A Comparative Study of Local Matching Approach for Face Recognition

Jie Zou, *Member, IEEE*, Qiang Ji, *Senior Member, IEEE*, and George Nagy, *Fellow, IEEE*

***Abstract*—In contrast to holistic methods, local matching methods extract facial features from different levels of locality and quantify them precisely. To determine how they can be best used for face recognition, we conducted a comprehensive comparative study at each step of the local matching process. The conclusions from our experiments include: 1) additional evidence that Gabor features are effective local feature representations and are robust to illumination changes; 2) discrimination based only on a small portion of the face area is surprisingly good; 3) the configuration of facial components does contain rich discriminating information and comparing corresponding local regions utilizes shape features more effectively than comparing corresponding facial components; 4) spatial multiresolution analysis leads to better classification performance; 5) combining local regions with Borda count classifier combination method alleviates the curse of dimensionality. We implemented a complete face recognition system by integrating the best option of each step. Without training, illumination compensation and without any parameter tuning, it achieves superior performance on every category of the FERET test: near perfect classification accuracy (99.5%) on pictures taken on the same day regardless of indoor illumination variations, and significantly better than any other reported performance on pictures taken several days to more than a year apart. The most significant experiments were repeated on the AR database, with similar results.**

***Index Terms*—AR database, face recognition, FERET database, local matching method.**

I. Introduction

OVER the last decade, face recognition has become one of the most active applications of visual pattern recognition due to its potential value for law enforcement, surveillance, and human-computer interaction. Although face recognition systems show striking improvement in successive competitions [35], [36], the face recognition problem is still considered unsolved. Modern face recognition methods can be generally divided into two categories: *holistic matching methods* and *local matching methods*.

Manuscript received October 27, 2006; revised June 19, 2007. This work was supported in part by the Task Support Working Group (TSWG) of the U.S. under Grant N41756-03-C-4028. Portions of the research in this paper used the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Manuel Samuelides.

J. Zou was with Rensselaer Polytechnic Institute, Troy, NY 12180 USA. He is now with the National Library of Medicine, Bethesda, MD 20894 USA (e-mail: jzou@mail.nlm.nih.gov).

Q. Ji and G. Nagy are with Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: qji@ecse.rpi.edu; nagy@ecse.rpi.edu).

Digital Object Identifier 10.1109/TIP.2007.904421

After the introduction of Eigenfaces [21], [44], holistic matching approaches, that use the whole face region as the input to a recognition system, were extensively studied. The principle of holistic methods is to construct a subspace using principal component analysis (PCA) [21], [44], [46], linear discriminant analysis (LDA) [4], [13], [26], [41], [48], or independent component analysis (ICA) [3]. The face images are then projected and compared in a low-dimensional subspace in order to avoid the curse of dimensionality.

Recently, local matching approaches have shown promising results not only in face recognition [2], [15], [16], [20], [29], [33], [42], [47], [53] but also in other visual recognition tasks [45]. The general idea of local matching methods is to first locate several facial features (components), and then classify the faces by comparing and combining the corresponding local statistics.

Heisele *et al.* compared component (local) and global (holistic) approaches and observed that "the component system outperformed the global systems for recognition rates larger than 60%" [18]. Due to increasing interest, in recent surveys stand-alone sections were specifically devoted to local matching methods [43], [54].

Careful comparative studies of different options in a holistic recognition system have been reported in the literature [38]. We believe that a similar comparative study on the options at each step in the local matching process will benefit face recognition through localized matching. Although several stand-alone local matching methods have been proposed, we have not found any studies on comparing different options. The aim of this paper is to fill this blank by presenting a general framework for the local matching approach, and then reviewing, comparing and extending the current methods for face recognition through localized matching.

Illumination compensation is an important issue in face recognition. The most common approach is to use illumination insensitive features, such as Gabor features and local binary pattern features. However, as pointed out by Adini *et al.*, illumination insensitive features are insufficient to overcome large illumination variations [1]. Prompted by the work of Belhumeur and Kriegman [5], the second approach, modeling illumination variations, has received increasing attention [17], [24]. We consider illumination compensation as an independent preprocessing step. Better illumination compensation methods will help every classification method. In this comparative study, we concentrate on classification. Modeling illumination is beyond the scope of this paper.

We adopt FERET frontal face images [34], [35], the most widely adopted benchmark, to conduct this comparative study. Even though impressive progress has been made on frontal face

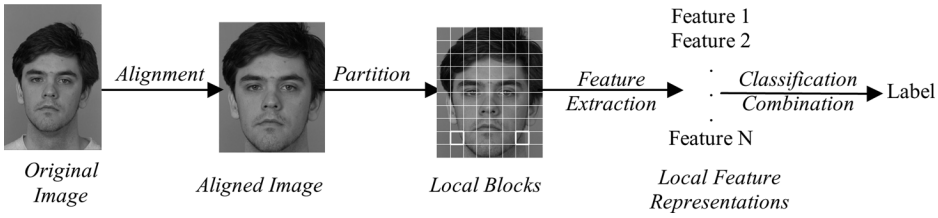


Fig. 1. Diagram of local matching face recognition.

recognition, the best reported recognition accuracy of FERET Dup1 and Dup2 probes is still below 75%, far from adequate for many practical applications. This relatively low recognition accuracy allows us to compare the performances of different options in the local matching process. Most of the important experiments are also duplicated on the AR face database [28]. Results from these two databases agree with each other.

Based on the results of the comparative experiments, we implement a complete local matching face recognition system by integrating the best option in each step of the local matching process. The resulting face recognition system achieves superior performance on the FERET test.

The rest of the paper is organized as follows. In Section II, we review the existing local matching face recognition methods. In Section III, the general framework of the local matching process is divided into three steps: alignment and partition, local feature extraction, and classification and combination. We list and discuss different options in each step. A set of hypotheses is also raised. They are empirically justified in Section V. In Section IV, we briefly describe the FERET and the AR databases. Section V is the detailed comparative experimental study, which answers the questions raised in Section III. Discussion and conclusions constitute Section VI.

II. Literature Review of Local Matching for Face Recognition

In the mid 1990s, researchers began to pay attention to local facial features and proposed several local matching approaches to face recognition. Pentland *et al.* extended the eigenface technique to a layered representation by combining eigenfaces and other eigenmodules, such as eigeneyes, eigennoses, and eigenmouths [33]. This modular eigenface approach was also studied and extended by several other researchers. Gottumukkal and Asari argued that some of the local facial features did not vary with pose, direction of lighting and facial expression and, therefore, suggested dividing the face region into smaller subimages [16]. A similar approach, named subpattern PCA or SpPCA, was studied by Chen and Zhu [8]. Tan and Chen realized that different parts of the human face may contribute differently to recognition and, therefore, extended SpPCA to adaptively weighted subpattern PCA [42]. Geng and Zhou made a similar observation, but chose to select several regions from all possible candidates instead of weighting them [15].

Wiskott *et al.* achieved good performance with the elastic bunch graph matching (EBGM) method [47] in the FERET test [35]. The elastic bunch graph is a graph-based face model with a set of jets (Gabor wavelet components) attached to each node of the graph. The algorithm recognizes new faces by first locating a set of facial features (graph nodes) to build a graph, which is then used to compute the similarity of both jets and topography.

Martinez warped the face into a "standard" (shape free) face, and divided it into six local regions. Within each local region, a probabilistic method was used to determine how "good" the match was. The final classification was based on the linear combination of the probabilities of the six local regions [29].

Local binary pattern (LBP) was originally designed for texture classification [31], and was introduced in face recognition in [2]. The face area was divided into 49 small (7×7) windows. Several LBP operators were compared and the $LBP_{8,2}^{U2}$ operator in 18×21 pixel windows was selected because it was a good tradeoff between recognition performance and feature vector length. The chi square statistic and the weighted chi square statistic were adopted to compare local binary pattern histograms.

