

Learning the best subset of local features for face recognition

Berk Gökberk*, M. Okan İrfanoğlu, Lale Akarun, Ethem Alpaydın

Department of Computer Engineering, Boğaziçi University, TR-34342, Bebek, Istanbul, Turkey

Received 28 October 2005; received in revised form 14 June 2006; accepted 6 September 2006

Abstract

We propose a novel, local feature-based face representation method based on two-stage subset selection where the first stage finds the informative regions and the second stage finds the discriminative features in those locations. The key motivation is to learn the most discriminative regions of a human face and the features in there for person identification, instead of assuming a priori any regions of saliency. We use the subset selection-based formulation and compare three variants of feature selection and genetic algorithms for this purpose. Experiments on frontal face images taken from the FERET dataset confirm the advantage of the proposed approach in terms of high accuracy and significantly reduced dimensionality.

© 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Face recognition; Face representation; Gabor wavelets; Feature subset selection; Genetic algorithms

1. Introduction

Face recognition has proved to be a difficult problem in computer vision. The main reason for this is that intra-personal variations caused by facial expressions, view point changes, and illumination variations are significant when compared to inter-personal variations. Many researchers have therefore focused on face representation techniques that are invariant to some of these variations [1,2]. These can be grouped into two, as holistic and local feature-based: In the first type, faces are represented as a whole and statistical techniques are used to extract features from faces [3–6]. The second type depends on the localization of salient facial features such as the eyes and the mouth [7–9]. There are also hybrid approaches which incorporate complementary knowledge from both [10]. Our work in this paper belongs to the second type, with the distinction that the salient local regions are not predicted but are learned from data.

The main idea in a feature-based face representation scheme is the extraction and analysis of local facial features. Salient facial features are first found and then used to code a face. Coding is generally carried out using geometric relationships between these points and extracting local image descriptions around these points. Among different alternatives, 2D Gabor-like filters are found to be very suitable as local descriptors because of their robustness against translation, rotation, and scaling [7,11,12]. 2D Gabor wavelets are selective to different orientations and spatial frequencies. Typically, features extracted by 2D Gabor wavelets have a very large dimensionality. It is therefore essential to analyze the contribution of each feature component to the recognition performance. Important parameters of 2D Gabor wavelets are: (1) spatial location of the kernel in the image; (2) kernel orientation; and (3) spatial kernel frequency.

Several studies have concentrated on examining the importance of the Gabor kernel parameters for face analysis. These include: the weighting of Gabor kernel-based features using the simplex algorithm for face recognition [13], the extraction of facial subgraph for head pose estimation [14], the analysis of Gabor kernels using univariate statistical techniques for discriminative region finding [15],

* Corresponding author. Tel.: +90 212 359 7183.

E-mail addresses: gokberk@boun.edu.tr (B. Gökberk),
irfanogl@boun.edu.tr (M.O. İrfanoğlu), akarun@boun.edu.tr (L. Akarun),
alpaydin@boun.edu.tr (E. Alpaydın).

the weighting of elastic graph nodes using quadratic optimization for authentication [8], the use of principal component analysis (PCA) to determine the importance of Gabor features [16], boosting Gabor features [17] and Gabor frequency/orientation selection using genetic algorithms [18].

In almost all previous studies, we see two fundamental assumptions: first, the contribution of each feature dimension is analyzed independently of others (*independence assumption*); and second, Gabor kernel placement over the face region is strongly affected by prior knowledge (*saliency assumption*). Placing the kernel at visually salient facial points, e.g., eyes, mouth, etc. is one of the frequently used methods. The first assumption of independence of features is not valid, and one should incorporate more complex methodologies to analyze the relationship between the features. Moreover, the effectiveness of the fiducial points should also be studied systematically, and a better solution would be to learn these locations from given training data for a given task. In our previous work, we have analyzed topographically important facial locations for both pose estimation and identity recognition [19], and used feature selection methods to extract optimal local image descriptor parameters for frontal face recognition [20]. We have also used such features to calculate bottom-up saliency in a selective attention-based face recognizer [21].

In this work, our aim is to relax the *independence* and *saliency* assumptions for face recognition by reformulating the optimal Gabor basis extraction problem as a feature subset selection problem. Doing this, we allow our approach to detect more complex relationships and correlations between feature dimensions, thus extracting a near-optimal Gabor basis. For this purpose, we have devised a two-stage subset selection mechanism: In the first stage, a genetic algorithm is used to find the most informative facial locations. In the second stage, a floating search method is used to learn the individual parameters, that is, frequency and orientation, of Gabor wavelet-based local descriptors.

The remainder of this paper is organized as follows: Section 2 describes the proposed approach and experimental results, including a sensitivity analysis, are presented in Section 3. We conclude and discuss future research directions in Section 4.

2. Proposed approach: learning the best features

We have designed a local feature-based face representation scheme for recognition. Multi-frequency and multi-orientation 2D Gabor wavelets are used as local feature extractors [7,11]. In order to find an efficient representation, these local image descriptors should be placed carefully over the face region. Moreover, depending on the locations of these image descriptors, useful frequencies and orientations should be found since specific parts of a face contain high frequency information (e.g., eyes) and some other parts contain low frequency information (e.g., cheeks). Orientation

selectivity also depends on the location of the Gabor kernels. Therefore, we consider the problem by dividing it into two consecutive stages: first, we determine the topographically important face regions, and then, we determine the optimal frequency and orientation parameters of 2D Gabor feature extractors at these locations. The order of selection is not important, and they are expected to converge to the same subset. However, selecting frequencies and orientations (*F/O*) first in the entire face region is not a good idea, since some facial regions prefer particular *F/O*s and others may prefer completely different *F/O*s. The overall diagram of the proposed approach is shown in Fig. 1.

Learning discriminative facial locations and obtaining optimal local feature extractor parameters is formulated as a feature subset selection problem. In feature selection, the aim is to select a subset from a given set such that the classification accuracy of the selected subset is maximized [22]. We use sub-optimal sequential and parallel subset selection algorithms in our system. As sequential selection algorithms, best-individual selection (BIF) algorithm, sequential forward selection (SFS), and sequential floating forward search (SFFS) algorithm are used [22]. BIF approach simply selects the best k features and performs well only if each local descriptor contributes independently to the discrimination performance. In order to consider complex feature dependencies, SFS and SFFS algorithms are used in our system. In SFS, at each step, we add the most significant feature with respect to the previously selected subset. SFFS algorithm takes this idea one step further by backtracking to remove the least useful features from an existing feature subset to overcome the nesting effect. As a parallel subset selection method, we use a genetic algorithm where a chromosome represents a subset and a chromosome's fitness is calculated according to the classification performance of its subset.

