Technical Report TR-CPS-96-16
Using Discriminant Eigenfeatures
for Image Retrieval

Daniel L. Swets
Augustana College
Computer Science Department
Sioux Falls, South Dakota 57197
swets@inst.augie.edu

John (Juyang) Weng
Michigan State University
Computer Science Department
East Lansing, Michigan 48824
weng@cps.msu.edu

March, 1996

Abstract

This paper describes the automatic selection of features from an image training set using the theories of multi-dimensional linear discriminant analysis and the associated optimal linear projection. We demonstrate the effectiveness of these *Most Discriminating Features* for view-based class retrieval from a large database of widely varying real-world objects presented as "well-framed" views, and compare it with that of the principal component analysis.

Index Terms—Principal component analysis, discriminant analysis, eigenfeature.

1 Introduction

The ability of computers to rapidly and successfully retrieve information from image databases based on the objects contained in the images has a direct impact on the progress of digital library technology [9]. The complexity in the very nature of two-dimensional image data gives rise to a host of problems that alphanumeric information systems were never designed to handle [1]. A central task of these multimedia information systems is the storage, retrieval, and management of images [15]. In many cases, the operator would like to base this retrieval on objects contained in the images of the database. As such, content-based image retrieval is fundamentally an object recognition problem.

The research emphasis to this end has historically been on the design of efficient matching algorithms from a manually designed feature set with hand-crafted shape rules (e.g., [7]). Hand-crafted shape rules can exploit the efficiency found in manually tuning features for a particular training image set. However, these rules have a severe limitation on the type of object classes that can be found by the image retrieval system. Objects greatly different than those for which the system was designed will not be retrieved accurately or efficiently. For example, features tuned to automatically find a human face will probably be useless for retrieving an image of a car.

An alternative to hand-crafting features is the approach in which the machine automatically determines which features to use. The representation of the system is at the signal level instead of at the knowledge (e.g., shape) level. In this type of framework, a training phase finds salient features to use in the subsequent recognition phase of the system. These types of approaches can deal directly with complex, real-world images [14] [20] [21] because the system is general and adaptive.

The efficient selection of good features, however, is an important issue to consider [2]. A well-known problem in pattern recognition is called "the curse of dimensionality"—more features do not necessarily imply a better classification success rate. For example, principal component analysis, also known as the Karhunen-Loève projection and "eigenfeatures," has been used for face recognition [20] and lip reading [3]. An eigenfeature, however, may represent aspects of the imaging process which are unrelated to recognition, such as the illumination direction. An increase or decrease in the number of eigenfeatures that are used does not necessarily lead to an improved success rate. The multivariate linear discriminant

analysis we will use addresses this important issue.

The application being studied in this paper is the query-by-example image retrieval problem. We base this problem on a user-defined labeling scheme, and produce a feature space that is tuned to tessellate the space covered by the samples using as few hyperplanes as possible. This feature space that is produced will not necessarily be good for a different labeling scheme. The labels used for a particular use scenario will determine the behavior patterns of the system we will describe.

2   Optimal Subspace Generation

We use the theories of optimal linear projection to generate a tessellation of a space defined by the training images. This space is generated using two projections: a Karhunen-Loève projection to produce a set of *Most Expressive Features* (MEFs), and a subsequent discriminant analysis projection to produce a set of *Most Discriminating Features* (MDFs).

In this work, as in [16] and [14], we require "well-framed" images as input for training and query-by-example test probes. By well-framed images we mean that only a small variation in the size, position, and orientation of the objects in the images is allowed. The automatic selection of well-framed images is an unsolved problem in general. Techniques have been proposed to produce these types of images, using, for example, pixel-to-pixel search [20], hierarchical coarse-to-fine search [21], or genetic algorithm search [18]. This reliance on well-framed images is a limitation of the work; however, there are application domains where this limitation is not overly intrusive. In image databases, for example, the human operator can pre-process the image data for objects of interest to be stored in the database.

