

Decomposed Eigenface Method along with Image Correction for Robust Face Recognition

Kazuma Shigenari Fumihiko Sakaue Takeshi Shakunaga *
 Department of Information Technology, Okayama University

Abstract

We have proposed a decomposition of the eigenface into two orthogonal eigenspaces and have shown that the decomposition is effective for realizing robust face recognition under various lighting conditions [10]. The present paper refines the decomposed eigenface method by introducing a projection-based image correction. The image correction technique is principally authorized when the object shape is fixed and a sufficient number of images are taken beforehand. However, the proposed technique can also be applied to a canonical eigenspace, which is constructed from several faces taken under various lighting conditions. Reflective noises, shadows and occlusions are detected and corrected by the projection of a facial image onto the canonical eigenface. Based on the newly proposed image correction, we develop herein a refined decomposed eigenface method. The experimental results indicate that the refinement works well for face recognition under various lighting conditions, as compared to the original decomposed eigenface method.

1 Introduction

Appearance-based face recognition can be resolved to the eigenface method[6], which is often identical to the subspace method[5, 2], if several facial images can be collected in the registration stage. The eigenface, however, cannot be stably composed when too few sample images are available or when the lighting conditions are highly similar over the sample images. In order to solve this problem, Shakunaga and Yamamoto[9] proposed the concept of virtual subspace, and Shakunaga and Shigenari[10] refined this concept to obtain the decomposed eigenface method. The decomposed eigenface method facilitates face recognition under various lighting conditions to cover cases in which too few images are available for registration.

The present paper improves the original decomposed eigenface method[10] by introducing a projection-based image correction.

Section 2 summarizes the decomposed eigenface method[10]. After briefly introducing the terminology in 2.1-2.3, registration and recognition schemes are described in 2.5. Refinement of the decomposed eigenface method is described in Section 3. Projection-based image correction is introduced in 3.1 using a canonical eigenspace (CS) which is constructed using several facial images taken under various lighting conditions. Examples of the image correction are also shown in 3.1.

The image correction enables the refinement of the CS and two schemes for registration and recognition, as summarized in 3.2 and 3.3.

Section 4 compares an experimentally obtained result to the result given by the original method for a database of facial images of 50 persons under 24 different lighting conditions.

2 Decomposed Eigenface Method

2.1 Normalized Image Space

Let an N -dimensional vector \mathbf{X} denote an image, and let $\mathbf{1}$ denote an N -dimensional vector in which any element is equal to 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 the 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 in the Normalized Image Space (NIS).

2.2 Canonical Space

In the present paper, a facial space is defined as a space composed from a set of frontal faces obtained from several persons under various lighting conditions. In order to simplify the problem, we assume that a good segmentation is readily accomplished for each normalized image, as shown in Fig. 4. Eigenspace analysis on the face space decreases the dimension of the face space with little loss of representability [2, 6]. In our experiments, a 45d eigenspace, called the canonical space (CS), is constructed from a canonical image set which consists of facial images of 50 persons under 24 lighting conditions.

Let $\bar{\mathbf{x}}$ and Σ denote the mean image and the covariance matrix on the canonical set, respectively. Let Λ denote a diagonal matrix of which the diagonal terms are eigenvalues of Σ in descending order, and let Φ denote a matrix of 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 n largest-eigenvalue eigenvectors, a projection of \mathbf{x} to the n -dimensional CS and the residual are respectively given as

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

and

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

Thus, a normalized image \mathbf{x} can be decomposed into the canonical component \mathbf{x}^* and the residual component $\mathbf{x}^\#$, which are orthogonal by definition.