2.1. Kernel location selection

We have designed three different methods to learn the important facial locations: *lattice-based sampling*, *landmark-based sampling*, and *dense sampling*. In the *lattice* and *landmark-based* methods, sparse sampling of Gabor kernels at several locations were performed. Figs. 2a and b show these sampling types, respectively. In *lattice-based* sampling (Fig. 2a), we place a rectangular lattice of size $N \times N$ over the central part of the face region. At each point in the lattice, M different Gabor kernel convolutions are carried out composed of v different frequencies and u different orientations with $M = u \times v$. The concatenation of the magnitudes of the complex outputs of Gabor convolutions forms a feature vector for the whole face.

In *landmark-based sampling*, salient facial landmarks are used. We have identified $S = 30$ salient locations over the face region commonly used by researchers as seen in Fig. 2b. The aim of constructing such a sampling scheme is

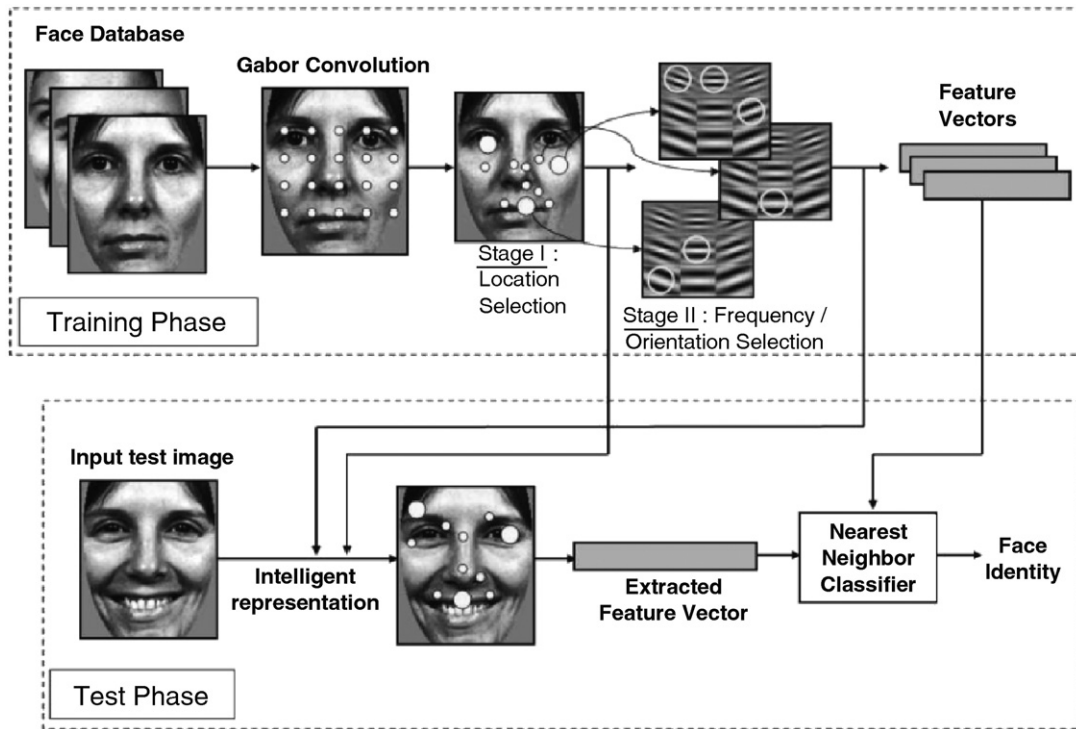


Fig. 1. Overall diagram of our approach.

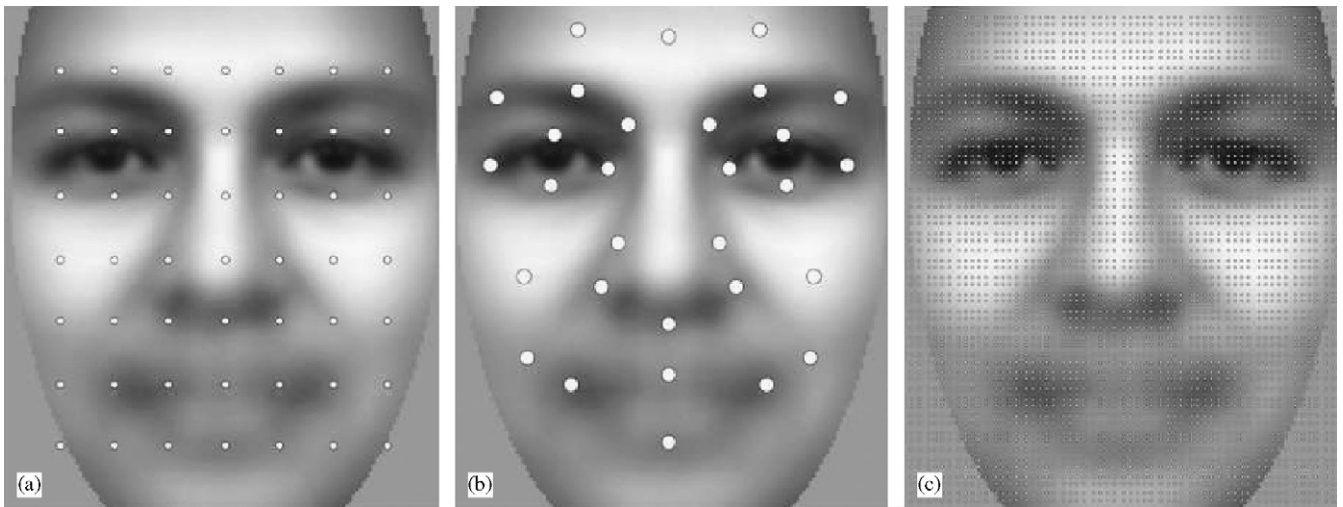


Fig. 2. Different sampling types shown over the mean image: (a) lattice-based sampling; (b) landmark-based sampling; and (c) dense sampling.

to test our prior information as to whether these points are really discriminative and to determine whether these points are really important for recognition. With *lattice-based* and *landmark-based* sampling, in order to determine the important locations among these points, we perform BIF, SFS, and SFFS-based subset selection. We consider each feature vector of the i th face location as a single dimension. Thus we have a search dimensionality of size $N \times N$ for lattice-based sampling and S for landmark-based sampling. As a

stopping condition, we have defined the cardinality of the resulting subset to a value d .

