# Object Recognition Using Subspace Methods

Daniel P. Huttenlocher, Ryan H. Lilien and Clark F. Olson

Department of Computer Science, Cornell University, Ithaca, NY 14853, USA

**Abstract.** In this paper we describe a new recognition method that uses a subspace representation to approximate the comparison of binary images (e.g. intensity edges) using the Hausdorff fraction. The technique is robust to outliers and occlusion, and thus can be used for recognizing objects that are partly hidden from view and occur in cluttered backgrounds. We report some simple recognition experiments in which novel views of objects are classified using both a standard SSD-based eigenspace method and our Hausdorff-based method. These experiments illustrate how our method performs better when the background is unknown or the object is partially occluded. We then consider incorporating the method into an image search engine, for locating instances of objects under translation in an image. Results indicate that all but a small percentage of image locations can be ruled out using the eigenspace, without eliminating correct matches. This enables an image to be searched efficiently for any of the objects in an image database.

## 1 Introduction

Appearance based recognition using subspace methods has proven successful in a number of visual matching and recognition systems (e.g. [2, 6, 4, 3]). The central idea underlying these methods is to represent images in terms of their projection into a relatively low-dimensional space which captures the important characteristics of the objects to be recognized. The most effective applications of these methods have been to problems in which objects are fully visible, against a uniform background, and are nearly correctly registered with each other. 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 [4]. 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 a common problem in any window-based matching technique, where background pixels included in a matching window can significantly alter the match.

In this paper we describe a new subspace recognition method that is designed to handle objects which appear in cluttered images and may be partly hidden from view, without prior segmentation of the objects from the background or registration of the image. This method is based on using subspace techniques to approximate the generalized Hausdorff measure [1], which measures the degree

of resemblance between binary images (bitmaps). We present some simple experiments which demonstrate that the method performs well when the background is unknown or when the object is partially occluded, including in cases where methods based on the SSD break down. More importantly, we can detect when the approximation to the generalized Hausdorff measure is likely to select an incorrect match. In addition, we need not assume that the location of the unknown object in the image is known. Instead we can incorporate our eigenspace matching methods into an image search engine, enabling the vast majority of image locations to be ruled out for all of the models in a large database.

## 2 Subspace approximation of SSD

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 $\lambda_i$, $1 \leq i \leq N$, 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 just the weighted sum of the eigenvectors). It is straightforward to show that $||x_m - x_n||^2 = ||g_m - g_n||^2$ [3], 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 all $N$ eigenvectors,

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

for $k << N$ (where $g_{m_i}$ denotes the $i$-th element of the vector $g_m$). 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. The SSD, $||x_m - x_n||^2$, is then approximated as $||f_m - f_n||^2$. As this representation uses just the $k$ predominant eigenvectors, 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).

## 3 Approximating the Hausdorff Fraction

In this section we describe a subspace method for approximating the generalized Hausdorff measure. Note that we are now restricting the discussion to binary images, which can be thought of as representing sets of feature points on a grid (i.e., a binary image is 1 for points in the set and 0 otherwise). First we review the generalized Hausdorff measure. 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 [1, 5] 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 classical distance uses the maximum element rather than some chosen rank.

The generalized Hausdorff measure is asymmetric (as is the classical 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 form 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. One means of using the measure is to specify a fixed distance, $d$, and then determine the resulting fraction of points that are within that distance. In other words, to 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, $d$. This has been termed "finding the fraction for a given distance." It measures how well two sets match, with larger fractions being better matches.

The subspace method that we present in this paper is based on finding the fraction for a given distance. Assume that the points of $\mathcal{P}$ and $\mathcal{Q}$ have integral coordinates and let $P$ be a binary image denoting the set $\mathcal{P}$, with a 1 in $P$ corresponding to a point that is in $\mathcal{P}$ and a zero otherwise. Similarly for $Q$ and $\mathcal{Q}$. We are interested in what fraction of the 1's in $P$ are near (within $d$ of) 1's in $Q$. 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 fraction for a given 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. This measure has also been termed

the Hausdorff fraction. It is the fraction of points in $P$ that lie within distance $d$ of points in $Q$.

