LH-008

RECOGNITION OF 3D OBJECTS IN VARIOUS CAPTURING CONDITIONS USING APPEARANCE MANIFOLD WITH EMBEDDED COVARIANCE MATRIX

Lina　Tomokazu Takahashi　Ichiro Ide　Hiroshi Murase

1. Introduction

3D objects are visually complex and highly dependent on environmental conditions. Therefore, to visually learn the appearance of a 3D object, it is necessary to figure it in such way that can fully represent its characteristics. In general, capturing the characteristics of a 3D object can be done by using several combinations of 2D images or by constructing a high-cost 3D shape model. Here, we focus on an appearance-based approach that uses combinations of images to capture the appearance variability of a 3D object. One primary advantage of appearance-based methods is that it is not necessary to create representations for objects since, for a given object, its model is now implicitly defined by the selection of the sample images of the objects [1].

Appearance-based approach is usually combined with the concept of Principal Component Analysis (PCA). This concept enables a method to efficiently present a series of sample images in a low-dimensional feature description, called the eigenspace. For years, the eigenspace has provided an efficient and easy way to solve many recognition problems. Some of the earlier works in this domain include the application of PCA method in eigenpictures of Kirby and Sirovich [2] and eigenfaces of Turk and Pentland [3] for characterizing human face. Later, Moghaddam [4] proposed Probabilistic PCA method which formulates a maximum-likelihood framework for target detection. Further, Murase and Nayar [5] introduced Parametric Eigenspace method which uses a simple manifold to capture pose changes of an object in eigenspace. Addressing a different problem, Martinez [6] developed the Probabilistic Approach for recognizing partially occluded human faces.

Just as the appearance of an object highly depends on the image conditions, the image's position in eigenspace relies on the object's appearance. For handling changes caused by pose and illumination variability, Murase and Nayar's Parametric Eigenspace method could give more satisfactory results than the traditional eigenspace method. Unfortunately, naturally the presence of various types of noise could not be avoided. It might occur during the camera-captured process or as an error product of a segmentation process. In many cases, appearances of objects in images depend not only to pose changes and illumination, but also to geometric distortion and quality degradation effects.

Figure 1. Image samples of 3D objects with various geometric distortions (translation, rotation) and quality degradations (blur)

Graduate School of Information Science, Nagoya University

Fig. 1 shows some image samples of 3D objects with translation, rotation, and blur effects. In eigenspace, when these significant variations exist, the eigenpoint position is changed relatively far from the learning samples. Thus, relying the learning and recognition process on a simple appearance manifold such as in Parametric Eigenspace method tends to fail.

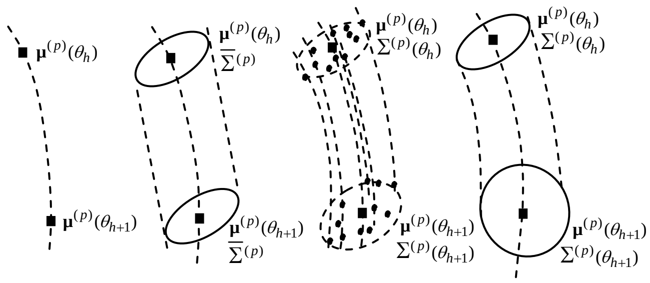
To overcome this problem, we propose the construction of an appearance manifold with embedded covariance matrix. The basic idea is to eliminate the eigenpoint-to-eigenpoint correspondences of each learning pose class and then to construct the correspondences from the covariance matrices directly. We develop a manifold by embedding the covariance correlation information of eigenpoint distribution in every viewpoint.

Here, we present various techniques to construct the appearance manifold with embedded covariance matrix. Each method uses a different type of construction process, although in general they use the same basic steps. We propose the appearance manifold with Constant Covariance matrix (CC) method which has a constant value of covariance matrix for every viewpoint, and propose the appearance manifold with learning-Point Interpolation (PI) and the appearance manifold with Covariance Interpolation (CI) which perform view-dependent covariance matrix value for each pose position. By using these appearance manifold models, the robustness of the system will be increased, since it could analyze image condition such as pose changes, while the embedded view-dependent covariance matrix could define the distribution information of the eigenpoints in eigenspace.

2. Eigenspace Representation

Generally, appearance-based approaches deal with a set of learning images in various capturing conditions. These images are represented in a very high dimensional space, thus, they could not be applied directly due to efficiency reasons. Here, PCA is used to efficiently represent a collection of images by reducing their dimensionality.

