Development of Pose-Invariant Face Recognition System for Mobile Robot Applications

Taigun Lee*, Sung-Kee Park*, Munsang Kim*, Mignon Park**

* Intelligent Robotics Research Center, Korea Institute of Science and Technology, Seoul, Korea
(Tel : +82-2-958-5619; E-mail: {dungding, skee, munsang}@kist.re.kr)
** Department of Electrical and Electronic Engineering , Yonsei University, Seoul, Korea
(Tel : +82-2-2123-2868; E-mail: mignpark@yonsei.ac.kr)

Abstract: In this paper, we present a new approach to detect and recognize human face in the image from vision camera equipped on the mobile robot platform. Due to the mobility of camera platform, obtained facial image is small and pose-various. For this condition, new algorithm should cope with these constraints and can detect and recognize face in nearly real time. In detection step, 'coarse to fine' detection strategy is used. Firstly, region boundary including face is roughly located by dual ellipse templates of facial color and on this region, the locations of three main facial features- two eyes and mouth-are estimated. For this, simplified facial feature maps using characteristic chrominance are made out and candidate pixels are segmented as eye or mouth pixels group. These candidate facial features are verified whether the length and orientation of feature pairs are suitable for face geometry. In recognition step, pseudo-convex hull area of gray face image is defined which area includes feature triangle connecting two eyes and mouth. And random lattice line set are composed and laid on this convex hull area, and then 2D appearance of this area is represented. From these procedures, facial information of detected face is obtained and face DB images are similarly processed for each person class. Based on facial information of these areas, distance measure of match of lattice lines is calculated and face image is recognized using this measure as a classifier. This proposed detection and recognition algorithms overcome the constraints of previous approach [15], make real-time face detection and recognition possible, and guarantee the correct recognition irregardless of some pose variation of face. The usefulness at mobile robot application is demonstrated.

Keywords: Face Detection, Recognition, Pose invariance, mobile robot vision, color map, pseudo convex hull, lattice lines

1. INTRODUCTION

This paper presents face detection and recognition algorithms using mobile robot vision system. This mobile robot is developed towards indoor personal services, so the importance of 'human-robot interaction' is emphasized. To obtain a lot of information of indoor environment, a stereo camera that is equipped on the top of robot plays an important role in various sensing works, such as visual navigation, human detection and face recognition.

Considering this vision system from a viewpoint of face detection and recognition, image condition is quite a poor because fixed focal length of camera is originally designed for visual navigation. As people may be observed away from the robot's working range, the portion of facial area from entire camera image is tiny and its size small as well. Moreover, the facial images which are obtained from robot's camera are highly pose-variant due to the mobility of robot. Therefore a new algorithm for solving this specific and difficult problem of both detecting and recognizing a small face, irrespective of pose is needed.

The main concern of this paper is handling above two challenging problems (recognition of small and pose-variant faces) simultaneously and presents a solution to real-time application. For face detection, the revised 'coarse to fine' strategy [13] is used. Since too small for directly detecting facial features -two eyes, nose, mouth, etc.- in detail, facial area is rather roughly located by dual ellipse templates of facial color. And then using simplified facial feature maps of characteristic chrominance and the facial intensity patterns, the locations of three main facial features (two eyes and mouth) are estimated. In recognition of pose various face, novel region of interest (pseudo convex-hull) based on facial features is suggested and the pose invariant information of small face is parameterized using random intensity-lattice lines.

Previous researches are presented in section 2. Our algorithms of face detection and recognition are followed in section 3. And we show experimental results in section 4, and conclusion in section 5.

2. PREVIOUS RESEARCHES

In the human-computer interaction, face-interface technique is important for the reason that minimal help of user is needed to identify people. Recently, many efforts on practical technology have reported for more reliable and robust human identification.

Main goal of face detection is locating position of face in uncontrolled environment, whether face detection is considered as a prior step of face recognition. In our case, both face detection and its feature location of variant pose must be considered. Related researches to these works are classified as two approaches, feature-based or view-based approach.

The typical feature-based approaches are that the face or facial features obtained from low level vision algorithms are similar to characteristics of facial geometry (or templates) [1,2], which are advantageous for real-time detection and for coping with the shape or pose variance of object. But, those methods may not work at un-predefined or complex environments and also not in variant light conditions of indoors and outdoors.

