Region-Based Representations for Face Recognition

BENJAMIN J. BALAS and PAWAN SINHA
Massachusetts Institute of Technology

Face recognition is one of the most important applied aspects of visual perception. To create an automated face-recognition system, the fundamental challenge is that of finding useful features. In this paper, we suggest a new class of image features that may be a useful addition to the set of representational tools for face-recognition tasks. Our proposal is motivated by the observation that rather than relying exclusively on traditional edge-based image representations, it may be useful to also employ region-based strategies that can compare noncontiguous image regions. The spatial homogeneity within regions allows for enhanced tolerance to geometric distortions and greater freedom in the choice of sample points. We first show that under certain circumstances, comparisons between spatially disjoint image regions are, on average, more valuable for recognition than features that measure local contrast. Second, we learn "optimal" sets of region comparisons for recognizing faces across varying pose and illumination. We propose a representational primitive—the dissociated dipole—that permits an integration of edge-based and region-based representations. This primitive is then evaluated using the FERET database of face images and then compared to established local and global algorithms.

Categories and Subject Descriptors: I.5.4 [**Pattern Recognition**]: Applications

General Terms: Design

Additional Key Words and Phrases: Face recognition

1. INTRODUCTION

The core challenge in constructing any computational model of object recognition lies in determining what representational primitive to use as a description of input images. Raw pixels are the simplest option, but are subject to wide variation when objects undergo very simple transformations (such as, changes in illumination and viewing angle), making them a poor choice of primitive features to subserve object recognition. A better choice of features should not suffer as greatly from extrinsic sources of variation and should confer some amount of stability and useful diagnosticity to the "back end" classification procedure, while remaining relatively simple to construct from raw input. For the past few decades, edge-based representations have been considered promising from this perspective.

1.1 Edge-Based Representations of Objects

Many computational models of recognition make use of edge-finding operators to represent images for recognition. These operators are usually modeled either as wavelets or derivative-of-gaussian functions.

Authors address: Benjamin J. Balas and Pawan Sinha, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[Young et al. 2001; Young and Lesperancy 2001; Daugman 1985]. They possess many useful qualities that make them attractive primitive features. First, measurements of image contrast are inherently more stable to changes in overall luminance (caused, for instance, by global illumination variations) than raw luminance values. Second, these operators are also useful in that they yield a "sparse code" for natural images.

Natural images are known to be highly spatially redundant. That is, real-world images do not frequently exhibit rapid changes in luminance. Edges are relatively infrequent, meaning that one can usually predict the luminance of any given image pixel by looking at its immediate neighbors [Attneave 1954; Kersten 1987]. One of the advantages of this redundancy is that it is possible to construct efficient coding schemes for natural scenes. Oriented waveletlike features have been shown to be an optimal tool for constructing just such a sparse code [Bell and Sejnowski 1997; Olshausen 1996; Olshausen and Field 1997]. These operators qualitatively resemble the receptive fields of neurons in the early primate visual system as well, suggesting that the "goal" of V1 may be to represent the visual world in terms of an efficient code [Field 1994].

Recognition systems based on edge representations are often quite successful and have the added benefit of being biologically motivated, to some degree. Edge-based models of human performance also appear to provide a good account of the recognition of generic objects [Biederman 1987; Biederman and Ju, 1988] and have been shown to support highly accurate face-recognition algorithms, as well [Lades et al. 1993; Wiskott et al. 1997]. However, there are some important reasons to think that it may be useful to consider alternate representational tools as well, especially in the domain of faces.

First, it is well-known that images of famous individuals are quite poorly recognized if the contours and edges in that face are all that a subject is given as input [Davies et al. 1978; Leder 1996; 1999]. If edges and contours were really the primary representational tool for recognition, images that isolate those features should be easily recognizable. The fact that such images are difficult to recognize indicates that edge and contour features are not sufficient for recognition. Second, we also point out that faces are quite easy to recognize when they are extensively blurred [Sinha 2002a; Yip and Sinha 2002]. Blurring removes high spatial frequency content, namely edges and contours. In this case, if edgelike features were vital to the recognition of faces, their removal should substantially impair subjects' ability to accurately recognize the stimuli. Given that this does not occur, we can infer that fine edge information is not a critical prerequisite for face recognition.

Given the limitations of pixel and edge-based primitives, we propose the use of novel image measurements that implement a region-based representational strategy. It is important to point out that our proposal relates to the basic vocabulary of image measurements, rather than to more global representations that can be constructed via procedures, such as, principal components analysis (PCA) [Moghaddam et al. 2000; Turk and Pentland 1991; Belhumer et al. 1997] or independent components analysis (ICA) [Bartlett et al. 1998]. These techniques are agnostic about what the numbers that they operate on correspond to. Typically, these methods operate directly on image pixels, but there is no reason they could not operate on edge or region primitives. Global techniques are useful tools to employ given any initial collection of image features, be they pixel values, "edgels" or another class of measurements, such as the one we propose below. Our interest in this work is primarily to identify useful basic measurements, which can be used as the foundation for any classification scheme.

1.2 Region-Based Representations of Objects

As reviewed above, edge-based features have received much attention in computational vision. However, they constitute only one-half of a representational dichotomy. The "dual" of a boundary (or edge)-based representation is a region-based one. Past work has suggested that early stages of the visual system organize the visual input in terms of surfaces [Nakayama et al. 1995). Furthermore, work in computer vision has shown that for the task of image parsing, region-growing techniques are generally better in noisy images where edges are extremely difficult to detect [Malik et al. 1999]. This is a direct consequence of the fact that region-based representations benefit from the spatial redundancy found in natural images. In a separate paper [Balas and Sinha, in press], we have argued for the use of a representational primitive that provides a convenient way for spanning both boundary and region-based information. The development of this primitive was motivated by the finding that receptive fields with disjoint excitatory and inhibitory lobes emerge from a recognition, rather than a high-fidelity reconstruction, criterion.

In this paper, we apply this primitive to two different face-recognition tasks in an effort to understand what features are most useful for recognizing faces under different circumstances. We demonstrate that representing faces in terms of the luminance differences between spatially disjoint regions can outdo the average performance of purely edge-based or purely global features under specific circumstances (such as when viewing angle of a face varies), but that in other situations (such as when illumination varies), localized features perform better than both nonlocal and global features. We subsequently explore strategies for determining the most relevant region comparisons (which may approximate edgelike features or be highly nonlocal) to use when representing faces for recognition.

Having noted the emergence of a novel operator in our previous experiments, our goal presenting this paper is twofold. First, we wish to determine how the utility of this operator (and others) changes as a function of the problem domain. Second, we attempt to translate our previous work with nonlocal operators into the development of an automated face recognition that is based on generic region comparisons. This allows us to explicitly compare the utility of nonlocal operators to other recognition techniques in the proper context. We present region-based representation as an overall strategy for face encoding that subsumes edge-based representations and introduces novel features (such as our nonlocal operator), as well. Although we focus here on faces, we believe that this representational strategy can apply across other object domains as well.

This paper is organized as follows. In the next section, we provide an overview of the basic representational primitive, called a “dissociated dipole,” for region-based encoding. The dipole model we present here was developed previously as a means of adding nonlocality to a simple difference-of-gaussians (DOG) framework. The presentation of the operator is followed by three experiments, all of which are designed to evaluate the effectiveness of this representation scheme compared to strictly local representations, as well as global pixel-based techniques. The first experiment assesses how much identity information arbitrary nonlocal region comparisons contain compared to arbitrary local comparisons. Developing a region-based representation for faces is only worth doing if we can show that there is some useful information beyond local (edge-based) region comparisons. Performance with a global template-based algorithm is also assessed so that we may compare both differential operators to procedures that use raw pixel values. In the second experiment, we explore the utility of generic region comparisons at a finer grain. We determine optimal region comparisons for different face-recognition tasks (tolerance to pose and illumination variations) and assess the contributions of local and nonlocal contrast encoding to each. Finally, in our third experiment, we provide baseline results for a region-based recognition system as tested on a standard database of faces and compare these results to established local and global algorithms.

2. THE REPRESENTATIONAL PRIMITIVE

Our proposed representational primitive is a simple difference operator since, as stated previously, measurements of image contrast are useful for providing a measure of invariance to global illumination changes. What distinguishes this primitive from the previously proposed ones is that the differences are computed across potentially nonlocal regions. If we only wish to execute comparisons between

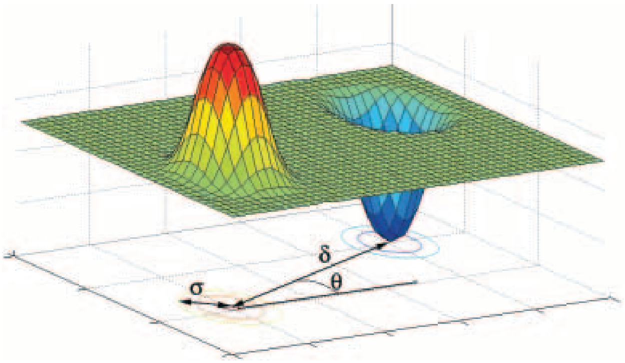

Fig. 1. A schematic of our dipole operator. The most important aspect of this operator is the decoupling of the distance between lobes (δ) from the width of the lobes (σ).

neighboring image regions, we would be able to use sufficiently large edgelike DOG features. However, since we do not know *a priori* that useful information will be concentrated in such comparisons, we must allow for the possibility that comparisons between distant image regions might be desirable as well.

