MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

6—1 Virtual Subspace Method for Robust Face Recognition Independent of Lighting Conditions

Takeshi SHAKUNAGA Nobumasa YAMAMOTO*

Department of Information Technology
Okayama University

Abstract

A concept of virtual subspace is introduced for realizing a robust face recognition independent of the lighting conditions. The virtual subspace is a paradoxical concept because it can be constructed even if only one image is taken. Furthermore, the virtual subspace is gradually converged to the real subspace when face images are subsequently taken. The virtual subspace is defined as an eigenspace composed from a synthesized image set which are supposed to be taken in a variety of lighting conditions. An integration algorithm is also proposed for updating the virtual subspace when additional images are available. In the experiments, we show the effectiveness of the virtual subspace method in comparison with both the conventional subspace method and the nearest neighbor discrimination.

1 Introduction

It is known that the face recognition problem can be reduced to the subspace method[6][3], if a lot of face images can be collected for registration. When the number of sample images is small, however, a subspace cannot be stably composed and the subspace method cannot work well. Even if the number of images is not small, the subspace method often cannot work when the lighting conditions are very similar over the sample images. In these situations, some improvement is necessary to the subspace method for realizing the robust recognition.

A concept of virtual subspace is introduced for realizing the subspace construction even if only one image can be used for the subspace construction. We combine a lighting estimation and a lighting compensation to synthesize virtual face images in a variety of lighting conditions. The virtual subspace is constructed over the synthesized images by eigenspace analysis.

*Address: Tsushima naka 3-1-1, Okayama 700-8530, Japan. E-Mail:shaku@it.okayama-u.ac.jp

2 Subspace Method and Virtual Subspace

2.1 Canonical eigenspace

In this paper, a face space is defined as a space composed from a set of frontal faces, which includes a lot of persons under a lot of lighting conditions. To simplify the problem, we assume that the face direction is fixed and a good segmentation is readily accomplished for each input image, as shown in Fig. 7.Eigenspace analysis (principal component analysis) on the face space decreases the dimension of the face space with a little loss of the representability [3][1]. In our experiments, the face eigenspace is constructed from an face image set of 50 persons in 24 lighting conditions. Let us call the eigenspace the canonical (eigen)space from now on. The dimension of the canonical space is set to 45.

Let $\mathbf{x}_{pl}$ denote a face of the p-th person under the l-th lighting condition. The mean vector and the covariance matrix are then calculated by

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

$$\Sigma = \frac{1}{PL}\sum_{p=1}^{P}\sum_{l=1}^{L}(\mathbf{x}_{pl}-\overline{\mathbf{x}})(\mathbf{x}_{pl}-\overline{\mathbf{x}})^T, \tag{2}$$

where P and L are numbers of persons and lighting conditions, respectively. Let Φ and Λ denote eigenvectors and a diagonal matrix of which diagonal elements are eigenvalues in the descending order, respectively. That is,

$$\Lambda = \Phi^T\Sigma\Phi.$$

Using a submatrix Φ_n of Φ, of which corresponding eigenvalues are in the largest n ones of Φ. Then a projection of $\mathbf{x}$ into the n-dimensional canonical space is defined by

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

Two distances can be defined in the canonical space[1]. They are called DFFS(distance-from-feature-space) and DIFS(distance-in-feature-space), respectively:

$$\begin{aligned} DFFS(\mathbf{x}) &= \sqrt{\|\mathbf{x}-\overline{\mathbf{x}}\|^2 - \|\Phi_n \mathbf{x}^*\|^2} \\ DIFS(\mathbf{x}) &= (\mathbf{x}^*)^T \Sigma_n^{-1} \mathbf{x}^* \end{aligned}$$

where Σ_n is a submatrix of Σ, of which corresponding eigenvalues are among the largest n eigenvalues of Σ.

2.2 Subspaces and subspace method

When a lot of images are taken, a subspace can be constructed for each person in the similar way as shown in Section 2.1. Then a projection of $\mathbf{x}$ into the m-dimensional subspace for the p-th person is defined by

$$\tilde{\mathbf{x}}_\mathbf{p} = \Phi_{pm}{}^T(\mathbf{x} - \overline{\mathbf{x}_p}) \quad \text{where} \quad \overline{\mathbf{x}_p} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{x}_{pl}.$$

Given an unknown face $\mathbf{x}$, the subspace method selects a person, p^* by

$$p^* = \arg \min_{1 \le p \le P} DFFS_p(\mathbf{x})$$

$$\text{where} \quad DFFS_p(\mathbf{x}) = \sqrt{\|\mathbf{x} - \overline{\mathbf{x}_p}\|^2 - \|\Phi_{pm}\tilde{\mathbf{x}}_p\|^2}.$$

