Appearance Manifold with Covariance Matrix for 3-D Object Recognition

Lina, Tomokazu Takahashi, Ichiro Ide, Hiroshi Murase

Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan

E-mail: {lina, ttakahashi}@murase.m.is.nagoya-u.ac.jp, {ide, murase}@is.nagoya-u.ac.jp

Abstract The authors present a robust 3-D object recognition system for recognizing noisy images. Since a recognition system usually deals with objects taken from various viewpoints, their appearance will vary from one viewpoint to another. Generally, the appearance of an object changes along with the changes of image conditions, and so does its position in the eigenspace. Such changes may cause an inaccurate recognition of an object. In this paper, we propose a novel object recognition method where covariance matrix calculation is embedded in parameterized appearance manifold. The appearance manifold will capture object characteristics along the pose rotations where the covariance matrix calculation will give the sample distribution information. Specifically, we propose the Appearance Manifold with Constant Covariance matrix (AMCC) and Appearance Manifold with View-dependent Covariance matrix (AMVC) methods. Experimental results showed that our approach could enhance the recognition performance, as well as perform robust recognition of 3-D objects under varying viewpoints and translation effects.

Key words 3-D object recognition, appearance manifold, covariance matrix, parametric eigenspace

1. Introduction

Object recognition is one of the most active research areas in computer vision. An object recognition system is efficiently organized to recognize an object of interest by comparing an image with models that already exist in the database gallery. It is argued that 3-D recognition can be accomplished using linear combinations of as few as four or five 2-D viewpoint images [1][2]. Unfortunately, with this traditional approach, an image is represented as a very high dimensional feature, that may cause inefficiency in its application. One most attractive and popular approach to handle this problem is the Principal Component Analysis (PCA) method which transforms an image represented by a high dimensional feature into low dimensional feature representation, called the eigenspace representation.

The eigenspace representation, which is a collection of points in the eigenspace, is very sensitive to image conditions – background noise, image shift, occlusion of objects, scaling of the image, and illumination changes [3]. Generally, the appearance of an object changes along with the changes of image conditions, and so does its eigen-point position. In the traditional recognition system whose works are based on a template matching technique, such changes may cause an inaccurate recognition of an object. Earlier works have proposed many methods to handle this problem, such as Murase and Nayar with Parametric Eigenspace (PE) [4], Ohba and Ikeuchi with Eigen Window [3], and Moghaddam and Pentland with Probabilistic Visual Learning [5].

We put our focus on Murase and Nayar's method which gave high recognition capability in recognizing 3-D

objects with its parameterized manifold in eigenspace. As the appearance of an object varies from one viewpoint to another, the PE method proposed the use of a parameterized manifold in eigenspace to capture object's changes which cover their pose and illumination direction. The PE method has shown high recognition capability in recognizing 3-D objects (see [4]). However, when the problem of image shifting and occlusion of objects are included in the system, PE method could not give a satisfying recognition result.

Our objective is to develop a robust 3-D object recognition system for recognizing noisy images. Since a 3-D object recognition system usually deals with objects taken from various viewpoints, it may also deals with the changes of object's appearances. When the appearance of an object is changed, its position in the eigenspace also changes. This condition might cause an inaccurate recognition of an object. One promising way is to use an appearance manifold parameterized by the object's pose and also add class-density information, such as mean vector and covariance matrix, to the system. The appearance manifold will capture object characteristics along the pose rotations, while the covariance matrix calculation will give the information of sample distribution.

In this paper, we propose a novel object recognition method where covariance matrix calculation is embedded in a parameterized appearance manifold. We propose two methods: the Appearance Manifold with Constant Covariance matrix (AMCC) and the Appearance Manifold with View-dependent Covariance matrix (AMVC) method. The AMCC method uses covariance matrices with constant values obtained from the average value of all covariance matrices in the manifold. While in the AMVC method, the covariance matrices change as function of viewpoint for each manifold.

The paper is organized as follows: we give a brief description of the PE method in section 2. Then, introduce our appearance manifold with covariance matrix methods (AMCC and AMVC) in section 3. Section 4 covers the experiments and analysis of the proposed methods. Finally, conclusion and future works are presented in section 5.

2 . Parametric Eigenspace Representation

In this section we will give a brief description of the PE representation. The PE method provides an efficient way to represent object appearance that is parameterized by its variables such as pose and illumination.