Dense sampling uses full convolutions at each pixel as shown in Fig. 2c. This dramatically enlarges the cardinality of the feature set. SFS- and SFFS-based algorithms become infeasible for this search space. In order to cope with this problem, we have employed a GA-based subset selection algorithm. In our GA formulation, each gene in a chromosome represents the position of a Gabor kernel. We define

the dimensionality of the selected subset as d ; so, each chromosome consists of d genes. The fitness function depends on the classification accuracy of the selected subset.

2.2. Kernel parameter selection

Once we find the *locations* of features, we determine the most useful orientations and frequencies of the Gabor kernels at the selected locations, using SFFS. The first stage returns a subset X_{loc} of dimensionality $d \times M$. In the second stage of frequency and orientation selection, we search for a subset X_{fo} of X_{loc} where $|X_{fo}| \ll |X_{loc}|$. Note that each dimension corresponds to a specific frequency and orientation pair of the outputs of a previously selected Gabor kernel at some specific location. Again, the feature selection criterion in SFFS is the supervised classification accuracy of the selected subset.

3. Experimental results

In our experiments, we have used a subset of the FERET face database [23] which contains subjects having four images. The database contains normalized frontal images of 146 subjects. Each subject has four gray scale images of resolution 150×130 . Faces contain facial expression and illumination variations. Each session contains two training, one validation, and one test image and therefore there are six possible experimental sessions: $\{S_1, S_2, \dots, S_6\}$. After training with two images per person, the validation set is used to determine when to stop training (that is, adding features) and the test set is used to report the final accuracy. The classifier is the nearest neighbor classifier. We use 6-fold paired t -test to compare the accuracies for statistically significant difference. S_2 and S_5 are special: in the selected subset of frontal FERET face images, some subjects have two images with eyeglasses and two images without eyeglasses. In configurations S_2 and S_5 , the training set contains either two images of a subject with eyeglasses, or without eyeglasses. Therefore, the validation and test sets are different from the training set because of the presence/absence of the eyeglasses. This property makes these configurations very challenging since validation and test sets are very different from training sets.

3.1. Kernel location selection

3.1.1. Lattice-based sampling

First experiments on kernel location selection were carried out using the lattice-based sampling method. A 7×7 lattice is positioned over the face. Gabor kernels are 15×15 pixels wide, and contain five frequencies and eight orientations [7]. At each lattice point i , we have extracted the local feature vector, $v_i = \{m_{0,0}, m_{0,1}, m_{0,2}, \dots, m_{p,q}\}$, $p = 0, \dots, 4$; $q = 0, \dots, 7$ of dimensionality $|v_i| = 40$ using multi-frequency and multi-orientation Gabor kernels where

$m_{i,j}$ denotes Gabor convolution magnitudes. Combining all local feature vectors, we obtain a global feature vector, $\Phi = \{v_1, v_2, \dots, v_k\}$ where $k = 49$ for lattice-based sampling. The cardinality of Φ is $|\Phi| = 49 \times 40$. Let Φ_{LOC} be the selected subset of dimensionality d , $\Phi_{LOC} = \{v_i : i \in 1, \dots, k\}$, where d is set to 15 in our experiments. Notice that we treat each local feature vector v_i as a single feature dimension in the subset selection formalism. Fig. 3 shows the selected kernel locations in the subset Φ_{LOC} graphically. The top row shows the 15 selected kernel locations using BIF-based subset selection for each of the six different configurations. These locations are found using the validation sets in all experiments. Similarly, the second and third rows show the most important kernel locations for SFS and SFFS algorithms, respectively.

Looking at the BIF results, we see that most of the kernels are located at the upper part of the face, and are highly symmetric, except in sessions S_2 and S_5 . These results comply with the findings of previous works and are expected. Eyes, eyebrows, and forehead seem to have more discriminating information. The bias towards the upper facial locations is due to the significant expression variations around the mouth region in the dataset. The eyeglasses problem in S_2 and S_5 leads to the positioning of some kernels around the mouth region.

The symmetry property is not present in SFS and SFFS, since they evaluate the importance of a new candidate feature with respect to the existing subset, and take feature dependencies into account. This is an advantage of SFS and SFFS over BIF: They avoid redundant, symmetric features. As with BIF, points are largely scattered outside the variable mouth region and we see the importance of the face outline in sessions S_1, S_3, S_4 , and S_6 . Visual inspection reveals that kernel locations found by SFS and SFFS approaches are very close. Though locations vary between configurations, for the same configuration, locations found by SFS and SFFS tend to coincide.

Classification accuracies of lattice-based sampling approach for each experimental session are shown in Table 1. For each subset selection method, test set results are presented together with their mean, standard deviation and median statistics. We include the median statistic in order to stress the outlier effect of sessions S_2 and S_5 , and we also report statistics over four sessions, excluding S_2 and S_5 . Remember that subset selection is based on classifier accuracy and selection criterion function is calculated on the validation set. Using the 6-fold paired t -test, SFS and SFFS methods are statistically significantly more accurate than BIF, while SFS and SFFS are statistically equivalent (with 95% confidence) again proving wrong the independence assumption.

