Face Recognition Under Varying Illumination Based on MAP Estimation Incorporating Correlation Between Surface Points

Mihoko Shimano[1], Kenji Nagao[1],
Takahiro Okabe[2], Imari Sato[3], and Yoichi Sato[2]

[1] Panasonic Tokyo (Matsushita Electric Industrial Co., Ltd.,)
4–3–1 Tsunashima-higashi, Kohoku-ku, Yokohama City, Kanagawa 223–8639, Japan
shimano.mhk@jp.panasonic.com
[2] Institute of Industrial Science, The University of Tokyo,
4–6–1 Komaba, Meguro-ku Tokyo 153–8505, Japan
{takahiro, ysato}@iis.u-tokyo.ac.jp
[3] National Institute of Informatics,
2–1–2 Hitotsubashi, Chiyoda-ku Tokyo 101–8430, Japan
imarik@nii.ac.jp

Abstract. In this paper, we propose a new method for face recognition under varying illumination conditions using a single input image. Our method is based on a statistical shape-from-shading method which combines the strengths of the Lambertian model and statistical information obtained from a large number of images of different people under varying illumination. The main advantage of our method over the previous methods is that our method explicitly incorporates a correlation between surface points on a face in the MAP estimation of surface normals and albedos, so that a new image of the same face under novel illumination can be synthesized correctly even when the face is partially shadowed. Furthermore, our method introduces pixel grouping and reliability measure in the MAP estimation in order to reduce computational cost while maintaining accuracy. We demonstrate the effectiveness of our proposed method via experiments with real images.

1 Introduction

Face recognition has become one of the most actively studied areas in computer vision, and a number of methods have been proposed to recognize a person's face from input images [6, 21]. This is because face recognition technologies can be effectively used for a wide range of applications. One such application is the identification of a person with face recognition when only one image of the person is available beforehand, e.g., a picture on a driver's license or a passport. We believe this example is important because it is not always possible to provide a large number of training images for each person in real applications.

The appearance of a human face is highly dependent on many factors including the pose of a face, illumination conditions, and facial expression. In this

P.J. Narayanan et al. (Eds.): ACCV 2006, LNCS 3851, pp. 569–580, 2006.

work, we deal with the problem of recognizing a person's face under varying illumination conditions when only one training image is available for the person. Therefore, we do not consider appearance variations due to other factors such as poses or facial expression.

The task of face recognition becomes easier and robuster if a sufficient number of training images taken under different conditions are available for each person, so that we can model appearance variation for the person accurately. For instance, it is known that appearance variation of a human face due to illumination change is represented approximately with a low-dimensional linear subspace [4, 7, 2], and, as a result, existing methods for face recognition under varying illumination work fairly well as long as we have enough training images taken under different lighting conditions (for instance, [13, 11, 3, 7, 12, 17, 14, 15]).

On the other hand, face recognition under varying illumination becomes a challenging task when only one training image is available for each person. It is not trivial, and even may be impossible, to predict how the appearance of a person's face varies with different lighting conditions if we are given no information other than a single input image.

Several methods proposed recently [1, 18, 16, 5, 22, 10] can be used for solving this challenging problem. They are based on the idea of using a statistical model obtained from a set of images or laser-scanned images of different persons. With these methods, the shape and reflectance properties of a face are estimated from a single input image by using a statistical model of human faces. This is the key difference from conventional shape-from-shading techniques which estimate the shape of an object with more explicit assumptions such as the integrability constraint [8] and the assumption of face symmetry [20].

Although these methods have been used successfully to predict appearance variations in human faces for different lighting conditions, they share a common difficulty with the exception of Sim and Kanade's method [18]. That is, reflection components other than those represented by a simple reflection model such as the Lambertian model or the Phong model cannot be reproduced accurately with these methods. For instance, both Atick et al. [1] and Zhou et al.[22] assume that reflection on a human face can be represented with the Lambertian model. Blanz and Vetter [5] describe the shading observed on faces by using the Lambertian model and the Phong model. Matthews and Baker [10] represent appearance variations due to lighting as a linear combination of basis images. However, it is reported that reflections on human faces often deviate significantly from simple reflection models such as the Lambertian model and the Phong model [9].

