Towards 3-Dimensional Pattern Recognition

Ken-ichi Maeda, Osamu Yamaguchi, and Kazuhiro Fukui

Corporate Research & Development Center, TOSHIBA Corporation
1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, 212-8582, Japan
{ken.maeda,osamu1.yamaguchi,kazuhiro.fukui}@toshiba.co.jp

Abstract. 3-dimensional pattern recognition requires the definition of a similarity measure between 3-dimensional patterns. We discuss how to match 3-dimensional patterns, which are represented by a set of images taken from multiple directions and approximately represented by subspaces. The proposed method is to calculate the canonical angles, in particular the third smallest angle between two subspaces. We demonstrate the viability of the proposed method by performing a pilot study of face recognition.

1   Introduction

We aim to extend pattern recognition coverage from 2-dimensional (2-D) applications to 3-dimensional (3-D). One expected application of such 3-D pattern recognition is differentiation between an object and its photograph as well as identification of the object.

In general, for pattern recognition, we define a similarity measure between two patterns and use this as the criterion for performing recognition. Thus, for our purpose, we need a definition of similarity between two 3-D patterns.

A set of images taken from multiple directions are used to describe a 3-D object[1] and the set is approximately represented by a subspace. A subspace representation of 3-D objects was introduced, for example, in the *parametric eigenspace representation* [1], although the subspace representation of a pattern set had been known as the *subspace method* [2]. In most cases, the *principal component analysis* (PCA) is used for making the subspace as an approximation of the distribution of the patterns in the set. However, the pattern matching method in [1] is limited to measure the nearest distance between an input represented by a vector and a reference that is the nearest vector in a class.

The largest problem with such a framework is that a single photograph which happens to be identical to an image that is captured from a particular direction matches exactly even though the actual objects are different. More concretely, it is impossible to differentiate between an object and its photograph using conventional methods.

---

[1] 3-D modeling provides a representation of 3-D objects, but it takes a long time to construct such models for real objects. It is therefore currently impractical to acquire real-time results with this approach.

We present a 3-D pattern matching method to solve this problem. The proposed method is an extension of a 2-D pattern matching method, the *mutual subspace method* (MSM) [3]. The extension is to calculate angles between two subspaces, in particular the third smallest one rather than the smallest one that is used in the MSM. We also show the results of a pilot study where faces and their photographs are dealt with.

2   Approach to the Task

2.1   Use of the Subspace Method Framework

It is desirable that the 3-D pattern matching method is an extension of a 2-D method. Among existing 2-D methods, we chose the subspace method framework [2] because most of its calculations are linear operations that are executable at high speed, in particular with the SIMD instructions of today's CPUs. The subspace method framework has been widely applied in various pattern recognition tasks, such as character, speech, and face recognition. However, if we try to apply the subspace method directly to the 3-D problem, a 3-D model is required: which is expensive.

Among the subspace methods, the MSM has the most interesting properties. It was initially developed in order to provide greater tolerance for handprinting deformation of Kanji in early 80's, and outperformed the classical subspace method in a Kanji recognition experiment [3]. It was later applied to face recognition, which is originally a form of 3-D recognition, and achieved satisfactory accuracy [4].

2.2   Review of the Mutual Subspace Method

The MSM[2] was the first method to utilize the angle between two subspaces for defining the similarity between an input and a reference[3], though they are originally two *sets* of vectors. Given subspaces, U and V, the angle between these is defined as the minimum angle between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, where $\boldsymbol{u} \in U$ and $\boldsymbol{v} \in V$, according to [7]. Let θ be the angle. Then θ is calculated as

$$\cos^2 \theta \doteq \sup_{\substack{\boldsymbol{u} \,\in\, U,\, \boldsymbol{v} \,\in\, V \\ \|\boldsymbol{u}\| \,\neq\, 0,\, \|\boldsymbol{v}\| \,\neq\, 0}} \frac{(\boldsymbol{u}, \boldsymbol{v})^2}{\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2}. \tag{1}$$

This is also used as the definition of the similarity of the subspaces, i.e. the sets of the vectors, in the MSM. In order to calculate the similarity, we apply the following theorem.

---

[2] Since the first MSM paper [3] is only in Japanese, we summarize the major topics in this subsection.

