# Automatic Real-Time Localization of Frowning and Smiling Faces under Head Rotation Variations

JOUNI EROLA, YULIA GIZATDINOVA, AND VEIKKO SURAKKA
Department of Computer Sciences
University of Tampere
Kanslerinrinne 1, 33014
FINLAND

*Abstract:* - A new method for real-time face localization from a streaming color video was developed. The method consisted of three stages. First, the face-like skin-colored image region was segmented from the background and transformed into the grey scale representation. Second, the cropped image was convolved with Sobel operator in order to extract local oriented edges at 16 orientations. The extracted local oriented edges were grouped together to form regions of interest which represent landmark candidates. Further, the candidates were matched against edge orientation model to verify the existence of the landmark in the image. Finally, the located landmarks were next spatially arranged into the face–like constellations. The best face-like constellation of the landmark candidates was defined by a new scoring function. The test results demonstrated that the proposed method located expressive faces with high rates in real time from facial images under controlled head rotation variations.

*Key-Words:* - Face localization, Facial landmarks, Sobel edge detection, Frontal-view geometrical face model, Facial expression, Head rotation.

## 1 Introduction

In automatic face detection, different feature detectors are applied in order to find a face-like region in static images or video frames. If it is known in advance that face is shown in the image, the task comes to find a true face location. The found face location is further delivered as an input to various systems of automatic face analysis such as face and facial expression recognition and perceptual vision-based user interfaces. To ensure that these systems work in real time efficiently and robustly, face detection is aimed to provide high speed and accuracy of the detection process.

The main challenge in face detection is to find a face representation that remains robust with respect to various changes in facial appearance since face varies noticeably with changes in environmental conditions (e.g. illumination, out- and in-plane head rotations, scene complexity, resolution, occlusions, etc.), ethnicity (i.e. skin color), gender, and facial expressions (i.e. emotional and social signals in the face). Many attempts have been undertaken to automatically detect faces from static images and video [10,22]. We can roughly classify the existing techniques of face detection as belonging to appearance- or feature-based approaches. The appearance-based approach uses holistic features of the image and considers a face as a whole. Methods of Principle Component Analysis (PCA) have been widely adopted for the purpose of face detection [9,21]. Generally, PCA-based methods can handle faces with nearly the same pose, constant illumination, and moderate facial expressions. Apart from PCA-based methods, there are also learning-based methods like boosted classifiers [19], support vector machines [12], and neural networks [14] which have to be trained on the representative sets of face and non-face images.

The feature-based approach to face detection utilizes local features of the image. This approach can overcome the constraints placed by illumination change, head rotations, and facial expressions. This is due to the fact that it is based on modeling local texture information around individual facial landmarks and modeling global shape information on spatial arrangement of the located landmark candidates. Facial landmarks are typically those which are the most distinctive for humans - eyebrows, eyes, nose, and mouth. These landmarks encode critical information on facial expressions and head movements that is used in automatic face analysis. In practice, feature-based face detection includes a selection of feature representation and a design of feature detector. Different features can be detected from the image or video frame, for example, edges, colors, points, lines, contours, etc.

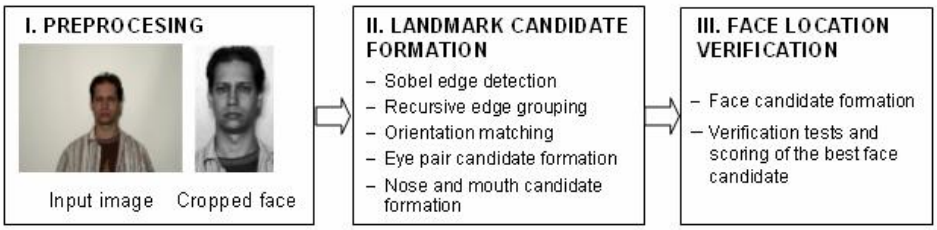


**Figure 1.** Data flow in our system of real-time face localization.

These features provide a meaningful and measurable description of the face as they represent specific visual patterns used to identify corresponding structures between images. Extensive work has been focused on shape representation of facial landmarks [2,3,4,15]. Many proposed face detectors utilize edges [6], grey-scale values [17], and their combinations [16]. A multi-resolution and multi-orientation representation of the image has been widely adopted for the purpose of landmark detection and demonstrated to be effective in face detection under expression and small pose variations [7,20,23]. However, these methods are generally computationally expensive and, therefore, are not applicable for the task of real-time face and facial feature detection.

