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

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

Abstract—View-based recognition methods, such as those using eigenspace techniques, have been successful for a number of recognition tasks. Such approaches, however, are somewhat limited in their ability to recognize objects that are partly hidden from view or occur against cluttered backgrounds. In order to address these limitations, we have developed a 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 eigenspace view matching techniques, while also being tolerant of partial occlusion and background clutter. The method applies to binary feature maps, such as intensity edges, rather than directly to intensity images.

Index Terms—Model-based recognition, Hausdorff matching, subspace methods, image matching.

1 Introduction

In this paper, we describe a subspace recognition method (cf. [3], [4], [5], [7], [8]) that handles clutter and partial occlusion by using the generalized Hausdorff measure [1], rather than the sum squared difference (SSD), as the underlying image comparison measure. This is in contrast to other subspace approaches that handle clutter and partial occlusions by decomposing objects into small subregions (e.g., [2]). Our method is based on using an eigen-decomposition to approximate the computation of the generalized Hausdorff measure. This provides the compact storage and fast indexing of eigenspace methods while having the robustness to partial occlusion of the generalized Hausdorff measure.

In the following section, we describe how to approximate the generalized Hausdorff measure using an eigen-decomposition. We then present some results contrasting our approach with a traditional SSD-based subspace matching approach. A key difference with most other subspace methods is that the generalized Hausdorff measure is defined for binary images. Thus, our technique operates on features extracted from images, such as edges or the sign of the Laplacian. Finally, we illustrate how the Hausdorff eigenspace approach can be incorporated into an image search engine such as that in [6] in order to quickly search a large image for instances of any stored model view.

2 Approximating the Hausdorff Fraction

First, we briefly review the generalized Hausdorff measure and, then, introduce a subspace approximation for the 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], [6] as

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

where $f^{\text{th}}_{p \in \mathcal{P}} g(p)$, $0 < f \leq 1$, denotes the fth quantile of $g(p)$ over the set $\mathcal{P}$. For example, the 1th quantile is the maximum and the $\frac{1}{2}$th quantile is the median. Equation (1) generalizes the classical Hausdorff distance, which uses the maximum rather than a quantile.

There are two complementary ways in which the generalized Hausdorff measure has been used for image matching problems. In this paper, we consider the case in which the distance, d, is fixed, and the fraction f measures the quality of the match between the sets $\mathcal{P}$ and $\mathcal{Q}$. In other words, the measure is 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 measure is called the *Hausdorff fraction* because it measures the fraction of points within the given distance, d.

For digital images, the points of the two sets $\mathcal{P}$ and $\mathcal{Q}$ have integer 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 1s of radius d). The Hausdorff fraction can then be expressed as

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

where $\#(S)$ denotes the number of 1s in a binary image S and $\wedge$ denotes the logical and (or the product) of two bitmaps. The result is the fraction of 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 presented here is thus also an approximation to binary correlation.

2.1 The Eigenspace Approximation

Given a collection of images, $I_1, \ldots, I_N$, let x_m be the representation of I_m as a column vector. Consider the matrix $X = [x_1 - c, \ldots, x_N - c]$, where c is the average of the x_ms. X is an $M \times N$ matrix where M is the number of image pixels and N is the number of images in the collection. The eigenvectors of XX^T are an orthogonal basis in terms of which the x_ms can be rewritten (and other, unknown, images as well). Let λ_i, $1 \leq i \leq M$, 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 $[e_1, \ldots, e_M]$ of the eigenvectors. Then, $g_m = E^T(x_m - c)$ is the rewriting of x_m in terms of the orthogonal basis defined by the eigenvectors of XX^T.