Zhang *et al.* proposed local Gabor binary pattern histogram sequence (LGBPHS) by combining Gabor filters and the local binary operator [53]. The face image was first filtered with Gabor filters at five scales and eight orientations. The local binary operator was then applied to all 40 Gabor magnitude pictures (Gabor filtering results) to generate the local gabor binary pattern histogram sequence. The LGBPHS of the unknown face was compared with the LGBPHS of the reference faces using histogram intersection and weighted histogram intersection.

III. General Framework for Face Recognition With Local Matching

As shown in Fig. 1, the local matching process can be divided into three major steps: alignment and partitioning, feature extraction, and classification and combination. Existing local matching systems adopt different methods in each of these three steps. Besides comparing overall system performance, we compare the different options in each step, so that their advantages and disadvantages can be analyzed in detail.

A. Alignment and Partitioning

As in the holistic approaches, the first step of most local matching methods is to align the face images. The aligned faces are then partitioned into local blocks. The alignment and partition steps are usually related, so we discuss them together.

The alignment and partition methods can be divided into three categories. The methods in the first category locate a few local facial components, such as eyes, nose, mouth, and so on. The face is partitioned into these facial components and in the subsequent recognition process the corresponding facial components are compared. This category of methods abandons the *shape* information (the geometric configuration of facial components),

and compares only the *appearance* (photometric cues) of the corresponding components. We believe that the configuration of facial components is very valuable and should play a part in automated face recognition.

The second category of alignment methods warps the face into a "standard" (shape free) face. Warping the face to a standard face is reported to benefit the holistic methods [10] because most holistic methods stack the face pixels into 1-D vectors. Precise registration of corresponding pixels is required in the following dimensionality reduction process. After warping, shape information can be incorporated into identification by separating shape and appearance features and then comparing them independently. For example, in [23], the shape of a new face is fitted by the active shape model [9]. The face is then warped into a "standard" (shape free) face to extract shape and appearance (gray-level) features separately. These two kinds of features are fed into shape and appearance classifiers, and the final classification combines the two.

Although warping is beneficial for holistic methods, applying it to local matching methods requires careful reconsideration. As will be discussed in the following several paragraphs, misalignment of local patches usually reflects the fact that the two faces have different global shape, and, therefore, misalignment can be utilized in the recognition. Additionally, although warping is designed to transform only the global shape of the face into a "standard" face, it also deforms the local facial components. These local features, such as eye corners, mouth corners, nostrils, and eyebrows, carry important discriminating information; therefore, we consider deforming them counterproductive.

The third category of approaches aligns the face into a common coordinate system (instead of warping it into a standard face) by a similarity transform (translation, rotation and scaling) based on certain detected fiducial points. The face is then divided into local regions. The recognition step compares the corresponding local regions (centered at the same coordinates), instead of corresponding facial components.

We emphasize the difference between local *components* and local *regions*. Local components are areas occupied by the facial components, such as eyes, noses and mouths, and centered independently at the component centers. Local regions are local windows centered at designated coordinates of a common coordinate system.

Although for a given face image, the local component, e.g., left eye, and the corresponding local region, the left eye region, may be only a few pixels from each other, we will show that comparing local regions is better than comparing local components. The corresponding local regions of faces with different shapes (facial component configurations) often cover different facial features; therefore, good false matches are rare. This is illustrated in Fig. 2, where the four pictures of two faces are aligned based on two eye centers. The faces are divided into local regions by the white grids. We can see that local regions of the same person cover the same facial components, but it is clear that the top of the eyebrows and mouths of these two faces are in different grid cells (local regions). Comparing corresponding local regions implicitly utilizes the geometrical configuration of facial components, i.e., holistic shape information, and helps to eliminate many invalid candidates. Alignment by similarity transform does not alter local facial components, which further contributes to accurate classification. In Section V, we will verify the conjecture that comparison of corresponding local regions is better than comparing corresponding facial components.

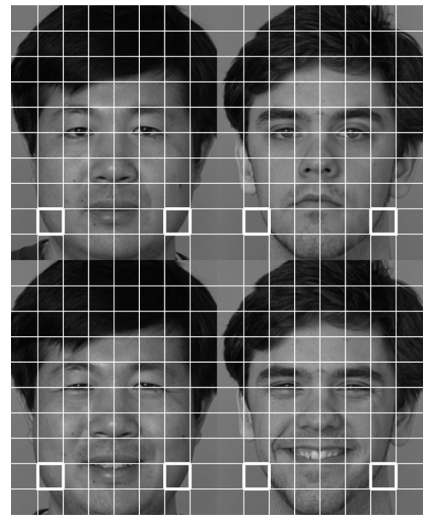


Fig. 2. Four pictures of two faces are aligned based on two eye centers. Comparing corresponding local regions instead of corresponding facial features uses shape information implicitly (see text for explanation).

In many existing face recognition methods (both holistic and local), face images are cropped, keeping only internal components, such as eyes, nose, and mouth, and abandoning external features, such as cheek contour and jaw line. Perhaps most researchers assume that internal components and their mutual spatial configuration are the critical constituents of a face, and the external features are too variable. However, a remarkable illusion presented by Sinha and Poggio suggests that the human visual system makes strong use of the overall head shape [40]. In the example in Fig. 2, the regions marked with thick boxes don't cover much of the right face, but they are useful for distinguishing thin faces from round faces. We shall test in Section V how much these "external" local regions contribute to the final classification.

The alignment of the faces depends on the accuracy of the fiducial points, which will, therefore, affect the recognition performance. We shall study experimentally how the displacements of fiducial points affect the recognition performance in Section V.

B. Local Feature Representation

We discuss three commonly used local feature representations: eigen (PCA) features, Gabor features, and local binary pattern (LBP) features.

Eigen features are probably the most widely adopted local features, possibly due to the early success of the holistic eigenface method. However, as has been clearly articulated both for general pattern classification problems [11] and specifically for face recognition [4], PCA is designed for representing patterns, while classification requires discriminative features. Another

drawback of eigen features is that they require a training set to construct a subspace.

Gabor filters, which are spatially localized and selective to spatial orientations and scales, are comparable to the receptive fields of simple cells in the mammalian visual cortex [27]. Because of their biological relevance and computational properties, Gabor filters have been adopted in face recognition [26], [47], [49]. Since Gabor filters detect amplitude-invariant spatial frequencies of pixel gray values, they are known to be robust to illumination changes.

Local binary patterns, which have been successfully applied in texture classification [31], are, by definition, invariant under any monotonic transformation of the pixel gray values. However, LBP is significantly affected by nonmonotonic gray value transformations. Unfortunately, in contrast to the flat surfaces where texture images are usually captured, faces are not flat; therefore, nonmonotonic gray value transformations (manifested by shadows and bright spots) occur, and change their positions depending on the illumination. LBP can be expected to have problems dealing with illumination variations in face recognition.

Local features at different spatial scales carry complementary discriminating information, and should be utilized to improve the classification performance. Gabor filters adopted in face recognition are usually computed at five different scales and, therefore, already analyze local features at multiresolution. LBP is also able to represent local features at multiresolution. However, multiresolution LBP has not yet been adopted in face recognition. We shall study how local features at different spatial scales contribute to the final classification.

A separate issue is the role of facial components in helping to identify faces. Psychophysical experiments typically indicate the importance of eyes followed by the mouth and then the nose. Sadr *et al.*, on the other hand, emphasize the importance of eyebrows [39]. It is, therefore, of interest to find out how automated systems rank facial components.

In Section V, we compare experimentally the three local features, investigate the effectiveness of multiscale analysis, and determine which parts of the face are most discriminative.

C. Classification and Combination

In most face recognition applications, there are many classes, but very few training samples per class. It is also common for some classes to have only gallery (reference) samples and no training samples at all. In view of the difficulty of estimating the parameters of sophisticated classifiers, the simple nearest neighbor classifier is usually adopted. The key to classification then is the similarity or distance function. Many similarity measures for both histogram and vector features have been proposed and studied [6], [37]. Instead of duplicating comparative studies of these distance functions, we will pick the ones adopted and recommended by previous comprehensive studies.