To this end, we define simple dipole features comprising a positive and negative lobe that are capable of comparing image regions separated by an arbitrary distance. A generic bilobed operator of this type can be modeled as follows:

$$\frac{1}{\sqrt{2\pi}\,|\Sigma_1|^{1/2}} e^{\frac{-(x-\mu_1)^t \Sigma_1^{-1}(x-\mu_1)}{2}} - \frac{1}{\sqrt{2\pi}\,|\Sigma_2|^{1/2}} e^{\frac{-(x-\mu_2)^t \Sigma_2^{-1}(x-\mu_2)}{2}} \qquad (1)$$

For our purposes, we shall only consider operators with diagonal covariance matrices Σ_1 and Σ_2. Further, the diagonal elements of each matrix Σ shall be equal, yielding isotropic Gaussian lobes. For this simplified case, the above equation can be expressed as follows:

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} - \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-(x-\mu_2)^2}{2\sigma_2^2}} \qquad (2)$$

We also introduce a parameter δ to represent the separation between two lobes. This is simply the Euclidean norm of the difference between the two means.

$$\delta = \|\mu_2 - \mu_1\| \qquad (3)$$

Figure 1 schematically shows a prototypical dipole operator. These dipole operators allow us to explore a wide range of feature subtypes. For example, to create a local-oriented operator, we shall set $\sigma 1 = \sigma 2$, and set the distance δ equal to 3σ. Nonadjacent comparisons can be carried out by allowing the distance δ to exceed 3σ (once again assuming equal spatial constants for the two lobes). Importantly, δ and σ are decoupled in this model. This allows us to vary these parameters separately, such that with this set of simple features in hand, we can assess how recognition performance changes as a function of both spatial scale (σ) and the locality/nonlocality of the operators (δ).

3. EXPERIMENT 1

In our first experiment, we evaluate the hypothesis that long-range luminance comparisons between image regions contain information diagnostic of identity. It is important for us to determine if this is the case, as our decoupling of σ and δ is otherwise unnecessary. We examine the utility of nonlocal comparisons in two ways.

First, we shall assess the recognition performance of adjacent and nonadjacent comparisons as a function of spatial scale. Specifically, given a particular value of σ, we wish to determine the effectiveness of nonlocal comparisons (in which $\delta \gg 3$) compared to strictly local measurements (in which $\delta \approx 3$). We shall also compare both of these feature sets to a representation that uses global templates for processing raw pixel values.

Second, we shall examine the performance of both feature sets under an "ordinal encoding" scheme in which we only retain the direction of contrast for each measurement. A large proportion of neurons in V1 tend to exhibit such nonlinear response patterns and ordinal encoding provides a simple means of modeling that behavior. The question here is whether local or nonlocal comparisons prove more robust to the imposition of a contrast threshold. We shall conduct both of these analyses on two distinct face databases, which contain faces that vary in pose and illumination, respectively. The performance of a global pixel-based system will not be considered here as it is difficult to meaningfully define the impact of a contrast threshold on the encoding of absolute pixel values.

3.1 Stimuli

We use faces drawn from two databases for this experiment. The first of these is the ORL database [Samaria and Harter 1994]. The images in this database are all grayscale, 112 × 92 pixels in size, and there are ten unique images of each of the 40 individuals. Faces vary primarily in pose, but there is some variation of position and expression as well. We do not explicitly normalize face position, because we are interested in determining how robust our proposed features are to both changing viewing angles and translation. We divided this database into training and test sets. The training set was composed of 200 faces (five randomly chosen instances per individual) and the test set was composed of the remaining 200 faces.

Our second set of faces is the Harvard face database, which contains a total of 600 96 × 84 grayscale images. This database contains images of each individual under a range of lighting conditions, including very extreme lighting angles. In this database, face position is explicitly normalized allowing us to examine the effects of varying illumination in isolation. There are 10 unique individuals in this database, with 60 images of each individual. Our training set of images comprised six images per individual. In order to be able to better assess the generalization abilities of the representations, we chose the training images to be those with the least extreme lighting angles. The test set comprised the remaining 540 images, including those with extremely eccentric lighting directions. As with the ORL database, no preprocessing was carried out (Figure 2).

3.2 Procedure

Our first aim is to determine the recognition performance that can be achieved with our general DOG features when δ exceeds 3σ. To this end, we first construct DOG operator banks that enact local comparisons ($\delta = 3\sigma$). 10 banks of 50 operators each were created at each of four spatial scales ($\sigma = 1$, 2, 4, or 8 pixels). Ten such features are displayed in the left panel of Figure 3.

To create a set of nonadjacent features, we use the same Gaussian subunits that we selected for these adjacent comparisons, but "rewire" the connections between subunits at random so that some features will be composed of widely separated image regions. The advantage to this procedure (relative to one where an arbitrary set of nonlocal operators is used) is that exactly the same image pixels are being used in both the adjacent and nonadjacent cases. The disadvantage is that we are unable to parametrically vary δ. Instead, we are left with a distribution of distances between region pairs. However, this is sufficient to evaluate whether good recognition performance can be achieved at each

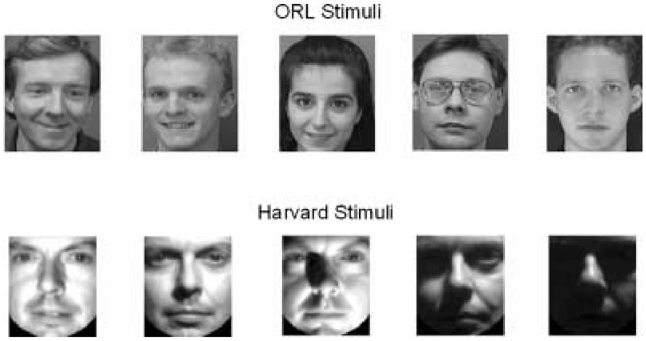

Fig. 2. Examples of faces drawn from the ORL database (top row) and the Harvard database (bottom row). We note the extreme variance of lighting conditions in the Harvard database. Each image in the bottom row depicts the same individual seen under increasingly severe lighting angles. The leftmost image in the bottom row is a typical training image, while the rest are representative of test images.

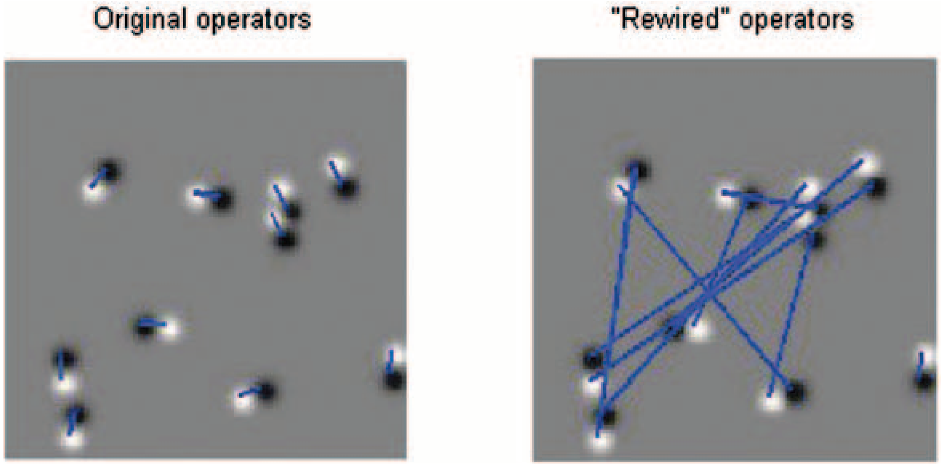

Fig. 3. A set of features for comparing adjacent image regions (left) and the "rewired" nonadjacent features (right). Note that the same operator subunits are used in both cases, but the connections between subunits have been altered to enact nonadjacent comparisons. In this manner, the same image pixels are under consideration in both feature sets, but the comparisons between them are being altered.

scale as δ increases. All we require is that the rewiring procedure produces values of δ that are typically larger than 3σ.

To determine if this is, indeed, the case, we created 100 sets of rewired operator banks (50 operators per bank) and measured the average separation between the subunits of each operator. The distance between image regions was significantly greater than 3σ at all spatial scales after rewiring (one-tailed t-test, $p < 0.0001$). At large values of σ, the average postwiring value of δ does not exceed this threshold by much, but the relationship holds across all the data points we recorded here.

Our second hypothesis concerns the effectiveness of local and nonlocal region comparisons under an ordinal code. In this scheme, only the direction of contrast is recorded for each pair of regions. The magnitude of that difference is discarded.

