The 9th Scandinavian Conference
on Image Analysis
Uppsala, Sweden, June 1995

Image Spotting of 3D Objects using Parametric Eigenspace Representation

Hiroshi Murase
NTT Basic Research Labs
3-1, Morinosato Wakamiya, Atsugi
Kanagawa, 243-01, JAPAN
murase@siva.ntt.jp

Shree K. Nayar
Columbia University
New York, N.Y. 10027
USA
nayar@cs.columbia.edu

Abstract

This paper proposes a novel method to detect three-dimensional objects in arbitrary poses and sizes from a complex image and simultaneously measure their poses and sizes. We refer to this process as image spotting. In the learning stage, for a sample object to be learned, a set of images is obtained by varying pose and size. This large image set is compactly represented by a manifold in compressed subspace spanned by eigenvectors of the image set. This representation is called the parametric eigenspace representation. In the image spotting stage, a partial region in an input image is projected to the eigenspace, and the location of the projection relative to the manifold determines whether this region belongs to the object, and what its pose is in the scene. This process is sequentially applied to the entire image at different resolutions. Experimental results show that this method accurately detect the target objects.

1. Introduction

Image spotting of three-dimensional (3D) objects has wide applications such as visual search of a target in security systems or target detection in recognition systems. There are two approaches used for image spotting. One uses local features such as edges or corners and matches them with 3D models [1-4]. This method might handle 3D rotation and scaling of objects, however, extraction of geometric features from noisy natural scenes is not easy. The other approach uses template matching such as image correlation (matched filtering) or image subtraction. This approach is insensitive to noise and small distortions. Our approach is based on this approach.

Template matching is a fundamental task in image processing. Even if we limit the discussion to searching problems, many vision algorithms using template matching have been proposed. For example, feature detection using template matching in pyramids [5,6] or using matched-filter[7] were proposed, however, these methods are developed for

* This research was conducted at Information Science Research Laboratory of NTT Basic Research Laboratories.

two-dimensional template matching, so they can not directly deal with 3D objects in a 3D scene. A 3D object has many appearances depending on the pose and distance between the camera and the object. If we store all variations of the object appearance and sequentially match them with the whole subpart of the input image using conventional template matching, a vast amount of memory size and computation time is required. Our method is related to this exhaustive template matching, however, we use new compact representation that makes the computation of image correlation quick and efficient. This representation is called parametric eigenspace. This approach makes it possible to detect a 3D object in an arbitrary pose and position in the scene.

The idea of parametric eigenspace was first applied for isolated object recognition[14,15] and tracking[16]. We extend this idea to image spotting, which solves the complex situation that the object has a complicated background. This representation uses two main ideas: KL (Karhunen-Loeve) expansion and manifold representation. The KL expansion is a well-known technique to approximate images in the low dimensional subspace spanned by eigenvectors of the image set. This technique is based on principal component analysis[10,11], and it has been applied to pattern recognition problems such as character recognition[12] and human face recognition[8,9]. We call this subspace the eigenspace. Calculation in this eigenspace reduces computation time. Secondly, appearance manifold conveniently represents continuous appearance-change due to the change of parameters such as object pose or object size. The combination of the above two ideas yields a new continuous and compact representation of 3D objects. We used this representation for partial image matching, and hierarchical matching at image resolutions to detect target objects.

2. Learning object models

The appearance of an object depends on its shape, its reflectance properties, its pose, its distance form the camera, and the illumination condition. The first two parameters are intrinsic properties of the object that do not vary. The correlation method is relatively robust to illumination variations when a brightness normalization process is used. On the other hand, object pose and camera distance can vary substantially from one scene to the next. Here, we represent an object using the parametric eigenspace representation that is parameterized by its pose and its distance from the camera.