2.1   The Most Expressive Features (MEF)

Each input subimage can be treated as a high dimensional feature vector by concatenating the rows of the subimage together, using each pixel as a single feature. Thus each image is considered as a sample point in this high-dimensional space. Image instances of a particular object can be represented by an n-dimensional random vector $\mathbf{X}$. $\mathbf{X}$ can be expanded exactly by $\mathbf{X} = V\mathbf{Y}$, where the columns of the $n \times n$ square matrix V are orthonormal basis vectors; $\mathbf{Y}$ is a random feature vector of the image $\mathbf{X}$. Without loss of generality, $\mathbf{Y}$ can be considered

as a zero-mean vector, since we could always redefine $\mathbf{Y} - E\mathbf{Y}$ as the new feature vector, and $\mathbf{X} - E\mathbf{X} = V(\mathbf{Y} - E\mathbf{Y})$.

2.1.1 Principal component analysis

This dimension n of $\mathbf{X}$ is usually very large, on the order of several thousand for even small image sizes. Since we expect that a relatively small number of features are sufficient to characterize a class, it is efficient and reasonable to approximate $\mathbf{X}$ using $m < n$ columns of V to give $\hat{\mathbf{X}}(m) = \sum_{i=1}^{m} y_i \mathbf{v}_i$, where the $\mathbf{v}_i$'s are the column vectors of V.

Let the effectiveness of the approximation be defined as the mean-square error $\|\mathbf{X} - \hat{\mathbf{X}}(m)\|^2$. Then we can use the proven result [6] [10] [12] that the best vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_m$ to use are the unit eigenvectors associated with the m largest eigenvalues of the covariance matrix of $\mathbf{X}$, $\Sigma_{\mathbf{X}} = E\left[(\mathbf{X} - \mathbf{M}_{\mathbf{X}})(\mathbf{X} - \mathbf{M}_{\mathbf{X}})^t\right]$, where $\mathbf{M}_{\mathbf{X}}$ is the mean (expected) vector of $\mathbf{X}$. During the training phase, $\Sigma_{\mathbf{X}}$ is replaced by the sample scatter matrix. Then the features $y_1, y_2, \cdots, y_m$ can be easily computed from $y_i = \mathbf{v}_i^t(\mathbf{X} - \mathbf{M}_{\mathbf{X}})$, $i = 1, 2, \cdots, m$. Since these y_i's give the minimum mean-square error, we call them the *Most Expressive Features* (MEF) in that they best express the population in the sense of linear transform as evidenced in reconstruction [11].

This projection, also called the Karhunen-Loève projection and principal component analysis [8], has been used to represent [11] and recognize [20] [16] face images, and for planning the illumination of objects for future recognition tasks [14].

To determine m, the number of features to use, we first rank the eigenvalues of $\Sigma_{\mathbf{X}}$, $\lambda_1, \lambda_2, \cdots, \lambda_n$, in non-increasing order. The residual mean-square error in using $m < n$ features is simply the sum of the eigenvalues not used, $\sum_{i=m+1}^{n} \lambda_i$. So we can choose m such that the sum of these unused eigenvalues is less than some fixed percentage P of the sum of the entire set. So we let m satisfy $(\sum_{i=m+1}^{n} \lambda_i)/(\sum_{i=1}^{n} \lambda_i) < P$. If $P = 5\%$, a good reduction in the number of features is obtained while retaining a large proportion of the variance present in the original feature vector [8] [20], thereby losing very little of their original population-capturing power.

2.1.2 Computational Considerations

We can approximate the covariance matrix $\Sigma_{\mathbf{X}}$ with the sample scatter matrix $S = UU^t$, where $U = [\mathbf{U}_1 \mathbf{U}_2 \cdots \mathbf{U}_k]$, and $\mathbf{U}_i = \mathbf{X}_i - \overline{\mathbf{X}}$, for k training samples. If $k < n$, as is

3

typically the case when dealing with a small number of training samples relative to the image dimension, we can find the eigensystem of the smaller $k \times k$ matrix $U^t U$. This means that $U^t U \mathbf{w}_i = \lambda_i \mathbf{w}_i$, with eigenvalue λ_i and associated eigenvector $\mathbf{w}_i$. Pre-multiplying by U gives $U U^t U \mathbf{w}_i = \lambda_i U \mathbf{w}_i$. Then $\mathbf{v}_i = U \mathbf{w}_i$ is the eigenvector of $S = U U^t$ with eigenvalue λ_i. If the number of samples available is more than the image dimensions, then the eigensystem of $U U^t$ can be computed directly.

