Proceedings of the Fifth IASTED International Conference
Signal Processing, Pattern Recognition, and Applications
February 13–15, 2008, Innsbruck, Austria
ISBN Hardcopy: 978-0-88986-717-8 / CD: 978-0-88986-718-5

EMBEDDING VIEW-DEPENDENT COVARIANCE MATRIX IN OBJECT MANIFOLD FOR ROBUST RECOGNITION

Lina, Tomokazu Takahashi, Ichiro Ide, Hiroshi Murase
Graduate School of Information Science,
Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Aichi, Japan
{lina, ttakahashi}@murase.m.is.nagoya-u.ac.jp, {ide, murase}@is.nagoya-u.ac.jp

ABSTRACT
Variations in camera-captured images usually occur naturally. For example, the appearance of an object usually differs for every pose and degradation effect might occur during the capturing process. While we could use a simple manifold to represent the variability of pose, relying on the simple manifold technique to deal with both pose and degradation problems is not possible, since a simple manifold does not take into account the information of sample distributions in feature space. In this paper, we propose a technique which embeds view-dependent covariance matrix in object manifold to develop a robust 3D object recognition system. Here, the view-dependent covariance matrices were obtained in an efficient way by interpolating eigenvectors and eigenvalues along the manifold. Experiment results showed that our developed 3D object recognition system could accurately recognize 3D objects even from images which are influenced by geometric distortions and quality degradation effects.

KEY WORDS
Object recognition, appearance manifold, covariance matrix, eigenspace

1. Introduction

Recognizing a 3D object from its 2D images raises many challenges. For more than a decade, the main issue for appearance-based approach is to solve the pose problem. It is not surprising that the performances of object recognition systems drop significantly when large pose variations are present in input images [1]. Therefore, various attempts have been made to handle this issue.

Earlier methods focused on constructing invariant features [2], synthesizing a prototypical frontal view [3], or classify pose problems [4]. Following the popular eigenface approach which was proposed by Turk and Pentland [5], many techniques have been extended to explore recognition in eigenspace domain. Recently, researchers improved the previous eigen model with the use of appearance manifold in eigenspace in order to achieve pose-invariant recognition. Addressing various problems, many types of appearance manifolds have been developed, such as the simple appearance manifold in [6-7] which could handle pose and illumination variations, the appearance manifold with probabilistic techniques [8-9] for handling various facial changes, the layer-transparent manifold in [10] for recognizing occluded objects, etc.

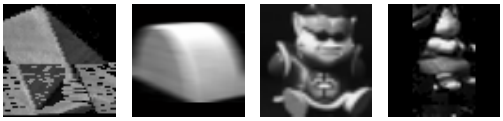


Figure 1. Samples of 3D objects with geometric distortions (translation, rotation) and quality degradations (motion blur)

We found that the work of Murase and Nayar in [7] is more favorable, due to its simplicity and applicability to more general pose variation problems. Thus, in this paper, we put our focus on their work. However, the disadvantage of their Parametric Eigenspace approach is the model only works well when the input images have no degraded effects. Unfortunately, this assumption is not realistic in real-world applications. Some degradation effects usually occur and contaminate the original images during the capturing process or segmentation process. Thus, relying on a simple manifold to handle this problem is not sufficient. Fig. 1 shows some image samples of 3D objects with some geometric distortion and quality degradation effects.

We have showed that to overcome this issue, constructing an appearance manifold with embedded covariance matrix is very useful. By using this appearance manifold model, the robustness of the system will be increased, since the manifold could capture the pose changes and the embedded covariance matrix could define the sample distribution information of every pose along the manifold. Moreover, since the appearance of an object in the captured image is different for every different pose, the covariance matrix value is also different for every pose. Thus, it is necessary to construct covariance matrix which is view-dependent.

Considering the models previously proposed by us in [11], a major limitation still exists, such as the need to critically control the correspondence between learning points in the interpolation process for manifold construction. When a huge number of learning points exists in the system, the controlling process becomes costly and time consuming.

In this paper, we propose the View-dependent Covariance matrix by Eigenvector Interpolation (VCEI) method where every mean vector and covariance matrix has different value for each training pose. The advantage of our proposed method is that it is not necessary to perform controlling process on the training images. In order to cover the untrained poses, we construct the appearance manifold by interpolating only the eigenvectors and eigenvalues of two consecutive trained poses. Thus, the appearance manifold will be noise-invariant and efficient.