2.3 Concept of virtual subspace

The subspace method [7][4] works well when each subspace is constructed from sufficient images which cover a variety of lighting conditions. However, this requirement is often unsatisfied when the number of images are too small or when the lighting conditions are very similar over the image set. In these situations, the conventional subspace method cannot work because subspaces cannot be stably constructed. If a subspace can be constructed from such a critical situation, the subspace method can be applied to wider applications. We would like to propose a concept of the virtual subspace for the purpose. The virtual subspace should be a virtual concept of the subspace, the following two requirements should be satisfied: (1) It should be constructed from a single image while the real subspace cannot be constructed directly from a single image. (2) It should be converged to the real subspace when subsequent images are taken.

To satisfy the first requirement, we synthesize a set of images in virtual lighting conditions from a single image, and construct a virtual subspace from the set. In this process, the image synthesis is based on the lighting estimation followed by the lighting compensation in the canonical space. The algorithm details are shown in Section 3. To satisfy the second requirement, an integration algorithm is defined and examined over the test data in Section 4.

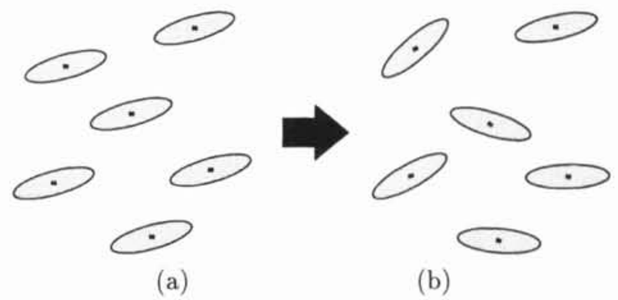


Figure 1: Virtual and real subspaces.

Figure 1 shows a relation between the virtual and the real subspaces. The virtual subspaces are created using the average distribution of the average person in a variety of lighting conditions, as illustrated in Fig. 1(a). On the other hand, the real subspaces are created from a lot of images of individual persons, as shown in Fig. 1(b). They have individual distributions which are affected by both the geometric and photometric properties of the individual faces.

3 Virtual Subspace Construction from a Single Image

3.1 Face mapping in canonical eigenspace

Let us define a mapping function in the canonical eigenspace, which transforms a face taken in a lighting condition to one in another condition. In the strict meaning, the mapping function depends on persons as well as on lighting conditions because both the geometric and photometric properties of faces are dependent on persons. While the strict lighting estimation seems too expensive, such an estimation is not necessary for our purpose. What we would like is to make a mapping function in the average meaning.

A mapping function can be created from the learning set which was used for composing the canonical space. For a simple implementation, the mapping is assumed to be a linear transformation of a face vector $\mathbf{x}^*$. Let $F^*(l_1, l_2)$ denote a matrix for face mapping from a lighting condition l_1 to a 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 the squared sum of errors,

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

Once the matrix $F^*(l_1, l_2)$ is specified, a face image in l_1 can be converted to an image in l_2 by a linear mapping.

3.2 Estimation of lighting condition

Our next point is posed in an estimation of lighting condition from a given face image. We suppose that a face is well segmented a priori, and a dictionary image set can be used again for the estimation. The estimation is reduced to the nearest neighbor discrimination in the canonical space. 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. Let $\lambda(\mathbf{x}^*)$ denote the lighting estimation function from now on, which selects a lighting condition by the nearest neighbor criterion.

3.3 Face mapping in image space

Combining the mapping F^* and the lighting condition estimation $\lambda(\mathbf{x}^*)$, we can propose a mapping function in the image space with compensating the lighting conditions. A simple mapping is composed using F^* and $\lambda(\mathbf{x}^*)$. Suppose that a face image $\mathbf{x}$ is transformed to one in a lighting condition l. Figure 2 shows a scheme of the compensation. In this scheme, $F^*(\lambda(\mathbf{x}^*), l)\mathbf{x}^*$ indicates a compensated face of $\mathbf{x}^*$ in the canonical space, which is expected to be converted from an image taken in the lighting condition l. Using this compensation, a simple mapping is defined in the image space by

$$F(\mathbf{x}, l) = \Phi_n F^*(\lambda(\mathbf{x}^*), l)\mathbf{x}^* + \Delta\mathbf{x} + \overline{\mathbf{x}}$$

$$\text{where} \quad \Delta\mathbf{x} = \mathbf{x} - \overline{\mathbf{x}} - \Phi_n \mathbf{x}^*.$$

It should be noted that the residual of the coding to the canonical eigenspace is reserved in the second term.

