Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ACADEMIC PRESS

Computer Vision and Image Understanding 91 (2003) 6–21

Computer Vision and Image Understanding

www.elsevier.com/locate/cviu

Face recognition: component-based versus global approaches

Bernd Heisele,[a,b,*] Purdy Ho,[c] Jane Wu,[b] and Tomaso Poggio[b]

[a] Honda Research Institute US, 145 Tremont St., Boston, MA 02111, USA
[b] Center for Biological and Computational Learning, M.I.T., Cambridge, MA, USA
[c] Hewlett-Packard, Cambridge, MA, USA

Received 15 February 2002; accepted 11 February 2003

Abstract

We present a component-based method and two global methods for face recognition and evaluate them with respect to robustness against pose changes. In the component system we first locate facial components, extract them, and combine them into a single feature vector which is classified by a support vector machine (SVM). The two global systems recognize faces by classifying a single feature vector consisting of the gray values of the whole face image. In the first global system we trained a single SVM classifier for each person in the database. The second system consists of sets of view-specific SVM classifiers and involves clustering during training. We performed extensive tests on a database which included faces rotated up to about 40° in depth. The component system clearly outperformed both global systems.

Keywords: Face recognition; Face detection; Support vector machines

1. Introduction

Over the past 20 years numerous face recognition papers have been published in the computer vision community; a survey can be found in [1]. The number of real-world applications (e.g., surveillance, secure access, human/computer interface)

* Corresponding author. Fax: 1-617-338-4909.
E-mail addresses: heisele@ai.mit.edu (B. Heisele), purdyho@alum.mit.edu (P. Ho), jia_wu@mit.edu (J. Wu), tp@ai.mit.edu (T. Poggio).

1077-3142/$ - see front matter
doi:10.1016/S1077-3142(03)00073-0

and the availability of cheap and powerful hardware also lead to the development of commercial face recognition systems. Despite the success of some of these systems in constrained scenarios, the general task of face recognition still poses a number of challenges with respect to changes in illumination, facial expression, and pose.

In the following, we give a brief overview on face recognition methods. Focusing on the aspect of pose invariance, we divide face recognition techniques into two categories: (i) global approach and (ii) component-based approach.

(i) In this category a single feature vector that represents the whole face image is used as input to a classifier. Several classifiers have been proposed in the literature, e.g., minimum distance classification in the eigenspace [2,3], Fisher's discriminant analysis [4], and neural networks [5]. A comparison between various state-of-the-art global techniques including eigenfaces, Fisher's discriminant analysis, and kernel PCA can be found in [6,7]. Global techniques work well for classifying frontal views of faces. However, they are not robust against pose changes since global features are highly sensitive to translation and rotation of the face. To avoid this problem, an alignment stage can be added before classifying the face. Aligning an input face image with a reference frontal face image requires computing correspondences between the two face images. These correspondences are usually determined for a small number of prominent points in the face like the center of the eye, the nostrils, or the corners of the mouth. Based on these correspondences the input face image can be warped to a reference face image. An affine transformation is computed to perform the warping in [8]. Active shape models are used in [9] to align input faces with model faces. A semi-automatic alignment step in combination with support vector machine (SVM) classification was proposed in [10]. Due to self-occlusion, automatic alignment procedures will eventually fail to compute the correct correspondences for large pose deviations between input and reference faces. An alternative, which allows a larger range of views, is to combine a set of view-specific classifiers, originally proposed in a biological context in [11]. In [12], an eigenface approach was used to recognize faces under variable pose by grouping the training images into several separate eigenspaces, one for each combination of scale and orientation. Combining view-specific classifiers has also been applied to face detection. The system presented in [13] was able to detect faces rotated in depth up to $\pm 90°$ with two naïve bayesian classifiers, one trained on frontal views, the other one trained on profiles.