The remainder of this paper is organized as follows: we describe the construction process of manifold in Section 2. Section 3 presents the description of embedding process of view-dependent covariance matrix in an object manifold. Section 4 shows and discusses the performance evaluation in recognizing 3D objects. Finally, Section 5 presents our conclusion.

2. Manifold of Object

Generally, the appearance-based approaches deal with a set of learning images in various capturing conditions. Since these images are usually high-dimensional images, 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*.

Next, the first *k* eigenvectors will be used to project *S* learning samples of *P* objects with *H* poses. Thus, with $\mathbf{x}_s^{(p)}(\ _h)$ the *s* sample image of object *p* with horizontal viewpoint $_h$ and $\mathbf{e}_i$ are the eigenvectors, the new feature vectors $\mathbf{g}_s^{(p)}(\ _h)\ ^k$ are defined by $\mathbf{g}_s^{(p)}(\ _h)\ [\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_k]^T(\mathbf{x}_s^{(p)}(\ _h)\ \mathbf{c})$ where $\mathbf{c}$ is the mean vector of sample images. These eigenvectors $\mathbf{e}_i$ were obtained by solving the eigen decomposition $_i\mathbf{e}_i\ \mathbf{Q}\mathbf{e}_i$, where $\mathbf{Q}$ is the auto correlation matrix of the training set and $_i$ is the eigenvalue associated with the eigenvector $\mathbf{e}_i$. Note that in this section, the eigenvectors and eigenvalues are used only to construct the eigenspace. Meanwhile, later in the next sections, the eigenvectors and eigenvalues are derived from the covariance matrix of each training-pose.

In the Simple Manifold (SM) method, after transforming learning images onto the eigenspace, the manifold of an object could be obtained by interpolating the mean vector of training images of one pose to its consecutive poses. Practically, we can simply apply an interpolation method to construct the manifold between training images in two-consecutive poses. Fig. 2 shows the illustration of the construction process of a simple manifold of an object in the feature space.

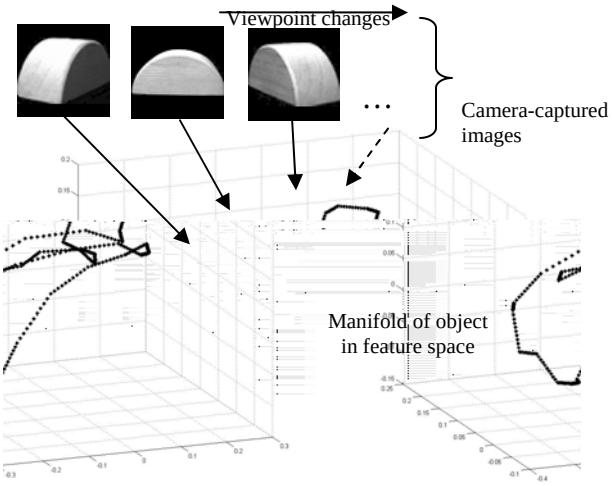


Figure 2. A simple manifold of an object in feature space

3. Embedding View-dependent Covariance Matrix in Object Manifold

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.

The construction process of View-dependent Covariance matrix by Eigenvector Interpolation (VCEI) method consists of two stages of interpolation process: the interpolation of mean vectors and the interpolation of eigenvectors and eigenvalues. The mean vector is used to represent the center point of samples in each learning pose, while the eigenvectors and eigenvalues represent the distribution of samples in each pose. The interpolation process is useful to cover the information of untrained poses.

Basically, the interpolation process of the mean vector can be done by simply using one of the several existing interpolation algorithms. Meanwhile, the interpolation process of eigenvectors and eigenvalues are done based on high-dimensional rotation theory. As the eigenvectors and eigenvalues can be considered as axes directions and lengths of a hyper-ellipsoid in the eigenspace, we consider obtaining the covariance matrices of untrained poses by rotating the hyper-ellipsoids of two-consecutive trained poses. Fig. 3 shows the 2D illustration of the interpolation process of eigenvectors and eigenvalues in the feature space. Meanwhile, Fig. 4 shows the construction process of an appearance manifold with view-dependent embedded covariance matrix. Here, we

only use the horizontal pose parameter ($_h$) to construct the appearance manifold.

The algorithm for interpolating the eigenvectors and eigenvalues are summarized as follows:

Input: $\mathbf{E}_0$ and $\mathbf{E}_1$ are matrices formed by aligning each pair of eigenvectors $\mathbf{e}_{0j}$ and $\mathbf{e}_{1j}$, while $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ are matrices formed by aligning each pair of eigenvalues $_{0j}$ and $_{1j}$ (j=1, 2, .., k). The covariance matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ represent the sample distribution of two consecutive poses.

