Adaptive Eye Location Using FuzzyART

Jo Nam Jung, Mi Young Nam, and Phill Kyu Rhee

Dept. of Computer Science & Information Engineering, Inha University,
253 Yong-Hyun dong, Incheon, South Korea
jjn10@korea.com, rera@im.inha.ac.kr, pkrhee@inha.ac.kr

Abstract. In this paper we propose a method of locating face and eyes using context-aware binarization. Face detection obtains the face region using neural network and mosaic image representation. Eye location extracts the location of eyes from the detected face region. The proposed method is composed of binarization, connected region segmentation by labeling, eye candidate area extraction by heuristic rules that use geometric information, eye candidate pair detection, and eye area pair determining by ranking method. Binarization plays an important role in this system that converts a source image to a binary image suitable for locating eyes. We consider edge detection based and image segmentation based binarization methods. However, each method alone cannot be used a solution in general environment because these are influenced by the factors such as light direction, contrast, brightness, and spectral composition. We propose a hybrid binarization using the concept of illumination context–awareness that mixes two binarization methods in general environment.

1 Introduction

Face and eyes are considered as a way of communication among people for a long time. For communication between person and computer, many researchers have studied face image processing such as face tracking, face detection, recognizing face and facial expression, lip reading, eye blinking, etc. Therefore, many algorithms and techniques are invented, but it remains a difficult problem yet, [1], [5]. The automatic face processing becomes a significant topic towards developing an effective HCI (Human-Computer Interface) [1], [6], [7]. Eye location with the information of face region consists of largely preprocessing, binarization, connected region segmentation by labeling, detecting of candidates for eye region and eye pair using heuristic rules based on geometric information and determining eye pair by Ranker. Binarization method that converts an original image to a binary image suitable for eye location is considered as edge detection and image segmentation. However, each method is dependent on the environmental factors such as light direction, contrast, brightness, and spectral composition. For solving this problem we propose a new method of mixing two binarization method. We introduce a concept of the context-aware binarization for solving this problem. The changes of illumination environment can be detected by analyzing the input images. We assume that the illumination environment changes continuously. We apply this methodology to face detection and eye location. In section 2, we present binarization. We present face detection in section 3 and locating of eye regions section 4. Finally, we give experimental results and conclusions.

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3611, pp. 109 – 118, 2005.

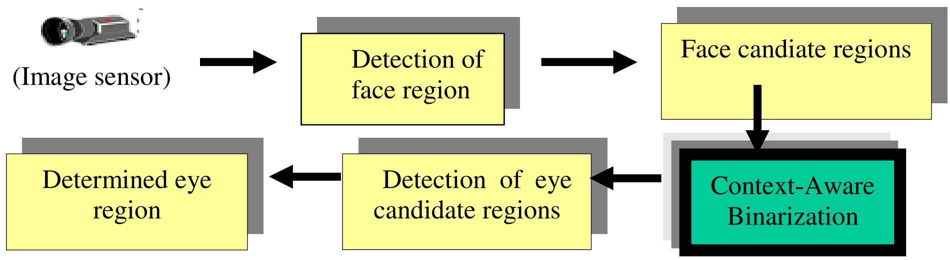


Fig. 1. Block diagram of eye locating process

1.1 Edge Detection

Image edges are defined as local variations of image intensity. This variation is gotten by edge detector operators. The magnitude of image gradient $\nabla f(x, y)$ is given by [3],

$$e(x, y) = \left| f_x(x, y) + f_y(x, y) \right| \tag{1}$$

Edge Image is obtained by an edge detector using *e(x,y)*. This image carries information about the edge magnitude. If the edge detector output is large, a local edge is present.

1.2 Image Segmentation

When an image contains an object having homogeneous intensity and a background with a different intensity level, the image can be segmented in two regions using image segmentation by threshold [3]. The following equation is the definition of segmentation adopted in this paper. where T is threshold whose choice can be based on the image histogram. The gray level of facial components such as hair, eyebrow, pupil, nostril, lip, etc. is darker than that of the skin. The threshold T can be obtained

$$E(x, y) = \begin{cases} 1 & \text{if} \quad e(x, y) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

nostril, lip, etc. is darker than that of the skin. The threshold T can be obtained using this property. The threshold is used to discriminate pixels belonging to eye region from skin region. Therefore, if we use threshold as a fixed value we cannot obtain a good image because we always experiment in different environments. For solving this problem we experiment a method getting a threshold *T* using the statistical information of the edge magnitude and the property that the larger the edge magnitude is, the higher possibility of being edge is.

