MVA'98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

13—19

Hierarchical Dictionary Constructing Method for the Parametric Eigenspace Method

Toru Abe *
School of Information Science
Japan Advanced Institute of Science and Technology, Hokuriku

Tomohiko Nakamura†
NTT Software Laboratories
Nippon Telegraph and Telephone Corp.

Abstract

The parametric eigenspace method is an object recognition method based on visual learning approach with image coding technique. In this paper, to improve matching efficiency, a novel approach to construct hierarchical dictionary for the parametric eigenspace method is proposed. In the proposed constructing method, learning image set is classified hierarchically, and tree-structured dictionary is constructed from the classification results. For the learning image set obtained by varying one parameter value, the proposed method determines the class boundaries that are optimal in the sense of discrimination and least squares. Experimental results show that the tree-structured dictionaries made by the proposed method improve matching efficiency without a decrease in recognition accuracy.

1 Introduction

The parametric eigenspace method[1] is an object recognition method based on visual learning approach with image coding technique. The eigenspace method[2, 3] uses principal component analysis to compress an image data into a low-dimensional feature vector, and represent the image as a point in a low-dimensional space i.e. the eigenspace. In the parametric eigenspace method, for each object of interest, a learning image set obtained by varying parameter value (e.g. pose, illumination, etc.) is represented as a parametric manifold in the eigenspace, and this manifold is used as a dictionary for the recognition. Given an unknown input image, the recognition system projects the input image to the eigenspace, and recognizes the object in the input image by matching with a point in the dictionary.

In comparison with other visual learning methods, the parametric eigenspace method makes it possible for the recognition system to reduce the dictionary size and the matching process time. Furthermore, if the dictionary is constructed from the learning images that were obtained by varying parameter value continuously, this method is capable of estimating the parameter value from an input image.

Generally, in a visual learning method, every image corresponded to all appearances of the object should be registered in the dictionary, and consequently an input image must be compared with all learning images in the dictionary. Therefore, to make the parametric eigenspace method fit for practical use, it requires further reduction in the dictionary size and the matching process time.

In this paper, to improve matching efficiency, we propose a novel approach to construct hierarchical dictionary for the parametric eigenspace method. In the proposed method, we classify the learning images in the eigenspace hierarchically, and make the mean image of each class represent all images in the class. For the learning image set obtained by varying one parameter value (in this paper, we assume that object is rotated about one axis), we determine the class boundaries that are optimal in the sense of discrimination and least squares. From the classification results, we construct a tree-structured dictionary in which each node has the mean image of the class. To reduce the time for recognizing the object and estimating the parameter value, we search the tree-structured dictionary for the learning image that corresponds to an input image.

2 Parametric Eigenspace Method

Suppose that learning and input images for each object are obtained in the environment shown in Figure 1. In that environment, the object is rotated about z-axis at an angle of θ. With the dictionary constructed from the learning image set, given an unknown input image, the recognition system identifies the object and estimates its pose θ in the input image.

*Address: 1-1 Asahidai, Tatsunokuchi-machi, Ishikawa 923-1292 Japan. E-mail: beto@jaist.ac.jp

†Address: 3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585 Japan. E-mail: tomohiko@slab.ntt.co.jp

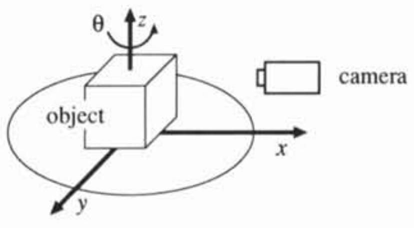

Figure 1: Experimental environment.

In order to construct the dictionary, for an object, we obtain R learning images (each image consists of N pixels) and regard each learning image as an N-dimensional vector $\mathbf{x}_r$ $(r = 1, 2, \cdots, R)$. From all $\mathbf{x}_r$, we can define the covariance matrix of the image set:

$$\mathbf{Q} = \mathbf{X}\mathbf{X}^T$$

where:

$$\mathbf{X} = [\mathbf{x}_1 - \mathbf{h},\ \mathbf{x}_2 - \mathbf{h},\ \cdots,\ \mathbf{x}_R - \mathbf{h}]$$

