Feature-Based Affine-Invariant Localization of Faces

M. Hamouz, J. Kittler, *Member, IEEE,*
J.-K. Kamarainen, P. Paalanen,
H. Kälviäinen, *Member, IEEE,*
and J. Matas, *Member,*
IEEE Computer Society

Abstract—We present a novel method for localizing faces in person identification scenarios. Such scenarios involve high resolution images of frontal faces. The proposed algorithm does not require color, copes well in cluttered backgrounds, and accurately localizes faces including eye centers. An extensive analysis and a performance evaluation on the XM2VTS database and on the realistic BioID and BANCA face databases is presented. We show that the algorithm has precision superior to reference methods.

Index Terms—Face localization, face authentication.

✦

1 INTRODUCTION

THIS paper focuses on face localization, with emphasis on authentication scenarios. We refer to flagging of face presence and rough estimation of the face position (e.g., by a bounding box) as "face detection" and to a precise localization including the position of facial features as "face localization." In contrast to face detection, localization algorithms make an assumption that a face is present and often operate in the regions of interest given by a detector. In the literature, face detection has received significantly more attention than localization. Generic face detection and localization are challenging problems because faces are nonrigid and have a high degree of variability in size, shape, color, and texture. The accuracy of localization heavily influences recognition and verification rates as demonstrated in [1], [2].

In the authentication scenarios, at least one large face is present in a complex background. Under this assumption, our algorithm does not require a face detection to be performed first. The method searches the whole image and returns the best localization candidate. The algorithm is feature-based, where facial features (parts) are represented by Gabor-filter-based complex-valued statistical models. Triplets formed of the detected feature candidates are tested for constellation (shape) constraints and only admissible hypotheses are passed to a face appearance verifier. The appearance verifier is the final test designed reliably to select the true face location based on photometric data. Our experiments show that the proposed method outperforms baseline methods with regard to localization accuracy using a nonambiguous localization criterion.

- *M. Hamouz and J. Kittler are with the Centre for Vision, Speech, and Signal Processing, School of Electronics and Physical Sciences, University of Surrey, Guildford, GU2 7XH, UK.*
  E-mail: {m.hamouz, j.kittler}@surrey.ac.uk.
- *J.-K. Kamarainen, P. Paalanen, and H. Kälviäinen are with the Department of Information Technology, Lappeenranta University of Technology, PO Box 20, FI-53851, Lappeenranta, Finland.*
  E-mail: {jkamarai, paalanen, kalviai}@lut.fi.
- *J. Matas is with the Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Karlovo namesti 13, Praha 2, 121 35, Czech Republic.*
  E-mail: matas@cmp.felk.cvut.cz.

2 STATE-OF-THE-ART IN DETECTION AND LOCALIZATION

In order to detect and localize a face, a face model has to be created. According to Hjelmås and Low [3], methods can be classified as *image-based* and *feature-based*. We add one more category: *warping methods*.

In **image-based methods**, faces are typically treated as vectors in a high-dimensional space and the face class is modeled as a manifold. The manifold is delimited by various pattern recognition techniques. The face model is holistic, i.e., the face is not decomposed into parts (features). Faces of different scales and orientations are detected with a "scanning window." Since it is not possible to scan all possible scales and orientations, discretization of scale, and orientation[1] has to be introduced. Perspective effects, e.g., the foreshortening of one dimension in oblique views are typically ignored. The face/nonface classifier has to learn all possible appearances of faces misaligned within a range of scales and orientations defined by the discretization. As a result, not only the precision of localization decreases (since the classifier cannot distinguish between slightly misaligned faces) but the cluster of face representation becomes much less compact and thus more difficult to learn.

A fast and reliable detection method was proposed by Viola and Jones [4]. This method has had significant impact and a body of follow-up work exists. Other image-based approaches are [5], [6], [7]. The method of Jesorsky et al. [7] is described in detail as it serves as a baseline since it focuses on face localization for face authentication. The method uses Hausdorff distance on edge images in a scale and orientation independent manner. The detection and localization involves three processing steps, where first a coarse detection is performed, then an eye-region model is used in a refinement stage and finally the pupils are sought by a Multilayer Perceptron.