PCA represents a linear transformation that maps the original n-dimensional space onto a k-dimensional feature subspace where normally $k<<n$. The first k eigenvectors will be used to project S learning samples of P objects with H poses. Thus, with $\mathbf{x}_s^{(p)}(\theta_h)$ the s sample image of object p with horizontal viewpoint θ_h and $\mathbf{e}_i$ are the eigenvectors, the new feature vectors $\mathbf{g}_s^{(p)}(\theta_h) \in \Re^k$ are defined by $\mathbf{g}_s^{(p)}(\theta_h) = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_k]^T \left(\mathbf{x}_s^{(p)}(\theta_h) - \mathbf{c}\right)$. These eigenvectors $\mathbf{e}_i$ were obtained by solving the eigenstructure decomposition $\lambda_i \mathbf{e}_i = \mathbf{Q}\mathbf{e}_i$, where $\mathbf{Q}$ is the auto correlation matrix and λ_i the eigenvalue associated with the eigenvector $\mathbf{e}_i$.



(a) PE (b) CC (c) PI (d) CI

Figure 2. Construction models of appearance manifold

3. Appearance Manifold with Embedded Covariance Matrix in Eigenspace

This section describes the process of constructing the appearance manifold with embedded covariance matrix in eigenspace and the recognition process of input images using the Mahalanobis distance measurement.

3.1 Construction Process of Appearance Manifold with Embedded Covariance Matrix

In this section, we present various techniques to construct the appearance manifold with embedded covariance matrix. Fig. 2 shows the four types of construction models for the appearance manifold: the simple manifold used in the Parametric Eigenspace (PE) method, the appearance manifold with Constant Covariance matrix (CC) method, and the other two appearance manifold with view-dependent covariance matrix methods, called the appearance manifold with learning-Point Interpolation (PI) method and the appearance manifold with Covariance Interpolation (CI) method.

Although each method uses a different type of construction technique for the appearance manifold, in general they use the same basic steps (see [7]). Fig. 3 shows the construction process of appearance manifold with embedded covariance matrix. Here, we only use the horizontal pose parameter (θ_h) to construct the appearance manifold.

After transforming learning images to the eigenspace, first, calculate the mean vector $\boldsymbol{\mu}^{(p)}(\theta_h)$ and the covariance matrix $\Sigma^{(p)}(\theta_h)$ for each object p for horizontal viewpoint θ_h. The mean vector is typically estimated using :

$$\boldsymbol{\mu}^{(p)}(\theta_h) = \frac{1}{S}\sum_{s=1}^{S} \mathbf{g}_s{}^{(p)}(\theta_h) \quad (1)$$

where S is the number of learning samples from each class, and $\mathbf{g}_s{}^{(p)}(\theta_h)$ is the image sample s from class viewpoint θ_h and object p in eigenspace. On the other hand, the covariance matrix is typically estimated by :

$$\Sigma^{(p)}(\theta_h) = \frac{1}{S-1}\sum_{s=1}^{S} \left(\mathbf{g}_s{}^{(p)}(\theta_h) - \boldsymbol{\mu}^{(p)}(\theta_h)\right)\left(\mathbf{g}_s{}^{(p)}(\theta_h) - \boldsymbol{\mu}^{(p)}(\theta_h)\right)^T \quad (2)$$

.

Then, create $\tilde{\boldsymbol{\mu}}^{(p)}(\theta)$ as a continuous manifold of the mean vector and $\tilde{\Sigma}^{(p)}(\theta)$ for the covariance matrix. The processes of creating these manifolds might be different from one method to another.

Parametric Eigenspace (PE) : The PE method uses a simple manifold obtained from the interpolation of the mean vector of the eigenpoints in two consecutive poses and applies the values of identity matrix for its covariance matrices. The construction model of the appearance manifold in the PE method is depicted in Fig. 2(a).

Constant Covariance (CC) : Next, Fig. 2(b) shows the appearance manifold with Constant Covariance matrix (CC). After calculating the mean vectors and covariance matrix values for each learning pose in Eq. (1) and Eq. (2), apply an interpolation method for the mean vector of two consecutive learning poses to obtain the manifold of mean vector $\tilde{\boldsymbol{\mu}}^{(p)}(\theta)$. On the other hand, the manifold of covariance matrix $\tilde{\Sigma}^{(p)}(\theta)$ contains the same value for every viewpoint θ_h by applying the average covariance matrix :

$$\overline{\Sigma}^{(p)} = \frac{1}{H}\sum_{h=1}^{H} \Sigma^{(p)}(\theta_h) \quad (3)$$

with H number of viewpoint classes for each object.

