Illumination Planning for Object Recognition Using Parametric Eigenspaces

Hiroshi Murase and Shree K. Nayar

Abstract—**This correspondence presents a novel approach to the problem of illumination planning for robust object recognition in structured environments. Given a set of objects, the goal is to determine the illumination for which the objects are most distinguishable in appearance from each other. Correlation is used as a measure of similarity between objects. For each object, a large number of images is automatically obtained by varying pose and illumination direction. Images of all objects, together, constitute the planning image set. The planning set is compressed using the Karhunen–Loeve transform to obtain a low-dimensional subspace, called the eigenspace. For each illumination direction, objects are represented as parametrized manifolds in eigenspace. The minimum distance between the manifolds of two objects represents similarity between the objects in the correlation sense. The optimal source direction is therefore one that maximizes the shortest distance between object manifolds. Several experiments have been conducted using real objects. Results produced by the illumination planner have been used to enhance the performance of an object recognition system.**

Index Terms—**Illumination planning, object recognition, correlation, image compression, principal component analysis, appearance matching, parametric appearance representation, pose invariance**

I. Introduction

In structured environments, vision systems are used to perform a variety of tasks, such as inspecting manufactured parts, recognizing objects and sorting them, or aiding a robot in assembly operations. In each of these cases, the illumination of the environment can be selected to enhance the reliability and accuracy of the vision system. Currently, illumination parameters are selected by human operators using the trial and error approach. The resulting illumination is seldom one that maximizes the performance of the vision system.

Lately, automatic illumination planning has emerged as a topic of research interest. Most of this work focuses on determining light source positions that maximize the detectability of image features such as edges. Cowan and Bergman [2] used CAD (geometric) models of objects to compute source positions for which all brightness values in the image lie within the sensor's dynamic range. The positions, orientations, and reflectance parameters of the objects are assumed to be known. Using the same assumptions, Cowan and Nitzan [3] computed source positions that ensure that the brightness contrast at selected edges of objects exceeds a threshold value. Recently, Yi *et al.* [12], [13] used the Torrance and Sparrow reflectance model to obtain accurate predictions of the brightness of object points. Yi propagates errors due to noise in image brightness to obtain errors in the positions of line segments in the image. The planning problem is then to determine the source direction that maximizes the accuracy of edge positions.

Addressing a different problem, Sakane *et al.* [10] determine optimal source directions for a photometric stereo system. They use the accuracy of computed surface normals and the range of computable normals as criteria for selecting optimal source directions. Recently, Batchelor [1] proposed an expert system that uses the knowledge of illumination experts to suggest the best illumination for a given vision application. The illumination plan proposed by an expert is based on his or her experience and not on a careful theoretical analysis of the problem. Hence the suggested illumination is not guaranteed to be optimal.

Here, we present a new approach to illumination planning for object recognition in structured environments. Object *appearance* is used as the criterion for finding optimal illumination. Given a set of objects, the goal is to determine the source direction that makes the objects maximally different from each other in the *correlation* sense. The resulting source direction can then be used to optimize the performance of a correlation-based[1] recognition system. In contrast to previous work on illumination planning, we assume that the pose of each object is *unknown*. Our objective is to find a source direction that is optimal over all poses of the objects. Unlike previous approaches, the method does not rely on the availability of CAD models. No assumptions are made with respect to geometric or reflectance properties of the objects. Instead, a sample of each object is used during the planning stage.

For each object, a large set of brightness images is automatically obtained by varying its pose and the illumination direction. One way to accomplish image-based illumination planning is to compare all images (corresponding to different poses and illumination directions) of each object with images of other objects. Images can be compared by computing their correlation. The optimal illumination direction is one for which images of each object are *minimally* correlated with those of other objects. Given the large number of images we are dealing with, this approach is clearly impractical from a computational perspective. However, we accomplish essentially the same task but in a very efficient manner. Images in the planning set are correlated to a large degree since variations in object pose and illumination between consecutive images are small. We take advantage of this strong correlation and compress the planning image set using the Karhunen–Loeve transform [8], a technique widely used for image compression and pattern recognition. The result is a low-dimensional subspace, called the eigenspace.[2]

For each source direction, images (for different poses) of each object are mapped to the eigenspace to obtain a manifold. This manifold is parametrized by object pose. The shortest distance between the manifolds of two objects represents poses of the objects for which their appearances are most similar, i.e., correlation is maximum. Our goal therefore is to find the source direction for which the manifolds of all objects are most distant from each other. This is the optimal source direction for the entire set of objects; it represents the illumination that makes the objects appear most different from each other, irrespective of their poses. Though we formulate the planning problem as one of finding the optimal source direction, several other source characteristics—such as size, distance, and spectral distribution—can be incorporated into the planning process.

Once the optimal illumination has been planned, its validity must be verified. For this, we have used an object recognition system that

Manuscript received April 27, 1993; revised May 17, 1994. This research was conducted at the Center for Research in Intelligent Systems, Department of Computer Science, Columbia University, New York, NY. It was supported in part by the David and Lucile Packard Fellowship and in part by ARPA under Contract DACA 76-92-C-0007. Recommended for acceptance by Associate Editor Y. Shirai.

H. Murase is with the NTT Basic Research Labs, Atsugi-shi, Kanagawa, 243-01 Japan.

S. K. Nayar is with the Department of Computer Science, Columbia University, New York, NY 10027 USA.

IEEE Log Number 9406171.

[1] Correlation, or template matching, remains one of the most widely used recognition strategies in the industrial arena. Finding optimal illumination for this task is therefore a problem of significant practical relevance.

[2] This subspace has previously been used in vision to classify handwriting [4] and human faces [11], as well as for recognition and pose estimation of 3D objects [6].

Fig. 1. The effect of illumination direction on object appearance.

identifies 3D objects and computes their poses from brightness images [6]. Experiments using this system show that the planned illumination direction produces the highest recognition rate compared to any other source direction. We conclude with a discussion on the merits and limitations of the method.

II. Illumination Planning

The appearance of an object depends on its shape, its reflectance properties, its pose, and the illumination conditions. The first two factors are intrinsic properties of the object that do not vary. On the other hand, object pose and illumination can vary substantially from one scene to the next. In most machine vision applications, the pose of an object is not within the control of the vision system; objects show up in the scene with arbitrary poses. That leaves us with illumination. In structured environments, such as industrial assembly lines, illumination of the scene can be controlled to provide the "best" images of the objects of interest. Fig. 1 shows images of a manufactured part obtained using different illumination directions. These images illustrate that object appearance is very sensitive to the direction of illumination. Our objective is to determine the illumination direction that makes a set of such objects most distinguishable from each other, irrespective of their poses.

