View-Based Recognition Using an Eigenspace Approximation to the Hausdorff Measure

Daniel P. Huttenlocher, Ryan H. Lilien, Clark F. Olson
Department of Computer Science
Cornell University
Ithaca, NY 14853
{dph,lilien,clarko}@cs.cornell.edu

Abstract

View-based recognition methods, such as those using eigenspace techniques, have been successful for a number of recognition tasks. Currently, however, such approaches are relatively limited in their ability to recognize objects which are partly hidden from view or occur against cluttered backgrounds. In order to address these limitations, we have developed a new view matching technique based on an eigenspace approximation to the generalized Hausdorff measure. This method achieves the compact storage and fast indexing that are the main advantages of previous eigenspace view matching techniques, while also being tolerant of partial occlusion and background clutter.

Our approach is based on comparing features extracted from views, such as intensity edges, rather than directly comparing the views themselves. The underlying comparison measure that we use is the Hausdorff fraction, as opposed to the sum of squared differences (SSD) which is employed by most eigenspace matching techniques. The Hausdorff fraction is quite insensitive to small variations in feature location as well as to the presence of clutter or partial occlusion. In this paper we define an eigenspace approximation to the Hausdorff fraction and present some simple recognition experiments which contrast our approach with prior work on eigenspace image matching. We also show how to efficiently incorporate our technique into an image search engine, enabling instances from a set of model views to be identified at any location (translation) in a larger image.

1 Introduction

Appearance-based approaches to object recognition have been successful in a number of tasks (e.g. [7, 13, 11, 10]). The central idea underlying such methods is to represent objects as collections of views, and to use an efficient encoding scheme for storing and retrieving the views. The most common encoding scheme is based on representing each view using a relatively low-dimensional space which captures important characteristics of the entire set of views. This low-dimensional space is generally formed using an eigen-decomposition (or principal components analysis) to define a subspace which provides a reasonable approximation to the set of stored views. Each stored model view is represented in terms of its projection into this subspace, which is quite compact in comparison with the number of pixels in each view. An unknown object is recognized by projecting its image into the subspace and then finding the closest model view(s) in the subspace using some similarity measure.

Subspace methods are attractive when there is a relatively large database of model views, because the set of model views can be represented using a small number of coefficients each, rather than the thousands of pixels in each image. This both saves storage and speeds the process of finding the closest matching images in the database. Moreover, when the subspace is relatively low-dimensional (e.g. $25 - 30$ dimensions) there are methods for finding approximate closest matches in time logarithmic in the number of items in the database (e.g. [1]). Subspace methods can also be viewed as a form of generalization or learning. To the extent that a subspace captures the important characteristics of a given set of images while omitting the unimportant characteristics, it can be insensitive to unimportant variations in the images.

The most effective applications of subspace methods have been limited to tasks where the objects that are to be recognized appear fully visible (i.e. not partially occluded), are against a uniform background and where the images are nearly correctly registered with each other in advance. For example, a particularly successful application is the recognition of faces from mugshots, where the head is generally about the same size and location in the image, and the background is a fixed color (e.g., [11]). The main reason for these limitations is that when extraneous information from the background of an unknown image is projected into the subspace, it tends to cause incorrect recognition results. This is analogous to the problem that occurs with template matching techniques, using measures such as the sum of squared differences (SSD) or correlation, where background pixels included in a matching window can significantly alter the correlation value and cause incorrect matches. One standard way of addressing this problem in template matching is to use sub-regions of the views, such that the regions do not contain any background. A similar approach has also been taken in eigenspace matching [10, 8]. One drawback, however, is that sub-regions are generally less distinctive and thus can lead to more possible matches being found. This issue of distinctiveness has been addressed in [9] where they use a selection procedure for image regions based on a minimum description

length principle.

In this paper we describe a subspace recognition method that handles clutter and partial occlusion by using a robust image comparison measure, rather than by using sub-regions of views. Our method is based on using an eigen-decomposition to approximate the computation of the generalized Hausdorff measure [5]. The Hausdorff measure has been used to determine the degree of resemblance between binary images (bitmaps). It has been effective in template matching recognition methods, even in the presence of significant background information in the match window [4, 3]. Much of the power of Hausdorff-based measures comes from the fact that they are robust to outlying data points [5]. The major contribution of the method that we report in this paper is that it combines the robustness of Hausdorff-based measures for identifying partially occluded objects in clutter, with the speed of subspace methods for recognizing sets of object models. The recent work of [2] has also developed a robust image comparison measure using eigenspaces, however the computation is considerably more expensive than our method (and thus far their technique has been applied to object tracking as opposed to view based recognition).

We present some simple experiments demonstrating that our method performs well when the background is unknown or when the object to be recognized is partially occluded, including in cases where prior eigenspace methods based on the SSD break down. In addition, our method can be extended to handle the image search problem, where the locations of objects to be recognized in an image are not known. We show how to incorporate the Hausdorff eigenspace method into an image search engine that identifies the locations (translations) in an image where any of the stored model views yield a good match. Experiments indicate that searching images using this approach can reliably rule out the vast majority of image locations for all of the models in the view set, without losing the correct match. Moreover, these experiments show that the Hausdorff eigenspace techniques provide considerble speed up over previous image search methods based on the generalized Hausdorff measure [5], when the task involves a set of more than about 200 stored model views.

In the following section we discuss subspace matching methods in more detail, focusing on the use of previous subspace techniques that are based on the SSD. We then describe the generalized Hausdorff measure and how it can be approximated using subspace methods. We also consider the error or uncertainty that is introduced by the projection of an image into an eigenspace. This error analysis is applicable to any use of subspace techniques. Following the error analysis, we present an empirical investigation of the accuracy of the approximation and contrast the Hausdorff eigenspace matching approach with an SSD-based approach. Finally, we consider how to incorporate the Hausdorff eigenspace approach into an image search engine such as that in [12], in order to search an image for instances of any of the objects in a set of views. We constrast the efficiency of this approach with previous Hausdorff-based image search techniques.

2 Subspace Representations and Approximating the SSD

In this section we review the use of eigenspace methods for grey-level matching (e.g., [10, 11]). Let I denote a two-dimensional image with N pixels, and let x be its representation as a (column) vector in scan line order. Given a set of training or model images, I_m, $1 \leq m \leq M$, define the matrix $X = [x_1 - c, \ldots, x_M - c]$, where x_m denotes the vector representation of I_m, and c is the average of the x_m's. The average image is subtracted from each x_m so that the predominant eigenvectors of XX^T will capture the maximal variation of the original set of images. In many applications of subspace methods, the x_m's are normalized in some fashion prior to forming X, such as making $\|x_m\| = 1$, to prevent the overall brightness of the image from affecting the results.

The eigenvectors of XX^T are an orthogonal basis in terms of which the x_m's can be rewritten (and other, unknown, images as well). Let λ_i, $1 \leq i \leq N$, denote the ordered (from largest to smallest) eigenvalues of XX^T and let e_i denote each corresponding eigenvector. Define E to be the matrix $[e_1, \ldots, e_N]$. Then $g_m = E^T(x_m - c)$ is the rewriting of $x_m - c$ in terms of the orthogonal basis defined by the eigenvectors of XX^T. The original x_m is then just the weighted sum of the eigenvectors

$$x_m = \sum_{i=1}^{N} g_{m_i} e_i + c$$

where g_{m_i} is the ith term of g_m.

