Face Occlusion Detection by using B-spline Active Contour and Skin Color Information

Gahyun Kim, Jae Kyu Suhr, Ho Gi Jung, and Jaihie Kim

School of Electrical and Electronic Engineering, Yonsei University,
Biometrics Engineering Research Center

{ghrapture, lfisbf, hgjung, jhkim}@yonsei.ac.kr

Abstract— **This paper proposes a face occlusion verification method for an automated teller machine (ATM) application. The proposed method mainly consists of three steps. Firstly, a head and shoulder shape is detected by applying B-spline active contour to motion edges. This motion edge is generated by a kurtosis-based frame selection and distance transformation-based motion edge detection. Secondly, a face area is estimated by fitting an ellipse to the detected head and shoulder shape. Finally, occlusion of the face area is determined by measuring skin color area ratio (SCAR) of whole face area and facial component areas. Experimental results show that the proposed head and shoulder detection method has 94.8% detection rate even though there are various types of severe occlusions in faces, and the proposed occlusion verifier has 86.7% verification rate.**

Keywords— **face detection, face occlusion, face verification, B-spline activve contour, skin color**

I.   INTRODUCTION

The detection and recognition of partially occluded face have been extensively researched since they can be used in a numerous applications such as biometrics and visual surveillance [1]. There is one more important research topic when dealing with an occluded face, which is face occlusion verification. This research determines whether a face is occluded or not and which part of face is occluded. It can be used for an automated teller machine (ATM) application [2]-[4].

The process of verifying face occlusion consists of two main algorithms. One is face detection and the other is occlusion verification. Face detection algorithms for this task should be robust against partial occlusions and there are two approaches: facial component-based and shape-based. Facial component-based approach [5],[6] detects facial components such as eyes, nose and mouth and determines a face area based on the component detection result. This approach provides face locations as well as facial component locations. However, this can have difficulties when dealing with non-frontal faces and it is hard to detect facial components in a low resolution image. Shape-based approach [4] detects faces based on the knowledge of head, neck and shoulder shapes. Since these shapes are not severely changed with different face pose, it can detect non-frontal faces even with a low resolution image.

However, it needs an additional effort to determine locations of facial components.

Occlusion verification algorithms can be categorized into three approaches: classifier-based, component presence-based and skin color area ratio (SCAR)-based. Classifier-based approach [7] determines face occlusion by using a pattern classification method. The method in [7] divided a face area into upper and lower parts, and trained occluded parts by using principal component analysis (PCA) and support vector machine (SVM). This method works well with frontal and aligned face images. However, its performance can be severely degraded when dealing with non-frontal or miss-aligned face images. Component presence-based approach [3] detects facial components in a face area and determines face occlusion based on the presence of facial components at appropriate locations. This can exactly determine which components are occluded. However, it can hardly be used with a low resolution face image. SCAR-based approach [4] extracts skin color pixels in a face area and determines face occlusion by calculates a ratio between the number of skin color pixels and the number of non-skin color pixels. This is robust against pose, alignment and image resolution, but its performance depends on illumination condition and camera characteristics.

This paper proposes a new face occlusion verification method that combines a shape-based face detection and SCAR-based occlusion verification. This method consists of three steps: 1) B-spline active contour-based head and shoulder detection, 2) ellipse fitting-based face localization, 3) global and local SCAR-based face occlusion verification. Firstly, a frame difference image is generated with an appropriate time interval which is determined by a kurtosis-based frame selection. Moving edges are detected by combining a distance transformation of frame difference image and edge detection result of current frame. Secondly, an exact face area is determined by fitting an ellipse to points on head shape of B-spline active contour estimated in the first step. Lastly, SCARs are calculated in a whole face area, eye area and mouth area, and then face occlusion verification is conducted by using these SCARs. Experimental results show that the proposed head and shoulder detection method has 94.8% detection rate even though there are various types of severe occlusions in faces, and the proposed occlusion verifier has 86.7% verification rate.

This paper is organized as follows: Section 2 introduces B-spline active contour-based head and shoulder detection method. Section 3 describes an ellipse fitting based face localization and SCAR-based face occlusion verification.. In Section 4, experimental results and analysis are presented. Finally, conclusions are given in Section 5.