Given two binary images, $I_m$ and $I_n$, if we let $x_m$ be the representation of $I_m$ as a column vector and $x_n$ be the representation of $I_n^d$ (the dilated $I_n$) then $\Phi_d(I_m, I_n)$ can be computed as follows,

$$\Phi_d(I_m, I_n) = \frac{x_m^T x_n}{||x_m||^2}$$

The Hausdorff fraction, $\Phi_d$, can be approximated using the subspace approximation to the correlation. First we look at the relation between the correlation 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 the new coordinate system defined by the eigenvectors $E$ of $XX^T$.

$$\begin{aligned} x_m^T x_n &= (x_m - c + c)^T (x_n - c + c) \\ &= (x_m - c)^T (x_n - c) + (x_m - c)^T c + (x_n - c)^T c + ||c||^2 \\ &= g_m^T g_n + x_m^T c + x_n^T c - ||c||^2 \end{aligned}$$

The last step follows from $g_m^T g_n = (x_m - c)^T E E^T (x_n - c) = (x_m - c)^T (x_n - c)$ (i.e., dot products are preserved under an orthogonal change of basis).

As in the SSD-based eigenspace methods, we approximate $g_m$ and $g_n$ using just the first $k$ coefficients, which 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 the cross terms are zero. Hence the error in using $f_m^T f_n$ as an approximation for $g_m^T g_n$ is $r_m^T r_n$. While we cannot compute this error term efficiently we can bound its magnitude by $||r_m|| \cdot ||r_n||$ which can be computed efficiently. Therefore the true correlation $x_m^T x_n$ lies in the range $f_m^T f_n + x_m^T c + x_n^T c - ||c||^2 \pm ||r_m|| \cdot ||r_n||$.

To construct the Hausdorff eigenspace for a set of *binary* "model" images, $x_1, \ldots, x_M$, form the matrix $X = [x_1 - c, \ldots, x_M - c]$, where $c$ is the centroid of the $x_m$'s. Compute and save the first $k$ eigenvectors of $XX^T$ (i.e., those corresponding to the $k$ largest eigenvalues). 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 $x_m$. This is all 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 $x_n^T c$. An explicit search of all of the models can be performed by computing the approximation to the Hausdorff fraction for each $x_m$ and the (dilated) unknown $x_n$,

$$\hat{F}_m = \frac{f_m^T f_n + x_m^T c + x_n^T c - ||c||^2}{||x_m||^2} \tag{3}$$

Note that each of the terms in this expression was computed and stored in forming the eigenspace, except for $f_m^T f_n$. Thus the matching 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.

(a) (b)

**Fig. 1.** Error introduced by the subspace approximation. (a) The dilated edges of an unknown image. (b) The edges after they have been projected into the eigenspace and then reconstructed using only the first 76 eigenvectors.

### 3.1 Error in the Approximation

The amount of error in $\hat{F}_m$ as an approximation to $\Phi_d$ can be bounded by $\varepsilon_m = (\|r_m\| \cdot \|r_n\|)/(\|x_m\|^2)$, so $\Phi_d$ lies in the interval $[\hat{F}_m - \varepsilon_m, \hat{F}_m + \varepsilon_m]$ (of course the true fraction can never be greater than 1).

One issue with approximating the Hausdorff fraction is that the unknown image is not necessarily well approximated by the eigenspace, because all of the model images are undilated and the 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. Empirically we have determined that there is a smaller residual error in the reconstructed images when the model images are dilated and the unknown image is not dilated than when the reverse is the case. Thus, we approximate the Hausdorff fraction $\Phi_d(P, Q) = \#(P \wedge Q^d)/\#(P)$ by $\#(P^d \wedge Q)/\#(P)$. This approximation is quite good for small $d$, which is generally the case as we use $d = 1$ in order to allow for uncertainty in the edge pixel locations.

In practice, the errors in the estimated fraction are considerably smaller than the error bound given above would predict. This is because the error bound is the worst case, which occurs when the two vectors are pointing in exactly the same direction and all of the errors multiply together. 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 [3]. This set of images consists 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 downsampled these images to $64 \times 64$ 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. Fig. 1 shows an example of a dilated image from this data set and the reconstruction of this image after projecting it into the eigenspace. Fig. 2 shows a plot of the approximate Haus-

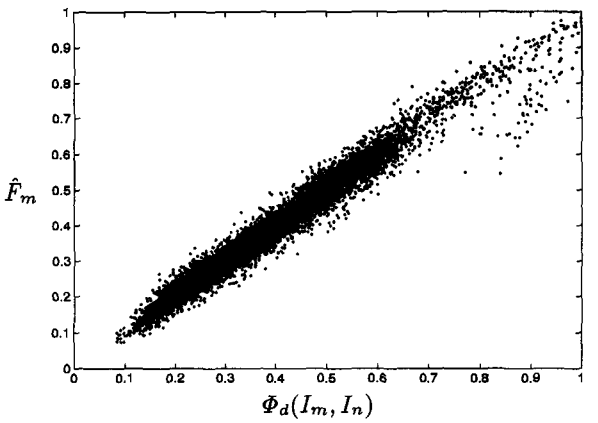



**Fig. 2.** Plot of the correct fraction versus the estimated fraction in the image subspace for an experiment with 100 model images and 100 unknown images.