2 Face Detection

The face location algorithm detects a face in an image using mosaic image representation and back-propagation neural network [4]. Because consecutive images obtained from image sensor usually contain a lot of noises, we perform histogram equalization, mosaic image, and normalization as a preprocessing step. When the histogram of some image is biased to one side, histogram equalization can make the histogram distribution uniform. This improves the contrast of image and makes image features more distinguishable. The input image is converted into a low resolution image called mosaic image [5]. The mosaic image representation provides fast noise-insensitive processing. From this mosaic image, all possible 8×8 mosaic called octet faces (the second figure of Fig. 2) are extracted. Each cell of octet is normalized in the 0 ~ 1 range of a real value. The octet is checked by the back-propagation neural network to extract the most suitable face region. The first figure of Fig. 2 is the original image captured from image sensor, and the second is mosaic image representation. This algorithm returns the coordinates of face region.

(a) Original image (b) Octet face

Fig. 2. The original image and the octet face

Fig. 3 shows examples of the result of face location. Two white rectangles of the second image mean that the number of possible face candidate is two.

Fig. 3. Examples of face detection

3 Locating of Eye Regions

Fig. 4 roughly describes the model that we propose for locating face and eye regions. Eye location begins with inputting each face candidate regions into eye location process given in Fig. 4. However, if image is captured in weak illumination, generally face candidate region becomes a small intensity region, i.e. Face candidate region's image contrast become poor and the subjective image quality become low. For enhancing image quality, histogram equalization that modifies its histogram performs in face candidate regions [4].

We found that binarization processed by edge detection algorithm is efficient when the candidate region is a dark image and binarization by segmentation algorithm is efficient when the candidate region is normal or bright from experiment. Input face candidate region is analyzed using neural network in the context awareness module. The binarization is performed differently according to the analyzed illumination condition. If the image illumination condition is dark, an input face candidate region is binarized by the edge detection method. If the condition is normal or bright, it is done by the segmentation method.

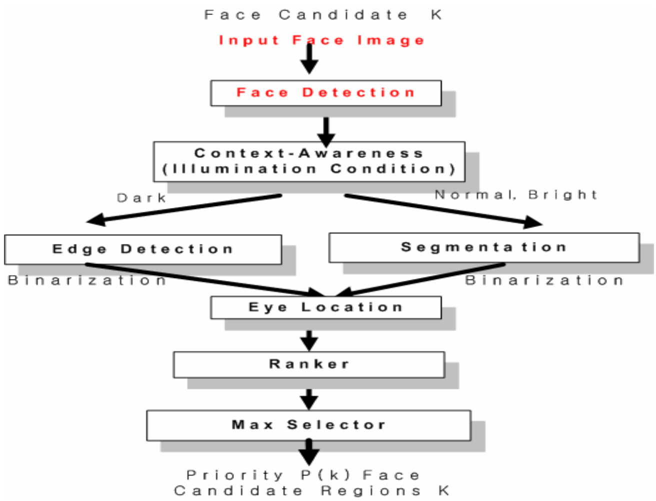


Fig. 4. The block diagram of face and eye location

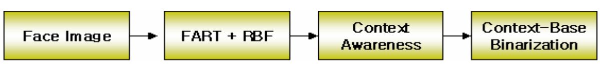
3.1 Face Candidate Regions

Eye locating is executed in face candidate regions detecting by mosaic and back propagation neural network. However, face candidate region exists more than one. This means that face candidate region is the completely wrong region and whether it can contain a pair of eyes or not.

(a) (b)

Fig. 5. Examples of face candidate regions

3.2 Binarization and Labeling



3.2.1 Context Awareness Using FART

In the proposed approach, illumination classification is dealt with as local targets to be detected in an image. We formulate the object detector generator to learn to discriminate object pattern witch can be measured formally, not intuitively as most previous approaches did. The illumination estimator classifies a face pattern into one of the face illuminant set. The facial pose estimation can be used to verify the detection of context-based faces. Gong and colleagues use non-linear support vector regression to estimate facial pose. In this paper, we propose how combined supervised learning and unsupervised learning. The proposed learning based classifier generator has advantages over previous approaches.