II.   HEAD AND SHOULDER DETECTION

This section introduces a head and shoulder detection method that applies B-spline active contour to motion edges produced by frame difference image. One of the most widely used method for generating frame difference images is three-sigma rule [8]. This method can be simply implemented and its computational cost is low. However, it produces false difference pixels when there is no moving object. Therefore, we propose a kurtosis based frame selection method to find an appropriate frame interval that produces enough object motion. Kurtosis   in (1) is a measure of the peakness of the probability distribution [9].

$$\gamma_2 = \frac{\mu_4}{\sigma^4} \tag{1}$$

μ_4 is the fourth moment of mean and σ is a standard deviation. The more γ_2 is close to 3, the more the distribution is similar to normal distribution. A distribution that is more outlier-prone than the normal distribution has γ_2 greater than 3, and γ_2 is less than 3 in case of a distribution that is less outlier-prone than the normal distribution. Since distribution of intensity difference between consecutive images obtained from a static camera follows white Gaussian noise (WGN) in case where there is no moving object, its kurtosis is close to 3. Fig. 1 (a) and (c) show the frame difference image with no moving object and its distribution, respectively. In this case the kurtosis is 3.49. Whereas, the distribution does not follow WGN and kurtosis increases when moving objects appear. Fig. 1 (b) and (d) show the frame difference image with a moving object and its distribution, respectively. In this case the kurtosis is 45.31. Therefore, we use the kurtosis of frame difference to determine frame interval based on the amount of motion. If the kurtosis is larger than a pre-determined threshold (λ =10), it is considered that there is enough amount of motion so that the head and shoulder detector is applied to the frame difference image. Otherwise, the head and shoulder detector is not applied since it is considered that there is not enough motion.

If B-spline active contour is directly applied to the frame difference images, the resulting contours are likely to show a lag as shown in Fig. 2. In this figure, (a) and (b) show the binarized frame difference image and corresponding head and shoulder countour, respectively. This lag is due to the thick frame differences caused by a large amount of object motion. Therefore, we detect moving edges, $B(i, j)$ by combining a distance transformation of frame difference image, $DT(i, j)$ and edge detection result of current frame, $E(i, j)$ as shown in (2).
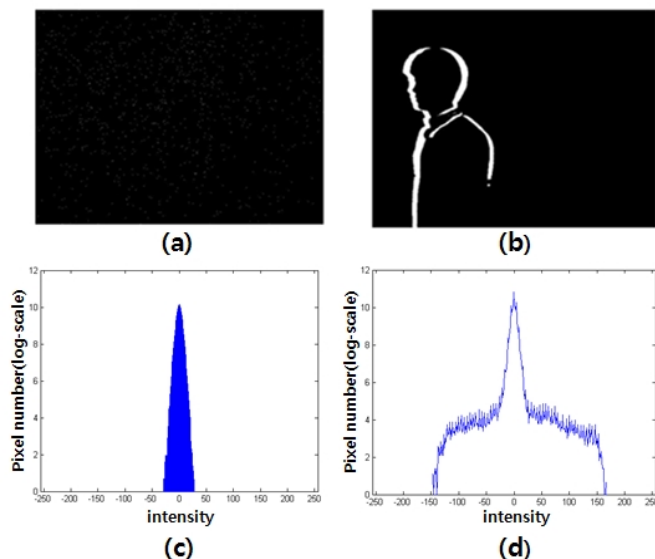

Figure 1. (a) and (b) are frame difference images without and with a moving object, respectively. (c) and (d) are log-scale histograms of the frame difference images without and with a moving object, respectively. Kurtosis of (c) and (d) are 3.49 and 45.31, respectively.

$$B(i, j) = \begin{cases} 1, & E(i, j) = 1 \quad \wedge \quad DT(i, j) < \theta \\ 0, & otherwise \end{cases} \tag{2}$$

$E(i, j) = 1$ means that pixel at (i, j) is classified as an edge pixel and θ is a pre-determined threshold which is set to 5. Fig. 3 shows the motion edge extraction process in detail. In this figure, (a), (b) and (c) are a frame difference image, its distance transformed image and edge detection result of current frame, respectivley. (d) shows the final motion edge extraction result by combining (b) and (c) via (2). It can be easily noticed that the motion edges in current frame are correctly extracted.