3.1.2. Landmark-based sampling

The same set of experiments were carried out for landmark-based sampling. Fig. 4 shows the locations of

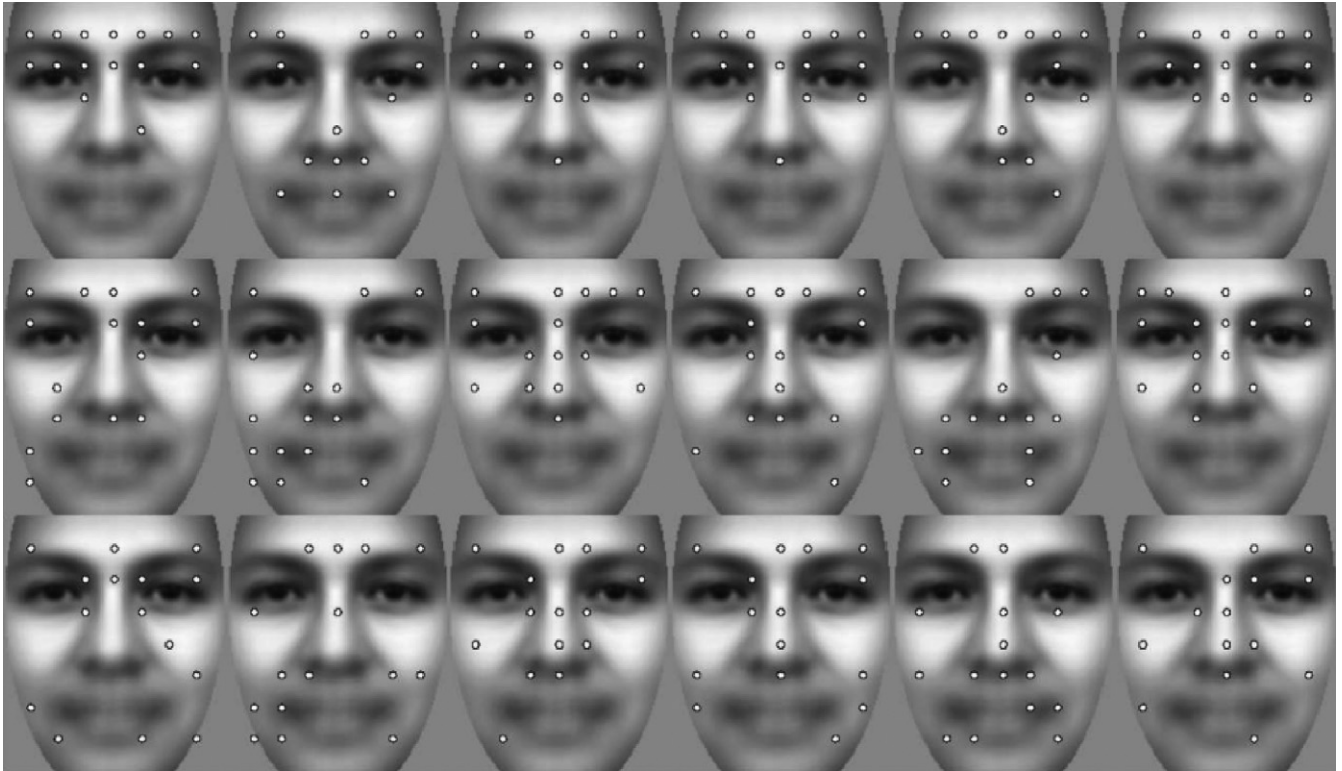


Fig. 3. Selected Gabor kernel locations for lattice-based sampling, found by BIF (top row), SFS (middle row), and SFFS-based (bottom row) subset selection methods for six different configurations, S_1, \dots, S_6 .

Table 1

Classification accuracies of lattice, landmark, and dense sampling methods on six different sessions

	Lattice-based			Landmark-based			Dense sampling
	BIF	SFS	SFFS	BIF	SFS	SFFS	GA
S_1	85.52	84.83	88.28	80.69	84.83	85.52	88.28
S_2	52.41	60.00	55.17	51.72	49.66	53.79	74.48
S_3	85.62	86.30	85.62	78.77	89.04	86.30	86.30
S_4	83.56	86.30	86.99	84.93	78.77	80.82	90.41
S_5	45.21	50.00	49.32	34.25	59.59	50.68	57.53
S_6	82.19	90.41	86.99	84.25	84.93	84.25	90.41
Mean(6)	72.42	76.31	75.39	69.10	74.47	73.56	81.24
STD(6)	18.47	16.91	18.05	21.09	16.03	16.65	13.05
Median(6)	82.88	85.56	86.30	79.73	81.80	82.53	87.29
Mean(4)	84.22	86.96	86.97	82.16	84.39	84.22	88.85
STD(4)	1.65	2.40	1.09	2.93	4.23	2.42	1.97
Median(4)	84.54	86.30	86.99	82.47	84.88	84.89	89.35

For each method, test set performances are given with their mean, standard deviation and median statistics. Statistics with number four are calculated by excluding S_2 and S_5 .

selected kernels in the set Φ_{LOC} for landmark-based sampling. As in the lattice case, BIF approach favors the upper face region by selecting symmetric locations around eyes, eyebrows and forehead. We see that the lower part of the nose also contributes to the subset. With SFS and SFFS, although the contribution of the nose region and cheeks are

more visible, forehead, eyes, and eyebrows are generally found to be informative.

The classification performance of landmark-based sampling is shown in Table 1. Again, we see that SFS and SFFS are significantly more accurate than BIF and that SFS and SFFS are statistically equivalent. An important

