Learning and Recognition of 3-D Objects from Brightness Images *

Hiroshi Murase
NTT Basic Research Labs
3-9-11 Midori-cho, Musashino-shi
Tokyo 180, Japan

Shree K. Nayar
Department of Computer Science
Columbia University
New York, N.Y. 10027

Abstract

We address the problem of automatically learning object models for recognition and pose estimation. In contrast to the traditional approach, the recognition problem is formulated here as one of matching visual appearance rather than shape. The appearance of an object in a two-dimensional image depends on its shape, reflectance properties, pose in the scene, and the illumination conditions. While shape and reflectance are intrinsic properties and are constant for a rigid object, pose and illumination vary from scene to scene. We present a new compact representation of object appearance that is parametrized by pose and illumination. For each object of interest, a large set of images is obtained by automatically varying pose and illumination. This large image set is compressed to obtain a low-dimensional subspace, called the eigenspace, in which the object is represented as a hypersurface. Given an unknown input image, the recognition system projects the image onto the eigenspace. The object is recognized based on the hypersurface it lies on. The exact position of the projection on the hypersurface determines the object's pose in the image.

Introduction

For a vision system to be able to recognize objects, it must have models of the objects stored in its memory. In the past, vision research has emphasized on the use of geometric (shape) models [1] for recognition. In the case of manufactured objects, these models are sometimes available and are referred to as computer aided design (CAD) models. Most objects of interest, however, do not come with CAD models. Typically, a vision programmer is forced to select an appropriate representation for object geometry, develop object models using this representation, and then manually input this information into the system. This procedure is cumbersome and impractical when dealing with large sets of objects, or objects with complicated geometric properties. It is clear that recognition systems of the future must be capable of *learning* object models without human assistance.

Visual learning is clearly a well-developed and vital component of biological vision systems. If a human is handed an object and asked to visually memorize it, he or she would rotate the object and study its appearance from different directions. While little is known about the exact representations and techniques used by the human mind to learn objects, it is clear that the overall appearance of the object plays a critical role in its perception. In contrast to biological systems, machine vision systems today have little or no learning capabilities. Hence, visual learning is now emerging as an topic of research interest [6]. The goal of this paper is to advance this important but relatively unexplored area of machine vision.

*This paper was presented at the 1993 AAAI Conference held in Washington D.C. This research was conducted at the Center for Research in Intelligent Systems, Department of Computer Science, Columbia University. It was supported in part by the David and Lucile Packard Fellowship and in part by ARPA Contract No. DACA 76-92-C-0007.

Here, we present a technique for automatically learning object models from images. The appearance of an object is the combined effect of its shape, reflectance properties, pose in the scene, and the illumination conditions. While shape and reflectance are *intrinsic properties* that do not vary for a rigid object, pose and illumination vary from scene to scene. We approach the visual learning problem as one of acquiring a compact model of the object's appearance under different illumination directions and object poses. The object is "shown" to the image sensor in several orientations and illumination directions. The result is a very large set of images. Since all images in the set are of the same object, any two consecutive images are correlated to large degree. The problem then is to compress this large image set into a low-dimensional representation of object appearance.

A well-known *image compression* or coding technique is based on principal component analysis. Often referred to as the Karhunen-Loeve transform [5] [2], this method computes the eigenvectors of an image set. The eigenvectors form an orthogonal basis for the representation of individual images in the image set. Though a large number of eigenvectors may be required for very accurate reconstruction of an image, only a few eigenvectors are generally sufficient to capture the significant appearance characteristics of an object. These eigenvectors constitute the dimensions of what we refer to as the *eigenspace* for the image set. From the perspective of machine vision, the eigenspace has a very attractive property. When it is composed of all the eigenvectors of an image set, it is optimal in a *correlation* sense: If any two images from the set are projected onto the eigenspace, the distance between the corresponding points in eigenspace is a measure of the similarity of the images in the l^2 *norm*. In machine vision, the Karhunen-Loeve method has been applied primarily to two problems; handwritten character recognition [3] and human face recognition [8], [9]. These applications lie within the domain of pattern classification and do not use complete parametrized models of the objects of interest.