Our method is most closely related to the method by Sim and Kanade [18], in that both methods estimate surface normals and albedos of surface points on a face from a given single input image via Maximum A Posteriori (MAP) estimation based on statistical information obtained from a set of images of different persons. Because reflection components other than those representable with the Lambertian model can also be estimated via MAP estimation, both methods have strengths in comparison with other statistical shape-from-shading methods in that subtle reflection components such as highlights and interreflections can be reproduced for novel lighting conditions.

The key difference between our method and Sim and Kanade's method is that our method explicitly incorporates the correlation between surface points on a face in MAP estimation of surface normals and albedos. This contributes to the distinct advantages of our method. For instance, a new image of the same face under novel illumination can be synthesized correctly with our method even when the face is partially shadowed, while Sim and Kanade's method fails to do so since each pixel is treated independently. In addition, we introduce the idea of pixel grouping and reliability measure in the MAP estimation in order to reduce computational cost while maintaining accuracy.

2 Our Proposed Method

Our method consists of three steps: i) *learning step* and ii) *modeling step* and iii) *rendering step*. In the *learning step*, our method computes statistics about human faces, i.e., the surface normal including albedo and an error term corresponding to reflection components other than the diffuse component, from a set of images of multiple people taken under varying illumination conditions. The set of images used for learning the statistics are referred to as bootstrap images. In the *modeling step*, a single image of a novel face, e.g., a picture on a driver's license or a passport, is used for predicting appearance variations for this person based on the statistics obtained in the learning step. We refer to this single image as a *training* image. Using this *training* image, our method first estimates the light source direction under which this *training* image was taken. Then, our method estimates the surface normal of the face via MAP estimation by using both the estimated light source direction and the learned statistics. Finally, in the *rendering step*, the error term for a novel illumination condition is computed by MAP estimation, and added to the diffuse reflection component to render the image of the face under the novel lighting. Images rendered this way are then used for face recognition under varying lightings. We will explain each of these steps in details.

2.1 Reflectance Equation

Our method assumes that a set of bootstrap images for the learning step and a single training image for the modeling step are taken under point light sources at infinity, that is, directional light sources. Then, the intensity i_p at the p-th pixel is represented as the sum of the diffuse component and the remaining component like in Sim and Kanade's method [18]

$$i_p = \boldsymbol{n}_p^T \boldsymbol{s} + e_p(\boldsymbol{s}), \tag{1}$$

where $\boldsymbol{n}_p = (n_{px}, n_{py}, n_{pz})^T$ is a product of the albedo and surface normal at the p-th pixel, and $\boldsymbol{s} = (s_x, s_y, s_z)^T$ is a product of the intensity and direction of a directional light source. The error term $e_p(\boldsymbol{s})$ describes reflection components other than diffuse reflection such as highlights, interreflections, and shadows.

To take into account the correlation between surface points explicitly, we represent the intensities of P pixels in each image as

$$\boldsymbol{i} = S^T \boldsymbol{b} + \boldsymbol{e}(\boldsymbol{s}), \tag{2}$$

$$\begin{pmatrix} i_1 \\ i_2 \\ \vdots \\ i_P \end{pmatrix} = \begin{pmatrix} \boldsymbol{s}^T & 0 & \dots & 0 \\ 0 & \boldsymbol{s}^T & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \boldsymbol{s}^T \end{pmatrix} \begin{pmatrix} \boldsymbol{n}_1 \\ \boldsymbol{n}_2 \\ \vdots \\ \boldsymbol{n}_P \end{pmatrix} + \begin{pmatrix} e_1(\boldsymbol{s}) \\ e_2(\boldsymbol{s}) \\ \vdots \\ e_P(\boldsymbol{s}) \end{pmatrix}.$$

Thus, $\boldsymbol{i}$, S, $\boldsymbol{b}$, and $\boldsymbol{e}$ are a P-dimensional vector, a $3P \times P$ matrix, a $3P$-dimensional vector, and a P-dimensional vector respectively.