In **feature-based methods** [8], [9], [10], a face is represented by a constellation (shape) model of parts (features), together with models of local appearance. This usually implies that a priori knowledge is needed in order to create the model of a face (selection of features), although an attempt to select salient local features automatically was reported [11].

Warping methods. The distinguishing characteristic of this group is that the facial variability is decomposed into a shape model and shape-normalized texture model as in the Active Shape Models (ASMs) [12], Active Appearance Models (AAMs) [13], and the Dynamic Link Architectures [14], [15], [16]. These methods are well-suited for precise localization. However, no extensive evaluation results have been published.

3 PROPOSED METHODOLOGY

Our method searches for correspondences between detected features and a face model. We chose this representation to avoid the discretization problems experienced by the scanning window technique mentioned above. The use of facial features has the following advantages. First, they simplify illumination distortion modeling. Second, the overall modeling complexity of local features is smaller than in the case of holistic face models. Third, the method is robust, because local feature models can use redundant representations. Fourth, unlike the scanning window techniques, the precision of localization is not limited by the discretization of the search space. It is true that feature-based methods need higher resolution than image-based methods; however, in authentication scenarios this requirement is met.

Since the constellation of local features is not discriminatory enough for a reliable separation of faces from a cluttered background, sophisticated pattern recognition approaches, commonly used in the scanning-window methods, are applied in the final decision on the face location. These approaches should be applied to

---

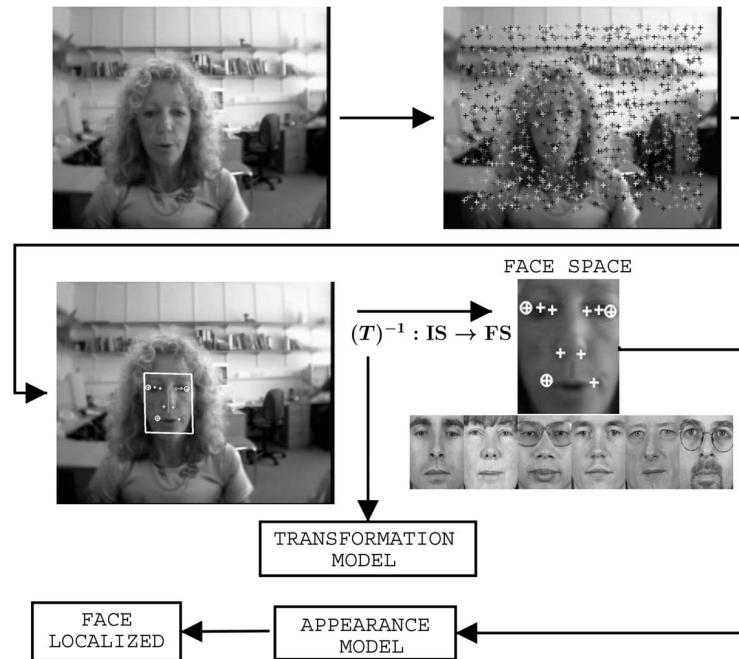1. Orientation means in-plane rotation in this paper.

Fig. 1. Structure of the algorithm, T is the transformation from the face-space into the image-space.

geometrically normalized data for the reasons mentioned above. A natural way is to introduce a scale-and-orientation invariant space. The face model can be then trained using registered templates and this maintains the localization accuracy, since a slightly misaligned face would likely be disregarded. To estimate real scale and orientation of faces in the scene, detected features can be exploited as done in stereo. As feature detectors are not error-free, we also have to expect a large number of false positives, especially in a cluttered background. The above arguments lead to the following sequence of steps: Feature Detection $\rightarrow$ Face Hypothesis Generation $\rightarrow$ Registration (scale and orientation normalization) $\rightarrow$ Appearance Verification. More detailed description of the algorithm follows.