[3] The concept of using the angle between subspaces was also discussed in [5] and [6], but the aim was other than defining similarity between an input and a reference.

Theorem 1:
Let U and V be two subspaces and P and Q be orthogonal projection operators onto U and V, respectively. Then the angle between U and V is calculated as the maximum eigenvalue of PQP or QPQ [3] [8] [9]. Let μ and ν be the maximum eigenvalues of PQP and QPQ, i.e.

$$PQP\boldsymbol{x} = \mu\boldsymbol{x}, \tag{2}$$

and

$$QPQ\boldsymbol{x} = \nu\boldsymbol{x}, \tag{3}$$

respectively. Then

$$\cos^2\theta = \|QP\|^2, \tag{4}$$
$$= \|PQ\|^2, \tag{5}$$
$$= \mu, \tag{6}$$
$$= \nu, \tag{7}$$

where the norm of an operator A is defined as

$$\|A\| \doteq \sup_{\|\boldsymbol{z}\|\neq 0} \frac{\|A\boldsymbol{z}\|}{\|\boldsymbol{z}\|}. \tag{8}$$

According to this theorem, μ or ν itself can be used as the similarity in the sense that it is the angle between the two subspaces. However, as the eigenvalue calculation of PQP or QPQ is costly, the matrices basically being large, we actually calculate the eigenvalue of a smaller matrix X whose eigenvalues are identical.

Let $X = (x_{ij})$ be

$$x_{ij} = \sum_{m=1}^{M} (\boldsymbol{\psi}_i, \boldsymbol{\phi}_m)(\boldsymbol{\phi}_m, \boldsymbol{\psi}_j), \tag{9}$$

or

$$x_{ij} = \sum_{n=1}^{N} (\boldsymbol{\phi}_i, \boldsymbol{\psi}_n)(\boldsymbol{\psi}_n, \boldsymbol{\phi}_j), \tag{10}$$

where $\{\boldsymbol{\phi}_m\}_{m=1}^{M}$ and $\{\boldsymbol{\psi}_n\}_{n=1}^{N}$ are the bases of U and V, respectively. Then the eigenvalues of X are equal to those of PQP and QPQ [3].

In the sense that an input is represented by a subspace, the MSM is an extension of the classical subspace method where the input is represented by a vector: which is also a basis of a 1-D subspace.

2.3   Use of Other Canonical Angles

Canonical Angles in Previous Methods. The *canonical angles* are defined as the angles between two subspaces which are measured from the directions that are orthogonal to one another [9]. The angle calculated in the MSM is the

smallest one, calculated as the largest eigenvalue. It is sufficient for application to 2-D pattern matching, but there exists a serious problem for 3-D pattern matching, as described earlier.

The problem can be modeled as follows; if we imagine a case where two subspaces share a smaller subspace, e.g. two 2-D subspaces sharing a line, as shown in Fig. 1, the smallest angle is zero. In such a case, the two patterns will be evaluated as identical in the MSM framework.
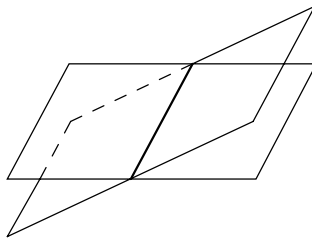

Fig. 1. Two 2-D subspaces in 3-D space.

As an alternative definition of distance between two subspace, Oja and Park-kinen proposed to use the smallest eigenvalue, $\mu_{\min}$, i.e. the largest canonical angle [5]. This method does not have the above problem, because it uses another angle of the shared line in Fig. 1.

Let P and Q be the projections to the subspaces U and V, respectively. Then the equation in [5] is

$$PQP\boldsymbol{x} = \mu_{\min}P\boldsymbol{x}, \tag{11}$$

where

$$\|P\boldsymbol{x}\| = 1. \tag{12}$$

As the eigenvectors of PQP are on the subspace U, (11) is equivalent to (2). But for the condition of (12), the smallest eigenvalue would be zero in most cases because it corresponds to a direction that is out of the subspaces, so it would not represent the canonical angle.