Learning-Point Interpolation (PI) : Fig. 2(c) shows the appearance manifold with learning-Point Interpolation (PI) method. It obtains the appearance manifold by interpolating every learning-eigenpoint in each pose class to other learning-eigenpoints in the consecutive pose classes that has the same characteristics, such as same degradation effects. After creating those manifolds for each eigenpoint, generate the new eigenpoints for every in-between pose class, and then calculate their mean vectors and covariance matrices for every pose class.

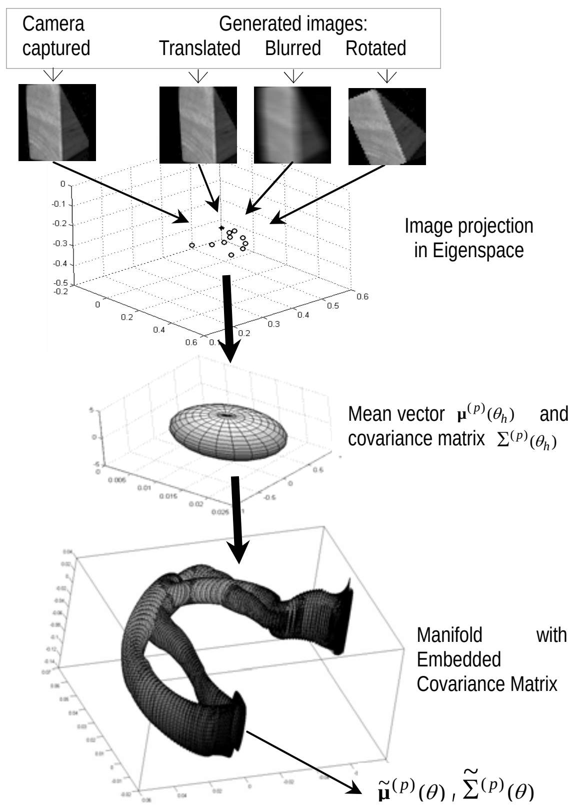


Figure 3. Construction Process of Appearance Manifold with Embedded Covariance Matrix

Covariance Interpolation (CI) : Finally, Fig. 2(d) shows another type of appearance manifold, called the CI method. This method uses the appearance manifold embedded with view-dependent covariance matrix that changes along with the function of viewpoints. The mean vector $\tilde{\boldsymbol{\mu}}^{(p)}(\theta)$ could be obtained by applying an interpolation method between two consecutive mean vectors $\boldsymbol{\mu}^{(p)}(\theta_h)$ and $\boldsymbol{\mu}^{(p)}(\theta_{h+1})$. Then, the covariance matrix $\tilde{\Sigma}^{(p)}(\theta)$ could be obtained by interpolating the covariance matrices $\Sigma^{(p)}(\theta_h)$ and $\Sigma^{(p)}(\theta_{h+1})$, respectively.

3.2 Recognition Process using Mahalanobis Distance Measurement

In order to recognize an input image $\mathbf{z}$, we calculate the Mahalanobis distance measurement defined in this formula:

$$d^{(p)}(\mathbf{z}) = \min_{\theta}\left(\mathbf{z} - \tilde{\boldsymbol{\mu}}^{(p)}(\theta)\right)^T \left(\tilde{\Sigma}^{(p)}(\theta)\right)^{-1} \left(\mathbf{z} - \tilde{\boldsymbol{\mu}}^{(p)}(\theta)\right) \qquad (3)$$

The Mahalanobis metric provides a sufficient way to classify images based on their related characteristics and likelihood in each pose class.

4. Application in 3D Object Recognition

To evaluate the performance of our proposed methods, explained in Section 3, we developed an application in 3D object recognition. The developed system was used to recognize seven objects with various horizontal pose positions and influenced by geometric and quality-degradation effects, such as translation, rotation, and motion blur. Samples of 3D objects we used in experiments are shown in Fig. 4.

In the learning stage, a total of 6,552 images were normalized into 32 x 32-pixels grayscale images. Each object consists of 36 poses with 10-degree intervals of horizontal positions (0°, 10°, 20°,…,350°), and each pose consists of an original camera-captured image and 25 generated images. Those generated images were obtained by composing artificial noises, such as left and right translations (3, 6, 9, 12, 15 pixels), clockwise and counter-clockwise rotations (5°, 10°, 15°, 20°, 25°), and motion blur (5%, 10%, 15%, 20%, 25%).