In this paper, we develop a continuous and compact representation of object appearance that is parametrized by the variables, namely, object pose and illumination. This new representation is referred to as the *parametric eigenspace*. First, an image set of the object is obtained by varying pose and illumination in small increments. The image set is then normalized in brightness and scale to achieve invariance to image magnification and the intensity of illumination. The eigenspace for the image set is obtained by computing the most prominent eigenvectors of the set. Next, all images in the set (the learning samples) are projected onto the eigenspace to obtain a set of discrete points. These points lie on a *hypersurface* that is parametrized by object pose and illumination. The hypersurface is computed from the discrete points by interpolation.

Each object is represented as a parametric hypersurface in two different eigenspaces. The *universal eigenspace* is computed by using the image sets of all objects of interest to the recognition system, and the *object eigenspace* is computed us-

ing only images of the object. Recognition and pose estimation can be summarized as follows. Given an image consisting of an object of interest, we assume that the object is not occluded by other objects and can be segmented from the remaining scene. The segmented image region is normalized in scale and brightness, such that it has the same size and brightness range as the images used in the learning stage. This normalized image is first projected onto the universal eigenspace to identify the object. After the object is recognized, the image is projected onto the object's eigenspace and the location of the projection on the object's parametrized hypersurface determines its pose in the scene.

The fundamental contributions of this paper can be summarized as follows. (a) The parametric eigenspace is presented as a new representation of object appearance. (b) Using this representation, object models are automatically learned from appearance by varying pose and illumination. (c) Both learning and recognition are accomplished without prior knowledge of the object's shape and reflectance. Several experiments have been conducted using objects with complex appearance characteristics and the results are very encouraging.

Visual Learning of Objects

Normalized Image Sets

While constructing image sets we need to ensure that all images are of the same size. Each digitized image is first segmented (using a threshold) into an object region and a background region. The background is assigned a zero brightness value and the object region is re-sampled such that the larger of its two dimensions fits the image size we have selected for the image set representation. We now have a scale normalized image. This image is written as a vector $\hat{\mathbf{x}}$ by reading pixel brightness values in a raster scan manner:

$$\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, ..., \hat{x}_N]^T \quad (1)$$

The appearance (brightness image) of an object depends on its shape and reflectance properties. These are intrinsic properties that do not vary for a rigid object. The object's appearance also depends on its pose and the illumination conditions. Unlike the intrinsic properties, object pose and illumination are expected to vary from scene to scene. Here, we assume that the object is illuminated by the ambient lighting of the environment as well as one additional distant light source whose direction may vary. Hence, all possible appearances of the object can be captured by varying object pose and the light source direction with respect to the viewing direction of the sensor. We denote each image as $\hat{\mathbf{x}}^{(p)}_{r,l}$ where r is the rotation or pose parameter, l represents the illumination direction, and p is the object number. The complete image set obtained for an object is referred to as the **object image set** and can be expressed as:

$$\left\{ \hat{\mathbf{x}}^{(p)}_{1,1}, ..., \hat{\mathbf{x}}^{(p)}_{R,1}, \hat{\mathbf{x}}^{(p)}_{1,2}, ..., \hat{\mathbf{x}}^{(p)}_{R,L} \right\} \quad (2)$$

Here, R and L are the total number of discrete poses and illumination directions, respectively, used to obtain the image set. If a total of P objects are to be learned by the recognition system, we can define the **universal image set** as the union of all the object image sets:

$$\{ \hat{\mathbf{x}}^{(1)}_{1,1}, ..., \hat{\mathbf{x}}^{(1)}_{R,1}, \hat{\mathbf{x}}^{(1)}_{1,2}, ..., \hat{\mathbf{x}}^{(1)}_{R,L}, \quad (3)$$
$$\hat{\mathbf{x}}^{(2)}_{1,1}, ..., \hat{\mathbf{x}}^{(2)}_{R,1}, \hat{\mathbf{x}}^{(2)}_{1,2}, ..., \hat{\mathbf{x}}^{(2)}_{R,L},$$
$$\vdots$$
$$\hat{\mathbf{x}}^{(P)}_{1,1}, ..., \hat{\mathbf{x}}^{(P)}_{R,1}, \hat{\mathbf{x}}^{(P)}_{1,2}, ..., \hat{\mathbf{x}}^{(P)}_{R,L} \}$$

We assume that the imaging sensor used for learning and recognizing objects has a linear response, i.e. image brightness is proportional to scene radiance. We would like our recognition system to be unaffected by variations in the intensity of illumination or the aperture of the imaging system. This can be achieved by normalizing each of the images in the object and universal sets, such that, the total energy contained in the image is unity. This brightness normalization transforms each measured image $\hat{\mathbf{x}}$ to a normalized image $\mathbf{x}$, such that $\| \mathbf{x} \| = 1$. The above described scale and brightness normalizations give us normalized object image sets and a normalized universal image set. In the following discussion, we will simply refer to these as the object and universal image sets.