2.3 Eigen-projection and Eigen-residual

The orthogonal components \mathbf{x}^* and $\mathbf{x}^\#$ enable us to decompose the eigenface (EF) in NIS. That is, as shown

*Address: Okayama-shi, Okayama 700-8530, Japan. E-mail: {sakaue, shaku}@chino.it.okayama-u.ac.jp

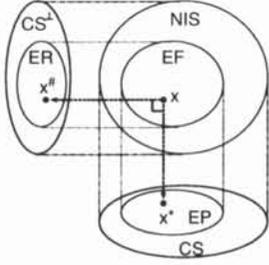


Figure 1: Decomposition of the EF to the EP and the ER.

in Fig. 1, two eigenspaces can be constructed independently in CS and in the orthogonal complement 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 . The EP and the ER are constructed by the eigenspace analysis in CS and CS^\perp .

For the EP construction, the mean vector $\bar{\mathbf{x}}_p^*$ and the covariance matrix Σ_p^* are calculated as

$$\bar{\mathbf{x}}_p^* = \frac{1}{L} \sum_{l=1}^L \mathbf{x}_{pl}^* \quad (1)$$

and

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

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

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

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

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

2.4 Virtual Eigen-projection

Virtual eigenspace is defined as a virtualized concept of eigenspace [10]. When at least one image is registered for a person, a virtual eigenspace can be directly constructed over a set of virtual images that are synthesized by lighting estimation followed by lighting transformation. Virtual eigenspace converges to the real eigenspace when additional images are taken.

We can also synthesize a set of images in CS from a single projection. Therefore, an eigenspace can be constructed in CS by PCA. Let us call the virtual eigenspace a virtual eigen-projection (VEP). The image synthesis is based on the following lighting estimation and lighting transformation:

(1) Lighting estimation in CS

We can estimate the lighting condition under which each input image is taken. In our current implementation, a nearest-neighbor discrimination is performed for 24 lighting conditions.

(2) Lighting transformation in CS

We can prepare a linear transformation in CS which approximately transforms a facial image taken under one lighting condition to a facial image taken under another condition. The transformation is constructed from a canonical image set which is used for CS construction. Using the lighting transformation, a set of images is constructed.

A VEP can be constructed in the CS from the synthesized image set, even if only one image is registered for each person. When additional images are registered, the VEP is updated by nearest-neighbor selection of the virtual images.

2.5 Registration and Recognition Schemes

In the registration stage, an input facial image \mathbf{x} is decomposed to \mathbf{x}^* and $\mathbf{x}^\#$. Then, the EP and the ER are created independently in CS and in CS^\perp , respectively. The Virtual Eigen-projection (VEP) is constructed as described in 2.4.

In the recognition stage, we can realize the face identification by combining the two eigenspaces. Given an unknown face \mathbf{x} , two similarity measures are defined by normalized correlations in CS and CS^\perp , where $C(\mathbf{x}, \mathbf{y})$ shows a normalized correlation of \mathbf{x} and \mathbf{y} :

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

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

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

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

(3) Combined similarity of C1 and C2:

Because $C1$ and $C2$ are calculated independently in CS and CS^\perp , they can be combined as shown below

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

where $\hat{p}_i = \arg \max_{1 \leq p \leq P} C_i p(\mathbf{x})$.

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

$$\arg \max_{1 \leq p \leq P} C_i p(\mathbf{x}).$$

3 Refinement of Decomposed Eigenface Method

3.1 Projection-based Image Correction

As discussed in 2.2, a projection of an image \mathbf{x} to CS is given as $\mathbf{x}^* = \Phi_n^T (\mathbf{x} - \bar{\mathbf{x}})$. The residual $\mathbf{x}^\#$ is then expressed as $\mathbf{x}^\# = \mathbf{x} - \bar{\mathbf{x}} - \Phi_n \mathbf{x}^*$.

In this section, we refine the original decomposed eigenface method by image correction prior to the decomposition. For this purpose, let us define the relative residual r_i for the i -th pixel of \mathbf{x} as

$$r_i = \frac{\mathbf{e}_i^T \mathbf{x}^\#}{\mathbf{e}_i^T (\bar{\mathbf{x}} + \Phi_n \mathbf{x}^*)},$$

where \mathbf{e}_i is a unit vector of which only the i -th element is 1 and the other elements are 0.