The objects need not be illuminated only by the source whose optimal direction we wish to determine. They may be illuminated, in addition, by other sources with unknown but constant intensities and directions. The planning system is unaffected by the existence of such ambient lighting. Although we have posed the planning problem as one of finding the optimal source direction, the approach can also be used to determine optimal source position. In fact, since the planning method uses 2D images and not 3D object models, other source characteristics (such as source size and color) as well as sensor characteristics (such as spectral response and optical settings) can be incorporated into the planning process. The only requirement is that these source and sensor characteristics be varied during the image acquisition stage of planning. The planning system described below is fully automated.

A. Planning Image Set

While constructing the planning image set we need to ensure that all object images are of the same size. Each digitized image is segmented into an object region and a background region. The background is assigned zero brightness value, and the object region is resampled such that the larger of its two dimensions fits the size we have selected for image representation. The result is an image that is normalized with respect to scale and thus invariant to the magnification of the imaging system. This image is written as a vector $\hat{\mathbf{x}}$ by reading pixel values by raster scan:

$$\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N]^T. \tag{1}$$

The above vector represents an unprocessed brightness image. Alternatively, processed images such as smoothed images, first derivatives, second derivatives, Laplacian, or the power spectrum of the image

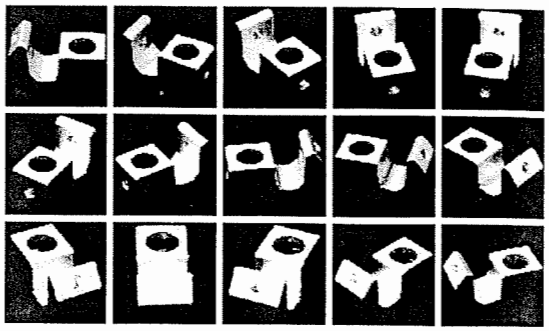


Fig. 2. Image set for the object shown in Fig. 1 obtained by varying pose, for a given illumination direction. Only 15 of the 45 images in the set are shown.

may be used in place of the above brightness image. The image type is selected based on its ability to capture distinct appearance characteristics of the objects of interest. Here, for the purpose of developing the illumination planning method we use raw brightness images, bearing in mind that the planning methodology, as well as the recognition technique described in Section III, are directly applicable to any other image type.

It is desirable that the illumination planning system be unaffected by variations in the intensity of illumination or the aperture of the imaging system. This can be achieved by normalizing each image 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:

$$\mathbf{x} = [x_1, x_2, \cdots, x_N]^T \tag{2}$$

where $\mathbf{x} = \hat{\mathbf{x}} / \| \hat{\mathbf{x}} \|$.

We denote each normalized image as $\mathbf{x}_{r,l}^{(p)}$, where r is the rotation or pose parameter, l represents the illumination direction, and p is the object number. The image set obtained by varying the pose of an object for a given illumination direction l can be written as

$$\mathbf{X}_l^{(p)} \triangleq \left\{ \mathbf{x}_{1,l}^{(p)}, \mathbf{x}_{2,l}^{(p)} \cdots, \mathbf{x}_{R,l}^{(p)} \right\} \tag{3}$$

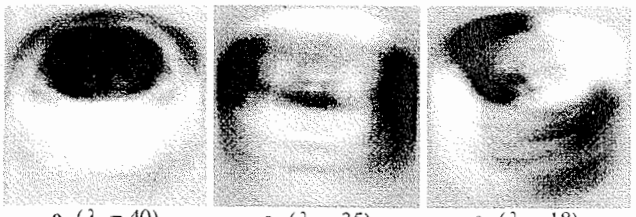
where R is the total number of discrete poses used for each object. Let P be the total number of objects and L be the total number of illumination directions. Then, the *planning image set* for the entire set of objects is

$$\{ \mathbf{X}_1^{(1)}, \cdots, \mathbf{X}_L^{(1)}, \mathbf{X}_1^{(2)}, \cdots, \mathbf{X}_L^{(2)}, \mathbf{X}_1^{(P)}, \cdots, \mathbf{X}_L^{(P)} \}. \tag{4}$$

In our experiments, we have used a motorized turntable to vary object pose. This gives us pose variations about a single axis. We have used several light sources positioned in a plane around the turntable. For each object illumination direction is automatically varied, and for each illumination direction a set of images is obtained by rotating the object. This setup was used to demonstrate the proposed planning method. However, the method itself is extensible to arbitrary object rotations and light source directions in three dimensions. Fig. 2 shows some of the images in the set obtained by varying the pose of the part shown in Fig. 1 for a given illumination direction.

B. Universal Eigenspace

Consecutive images in the planning image set are correlated to a degree since pose and illumination variations between these images are not large. Our objective is to take advantage of this correlation and compress the large planning set into low-dimensional representations of each object's appearance. A suitable compression technique is the Karhunen–Loeve method [8], where the eigenvectors of an image set are computed and used as orthogonal basis functions for representing



$\mathbf{e}_1$ ($\lambda_1 = 40$) $\mathbf{e}_2$ ($\lambda_1 = 35$) $\mathbf{e}_3$ ($\lambda_1 = 18$)

Fig. 3. Eigenvectors corresponding to the three largest eigenvalues, computed for the image set shown in Fig. 2.

individual images. Although, in general, all eigenvectors of an image set are required for perfect reconstruction of any particular image, only a few are sufficient for illumination planning. These eigenvectors constitute what we refer to as the *universal eigenspace*.

To compute the universal eigenspace, we first subtract the average c of all images in the planning set from each image in the set. 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. A new image set is obtained by subtracting the average c from each image in the planning set:

$$\mathbf{Y} \triangleq \left\{ \mathbf{x}_{1,1}{}^{(1)} - \mathbf{c},\ \mathbf{x}_{2,1}{}^{(1)} - \mathbf{c},\ \cdots,\ \mathbf{x}_{R,L}{}^{(P)} - \mathbf{c} \right\}. \quad (5)$$

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

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

$\mathbf{Q}$ 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 computed by solving the following well-known eigenvector decomposition problem:

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

All N eigenvectors of $\mathbf{Q}$ constitute a complete eigenspace. Any two images from the planning image set, when projected to this space, give two discrete points. We will show later that the distance between these points is a measure of correlation between the two images. Although all N eigenvectors of the planning image set are needed to represent images exactly, only a small number ($k \ll N$) of eigenvectors are generally sufficient for capturing the primary appearance characteristics of objects. These k eigenvectors correspond to the largest k eigenvalues of $\mathbf{Q}$ and constitute the universal eigenspace.