The image sets can be obtained in several ways. We assume that we have a sample of each object that can be used for learning. One approach then is to use two robot manipulators; one grasps the object and shows it to the sensor in different poses while the other has a light source mounted on it and is used to vary the illumination direction. In our experiments, we have used a turntable to rotate the object in a single plane (see Fig. 1). This gives us pose variations about a single axis. A robot manipulator is used to vary the illumination direction. If the recognition system is to be used in an environment where the illumination (due to one or several sources) is not expected to change, the image set can be obtained by varying just object pose.

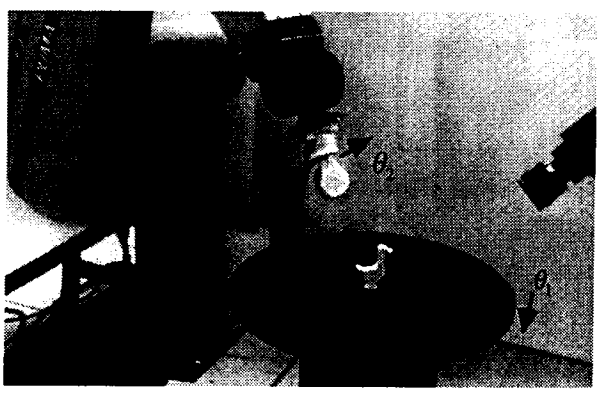


Figure 1: Setup used for automatic acquisition of object image sets. The object is placed on a motorized turntable.

Computing Eigenspaces

Consecutive images in an object image set tend to be correlated to a large degree since pose and illumination variations between consecutive images are small. Our first step is to take advantage of this correlation and compress large image sets into low-dimensional representations that capture the gross appearance characteristics of objects. A suitable compression technique is the Karhunen-Loeve transform [2] where the eigenvectors of the image set are computed and used as orthogonal basis functions for representing individual images.

Two types of eigenspaces are computed; the universal eigenspace that is obtained from the universal image set, and object eigenspaces computed from individual object image sets. To compute the universal eigenspace, we first subtract the average of all images in the universal set from each image. This ensures that the eigenvector with the largest eigenvalue represents the dimension in eigenspace in which the variance of images is maximum in the correlation sense. In other words, it is the most important dimension of the eigenspace. The average of all images in the universal image set is determined

as:

$$\mathbf{c} = \frac{1}{RLP} \sum_{p=1}^{P} \sum_{r=1}^{R} \sum_{l=1}^{L} \mathbf{x}_{r,l}^{(p)} \quad (4)$$

A new image set is obtained by subtracting the average image $\mathbf{c}$ from each image in the universal set:

$$\mathbf{X} \triangleq \{ \mathbf{x}_{1,1}^{(1)} - \mathbf{c}, \ldots, \mathbf{x}_{R,1}^{(1)} - \mathbf{c}, \ldots, \mathbf{x}_{R,L}^{(P)} - \mathbf{c} \} \quad (5)$$

The matrix $\mathbf{X}$ is $N \times M$, where $M = RLP$ is the total number of images in the universal set, and N is the number of pixels in each image. To compute eigenvectors of the image set we define the *covariance matrix* as:

$$\mathbf{Q} \triangleq \mathbf{X}\mathbf{X}^{\mathrm{T}} \quad (6)$$

The covariance matrix is $N \times N$, clearly a very large matrix since a large number of pixels constitute an image. The eigenvectors $\mathbf{e}_i$ and the corresponding eigenvalues λ_i of $\mathbf{Q}$ are to be determined by solving the well-known eigenvector decomposition problem:

$$\lambda_i \mathbf{e}_i = \mathbf{Q}\mathbf{e}_i \quad (7)$$

All N eigenvectors of the universal set together constitute a complete eigenspace. Any two images from the universal set, when projected onto the eigenspace, give two discrete points. The distance between these points is a measure of the difference between the two images in the correlation sense. Since the universal eigenspace is computed using images of all objects, it is best tuned for discriminating between images of different objects.