Considering ordinal codes is worthwhile for two reasons. First, the output of many neurons (including V1 cells) is decidedly non-linear [DeAngelis et al. 1993; Anzai et al., 1995]. Second, by not encoding precise differences in brightness, an ordinal code gains invariance to small changes in interregion contrast Sadr et al. [2002].

We can employ the same DOG features described above by modifying their output slightly. Rather than using the complete encoding of the luminance difference between the subunits, we manipulate these measurements to incorporate a Michelson contrast threshold of varying magnitude. We define the unsigned Michelson contrast between two regions with luminances L_1 and L_2 by the following formula:

$$C_m(L_1, L_2) = \left| \frac{L_1 - L_2}{L_1 + L_2} \right| \tag{4}$$

We note that it is unclear that this is the relevant measure of contrast to use when considering nonadjacent image regions [Peli 1990]. For our purposes, it is sufficient to convert our complete encoding into a trinary ordinal encoding of the images. Briefly, for each comparison in our operator banks, we convert the measured luminance difference into an ordinal measurement via the following formula:

$$ord(L_1, L_2) = \begin{cases} sign(L_1 - L_2) & if\ C_m \geq C_{thresh} \\ 0 & if\ C_m < C_{thresh} \end{cases} \tag{5}$$

Given that most adjacent regions will have similar luminances, we expect that a nonzero threshold will nullify the outputs of most local operators. Even with a threshold of zero, we expect that the outputs of local operators will be unstable. Near-zero outputs are “fragile,” in the sense that a small change in luminance to one region or another could result in a sign change. This means that even if local operators are providing contrast information in an ordinal encoding scheme, that information may be very unreliable.

In all cases, recognition is performed by carrying out the luminance comparisons in a particular filter bank for each training image and storing the output values in a feature vector. Each test image is encoded in the same way. The resulting feature vector is labeled based on its nearest neighbor in the training set, as determined by an L_2 norm. As mentioned previously, 10 operator banks were constructed at each spatial scale. Each operator bank is used to create both “local” and “rewired” representations of the images, as well as “complete” and “ordinal” encodings of the output. When employing ordinal encoding, we also parametrically vary the contrast threshold from 0 to 0.6. For both face databases, this allows us to examine the role of different region comparisons and encoding strategies while keeping the actual raw input identical across simulations.

For the sake of comparison to an established algorithm that uses global pixel-based features, we also evaluate the performance of an Euclidean classifier based on linear discriminant analysis (LDA) of the faces in the ORL and Harvard databases. In each case, the labeled training images are used to determine a set of global basis functions that maximize the ratio of between- to within-individual distances. Training and test images are projected into this lower-dimensional space (200 dimensions for the ORL database and 60 for the Harvard database, corresponding to the number of training images) and test images are then classified according to the nearest neighbor under an L_2norm.

3.3 Results

3.3.1 *Interaction between Lobe Scale and Separation.* We first examine how recognition performance is affected both by the size of operator lobes and the distance between them. In Figure 4, we present recognition rates obtained from the ORL and Harvard databases for both kinds of operators as a function of the spatial constant of the subunit Gaussians.

When recognition is performed using the ORL database (which contains mainly variations in face pose and expression), better results are achieved using a “rewired” representation for small values of the spatial constant. As the operator lobes grow larger, however, the difference between “local” and

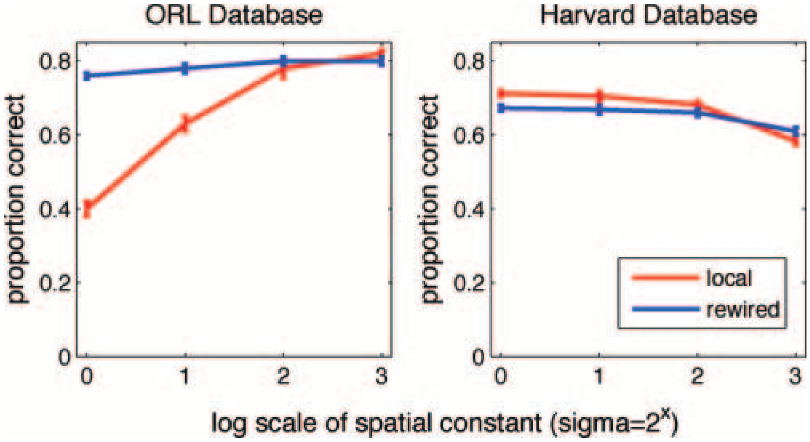

Fig. 4. Recognition performance as a function of spatial scale for both the ORL database (left panel) and the Harvard database (right panel). Adjacent comparison results are in red and nonadjacent results are in blue. Mean ± SE (open circles) is displayed.

"rewired" features disappears. By contrast, when we consider the results from the Harvard database, we see a completely different pattern. Now it is "local" features that are superior to "rewired" ones, and the highest recognition rates for both feature types are obtained at the smallest spatial scales.

Based on these results, we can draw two conclusions. First, nonlocal contrast measurements can support better recognition results than local measurements under some circumstances. The pattern of results from the ORL database suggests that nonlocal features are as good or better than local features when face images are subject to pose variation. Thus, region-based representations that decouple σ and δ are a useful tool for classifying faces subject to varying views. Second, the results from the Harvard database indicate that the best choice of features to accomplish a recognition task is highly dependent on the sources of variation in the images. When illumination varies from image to image, small-scale local features (instead of nonlocal or large-scale measurements) are, on average, the better tool for accurate classification.

How do these results compare to recognition carried out using LDA? In the case of the ORL database, we achieved 80.5% accuracy using global templates. This is comparable to the best performance of both the local and "rewired" features in Figure 4. It would seem for this task that local, global, and nonlocal features all provide useful information for recognition. Nonlocal features support accurate recognition across a range of spatial scales, however, whereas local operators only provide useful information as they grow very large. When we consider the Harvard database, we find that LDA performance is only 23%. This is very poor compared to the results from both the local and "rewired" features displayed above. Local features clearly provide the best performance of these three alternatives, while nonlocal features also seem to support relatively good performance.

For both kinds of differential operators, as the spatial scale increases, we see that the difference in performance levels between "local" and "rewired" features grows smaller. As discussed previously, this is likely because of the fact that our rewiring procedure results in mean values of δ that are only slightly above the separation of local features when σ is large. Regardless of this weakness of our rewiring procedure, our results highlight the usefulness of a generic region-based framework for at least one setting of the face-recognition task. Furthermore, we note that the qualitatively different recognition profiles of local, nonlocal, and global features across these two recognition tasks suggest that a broad, multifeature algorithm may be more useful than any of these operators considered alone. Combining useful aspects of each operator within a given task might lead to improved performance. In

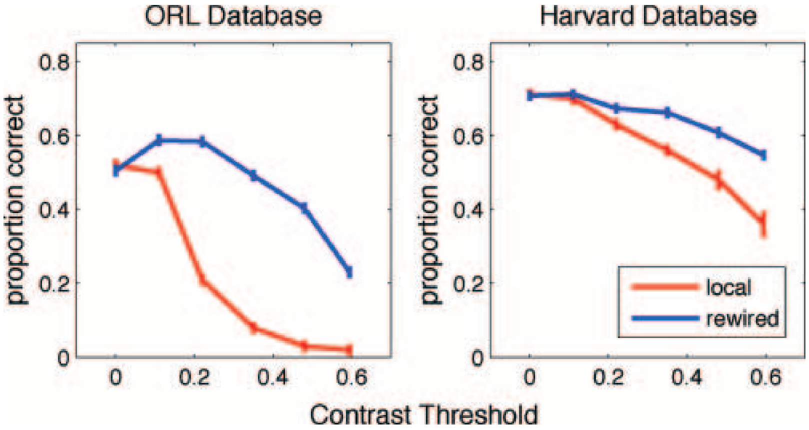

Fig. 5. Recognition performance as a function of the contrast threshold imposed during ordinal encoding of luminance differences. Results obtained from the ORL database are on the left; Harvard database results are on the right. $\sigma = 2$ pixels in both cases, Mean ± SE (open circles) is displayed in both panels.

this context, our region-based operator is not a new "optimal" tool for recognition, but rather a novel piece of a potentially larger system that carries its own distinct information, useful for face classification.

3.3.2 *Lobe Separation and Ordinal Encoding*. We next examine the performance of both "local" and "rewired" features under the constraint of ordinal encoding. First, we seek to determine recognition performance for each of these features when the direction of a luminance difference is the only information recorded by each differential operator. Second, we shall examine the influence of an increasing contrast threshold on each operator type's accuracy. This will give us the ability to determine which encoding strategy might be more useful under a rough approximation of the constraints imposed by neural hardware. Our prediction is that increasing the contrast threshold required to generate a nonzero response will impair "local" feature performance more substantially than "rewired" features. We perform this analysis by fixing the spatial scale of the operators at two pixels, because roughly similar recognition rates were obtained using the ORL and Harvard databases at this value (see Figure 4). We use the same outputs from our first analysis, but impose varying levels of the contrast threshold upon those outputs to obtain an ordinal representation of image structure. The results of our analysis are presented in Figure 5. LDA performance is not assessed in this experiment, as it is not clear how to generalize the ordinal code imposed on our operators to a representation that encodes absolute pixel values rather than contrast between image regions.