3.2.2 Context Awareness by FART+RBF

The RBF networks, just like MLP networks, can therefore be used in classification and/or function approximation problems. In the case of a RBF network, we usually prefer the hybrid approach, described below [10], [11], [12]. The RBFNs, which have a similar architecture to that of MLPs, however, achieve this goal using a different strategy. One cluster center is updated every time an input vector x is chosen by FuzzyART from the input data set. The cluster nearest to x has its position updated using

$$W_{ji}(t+1) = \beta(I \wedge W_{ji}(t)) + (1-\beta)W_{ji}(t) \tag{3}$$

FuzzyART is a variant of ART system derived from the first generation of ART, namely ART1. It is a synthesis of ART algorithm and Fuzzy operator. ART1 can only accept binary input pattern, but FuzzyART allows both binary and continuous input patterns [9], [13]. The feature space of object instance with multiple viewing angles must be clustered properly so that the location error can be minimized. However, the classification of multiple viewing points is very subjective and ambiguous. Thus, we adopt FuzzyART and RBF methods for achieving an optimal pose classification architecture. Executed step is as following to FuzzyART [9], [13], [14]. In this Paper, Clustering's performance improves by studying repeatedly about done data.

The cluster center is moved closer to x because this equation minimizes the error vector. Each hidden unit calculates the distance of the input vector from the corresponding Gaussian:

$$\phi_j(x) = \exp\{-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\} \tag{4}$$

In this paper, centers are obtained from unsupervised learning (clustering), FuzzyART Algorithm. The weights between the hidden units and the output layer, denoted by w_{kj}, are regular multiplication weights (as in a MLP).

$$y_k(x) = \sum_{j=1}^{M} w_{kj}\phi_j(x) + w_{k0} \tag{5}$$

Where x is the input vector, m_j is the jth prototype vector, σ_j is the width of the Gaussian of that prototype or cluster centre. There are various approaches for training RBF networks. In this paper, centers are obtained from unsupervised learning (clustering), FuzzyART algorithm. Clustering (FuzzyART algorithm) and LMS are iterative. This is the most commonly used procedure. Typically provides good results. After finding a suitable cluster using clustering algorithm, do laying center on this. The winning node μ_j is what FuzzyART is its best match for the input pattern. Hidden node's center determined by unsupervised learning, FART.

As showed Fig. 6, the idea is to train the network in two separate stages - first, we perform an unsupervised training (FuzzyART) to determine the Gaussians' parameters

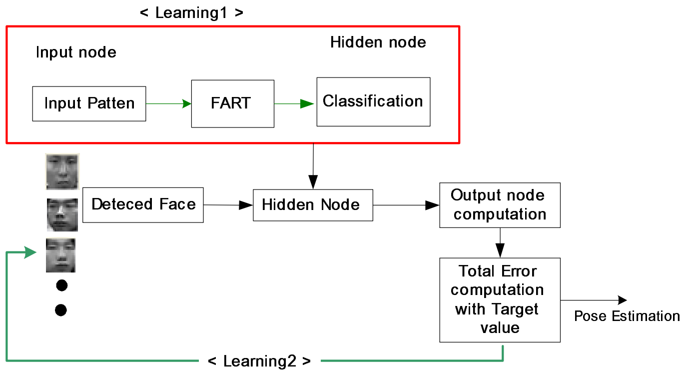


Fig. 6. Training System Architecture

(j, j). In the second stage, the multiplicative weights w_{kj} are trained using the regular supervised approach. Input pattern is vectorized for grayscale image size of 20x20 pixels, input node had mosaic of size of 10x10 pixels. The transformation from the input space to the hidden unit space is non-linear, whereas the transformation from the hidden-unit space to the output-space is linear. RBF classifier expand input vectors into a high dimensional space. RBF network has architecture that of the traditional three-layer back-propagation. In this paper, hidden units is trained using FuzzyART network and basis function used are Gaussians. The proposed network input consists of n normalized and rescaled size of 1/2 face images fed to the network as 1 dimension vector. And input unit has floating value [0, 1]. The vector value is normalized. In case learn by FuzzyART, performance is best in case used picture itself by input node vectorized.

3.2.3 Binarization

The binarization method explained in section 2. The labeling algorithm examines a state of connection between a pixel and its 8-neighborhood pixels, and labels the objects in the image. The labeling algorithm is used to obtain connected and isolated regions [2]. Fig. 7 (a) shows the result of image segmentation by thresholding in a face region. Fig. 7 (b) shows that facial components such as eye, eyebrow, and nose are labeled by the labeling algorithm. The labeled facial components are recorded as eye candidates.