Then, a pixel-wise correction is defined as follows. When $|r_i| \geq r_\theta$ for a threshold r_θ , the i -th pixel of \mathbf{x} is replaced by $e_i^T(\bar{\mathbf{x}} + \Phi_n \mathbf{x}^*)$. The image correction makes an intensity value to be consistent with the projection. For example, regions obscured by shadows and reflections from eyeglasses are removed by the image correction.

The image correction does not satisfy the normality of the image. Therefore, the corrected image should be re-normalized when all of the pixels are checked and corrected.

The projection-based correction changes outliers to inliers. When more than a few pixels are corrected, \mathbf{x}^* also changes to some extent. Therefore, a few iterations of the image correction should be performed in order to obtain better noise suppression. After a few iterations, \mathbf{x} converges to an image containing little noise and few shadows.

Figure 2 shows an example of image correction. The original image in the top row is gradually corrected in both the projection and residual domains. In particular, reflections from the person’s glasses are suppressed.

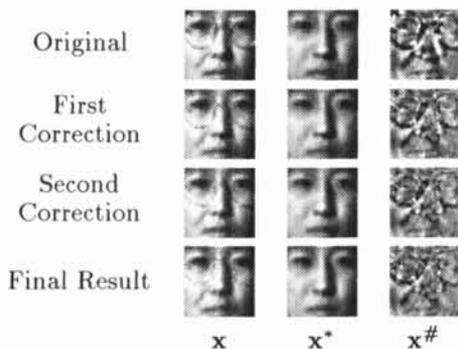


Figure 2: Example of iterative image correction.

3.2 Refinement of Canonical Space

Because the image correction can also be applied to any image in the canonical image set, CS is reconstructed from the corrected image set.

Figure 3 shows the most significant five eigenfaces of CS before and after image correction. The image correction refines the canonical spaces because several reflections and shadows are removed from the canonical image set. After reconstruction of the canonical space, lighting transformation matrices are also reestimated for the construction of the VEP[10].

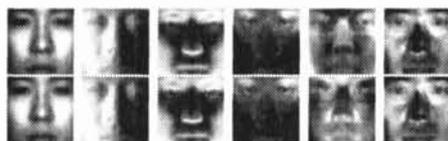


Figure 3: Comparison of average face and the most significant five bases: Upper and lower rows show images before and after image correction, respectively.

3.3 Refined Registration/Recognition Schemes

In the registration scheme, the VEP is constructed in the manner described in 2.5 after image correction.

Because both the CS and the registered image are corrected, each VEP includes much less noise than the original method. No further noise reduction is necessary for the construction of the ER because the image correction repairs the ER.

In the recognition scheme, the subspace method is applied to an image after the image correction. Because the VEP and the ER include less noise, the proposed recognition scheme works better than the original method.

4 Experimental result

4.1 Data Specifications

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

Table 1: Data specifications

| | Canonical set | Test images |
|------------------------------|---------------|-------------|
| # of persons | 50 | 50 |
| # of lighting conditions | 24 | 24 |
| Image size | 32×32 | 32×32 |
| # of persons wearing glasses | 9 | 15 |

As shown in 2.2, CS is created from the canonical image set, which consists of 1200 images of 50 persons. For each person, images were taken under 24 lighting conditions, which were controlled by changing the position of the light. In the canonical set, nine persons wore glasses. Figure 4 shows the averages of the canonical images taken under the 24 lighting conditions. The remaining 50 persons were used for the test data, in which 15 persons wore glasses. Figures 5 (a) and (b) show seven examples of canonical images and seven example test images, respectively, taken under fixed lighting conditions.



Figure 4: Averages of canonical images under 24 lighting conditions.