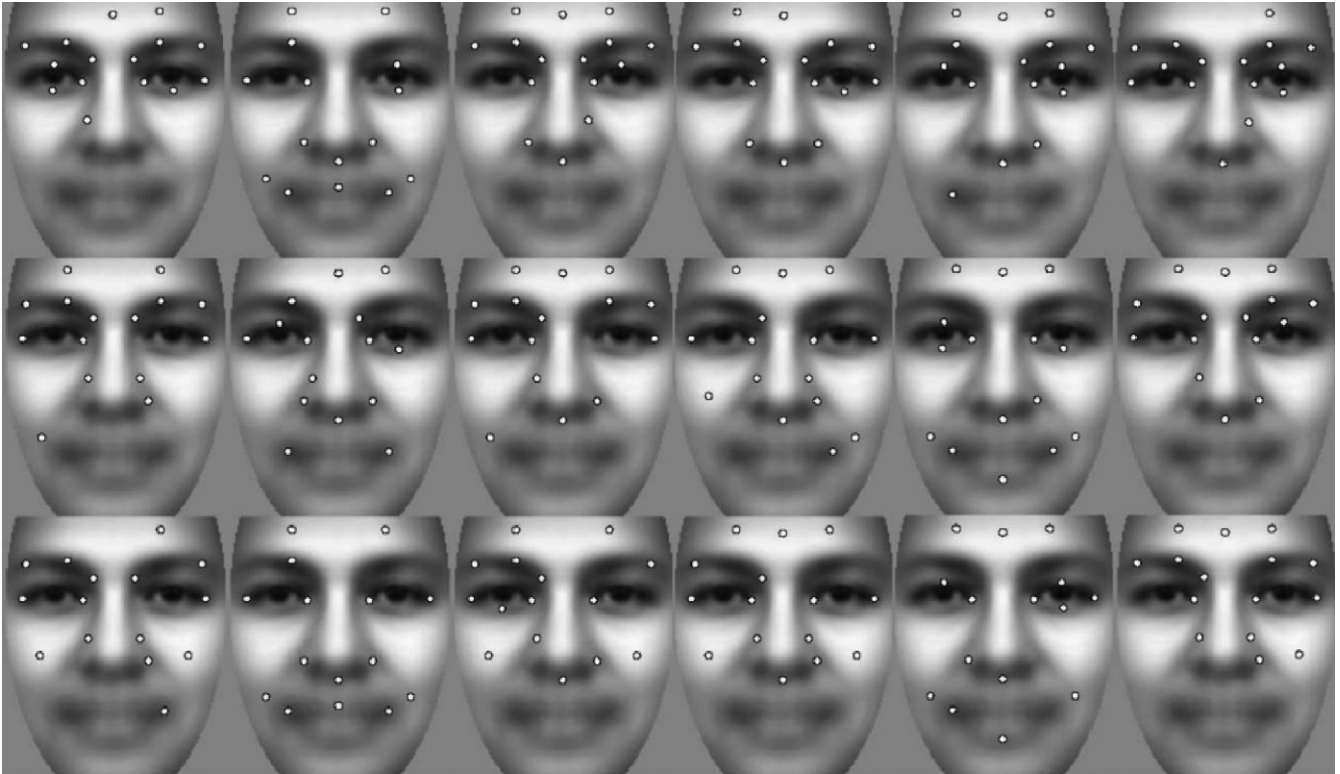


Fig. 4. Selected Gabor kernel locations for landmark-based sampling, found by BIF (top row), SFS (middle row), and SFFS-based (bottom row) subset selection methods for six different configurations.

observation is that lattice-based sampling is more accurate than landmark-based sampling. This indicates that our prior beliefs in saliency regions is not always correct and that it is better to extract salient locations from data.

3.1.3. Dense sampling

Parallel search for a subset Φ_{LOC} is done via constructing genetic chromosomes of size $|\Phi_{LOC}|$ where each gene points to a location in the face image. As in previous experimental settings, $|\Phi_{LOC}|$ is set to 15. As the fitness function, we have used the recognition performance of the subset on the validation set. The single-point crossover operator was implemented to produce new individuals. Since we have the (x, y) coordinates in genes, the mutation operator is implemented as a displacement vector, where the gene to be mutated is displaced by a vector $\eta = \{\eta_x, \eta_y\}$. The norm $|\eta|$ is gradually decreased at each iteration for better convergence. In both operators, we require that the coordinates of face points in a single chromosome do not overlap by more than a specified amount in order to extract independent local information and this distance is selected to be 20 pixels. The probability of crossover and mutation are selected to be $P_c = 0.5$ and $P_m = 0.05$, respectively. The selection of a new population is based on the probability distribution of fitness values. For quick convergence, elitism is employed, where the elitism ratio is 0.05. The initial population size

is 1600. GA terminates when there is no improvement on the accuracy of the best individuals for a specified time interval.

In Fig. 5, the 15 feature points found by the best individuals of GAs are shown. From the figures, it is clear that the outline of the face, the outline of the nose region, and eyes and eyebrows contribute to the most discriminative subset Φ_{LOC} . Almost in all configurations except S_2 and S_5 , cheeks, mouth region and the center area of the forehead are absent. In S_1 and S_2 , there is a feature point outside the face area. This may happen because of two reasons: (i) the sub-optimal convergence of the GA algorithm, (ii) the selected point does not positively or negatively contribute to the recognition performance (i.e., effectively there are 14 useful points.) This explanation is verified in Section 3.2.

The recognition performances of GA-based location selection are also shown in Table 1. Although the second and fifth configurations again perform poorly, the recognition accuracies are quite high.

3.1.4. Comparison of three methods

The comparison of classification accuracies of lattice, landmark and dense sampling methods using 6-fold paired t -test shows that dense sampling using GA performs the best. Lattice-based sampling is found to be statistically more accurate than landmark-based sampling, and between

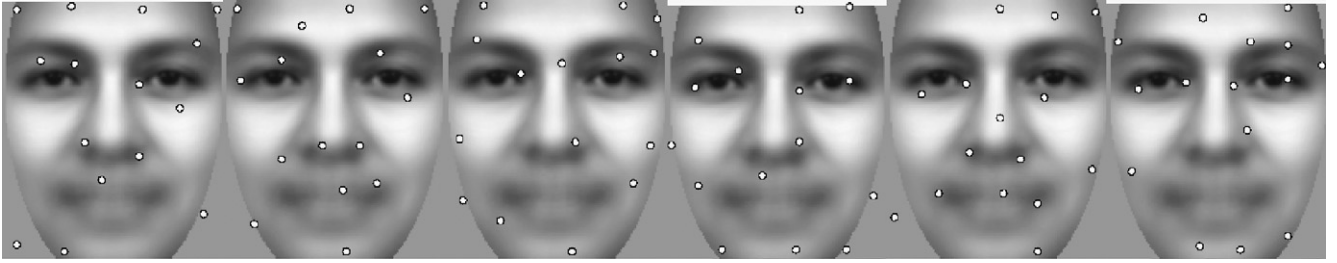


Fig. 5. Selected kernel positions found by GA for six different sessions.