Determining the eigenvalues and eigenvectors of a large matrix such as $\mathbf{Q}$ is a non-trivial problem. It is computationally very intensive and traditional techniques used for computing eigenvectors of small matrices are impractical. Since we are interested only in a small number (k) of eigenvectors, and not the complete set of N eigenvectors, efficient algorithms can be used. In our implementation, we have used the *spatial temporal adaptive* (STA) algorithm proposed by Murase and Lindenbaum [4]. This algorithm was recently demonstrated to be substantially more efficient than previous algorithms. The result is a set of eigenvalues $\{ \lambda_i \mid i = 1, 2, \ldots, k \}$ where $\{ \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \}$, and a corresponding set of eigenvectors $\{ \mathbf{e}_i \mid i = 1, 2, \ldots, k \}$. Note that each eigenvector is of size N, i.e. the size of an image. These k eigenvectors constitute the universal eigenspace; it is an approximation to the complete eigenspace with N dimensions. We have found from our experiments that less than ten dimensions are generally sufficient for the purposes of visual learning and recognition (i.e. $k \leq 10$). Later, we describe how objects in an unknown input image are recognized using the universal eigenspace.

Once an object has been recognized, we are interested in finding its pose in the image. The accuracy of pose estimation depends on the ability of the recognition system to discriminate between different images of the same object. Hence, pose estimation is best done in an eigenspace that is tuned to the appearance of a single object. To this end, we compute an object eigenspace from each of the object image sets. In this case, the average $\mathbf{c}^{(p)}$ of all images of object p is computed and subtracted from each of the object images. The resulting images are used to compute the covariance matrix $\mathbf{Q}^{(p)}$. Once again, we compute only a small number ($k \leq 10$) of the largest eigenvalues $\{ \lambda_i^{(p)} \mid i = 1, 2, \ldots, k \}$ where $\{ \lambda_1^{(p)} \geq \lambda_2^{(p)} \geq \ldots \geq \lambda_k^{(p)} \}$, and a corresponding set of eigenvectors $\{ \mathbf{e}_i^{(p)} \mid i = 1, 2, \ldots, k \}$. An object eigenspace is computed for each object of interest to the recognition system.

Parametric Eigenspace Representation

We now represent each object as a hypersurface in the universal eigenspace as well as its own eigenspace. This new representation of appearance lies at the core of our approach to visual learning and recognition. A parametric hypersurface for the object p is constructed in the universal eigenspace as follows. Each image $\mathbf{x}_{r,l}^{(p)}$ (learning sample) in the object image set is projected onto the eigenspace by first subtracting the average image $\mathbf{c}$ from it and finding the dot product of the result with each of the eigenvectors (dimensions) of the universal eigenspace. The result is a point $\mathbf{g}_{r,l}^{(p)}$ in the eigenspace:

$$\mathbf{g}_{r,l}^{(p)} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]^{\mathrm{T}} (\mathbf{x}_{r,l}^{(p)} - \mathbf{c}) \quad (8)$$

Once again the subscript r represents the rotation parameter and l is the illumination direction. By projecting all the learning samples in this manner, we obtain a set of discrete points in the universal eigenspace. Since consecutive object images are strongly correlated, their projections in eigenspace are close to one another. Hence, the discrete points obtained by projecting all the learning samples can be assumed to lie on a k-dimensional hypersurface that represents all possible poses of the object under all possible illumination directions. We interpolate (using cubic splines, for instance) the discrete points to obtain this hypersurface. The resulting hypersurface can be expressed as $\mathbf{g}^{(p)}(\theta_1, \theta_2)$ where θ_1 and θ_2 are now *continuous* rotation and illumination parameters. The above hypersurface is a compact representation of the object's appearance.