Canonical Angles for 3-D Recognition. We have found a possibility of further extension of the subspace method framework, using other angles[4], referring to the two previous methods. The methods adopted the opposite definitions of subspace distance; the smallest angle in the MSM and the largest angle in [5]. However we can have more canonical angles by solving the eigenvalue problem, (2) or

$$X\boldsymbol{y} = \mu\boldsymbol{y}. \tag{13}$$

Now the question is determining which one is suitable for 3-D recognition.

---

[4] A similar approach was presented in [10], but the meanings of the other angles and similarity were not clearly discussed.

In order to simplify the issue of 3-D recognition, we assume the following conditions, so that a linear approximation can be applied:

- The moving object is captured using a fixed camera.
- Motion primarily consists of small rotations.
- An appropriate normalization is performed.

These conditions ensure that variations in the recognition patterns are small.

When variation, e.g. lateral rotation, in the patterns is small, the distribution of the patterns caused by the variation can be assumed to be on a 2-D subspace, reflecting the dimension of the pattern itself and that of the variation. A different variation, e.g. vertical rotation[5], makes another 2-D subspace that shares a dimension of the pattern itself with the previous one.
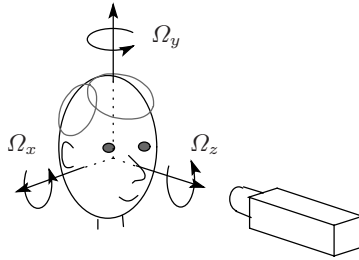

Fig. 2. Rotation of a 3-D object – a face –.

We should consider both lateral and vertical rotations, so that the rotations of a different direction or multiple directions can be described as a combination of these two rotations. Therefore a 3-D subspace is a good approximation of distributions due to rotation[6].

Given two 3-D objects that are identical, the 3-D subspaces for them should also be identical, or more precisely, the 3-D intersecting parts of two higher-dimensional subspaces should be identical. This means that the three smallest canonical angles should be zero. Since some actual objects like human faces have other variations, the canonical angles may have small values. However the three angles are still good measures of the similarity between the objects.

Referring to the definition of distance used in [5], we propose to use just the third smallest angle or the third largest eigenvalue among the three. The fourth largest eigenvalue should be nearly zero because we assume that the distribution is approximated by a 3-D subspace. The second largest eigenvalue can be 1 or near 1 if the photograph has a cylinder shape whose side view is the same as that of the object. This makes the third largest more appropriate.

---

[5] There is yet another rotation for 3-D objects; a rotation around the axis of the lens. However, considering the task of differentiating an object from its photograph, this the rotation is of no use, we take only Ω_x and Ω_y into account, and not Ω_z (see Fig. 2).

[6] Since the actual distribution lies in a higher-dimensional subspace due to many variations such as facial expressions, the input and reference subspaces should be represented in higher dimensions, e.g. in five or seven dimensions, shown empirically in [11].

The proposed method corresponds to calculating the largest canonical angle between two 3-D subspaces that are the closest part of the two original subspaces. Assume that $\{\mu_k\}_{k=1}^K$ are the eigenvalues of matrix X in (9) or (10) which are sorted in descending order. Then $\mu_3(= \cos^2\theta_3)$ is the answer. Since the calculation of the eigenvalues is in descending order and we need only the third largest one, the condition of (12) is no longer required.

3   A Pilot Study on Differentiating an Object from Its Photograph

We performed a pilot study on differentiating an object from its photograph in face recognition, which we consider to be a typical application in this paper (see Fig. 3). As described earlier, it is impossible to make this differentiation if we use a single image that is taken head-on.
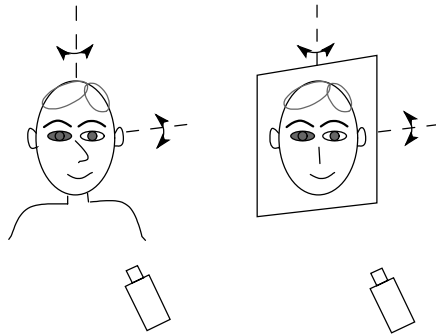

Fig. 3. Differentiation between an object – a face – and its photograph.

If we use the images of a 3-D object taken from multiple directions, some images may exhibit occlusions, or the lighting and shading conditions may be different among the images. Differently, all the images of a 2-D object are affine transformed ones of any single image. Even with the difference between these two situations, the largest eigenvalues of the two cases in the MSM framework are identical.