In this paper, we introduce a feature-based method of face localization from streaming color video. It is based on the multi-orientation image representation that helps in composing facial landmarks. This is followed by spatial arrangement of the located landmark candidates. We present a new scoring function designed to define the best face-like constellation of the candidates. The landmarks to be located in the face are centers of the eyes, nose tip, and center of the mouth. In the subsequent sections we demonstrate how our proposed method successfully overcomes speed limitations of the feature-based face detection methods.

# 2 Face Localization

The method for real-time face localization consists of three stages depicted in Fig. 1. Given a facial video, the first stage finds a rough approximation of the face location in each video frame. The second stage forms and extracts all possible landmark candidates from the cropped face region. The last stage decides whether constellations formed from the located landmark candidates meet requirements placed by the geometrical frontal-view facial model, and if so, defines the location of the best-scored face candidate. Below we explain each stage of the method in more detail.

## 2.1 Preprocessing

On this stage, a face-like image region is segmented from the background. To do that we first apply the procedure of histogram equalization to each video frame for all three RGB color channels. This calculation is not computationally expensive and is widely used to allow areas of low local contrast to gain a higher contrast without affecting the global contrast of the whole image [8]. After this, the original RGB image is transferred into YCbCr chromatic color space as proposed in [11].

Next, we segment the skin-colored regions of the image similarly to the method proposed in [1]. This procedure allows for noisy regions of the image to be discarded at the early stage of the processing and, therefore, focuses the following stages of the method on those parts of the image in which the face is more likely to be located. The Gaussian-fitted skin color model is used for this purpose. The idea that lies behind is that a distribution of skin color for different people is clustered in the chromatic color space and can be represented by a Gaussian model. The likelihood $P$ of a skin color for any pixel $(x,y)$ of the image thus can be obtained with a Gaussian-fitted skin color model:

$$P = e^{\left(-\frac{1}{2 \cdot Cv} \cdot \left((Cb - Cb_{mean})^2 + (Cr - Cr_{mean})^2\right)\right)}, \quad (1)$$

where $Cv$ is a covariance matrix; $Cb$ is a blue chromatic value; $Cr$ is a red chromatic value; $Cb_{mean}$ is an average blue chromatic value; and $Cr_{mean}$ is an average red chromatic value.

The skin-colored regions are next segmented from the rest of the image through a thresholding process. The parameters for a thresholding are selected experimentally using a small image set from the database. Finally, the received regions are cropped from the background and the procedure of histogram equalization is applied for them. This stage outputs 8-bit grey scale image of the face. If a face is not extracted, the following steps of the method are applied to the whole image converted into 8-bit grey scale representation.

### 3.2 Landmark Candidate Formation

On the next stage, the cropped face image is convolved with Sobel operator [8] in order to extract local oriented edges at 16 orientations. The edge points are further recursively grouped together to form regions of interest which represent candidates for facial landmarks. The shapes of the bounding boxes which are placed over the located regions of interest are analyzed next. If a candidate is bounded by a box which has height much bigger than its width, the candidate is eliminated. After this, among the located candidates there still exist many noisy regions which have to be eliminated. In order to define regions which represent landmark candidates we analyze local properties of the located regions of interest. As it has been demonstrated earlier [7], regions of facial landmarks have a characteristic distribution of local oriented edges with two horizontal dominants (Fig. 2*a* and 2*b*). On the other hand, non-landmark regions which are, for example, elements of face, hair, clothing, and decoration typically do not have a characteristic structure of the oriented edges. These regions demonstrate a random distribution of the oriented edges (Fig. 2*c* and 2*d*). This local property of the located regions of interest allowed us to discard noisy regions while preserving regions which contain facial landmarks.

