Decomposed eigenface for face recognition under various lighting conditions

Takeshi Shakunaga
Okayama University

Kazuma Shigenari
Okayama University

Decomposed Eigenface for Face Recognition under Various Lighting Conditions

Takeshi Shakunaga Kazuma Shigenari
Department of Information Technology, Okayama University
Okayama-shi 700-8530, Japan
{shaku,shige}@chino.it.okayama-u.ac.jp

Abstract

Face recognition under various lighting conditions is discussed to cover cases when too few images are available for registration. This paper proposes decomposition of an eigenface into two orthogonal eigenspaces for realizing robust face recognition under such conditions. The decomposed eigenfaces consisting of two eigenspaces are constructed for each person even if only one image is available. A universal eigenspace called the canonical space (CS) plays an important role in creating the eigenspaces by way of decomposition, where CS is constructed a priori by principal component analysis (PCA) over face images of many people under many lighting conditions. In the registration stage, an input face image is decomposed to a projection image in CS and the residual of the projection. Then two eigenspaces are created independently in CS and in the orthogonal complement $CS^{\perp}$. Some refinements of the two eigenspaces are also discussed. By combining the two eigenspaces, we can easily realize face identification that is robust to illumination change, even when too few images are registered. Through experiments, we show the effectiveness of the decomposed eigenfaces as compared with conventional methods.

1. Introduction

A human face changes in appearance with lighting conditions and difficulty is encountered in controlling lighting conditions in natural environments where face images are taken. These facts suggest that, in order to realize robust face recognition, we should construct a face recognition algorithm that is insensitive to lighting conditions. Meanwhile, appearance-based face recognition can be resolved into the eigenface method [10], which is often identical to the subspace method [6, 3]. Eigenfaces are widely used both for the detection of (unknown) faces in an image and for person identification. When intended for detection of faces in an image, eigenfaces are constructed from many people, and when intended for person identification, each eigenface should be constructed from face images of the individual. In the present paper we focus on the second purpose, and assume that an eigenface(EF) always means an eigenspace constructed from face images of the individual for person identification. If many face images can be collected in the registration stage, the EF can be constructed by PCA. Recently, Yuille et al. [12] have shown that an EF can also be constructed by photometric Singular Value Decomposition when many images of an individual are taken under various illumination conditions. However, an EF for an individual cannot be stably constructed when too few sample images are available or when the lighting conditions in the sample images are very similar. In these situations, some refinements are required for the eigenface method to realize illumination-insensitive identification. This paper analyzes the eigenface approach and proposes a decomposition of EF for the foregoing purpose.

2. Eigenface in Normalized Image Space

2.1. Normalized image space (NIS)

Let an N-dimensional vector $\mathbf{X}$ denote an original image composed of N pixels, and let $\mathbf{1}$ denote an N-dimensional vector in which each element is 1. The normalized image $\mathbf{x}$ of an original image $\mathbf{X}$ is defined as $\mathbf{x} = \mathbf{X}/(\mathbf{X}^T\mathbf{1})$. After normalization, $\mathbf{x}$ is normalized in the sense that $\mathbf{x}^T\mathbf{1} = 1$. Any nonzero image $\mathbf{X}(\neq \mathbf{0})$ can be mapped to a point within the Normalized Image Space (NIS). This normalization is very natural, in that total energy is normalized in each image. NIS is closed against any averaging operation and provides an image representation that is invariant to changes in light intensity if the original image includes neither saturated points nor shadow regions.

0-7695-1272-0/01 $10.00 © 2001 IEEE

2.2. Eigenface construction in NIS

In previous studies, eigenspaces have often been constructed around the origin of an image space [12] when photometric analysis is essential and inevitable. In contrast, eigenspaces are often constructed around mean images [6, 10, 5, 1, 8] when the results are to be used for recognition. This paper follows the latter convention, because it concerns appearance-based face recognition.

The eigenspace should be two-dimensional in NIS when the subject has only Lambertian surfaces without shadows. Although human faces are not purely Lambertian, they are mostly Lambertian. Hence, face recognition can be resolved into low-dimensional eigenspace construction around the mean image in NIS.