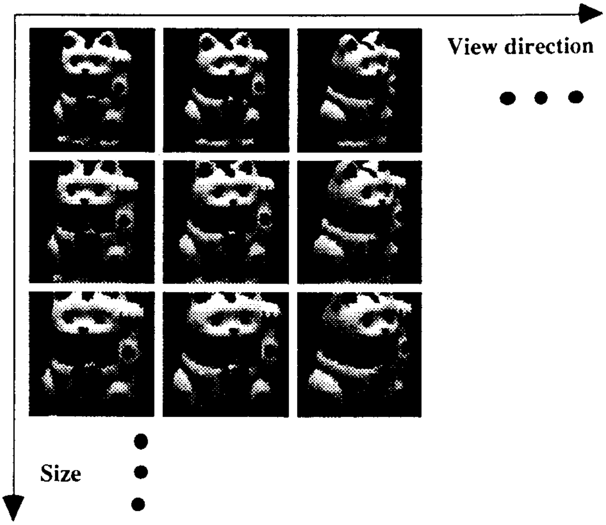
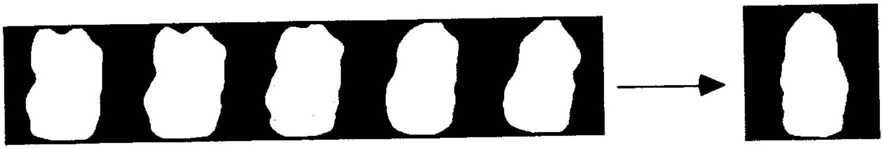


Fig.1. A learning image set.

2.1 Search window

First, for a given object sample to be learned, we collect a set of images by varying pose using a computer controlled turntable. Then we



(a) Object region of the learning images

(b) Search window

Fig.2. A search window.

segment the object region from each image and normalize its size to some fixed rectangle. Next, we generate several sizes of images (i.e., scale factor 1, 1.1, 1.2, .., α , where α =1.5) for each pose. These images are used for object learning. We refer to this image set as the learning image set (Fig. 1). Using all generated images, we design the search window. The window is the AND area of the object region of all images in the learning image set. Fig. 2 shows an example of the search window made using the learning image set. If the AND area becomes too small due to the shape of the object, the pose angle range is divided to several parts and the procedure is applied separately. This search window is introduced to eliminate background region and extract only parts of the object region in the leaning stage. In the image spotting stage, this search window is used to scan the whole input image.

2.2 Eigenspace

Each learning image is first masked by the search window, then represented by the N dimensional vector $\hat{\mathbf{x}}_{r,s}$ $(r = 1,\cdots,R,\ s = 1,\cdots,S)$, where the element of the vector is a pixel value of the image inside the window, N is a number of the pixels, r is a pose parameter, and s is a size parameter. Here, R and S are the total number of discrete poses and sizes, respectively. We normalize the brightness to be unaffected by variations in intensity of illumination or the aperture of the imaging system. This can be achieved by normalizing each image, such that, the total energy constrained in the image is unity. This brightness normalization transforms each measured image $\hat{\mathbf{x}}_{r,s}$ to a normalized image $\mathbf{x}_{r,s}$ where

$$\mathbf{x}_{r,s} = \frac{\hat{\mathbf{x}}_{r,s}}{\left\|\hat{\mathbf{x}}_{r,s}\right\|}.$$

The covariance matrix of this normalized image vector set is

$$Q = \frac{1}{RS}\sum_{s=1}^{S}\sum_{r=1}^{R}(\mathbf{x}_{r,s} - \mathbf{c})(\mathbf{x}_{r,s} - \mathbf{c})^{\mathrm{T}}.$$

Here $\mathbf{c}$ is the average of all images in the learning set determined as

$$\mathbf{c} = \frac{1}{RS}\sum_{s=1}^{S}\sum_{r=1}^{R}\mathbf{x}_{r,s}.$$

The eigenvectors $\mathbf{e}_i$ (i=1,..k) and the corresponding eigenvalues λ_i of Q can be determined by solving the well-known eigenvalue decomposition problem:

$$\lambda_i\mathbf{e}_i = Q\mathbf{e}_i.$$

Although all N eigenvectors of the planning image set are needed to represent images exactly, only a small number ($k << N$) of eigenvectors are generally sufficient for capturing the primary appearance characteristics of objects. The k-dimensional eigenspace spanned by the eigenvectors :

$$\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k\}$$
$$(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k)$$

is an optimal subspace to approximate the original leaning image set in the sense of l^2 norm. Computing the eigenvectors of a large matrix such as Q can prove computationally very intensive. Efficient algorithms for this are summarized in [11,13]. Fig. 3 shows eigenvectors for the object shown in figure 1.