2.2 Computing Statistics

Let us assume that a set of bootstrap images consists of images of L people taken under J known directional light sources $\boldsymbol{s}_j$ ($j = 1, 2, ..., J$). For the l-th person, a set of images $I^{(l)}$ taken under J light sources are represented by

$$I^{(l)} = B^{(l)T} S' + E^{(l)}. \tag{3}$$

Here, $I^{(l)} = (\boldsymbol{i}_1^{(l)}, \boldsymbol{i}_2^{(l)}, \dots, \boldsymbol{i}_J^{(l)})$, $B^{(l)} = (\boldsymbol{n}_1^{(l)}, \boldsymbol{n}_2^{(l)}, \dots, \boldsymbol{n}_P^{(l)})$, $S' = (\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_J)$, and $E^{(l)} = (\boldsymbol{e}^{(l)}(\boldsymbol{s}_1), \boldsymbol{e}^{(l)}(\boldsymbol{s}_2), \dots, \boldsymbol{e}^{(l)}(\boldsymbol{s}_J))$ respectively. We assume that the illumination intensity $|\boldsymbol{s}_j|$ of the bootstrap images are same (We describe $|\boldsymbol{s}_j|$ as 1 in the rest of this paper.).

Then, we consider $E^{(l)}$ as Gaussian noise[1], and compute the least-squares solution of $B^{(l)}$ as

$$B^{(l)} = (S' S'^T)^{-1} S' I^{(l)}, \tag{4}$$

and the residuals, that is, the error $E^{(l)}$ as

$$E^{(l)} = I^{(l)} - B^{(l)T} S'. \tag{5}$$

Finally, we compute the statistics of the surface normal and the error term from the estimated surface normals and error terms for all people in the bootstrap images. With regard to the surface normals $\boldsymbol{b}$, all matrices $B^{(l)}$ for L people are converted into $3P$-dimensional vectors in a raster-scan manner, and then the mean vector $\mu_{\boldsymbol{b}}$ ($3P$-dimensional vector) and the covariance matrix $C_{\boldsymbol{b}}$ ($3P \times 3P$ matrix) are computed. For the statistics of the error term $\boldsymbol{e}(\boldsymbol{s})$, the mean vector $\mu_{\boldsymbol{e}}(\boldsymbol{s}_j)$ is computed from L error vectors $\boldsymbol{e}^{(l)}(\boldsymbol{s}_j)$ for each light source $\boldsymbol{s}_j$. The $PJ \times PJ$ covariance matrix $C_{\boldsymbol{e}}$ is computed from L error matrices $E^{(l)}$ in the same way as $C_{\boldsymbol{b}}$ is computed from $B^{(l)}$.Note that the difference between our

[1] In order to carefully recover $B^{(l)}$, we removed outliers such as highlights and shadows contained in bootstrap images.

proposed method and the previous method is that we incorporate the correlation between surface points rather than treating each point independently.

2.3 Modeling from a Single Image

The modelling step consists of two sub-steps; estimation of the light source direction under which a single training image of a subject was taken, and the estimation of the surface normals of the subject. We explain these sub-steps in detail.

Estimating Illumination

We estimate the illumination intensity and direction under which a single training image $\boldsymbol{i}$ was taken. Our method assumes that the subject has the average face shape and albedos which are represented by the Lambertian model, and computes the least-squares solution of the illumination[2]. We extended the illumination estimation method proposed by Sim and Kanade [18] so that we are able to take variations in intensity of illumination into consideration. Although it is assumed that the illumination intensities of the training image and the bootstrap images are same in Sim and Kanade's method [18], it is often the case that intensity of illumination changes between the bootstrap images and the training image.

The illumination intensity and direction of a single training image is estimated by using the average B_{avr} of the computed matrices $B^{(l)}$ for all people in the bootstrap images $\boldsymbol{i}_j^{(l)}$

$$\boldsymbol{s} = B_{avr}^{T+}\boldsymbol{i} = (B_{avr}B_{avr}^T)^{-1}B_{avr}\boldsymbol{i}. \tag{6}$$

Then the ratio α of the estimated illumination intensity to that of the bootstrap images is computed as $\alpha = |\boldsymbol{s}|$.