Hereinafter, we assume that any original image has already been normalized. Let $\mathbf{x}_{pl}$ denote a face of the p-th person under the l-th lighting condition. When many images of the p-th person are taken, an eigenface (EF) is constructed in the following manner. The mean vector and the covariance matrix are first calculated for person p in

$$\overline{\mathbf{x}}_p = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_{pl}$$

$$\Sigma_p = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{x}_{pl}-\overline{\mathbf{x}}_p)(\mathbf{x}_{pl}-\overline{\mathbf{x}}_p)^T$$

where L is the number of lighting conditions. Let Λ_p denote a diagonal matrix in which the diagonal terms are eigenvalues of Σ_p in descending order, and let Φ_p denote a matrix in which the i-th column is the i-th eigenvector of Σ_p. Then PCA implies $\Lambda_p = \Phi_p^T\Sigma_p\Phi_p$. Using a submatrix Φ_{pm} of Φ_p, which consists of the most principal (first) m eigenvectors, the projection of $\mathbf{x}$ to the m-dimensional p-th EF is given by

$$\tilde{\mathbf{x}}_p = {\Phi_{pm}}^T(\mathbf{x}-\overline{\mathbf{x}}_p).$$

In our problems m is a small number, because human faces are mostly Lambertian.

3. Decomposed Eigenface

3.1. Canonical space

In this paper, we assume that a face space is composed of frontal faces. We also assume that the face direction is fixed and good segmentation is readily accomplished by structure analysis [7] of each normalized image, as shown in Fig. 5.

As is well known, PCA reduces the dimension of the face space with little loss of representability [3, 10].

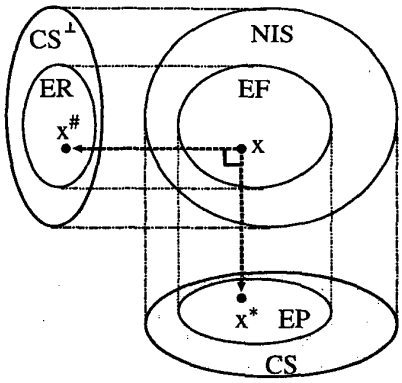


Figure 1. Decomposition of EF to EP and ER using CS and CS$^\perp$.

In our experiments, the face eigenspace is constructed from a face image set called the canonical set, which comprises face images of 50 persons taken under 24 lighting conditions for each person. Let us call the constructed eigenspace the canonical (eigen)space, or CS for short. The dimension of CS is set to 45. Note that CS is distinguished from EF in this paper (Fig. 1).

The mean vector and the covariance matrix are written as

$$\overline{\mathbf{x}} = \frac{1}{PL}\sum_{p=1}^{P}\sum_{l=1}^{L}\mathbf{x}_{pl}$$

$$\Sigma = \frac{1}{PL}\sum_{p=1}^{P}\sum_{l=1}^{L}(\mathbf{x}_{pl}-\overline{\mathbf{x}})(\mathbf{x}_{pl}-\overline{\mathbf{x}})^T$$

where P and L are the number of persons and the number of lighting conditions, respectively.

Let Λ denote a diagonal matrix in which the diagonal terms are eigenvalues of Σ in descending order, and let Φ denote a matrix in which the i-th column is the i-th eigenvector of Σ. Then PCA implies $\Lambda = \Phi^T\Sigma\Phi$. Using a submatrix Φ_n of Φ, which consists of the most principal n eigenvectors, the projection of $\mathbf{x}$ to the n-dimensional CS is represented by

$$\mathbf{x}^* = {\Phi_n}^T(\mathbf{x}-\overline{\mathbf{x}}).$$

The residual $\mathbf{x}^\#$ is then expressed as

$$\mathbf{x}^\# = \mathbf{x}-\overline{\mathbf{x}}-\Phi_n\mathbf{x}^*.$$

Let CS$^\perp$ denote the orthogonal complement of CS. Then an image $\mathbf{x}$ in NIS can be orthogonally decomposed into a canonical component $\mathbf{x}^*$ in CS and a residual component $\mathbf{x}^\#$ in CS$^\perp$. These are orthogonal and complementary, and can be effectively processed as described in the following sections.

3.2. Eigen-projection and eigen-residual

As formulated in 2.2, EF is an eigenspace constructed in NIS. We also show that any image $\mathbf{x}$ can be decomposed to two orthogonal components $\mathbf{x}^*$ and $\mathbf{x}^\#$ by projecting $\mathbf{x}$ to CS. The orthogonal components enable us to decompose EF. That is, as shown in Fig. 1, two eigenspaces can be constructed independently in CS and in $CS^\perp$. The first eigenspace, called an eigen-projection (EP), is constructed from the canonical components in CS. The second eigenspace, called an eigen-residual (ER), is constructed from the residual components in $CS^\perp$. EP and ER are constructed in a manner similar to that employed for EF construction in NIS as described in 2.2.