Efficient algorithms for computing eigenvectors of very large matrices (such as $\mathbf{Q}$) are described in [8] and summarized in [6]. In our experiments, we have used the algorithm presented in [5]. The number k of eigenvectors to be computed is selected empirically based on our experimental results on planning as well as recognition. For the objects we have used in our experiments, universal eigenspaces with less than 10 dimensions ($k < 10$) are found to be adequate. As an example, Fig. 3 shows three eigenvectors (shown as images) computed for the image set shown in Fig. 2. They are shown in descending order of their eigenvalue magnitudes.

C. Parametric Eigenspace Representation

Our objective is to get a measure of how well the set of objects can be discriminated under illumination from each of the source directions. The image set $\mathbf{X}_l{}^{(p)}$ includes images of the object p, obtained for different poses r, while it is illuminated by the source l. Each image $\mathbf{x}_{r,l}{}^{(p)}$ in $\mathbf{X}_l{}^{(p)}$ is projected to the universal eigenspace.

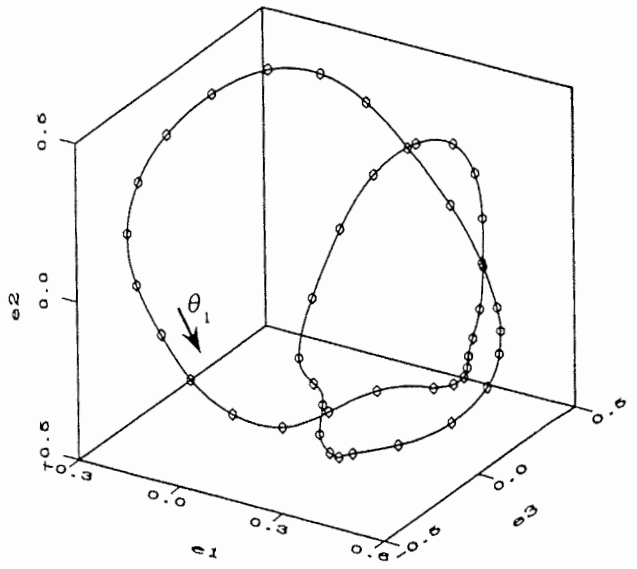


Fig. 4. Curve in universal eigenspace obtained by projecting the object image set shown in Fig. 2.

This is done by subtracting the average image c from $\mathbf{x}_{r,l}{}^{(p)}$, then finding the dot product of the result with each of the k eigenvectors, or dimensions, of the universal eigenspace. The result is a single point in eigenspace:

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

By projecting all the planning samples in $\mathbf{X}_l{}^{(p)}$, we get a set of discrete points in universal eigenspace. Pose variation between any two consecutive images in $\mathbf{X}_l{}^{(p)}$ is small. As a result, consecutive images are strongly correlated and their projections in eigenspace are close to one another.[3] The reasons for this are given in the next section. The discrete points obtained by projecting all samples in $\mathbf{X}_l{}^{(p)}$ can be assumed to lie on a manifold that represents all possible poses of the object for the illumination direction l. The discrete points are interpolated to obtain the manifold $\mathbf{g}_l{}^{(p)}(\theta_1, \theta_2, \theta_3)$, where θ_1, θ_2, and θ_3 are the three continuous rotation parameters needed to describe pose in three-dimensional space. The above manifold is referred to as the *parametric eigenspace representation*; it is a compact representation of the appearance of object p when illuminated by source l. In our experiments, we rotate the object about a single axis. This variation in pose is sufficient for objects that have a finite number of stable configurations when placed on a planar surface. Thus, the manifold is reduced to a curve with a single parameter, θ_1:

$$\mathbf{g}_l{}^{(p)}(\theta_1). \quad (9)$$

Fig. 4 shows the parametrized eigenspace representation of the object shown in Fig. 1. The eigenvectors (dimensions) of this eigenspace were computed using a planning set that includes two object image sets. 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 points shown on the curve correspond to projections of the images shown in Fig. 2. The continuous curve passing through the discrete points is parametrized by rotation θ_1 and is obtained by cubic spline interpolation [9].

[3] This assumption holds well except when the object is either highly specular or has high-frequency texture. In such cases, an incremental pose variation can cause dramatic changes in image brightness.

D. Distance in Eigenspace and Template Matching

Before proceeding to determine the optimal source direction, it is worth describing some relevant properties of the eigenspace representation. 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 the two images in the eigenspace. It is well known in pattern recognition theory [8] that each of the images can be expressed in terms of its projection as

$$\mathbf{x}_m = \sum_{i=1}^{N} g_{m_i} \mathbf{e}_i + \mathbf{c} \tag{10}$$

where $\mathbf{c}$ is once again the average of the entire image set. The above expression states that the image $\mathbf{x}_m$ can be exactly represented as a weighted sum of all N eigenvectors of the image set. The individual weights g_{m_i} are the coordinates of the projection of the image in eigenspace. Note that our eigenspace is composed of only k eigenvectors. Since these eigenvectors correspond to the largest eigenvalues, they represent the most prominent principal components of the image set. Hence, $\mathbf{x}_m$ can be approximated by the first k terms in the above summation:

$$\mathbf{x}_m \approx \sum_{i=1}^{k} g_{m_i} \mathbf{e}_i + \mathbf{c}. \tag{11}$$

As a result of the brightness normalization described in Section II–A, $\mathbf{x}_m$ and $\mathbf{x}_n$ are unit vectors. The similarity between the two images can be determined by finding the *sum of squared difference* (SSD) between brightness values in the images. This measure is extensively used for template matching in a variety of industrial applications. We known that the SSD measure can be related to correlation as

$$\begin{aligned} \| \mathbf{x}_m - \mathbf{x}_n \|^2 &= (\mathbf{x}_m - \mathbf{x}_n)^{\mathrm{T}} (\mathbf{x}_m - \mathbf{x}_n) \\ &= 2 - 2\,\mathbf{x}_m{}^{\mathrm{T}} \mathbf{x}_n \end{aligned} \tag{12}$$

where, $\mathbf{x}_m{}^{\mathrm{T}} \mathbf{x}_n$ is the correlation between the images. Maximizing correlation, therefore, corresponds to minimizing the SSD and thus maximizing similarity between the images. Alternatively, the SSD can be expressed in terms of the coordinates $\mathbf{g}_m$ and $\mathbf{g}_n$ in eigenspace using (11):

$$\| \mathbf{x}_m - \mathbf{x}_n \|^2 \approx \| \sum_{i=1}^{k} g_{m_i} \mathbf{e}_i - \sum_{i=1}^{k} g_{n_i} \mathbf{e}_i \|^2 . \tag{13}$$

Note that the eigenvectors are orthogonal; $\mathbf{e}_i{}^{\mathrm{T}} \mathbf{e}_j = 1$ when $i = j$, and 0 otherwise. Using this, the right-hand side of the above expression can be simplified to get

$$\| \mathbf{x}_m - \mathbf{x}_n \|^2 \approx \| \mathbf{g}_m - \mathbf{g}_n \|^2 . \tag{14}$$

The above relation implies that the square of the Euclidean distance between the points $\mathbf{g}_m$ 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 the eigenspace to determine optimal illumination in the next section.