Since the head and shoulder detection result can be erroneous when head and shoulder shape partially appears in images, we need to determine if the entire head and shoulder shape appears in images. This is conducted by using the minimum and maximum x coordinates of motion edges. The entire shape is considered to be inside of images when the minimum x-coordinate of the motion edges is larger than five and its maximum x-coordinate is smaller than the value of image width minus five.

Figure 2. A lag problem of head and shoulder detection. (a) and (b) show the binarized frame difference image and corresponding head and shoulder countour, respectively.
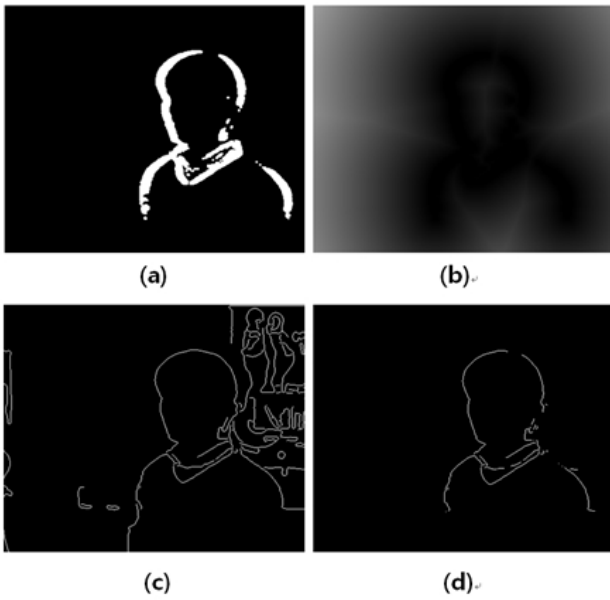
Figure 3. Motion edge detection procedure. (a), (b) and (c) are a frame difference image, its distance transformed image and edge detection result of current frame, respectivley. (d) shows the final motion edge extraction result by combining (b) and (c).

Once the entire head and shoulder shape appears in images, it is necessary to generate a mean shape template. A mean shape template is an initial shape of B-spline active contour which is used as a seed for contour fitting. To generate a mean shape template, we assume two conditions. One is that the motion edges are extracted from a single moving person and the other is that the height of head is 60% of the width of shoulder as shown in Fig. 4. Under these two assumptions, the center location (x and y) and scale factor (s) of mean shape template are chosen as (3).

$$\begin{cases} x = x_{mid} \\ y = y_{top} + 0.6w \\ s = 0.75 \times w / h_{template} \end{cases} \tag{3}$$

x_{mid}, w and $h_{template}$ are x-coordinate of motion edge center, width of shoulder and height of the template, respectively. The orientation of mean shape template is set to zero degree. The scale factor (s) is set to 75% of shoulder width (s) to make a mean shape template slightly larger than a real head and shoulder size. This larger mean shape template size can avoid contour fitting to be disturbed by motion edges within a face area.
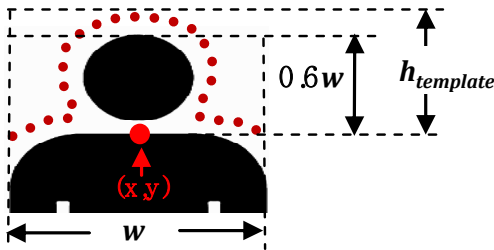

Figure 4. Mean shape template generation procedure.