For the EP construction, the mean vector and the covariance matrix are given by

$$\overline{\mathbf{x}_p^*} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_{pl}^* \tag{1}$$

$$\Sigma_p^* = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{x}_{pl}^* - \overline{\mathbf{x}_p^*})(\mathbf{x}_{pl}^* - \overline{\mathbf{x}_p^*})^T. \tag{2}$$

Let Φ_p^* and Λ_p^* denote the eigenvectors and the diagonal matrix, respectively. Then PCA implies $\Lambda_p^* = \Phi_p^{*T}\Sigma_p^*\Phi_p^*$. Using a submatrix Φ_{pm}^* of Φ_p^*, which consists of the first m eigenvectors, the projection of $\mathbf{x}^*$ to the p-th EP is given by

$$\tilde{\mathbf{x}}_p^* = {\Phi_{pm}^*}^T(\mathbf{x}^* - \overline{\mathbf{x}_p^*}). \tag{3}$$

ER can also be constructed in the same way as described for EP. We can define $\overline{\mathbf{x}_p^\#}$, $\Sigma_p^\#$, $\Phi_p^\#$, $\Lambda_p^\#$, and $\Phi_{pm}^\#$ in the same way. Consequently, the projection of $\mathbf{x}^\#$ to the p-th ER is given by

$$\tilde{\mathbf{x}}_p^\# = {\Phi_{pm}^\#}^T(\mathbf{x}^\# - \overline{\mathbf{x}_p^\#}).$$

3.3. Characteristics of EP and ER

The decomposition of EF facilitates the refinement of EP in CS and the refinement of ER in $CS^\perp$. Since CS is constructed from numerous face images which cover many persons and many lighting conditions, $\mathbf{x}^*$ contains both geometric and photometric properties of the face. Although the geometric properties are approximately encoded in CS from one image, the photometric properties cannot be encoded from one image, because they are also dependent on the individual 3d shape; additional information is necessary for attaining full coding of the photometric properties. In order to complete the coding in CS, we introduce a concept of virtual eigenspace in 4.1 and a lighting transformation in 4.2.

Meanwhile, $\mathbf{x}^\#$ is a residual of the projection. Although the residual seems to contain a lot of noise, it also contains a lot of personal properties which cannot be encoded by the projection to CS. Since CS is constructed from a finite number of people, $\mathbf{x}^*$ often loses personal properties which characterize the individual. The personal properties are included in $\mathbf{x}^\#$ and often work well for person identification. For example, if $\mathbf{x}$ is an image of a person wearing characteristic eyeglasses, $\mathbf{x}^\#$ should contain a lot of information about the glasses (Fig. 4). However, in order to utilize the personal properties, we must suppress noise. In 5, we discuss a noise suppression method whereby the personal properties are not suppressed.

4. Refinement of EP construction

4.1. Concept of virtual eigenspace

Let us introduce the concept of virtual eigenspace. Virtual eigenspace is defined as a virtualized concept of eigenspace and should satisfy the following two requirements:

(a) Virtual eigenspace could be constructed from a single image, whereas the real eigenspace cannot be constructed directly from a single image.

(b) Virtual eigenspace should converge to the real eigenspace when additional images are taken.

We would like to construct a virtual eigenspace in CS in order to cover many lighting conditions, even if too few images are registered for each person. Let us call the virtual eigenspace a virtual eigen-projection (VEP), in order to distinguish it from the real EP. To satisfy (a), we synthesize a set of images in CS under many lighting conditions from a single image, and construct an eigenspace from the set as discussed in 4.3(1). In this process, image synthesis is based on the lighting estimation and the lighting transformation that are discussed in 4.2.

VEPs are constructed by means of the average distribution of a set of persons under a variety of lighting conditions. In contrast, real EPs have individual distributions affected by both the geometric and photometric properties of individual faces.

It seems difficult to completely satisfy (b) in an open set of face images. However, we can define VEP as shown in 4.3(2) so as to satisfy (b) in a finite set of images.

