Face Occlusion Detection Using Skin Color Ratio and LBP Features for Intelligent Video Surveillance Systems

Pengfei Ji
Chongquing Key Lab. of
Computational Intelligence,
Chongquing University of Posts
and Telecommunications, China
Email: jipfchn@163.com

Yonghwa Kim, Yong Yang
Dept. of Information &
Communication Engineering,
Inha University, Incheon 402-751,
Korea
Email: yonghwa@inha.edu

Yoo-Sung Kim
Dept. of Information &
Communication Engineering,
Inha University, Incheon 402-751,
Korea
Email: yskim@inha.ac.kr

Abstract—A face occlusion detection scheme which is based on both skin color ratio (SCR) and Local Binary Pattern (LBP) feature, is proposed. The proposed method mainly consists of four steps: foreground extraction, head detection, feature extraction, and occlusion detection. First, foreground is extracted by codebook background subtraction algorithm. Then, the head region is located using HOG head detector. After that, the skin-color ratio and LBP feature are extracted. Finally, SVM is trained based on LBP feature. The recognition result of SVM and the result of skin-color ratio feature are merged by weighted voting strategy, and then occluded faces are classified as three categories: concealed, partially concealed, and visible. Experimental results show that the proposed detection system can achieve desirable results in intelligent video surveillance systems.

I. INTRODUCTION

Nowadays, intelligent video surveillance system has been a very active research field, and has many practical applications, such as face identification, behavior recognition, abnormal event detection, and so on. It greatly reduced the manual inspection and verification to improve the effectiveness. Face occlusion detection is an essential component of intelligent video surveillance system. This is necessary for the prevention of suspicious behavior in security zones, such as bank, government, and other sensitive areas. With the increasing requirements for security, face occlusion detection has become a hot research topic in pattern recognition and image processing. If the surveillance system can automatically detect a person with concealed face, it would be helpful for making more secure society. For example, the system can be widely used for automated teller machine (ATM) security.

Concealed face detection is a very challenging task in image understanding field due to the complexity of environment condition, such as low resolution, illumination changes, and object movement, etc. Under complex conditions, there is no guarantee of image quality, so face occlusion detection becomes extraordinary difficult. Furthermore, the designed system requests reduced computation and high reliability, so that it can be able to realize the real-time processing requirement. Although various methods have been proposed to detect occluded face, but most of them are for ATM, this fact leads to a problem that these methods cannot be used directly for video surveillance systems.

In order to solve the above-mentioned problems, we proposed an efficient approach that combines skin color ratio feature and texture feature for concealed face detection. The method consists of the main four steps: (1) Extraction of moving foreground by using codebook model which is a real-time segmentation algorithm, (2) Head detection based on HOG head detector, which only focus on the foreground moving region, (3) Skin color feature and LBP feature are extracted, (4) The two assessment methods are fused by weighted voting and the final recognition result is obtained.

The main contributions of this work can be stated as follows: (1) A novel framework of face occlusion detection in the application of the real-world video surveillance environment is proposed, (2) The skin color feature and texture feature with LBP operator are merged by weighted voting strategy in order to improve detection performance.

The rest of this paper is organized as follows: Section 2 introduces the related works on face occlusion detection. Section 3 describes the details of the proposed method. Section 4 presents the experimental results and analysis. Section 5 draws a conclusion with the future plan.

II. RELATED WORKS

Generally, SCR-based methods [1]–[4] adopt various skin models, extract color information, and then calculate skin ratio. Lin *et al.* proposed a method based on the combination of ellipse fitting and skin area ratio to determine whether the face is concealed or not [1]. Kim *et al.* presented a face occlusion verification method that combines a shape-based face detection and skin color [2]. Zhang *et al.* used Adaboost to combine skin color detection and face templates matching to produce the strong classifier for occlusion detection [3]. Charoenpong *et al.* proposed a face occlusion detection method from a viewpoint of face by skin color ratio of two parts of head re-

gion [4]. These methods present an effective detection system for bank ATM application. However, it can only handle the detection problem in which the target is close to the camera. Therefore, these methods cannot be directly used in intelligent video surveillance environments.