The original vector x_m is then simply the weighted sum of the eigenvectors

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

where g_{m_i} is the ith element of the vector g_m. A good approximation (in the least squares sense) to x_m is obtained using only the first k terms in this summation rather than all M terms,

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

The central idea underlying subspace methods is to use this approximation. Let $f_m = (g_{m_1}, \ldots, g_{m_k}, 0, \ldots, 0)$, the first k elements of g_m. Each image I_m (actually, its corresponding column vector x_m) is then represented by f_m, which can be viewed as a point in the space defined by the k eigenvectors $(e_1, \ldots, e_k)$. This can be

- *D.P. Huttenlocher is with the Department of Computer Science, Cornell University, Ithaca, NY 14853. E-mail: dph@cs.cornell.edu.*
- *R.H. Lilien is with Dartmouth University, 6211 Sudikoff Lab, Hanover, NH 03755. E-mail: ryan.h.lilien@dartmouth.edu.*
- *C.F. Olson is with NASA JPL, Mail Stop 125-209, 4800 Oak Grove Dr., Pasadena, CA 91109. E-mail: olson@robotics.jpl.nasa.gov.*

Manuscript received 10 Feb. 1997; revised 3 Feb. 1999.
Recommended for acceptance by K. Ikeuchi.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107709.

TABLE 1
Summary of Results for the Subspace Image Matching Experiments

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)

The first column is for the normalized correlation of the gray-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 a subspace approximation with 76 coefficients.

thought of as projecting the vector g_m into the subspace defined by these k eigenvectors (hence the name subspace method).

Given two binary images I_m and I_n, as above, let x_m be the representation of I_m as a column vector and, further, let x'_n be the representation of I_n^d (throughout, we use primes to denote vectors corresponding to dilated images). The Hausdorff fraction $\Phi(I_m, I_n^d)$ in (2) can then be written as

$$\Phi(I_m, I_n^d) = \frac{x_m^T x'_n}{\|x_m\|^2} \tag{3}$$

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

Given this formulation of the Hausdorff fraction Φ, we now look at how it can be approximated using a subspace approach. As just described above, given some set of images, and the corresponding eigenvectors, E of XX^T, consider the rewriting of x_m and x'_n in the coordinate system defined by these eigenvectors,

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

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).

Now, using the subspace approximation, replacing g_m with f_m and g'_n with f'_n (the first k coefficients) yields

$$\hat{x}_m^T \hat{x}'_n = f_m^T f'_n + x_m^T c + x'^T_n c - \|c\|^2. \tag{4}$$

Note that $\hat{x}_m$ is no longer in general a binary vector. However, $\hat{x}_m^t \hat{x'}^T_n \approx x_m^T x'_n$, i.e., the dot product, is still an approximation of the dot product of the complete binary vectors.

2.2 Using the Subspace Approximation

We now describe the steps for constructing the eigenspace given a set of *binary* model views, $x_1, \ldots, x_N$. First, form the matrix $X = [x_1 - c, \ldots, x_N - c]$, as above, where c is the centroid of the x_ms. Compute and save the first k eigenvectors of XX^T (i.e., those corresponding to the k largest eigenvalues). Note that it is generally more efficient to compute the eigenvectors of $X^T X$, since this matrix is usually considerably smaller (fewer images in the model set than pixels in each image). 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.

For each of the x_ms, compute $f_m = (g_{m_1}, \ldots, g_{m_k})$, where $g_{m_i} = e_i^T (x_m - c)$. Then, compute the scalars $x_m^T c$ and $\|x_m\|^2$. Store the vector f_m and these two scalars for each x_m. This, in addition to the k eigenvectors with the largest eigenvalues, is all the information needed to match the set of model views to unknown images.

An unknown image is processed by dilating it by d, forming the vector x'_n from this dilated image, and computing the $k$vector f'_n and the scalar $x'^T_n c$. Then, the set of model views is compared to the unknown image by computing the subspace approximation to the Hausdorff fraction, Φ, for each x_m and the (dilated) unknown x'_n. Using (3) and (4), this approximation is

$$\hat{\Phi}(x_m, x'_n) = \frac{f_m^T f'_n + x_m^T c + x'^T_n c - \|c\|^2}{\|x_m\|^2}. \tag{5}$$

Note that the scalars $x_m^T c$, $\|x_m\|^2$, and $\|c\|^2$ were precomputed and stored in forming the eigenspace and the scalar $x'^T_n c$ is computed once for an unknown image. Thus, the only term that is computed for each pair of model view and unknown image is $f_m^T f'_n$. Hence, 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 in the following section.

