Chapter 5
From the Subspace Methods to the Mutual Subspace Method

Ken-ichi Maeda

Abstract. The *Subspace Method* [25, 21] is a classic method of pattern recognition, and has been applied to various tasks. The *Mutual Subspace Method* [19] is an extension of the *Subspace Methods*, in which *canonical angles* (*principal angles*) between two subspaces are used to define *similarity* between two patterns (or two sets of patterns). The method is applied to face recognition and character recognition in Toshiba Corporation. The *Karhunen-Loève eigenvalue method* or *Principal Component Analysis* (*PCA*) [8, 13, 17] is a well-known approach to form a subspace that approximates a distribution of patterns, and it was introduced as a tool of pattern recognition [10, 24]. The extension from the *Subspace Methods* to the *Mutual Subspace Method* corresponds to the difference between *PCA* and *Canonical Correlation Analysis* (*CCA*) [9][1]. In this chapter, the *Mutual Subspace Method*, its mathematical foundations and its applications are described.

5.1 Introduction

Similarity definition is the most important factor for pattern recognition since no two objects are identical in the real worlds. Similarity can be defined in various ways, but here we take one based on the so-called *pattern matching*. The easiest example is *template matching*, e.g., as shown in an optical character reader (OCR) patent by Tauschek [22] (see Figure 5.1).

A more mathematical interpretation of such a simple matching is as follows: An image is a set of light intensity, $f(x,y)$, where (x,y) is a position. Sometimes we write $f(\boldsymbol{r})$, where $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j}$. We can appropriately quantise each position and divide whole area into I meshes. Thus an image is represented with a vector,

Ken-ichi Maeda
Corporate Research & Development Center, Toshiba Corporation, Japan
e-mail: ken.maeda@toshiba.co.jp

[1] A text book by Bishop [1] is recommended to learn about the difference between *PCA* and *CCA*, as well as to learn about more general statistical background of pattern recognition.

R. Cipolla, S. Battiato, & G.M. Farinella (Eds.): Computer Vision, SCI 285, pp. 135–156.
springerlink.com

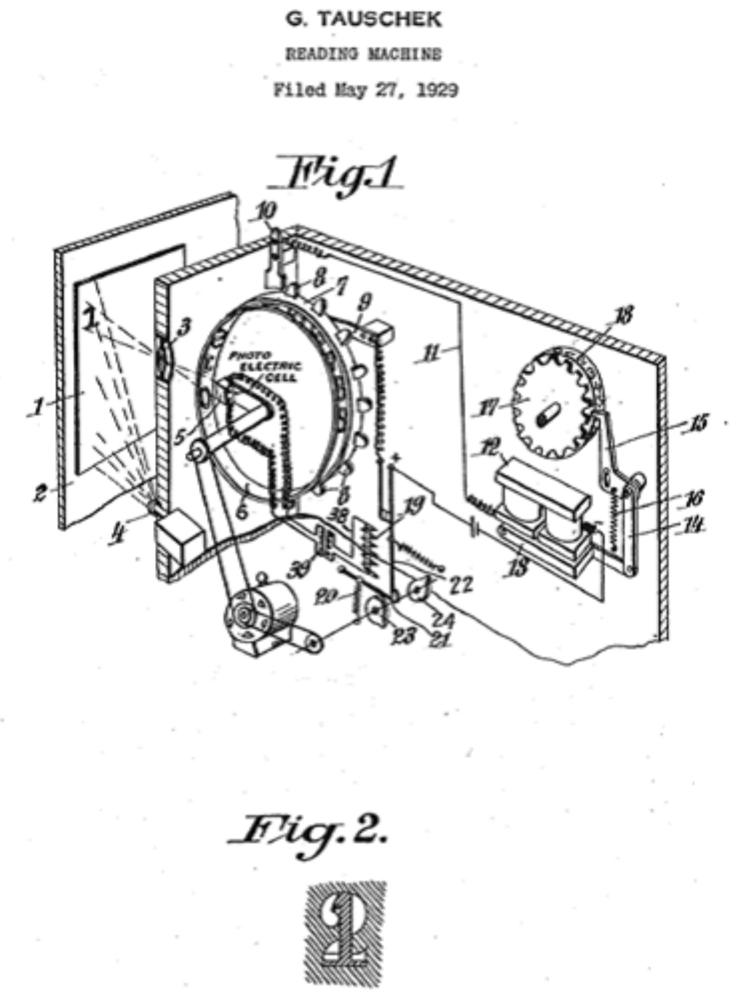


Fig. 5.1 An example of template matching: Tauschek's OCR patent (US Patent 2,026,329).

$$\boldsymbol{f} = (f_1, f_2, \cdots, f_i, \cdots, f_I)^t, \tag{5.1}$$

where f_i is the intensity of the i-th mesh (see Figure 5.2). We call this "*vector representation* of image pattern (or just pattern)."

There exist no two objects that are identical in the real world, (i.e., some mesh values should be different between any two patterns). Thereby a set of patterns which belong to the same class are distributed in a certain area.

If the illumination becomes a-times brighter, each mesh value of a pattern becomes a-times larger. Thus the area has a shape of cone, as shown in Figure 5.3.

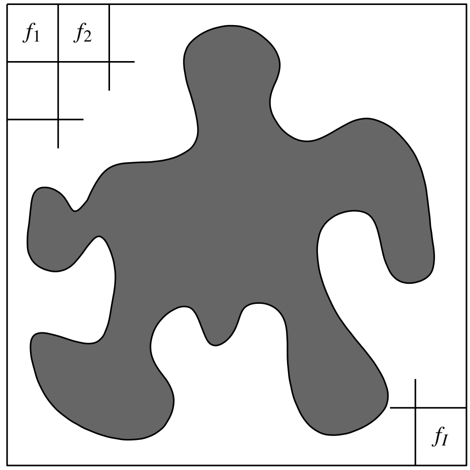


Fig. 5.2 Vector representation of image pattern: each mesh value is light intensity of the position.

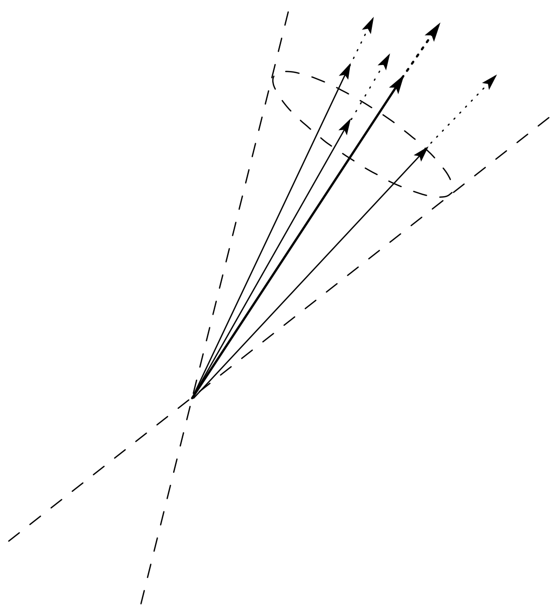


Fig. 5.3 Cone shape pattern distribution in a class.