4.2. Lighting transformation in CS

Let us introduce a mapping function in CS which approximately transforms a face taken under one lighting condition to a face under another lighting condition. In the strict meaning, the mapping function depends on the person as well as on lighting conditions, because both the geometric and photometric properties are dependent on the person [2]. Although a lighting transformation can be effected if many samples are available for each person [12], our objective is different. Let us effect a linear transformation in CS, which approximates the transformation in a least-squared sense over the canonical set used in 3.1. Let $F(l_1, l_2)$ denote a face mapping from a lighting condition l_1 to another lighting condition l_2. That is,

$$\mathbf{x}^*_{pl_2} = F(l_1, l_2)\mathbf{x}^*_{pl_1}.$$

A matrix $F(l_1, l_2)$ can be estimated by minimizing a squared sum of errors,

$$\sum_{p=1}^{P}(\mathbf{x}^*_{pl_2} - F(l_1, l_2)\mathbf{x}^*_{pl_1})^T(\mathbf{x}^*_{pl_2} - F(l_1, l_2)\mathbf{x}^*_{pl_1}).$$

In order to stabilize the transformation, only a square of $n'(< n)$ terms in the upper left of F are estimated with the other ij-term fixed by $F_{ij} = \delta_{ij}$, where δ_{ij} indicates Kronecker's delta function. Once the matrix $F(l_1, l_2)$ is estimated, a face image in l_1 can be converted to an image in l_2 by means of linear mapping.

We also have to estimate lighting condition from a given face image. We suppose that a face is well segmented a priori, using structure analysis [7] , and a dictionary image set can be used again for the estimation. In this case, a mean vector

$$\overline{\mathbf{x}^*_l} = \frac{1}{P}\sum_{p=1}^{P}\mathbf{x}^*_{pl}$$

is regarded as a registered image for the lighting condition l. Then the lighting estimation is accomplished by the nearest neighbor discrimination in CS. Let $\lambda(\mathbf{x}^*)$ denote the lighting estimation function, which selects a lighting condition from $\{\overline{\mathbf{x}^*_l}\}$.

Thus, a simple mapping can be effected in CS using F and $\lambda(\mathbf{x}^*)$. Suppose that a face image $\mathbf{x}^*$ is transformed to an image under a lighting condition l. This operation is realized as shown in Fig. 2 by

$$\mathbf{f}^*(\mathbf{x}^*, l) = F(\lambda(\mathbf{x}^*), l)\mathbf{x}^*.$$

4.3. Virtual eigen-projection construction

(1) From one image

We shall show how to construct an eigen-projection

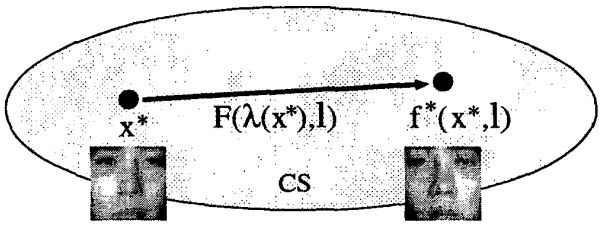


Figure 2. Lighting transformation.

when only one image of the p-th person is taken. VEP can be directly constructed in CS over the synthesized images by an algorithm similar to that shown in 2.2. First, the mean vector and the covariance matrix are given by

$$\overline{\mathbf{x}^*_p} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{f}^*(\mathbf{x}^*_p, l) \tag{4}$$

$$\Sigma^*_p = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{f}^*(\mathbf{x}^*_p, l) - \overline{\mathbf{x}^*_p})(\mathbf{f}^*(\mathbf{x}^*_p, l) - \overline{\mathbf{x}^*_p})^T. \tag{5}$$

Let Φ^*_p and Λ^*_p denote the eigenvectors and the diagonal matrix, respectively. Then PCA implies $\Lambda^*_p = \Phi^{*T}_p\Sigma^*_p\Phi^*_p$. Using a submatrix Φ^*_{pm} of Φ^*_p, which consists of the first m eigenvectors, the projection of $\mathbf{x}^*$ to the p-th VEP is given by Eq. (3).

(2) From K images

Let us discuss how to construct an eigen-projection when K images of the p-th person are taken. Two notations should be defined for the discussion; let $\mathbf{x}_k$ denote the k-th input image, and

$$\kappa(K, l) = \arg\max_{1\le k\le K} C(\mathbf{f}^*(\mathbf{x}^*_k, l), \mathbf{x}^*_k),$$

where C represents a normalized correlation,

$$C(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}/(\mathbf{x}^T\mathbf{x}\mathbf{y}^T\mathbf{y})^{1/2}. \tag{6}$$

Because the linear transformation is only an approximation of the real transformation, the better course is to choose the nearest input image for each virtual lighting condition. κ is used for selection of the nearest neighbor of the lighting condition, in order to compensate for the weak point of the linear transformation.

