# FACE RECOGNITION USING SURFACE FEATURES IN XYI SPACE

*Noriji Kato, Motofumi Fukui, and Hirotsugu Kashimura*

Corporate Research Laboratory, Fuji Xerox Co., Ltd.
430 Sakai, Nakai-machi, Ahigarakami-gun, Kanagawa, 259-0157 Japan
{noriji.kato, motofumi.fukui, kashimura.hirotsugu}@fujixerox.co.jp

## ABSTRACT

We propose a face recognition algorithm that utilizes novel surface features in (x, y, I(x,y)) space. A face image is considered as a surface in XYI space, and the surface is segmented into definite number of regions by using Gaussian Mixture Model. Parameters of each Gaussian distribution are determined by maximizing log-likelihood function, and stored as features of individual face image. In recognition process, the log-likelihood is used as similarity between a test image and the stored features. The face recognition performance of our algorithm is evaluated with FERET database. Our algorithm achieves identification rate of 95.4% and equal error rate of 1.4%, which are superior to other algorithms based on eigenface features and Gabor wavelet features.

## 1. INTRODUCTION

Many researchers believe that human visual system can reconstruct 3-dimensional information of objects from their 2-dimensional images projected on retina[4]. However, current approaches for face recognition do not make use of such 3-dimensional information explicitly. For example, subspace projection[7] and Gabor Wavelet filtering [8], which are widely used to extract discriminative features, are 2-dimensional image processing.

In this paper, we incorporate 3-dimensional information extracted 2-dimensional face images into face recognition algorithm. The algorithm enables us to utilize geometrical structure of faces in addition to conventional 2-dimensional features. Especially information about the geometrical structure would improve recognition performance for noisy or low resolution images, because facial pattern is no longer discriminative feature for those images.

Generally, a 2-dimensional image projected from a 3-dimensional object is determined from normal vector of an object surface, reflectance of the surface, and the directions of light sources. By using this relation, depth information of the object is restored from the 2-dimensional shade pattern. This problem is called Shape from Shading (SFS) [3], and the depth information can be calculated under an ideal condition. Although several applications of SFS to face images have been reported[1, 9], it is difficult to extract depth information stably, because geometrical structure of a face is very complicated and its reflectance is not uniform.

Therefore, instead of restoring depth information, we attempt to extract surface features that are useful to discriminate individual persons. In our approach, a face image is considered as a surface in (x, y, intensity) space [5]. We call this space XYI space. Then, the surface is segmented into definite number of regions so that all pixels in each region have similar intensity. We consider the segmented regions as surface features of a face, because similar intensity means similar normal vector of real 3-dimensional surface under the Lambertian model.

To represent a surface in XYI space, we adopt Gaussian Mixture Model, because it can capture local covariance of data distribution well. First, parameters of Gaussian Mixture Model, such as centers, covariant matrices, and mixture coefficients, are determined by maximizing log-likelihood function. After that each pixel is segmented into a class whose posterior probability is the highest. We consider the parameters of Gaussian Mixture Model as features of the face image, and utilize for face recognition.

## 2. SURFACE FEATURE IN XYI SPACE

### 2.1. Surface Segmentation by Gaussian Mixture Model

We consider to segment a surface in XYI space by Gaussian Mixture Model. The surface is treated as a set of samples in XYI space, and the *i*-th sample is expressed by

$$\mathbf{x}_i = (x_i, y_i, I(x_i, y_i)), \tag{1}$$

where $x_i$, $y_i$ and $I(x_i, y_i)$ are *x* coordinate, *y* coordinate, and intensity of the *i*-th pixel of a face image. The set of $(x_i, y_i)$ covers all pixel positions in the image. Then, we model probability $p(\mathbf{x}_i; \Theta)$ of *i*-th sample as linear combination of *m* Gaussian distributions as follows.

$$p(\mathbf{x}_i; \boldsymbol{\Theta}) = \sum_{j=1}^{m} \xi_j p(\mathbf{x}_i, \theta_j) \tag{2}$$

Here, $\mathbf{\Theta} = (\theta_1, \ldots, \theta_m)$ is a set of Gaussian distribution parameters, where $\theta_j$consists of a center $\mu_j$ and a covariant matrix $\mathbf{\Sigma}_j$ of the *j*-th Gaussian distribution. $\xi_j$ is the mixing coefficients of the *j*-th Gaussian distribution. The conditional probability of the $\mathbf{x}_i$for the *j*-th Gaussian distribution is expressed by

$$p(\mathbf{x}_i, \theta_j) = \frac{1}{(2\pi)^{3/2}} \exp \{ -\frac{1}{2}(\mathbf{x}_i - \mu_j)^T \mathbf{\Sigma}_j^{-1} (\mathbf{x}_i - \mu_j) \} . \quad (3)$$

The parameter of each Gaussian distribution is determined by maximizing sum of log-likelihood over all samples on the surface. After parameter determination, each sample is classified into a class whose posterior probability is the highest.