(ii) An alternative to the global approaches is to classify local facial components. The main idea of component-based recognition is to compensate for pose changes by allowing a flexible geometrical relation between the components in the classification stage. In [14], face recognition was performed by independently matching templates of three facial regions (both eyes, nose, and mouth). The configuration of the components during classification was unconstrained since the system did not include a geometrical model of the face. A similar approach with an additional alignment stage was proposed in [15]. In an effort to enhance the robustness against pose changes the originally global eigenface method has been further developed into a component-based system [12] where PCA is applied to local facial components (eyes, nose, and mouth). Elastic grid matching described in [16] uses Gabor wavelets to extract features at grid points and graph matching for the proper positioning of the grid. The recognition was based on wavelet coefficients that were computed on the nodes of a 2-D elastic graph. In [17], a window was shifted over the face image and the DCT coefficients computed within the window were fed to a 2-D Hidden Markov Model. A probabilistic approach using part-based matching has been proposed in [18] for expression invariant and occlusion tolerant recognition of frontal faces.

We present two global approaches and a component-based approach to face recognition and evaluate their robustness against pose changes. The first global method consists of a straightforward face detector which extracts the face from an input image and propagates it to a set of SVM classifiers that perform the face recognition. By using a face detector we achieve translation and scale invariance. In the second global method we split the images of each person into view-specific clusters. We then train view-specific SVM classifiers on each single cluster. In contrast to the global methods, the component-based system uses a face detector that detects and extracts local components of the face. The detector consists of a set of SVM classifiers that locate learned facial components and a single geometrical classifier that checks if the configuration of the components matches a learned geometrical face model. The detected components are extracted from the image, normalized in size, and fed to a set of SVM classifiers.

The outline of the paper is as follows: Section 2 gives a brief overview on SVM learning and on strategies for multi-class classification with SVMs. In Section 3 we describe the two global methods for face recognition. Section 4 is about the component-based system. Section 5 contains experimental results and a comparison between the global and component systems. Section 6 concludes the paper and suggests future work.

2. Support vector machine classification

We first explain the basics of SVMs for binary classification [19]. Then we discuss how this technique can be extended to deal with general multi-class classification problems.

2.1. Binary classification

SVMs belong to the class of maximum margin classifiers. They perform pattern recognition between two classes by finding a decision surface that has maximum distance to the closest points in the training set which are termed support vectors. We start with a training set of points $\mathbf{x}_i \in n$, $i = 1, 2, \ldots, N$ where each point $\mathbf{x}_i$ belongs to one of two classes identified by the label $y_i \in \{-1, 1\}$. Assuming linearly separable data,[1] the goal of maximum margin classification is to separate the two classes by a

[1] For the non-separable case the reader is referred to [19].

hyperplane such that the distance to the support vectors is maximized. This hyperplane is called the optimal separating hyperplane (OSH). The OSH has the form:

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b, \tag{1}$$

The coefficients α_i and the b in Eq. (1) are the solutions of a quadratic programming problem [19]. Classification of a new data point $\mathbf{x}$ is performed by computing the sign of the right-hand side of Eq. (1). In the following we will use:

$$d(\mathbf{x}) = \frac{\sum_{i=1}^{\ell} \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b}{\| \sum_{i=1}^{\ell} \alpha_i y_i \mathbf{x}_i \|} \tag{2}$$

to perform multi-class classification. The sign of d is the classification result for $\mathbf{x}$, and $|d|$ is the distance from $\mathbf{x}$ to the hyperplane. Intuitively, the farther away a point is from the decision surface, i.e., the larger $|d|$, the more reliable the classification result.

The entire construction can be extended to the case of nonlinear separating surfaces. Each point $\mathbf{x}$ in the input space is mapped to a point $\mathbf{z} = \Phi(\mathbf{x})$ of a higher dimensional space, called the feature space, where the data are separated by a hyperplane. The key property in this construction is that the mapping $\Phi(\cdot)$ is subject to the condition that the dot product of two points in the feature space $\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})$ can be rewritten as a kernel function $K(\mathbf{x}, \mathbf{y})$. The decision surface has the form:

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b,$$

again, the coefficients α_i and b are the solutions of a quadratic programming problem. Note that $f(\mathbf{x})$ does not depend on the dimensionality of the feature space.

An important family of kernel functions is the polynomial kernel

$$K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^d,$$

where d is the degree of the polynomial. In this case the features of the mapping $\Phi(\mathbf{x})$ are all the possible monomials of input features up to the degree d.

2.2. Multi-class classification

There are a number of strategies for solving q-class problems with binary SVM classifiers (see, e.g. [20]). Popular are the one-vs-all and the pairwise approach:

(i) In the one-vs-all approach q SVMs are trained. Each of the SVMs separates a single class from all remaining classes [21,22].