Feature detection. Since feature detection is the initial step, the speed, accuracy, and reliability of the whole detection/localization system will critically depend on the accuracy and reliability of the detected feature candidates.

We regard facial features to be detectable subparts of faces, which may appear in arbitrary poses and orientations as faces themselves. When modeling geometric variations, every additional degree of freedom increases the computational complexity of the detector. It is reasonable to assume that faces are flat objects apart from the nose. In the authentication scenarios, in-depth rotation is small and faces are close to frontal (person standing/sitting in front of the camera). For such situations, 2D affine transformation is a sufficient model. Since facial features are much smaller than the whole face, it is justifiable to approximate the feature pose variations merely by a similarity transformation. Any resulting inaccuracies (errors introduced by this approximation) have to be absorbed by the statistical appearance part of the feature model itself.

We aim to detect 10 points on the face: eye inner and outer corners, the eye centers, nostrils, and mouth corners. The PCA-based detector from our earlier work [17], [18] was replaced by a detector using Gabor filters [19], [20], [21]. Based on the invariant properties of Gabor filters a simple Gabor feature space, where illumination, orientation, scale, and translation invariant recognition of objects can be established, was proposed [22], [21]. A response matrix in a single spatial location can be constructed by calculating the filter responses for a finite set of different frequencies and orientations. The scale and orientation invariance is achieved via column or row-wise shifts of the Gabor feature matrix and the illumination invariance by normalizing the energy of the matrix.

In order to capture appearance variation over a whole population, a statistical model of the Gabor response matrix for each facial part (feature) is used. Initially, we opted for a cluster-based method called subcluster classifier [22], [23] which was eventually replaced by a Gaussian mixture model (GMM) [24]. The method of Figueiredo and Jain [25], using automatic selection of the number of components, provided the best convergence properties and classification results in the experiments conducted. Let us stress that the elements of the response matrix are complex numbers. A complex Gaussian distribution was previously studied in the literature and used in several applications [26]. We verified experimentally that complex-valued representation significantly outperforms the magnitude-only model [22]. In experiments, we output 200 best candidates per feature.

Transformation (Constellation) model. The task of face registration and the elimination of the capture effects calls for an introduction of a reference coordinate system. In our design, this space is affine and it is represented by three landmark points positioned on the face in the two-dimensional image space. We call this coordinate system "face space," for more details, see [18]. Any frontal face can be then represented canonically in the frame defined by the three points. As most of the shape variability and the main image acquisition effects (scale, orientation, translation) are removed, the faces become photometrically tightly correlated and the modeling of face/background appearance is therefore more effective. In such a representation, features exhibit a small positional variation of their face-space coordinates. A triplet of such corresponding features defines an affine transformation the whole face has to undergo in order to map in the face space. The transformation also provides evidence for/against the face hypothesis, since nonfacial points (false alarms) result in hypothesized transformations that have not been encountered in the training set. Not all triplets of features are suitable for the estimation of the transformation. Triplets which lead to a poorly conditioned transformation are excluded. For the 10 detected features, only 58 out of 120 all possible triplets were well-conditioned. The probability $p(T|face)$, T being the transformation from the face-space into the image-space, is estimated from the training set. Probability $p(T|nonface)$ is assumed uniform, for more details see [18]. The structure of the full algorithm is shown in Fig. 1.
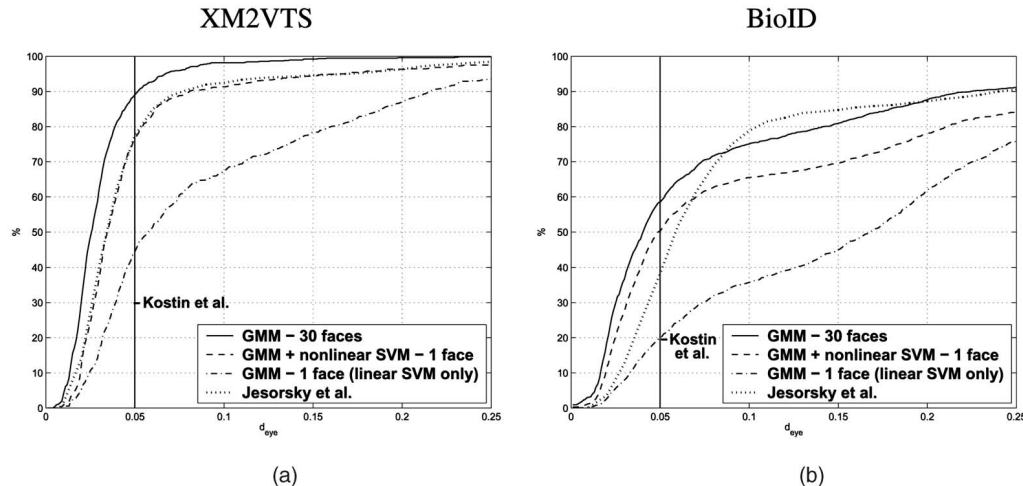
Fig. 2. Cumulative histograms of d_{eye} (see (1)).

