Image-based Fraud Detection in Automatic Teller Machine

WenTao Dong and YoungSung Soh

Department of Information Engineering, MyongJi University，Kyunggido, Korea

Summary
Some criminals access ATM(Automatic Teller Machine) using other's bank card by disguising themselves with masks, hats, and sunglasses. This paper proposes a novel and efficient method to detect such activities by analyzing the video from the camera mounted inside ATM. We first extract moving objects based on MOG(Mixture of Gaussians) background models and detect face region in HSV color space. We then obtain facial features by detecting face boundary, suppressing the outlier regions(non-facial features) inside the boundary, and analyzing projection peaks of survived regions. The presence of fraud is determined based on the relative location of features detected..

Key words:
Facial feature extraction, Fraud detection, Background model.

1. Introduction

The surveillance system plays a great role in improving the security of ATM. It records the customer's face information which is useful to trace who did the transactions. However, sometimes criminals swindle out of other's bank card and access ATM. They usually disguise themselves by wearing hats, masks, helmets, sunglasses and so on. So if the surveillance system can pinpoint the disguised among customers and give alarms when suspicious customer is found, it will prevent the crime and enhance the safety. Many approaches were proposed on face detection, facial feature detection, and face recognition. An excellent survey of these fields can be found in [1, 2]. In the survey, many application areas were summarized. However, fraud detection in ATM does not seem to be tried.

In this paper, we present a novel and efficient fraud detection system where we detect face using color features, extract facial features based on generic features such as edges and structural matching, and determine the presence of fraud based on extracted features.

2. Related Research

Tremendous amount of methods were proposed for face detection, facial feature detection, and face recognition. In this paper we focus only on the detection of face and facial features. Face recognition is out of scope of this research. We may broadly categorize face detection methods into three approaches. They are feature-based, template matching-based, and appearance- or image-based approaches. In feature-based approach, features such as colors are used. Template matching-based approach uses whole face or deformable feature templates to find the face. Appearance-based approach uses a large set of samples to train the system extensively. Facial feature extraction techniques can also be divided into three broad approaches: generic method, feature template-based and structural matching approaches. Generic method uses basic features such as edges, lines and curves to detect facial features. Feature template-based approach[3, 4] uses predefined parameterized templates and matches them to an image to detect features such as eyes. Structural matching approach[5, 6] adopts flexible statistical shape and appearance models to extract facial features. ASM(Active shape model) and AAM(Active appearance model) in[7, 8] are examples of this approach.

3. The Proposed Method

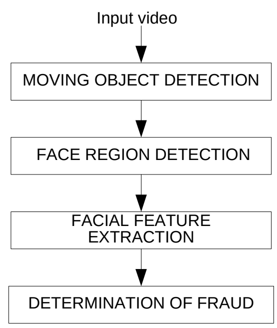


Figure 3.1 Block diagram of the proposed method

Fig. 3.1 shows the overall flow of the proposed method. Given an input video, we first detect moving objects, find face area, extract facial features, and determine the presence of fraud. We explain each of these steps in detail.

Manuscript received November 5, 2006.
Manuscript revised November 25, 2006.

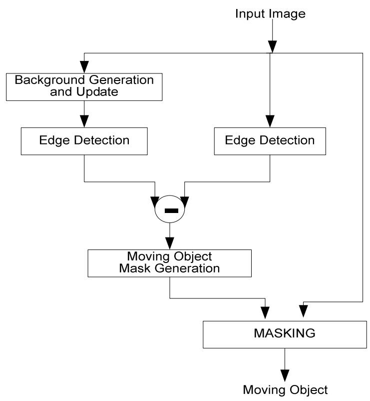


Figure 3.2 Moving object detection

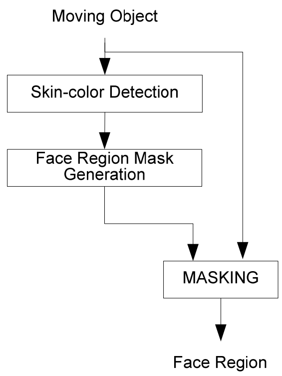


Figure 3.3 Face region detection

3.1 Moving Object Detection

Fig. 3.2 illustrates how to detect moving objects. We first generate and update background based on modified MOG(Mixture of Gaussians)[9]. We then obtain moving edge using equation (1).

$$ME_n = |\Phi(F_n) - \Phi(B)| \tag{1}$$

, where ME_n is the moving edge of nth frame,Φ the well-known Canny edge detector, F_n the current frame, and B the background produced via modified MOG. The ME_n in equation (1) is more insensitive to illumination changes than moving edges presented in [10].