2.2   The Most Discriminating Features (MDF)

Although the MEF projection is well-suited to object representation, the features produced are not necessarily good for discriminating among classes defined by the set of samples. The MEFs describe some major variations in the class, such as those due to lighting direction; these variations may well be irrelevant to how the classes are divided. Figure 1 shows an example of a two-dimensional case where the MEF projection cannot separate classes in the population. In this figure, the MEF projection onto the principal component $\mathbf{Y}_1$ is unable
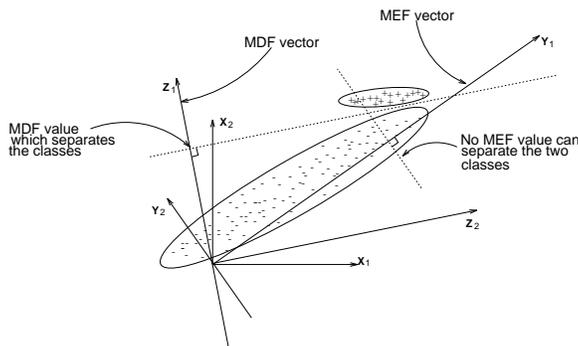

Figure 1. Problems with the MEF projection for class separation.

to separate the two obvious classes. A projection onto $\mathbf{Z}_1$, however, gives a clear separation. This clear separation is provided by the discriminant analysis procedure. The features used to effect this clear separation are called the *Most Discriminating Features* (MDFs) because in a linear projection sense, they optimally discriminate among the classes represented in the training set, in the sense explained below.

2.2.1   Multivariate linear discriminant analysis

Let W be a projection matrix that projects a vector into the MDF subspace. Vector $\mathbf{Z} = W^t \mathbf{Y}$ is a new feature vector from samples of c classes with class means $\mathbf{M}_i$, $i = 1, 2, \cdots, c$.

Then the within-class scatter matrix is defined as [8] $S_w = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (\mathbf{Y}_j - \mathbf{M}_i)(\mathbf{Y}_j - \mathbf{M}_i)^t$ for n_i samples from class i.

For a grand mean vector $\mathbf{M}$ for all samples from all classes, the between-class scatter matrix is defined as $S_b = \sum_{i=1}^{c} (\mathbf{M}_i - \mathbf{M})(\mathbf{M}_i - \mathbf{M})^t$.

In discriminant analysis, we want to determine the projection matrix W that maximizes the ratio $\frac{\det\{S_b\}}{\det\{S_w\}}$. In other words, we want to maximize the between-class scatter while minimizing the within-class scatter.

It has been proven [5] [22] that this ratio is maximized when the column vectors of projection matrix W are the eigenvectors of $S_w^{-1} S_b$ associated with the largest eigenvalues. Then the scalar components in $\mathbf{Z}$ are feature values of the given samples and the column vectors of W are the MDF feature vectors.

2.2.2 Computational Considerations

Because the matrix $S_w^{-1} S_b$ need not be symmetric, the eigensystem computation could be unstable. To avoid this problem, the following method diagonalizes two symmetric matrices, and produces a stable eigensystem computation procedure.

Compute H and Λ such that $S_w = H\Lambda H^t$, where H is orthogonal and Λ is diagonal. Then $\left(H\Lambda^{-\frac{1}{2}}\right)^t S_w H\Lambda^{-\frac{1}{2}} = I$. Now compute U and Σ such that $\left(H\Lambda^{-\frac{1}{2}}\right)^t S_b H\Lambda^{-\frac{1}{2}} = U\Sigma U^t$ where U is orthogonal and Σ is diagonal. Then

$$S_b = H\Lambda^{\frac{1}{2}} U\Sigma U^t \Lambda^{\frac{1}{2}} H^t \tag{1}$$

$$S_w = H\Lambda^{\frac{1}{2}} U I U^t \Lambda^{\frac{1}{2}} H^t, \tag{2}$$

Defining $\nabla = H\Lambda^{-\frac{1}{2}} U$, ∇ diagonalizes S_b and S_w at the same time. Since

$$S_w^{-1} = H\Lambda^{-1} H^t, \tag{3}$$

Equations 1 and 3 give

$$\begin{aligned} S_w^{-1} S_b &= H\Lambda^{-1} H^t H\Lambda^{\frac{1}{2}} U\Sigma U^t \Lambda^{\frac{1}{2}} H^t \\ &= H\Lambda^{-\frac{1}{2}} U\Sigma U^t \Lambda^{\frac{1}{2}} H^t \\ &= \nabla\Sigma\nabla^{-1} \end{aligned}$$

That is, ∇ consists of the eigenvectors of $S_w^{-1} S_b$ and Σ contains the eigenvalues of $S_w^{-1} S_b$. Thus using the symmetric properties of the component scatter matrices, we have a method for finding the eigensystem of $S_w^{-1} S_b$ that is stable.