On the other hand, learning-based methods [5]–[9] require feature extraction to train the classification model, and then achieve classifying recognition. Yoon *et al.* proposed a method based on principal component analysis (PCA) and support vector machine (SVM). By applying PCA and SVM to extract the feature points and classify that the feature points are near the normal or concealed face [5]. Min *et al.* presented a novel learning based method for scarf detection. The method exploits Gabor wavelet feature and SVM classifier to distinguish the normal and scarf faces [6]. Priya *et al.* proposed a Mean Based Weight Matrix (MBWM) algorithm, which is an improved form of LBP, and the SVM classifier to detect face occlusion [7]. Bianco *et al.* designed a recognizability of concealed face system, in which Adaboost algorithm is taken to build the facial feature detectors [8]. Kim *et al.* also proposed a face occlusion detection method based on gradient map and SVM, used especially for the partially concealed face [9]. Although good results are reported in these methods, the robustness cannot be guaranteed under complex conditions. In particular, surveillance systems cannot always produce high quality images. Therefore, the proposed method should handle both high quality and low quality images to meet the result of robustness.

In addition, component-based methods [10], [11] use facial component regions to detect face occlusion. Suhr *et al.* proposed a face occlusion detection system, which combine the eye-mouth combinations and geometric constraints [10]. Eum *et al.* presented a face recognizability evaluation to use facial components and the exceptional occlusion handling (EOH) [11]. This method is insensitive to illumination condition and achieves the robustness against facial postures, but it needs a high computational cost.

Nowadays, deep learning is a hot topic in machine learning, and CNN is one of deep learning methods, which can learn hierarchical features from low-level data [12]. Xia *et al.* proposed a robust and effective facial occlusion detection method based on CNN and multi-task learning [13]. The trained CNN model is applied in extracting facial features and predicting face occlusion. However, the disadvantages are complex structure, and it requires large-scale and complete training data.

III. FACE OCCLUSION DETECTION SCHEME USING SKIN COLOR RATIO AND LOCAL BINARY PATTERN

This section mainly introduces the details of our proposed method. The processing procedure includes the following steps: foreground extraction, head detection, feature extraction, and occlusion detection. The proposed system architecture is shown in Fig. 1.