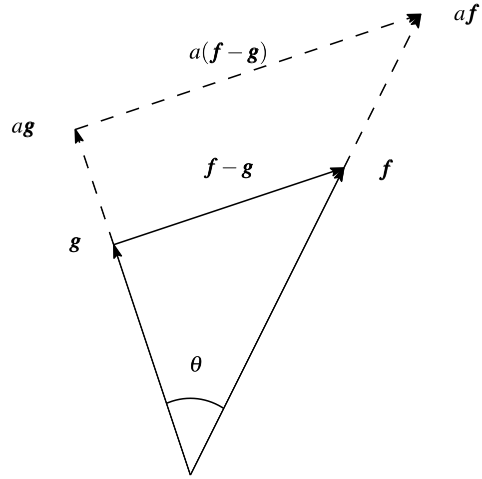
On one hand, the simplest way to define a similarity between two patterns is to use the distance between the two vectors that represent the patterns, as

$$\boldsymbol{d} = \boldsymbol{f} - \boldsymbol{g}\,. \tag{5.2}$$

However, if the distance is used, it directly reflects the change of illumination; if the illumination becomes a-times brighter, the distance is also a-times longer.

On the other hand, another way to define a similarity is to use the angle between the two vectors, as

Fig. 5.4 Comparison of similarity definitions: distance or angle.



$$\cos^2\theta = \frac{(\boldsymbol{f},\boldsymbol{g})^2}{(\|\boldsymbol{f}\|^2\,\|\boldsymbol{g}\|^2)}\,. \tag{5.3}$$

We call this "the *simple similarity* between $\boldsymbol{f}$ and $\boldsymbol{g}$." A merit of using the angle is that it is unchanged under the illumination change. See Figure 5.4 for comparison of these definitions.

However such a simple matching method has a problem caused by change in shape or position even though the amount of change is small. Even a single bit shift as shown in Figure 5.5 may make a serious change in similarity; the inner product of these two patterns is 0!

We have to introduce a more powerful matching method in order to solve the problem. The *Subspace Methods* and the *Mutual Subspace Method* were invented in order to define similarities which were more stable against pattern variations and more discriminative against similar classes.

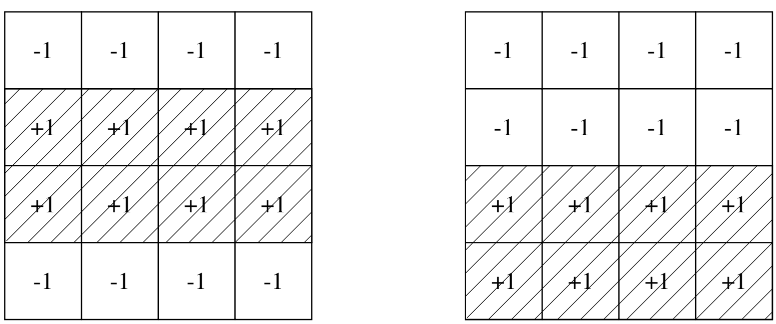


Fig. 5.5 An example of problem with simple matching: A single bit shift makes the inner product value 0.

5.2 The Subspace Methods

5.2.1 A Brief History of the Subspace Methods

A solution to the problem with *template matching* is to use the *Subspace Methods*[2]. The *Subspace Methods* are a kind of the *pattern matching method*, where a more sophisticated matching than that in *template matching* is employed, i.e., the *dictionary* or the reference pattern is represented with not a single pattern but a subspace. *Karhunen-Loève expansion* (*K-L expansion*), which is employed to make the subspace, was independently introduced by Iijima [10] and Watanabe [24] in pattern recognition at almost the same time[3]. The name of "the *Subspace Method*" was given by Watanabe et al. [25].

Iijima's method was originally called "the *Multiple Similarity Method* [11, 12]," and corresponding Watanabe's method was called "*CLAFIC* [25]." They are slightly different, but the process are almost the same:

- use a subspace as a dictionary of each class,
- collect a set of known patterns of each class and perform *K-L expansion* to make the dictionary[4] and
- calculate measures for all classes to classify an unknown pattern.

It is interesting that the *K-L method* itself was also independently invented by Karhunen [13] and Loève [17] for the research of stochastic process at almost the same time. A discrete form was proposed as *Principal Component Analysis* (*PCA*) by Hotelling [8] earlier than them.

The *K-L method* is occasionally employed also in order to reduce dimensionality; which is called "*K-L transformation*." Watanabe's corresponding method is called "*SELFIC* [24]." Almost the same method was revived by Turk and Pentland [23] for face recognition.

The *Subspace Methods of pattern recognition* have been investigated also by many researchers, such as Fu and Yu [5], Kittler and Young [15] and Kohonen [16]. There are a number of variations in the *Subspace Methods*, such as "*MOSS*," "the *Learning Subspace Method*," "the *Orthogonal Subspace Methods*," etc. For further detail, refer to Oja's text book. For those interested in history, refer to Grenander's book [7].

5.2.2 Basic Idea

As discussed before, no two patterns are identical in the real world. The question is what should be the representative of a class, or the reference pattern: We call it "a *dictionary* of the class." The simplest idea is to assign the mean vector as the dictionary. An alternative is to use all sample vectors. The former is too simple and the

[2] A good text book [21] is unfortunately out of print, but it may be found in a library.
[3] Iijima called it "*Mode Function Expansion* [10]."
[4] Iijima et al. [11] also showed another way to make a subspace using derivation.

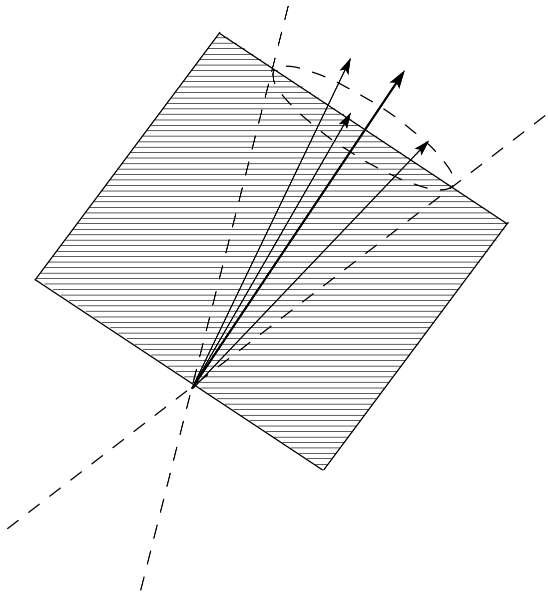


Fig. 5.6 Pattern distribution approximation by a subspace.

latter requires a large memory to store all vectors. Viewing the pattern distribution in a class shown in Figure 5.3, we find the cut plane is an ellipse. The ratio of the major axes and the minor axes of the ellipse is usually large.