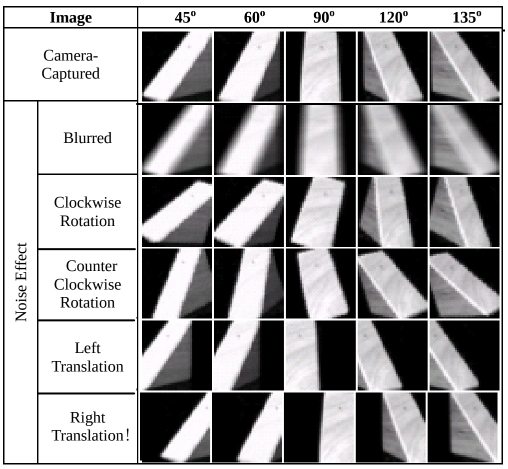


Figure 4. Samples of 3D objects used in the experiments

Next, those images were projected into the eigenspace, and the appearance manifolds were created based on each construction method, as explained in section 3.1. We applied spline interpolation technique to obtain smooth manifold of mean vectors and linear interpolation for covariance matrices.

Finally, we tested the system with input images that were different from the learning images (5°, 15°, 25°…355°) in horizontal poses and influenced by various types of degradation effects. For classification, we employed the Mahalanobis distance, as described in section 3.2. Figures 5, 6, and 7 show a series of the recognition accuracies of four methods in recognizing images influenced with translation effects, rotation effects, and motion blur effects, respectively.

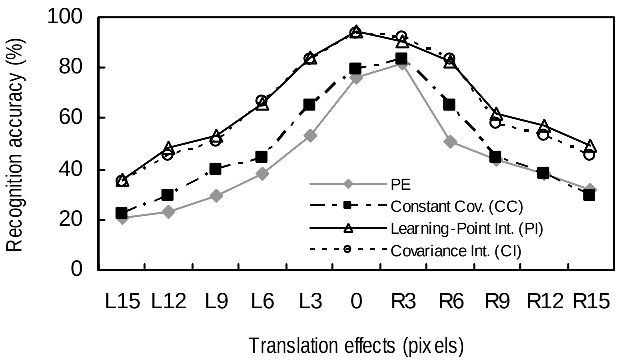


Figure 5. Recognition accuracies of images with left (L) and right (R) translation effects

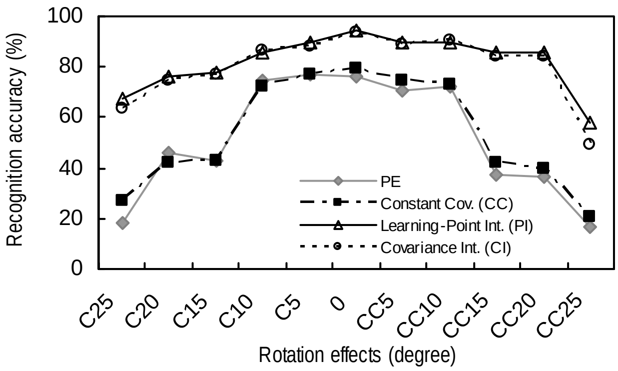


Figure 6. Recognition accuracies of images with clockwise (C) and counter-clockwise (CC) rotation effects

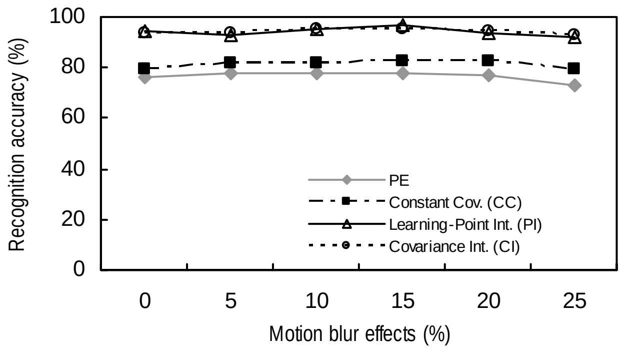
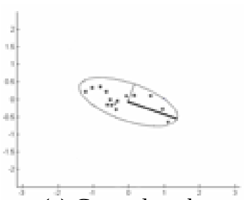


Figure 7. Recognition accuracies of images with motion blur effects

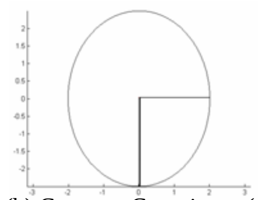
All figures indicate that the PI method and CI method, with their view-dependent covariance matrices, always achieved higher recognition accuracies than the PE method and CC method. For recognizing non-degraded images, the PI method achieved the highest recognition results with 94.05% recognition accuracy, while the CI method, CC method, and PE method achieved 93.65%, 79.37%, and 76.19%, respectively.