It is straightforward to show that $\|x_m - x_n\|^2 = \|g_m - g_n\|^2$ (cf. [10]), because distances are preserved under an orthonormal change of basis. That is, the sum of squared differences (SSD) of two images can be computed using the distance between the eigenspace representations of the two images.

The central idea underlying the use of subspace methods is to approximate x_m using just those eigenvectors corresponding to the few largest eigenvalues, rather than using all N eigenvectors. This low-dimensional representation is intended to capture the important characteristics of the set of training images. Let $f_m = (g_{m_1}, \ldots, g_{m_k}, 0, \ldots, 0)$ and $r_m = (0, \ldots, 0, g_{m_{k+1}}, \ldots, g_{m_N})$, so that $g_m = f_m + r_m$. That is, f_m is the vector of coefficients corresponding to the first k terms in the sum, and r_m is the vector of remaining coefficients, where $k << N$. Then x_m can be approximately reconstructed using just the first k coefficients:

$$x_m \approx \hat{x}_m = \sum_{i=1}^{k} f_{m_i} e_i + c \ .$$

In some applications, the few largest eigenvectors are also not used in constructing the approximation $\hat{x}$, because they capture properties that are common to the entire set of images.

The SSD, $\|x_m - x_n\|^2$, is then simply approximated as $\|f_m - f_n\|^2$. As this representation uses only the k eigenvectors with largest eigenvalues, it is not necessary to compute all N eigenvalues and eigenvectors of XX^T (which would be quite impractical as N is usually many thousands). One approach, when the number of model views is smaller than the number of pixels in each view, is to compute the eigenvectors of X^TX instead (this is done in [10]).

The search for a model, x_m, that is most similar to some unknown x_n, can, in theory, be performed in $O(\log M)$ time. In practice this is efficient currently for up to about a 25-dimensional space (i.e., up to about $k = 25$ eigenvectors) using the approximation method of Arya et al. [1]. In addition, the points in the subspace can be viewed as samples of some underlying manifold representing all possible views of a given object or set of objects. In [10] this manifold is approximated and used in computing the distance between an unknown image and the set of model views.

3 Approximating Binary Correlation and the Hausdorff Fraction

In this section we describe a subspace method for approximating the generalized Hausdorff measure. The Hausdorff measure is defined for sets of points, and thus we are now restricting the discussion to binary images which represent sets of feature points on a grid (i.e., a binary image that is 1 for points that are in the set and 0 otherwise). First we review the generalized Hausdorff measure, and then consider a subspace approximation.

Given two point sets $\mathcal{P}$ and $\mathcal{Q}$, with m and n points respectively, and a fraction, $0 \leq f \leq 1$, the *generalized Hausdorff measure* is defined in [5, 12] as

$$h_f(\mathcal{P}, \mathcal{Q}) = \underset{p \in \mathcal{P}}{f^{\text{th}}} \min_{q \in \mathcal{Q}} \|p - q\|, \tag{1}$$

where $f^{\text{th}}_{p \in \mathcal{P}} g(p)$ denotes the f-th quantile value of $g(p)$ over the set $\mathcal{P}$. For example, the 1-th quantile value is the maximum (the largest element), and the $\frac{1}{2}$-th quantile value is the median. Equation (1) generalizes the classical Hausdorff distance, which *maximizes* over $p \in \mathcal{P}$. In other words, the generalized measure uses an arbitrary percentile (rank) of distances rather than the maximal distance as used in the classical measure.

The generalized Hausdorff measure is asymmetric (as is the classical Hausdorff distance). Given a fraction, f, and two point sets, $\mathcal{P}$ and $\mathcal{Q}$, $h_f(\mathcal{P}, \mathcal{Q})$ and $h_f(\mathcal{Q}, \mathcal{P})$ can attain very different values. For example, there may be points of $\mathcal{P}$ that are not near any points of $\mathcal{Q}$, or vice versa. We can also use a bidirectional form of this measure, $h_{fg}(\mathcal{P}, \mathcal{Q}) = \max(h_f(\mathcal{P}, \mathcal{Q}), h_g(\mathcal{Q}, \mathcal{P}))$. The bidirectional measure is not robust to large amounts of image clutter, but it is useful in uncluttered images and for verification of hypotheses.

The generalized Hausdorff measure has been used for a number of matching and recognition problems. In particular, there are two complementary ways in which the measure has been used:

1. Specify a fixed fraction, f, and then determine the distance, $d = h_f(\mathcal{P}, \mathcal{Q})$. In other words, find the smallest distance, d, such that $k = \lceil fm \rceil$ of the points of $\mathcal{P}$ are within d of points of $\mathcal{Q}$. We call this the *fractional Hausdorff distance*, because it is analogous to the Hausdorff distance, but considering only a fixed fraction of the points rather than all the points. Intuitively, the fractional Hausdorff distance measures how well the best subset of size $k = \lceil fm \rceil$ of $\mathcal{P}$ matches $\mathcal{Q}$. It is one way of determining how well two sets match, with smaller distances being better matches.

2. Specify a fixed distance, d, and then determine the resulting fraction of points that are within that distance. In other words, find the largest f such that $h_f(\mathcal{P}, \mathcal{Q}) \leq d$. Intuitively, this measures what portion of $\mathcal{P}$ is near $\mathcal{Q}$, for some fixed neighborhood size, d. We call this the *Hausdorff fraction*, because it measures the fraction of points within some given distance. It is a second way of determining how well two sets match, with larger fractions being better matches.

In this paper we use the second of these measures, the Hausdorff fraction. This fraction specifies for a given distance d the fraction of points in one set that are within distance d of points in the other set. For digital images, the points of the two sets $\mathcal{P}$ and $\mathcal{Q}$ have integral coordinates. Thus we let P be a binary image denoting the set $\mathcal{P}$, with each 1 in the binary image P corresponding to a point in $\mathcal{P}$ (and zero otherwise). Likewise for Q and $\mathcal{Q}$). Let Q^d be the dilation of Q by a disk of radius d (i.e., each 1 in Q is replaced by a "disk" of 1's of radius d). The Hausdorff fraction, for distance d, is then

$$\Phi_d(P, Q) = \frac{\#(P \wedge Q^d)}{\#(P)} \tag{2}$$

where $\#(S)$ denotes the number of 1's in a binary image S, and $\wedge$ denotes the logical and (or the product) of two bitmaps. That is, we simply replace each point of Q with a disk (specifying all the points within distance d of that point). Then we compute the logical and of that dilated image with the other image. The result is all the points in P that are within distance d of points in Q. Note the asymmetry of the measure: one set is dilated and the other is not. Furthermore, note that when the dilation is zero the Hausdorff fraction is simply a normalized binary correlation. The eigenspace approximation to the Hausdorff fraction developed below is thus also an approximation to the binary correlation.

Given two binary images, I_m and I_n, we let x_m be the representation of I_m as a column vector and x'_n be the representation of I_n^d (throughout we use primes to

denote vectors corresponding to dilated images). The Hausdorff fraction $\Phi_d(I_m, I_n)$ can then be computed as

$$\Phi_d(I_m, I_n) = \frac{x_m^T x'_n}{\|x_m\|^2}$$

because x^m and x'_n are both binary vectors and thus their dot product is the number of ones in the logical and.