A. Foreground extraction

Extracting foreground accurately is the first step for face occlusion analysis. By foreground extraction, the moving foreground is segmented in the video sequence. There are many ways for foreground segmentation, such as Gaussian Mixture Model (GMM) and background subtraction. Compared with the above methods, the codebook model can detect foreground more accurately and efficiently, and handle scenes containing moving backgrounds or illumination variations [14]. It is a kind of adaptive background subtraction algorithm by building a codebook for each pixel. So the information of the foreground areas in current frame is obtained with codebook model. Some examples of foreground segmentation using codebook model are shown in Fig. 2.
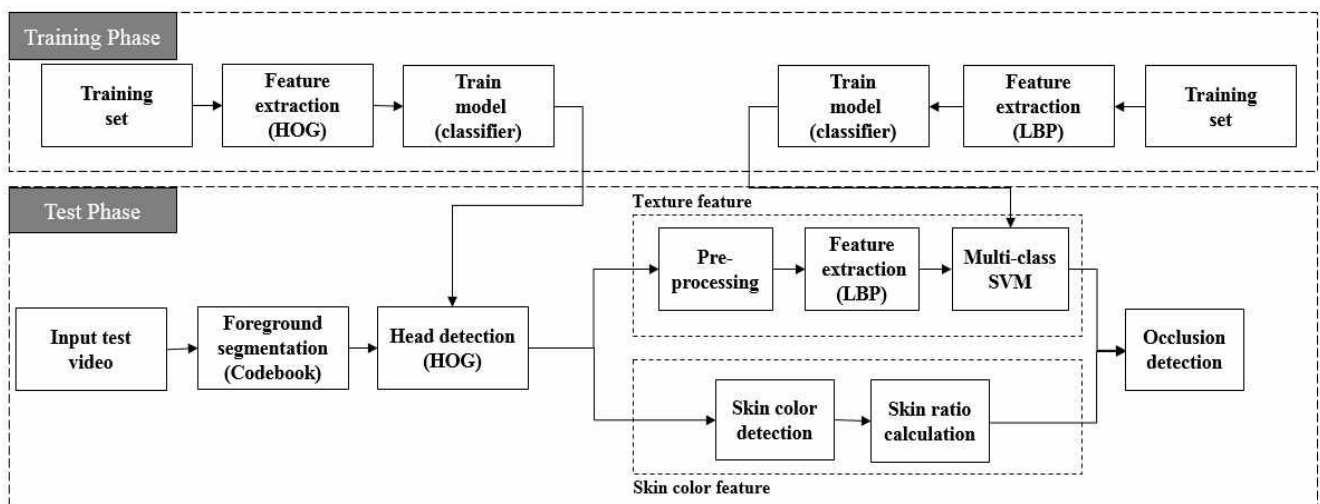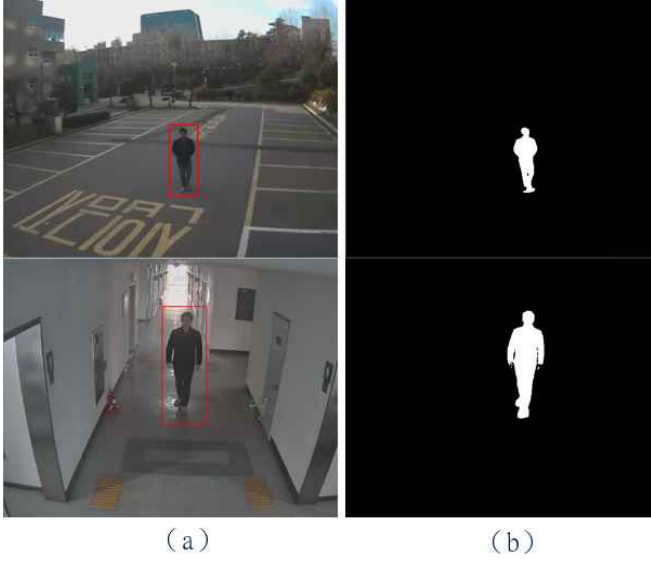

Fig 1. The framework of face occlusion detection system

Fig 2. Foreground extraction results: (a) original (b) Codebook

B. Head detection

In our method, head detection is accomplished by combining HOG and SVM method proposed by Dalal and Triggs [15] and the part-based models proposed by Felzenszwalb *et al.* [16]. HOG can well describe the edge or local shape information, so it is suitable to depict characteristics of head. Meanwhile, this method is capable of handling head detection with challenging illumination conditions and low resolution.

While the HOG is one of the popular descriptors in image processing, it needs a high computational cost. To address the problem, GPU technique is used to accelerate process by using a parallel implementation of the HOG algorithm. Comparing CPU implementation, GPU based on parallel architecture has a better computational performance [17], [18]. The result of head detection demonstrates that our implementation using GPU can achieve a speedup of over 5 times, which allows for real-time detection. Moreover, the scanning area only focuses on the moving foreground region, and the probability of false and failure detections is significantly reduced. Examples of head detection are shown in Fig. 3.

Fig 3. Some results of head detection

C. Feature extraction

The analysis of skin color is adopted to discriminate skin and non-skin pixels. An unconcealed face includes a large number of skin color pixels. The whole process of skin color detection consists of two steps: color space transform and skin ratio calculation.