1. Sort the eigenvectors $\mathbf{E}_0$ and $\mathbf{E}_1$ in the decreasing order according to their eigenvalues $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$ to obtain $\mathbf{E}_0$ and $\mathbf{E}_1$, and also $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_1$
2. Check the angle between the corresponded axes so that it is less than or equal to 0.5 :
   if $\mathbf{e}_{0j}^T \mathbf{e}_{1j}$ 0 then invert $\mathbf{e}_{1j} (j \quad 1,2,...,k)$
3. For covariance matrix $\boldsymbol{\Sigma}_x$, do calculation for the eigenvalue interpolation with
   $_{xj} \quad 1 \quad x\sqrt{\ _{0j}} \quad x\sqrt{\ _{1j}}^{\ 2}$
4. For covariance matrix $\boldsymbol{\Sigma}_x$, do calculation for the eigenvector interpolation with $\mathbf{E}_x \quad \mathbf{R}(x\ )\mathbf{E}_0$
   where $\mathbf{R}$ represents an interpolated rotation when $0 \quad x \quad 1$ and $_1,..., _m$ represents the parameter vector of rotation angles to define the rotation matrix. Here, $m \quad n/2$ since the rotation angles always come in pairs in the complex conjugate roots process.
   (a) Define the rotation matrix by
   $\mathbf{R}(\ ) \quad \mathbf{E}_1\mathbf{E}_0^T$
   (b) Diagonalize $\mathbf{R}(\ )$ with Special Orthogonal (SO) rule $\mathbf{R}(\ ) \quad \mathbf{UD}(\ )\mathbf{U}$
   where $\mathbf{U}$ is the conjugate transpose of $\mathbf{U}$
   (c) Process complex conjugate roots:
   if $n \quad 2m$ then
   $\mathbf{D}(\ ) \quad diag(\mathbf{e}^{i\ _1}, \mathbf{e}^{\ i\ _1},..., \mathbf{e}^{i\ _m}, \mathbf{e}^{\ i\ _m})$
   however, if $n \quad 2m \quad 1$ then
   $\mathbf{D}(\ ) \quad diag(1, \mathbf{e}^{i\ _1}, \mathbf{e}^{\ i\ _1},..., \mathbf{e}^{i\ _m}, \mathbf{e}^{\ i\ _m})$
   where $\mathbf{e}^i \quad \cos \quad i\sin$
   (d) Apply linear interpolation technique to obtain
   $\mathbf{R}(x\ )$

Output: Covariance matrix for untrained poses $\boldsymbol{\Sigma}_x \quad \mathbf{E}_x\boldsymbol{\Lambda}_x\mathbf{E}_x^T$ where $\boldsymbol{\Lambda}_x \quad diag(\boldsymbol{\lambda}_x)$

In the recognition stage, in order to recognize an input image $\mathbf{z}$, we calculate its distance to the manifold of object p by Mahalanobis distance measurement defined in this formula:

$$d^{(p)}(\mathbf{z}) \quad \min \ (\mathbf{z} \quad \tilde{\boldsymbol{\mu}}^{(p)}(\ ))^T(\tilde{}^{(p)}(\ ))^{\ 1}(\mathbf{z} \quad \tilde{\boldsymbol{\mu}}^{(p)}(\ )) \quad (1)$$

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

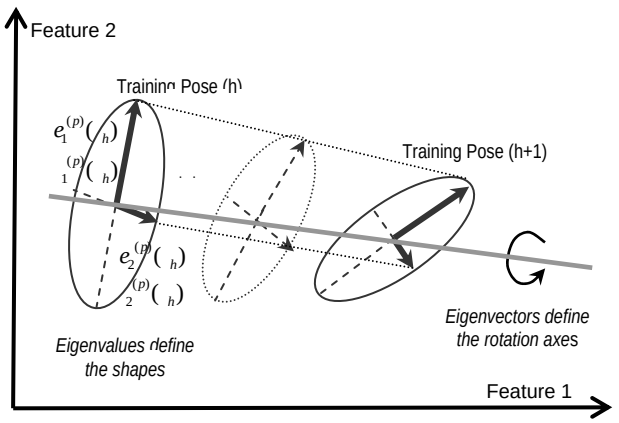


Figure 3. Illustration of an interpolation process of eigenvectors and eigenvalues in 2D feature space