### 2.2. Parameter Determination by EM Algorithm

The EM algorithm is widely used to determine parameters of various mixture models[2]. The algorithm maximizes log-likelihood for complete data that consists of an observable $\mathbf{x}_i$ and a hidden variable *j* from which the observable was generated. The EM algorithm iterates between the E-step and the M-step to maximize the complete likelihood. The E-step calculates the expectation value of the complete likelihood $Q\left(\mathbf{\Theta} \mid \Theta^{(t)}\right)$ assuming the hidden variable *j* is determined, and the M-step maximizes the expectation value. In the case of (2), $Q\left(\mathbf{\Theta} \mid \Theta^{(t)}\right)$ at step *t* is calculated in the E-step as

$$Q\left(\mathbf{\Theta} \mid \Theta^{(t)}\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} q^{(t)}\left(j \mid \mathbf{x}_i\right) log \left\{\xi_j p(\mathbf{x}_i, \theta_j)\right\}, \quad (4)$$

where *n* is the total number of pixels. $q^{(t)}\left(j \mid \mathbf{x}_i\right)$ is the posterior probability of the *j*-th class that is calculated by using the estimated parameters $\Theta^{(t)}$ at step *t* as follows.

$$q^{(t)}\left(j \mid \mathbf{x}_i\right) = \frac{\xi_j^{(t)} p(\mathbf{x}_i, \theta_j^{(t)})}{\sum_{l=1}^{m} \xi_l^{(t)} p(\mathbf{x}_i, \theta_l^{(t)})} \quad (5)$$

In the M-step, the parameters are updated by maximizing (4). The update equations are

$$\xi_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} q^{(t)}\left(j \mid \mathbf{x}_i\right) \quad (6)$$

$$\mu_j^{(t+1)} = \frac{1}{n\xi_j^{(t+1)}} \sum_{i=1}^{n} q^{(t)}\left(j \mid \mathbf{x}_i\right) \mathbf{x}_i \quad (7)$$

$$\mathbf{\Sigma}_j^{(t+1)} = \sum_{i=1}^{n} q^{(t)}\left(j \mid \mathbf{x}_i\right) (\mathbf{x}_i - \mu_j^{(t+1)})(\mathbf{x}_i - \mu_j^{(t+1)})^T . \quad (8)$$

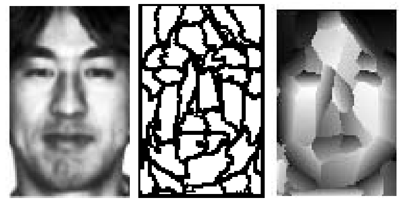

**Fig. 1**. Surface feature extraction by segmentation: A face image (left), segmentation of the image (middle), the most probable image by the model (right). The number of segmented regions is 36.

The parameters of the *j*-th Gaussian distribution can be determined by repeating these E steps and M steps.

### 2.3. Extracted Features

After determination of the parameters, we can obtain posterior probability for each sample from (5). By using the posterior probability, the sample $\mathbf{x}_i$ is classified into class $C_j$ as follows.

$$\mathbf{x}_i \in C_j \, , j = \arg_k \max q\left(k \mid \mathbf{x}_i\right) \quad (9)$$

Figure 1 shows an example of surface segmentation by our algorithm. A face image in fig.1(a) is segmented into 36 regions as shown in fig.1(b). To show extracted surface features graphically, we generate the *most probable image* as fig.1(c) by determining intensity of each pixel so that its log-likelihood (4) becomes maximum. From fig 1(c), it is found that the Gaussian Mixture Model captures geometrical structure of the face, such as contour, prominent of the nose, and hollows around the eyes.

It is remarkable point that our algorithm can extract contour of the face stably. I think that it would be advantage of our algorithm, since most of current approaches do not utilize contour information for face recognition[8, 10].