Since in local matching approaches, faces are partitioned into local components or local regions, an unavoidable question is how to combine these local features to reach the final classification. Nearly all of the existing local matching methods choose to combine local features before classification. The local features are either simply concatenated into a longer global feature vector or combined linearly by assigning weights to them.

An alternative approach for combining local features is to let them act as individual classifiers and to combine these classifiers for final classification. Many classifier combination methods have been studied in the literature, from static combiners, such as majority vote, sum rule, and Borda count to trainable combiners, including logistic regression and the more recent AdaBoost [14], [19], [22].

Kittler *et al.*compared several static combiners and showed that the Sum Rule outperformed others [22]. In face recognition, the posterior probabilities of the classes required in the sum rule are usually unavailable. if the posterior probabilities are assumed to be proportional to the rankings (rank orders), the sum rule becomes the Borda count. The Borda count has been long studied in political science as an election scheme [7], and has also been applied in pattern recognition by Ho *et al.* [19]. Although thorough theoretical justification in general cases is difficult to reach, empirical studies show that Borda count is a simple and effective method for combining classifiers [19].

Trainable classifier combination schemes require a large number of training samples, which are usually unavailable in face recognition. Therefore, only a few trainable combiners have been studied in face recognition [25], [52]. Even static classifier combination is more computation intensive than feature combination because classification must be performed with each individual local feature before combining them.

Comprehensive comparison of all classifier combination schemes is beyond the scope of this paper. We compare only the feature vector/histogram concatenation (a typical static feature combination method) with the Borda count (a static classifier combination method). This comparison sheds some light on the respective advantages and disadvantages of feature combination and classifier combination. An extreme paradigm of trainable combining methods is feature/classifier *selection*. We shall also briefly examine whether classifier selection benefits the final classification.

IV. FERET AND AR DATABASES

The FERET database [34], [35] is the most widely adopted benchmark for the evaluation of face recognition algorithms. For ease of replicating our experiments and comparing our results to the other reported results, we adopted the FERET database and followed the FERET face recognition evaluation protocol. We briefly introduce the FERET database and the FERET tests. Please refer to [35] or [54, Section 5.1] for detailed information about the database and terminology.

In the FERET database, all frontal face pictures are divided into five categories: Fa, Fb, Fc, Dup1, and Dup2. Fb pictures were taken at the same day as Fa pictures and with the same camera and illumination condition. Fc pictures were taken at the same day as Fa pictures but with different cameras and illumination. Dup1 pictures were taken on different days than Fa pictures but within a year. Dup2 pictures were taken at least one year later than Fa pictures.

Gallery is a set of labeled images of individuals. An image of an unknown face presented to the recognition algorithm is called

a *probe*. The algorithm compares the probe to each of the gallery samples and labels the probe as the most similar gallery sample. *Training* samples are another set of images, which may be used to train the classifiers. In the FERET tests, 1196 Fa pictures are gallery samples. 1195 Fb, 194 Fc, 722 Dup1, and 234 Dup2 pictures are named as Fb, Fc, Dup1, and Dup2 probes, respectively. There are 736 training samples. In the FERET tests, there is only one image per person in the gallery. Therefore, the size of the gallery, 1196, indicates the number of classes. Four series of tests of different degrees of difficulty are conducted, one with each set of probes.

Thanks to the *CSU Face Identification Evaluation System* [6], [55], we can easily collect the metadata of the FERET database and duplicate the FERET tests. FERET provides the positions of two eyes, nose and mouth for every picture. The nose and mouth coordinates of nine pictures are missing, so we manually assign their coordinates.

The *cumulative match curve* is used in the FERET tests to compare the performance of different algorithms. The horizontal axis of the graph is the rank, and the vertical axis is the percentage of the correct matches. The cumulative match curve shows the percentage of correctly identified probes within rank n. The usual definition of *classification accuracy* is the percentage of correct matches of rank 1.

We follow exactly the protocol of the FERET test, except for a slight modification for Dup2 probes. In the original FERET test, the gallery for Dup2 probe was a subset of 864 Fa pictures. In our experiment, we use all 1196 Fa pictures as the gallery.[1] Therefore, our test on Dup2 probes is more difficult than the original FERET test.

In order to test results across different databases, we duplicate most of the important experiments on another well-known database, the AR face database [28]. The AR database, created by Martinez and Benavente, contains over 3000 mug shots of 135 individuals (76 males and 59 females) with different facial expressions, illumination conditions and occlusions. Each subject has up to 26 pictures in two sessions. The first session, containing 13 pictures, named from 01 to 13, includes neutral expression (01), smile (02), anger (03), screaming (04), different lighting (05–07), and different occlusions under different lighting (08–13). The second session exactly duplicates the first session two weeks later.

We used 135 "01" pictures, one from each subject, as gallery. Our experiments are conducted on three probe sets: AR2–3 ("02" and "03" pictures, different expressions), AR5–7 ("05," "06," and "07" pictures, different lighting conditions), and AR14–16 ("14," "15," and "16" pictures, different expressions taken in two weeks later). No fiducial points are provided in the AR database. We manually entered the coordinates of two eyes, nose and mouth for every picture.

V. Experimental Comparative Study

A. Which is the Best Local Feature Representation?

In Section III, we discussed three local feature representations: Eigen (PCA) features, Gabor features and local binary pattern (LBP). Here, we experimentally compare them. We align the faces based on eye and mouth positions (three fiducial points) with a similarity transform. The original image is cropped to 203×251 pixels (Fig. 3). After alignment, a 37×37 local patch for each of four local components (two eyes, nose, and mouth) is cropped. The comparison of local feature representations is conducted on these four 37×37 local patches (components).

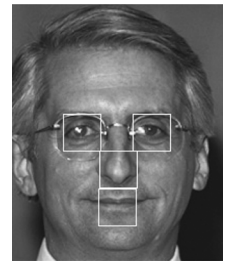


Fig. 3. Aligned face image. The four white boxes indicate the four 37×37 facial components.

In the commonly adopted PCA representation, histogram equalization is first conducted on each block, and then the pixel grey values are normalized to have a mean of zero and a standard deviation of one. Eigen vectors are computed with 736 FERET training samples. 60% of the leading eigen vectors are kept to construct the PCA subspace. The faces are compared with the cosine of the angle between the images after they have been projected into the PCA space and have been further normalized by the variance estimates (i.e, MahCosine reported in CSU face recognition evaluation [6]).

In Gabor feature representation, only the Gabor magnitudes are used because the Gabor phases change linearly with small displacements. Five scales $\lambda \in \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$ (in pixels) and eight orientations $\theta \in \{0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8\}$ of Gabor filters are adopted. We call the eight Gabor filters with the same wavelength (λ) and position (x, y), but different orientations (θ), a *Gabor jet*. Because Gabor magnitude changes only slowly with displacement, we space the Gabor jets of the same wavelength uniformly one wavelength apart. In a 37×37 local area, there are 110 Gabor jets: 64 jets for $\lambda = 4$, 25 jets for $\lambda = 4\sqrt{2}$, 16 jets for $\lambda = 8$, four jets for $\lambda = 8\sqrt{2}$, and one jet for $\lambda = 16$. Corresponding jets are compared with the normalized inner product and the results are combined by the Borda count.

In LBP representation, following the recommendation of Ahonen *et al.* [2], the $LBP_{8,2}^{U2}$ operator in 19×19 windows is adopted. Four windows can be placed in a 37×37 local area and, therefore, generate a histogram of 236 (59×4) bins. The histograms are compared with the chi square statistic.

The FERET test is conducted with the four local facial components, and the classification accuracies of these three local feature representations are reported in Table I. The PCA feature generally yields the worst accuracies. LBP is a very good local feature representation for Fb probes. However, as expected, it performs very badly on Fc probes. Except for illumination changes, Fc probes are nearly identical to their Fa counterparts. The experimental results on Fc probes, especially the nose, which probably is the most vulnerable

[1]We are not aware which 864 Fa pictures were selected.