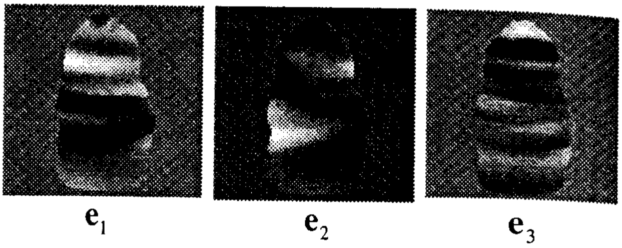


Fig.3. Eigenvectors for a learning image set shown in figure 1.

2.3 Correlation and distance in eigenspace

In this section, we discuss the relation between image correlation and distance in eigenspace. Consider two images $\mathbf{x}_m$ and $\mathbf{x}_n$ that belong to the image set used to compute an eigenspace. Let the points $\mathbf{g}_m$ and $\mathbf{g}_n$ be the projections of two images in eigenspace. Each image can be expressed in terms of its projection as:

$$\mathbf{x}_m = \sum_{i=1}^{N} g_{mi}\mathbf{e}_i + \mathbf{c},$$

where $\mathbf{c}$ is once again the average of the entire image set. Note that our eigenspaces are composed of only k eigenvectors. Hence, $\mathbf{x}_m$ can be approximated by the first k terms in the above summation:

$$\mathbf{x}_m \approx \sum_{i=1}^{k} g_{mi}\mathbf{e}_i + \mathbf{c}.$$

As the result of the brightness normalization described in section 3.2, $\mathbf{x}_m$ and $\mathbf{x}_n$ are unit vectors. The SSD (sum-of-squared-difference) measure between the two images is related to correlation as:

$$\begin{aligned} \| \mathbf{x}_m - \mathbf{x}_n \|^2 &= (\mathbf{x}_m - \mathbf{x}_n)^{\mathrm{T}}(\mathbf{x}_m - \mathbf{x}_n) \\ &= 2 - 2\mathbf{x}_m{}^{\mathrm{T}}\mathbf{x}_n, \end{aligned}$$

where $\mathbf{x}_m{}^{\mathrm{T}}\mathbf{x}_n$ is the correlation between the images. Alternatively, the SSD can be expressed in terms of the coordinates $\mathbf{g}_m$ and $\mathbf{g}_n$ in eigenspace:

$$\begin{aligned} \| \mathbf{x}_m - \mathbf{x}_n \|^2 &\approx \| \sum_{i=1}^{k} g_{mi}\mathbf{e}_i - \sum_{i=1}^{k} g_{ni}\mathbf{e}_i \|^2 \\ &= \| \mathbf{g}_m - \mathbf{g}_n \|^2. \end{aligned}$$

So we have:

$$\| \mathbf{g}_m - \mathbf{g}_n \|^2 \approx 2 - 2\mathbf{x}_m{}^{\mathrm{T}}\mathbf{x}_n.$$

This relation implies that the square of the Euclidean distance between the point $\mathbf{g}_m$ and $\mathbf{g}_n$ is an approximation of the SSD between the images $\mathbf{x}_m$ and $\mathbf{x}_n$. In other words, the closer the projections are in eigenspace, the more highly correlated are the images. We use this property of eigenspace to calculate image correlation efficiently.

2.4 Parametric manifold

The next step is to construct the parametric manifold for the object in eigenspace. Each image $\mathbf{x}_{r,s}$ in the object image set is projected to the eigenspace by finding the dot product of the result with each of the eigenvectors of the eigenspace. The result is a point $\mathbf{g}_{r,s}$ in the eigenspace:

$$\mathbf{g}_{r,s} = \left[\mathbf{e}_1 \cdots \mathbf{e}_k\right]^T \mathbf{x}_{r,s}.$$