View based approach is representing facial image as a holistic arrays of distinctive templates, so the estimation of face location is obtained by calculating correlations between face array and standard face template (i.e. face database images; face DB). Some popular approaches are 1) neural networks methods using statistical represent of normalized gray facial image [3], 2) eigenspace representing of facial image vector (decreased dimension) by principal component approach (PCA) with KL transform[4], 3) Fisher linear discriminant (FLD) to project samples [5]. However, while its merit for high precision rate of detection and its fitness to low resolution image, those have time-consuming problem.

In recent years, face recognition is an active field of study, and also many researches and commercial results have been reported. Especially, if concentrated on pose invariant face recognition, relative researches are classified also into feature-based and view-based approaches.

From a typical example of feature-based approach [6], this technique extracts vectors of geometric descriptors of biometric feature components such as eyebrow thickness, nose anchor points, etc. and these vectors are compared with the stored face model vectors. This approach requires correspond problem, so model generation and matching with non-frontal-parallel pose are more complex and can be achieved at more expense of computations.

Because of the fact that view-based approach does not use detailed biometric information of face, this approach has advantage of inherent simplicity compared with feature-based approach. So, view based methods can handle any variation in face appearance due to changes in pose and lighting, just by storing many different views into face DB. Like view-based detection techniques, image based metric, correlation, is also used to match sampled image with the set of model images. Typical methods are 1) multi-view image (low dimensional) coding of face image with subspace PCA [7], 2) elastic face transform using distinctive feature points modeling of face with neural networks [8], 3) stochastic modeling with Hidden Markov Model [9], which is frequently used technique in speech recognition system.

In additional, the improvement techniques of FR in video-based face processing, face recognition using infra-red images, and multi-modal based person recognition system incorporated with speech recognition are reported [10]

3. PROPOSED METHODS

3.1 Outline of our detection and recognition procedure

First, outlines of our face detection algorithm are as follows. In first step, segmentation of color image for making binary image is performed on the basis of skin color characteristics. For segmenting facial color from various range of skin appearances from different races and light conditions, YCbCr color space is used for simple and compact clustering. The second step, multi-resolution pyramid image [11] by wavelet transform is used for quickly searching the scaled-down image. Wavelet representations of low frequency image have showed good experimental results that the locations of facial area lump in binary image are well preserved during wavelet transform. In next, location of facial features can be obtained from the creation of facial feature map and color chrominance characteristics of features. Considering the average orientation of eye and mouth blobs, geometry of eye-mouth triangle, and presence of a face boundary around the triangle, fine location of each feature is estimated and chosen.

In the beginning of face recognition, the convex-hull facial sub-area includes previously obtained facial features and mainly interested region of face. After preprocessing (gray scale and histogram-equalization), various-pose face images are collected, five or ten images per each person, for the purpose of creating facial database. For handling 3D face pose in 2D facial images, random intensity-lines are scattered over sub-area and face information is obtained from these. And then, recognition step is carried out for input sample image. For all various-pose models and sample image, lattice lines are randomly distributed and distance measure metric between lattice lines of sample and that of each model class are computed. Using this metric, a measure of confidence by nearest-neighborhood classification is defined and the class which maximum sum of confidence measures belongs to is selected for result of recognition.

3.2 Face detection algorithm

The goal for face detection in this research is the estimation of positions of main three facial features-two eyes and mouth, which are basic requirements for next face recognition step. This step is composed of two procedures; the coarse location of face regions and fine location of facial features.

In the first step of face detection, segmentation of color image is performed on the basis of skin color characteristics. For segmenting facial colors from various and large range of skin appearance from different races to different light conditions, we use YCbCr color space[11,12] for simple and compact clustering.

Transforming from RGB space to YCbCr space, we saw that the cluster of skin color is more compact. Fig.1 shows that the skin color distribution in YCbCr space is more compact and easy to cluster facial color class from non color ones.

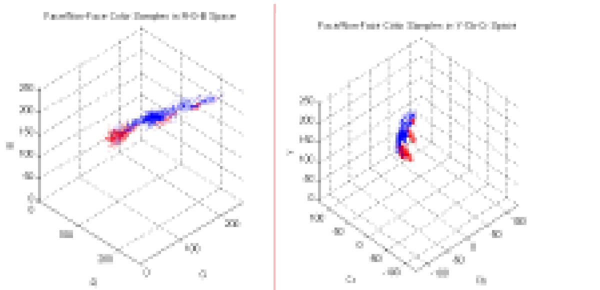


Fig.1. Facial color distribution of RGB and YCbCr space

From such distributions in Fig.1, facial color are segmented using the bounding line conditions (Eq.1~3) in Cb-Cr planes.