TABLE I
CLASSIFICATION ACCURACIES (%) OF LOCAL COMPONENTS ON THE FERET DATABASE

	Left Eye				Right Eye			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
PCA	30.4	41.2	9.8	6.8	28.5	25.3	10.1	9.8
Gabor	71.1	81.4	36.7	31.6	71.3	74.2	44.5	50.4
LBP	80.4	38.7	26.2	13.7	79.8	9.3	35.5	41.0
	Nose				Mouth			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
PCA	35.5	10.3	9.3	5.6	21.2	33.0	12.7	6.0
Gabor	88.5	51.0	43.5	45.3	41.1	51.5	32.1	22.2
LBP	81.8	0.5	27.3	12.4	52.8	5.2	33.8	17.5

TABLE II
CLASSIFICATION ACCURACIES (%) OF LOCAL COMPONENTS ON THE AR DATABASE

	Left Eye			Right Eye		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
PCA	44.8	63.5	47.5	45.2	57.5	41.4
Gabor	71.1	90.4	69.2	76.3	94.3	69.7
LBP	80.7	42.5	72.8	80.4	55.1	68.1
	Nose			Mouth		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
PCA	50.7	33.6	50.0	38.9	68.1	53.3
Gabor	91.5	88.9	88.1	46.3	91.4	58.9
LBP	81.1	39.5	79.7	51.1	24.4	61.9

facial component under varying lighting conditions, clearly indicate that LBP is inadequate for nonmonotonic illumination changes. The Gabor features achieve a slightly lower accuracy on Fb probes compared to LBP features, but are much more robust to illumination changes and achieve significantly better performance on the other three kinds of probes than any of the other feature representations.

The experimental results on the AR database, shown in Table II, confirm the findings from the FERET database. If there are no significant lighting condition changes, LBP is an excellent local feature, achieving high accuracies for AR2–3 (same day, different expressions, no lighting changes) and AR14–16 (two weeks later, different expressions, but similar lighting) probes. However, the accuracies drop significantly for AR5–7 probes, where the only variation, compared to AR1 gallery pictures, is lighting conditions. On the other hand, Gabor features are robust to lighting condition changes.

B. Which Facial Component is the Best for Machine Face Recognition?

As mentioned in Section III, psychophysical experiments indicate that eyes or eyebrows are the best facial components for face recognition by humans, followed by the mouth, and then the nose. Our experiments on machine classification rank them differently. In the FERET experiments, for Fb probes, the nose consistently achieves the best classification accuracy with all three methods (although the tip of the nose may be less clearly defined than eye centers and, therefore, likely to be more error prone in manual marking of the fiducial points). Even for Dup1 and Dup2 probes, where the lighting is similar to that of Fa gallery pictures, the nose does not seem to be a bad choice. On the other hand, for Fc probes, the nose consistently results in the worst accuracy with all three methods. Similarly, in the AR experiments, the nose achieves the best classification accuracy for AR2–3 and AR14–16 probes, but worst for AR5–7 probes.

This suggests that human face recognition and current machine face recognition are quite different. It is well known that humans are good at gestalt tasks, apply to recognition a rich set of contextual constraints, and have superior noise filtering abilities [32]. However, our ability to quantify features is limited to approximately seven scales [30]. Machines are the opposite: very bad on using contextual information, and on dealing with noise, such as illumination, pose and expression changes, but able to evaluate facial features in nearly unlimited fine detail. Machines, therefore, favor facial components which contain least noise. Although the nose may not be the most discriminating facial component, it is probably less noisy than eyes and mouth when there is not much illumination change (people cannot easily move their nose after all). Machines can use their excellent feature quantification ability to detect the subtle difference between two noses.

TABLE III
CLASSIFICATION ACCURACIES (%) OF LOCAL REGIONS ON THE FERET DATABASE

	Left Eye Region (67, 125)				Right Eye Region (135, 125)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
PCA	34.0	43.8	**9.4**	6.8	35.5	32.5	13.3	14.5
Gabor	74.1	84.5	38.5	35.9	73.3	76.8	46.7	53.8
LBP	81.5	39.2	28.0	14.5	82.8	**8.8**	37.8	44.4
	Nose Region (101, 153)				Mouth Region (101, 197)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
PCA	42.0	22.7	13.3	6.0	24.0	34.5	13.2	6.0
Gabor	**87.9**	64.4	47.2	48.3	41.6	52.1	34.6	25.6
LBP	84.9	0.5	33.4	17.1	**51.6**	5.7	36.7	20.5

We are surprised to see that machines can achieve 88% accuracy on a face recognition problem of 1196 classes based only on a small 37×37 local patch of the nose (less than 5% of the head region). Although we are not able to find pertinent psychophysical experiments in the literature, it is hard to believe that humans can do as well.

C. Local Facial Components versus Local Facial Regions

As suggested in Section III, it is preferable to compare local regions instead of local components, because region comparison implicitly uses shape information. The following experiment justifies this argument.

After face alignment in the common coordinate system, the average positions of left eye, right eye, nose and mouth fiducial points of 736 FERET training samples are approximately at coordinates (67, 125), (135, 125), (101, 153), (101, 197). We call the local regions centered at these four coordinates left eye region, right eye region, nose region, and mouth region. An experiment similar to that described in Section V-A is conducted on these four local regions. The classification accuracies on the FERET and the AR databases are shown in Tables III and IV, respectively. Again, LBP achieves good performance on Fb, AR2–3 and AR14–16 probes, but the Gabor feature has better overall performance.

Many of the 48 classification results in Table III are significantly better than the corresponding results in Table I, except 4 (bold) that are slightly worse. Similarly, most numbers, except 4 (bold), in Table IV are also larger than the corresponding numbers in Table II. We believe that the improvement is due to the discriminating power of the geometric configuration of facial components or shape features. Comparing corresponding local regions appears to be an effective way to utilize it.

D. Should we Abandon "External" Regions?

As mentioned in Section III, Sinha and Poggio suggested that the human visual system makes strong use of the overall head shape [40]. The local regions centered at (35, 197) and

TABLE IV
CLASSIFICATION ACCURACIES (%) OF LOCAL REGIONS ON THE AR DATABASE

	Left Eye Region (67, 125)			Right Eye Region (135, 125)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
PCA	55.9	76.0	56.9	58.5	71.4	57.5
Gabor	74.8	91.1	72.2	78.9	96.8	72.5
LBP	**78.9**	44.4	74.2	82.2	**54.6**	70.8
	Nose Region (101, 153)			Mouth Region (101, 197)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
PCA	56.7	53.3	54.4	39.6	80.2	58.3
Gabor	92.6	93.3	88.6	**45.2**	93.1	61.9
LBP	84.4	**33.3**	83.9	53.7	29.6	61.9

TABLE V
CLASSIFICATION ACCURACIES (%) OF CHEEK CONTOUR EXTERNAL REGIONS ON THE FERET DATABASE

	Left Cheek Contour (35, 197)				Right Cheek Contour (163, 197)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
PCA	8.2	5.2	1.8	0.9	7.4	7.2	2.5	2.1
Gabor	39.6	36.1	14.8	4.7	36.6	41.2	21.2	12.8
LBP	51.1	11.9	15.7	1.7	47.2	9.8	19.8	5.1

TABLE VI
CLASSIFICATION ACCURACIES (%) OF CHEEK CONTOUR EXTERNAL REGIONS ON THE AR DATABASE

	Left Cheek Contour (35, 197)			Right Cheek Contour (163, 197)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
PCA	14.4	32.6	13.3	17.8	36.3	14.4
Gabor	48.9	52.6	42.5	53.7	72.8	47.8
LBP	57.4	58.5	35.8	64.1	54.1	36.7

TABLE VII
CLASSIFICATION ACCURACIES (%) WITH MULTIRESOLUTION LBP FEATURES ON THE FERET DATABASE

	Left Eye Region (67, 125)				Right Eye Region (135, 125)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
$LBP_{8,2}^{U2}$	81.5	39.2	28.0	14.5	82.8	8.8	37.8	44.4
$LBP_{8,4}^{U2}$	84.0	36.1	30.1	17.5	83.8	5.7	41.3	51.7
Combined	*86.6*	*44.8*	*33.9*	*20.1*	*85.7*	*7.7*	*43.5*	*52.6*
	Nose Region (101, 153)				Mouth Region (101, 197)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
$LBP_{8,2}^{U2}$	84.9	0.5	33.4	17.1	51.6	5.7	36.7	20.5
$LBP_{8,4}^{U2}$	87.5	0.5	36.1	20.5	50.8	10.3	32.8	19.7
Combined	*89.6*	***0.5***	*38.5*	***19.7***	*54.5*	*7.7*	*38.0*	*23.9*

(163, 197) correspond to left and right cheek contour segments. Table V shows the classification accuracies of the two local regions with the three local feature representations on the FERET database.