In these notations, a synthesized canonical component is calculated for the l-th lighting condition as $\mathbf{f}^*(\mathbf{x}^*_{\kappa(K,l)}, l)$. This is one of the natural generalizations of $\mathbf{f}^*(\mathbf{x}^*, l)$. Consequently, we can construct $\overline{\mathbf{x}^*_p}$ and Σ^*_p for K images as follows:

$$\overline{\mathbf{x}^*_p} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{f}^*(\mathbf{x}^*_{\kappa(K,l)}, l) \tag{7}$$

$$\Sigma_p^* = \frac{1}{L}\sum_{l=1}^{L}(\mathbf{f}^*(\mathbf{x}^*_{\kappa(K,l)}, l) - \overline{\mathbf{x}_p^*})(\mathbf{f}^*(\mathbf{x}^*_{\kappa(K,l)}, l) - \overline{\mathbf{x}_p^*})^T. \quad (8)$$

The eigen-projection is constructed directly by PCA of the updated Σ_p^*. Note that Eqs. (7) and (8) are equivalent to Eqs. (4) and (5) when $K = 1$. Also note that Eqs. (7) and (8) become equivalent to Eqs. (1) and (2) when $K = L$ and all the test images are registered for the person p.

4.4. Examples and convergence

Figure 3 shows an example of successive VEP construction fora particular person. In each row, the leftmost image shows a mean VEP image for the person. The remaining four images show the four most principal components (PC1-4).

In Fig. 3 the first 2 rows, (a) and (b), show VEPs constructed from a single image of the person. Rows (a) and (b) are different, because the input images were taken under different lighting conditions. Rows (c) and (d) show VEPs constructed from two and five input images, respectively, both including images used in (a) and (b). VEP appears to converge to the real EP, as shown in (e), which is constructed from 24 images. VEP is authorized in a finite set of images as follows: We have effected the normalized correlations of PCi(i=1,2,3,4) between VEP and the real EP. The absolute values of the correlations statistically converge to 1 in our database as the number of registered images increases. Thus, we have confirmed that the VEP gradually converges to the real EP by input of additional images.

Figure 3 shows that PC1 and PC2 exhibit Lambertian properties, whereas PC3 and PC4 contain a lot of personal properties but few Lambertian properties.

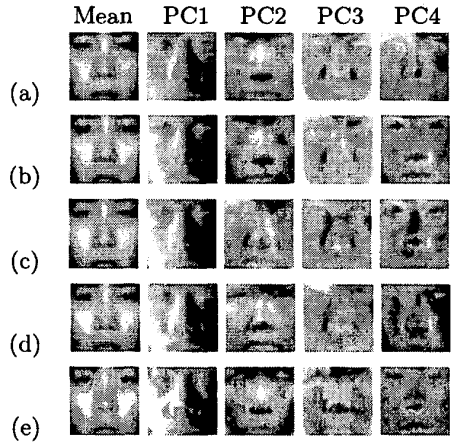


Figure 3. Successive VEP construction.

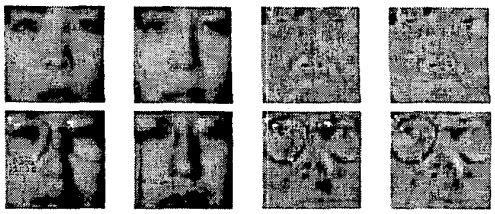


Figure 4. Examples of noise suppression on $\mathbf{x}^{\#}$:(a) and (b) show x and $\mathbf{x}^*$, (c) and (d) $\mathbf{x}^{\#}$ before/after noise suppression.

5. Refinement of ER construction

5.1. Noise suppression of $\mathbf{x}^{\#}$

The residual component $\mathbf{x}^{\#}$ includes a lot of personal properties which cannot be encoded in the canonical component $\mathbf{x}^*$. On the other hand, the residual component also includes noise, such as spike reflections and shadows. In order to utilize the fruitful personal information, we must suppress the noise.

Let $x_j^{\#}(j = 1...N)$ denote the intensity of the j-th pixel of $\mathbf{x}^{\#}$, where N is the number of pixels. The average and standard deviation of all the pixels in $\mathbf{x}^{\#}$ are calculated by

$$\overline{x^{\#}} = \frac{1}{N}\sum_{j=1}^{N} x_j^{\#}$$

$$\sigma_{x^{\#}} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(x_j^{\#} - \overline{x^{\#}})^2}.$$

A function for spike noise suppression is accomplished by changing $x_j^{\#}$ to $\overline{x^{\#}}$ when

$$|x_j^{\#} - \overline{x^{\#}}| \geq w\sigma_{x^{\#}}.$$

Note that the other pixels should retain their original values, because they contain rich personal properties which cannot be encoded in CS.