1) Color Space Transform: RGB is a basic color model. Because of the sensitivity to illumination variations, it is hardly separated from chrominance and luminance information [19]. Compared with RGB, YCbCr has a clear distinction between chrominance and luminance components. Thus the first step is to convert color space from RGB to YCbCr representation. Using (1), we transform RGB to YCbCr color space.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.533 & 24.966 \\ -37.797 & -74.203 & 112.0 \\ 112.0 & -93.786 & -18.241 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

2) Skin ratio calculation: In chromatic space, the skin color mainly focuses on a narrow range. Obviously, we can determine the skin pixel by judging whether the YCbCr value is in a certain range. As shown in (2), the skin ratio is defined as the skin pixels of the head region.

$$S = \frac{number\ of\ skin\ pixels}{total\ number\ of\ head\ region\ pixels} \quad (2)$$

Where, S is the skin ratio. Because it is hard to ensure the detection result with only one frame, we can utilize multiple frames to ensure high accuracy. Consequently, the calculation of the ratio comes out from the combination of multi-frames.

For the reason that the camera will be a certain distance from the people in real scenes, it makes the face appear smaller and vaguer. However, with the movement of the body, the size of head region is also varying constantly. To assess the impact, scoring method is used for the evaluation process. Large size can obtain a higher score than head image of small size, but the performance is going to reach a specific limitation [20]. Given a specific value μ, it is suitable for the width and height. Using (3), we represent the score related to the size.

$$Q = min\left\{1, \frac{width}{\mu} \times \frac{height}{\mu}\right\} \quad (3)$$

Therefore, this method, which applies the head size as criterion, is used to realize multi-frame selection. When the head size is larger than a preset threshold, these frames are selected in video sequence. Let N be the number of frames, and we obtain the final ratio to integrate the scores. It can be expressed as

$$R = \frac{\sum_{i=0}^{N} S_i Q_i}{\sum_{i=0}^{N} Q_i} \quad (4)$$

where, S_i is the skin ratio of i-th frame, and Q_i is the score of i-th frame.

Based on fusion strategy, we need to know the classification result for the detected N frames. Thus we introduce two ratio threshold parameters τ_1 and τ_2, which will be discussed in the following section. For every frame, the

occlusion result can be determined by comparison of two parameters. Let X_i be the number of detection result for each class in all frames, $i= \{1,2,3\}$, which denote concealed, partially concealed, and visible, respectively. Assume p_i is the likelihood for every class, it is calculated by using (5).

$$p_i = \frac{X_i}{N} \qquad (5)$$

In practice, the above skin color method is sensitive to light condition. And when face covers by objects of colors similar to the skin color, its effect is not ideal. To overcome this problem, we extract texture feature with LBP operator [21]. As a popular texture descriptor, it describes well the image texture information. Therefore, we can effectively distinguish them by the texture. The operator labels the pixels of an image by thresholding the neighborhood of each pixel with the center value and forming the result of binary pattern. Then, a pattern code is computed as follow

$$LBP_{(P,R)} = \sum_{n=0}^{P-1} s(g_n - g_c)2^n \qquad (6)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \qquad (7)$$

where, g_c corresponds to the gray level of the center pixel, g_n to the grey values of the neighbor pixels. The notation (P, R) indicates P sampling points on a circle of R radius. This paper uses the uniform LBP, and the selection of P and R will be discussed in the next section.

The feature extraction process includes the following steps. The first step is pre-procession, such as normalization and histogram equalization. Second, we divide the image into several sub-blocks, and the feature histogram is extracted from each sub-block. Then, all the local histograms are concatenated into a feature vector. The more the block number is, the higher the computational complexity is. Therefore, a suitable block number should be selected. And this issue is explored future in the next section. Fig. 4 shows the process of feature extraction.
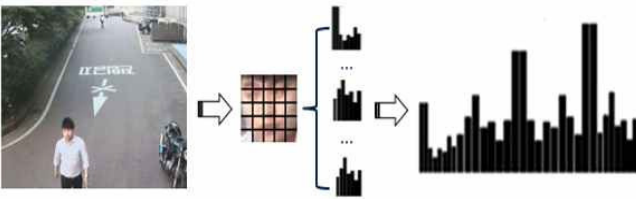

Fig 4. LBP feature extraction

In this paper, the detection of concealed faces can be considered as a multi-class problem: concealed, partially concealed, visible. Therefore, we use a multi-class SVM to fulfill the concealed face classification. Furthermore, assume X_i is the number of detection result for each class in all frames, which represent concealed, partially concealed, visible respectively, $i=\{1,2,3\}$. Let p_i' is the likelihood for each class and it is calculated by using (8).