The most discriminating features in these two "external" local regions are the cheek edges (positions and orientations). Comparing Tables III and V, PCA features do not seem to be able to capture these kinds of edge features effectively, and the accuracies of these two "external" local regions are significantly worse than those of the "internal" local regions. On the other hand, both Gabor and LBP features are sensitive to high gradients and their orientations. When these two features are used, these "external" local regions also achieve good classification accuracies (comparable to that of the mouth region for Fb probes). The accuracies for Dup1 and Dup2 probes gradually drop, indicating that human faces do change their external shapes as time passes, and the longer the time interval, the more significant the change. Similar conclusions can also be drawn from the experimental results on the AR database (Table VI). We still see no justification to ignore "external" local regions, especially when it is known that the gallery pictures and probe pictures were taken at about the same time.

E. Local Features at Different Spatial Scales

We suggested in Section III that local features at different spatial scales carry complimentary discriminating information.

TABLE VIII
CLASSIFICATION ACCURACIES (%) WITH MULTIRESOLUTION GABOR FEATURES ON THE FERET DATABASE

	Left Eye Region (67, 125)				Right Eye Region (135, 125)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
$\lambda=4$	65.4	77.8	27.7	22.6	62.7	70.1	32.7	35.0
$\lambda=4\sqrt{2}$	53.0	62.4	20.5	16.7	52.8	45.9	25.9	32.5
$\lambda=8$	41.7	53.1	11.8	10.3	42.2	38.7	13.7	16.7
$\lambda=8\sqrt{2}$	31.8	33.0	7.8	8.1	32.0	25.8	10.4	14.1
$\lambda=16$	19.5	6.2	3.5	1.3	20.2	2.6	5.1	6.4
Combined	*74.1*	*84.5*	*38.5*	*35.9*	*73.3*	*76.8*	*46.7*	*53.8*
	Nose Region (101, 153)				Mouth Region (101, 197)			
	Fb	Fc	Dup1	Dup2	Fb	Fc	Dup1	Dup2
$\lambda=4$	70.1	58.8	31.7	28.6	35.3	46.9	24.7	12.8
$\lambda=4\sqrt{2}$	67.3	21.1	26.9	15.8	27.4	17.0	17.5	8.5
$\lambda=8$	63.0	7.2	17.3	11.5	22.9	9.3	14.1	9.4
$\lambda=8\sqrt{2}$	50.5	1.0	13.0	9.0	17.8	6.2	11.8	7.3
$\lambda=16$	22.8	0.0	2.5	0.4	7.3	1.5	4.3	0.4
Combined	*87.9*	*64.4*	*47.2*	*48.3*	*41.6*	*52.1*	*34.6*	*25.6*

TABLE IX
CLASSIFICATION ACCURACIES (%) WITH MULTIRESOLUTION LBP FEATURES ON THE AR DATABASE

	Left Eye Region (67, 125)			Right Eye Region (135, 125)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
$LBP_{8,2}^{U2}$	78.9	44.4	74.2	82.2	54.6	70.8
$LBP_{8,4}^{U2}$	80.0	49.4	75.0	81.1	52.6	70.8
Combined	*84.4*	***48.1***	*77.2*	*82.2*	*55.8*	*75.3*
	Nose Region (101, 153)			Mouth Region (101, 197)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
$LBP_{8,2}^{U2}$	84.4	33.3	83.9	53.7	29.6	61.9
$LBP_{8,4}^{U2}$	92.2	43.5	87.2	54.4	30.4	62.8
Combined	***91.9***	***40.7***	*88.1*	*56.3*	*32.8*	*64.4*

TABLE X
CLASSIFICATION ACCURACIES (%) WITH MULTIRESOLUTION GABOR FEATURES ON THE AR DATABASE

	Left Eye Region (67, 125)			Right Eye Region (135, 125)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
$\lambda=4$	69.6	88.4	66.7	78.5	91.9	70.8
$\lambda=4\sqrt{2}$	49.6	50.9	49.7	53.3	63.2	51.9
$\lambda=8$	31.1	37.3	41.1	35.6	44.9	38.3
$\lambda=8\sqrt{2}$	29.3	36.8	30.8	29.6	41.5	37.5
$\lambda=16$	18.9	20.5	25.6	21.5	23.7	25.8
Combined	*74.8*	*91.1*	*72.2*	*78.9*	*96.8*	*72.5*
	Nose Region (101, 153)			Mouth Region (101, 197)		
	AR2-3	AR5-7	AR14-16	AR2-3	AR5-7	AR14-16
$\lambda=4$	81.1	89.1	75.6	41.9	88.1	51.7
$\lambda=4\sqrt{2}$	76.3	72.6	69.4	37.8	70.4	41.1
$\lambda=8$	59.3	51.6	65.0	26.3	49.1	35.3
$\lambda=8\sqrt{2}$	51.9	13.8	52.8	21.9	13.8	31.4
$\lambda=16$	34.8	3.2	35.6	16.3	5.9	20.0
Combined	*92.6*	*93.3*	*88.6*	*45.2*	*93.1*	*61.9*

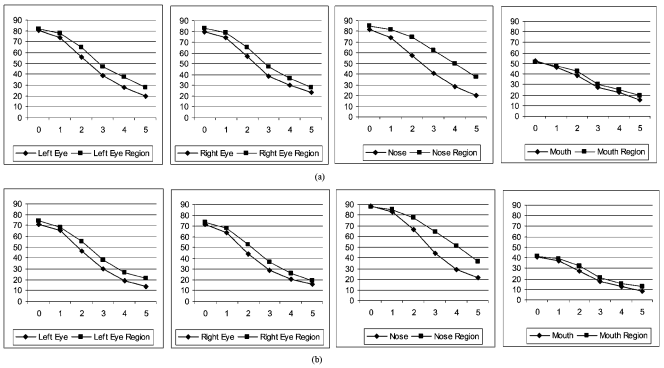


Fig. 4. Recognition accuracies (%) with respect to displacements of fiducial points. (a) with LBP features; (b) with Gabor features.

We test the recognition accuracy of Gabor filters at each of the five usual scales, and in combination. Although LBP is capable of multiresolution analysis, it has not been studied in face recognition yet. In the previous experiments, we use the $LBP_{8,2}^{U2}$ operator. We now compare it to the $LBP_{8,4}^{U2}$ operator and the combination of $LBP_{8,2}^{U2}$ and $LBP_{8,4}^{U2}$. (The two histograms are simply concatenated). Tables VII and IX show the recognition

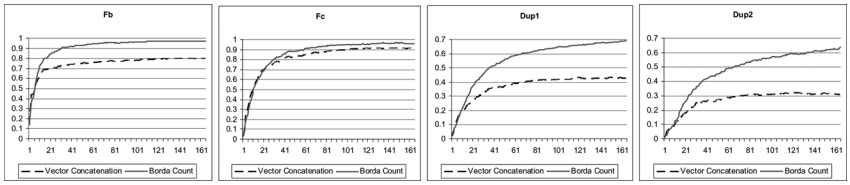


Fig. 5. Comparison between feature vector concatenation and Borda count combination with 165 Gabor jets on the FERET database.