Estimating Surface Normals and Albedos

Taking into account the correlation between surface points, we recover surface normals and albedos by MAP estimate as $\boldsymbol{b}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{b}} P(\boldsymbol{b}|\boldsymbol{i})$. According to the Bayes' rule,

$$\boldsymbol{b}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{b}} P(\boldsymbol{i}|\boldsymbol{b})P(\boldsymbol{b}). \tag{7}$$

Because we assume that the probability density functions (PDFs) of $\boldsymbol{b}$ and $\boldsymbol{e}$ are Gaussian distributions, $P(\boldsymbol{b})$ is described by $\mu_{\boldsymbol{b}}$ and $C_{\boldsymbol{b}}$, and $P(\boldsymbol{i}|\boldsymbol{b})$ is described by the mean $S^T\boldsymbol{b} + \mu_{\boldsymbol{e}}(\boldsymbol{s})$ and the covariance $\Sigma_{\boldsymbol{e}}$. Here, we calculate the mean $\mu_{\boldsymbol{e}}(\boldsymbol{s})$ based on kernel regression method by using the known illumination $\boldsymbol{s}_j$, and the elements of $\Sigma_{\boldsymbol{e}}$ are also interpolated from the computed $C_{\boldsymbol{e}}$ as

$$\mu_{\boldsymbol{e}}(\boldsymbol{s}) = \alpha\beta\frac{\sum_{j=1}^{J} w_j \mu_{\boldsymbol{e}}(\boldsymbol{s}_j)}{\sum_{j=1}^{J} w_j}, \quad \sigma_{\boldsymbol{e}}(s)^2 = \alpha^2\beta^2\frac{\sum_{j=1}^{J} w_j \sigma_{\boldsymbol{e}}(s_j)^2}{\sum_{j=1}^{J} w_j}, \tag{8}$$

[2] In order to carefully recover the illumination, we removed outliers such as highlights and shadows contained in a training image.

where $w_j = \exp(-(D(\boldsymbol{s}, \boldsymbol{s}_j)/\sigma_j)^2/2)$, $D(\boldsymbol{s}, \boldsymbol{s}_j) = |\boldsymbol{s}/\alpha - \boldsymbol{s}_j|$, and β is a coefficient which sets the norm of the estimated illumination vector to 1. The coefficient β is defined by $\boldsymbol{s}/\alpha = \beta(\sum_{j=1}^{J} w_j \boldsymbol{s}_j)/\sum_{j=1}^{J} w_j$. Substituting the above vectors and matrices into equation (7), $\boldsymbol{b}_{\mathrm{MAP}}$ is given by

$$\boldsymbol{b}_{\mathrm{MAP}} = (S\Sigma_{\boldsymbol{e}}^{-1}S^T + C_{\boldsymbol{b}}^{-1})^{-1}(S\Sigma_{\boldsymbol{e}}^{-1}(\boldsymbol{i} - \mu_{\boldsymbol{e}}) + C_{\boldsymbol{b}}^{-1}\mu_{\boldsymbol{b}}). \tag{9}$$

2.4 Rendering for Novel Lightings

In order to synthesize an image under a novel illumination condition, we estimate the error terms under the illumination condition, considering both the correlation between surface points and between illumination directions. Because we assume a jointly Gaussian distribution for the PDF of the error terms, by using the actual error $\boldsymbol{e} = \boldsymbol{i} - S^T\boldsymbol{b}_{\mathrm{MAP}}$, the MAP estimate of the error terms under a novel illumination condition S_{new} is given by

$$\boldsymbol{e}_{\mathrm{MAP}} = \mu_{\boldsymbol{e}_{\mathrm{new}}} + R^T\Sigma_{\boldsymbol{e}}^{-1}(\boldsymbol{e} - \mu_{\boldsymbol{e}}), \tag{10}$$

where $\mu_{\boldsymbol{e}_{\mathrm{new}}}$ and R is the mean error under the novel lighting and the covariance of the error terms between the lighting of the training image and the novel lighting respectively. These quantities are also interpolated from $\mu_{\boldsymbol{e}}$ and $\Sigma_{\boldsymbol{e}}$.