In a similar manner, a hypersurface is also constructed in the object's eigenspace by projecting the learning samples onto this space:

$$\mathbf{f}_{r,l}^{(p)} = \left[\mathbf{e}_1^{(p)}, \mathbf{e}_2^{(p)}, \ldots, \mathbf{e}_k^{(p)}\right]^{\mathrm{T}} (\mathbf{x}_{r,l}^{(p)} - \mathbf{c}^{(p)}) \quad (9)$$

The discrete points $\mathbf{f}_{r,l}^{(p)}$ are interpolated to obtain the hypersurface $\mathbf{f}^{(p)}(\theta_1, \theta_2)$. This continuous parameterization enables us to find poses of the object that are not included in the learning samples. It also enables us to compute accurate pose estimates under illumination directions that lie in between the discrete illumination directions used in the learning stage.

Recognition and Pose Estimation

Consider an image of a scene that includes one or more of the objects we have learned. We assume that the objects are not occluded by other objects in the scene when viewed from the sensor direction, and that the image regions corresponding to objects have been segmented away from the scene image. First, each segmented image region is normalized with respect to scale and brightness as described in the previous section. This ensures that (a) the input image has the same dimensions as the eigenvectors (dimensions) of the parametric eigenspace, (b) the recognition system is invariant to object magnification, and (c) the recognition system is invariant to fluctuations in the intensity of illumination.

A normalized image region $\mathbf{y}$ is first projected to the universal eigenspace to obtain a point:

$$\mathbf{z} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]^{\mathrm{T}} (\mathbf{y} - \mathbf{c}) = [z_1, z_2, \ldots, z_k]^{\mathrm{T}} \quad (10)$$

The recognition problem then is to find the object p whose hypersurface the point $\mathbf{z}$ lies on. Due to factors such as image noise, aberrations in the imaging system, and quantization effects, $\mathbf{z}$ may not lie exactly on an object hypersurface. Hence,

we find the object p that gives the minimum distance $d_1^{(p)}$ between its hypersurface $\mathbf{g}^{(p)}(\theta_1, \theta_2)$ and the point $\mathbf{z}$:

$$d_1^{(p)} = \min_{\theta_1, \theta_2} \| \mathbf{z} - \mathbf{g}^{(p)}(\theta_1, \theta_2) \| \quad (11)$$

If $d_1^{(p)}$ is within some pre-determined threshold value then the input image is of object p. If not, we conclude that input image is not of any of the objects used in the learning stage.

Once the object in the input image $\mathbf{y}$ is recognized, we project $\mathbf{y}$ to the eigenspace of the object. This eigenspace is tuned to variations in the appearance of a single object and hence is ideal for pose estimation. Mapping $\mathbf{y}$ to the object eigenspace again results in a point $\mathbf{z}^{(p)}$. The pose estimation problem may be stated as follows: Find the rotation parameter θ_1 and the illumination parameter θ_2 that minimize the distance $d_2^{(p)}$ between the point $\mathbf{z}^{(p)}$ and the hypersurface $\mathbf{f}^{(p)}$ of the object p:

$$d_2^{(p)} = \min_{\theta_1, \theta_2} \| \mathbf{z} - \mathbf{f}^{(p)}(\theta_1, \theta_2) \| \quad (12)$$

The θ_1 value obtained represents the pose of the object in the input image. Note that the recognition and pose estimation stages are computationally very efficient, each requiring only the projection of an input image onto a low-dimensional (generally less than 10) eigenspace. Customized hardware can therefore be used to achieve real-time (frame-rate) performance.

Experimentation

Fig. 1 in the introduction shows the set-up used to conduct the experiments reported here. The object is placed on a motorized turntable and its pose is varied about a single axis, namely, the axis of rotation of the turntable. The turntable position is controlled through software and can be varied with an accuracy of about 0.1 degrees. Most objects have a finite number of stable configurations when placed on a planar surface. For such objects, the turntable is adequate as it can be used to vary pose for each of the object's stable configurations. We assume that the objects are illuminated by ambient light as well as one additional source whose direction can vary (see Fig. 1).