In both panels of Figure 5, recognition performance degrades sharply as the contrast threshold increases when adjacent image regions are compared. When nonadjacent region comparisons are used, recognition performance degrades more gracefully. For fairly low levels of the threshold (approximately 0.25) we note that there is a substantial difference between "local" and "rewired" representations. This is especially evident in the results obtained from the ORL database.

Further, we note that ordinal encoding actually proves slightly superior to complete encoding in some circumstances. Using the Harvard database, "rewired" performance increases roughly 5% when an ordinal code is introduced. As we move away from a nonzero threshold in both figures, there is also a slight increase in recognition performance for the rewired operators. This lends some credence to our earlier point describing how failure to completely encode some aspect of an image provides some invariance to changes in that feature. This is also interesting in that it has already been suggested that local, ordinal measures of contrast are powerful tools for face detection given variation in illumination

[Sinha 2002b]. These results suggest that their utility may extend to individuation, especially when nonlocal measurements are also considered.

We conclude that these results support our hypothesis that nonlocal contrast measurements are less impaired than local measurements by ordinal encoding in both pose-varying and illumination-varying scenarios.

3.4 Discussion

Using two different face databases, we have found evidence that supports our initial hypotheses concerning region-based representations of face images for recognition. Long-range region comparisons are useful across multiple scales for recognition when faces vary in pose and expression, indicating that the adoption of a general region-based framework has important merits. Long-range comparisons between regions are also more robust than local comparisons to the imposition of a simple ordinal code across varying pose and illumination. Using a flexible operator that spans edge- and region-based processing, we have also noticed that different sources of variation in the appearance of an individual face (pose versus illumination) call for markedly different features. When illumination varies across images, local contrast measurements at a small scale prove most valuable.

Although we began these experiments with the hypothesis that nonadjacent region comparisons would be particularly useful tools for recognition, the finding that adjacent comparisons have value does not invalidate our main idea. By considering region comparisons to be fundamental, we allow for a spectrum of representational tools ranging from traditional edgelike features to the spatially disjoint structures we have found to be useful for recognizing the ORL faces. Our goal is not to argue against local contrast information, but rather to argue for augmenting our set of representational tools to include long-range (nonlocal) measurements. Indeed, we suggest that local, nonlocal, and global representations may all contain useful (and unique) information for classification.

We must also point out that we have not made any claims about our sampling procedure producing state-of-the-art recognition results. Indeed, we are not suggesting that randomly sampling features from face images is a useful way to build a recognition system. Our use of random feature sampling is meant to provide a means by which we can assess the utility of adjacent and nonadjacent region comparisons. What we observe is that, on average, such measurements are less useful than nonadjacent comparisons in some scenarios and more useful in others. In neither case, do we mean to imply that there is no useful information in other representational schemes.

This discussion highlights an important weakness of our random sampling procedure. At present, we have only been able to make observations about the average performance of various differential operators. It would be better to make a statement about which features are actually the best for recognition in a particular situation. While the average performance of one set of operators may be poor, it could also be the case that there is a subset of such operators that support extremely good performance. In randomly sampling features, we will not find such features, because their utility will be washed out by the generally negative contributions of other members of the set. In principle, we would like to find a way to isolate the best features, such that we can examine the qualities of an “optimal” face representation under a general region-based model.

In our second set of analyses, we attempt to carry out this task. Our goal is to identify which luminance comparisons are most informative of identity, given a particular recognition problem and discuss how the nature of those comparisons varies as a function of the problem domain. Given the results of our first experiment, we hypothesize that for the ORL database longer-range comparisons of luminance should dominate the set of “optimal” features. Conversely, the same analysis of the Harvard database should yield “optimal” features that primarily carry out local comparisons.

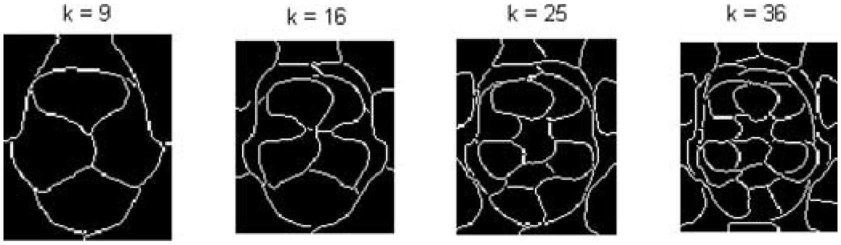

Fig. 6. Examples of the average "training face" from the ORL database clustered into regions of approximately uniform brightness. The number of clusters requested of the algorithm is displayed above each face.

4. EXPERIMENT 2

We present a procedure for selecting and evaluating a set of luminance comparisons between regions identified in face images. We break this problem into two broad steps: region identification and feature evaluation. We assume that, in all cases, we are provided with a labeled set of training images that we may use to guide both processes. We shall first describe the procedure we use to locate and rank region comparisons and then we present the results of this analysis on both the ORL and Harvard databases.

4.1 Methods

4.1.1 *Region Identification.* Given a set of labeled face images, how can we learn the best set of luminance comparisons for classification? If we are to consider comparisons at an arbitrary number of scales, orientations, and distances, we have an intractably large set of features to choose from. We, therefore, need a principled way of winnowing this large space of possibilities down to a more manageable size.

Given the spatial redundancy of natural images, it will be particularly useful for us to make measurements where information is concentrated. To that end, the first step in our algorithm is to find a set of face regions with relatively uniform luminance. Once these regions have been identified, we will only consider the comparisons that result between pairs of regions. Comparisons within a region would be generally useless, as we know that they should be near zero. In only enacting comparisons between regions, we are explicitly sidestepping the spatial redundancy of the face image by placing similar neighboring pixels into one group.

We shall not carry out this segmentation on each unique face image, but rather on the *average face*, obtained by taking the mean luminance of each pixel across our entire database. This saves time by only requiring segmentation of one face instead of many. Also, averaging faces together results in a relatively smooth image that highlights consistencies across the set while removing noise. Aligning the faces with respect to eye position and/or size will result in a cleaner end result. Here, we have not carried out any such preprocessing steps, as we are interested in the system's ability to discover features invariant to changes in position or viewing angle without the benefit of independent normalization. In the case of the ORL database, this results in a "fuzzy" grand average face, while the Harvard database yields a sharper grand average.

We find regions in the average face using the k-means clustering algorithm. We represent each point in the average face by its luminance and spatial location (x, y coordinates). Including both the grayscale value and location of the pixel in the clustering procedure ensures that pixels with similar luminances will be clustered together, while enforcing spatial contiguity within a cluster. Some examples of this clustering, on the average face, extracted from the ORL database, are shown in Figure 6.

The k-means algorithm requires that the user specify how many clusters should be returned. There are no principled guidelines as to how to choose this value, so we have adopted a practical criterion. In general, we would like to have as many regions as possible, so that we might examine a larger number of pairwise comparisons between the regions. However, the number of pairwise comparisons that exist between k regions is equal to (1/2)*$k(k-1)$. We wish to keep this number from growing too large and we also wish to keep the region size from growing too small. To this end, we have used an intermediate value of $k = 25$. There is little qualitative variation across values of this parameter ranging from 16 to 30, leading us to believe that this value does not need to be fine tuned.

4.1.2 *Feature Evaluation.* Once we have identified a list of x, y coordinates that represent the centers of k regions of uniform luminance, we proceed by identifying which pairwise region comparisons are the most useful for classification. For $k = 25$, we have a set of 300 pairwise comparisons to consider. We execute each of these comparisons using the same DOG operators we have already employed. A Gaussian lobe is placed at the centroid of each region and we record the luminance difference between the two lobes. We use complete encoding of these luminance differences at present, but extending this method to ordinal encoding is straightforward. The spatial constant of the lobes, σ, is a free parameter in this algorithm. We have chosen a value of $\sigma = 2$ pixels for this analysis, primarily because it offered good performance in our previous experiments for both adjacent and nonadjacent comparisons with complete or ordinal encoding.

We wish to rank these 300 features in descending order of recognition performance. Unfortunately, the contribution of any one feature in isolation is so small that we cannot simply include or exclude individual measurements one at a time and use fluctuations in recognition accuracy as our measure of efficacy. The combined contribution of many measurements is necessary to obtain a range of performance values that we can use for feature ranking.

To that end, we adopt a sampling procedure that can be used to evaluate an arbitrarily large collection of image features. We first initialize a set of weights, one for each feature in our collection. These weights will all be zero at the beginning of the evaluation. We then randomly select n features from our full set. We measure recognition accuracy with that subset of n features using our training and test images, as described in Experiment 1. The weights of all these n features are then incremented by the percentage of faces correctly identified. We repeat this procedure for several iterations and also record the number of times each feature was sampled.