If the pattern distribution has such a large ratio of ellipse shape, the distribution can be approximated by a plane (i.e., a subspace). An example of the approximation is shown in Figure 5.6. Instead of calculating the angle between an unknown input pattern and all known sample patterns, we can define a similarity as the angle between the unknown pattern and the subspace, as shown in Figure 5.7 and Equation (5.5).

$$S_{MS}^{(l)}[\boldsymbol{f}] \stackrel{\text{def}}{=} \cos^2\theta \tag{5.4}$$

$$= \sum_{m=1}^{M} \frac{(\boldsymbol{f}, \boldsymbol{\varphi}_m)^2}{\|\boldsymbol{f}\|^2}. \tag{5.5}$$

where $\boldsymbol{f}$ is an unknown input pattern and $\{\boldsymbol{\varphi}_m^{(l)}\}_{m=1}^{M}$ are an orthonormal basis of a subspace for class 'l.' We call this "the *multiple similarity* of $\boldsymbol{f}$ for class 'l.' "

The angle between an unknown input pattern and a dictionary subspace is the angle between the unknown pattern and the nearest vector in the subspace. Thereby a subspace dictionary is equivalent to preparing infinite number of sample patterns.

5.2.3 Subspace Construction

Now the problem is how to make the subspace, more practically how to find the orthonormal basis of the subspace. We follow Iijima's idea [10] for this purpose.

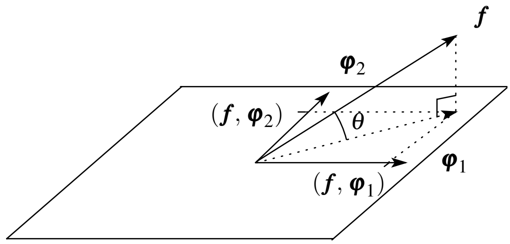


Fig. 5.7 Similarity definition by the angle between a pattern and a subspace.

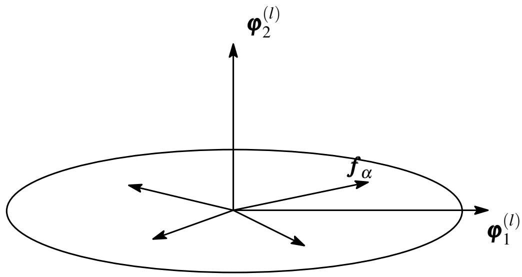


Fig. 5.8 Cut plane of the pattern distribution of class '*l*.'

Figure 5.8 shows the cut plane of the pattern distribution of class '*l*.' We may think the figure shows an orthogonal projection of the pattern distribution to the complementary subspace that is orthogonal to the standard pattern. The problem is interpreted as how to find $\{\boldsymbol{\varphi}_m^{(l)}\}_{m=1}^M$.

Let $\{\boldsymbol{f}_\alpha\}$ and $\{w_\alpha^{(l)}\}$ be a set of sample patterns and their probability of existence in a class '*l*,' respectively. Then the mean value of the *simple similarities* between a pattern $\boldsymbol{f}$ and the sample patterns $\boldsymbol{f}_\alpha$ is

$$S^{(l)}[\boldsymbol{f}] = \sum_\alpha w_\alpha^{(l)} \frac{(\boldsymbol{f}, \boldsymbol{f}_\alpha)^2}{\|\boldsymbol{f}\|^2 \, \|\boldsymbol{f}_\alpha\|^2} \tag{5.6}$$

$$= \sum_{m=1}^{M} \frac{\lambda_m^{(l)} (\boldsymbol{f}, \boldsymbol{\varphi}_m^{(l)})^2}{\|\boldsymbol{f}\|^2} . \tag{5.7}$$

where $\{\lambda_m^{(l)}\}_{m=1}^M$ and $\{\boldsymbol{\varphi}_m^{(l)}\}_{m=1}^M$ are eigenvalues and eigenvectors of the following $K^{(l)}$, respectively.

$$K^{(l)} = \sum_\alpha w_\alpha^{(l)} \frac{< \boldsymbol{f}_\alpha, \boldsymbol{f}_\alpha >}{\|\boldsymbol{f}_\alpha\|^2} \tag{5.8}$$

$$= \sum_{m=1}^{M} \lambda_m^{(l)} < \boldsymbol{\varphi}_m^{(l)}, \boldsymbol{\varphi}_m^{(l)} > . \tag{5.9}$$

where $< \bullet, \bullet >$ denotes dyad or Neumann-Schatten product.

Now the problem is to find the $\boldsymbol{\varphi}$ that maximises $S^{(l)}[\boldsymbol{\varphi}]$. The $\boldsymbol{\varphi}$ should satisfy $\delta S^{(l)} = 0$. Let $\boldsymbol{p}_\alpha$ be $\boldsymbol{f}_\alpha / \|\boldsymbol{f}_\alpha\|$. Then Equation (5.6) is

$$S^{(l)}[\boldsymbol{\varphi}] = \sum_{\alpha} w_{\alpha}^{(l)} \frac{(\boldsymbol{\varphi}, \boldsymbol{p}_{\alpha})^2}{\|\boldsymbol{\varphi}\|^2}. \tag{5.10}$$

Calculating variation of Equation (5.10),

$$\|\boldsymbol{\varphi}\|^2 \delta S^{(l)} + 2S^{(l)}(\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}) = 2\sum_{\alpha} w_{\alpha}^{(l)} (\boldsymbol{\varphi}, \boldsymbol{p}_{\alpha})(\boldsymbol{p}_{\alpha}, \delta\boldsymbol{\varphi}) \tag{5.11}$$

$$= 2(K^{(l)}\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}). \tag{5.12}$$

Since local maxima should satisfy $\delta S^{(l)} = 0$,

$$(S^{(l)}\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}) = (K^{(l)}\boldsymbol{\varphi}, \delta\boldsymbol{\varphi}), \tag{5.13}$$

for any $\delta\boldsymbol{\varphi}$. Thus we have an eigenvalue problem,

$$S^{(l)}\boldsymbol{\varphi} = K^{(l)}\boldsymbol{\varphi}, \tag{5.14}$$

which is known as the *Karhunen-Loève eigenvalue method* or *PCA*.