(ii) In the pairwise approach $q(q-1)/2$ machines are trained. Each SVM separates a pair of classes. The pairwise classifiers are arranged in trees, where each tree node represents an SVM. A bottom-up tree, similar to the elimination tree used in tennis tournaments, was originally proposed in [23] for recognition of 3-D objects and was applied to face recognition in [24]. A top-down tree structure has been published in [25].

There is no thorough theoretical analysis of multi-class techniques for SVMs with respect to recognition performance. Experiments on person recognition show similar classification results for the two strategies [26]. A more recent comparison between several multi-class techniques [20] favors the one-vs-all approach because of its simplicity and excellent classification performance. Regarding the training effort, the one-vs-all approach is preferable over the pairwise approach since only q SVMs have to be trained compared to $q(q-1)/2$ SVMs in the pairwise approach. The run-time complexity of the two strategies is similar: The one-vs-all approach requires the evaluation of q, the pairwise approach the evaluation of $q-1$ SVMs. We opted for one-vs-all since it seems at least on par with other approaches regarding the classification rate and because it requires the training of only q classifiers.

3. Global approach

Both global systems described in this paper consist of a face detection stage, where the face is detected and extracted from an input image and a recognition stage where the person's identity is established.

3.1. Face detection

We developed a face detector similar to the one described in [27]. In order to detect faces at different scales we first computed a resolution pyramid for the input image and then shifted a 58×58 window over each image in the pyramid. We applied two preprocessing steps to the gray images to compensate for certain sources of image variations [28]. A best-fit intensity plane was subtracted from the gray values to compensate for cast shadows. Then histogram equalization was applied to remove variations in the image brightness and contrast. The resulting gray values were normalized to be in a range between 0 and 1 and were used as input features to a second-degree polynomial SVM classifier. Some detection results are shown in Fig. 1.

The training data for the face detector was generated by rendering seven textured 3-D head models [29]. The heads were rotated between $-30°$ and $30°$ in depth and illuminated by ambient light and a single directional light pointing towards the center of the face. We generated 2457 face images of size 58×58 pixels, some examples are shown in Fig. 2. The negative training set initially consisted of 10,209 58×58 non-face patterns randomly extracted from 502 non-face images. We expanded the training set by bootstrapping [28] to 13,655 non-face patterns.

3.2. Recognition

We implemented two global recognition systems. Both systems were based on the one-vs-all strategy for SVM multi-class classification described in the previous section.

The first system had a linear SVM for every person in the database. Each SVM was trained to distinguish between all images of a single person (labeled $+1$) and all other images in the training set (labeled -1). For both training and testing we first

Fig. 1. Examples of the global face detector applied to real images. Shown are pairs of the original image and the extracted face part.

Fig. 2. Examples of synthetic faces used for training the face detector.

ran the face detector on the input image to extract the face. We re-scaled the face image to 40×40 pixels and converted the gray values into a feature vector.[2] Given a set of q people and a set of q SVMs, each one associated to one person, the class label y of a face pattern $\mathbf{x}$ is computed as follows:

$$y = \begin{cases} n & \text{if } d_n(\mathbf{x}) + t > 0, \\ 0 & \text{if } d_n(\mathbf{x}) + t \leqslant 0, \end{cases} \tag{3}$$

with

$$d_n(\mathbf{x}) = \max \{d_i(\mathbf{x})\}_{i=1}^{q}.$$

where $d_i(\mathbf{x})$ is computed according to Eq. (2) for the SVM trained to recognize person i. The classification threshold is denoted as t. The class label 0 stands for rejection.

[2] We applied the same preprocessing steps to generate the features as for the face detector described.