When we have completed all of our iterations, we may calculate the rank of each feature. We divide the final weight of each feature by the number of times it was sampled. The result is a score that reflects the average recognition performance obtained when a particular feature participated in representing the images.

4.2 Results

4.2.1 *ORL Database.* In Figure 6, we present the results of this analysis for the full ORL database using 200 training and 200 test images. We sampled $n = 50$ comparisons from our full set of 300 features at each iteration, but feature ranking is consistent for values of n between 20 and 100.

The graph of sorted feature scores (Figure 7) reveals that there is not a great deal of variation in the weighted accuracy scores we obtain. This helps explain the very small levels of standard sampling error that we have observed throughout these experiments. When features are combined into large enough subsets, the resulting recognition rate is very consistent.

We note two points of inflection at the left and right of the graph, however. These indicate that there exist two subsets of features that are markedly better and worse, respectively, than the rest of the features. This is encouraging, because it suggests that accurate classification is possible with a very small set of luminance comparisons. We test this by choosing the top 20 features determined by our ranking and perform recognition on the full ORL database (five training, and five test images per

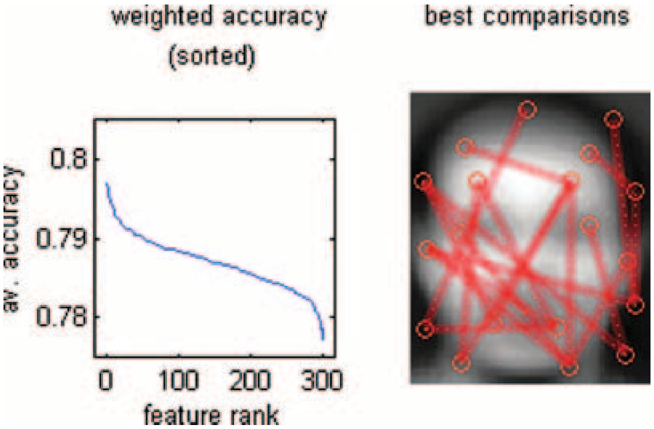

Fig. 7. Evaluation of the 300 pairwise comparisons extracted from segmenting the average ORL face. At left, the sorted scores of each feature are presented. There is not a great deal of variation across these features, but we note clear inflection points at the left and right of the graph. These suggest that relatively small subsets of both "good" and "bad" features exist in this full set of 300. We select the 20 comparisons with the highest rank and plot those comparisons schematically in the right panel.

individual) using only the "best" comparisons (Figure 7). Performance with this subset of features is 88%, when we look for the correct individual as the top match, and increases to 92.5%, when we look for the correct individual in the top three matches. For comparison, if we choose a random subset of 20 features for recognition, we only attain 75.5% performance, on average, using the top match, and this increases to 78% when we use the top three matches. Performance with the "best" features is significantly better than this random sampling ($p < 0.0001$, one-tailed t-test). In both cases, using the "best" features is superior.

Given this supporting evidence suggesting that our sampling procedure results in an "optimal" selection of luminance comparisons, we may now discuss what features appear to be the most useful for classification. Our first observation is that the distance over which the best comparisons are carried out appears to be rather large. The mean distance between regions in our set of best comparisons is 64.03 pixels. We compare this to a distribution of mean interregion distances obtained by sampling 10,000 random subsets of 20 features from our original set of 300 (Figure 8). We see that the mean distance between regions in our optimal set is indeed substantially greater than what we expect, on average. This supports our earlier observation that nonadjacent luminance comparisons are particularly useful for classification using this data set, as they overcome the sparsity of data produced by adjacent comparisons.

4.2.2 *Harvard Database.* In Figure 9, we present the results of the same analysis conducted for the Harvard database. Sixty training and 540 test images were used for this analysis, with the same sampling parameters as described for the ORL faces. Once again, the graph of the sorted feature rank is displayed and the "best" features are plotted on the average face.

Performance with the 20 "best" features is 82%, when we look for the correct individual for each target as the top match, and increases to 85%, when we look for the correct individual in the top three matches. Choosing a random subset of 20 features for recognition only yields 72% performance, on average, using the top match increasing to 74.5% when we use the top three matches. As in the ORL analysis, performance with the "best" features is significantly better than random sampling ($p < 0.0001$, one-tailed t-test).

The mean inter-region distance for the best Harvard features is less than we would expect by chance. This is in contrast to the ORL features, which were biased toward longer-range comparisons. Specifically,

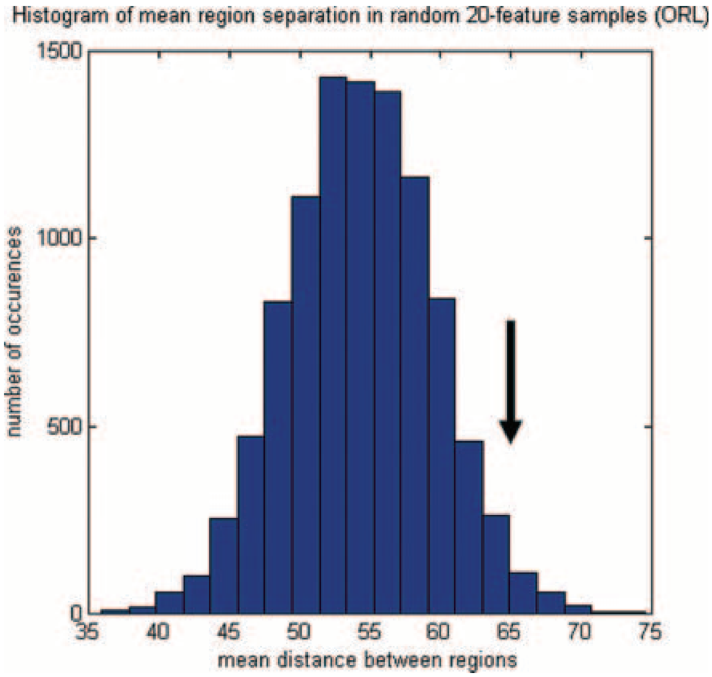

Fig. 8. Distributions of mean interregion distances obtained from 10,000 random 20-feature subsets of our initial feature sets of 300 interregion comparisons obtained from the ORL images. The mean interregion distance between lobes of each set of 20 optimal features is marked by the arrow. Note that the mean length is larger than expected by chance.

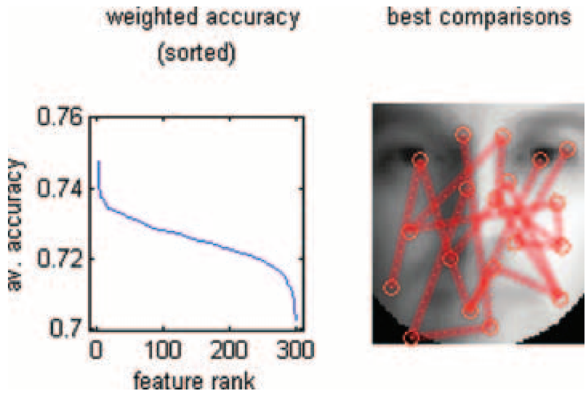

Fig. 9. Evaluation of the 300 pairwise comparisons extracted from segmenting the average Harvard face. At left, the sorted scores of each feature are presented. The 20 comparisons with the highest rank are plotted in the right panel.

the mean interregion distance for the 20 best Harvard features is 43 pixels. We construct a distribution of mean interregion distances in the same manner as we did for the ORL database, and find that this value is on the lower end of expected mean separations (Figure 10). This is consistent with our finding that short-range features are more useful than their long-range counterparts when illumination varies a great deal across our test images.

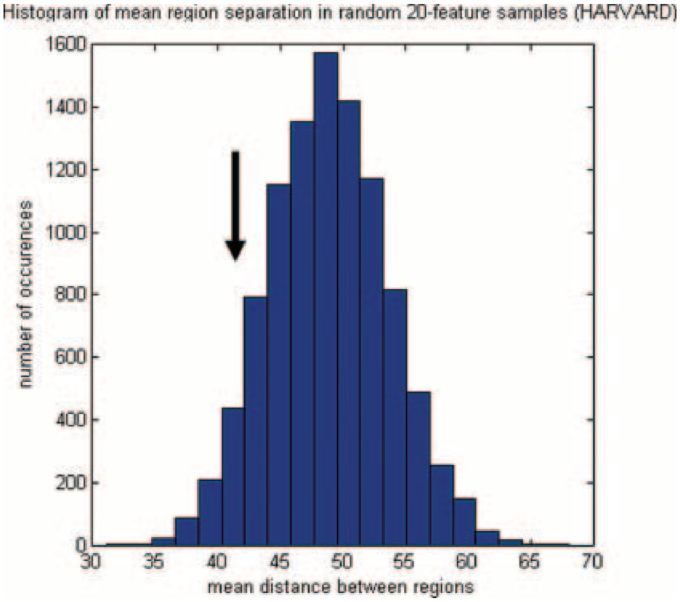

Fig. 10. Distributions of mean interregion distances obtained from 10,000 random 20-feature subsets of our initial feature sets of 300 interregion comparisons obtained from the Harvard images. The mean interregion distance between lobes of each set of 20 optimal features is marked by the arrow. Note that the mean length is smaller than expected by chance.