2.2.3 The *DKL* projection

The discriminant analysis procedure breaks down, however, when the within-class scatter matrix S_w becomes degenerate. This can happen when the number of samples is smaller than the dimension of the sample vectors. Using discriminant analysis directly on the images will generally render S_w non-invertible due to the high dimension of the input vector relative to the number of training samples. For example, a very small image of 64×64 when vectorized turns into a 4096-dimensional sample vector. The number of training samples available is usually much smaller than this large dimension.

Our resolution to this problem is to perform two projections instead of one. The discriminant analysis projection is performed in the space of the Karhunen-Loève projection (i.e., MEF space), where the degeneracy does not occur.

We first project the n-dimensional image space onto the m-dimensional MEF space. We choose m such that for s training samples from c classes, $m + c \leq s$. In fact, it is impossible for $m > s - 1$ since there are a maximum of $s - 1$ non-zero eigenvalues in the Karhunen-Loève projection. But we further constrain the final dimension of the MEF space to be less than the rank of S_w in order to make S_w non-degenerate. On the other hand, m cannot be smaller than the number of classes, c.

Since there are at most $c - 1$ non-zero eigenvalues of $S_w^{-1} S_b$, we choose $k \leq c - 1$ to be the final dimension of the MDF space. So we relate the dimensions of the MEF- and the MDF- spaces as $k + 1 \leq c \leq m \leq s - c$.

Thus, the new overall discriminant projection is decomposed into two projections, the Karhunen-Loève projection followed by the discriminant projection. We call this new projection the Discriminant Karhunen-Loève projection (DKL projection).

Definition 1 (DKL projection) *The DKL projection to the Most Discriminating Feature (MDF) space is $\mathbf{Z} = W^t V^t \mathbf{X}$, where V is the projection matrix from the image space to the MEF space, and W is the projection matrix from the MEF space to the MDF space.*

2.2.4 Explanation of the MDFs

The Most Discriminating Features are not directly tied to the absolute intensity values of the input images. Figure 2 shows a set of MEF and MDF features obtained from a large training set of human faces. As can be seen from the figure, the features encapsulated in

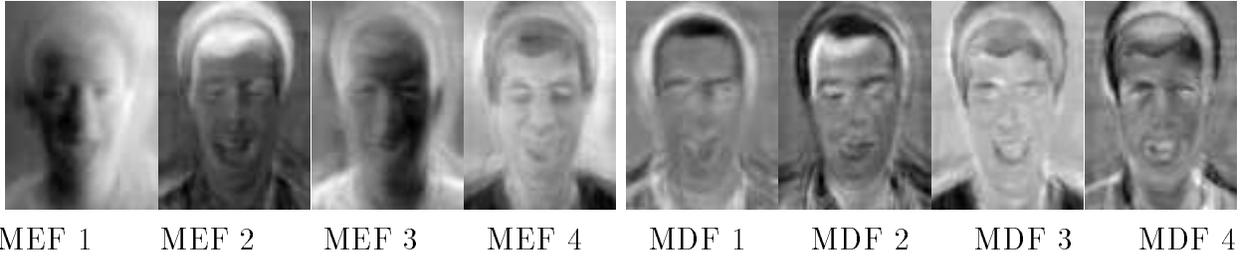MEF 1   MEF 2   MEF 3   MEF 4   MDF 1   MDF 2   MDF 3   MDF 4

Figure 2. A sample of MEF and MDF vectors treated as images. The MEF vectors show the tendency of the principal components to capture major variations in the training set, such as lighting direction. The MDF vectors show the ability of the MDFs to discount those factors unrelated to classification. The training images used to produce these vectors are courtesy of the Weizmann Institute.

the MDF vectors show directional edges found in the training set and have discarded the imaging artifacts, such as lighting direction, to a large extent.