Fig. 6. Selected frequency and orientation pairs at the selected kernel locations. Filled circles on each oriented line represent the selected kernel frequency where innermost circles are for low frequencies and outermost frequencies are for high frequencies. Oriented lines represent the kernel orientations.

lattice-based and landmark-based sampling, SFS or SFFS on lattice-based sampling is the most accurate. These results indicate that our prior beliefs as to the saliency of certain regions for discrimination (as in landmark-based sampling) are not true and that it is better to allow a general sampling from a grid (as in lattice-based sampling) and it is even better to allow a more general sampling from the whole image (as in dense sampling).

3.2. Kernel frequency and orientation selection

In the previous section, we have identified the discriminative facial positions and represent them as Φ_{LOC} in subset selection formalism. Now, our aim is to select the useful frequency and orientation pairs from Φ_{LOC} to construct the subset Φ_{FO} , where $\Phi_{FO} \subset \Phi_{LOC}$. Since *dense sampling method* is the top performer in the previous part, we will

Table 2
Classification accuracies for SFFS-based frequency/orientation selection

S_1	86.90
S_2	55.17
S_3	86.99
S_4	85.62
S_5	47.26
S_6	89.73
Mean \pm SD(6)	75.28 \pm 18.85
Median(6)	86.26
Mean \pm SD(4)	87.31 \pm 1.73
Median(4)	86.95

continue our experiments using its output as our input set in this section. Recall that Φ_{LOC} consists of local feature vectors v_i , each v_i contains magnitudes from Gabor kernel convolutions, $m_{p,q}$, and $|\Phi_{LOC}|$ is $15 \times 40 = 600$.

Frequency and orientation (F/O) selection is carried out using the SFFS algorithm since our experiments have shown that it has the best trade-off between complexity and accuracy. The termination condition is determined empirically by observing the behavior of the classification rate on the validation set, and the dimensionality of the subset Φ_{FO} is set to a value where the classification performance does not improve significantly for a specified time. Our experiments have shown that the target dimensionality of 200 is sufficient for best accuracy on the validation set, which implies a decrease of complexity to one-third. Fig. 6 shows the selected F/O pairs at their specific facial locations for each of the six sessions. In general, we see that the selected kernel orientations are correlated with the underlying characteristics of facial texture. This is more obvious in locations where non-complex local facial directions are present, i.e., at the outline of faces. In terms of frequencies, some positions favor low frequencies, some high frequencies, and in some places, both of them are used together. In S_1 and S_5 , there were selected points outside of the face area. In the F/O selection phase, none of the F/O 's in S_1 were selected (see Fig. 6(a)). However, in S_5 in Fig. 6(e), all of the F/O were selected which is due to the problematic training/validation sets in S_5 .

The recognition performances of each subset Φ_{FO} are shown in Table 2. Comparing these with the bottom part of Table 1, using the 6-fold paired t -test, Φ_{FO} and Φ_{LOC} are found to be statistically equivalent. Although the average test set performances do not illustrate this equivalence due to the problematic sessions S_2 and S_5 , median statistics with recognition accuracies, 87.29% and 86.26% for Φ_{LOC} and Φ_{FO} , respectively, indicate the similar performance. This indicates that the dimensionality can be decreased from 600 to 200 without losing from accuracy. As mentioned previously, S_2 and S_5 lead to a performance degradation due to the eyeglasses. Fig. 7 shows some misclassified faces from sessions S_2 , S_3 , S_4 , and S_5 . Notice that in addition to the eyeglasses,

expression variations that cause high texture variations in the mouth region, especially teeth, are the other sources for the misclassifications.

3.3. Learning the number of important locations

In our system, we have fixed the number of local image descriptors to a constant value, $d = 15$. In order to investigate the effect of d , a different methodology is used where the number of Gabor kernels is variable. This variability is introduced into the system by adding an additional penalty term to the fitness function of the GA-based subset selection algorithm that penalizes the cardinality. Therefore, among different chromosomes having the same classification accuracy, the one with the least cardinality will be favored. Table 3 shows classification accuracies and the found cardinalities of the selected subsets for dense and lattice-based sampling. The results show that the classification accuracy of the variable-size extension of a kernel selection module is statistically similar to its fixed ($d = 15$) version while using subsets of cardinality $d \sim 8 - 11$. However, as in the previous section, dense sampling-based selection outperforms lattice-based selection, and with S_2 and S_5 , the problem gets more difficult and we need more locations because of the outliers. These results indicate that sampling 15 points from the face is sufficient.

3.4. Joint-learning of location, frequency, and orientation

In the proposed architecture, the topographical locations of Gabor kernels, and their F/O parameters are learned consecutively in a two-stage subset selection methodology. The separation of these two tasks allows us to find near optimal solutions in a significantly faster way. However, one may consider the application of subset selection directly on the whole feature variables, thus learning the locations together with frequency and orientation parameters in one step. Although such an approach is infeasible by using sequential selection methods, GA-based parallel search can be a viable solution. However, it should be noted that even GA-based selection from all feature variables has significant time complexity when compared to the two-stage selection. We have performed a joint-learning of location, frequency, and orientation using a dense sampling-based GA algorithm where the dimensionality of the selected subset is allowed to vary by introducing a penalty term as in Section 3.3. Classification accuracies and the cardinalities of the found subsets are shown in Table 4. The joint subset selection search performed statistically worse than the two-stage selection scheme. This is to be expected, since it is very difficult to find the optimal parameters in such a large search space. The mean classification accuracy of the four experimental sessions (excluding S_2 and S_5) decrease from 87.31% to 76.37% on the test set.



Fig. 7. Examples of misclassified face images. In each set, first two columns display the training set images, third column is the probe image, and last column shows the nearest face image found by the algorithm.

Table 3

Classification accuracies for variable-size kernel location selection on the test set

	Dense sampling-based location sel.		Lattice-based location sel.	
	Accuracy	#Locs	Accuracy	#Locs
S_1	85.60	8	83.00	10
S_2	64.39	15	50.00	15
S_3	85.00	9	83.00	8
S_4	88.00	11	81.50	10
S_5	66.50	18	47.30	11
S_6	88.40	8	89.70	10
Mean \pm SD(6)	79.65 \pm 11.10		72.42 \pm 18.65	
Median(6)	85.30		82.25	
Mean \pm SD(4)	86.75 \pm 1.70		84.30 \pm 3.67	
Median(4)	86.80		83.00	

#Locs denotes the number of selected points over the face image.