Based on YCbCr distribution, binary image of segmented facial color (FC) is as below Fig.2.

$$C_{r-FC} \;:\; r_i = \sqrt{Cr_i^2 + Cb_i^2}\;,\;\; r_i > 0,\; r_i < 70 \qquad (1)$$

$$C_{\phi-FC} \;:\; \phi_i = A\tan 2(Cr, Cb)\,,\;\; 1.6 < \phi_i < 3.0 \;\text{(rad)} \qquad (2)$$

$$FC_i = \bigcup [r_i \cap \phi_i]_{CFC} \qquad (\forall r_i \in C_{r-FC}\,,\;\; \forall r_i \in C_{\phi-FC}) \qquad (3)$$

Second step, we use multi-resolution image using wavelet transform [12], for the reason of searching the scaled-down image to real time computation. Especially, low frequency image of wavelet multi-scale representation have showed good experimental results that the locations of facial area lump in binary FC image are finely preserved in wavelet transform process. The ‘coarse detection’ of face is performed as like.

Fig.2. Binary and gray image based on skin color

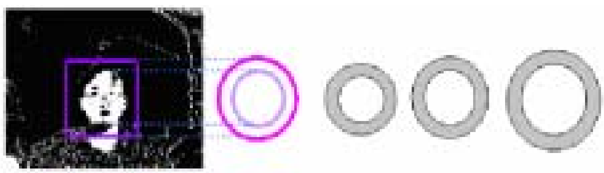
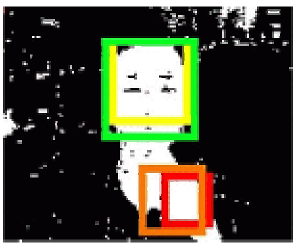


Fig. 3. Multi-scale dual ellipse templates for the coarse detection of face



In case of multiple FC areas, upper regions are selected as more likely face region than lower regions.
(Upper region’s weighting value is high because nearby facial boxes are more closed together)

Fig. 4 Refinement of candidate facial area

For above transformed image, double-ellipse templates are applied to. First, by considering the appropriate face sizes that can be detected in our image, multiple ellipse sets are predefined. Then, as Fig.3, dual-ellipse templates are created like that inner ellipse whose rate of facial colors of pixels is greater than defined threshold and outer ellipse pixels whose rate of non-facial colors are greater than its threshold.

$$E_i \equiv (x_i, y_i)_{pos} \in C_{ellipse} \qquad (4)$$

$$C_{ellipse} = \left\{ (Th_{FC,\min} < E_{inner} < Th_{FC,\max}) \cap (Th_{NFC,\min} < E_{outer}) \right\}$$

With these templates scanned over entire image, multiple satisfied positions and its candidate facial area are obtained. Next, observing all satisfied positions, nearby- located facial area is merged and weighting of this candidate area is increased. Most weighted rectangular area and its position are determined as a candidate facial area. From this operation, in case of the interference of similar face-like region such as Fig.4, the feasibility to correct detect facial area is augmented.

In third step, the location of principal facial features such as two eyes and mouth are estimated by constructing chromatic eye map and mouth map [16].

In this obtained face candidate area, maximum and minimum values of Cb and Cr are found. Then normalized chrominance image (Cb & Cr) of this area can be obtained and the luminance image with gray scale morphology technique (Y-map) is also obtained. Applying modification rule (Eq. 5,6) to above images, chrominance components of facial features can be distinct for segmenting and pixel-labeling. In addition, from additional process of depressing noise and adjusting parameters, some candidate feature regions are more highlighted to be seen clearly.

$$EyeMap = \frac{1}{3}\{(Cb) - (Cr) + (Cb/Cr)\} AND \{Y(x,y) \oplus g_\sigma(x,y)\}$$

$$MouthMap = \alpha\{(Cr) - \beta(255 - Cb)\}\,,\quad \alpha = 1.5, \beta = -1.3 \quad (5),(6)$$