2.2.5 The Clustering Effect of the MDF subspace using the DKL projection

To show how the MDF subspace effectively discounts factors unrelated to classification, an experiment was performed to obtain the MEF and the MDF vectors for a collection of training images. Figure 3 shows samples of the data in the space of the best (in terms of the largest eigenvalues) two MEFs and MDFs for the experiment. From Figure 3, it is clear that

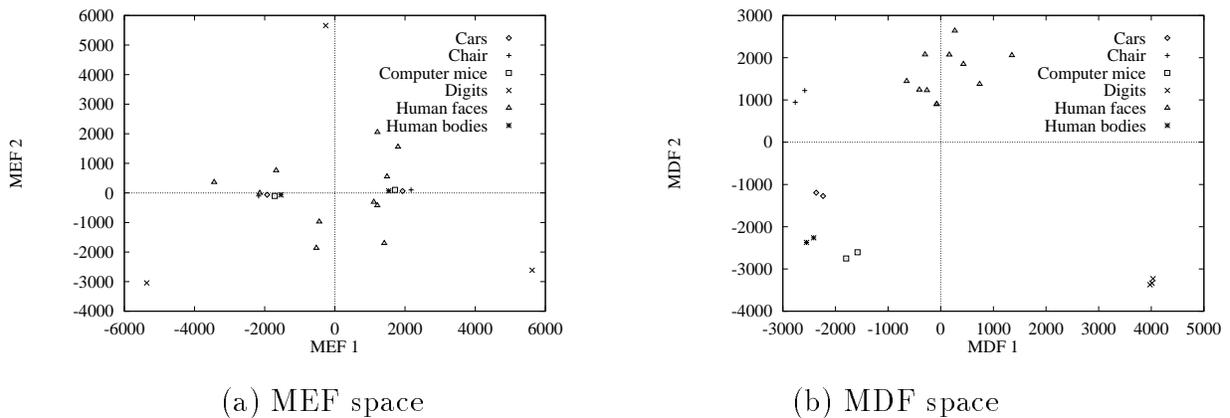(a) MEF space   (b) MDF space

Figure 3. Distribution of some samples using the best two features in the MEF and the MDF spaces. In the MDF subspace, objects of the same class are clustered much more tightly than in the MEF space.

the MDF subspace has a significantly higher capability than the MEF subspace to correctly classify an input image that is considerably different from those used in training, because

classes in the MDF subspace have larger between-class distance and smaller within-class scatter.

Though the human operator may assign a semantic "closeness" measure to particular sets of classes (e.g., a bus is closer to a car than a face), the linear discriminant analysis procedure does not take this into account. The distance between clusters of points in Figure 3 is not intended to proportionally carry any meaning for semantic closeness of classes. The linear discriminant analysis procedure merely tries to separate each class from the others as well as possible.

2.3   Image Matching

The set of Most Expressive Features and Most Discriminating Features are generated for each image in the training set and stored in the recognition module. When an image query is presented to the recognition module, it is projected to these same subspaces. A simple Euclidean distance in this feature space is computed to find a the set of k nearest neighbors for retrieval. It has been shown that the probability of error for this nearest neighbor decision rule is bounded above by twice the Bayes probability of error [4] if we have an infinite number of samples. This simple measure of similarity is used because it does not require estimation of the distribution function, which is impractical for our high dimensional space.

3   Results

In this section, we demonstrate the ability of the MDF space to tolerate within-class variations and to discount such imaging artifacts as lighting direction.

3.1   A Comparison Between the MEF and the MDF feature spaces

To demonstrate the superiority of the MDF space over the MEF space, we designed an experiment to compare the performance of the system using the MEF space alone versus using the MDF space. The number of features from each subspace was varied to show how the system performed. Figure 4 shows the correct recognition rates for the system using these two subspaces. The images for this experiment came from the Weizmann Institute and contained well-framed faces with different expressions taken under different lighting
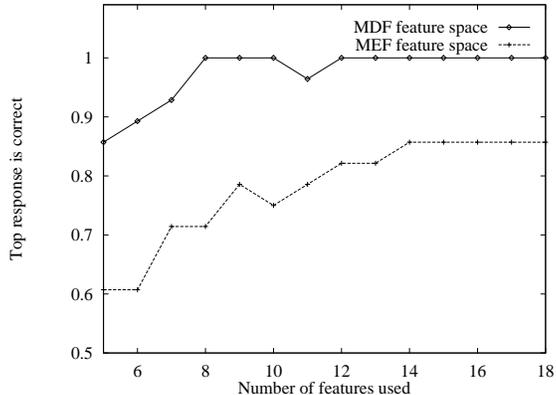
Figure 4. The performance of the system for different numbers of MEF and MDF features, respectively. The number of features from the subspace used was varied to show how the MDF subspace outperforms the MEF subspace. 95% of the variance for the MDF subspace was attained when 15 features were used; 95% of variance for the MEF subspace did not occur until 37 features were used. Using 95% of the MEF variance resulted in an 89% recognition rate, and that rate was not improved using more features.

