three frequency subbands.

In order to validate the effectiveness of the proposed solution in terms of privacy preserving nature, we adopt evaluation methodology presented in [10]. In order words, sample images are encrypted using the proposed scheme with various scrambling window sizes, and automatic FR algorithms are applied to see if it prevents successful identification. For evaluation, we build a color image database, containing nearly frontal face images of 68 identities (68 gallery, 340 training, and 1710 probe images) manually selected from 'talking' and 'lighting' sets of CMU PIE database [11], which is publicly available. These images cover a wide range of facial variations in varied illumination conditions, ethnic groups, and lip movements. Sample images are normalized to 192x192 resolution by locating two eye coordinates using the MPT eye detection tool, followed by aligning, and cropping to place the center of eye regions on a specific pixel.

We made use of following two widely used benchmark FR algorithms: i) Principal Component Analysis [12] with nearest-neighbor classifier (PCA-NN), and ii) Local Binary Pattern [13] with nearest-neighbor classifier (LBP-NN). For similarity measurement, the Euclidean distance is used for PCA-NN, while the Chi-Square distance is used for LBP-NN. For LBP-NN, the LBP feature is extracted on the basis of 16x16 pixel blocks by sampling 8 equally spaced pixels on a circle of radius 2.

Fig. 2 illustrates different scrambling block sizes used in this experiment along with full facial encryption case. We examine two recognition scenarios as described in Fig. 3 where the selective encryption only occurs in probe images in the scenario 1, whereas it is applied to all gallery, training, and probe images in the scenario 2. These scenarios represent different types of privacy invasions, where the scenario 1 simulates an attempt that an attacker applies FR methods on common database (unaltered images) without prior knowledge about the encryption technique. On the other hands, the scenario 2 assumes that an attacker can reproduce similar alterations to images either by getting access to the scrambled database or by applying the same encryption technique.

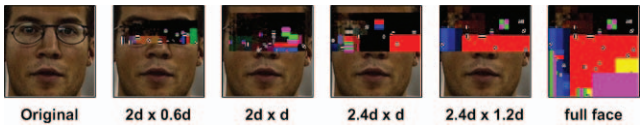

Fig. 2: Scrambling block size configurations represented with respect to the eye distance d

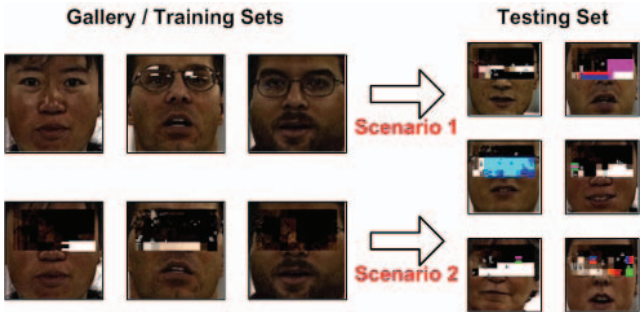

Fig. 3: Scrambling scenarios under consideration

Table 2 summarizes the rank 1 recognition (best match) results for various experimental setups. For non-protected probe images, PCA-NN and LBP-NN yield recognition rate of 71.1 and 88.2, respectively. The correct identification rate obtained by entire facial region scrambling results in below 5 percent for all cases, demonstrating the fundamental effectiveness of the utilized encryption module. In the scenario 1, the performance of both PCA-NN and LBN-NN decreases consistently as the size of protected eye region grows. Around 30 percent of recognition rate is realized for both methods with the largest block size 2.4d x 1.2d, demonstrating feasibility of the eye region based solution. It is worthwhile to mention that local feature based LBP-NN is relatively robust than global feature based PCA-NN since it can effectively take advantage of local information from non-protected regions. In the scenario 2, the overall recognition accuracies of both methods are even lower than those of the scenario 1. This fact indicates that the the proposed solution is robust against the attack that invokes the same alteration to both gallery and training sets.