E. Optimal Illumination Direction

Consider two objects, say p and q, from the set used to compute the universal eigenspace. For each light source direction l, we compute parametric curves for the two objects:

$$\mathbf{g}_l{}^{(p)}(\theta_1{}^{(p)}) \quad \text{and} \quad \mathbf{g}_l{}^{(q)}(\theta_1{}^{(q)}). \tag{15}$$

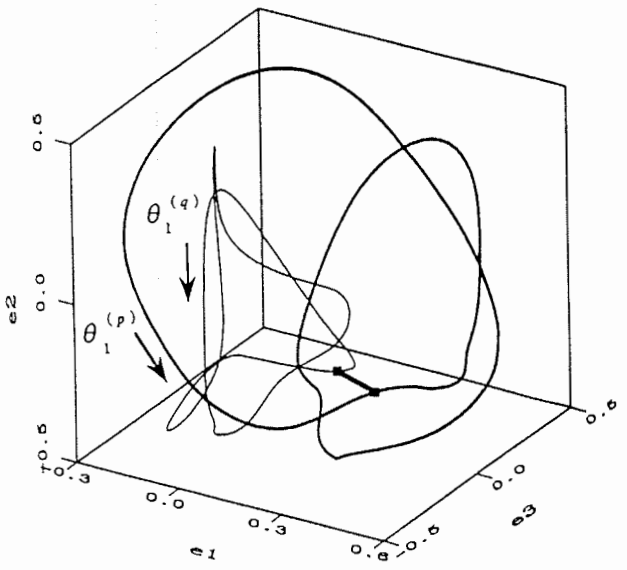


Fig. 5. Parametric eigenspace curves of two different objects obtained for a given illumination direction. The shortest distance (line segment) between the two curves represents the worst-case poses for which the objects appear most similar in the correlation sense.

Here, the parameters $\theta_1{}^{(p)}$ and $\theta_1{}^{(q)}$ represent rotations of p and q, respectively. The shortest Euclidean distance between the two curves is computed as

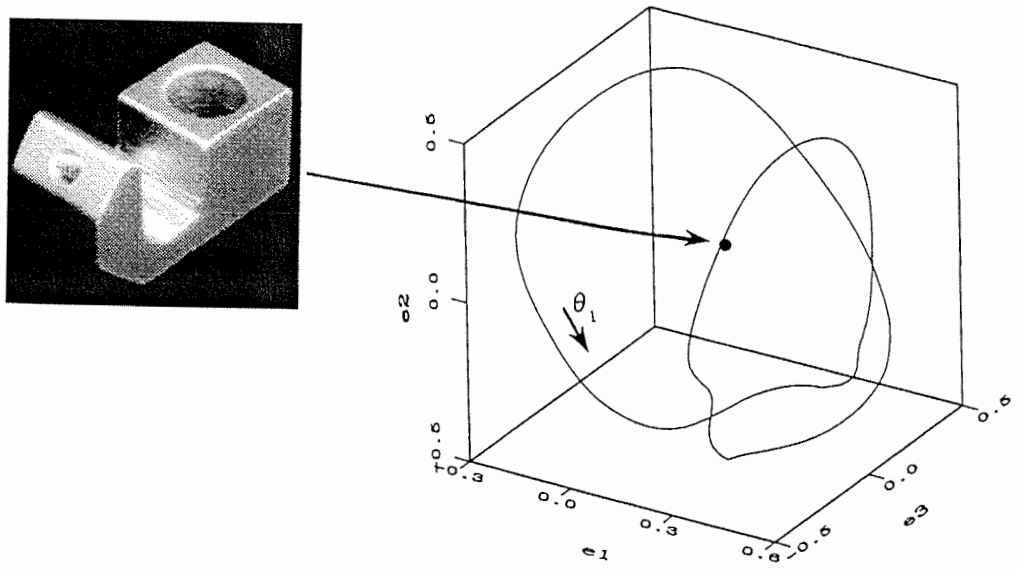
$$d_l{}^{(p,q)} = \min_{\theta_1{}^{(p)}, \theta_1{}^{(q)}} \| \mathbf{g}_l{}^{(p)}(\theta_1{}^{(p)}) - \mathbf{g}_l{}^{(q)}(\theta_1{}^{(q)}) \| . \tag{16}$$

The $\theta_1{}^{(p)}$ and $\theta_1{}^{(q)}$ values that produce the minimum distance $d_l{}^{(p,q)}$ correspond to poses of the two objects for which the objects appear most similar (in correlation) when illuminated by source l. The illumination planning problem is formulated as follows: Find the source direction $\tilde{l}$ that maximizes the minimum distance $d_l{}^{(p,q)}$ between the object curves. This *max-min* strategy yields the safest illumination direction for the worst-case poses that make the two objects appear most similar.

The above example includes only two objects. The *max-min* strategy is easily extended to a set of P objects. For a given illumination direction l, we now have P curves in universal eigenspace. The minimum distance $d_l{}^{(p,q)}$ is computed for all pairs of objects, resulting in P^2 minimum distances. The minimum of all these distances, say d_l, represents the worst case for the entire object set. The source direction $\tilde{l}$ that maximizes d_l is then the *optimal source direction* for the object set. Fig. 5 shows eigenspace curves of two objects used in the experiments, for a particular illumination direction. The solid-line segment illustrates the shortest distance between the two curves. If in a particular application the poses of the objects are fixed, the eigenspace representation of each object, for a given illumination, is reduced from a curve to a point. In that case, the optimal source direction maximizes the minimum distance between points in eigenspace that represent different objects.

III. Correlation-Based Object Recognition

In this section we describe an object recognition system that is based on the parametric eigenspace representation. We first presented this system in [6], where it was demonstrated as an effective approach for recognizing a variety of complex 3D objects. In the experimental section, it is used to evaluate the performance of the illumination planning method described above.



(a) (b)

Fig. 6. (a) An input image. (b) The input image is mapped to a point in universal eigenspace. The location of the point determines the object and its pose in the scene.