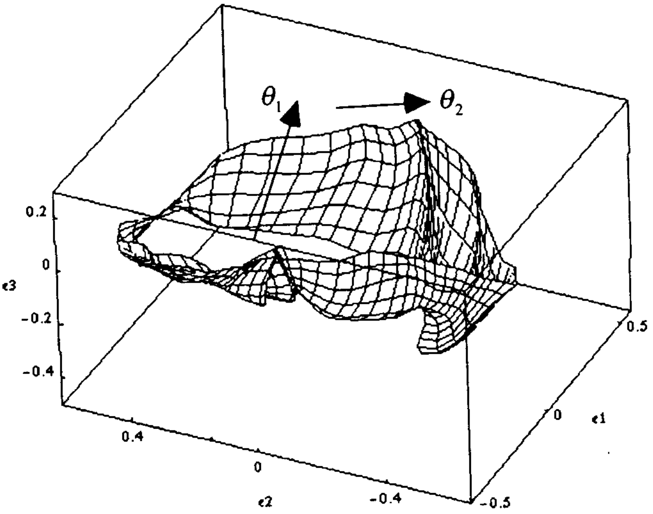


Fig.4. A parametric eigenspace representation for the object shown in figure 2.

Once again the subscript r represents the rotation parameter and s is the size parameter. By projecting all the learning samples in this way, we obtain a set of discrete points in 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 manifold that represents all possible poses and a limited range of object size variation. We interpolate the discrete points to obtain this manifold. In our implementation, we have used a standard cubic spline interpolation[17]. This interpolation makes it possible to represent appearance between sample images. The resulting manifold can be expressed as: $\mathbf{g}(\theta_1, \theta_2)$ where θ_1 and θ_2 are the continuous rotation and size parameters. The above manifold is a compact representation of the object's appearance. Fig.4 shows the parametric eigenspace representation of the object shown in Fig.1. The figure shows only three of the most significant dimensions of the eigenspace since it is difficult to display and visualize higher dimensional spaces. The object representation in this case is a surface since the object image set was obtained using two parameters. If we add more parameters such as rotations in other axes, this surface becomes high dimensional manifold.

3 Image spotting

3.1 image spotting using the parametric eigenspace

Consider an image of a scene that includes one or more of the objects that we have learned, on a complicated background. We assume that the objects are not occluded by other objects in the scene when viewed from the camera direction.

First, the search window is scanned on the whole input image area ($1 \leq x \leq X$; $1 \leq y \leq Y$) and a sequence of the subimages is made. Here, X and Y are sizes of the input image. The search window eliminates the background effect and extracts only subpart of the input images, namely, inside the object region. Each subimage is normalized respect to brightness as described in the previous section. The normalized subimage at position *(x,y)* is represented by vector $\mathbf{p}(x,y)$. Next, $\mathbf{p}(x,y)$ is projected into the eigenspace by

$$\mathbf{h}(x,y)=\left[\mathbf{e}_1\cdots\mathbf{e}_k\right]^{\mathrm{T}}\mathbf{p}(x,y).$$

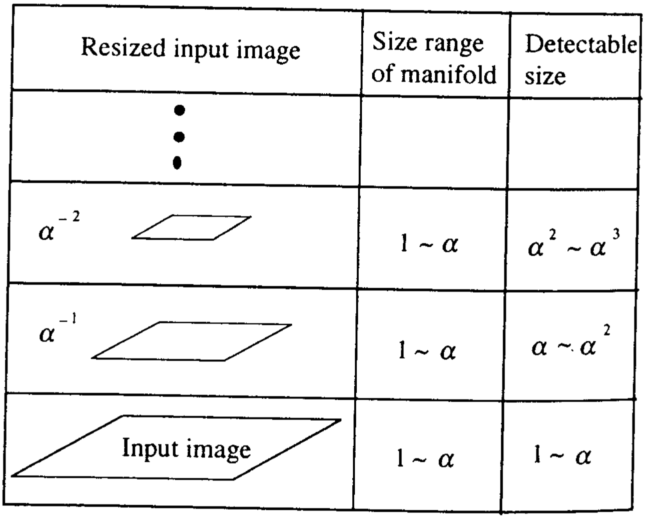
If this subimage belongs to the learned object, the projected point $\mathbf{h}(x,y)$ will be located on the manifold $\mathbf{g}(\theta_1,\theta_2)$. Next, we compute the distance between the projected point and the manifold, using:

$$d(x,y)=\min_{\theta_1,\theta_2}\left\|\mathbf{h}(x,y)-\mathbf{g}(\theta_1,\theta_2)\right\|.$$