Given this formulation of the Hausdorff fraction Φ_d, we now look at how it can be approximated using a subspace approximation to the dot product. First we look at the relation between the dot product of two images and their representations in eigenspace, where, as above, g_m and g'_n are the rewriting of x_m and x'_n in a new coordinate system defined by the eigenvectors E of XX^T.

$$\begin{aligned} x_m^T x'_n &= (x_m - c + c)^T (x'_n - c + c) \\ &= (x_m - c)^T (x'_n - c) + (x_m - c)^T c + (x'_n - c)^T c + \|c\|^2 \\ &= g_m^T g'_n + x_m^T c + x'^T_n c - \|c\|^2 \end{aligned}$$

The last step follows from $g_m^T g'_n = (E^T(x_m - c))^T E^T (x'_n - c) = (x_m - c)^T E E^T (x'_n - c) = (x_m - c)^T (x'_n - c)$ (i.e., dot products are preserved under an orthogonal change of basis).

We wish to approximate g_m and g'_n using just the first k coefficients, which, as above, we denote by f_m and f'_n. Thus we note that $g_m^T g'_n = (f_m + r_m)^T (f'_n + r'_n) = f_m^T f'_n + r_m^T r'_n$, because all of the cross terms are zero.

Note that the reconstruction $\hat{x}_m$ using just the first k coefficients:

$$\hat{x}_m = \sum_{i=1}^{k} f_{m_i} e_i + c$$

is no longer in general a binary vector. However $x_m^T x'_n \approx \hat{x}_m^t \hat{x'}_n^T$, i.e., the dot product is still an approximation of the dot product of the complete binary vectors. The quality of this approximation depends on the magnitude of the residuals, r_m and r'_n.

3.1 Subspace Approximation of the Hausdorff Fraction

We now describe the steps for constructing the eigenspace given a set of *binary* model views, $x_1, \ldots, x_M$. First, form the matrix $X = [x_1 - c, \ldots, x_M - c]$, as above, where c is the centroid of the x_m's. Do not normalize the x_m's in any fashion. Compute and save the first k eigenvectors of XX^T (i.e., those corresponding to the k largest eigenvalues)[1]. For each of the x_m's, compute $f_m = (g_{m_1}, \ldots, g_{m_k})$, where $g_{m_i} = e_i^T (x_m - c)$. Then compute $x_m^T c$ and $\|x_m\|^2$. Save this vector and two scalars for each

[1]It is often much more efficient to compute the eigenvectors of $X^T X$, since it is usually much smaller. If e is an eigenvector of $X^T X$, then Xe is an eigenvector of XX^T and the ordering of the eigenvectors by eigenvalue is the same.

x_m. This, in addition to the k eigenvectors with the largest eigenvalues, is all of the information needed to match the set of models to each unknown image.

Once the above information has been computed and saved for each model image, an unknown image is processed by dilating it by d, forming the vector x'_n from this dilated image, and computing f'_n and $x'^T_n c$.

An explicit search of all of the models can be performed by computing the approximation to the Hausdorff fraction, Φ_d, for each x_m and the (dilated) unknown x'_n,

$$\hat{F}_m = \frac{f_m^T f'_n + x_m^T c + x'^T_n c - \|c\|^2}{\|x_m\|^2} \tag{3}$$

Note that each of the terms in this expression was computed and stored in forming the eigenspace or is computed once per unknown image, except for $f_m^T f'_n$. Thus the matching a given view in the eigenspace to an unknown image only requires a dot product of two k length vectors (just as in the traditional eigenspace matching techniques), plus a division and a few additions.

One issue with approximating the Hausdorff fraction is that the unknown images may not be well approximated by the eigenspace, simply because all of the model views are undilated whereas each unknown image is dilated. For "thin" features like intensity edges, the dilated images are quite different in appearance and thus are not necessarily well represented by the eigenspace. For edge features better performance is achieved if the subspace is created using both dilated and undilated versions of each model view (i.e., using both x_m and x'_m to represent each stored model view I_m). This approach is taken for the experiments reported below.

4 The Error Introduced by Subspace Approximations

We now turn to the question of how much error is introduced in using a subspace representaiton to approximate the SSD and the Hausdorff fraction. This error can be used to determine whether the best matching view is sufficiently better than the next best match to be reported as the single best match. In particular, the difference between the first and second best match in the subspace can be compared with the difference between the true distances and the approximate distances. If the approximation error is larger than the difference between the two best matches, then these matches are indistinguishable given the approximation.

First we consider the error in the SSD approximation. Expanding out the SSD yields,

$$\begin{aligned} \|x_m - x'_n\|^2 &= \|g_m - g'_n\|^2 \\ &= \|(f_m + r_m) - (f'_n + r'_n)\|^2 \\ &= \|f_m - f'_n\|^2 + \|r_m - r'_n\|^2 \end{aligned}$$

The last step is because all of the cross terms involving f and r are zero.

Thus there is an error of $\|r_m - r'_n\|^2$ when using $\|f_m - f'_n\|^2$ to approximate $\|x_m - x'_n\|^2$. Of course directly determining the value of this error would defeat the goal of efficient computation, as it would be necessary to compute all N eigenvectors rather than just k of them. However we can bound the error using just the coefficients of the k largest eigenvectors,

$$\|r_m - r'_n\|^2 = \|r_m\|^2 + \|r'_n\|^2 - 2r_m^T r'_n$$

First we note that $\|r_m\|^2 = \|x_m - c\|^2 - \|f_m\|^2$, since $\|g_m\|^2 = \|x_m - c\|^2$ (it is just an orthogonal change of basis) and $\|g_m\|^2 = \|f_m\|^2 + \|r_m\|^2$. Thus $\|r_m\|^2$ can be computed from just the original image, the centroid of the images and the projection using the k largest eigenvectors.

We also note that $|2r_m^T r'_n| \leq 2\|r_m\| \cdot \|r'_n\|$. Thus the total error is in the range,

$$\|r_m\|^2 + \|r'_n\|^2 \pm 2\|r_m\| \cdot \|r'_n\|$$

or equivalently, $\|x_m - x'_n\|$ lies in the range

$$\|f_m - f'_n\| + \|r_m\|^2 + \|r'_n\|^2 \pm 2\|r_m\| \cdot \|r'_n\|$$

The quantity $\|r_m\| \cdot \|r'_n\|$ is a relatively loose bound on the magnitude of $r_m^T r'_n$. A tighter bound would also take into account the angle between the two vectors. While this angle cannot be computed efficiently, it may be possible to use the distribution of angles between two randomly chosen $(N-k)$-vectors to produce a distribution of estimated error magnitudes.

In order to compute the error ranges efficiently, for each x_m, $1 \leq m \leq M$, $\|r_m\|$ and $\|x_m - c\|$ can be computed and stored along with the k nonzero coefficients of f_m. Then for a given image x'_n, we compute $\|r'_n\|$ and $\|x'_n - c\|$ when f'_n is computed.

For the Hausdorff fraction, first we consider the error in using $f_m^T f'_n$ as an approximation for $g_m^T g'_n$ is $r_m^T r'_n$. As above, while we cannot compute this term efficiently we can bound its magnitude by $\|r_m\| \cdot \|r'_n\|$ which can be computed efficiently.