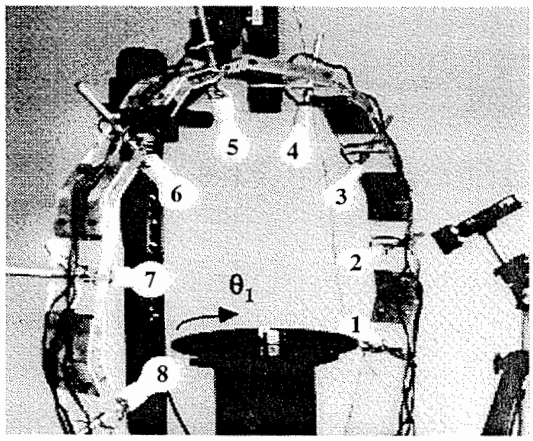
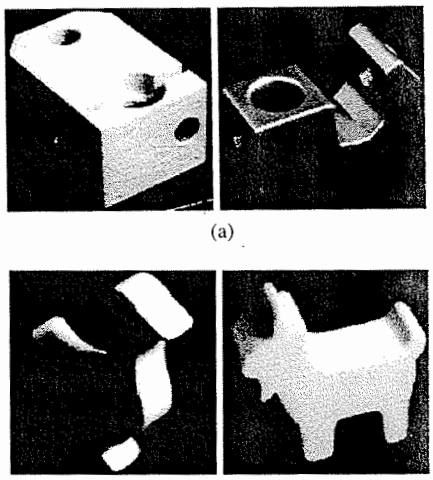


Fig. 7. Device used to obtain planning image sets. Each object is placed on the motorized turntable and illuminated from different directions.



(a) (b)

Fig. 8. Two pairs (**A** and **B**) of objects used to test the illumination planning method.

Consider an image of a scene that includes one or more of the objects we have used to compute the universal eigenspace. We assume that the objects are not occluded by other objects in the scene when viewed from the sensor direction, and that image regions corresponding to objects have been segmented away from the scene image. In a variety of industrial applications, objects appear in the field on the vision system without occluding each other. For segmentation, one of several existing algorithms may be applied. In some applications a constant brightness background can be used, in which case a simple thresholding operation yields object regions. If the objects are in motion, spatio-temporal image filters (see [6] for examples) may be applied to recover the moving (object) regions.

The first step is to normalize the segmented image regions with respect to scale and brightness, as described in Section II–A. This normalization ensures that image regions have the same size and brightness range as the eigenvectors that constitute the universal eigenspace. The normalization also renders the recognition system invariant to imaging optics (magnification[4] and aperture) and the intensity of illumination. An image region, after normalization, is referred to as the *input image* $\mathbf{y}$.

For recognition, the average $\mathbf{c}$ of the entire image set, used to compute the universal eigenspace, is subtracted from the input image $\mathbf{y}$. The resulting image is projected to the universal eigenspace to obtain a point $\mathbf{z}$.

The recognition problem is to find the object p whose eigenspace representation (manifold in general and curve in our case) the point $\mathbf{z}$ lies on. Here, the source direction l is known *a priori*, and so are the object curves in eigenspace for the direction l. Due to factors such as image noise, aberrations in the imaging system, and digitization effects, $\mathbf{z}$ may not lie exactly on an object curve. Therefore, we find the object p that gives the minimum distance ${h_1}^{(p)}$ between its curve

[4]The image projection model is assumed to be weak-perspective; orthographic projection followed by scaling.

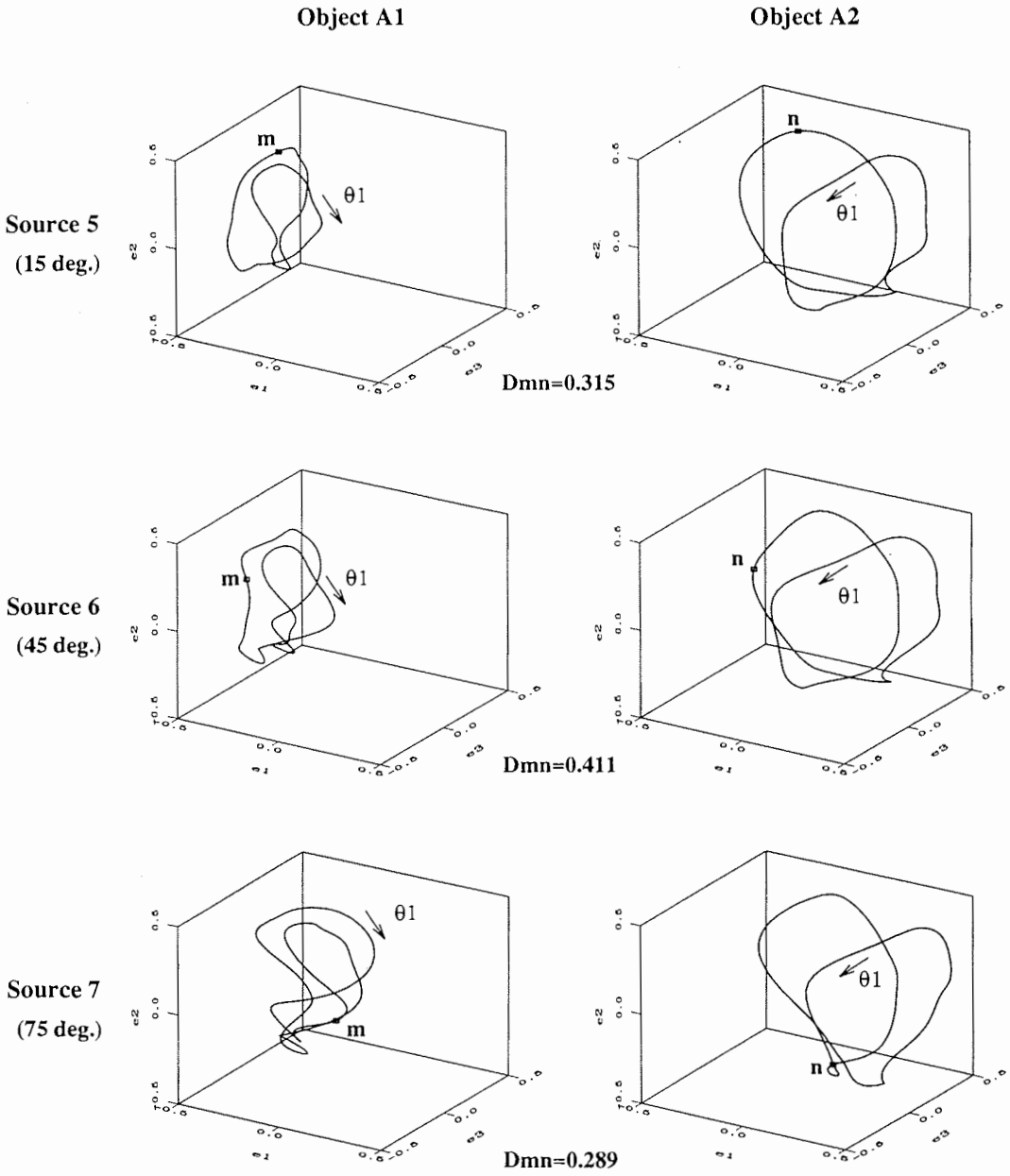


Fig. 9. Curves in universal eigenspace obtained for object pair A for three different sources. For display, only the three most important dimensions of universal eigenspace are shown. Source 6 (45°) produces the maximum distance between the closest points on the parametric curves of the two objects.