3 Experimental Results

We now consider some simple experiments that illustrate the matching performance of the Hausdorff eigenspace technique. We are particularly interested in comparing this technique with SSD-based eigenspace matching techniques using gray-level images when the background is unknown or when the object is partially occluded. The experiments reported here (summarized in Table 1) use the image set from [4]. We used 30 evenly spaced views of each of 20 objects as the set of model views and 30 different evenly spaced views of the same objects as the test images. Each of the

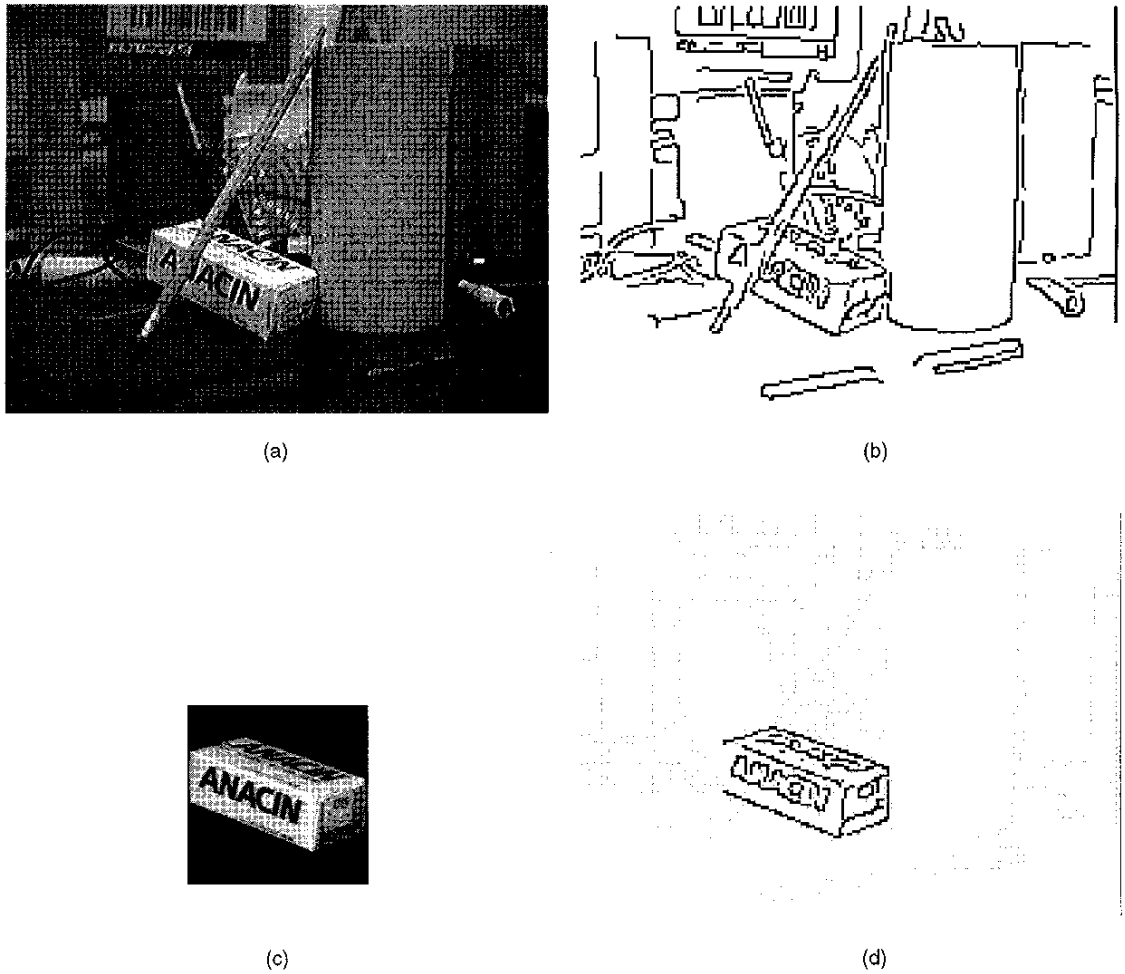

Fig. 1. A cluttered image with some occlusion that was used to test the image search. (a) The original image. (b) The edges detected in the image. (c) The best matching view of the Anacin box. (d) The edges of the Anacin box overlaid on the full edge image at the location of the best match.