Figure 3 shows the lighting compensation process along with an example. Suppose an input image $\mathbf{x}$, shown in Fig. 3 (a), to be transformed to an image in the lighting condition l. For reference, a real image $\mathbf{y}$ taken in the lighting condition l is shown in Fig. 3 (c).

In the compensation process, $\mathbf{x}$ is encoded to $\mathbf{x}^*$ in the canonical space as shown in (d). By the lighting compensation in the space, $F^*(\lambda(\mathbf{x}^*), l)\mathbf{x}^*$ is created as shown in (e), while $\mathbf{y}^*$ is shown in (f). Finally, $F(\mathbf{x}, l)$ is created in the image space as shown in (b) while $\mathbf{y}$ is shown in (c). This shows that the lighting compensation works well to simulate a lighting conversion.

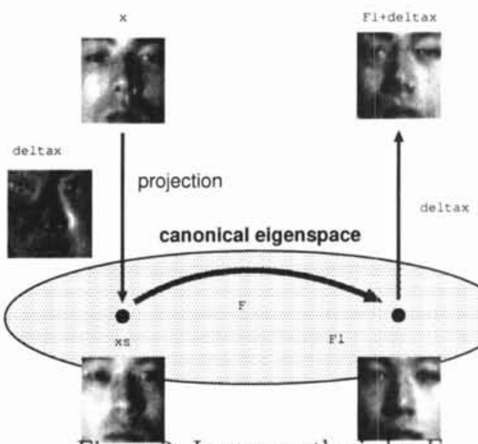


Figure 2: Image synthesis by F.

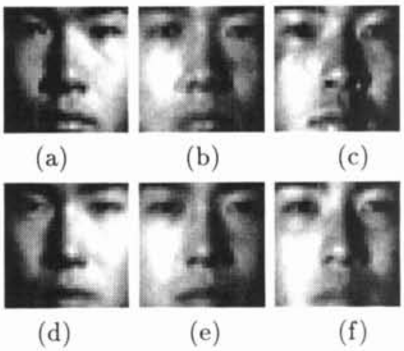


Figure 3: An example of lighting compensation.

3.4 Virtual subspace construction

A virtual subspace can be constructed from the synthesized images by the direct application of the algorithm as shown in 2.2. In the subspace construction, the following definitions are used instead of Eqs. (1) and (2).

$$\overline{\mathbf{x}_p} = \frac{1}{L}\sum_{l=1}^{L} F(\mathbf{x}, l) \tag{3}$$

$$\Sigma_p = \frac{1}{L}\sum_{l=1}^{L} (F(\mathbf{x}, l) - \overline{\mathbf{x}_p})(F(\mathbf{x}, l) - \overline{\mathbf{x}_p})^T, \tag{4}$$

With these modifications, a virtual subspace can be synthesized by the algorithm shown in 2.2. Figure 4 illustrates a scheme of the virtual subspace construction. Because the synthesized images are made by the lighting compensation on the canonical eigenspace, the virtual subspace is never identical to the real subspace. The virtual subspace, however, is a reasonable approximation of the real subspace because the mapping is authorized in the average meaning.

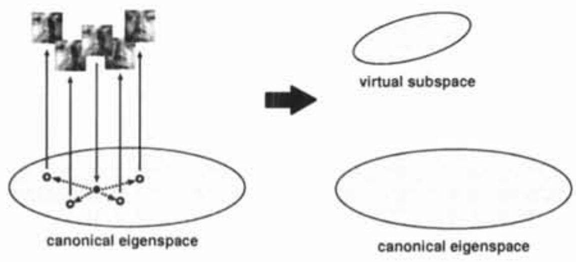


Figure 4: Virtual subspace construction.

4 Subsequent virtual subspace construction

4.1 Subspace construction from images

Let us discuss how to construct a virtual subspace when K images are taken for the p-th person. For the purpose, some more notations should be defined. Let $\mathbf{x}_k$ denote the k-th input image, $\mathbf{x}_{kl}^* = F^*(\lambda(\mathbf{x}_k^*), l)\mathbf{x}_k^*$ and

$$\kappa(K,l) = \arg\min_{1\le k\le K} DIFS(\mathbf{x}_{kl}^* - \mathbf{x}_k^*).$$