5.3 The Mutual Subspace Method

5.3.1 Basic Idea

The reference patterns are represented with subspaces in the *Subspace Methods*. Thinking about the difference between the *simple similarity* and the *multiple similarity*, it is dissymmetric to replace just one of two vectors by a subspace. What about representing both the input and reference patterns with subspaces?

The use of subspaces in the *Subspace Methods* improved the stability of similarity against the changes of pattern positions or shapes. If both the input and reference patterns are represented with the subspaces, we may expect more stability against the changes.

Before extending the definition of similarity from the angle between a vector and a subspace to that between two subspaces, we should define what the angle between two subspaces is. Given the 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 Dixmier [4]. See Figure 5.9 for a more concrete image. Let θ be the angle. Then

$$\cos^2\theta \stackrel{\text{def}}{=} \sup_{\substack{\boldsymbol{u}\in U,\, \boldsymbol{v}\in V \\ \|\boldsymbol{u}\|\neq 0,\, \|\boldsymbol{v}\|\neq 0}} \frac{|(\boldsymbol{u}, \boldsymbol{v})|^2}{\|\boldsymbol{u}\|^2\,\|\boldsymbol{v}\|^2}. \tag{5.15}$$

We use this value also as the definition of similarity. For actual calculation of the similarity, we apply the following theorem.

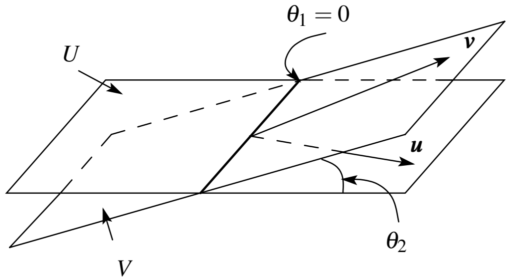


Fig. 5.9 Angle between two subspaces: Compare to angle between a vector and a subspace shown in Fig. 5.7

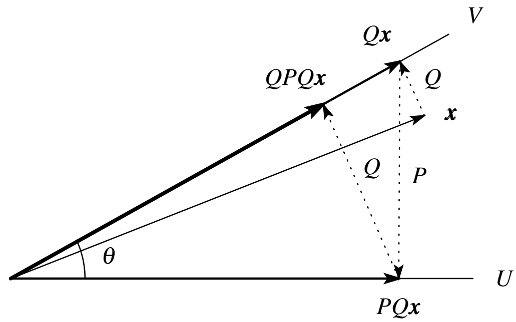


Fig. 5.10 How to calculate the angle between two subspaces: The length of $QPQ\boldsymbol{x}$ is $\cos^2\theta$ times the length of $Q\boldsymbol{x}$.

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 [19,2,3]. Let μ and ν be the maximum eigenvalues of PQP and QPQ, i.e.,

$$PQP\boldsymbol{x} = \mu\boldsymbol{x}, \tag{5.16}$$

and

$$QPQ\boldsymbol{x} = \nu\boldsymbol{x}, \tag{5.17}$$

respectively. Then

$$\begin{aligned} \cos^2\theta &= \|QP\|^2 && (5.18)\\ &= \|PQ\|^2 && (5.19)\\ &= \mu && (5.20)\\ &= \nu, && (5.21) \end{aligned}$$

where the norm of an operator A is defined as

$$\|A\| \stackrel{\text{def}}{=} \sup_{\|\boldsymbol{z}\|\neq 0} \frac{\|A\boldsymbol{z}\|}{\|\boldsymbol{z}\|}. \tag{5.22}$$

(see Figure 5.10 for intuitive understanding).

According to this theorem, we may use the eigenvalue, μ or ν, as the similarity instead of the angle, $\cos^2\theta$. However, as the eigenvalue calculation of PQP or QPQ is costly due to the large size of the matrices, we practically translate the eigenvalue problem into that of a smaller matrix X whose eigenvalues are identical to PQP or QPQ,

$$X\boldsymbol{z} = \lambda\boldsymbol{z}, \tag{5.23}$$

$$X = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{in} \\ \vdots & & \vdots & & \vdots \\ x_{m1} & \dots & x_{mj} & \dots & x_{mn} \end{pmatrix}, \tag{5.24}$$

where

$$x_{ij} = \sum_{m=1}^{M} (\boldsymbol{\psi}_i, \boldsymbol{\varphi}_m)(\boldsymbol{\varphi}_m, \boldsymbol{\psi}_j), \tag{5.25}$$

or

$$x_{ij} = \sum_{n=1}^{N} (\boldsymbol{\varphi}_i, \boldsymbol{\psi}_n)(\boldsymbol{\psi}_n, \boldsymbol{\varphi}_j), \tag{5.26}$$

where $\{\boldsymbol{\varphi}_m\}_{m=1}^{M}$ are a set of basis of U whose dimension is M, and $\{\boldsymbol{\psi}_n\}_{n=1}^{N}$ are of V whose dimension is N.

5.3.2 Application to Chinese Character Recognition

The *Mutual Subspace Method* was originally developed for Chinese Character (Kanji) recognition, in particular hand-printed Kanji recognition. Examples of hand-printing are shown in Figure 5.11.

The problems of hand-printed Kanji recognition are:

- many classes (more than 1,000, maybe 200,000),
- hand-printing variations (such as shown in Figure 5.11) and
- existence of similar classes (such as shown in Figure 5.11).

The existence of many classed makes it difficult to collect the enough number of sample patterns for making the dictionaries using the *K-L method*[5]. The *Mutual Subspace Method* was mainly a solution to this problem; we expected a more robust similarity against the variations without collecting so many patterns.

The next step is to make input subspace. If we assume that the major variations of patterns are only position shift in x and y directions and that the amount of shift is small, we can approximate the variations using derivations in x and y directions. We

[5] We empirically know that the number of required training patterns is the cube of the subspace dimension. If we make 1,000 10-dimensional subspaces, we need 1,000,000 patterns for training.

Fig. 5.11 Examples of hand-printed Kanji: 3 different classes that mean 'know,' 'weave' and 'fiber.'

can make a 3-dimensional subspace using original and the differentiated patterns in x and y directions[6]. Thus we can make a subspace without the *K-L method*. We call it "the *differential method* of dictionary."