1,200 images is split into a foreground (object) and a background (which has intensity value zero).

For each recognition experiment, all 600 test views (not used in constructing the eigenspace) were classified as being one of the 20 objects, based on the closest matching model view in the eigenspace. That is, a trial was considered successful if the best matching model view was from the same object as the test view, regardless of the viewpoint of the test view and the best matching model view. For the gray-level SSD-based matching, we used a standard subspace technique, normalizing the model vectors to magnitude 1 and using the coefficients for the 76 largest eigenvalues. (Note that Murase and Nayar [4] use a more sophisticated 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.) For the binary matching, we computed edge maps and used the technique described above, again with the 76 largest coefficients.

When the original (unperturbed) test views are used, the SSD subspace matching technique yields perfect performance, while the Hausdorff subspace matching technique is successful 96 percent of the time (575 of 600 trials). One model, a Tylenol box, accounted for 16 of these unsuccessful trials, with three other models accounting for the remaining unsuccessful trials. It should be noted that using the true Hausdorff fraction Φ (without the subspace approximation) did not exhibit perfect performance either (it was also successful in 96 percent of the trials). In cases where the true Hausdorff fraction was unsuccessful, it was typically due to a test view that had dense edges. In such cases, a very high fraction of pixels in the model view are near image pixels in the test view and, thus, match very well. This is due to 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.

When comparing uncluttered images, such as those used in this experiment, better results are obtained using the *bidirectional* Hausdorff fraction $\min(\Phi(I_m, I_n^d), \Phi(I_n, I_m^d))$. 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. The subspace approximation of the *bidirectional* Hausdorff fraction was successful 99 percent of the time (593 of 600 trials).

We next considered the case in which the test views were modified so that the background intensity (which was zero in the original images) was changed to a uniform nonzero value. The overall image was still normalized to have unit length for the gray-level matching using the SSD. The edges of the test views were recomputed after the change of background intensities for the binary matching. When the background of the test views was changed changed to 50, the gray-level technique was successful 94

percent of the time (564 of 600 trials). When the background value was changed to 100, the gray-level technique was successful only 41 percent of the time (248 of 600 trials). Thus, the gray-level technique, not surprisingly, is 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 percent success in both cases (567 and 568 successful trials, respectively) despite some changes in the edges due to the substantial changes in background intensity.

For the final experiment, we again used test views with a black background, but where each object was partly occluded. We simulated occlusion of 25 percent of the object by setting the upper, left quarter of the image to a black background in the gray-level images and by erasing the edge pixels in this region for the edge images. In this experiment, the gray-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 gray-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 gray-level SSD and for both the directed and bidirectional Hausdorff fractions. For comparison, results are also given for the approximation to the normalized correlation (simply Hausdorff matching with a dilation d of zero). The main overall result is that the edge-based measures, either Hausdorff or correlation, suffer much less than the gray-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 themselves.

The second overall result seen in Table 1 is that the Hausdorff matching techniques have uniformly good performance, whereas the gray-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).

4 Image Search

In many recognition tasks, the positions of objects that may be present in the image are unknown. Moreover, current segmentation methods cannot generally determine the regions of an image that correspond to separate objects. 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 using the Hausdorff matching methods reported in [1], [6], which have been heavily optimized.

We first consider the simple case 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 for ruling out a location that is 0.05 smaller than that specified by the maximal amount of occlusion allowed (because our experiments indicate that the true fraction is nearly always within 0.05 of the estimated one).

Fig. 1 shows an example of an image where a model (an Anacin box) was partially occluded. We allow for 25 percent mismatch in this case (due to the partial occlusion) and, thus, set the threshold at 0.7, also allowing for an error of 0.05 in the approximate Hausdorff fraction. 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. Over the entire image, 99.3 percent of the locations (translations) of the model have a match with a fraction of less than 0.7. Experiments with images like these indicate that the subspace matching techniques can be used to eliminate nearly all of the possible positions of the model views in a large unknown image without performing full comparisons of model views against the image at these positions. In the rest of this section, we report some experimental results using the Hausdorff subspace approach to perform such pruning.