## 3. DEFORMABLE TEMPLATE

In our algorithm, the Gaussian Mixture parameters are stored as the templates of a person. For recognition, similarity between an input face image and stored templates has to be measured. We can use the log-likelihood in (4) as the similarity.

In actual situation, face images of the same person vary a lot according to their expression, posture, etc. Therefore, *deformable templates* that absorb variation of the image will improve recognition performance[8]. In our algorithm, the *deformable templates* can be realized by the M-step calculation which allows movement of the center of each Gaussian distribution.

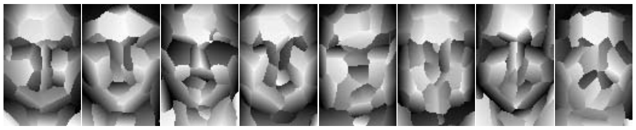

**Fig. 2**. Examples of the most probable images generated from the surface features.

The M-step calculation to deform the templates is slightly different from that in feature extraction. The difference is that a penalty term is added to (4) to prevent large movement of the centers. The modified log-likelihood is expressed by

$$\begin{aligned} Q'\left(\boldsymbol{\Theta} \mid \Theta^{(t)}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{m} q^{(t)}\left(j \mid \mathbf{x}_i\right) log\left\{\xi_j p(\mathbf{x}_i, \theta_j)\right\} \\ &+ \eta \sum_{i=1}^{n}\sum_{j=1}^{m} q^{(t)}\left(j \mid \mathbf{x}_i\right)\left(\theta_j - \theta_{0j}\right)^2 \end{aligned} \tag{10}$$

where $\eta$ is a positive constant and $\theta_{0j}$ is the initial center of the *j*-th Gaussian distribution that is stored as a template. In the M-step, centers of the Gaussian distributions are updated by maximizing (10) as follows.

$$\begin{aligned} \mu_j^{(t+1)} &= \left(1 + 2\eta\boldsymbol{\Sigma}_j\right)^{-1} \\ & \left\{\frac{1}{n\xi_j^{(t+1)}}\sum_{i=1}^{n} q^{(t)}\left(j \mid \mathbf{x}_i\right)\mathbf{x}_i + 2\eta\boldsymbol{\Sigma}_j\mu_{0j}\right\} \end{aligned} \tag{11}$$

Here, we do not update the covariant matrix, because it represents the surface feature. The amount of the movement is controlled by $\eta$. That is, smaller $\eta$ allows larger movement.

By repeating above E step and M step several times for an input image, the center of Gaussian distribution moves and compensates the variation of the face image.

## 4. EXPERIMENTAL RESULTS

### 4.1. FERET database

To test our approach, we use fa and fb frontal face images of the FERET database which has been widely used to evaluate face recognition algorithm[6]. There are 1196 fa images and 1195 fb images. The fa and fb images of the same person vary only in expression. All images are normalized so that distance between eyes is 30 pixels, and cropped into the images of 64-pixel width and 100-pixel height so that the images include contour information of the faces. The eye positions are supplied with the FERET database. The cropped images are smoothed by 3x3 Gaussian filter, and are histogram equalized.

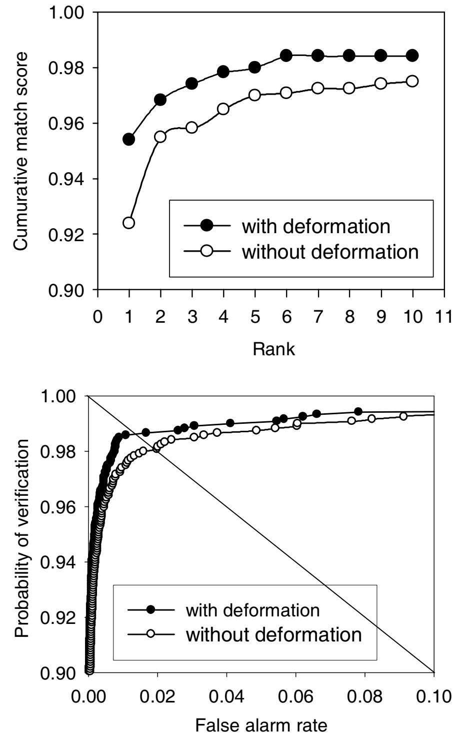


**Fig. 3**. Identification rate(upper) and ROC curve(lower)

### 4.2. Feature Extraction

We use fa images as gallery images, and fb images as probe images. For all gallery images, the Gaussian Mixture parameters are determined and stored as the surface features. The number of Gaussian distributions are fixed at 36. At the beginning of the parameter determination, the cropped image is divided into 6x6 rectangle regions of the same size, each of which corresponds to individual Gaussian distribution. Then, the center and covariance of each region are calculated and set as initial Gaussian distribution parameters. The mixture coefficients are set to the equal value. After setting of initial parameters, 20 times of E-step and M-step are repeated to determine the parameters.