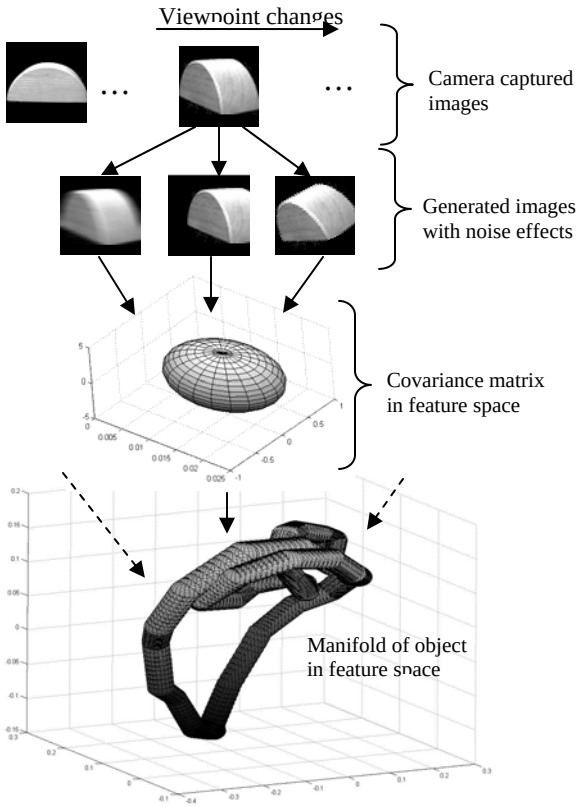


Figure 4. An appearance manifold with view-embedded covariance matrix of a single object in feature space

4. Performance Evaluation

To evaluate the performance of our proposed method, we developed an application system for 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.

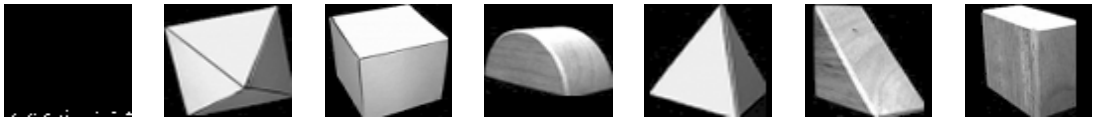


Figure 5. Samples of 3D objects used in the experiments

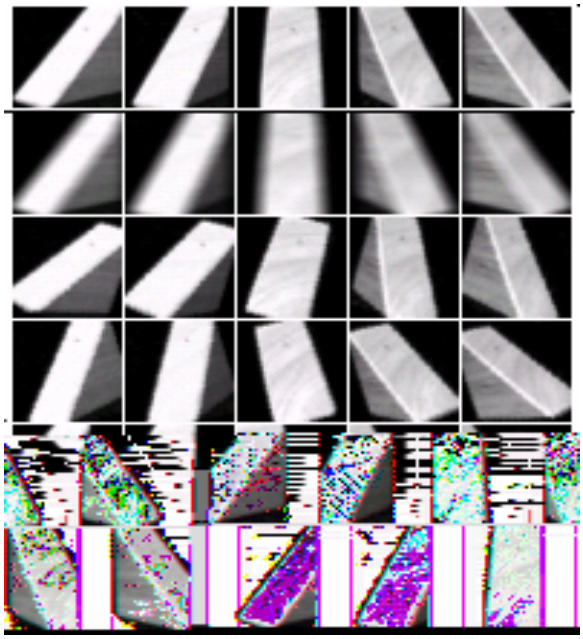


Figure 6. Samples of training images of an object

Samples of 3D objects used in our experiments are shown in Fig. 5, while Fig. 6 shows the samples of training images of an object.

In the learning stage, a total of 6,552 images were normalized into 32x32pixels grayscale images. Each object consists of 36 poses with 10-degree intervals of horizontal positions (0^o, 10^o, 20^o, …, 350^o), and each pose consists of an original camera-captured image and 25 generated images. The generated images were obtained by composing artificial noises, such as left and right translations (3, 6, 9, 12, 15pixels), clockwise and counter-clockwise rotations (5^o, 10^o, 15^o, 20^o, 25^o), and motion blur (5%, 10%, 15%, 20%, 25%).