$\mathbf{g}_l^{(p)}(\theta_1)$ and the input $\mathbf{z}$:

$$h_1^{(p)} = \min_{\theta_1} \| \mathbf{z} - \mathbf{g}^{(p)}(\theta_1) \| . \quad (17)$$

If $h_1^{(p)}$ is less than a small threshold value, we conclude that the input image is of object p. The value of θ_1 that corresponds to $h_1^{(p)}$ represents the pose of the object in the scene. Fig. 6(a) shows an input image of the object whose parametric curve is shown in Fig. 4. This input image is not one of the images in the planning image set used to compute the universal eigenspace. In Fig. 6(b), the input image is mapped to eigenspace and is seen to lie close to the parametric curve of the object.

IV. Experiments

First, the setup used to acquire the planning image set is described. The universal eigenspace is computed from the planning set. The eigenspace representations of different objects are used to determine the optimal source direction. Finally, the recognition system presented in the previous section is used to verify the optimal illumination.

A. Automatic Acquisition of Image Sets

If an object's geometry and reflectance are known *a priori*, its images under different poses and illumination conditions can be synthesized using image rendering techniques such as radiosity or ray tracing. Here, we have not assumed that object models are available. Therefore, we need a mechanism that automatically varies object pose and illumination and generates image sets. Fig. 7 shows the setup we have developed for illumination planning. 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 via software and can be varied with an accuracy

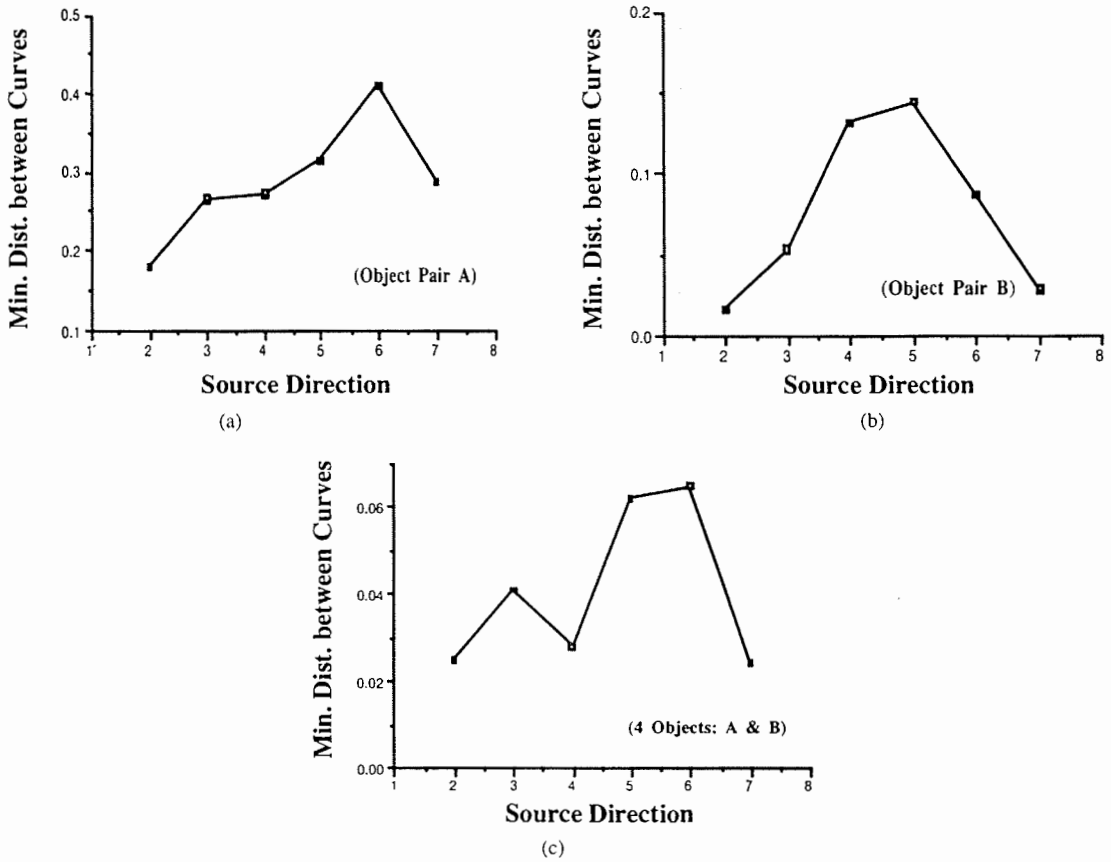


Fig. 10. The minimum distance between parametric curves plotted as a function of source number for (a) object pair **A**, (b) object pair **B**, and (c) the set including pairs **A** and **B**.

of 0.1°. 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.

The objects are illuminated by the ambient lighting conditions of the environment that do not vary during the acquisition of image sets. This ambient illumination is of relatively low intensity. In addition, eight incandescent light bulbs (100 W each) are used to illuminate the objects from different directions. Of these, only six sources were used since sources 1 and 8 generate strongly self-shadowed images of the objects. The light bulbs are uniformly distributed in a plane around the turntable, and the angle between adjacent light bulbs is 30°. These light sources are activated through software. The planning problem is to find the optimal light source among the six. Images are sensed using a 512 × 480 pixel CCD camera, and are digitized using an Analogics frame-grabber board.

B. Planning Results

Fig. 8 shows the two pairs of objects, namely, **A** and **B**, used in the experiments. For each of the six light sources (2–7), each object was placed on the turntable and images obtained for 45 different poses (8° increments of the turntable). For each of the object pairs, therefore, a planning set with 720 images was obtained. Each image is automatically segmented and normalized in scale and brightness as described in Section II–A. Each normalized image is 128 × 128 pixels in size.