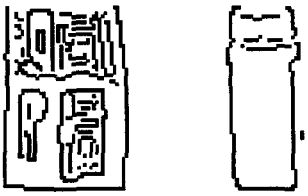
dorff fraction versus the true Hausdorff fraction for 10,000 pairs of model images with unknown images (that were not part of the training set).

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 model image was exactly the same as the unknown image, 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. Of course, we will never reach this extreme since the model images are dilated and the unknown images are not.

## 4 Matching experiments

We now consider some experiments to evaluate the discrimination ability of these matching techniques. We are particularly interested in comparing the performance of these techniques with previous techniques using grey-level images (e.g., [3]) when the background is unknown or when the object is partially occluded. These experiments used the image set from [3], with 30 evenly spaced views of each of 20 objects as the model set and 30 other evenly spaced views of each of the same objects as the set of unknown images. The backgrounds in all the images are dark black.

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 for an unknown image, the



(a) (b)

**Fig. 3.** An example where the directed Hausdorff fraction yields an incorrect match, but the bidirectional measure does not. (a) The unknown image. (b) The incorrect match.

model image with the minimum approximate SSD computed using the method described in Section 2 For the binary matching we computed edge maps for each image and selected the model image with the largest approximate Hausdorff fraction $\hat{F}_m$ as the best match for each unknown image.

First it should be noted that using the actual Hausdorff fraction, $\Phi_d$, to select the best matching view did not exhibit perfect performance in selecting the correct object. It was successful in 96% of the trials (575 of 600). The reason that the true Hausdorff fraction was unsuccessful was typically due to unknown images that had dense edges, such that the fraction of model pixels that were near image pixels was very high. This is because of the asymmetry of the Hausdorff distance, which only measures how well the model is accounted for by the image, and not vice versa. Fig. 3 shows an example. In this case, the sparse edges of the incorrect match were well matched by the unknown image, but reverse was not true. The bidirectional Hausdorff measure yields better results for this case (99% correct), since the images are uncluttered. This is analogous to the SSD performing better in uncluttered images; both the SSD and the bidirectional Hausdorff measure take advantage of the excess clutter to rule out possible matches, which results in neither being robust to significant image clutter.

Using the unperturbed images the grey-level matching techniques have perfect performance, while the Hausdorff subspace matching techniques are successful in 551 of 600 trials. Of the 49 unsuccessful trials, 23 were also unsuccessful when the true Hausdorff fraction was used to find the best match. One model accounted for 28 of the unsuccessful trials, with 8 other models accounting for the remaining unsuccessful trials. It is important to note that we can detect when the approximate Hausdorff match is likely to be incorrect. 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 .234 on average. In contrast, for the unsuccessful trials this difference was .015 on average, with a maximum value of .090. We should thus consider not only the match with the largest approximate Hausdorff fraction, but also any matches with approximate Hausdorff fractions that are nearly as large. The subspace version of the bidirectional measure was

| Image change | Grey-level | Directed Hausdorff | Bidirectional Hausdorff |
|---|---|---|---|
| Unperturbed | 100% (600) | 92% (551) | 98% (589) |
| Background=50 | 94% (564) | 93% (556) | 97% (580) |
| Background=100 | 41% (248) | 90% (542) | 91% (546) |
| Shift by 50 | 95% (568) | 92% (551) | 98% (589) |
| Shift by 100 | 48% (291) | 92% (551) | 98% (589) |
| 25% occlusion | 52% (314) | 87% (524) | 94% (565) |
| 50% occlusion | 49% (291) | 83% (501) | 85% (507) |

**Table 1.** Summary of results for the subspace image matching experiments using the normalized correlation of grey-level images and the Hausdorff fraction of edge images. The results show the percentage (number) of trials out of 600 that were successful.

successful in 589 of 600 trials.

We next considered unknown images where the background had been changed to a uniform non-zero value. The overall image was still normalized to be a vector of unit length. When the background of the unknown images was changed changed to 50, the grey-level techniques were successful in 564 of 600 trials. When the background value was changed to 100, the grey-level techniques were successful in only 248 of 600 trials. These changes yielded little difference for the Hausdorff techniques, yielding 556 and 542 successful trials, respectively. When the grey-levels in the entire image were shifted up by 50 and 100 values, the grey-level techniques were successful in 568 and 291 trials, respectively. Such a shift had no effect on the the Hausdorff matching techniques.

Finally we returned to images with a uniform background, but in which the object had been occluded. We occluded 25% of the object by setting the upper, left quarter of the image to the background color 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 524 trials. When the entire left half of the image was occluded, the grey-level techniques yielded 291 successful trials and the Hausdorff techniques yielded 501 successful trials.