$$p_i' = \frac{X_i}{N} \qquad (8)$$

D. Occlusion detection

The system uses the idea of decision level fusion to improve performance. On the phase of decision fusion, the weighted voting strategy is used to fuse the result for all levels to get target detection output. First, we determine whether it is the concealed face by skin color ratio in the skin color method, and whether it is the concealed face based on LBP. Then we assign different weight according to the contribution of various assessment methods. Using (9), we calculate the weighted vote result.

$$W_i = p_i w_1 + p_i' w_2 \qquad (9)$$

where W_i is the weighted results, $i=\{1,2,3\}$, W_i denote the weighted result in three classes: concealed, partially concealed, visible, respectively. p_i and p_i' can be obtained by using equation (5) and (8). w_1 and w_2 are the weights of i-th method, respectively, and they are measured by

$$w_x = \frac{n_x}{\sum_{x=1}^{k} n_x} \qquad (10)$$

where k is the number of methods. Because the system has two methods, namely, k=2. n_x is the recognition rate based on various assessments on test video set, $x = \{1,2\}$. The idea of weight distribution comes from [22]. Eventually, the largest W_i will be chosen to determine the concealed face classification.

IV. EXPERIMENTS

To evaluate the system performance, experiment inputs are taken on a dataset with 138 video samples, which are collected from indoor or outdoor surveillance cameras. The video set includes three classes: concealed, partially concealed, visible. We chose 60 video samples as the training set, and the rest is used as test set. The video sequence has a resolution of 720×576 at 15 frames per second. The face covers by some objects such as hat, mask, sunglasses, book, etc. The sample set contains the different times of the day. The experiment is carried out with an Intel(R) Core(TM) i5-3230M CPU @2.6GHz and NVIDIA GeForce 710M. Fig. 5 illustrates some examples of face occlusion in surveillance cameras.

Fig 5. Examples from surveillance camera for the concealed face

A. Head detection

In order to evaluate the head detection accuracy and acceleration effect based on GPU implementation, experiments are taken. In experiments, head samples are collected from surveillance cameras, and then the head detector is trained on manually annotated data. The data set has approximately 5,000 head images as positives training samples and randomly chosen windows that do not contain any head area as negative training samples from surveillance videos. The size will vary depending on the distance between the people and camera, and is normalized to 64×64 pixels.

The algorithm is tested on video samples with different resolutions. Table I shows the processing time with the GPU and CPU implementations. It can be concluded that GPU implementation speeds up more than 5 times for the processing time over CPU implementation. And the GPU implementation shows more significant effect of acceleration for higher resolution image. On the other hand, to evaluate the accuracy of head detection, we carried out the experiments on several video clips. The experimental results are summarized in Table II. Hit rate can be calculated as the number of detected heads divided by total number of heads, while the false alarm rate is the number of false heads divided by total number of heads [23]. The results show that our approach has high detection rate and can reduce the false alarm rate effectively.

TABLE I.

THE COMPARISON OF PROCESSING TIMES FOR DIFFERENT RESOLUTION

Resolution	CPU Time(ms)	GPU Time(ms)	Overall Speedup
320*240	126.10	24.63	5.12
720*576	653.59	97.08	6.73
1280*1024	2083.33	299.40	6.94

TABLE II.

ANALYSIS OF HIT RATE AND FALSE ALARM RATE

Total frames	1462
Total number of heads	1512
Number of detected heads	1373
Number of false heads	16
Hit rate(HR)	90.8%
False alarm rate(FAR)	1.0%

### B. Threshold determination in skin color ratio

The skin ratio threshold is important factor that directly affects the accuracy of face occlusion detection. In order to achieve higher results, the optimal threshold should be determined. In our experiment, we use training video set to decide the optimal threshold. The result is illustrated in Fig 6. As seen in the Fig 6(a), it shows that along with the increasement of skin color ratio, the recognition rate of visible face is declining gradually, and partially concealed face is increasing at the same time. There is a peak value of average recognition rate at 0.34. On the other hand, a conclusion can be drawn from Fig. 6(b), which is similar to the above conclusion, and the peak point appears at 0.10. Then, let τ_1 and τ_2 denote two thresholds respectively, that is, the ratio thresholds are set to be $\tau_1 = 0.34$ and $\tau_2 = 0.10$.