With ME_n, we generate moving object mask that can cover the moving part as much as possible. This is done by a series of operations including particle noise removal, morphological bridging, hole filling, and morphological closing and opening. By applying the object mask to the current frame, we can extract the moving object.

3.2 Face Region Detection

Fig. 3.3 shows how to detect face region. We use the HSV color space to detect face region. Some researchers[2] use both chromatic components, H and S, in isolating face colors. In our application, since we use the camera with relatively poor quality, face region frequently shows washed out look. Thus restricting the S within some range makes the result worse. We use only H with the value range described in equation (2).

$$H_{Skin} = (H>325^o \| H<35^o) \tag{2}$$

Next step is to generate face region mask. By applying equation (2) to the current image, we may obtain many face regions of different size and location. We choose the one with maximum size among ones near the center of the image. We use the location restriction because, in our application, face is usually located around the center of the image. We then apply hole filling and morphological closing and opening to the selected region to generate the face region mask. By applying the face region mask to the moving object, we obtain face region

3.3 Facial Feature Extraction

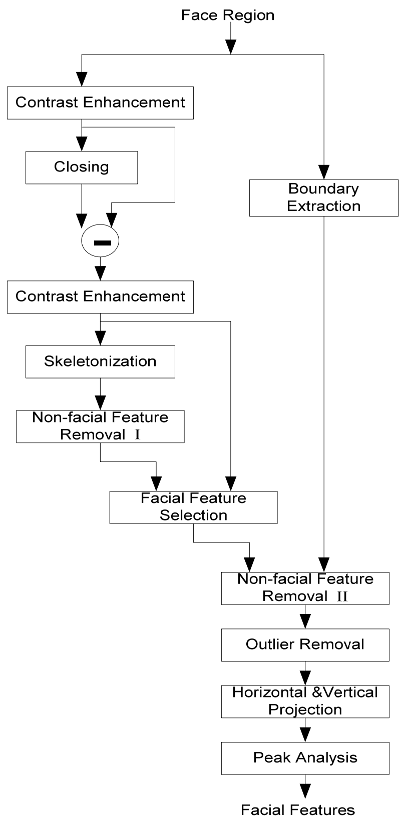


Figure 3.4 Facial feature extraction

In order to extract facial features, we use only non-chromatic part of HSV, i.e., V, of detected face region. Taking face region as an input, we first enhance the contrast to compensate for poor quality. We then apply morphological closing to the enhanced to fill up holes caused mostly by facial features such as eyes, nose, and mouth. Taking the difference between the enhanced and the hole-filled, we get facial feature parts possibly with non-facial parts. We do the contrast enhancement again to bring up facial features more clearly. We then obtain skeletons to distinguish between facial and non-facial features. The skeletons of facial features are usually oriented horizontally and have lengths within some range. Thus, by analyzing the orientation and the length of skeletons, we can classify skeletons into two groups: skeletons of facial features and skeletons of non-facial features. In this work, we assume that face is frontal and upright in most of the time. We keep only skeletons of facial features and project them back to the image just before skeletonization to select only facial features. However, it is quite probable that we still have some non-facial features. Thus we use boundary information of detected face region. We remove selected facial features on or outside the boundary. If the detected face region is relatively larger than real face region, we might still have non-facial features which are located relatively far away from the group of actual facial features. To remove these outliers, we compute mutual distances among features.

Let N be the total number of facial features and D_{ij} the distance between i*th* and j*th* facial features.

Then the average distance is defined as

$$\mu = \sum_{i=1}^{N} \sum_{j=i+1}^{N} D_{ij} / C(N,2) \tag{3}$$

, where C(N,2)= $\frac{n(n-1)}{1 \cdot 2}$ and is the number of combinations of mutual distances.

Similarly standard deviation is defined as

$$\sigma = \sqrt{\frac{1}{C(N,2)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} (D_{ij} - \mu)^2} \tag{4}$$

We say that if $D_{ij} > \mu + \sigma, (i < j)$,then i*th* and j*th* facial features become candidates of an outlier. If j*th* feature is nominated as an outlier candidate many times and if it is near the boundary of face, we consider it as an outlier and remove it. To find the locations of facial features such as eyes and mouth, we project facial feature regions both horizontally (into y axis) and vertically (into x axis). For the horizontal projection, we divide the projection into two equal-sized parts and try to find a peak in each half. The peak in the upper half corresponds to y location of eyes and the peak in the lower half to the y location of mouth. Similarly we divide vertical projection into four equal-sized parts and try to find peaks in the first and the last parts. The peak in the first part corresponds to x location of left eye and the peak in the last part to x location of right eye. The middle of x locations of left and right eyes becomes the x location of mouth.