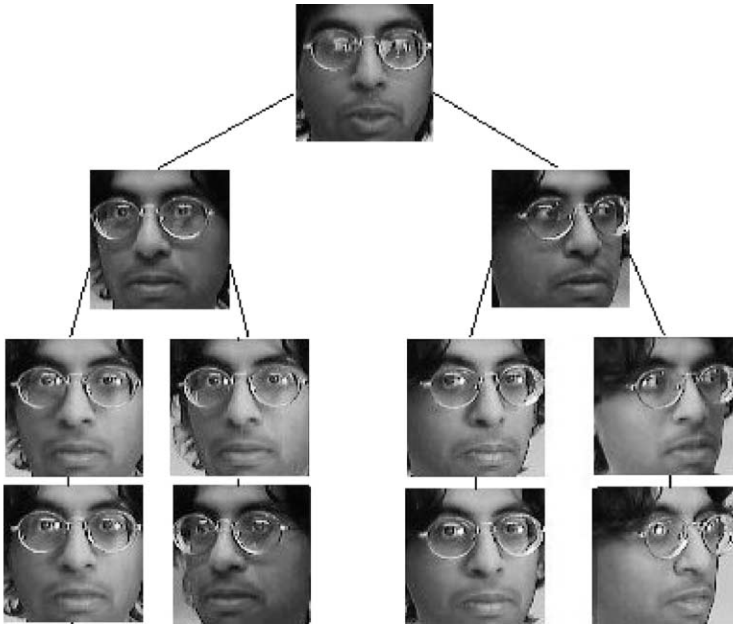

Fig. 3. Binary tree of face images generated by divisive clustering.

Changes in the head pose lead to strong variations in the images of a person's face. These in-class variations complicate the recognition task. For this reason, we developed a second method in which we split the training images of each person into clusters by a divisive cluster technique [30]. The algorithm starts with an initial cluster including all face images of a person after preprocessing. The cluster with the highest variance is split into two by a hyperplane. The variance of a cluster is calculated as:

$$\sigma^2 = \min\left\{\frac{1}{N}\cdot\sum_{m=1}^{N}\|\mathbf{x}_n - \mathbf{x}_m\|^2\right\}_{n=1}^{N},$$

where N is the number of faces in the cluster. After the partitioning has been performed, the face with the minimum distance to all other faces in the same cluster is chosen to be the average face of the cluster. Iterative clustering stops when a maximum number of clusters is reached.[3] The average faces can be arranged in a binary tree. Fig. 3 shows the result of clustering applied to the training images of a person in our database. The nodes represent the average faces; the leaves of the tree are some example faces of the final clusters. As expected, divisive clustering performs a view-specific grouping of faces.

[3] In our experiments we divided the face images of a person into four clusters.

We trained a linear SVM to distinguish between all images in one cluster (labeled $+1$) and all images of other people in the training set (labeled -1).[4] Classification was done according to Eq. (3) with q now being the number of clusters of all people in the training set.

4. Component-based approach

The global approach is highly sensitive to image variations caused by facial rotations. The component-based approach avoids this problem by independently detecting parts of the face. For small rotations, the changes in the components are relatively small compared to the changes in the whole face pattern. Changes in the 2-D locations of the components due to pose changes are accounted for by a learned, flexible face model.

4.1. Detection

We implemented a two-level, component-based face detector which is described in detail in [31]. In the following we give a brief overview of the system.

The principles of the component-based detection system are illustrated in Fig. 4. On the first level, component classifiers independently detected facial components. On the second level, a geometrical configuration classifier performed the final face detection by combining the results of the component classifiers. Given a 58×58 window, the maximum continuous outputs of the component classifiers within rectangular search regions around the expected positions of the components were used as inputs to the geometrical configuration classifier. The search regions have been calculated from the mean and standard deviation of the components' locations in the training images. We also provided the geometrical classifier with the X–Y locations of the maxima of the component classifier outputs relative to the upper left corner of the 58×58 window. The 14 facial components used in the detection system are shown in Fig. 5a, their dimensions are given in Table 1. The shapes and positions of the components have been automatically determined from the training data in order to provide maximum discrimination between face and non-face images; see [31] for details about the learning algorithm.

Training the component-based detector required the extraction of corresponding components from a large number of training images. To automate the extraction process we used a set of seven textured 3-D head models with known point-wise 3-D correspondences. As described in the previous section we rendered the head models under varying pose and illumination. Knowing the correspondences between the images we could locate and extract the 14 components from each of the synthetic images to build a positive component training set. The negative component training set was extracted from the same non-face patterns used for training the global face

[4] This is not exactly a one-vs-all classifier since images of the same person but from different clusters were omitted.