It is a good point to think of a *blurring filter*. The *blurring filter* is effective to make matching easier because it reduces high frequency noises. The most common blurring filter is made with the *Gaussian function*. Let $G(\boldsymbol{r}, \sigma)$ be a *Gaussian function*,

$$G(\boldsymbol{r}, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{\|\boldsymbol{r}\|^2}{2\sigma^2}), \tag{5.27}$$

and $f(\boldsymbol{r})$ be a pattern. Then the *blurring process* is as follows,

$$f(\boldsymbol{r}_i, \sigma) = \int G(\boldsymbol{r}_i - \boldsymbol{r}, \sigma)\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}. \tag{5.28}$$

However, since there are mutually similar classes as discussed above, simple blurring may cause degradation of the recognition accuracy. In order to intensify small differences among the mutually similar classes, Maeda et al. [18] introduced order m derivation in x direction and order n in y direction,

[6] We need the Gram-Schmidt procedure in order to make the orthonormal basis of the subspace.

$$\frac{\partial^{m+n}}{\partial x^m \, \partial y^n} f(\boldsymbol{r}, \sigma) = \int \frac{\partial^{m+n}}{\partial x^m \, \partial y^n} \, G(\boldsymbol{r} - \boldsymbol{r}', \sigma) \, f(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}' \tag{5.29}$$

$$= \frac{\sqrt{m!n!}}{(-\sqrt{2})^{m+n}} \, f_{mn}(\boldsymbol{r}, \sigma) \, , \tag{5.30}$$

where

$$f_{mn}(\boldsymbol{r}, \sigma) = \int \frac{1}{\sqrt{m!n!} \, (\frac{\sigma}{\sqrt{2}})^{m+n}} \, G(\boldsymbol{r} - \boldsymbol{r}', \sigma) \, H_m(\frac{x - x'}{\sigma}) \, H_n(\frac{y - y'}{\sigma}) \, f(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}' \, . \tag{5.31}$$

H_m is the *Hermite Polynomial* of order m. – Isn't this process the same as the *differential method* of dictionary?

The integral kernel of Equation (5.31) is the *Gaussian weighted Hermite Polynomials*. We call this "the *Gauss-Hermite Kernel*." The 1-dimensional shapes are shown in Figure 5.12.

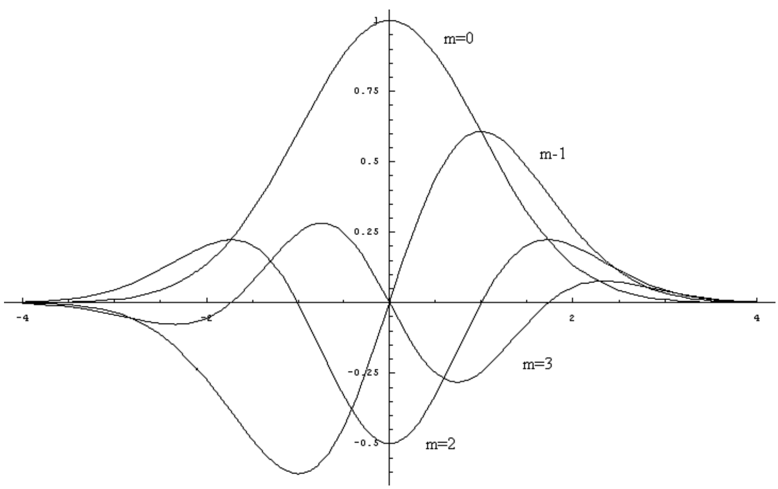


Fig. 5.12 1-dimensional shapes of the *Gauss-Hermite Kernel*: The shapes resemble the *Gabor Functions*.

Since *Hermite Polynomials* are mutually orthogonal with a Gaussian weight, we can have a set of orthonormal basis of a subspace which represents an input pattern[7].

The experimental results on hand-printed Kanji recognition, just with 500 patterns for each class shown in Figure 5.11, using the *simple similarity* (*SS*), the *Subspace Method* (*SM*) and the *Mutual Subspace Method* (*MSM*) showed the effectiveness of the proposed method (see Table 5.1).

[7] In most cases, we also need the *Gram-Schmidt* procedure since we usually define the inner product with no weight.

Table 5.1 Hand-printed Kanji recognition result for the three mutually similar classes: 500 patterns each.

Method	*SS*	*SM*	*MSM*
Recognition Rate	75.9 %	80.2 %	82.0 %

In *SS* and *MSM* only the mean pattern of a class was used to make the dictionary of the class whereas all patterns were used in *SM*. In *MSM* the x and y derivatives of the mean pattern are also used to construct a subspace instead of *K-L expansion*. Even though only the mean pattern was used in the *MSM*, its recognition rate was the best.

5.3.3 Application to Face Recognition

Face recognition is an easy-to-use authentication method. A problem of this method is its relatively low accuracy in comparison with other biometric methods, such as finger print recognition, iris recognition and vein recognition. However it still has several advantages over others, such as remote sensing, unrealisedness and many objects at a time.

Difficulties in face recognition are mainly caused by

- change of face directions,
- non-rigid objects,
- aging of objects and
- change of illuminations.

The first two items make relatively short-term variation whereas the last two are long-term.

We focus on the first two problems and show a way to overcome them: The most important point is how to define a similarity that is robust against the variation of input pattern. The input patterns vary according to a probabilistic distribution caused by change of face directions and facial expressions. If we use the moving picture, or an image sequence, as the input instead of a single photograph, we can make a subspace, which approximates the distribution. Thus we can apply the *Mutual Subspace Method*, with which we may expect a more stable similarity against the changes of facial directions and facial expressions. Figure 5.13 shows the conceptual scheme of the *Mutual Subspace Method* applied to face recognition.

We made experiments with 100 persons using the *Subspace Method* and the *Mutual Subspace Method*. The experimental results are shown in the ROC curves in Figure 5.14. The x-axis is the false acceptance rate (FAR) and the y-axis is the false rejection rate (FRR). The equal error rate (EER) of the *Mutual Subspace Method* was reduced to $1/4$ of that of the *Subspace Method*.

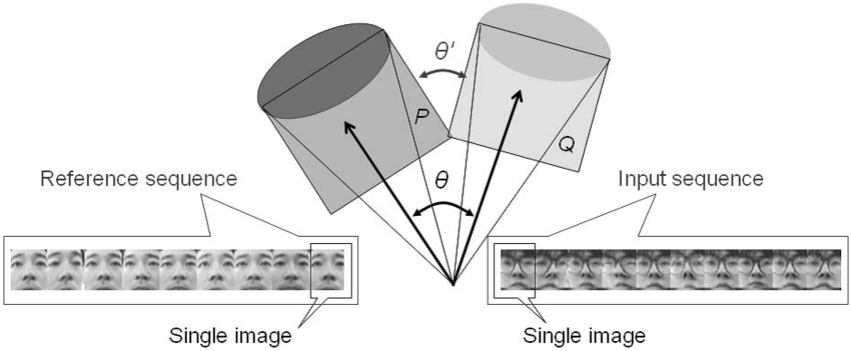


Fig. 5.13 Conceptual scheme of the *Mutual Subspace Method* applied to face recognition: The similarity is defined as $\cos^2\theta'$, where θ' is the canonical angle of two subspaces, instead of $\cos^2\theta$, where θ is the angle between two single images.