When recognizing images with various geometric distortion and quality degradation effects, the PI method and CI method still achieved high recognition results. PI method and CI method could give satisfactory results in recognizing objects with up to 6 pixels (20% of image size) translation from its center, up to 20^o of rotation effects, while able to maintain their accuracies to recognize up to 25% blurred images. Meanwhile, CC method could only give satisfactory results in recognizing objects with up to 3 pixels (10% of image size) translation from its center, 10^o of rotation effects, and 20% blurred images.

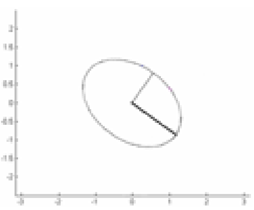
The verification results are shown in Fig. 8 below, which shows the construction of covariance matrix along with its first and second eigenvectors directions. In this case, the covariance matrix constructions were obtained by slicing the appearance manifold of each method on an unlearned 45^o viewpoint, where each appearance manifold was constructed from two extremely different learning viewpoints (0^o and 90^o). Fig. 8(a) shows the ground truth of a covariance matrix construction, obtained from real image projections, while Fig. 8(b), Fig. 8(c), and Fig. 8(d) show the construction results of covariance matrix from CC method, PI method, and CI method, respectively.



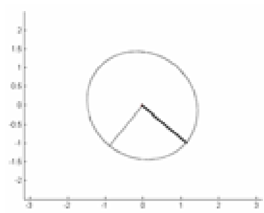
(a) Ground truth



(b) Constant Covariance (CC)



(c) Learn-Point Interpolation (PI)



(d) Covariance Interpolation (CI)

Figure 8. Construction results of covariance matrix of an unlearned viewpoint along with its first and second eigenvectors directions, sliced from the appearance manifold

Fig. 8(c) of PI method clearly presents the most similar construction results of covariance matrix with the results of real image projections which depicted in Fig. 8(a). This verifies that PI method has the highest recognition capability among the other methods. Following the PI method is the CI method with its construction results depicted in Fig. 8(d). Finally, Fig. 8(b) shows the construction result of covariance matrix of CC method with its less similar shape and direction which confirms its weak recognition capability compared with the PI method and the CI method with their view-dependent covariance matrix.

5. Conclusion and Future Works

We proposed the construction of appearance manifold with embedded covariance matrix and developed its application to recognize 3D objects from images that are influenced by geometric distortions and quality-degraded effects. As a result, our proposed appearance manifold with embedded covariance matrix could outperform the accuracy of the simple appearance manifold such as in PE method. Among these methods, our view-dependent covariance matrix methods (PI method and CI method) could provide more robust recognition capability compared with that of constant covariance matrix of CC method.

Our future works include recognizing 3D objects from images that are influenced by other types of effects, as well as developing a recognition system that uses fewer learning samples by implementing a larger interval of viewpoint orientations.

References

[1] A.M.Martinez and A.C. Kak, "PCA Versus LDA," IEEE Trans. PAMI, vol. 23, no. 2, pp. 228 - 233, Feb. 2001.

[2] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Face," IEEE Trans. PAMI, vol. 12, no. 1, pp. 103 - 108, Dec. 1990.

[3] M. Turk and A. Pentland, "Face Recognition using Eigenfaces," Proc. IEEE conf. Computer Vision and Pattern Recognition, pp. 586 - 591, Jun. 1991.

[4] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," IEEE. Trans. PAMI, vol. 19, no. 7, pp. 696 - 710, Jul. 1997.

[5] H. Murase and S.K. Nayar, "Illumination Planning for Object Recognition using Parametric Eigenspaces," IEEE Trans. PAMI, vol. 16, no. 12, pp. 1219 - 1227, Dec. 1994.

[6] A.M. Martinez, "Recognition of Partially Occluded and/or Imprecisely Localized Faces Using a Probabilistic Approach", Proc. IEEE conf. Computer Vision and Pattern Recognition, vol. 1, pp. 712 - 717, Jun. 2000.

[7] Lina, T. Takahashi, I. Ide, H. Murase, "Appearance Manifold with Embedded Covariance Matrix for Robust 3D Object Recognition," Proc. IAPR conf. Machine Vision Applications 2007, pp. 504 – 507, May 2007.