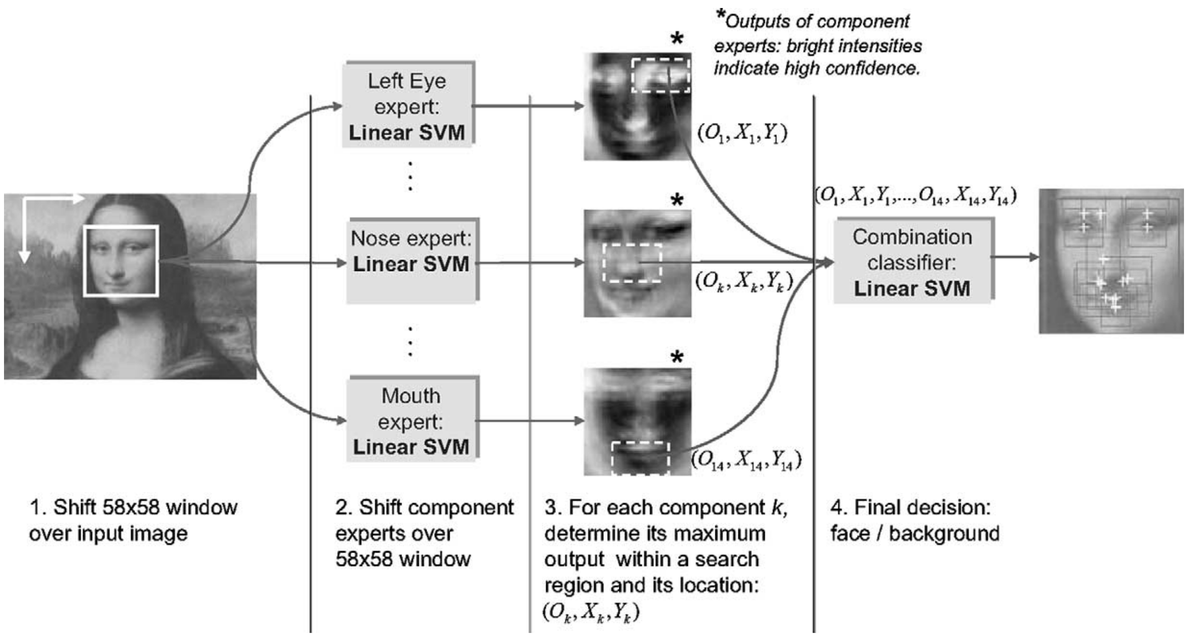

Fig. 4. System overview of the component-based face detector using four components.

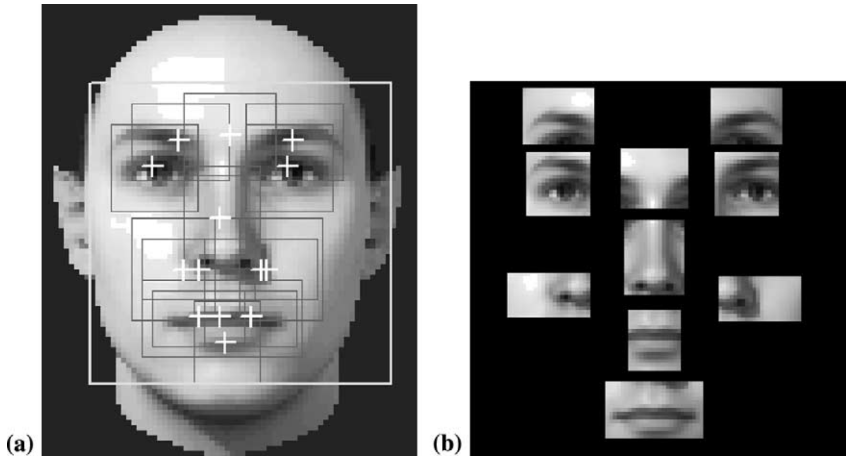

Fig. 5. (a) The 14 components of our face detector. The centers of the components are marked by a white cross. The 10 components that were used for face recognition are shown in (b).

Table 1
Size of the 14 components of the component-based detector

	Eyebrows	Eyes	Nose bridge	Nose	Nostrils	Cheeks	Mouth	Lip	Mouth corners
Width	19	17	18	15	22	21	31	13	18
Height	15	17	16	20	12	20	15	16	11

detector. We trained 14 linear SVMs on the component data and applied them to the whole training set in order to generate the training data for the geometrical classifier. In a final step, we trained the geometrical classifier, which was again a linear SVM, on the X–Y locations and continuous outputs of the 14 component classifiers.