The subspace approximation to the Hausdorff fraction can be integrated into a multiresolution search strategy to achieve a substantial speedup over a separate search for each model view (as was done in [6]). The basic idea behind multiresolution strategies for Hausdorff matching is to exploit the fact that if there is not a good match at a particular location, then this can be used to eliminate other nearby locations from consideration. When 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, d. 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 d.

We can formulate an efficient search strategy using this observation by considering a hierarchical cell decomposition of the search space. The translations are divided into cells of uniform size (which are recursively divided into similarly uniform cells).

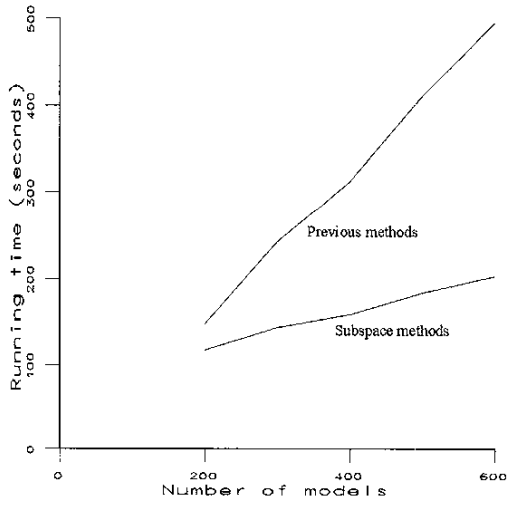

Fig. 2. The time required by the subspace methods grows far less quickly than previous Hausdorff matching techniques as the number of model images in the database grows.

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, d. 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. For more details regarding this cell decomposition, see [6].

As 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. Therefore, we set the threshold for ruling out a cell lower than the actual threshold that we are interested in in order to be reasonably certain that we do not rule out any cells that we should not. We again use an estimate of 0.05 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 compute the true Hausdorff fracdon rather than using the eigenspace approximation (as there will be few such cells that remain and each such cell will only have a small number of possible matching model views).

Fig. 2 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 [1], 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 technique is not as heavily optimized and has additional overhead associated with mapping subimages of the unknown image into the subspace, the time required by this technique grows slowly with the number of objects in the database. As the set of models grows large, the subspace image search method thus outperforms the previous techniques by a considerable margin. From the graph, it can be seen that, for 200 model views, the subspace method already has about a 20 percent speed advantage over the method which considers each model view independently. When the model set reaches 600 views, the speed advantage is about 300 percent.

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 [8], but this method yields very large catalogs of model images. Alternately, one 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. One can then consider cells of this transformation space as above in the multiresolution search strategy. Such an approach to Hausdorff matching is taken in [6], without the use of a subspace approximation. For such larger search spaces, a similar speed up would be observed when using the subspace approximation to the Hausdorff fraction.

ACKNOWLEDGMENTS

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 U.S. National Science Foundation Presidential Young Investigator grant IRI-9057928.

REFERENCES

[1] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, Sept. 1993.

[2] K. Ohba and K. Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen-Windows for Stable Verification of Partially Occluded Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 1,043-1,048, Sept. 1997.

[3] J. Krumm, "Eigenfeatures for Planar Pose Measurement of Partially Occluded Objects," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 55-66, 1996.

[4] H. Murase and S.K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," *Int'l J. Computer Vision*, vol. 14, pp. 5-24, 1995.

[5] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[6] W.J. Rucklidge, "Locating Objects Using the Hausdorff Distance," *Proc. Int'l Conf. Computer Vision*, pp. 457-464, 1995.

[7] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[8] S. Yoshimura and T. Kanade, "Fast Template Matching Based on the Normalized Correlation by Using Multiresolution Eigenimages," *Proc. Int'l Conf. Intelligent Robots and Systems*, vol. 3, pp. 2,086-2,093, 1994.