(a)Image segmentation

(b) Labeling

Fig. 7. Example of labeling

3.3 Detection of Candidates of Eye and a Pair of Eyes

Candidates of eye and a pair of eyes are detected by heuristic rules based on eye's geometrical information in face. Eye candidate's width has much correlation with face candidate's height. However, because the height of eye candidate has an extreme changeability on condition of blinking one's eyes, it is difficult to set a correlation between eye candidate's height and face candidate region. Eye detection has a tendency to acquire the boundary of eye. However, image segmentation have a more concern in pupil regions than the boundary of eyes, so the connected regions by labeling have a tendency to become more smaller and more nearer a square than ones obtained by edge detection. Therefore, rules used in eye location differ as to binarization methods. The example of detecting candidates of eye region is shown in Fig. 8.

(a) Connected region segmentation by labeling

(b) Candidates of eye region

Fig. 8. Example of detecting candidates of eye region

Once eye candidate regions are detected, candidates of a pair of eyes are detected. First, because two eyes locate in a similar position by y -axis, eye candidate regions are sorted by y -axis. Candidates of a pair of eyes satisfy the following rules- the gradient of two eyes, comparison between two eye's size and distance between two eyes. These rules also differ as to binarization method similar to the rule of eye candidate regions.

3.4 Ranker

After eye location process, Ranker calculates an energy of each candidate of a pair of eyes detecting in each face region F_k . Each energy obtained by Ranker inputs to *Max* and it selects a max value among those. A candidate of a pair of eyes whose energy is equal to this selected value become a pair of eyes. Fig. 9 shows a template of a pair of eyes that are represented with terms of the size, shape, positional information of eyes, eyebrows and mouth. Using this template, Ranker calculates an energy of a pair of eyes.

Eq (6) is an equation representing a template given in Fig. 8. E_L, E_R, E_M which are the energy, using terms of image intensity, edge magnitude, the region information obtained by connected region segmentation and so forth, calculates in the left eye, right eye, mouth respectively. In addition, E_{LR}, E_{LM} and E_{RM} are the energy between two eyes, the left eye and the mouth, the right eye and the mouth, respectively. Finally, E_{LL} is the energy between the left eye and the left eyebrow, and E_{RR} is the energy between the right eye and the right eyebrow.

$$E = k_1 E_L + k_2 E_R + k_3 E_M + k_4 (E_{LR} + E_{LM} + E_{RM} \\ + k_5 (E_{LL} + E_{RR}) \tag{6}$$

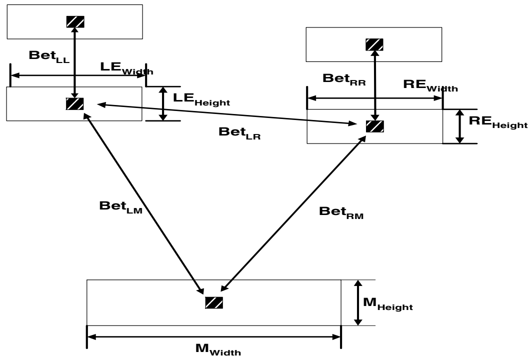


Fig. 9. The template of a pair of eyes

3.5 Determination a Pair of Eyes

There can exist more than one candidate for a pair of eyes obtained in each candidate face region. Now, determined eyes have a maximum value of energy by Ranker. Therefore if there exist *m* candidates for a pair of eyes obtained in a face candidate region *Fi*, each candidate for a pair of eyes is expressed as Eye^i_k, $1<=i<=m$ and its energy calculated by Ranker is define as $E(Eye^i_k)$. Therefore, the energy of a determined pair of eyes in face candidate region *Fi* are defined as below Eq (7),

$$E(k) \;=\; \underset{1\le i\le m}{MAX}\left\{P(Eye^i_k)\right\} \tag{7}$$

and finally, the determined eyes of this facial image should satisfy Eq (8).

$$\underset{1\le k\le m}{MAX}\left\{E(k)\right\} \tag{8}$$

Fig. 10 shows a process of eye location. Fig. 10 (c) are the results of executing the binarization, connected region segmentation by labeling and conditioning the rule that will be satisfied by a candidate for a pair of eyes. we determine as eye regions because of it having maximum value among them.

(a) Face candidates (b) Labeling (c) Candidates of eye region (d) Located eye regions

Fig. 10. A process of eye location

4 Experimental Results and Conclusion

Our proposed method of locating eye regions was developed with MFC(Microsoft foundation class) in Pentium Ⅳ 2.4GHz CPU PC having Windows XP as operating system. Compiler used was Visual C++ 6.0. Experimental images were captured from image sensor with 320×240 size and 256 gray levels and were collected totally 400 frames from 35 persons(see Table 1).