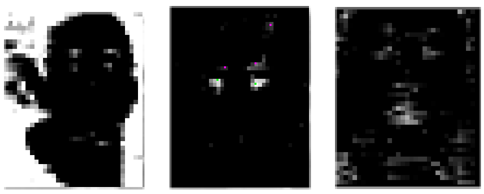


Fig. 5 Examples of chromatic Eye Maps and Mouth Map

Next, three most candidate features’ region in facial area are selected, by means of considering geometry and suitability of facial pattern. First of all, component labeling of prominent feature regions is performed. In following work, each region’s area, location, region-connecting vector and connecting lattice line which represents luminance pattern of region’s interval are respectively obtained. For all combinations of region pairs, their suitability to standard facial form and similarity to facial pattern are verified. Most appropriate pair of feature region is chosen as candidate eye feature regions.

Most candidate mouth region is chosen through as below procedures. To chrominance image composed for mouth feature, five lattice lines are made which points to perpendicular direction of candidate eye region's vector. These lines starts from middle pixel of candidate eye regions and rotated $\pm 5°$ respectively, for the reason of treating the inaccuracy of the estimation result of eye pair and occurrence of some asymmetry of facial figure. While tracking these lattice lines in mouth map image, proper pixel location which has most Cr intensity and is fit for facial geometry is determined as candidate mouth regions.

3.3 Face recognition algorithm

In this paper, modified line-based face recognition method is proposed, well-suited to combine with face detection step. This lattice line based method [15] can handle face rotation problem and is robust to pose, scale variations. Also, otherwise other view-based approaches, small number of database images is required and less processing times for training or recognition is needed. However, the prerequisite of original research -precise boundary pixels of face- is difficult to obtain in our condition; monitoring people on mobile robot platform and varying background of face. Therefore, based on the previous detection result of three main features' location, effective facial range area which is composed by 'convex hull geometric' is used instead. And our modified method is verified from consistency in cases of both given boundary conditions and unknown boundary with given 3 feature locations. Its robustness to pose variance is also showed.

First step is the preprocessing of facial area using pseudo-convex hull definition. Firstly, three point-connecting triangle from facial feature locations is created. Based on distance between two eyes, circular areas are defined in three locations of features and three tangent lines contacting with these area is made. From this way, pseudo convex hull area is defined (Fig.6); this area contains pure and essential pattern information of face. This process eliminates a cause of misrecognition that non-facial region of typical rectangular patch image in previous researches (ex: corner region, hair region, backgrounds, etc.). And this area is also meaningful for an alternative condition of the premise of original paper; facial boundary location of pixels.

Fig. 6 Example of pseudo convex hull area

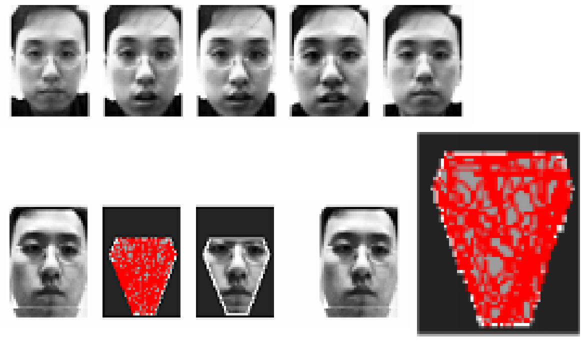


Fig.7 Facial DB, processed PCH area, intensity lattice lines

In the second step, for detected facial area, face view patterns are represented as a combination of random lattice lines in their pseudo-convex hull area. Basic motivation is as follows. Though each intensity-lattice line contains a bit of facial view pattern, but the combination of random line segments can give good descriptions of view patterns, so correct face classification can be expected. For instance, as below facial database examples, area of pseudo-convex hull (PCH) is formed. In this area, as like Fig.7, intensity-lattice lines are arranged and extract the grayscale intensity information of facial patterns. For more detailed equations of constructing random lattice lines, refer to [15]. The pattern information of facial DB from above process is stored in offline works, and same process is applied to detected face area.