$$\mathbf{h} = \frac{1}{N}\sum_{r=1}^{N}\mathbf{x}_r$$

The eigenvalues λ_n $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N)$ and the corresponding eigenvectors $\mathbf{e}_n$ $(n = 1, 2, \cdots, N)$ of $\mathbf{Q}$ are determined by solving the eigenstructure decomposition problem:

$$\lambda_n \mathbf{e}_n = \mathbf{Q}\mathbf{e}_n$$

Though all N eigenvectors are needed to restore $\mathbf{x}_r$ exactly, a small number $(k \ll N)$ is sufficient for capturing the primary appearance characteristics of $\mathbf{x}_r$. We select k such that the first k eigenvectors of $\mathbf{Q}$ capture important appearance variations in the images, that is:

$$W_k = \sum_{n=1}^{k}\lambda_n \Big/ \sum_{n=1}^{N}\lambda_n \ \geq\ T_s$$

where the threshold T_s is close to, but less than, unity[4]. Consequently, by using Equation (1), we can project N-dimensional vector $\mathbf{x}_r$ to the k-dimensional eigenspace:

$$\mathbf{g}_r = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k]^T(\mathbf{x}_r - \mathbf{h}). \qquad (1)$$

The dictionary consists of $\mathbf{g}_r$ $(r = 1, 2, \cdots, R)$ that are the learning image set projected to the eigenspace. Given an unknown input image, the recognition system projects the input image to the eigenspace, and recognizes the object in the input image by matching with a point in the dictionary.

The dictionaries are constructed in two different eigenspaces. One is the universal eigenspace that is determined by all learning images of all objects of interest, and the other is the object eigenspace that is determined by learning images of an object. The dictionary constructed in the universal eigenspace is used for identifying the object in the input image. For estimating θ of each object, the dictionary constructed in each object eigenspace is used.

Figure 2 shows examples of learning image set and Figure 3 shows their representations in the eigenspaces. These learning image sets were obtained by varying θ through 360°. Such learning image set is represented as a closed curve in the eigenspace.

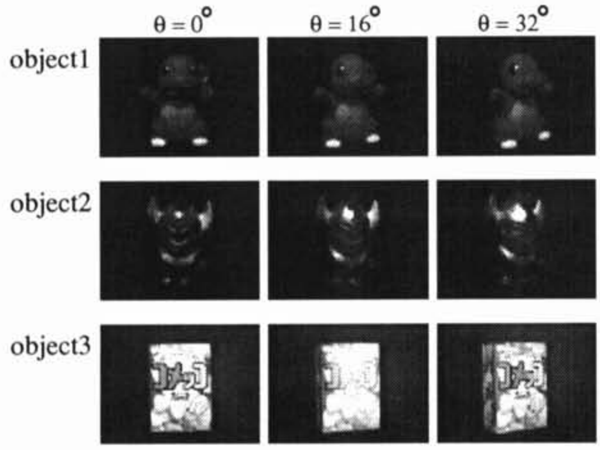

Figure 2: Examples of learning image set obtained in Figure 1 environment.

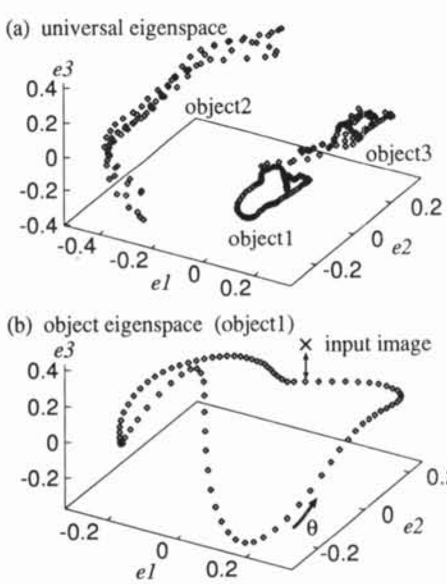

Figure 3: Learning image sets representations in the universal and the object eigenspaces ($k = 3$).

3 Hierarchical Dictionary Constructing Method

To improve matching efficiency in the parametric eigenspace method, we construct tree-structured dictionary. In the proposed method, we classify the learning image set in the eigenspace hierarchically, and at each level we make the mean image of each class represent all images in the class.