In order to classify the located candidates and find their proper spatial arrangement, the proposed method applies a set of verification rules which are based on face geometry. The knowledge on face geometry is taken from the anthropological study by Farkas [5]. This thorough study examined thousands of Caucasians, Chinese, and African-American subjects in order to determine characteristic measures and proportion indexes of the human face. We performed several tests to verify the existing facial measures, calculate new measures, and built a frontal-view geometrical face model depicted in Fig. 3. The anthropometric facial features and measures of the model are described in Table 1. Center points of facial landmarks are calculated as mass centers of the located edge regions. It has been

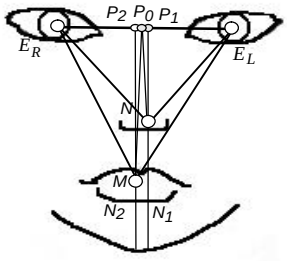


**Figure 3.** Frontal-view geometrical face model.

demonstrated that facial measures from the table can slightly vary between subjects of different gender, age, and race [5]. Therefore, we define constrains for facial measures as intervals between minimum and maximum values for a given measure. All distance constrains from the table are defined as percentages of the interocular distance $d(E_R, E_L)$. As our preliminary tests demonstrated, this set of facial measures and their corresponding constrains achieved good results in composing face-like constellations from the located landmark candidates.

The classification of the landmark candidates proceeds as follows. The eye pair candidates are found first as any two candidates aligned nearly horizontally. After this step, all possible nose and mouth candidates for each found eye pair candidate are independently searched for using constrains of the frontal-view geometrical face model.

### 3.3 Face Location Verification

On the last stage, the defined eye-nose and eye-mouth candidates are combined together into a complete face candidate so, that the eye pair is the same for both eye-nose and eye-mouth candidates. Each found face candidate consisting of four facial landmarks is given a score. A score is calculated as a sum of intermediate scores which show how well a face candidate performs verification tests. The verification tests are fuzzy rules defined as follows: *Test 1* checks the horizontality of the eye pair candidate. The eye pair candidate that has the most horizontal position in the image as compared to

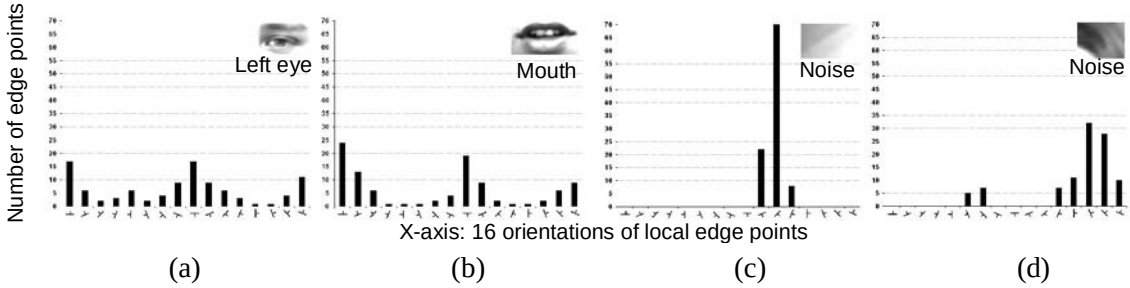


**Figure 2.** Facial landmarks with a characteristic distribution of the oriented edges (*a* and *b*) and noisy regions with a random distribution of the oriented edges (*c* and *d*).

**Table 1**: Features and measures of the frontal-view geometrical face model (and their constrains).

| Feature/measure | Feature description |
|---|---|
| $E_R$ , $E_L$ , $N$ , $M$ | Centers of the landmarks |
| $d(E_R, E_L)$ | Interocular distance |
| $N_1$, $N_2$ | Perpendiculars to $d(E_R, E_L)$ |
| $P_1$, $P_2$ | Cross points of $d(E_R, E_L)$ and $N_1$ and $N_2$, correspondently |
| $P_0$ | Middle point between $P_1$ and $P_2$ |
| $d(N, M)$ | Nose-mouth distance (30-110% of $d(E_R, E_L)$ ) |
| $\angle NP_0M$ | Nose-mouth angle (0-16°) |
| $\angle E_RNP_1$, $\angle E_LNP_1$ $\angle E_RMP_2$, $\angle E_LMP_2$ | Eye-nose and mouth-eyes angles (0-13°) |
| $d(E_R, N)$, $d(E_L, N)$ | Eye-nose distance (25-120% of $d(E_R, E_L)$ ) |
| $d(E_R, M)$, $d(E_L, M)$ | Eye-mouth distance (60-160% of $d(E_R, E_L)$ ) |