In the final step, classification process is performed. Basic distance measure for classification is the distance between two lattice lines is defined as,

$$D(\mathbf{L}_{r,s},\mathbf{L}_{m,n}) = \sum_{q=1}^{l}\left((L_{r,s}^{(q)} - (L_{m,n}^{(q)} + \Delta))^2\right) \qquad (7)$$

Considering a set of lattice lines sampled from one or more face views in detected face image, these unseen (un-stored in face DB) line set can map into certain class on the basis of nearest-neighbor classifier. Concretely, cumulative l_1 -norm error statistic and maximum cumulative error statistic are defined, in pursuit of minimizing sum of distance measure of lattice lines [15].

These line combination shows good result for recognition of face in and out of imaging plane, semi-frontal-lateral view. And the use of simpler 1D line segments as the 2D image representation achieves good efficiency in quasi-real time system. In all images of each class, similarity measure is obtained and total sum result of each class is computed.

Maximum likelihood class is determined from these classification results. Also, novel decision rule of confidence measure factor (Eq.8,9) is defined so as to avoid some easily error case; if difference between most and second largest compounded confidence measure is small, re-evaluation process of confidence measure is performed, like as median filtering, for depressing the error that some peak values exaggerate total sum.

$$\mathrm{CMF} = [\mathrm{TC}_g - \mathrm{TC}_j^{(2)}] / \mathrm{TC}_j^{(2)} \tag{8}$$

If $\mathrm{CMF} < T_{thresh} = 0.65$

$$new\mathrm{TC}_i = \sum_{i=1}^{20} \mathop{\mathbf{Max}}_{j=1\sim N_k} \left[\left(\frac{D_{\max} - D_j}{D_{\max} - D_{\min}} w_1 \right)^{p1} \right] \tag{9}$$

$k = \mathbf{Max}\left[new\mathrm{TC}_i, new\mathrm{TC}_j \right]$, k: final estimated class (10)

4. EXPERIMENTAL RESULTS

4.1 Results of face detection

Compared with previous method [11] which use fixed templates for feature detection, this three step algorithm shows its robustness in detecting pose- and scale-variant face in real applications.

Fig. 8 face detection result from a distance of 1.0m (320x240)

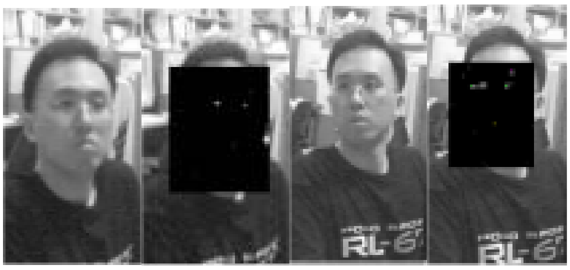


Fig. 9 Facial feature detection in various pose image

Table 1 Average correct feature detection rate with respect to distance from camera and pose variation of face ($f = 4.6$)

Pose / Dist.	± 10°	± 20°	± 30°
0.6(m)	93 %	85 %	74 %
0.9(m)	89 %	77 %	65 %
1.2(m)	81 %	65 %	52 %

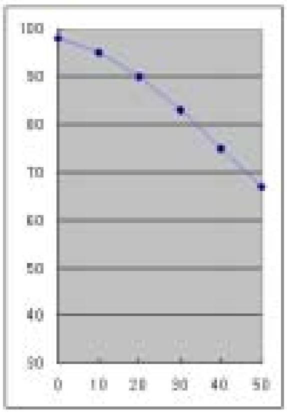
4.2 Results of face recognition

First, the construction of facial DB is needed as follows. For this various images are collected; by changing pose and light condition. About desired number of people, some representative illumination variation images and pose variation images (rotate and node of face) are selected, normalized with histogram-equalized, pseudo-convex hull processed based on each facial feature locations, and grouped & stored into each facial class. At this, random line recognition algorithm to real detected face image is applied. Proper number of lines in each images of face DB examples are 50~100, proper number of lines in detected images are 40~70.

Two type of experimental face samples by detection process are tested, set of standard frontal face images and set of random pose face images. About these face image sets, face recognition is performed and its result is analyzed.

From our test results, this new recognition method has

Fig. 10 Facial DB for face recognition and recognition example of varying pose image samples



Correct recognition rate are much affected by correct detection of facial feature location.