Thus, a new image under a novel lighting condition is synthesized as

$$\boldsymbol{i}_{\mathrm{new}} = S_{\mathrm{new}}^T\boldsymbol{b}_{\mathrm{MAP}} + \boldsymbol{e}_{\mathrm{MAP}} \tag{11}$$

by using the estimated surface normals and error terms.

3 Reduction of Computational Cost

In addition to the correlation between surface points, we introduce two important improvements for reducing computational cost of our method while maintaining accuracy.

3.1 Grouping Pixels

The computational cost of our method increases approximately at $O(P^3)$ for the number of pixels since our method incorporates correlation between pixels rather than treating them independently. Thus, in order to reduce the calculation cost and make our method more tractable, we divide an image into subareas and consider the correlation between pixels in each area.

It is worth noting that the grouping used for estimating surface normals and that used for estimating error terms are not necessarily the same. In our experiments, surface normals and albedos were estimated by incorporating the correlation between surface points as we described above. On the other hand, the error terms were treated independently with respect to surface points, and only the correlation between illumination conditions is taken into account, because the $PJ \times PJ$ covariance matrix requires a large amount of computational cost.

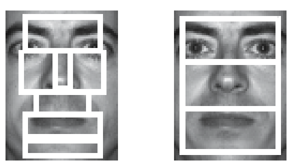
3.2 Considering Reliability

In order to correctly recover surface normals, we further introduce a reliability measure representing how reliable each pixel value. Based on the reliability measure, the surface normals at shadowed regions or at pixels with noise are estimated from reliable pixels rather than other unreliable pixels in the training image. In practice, the value of the error variance $\Sigma_{\boldsymbol{e}}$ are used as this reliable measure; if the variance of the error term at a pixel is high, the contribution of the pixel is decreased for calculation of the correlation $C_{\boldsymbol{b}}$.

The threshold for this reliable pixel selection can be determined by preliminary experiments. In addition, this reliable pixel selection results in reducing the computational cost of our method.

4 Experiments

In order to evaluate the performance of our proposed method, we conducted experiments on face recognition. We used the same statistical model for all of our experiments, and the model was obtained by using the Yale database B [3] which contains 640 images of 10 individuals (each person has 64 different images) under various lighting conditions per pose. Among 640 images in a frontal pose, we omitted 24 images for each of 10 people which contain an excessive amount of shadows due to extreme lighting conditions and used the other 40 images for each person for computing statistics about human faces (400 images in total). Each image was manually cropped and resized to 40×30 pixels with aligned eye positions. We used two different segmentations for grouping pixels in our experiments on face recognition as shown in Figure 1.



(a) (b)

Fig. 1. Segmentations used for grouping pixels in our experiments: (a) 6 regions and (b) 3 regions. Note that left and right cheeks compose one region although they are not adjacent in (a).

4.1 Image Synthesis

Figure 2 shows one example of synthesized images by using our proposed method. The training image of a face illuminated from right (Figure 2 (a)) was used for synthesizing images under different lightings. As a ground truth, a real image of the same face taken under frontal illumination is shown in Figure 2 (d). Figure 2 (b) shows the image synthesized by our proposed method for frontal illumination with the grouping shown in Figure 1 (b). We can see both the diffuse reflection

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

Fig. 2. Example of synthesized images with our method: (a) training image of a face illuminated from right, (b) synthesized image taken under frontal illumination with our method incorporating correlation between surface pixels, (c) synthesized image taken under frontal illumination without correlation , (d) real image taken under frontal illumination.

component and highlights are correctly synthesized even at surface points in shadows, e.g., the shadow cast by the nose and the attached shadow on the cheek. In contrast, it can be clearly seen that the image synthesized by using Sim and Kanade's method in Figure 2 (c) has problems for dealing with surface points in shadows.

4.2 Face Recognition

Two databases of face images taken under different illumination were used for our tests: our own database which contains frontal face images of 12 individuals illuminated from 11 different lighting directions and CMU-PIE database [19] which contains frontal images of 68 individuals illuminated from 21 different lighting directions.