Table 1 summarizes the number of objects, light source directions, and poses used to acquire the image sets used in the experiments. For the learning stage, a total of 4 objects were used. These objects (cars) are shown in Fig. 2(a). For each object we have used 5 different light source directions, and 90 poses for each source direction. This gives us a total of 1800 images in the universal image set and 450 images in each object image set. Each of these images is automatically normalized in scale and brightness to obtain a 128×128 pixel image. The universal and object image sets are used to compute the universal and object eigenspaces. The parametric eigenspace representations of the four objects in their own eigenspaces are shown in Fig. 2(b).

Table 1: Image sets obtained for the learning and recognition experiments. The 1080 test images used for recognition are different from the 1800 images used for learning.

Learning samples	Test samples for recognition
4 objects	4 objects
5 light source directions	3 light source directions
90 poses	90 poses
1800 images	1080 images

A large number (1080) of images were also obtained to test the recognition and pose estimation algorithms. All of these images are different (in pose and illumination) from the ones used in the learning stage. Each test image is first normalized in scale and brightness and then projected onto the universal eigenspace. The object in the image is identified by finding the closest hypersurface. Unlike the learning process, recognition is computationally simple and can be accomplished on a Sun SPARC 2 workstation in less than 0.2 seconds.

Fig. 3(a) illustrates the sensitivity of the recognition rate (percentage of correctly recognized test images) to the number of dimensions of the universal eigenspace. Clearly, the discriminating power of the universal eigenspace is expected to increase with the number of dimensions. For the objects used, the recognition rate is poor if less than 4 dimensions are used but approaches unity as the number of dimensions approaches 10. In general, however, the number of dimensions needed for robust recognition is expected to increase with the number of objects learned by the system. It also depends on the appearance characteristics of the objects used. From our experience, 10 dimensions are sufficient for representing objects with fairly complex appearance characteristics such as the ones shown in Fig. 2.

Finally, we present experimental results related to pose estimation. Once again we have used all 1080 test images of the 4 objects. Since these images were obtained using the controlled turntable, the actual object pose in each image is known. Fig. 3(b) shows the histogram of the errors (in degrees) in the poses computed for the 1080 images. Here, 450 learning samples (90 poses and 5 source directions) were used to compute 8-dimensional object eigenspaces. This result demonstrate remarkable accuracy; the absolute pose error computed using all 1080 images is 0.5 degrees.

Conclusion

In this paper, we presented a new representation for machine vision called the parametric eigenspace. While representations previously used in computer vision are based on object geometry, the proposed one describes object appearance. We presented a method for automatically learning an object's parametric eigenspace. Such learning techniques are fundamental to the advancement of visual perception. We developed efficient object recognition and pose estimation algorithms that are based on the parametric eigenspace representation. The learning and recognition algorithms were tested on objects with complex shape and reflectance properties. A statistical analysis of the errors in recognition and pose estimation demonstrates the proposed approach to be very robust to factors, such as, image noise and quantization. We believe that the results presented in this paper are applicable to a variety of vision problems. This is the topic of our current investigation.

Acknowledgements

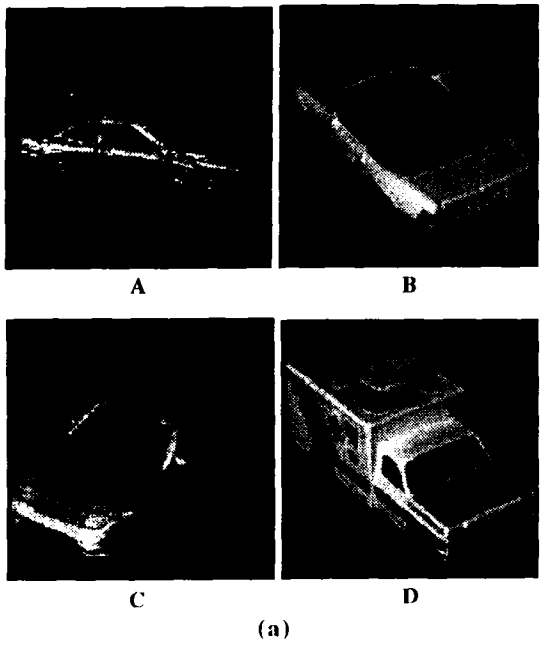
The authors would like to thank Daphna Weinshall for several useful comments on the paper.

References

[1] R. T. Chin and C. R. Dyer, "Model-Based Recognition in Robot Vision," *ACM Computing Surveys*, Vol. 18, No. 1, pp. March 1986.