Table 1. Final results o f different methods of eye location

Glasses	Number of frame	Method of binarization	Total sum	
			Number of Success	Successful rate
Not wearing	247	E	243	98.38%
		S	231	93.52%
		E+S	243	93.38%
Wearing	153	E	116	75.82%
		S	138	90.20%
		E+S	143	93.46%
Total sum	400	E	359	89.75%
		S	369	92.25%
		E+S	386	96.50%

This successful rate of eye locations largely depends on the environmental factors image quality, illumination, glasses and hair of head and so forth. In this experiments, only the factor of glasses was considered and experimental image s were classified by the factor of glasses. These classified images were separately experimented with different methods edge(E), image segmentation (S) and combined method(E+S) and were compared with each result. According to the above result using combined method was superior to single method. In a stable illumination, in case not wearing glasses we could get successful rate using only edge method. However, if glasses were worn, successful rate of edge method was lower than that of segmentation method and at the same time especially in case black glasses worn frame edge method made the rate of success remarkably lowered. Consequently, using a single method is suitable for a particular environment. Nevertheless, because the environment getting images isn't always stable in the real world, if we can't select appropriate method we can't get a good result. So, in this paper we combined two binarization method for locating eye regions and can get a good result in a general environment as well as a particular one. In the future's experiments, we will have experimented on locating eye regions in the more real world's environments an object near eyes and the reflection of a beam of light off a glasses and so forth.

References

1. Lars-Peter Bala., Kay Talmi., and Jin Liu.: Automatic Detection and Tracking of Faces and Facial Features in Video Sequences. Picture Coding Symposium 97, (Sept. 1997) 251-256
2. R. L. Lumina., G. Shapiro., and O. Zuniga.: A New Connected Components Algorithm for Virtual Memory Computers. *Computer Vision, Graphics, and Image Processing.*, Vol. 22 (1983) 287-300
3. I. Pitas, *Digital image processing algorithms.* Prentice Hall (1992)
4. R. Klette., and P.Zamperoni.: *Handbook of image processing operators.* John Wiley&Sons, 1996

5. Zhou ZH., and Geng X.: Projection functions for eye detection. Pattern Recognition 37(5), no.(5) (2004) 1049-1056
6. Bin Chen., Zhi-Qiang Liu., and Xiang-Hua Zhu.: Eye Location In Human Face Images Using Fuzzy Integral. IEEE Trans. On Proceedings of the Second International Conference on` Machine Learning and Cybernetics, Vol.14. no.2-5 (Nov.2003) 2500-2505
7. Li Weijun., Xu jian., and Wang Shoujue.: A Fast Eye Location Algorithm Based on Geometric Complexity. Proceeding of the 5th World Congress in Intelligent Control and Automation, Hangzhou, P.R. China, Vol. 5. no. 6 (2004) 4105-4107
8. Qu Ying-Dong., Cui Cheng-Song., Chen San-Ben and Li Jin-Quan.: A fast subpixel edge detection method using Sobel-Zernike moments operator. Image and Vision Computing, China, Vol. 23. Issue 1, no.1 (Janu. 2005) 11-17
9. T.Kasuba: Simplefied Fuzzy ARTMAP. AI Expert, November (1993) 18-25
10. Shakunaga, T.: An Object Pose Estimation System Using A Single Camera. Proc. of IEEE international conference, Vol.2. (1992) 7~10
11. Paulino, A. Araujo, H.: Pose Estimation for central catadioptric system. An analytic Approach. Proc. Of Pattern recognition (2002) 969-699
12. Huang, J. Shao, X Wechsler, H.: Face Pose discrimination using support vector machanism(SVM). Proc. of Pattern Recognition, Vol.1. (1998) 154-156
13. G.A. Carpenter et al: Fuzzy ARTMAP : A neural network architecture for inceremetal supervised learning of analog multidimensional maps. IEEE Trans. Neural Networks, Vol.3, No. 5. September (1992) 698-712
14. Ramuhalli, P., Polikar, R., Udpa L., Udpa S.: Fuzzy ARTMAP network with evolutionary learning. Proc. of IEEE 25th Int. Conf. On Acoustics, Speech and Signal Processing (ICASSP 2000), Vol. 6. Istanbul, Turkey (2000) 3466-3469