If the ratio is more than τ_1, this case is called visible. If the ratio is less than τ_2, it's called concealed. In addition to this, all other cases belong to partially concealed. Finally, we have tested its performance on the test video set. The average accuracy can reach to 87 %.
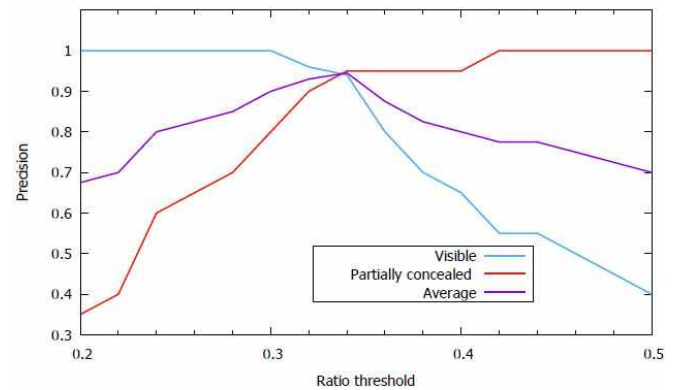
(a)
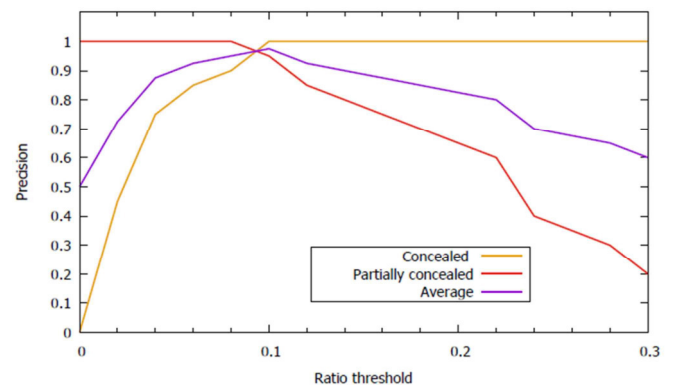
(b)

Fig 6. Precision of various ratio threshold on training video set.

C. Parameter determination for LBP

In order to analyzing the influences of some parameters, a series of experiments is carried out. In our experiment, we manually crop concealed & unconcealed face images from the training video set. All these concealed faces are normalized to 90×90 pixels. Then multi-class SVM is trained to detect the concealed face. In testing phase, we use the trained classifier to detect whether the face is concealed or not in every frame.

Taking into account the feature dimension, we choose uniform patterns of LBP operator. We adopt cross validation to determine some parameters. Let $N*N$ and R denote the number of blocks and sample radius. In the experiment, the parameters are set to $N=\{1,2,3,5,6,9\}$, $R=\{1,2\}$. Fig. 7 gives the comparison with the recognition rate of different parameters. From this figure, it indicated that the recognition rate increased with the incensement of block number. But the feature dimension has doubled and redoubled when the number of blocks increased. Therefore, we select $N=5$, $R=1$ as the best choice, which can both maintain a relatively high precision and lower dimension. The trained classifier can be used to classify the unknown face samples. Similarly, we have tested it on test video set. The average accuracy can reach to 70%.
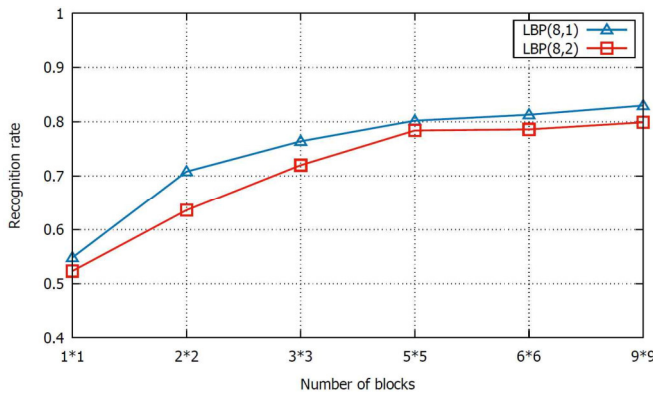