Figure 4 shows two examples of noise suppression. Input images $\mathbf{x}$ in column (a) are decomposed to $\mathbf{x}^*$ shown in (b) and to $\mathbf{x}^{\#}$ by the projection to CS. Then the residuals shown in (c) are refined to (d) by noise suppression. Although little change occurs in the first example, in the second example much spike noise is suppressed in the eyeglasses.

Assuming symmetry of the face, we can utilize a mirror image of $\mathbf{x}^{\#}$, because all faces in our database are frontal faces. Let $\mathbf{x}_1^{\#}$ and $\mathbf{x}_2^{\#}$ denote $\mathbf{x}^{\#}$ and its mirror image. Suppose a more general situation, where

a set of images $\{\mathbf{x}_i^{\#}(i = 1..K)\}$ is provided for one person. Let $\mathbf{x}_{K+i}^{\#}$ denote a mirror image of $\mathbf{x}_i^{\#}$. After noise suppression of each image, the images are used for the construction of ER.

6. Recognition Scheme and Experiments

6.1. Discriminant Functions

In face recognition based on eigenfaces, distance-from-feature-space(DFFS) [5] is often used as a discriminant function. However, from a preliminary experiment, we find that in NIS a normalized correlation works better than the DFFS distance. Since we have three types of eigenspaces, three types of similarities can be defined by use of a normalized correlation $C(\mathbf{x}, \mathbf{y})$ as defined in Eq. (6). Given an unknown face $\mathbf{x}$, the similarities are represented as follows:
(1) Similarity between $\mathbf{x}$ and Φ_{pm} in NIS:

$$C0_p(\mathbf{x}) = C(\Phi_{pm}\tilde{\mathbf{x}}_p + \overline{\mathbf{x}}_p, \mathbf{x}).$$

(2) Similarity between $\mathbf{x}^*$ and (V)EP Φ_{pm}^* in CS:

$$C1_p(\mathbf{x}) = C(\Phi_{pm}^*\tilde{\mathbf{x}}_p^* + \overline{\mathbf{x}_p^*}, \mathbf{x}^*).$$

(3) Similarity between $\mathbf{x}^{\#}$ and ER $\Phi_{pm}^{\#}$ in $\mathrm{CS}^{\perp}$:

$$C2_p(\mathbf{x}) = C(\Phi_{pm}^{\#}\tilde{\mathbf{x}}_p^{\#} + \overline{\mathbf{x}_p^{\#}}, \mathbf{x}^{\#}).$$

Because $C1$ and $C2$ are calculated independently in CS and $\mathrm{CS}^{\perp}$, they can be combined to derive a new similarity

$$C3_p(\mathbf{x}) = \frac{C1_p(\mathbf{x})}{C1_{\widehat{p}_1}(\mathbf{x})} + \frac{C2_p(\mathbf{x})}{C2_{\widehat{p}_2}(\mathbf{x})},$$

$$\text{where} \quad \widehat{p}_i = \arg\max_{1 \le p \le P} Ci_p(\mathbf{x}).$$

A simple discrimination rule is created for $Ci(i = 0, 1, 2, 3)$, to select a person by

$$\arg\max_{1 \le p \le P} Ci_p(\mathbf{x}).$$

6.2. Data specification

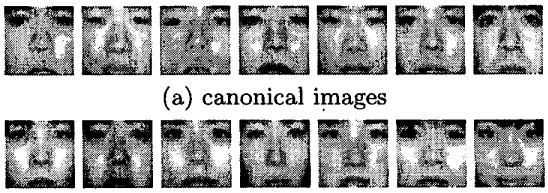
Data specification is summarized in Table 1. Facial images were taken from a fixed camera located in our laboratory. Each of 100 persons looked forward while sitting in a chair located a fixed distance from the camera. The chair was fixed in order to obtain the frontal facial images of each person.

As shown in 3.1, CS is created from the canonical set, which consists of 1200 images of 50 persons. For each person, images are taken under 24 lighting conditions, which were controlled by changing the position of a light. In the canonical set, 9 persons wear glasses. Figure 5 shows averages of the canonical images taken under the 24 lighting conditions. The remaining 50 persons are used for the test data, in which 15 persons wear glasses. Figures 6 (a) and (b) show 7 examples of canonical images and 7 examples of the test images, respectively, taken under a fixed lighting condition.

Table 1. Data specification

	Canonical set	Test images
# of persons(P)	50	50
# of lighting conditions(L)	24	24
Image size(N)	32×32	32×32
persons wearing glasses	9	15

Figure 5. Averages of canonical images taken under 24 lighting conditions.