After projecting the images onto the eigenspace, the appearance manifolds were created based on the simple manifold (SM) method and appearance manifold with View-dependent Covariance matrix by Eigenvector Interpolation (VCEI) method. In the SM method, the mean vectors became the center of the manifold and an identity matrix was applied as the covariance matrices for every pose along the manifold. Meanwhile, in the VCEI method, the mean vectors were set as the center of the manifold, but the covariance matrices were obtained by interpolating the eigenvectors and eigenvalues as described in Section 3. Here, we applied linear interpolation technique to construct the manifold of mean vectors.

Finally, we tested the system with input images that were different from the learning images (5^o, 15^o, 25^o, …, 355^o) in horizontal poses and influenced by various types of degradation effects. For classification, we employed the Mahalanobis distance measurement shown in Eq.1 for the VCEI method.

Fig. 7, Fig. 8, and Fig. 9 show a series of the recognition accuracies of two appearance manifold methods in recognizing images influenced with translation effects, rotation effects, and motion blur effects, respectively. All figures indicate that the proposed VCEI method achieved higher recognition accuracies than the SM method for all types of degradation effects. For recognizing non-degraded images, the VCEI method achieved 92.06% recognition accuracy, while the SM method only achieved 76.19%. When recognizing images with various geometric distortion and quality degradation effects, the VCEI method also could still achieve high recognition results, especially in blur and rotation cases.

From our observation, we also found that the SM method gave good recognition results only for objects having distinct appearance and major differences in shape. Meanwhile, the VCEI method could recognize objects with similar appearances. Fig. 10 shows the example of cases where the SM method failed but the VCEI method succeeded.

To observe the overall performance stability of both the SM and the VCEI methods, we show the average recognition accuracy for each method related to each degradation effect in Fig. 11. For the translation effect case, the average recognition accuracy of the VCEI method was 70.57%, higher than that of the SM method with 49.41%. In the case of the rotation effect, the average recognition accuracy of the VCEI method was 83.73%, still higher than the average recognition accuracy of the SM method. Finally, for the motion blur case, the VCEI method also gave higher average recognition accuracy with 93.02% compared with that of the SM method with 76.75%. From Fig. 11, we could also learn that for both the SM and the VCEI methods, the robustness level of both methods to the motion blur effect was the highest, followed by the rotation effects and then the translation effects. Thus, images which are affected with the translation effects seemed to be the most difficult to recognize. Meanwhile, images which are influenced with the motion blur effects seemed to be the easiest to recognize.

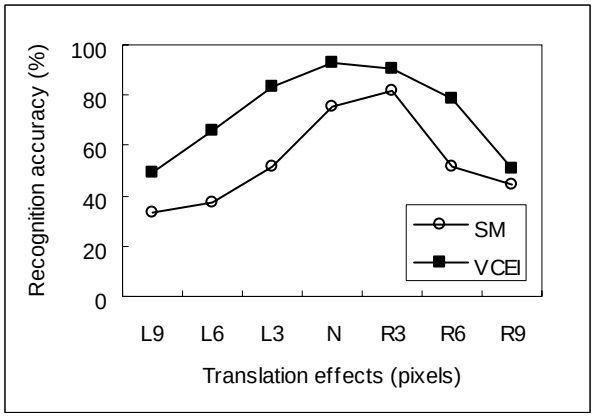


Figure 7. Recognition accuracies of images with translation effects

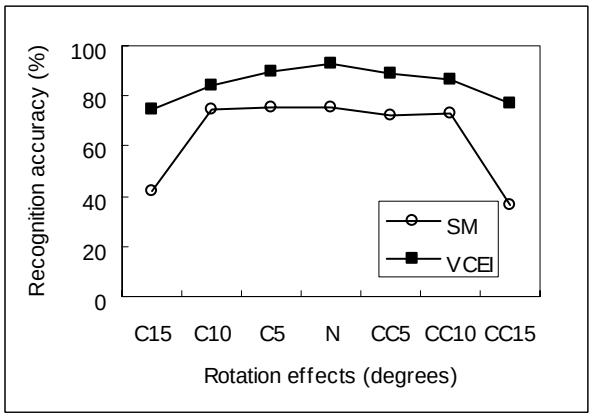


Figure 8. Recognition accuracies of images with rotation effects

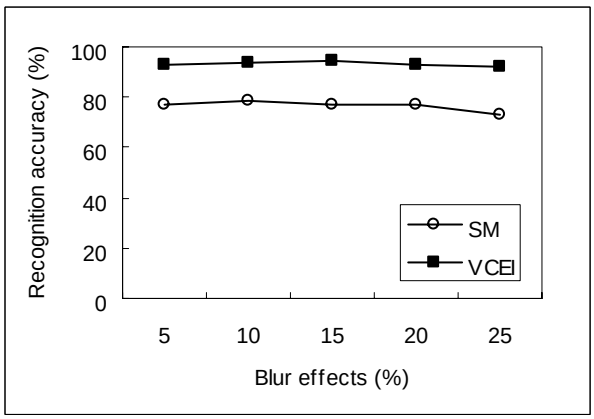
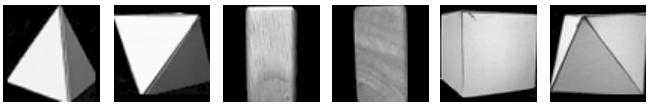