Fig 7. Comparison with the recognition rate of different parameter

D. Fusion strategy

For the purpose of comparing these two methods, the fusion experiment is carried out. In the fusion strategy, the average accuracy is above 95% by combining the above two methods in the test video set. Fig. 8 summarizes the confusion matrix. As can be seen from the chart below, our fusion strategy can improve the recognition rate obviously. In SCR approach, the concealed face was often classified as partially concealed face, primarily because of the individual difference of human beings, such as short and thin hair, which largely affects the skin color ratio. In LBP method, the partially concealed face has the poor accuracy of recognition, and this is due to the diversity of partially concealed face. Therefore, the fusion strategy verifies its effectiveness.

V. CONCLUSIONS

This paper proposed a novel scheme to detect concealed face for intelligent video surveillance systems. It combines texture feature & skin color feature to detect concealed face. To ensure the accuracy, we have studied a multi-frame detection algorithm. There are two parts in this method. First, we introduced and analyzed the codebook foreground segmentation model, and the head re-
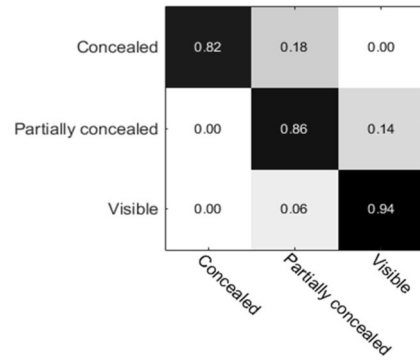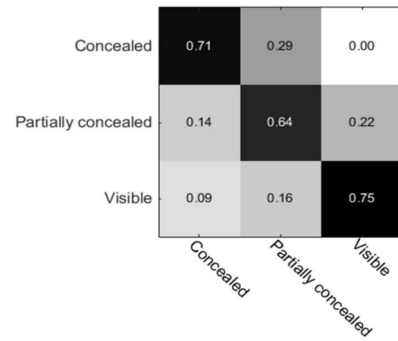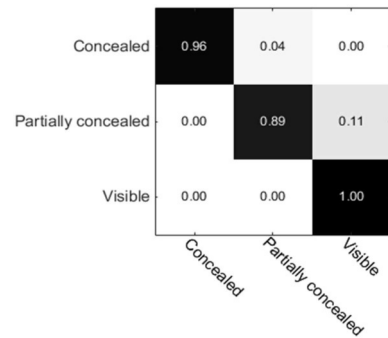
Fig 8. Confusion matrices for three occluded cases on test video set.
(a) SCR (b) LBP (c) Fusion

gion is located by using HOG with GPU implementation. Then the decision fusion is used to validate concealed face by combining LBP and skin color. In the future work, we are planning to improve overall system performance through robust method for feature extraction, and solve head pose estimation problem.

REFERENCES

[1] D. T. Lin and M. J. Liu, "Face occlusion detection for automated teller machine surveillance," in Advances in Image and Video Technology. Springer, pp. 641–651, 2006.

[2] G. Kim, J. K. Suhr, H. G. Jung, and J. Kim, "Face occlusion detection by using b-spline active contour and skin color information," in 11th International Conference on IEEE Control Automation Robotics & Vision (ICARCV), pp. 627–632, 2010.

[3] X. Zhang, L. Zhou ,T. Zhang, and J. Yang, "A novel efficient method for abnormal face detection in ATM," in International Conference on. IEEE Audio, Language and Image Processing (ICALIP, pp. 695–700), 2014.