After generating a mean shape template, the head and shoulder contour is fitted to the motion edges in order to find out its accurate location, scale and orientation. The B-spline active contour is fitted to motion edges by minimizing difference between observation contour Q_f and contour Q ($=WX+Q_0$) which is represented by a shape vector X as shown in (4).

$$\min_X \left\| \mathbf{WX} + \mathbf{Q_0} - \mathbf{Q_f} \right\|^2 \tag{4}$$

This contour fitting is an iterative process and it can be done by follow procedure proposed in [10]. Firstly, normal vectors are generated on the points that are spaced apart on the mean shape template in each iteration. Secondly, distances between the points spaced apart on the mean shape template and the intersection points of normal vectors and motion edges are calculated. Lastly, the shape vector is updated by minimizing sum of these distances. The number of iteration is set to 25. The mean shape template of N-th frame is generated by using the resulting shape vector of (N-1)-th frame.

III. FACE OCCLUSION VERIFICATION

Once the head and shoulder shape is estimated, we need to determine a face area and verify face occlusion in that area. To determine an exact face area, an ellipse fitting method in [11] is applied to the points on the mean shape template which compose a head shape. After finding a face area, face occlusion verification is conducted by using a skin color area ratio (SCAR) in (5) [4]. The method in [12] is used for detecting skin color pixels.

$$SCAR = \frac{\text{the number of skin color pixel in face area}}{\text{the number of skin color pixel in ellipse area}} \tag{5}$$

Since an unoccluded face includes a large number of skin color pixels, SCAR has a large value. Otherwise, this has a small value. Based on this knowledge, a face area is classified
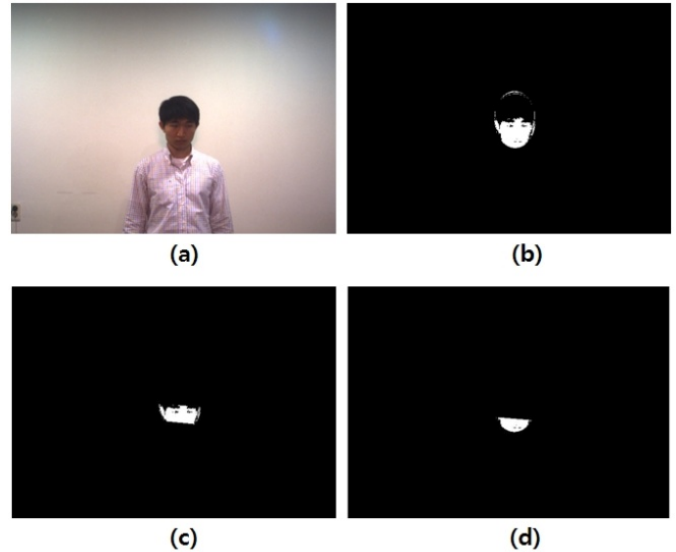

Figure 5. An example of tilted head. (a) is an original image. (b), (c) and (d) are face, eye and mouth areas, respectively. (b), (c) and (d) have SCARs of 0.404, 0.7613 and 0.9237, respectively.
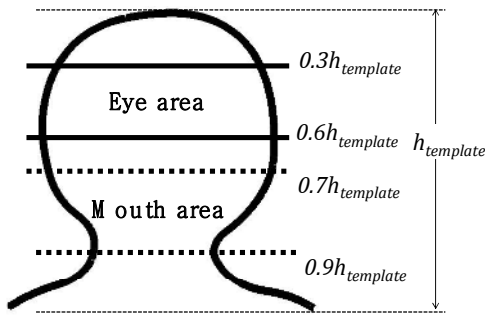
Figure 6. Eye and mouth area.

as an unoccluded face if its SCAR is larger than a pre-determined threshold (τ_{face} =0.6). However, it is difficult to classify a face area with SCAR smaller than τ_{face} as an occluded face because face areas of tilted heads are likely to have small SCAR as shown in Fig. 5 (a) and (b). In this case, the face is identifiable but its SCAR (=0.404) is less than τ_{face}. To overcome this problem, we verify face occlusion once more by using local SCARs obtained from eye and mouth areas for the face areas which have SCAR less than τ_{face}. If the SCAR of eye area is less than τ_{eye}, it is classified as an eye-occluded face. If the SCAR of mouth area is less than τ_{mouth}, it is classified as a mouth-occluded face. If these two SCARs are larger than the thresholds, it is classified as an unoccluded face. The eye and mouth areas are determined as shown in Fig. 6. Fig. 5 (c) and (d) show the eye and mouth areas, respectively. In this case, SCARs of the eye and mouth areas are 0.7613 and 0.9237, respectively, so that the face area in this figure is classified as an unoccluded face.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