The 45 pose images of each object, taken for each of the light sources, are projected to universal eigenspace to get a set of discrete points. These points are interpolated using a standard cubic spline interpolation algorithm [9] to obtain a parametric curve. The parameter of the curve is object pose (θ_1). Fig 9 shows the curves of the two objects in pair **A**, for three different light sources. Since it is difficult to visualize the curves of two objects in the same space, they are displayed separately in Fig. 9. The measure $D_{m,n}$ is the same as $d_l^{(p,q)}$ in expression (16). It is computed for each source and is the minimum distance between the curves of the two objects in eight-dimensional eigenspace. It therefore represents the worst-case poses of the objects for each source direction. Note that these poses need not be among the ones present in the planning image set. Since the curves are obtained by interpolation, the worst-case poses may lie in between the discrete poses used for planning. Fig. 10(a) and (b) shows $D_{m,n}$ plotted as a function of source number for object pairs **A** and **B**, respectively. We see that source 6 (at 45° with respect to the vertical axis) is found to be optimal for object pair **A**, while source 5 (at 15°) is optimal for object pair **B**. Fig. 10(c) shows results for an object set that includes all four of the above objects (pairs **A** and **B**). Here, all 720 images of the four objects were used to compute the universal eigenspace, and for each illumination direction four appearance curves were computed. As seen from Fig. 10(c), source 6 is optimal in this case.

The optimal source direction determined by the illumination planner is meaningful only if it can be used to accomplish a vision task. We have used the correlation-based recognition system presented in Section III to verify the above results. For each light source, we obtained 45 test images of each object, to use as inputs to the recognition system. All of these test images are different from ones used for illumination planning; they correspond to object poses that lie in between the poses used for planning. Each test image is first

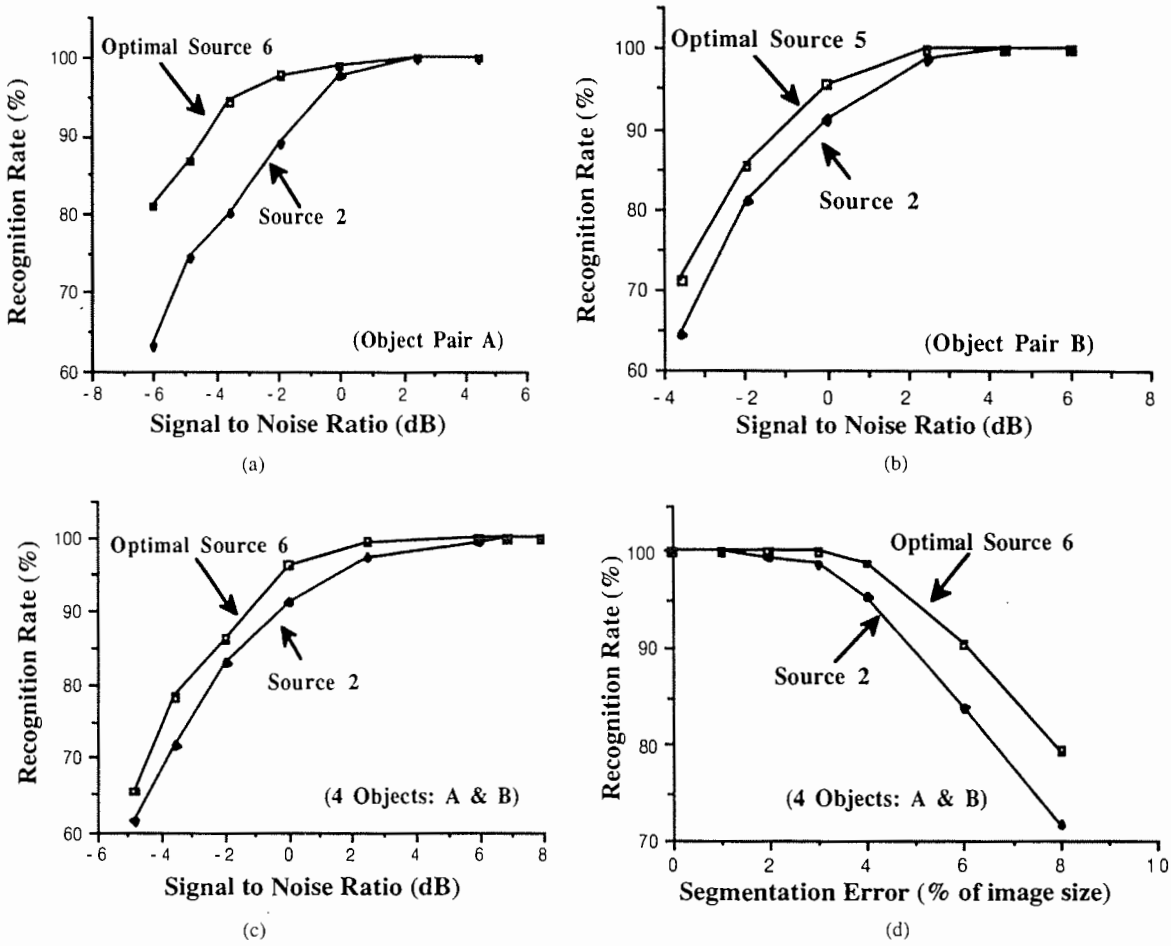


Fig. 11. Recognition rates for optimal and suboptimal source directions for (a) object pair **A**, (b) object pair **B**, and (c) the set including pairs **A** and **B**. For each object a total of 45 test images were used. As the noise level in the input images is increased, the recognition rates decrease but the optimal source always produces the highest rate. (d) Sensitivity of recognition rate to segmentation errors.

normalized in scale and brightness and then projected to universal eigenspace. The object in the image is identified by finding the curve that is closest to the input point in universal eigenspace.

We define *recognition rate* as the percentage of test images for which the object in the image is correctly recognized *and* the computed pose is within 6°[5] of the actual pose. Fig. 11(a) compares recognition rates computed using the optimal source 6 and the suboptimal source 2, for object pair **A**. To test the sensitivity of the optimal source, we added white noise to the test images. The noise level is given in decibels of signal to noise ratio, i.e., 10 $log_{10}(S/N)$; a noise level of -10 dB corresponds to noise that is 10 times the signal. Note that the noise levels added to the test images are substantial. As noise increases, the recognition rates naturally decay but the optimal source 6 consistently produces higher recognition rates than source 2 (used as an example nonoptimal source). Fig. 11(b) shows similar results for object pair **B**; optimal source 5 suggested by the planner produces consistently higher recognition rates than source 2 as image noise is increased. Fig. 11(c) shows the validity of optimal source 6 for the set including all 4 objects. These results demonstrate the robustness of the source selected by the illumination planning method to image noise.