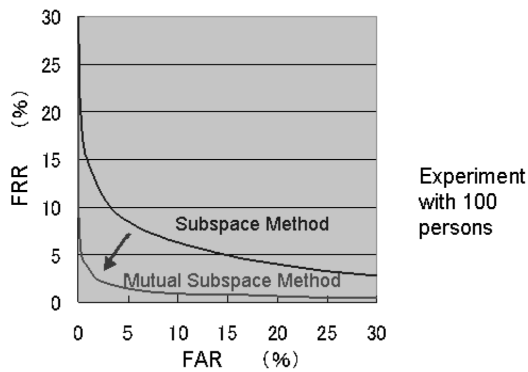


Fig. 5.14 Experimental results of face recognition using the *Subspace Method* and the *Mutual Subspace Method*: FAR-FRR ROC curves. EER was reduced to 1/4.

5.3.4 Application to 3-D Face Recognition

Another problem of face recognition is potential deception using photograph. The *Mutual Subspace Method* discussed above uses only the smallest canonical angle to define a similarity. Since there is no remarkable difference between a frontal face and a photograph of the same face, their vectors are almost identical; which means even if moving pictures are used, the smallest canonical angle between photograph input and face dictionary of the same person is almost 0. Figures 5.15 and 5.16 show the picture sequences of a moving face and a moving photograph, respectively. The leftmost of the face picture sequence is the frontal face, which is almost the same as that of the photograph.

A more mathematical discussion is as follows: Let U and V be the subspaces representing a face and its photograph as shown in Figure 5.9. Then the both frontal patterns are on the line of the intersection of the two subspaces, so the angle between them are almost 0. However, in this case, there is another angle, θ_2. We expect that

Fig. 5.15 Picture sequence of a moving face.

Fig. 5.16 Picture sequence of a moving photograph.

the angle reflects the difference between face, which is a 3-dimensional (3-D) object, and its photograph, which is 2-dimensional (2-D).

Maeda et al. [19] proposed to use the third smallest angle. Table 5.2 shows the result of a pilot study, in which only the similarities defined with the smallest and the third smallest canonical angles were measured and the reference pattern was made only for the person, P0.

Table 5.2 Similarity using the smallest and the 3rd smallest canonical angles (modified from Maeda et al. [19]).

	Smallest CA		3rd smallest CA	
Person ID	Face	Photo	Face	Photo
P0	*0.989*	*0.977*	*0.937*	*0.204*
P1	0.702	0.591	0.256	0.165
P2	0.707	0.619	0.520	0.237
P3	0.786	0.741	0.488	0.123
P4	0.701	0.665	0.457	0.075
P5	0.643	0.626	0.459	0.124
P6	0.730	0.612	0.227	0.055
P7	0.554	0.678	0.334	0.238
P8	0.750	0.732	0.557	0.246
P9	0.716	0.600	0.545	0.154
P10	0.772	0.648	0.435	0.075

According to the result, the similarity using the smallest canonical angle does not have capability to differentiate face and its photograph because the similarity value for the photograph is 0.977 where as that for the face is 0.989; the difference is quite small. However the similarity values using the third smallest canonical angle have enough difference for this purpose; 0.204 and 0.937, respectively.

Now the problem is to determine the angle or angles that is the best for differentiating a face and its photograph with enough authentication accuracy. We first should find the dimension of the subspace that approximates an input pattern distribution. Assuming that a face is a rigid object and that we employ some appropriate position and size normalisation, the major variations of the input are caused by face rotation.

Figures 5.17 and 5.18 show rotations of a face and a face photograph. Among these rotations, we focus on Ω_x and Ω_y because Ω_z is not useful for differentiating

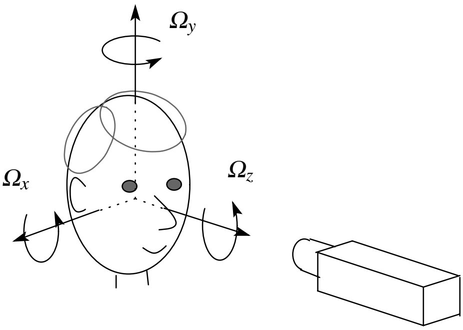


Fig. 5.17 Rotation of a face.

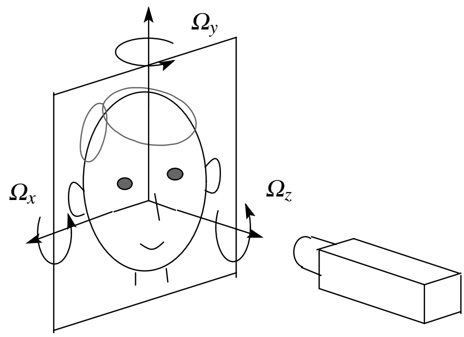


Fig. 5.18 Rotation of a face photograph.

a face and a photograph. Consequently we may suppose that the dimension of the subspace is 3; frontal face, its vertical rotation and horizontal rotation.[8]

Figure 5.19 shows the shape of the manifold (depicted by ⋄ and +), on which rotated face vectors exist, and the approximation with a 3-D pyramid, which leads to a 3-D subspace. For intuitive understanding, Figure 5.20 shows the case we consider only vertical rotation. In this case the approximation is a 2-D subspace, and it appears more like a *subsapce*. Now that we understand the variation distribution of the face input is approximated with a 3-dimensional subspace, we still wonder if the third smallest angle is the best; what about using the second or the fourth smallest, what about using the angles in combination instead of the single angle on its own? Maeda et al. [20] made experiments to find the difference between the second and the third smallest angles, using convex curved photographs (see Figure 5.21).

A convex curved photograph appears more like the actual face than a flat photograph, and it is an easy means if it works for deception. Also the mean and the product of the similarities made with up to the third smallest angle and the product of the similarities were used in the experiments.

Figures 5.22, 5.23, 5.24 and 5.25 show the FRR and FAR curves using the similarities defined as the $\cos^2$ of the smallest, the second smallest, the third smallest and the fourth smallest angles on their own. Figures 5.26 and 5.27 show the results

[8] We also assume the amount of rotation is small, so that the linear approximation can be applied.