3.4 Determination of Fraud

Table 1 Average Lengths of Facial Features in De Silva et al.'s Algorithm [7]

	Head height	Eye separation	Eye to nose	Eye to mouth
Average length	1.972	0.516	0.303	0.556

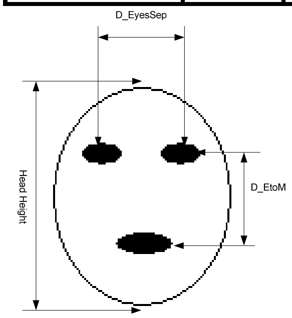


Fig 3.5 Simplified face model

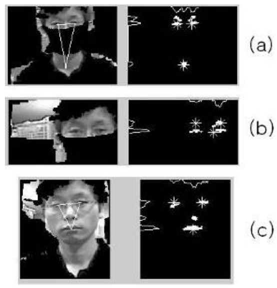


Fig 3.6 Various cases of triangle

Fig. 3.5 shows the simplified facial features. Since eyes and mouth are the most significant facial features, other features such as nose, eyebrows are omitted. We define several entities as follows.

$$E_{Eyes} = \begin{cases} 1\text{, if either candidate of eyes is correctly located} \\ 0\text{, otherwise} \end{cases}$$

$$E_{Mouth} = \begin{cases} 1\text{, if mouth candidate is correctly located} \\ 0\text{, otherwise} \end{cases}$$

$$E_{D\_EyesSep} = \begin{cases} 1\text{, if}(\frac{\text{D\_EyesSep}}{\text{HeadHeight}})\text{ is close to the ratio }(\frac{0.516}{1.972}) \\ 0\text{, otherwise} \end{cases}$$

$$E_{D\_EtoM} = \begin{cases} 1\text{, if}(\frac{\text{D\_EtoM}}{\text{HeadHeight}})\text{ is close to the ratio }(\frac{0.556}{1.972}) \\ \quad\text{and}(\frac{\text{D\_EtoM}}{\text{D\_EyeSep}})\text{ is close to }(\frac{0.556}{0.516}) \\ 0\text{, otherwise} \end{cases}$$

, where D_EyesSep and D_EtoM are indicated in Fig 3.5 and the values used in computing the ratio are from Table 1 where the relative distances among facial features are shown[7]. We aggregate above four entities into one figure of merit E as follows

$$E = w_{Eyes} * E_{Eyes} + w_{Mouth} * E_{Mouth} + w_{D\_EyesSep} * E_{D\_EyesSep} + w_{D\_EtoM} * E_{D\_EtoM} \quad (5)$$

where $w_{Eyes}, w_{Mouth}, w_{D\_EyesSep}$ and w_{D_EtoM} are weights of E_{Eyes}, E_{Mouth}, $E_{D\_EyesSep}$ and E_{D_EtoM} respectively, with $w_{Eyes} + w_{Mouth} + w_{D\_EyesSep} + w_{D\_EtoM} = 1$. Next we check two more figures of merit: aspect ratio of bounding box and solidity of triangle formed by centers of two eyes and mouth.

$$F_{\text{Aspect\_raio}} = \frac{\textit{Width} \text{ of Detected FacialBoudingbox}}{\textit{Height} \text{ of Detected FacialBoudingbox}} \quad (6)$$

$$F_{\text{Solidity}} = \frac{\text{Actual area of } \text{Tri}_{\text{EtoM}}}{\text{Expected area of } \text{Tri}_{\text{EtoM}}} \quad (7)$$

, where Tri_{EtoM} is the triangle consisting of centers of two detected eyes and mouth. Fig. 3.6 shows various cases of triangle. Fig. 3.6(c) shows the triangle formed by two well-detected eye centers and a mouth center, whereas Fig. 3.6(a) illustrates the triangle formed by two well-detected eye centers and a wrong mouth center. In Fig. 3.6(b), we have only two well-detected eye centers, thus no triangle can be formed. In equation (7), actual area means the number of pixels with skin color inside the triangle, and expected area the area of whole triangle .Thus the solidity tells us the degree of correctness of feature extraction. It is clear that the actual area is far less than expected area, if either mouth is not detected right or it is not detected at all since E_{Mouth} and E_{D_EtoM} will be zero and, as a result, E will become very low. The final decision criterion is to check D defined as:

$$D = (E > E_{\text{thresh}}) \,\&\&\, (F_{\text{Aspect\_ratio}} > F_{\text{Ratio\_thresh}}) \,\&\&\, (F_{\text{Solidity}} > F_{\text{Solidity\_thresh}}) \quad (8)$$

, where E_{thresh}, $F_{\text{Ratio\_thresh}}$, $F_{\text{Solidity\_thresh}}$ are manually-set thresholds.