The correlation $x_m^T x'_n$ can be seen to be in the range

$$f_m^T f'_n + x_m^T c + x'^T_n c - \|c\|^2 \pm \|r_m\| \cdot \|r'_n\|$$

The amount of error in the overall approximation to the Hausdorff fraction is thus bounded by

$$\varepsilon_m = \frac{\|r_m\| \cdot \|r'_n\|}{\|x_m\|^2}$$

Note that each of the terms in this expression can be pre-computed for each model view and computed once for an image view. Thus the uncertainty interval can be computed easily as part of the matching process. The true Hausdorff fraction, $\Phi_d(I_m, I_n)$, lies in the interval $[\hat{F}_m - \varepsilon_m, \hat{F}_m + \varepsilon_m]$ (of course the true fraction can never be less

(a) (b) (c) (d)

Figure 1: Error introduced by the subspace approximation. (a) The edges of a model image. (b) The edges after projection into the subspace and reconstruction using only the first 76 eigenvectors. (c) The dilated edges of an unknown image. (d) The edges after projection into the subspace and reconstruction using only the first 76 eigenvectors.

than 0 or greater than 1), where $\hat{F}_m$ is the approximate fraction using the eigenspace, as defined in equation (3).

In practice, the actual errors in the approximation to the Hausdorff fraction are considerably smaller than the error bound given above. This is because the error bounds the worst possible case, where the two vectors are pointing in exactly the same direction and all of the errors multiply together, which is very unlikely. For cases where the true Hausdorff fraction is not large, the estimated fraction is typically very close to the true fraction (within $\pm.05$).

In order to examine the errors in the subspace approximation to the Hausdorff fraction, we ran an experiment using a subset of the image set from [10]. This set of images consists of views of 20 different three-dimensional objects. 60 views of each object were created by placing each object on a turntable and capturing an image at regularly spaced rotations of the turntable. We subsampled these images to 64×64 pixels and used the even numbered views as the model image set and the odd numbered views as the unknown image set. In these experiments we used the 76 most significant eigenvectors to approximate the set of training images. Figure 1 gives examples showing the reconstruction of both an undilated image and a dilated image after projecting them into the subspace. Figure 2 shows a plot of the approximate Hausdorff fraction versus the true Hausdorff fraction for 10,000 pairs of model images with unknown images (that were not part of the training set).

Note that as the true fraction $\Phi_d(I_m, I_n)$ becomes large, the approximate fraction $\hat{F}_m$ sometimes underestimates the correct value. The reason for this is that, in closely correlated images, r_m and r'_n will have similar directions, which results in $\hat{F}_m$ being less than $\Phi_d(I_m, I_n)$. In the extreme case, if the dilated unknown view was exactly the same as the model view, then $\Phi_d(I_m, I_n)$ would be underestimated by $\frac{||r_m||^2}{||x_m||^2}$ since r_m and r'_n would be the same.

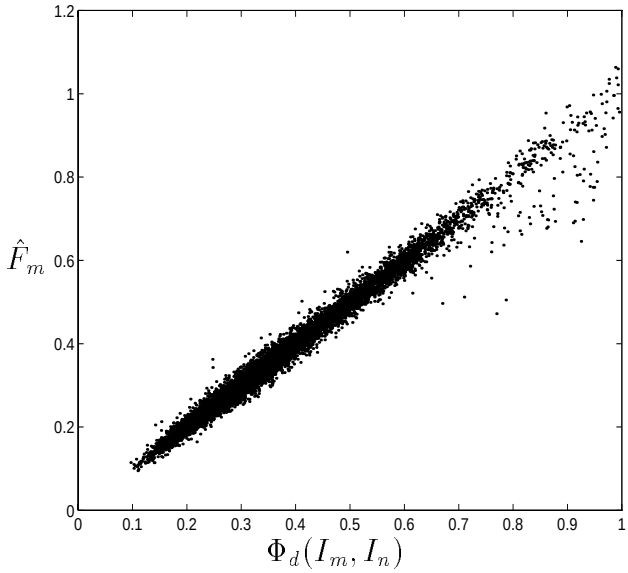


Figure 2: Plot of the correct fraction versus the estimated fraction in the image subspace for an experiment with 100 model images and 100 unknown images.

5 Experimental Results

We now consider some simple experiments which illustrate the matching performance of the Hausdorff eigenspace techniques. We are particularly interested in comparing the performance of these techniques with previous eigenspace matching techniques using grey-level images, when the background is unknown or when the object is partially occluded.

These experiments also used the image set from [10], with 30 evenly spaced views of each of 20 objects as the set of model views and 30 other evenly spaced views of each of the same objects as the set of unknown images. Each of the images is divided into the foreground (object) and the background, where the background has intensity value zero. Figure 3 shows an example view for each of the objects in the image set and the edges that were detected for each example.

Each of the 600 unknown views (not used in constructing the eigenspace) was classified as one of the 20 objects by finding the closest matching model view in the eigenspace. That is, a trial was considered successful if the best match was from the same object as the unknown, regardless of the viewpoint of the unknown image and the best matching model image. For the grey-level matching both the model images and unknown images were normalized such that each has a magnitude of one. We selected as the best match the model view with the minimum approximate SSD computed using the method described in Section 2[2]. For the binary matching we computed edge maps for each image and selected the model view with the largest approximate Hausdorff fraction $\hat{F}_m$ as the best match for each unknown image (or using binary correlation for those experiments).

First it should be noted that using the true Hausdorff fraction Φ_d (with no subspace approximation) did not exhibit perfect performance in selecting the correct object (it was successful in 96% of the trials). The reason that the true Hausdorff fraction was unsuccessful was typically due to unknown images that had dense edges, such that a very high fraction of pixels in the model view that were near image pixels in the unknown image. This is because of the asymmetry of the Hausdorff distance, which only measures the degree to which the model is accounted for by the image, and not vice versa. Figure 4 shows examples of correct and incorrect matches for the true Hausdorff distance. In the incorrect matches the sparse edges of the incorrect model view were well matched by the dense edges of the unknown image, but the reverse was not true. When comparing uncluttered images, like those used in this experiment, better results are obtained using the *bidirectional* Hausdorff fraction $\min(\Phi_d(I_m, I_n), \Phi_d(I_n, I_m))$. However, using the bidirectional fraction makes the measure more sensitive to clutter, because of the insistence that a high fraction of feature points in the unknown image lie near feature points of the model view. In the

[2]Note that Murase and Nayar use a more complicated method where each object is represented by a manifold in the eigenspace. This manifold is approximated from the points corresponding to individual views using a spline interpolation technique.