[2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, London, 1990.

[3] H. Murase, F. Kimura, M. Yoshimura, and Y. Miyake, "An Improvement of the Auto-Correlation Matrix in Pattern Matching Method and Its Application to Hand-



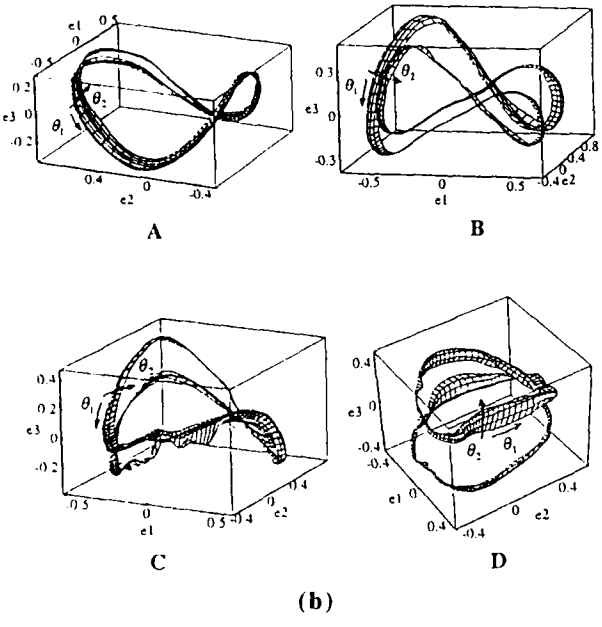
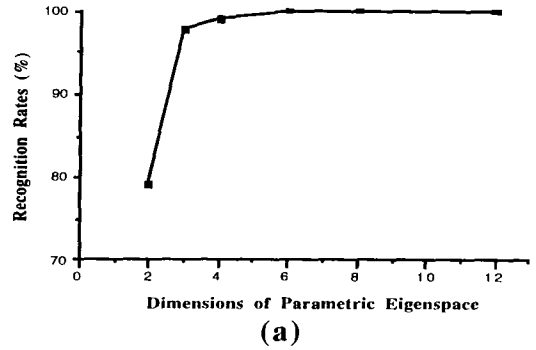


Figure 2: (a)The four objects used in the experiments. (b) The parametric hypersurfaces in object eigenspace computed for the four objects shown in (a). For display, only the three most important dimensions of each eigenspace are shown. The hypersurfaces are reduced to surfaces in three-dimensional space.



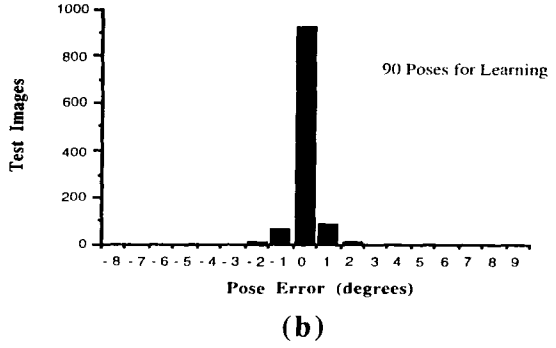


Figure 3: (a) Recognition rate plotted as a function of the number of universal eigenspace dimensions used to represent the parametric hypersurfaces. (b) Histogram of the error (in degrees) in computed object pose for the case where 90 poses are used in the learning stage. The average of the absolute error in pose for the complete set of 1080 test images is 0.5 degrees.

printed 'HIRAGANA'," *Trans. IECE*, Vol. J64-D, No. 3, 1981.

[4] H. Murase and M. Lindenbaum, "Spatial Temporal Adaptive Method for Partial Eigenstructure Decompisition of Large Images," *NTT Technical Report No. 6527*, March 1992.

[5] E. Oja, *Subspace methods of Pattern Recognition*, Research Studies Press, Hertfordshire, 1983.

[6] T. Poggio and F. Girosi, "Networks for Approximation and Learning," *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1481-1497, September 1990.

[7] W. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1988.

[8] L. Sirovich and M. Kirby, "Low dimensional procedure for the characterization of human faces," *Journal of Optical Society of America*, Vol. 4, No. 3, pp. 519-524, 1987.

[9] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.