Our component-based face detector was computationally more expensive than the global face detector. This was because the combined size of the 14 components was about 1.12 times the size of the face region used in the global detector. In addition, we had to locate the maxima of the responses of the component classifiers and compute the output of the geometrical classifier. In average, the component-based detector was about 1.2 times slower than the global detector although we used linear SVMs rather than the polynomial SVM used in global detection. If speed is of major concern, we would suggest roughly localizing the face with a fast global face detector or a skin detector and then apply the component-based detection.

4.2. Recognition

To train the face recognizer we first ran the component-based detector over each image in the training set and extracted the components. From the 14 original components we kept 10 for face recognition, removing those that either contained few

Fig. 6. Examples of component-based face detection. Shown are face parts covered by the 10 components that were used for face recognition.

gray value structures (e.g., cheeks) or strongly overlapped with other components. The 10 selected components are shown in Fig. 5b. Examples of the component-based face detector applied to images of the training set are shown in Fig. 6. To generate the input to our face recognition classifier we normalized each of the components in size and combined their gray values into a single feature vector.[5] As for the first global system we used a one-vs-all approach with a linear SVM for every person in the database. The classification result was determined according to Eq. (3).

5. Experiments

The training data for the face recognition system was recorded with a digital video camera at a frame rate of about 5 Hz. The training set consisted of about 10,000 gray

[5] Before extracting the components we applied the same preprocessing steps to the detected 40×40 face image as in the global systems.

Table 2
Average number of support vectors per SVM classifier

Experiment	Number of support vectors
Global linear SVMs	126
Global polynomial SVMs	147
Component linear SVMs	154

face images of 10 subjects from which about 1400 were frontal views. The resolution of the face images ranged between 80×80 and 130×130 pixels with rotations in azimuth up to about $\pm 40°$. Since our images were taken from a dense video sequence, they contained highly redundant information.[6] This was reflected in the training results of the SVMs given in Table 2. The number of support vectors, i.e., the training images based on which the decision function of the SVMs was computed, was small compared to the overall number of training examples.

The test set was recorded with the same camera but on a separate day and under different illumination and with different background. The set included 1544 images of all 10 subjects in our database. The rotation in depth was again up to about $\pm 40°$. Compared to commonly used databases in face recognition, like the PIE database from CMU [32] or the FERET database from NIST, our database included a relatively small number of subjects. Since the goal of this paper is to compare two fundamentally different approaches under similar conditions (i.e., same features, similar classifiers, same training, and test sets) rather than presenting a system which outperforms the best commercial face recognition system, we opted for a small database which made the experiments much easier. For a larger number of subjects the choice of binary classifiers, like SVMs, might not be appropriate since the computational complexity for training and classification is linear with the number of classes. We trained four different recognition systems on the 10,000 images: (1) global system using one linear SVM classifier per person, (2) global system using one second-degree polynomial SVM per person, (3) global system with one linear SVM for each cluster, and (4) component-based approach with one linear SVM classifier per person. The ROC curves for the four systems are shown in Fig. 7.

There are three interesting observations:

- The component system outperformed the global systems for recognition rates larger than 60%. This was the case although the face classifier itself (10 linear SVMs) was less powerful than the classifiers used in the global methods (10 non-linear SVMs in the global method without clustering, and 40 linear SVMs in the method with clustering).
- Clustering lead to a significant improvement of the global method. This is because clustering generates view-specific clusters that have smaller in-class variations than the whole set of images of a person. The global method with clustering and linear SVMs was also superior to the global system without clustering and a non-linear SVM. This shows that a combination of weak classifiers trained on properly

[6] The average normalized correlation was 0.55 between the extracted face images of one person.