First, an image set of an object is obtained in various poses. Then the image is normalized in brightness and scaled to achieve invariance to image magnification and illumination intensity. These normalized images can be written as a vector $\mathbf{x}$ by reading the number of pixels (N) in an image:

$$\mathbf{x} = [x_1, x_2, ..., x_N]^T \tag{1}$$

Let M be the number of the images in a learning set. By subtracting the average image c of all images, we obtain the learning set $\mathbf{Y}$:

$$\mathbf{Y} = [\mathbf{x}_1 - c, \mathbf{x}_2 - c, ..., \mathbf{x}_M - c] \tag{2}$$

Next, we define the covariance matrix by

$$\mathbf{Q} = \mathbf{Y}\mathbf{Y}^T \tag{3}$$

and determine the eigenvectors $\mathbf{e_i}$ and the corresponding eigenvalues $\lambda_\mathbf{i}$ by solving the following well-known eigenvector decomposition problem:

$$\lambda_i \mathbf{e}_i = \mathbf{Q}\mathbf{e}_i \tag{4}$$

For dimension reduction, simply ignore small eigenvalues and use only k corresponding eigenvectors using T threshold value:

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \geq T \tag{5}$$

where $k << N$.

Next, use the first k eigenvectors to project $\mathbf{x}_l^{(p)}$ as images of object p with viewpoint l into the eigenspace:

$$\mathbf{g}_l^{(p)} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_k]^T (\mathbf{x}_l^{(p)} - \mathbf{c}) \tag{6}$$

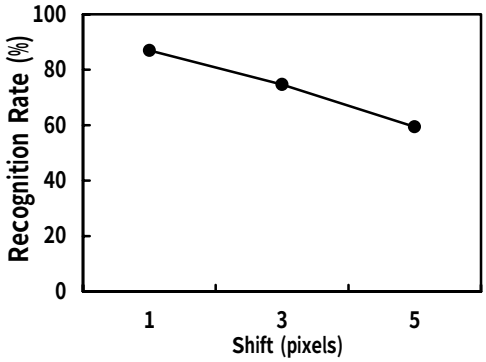

Figure 1. Recognition rates of the PE method in recognizing 3-D objects with translation effects.

By projecting all the learning samples in image set $\mathbf{x}_l^{(p)}$, we get a set of discrete points in the eigenspace. Pose variation between any two consecutive images in $\mathbf{x}_l^{(p)}$ is relatively small [3]. As a result, consecutive images are strongly correlated and could be represented in a smooth manifold in the eigenspace:

$$\tilde{\mathbf{g}}_l^{(p)}(\theta) \tag{7}$$

where θ is a continuous pose parameter.

To recognize an input image z, project z into the eigenspace:

$$\mathbf{h} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_k]^T (\mathbf{z} - \mathbf{c}) \tag{8}$$

then calculate distance d between the projected-image in the eigenspace $\mathbf{h}$ and the manifold $\tilde{\mathbf{g}}_l^{(p)}$:

$$d = \left\| \mathbf{h} - \tilde{\mathbf{g}}_l^{(p)} \right\| \tag{9}$$

Next, an input image is classified based on the minimum distance d.

3 . Appearance Manifold with Covariance Matrix

The PE method with its parameterized manifold, which covers object's pose and illumination direction, has shown high recognition capability in recognizing 3-D objects (see [4]). However, when the problem of image shifting and occlusion of objects are included in the system, the PE method could not give a satisfying recognition result. Fig. 1 illustrates the recognition rate of the PE method in recognizing ten 3-D objects with various horizontal positions influenced by various translation effects.

The learning images contained 32x32 pixels of an original-captured image and images generated with artificial noises, such as blur and shift, with ten degrees interval of horizontal viewpoints (0^o, 10^o, 20^o, …, 350^o). The generated images consist of 21 images with 5% until 25% blur effects and 10 images with one until five pixels left shift and right shift effects. While for testing, images were five degrees horizontal shifted from every learning images (5^o, 15^o, 25^o,…, 355^o) and influenced with one, three, and five pixels translation effects. We created the appearance manifold using the cubic spline interpolation method and then classified a test image based on its minimum Euclidean distance to the mean vector.

Fig. 1 shows that the recognition rate of the PE method for recognizing ten objects with various horizontal positions and one pixel translation effect was 86.94%. While for recognizing objects with three and five pixels translation noise effects, the recognition rate of the PE method was 74.72% and 59.44% respectively. These results proved that the PE method could not give a satisfying result for recognizing shifted images.

One promising idea to solve this problem is adding class-density information, such as mean vector and covariance matrix, to the parameterized appearance manifold. The appearance manifold will capture object characteristics along the pose rotations where the covariance matrix calculation will give the sample distribution information.