If D=1, we classify the face into “no fraud” and “fraud” otherwise.

4. Experimental Results

In this chapter, we present experimental results for each step of the proposed method.

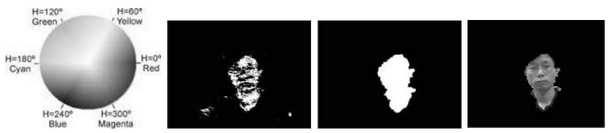
4.1 Moving Object Detection

(a) (b) (c) (d) (e)

Fig. 4.1 Results of moving object detection

Fig. 4.1 shows the result for moving object detection. Fig. 4.1(a) is the generated background, (b) the current image, (c) the difference of edges of (a) and (b), (d) the moving object mask after morphological operations and hole filling on (c), and (e) the detected moving object

4.2 Face Region Detection

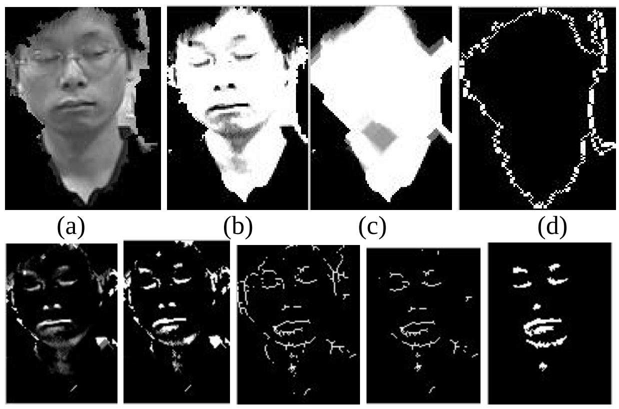


(a) (b) (c) (d)

Fig. 4.2 Results of face region detection

Fig. 4.2 illustrates the result for face region detection. Fig. 4.2(a) is H space cross section, (b) the skin-color region, (c) the face region mask obtained by applying morphological operations and hole filling to (b), and (d) the detected face region

4.3 Facial Feature Extraction



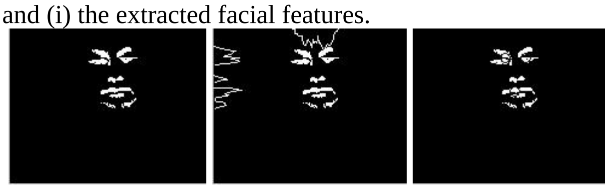
(a) (b) (c) (d)

(e) (f) (g) (h) (i)

Fig. 4.3 Results of facial feature extraction

Fig. 4.3 shows the result for facial feature extraction. Fig. 4.3(a) is the detected face region, (b) the contrast enhanced of (a), (c) the result of applying morphological closing to (b), (d) the boundary of(b), (e) the difference of

(b) and (c), (f) the contrast enhanced of(e), (g) the skeleton of (f), (h) the result of removing non-facial skeletons in (g), and (i) the extracted facial features.

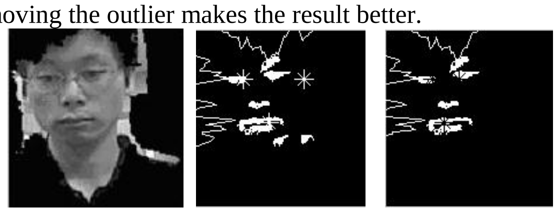


(a) (b) (c)

Fig. 4.4 Result of projection and peak analysis

Fig. 4.4 illustrates the result of projection peak analysis. Fig. 4.4(a) is the extracted facial features, (b) the horizontal and vertical projections, and (c) the detected mouth and eye locations marked with asterisks.

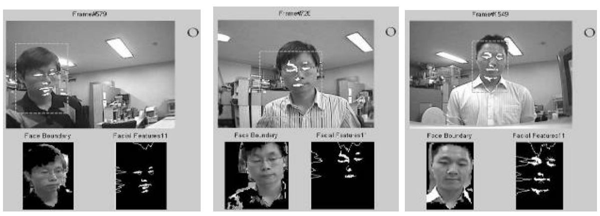
Sometimes we have outliers in extracted facial features. We remove them by analyzing mutual distances among facial features. Fig. 4.5 shows the result of outlier removal. Fig. 4.5(a) is the detected face region, (b) the location of mouth and eyes with outlier, (c) the location of mouth and eyes after removing the outlier. As can be seen, removing the outlier makes the result better.