Figure 9. Recognition accuracies of images with motion blur effects



(a) (b) (c) (d) (e) (f)

Figure 10. Sample of cases where SM method failed but VCEI method succeed. (Using SM method, object (b) was recognized as object (a), and objects (d) and (f) were recognized as objects (c) and (e), respectively.)

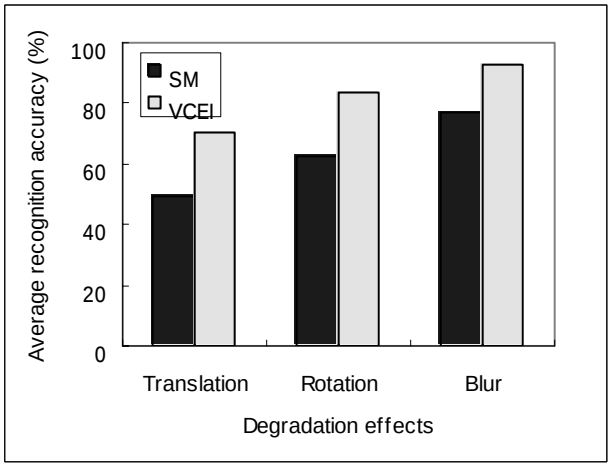


Figure 11. Average recognition accuracy for each degradation effect

5. Conclusion

We proposed the construction of appearance manifold with embedded View-dependent Covariance matrix by Eigenvector Interpolation (VCEI) method and developed its application to recognize 3D objects from images that are influenced by geometric distortions and quality-degraded effects. This method is based on the eigenvectors and eigenvalues interpolation to form a view-dependent covariance matrix model. In our proposed method, it is not necessary to control the correspondence between learning points in each pose to their consecutive poses. Thus, two advantages of this VCEI method are its robustness and efficiency.

Our future works include recognizing 3D human faces 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] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, Face recognition: a literature survey, *ACM Computing Surveys, 35(4)*, 2003, 399-458.

[2] L. Wiskott, J.-M. Fellous, and C. Von Der Malsburg, Face recognition by elastic bunch graph matching, *IEEE Transaction on Pattern Analysis and Machine Intelligence, 19(7)*, 1997, 775-779.

[3] A. Lanitis, C.J. Taylor, and T.F. Cootes, Automatic face identification system using flexible appearance models, *Image and Vision Computing, 13(5)*, 1995, 393-401.

[4] W. Zhao and R. Chellappa, SFS based view synthesis for robust face recognition. *Proc. 4th Conf. on Automatic Face and Gesture Recognition*, Washington DC, USA, 2000, 285-290.

[5] M. Turk and A. Pentland, Face recognition using eigenfaces, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Fort Collins, USA, 1991, 586-591.

[6] S.K. Nayar, S.A. Nene, and H. Murase, Real-time 100 object recognition system, *Proc. IEEE Conf. on Robotics and Automation*, Minneapolis, USA, 1996, 2321-2325.

[7] H. Murase and S.K. Nayar, Illumination planning for object recognition using parametric eigenspaces, *IEEE Transaction on Pattern Analysis and Machine Intelligence, 16(12)*, 1994, 1219-1227.

[8] B. Moghaddam and A. Pentland, Probabilistic visual learning for object representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence, 19(7)*, 1997, 696-710.

[9] A.M Martinez, Recognition of partially occluded and/or imprecisely localized faces using a probabilistic approach, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head, USA, 2000, 712-717.

[10] B.J. Frey, M. Jojic, and A. Kanan, Learning appearance and transparency manifolds of occluded objects in layers, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Madison, USA, 2003, 45-52.

[11] Lina, T. Takahashi, I. Ide, H. Murase, Appearance manifold with embedded covariance matrix for robust 3D object recognition, *Proc. IAPR Conf. on Machine Vision Applications 2007*, Tokyo, Japan, 504-507.