others gives the lowest score for a given face candidate. *Test 2* checks angles $\angle E_RNP_1$, $\angle E_LNP_1$, $\angle E_RMP_2$, and $\angle E_LMP_2$ - face candidate with small angles gets low scores, and vice versa. *Test 3* considers the result of the previous face localization. If the landmark center points in the previous frame are nearly the same as compared to those in the current frame (face is not moving), a face candidate gets a score which is lower than in the opposite case. *Test 4* checks sizes of the located landmark candidates – the biggest values give the lowest score for a face candidate. *Test 5* checks widths of the landmark candidates in the facial configuration. It has been validated that mouth is usually wider than eyes and nose has nearly the same width as eyes have [5]. The closer face candidate satisfies to this criterion, the lower score it gets from the test. *Test 6* utilizes the property of face symmetry and checks sizes of both eyes. If eyes have the same size, a face candidate gets the lowest score from this test. Each verification test is also given a weight. We performed a number of tests to define optimal weights for each test. This way, each test gives as its output a relative score for a given face candidate. In order to select the best-scored face-like constellation of the located landmarks, a new scoring function is introduced:

$$P_r = MAX - \frac{MAX}{P_{cand} / P_{\min}} \quad (2)$$

where *MAX* is a maximum score for a given face candidate (we used 100); $P_{cand}$ is a current score for a given face candidate; $P_{\min}$ is the lowest score achieved among all face candidates. This way, if we have several face-like constellations of the landmarks, we select the one that gives the highest score $P_r$.

A localization result is considered correct if a distance between manually annotated and automatically located landmark location met the requirement placed by a performance evaluation measure elaborated in [13]:

$$d_{eye} = \frac{\max\left(d(E_R, E'_R), d(E_L, E'_L)\right)}{d(E_R, E_L)} \quad (3)$$

where $d(a,b)$ is Euclidean distance between point locations *a* and *b*; $E_R$ , $E_L$ are manually annotated and $E'_R$, $E'_L$ are automatically located positions of facial landmarks. A successful localization was considered if $d_{eye} < 0.25$ which corresponds approximately to 1/4 of the annotated interocular distance $d(E_R, E_L)$ (a half of the width of the eye). After the locations of the landmarks are known, the location of the face in the image is also known.

# 4 Test Data

As the prospective application of the developed method lies in human-computer interaction, we assume that the input video includes some head rotations and facial expressions. For the purpose of method testing under these conditions, we created our own video database with neutral, frowning, and smiling faces under three controlled head rotations with angles of rotation 0°, 20°, and 30° in both right and left directions. We used a low-cost Canon Mini-DV camera with 720x568 pixel image resolution and 24-bit precision for color values. The potential impact of illumination, background, facial hair or eye-glasses was controlled to some extent in all video sequences and therefore ignored. No face alignment was performed.

The database consists of 10 Caucasian subjects (40% females) with average age of 30 years. Each video starts with neutral face, proceeds with facial expression, and ends up with neutral face. The level of the expression intensity varies among different subjects. In total, 150 video sequences were created with duration of about 7-8 seconds. The test data were annotated in advance by recording the true locations of the landmark centers in each frame for each test subject.

# 5 Results

The tests were run on the computer Dell Optiplex 745, Intel Core2 with 2133 MHz and 1 GB DDR2-memory in Win XP 2002 SP 2/Delphi-environment with DirectShow-interface. Fig. 4 shows the average rates of face localization under three expressions and three head rotations. The raw test data are shown in Table 2. The statistical analysis was done by using a two-way $3\times3$ (expression$\times$head rotation) repeated measures analysis of variance (ANOVA). The statistical analysis showed that head rotation had a statistically significant main effect on the face localization $F(2,18)=9.31$, $p<0.001$. Bonferroni-corrected pairwise post-hoc comparisons showed that the detection percentage was significantly lower when head was rotated by 30° $MD=30.81$, $p<0.05$ as compared with frontal head position. Difference between 20° and 30° head rotations was also statistically significant $MD=22.14$, $p<0.05$. There was no statistically significant difference between head rotation by 20° and frontal head position. ANOVA showed that there were no other significant main or interaction effects.