In this paper, we propose an appearance manifold with covariance matrix calculation that is parameterized by object's poses. Specifically, we propose the Appearance Manifold with Constant Covariance matrix (AMCC) and the Appearance Manifold with View-dependent Covariance matrix (AMVC) method. Fig. 2 illustrates the appearance manifold representation without covariance matrix calculation, while Fig. 3 illustrates our appearance manifold with covariance matrix representation.

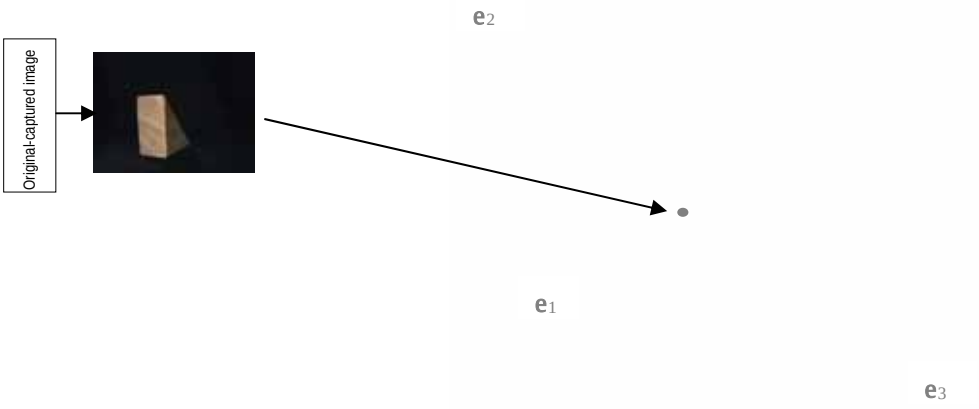

Figure 2. Appearance manifold representation.

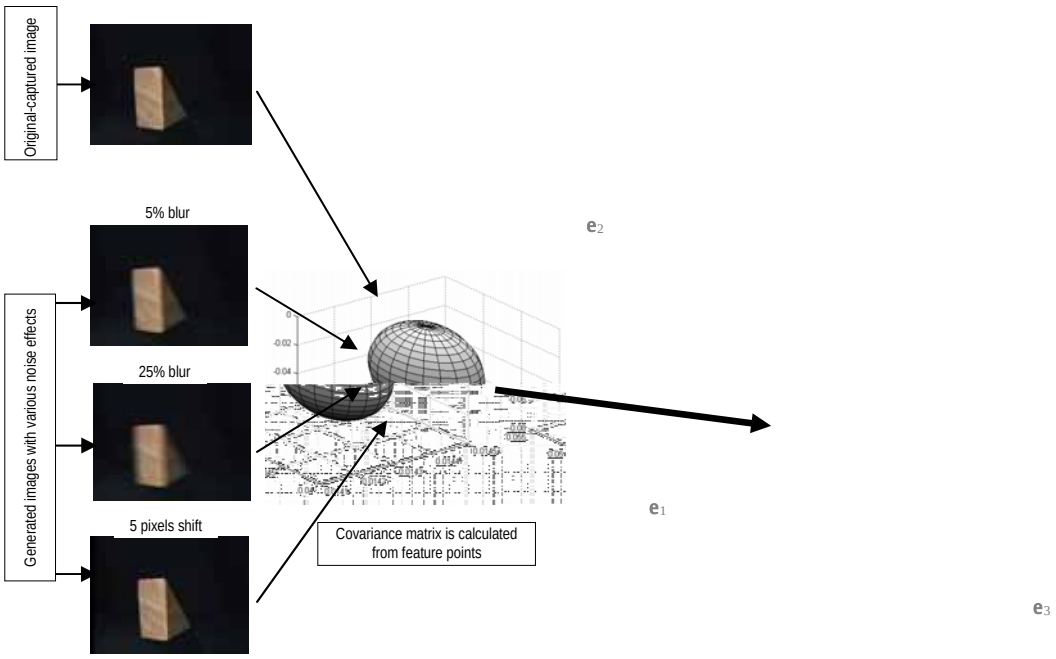

Figure 3. Appearance manifold with covariance matrix representation.