Table 1 summarizes the subspace results for the grey-level matching techniques and for both the directed and bidirectional Hausdorff matching techniques. The Hausdorff matching techniques have uniformly good performance, whereas the grey-level techniques break down when the background or the total brightness is changed and when the object is partially occluded.

## 5 Image search

In many applications the positions of the object(s) that may be present in an image are not known. Moreover, current segmentation methods cannot reliably

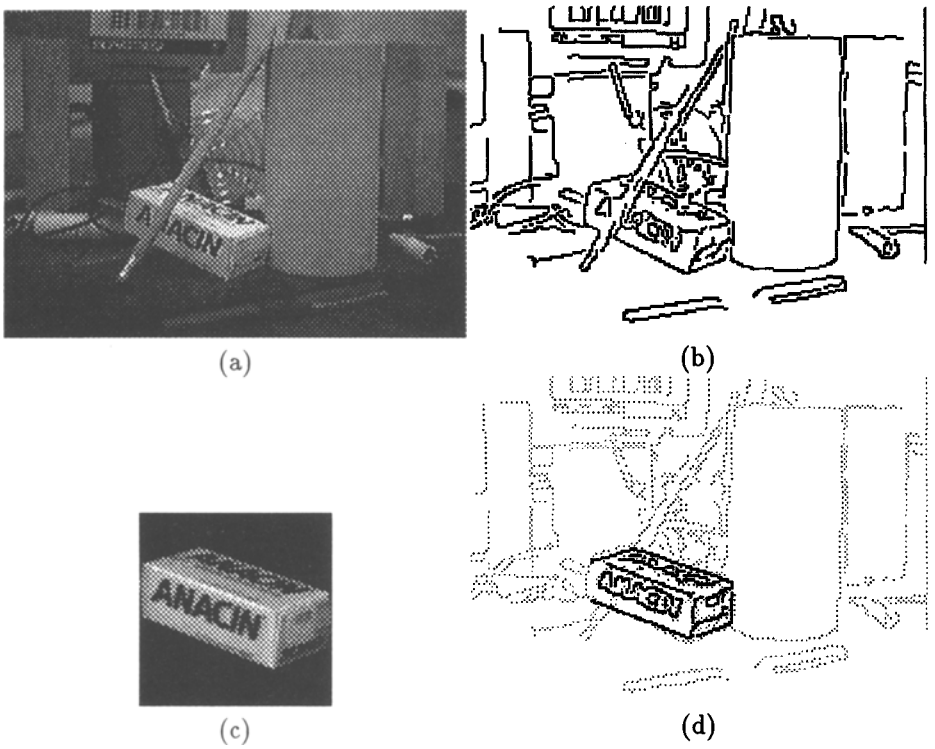



**Fig. 4.** A cluttered image with some occlusion that was used to test the image search. (a) The original image. (b) The edges detected in the image. (c) The best matching view of the Anacin box. (d) The edges of the Anacin box overlaid on the full edge image at the location of the best match.

determine the regions of an image that correspond to separate objects, except in very simple cases. In this section we consider the simple experiment of using the eigenspace approximations to rule out those locations (translations) in an unknown image that are a poor match in the subspace. 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 depend on the fact that the approximate Hausdorff fraction is nearly always close to (within $\pm 0.05$ of) the true fraction to avoid ruling out correct matches (see Fig. 2).

Fig. 4 shows an example of the kind of image that was searched in these experiments. The instance present in this image is the Anacin box, which is partially occluded. In this case 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. A number of additional images yielded similar results. These experiments indicate that the subspace matching techniques can be used to eliminate most of the possible positions of the model images in a large unknown image without performing the full correlation 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 Summary

We have considered a subspace method of approximating the Hausdorff fraction between two binary images. The use of edge images rather than grey-level images has yielded robustness to lighting changes and unknown backgrounds and the Hausdorff fraction is robust to clutter and occlusion. The use of subspace matching allows individual matches to be processed much faster than a system that considers the full image space. This combination of techniques thus yields a system with the speed of subspace methods and the robustness of the Hausdorff measure. In addition, we can incorporate these matching techniques into an image search engine. This allows us to perform matching between a library of model images and a large unknown image that may have clutter and occlusion and in which the positions of the model images are unknown.

## Acknowledgments

This work was supported in part by ARPA under ARO contract DAAH04-93-C-0052 and by National Science Foundation PYI grant IRI-9057928.

## References

1. 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.
2. 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.
3. 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.
4. 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.
5. W. J. Rucklidge. Locating objects using the Hausdorff distance. In *Proceedings of the International Conference on Computer Vision*, pages 457–464, 1995.
6. 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.