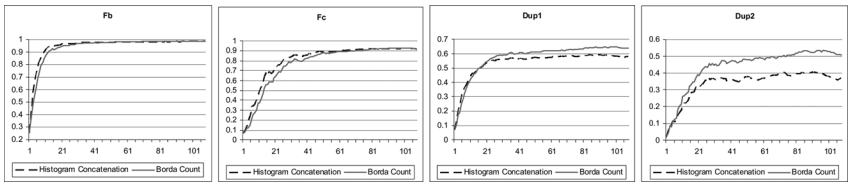


Fig. 6. Comparison between histogram concatenation and Borda count combination with 108 LBP histograms on the FERET database.

accuracies of $LBP_{8,2}^{U2}$, $LBP_{8,4}^{U2}$ and the two together on four local regions for the FERET and the AR databases, respectively. Tables VIII and X show the recognition accuracies of Gabor features with $\lambda = 4$, $\lambda = 4\sqrt{2}$, $\lambda = 8$, $\lambda = 8\sqrt{2}$, $\lambda = 16$, and the combination of these five on four local regions for the FERET and the AR databases.

Combining multiscale Gabor features or multiresolution LBP features generally achieves higher classification accuracy than the individual feature sets, with the exception of seven cases in Tables VII and IX (bold numbers). This confirms that local features at different scales do carry complementary discriminating information; therefore, multiresolution analysis is recommended.

Before we leave this section, we add an explanation of the seemingly anomalous Gabor feature results. In Tables VIII and X, it seems that the longer the wavelength, the worse the performance. This is actually not the case; there are merely fewer Gabor filters in a 37×37 local region when their wavelength is longer. For example, there is only 1 Gabor jet for $\lambda = 16$, but 64 jets for $\lambda = 4$. Combining more jets leads to higher accuracy.

F. Sensitivity to the Displacements of Fiducial Points

Local matching, like all face recognition methods, depends on the accuracy of the location of the fiducial points. In order to test sensitivity to displacements of the fiducial points, we add Gaussian random noise with mean 0 and standard deviation 1, 2, 3, 4, or 5 to the x-y coordinates of the FERET provided eye and mouth positions. The standard deviations of the perturbations are $\sqrt{2}$, $2\sqrt{2}$, $3\sqrt{2}$, $\lambda = 4\sqrt{2}$, and $5\sqrt{2}$, or about 2%, 4%, 6%, 8%, and 10% of the distance between the eyes. Fig. 4 shows the inverse relationship of the recognition accuracy with the perturbation of the LBP and Gabor features on Fb probes.

Fig. 4 also compares the performance of local components and local regions. In general, the accuracies of local components drop faster than those of their corresponding local regions. This is due to the fact that corresponding local regions are aligned based on all fiducial points, which effectively attenuates errors introduced by individual fiducial points. In practical applications, we, therefore, recommend detecting as many fiducial points as possible, and then aligning the faces using all available points. This not only utilizes the shape information (cf. Section V-C), but is also more robust to the displacements of individual fiducial points.

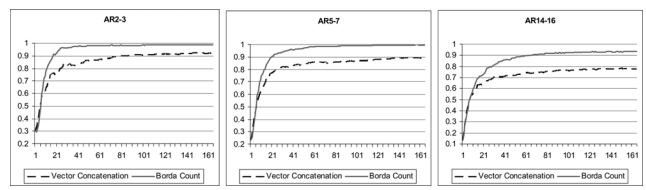


Fig. 7. Comparison between feature vector concatenation and Borda count combination with 165 Gabor jets on the AR database.

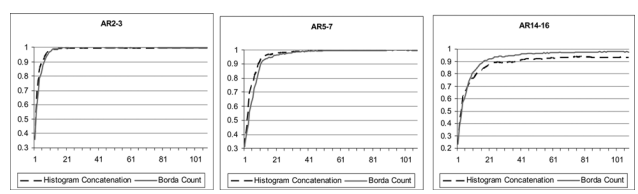


Fig. 8. Comparison between histogram concatenation and Borda count combination with 108 LBP histograms on the AR database.

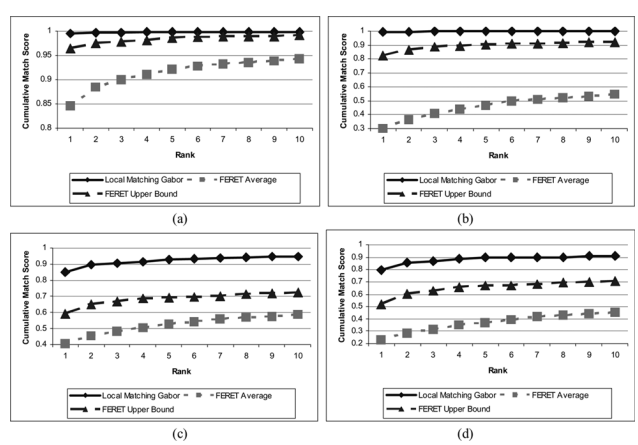


Fig. 9. Cumulative match curves on FERET (a) Fb, (b) Fc, (c) Dupl, and (d) Dup2 probes. FERET average and upper bound performance is estimated from the printouts of [35, Figs. 7 and 8].

(a) (b) (c) (d) (e) (f) (g)

Fig. 10. Errors made by local matching method with Gabor features on (a)–(f) Fb and (g) Fc probes. Left is the gallery Fa picture; right is the probe. The rankings the algorithm assigns to the correct candidates are (a) 2, (b) 3, (c) 2, (d) 20, (e) 2, (f) 21, and (g) 3. FERET-provided eye positions for (a) and (b) are not very accurate. Most of the errors are due to significant pose and expression changes.

G. Feature Combination versus Classifier Combination

In this section, we compare a static feature combination method (feature vector/histogram concatenation) with a static classifier combination method (the Borda count).

In order to combine more local regions, we now conduct the experiment with the whole face area. We use Gabor jets with $\lambda = 8\sqrt{2}$ to conduct this experiment. 165 Gabor jets are placed uniformly on the face area, at one wavelength intervals. Each Gabor jet produces a feature vector of eight Gabor magnitudes. We randomly shuffle the order of all available 165 Gabor jets. For feature combination, we concatenate the feature vectors and then compare them with the MahCosine distance measure. For classifier combination, we combine them with the Borda count. The experiment is repeated five times. Fig. 5 shows the average classification accuracy with respect to the number of Gabor jets combined for all four probes. Observing carefully the initial parts of the curves, we can see that feature concatenation is (slightly) better than Borda count when only a few Gabor jets are combined. However, its performance is soon saturated and outperformed by the Borda count.

We conducted a similar experiment with LBP features to test histogram feature combination. The face area is divided into 108 local windows of 19 by 19 pixels. $LBP_{8,2}^{U2}$ is applied on each window and produces a histogram feature of 59 bins. The order of all 108 histograms is randomly shuffled. For feature combination, we concatenate the histograms and then compare them with the chi square statistic. For classifier combination, we combine them with the Borda count. The experiment is repeated five times, yielding similar results (Fig. 6) as the Gabor features. The same experiments are duplicated on the AR database, and similar results are obtained as shown in Figs. 7 and 8.

The saturation of feature concatenation is due to the curse of dimensionality. Borda count classifier combination is much less vulnerable to this problem, especially for the difficult Dup1 and Dup2 probes. On the other hand, Borda count classifier combination does increase the computational load.

H. Overall Performance Comparison

Guided by the results of the above comparative studies, a complete local region matching face recognition system can be implemented by integrating the best option in each step. To make a fair comparison to the FERET evaluation and other reported performances, only the two FERET-provided eye positions are used for aligning faces. The original face image is similarity transformed to place the two eye centers at (67, 125) and (135, 125), respectively. The image is cropped to 203×251 pixels. We again space the Gabor jets uniformly one wavelength apart. This produces 4172 local regions in the available face area at five scales: 2420 at $\lambda = 4$, 1015 at $\lambda = 4\sqrt{2}$, 500 at $\lambda = 8$, 165 at $\lambda = 8\sqrt{2}$, and 72 at $\lambda = 16$. These Gabor jets are compared with normalized inner product and combined with Borda count.