3.1 Appearance Manifold with Constant Covariance Matrix (AMCC)

In AMCC, after transforming learning images to the eigenspace, the mean vector $\mu^{(p)}(\theta)$ and the covariance matrix $\sum^{(p)}(\theta)$ for each object p for viewpoint θ are calculated. The mean vector is typically estimated using:

$$\mu^{(p)}(\theta) = \frac{1}{s}\sum_{i=1}^{s} g_i{}^{(p)}(\theta) \qquad (10)$$

where s is the number of learning samples from each class, and $g_i{}^{(p)}(\theta)$ is the image sample i from class viewpoint θ and object p. The covariance matrix is typically estimated by:

$$\Sigma^{(p)}(\theta) = \frac{1}{s-1}\sum_{i=1}^{s}\left(g_i{}^{(p)}(\theta) - \mu^{(p)}(\theta)\right)\left(g_i{}^{(p)}(\theta) - \mu^{(p)}(\theta)\right)^T \qquad (11)$$

Next, create $\tilde{\mu}^{(p)}(\theta)$ as the manifold of the mean vector and $\tilde{\Sigma}^{(p)}(\theta)$ as the manifold of the covariance matrix.

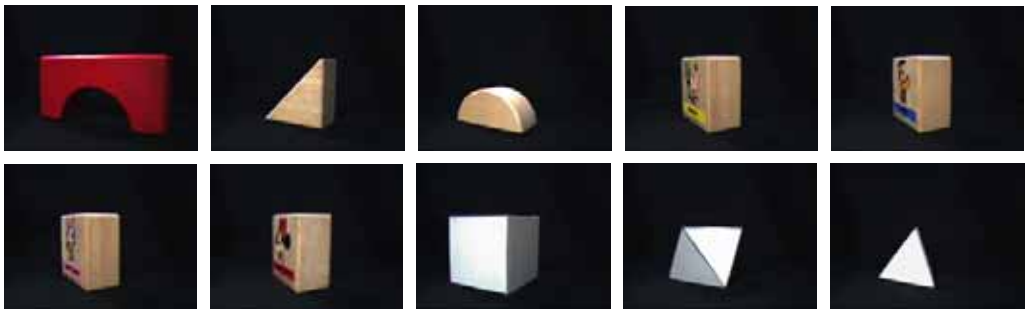

Figure 4. 3-D Basic shape objects.

Figure 5. 3-D Basic shapes objects with 5%, 10%, 15%, 20%, and 25% blur effects.

The manifold of $\tilde{\mu}^{(p)}(\theta)$ is obtained by applying an interpolation method between two consecutive mean vectors. While the manifold of covariance matrix $\tilde{\Sigma}^{(p)}(\theta)$ contains the same value for every viewpoint θ by applying the average covariance matrix:

$$\overline{\Sigma}^{(p)} = \frac{1}{w}\sum_{i=1}^{w}\Sigma_i^{(p)}(\theta) \tag{12}$$

with w the number of viewpoint class for each object.

Since we have the parameter of mean vector and covariance matrix in the appearance manifold, the sufficient distance to use in this calculation is the Mahalanobis distance. We use the Regularized Mahalanobis distance [6] measurement to classify an object z:

$$d^{(p)}(z) = \left(z - \tilde{\mu}^{(p)}(\theta)\right)^T \left[(1-\lambda)\left(\tilde{\Sigma}^{(p)}(\theta) + \varepsilon I\right)^{-1} + \lambda I\right]\left(z - \tilde{\mu}^{(p)}(\theta)\right) \tag{13}$$

where λ and ε are learning parameters. Parameter λ is in the interval [0,1] and it controls the tradeoff between Mahalanobis and Euclidean distances. If $\lambda = 0$ then the Regularized Mahalanobis distance is the Mahalanobis distance, while if $\lambda = 1$ then it becomes the Euclidean distance. The next parameter ε is used to stabilize the learning process by converting a singular matrix to a non-singular one.

3.2 Appearance Manifold with View-Dependent Covariance Matrix (AMVC)

In AMVC, after transforming learning images for each viewpoint θ to the eigenspace, the $\mu^{(p)}(\theta)$ mean vector and the covariance matrix $\Sigma^{(p)}(\theta)$ for every viewpoint-class for each object are calculated. Next, a continuous curve which is parameterized by viewpoint rotation (θ) is developed using an interpolation method. Thus, we have the manifold $\tilde{\mu}^{(p)}(\theta)$ of mean vector with $\tilde{\Sigma}^{(p)}(\theta)$ covariance matrix as a function of the viewpoint θ.

Finally, we use the Regularized Mahalanobis distance to classify an object z.

4 . Experimental Results and Analysis

To demonstrate the performance of our method, we conducted experiments to recognize ten objects with various 3-D basic shapes. Fig. 4 illustrates the object set used in the experiment.