conditions. Each individual in the set of images used had two expressions; each expression image was taken with three different lighting conditions. An example of the pool of available images for a sample individual is given in Figure 5. This set of images seemed particularly

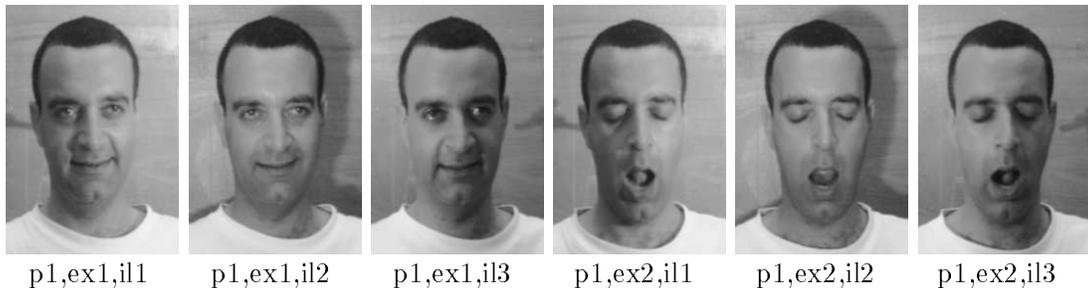p1,ex1,il1　　p1,ex1,il2　　p1,ex1,il3　　p1,ex2,il1　　p1,ex2,il2　　p1,ex2,il3

Figure 5. A sample of the Weizmann Institute face data. The frontal images for an individual contain two expressions; for each expression, a set of three images with differing lighting conditions forms the set of images available for an individual.

suited to testing the ability of the MDF space to discount factors unrelated to classification residing in labeled image samples (such as lighting direction); thus it outperforms MEF space in recognition. The MEF space has a distance metric that attempts to preserve the Euclidean distance metric. This metric is not tuned to any particular labeling scheme. The MDF space is tuned to the labeling scheme given by the user. Of course, this labeling scheme may not be good for a different labeling scheme. This experiment showed an improved performance with the MDF space, due in large part to the fact that for the labeling scheme chosen for this

9

experiment, we had enough training images with sufficient within-class variation. With fewer training images, or training images that do not sufficiently capture the desired variations, the performance difference between the MEF and the MDF spaces will be smaller. For this experiment, a disjoint test set was utilized. This test set was formed by randomly choosing one image from the set of images available for each individual. Therefore, each individual was trained with all of the different lighting conditions except for the expression image in the test set.

3.2 Combination database: Faces and Other Objects

To show the general applicability of the method, we have trained the system on a diverse set of objects from natural scenes, ranging from human faces to street signs to aerial photographs. A list of some examples from the various classes learned is given in Figure 6. The classes

isign   monitor   phone   renault   seated   sharpener   sidewalk   dbody
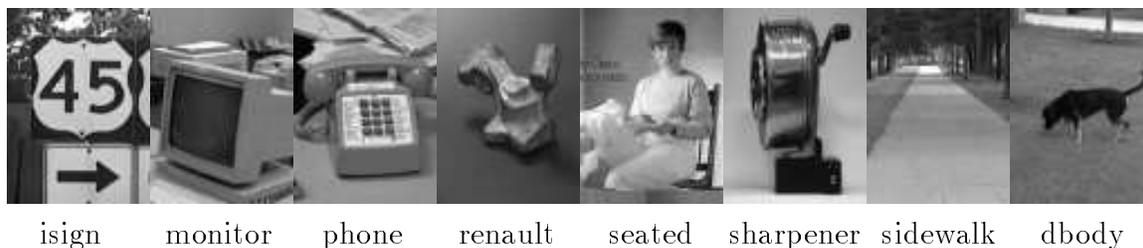
Figure 6. Representatives of training images from some of the various classes learned. Objects of interest are at the center of the fovea images. In the learning phase for this experiment, the training images were generated using manual extraction of the areas of interest.

were established by labeling the images with the name of the object in the image (or in the case of faces, naming the individual in the image). The reason for this choice of labeling scheme is because the main usage of the database was to retrieve images classified by object name. Labeling is application dependent—no single labeling scheme can possibly fit all applications. Given a labeling scheme designed by the user, the system automatically finds the best subspace to provide the capability indicated by the labels given to the training images. Each stored image can maintain pointers to a relational database to provide retrievals under various other desired organizations [17] such as gender, age group, *etc*.