Performance evaluation was conducted by using 120 video sequences of 8 subjects (40 video sequences for three cases: no occlusion, mouth occlusion by a mask and eye occlusion by sunglasses). The image resolution and frame rate are 640 x 480 pixels and 15 frame per second, respectively. The video sequences were acquired according to the scenario of ATM user in Fig. 7. Image acquisition was started when users appear in front of ATM machine by one meter. The scenario in Fig. 7 approximately took ten seconds.

120 video sequences include 12,120 frames (4,040 frames per each occlusion case). Among them, the proposed head and shoulder detection method succeeded in 11,490 frames so that the detection rate is 94.8%. This shows that this method can successfully detect head and shoulder contours regardless of partial occlusions. Fig. 8 shows the head and shoulder detection results of three occlusion cases.

To apply the proposed face occlusion verification method to a video sequence, we calculate SCARs of face areas for all the images in a video and find the face area that produces the maximum SCAR value. If this maximum SCAR value is larger than the threshold (τ_{face}), it is classified as an unoccluded face.

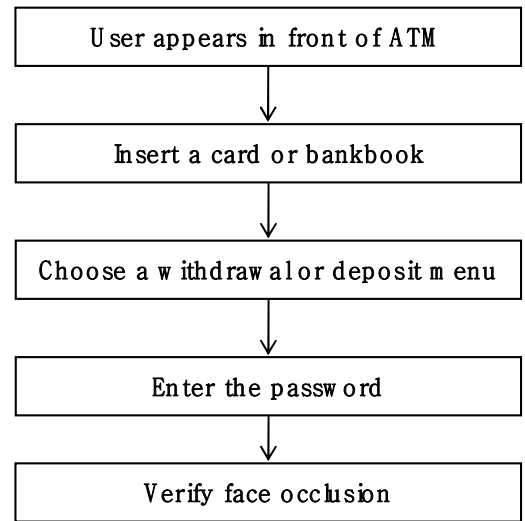Otherwise, SCARs of eye and mouth areas are calculated to

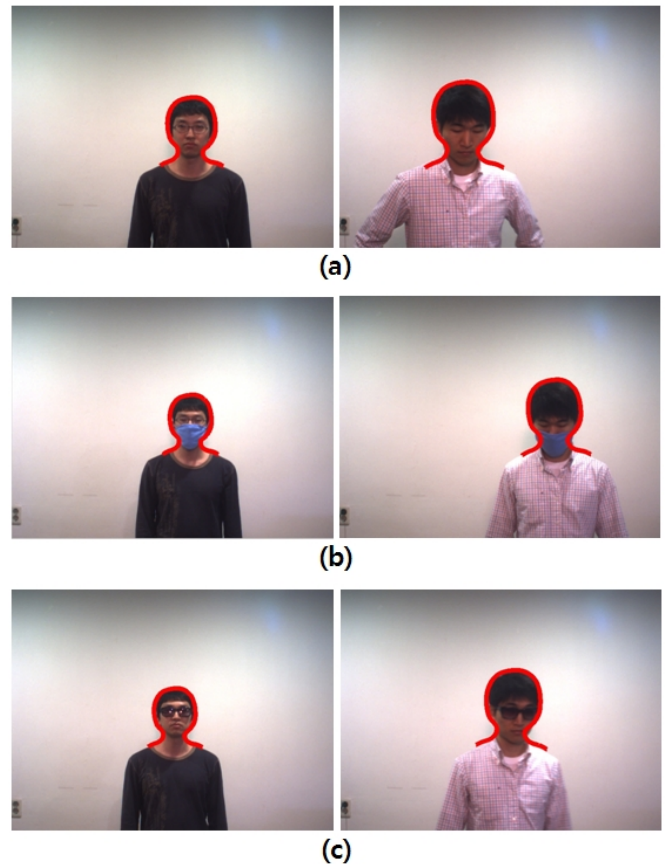Figure 7. Scenario for database acquisition.