In order to make the mean image of each class C_m $(m = 1, 2, \cdots, M)$ represent all images in C_m, it is necessary to classify the image set into the classes in which all images included in C_m have similar feature and adjacent parameter θ. As shown in Figure 4 (a), this can be restated that θ corresponded to the images in C_m should be limited within the range of $\theta_m \sim \theta_{m+1}$. However, with ordinary classification method that does not take account of parameter continuity, the image set may be classified as shown in Figure 4 (b) and parameter continuity is not preserved. For this reason, in the proposed method, we classify learning images along the curve in the eigenspace, and keep continuity of θ in C_m. In addition, to determine the class boundaries that are optimal in the sense of discrimination and least squares, we extend the discriminant analysis method[5] to multi-dimension and multi-class.

As shown in Figure 5, R learning images $\mathbf{g}_r$ in the eigenspace are numbered in order of θ, and classified into C_m. Let the first boundary be set at $\mathbf{g}_1$, and $k_m = r$ denote that the mth boundary is set at $\mathbf{g}_r$. Boundaries k_m and classes C_m are represented as:

$$1 = k_1 < k_2 < \cdots < k_M \leq R$$
$$C_m = [k_m, k_{m+1} - 1] \ (m = 1, 2, \cdots, M)$$

where $k_{M+1} - 1 = R$. The mean image and the variance of all $\mathbf{g}_r$ are independent of k_m:

$$\mathbf{h}_T = \frac{1}{R} \sum_{r=1}^{R} \mathbf{g}_r, \quad \sigma_T^2 = \frac{1}{R} \sum_{r=1}^{R} \|\mathbf{g}_r - \mathbf{h}_T\|^2$$

C_m has ω_m probability of including $\mathbf{g}_r$, that is:

$$\omega_m = \frac{1}{R} \sum_{r \in C_m} 1$$

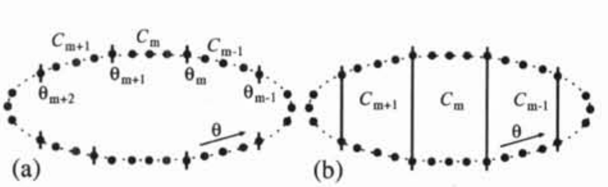

Figure 4: Classification of learning images in the eigenspace.

The mean image and the variance of $\mathbf{g}_r$ in C_m are:

$$\mathbf{h}_m = \frac{1}{R\omega_m} \sum_{r \in C_m} \mathbf{g}_r$$

$$\sigma_m^2 = \frac{1}{R\omega_m} \sum_{r \in C_m} \|\mathbf{g}_r - \mathbf{h}_m\|^2$$

The goodness of boundary loci k_m is evaluated with the separation of C_m:

$$\lambda_M(k_1, \cdots, k_M) = \frac{\sigma_B^2(k_1, \cdots, k_M)}{\sigma_W^2(k_1, \cdots, k_M)} \qquad (2)$$

where:

$$\sigma_W^2(k_1, \cdots, k_M) = \sum_{m=1}^{M} \omega_m \sigma_m^2$$
$$\sigma_B^2(k_1, \cdots, k_M) = \sum_{m=1}^{M} \omega_m \|\mathbf{h}_m - \mathbf{h}_T\|^2$$

σ_W^2, σ_B^2 and σ_T^2 have the following relation:

$$\sigma_W^2 + \sigma_B^2 = \sigma_T^2$$

Consequently, Equation (2) is equivalent to the evaluation with the following η_M:

$$\eta_M(k_1, \cdots, k_M) = \sigma_B^2(k_1, \cdots, k_M)/\sigma_T^2$$

For this reason we set k_m at k_m^*, those are:

$$\eta_M^*(k_1^*, \cdots, k_M^*) = \max_{1 \leq \{k_m\} \leq R} \sigma_B^2(k_1, \cdots, k_M)/\sigma_T^2$$

k_m^* are the optimal boundaries in the sense of not only discrimination but also least squares.