Figure 2 shows feature extraction results. It is found that our algorithm captures surface features, which varies significantly among the persons.

### 4.3. Identification and Verification Performance

Face recognition performance of our algorithm is evaluated by both identification and verification test. In both tests, similarity between a probe image and features extracted from a gallery image is calculated. In similarity calculation, template deformation is performed once by M-step, where parameter $\eta$ is set as 0.01.

| | our algorithm | Eigenface | USC. | UMD. |
|---|---|---|---|---|
| algorithm | surface feature | PCA | Gabor wavelet | Fisherface |
| Identification rate | 95.4% | 79.7% | 95.0% | 96.2% |
| Equal error rate | 1.4% | 6.7% | 2.5% | 1.2% |

**Table 1**. Comparison identification and verification performance.

Figure 3 shows identification and verification performance of our algorithm. It is found that deformable template improve both identification and verification performances.

Since we use the same test set as the FERET competition [6], we can compare the performance of our algorithm with other algorithms that are tested in the competition. In table1, we show probability of identification and equal error rate of our algorithm and three algorithms reported in [6]. The equal error rate is the point where the probability of false acceptance is equal to the probability of false verification, and used to measure verification performance.

Our algorithm shows comparable performance with result of the University of Maryland (UMD) group[10], which is the best performance reported in [6]. Although our algorithm does not consider inter-personal variation unlike UMD, our algorithm shows good performance. Furthermore, our algorithm shows better performance than the University of Southern California (USC) group[8] and Eigenface algorithm. This result means that the surface features are better features than Gabor wavelet and Eigenface to discriminate individual faces.

## 5. CONCLUSION

We introduced the surface features for face recognition, which are extracted by segmenting a surface in XYI space. The Gaussian Mixture Model was utilized for the segmentation. We compared the face recognition performance of our algorithm with other algorithm. Our algorithm outperformed the algorithm based eigenface features and Gabor wavelet features for FERET fa and fb database.

One of the major concern is that the XYI surface would be hardly influenced by lighting condition. Some shading correction technique will be necessary to solve this problem. In future research, we will work on this issue.

## 6. REFERENCES


[1] J. Atick, P. Griffin and N. Redlich, "Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional Face Surfaces from Single Two-Dimensional Images," **Neural Computation**, vol. 8, pp. 1321–1340, 1996.

[2] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," **Journal of the Royal Statistical Society Series B**, vol. 39, no. 1, pp. 1–38, 1977.

[3] Ikeuchi and B. Horn, "Numerical Shape from Shading and Occluding Boundaries," **Artificial Intelligence**, vol. 17, pp. 141–184, 1981.

[4] D. Marr, **Vision**, Cambridge: The MIT Press, 1986.

[5] B. Moghaddam, C. Nastar and A. Pentland, "Bayesian Face Recognition using Deformable Intensity Surfaces," in **Proc. IEEE Conf. on Computer Vision and Pattern. Recognition**, 1996, pp. 638–645.

[6] S. Rizvi, P. J. Phillips and H. Moon, "A verification protocol and statistical performance analysis for face recognition algorithms," in **Proc. IEEE Conf. on Computer Vision and Pattern Recognition**, 1998, pp. 833–838.

[7] M. Turk and A. Pentland, "Eigenfaces for recognition," **J. of Cognitive Neuroscience**, vol. 3, no. 1, pp. 71–86, 1991.

[8] L. Wiskott, J. Fellous, N. Kruger and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," **IEEE Trans. on Pattern Analysis and Machine Intelligence**, vol. 19, no. 7, pp. 477–500, 1996.

[9] W. Zhao and R. Chellappa, "Illumination-insensitive Face Recognition using Symmetric Shape from-Shading," in **Proc. IEEE Conf. on Computer Vision and Pattern. Recognition**, 2000, pp. 1286–1293.

[10] W. Zhao, R. Chellappa and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in **Int. Conf. on Automatic Face and Gesture Recognition**, 1978, pp. 336–341.