All of the tests were conducted as follows. First, one image for each individual was used as a training image, and 40 images under different illumination (5 images only in the first experiment) were synthesized by using the training image and the statistical model learned from the Yale database B. Those 41 images were then used to generate the subspace for each individual. The rest of the images in the database were used as a test image and classified by searching for the subspace with the closest Euclidean distance to the test image.

In the first experiment, we compared the performance of our method with the most closely related method by Sim and Kanade [18] by using our own database. An image taken under frontal lighting was used as the single training image for each person for generating the person's subspace by Sim and Kanade's method and our proposed method. Table 1 shows recognition rates achieved by these two methods, and it shows significant improvement in recognition accuracy by incorporating correlations between surface points in MAP estimation as in our method.

In the second experiment, we used the CMU-PIE database, and the image of each person illuminated from the side was used as a training image. The result is shown in Table 2. As in the first experiment, recognition accuracy was significantly improved from 68% to 86% by incorporating correlations between surface points. Our method works well because both the diffuse reflection component

Table 1. Performance comparison of Sim and Kanade's method and our proposed method by using our face image database of 12 individuals

Methods	Recognition rate [%]
Sim and Kanade's method (without correlation)	88
Our method (with correlation in a group)	94

Table 2. Performance comparison of Sim and Kande's method and our proposed method by using CMU-PIE database

Methods	Recognition rate [%]
Sim and Kanade's method (without correlation)	68
Our method (with correlation in a group)	86

Table 3. Performance improvement by grouping pixels (3 areas) and the use of reliability measure in our method

Methods	Recognition rate [%]
Sim and Kanade's method (without correlation)	74
Our method (with correlation in a group without reliability)	81
Our method (with correlation in a group with reliability)	83

and highlights are correctly synthesized even at surface points in shadows as shown in Figure 2.

In the third experiment, we evaluated the effectiveness of pixel grouping and the reliability measure introduced in Section 3. The result is shown in Table 3. This experiment was done by using our database as in the first experiment except that face images illuminated from the side were used as a training image this time. First, we can see that the recognition rate was improved by almost 10% from 74% (*without correlation*) to 83% by incorporating correlations in MAP estimation. This also demonstrates that face recognition can be performed efficiently by using pixel grouping together with the reliability measure.

In the fourth experiment, we compared our method with Zhou et al.'s method [22] which is one of the most recently proposed methods for the same problem setting, i.e., face recognition under varying lightings by using a single training image. In order to compare the performance of our method with that of Zhou et al.'s method, we conducted experiments under the same condition as reported in [22]. The results of Zhou et al.'s method were taken from [22]. As we can see in Table 4, our method outperformed Zhou et al.'s method significantly. The recognition rate of our method (100%) is higher than that of Zhou et al's method (59%), when we used "f13" as a training set and "f16" as a test set, that is, both the training set and the test set contain face images illuminated from the same side. When we used "f08" under frontal illumination as a training set and "f15"

Table 4. Performance comparison of Zhou's method and our method by using CMU-PIE database

Methods	Recognition rate [%]	
	f13/f16(training/test)	f08/f15
Zhou et al.'s method	59	33
Our method	100	99

illuminated sideways as a test set, our method achieved high recognition rate (99%) in contrast with that of Zhou et al.'s method (33%).

The reason for our method's outperforming Zhou et al.'s method is attributed to the following two points. First, our method statistically models reflection components other than the diffuse component such as specular highlights and shadows, while Zhou et al.'s method assumes the Lambertian model. Second, our method takes into account correlations among surface points on a face so that a new image of the same face under novel illumination can be synthesized even when a single training image is partially shadowed.

5 Conclusions

In this work, we proposed a new method based on statistical shape from shading for face recognition under varying lighting conditions using a single training image for each person. Our method first learns a statistical model about human faces by using a set of training images of multiple people taken under varying illumination conditions. Then, the shape and albedo of a novel face are estimated via MAP estimation using the obtained statistical model and a single training image of the novel face. Finally, images of the face under novel lighting conditions are generated by using the obtained shape and albedo together with the error term estimated via MAP estimation.