For this experiment, the database consisted of predominantly pairs of images to describe each class. Most classes in the database were represented by two images, and 19% of the classes had three or more images, up to twelve for some objects. Each image consisted of

a well-framed object of interest. The different images from each class were taken either in a different setting or from a slightly different angle; where possible a change in the lighting arrangement was used to provide variation in the training images.

Following training, the system was tested using a disjoint test set. A summary of the makeup of the test and the results are shown in Table 1. The instances where the retrieval

Number of samples	Number of classes
2	504
3	89
4	7
13	1

(a) Makeup of training set

Training Images	1316 images from 504 classes
Test images	298 from 298 classes
Top choice correct	90%
Correct in top 15	98%

(b) Results of test

Table 1. Summary of large natural scene experiment. The training images were drawn at random from the pool of available images, with the remaining images serving as a disjoint set of test images. (a) The training set contained predominantly classes with two training images in each class. This table shows the number of classes that contain the corresponding number of training images. (b) A list of 298 test images from the disjoint test set were given to the system to find a closest match.

failed were due in large part to significant differences in object shape and three-dimensional (3D) rotation. Figure 7 shows an example of a face image which the system failed to properly retrieve. This search probe failed because the training data for this class did not include

(a) Search probe   (b) Training images

Figure 7. Example of a failed search probe. The retrieval failed to select the appropriate class due to a lack of 3D rotation in the set of training images.

any 3D object rotation and the search probe did. The recognition results for this system rely heavily on the assumption that the training images are representative of image class variation which will be seen during the recognition phase.

3.3 The power for handling within-class variation using the MDF subspace

The capability of the system for handling large within-class variation is demonstrated in Figure 8. Each of these search probes retrieved samples from the correct class defined by the

(a) List of training images
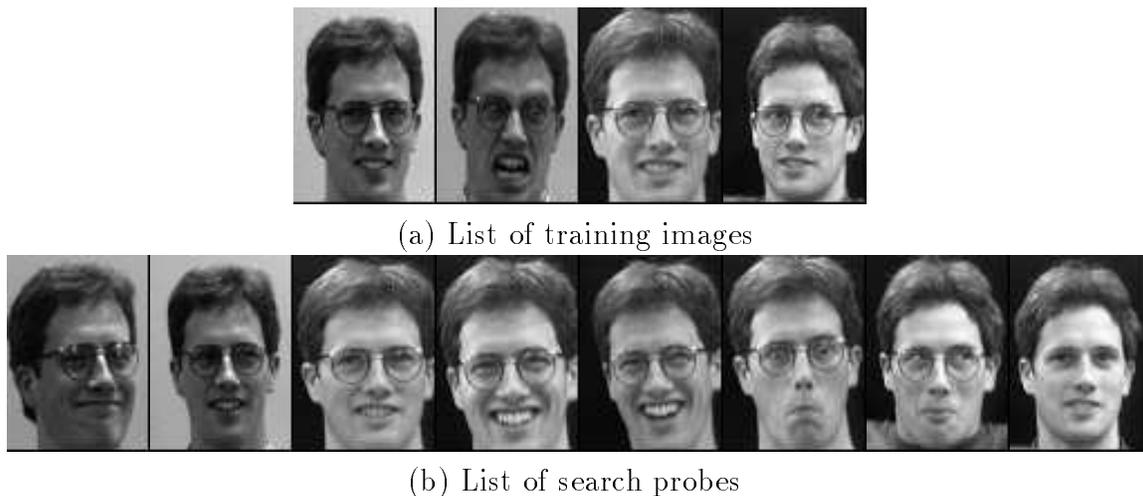

(b) List of search probes

Figure 8. Example of how well within-class variation is handled. The system correctly retrieved images from the class defined by the training samples for each of the search probes.

training exemplars. This example supports the claim that those features unimportant for subclass selection are weighted down appropriately or discarded by the Most Discriminating Feature selection process. For example, the change in the shape of the mouth among the different expressions is unimportant for determining to which subclass these images belong. As long as the variation in a class is sufficiently present in the training set for MDF feature computation, the approach performs well.