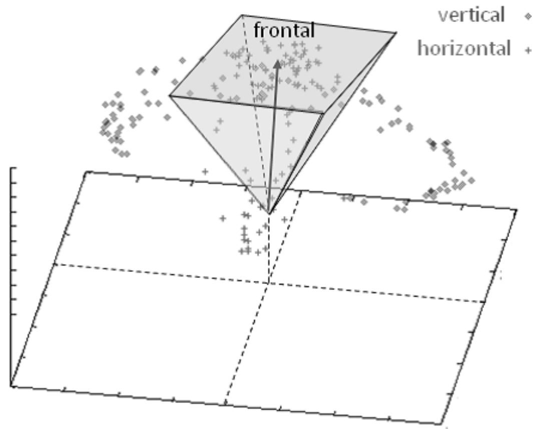


Fig. 5.19 The manifold of vertical and horizontal face rotation and its approximation with a 3-D pyramid.

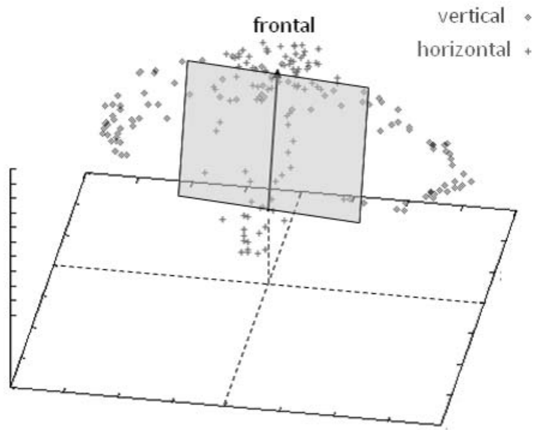


Fig. 5.20 The approximation with a 2-D subspace considering only vertical rotation.

using the mean of the similarities up to the third and the product of the similarities up to the third.

We can confirm that the smallest canonical angle is not effective to differentiate a face and its photograph since the Self FRR and Self Photograph FAR curves cross (see Figure 5.22). Also the Self FRR and Self Convex FAR curves cross. The fourth smallest canonical angle is not effective to identify a face since the Self Face FRR and Others Face FAR curves cross (see Figure 5.25). As a result, we should use the second and/or the third smallest canonical angle, at least.

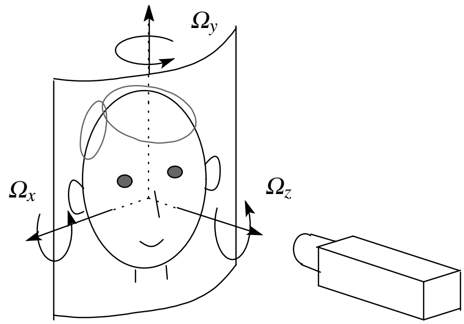


Fig. 5.21 Rotation of a convex curved photograph.

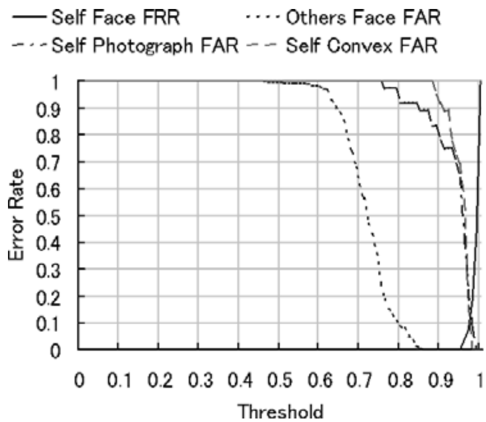


Fig. 5.22 FRR and FAR using the smallest canonical angle.

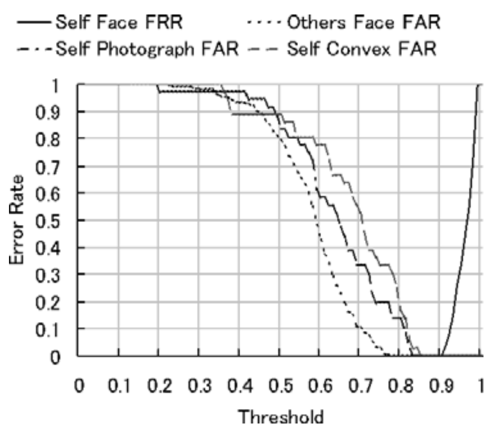


Fig. 5.23 FRR and FAR using the 2nd smallest canonical angle.

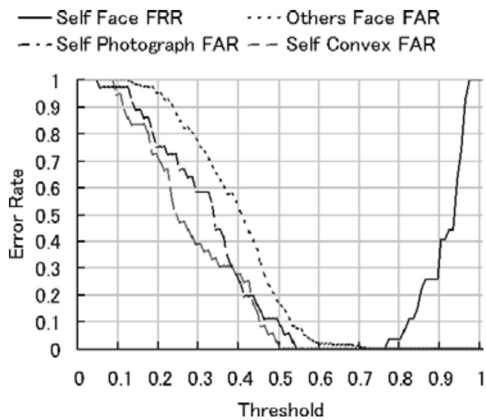


Fig. 5.24 FRR and FAR using the 3rd smallest canonical angle.

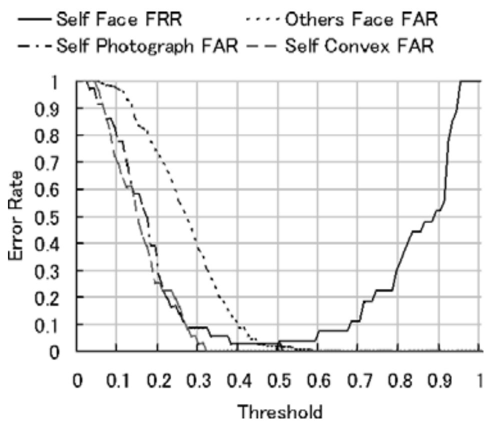


Fig. 5.25 FRR and FAR using the 4th smallest canonical angle.