If the distance *d(x,y)* is less than some pre-determined threshold value, the position (x,y) is a candidate for the object. After finding the candidate, the minimum peak of the distance around this position is searched, because the distance of the subimage at (x,y) is similar to that of the subimage around this position since these images are correlated to each other. Finally, we can conclude that the position that minimizes the distance is of the object. The pose and size parameters can be estimated by the parameters θ_1 and θ_2 that minimize the distance.

3.2 Hierarchical image spotting

We assume weak perspective image projection. This means the size of the object is a function of the distance between a camera and the object. As shown in the previous section, the parameter θ_2 in the manifold can deal with size variation of the object region. However, the dynamic range should be limited, because the effective window area, that is used for correlation, becomes small if we cover a large range of the size parameter using the parametric eigenspace representation. In our experiment, we set the dynamic range of the size parameter of the manifold to around 1.5 ($1 \le \theta_2 \le 1.5$). This range of the size variation is not enough for many applications. Here, to cover a wider range of size variation, we apply this process hierarchically. The input image is resized like 1, α^{-1}, α^{-2}, ... , and the same image spotting procedure is applied for each resized image. Here, α is set to the maximum value of the parameter θ_1 As the result, this method can cover size variation continuously. Fig. 5 shows the range to cover the size



Resized input image	Size range of manifold	Detectable size
⋮		
α^{-2}	$1 \sim \alpha$	$\alpha^{2} \sim \alpha^{3}$
α^{-1}	$1 \sim \alpha$	$\alpha \sim \alpha^{2}$
Input image	$1 \sim \alpha$	$1 \sim \alpha$

Fig.5. Hierarchical scaling of the input image for arbitrary size.

for each resized input image.

3.3 Computational cost

Here, we discuss the computational cost by estimating the number of operations such as multiplications for each calculation of distance d(x,y). Consider the dimension in the search window as N, and the number of the templates that corresponds to the number of the possible poses and sizes as M. And k is the dimension of the eigenspace. In general, N>M>>k. If we ignore the second order, the number of operations for each step for the conventional correlation technique is as follows: (1) N multiplications for normalization. (2) NM multiplications for the correlation. (3) M comparisons to find the maximum correlation. The total number of operations for each calculation of distance d(x,y) is N+NM+M. On the other hand, if we apply the parametric eigenspace method, the operation in each step is as follows. (1) N multiplications for normalization. (2) Nk multiplications for the projection. (3) Mk multiplications for distance calculation. (4) M comparisons to find the minimum distance. The total number of the operations is N+Nk+Mk +M, hence, the dominant factor is (N+M)k. Assume, N is 6,500 (the pixel number of the search window), and M is 3,600 (R=360, S=10), and k is 10. These numbers are picked from the example in our experiments (See section 5). The results show 111,100 operations for our method, though 23,410,100 operations for the exhaustive correlation method. At the same time the memory size for the templates can be reduced from NM to Nk words.

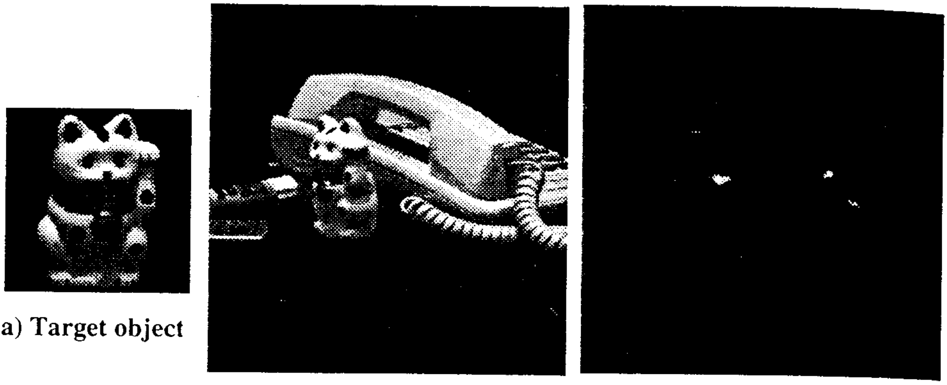
4. Experiments

We have conducted several experiments using complex objects to verify the effectiveness of the parametric eigenspace representation. This section summarizes some of our results.