In these notations, the virtual image in the l-th lighting condition is defined by $F(\mathbf{x}_{\kappa(K,l)}, l)$. This definition is one of natural generalizations of $F(\mathbf{x}, l)$, and leads the definitions of $\overline{\mathbf{x}_p}$ and Σ_p as follows:

$$\overline{\mathbf{x}_p} = \frac{1}{L}\sum_{l=1}^{L} F(\mathbf{x}_{\kappa(K,l)}, l) \tag{5}$$

$$\Sigma_p = \frac{1}{L}\sum_{l=1}^{L}(F(\mathbf{x}_{\kappa(K,l)}, l) - \overline{\mathbf{x}_p})(F(\mathbf{x}_{\kappa(K,l)}, l) - \overline{\mathbf{x}_p})^T. \tag{6}$$

It is noted that Eqs. (5) and (6) are equivalent to Eqs. (3) and (4) when $K = 1$. Figure 5 illustrates the virtual subspace integration when two input images are registered. The residual term, $\Delta\mathbf{x}_{\kappa(K,l)}$, is not constant but optimally selected by $\kappa(K, l)$.

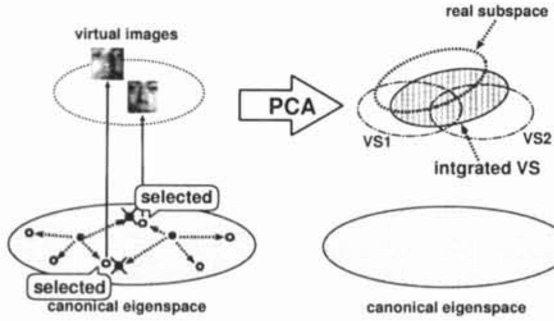


Figure 5: Integration of virtual subspaces.

Figure 6: Examples of subsequent construction of virtual subspace.

4.2 Examples of subsequent construction

An example of subsequent construction is shown in Fig. 6. In each row, the leftmost image shows an average image in the subspace. The rest three images show the three most principal components.

The upper 4 rows in Fig. 6 show virtual subspaces for the same person. The first two rows show virtual subspaces composed from a single image of a particular person. They are different because different input images are used in the two rows. The third row shows a virtual subspace composed from two input images which are used for the first two rows. The fourth row shows a virtual subspace after five input images are taken. The fifth and sixth rows show virtual subspaces of another person. They are composed from one image and five images, respectively.

These examples show that the virtual subspace gradually changes to the real subspace by additional input images. Both the average image and the principal components are updated from ones with the average properties to ones with the individual properties.

5 Experiments

5.1 Data specification

Data specification is summarized in Table 1.Facial images were taken from a fixed camera in the laboratory under natural lighting conditions. The 100 persons, looking forwards, were sitting on a chair in

Figure 7: Average faces under 24 lighting conditions.

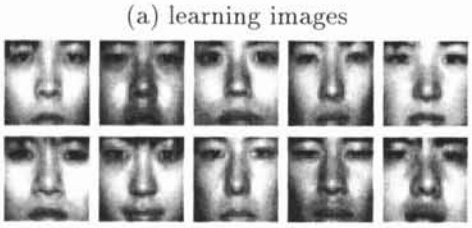
the fixed distance from the camera. The chair was fixed to get the frontal facial images for each person.

The canonical eigenspace is created from 1200 images of 50 persons under 24 lighting conditions. In the learning set, 9 persons are with glasses. Figure 7 shows 24 images of average of 50 person. The rest 50 persons are used as test data. 15 persons are with glasses in the test set. Figure 8 (a) shows 10 example images in the learning set under a particular lighting condition. Figure 8 (b) shows 10 example images in the test set under the same lighting condition.

Table 1: Data specification

	Learning samples	Test samples
Number of persons	50	50
Number of lighting conditions	24	24
Image size(face)	32×32	32×32
with glasses(persons)	9	15

(a) learning images



(b) test images

Figure 8: Examples of learning and test images.

5.2 Single image registration for each

For each person to be registered, only one image is randomly selected for the registration from 24 images of each person. Then the discrimination experiment has been accomplished over the rest 23 images of all the registered person. This process was repeated in a hundred times with randomly changing a registered image for each person. The discrimination rate with the virtual subspace method is 71.8 %, about 12 % more effective than the nearest neighbor method. Experimental results show that the virtual subspace method is effective even when only one image is registered for each person.

5.3 Subsequent image registration for each

For each registered person, a fixed number of images are randomly selected from 24 images of each person included in the data base. Then the discrimination experiment has been accomplished over the rest images of all the registered person. This process was repeated in a hundred times with changing registered images for each person.

Table 2 shows average discrimination rates. Five methods are compared for the same learning and test sets. In the table, NN indicates the nearest neighbor discrimination. CNN indicates the nearest neighbor with the lighting compensation. SS indicates the conventional subspace method, which is effective when $n > 1$ where n shows the number of registered images.