We experimented with using μ_3 as well as μ_1. Due to the normalization process [4], the variation in the normalized 3-D patterns is small as shown in Fig. 4. The experimental set therefore fulfills the conditions assumed in the last section.

We registered a set of face images (subject P0), and performed recognition for 11 people's faces including the registered person's and their photographs. Figure 5 shows an example set of face images, and Figure 6 shows an example set of photograph images of the same subject. Table 1 shows the experimental results for face inputs, and Table 2 shows the experimental results for photograph inputs.

Fig. 4. Normalized recognition area.
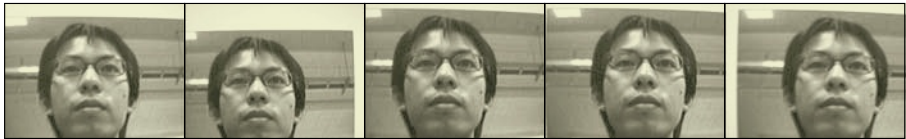

Fig. 5. Example set of face images.

Fig. 6. Example set of photograph images.

Table 1. Recognition results for face inputs.

Person	μ_1	μ_3
P0	0.989	0.937
P1	0.702	0.256
P2	0.707	0.520
P3	0.786	0.488
P4	0.701	0.457
P5	0.643	0.459
P6	0.730	0.227
P7	0.554	0.334
P8	0.750	0.557
P9	0.716	0.545
P10	0.772	0.435

Table 2. Recognition results for photograph inputs.

Person	μ_1	μ_3
P0	0.977	0.204
P1	0.591	0.165
P2	0.619	0.237
P3	0.741	0.123
P4	0.665	0.075
P5	0.626	0.124
P6	0.612	0.055
P7	0.678	0.238
P8	0.732	0.246
P9	0.600	0.154
P10	0.648	0.075

The experimental results show that μ_1 is not suitable for differentiating between the face and the photograph since the values for P0's face and photograph are 0.989 and 0.977, respectively. It is difficult to determine a threshold of rejection between such similar values.

Conversely, μ_3 makes a good criterion because the values for the face and the photograph are 0.937 and 0.204, respectively. As well as μ_1, μ_3 is also effective for rejecting other subjects (P1, ..., P10).

4   Conclusion

We have shown that the MSM framework can be extended to 3-D object recognition by using the third largest eigenvalue or the third smallest canonical angle. The viability of the proposed method has been demonstrated by a face recognition pilot study. Quantitative evaluation of this method with a large data set will be considered in future work.

Acknowledgments

References

1. Murase, H. and Nayar, S. K.: Illumination planning for object recognition in structured environments. *Proc. of CVPR* (1994) 31–38
2. Oja, E.: *Subspace Method of Pattern Recognition*, Research Studies Press (1983)
3. Maeda, K. and Watanabe, S.: A Pattern Matching Method with Local Structure *Trans. IECE* **J68-D 3** (1984) 345–352 (in Japanese)
4. Yamaguchi, O., Fukui, K., and Maeda, K.: Face Recognition Using Temporal Image Sequence. *Proc. of FG98* (1998) 318–323
5. Oja, E. and Parkkinen, J.: On Subspace Clustering. *Proc. of 7th ICPR* (1984) 692–695
6. Riittinen, H.: Short-cut Algorithms for the Learning Subspace Method. *Proc. of ICASSP 84* (1984) 17.2.1–17.2.4
7. Dixmier, M.: Etude sur les Varietes et les Operaterns de Julia, avec Quelques Applications. *Bull. Soc. Math. France* **77** (1949) 11–101
8. Björck, Å. and Golub, G. H.: Numerical Methods for Computing Angles between Linear Subspaces. *Mathematics of Computation* **27** (1975) 579–594
9. Chatelin, F.: *Valeurs Propres de Matrices*, Masson (1988)
10. Fukui, K. and Yamaguchi, O.: Face Recognition Using Multi-viewpoint Patterns for Robot Vision. *Proc. of ISRR03* (2003)
11. Epstein, R. et al: 5 ± 2 Eigenimages Suffice: An Empirical Investigation of Low-Dimensional Lighting Models. *Proc. of IEEE Workshop on Physics-based Modeling in CV* (1995) 108-116