For example, we have demonstrated three kinds of target objects, a toy cat, a juice can, and a human face. In the learning step, the object is placed on a motorized turntable and its pose is varied about a single axis, namely, the axis rotation of the turntable. Most objects have a finite number of stable configurations when placed on a planar surface. For such objects, a turntable is adequate as it can be used to vary pose for each of the object's stable configuration. For a human face, we used a rotating stool instead of a turntable. When learning, we used a black background to make it easy to segment the object region from the background. Images of the object are sensed using a 512x480 pixel CCD camera and are digitized as 8bits per pixel. We took 45 images of different poses for each object for learning. The object region is segmented by the simple thresholding technique and its size is normalized to 128x128 pixels. Then, we compute a search window and the parametric eigenspace representation for each object (see Fig. 2). In this example, the number of pixels of the window is 5400.

Next, we constructed a manifold in the eigenspace according to the procedure in section 3, and densely resampled the manifold by 360 poses for the rotation parameter and 10 steps for the size parameter. Totally we have 3,600 resampled points on the manifold, which are used for searching the manifold. In our experiments, calculating the distance from the manifold was achieved by finding the nearest neighbor distance from these resample points.

To test the algorithm, we used 20 images where the target object was placed on the complicated background. We applied the procedure written in section 4. Fig. 6 (c) shows the distance map for the image example shown in Fig. 6 (b) and target object shown in Fig.6 (a). Here, a white pixel means an area of small distance value, and the object is



(a) Target object

(b) Input image

(c) Distance from the manifold

Fig.6. An example of image spotting. (a) a target object,
(b) An input image, (c) Distance map from the manifold.

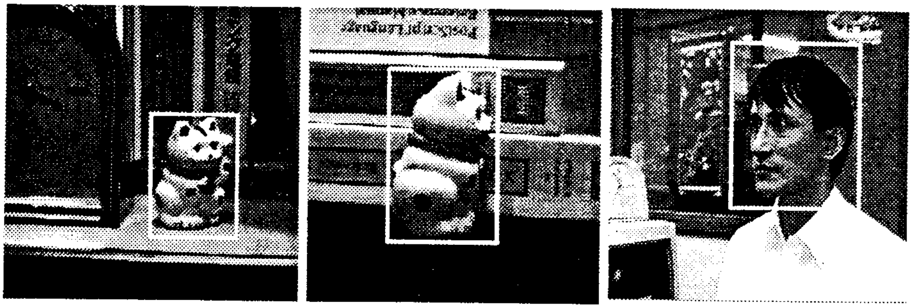


Fig.7. Results of image spotting.

possibly there. The computation time is 2 minutes using a SUN workstation SS10. Fig. 7 shows the results for several input images.

We evaluated the number of dimensions of the eigenspace by changing the number of dimensions. If the number of dimensions is too low, the representation is less accurately approximated, and many positions that do not belong to the object have a small distance. This causes errors for image spotting. If the number of dimensions is large enough, the correct position of the object is detected accurately. We tested this algorithm for 20 test images, and found a 10 dimensional eigenspace is enough for image spotting of these objects.

Image correlation methods are robust to noise. We tested for noisy images (i.e. SNR=20dB). Our method works well for these images. The method is also working well for small occlusion, because this method uses global features. Consequently, the parametric eigenspace method is robust for noise and small occlusion.

We can compute an object's pose at the same time of image spotting, by estimating the pose parameter that minimizes the distance from the manifold. It was shown that the accuracy of the pose estimation for an isolated object [14] is high using the parametric eigenspace method. In our case, however, the strong feature of the object boundary can not be used, because the boundary can be obtained after segmentation and pose estimation. We evaluate the pose estimation error for both cases for the same object: (1) Isolated objects and using image matching including object boundary, (2) not isolated object and using only partial matching of the object region. Fig. 8 shows a histogram of the error for both cases for 45 test images. The accuracy of the pose estimation without using boundary information is still high, though the accuracy is a little lower than that using boundary information.