4.3 Discussion

We have constructed sets of optimal luminance comparisons between approximately uniform regions under a recognition criterion. A small subset of useful features can be discovered from an arbitrarily large starting set of interregion comparisons. Such an optimal subset is shown to be statistically superior to randomly selected subsets of equal size.

The subset of features recovered using the ORL database is also composed of more long-range region comparisons than we expect by chance. In contrast, the features recovered using the Harvard database are composed of more short-range comparisons than we would expect by chance. We point out that both sets of comparisons are markedly different than many that have been proposed for face detection [Sinha 2002b; Viola and Jones 2001]. However, given that the problem of individuation (person A versus person B) is quite different from the problem of detection (face versus nonface), the specific features that are useful for detection may simply be too universal across the population of faces to be of any great use for individuation. What is most interesting to us here is how the structure of the optimal solution changes as the problem domain is changed. The emergence of small-scale local features as a robust solution to recognition under varying illumination is noteworthy in that it agrees with recent results concerning the optimality of gradient-based features for such tasks [Chen et al. 2000]. Local comparisons between small regions can be thought of as samples of the image gradient. It, therefore, makes a great deal of sense that these features should emerge in this framework as a solution to the problem of changes in lighting. The emergence of nonlocal features given variation in position and viewing angle also seems reasonable. When a face translates or rotates in depth, edges either translate a great deal or vanish as a result of self-occlusion of the face. This makes local features very unstable. By comparison, even though the boundaries of a uniform region within the face may change (such as, the cheeks or forehead), the region's average luminance remains fairly coherent. Nonlocal features that make comparisons between these more robust image features are thus more stable than edgelike features.

These results make some interesting predictions about human recognition of faces that vary in pose and illumination. There is a growing body of evidence that the human visual system is able to extract nonlocal relationships and uses those relationships for a wide range of tasks [Burbeck and Pizer 1995; Kohly and Regan 2000; Danilova and Mollon 2002]. We have discussed the implications of these findings elsewhere [Balas and Sinha, in press]. Presently, we wish to focus more specifically on the problem of determining the perceptual relevance of local and nonlocal comparisons for face recognition. There have been many studies concerning the role of local versus global features in human face performance [Schwaninger et al. 2002; Le Grand et al. 2003], but little attention to the possibility that nonlocal comparisons between face regions are of any importance. Our findings in this second analysis suggest that for faces that vary in pose, nonlocal comparisons contain more diagnostic information than local comparisons. It may be the case that human observers implicitly make use of nonlocal contrast for recognition across pose, but not for recognition across different illumination conditions.

A good test of these predictions could be carried out within the "Bubbles" paradigm [Gosselin and Schyns 2001]. "Bubbles" images have been constructed for a wide range of face discrimination tasks [Schyns et al. 2002; Chauvin et al. 2005] and provide a straightforward means of determining what face regions are most relied upon by human observers to carry out various discrimination tasks. To test our predictions regarding the use of nonlocal information in recognition across pose and illumination, it would be interesting to employ masks with particular distributions of Bubbles across the image. If one must perform a same/different recognition task with masked faces varying in pose, our data suggests that widely spaced Bubbles would be more useful than closely spaced pairs. Conversely, if the faces vary in illumination, we would predict that closely spaced pairs of Bubbles would be best.

Demonstrating the perceptual utility of local and nonlocal region comparisons could help motivate more detailed computational models of region-based face recognition. Interesting work has already been done adapting global [Moghaddam et al. 2000] features, local features of "intermediate complexity" [Ullman et al. 2002] and edge features [Wiskott et al. 1997] to solve various face-recognition tasks. Augmenting these models with region-based representations may prove very fruitful. In our final experiment, we describe and implement a region-based recognition system to provide a baseline for future efforts. Although we do not expect to achieve state-of-the-art results at present, we shall at least be able to compare a "bare-bones" region-based recognition system to standards, such as pixel-based LDA systems and established edge-based algorithms in the context of a standard face database.

5. EXPERIMENT 3

In this final section, we evaluate the performance of a region-based face recognition system on the FERET database [Phillips et al. 2000]. In our previous analyses, we have described various methods for determining what region comparisons are most useful for solving particular recognition problems (pose versus illumination variation). Here we describe a method for representing an arbitrary set of images with a set of binary region comparisons. We assume no prior knowledge concerning what transformations faces undergo between the training and test sets. Our results in this section are meant to provide a benchmark for the performance of a region-based representation. The effectiveness of future improvements and embellishments of the framework can be easily compared both to our initial efforts and other standard algorithms that make use of either global or local representational schemes.

5.1 Methods

5.1.1 *Stimuli.* We use the standard training, gallery, and probe sets of face images contained in the FERET database. All images in the database underwent histogram equalization and normalization with respect to eye position and size [Bolme et al. 2003].

5.1.2 *Procedure.* Our first step is to determine a set of roughly uniform regions we can use as the center points for our two-lobed region comparisons. We accomplish this by constructing an average face from the normalized training images and performing k-means clustering over the pixels in that image based on intensity and x, y position. We set $k = 25$ and adjust the weighting of x, y coordinates relative to intensity so that spatial contiguity is enforced. This is best accomplished by starting with equal weighting of these features and increasing the weight of the x, y coordinates gradually until all regions are perfectly contiguous. In the case of the FERET training images, a 2:1 ratio of position weights to intensity weights was required to form fully connected regions.

The second step is to construct a set of bilobed features that will execute intensity comparisons between each pair of regions. Regions that were outside the face area were discarded from consideration. For the remaining region pairs (210 in all), a bilobed feature was constructed for each pair by placing a positively weighted Gaussian lobe at the x, y centroid of one region and a negatively weighted Gaussian lobe at the x, y centroid of the other. A spatial constant of two pixels was used for each Gaussian lobe. This value was selected because each region could be almost completely covered by a Gaussian of this size. In general, we suggest choosing the value of the spatial constant such that this holds true.

Finally, the full set of bilobed features were applied to the training images and we carried out PCA to determine a Mahalanobis space for classification of the probe images given a new gallery of face images.

To review, our region-based representation is developed as follows, assuming a set of unlabeled training images:

1. Construct an average face from the training images.
2. Perform k-means clustering on the average face, using intensity at each pixel and x, y coordinates to find 25 regions. The values of the x, y coordinates should be minimally weighted to enforce strict spatial contiguity.
3. Discard any regions that do not overlap with the face area within the average image.
4. For each remaining unique pair of regions, construct bilobed difference-of-Gaussian features with each lobe centered on the x, y centroid of its assigned region.
5. Apply the full set of bilobed features to each gallery image and probe image.
6. Use gallery images to classify probe images using whatever space and distance metric one prefers. (Here we implement a Mahalanobis cosine metric.)

5.2 Results

We present recognition rates for three distinct probe sets taken from the FERET database. These are the "fb," "dup1," and "dup2" probe sets. Relative to the 1196 gallery images, the "fb" probe images are composed of 1195 additional images of the same individuals with a nonneutral facial expression. The "dup1" and "dup2" probe images contain images of the individuals in the gallery photographed at a later date. The "dup1" images were taken between 0 and 1031 days after the gallery images (median elapsed time between photographs is 72 days). The "dup2" images were taken between 540 and 1031 days after the gallery images (median elapsed time between photographs is 569 days). There are 722 images in the "dup1" set and 234 images in the "dup2" set. Cumulative match scores for our region-based representation using cosine classification in a PCA-Mahalanobis space are displayed in Figure 11.

Performance of our region-based representation compares favorably to other simple benchmarks such as LDA classification on raw pixel values [Beveridge et al. 2001]. Even when we use a less sophisticated distance metric that does not incorporate PCA or Mahalanobis cosine distance (such as a simple L2 norm), the cumulative match scores are slightly above this baseline in some cases (Figure 12). We find these results encouraging, as it suggests that characterizing a face in terms of the relationships between a small set of uniform regions is a plausible alternative to global image representations.