Scrambled Block Size	Scenario 1		Scenario 2	
	PCA-NN	LBP-NN	PCA-NN	LBP-NN
Original	71.1	88.2	71.1	88.2
2.0d x 0.6d	56.4	79.1	25.5	54.9
2.0d x 1.0d	47.2	66.6	14.4	34.0
2.4d x 1.0d	35.9	47.0	10.9	16.4
2.4d x 1.2d	24.4	31.2	9.9	18.3
Full Face	3.2	1.3	3.0	4.85

Table 2: Face Recognition results for various scrambling block sizes and attack scenarios

Considering the privacy protection capability, the optimal eye region size is found to be 2.4d x 1.2d. This block size allows us to prevent successful identification of subjects in certain degree (in

our case, recognition rate remains between 9.9 and 31.2 percent) by only scrambling approximately a half size of regions compared to the method based on full face. This is huge advantage for deployment of the system into an application that requires real time operability and reduced computational power, given that scrambling operation typically requires a significant amount of computations, which has direct impact on production costs. For complexity analysis, we measure the average encoding delay to produce eye region protected image and fully protected image. On Core 2 Duo 2.53Hz CPU with 4GB RAM running Windows 7 operating system, the encoding delay per frame (192x192) reported for eye region solution and facial region solution are 34.94 ms (millisecond) and 46.20 ms, respectively. The reduced encoding delay observed with the eye region solution validates our claim.

The aforementioned results indicate that the proposed eye region based scheme effectively removes discriminative features in facial region to disallow successful identification at reduced computational costs. We believe this scheme can be considered as a feasible candidate for cost-effective privacy protection applications, where degrading the visual quality of the eye region can provide sufficient level of security. In addition, this approach can be used as a sub-mode within an adaptive framework that allows for reconfiguration of encryption block size depending on security/computation requirements imposed by users. Fig. 4 illustrates some output frames acquired by applying the proposed eye region scrambling scheme in two different experimental conditions.

Fig. 4: Eye region scrambling results using test images from CMU PIE database

4. CONCLUSIONS

In this paper, a novel eye region based privacy protection scheme is introduced by combining an automatic eye detection module with a privacy enabling JPEG XR encoder module. A probabilistic model of image generation is used in conjunction with a skin-tone detector to locate eye regions in real time. The incorporation of the novel RGB to grayscale conversion and the skin-tone segmentation into the eye detection module improves its robustness towards illumination and background variations. The utilized JPEG XR encoder effectively deteriorates the visual quality of located eye region at low computational costs. Experimentation results reported in this paper using two benchmark face recognition algorithms indicate that the proposed scheme successfully prevent correct identification of subject, by removing discriminative features in eye region. The proposed scheme effectively reduces computational complexity associated with the encryption process compared to the entire facial region based approach.

5. REFERENCES

[1] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *Circuits Syst. for Video Technol., IEEE Trans. on,*, vol. 18, no. 8, pp. 1168 –1174, aug. 2008.

[2] H. Sohn, W. De Neve, and Y. Ro, "Privacy protection in video surveillance systems: Analysis of subband-adaptive scrambling in JPEG XR," *Circuits Syst. for Video Technol., IEEE Trans. on,*, vol. 21, no. 2, pp. 170 –177, feb. 2011.

[3] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B.V.K. Vijayakumar, "Partial holistic face recognition on frgc-ii data using support vector machine," in *Computer Vision and Pattern Recognition Workshop(CVPRW), 2006. Conference on*, june 2006, p. 48.

[4] "ITU-T Rec. T.832 and ISO/IEC 29199-2 : Information technology - JPEG XR image coding system - part 2: Image coding specification," 2009.

[5] I. Fasel, B. Fortenberry, and J. Movellan, "A generative framework for real time object detection and classification," *Comput. Vis. Image Underst.*, vol. 98, pp. 182–210, April 2005.

[6] Sanjay Kr. Singh, D. S. Chauhan, Mayank Vatsa, and Richa Singh, "A robust skin color based face detection algorithm, tamkang," *Journal of Science and Engineering*, vol. 6, pp. 227–234, 2003.

[7] J. Lu and K.N. Plataniotis, "On conversion from color to gray-scale images for face detection," in *Computer Vision and Pattern Recognition Workshop(CVPRW), 2009. Conference on*, june 2009, pp. 114 –119.

[8] D. Chai and K.N. Ngan, "Face segmentation using skin-color map in videophone applications," *Circuits Syst. for Video Technol., IEEE Trans. on,*, vol. 9, no. 4, pp. 551 –564, jun 1999.

[9] Y. Wang and B. Yuan, "A novel approach for human face detection from color images under complex background," *Pattern Recognition*, vol. 34, no. 10, pp. 1983–1992, 2001.

[10] F. Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," *SPIE*, vol. 8063, no. 1, pp. 806302, 2011.

[11] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database of human faces," Tech. Rep. CMU-RI-TR-01-02, Robotics Institute, Pittsburgh, PA, January 2001.

[12] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 28, no. 12, pp. 2037–2041, 2006.