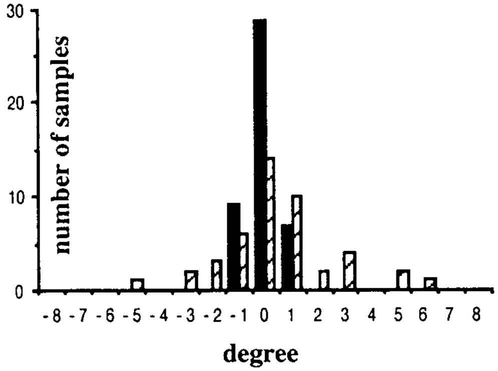


Fig. 8. Histogram for pose estimation error.

5. Conclusion

In this paper, we described an image spotting method for a three-dimensional object with a complicated background. This new image representation is called the parametric eigenspace method. The method detects an object in an arbitrary pose and size in a natural scene based on 2D image correlation, and simultaneously computes the pose and size of the object. There are a variety of the appearances for a 3D object depending on its pose and position. We represent them using a compact image representation based on two key ideas. One is the KL transform, which approximates the appearance of the object image set using a small number of eigenvectors to reduce the computation time and memory size. The other is a parametric representation, which represents the continuous change of appearance by varying pose and position by a manifold to compute object pose and position. Image correlation using this representation is hierarchically computed for different sizes of input images to cover a large dynamic size range of the object.

Experimental results show this method can accurately spot the target object. We have also shown this method reduces the computational cost compared with the exhaustive correlation method, and is also robust to noise in input images. Future research will concentrate on recognizing objects with large occlusion.

Acknowledgment We would like to thank Dr. K. Ishii and Dr. S. Naito for encouragement.

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] P. J. Besl and R. C. Jain, Three-dimensional object recognition, ACM Computing Surveys, Vol. 17, No. 1, pp. 75-145, 1985.

[3] T. Poggio and S. Edelman, A network that learns to recognize three-dimensional objects, Nature, Vol. 343, pp. 263-266, 1990.

[4] J. J. Weng, N. Ahuja, T. S. Huang, Learning recognition and segmentation of 3D objects from 2D images, IEEE ICCV, pp.121-128, 1993.

[5] S. L. Tanimoto, Template matching in pyramids, Computer Vision, Graphics and Image Processing, 16, pp. 356-369, 1981.

[6] A. Rosenfeld and G. J. Vanderbrug, Coarse-fine template matching, IEEE Transactions on System, Man, and Cybernetics, pp. 104-107, 2, 1977.

[7] Z. Q. Liu and T. M. Caelli, Multiobject pattern recognition and detection in noisy backgrounds using a hierarchical approach, Computer Vision, Graphics and Image Processing, 44, pp. 296-306, 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. Tùrk and A. P. Pentland, Face recognition using eigenfaces, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, June 1991.

[10] K. Fukunaga, Introduction to statistical pattern recognition, Academic Press, London, 1990.

[11] E. Oja, Subspace methods of pattern recognition, Research Studies Press, Hertfordshire, 1983.

[12] H. Murase, F. Kimura, M. Yoshimura, and Y. Miyake, An improvement of the auto-correlation matrix in pattern matching method and its application to handprinted `HIRAGANA', Trans. IECE, Vol. J64-D, No. 3, 1981.

[13] H. Murase and M. Lindenbaum, Spatial temporal adaptive method for partial eigenstructure decomposition of large images, IEEE Transactions on Image Processing, May, 1995.

[14] H. Murase and S. K. Nayar, Learning object models from appearance, AAAI-93, American Association for Artificial Intelligence, pp. 836-843, July, 1993.

[15] H. Murase and S. K. Nayar, Illumination planning for object recognition Using Parametric Eigenspace, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, No.12, pp.1219-1227, 1994.

[16] S.K.Nayar, H.Murase, and S.A. Nene, Learning, and Positioning, and Tracking Visual Appearance, IEEE International Conference on Robotics and Automation, May, 1994.

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