(Y axis : Average correct recognition rate)
(X axis : mismatched rate of feature location with respect to its size, not be rate of feature detection)

Fig.11 Average correct recognition rate combined with facial feature detection rate of previous step

many advantages compared to previous PCA-based face recognition algorithms [11]. First, the correct classification rates are obtained regardless of either standard frontal or random pose form of face image. Also, even if the location of detected facial features are obtained imprecisely for the reason of camera property that just can acquire small facial area, correct recognition rate are well maintained rather than previous method. Additionally, the processes of scaling face images and aligning the location of its facial features are not needed.

5. CONCLUSION

In this research, novel face detection and recognition algorithm that are suitable for mobile robot are proposed.

For detection of face, our new method of locating facial features is distinguished for a applicable solution to pose variant image, which is difficult for previous fixed, scalable template matching method. Also, this algorithm can be used at detection of wide, distant-standing face. Also, our recognition algorithm shows good performances to the pose variance of face due to mobility of robot vision system. This algorithm has property of enduring light changes in some extent. But, if more devised illumination-invariant recognition technique can be combined with this algorithm, this can be more general recognition system. To guarantee high correct recognition rates of human face, precise facial feature-detection results must be need above all. These algorithm leaves room for improvement that includes multi-view facial pose and perturbation of facial features' location and these questions will be more considered in the future.

REFERENCES

[1] R. Brunelli and T. Poggio, "Face recognition: Feature ver-sus templates", IEEE Trans. Pattern Analysis and. Machine Intell., vol.15, 1993, pp.1042 ~1052.
[2] Cox, I.J.; Ghosn, J.; Yianilos, P.N. , "Feature-based face recognition using mixture-distance", Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, IEEE Computer Society Conference on , 1996, pp.209 -216
[3] H.A.Rowley, S.Baluja, and T.Kanade, "Neural Network-based Face Detection", IEEE Trans. Pattern Analysis and. Machine Intell., vol.20, no.1, pp.23 ~38, Jan.1998
[4] K.-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," Technical Report AI Memo 1521, Massachusetts Inst. of Technology AI Lab, 1994.
[5] M.H. Yang, N. Ahuja, and D. Kriegman, "Mixtures of Linear Subspaces for Face Detection," Proc. Fourth Int'l Conf. Automatic Face and Gesture Recognition, pp. 70-76, 2000
[6] R. Brunelli, "Estimation of Pose and Illuminant Direction for Face Processing", Image and Vision Computing, vol. 15, pp. 741-748, 1997.
[7] B. Moghaddam, A.P. Pentland, "Probabilistic visual learn-ing for object detection", IEEE Trans. Pattern Analysis and Machine Intell., vol.19, July. 1997
[8] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 775-?, 1997.
[9] F. Samaria and A. Harter, "Parametrisation of a Stochastic Model for Human Face Identification", Proc. Second IEEE Workshop Applications of Computer Vision, Sarasota, Fla., Dec. 1994.
[10] W.Zhao, R.Chellappa, A.Rosenfeld, P.J.Phillips, "Face Recognition : A Literature Survey", Tech.Report, 2000
[11] Taigun Lee, Sung-Kee Park and Munsang Kim, "Face Detection and Recognition with Multiple Appearance Models for Mobile Robot Application", Proceedings ICCAS, 2002, pp.
[12] C. Garcia and G. Tziritas, "Face Detection using Quantized Skin Color Regions Merging and Wavelet Packet Analy-sis", IEEE Trans. Multimedia, vol.1, no.3, Sep. 1999
[13] Francois Fleuret Donald Geman, "Coarse-to-Fine Face Detection", Avant-Projet IMEDIA, June 2000
[14] M.H.Yang, D.Kriegman, N.Ahuja, "Detecting face in Images: Survey", IEEE Trans. Pattern Analysis and Machine Intell. vol.24, Jan. 2002, pp.33-58
[15] Olivier de Vel, S.Aeberhand, "Line-based FR under varying pose", IEEE Trans. Pattern Analysis and Machine Intell. vol.21, Oct. 1999, pp.1081-1088
[16] R.L.Hsu, Mohamed Abdel-Mottaleb, Anil K. Jain, "Face detection in Color images", IEEE Transaction on Pattern Analysis and Machine Intellligence, vol.24, no.5, may 2002, pp.696~
[17] X.Zhu, J.Fan, A.K.Elmagarmid, Towards Facial Feature Extraction and Vefication for Omni-Face Detection in Video/Images, IEEE Int'l, Conf. on Image processing, 2002, pp.113~116