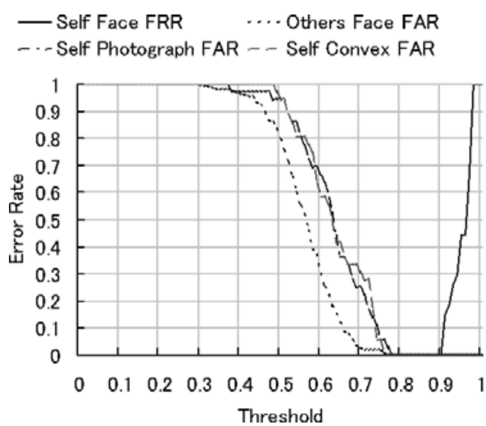
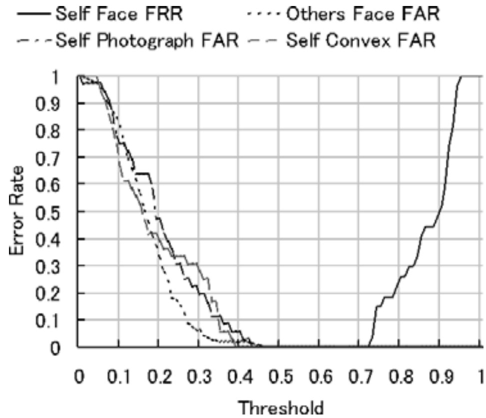


Fig. 5.26 FRR and FAR using the mean of the similarities.

Fig. 5.27 FRR and FAR using the product of the similarities.



Comparing Figures 5.23 and 5.24, the second smallest canonical angle is better at differentiating a person and others whereas the third is better at differentiating a face and its photograph, in particular for convex curved photograph; there is a trade-off between authentication and anti-deception.

The best one among all of the results is the product of the similarities up to the third canonical angle since the gap between the Self FRR and FARs are the widest among all (see Figure 5.27).

5.4 Conclusion

We have discussed a series of pattern matching methods that are based on subspaces. The methods have a long history started with mathematics and physics; the origins were 1930s and 1940s. Even after the *Subspace Methods* were introduced to pattern recognition, more than 40 years have already passed. Since the basic ideas were quite reasonable, the methods have survived for such a long time and have evolved against increasing difficulties in pattern recognition tasks. The *Mutual Subspace Method* is one of such evolutions.

We had to introduce some new mechanisms to overcome the difficulties in more complex recognition objects. For instance, we introduced a new feature extraction based on the *Gauss-Hermite Kernel* for hand-printed Kanji recognition, and we introduced multiple canonical angles for 3-D face recognition. By virtue of the newly introduced mechanisms, the subspace based matching methods have effectively worked and have achieved their aims. However, through the evolution, several basic concepts are common, such as assuming cone shape distributions, using subspaces and using angles for similarity definitions, etc.

We started this chapter with quickly following the history and viewed the evolution from the original *Subspace Method* to the *Mutual Subspace Method*. The evolution has not yet reached the final stage, and hopefully will continue. There already exist further extensions of the *Mutual Subspace Method*, such as the *Constrained Mutual Subspace Method* [6] and *Tensor Canonical Correlation* [14]. We expect further evolution for more difficult applications in the future.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
2. Björck, Å., Golub, G.H.: Numerical Methods for Computing Angles between Linear Subspaces. Mathematics of Computation 27, 579–594 (1975)
3. Chatelin, F.: Valeurs Propres de Matrices, Masson, Paris (1988)
4. Dixmier, M.: Etude sur les Varietes et les Operaterns de Julia, avec Quelques Applications. Bull. Soc. Math. France 77, 11–101 (1949)
5. Fu, K.S., Yu, T.S.: Statistical Pattern Classification Using Contextual Information. John Wiley and Sons, New York (1980)
6. Fukui, K., Yamaguchi, O.: Face Recognition using multi-viewpoint patterns for robot vision. In: Proceedings of the 11th International Symposium of Robotics Research (ISRR 2003), pp. 192–201 (2003)
7. Grenander, U.: Abstract Inference. John Wiley, New York (1981)
8. Hotelling, H.: Analysis of a Complex Statistical Variables into Principal Components. J. Educ. Psych. 24, 417–441, 498–520 (1933)
9. Hotelling, H.: Relations between Two Sets of Variates. Biometrica 28, 321–377 (1936)
10. Iijima, T.: Basic Theory on Feature Extraction of Visual Pattern. J. IEICE 46, 1714 (1963)
11. Iijima, T., Genchi, H., Mori, K.: A Theoretical Study of Pattern Recognition by Matching Method. In: Proceedings of the First USA-Japan Computer Conference, pp. 42–48 (1972)
12. Iijima, T., Genchi, H., Mori, K.: A Theory of Character Recognition by Matching Method. In: Proceedings of the 1st International Conference on Pattern Recognition, pp. 50–56 (1973)
13. Karhunen, K.: Zur Spektraltheorie stochastischer Prozesse. Ann. Acad. Sci. Fennicae 34 (1946)
14. Kim, T., Wong, S., Cipolla, R.: Tensor Canonical Correlation Analysis for Action Classification. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition (2007)
15. Kittler, J., Young, P.C.: A New Approach to Feature Selection Based on the Karhunen-Loeve Expansion. Pattern Recognition 5(4), 335–352 (1973)
16. Kohonen, T.: Associative Memory – a System-Theoretical Approach. Springer, Heidelberg (1977)
17. Loève, M.: Functions alèatoires du second ordre. Processus stochastique et mouvement Brownien (Lèvy, P.), 366–420, Gauthier-Villars, Paris (1948)
18. Maeda, K., Kurosawa, Y., Asada, H., Sakai, K., Watanabe, S.: Handprinted Kanji Recognition by Pattern Matching Method. In: Proceedings of the International Conference on Pattern Recognition, pp. 789–792 (1982)
19. Maeda, K., Yamaguchi, O., Fukui, K.: Towards 3-Dimensional Pattern Recognition. In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) SSPR&SPR 2004. LNCS, vol. 3138, pp. 1061–1068. Springer, Heidelberg (2004)
20. Maeda, K., Yamaguchi, O., Fukui, K.: A Fundamental Discussion on 3-Dimensional Pattern Recognition Using Canonical Angles between Subspaces for the Purpose of Differentiating Face and its Photograph. Systems and Computers in Japan 38, 11–20 (2007)

21. Oja, E.: Subspace Methods of Pattern Recognition. Research Studies Press, Letchworth (1983)
22. Tauschek, G.: Reading Machine, US Patent 2,026,329 (1935)
23. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. In Proceedings of the International Conference on Compute Vision and Pattern Recognition, 453–458 (1993)
24. Watanabe, S.: Karhunen-Loève Expansion and Factor Analysis. In: Trans. 4th Prague Conf. on Inf. Theory, Stat. Decision Functions, and Random Proc., pp. 635–660 (1965)
25. Watanabe, S., Lambert, P.F., Kulikowsky, C.A., Buxton, J.L., Walker, R.: Evaluation and Selection of Variables in Pattern Recognition. In: Tou, J. (ed.) Computer and Information Science II, pp. 91–122. Academic Press, New York (1967)