(a) (b) (c)

Fig. 4.5 Results of outlier removal

4.4 Determination of Fraud

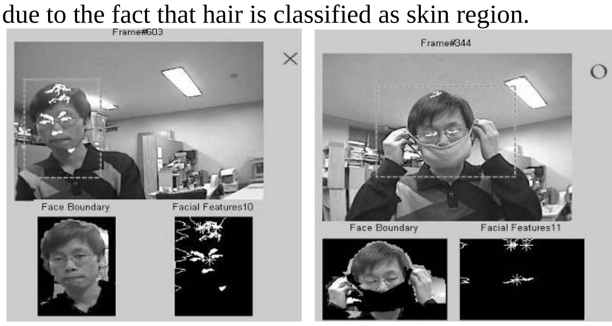


(a) (b) (c)

(d) (e) (f)

Fig. 4.6 Results of fraud determination

Fig. 4.6 shows the result for fraud determination. Each image has 3 sub images. Upper one is the input image with bounding box on detected face. Lower left is the detected face region and lower right is the extracted facial features. On top right of each image, either O or X is displayed. O means "no fraud" and X means "fraud". Fig. 4.6(a), (b), and (c) are "no fraud" examples and the rest are "fraud" examples. As can be seen, the presence of fraud is correctly determined. Sometimes the proposed method yields erroneous results. Fig. 4.7(a) is falsely determined as "fraud" because hair is considered to have skin color. This is due to the direct reflection of lights on top of the head. Fig. 4.7(b) shows the case where man wearing a mask is determined as "no fraud". This is also due to the fact that hair is classified as skin region.



(a) (b)

Fig. 4.7 Erroneous results of fraud determination

5. Discussion

In this paper, we propose a fraud detection system in ATM. We first detect moving object based on modified MOG. We then isolate face region in HSV color space and extract facial features using morphology, projection and peak analysis. Based on the relative location of facial features and the aspect ratio of face region, we determine the presence of fraud. For test video, we obtain good results. However, for some instances, we have erroneous results mostly due to the wrong detection of face region and this is intended for future research.

References

[1] B. L. E. Hjelmas, Face detection: a survey, Computer Vision and Image Understanding, vol. 83, pp. 236-274, 2001.

[2] W.Y. Zhao, R. Chellappa, A. Rosenfeld and J.P. Phillips, Face Recognition: A Literature Survey, ACM Computing Surveys, December Issue, pp. 399-458, 2003

[3] Bernhard Froba, Andreas Ernst, and Christian Kublbeck, Real-time face detection, Proc. 4 th IASTED SIP, pp. 497-502, 2002

[4] K. Sobottka, I. Pitas, Face localization and facial feature extraction based on shape and color information, In Proceedings of the IEEE International Conference on

Image Processing, Lausanne, Switzerland, vol. III, pp. 483-486, 1996

[5] X. Xie, R. Sudhhakar, and H. Zhuang, On improving eye feature extraction using deformable templates, Pattern Recog. 27, pp. 791–799，1994,.

[6] Y. Saito, Y. Kenmochi, and K. Kotani, Extraction of a symmetric object for eyeglass face analysis using active contour model, in Proceedings of the 2000 International Conference on Image Processing, vol.2, pp.231-234 2000

[7] Kwok-Wai Wan, Kin-Man Lam, Kit-Chong Ng, An accurate active shape model for facial feature extraction, Pattern Recognition Letters , vol.26, pp 2409-2423, 2005

[8] J. Ahlberg, Face and Facial Feature Tracking using the Active Appearance Algorithm, 2nd European Workshop on AVBS, pp. 83-93, 2001.

[9]Hye-Seung A New Shadow Removal Method using Color Information and History Data, Korea Information Processing Society ,vol. 12-B, pp.395-402, 2005

[10] Alsaqre F.E, Baozong Y, Moving object segmentation from video sequences: an edge approach, 4th EURASIP Conference, vol. 1, 2-5 pp.193 – 199, July 2003

[11] L. C. De Silva, K. Aizawa, and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, IEICE Trans. Inform. Systems E78–D(9), pp. 1195–1207, 1995.

WenTao Dong received B.S. and M.S. in Computer Technology from Shanghai Normal University in 2001 and 2004, respectively. Now, he is a candidate for Ph.D in Information Engineering at MyongJi University in Korea.

Young Sung Soh received B.S. in Electrical Engineering from Seoul National University in 1978. He received M.S. and Ph.D in Computer Science from Univ. of South Carolina in 1986 and 1989, respectively. His main research focus is on computer vision and pattern recognition..