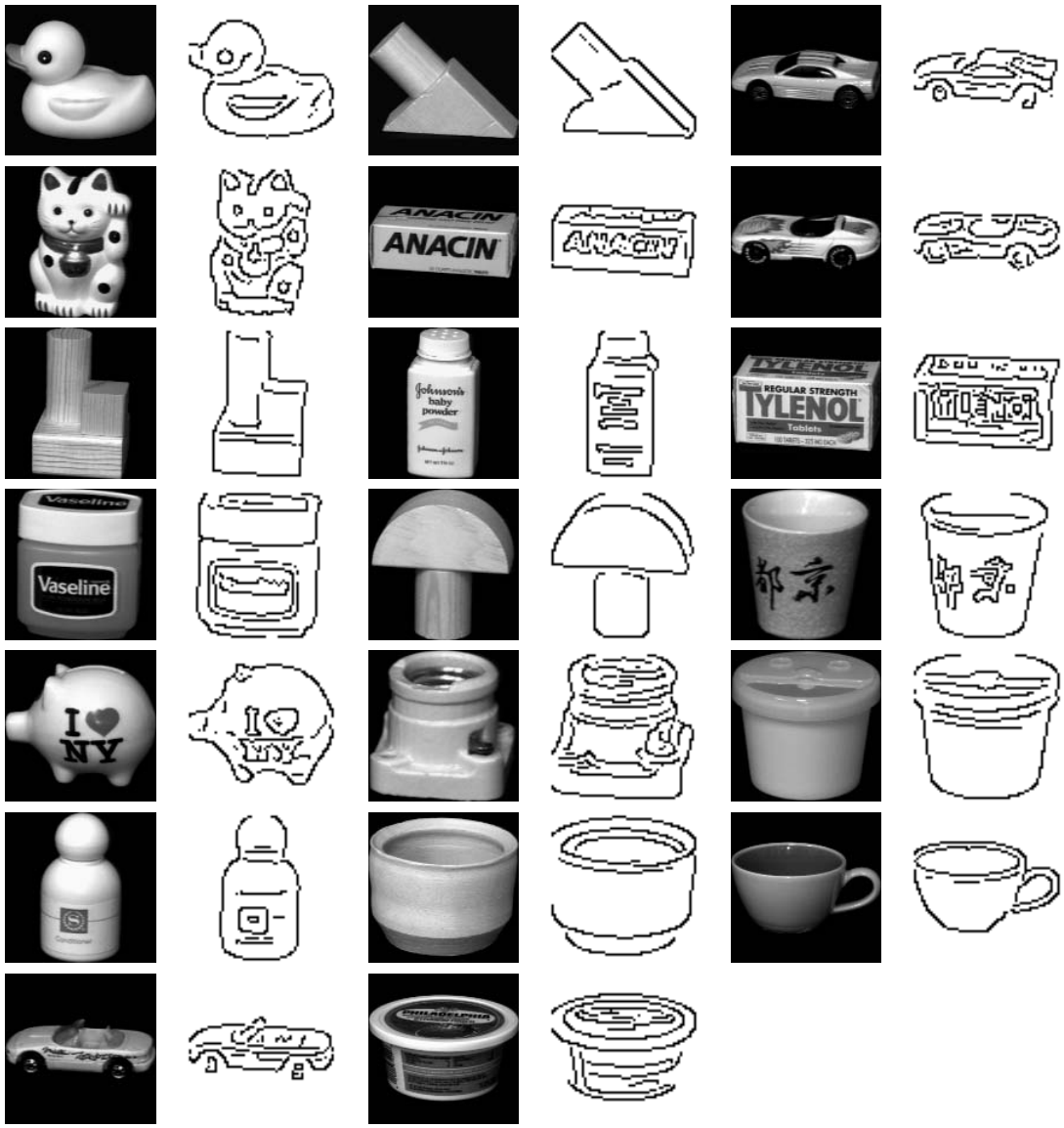


Figure 3: A single view of each of the objects and the edges found in each view.

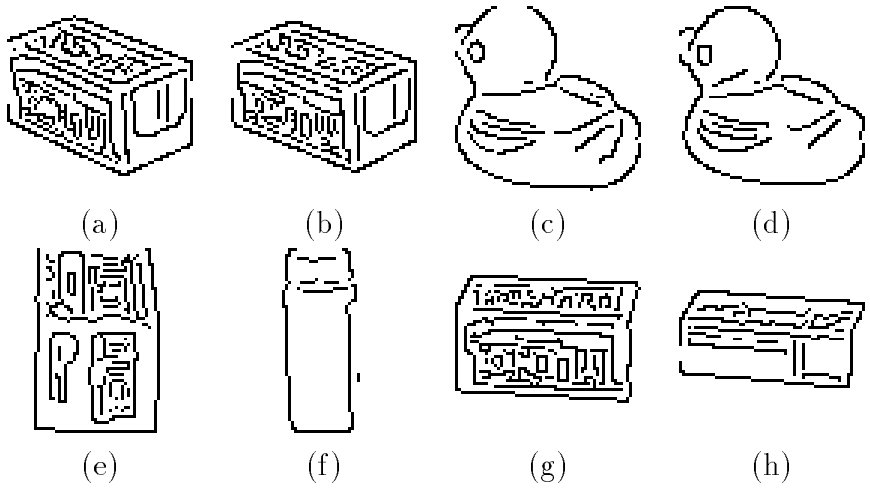


Figure 4: Examples of correct and incorrect matches that occur when the actual Hausdorff fraction is used. The correct matches are from slight rotations of the object. The incorrect matches are from different objects. (a) A view of a Tylenol box. (b) The best scoring match. (c) A view of a rubber duck. (d) The best scoring match. (e) A view of a Tylenol box. (f) The best scoring match. (g) A view of a Tylenol box. (h) The best scoring match.

experiments below, we report approximations to both the Hausdorff fraction and the *bidirectional* fraction.

When we use the original unperturbed test images, the grey-level matching techniques in the image subspace yield perfect performance, while the Hausdorff subspace matching technique is successful 96% of the time (575 of 600 trials). The Tylenol box accounted for 16 of the unsuccessful trials, with 3 other models accounting for the remaining unsuccessful trials. The subspace approximation of the *bidirectional* Hausdorff fraction was successful 99% of the time (593 of 600 trials).

It is important to note that for the approximate Hausdorff fraction we can generally detect when it is likely to be incorrect, by looking at the quality of the next-best-match. For successful trials, there was no match that was nearly as good as the correct match, whereas for the unsuccessful trials there generally were other close matches. In particular, for the successful trials, the difference between the largest $\hat{F}_m$ for a view of the correct object and the largest $\hat{F}_m$ for a view of any other object was .238 on average. In contrast, for the unsuccessful trials this difference was .017 on average, with a maximum value of .041.

Recall that the error in approximating the Hausdorff fraction is about .05. Thus for the unsuccessful trials there are multiple matches within the uncertainty of the approximation, whereas for the successful trials there are not. This provides empirical support for the error analysis above, which suggests that all matches within the

Image change	Grey-Level SSD	Directed Hausdorff	Bidirectional Hausdorff	Normalized Correlation
Unperturbed	100% (600)	96% (575)	99% (593)	89% (532)
Background=50	94% (564)	95% (567)	98% (585)	89% (535)
Background=100	41% (248)	95% (568)	95% (571)	88% (530)
25% occlusion	52% (314)	88% (528)	97% (583)	87% (523)
50% occlusion	51% (309)	85% (510)	90% (538)	84% (503)

Table 1: Summary of results for the subspace image matching experiments. The first column is for the normalized correlation of the grey-level images. The second column is for the Hausdorff fraction of the edge maps. The third column is for the bidirectional Hausdorff fraction described in the paper. The fourth column is for the normalized correlation of the edge maps. All results are using the subspace approximation with 76 coefficients.

uncertainty range of the best match should be considered. We find here that the best match is correct in all cases when there are no matches to views of other objects within this uncertainty range. Moreover, the best match turns out to be incorrect in all the cases where there are such close matches.

We next considered the case in which the unknown images were modified such that the background intensity (which was zero in the original images) was changed to a uniform non-zero value. The overall image was still normalized to have unit length for the grey-level matching using the SSD. The edges of the unknown images were recomputed after the change of background intensities for the binary matching. When the background of the unknown images was changed changed to 50, the grey-level techniques were successful 94% of the time (564 of 600 trials). When the background value was changed to 100, the grey-level techniques were successful only 41% of the time (248 of 600 trials). Thus the grey-level techniques, not surprisingly, are fairly sensitive to large changes in the background intensity, because all pixel differences contribute equally to the overall measure. These changes yielded little difference for the Hausdorff techniques, yielding 95% success in both cases (567 and 568 successful trials, respectively).