[4] T. Charoenpong, C. Nuthong, and U. Watchareeruetai, "A new method for occluded face detection from single viewpoint of head," in 11th International Conference on IEEE Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1-5, 2014.

[5] S. M. Yoon and S. C. Kee, "Detection of Partially Occluded Face using Support Vector Machines," in IAPR Workshop on Machine Vision Applications, pp. 546-549, 2002.

[6] R. Min, A. D'Angelo, and J. L. Dugelay, "Efficient scarf detection prior to face recognition," in Proceedings of the 18th European Signal Processing Conference, pp.259-263, 2010.

[7] G. N. Priya, and R. S. D. W. Banu, "Detection of occluded face image using mean based weight matrix and support vector machine," in Journal of Computer Science, Vol. 8, no. 7, pp.1184-1190, 2012.

[8] S. Bianco, G. Ciocca, G. C. Guarnera, A. Scaggiante, and R. Schettini, "Scoring recognizability of faces for security applications," in Proc. SPIE 9024, Image Processing: Machine Vision Applications VII, 90240L, 2014.

[9] J. Kim, Y. Sung, S. M. Yoon, and B. G. Park, "A new video surveillance system employing occluded face detection," in Lecture Notes in Computer Science, vol. 3533, pp. 65-68, 2005.

[10] J. K. Suhr, S. Eum, H. G. Jung, G. Li, G. Kim and J. Kim. "Recognizability assessment of facial images for automated teller machine applications." in Pattern Recognition, vol. 45, pp. 1899-1914, 2012.

[11] S. Eum, J. K. Suhr, and J. Kim. "Face Recognizability Evaluation for ATM Applications with Exceptional Occlusion Handling," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.82-89, 2011.

[12] C. B. Jin, S. Z. Li, T. D. Do, and H. Kim, "Real-Time Human Action Recognition Using CNN Over Temporal Images for Static Video Surveillance Cameras," in Advances in Multimedia Information Processing, Vol. 9315, pp. 330-339, 2015.

[13] Y. Xia, and F. Coenen, "Face Occlusion Detection Based on Multi-task Convolution Neural Network," in Proceedings of 12th International Conference on IEEE Fuzzy Systems and Knowledge Discovery (FSKD), pp.375-379, 2015.

[14] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," in Real-Time Imaging, Vol. 11, no. 3 pp. 172–185, 2005.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886 –893, 2005.

[16] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.

[17] M. Hirabayashi, S. Kato, M. Edahiro, K. Takeda, T. Kawano, and S. Mita, "GPU Implementations of Object Detection using HOG Features and Deformable Models," in Proceedings of the IEEE International Conference on Cyber-Physical Systems, Networks, and Applications, pages 106-111, 2013.

[18] M. Szkudlarek and M. Pietruszka, "Fast GPU and CPU computing for Head Position Estimation," in Proceedings of the Federated Conference on Computer Science and Information System, pp. 231-240, 2015. DOI: 10.15439/2015F410.

[19] D. Lee, J. Wang, K.N. Plataniotis, "Contribution of skin color cue in face detection applications," in: M.C. Emre, B. Smolka (Eds.), Advances in Low-Level Color Image Processing, Springer, Netherlands, pp. 367–407, 2014.

[20] K. Nasrollahi and T. B. Moeslund, "Complete face logs for video sequences using quality face measures," in IET International Journal of Signal Processing, vol.3, no. 4, pp. 289–300, 2009.

[21] T. Ojala and M. Pietikainen, "Unsupervised Texture Segmentation Using Feature Distributions," in Pattern Recognition, vol. 32, pp. 477-486, 1999.

[22] Z. Zhang, J. Iria, C. Brewster, and F. Ciravegna, "A comparative evaluation of term recognition algorithms," in Proceedings of the sixth international conference of Language Resources and Evaluation (LREC), pp. 2108-2113, 2008.

[23] Z. Zhang, H. Gunes and M. Piccardi, "Head detection for video surveillance based on categorical hair and skin colour models," in Proceedings of IEEE International Conference on Image Processing (ICIP), pp. 1137–1140, 2009.