4.2 Comparison

For personal registration, K images were randomly sampled from 24 images of each person in the test

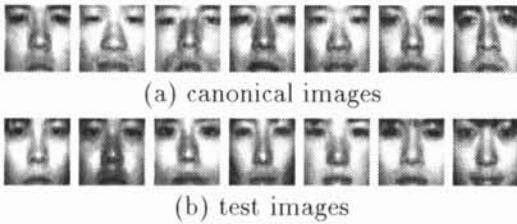


Figure 5: Examples of canonical/test image sets.

data. Therefore, the discrimination experiment was performed using 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 the average discrimination rates. Six methods were compared using the same canonical set and test set. The first three rows show the results of the original method, and the remaining three rows show the results of the refined method. In all of the methods, face symmetry was used in the registration stage. The dimension of the EP is $K + 3$, whereas the dimension of the ER is $\min(2K - 1, K + 2)$.

The table shows that the discrimination rates are improved by the image correction for all K and for all similarity measures. The best result is provided by $C3$, with recognition reaching 95.1% when only one image is registered for each person. This represents an improvement of three points over the original method. When five images are registered for each person, the result obtained using $C3$ reaches 99.9%.

Table 2: Discrimination rates for 50 persons [%]

| Method | Number of samples(K)/person | | | | |
|-----------------|-----------------------------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 |
| $C1$ (original) | 81.2 | 88.3 | 91.1 | 91.8 | 93.3 |
| $C2$ (original) | 89.0 | 96.2 | 98.2 | 99.1 | 99.3 |
| $C3$ (original) | 92.1 | 97.3 | 98.6 | 99.3 | 99.3 |
| $C1$ (refined) | 83.1 | 91.4 | 93.1 | 94.9 | 94.6 |
| $C2$ (refined) | 91.0 | 97.0 | 98.9 | 99.6 | 99.8 |
| $C3$ (refined) | 95.1 | 98.5 | 99.4 | 99.8 | 99.9 |

4.3 Recognition on AR database

In order to confirm the effectiveness of the refinement, we conducted experiments on the AR Face Database [4]. 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 three lighting conditions for each person. In order to obtain a clear comparison with the results in our database, 50 persons are randomly selected from the AR database, and the discrimination experiment is performed using $C3$. 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 the refinement can stably discriminate persons in the AR database as well as persons in our database.

5 Conclusions

Refinement of the decomposed eigenface method is established based on projection-based image correction. Experimental results show that the refinement improves the discrimination rates on both our own and

Table 3: Results for two databases [%]

| Method | P | $C3$ (original) | | $C3$ (refined) | |
|--------|----|-----------------|------|----------------|------|
| | | K=1 | K=2 | K=1 | K=2 |
| Ours | 50 | 92.1 | 97.3 | 95.1 | 98.5 |
| AR | 50 | 91.6 | 98.1 | 92.3 | 99.4 |

AR databases. The refined method can be applied to a face recognition under natural lighting conditions, even if the lighting condition is unknown or changes with time. The image correction method can also be applied to a wide range of applications, including face and object recognition.

Acknowledgment

This work has been supported in part 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] 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.
- [3] 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.
- [4] A. Z. Martinez and R. Benavente, The AR Face Database, CVC Technical Report #24, 1998.
- [5] E. Oja, Subspace Methods of Pattern Recognition, Research Studies Press Ltd., 1983.
- [6] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
- [7] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?", Proc. CVPR'96, pp.270-277, 1997.
- [8] 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.
- [9] T. Shkunaga and N. Yamamoto, "Virtual Subspace Method for Robust Face Recognition Independent of Lighting Conditions", Proc. IAPR Workshop on Machine Vision Applications (MVA 2000), 2000.
- [10] T. Shkunaga and K. Shigenari, "Decomposed Eigenface for Face Recognition under Various Lighting Conditions", Proc. CVPR2001, vol. 1, pp. 864-871, 2001.
- [11] T. Shkunaga and F. Sakaue, "Natural Image Correction by Iterative Projections to Eigenspace Constructed in Normalized Image Space", in Proc. ICPR2002, 2002.
- [12] W. Y. Zhao and R. Chellappa, "Illumination insensitive face recognition using symmetric shape-from-shading", Proc. CVPR2000, pp.286-293, 2000.