The main advantage of our method over the previous methods is that our method explicitly incorporates a correlation between surface points on a face in the MAP estimation of surface normals and albedos, so that a new image of the same face under novel illumination can be synthesized correctly even when the face is partially shadowed. Furthermore, our method introduces pixel grouping and reliability measure in the MAP estimation in order to reduce computational cost while maintaining accuracy. The performance of our proposed method was demonstrated via experiments on face recognition by using real images.

We manually specified areas grouping pixels in this work, and automatic segmentation remains as a part of our future work. And we will further investigate the treatment of the error term. For instance, one possibility is to decompose the error term into different components such as specular highlights and shadows, and then treat them independently. We are also interested in extending our method for modeling the variations due to other factors such as poses.

References

1. J. J. Atick, P. A. Griffin, and A. N. Redlich, "Statistical approach to shape from shading: reconstruction of three-dimensional face surfaces from single two-dimensional images," *Neural Computation*, Vol.8, No.6, pp.1321–1340, 1996.
2. R. Basri, D. Roth, and D. Jacobs, "Clustering appearances of 3D objects," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.414–420, 1998.
3. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.711–720, 1997.
4. P. N. Belhumeur, and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.270–277, 1996.
5. V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.9, pp.1063–1074, 2003.
6. R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces, A Survey," *Proc. of the IEEE*, Vol.83, pp.705–740, 1995.
7. A. S. Georghiades, D. J. Kriegman, and P. N. Belhumeur, "Illumination cones for recognition under variable lighting: faces," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.52–58, 1998.
8. B. K. P. Horn, "Height and gradient from shading," *Int. J. Computer Vision*, Vol.5, pp.37–75, 1990.
9. S. R. Marschner, S. H. Westin, E. P. F. Lafortune, K. E. Torrance, and D. P. Greenberg, "Image-based BRDF Measurement Including Human Skin," *Proc. 10th Eurographics Workshop on Rendering*, pp.139–152, 1999.
10. I. Matthews and S. Baker, "Active appearance models revisited," *Int. J. Computer Vision*, Vol.60, No.2, pp.135–164, 2004.
11. H. Murase and S. K. Nayar, "Visual learning and Recognition of 3-D Objects from Appearance," *Int. J. Computer Vision*, Vol. 14, pp. 5–24, 1995.
12. K. Nagao, "Face recognition by distribution specific feature extraction," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.I-1278–1285, 2000.
13. E. Oja, "Subspace Methods for Pattern Recognition," *Research Studies Press Ltd.*, 1980.
14. T. Okabe and Y. Sato, "Object Recognition Based on Photometric Alignment Using RANSAC" *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.I–221–228, 2003.
15. T. Okabe and Y. Sato, "Support Vector Machines for Object Recognition under Varying Illumination Conditions" *In Proc. Asian Conf. Computer Vision*, pp.724–729, 2004.
16. T. Shakunaga and K. Shigenari, "Decomposed Eigenface for Face Recognition under Various Lighting Conditions," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Vol.1, pp.864–871, 2001.
17. M. Shimano, K. Nagao, "Simultaneous Optimization of Class Configuration and Feature Space for Object Recognition," *Proc. Int. Conf. Pattern Recognition*, No.2, pp.7–10, 2004.
18. T. Sim and T. Kanade, "Combining Models and Exemplars for Face Recognition: An Illuminating Example," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition Workshop on Models versus Exemplars in Computer Vision*, 2001.

19. T. Sim, S. Baker, and M. Bsat, “The CMU Pose, Illumination, and Expression (PIE) Database,” *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp.53–58, 2002.
20. W. Zhao and R. Chellappa, “Illumination-insensive face recognition using symmetric shape-from-shading,” *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Vol.1, pp.286–293, 2000.
21. W. Zhao, R. Chellappa, J. Phillips and A. Rosenfeld, “Face Recognition in Still and Video Images: A Literature Surrey,” *ACM Computing Surveys*, Vol. 35, pp. 399–458, Dec. 2003.
22. S. K. Zhou, R. Chellappa, and D. W. Jacobs, ”Characterization of Human Faces under Illumination Variations Using Rank, Integrability, and Symmetry Constraints,” *Proc. European Conf. Computer Vision (ECCV2004)*, Vol.1, pp.588–601, 2004.