We duplicate the FERET test to evaluate the algorithm's performance. Twelve algorithms were reported in the FERET test, most of them holistic. Fig. 9 shows the cumulative match curves. Our method (local matching Gabor) achieves significantly more accurate classification than the best of the 12 participating algorithms.

Fb probes are considered as the easiest probes. The FERET average and upper bound accuracies are about 85% and 96%. Our algorithm achieves 99.5% accuracy. The six errors (out of 1195 probes) are shown in Fig. 10(a)–(f). Three of them are just barely missed by the algorithm (the correct candidates are in the second place).

Fc probes test the algorithm's robustness against illumination variations. Most algorithms in the FERET test do not compensate well: the average accuracy is only about 30%. The upper bound accuracy, achieved by EBGM algorithm, is about 82%. Our algorithm achieves 99.5% accuracy. The only error (out of 194 probes) is shown in Fig. 10(g), and the correct candidate is in the third place. We used Gabor jets without any preprocessing for illumination compensation. This suggests that Gabor filters are robust against indoor illumination changes.

Dup1 probe and Fa gallery pictures were taken at different days but within a year. The FERET average and upper-bound accuracies are about 40% and 59%, respectively. Our method achieves 85.0% accuracy (108 errors out of 722 probes).

Dup2 probe and Fa gallery pictures were taken at least one year apart, and were considered as the most difficult test. The FERET average and upper-bound accuracies are about 23% and 52%, respectively. Our method achieves 79.5% accuracy (48 errors out of 234 probes) in a more difficult experiment than the original FERET test.

We also collected all the classification accuracies reported by recent local matching face recognition methods on the FERET probes and listed them in Table XI. Similar to our findings, the accuracies on the Fc probes of the LBP method reported in [2] are relatively low, which is due to the fact that LBP features are inadequate in dealing with nonmonotonic illumination changes. Our method achieves much higher accuracies, especially on more difficult Dup1 and Dup2 probes.

I. Component Classifier Selection

There are 4172 component classifiers (Gabor jets) in the above Gabor local matching method. It is natural to ask how many component classifiers we really need and whether we can

TABLE XI
CLASSIFICATION ACCURACIES (%) OF SEVERAL LOCAL MATCHING METHODS ON THE FERET PROBES

Methods		Fb	Fc	Dup1	Dup2
EBGM [47]		95	82	59	52
LBP [2]	Non-weighted	93	51	61	50
	Weighted	97	79	66	64
LGBPHS [53]	Non-weighted	94	97	68	53
	Weighted	98	97	74	71
Best Results of [52]		97	-	-	-
Best Results of [25]		98	-	-	-
Ours		99.5	99.5	85.0	79.5

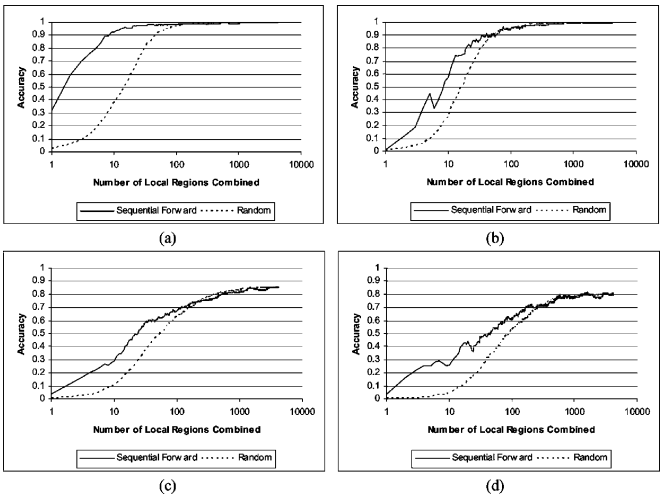


Fig. 11. Accuracies on (a) Fb probes, (b) Fc probes, (c) Dup1 probes, and (d) Dup2 probes with respect to the number of Gabor jets combined.

select a set of "optimal" component classifiers to achieve even better performance.

We tried a simple sequential forward wrapper approach to select a set of Gabor jets. Initially, the Gabor jet with the highest classification accuracy is selected. At each step, we select the Gabor jet which, when combined with the already-selected jets, achieves the highest classification accuracy. The process is iterated until all the Gabor jets are ordered. The ordering is conducted on the FERET-provided 736 training samples.

The ordered Gabor jets are then combined one by one to classify the probes. Fig. 11 plots the accuracies on four kinds of probes with respect to the number of Gabor jets combined. For comparison, the accuracies with respect to combining randomly ordered Gabor jets (averaged over five random trials) are also plotted.

Evidently, the number of Gabor jets required for good classification depends on the difficulty of the task. Dup1 and Dup2 probes are much more difficult than Fb and Fc probes. Accordingly, Fb and Fc probes require combining fewer than 200 Gabor jets to reach the peak accuracy, while Dup1 and Dup2 probes require about 1000 Gabor jets.

The sequential forward algorithm does help to find many "good" Gabor jets at the beginning. This could be very useful for practical applications, where computation time is also a critical factor. However, the sequential forward curve of Fb probes is quite different from the curves of the other three probes, and only for Fb probes does sequential forward ordering reach peak accuracy sooner than random selection. Furthermore, after combining about 150 local regions, the accuracies of the randomly selected local regions are about the same as those of the sequential forward selected local regions. Because many FERET training samples are Fa and Fb pictures, it is not surprising that sequential forward ordering is better than random ordering for Fb probes, but does not generalize well to other probes.

There are of course other more sophisticated feature selection techniques, e.g., floating algorithm, but with only a few (and not necessarily representative) training samples, selecting a subset of "optimal" component classifiers, applicable to all probes, is not an easy task. An alternative approach is to select different subsets for different probes. For example, when the lighting condition of the probe is significantly different from that of the gallery samples, the nose region should be avoided or at least weighted less than the eye regions. Probe-dependent feature selection has not been explored much [12]. In any case, trainable local region selection or weighting is certainly valuable and worth careful study.

VI. CONCLUSION

Although several psychophysical experiments suggest that human face recognition is a holistic process, we believe that at the current state of the art, local region matching is more appropriate for machine face recognition. Compared to holistic methods, local matching methods can extract facial features with different levels of locality and take advantages of fine feature quantification. On the other hand, local matching does require relatively high resolution images. There is certainly no point to partitioning a 7×10 pixel face image into local regions, even though such low resolution pictures are adequate for human recognition of familiar faces [50]. Our study is the first comprehensive comparative analysis of the details of face recognition through localized matching.

Among the three local feature representations, the eigen (PCA) feature cannot be recommended. Local Binary Pattern is a good local feature, but it is inadequate for nonmonotonic illumination changes, which often appear in facial regions such as the nose. The Gabor jet is our choice for local feature representation because of its robustness to illumination variations.

Machine discrimination based on only a small portion of the face area is surprisingly good, far better than human. However, machine filtering of "noise," such as pose, expression, and illumination, is still rudimentary. Although psychophysical experiments suggest that the configuration of facial components is at least as important as their appearance in human face recognition [51], current programs concentrate more on appearance than on facial component configuration. Our comparative study demonstrates that comparing corresponding local regions instead of corresponding local components is an effective way of exploiting variations of facial component configuration (although it may not be what the human vision system does).

Spatial multiresolution analysis may be part of human vision: Gabor filters have been shown to be comparable to the receptive fields of simple cells in the mammalian primary visual cortex. Although this has not been emphasized in the literature on automated face recognition, we demonstrate that multiscale analysis does improve the overall recognition accuracy.

Simply concatenating all local features is prone to the curse of the dimensionality. Classifier combination, such as the Borda

count, alleviates this problem, especially for difficult probes, at the expense of additional computation. Trainable combination methods require many training samples, which are seldom available in face recognition. We showed that a simple sequential forward selection method helps to identify some discriminating local regions. However, it exhibits overfitting when the training samples are not representative, and it is not very successful when accuracy is the primary goal. Feature selection or weighting with a limited number of training samples remains an interesting and important research topic in machine face recognition.