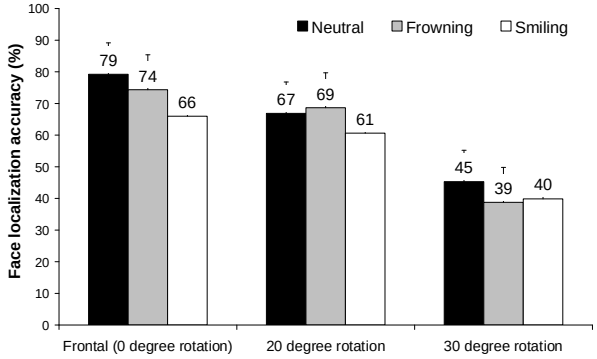


**Figure 4.** Average localization rates (%) of neutral, frowning, and smiling faces (all four landmarks were found) under three head rotations.

**Table 2**: Rates (%) of localization of neutral, frowning, and smiling faces under three head rotations.

| Subject | Neutral | | | Frowning | | | Smiling | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0° | 20° | 30° | 0° | 20° | 30° | 0° | 20° | 30° |
| **1** | 69 | 91 | 97 | 79 | 84 | 67 | 96 | 72 | 93 |
| **2** | 95 | 51 | 17 | 82 | 74 | 31 | 87 | 67 | 17 |
| **3** | 55 | 17 | 30 | 45 | 10 | 37 | 27 | 34 | 31 |
| **4** | 98 | 68 | 60 | 98 | 83 | 35 | 87 | 72 | 48 |
| **5** | 95 | 62 | 36 | 89 | 91 | 17 | 82 | 56 | 29 |
| **6** | 33 | 47 | 49 | 21 | 51 | 47 | 26 | 62 | 46 |
| **7** | 93 | 88 | 30 | 85 | 79 | 13 | 24 | 49 | 46 |
| **8** | 63 | 61 | 25 | 81 | 35 | 8 | 90 | 44 | 14 |
| **9** | 92 | 95 | 98 | 62 | 93 | 71 | 85 | 77 | 37 |
| **10** | 91 | 72 | 30 | 97 | 82 | 64 | 63 | 65 | 45 |

# 6 Discussion

The developed method demonstrated high speed performance in face localization from streaming color video in real time. The speed of the method was 20 frames per second that meets the requirement of real-time video processing defined in [18]. This way, the method is comparable to the best existing real-time face detectors [6,15,19] in terms of processing speed.

The local oriented edges served as basic features for expression-invariant representation of facial landmarks. The results confirmed that in the majority of expressive images a distribution of the local oriented edges had structure with two horizontal dominants as predefined in [7]. This property allowed discarding noisy regions and preserving regions of the landmarks. Thus, the method was able to locate landmarks from images with hair and shoulders. The use of frontal-view geometrical face model further improved the overall performance of the method. As Fig. 4 shows, the method was effective in locating faces with frontal and near-to-frontal head poses. However, head rotations by 30° significantly decreased face localization rates. This is explained by the fact that geometrical constrains from Table 1 were defined mainly for frontal-view geometrical face model. The landmarks were located correctly in case of 30° head rotations, but failed to compose the face-like constellations. Relaxation of geometrical constrains from Table 1 or development of new measures and constrains for a near-to-profile geometrical face model would improve the performance of the method in case of strong face rotations.

In case of frontal and near-to-frontal head positions, the method demonstrated sufficiently high rates in locating faces with all three tested expressions - neutral, frowning, and smiling expressions. This gives similar performance of the method for these particular expressions as compared to the results of previous studies which use similar approach to facial landmark localization [7]. As distinct from that study, we concentrated on face localization rather than on independent landmark localization meaning that all four facial landmarks needed to be correctly located in order to declare successful face localization. Face localization rate would be improved if we consider two or three correctly located landmarks as necessary and sufficient requirement for successful face localization, as it is done, for example, in [2]. This way, the method can also be applied for independent facial landmark localization, when it is allowed to miss some landmarks.

In summary, as compared to the existing feature-based methods of face localization, the method demonstrated similar or superior performance in terms of localization rates [7,12] and speed [6,15,19]. Besides robustness to facial expressions and small out-of-plane head rotations, the developed method demonstrated robustness to noise such as hair, and elements of clothing and decoration. Emphasizing simplicity, high speed, and low computation cost of the method, we conclude that it can be used in face localization as such and also in preliminary localization of regions of facial landmarks for their subsequent processing where coarse landmark localization is followed by fine feature detection. The method is simple and straightforward to be utilized, for example, as face or facial landmark detector in the mobile phone environment.

## 7 Acknowledgement

This work was financially supported by the University of Tampere and Academy of Finland (project numbers 177857 and 1115997).