Figure 8. Head and shoulder detection results of (a) unoccluded case, (b) mouth-occluded case and (c) eye-occluded case.

verify face occlusions. By using this approach, the proposed face occlusion verification method succeeded in 104 video sequences out of 120 video sequences. So, the occlusion verification rate is 86.7%. Fig. 9, 10 and 11 show the cases where the proposed method succeeded to verify face occlusion. It can be seen that face, eye and mouth areas are correctly
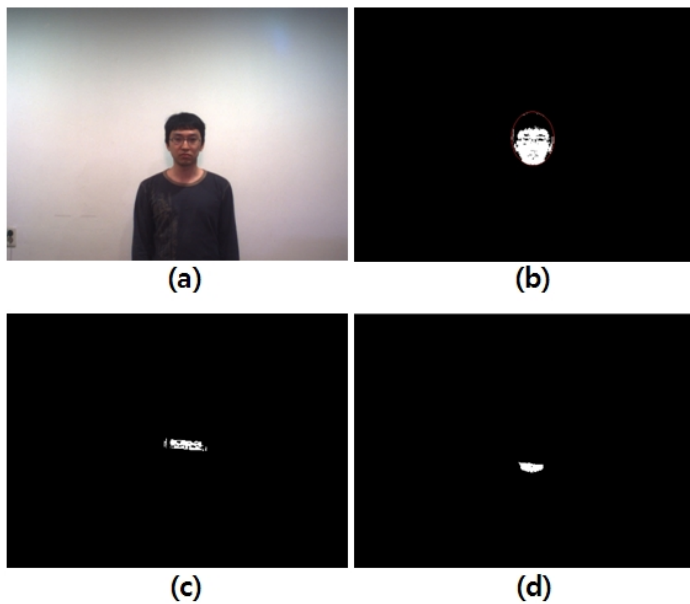
Figure 9. An example of unoccluded case. (a) is an original image. (b), (c) and (d) are face, eye and mouth areas, respectively.
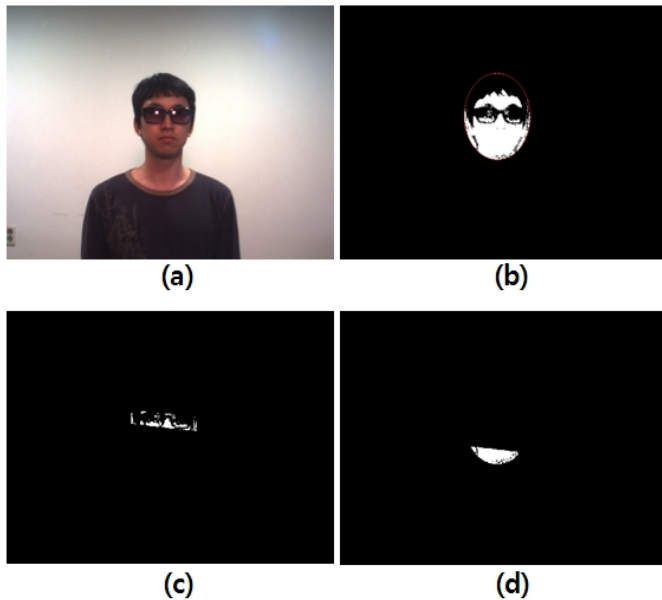

Figure 10. An example of mouth-occluded case. (a) is an original image. (b), (c) and (d) are face, eye and mouth areas, respectively.

Figure 11. An example of eye-occluded case. (a) is an original image. (b), (c) and (d) are face, eye and mouth areas, respectively.