Table 4

Classification accuracies for variable-size joint kernel location, frequency and orientation selection

Session	Test set	Dim.
S_1	77.40	139
S_2	47.26	138
S_3	80.14	110
S_4	73.29	116
S_5	43.15	115
S_6	74.66	101
Mean \pm SD(6)	65.98 \pm 16.32	
Median(6)	73.98	
Mean \pm SD(4)	76.37 \pm 3.04	
Median(4)	76.03	

Dim. denotes the overall cardinality of the selected subset.

3.5. Sensitivity analysis

The proposed architecture learns the discriminative facial locations and the parameters of the local Gabor feature extractors on perfectly aligned faces at the enrollment stage. Although accurate alignment can be performed at the enrollment phase, it is not guaranteed to have perfectly aligned faces at the identification phase. Depending on the accuracy of the face detection and facial landmark localization module, test faces may have some scale and rotation variations. In order to study the effect of face alignment and registration process on the recognition accuracy, we have performed sensitivity analysis by simulating possible variations at the preprocessing stage.

In the pre-processing stage, faces in the FERET database are aligned according to the coordinates of eyes and mouth. Left eye, right eye and the mouth center coordinates are used to find the in-plane rotation angle and the scale of the face. Once these two parameters are found, the input face is

transformed to a canonical position. In the sensitivity analysis, we assume that the landmark localization module which detects eye and mouth coordinates is not accurate. Then, we can have inaccurately registered faces at the identification phase.

In our tests, we add Gaussian noise to the groundtruth fiducial positions to simulate the behavior of the inaccurate localization module. For each test face, we have generated 10 different badly registered faces, each having different localization errors. Localization errors were produced by adding zero mean Gaussian noise with standard deviations (σ) starting from one pixel up to 10 pixels in both x and y directions. Fig. 8 shows sample images which are inaccurately registered with increasing σ error levels to the right. Note that although we have included errors with standard deviations up to 10 pixels, errors this high are highly unlikely.

Table 5 shows the recognition rates for five σ levels. For each σ , 10 synthetic test sets are formed. In these experiments, only the configurations S_1 , S_3 , S_4 , and S_6 are used. The recognition rates are reported for the location selection module and for the F/O selection module. If the test images are registered without any localization error ($\sigma = 0$), the identification rates are 88.85% and 87.31%, respectively. If the feature localization algorithm performs an inaccurate localization with three pixel standard deviation ($\sigma = 3$), the identification rate for location selection and F/O selection modules drop to 85.53% and 80.70%, respectively. The performance degradation for all error levels is plotted in Fig. 9. We see that performance degradation for small standard deviations is tolerable. Comparing the performance figures of Table 5 with Table 1, we observe that the benefits of learning discriminative facial locations over fixed landmark-based selection remain even with erroneous alignment up to $\sigma = 3$. Furthermore, we expect the performance figures of Table 1 to also degrade in the presence of alignment errors.



Fig. 8. Inaccurately registered faces of two subjects in the FERET database. At each row, the standard deviations (σ) of the Gaussian noise added to the groundtruth are 1, 3, 5, 7, and 10 pixels.

Table 5
The recognition accuracies for various localization error levels (σ)

	$\sigma = 0$	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$	$\sigma = 5$
LOC	88.8 ± 1.9	88.0 ± 2.6	86.5 ± 3.0	85.5 ± 2.7	83.9 ± 4.3	83.6 ± 3.3
FO	87.3 ± 1.7	85.1 ± 2.8	82.1 ± 3.4	80.7 ± 4.1	78.58 ± 4.4	78.7 ± 4.1

Average recognition accuracies and their standard deviations are reported both for location (second row), and for frequency/orientation selection module (third row).

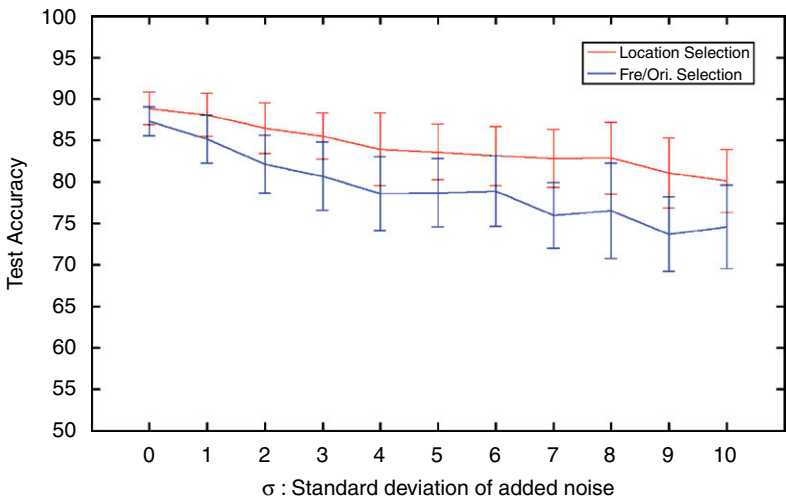


Fig. 9. The recognition rates for various localization error levels (σ). Average recognition accuracies and their standard deviations are plotted both for location selection module and for frequency/orientation selection module.

4. Conclusion and discussion

We present a new form of local feature-based face representation technique which is able to consider local feature dependencies present in faces and allows better feature extraction. Our main contribution is to reformulate the representation task as a subset selection problem. We have shown that it is possible to reach an accurate and simple facial feature set by learning informative locations from training data instead of assuming regions of saliency a priori, and by taking the dependencies of local features into account. To confirm this idea, we assume a perfect facial feature localization module by finding these points manually, and show that using these facial feature points (i.e., landmark-based sampling), one can get only sub-optimal identification performance.

Feature selection in high dimensions has proved to be very difficult and time consuming. In order to overcome this problem, we have devised a two-stage subset selection scheme. In the first stage, we use a GA-based selection technique to find the topographically discriminative facial locations. In the second stage, the best set of frequencies and orientations of Gabor filters at these locations are chosen using the SFFS algorithm. Frontal face recognition experiments on the FERET dataset show that topographical feature distribution around face outline, upper part of the face covering forehead, eyebrows and eye corners, and points in the periphery of nose have high recognition power. However, it is essential to employ a mechanism which takes the dependencies of features into account. Since our approach uses floating search and GA algorithms, time complexity of the training phase is high. This is due to our use of a fitness function based on classification accuracy; this increases complexity but finds the best features for high accuracy.