Finally we return to images with a uniform, black background, but in which the object was partly occluded. We simulated occlusion of 25% of the object by setting the upper, left quarter of the image to a black background in the grey-level images and by erasing the edge pixels in this region for the edge images. In this experiment, the grey-level techniques were successful in 314 trials, while the Hausdorff techniques were successful in 528 trials. When the entire left half of the image was occluded, the grey-level techniques yielded 309 successful trials and the Hausdorff techniques yielded 510 successful trials.

Table 1 gives a summary of the results for the eigenspace approximations to the

(a) (b) (c)

Figure 5: Tests were run on images with random clutter. (a) Test images not included in forming the subspace. (b) The test images with 20% random clutter added. (c) The best matching model images for the test images according to the approximate Hausdorff fraction.

grey-level SSD and to both the directed and bidirectional Hausdorff fractions. For comparison, results are also given for the approximation to the normalized correlation. Perhaps the most striking overall result is that the edge-based measures (Hausdorff or correlation) suffer much less than the grey-level measures as the background intensity is changed. This indicates that while edge detectors are sensitive to changes in illumination, they can be considerably less sensitive than the normalized intensity values. We believe that this suggests a view-based approach to recognition which makes use of features extracted from views (not simply edges, but multiple types of features) rather than the views themselvs.

The second overall result seen in Table 1 is that the Hausdorff matching techniques have uniformly good performance, whereas the grey-level techniques break down when the background is changed and when the object is partially occluded. The Hausdorff measure also performs significantly better than the normalized binary correlation of the edge maps. The improvement over binary correlation is to be expected, because the Hausdorff fraction handles small perturbations in the locations of image features (whereas for binary correlation either feature points are directly superimposed or they do not match).

In the next set of experiments we considered the effects of random edge clutter on the edge matching techniques. In these experiments we added random straight edge segments to each image until a specified fraction of the white space was covered by clutter (see Figure 5). In this case the performance of the Hausdorff matching techniques degraded as additional clutter was added, but even when 40% of the non-edge pixels were changed to clutter, the techniques identified the correct object in over

Clutter Percentage	Directed Hausdorff Measure	Normalized Correlation of Edges
0%	96% (575)	89% (532)
10%	94% (562)	72% (433)
20%	89% (533)	53% (318)
30%	82% (492)	44% (264)
40%	73% (437)	32% (193)

Table 2: Percentage and number of successful trials as a function of image clutter in the subspace experiments. The clutter percentage is the fraction of background pixels that were covered by straight edge segment clutter. Results are given both for the directed Hausdorff measure and the normalized correlation of the edge maps. The number of successful trials is out of 600 total trials.

70% of the trials. The performance of the normalized correlation of the edge maps degraded much faster, for instance at 10% occlusion the Hausdorff-based measure achieved 96% correct classification whereas the normalized binary correlation was only 72% correct.

6 Image search

In many recognition tasks the positions of objects that may be present in the image are unknown. Moreover, current segmentation methods cannot reliably determine the regions of an image that correspond to separate objects, except in simple cases. For this reason it is crucial to have methods for quickly searching an image for locations where there may be a match of one of the views in a set of model views. In this section we describe how to integrate the Hausdorff-based subspace matching technique into an image search engine. When the set of model views is larger than about 200, we obtain substantial speedup over techniques that separately search for each model view in an unknown image. These running time comparisons are done using the Hausdorff matching methods reported in [5, 12], which have been heavily optimized.

We first consider the simple experiment of using the eigenspace approximation to the Hausdorff fraction in order to rule out those locations (translations) in an unknown image that are a poor match in the subspace. Note that the subspace techniques need not rule out all of the incorrect translations of the model. As long as the vast majority of the locations and models are eliminated, without eliminating the correct matches, we can use standard techniques to check the remaining hypotheses. We rely on the fact that the approximate Hausdorff fraction is nearly always close to the true fraction as a heuristic to avoid ruling out correct matches. We use a

threshold fraction that is .05 smaller than that specified by the maximal amount of occlusion allowed (because the empirical data in Figure 2 illustrates that the true fraction is nearly always within .05 of the estimated one).

Figure 6 shows an example of a cluttered image that was used in these experiments. In this example, we are attempting to locate the Tylenol box, which had by far the worst performance of the 20 models in our previous experiments. Let's consider a case where we wish to find all translations of one of the 600 views of the 20 models. The image is 220×170 pixels and the model views are each 64×64 (as above). As there are about 25 thousans locations in the image to search, a brute force search for all 600 models would perform about 15 million comparisons of stored model views to the image.

In order to allow for some mismatch between the model views and the image, we allow for 20% of the model edges to be unmatched by the image (due to differences in the edge features from lighting, slight viewpoint differences, etc). In order to ensure that it will be unlikely that we rule out any such translation, we eliminate from consideration only those model/translation pairs that have $\hat{F}_m < 0.75$ (allowing for a .05 error in the approximation). In this image, the only model with such a match is the Tylenol box, which has a true Hausdorff fraction of 0.844 at the best match. By ruling out all locations where the best *approximate* fraction for all the model views in the subspace is less than 0.75, we are able to eliminate 98.7% of the translations in this image. Note that the best approximate match over the entire image, which is shown in Figure 6, is a view of the Tylenol box, with estimated fraction 0.836 (which is quite close to the true fraction of 0.844).

Figure 7 shows an example of an image where the model (the Anacin box) was partially occluded. We want to allow for 25% mismatch in this case (due to the small amount of occlusion) and thus set the threshold at 0.7. The best match shown in the figure yielded a true Hausdorff fraction of 0.702 and the subspace methods yield an estimated fraction of 0.727. When we eliminate all translations that yield a best estimated fraction below 0.7, 99.3% of the search space is pruned. Experiments with images like these indicate that the subspace matching techniques can eliminate most of the possible positions of the model images in a large unknown image without performing full comparisons of model views against the image at these positions. We thus expect these techniques to yield a considerable improvement in the speed of image matching techniques using the Hausdorff fraction.

6.1 Subspace Matching in an Image Search Engine

The subspace approximation to the Hausdorff fraction can be integrated into a multi-resolution search strategy to achieve additional speedup over a separate search for each model view. The basic idea behind multi-resolution strategies for Hausdorff matching is to exploit the fact that if there is not a good match at a particular location, then this fact can be used to eliminate other nearby locations from consideration. When