*References:*

[1] H. Chang and U. Robles, Face Detection, Project report, *http://www-cs-students.stanford.edu/robles/ee368/main.html*, 2000.

[2] D. Colbry, G. Stockman, and J. Anil, Detection of Anchor Points for 3D Face Verification, *CVPR'05*, Vol.3, pp. 118-126.

[3] T. Cootes, G. Edwards, and C. Taylor, Active Appearance Models, *Trans. Pattern Anal Mach. Intel.*, Vol.23, No.6, 2001, pp. 681-685.

[4] D. Cristinacce and T. Cootes, Feature Detection and Tracking with Constrained Local Models, *BMVC'06*, Vol.3, pp. 929-938.

[5] L. Farkas, *Anthropometry of the Head and Face*, (2 ed), Raven, New York, 1994.

[6] B. Fröba and C. Küblbeck, Robust Face Detection at Video Frame Rate Based on Edge Orientation Features, *FG'02*, pp. 342-347.

[7] Y. Gizatdinova and V. Surakka, Automatic Detection of Facial Landmarks from AU-Coded Expressive Facial Images, *ICIAP'07*, pp. 419-424.

[8] R. Gonzalez and R. Woods, *Digital Image Processing*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001.

[9] R. Gottumukkal and V. Asari, Real Time Face Detection from Color Video Stream Based on PCA Method, *AIPR'03*, pp. 146-150.

[10] E. Hjelmas and B. Low, Face Detection: A Survey*Comp. Vis. Image Understanding*, Vol.83, 2001, pp. 235-274.

[11] J. Martinkauppi, M. Soriano, and M. Laaksonen, Behavior of Skin Color under Varying Illumination Seen by Dierent Cameras at Dierent Color Spaces, *Mach. Vis. in Industrial Inspection*, V.9, No.4301, 2001, pp. 102-113.

[12] P. Michel and R. Kaliouby, Real Time Facial Expression Recognition in Video Using Support Vector Machines, *ICMI'03*, pp. 258-264.

[13] Y. Rodriguez, F. Cardinaux, S. Bengio, and J. Mariéthoz, Measuring the Performance of Face Localization Systems, *J. Image and Vis. Computing*, Vol.24, 2006, pp. 882-893.

[14] H. Rowley, S. Baluja, and T. Kanade, Neural Network-Based Face Detection, *Trans. Pattern Anal. Mach. Intel.*, Vol.20, No.1, 1998, pp. 23-38.

[15] S. Sclaroff and J. Isidoro, Active Blobs: Region-Based, Deformable Appearance Models, *Comp. Vis. Image Understanding*, Vol.89, No.2–3, 2003, pp. 197-225.

[16] D. Shaposhnikov, L. Podladchikova, and X. Gao, Classification of Images on the Basis of the Properties of Informative Regions, *Pattern Rec. Image Anal.*, Vol.13, No.2, 2003, pp. 349-352.

[17] K. Sobottka and I. Pitas, A Fully Automatic Approach to Facial Feature Detection and Tracking, *Lecture Notes In Comp. Science, AVBPA'97*, Vol.1206, pp. 77-84.

[18] M. Turk and M. Kölsch, Perceptual interfaces, *In G. Medioni and S.B. Kang, (Eds), Emerging Topics in Comp. Vis.*, Prentice Hall, chapter 9, 2004, 45 p.

[19] P. Viola and M. Jones. 2004. Robust Real-Time Face Detection, *Int. J. Comp. Vis.*, Vol.57, No.2, pp. 137–154.

[20] L. Wiskott, J. Fellous, N. Krüger, and C. der Malsburg, Face Recognition by Elastic Bunch Graph Matching, Trans. *Pattern Anal. Mach. Intel.*, Vol.19, No.7, 1997, pp. 775-779.

[21] M. Yang, N. Abuja, and D. Kriegman, Face detection using mixtures of linear subspaces, *FG'00*, pp. 70-76.

[22] M. Yang, D. Kriegman, and N. Ahuaja, Detecting Face in Images: A Survey, *Trans. Pattern Anal. Image Understanding*, Vol.24, 2002, pp. 34-58.

[23] D. Xi and S. Lee, Face Detection and Facial Component Extraction by Wavelet Decomposition and Support Vector Machines, *AVBPA'03*, pp. 199-207.