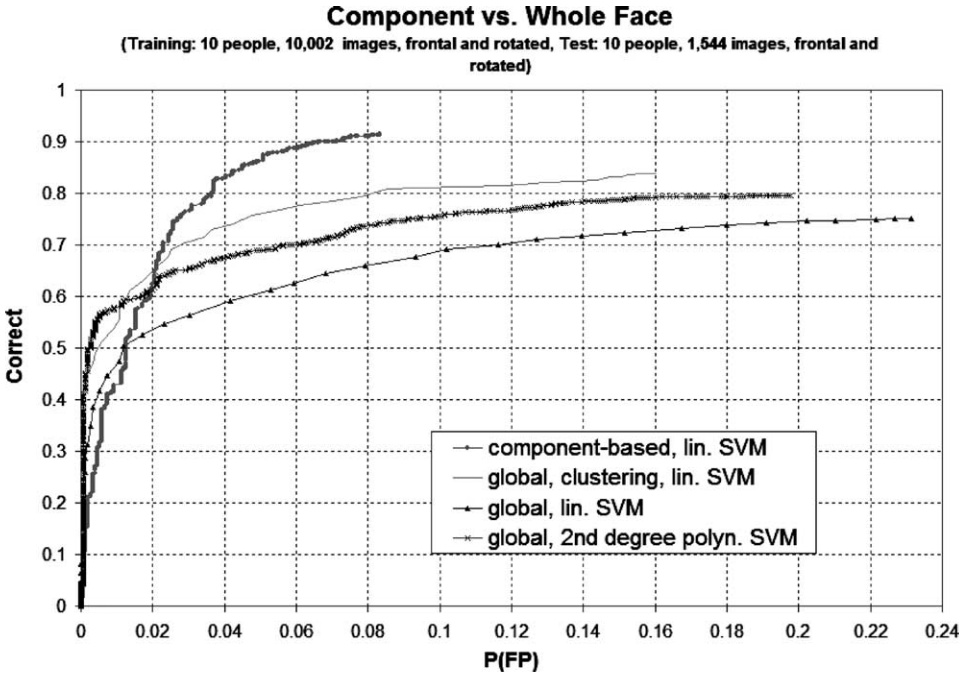

Fig. 7. ROC curves for the four systems.

chosen subsets of the data can outperform a single, more powerful classifier trained on the whole data.

- For low recognition rates the component classifier is slightly worse than the global classifiers. This was probably because of failures in the component detection stage. A visual analysis of the detection results showed that the component extraction failed for about 40 faces with strong rotation while the global detector was able to extract the faces properly. Some examples of misclassifications caused by false detections are shown in Figs. 8 and 9.

Fig. 8. Examples from the test set, which were correctly classified by component-based face recognition system.

Fig. 9. Examples of failures of the component-based face recognition caused by false detections of the components.

6. Conclusion and future work

We presented a component-based technique and two global techniques for face recognition and evaluated their performance with respect to robustness against pose changes. The component-based system detected and extracted a set of 10 facial components and arranged them in a single feature vector that was classified by linear SVMs. In both global systems we detected the whole face, extracted it from the image, and used it as input to the classifiers. The first global system consisted of a single SVM for each person in the database. In the second system we clustered the database of each person and trained a set of view-specific SVM classifiers. We tested the systems on a database which included faces rotated in depth up to about 40°. In the experiment the component-based system outperformed the global systems even though we used more powerful classifiers (i.e., non-linear instead of linear SVMs) for the global system. Some of the classification errors in the component-based recognition resulted from inaccurate extraction of the components. Improvement can be expected from our recent work on component detection [33] where we used pairwise conditional probabilities of the component positions to increase the localization accuracy. Despite some degree of pose invariance, the current component-based classifier cannot deal with the full range of poses (from frontal to profile views). To solve this problem it will be necessary to train view-specific component classifiers, e.g., two mouth classifiers trained on frontal and profile views, respectively. Another significant step towards achieving view invariance can be expected from the use of 3-D head models for training along the lines described in [31]. A preliminary study on combining 3-D morphable models [29] with component-based face recognition showed promising results on synthetic test data [34].

References

[1] R. Chellapa, C. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (5) (1995) 705–741.
[2] L. Sirovitch, M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. Soc. Am. A 2 (1987) 519–524.
[3] M. Turk, A. Pentland, Face recognition using eigenfaces, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1991, pp. 586–591.