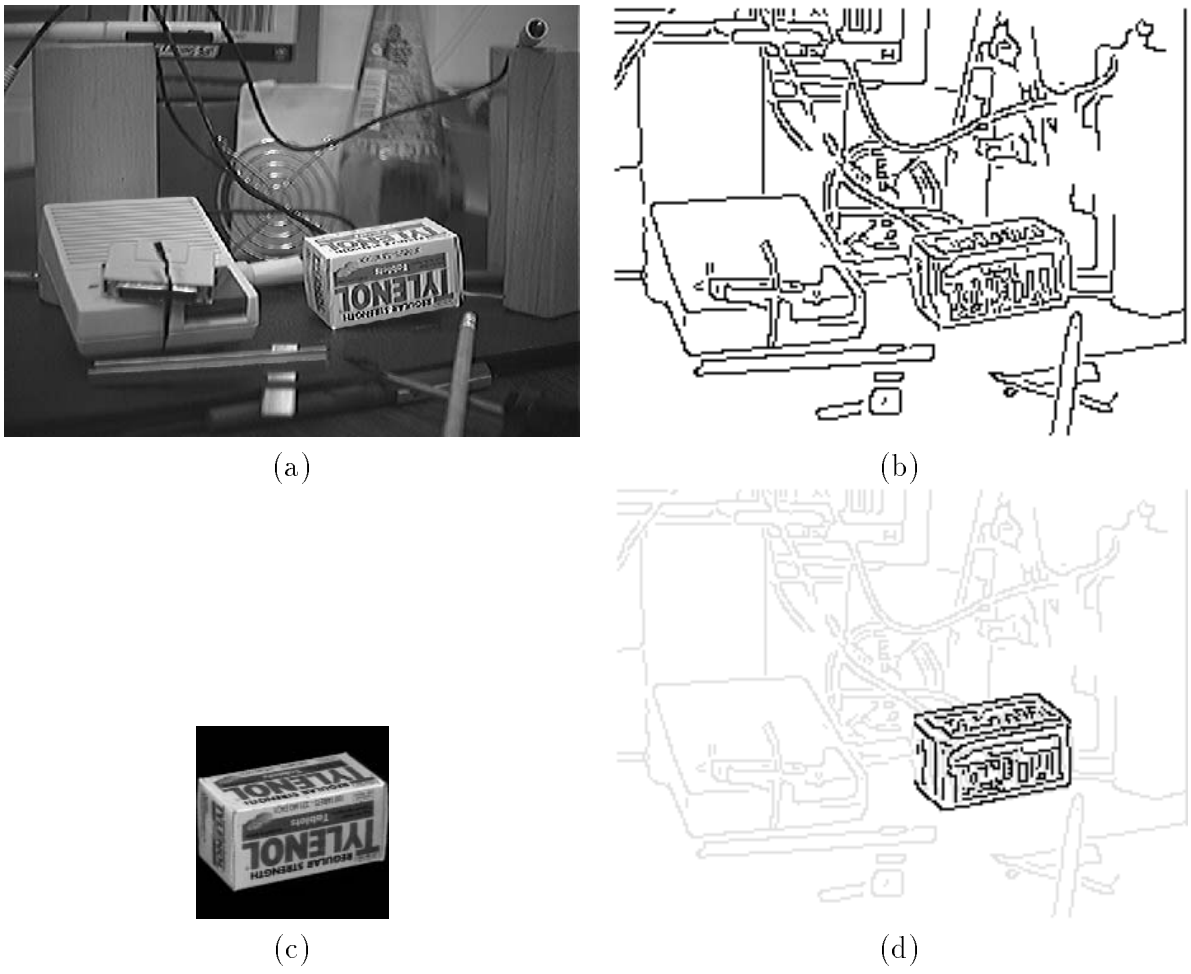
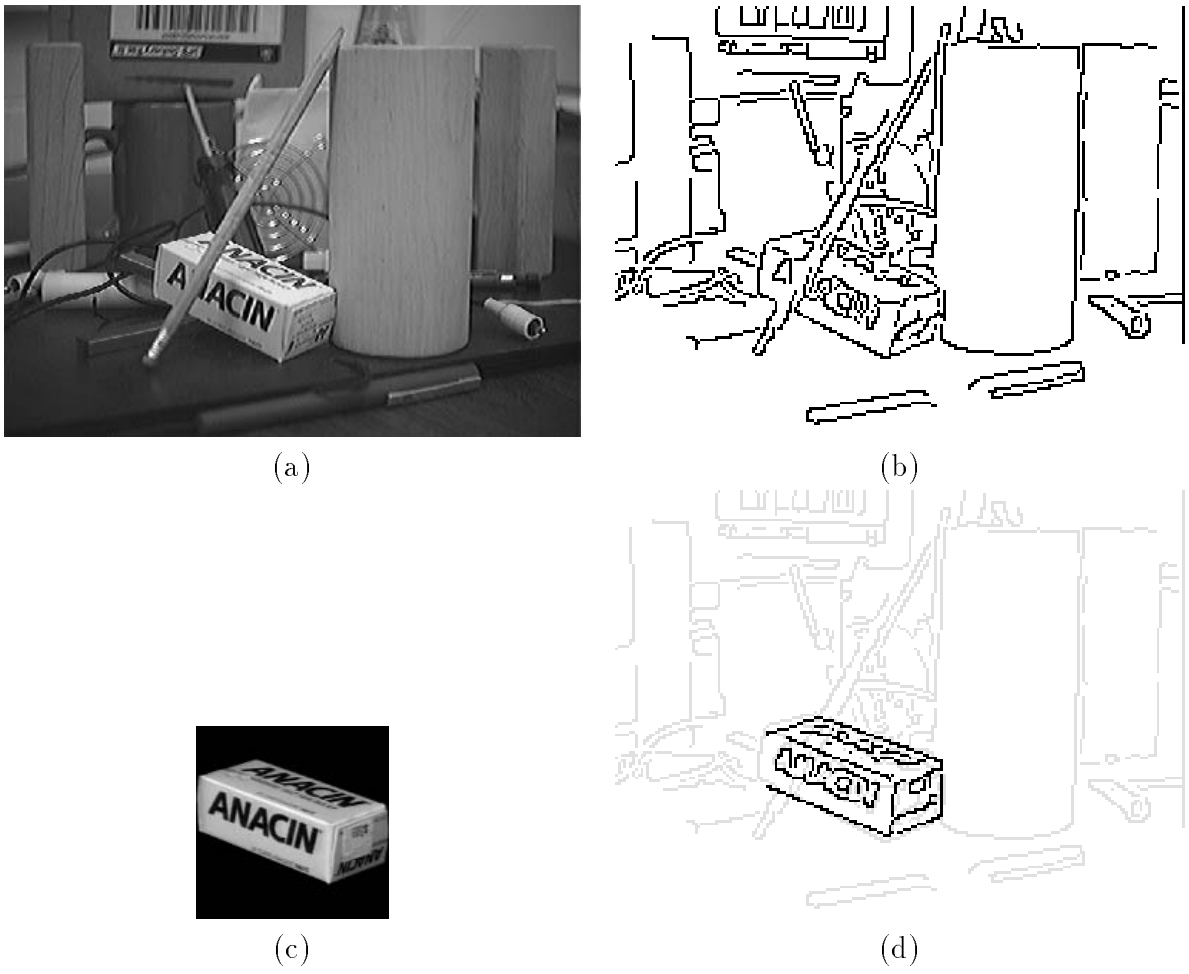


Figure 6: A cluttered image that was used to test the image search. (a) The original image. (b) The edges detected in the image. (c) The best matching view of the Tylenol box. (d) The edges of the Tylenol box overlaid on the full edge image at the location of the best match.



(a) (b)

(c) (d)

Figure 7: A cluttered image with some occlusion that was used to test the image search. (a) The original image. (b) The edges detected in the image. (c) The best matching view of the Anacin box. (d) The edges of the Anacin box overlaid on the full edge image at the location of the best match.

searching under translation, one strategy that can be used is to dilate the image by a disk with a radius greater than the desired error radius, δ. If a model does not match this highly dilated image at some position, then this position of the model and other positions close to it can be ruled out as possible matches in an image that is dilated only by δ. To make this concrete, let's say that we dilate the image with a disk of radius $\delta + \gamma$. If the match of the model to this dilated image at some translation does not surpass the specified Hausdorff fraction, then no translation within γ can possibly yield a match in the image when dilated by only δ. This thus allows us to rule out all translations that are within γ of the current translation.