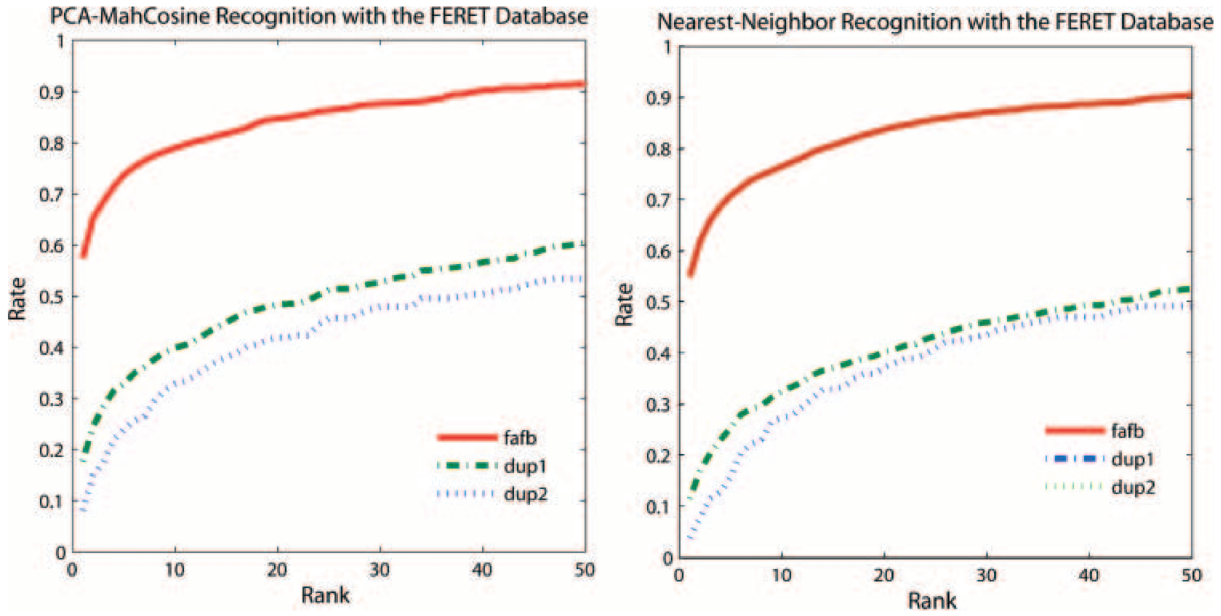

Fig. 11. Cumulative match scores for region-based performance on three FERET probe sets. The xaxis determines the maximum allowed position of the correct target face in the ranked order of gallery faces for classification to be considered correct. At left, is performance under a cosine distance metric in PCA-Mahalanobis space. At right, is performance under a Euclidean nearest-neighbor distance metric.

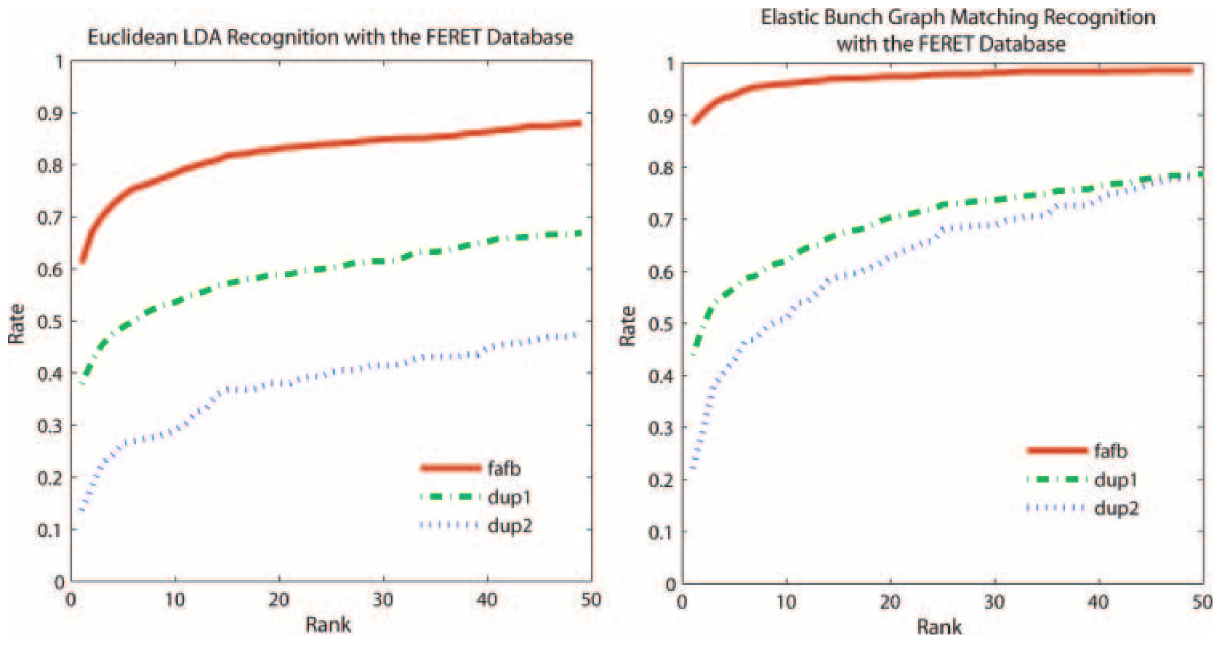

Fig. 12. At left, cumulative match scores for LDA classification on the three FERET probe sets. Performance is comparable to our region-based algorithm. At right, performance of an elastic bunch graphing matching (EBGM) algorithm on the same data set. Performance with the latter algorithm is superior to both LDA and region-based procedures.

We note, however, that when we compare our algorithm to an edge-based system based on elastic bunch graph matching (EBGM) [Wiskott et al. 1997], we are below their standard (Figure 12). This may seem puzzling given the rather limited success of strictly local representations throughout our analyses so far, but also may indicate useful steps we can take to refine our region-based model in the future.

EBGM is carried out by characterizing a set of fiducial points on each face image by a multiscale, multiorientation pyramid of Gabor functions. Since the relevant points on a face may appear in different locations across different images, a crucial step in the algorithm is to flexibly allocate each "jet" of local features to the correct location on the image. This is clearly a sophisticated procedure and demonstrates how the usefulness of a given feature can be augmented by additional processing. There are two aspects of this algorithm that may be useful to apply to our region-based algorithm. First, considering interregion comparisons at multiple scales may be worthwhile. We noted in experiment 1 that nonlocal features performed well across a range of spatial scales when applied to the ORL database and so it may be that considering nonlocal comparisons at multiple scales between regions is worth the extra complexity of the representation. Second, just as EBGM begins by determining where the fiducial points are on a new test image, it may be important to give our system the ability to locate face regions from scratch on each test image. Rather than doing one segmentation of an average face, it may be better to segment each face separately and consider how regions move about from image to image. Flexible comparisons between regions that can "float" from image to image may more accurately capture the information necessary for classification.

5.3 Discussion

The region-based representation that we propose for faces is certainly not competitive with the highest performing systems developed to date. However, we are optimistic concerning this representation for two reasons. First, experiments 1 and 2 have demonstrated that there is useful information for classification in generic bilobed region comparisons. Second, using a set of region comparisons obtained from training images in a simple and straightforward way, we see performance comparable to pixel-based LDA. In some regimes, our system appears to do slightly better, while in others (specifically the "dup1" probe set) it does worse. We note, however, that we are retaining only 210 fixed measurements of image contrast, compared to the global templates that must be retained for LDA recognition and the complex warping procedure that supports EBGM. This suggests to us that we have identified a potentially rich source of information for further development of this system.

It is clear that the use of region-based representations is at a very preliminary stage and these results should be interpreted as an introductory baseline rather than as a final word. Beyond the possible improvements mentioned in our discussion of EBGM, there are many other important issues left to resolve. It is not obvious that bilobed comparisons between distinct regions are an optimal means for encoding interregion face structure, for example. Higher-order comparisons may prove informative. Even if we choose to adopt a bilobed operator to characterize region contrast, it may be that tuning the size of each Gaussian lobe to the size of its assigned region is useful. We also have not explored many alternatives for extracting the initial set of face regions from training images. It could be the case that employing some sort of selection criteria, perhaps based on an estimate of uniformity within a region, could help strengthen the diagnosticity of the final representation. Developing our initial encoding scheme further and combining it with established edge-based or global template-based techniques may result in a powerful recognition engine.

6. CONCLUSIONS

We have demonstrated in three experiments that local, nonlocal, and global features can be highly useful for face classification. When face pose and position varies across images, nonlocal measurements provide rich information for classification, as opposed to a representation of the image by adjacent luminance comparisons. When illumination varies across images, local comparisons of luminance prove superior to both alternatives. In the case of illumination variation, we have found that the use of an ordinal code can actually increase recognition performance under some circumstances. In considering pose variation, we have observed that the introduction of nonadjacent region comparisons makes ordinal encoding a viable strategy. In both cases this is encouraging, as such a code may more accurately reflect the nature of neural image encoding. We also note that the relationship between local, nonlocal, and global features appears to be complex, with each basic feature vocabulary performing differently across these two tasks. This may suggest that a multifeature strategy is necessary, with different weight given to each feature, depending on its ability to accurately perform the task at hand.

To more deeply understand how generic region-based comparisons should be effectively used, we have also presented a learning procedure for identifying the optimal set of interregion luminance comparisons given some large set of candidate features. Applying this procedure to the ORL and Harvard databases, we have learned distinct feature subsets that allow us to perform accurate recognition with a small number of operators. Supporting our findings from experiment 1, the features learned from the ORL database are biased toward long-range comparisons. Conversely, the features learned from the Harvard database are predominantly short-range. This demonstrates that the use of our generic region-based framework is powerful enough to both reveal novel sources of diagnostic information and confirm the usefulness of previously proposed representational tools. Furthermore, this procedure makes interesting perceptual predictions about how information may be integrated across a face image when different matching tasks need to be solved.