estimated in these three cases. In Fig. 9, the face area was classified as an unoccluded face since SCARs of those three areas were larger than the corresponding thresholds. In Fig. 10, the face area was classified as an eye-occluded face because SCARs of face and eye areas were less than the corresponding thresholds. In Fig. 11 the face area was classified as a mouth-occluded face since SCARs of face and mouth areas were less than the corresponding thresholds. There were 16 cases where the proposed method failed to verify face occlusions. There were two main reasons for these failures. One is the incorrect localization of eye and mouth areas caused by head tilting and inaccurate ellipse fitting, and the other is the incorrect skin color detection, especially on the region of sunglasses.
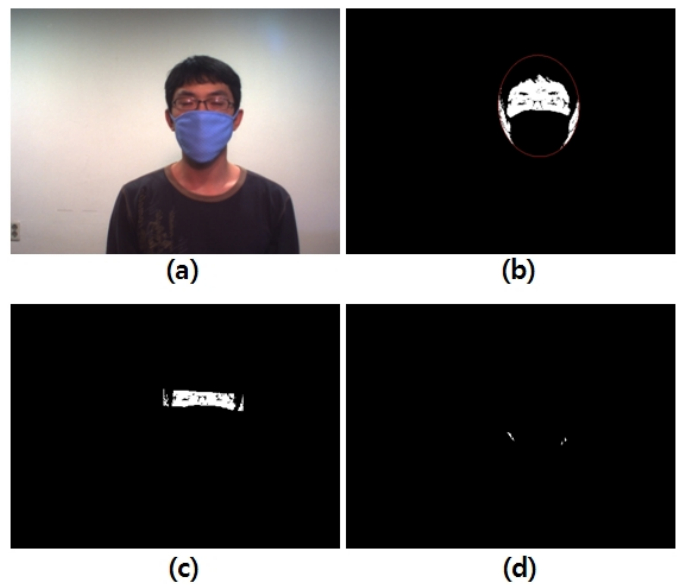
V.   CONCLUSION AND FUTURE WORK

This paper proposed a method for face occlusion verification method for an ATM application. The proposed head and shoulder detection method can accurately extract face areas even in cases of face occlusions and pose variations, and the global and local skin color area ratios-based occlusion verifier can effectively determine whether a face area is occluded and which part is occluded even under slight pose variations. In the future, we are planning to improve the occlusion verification performance by using a head pose estimator and face wearing detector.

REFERENCES

[1]   B. Heisele, T. Serre and T. Poggio. "A component-based framework for face detection and identification," International Journal of Computer Vision, vol. 74, no. 2, pp. 167-181, 2007.

[2]   Y. Tang, Z. He, Y. Chen, J. Wu, "ATM intelligent surveillance based on omni-directional vision," World congress on Computer Science and Information Engineering, pp. 660-664, 2009.

[3]   C. Wen, S. Chiu, Y. Tseng, C. Lu, "The Mask Detection Technology for Occluded Face Analysis in the Surveillance System," Journal of Forensic Science, vol. 50, no. 3, pp. 1-9, 2005.

[4]   D. Lin, M. Liu, "Face occlusion detection for automated teller machine surveillance," Lecture Notes in Computer Science, vol. 4319, pp.641-651, 2006.

[5]   B. Heisele, T. Serre, M. Pontil, T. Poggio, "Component-based face detection," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 657-662, December 2001.

[6]   K. Ichikawa, T. Mita, O. Hori, T. Kobayashi. "Component-based face detection method for various types of occluded faces," In Proceedings of

the 3rd International Symposium on Communications, Control and Signal Processing, pp. 538-543, 2008.

[7] S. Yoon, S. Kee, "Detection of partially occluded face using support vector machines," IAPR Conference on Machine Vision Applications, pp. 546-549, 2002.

[8] F. Pukelsheim, "The Three Sigma Rule," The American Statistician, vol. 48, no. 2, pp.88-91, 1994.

[9] K. P. Balanda, H. L. MacGillivray, "Kurtosis: a critical review," The American Statistician, vol. 42, no. 2, pp.111-119, 1988.

[10] A. Blake, M. Isard, "Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion," Springer-Verlag, 1998.

[11] R. Halir, J. Flusser, "Numerically stable direct least squares fitting of ellipses," In Proceedings of the Sixth International Conference in Central Europe on Computer Graphics, Visualisation and Interactive Digital Media, pp. 125-132, 1998.

[12] A. Cheddad, J. Condell, K. Curran, P. M. Kevitt, "A skin tone detection algorithm for an adaptive approach to steganography," Signal Processing, vol. 89, no. 12, pp. 2465-2478, 2009.