The size of the learning image was 32x32 pixels. The degree interval was 10 degrees of horizontal positions (0^o, 10^o, 20^o, …, 350^o). For each object in the learning stage, we trained the system with original-captured images and generated images with artificial noises, such as blur and translation effects. The generated images consist of 21

images with 5% until 25% blur effects and 10 images with one until five pixels left shift and right shift effects. Fig. 5 illustrates the example of learning images with 5%, 10%, 15%, 20%, and 25% blur effects. Next, a cubic spline interpolation method is used to form the appearance manifold.

The testing images were five degrees horizontal shifted from every learning images (5^o, 15^o, 25^o,..., 355^o) and influenced with one, three, and five pixels of translation effects. The PE method classified the test images based on its minimum Euclidean distance to the mean vector. However, for AMCC and AMVC, the minimum Regularized Mahalanobis distance with $\lambda = 0.1$ and $\varepsilon = 0.01$ was used for classifying the testing images.

Experiments were conducted to compare the classification accuracy from the PE, AMCC and AMVC methods. In our experiments, the covariance matrices of the PE method were equal to the identity matrix, but each class had a different mean vector. In AMCC, we use the identical covariance matrix for every class, based on the average covariance matrix. While in AMVC, the covariance matrices changed for each class based on the function of viewpoints.

Fig. 6 shows the recognition results of the PE, AMCC, and AMVC methods. Based on Fig. 6, our proposed AMVC method gave the best recognition rates compared with the AMCC and PE methods. Also, the AMVC method showed its robustness to the presence and the increment of the translation effects in the 3-D recognition system.

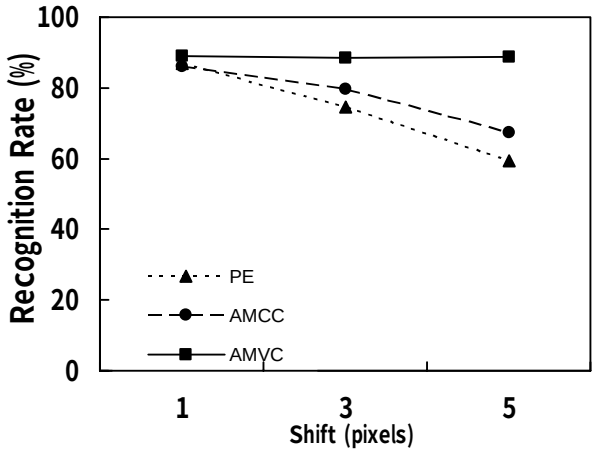

Figure 6. Recognition rates for 3-D objects with translation effects.

For images with one pixel translation effect, the recognition rate for the PE method was 86.94%, the recognition rate for the AMCC method was 86.11%, while our proposed AMVC method could give the highest recognition rate 89.17%. However, along with the increasing number of translated pixels, the recognition rates of all methods decreased. For images with 5 pixels translation effect, the recognition rate for the PE method was 59.44% and 67.50% for the AMCC method. While, our proposed AMVC method could maintain its recognition rates up to 88.89%, the highest recognition rates compared to the other two methods.

5 . Conclusion and Future Works

In this paper, we presented a novel method to recognize 3-D objects with noisy images. Recognition experiments showed that our proposed AMVC method, with its view-dependent covariance matrix, could enhance the recognition performance, as well as perform a robust recognition of 3-D objects under varying viewpoints and translation effects.

Future works include recognizing 3-D objects influenced with other type of noises, developing the recognition system using less learning image samples by changing the interval of viewpoint orientations and solving the segmentation problem in order to enhance the performance of the recognition system.

Reference

[1] S. Ullmann and R. Basri, “Recognition of Linear Combination of Models”, IEEE Trans. PAMI, Vol.13, No.10, pp.992-1007, 1991.

[2] T. Poggio and S. Edelman, “A Network that Learns to Recognize Three Dimensional Objects”, Nature, Vol.343, No.6255, pp.263-266, 1990.

[3] K. Ohba and K. Ikeuchi, “Detectability, Uniqueness, and Reliability of Eigen Windows for Stable Verification of Partially Occluded Objects”, IEEE Trans. PAMI, Vol.19, No.9, pp.1043-1048, 1997.

[4] H. Murase and S.K. Nayar, “Illumination Planning for Object Recognition Using Parametric Eigenspaces”, IEEE Trans. PAMI, Vol.16, No.12, pp.1219-1227, 1994.

[5] B. Moghaddam and A. Pentland, “Probabilistic Visual Learning for Object Representation”, IEEE Trans. PAMI, Vol.19, No.7, pp.696-710, 1997.

[6] J. Mao and A.K. Jain, “A Self-organizing Network for Hyperellipsoidal Clustering (HEC)”, IEEE Trans. NN, Vol.7, No.1, pp.16-29, 1996.