Finally, we have compared a basic region-based recognition system to widely used techniques for global and local image representation. Although our system does not match the highest standards set for recognition within the FERET database, even a nearest-neighbor classifier in Euclidean space using our dipole primitives performs favorably compared to pixel-based PCA and LDA systems. Both the region-based system and the global algorithm we test here do not achieve the performance level of the much more complex EBGM procedure, however. The superior performance of this latter algorithm indicates potentially valuable embellishments (such as, explicit incorporation of region deformation and movement) that may greatly increase the power of our region-based system.

In short, this work suggests that region-based strategies for face recognition may be a useful new tool for classification schemes. A region-based framework for face recognition provides a unified means of exploring both classical representation strategies that rely on edgelike features, as well as novel tools, such as, our nonadjacent comparisons. Future work to develop this general framework of image encoding into a stronger unsupervised face classification scheme is likely to prove quite fruitful. It may be possible to identify both the generic qualities of useful features for recognition (adjacent versus nonadjacent comparisons) and the specific features that will provide the best performance. By considering generic region-based encoding schemes, we are able to supplement our toolbox of processing strategies with powerful new methods. At the same time, we preserve our ability to utilize classic representational schemes that have previously been shown to be useful in several recognition tasks. Considering how to integrate our novel features with a range of additional tools (such as, local and global features) will likely result in more powerful algorithms for face and object classification.

ACKNOWLEDGMENTS

This work has benefited substantially from discussions with David Cox, Jim DiCarlo, and Gadi Geiger. BJB is a National Defense Science and Engineering Graduate Fellow. PS is supported by an Alfred P. Sloan Fellowship in neuroscience and a Merck Foundation Fellowship.

REFERENCES

ANZAI, A. BEARSE, M. A., FREEMAN, R. D., AND CAI, D. 1995. Contrast coding by cells in the cat's striate cortex: Monocular vs. binocular detection. *Visual Neuroscience*, *12*, 77–93.

ATTNEAVE, F. 1954. Some informational aspects of visual perception. *Psychol. Rev. 61*, 183–193.

BALAS, B. AND SINHA, P. (in press). Receptive field structures for recognition. *Neural Computation*.

BARTLETT, M. STEWART, LADES, H. MARTIN, AND SEJNOWSKI, T. J. 1998. Independent component representations for face recognition. In *Proceedings of the SPIE, Vol 3299: Conference on Human Vision and Electronic Imaging III.* 528–539.

BELHUMER, P. N., HESPANHA, J. P., AND KRIEGMAN, D. J. 1997. Eigenfaces vs. Fisherfaces. *IEEE Transactions on PAMI*, 19, 711–720.

BELL, A. J., AND SEJNOWSKI, T. J. 1997. The "independent components" of natural scenes are edge filters. *Vision Research 37*, 23, 3327–3338.

BEVERIDGE, J. R., SHE, K., DRAPER, B., AND GIVENS, G. H. 2001. A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition. In *Proc. IEEE Conf. CVPR 3*, 535–542.

BIEDERMAN, I. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review 94*, 115–147.

BIEDERMAN, I., AND JU, G. 1988. Surface vs. edge-based determinants of visual recognition. *Cognitive Psychology 20,* 38–64.

BOLME, D., BEVERIDGE, R., TEIXEIRA, M., AND DRAPER, B. 2003. The CSU face identification evaluation system: Its purpose, features and structure. *International Conference on Vision Systems.* Springer-Verlag, New York. 304–311.

BURBECK, C. A., AND PIZER, S. M. 1995. Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research 35,* 13, 1917–1930.

CHAUVIN, A., WORSLEY, K. J., SCHYNS, P. G., ARGUIN, M., AND GOSSELIN, F. 2005. Accurate statistical tests for smooth classification images. *J. Vis*. 5, 9, 659–667.

CHEN, H. F., BELHUMEUR, P. N., AND JACOBS, D. W. 2000. In search of illumination invariants *Proc. IEEE Conf. CVPR*, *2*, 254–61.

DANILOVA, M. V., AND MOLLON, J. D. 2003. Comparison at a distance. *Perception 32*, 4, 395–414.

DEANGELIS, G. C., OHZAWA, I., AND FREEMAN, R. D. 1993. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J. Neurophysiology 69*, 1091–1117.

DAUGMAN, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A 2*, 7, 1160–1169.

DAVIES, G., ELLIS, H., AND SHEPHERD, J. 1978. Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology 63*, 180–187.

FIELD, D. J. 1994. What is the goal of sensory coding? *Neural Computation 6*, 559–601.

GOSSELIN, F. AND SCHYNS, P. G. 2001. Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research 41*, 17, 2261–2271.

KERSTEN, D. 1987. Predicability and redundancy of natural images. *Journal of the Optical Society of America A 4*, 2395–2400.

KOHLY, R. P. AND REGAN, D. 2000. Coincidence detectors: Visual processing of a pair of lines and implications for shape discrimination.*Vision Research 40*, 17, 2291–2306.

LADES, M., VORBRUGGEN, J. C., BUHMANN, J., LANGE, J., V. D. MALSBURG, C., WURTZ, R. P., AND KONEN, W. 1993. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers 42*, 3, 300–310.

LEDER, H. 1996. Line drawings of faces reduce configural processing. *Perception 25*, 355–366.

LEDER, H. 1999. Matching person identity from facial line drawings. *Perception 28*, 1171–1175.

LE GRAND, R., MONDLOCH, C., MAURER, D., AND BRENT, H. 2003. Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience* 6, 1108–1112.

MALIK, J., BELONGIE, S., SHI, J., AND LEUNG, T. K. 1999. Textons, contours and regions: Cue integration in image segmentation, *ICCV*. 918–925

MOGHADDAM, B., JEBARA, T., AND PENTLAND, A. 2000. Bayesian face recognition. *Pattern Recognition 333*, 11, 1771–1782.

NAKAYAMA, K., HE, Z. J., AND SHIMOJO, S. 1995. Visual surface representation: A critical link between lower-level and higher-level vision. In *An Invitation to Cognitive Science, Visual Cognition* S. M. Kosslyn and D. N. Oshershon, Eds., MIT Press, Cambridge, MA.

OLSHAUSEN, B. A. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, London 381*, 607–609.

OLSHAUSEN, B. A., AND FIELD, D. J. 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research 37*, 23, 3311–3325.

PELI, E. 1990. Contrast in complex images. *J. Opt. Soc. Am. A 7*, 10, 2032–2040.

PHILLIPS, P. J., MOON, H., RIZVI, S. A., AND RAUSS, P. J. 2000. The FERET Evaluation Methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence 2*, 10, 1090–1103.

SADR, J., MUKHERJEE, S., THORESZ, K., AND SINHA, P. 2002. The fidelity of local ordinal encoding. In *Advances in Neural Information Processing Systems 14*. T. Dietterich, S. Becker, and Z. Ghahramani, Eds., MIT Press. Cambridge, MA.

SAMARIA, F. AND HARTER, A. 1994. Parametrisation of a stochastic model for human face identification. Paper presented at the *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL.

SINHA, P. 2002a. Identifying perceptually significant features for recognizing faces. *Proc. SPIE Electronic Imaging Symp. 4662*. 12–21.

SINHA, P. 2002b. Qualitative representations for recognition. Lecture Notes in Computer Science vol. 2525. Springer-Verlag, New York, 249–262.

SCHWANINGER, A., LOBMAIER, J., AND COLLISHAW, S. M. 2002. Component and configural information in face recognition. *Lecture Notes Computer Science*, Vol. LNCS 2525. 643–650. Springer-Verlag, New York.

SCHYNS, P. G., BONNAR, L., AND GOSSELIN, F. 2002. Show me the features! Understanding recognition from the use of visual information. *Psychological Science 13*, 402–409.

TURK, M. A. AND PENTLAND, A. P. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience 3*, 1, 71–86.

ULLMAN, S., VIDAL-NAQUET, M., AND SALI, E. 2002. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience 5*, 7, 682–687.

VIOLA, P. AND JONES, M. 2001. *Rapid object detection using a boosted cascade of simple features.* Paper presented at the *Accepted Conference on Computer Vision and Pattern Recognition.*

WISKOTT, L., FELLOUS, J. M., KRUGER, N., AND VON DER MALSBURG, C. 1997. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 7, 775–779.

YIP, A., AND SINHA, P. 2002. Role of color in face recognition. *Perception 31*, 995–1003.

YOUNG, R. A., LESPERANCE, R. M., AND MEYER, W. W. 2001. The Gaussian derivative model for spatial-temporal vision: I. Cortical model. *Spatial Vision 14*, 261–319.

YOUNG, R. A. AND LESPERANCE, R. M. 2001. The Gaussian derivative model for spatial-temporal vision: II. Cortical data. *Spatial Vision 14*, 321–389.

Received July 2005; revised November 2005 and February 2006; accepted February 2006