(a) canonical images

(b) test images

Figure 6. Examples of two image sets.

6.3. Recognition result for a whole face

For personal registration, K images were randomly sampled from 24 images of each person in the test data. Therefore, the discrimination experiment was carried out from the remaining $24-K$ images of each registered person. This process was repeated one hundred times while registered images for each person were varied.

Table 2 shows average discrimination rates. Eleven methods are compared by use of the same canonical and test sets. The first 8 rows show the results when only the whole faces are used, and the remaining 3

Table 2. Discrimination rates for 50 persons[%] : in table $h(K) = \min(2K-1, K+2)$

Method	Dim. of ES	Number of samples(K)/person 1	2	3	4	5
NN	45	35.8	50.2	63.1	63.6	68.7
$C0'$	$h(K)$	61.0	81.4	91.6	93.2	96.1
$C1'$	$h(K)$	47.1	77.1	90.0	92.7	96.2
$C2'$	$h(K)$	84.1	91.6	94.2	96.0	96.6
$C3'$	–	82.0	92.1	94.7	96.6	97.1
$C1$	K+3	81.2	88.3	91.1	91.8	93.3
$C2$	$h(K)$	89.0	96.2	98.2	99.1	99.3
$C3$	–	92.1	97.3	98.6	99.3	99.3
$C1^{\$}$	K+3	90.3	95.4	97.3	97.7	98.0
$C2^{\$}$	$h(K)$	89.0	96.0	98.0	99.2	99.5
$C3^{\$}$	–	96.4	98.9	99.5	99.5	99.7

rows show the results when, in addition to the whole faces, three features are used for recognition. In all the methods, face symmetry is also used in the registration stage.

In the table, NN indicates the nearest neighbor discrimination in CS. $C0'$, $C1'$, and $C2'$ indicate that the conventional eigenspace method is used for eigenface(EF), eigen-projection(EP) and eigen-residual(ER), respectively. In these methods, we construct $\min(2K-1, K+2)$ dimensional eigenspaces using face symmetry. $C3'$ shows a result when EP and ER are combined. These results indicate that the residual components are very effective for face recognition.

In the second class of methods, represented by $C1$, $C2$, and $C3$, the additional refinements discussed in Sections 4 and 5 are used instead of the fundamental construction discussed in 3.2. As compared with the results for $C1'$, the results for $C1$ are much improved because of VEP construction. The results for $C2$ are also improved from the results for $C2'$. Our best result is yielded with $C3$, with recognition reaching 92.1% even if only one image is registered for each person. This represents an improvement of about 30 points over the results obtained from the conventional eigenface method($C0'$). When five images are registered for each person, the results obtained with $C3$ reach 99.3%.

6.4. Recognition using three features

In the above experiments, only the whole faces are used for recognition, as shown by broken lines in Fig. 7. Let us also take into account three other features (eyes, nose, and mouth, as shown by rectangles in Fig. 7). The discrimination scheme is summarized as follows: Three canonical spaces are created in the same manner

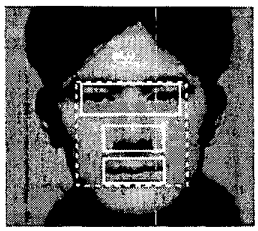


Figure 7. Three features used in $Ci^{\$}$.

as described in 3.1, and EP and ER are created in each CS and $\mathrm{CS}^{\perp}$ as described in Sections 4 and 5. Let $Ci_{\phi p}(\mathbf{x})$ denote a similarity $Ci(i = 0, 1, 2, 3)$ for the p-th eigenspace of the feature $\phi(= 1, 2, 3)$. Therefore, the most similar person $\widehat{p}_{i\phi}$ is defined for each i and ϕ as

$$\widehat{p}_{i\phi} = \arg \max_{1 \le p \le P} Ci_{\phi p}(\mathbf{x}).$$

Let $\phi = 0$ indicate a case of the whole face. A combined similarity is defined in

$$Ci_p^{\$}(\mathbf{x}) = \sum_{\phi=0}^{3} \frac{Ci_{\phi p}(\mathbf{x})}{Ci_{\phi \widehat{p}_{i\phi}}(\mathbf{x})}.$$

A simple discrimination rule is specified with $Ci^{\$}(i = 0, 1, 2, 3)$, which selects the person by

$$\arg \max_{1 \le p \le P} Ci_p^{\$}(\mathbf{x}).$$

In the last three rows of Table 2, we show average discrimination rates when the three features are used in combination with the whole face. Among the similarities, $C3^{\$}$ yields the best result. The discrimination rates are improved when the three features are also used. When just one image is registered for each person, the discrimination rate is 92.1% for the whole face ($C3$), and is improved to 96.4 % when the three features are also used ($C3^{\$}$). When five images are registered for each person, the discrimination rate reaches 99.7 %.