4 Conclusions and Future Work

The Most Discriminating Features described in this paper provides an effective feature space to be used for classification. This MDF space discounts factors unrelated to classification, such as lighting direction and facial expression when such variations are present in the training data. Respectable recognition results were obtained for a large database of images.

In the experiments described in this paper, a comparison must be made between a test probe and every image in the database. The average time required for a query image to

obtain a set of matches in this manner was 400.7 seconds (including projection) running on a Sun SPARC 20. When this Most Discriminating Feature subspace computation is placed into a hierarchy and the resulting spaces decomposed into a hierarchical Voronoi tessellation as described in [19], the average time required for a test probe fell to 9.1 seconds; at the same time, the recognition rate rose to 95% for an image from the correct class being retrieved as the top choice and 99% for the correct class being in the top 10 retrieved images.

The work reported in this paper only investigates the use of intensity images as input to the system. In order to make our system nearly insensitive to lighting conditions, it may also be valuable to use edge images as well as intensity images. It is desirable to investigate the utility of intensity in combination with edge map images [13] in the Most Discriminating Features space.

References

[1] J. R. Bach, S. Paul, and R. Jain, "A visual information management system for the interactive retrieval of faces," *IEEE Transactions of Knowledge and Data Engineering*, vol. 5, no. 4, p. 619ff, 1993.

[2] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proc. International Conference on Computer Vision*, pp. 500–507, 1995.

[3] C. Bregler and S. M. Omohundro, "Nonlinear manifold learning for visual speech recognition," in *Proc. International Conference on Computer Vision*, pp. 494–499, 1995.

[4] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions of Information Theory*, vol. IT-13, pp. 21–27, January 1967.

[5] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, vol. 8, pp. 376–386, 1938.

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 2nd ed., 1990.

[7] K. Ikeuche and T. Kanade, "Automatic generation of object recognition programs," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 1016–1035, 1988.

[8] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice-Hall, New Jersey, 1988.

[9] R. Jain and A. Hampapur, "Metadata in video databases," in *Sigmod Record: Special Issue on Metadata for Digital Media*, vol. 23, p. 27ff, ACM: SIGMOD, December 1994.

[10] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[11] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, January 1990.

[12] M. M. Loève, *Probability Theory.* Princeton, NJ: Van Nostrand, 1955.

[13] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object detection," in *Proc. International Conference on Computer Vision*, pp. 786–793, 1995.

[14] H. Murase and S. K. Nayar, "Illumination planning for object recognition in structured environments," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 31–38, June 1994. Seattle, Washington.

[15] E. Oomoto and K. Tanaka, "OVID: Design and implementation of a video-object database system," *IEEE Trans. Knowledge and Data Engineering*, vol. 5, p. 629ff, August 1993.

[16] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 84–91, June 1994. Seattle, Washington.

[17] D. L. Swets, Y. Pathak, and J. J. Weng, "A system for combining traditional alphanumeric queries with content-based queries by example in image databases," Tech. Rep. CPS-96-03, Michigan State University, January 1996.

[18] D. L. Swets, B. Punch, and J. J. Weng, "Genetic algorithms for object recognition in a complex scene," in *Proceedings, International Conference on Image Processing*, (Washington, D.C.), pp. 595–598, October 1995.

[19] D. L. Swets and J. J. Weng, "Efficient image retrieval using a network with complex neurons," in *Proceedings, International Conference on Neural Networks*, (Perth, Western Australia), November 1995. Invited Paper.

[20] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[21] J. Weng, N. Ahuja, and T. S. Huang, "Learning recognition and segmentation using the Cresceptron," in *Proc. International Conference on Computer Vision*, pp. 121–128, May 1993. Berlin, Germany.

[22] S. S. Wilks, *Mathematical Statistics.* Wiley, New York, 1963.

Affiliation of Authors

Daniel L. Swets is with the computer science department at Augustana College, 2001 South Summit, Sioux Falls, SD 57197, email swets@inst.augie.edu.

John (Juyang) Weng is with the computer science department at Michigan State University, East Lansing, Michigan 48824, email weng@cps.msu.edu.