The rightmost two columns are based on the virtual subspace method. While VSNN indicates the nearest neighbor discrimination over the separate virtual subspaces for each person, SVS indicates the subsequent virtual subspace method. It should be noted that the recognition cost with SVS is about $1/n$ of VSNN, while the discrimination rate is a little higher whenever the number of images is 2 through 5. The subsequent subspace method seems reasonable when taking into account the calculation cost.

Figure 9 shows the discrimination rates for SVS with changing both the dimension of subspaces and the number of registered images. This figure shows that best dimension changes dependent on the number of registered images for each person.

From the data analysis, we have found that incorrect discrimination increases when a person is taken with glasses. For persons without glasses, the discrimination rates reach 83.0 % and 93.2 % when one and two images are registered for each person. This suggests that the face recognition could be more improved, if we could suppress the noise caused by the glasses.

Table 2: Discrimination rates for plural registered images[%] (Whole face)

Method	NN	CNN	SS	VSNN	SVS
Dimension	45	19	n-1	4	4
1 image	59.3	62.6	–	71.8	71.8
2 images	73.4	76.1	49.2	82.5	83.6
3 images	81.0	83.2	65.6	87.7	87.7
4 images	85.8	87.4	77.3	90.8	91.0
5 images	88.5	89.8	84.5	92.6	93.3

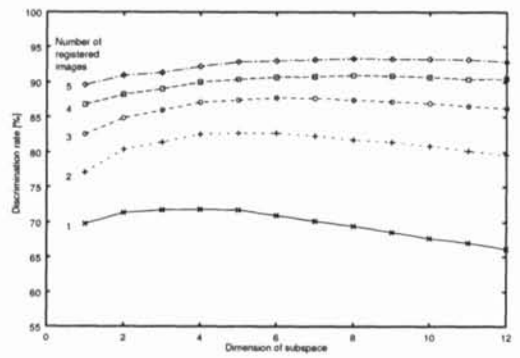


Figure 9: Discrimination rates for the virtual subspace method.

5.4 Discrimination with using four features

In the above experiments, only the whole faces are used for discrimination. Let us take into account four other features (left and right eyes, nose and mouth). The discrimination scheme is summarized as follows: Four independent feature spaces are created in the same way as mentioned in 2.2. For each feature $f(=1,2,3,4)$, the nearest person p_f^* is calculated.

$$p_f^* = \arg \min_{1 \le p \le P} DFFS_{f,p}(\mathbf{x}).$$

Let $f = 0$ indicate a case of the whole face. Then the final discrimination is accomplished due to

$$p^* = \arg \min_{1 \le p \le P} \sum_{f=0}^{4} \frac{DFFS_{f,p}(\mathbf{x})}{DFFS_{f,p_f^*}(\mathbf{x})}.$$

Table 3 shows average discrimination rates when the four features are used together with the whole face. In comparison with four other methods, SVS provides the best results. The discrimination rates are improved by combining four features. When only one image is registered for each person, the discrimination rate is 71.8% for a whole face, while it is improved to 75.1 % when the four features are combined. When two images are registered for each person, the rate is improved from 83.6 % to 87.6 % by combining the four features. When five images are registered for each person, the discrimination rate gets 97.0 %. The cumulative discrimination rate is 99.3 % when 5 nearest persons are selected.

Table 3: Discrimination rates for plural registered images[%] (Combined)

Method	NN	CNN	SS	VSNN	SVS
Dimension	45	19	n-1	4	4
1 image	63.6	67.5	–	75.1	75.1
2 images	78.7	79.4	50.1	86.6	87.6
3 images	86.1	86.1	66.7	91.9	92.5
4 images	90.4	89.9	78.6	94.2	95.6
5 images	93.1	92.5	86.4	96.3	97.0

6 Conclusions

A concept of virtual subspace is introduced to realize a robust face recognition independent of the lighting conditions. The virtual subspace is constructed from a single image, and gradually modified to the real subspace by the subsequent update. The virtual subspace method can be applied to a face recognition in the natural lighting condition, even if the lighting condition is unknown or changes from time to time.

Acknowledgment

This work was partly supported by Research for the Future Program, the Japan Society for the Promotion of Science (Project ID: JSPS-RFTF96P00501).

References

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

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

[3] Kirby, M. and Sirovich, L., "Application of the Karhunen-Loeve Procedure for the characterization of human faces", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, 1990.

[4] Adini, Y., Moses, Y. and Ulman, S., "Face recognition: The problem of compensating for changes in illumination direction," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 721-732, 1997.

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

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