[4] P. Belhumeur, P. Hespanha, D. Kriegman, Eigenfaces vs fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[5] M. Fleming, G. Cottrell, Categorization of faces using unsupervised feature extraction, in: Proc. IEEE IJCNN Int. Joint Conf. on Neural Networks, vol. 2, 1990, pp. 65–70.
[6] A.M. Martinez, A.C. Kak, Pca versus lda, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 228–233.
[7] M.-H. Yang, Face recognition using kernel methods, in: Neural Information Processing Systems (NIPS), Vancouver, 2001.
[8] B. Moghaddam, W. Wahid, A. Pentland, Beyond eigenfaces: probabilistic matching for face recognition, in: Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition, 1998, pp. 30–35.
[9] A. Lanitis, C. Taylor, T. Cootes, Automatic interpretation and coding of face images using flexible models, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 743–756.
[10] K. Jonsson, J. Matas, J. Kittler, Y. Li, Learning support vectors for face verification and recognition, in: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 208–213.
[11] T. Poggio, S. Edelman, A network that learns to recognize 3-D objects, Nature 343 (1990) 163–266.
[12] A. Pentland, B. Mogghadam, T. Starner, View-based and modular eigenspaces for face recognition, Technical Report 245, MIT Media Laboratory, Cambridge, 1994.
[13] H. Schneiderman, T. Kanade, A statistical method for 3d object detection applied to faces and cars, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2000, pp. 746–751.
[14] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Trans. Pattern Anal. Mach. Intell. 15 (10) (1993) 1042–1052.
[15] D.J. Beymer, Face recognition under varying pose, A.I. Memo 1461, Center for Biological and Computational Learning, M.I.T., Cambridge, MA, 1993.
[16] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 775–779.
[17] A. Nefian, M. Hayes, An embedded hmm-based approach for face detection and recognition, in: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, vol. 6, 1999, pp. 3553–3556.
[18] A.M. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intell. 24 (6) (2002) 748–763.
[19] V. Vapnik, Statistical Learning Theory, John Wiley and Sons, New York, 1998.
[20] R. Rifkin, Everything old is new again: a fresh look at historical approaches in machine learning, Ph.D. thesis, M.I.T., 2002.
[21] C. Cortes, V. Vapnik, Support vector networks, Mach. Learning 20 (1995) 1–25.
[22] B. Schölkopf, C. Burges, V. Vapnik, Extracting support data for a given task, in: U. Fayyad, R. Uthurusamy (Eds.), Proc. First Int. Conf. on Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, CA, 1995.
[23] M. Pontil, A. Verri, Support vector machines for 3-d object recognition, IEEE Trans. Pattern Anal. Mach. Intell. (1998) 637–646.
[24] G. Guodong, S. Li, C. Kapluk, Face recognition by support vector machines, in: Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000, pp. 196–201.
[25] J. Platt, N. Cristianini, J. Shawe-Taylor, Large margin dags for multiclass classification, Adv. Neural Inform. Process. Systems.
[26] C. Nakajima, M. Pontil, T. Poggio, People recognition and pose estimation in image sequences, Proc. IEEE-INNS-ENNS International Joint Conf. on Neural Networks, 2000, Vol. 4, pp. 4189–4195.
[27] B. Heisele, T. Poggio, M. Pontil, Face detection in still gray images, AI Memo 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA, 2000.
[28] K.-K. Sung, Learning and example selection for object and pattern recognition, Ph.D. thesis, MIT, Artificial Intelligence Laboratory and Center for Biological and Computational Learning, Cambridge, MA, 1996.
[29] V. Blanz, T. Vetter, A morphable model for synthesis of 3D faces, in: Comput. Graphics Proc. SIGGRAPH, Los Angeles, 1999, pp. 187–194.

[30] Y. Linde, A. Buzo, R. Gray, An algorithm for vector quantizer design, IEEE Trans. Commun. 28 (1) (1980) 84–95.
[31] B. Heisele, T. Serre, M. Pontil, T. Vetter, T. Poggio, Categorization by learning and combining object parts, in: Neural Information Processing Systems (NIPS), Vancouver, 2001.
[32] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression (PIE) database of human faces, Computer Science Technical Report 01-02, CMU, 2001.
[33] S.M. Bileschi, B. Heisele, Advances in component-based face detection, in: Proc. of Pattern Recognition with Support Vector Machines, First International Workshop, SVM 2002, Niagara Falls, 2002, pp. 135–143.
[34] J. Huang, V. Blanz, B. Heisele, Face recognition using component-based svm classification and morphable models, in: Proc. of Pattern Recognition with Support Vector Machines, First International Workshop, SVM 2002, Niagara Falls, 2002, pp. 334–341.