In Fig. 11(d), the effects of segmentation error on the planning result are explored. Again, the object region is first segmented from each of the 720 test images and scale normalized to fit a 128 × 128 pixel image, as described in Section II–A. Then segmentation errors are introduced in each normalized image by shifting the object region in a randomly selected direction ($+x$, $-x$, $+y$, or $-y$) by some percentage of the image dimension (128 pixels). The resulting image emulates one with segmentation error. In Fig. 11(d), the segmentation error, or percentage shift, is plotted along the horizontal axis. As this error increases recognition rate deceases. However, the optimal source is seen to always produce higher recognition performance than the suboptimal one.

V. Discussion

We have presented a method for determining illumination parameters that make a set of objects maximally different from each other in the correlation sense. The proposed approach was shown to be effective in improving the performance of a correlation-based recognition system. Such recognition systems are widely used in industry for object identification and classification.

The illumination planning method uses samples of the objects of interest and does not require that geometry or reflectance of the objects be known. An object could have complex geometric features or varying reflectance properties, or produce specular reflections, or even interreflections. Since illumination planning is based on object appearance, none of the above effects need be analyzed in isolation.

[5] This pose tolerance was selected arbitrarily. It is used to ensure that the optimal source yields the highest accuracy not only in object identification but also in pose estimation.

The parametric eigenspace representation enables us to determine an illumination that is optimal when the poses of the objects are unknown. The planning process is further simplified if the poses of the objects are in fact fixed. In this case, the eigenspace representation of each object is a point and only the distances between points need be computed to determine the optimal illumination.

In the experiments, a single parameter was used to represent object pose (rotation) for a given stable configuration. For certain applications, three degree of freedom (DOF) may be needed to describe object pose. In such cases, for any given illumination, object appearance is represented in eigenspace as a 3-DOF manifold. This, of course, involves the acquisition of a larger number of object images for each illumination. Further, illumination itself can be described using additional parameters, including source size, distance, color, and the number of sources. In [7], optimization of illumination color is described and demonstrated by experiments.

The planning approach can also be used to simultaneously optimize multiple parameters. The only requirement is that these parameters be varied during the acquisition of the planning image set. Clearly, for multiple parameters, acquiring image sets, computing parametric eigenspaces, and determining the optimal parameter values can be very time consuming. Therefore, our planning method may prove impractical when more than three illumination parameters must be jointly optimized. A lesser number of parameters, however, can be easily accommodated since illumination planning is typically done off-line and only once. As a result, it is generally not subject to severe time constraints.

References

[1] B. G. Batchelor, "A Prolog lighting advisor," *SPIE Proc. Intelligent Robots and Computer Vision VIII: Systems and Applications,* B. G. Batchelor, Ed., vol. 1193, 1993.

[2] C. Cowan and A. Bergman, "Determining the camera and light source location for a visual task," in *Proc. IEEE Int. Conf. Robotics and Automation,* pp. 509–514, 1989.

[3] C. Cowan, and D. Nitzan, "Automatic placement of vision sensors" in *Proc. 18th Annual NSF Conf. Design and Manufacturing Systems,* Georgia Tech University, Atlanta, GA, Jan., 1992.

[4] 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.

[5] H. Murase and M. Lindenbaum, "Spatial temporal adaptive method for partial eigenstructure decomposition of large images," NTT Tech. Rep. No. 6527, Mar. 1992. Also in *IEEE Trans. Image Processing*, in press.

[6] H. Murase and S. K. Nayar, "Visual learning and recognition of 3D objects from appearance," *Int. J. Comput. Vision*, vol. 14, 1995.

[7] ——, "Illumination planning for object recognition in structured environments," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 31–38, 1994.

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

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

[10] S. Sakane, M. Sato, and M. Kakikura, "Automatic planning of light source placement for an active photometric stereo system," in *Proc. IEEE Int. Workshop on Intelligent Robots and Systems,* pp. 559–566, 1990.

[11] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 586–591, June 1991.

[12] S. Yi, R. M. Haralick, and L. G. Shapiro, "Optimal sensor and light source positioning for machine vision," in *Proc. Int. Conf. Pattern Recognition,* 1990.

[13] S. Yi, "Illumination control expert for machine vision: A goal driven approach," Ph.D. dissertation, Dept. of Computer Science and Engineering, Univ. of Washington, Seattle, Mar. 1990.

A Fast Statistical Mixture Algorithm for On-Line Handwriting Recognition

Eveline J. Bellegarda, Jerome R. Bellegarda, David Nahamoo, and Krishna S. Nathan

***Abstract*— The automatic recognition of on-line handwriting is considered from an information theoretic viewpoint. Emphasis is placed on the recognition of unconstrained handwriting, a general combination of cursively written word fragments and discretely written characters. Existing recognition algorithms, such as elastic matching, are severely challenged by the variability inherent to unconstrained handwriting. This motivates the development of a probabilistic framework suitable to the derivation of a fast statistical mixture algorithm. This algorithm exhibits about the same degree of complexity as elastic matching, while being more flexible and potentially more robust. The approach relies on a novel front-end processor that, unlike conventional character or stroke-based processing, articulates around a small elementary unit of handwriting called a frame. The algorithm is based on 1) producing feature vectors representing each frame in one (or several) feature spaces, 2) Gaussian K-means clustering in these spaces, and 3) mixture modeling taking into account the contributions of all relevant clusters in each space. The approach is illustrated on a simple task involving a 81-character alphabet. Both writer-dependent and writer-independent recognition results are found to be competitive with their elastic matching counterparts.**

***Index Terms*— On-line handwriting recognition, statistical modeling, frame-based processing, mixture output distributions.**

I. Introduction

As they allow natural interactions between man and machine without the use of a keyboard, speech and handwriting are increasingly gaining acceptance as input to computer devices. Although speech is better suited for bulk data entry, handwriting is an attractive input medium in noisy acoustic environments, when a quiet operation is desired, or where privacy is a primary concern. Examples of applications where handwriting is the preferred input channel include form filling, personal note taking, document editing, record keeping, and meeting support. To address such applications, computer interfaces must include the capability of automatically recognizing natural handwriting.

Pen-based user interfaces most often capture handwriting through a special pen operating on an electronic tablet. The handwriting data recorded on the tablet is sent to a recognizer that outputs a string of typeset characters. Such operation requires the on-line recognition of the handwritten data. It is therefore to be distinguished from optical character recognition (OCR), that is always performed off-line. In OCR the basic input unit is a bitmap, whereas in on-line handwriting it is the trace the pen leaves on the tablet [1]. As a result, OCR devices operate on a static image of the handwriting, while on-line devices

Manuscript received August 19, 1993; revised December 15, 1993. Recommended for acceptance by Associate Editor R. Kasturi.

The authors are with IBM Research, T. J. Watson Research Center, Yorktown Heights, NY 10598 USA.

IEEE Log Number 9405790.