We built a local region matching face recognition system by assembling the best choices at each step. Without training and without any illumination compensation and parameter tuning, our system achieves superior performance on the FERET tests: near perfect (99.5% accuracy) on Fb and Fc probes, 85.0% accuracy (26% higher than the FERET upper bound performance) on Dup1 probes, and 79.5% accuracy (28% higher than the FERET upper bound performance) on Dup2 probes. This, to our knowledge, is the best performance ever reported in the literature.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments.

REFERENCES

[1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.

[2] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," *ECCV*, pp. 469–481, 2004.

[3] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Jun. 2002.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[5] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions?," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 1–16, 1998.

[6] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper, *The CSU Face Identification Evaluation System User's Guide: Version 5.0*. Denver, CO: Comput. Sci. Dept., Colorado State Univ., 2003.

[7] D. Black, *The Theory of Committees and Elections*, 2nd ed. London, U.K.: Cambridge Univ. Press, 1963.

[8] S. C. Chen and Y. L. Zhu, "Subpattern-Based principle component analysis," *Pattern Recognit.*, vol. 37, no. 5, pp. 1081–1083, 2000.

[9] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[10] I. Craw, N. Costen, Y. Kato, G. Robertson, and S. Akamatsu, "Automatic face recognition: Combining configuration and texture," in *Proc. Int. Workshop Automatic Face and Gesture Recognition*, 1995, pp. 53–58.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[12] H. K. Ekenel and R. Stiefelhagen, "Block selection in the local appearance-based face recognition scheme," presented at the CVPR Biometrics Workshop, 2006.

[13] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Amer.*, vol. 14, pp. 1724–1733, 1997.

[14] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Statist.*, vol. 28, no. 2, pp. 337–374, 2000.

[15] X. Geng and Z.-H. Zhou, "Image region selection and ensemble for face recognition," *J. Comput. Sci. Technol.*, vol. 21, no. 1, pp. 116–125, 2006.

[16] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular PCA approach," *Pattern Recognit. Lett.*, vol. 25, no. 4, pp. 429–436, 2004.

[17] R. Gross, I. Matthews, and S. Baker, "Appearance-Based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 449–465, Apr. 2004.

[18] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: Component-based versus global approaches," *Comput. Vis. Image Understand.*, vol. 91, no. 1, pp. 6–12, 2003.

[19] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.

[20] T.-K. Kim, H. Kim, W. Hwang, and J. Kittler, "Component-Based LDA face description for image retrieval and MPEG-7 standardisation," *Image Vis. Comput.*, vol. 23, pp. 631–642, 2005.

[21] M. Kirby and L. Sirovich, "Application of the Karhunen–Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[22] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[23] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic face identification system using flexible appearance models," *Image Vis. Comput.*, vol. 13, no. 5, pp. 393–401, 1995.

[24] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[25] S. C. Liao, Z. Lei, X. X. Zhu, Z. N. Sun, S. Z. Li, and T. Tan, "Face recognition using ordinal features," in *Proc. Int. Conf. Biometrics*, 2006, pp. 40–46.

[26] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[27] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, no. 11, pp. 1297–1300, 1980.

[28] A. M. Martinez and R. Benavente, The AR Face Database 1998, CVC Tech. Rep. #24.

[29] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.

[30] G. Miller, "The magical number seven plus or minus Two; some limits on our capacity for processing information," *Psych. Rev.*, vol. 63, pp. 81–97, 1956.

[31] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[32] S. E. Palmer, *Vision Science, Photons to Phenomenology*. Cambridge, MA: MIT Press, 1999.

[33] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and modular eigenspaces for face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 84–91.

[34] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput. J.*, vol. 16, no. 5, pp. 295–306, 1998.

[35] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[36] P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, "Face recognition vendor test 2002: Evaluation report," NISTIR 6965, 2003.

[37] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *Comput. Vis. Image Understand.*, vol. 84, no. 1, pp. 25–43, 2001.

[38] J. Ruiz-del-Solar and P. Navarrete, "Eigenspace-Based face recognition: A comparative study of different approaches," *IEEE Trans. Syst., Man, Cybern. C, Cybern.*, vol. 35, no. 3, pp. 315–325, 2005.

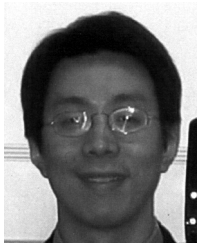
[39] J. Sadr, I. Jarudi, and P. Sinha, "The role of eyebrows in face recognition," *Perception*, vol. 32, pp. 285–293, 2003.

[40] P. Sinha and T. Poggio, "I think I know that face …," *Nature*, vol. 384, pp. 404–404, 1996.

[41] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

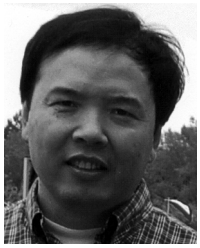
[42] K. Tan and S. Chen, "Adaptively weighted sub-pattern PCA for face recognition," *Neurocomputing*, vol. 64, pp. 505–511, 2005.

[43] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, pp. 1725–1745, 2006.
[44] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 13, no. 1, pp. 71–86, 1991.
[45] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual features of intermediate complexity and their use in classification," *Nature Neurosci.*, vol. 5, no. 7, pp. 682–687, 2002.
[46] N. Vaswani and R. Chellappa, "Principal components null space analysis for image and video classification," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1816–1830, Jul. 2006.
[47] L. Wiskott, J.-M. Fellous, N. Kruger, and C. v. d. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.
[48] C. Xiang, X. A. Fan, and T. H. Lee, "Face recognition using recursive fisher linear discriminant," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2097–2105, Aug. 2006.
[49] X. Xie and K.-M. Lam, "Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2481–2492, Sep. 2006.
[50] A. Yip and P. Sinha, "Role of color in face recognition," *Perception*, vol. 31, pp. 995–1003, 2002.
[51] A. W. Young, D. Hellawell, and D. C. Hay, "Configurational information in face perception," *Perception*, vol. 16, pp. 747–759, 1987.
[52] L. Zhang, S. Z. Li, Z. Y. Qu, and X. Huang, "Boosting local feature based classifiers for face recognition," presented at the IEEE Conf. Computer Vision and Pattern Recognition Workshop on Face Processing in Video, Washington, DC, 2004.
[53] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. ICCV*, 2005, pp. 786–791.
[54] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, 2004.
[55] 2007 [Online]. Available: http://www.cs.colostate.edu/evalfacerec/.



Jie Zou (M'00) received the B.S. and M.S. degrees from Shanghai Jiao Tong University, China, in 1994 and 1997, respectively, and the Ph.D. degree in computer engineering from Rensselaer Polytechnic Institute in 2004.

Since 2005, he has been with the Lister Hill National Center for Biomedical Communications, National Library of Medicine. His research interests are in the areas of information retrieval, pattern recognition, image processing, and computer vision.



Qiang Ji (SM'05) received the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 1998.

He is currently an Associate Professor with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute (RPI), Troy, NY. Prior to joining RPI in 2001, he was an Assistant Professor with the Department of Computer Science, University of Nevada, Reno. He also held research and visiting positions with Carnegie Mellon University, Pittsburgh, PA; Western Research Company; and the U.S. Air Force Research Laboratory. His research interests are in computer vision, probabilistic reasoning using probabilistic graphical models for decision making, situation awareness, and human computer interaction. He has published over 100 papers in peer-reviewed journals and conferences.

Dr. Ji is an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION, *Pattern Recognition Letters*, and the *Image and Vision Computing*.

George Nagy (F'03) received the B.Eng. and M.Eng. degrees from McGill University, Montréal, QC, Canada, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 1962 (on neural networks).

He was a member of the research staff at the IBM T. J. Watson Research Center, Yorktown Heights, NY, then Professor of computer science at the University of Nebraska, Lincoln. Since 1985, he has been a Professor of computer engineering at the Rensselaer Polytechnic Institute, Troy, NY. In addition to pattern recognition, document image analysis, and character recognition, his interests include computer vision, solid modeling, and finite-precision spatial computation.