Our proposed method focuses only on the face coding and assumes that face images are normalized with respect to eye coordinates previously. However, robustness of the algorithm has been studied when some variations such as rotation and scale are present in the data. We have shown that location and parameter selection are not highly sensitive to correct alignment and that the benefits our feature selection remain even when small errors are present.

Acknowledgments

This work is supported by Boğaziçi University research Grants 03K120250 and 02A104D. A preliminary version of this work is presented at the IEEE Int. Conference on Image Processing, August 2003, Barcelona [20]. E. Alpaydın is also supported by the Turkish Academy of Sciences, in the framework of the Young Scientist Award Program (EA-TÜBA-GEİP/2001-1-1).

References

- [1] M. Grudin, On internal representations in face recognition, *Pattern Recognition* 33 (7) (2000) 1161–1177.
- [2] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition—a review, *Comput. Vision Image Understanding*, in press.
- [3] B. Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition, *IEEE Trans. PAMI* 24 (6) (2002) 780–788.
- [4] A.M. Martinez, A.C. Kak, PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2) (2001) 228–233.
- [5] T.K. Kim, H. Kim, W. Hwang, J. Kittler, Independent component analysis in a local facial residue space for face recognition, *Pattern Recognition* 37 (9) (2004) 1873–1885.
- [6] Liu, H. Wechsler, Evolutionary pursuit and its application to face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (6) (2000) 570–582.
- [7] L. Wiskott, J.M. Fellous, N. Kruger, C. Malsburg, Face recognition by elastic bunch graph matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 775–779.
- [8] A. Tefas, C. Kotropoulos, I. Pitas, Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (7) (2001) 735–746.
- [9] M.J. Viola, P. Viola, Face recognition using boosted local features, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2003.
- [10] C. Liu, H. Wechsler, Independent component analysis of Gabor features for face recognition, *IEEE Trans. Neural Networks* 14 (4) (2003) 919–928.
- [11] J.G. Daugman, Complete discrete 2D Gabor transforms by neural networks for image analysis and compression, *IEEE Trans. Acoust. Speech Signal Process.* 36 (7) (1988) 1169–1179.
- [12] O. Ayinde, Y.H. Yang, Face recognition approach based on rank correlation of Gabor-filtered images, *Pattern Recognition* 35 (6) (2002) 1275–1289.
- [13] L. Wiskott, Phantom faces for face analysis, *Pattern Recognition* 30 (6) (1997) 837–846.
- [14] N. Kruger, M. Potzsch, C. Malsburg, Determination of face position and pose with a learned representation based on labelled graphs, *Image Vision Comput.* 15 (1997) 665–673.
- [15] P. Kalocsai, C. Malsburg, J. Horn, Face recognition by statistical analysis of feature detectors, *Image Vision Comput.* 18 (4) (2000) 273–278.
- [16] D.H. Liu, K.M. Lam, L.S. Shen, Optimal sampling of Gabor features for face recognition, *Pattern Recognition Lett.* 25 (2) (2004) 267–276.
- [17] P. Yang, S. Shan, W. Gao, S.Z. Li, D. Zang, Face recognition using ada-boosted Gabor features, in: *Proceedings of the 16th International Conference on Face and Gesture Recognition*, 2004.
- [18] X. Wang, H. Oi, Face recognition using optimal non-orthogonal wavelet basis evaluated by information complexity, in: *Proceedings of the 16th International Conference on Pattern Recognition*, 2002, pp. 164–167.
- [19] B. Gökberk, L. Akarun, E. Alpaydın, Feature selection for pose invariant face recognition, in: *Proceedings of the 16th International Conference on Pattern Recognition*, 2002, pp. 306–309.
- [20] B. Gökberk, M.O. İrfanoğlu, L. Akarun, E. Alpaydın, Optimal Gabor kernel selection for face recognition, in: *Proceedings of the International Conference on Image Processing*, 2003, pp. 677–680.
- [21] A.A. Salah, E. Alpaydın, L. Akarun, A selective attention based method for visual pattern recognition with application to handwritten digit recognition and face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (3) (2002) 420–425.
- [22] A. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (1) (2000) 4–37.
- [23] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, *Image Vision Comput.* 16 (5) (1998) 295–306.

About the Author—BERK GÖKBERK received his B.S. and M.S. degrees in computer science from Boğaziçi University, Istanbul, Turkey in 1999 and 2001, respectively. He is currently a Ph.D. student at the Department of Computer Engineering, Boğaziçi University. His research interests include biometrics, 2D/3D face recognition, pattern recognition, and computer vision.

About the Author—M. OKAN İRFANOĞLU received his B.S. degree in industrial engineering and M.S. degree in computer science both from Boğaziçi University, Istanbul, Turkey in 2002 and 2004, respectively. He is currently a Ph.D. student at the The Ohio State University Computer Sciences and Engineering Department.

About the Author—LALE AKARUN received the B.S. and M.S. degrees in electrical engineering from Boğaziçi University, Istanbul, Turkey, in 1984 and 1986, respectively, and the Ph.D. degree from Polytechnic University, Brooklyn, NY, in 1992. From 1993 to 1995, she was Assistant Professor of electrical engineering at Boğaziçi University, where she is now Professor of computer engineering. Her current research interests are in image processing, computer vision, and computer graphics.

About the Author—ETHEM ALPAYDIN received his Ph.D. degree in computer science from Ecole Polytechnique Federale de Lausanne, Switzerland in 1990 and did postdoctoral work at the International Computer Science Institute (ICSI), Berkeley in 1991. Since then he has been teaching in the Department of Computer Engineering at Boğaziçi University, Istanbul, where he is currently professor. He had visiting appointments at MIT in 1994, ICSI in 1997 (as a Fulbright scholar) and IDIAP, Switzerland in 1998. He received the young scientist award from the Turkish Academy of Sciences in 2001 and the scientific encouragement award from the Turkish Scientific and Technical Research Council in 2002. His book, *Introduction to Machine Learning*, has recently been published by The MIT Press. He is a senior member of the IEEE and the IEEE Computer Society.