At each level, the appropriate number of classes M^* is determined by:

$$Q(M) = \eta_M^* / \bar{\eta}_M^*$$
$$Q(M^*) = \max_{2 \leq M \leq R} Q(M) \qquad (3)$$

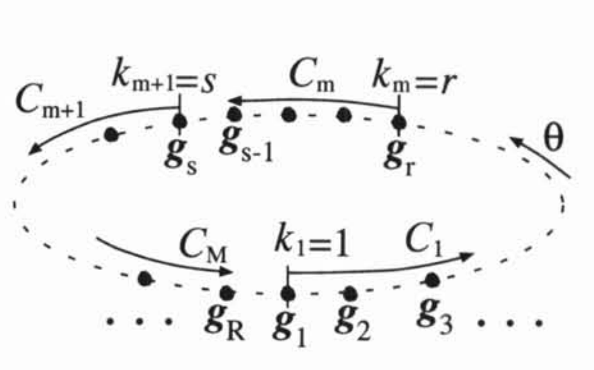

Figure 5: Learning image set and class boundaries in the eigenspace.

where R is the number of images that are classified and M is the number of classes. $\bar{\eta}_M^*$ is referred to as natural bias, and it represents the η_M of the classification in which all images are the same and k_m are set at regular intervals.

$$\bar{\eta}_M^* = 1 - \frac{\left(\frac{R}{M}\right)^2 - 1}{R^2 - 1}$$

4 Experiments

We have applied our proposed method to object recognition experiment. In the experiment, three objects shown in Figure 2 were used. For each object, 90 learning images were obtained (θ was set at every 4°), and other 180 images were obtained (θ was set at every 2°) and used for unknown input images. The dimensions of the object eigenspace for object1,2,3 and the universal eigenspace are 19, 30, 37 and 66 respectively. With three dictionaries DIC-qd, DIC-2d, DIC-2u, we identify the object and estimate the parameter θ in the input image. These dictionaries were constructed with the following methods, especially DIC-2d and DIC-2u were constructed as binary-tree.

- DIC-qd : the class boundaries were determined with our proposed method, and the number of classes at each level was determined by Equation (3).
- DIC-2d : the class boundaries were determined with our proposed method, and the number of classes at each level was fixed at two.
- DIC-2u : the class boundaries were determined with ordinary classification method (UPGMA method[6]) that does not take account of θ continuity, and the number of classes at each level was fixed at two.

In all cases, object identification succeeded, therefore we summarize the experimental result of θ estimating in Table 1. In Table 1, the counts of comparing the input image with the learning images in the dictionary and the estimation error of θ are listed.

To find the learning image that corresponds to an input image, sequential search needs 90 comparings and its maximum/mean estimation errors are 2°/1° respectively. Experimental results show that our proposed method improves matching efficiency without a decrease in recognition accuracy.

5 Conclusion

In this paper, to improve matching efficiency, we propose a novel approach to construct hierarchical dictionary for the parametric eigenspace method.

Table 1: Efficiency and accuracy with each dictionary.

	count of comparing max.(mean) [times]	estimation error of θ max.(mean) [deg]
sequential search	90 (90.0)	2 (1.00)
object1		
DIC-qd	20 (16.5)	2 (1.00)
DIC-2d	16 (13.1)	2 (1.00)
DIC-2u	20 (14.7)	10 (1.31)
object2		
DIC-qd	20 (18.7)	2 (1.00)
DIC-2d	18 (14.0)	178 (15.40)
DIC-2u	22 (15.3)	8 (1.40)
object3		
DIC-qd	28 (26.2)	2 (1.00)
DIC-2d	20 (13.8)	174 (14.20)
DIC-2u	26 (17.2)	8 (1.30)

Experimental results show that the tree-structured dictionaries made by the proposed method improve matching efficiency without a decrease in recognition accuracy.

In this paper, we assume that the object of interest is rotated about z-axis, however, this method is applicable to other cases in which the image set is obtained by varying another parameter. Furthermore, the proposed methods to determine the boundaries and the number of classes are capable of using for general classification problem.

References

[1] H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," Int. J. Comput. Vision, vol.14, no.1, pp.5–24, 1995.

[2] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," IEEE Trans. Pattern Anal. & Machine Intell., vol.12, no.1, pp.103–108, Jan. 1990.

[3] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.84–91, Jun. 1994.

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

[5] N. Otsu, "Mathematical Studies on Feature Extraction in Pattern Recognition," Researches of the Electrotechnical Laboratory, no.818, 1981.

[6] H.C. Romesburg, "Cluster Analysis for Researchers," Robert E.Krieger Publishing Company, Inc., 1983.