6.5. Recognition on AR and Yale databases

In order to confirm the effectiveness of our methods, we also conducted experiments on the AR [4] and Yale Face Databases. These experiments employ the same canonical spaces used in the above experiments. Table 3 shows the experimental results.

The AR database contains images of 68 persons taken under 3 lighting conditions for each person. In order to obtain clear comparison with the results in our database, 50 persons are randomly selected from the AR database, and the discrimination experiment is

Table 3. Results for 3 databases[%]

Method		$C3$	$C3$	$C3^{\$}$	$C3^{\$}$
Database	P	K=1	K=2	K=1	K=2
Ours	50	92.1	97.3	96.4	98.9
AR	50	91.6	98.1	94.2	99.3
Yale	15	95.6	96.7	95.6	100.0

carried out using $C3$ and $C3^{\$}$. In the experiment, one or two images are registered for each person and the remaining images are used for the test. Table 3 shows that our methods can stably discriminate persons in the AR database as well as persons in our database.

We also apply our methods to the Yale database, which contains images of 15 persons taken under 4 lighting conditions. From the four images of each person, one or two images are randomly sampled for registration and the remaining images are used for the test. As shown in the table, the discrimination results seem good. For illumination-insensitive face recognition, Zhao and Chellappa [13] propose an approach based on shape-from-shading. As compared with their method, our method provides better results on the Yale Database. When one image is registered for each person, our method employing $C3^{\$}$ yields 95.6%, whereas their method yields 83.3%. When two images are registered, our method yields 100 %.

7. Conclusions

The concept of decomposed eigenface is introduced in order to realize robust face recognition that is insensitive to lighting conditions. The basic concept is to decompose the eigenface to two eigenspaces, in CS and in $CS^{\perp}$, via projection from NIS to CS. The decomposition enables us to independently refine construction algorithms in CS and in $CS^{\perp}$. Experimental results show the effectiveness of both the basic concept of decomposition and the refinement algorithms. Both the VEP concept and the noise suppression on ER work well on the AR, Yale, and our own databases, even if the images are taken under different lighting conditions. The decomposed eigenface can be applied to face recognition under natural lighting conditions, even if the lighting condition is unknown or changes from time to time. We also hope that the concept of eigenspace decomposition can work effectively for other problems in image understanding and computer vision.

Acknowledgment

The authors would like to thank Shohei Oki, Nobumasa Yamamoto, and Kunihiro Nakagawa for providing helpful utilities and kind comments. This work was partly supported by Research for the Future Program, the Japan Society for the Promotion of Science (Project ID: JSPS-RFTF96P00501). It has been also partly supported by Japan Science and Technology Corporation under Ikeuchi CREST project.

References

[1] 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.

[2] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The Bas-Relief Ambiguity", Proc. CVPR'97, pp. 1040-1046, 1997.

[3] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103-108, 1990.

[4] A. Z. Martinez and R. Benavente, The AR Face Database, CVC Technical Report #24, 1998.

[5] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 696-710, 1997.

[6] E. Oja, Subspace Methods of Pattern Recognition, Research Studies Press Ltd., 1983.

[7] T. Shakunaga, K. Ogawa, and S. Oki, "Integration of Eigentemplate and Structure Matching for Automatic Facial Feature Detection", Proc. FG'98, pp. 94-99, 1998.

[8] T. Shakunaga and N. Yamamoto, "Virtual Subspace Method for Robust Face Recognition Independent of Lighting Conditions", Proc. MVA'00, pp. 205-210, 2000.

[9] A. Shashua, "Geometry and Photometry in 3D Visual Recognition", Ph.D. Thesis, MIT, 1992.

[10] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[11] T. Vetter, M. J. Jones, and T. Poggio, "A Bootstrapping Algorithm for Learning Linear Models of Object Classes", Proc. CVPR'97, pp. 40-46, 1997.

[12] A. L. Yuille, S. R. Epstein, and P. N. Belhumeur, "Determining Generative Models of Objects Under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability", International Journal of Computer Vision, vol. 35, no. 3, pp. 203-222, 1999.

[13] W. Y. Zhao and R. Chellappa, "Illumination-Insensitive Face Recognition Using Symmetric Shape-from-Shading", Proc. CVPR'00, pp. 286-293, 2000.