We can formulate an efficient search strategy using this observtion by considering a hierarchical cell decomposition of the search space. The translations are divided into cells of uniform size (which can recursively be divided into similarly uniform cells) [6, 12]. We then create a new image dilated by a disk with a radius equal to the distance from the center of the cell to the cell boundary plus the error allowed, δ. This allows an entire cell to be ruled out or expanded by only examining the translation at the center of the cell. For each cell that cannot be ruled out at this level, we divide the cell and apply the process recursively until the final cells consist of a single translation of the model, which are good matches between a model and the image according to the subspace approximation. The use of these techniques requires the computation of several dilations of the image, at different radii, but this is more than compensated by the reduction in the number of positions of the models that have to be examined with a brute force search.

Since we are using a subspace approximation to the Hausdorff fraction, we can only determine whether a match exists up to the error in this approximation. As above, if we set our threshold for ruling out a cell lower than the actual threshold we are interested in, we can be reasonably certain that we do not rule out any cells that we should not. We again use a heuristic of .05 error in the approximation. At the bottom level of the hierarchy, when we reach cells that contain a single translation which cannot be ruled out, we can compute the true Hausdorff fraction rather than using the eigenspace approximation, since there will be few such cells that remain, and each such cell will only have a small number of possible matching model views.

Figure 6.1 shows a running time comparison between our implementation of a hierarchical image search using the subspace Hausdorff matching techniques and a previous implementation of hierarchical search using the true Hausdorff fraction [5], both running on a SPARC-10. The previous system has been heavily optimized in order to efficiently rule out regions of the search space that do not need to be considered, but it does not use the subspace techniques for approximating the Hausdorff fraction. While the subspace techniques are not as heavily optimized and have additional overhead associated with mapping subimages of the unknown image into the subspace, the time required by these techniques grows slowly with number of objects in the database. As the set of models grows large, the subspace image search method outperforms the previous techniques by a considerable margin. From the graph it

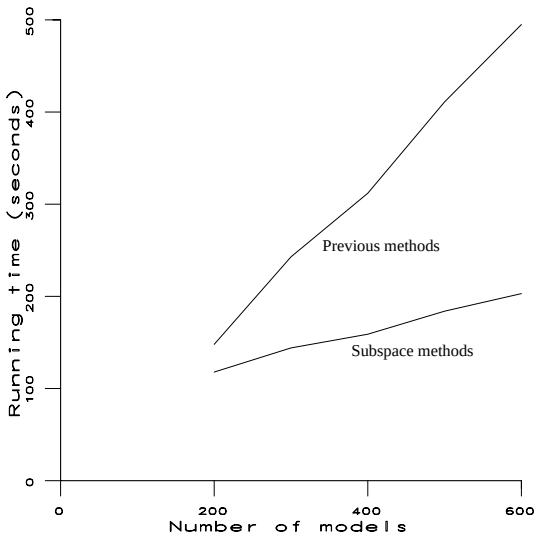


Figure 8: The time required by the subspace methods grows far less quickly than previous Hausdorff matching techniques as the number of model images in the database grows.

can be seen that for 200 model views the subspace method already has about a 20% speed advantage over the method which considers each model view independently. When the model set reaches 600 views, the speed advantage is about 300%. As noted above, a brute force comparsion of all 600 views would involve about 15 million matches of model views to the image. Therefore the effective speed of the subspace method is about 75,000 model view matches per second on a SPARC-10. This speed is achieved both by pruning the space of possible translations that are considered and by matching the eigen-coefficients rather than the complete model views.

While we have only considered searching over possible translations of the object models in an image, it is also possible to consider other transformations such as scaling, rotation or affine. One method by which this could be done is to include scaled and rotated versions of the model images in the database [14], but this method yields very large catalogs of model images. Alternately, we can explore the space of such transformation together with the space of possible translations. First, the transformation space is discretized such that no two adjacent transformations map any model pixel more than one pixel apart in the image. We can then consider cells of this transformation space as above in the multi-resolution search strategy. Such an approach to Hausdorff matching is taken in [12], without the use of a subspace approximation.

7 Summary

We have descried a new subspace method for approximating the Hausdorff fraction between two binary images, and have demonstrated the use of this method for view-based recognition. The use of edge features rather than grey-level images yields a view-based recognition technique that is relatively insensitive to lighting changes and to unknown backgrounds. The use of the Hausdorff fraction to compare feature maps provides robustness to clutter and occlusion. The eigenspace approximation to the Hausdorff fraction allows individual matches to be processed much faster than previous Hausdorff matching methods. Thus, overall this combination of techniques results in a system that has both the speed of subspace methods and the robustness of the Hausdorff measure.

Empirical results presented in the paper indicate that the Hausdorff based eigenspace method provides a substantial improvement over SSD-based methods, in situations where the background is unknown or cluttered or objects are partially occluded. In addition, these experiments suggest that it is possible to detect when the Hausdorff matching techniques are likely to have selected an incorrect match, based on the distance between the best match and the next-best match. Finally the paper showed how the Hausdorff eigenspace method can be used for image search, by integrating it with a multi-resolution search strategy in order to quickly identify possible instances of a set of model views at unknown locations in an image. Comparison with prior methods for performing Hausdorff matching (that did not use subspace techniques) shows considerable improvement in matching time when a set of a few hundred model views is used.

Acknowledgements

We are grateful to Hiroshi Murase and Shree Nayar for making their object model database available to us. This work was supported in part by DARPA under ARO contract DAAH04-93-C-0052 and by National Science Foundation PYI grant IRI-9057928.

References

[1] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 573–582, 1994.

[2] M. Black and A. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In *Proceedings of the European Conference on Computer Vision*, pages 239–342, LNCS volume 1064, 1996.

[3] B. Bremner, A. Hoogs, and J. Mundy. Integration of image understanding exploitation algorithms in the RADIUS testbed. in *Proceedings of the DARPA Image Understanding Workshop*, pp. 255–268, 1996.

[4] D. D. Fu, K. J. Hammond, and M. J. Swain. Using regularities of man-made environments for appropriate sensing and action. In *Proceedings of the IEEE Workshop on Context-Based Vision*, pp.22–29, 1995.

[5] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, September 1993.

[6] D. P. Huttenlocher, J. J. Noh, and W. J. Rucklidge. Tracking non-rigid objects in complex scenes. In *Proceedings of the International Conference on Computer Vision*, pages 93–101, 1993.

[7] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, January 1990.

[8] J. Krumm. Eigenfeatures for planar pose measurement of partially occluded objects. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 55–66, 1996.

[9] A. Leonardis and H. Bischof. Dealing with occlusions in the eigenspace approach. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 453–458, 1996.

[10] H. Murase and S. K. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.

[11] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[12] W. J. Rucklidge. Locating objects using the Hausdorff distance. In *Proceedings of the International Conference on Computer Vision*, pages 457–464, 1995.

[13] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[14] S. Yoshimura and T. Kanade. Fast template matching based on the normalized correlation by using multiresolution eigenimages. In *Proceedings of the International Conference on Intelligent Robots and Systems*, volume 3, pages 2086–2093, 1994.