The complexity of the brute force implementation is $O(n^3)$, where n is the total number of features detected in the image. In authentication scenarios, 360 degree orientation-invariance is not required as the head orientation is constrained. This allows the computational complexity to drastically decrease. After picking the first of the three correspondences, a region where the other two features of the triplet can lie is established as an envelope of all positions encountered in the training set. We call these regions *feature position confidence regions*. Since the probabilistic transformation model and the confidence regions were derived from the same training data, triplets formed by features taken from these regions have high transformation probability. All other triplets include at least one feature detector false alarm since such configurations did not appear in the training data. In our early experiments with the XM2VTS database, which contains images with uniform background, the use of confidence regions reduced the search by 55 percent [18]. In cluttered backgrounds, the reduction is drastic (speed-up factor up 1,000 times) since the majority of the nonfacial detected points lie outside the predicted regions.

Appearance Model. The appearance test is the final verification step which decides if the normalized data correspond to a face. Its output is a score which expresses the consistency of the image patch with the face class. Support Vector Machines (SVM) were successfully used in face detection in the context of the sliding window approach. In our localization algorithm, we use SVM as the means of distinguishing well-localized faces from background or misaligned face hypotheses using geometrically registered photometric data (by using face space) obtained from the previous stages of our algorithm. SVMs offer a very informative score (discriminant function value). We opted for cascaded classification: First, a linear SVM using 20×20 resolution templates preselects a fixed number of the fittest hypotheses. Then, a nonlinear third degree polynomial SVM in 45×60 resolution chooses the best localization hypothesis.

As training patterns for the first stage, face patches registered in the face space using manual ground truth features were used. The set of negative examples was obtained by applying the bootstrapping technique proposed by Sung and Poggio [27]. The first stage gives an ordered list of hypotheses, based on low resolution sampling. In our experiments, it was observed that this step alone is unable to distinguish between slightly misaligned faces. This is caused by the fact that downsampling to low resolution makes even slightly misaligned faces look almost identical. In order to increase the accuracy, we employ a high resolution classifier. Faces in higher resolution exhibit more visual detail and, therefore, a more complex classifier has to be used to cope with it. At this stage, we also redefined the face patch borders so that the patch contained mainly the eye region.

We tested several illumination correction techniques and adopted the zero mean and unit variance normalization. It is a very simple correction which does not actually remove any shadows from the face patches.

4 EXPERIMENTS

Commonly, the performance of face detection systems is expressed in terms of receiver operating curves (ROC), but the term "successful localization" is often not explicitly defined [28]. In localization, a measure expressing error in the spatial domain is needed. To allow us to make a direct comparison, we adopted the measure proposed by Jesorsky et al. [7]. The localization criterion is defined in terms of the eye center positions:

$$d_{eye} = \frac{max(d_l, d_r)}{||C_l - Cr||}, \qquad (1)$$

where C_l, C_r are the ground truth eye center coordinates and d_l, d_r are the distances between the detected eye centers and the ground truth ones. We established experimentally that, in order to succeed in the subsequent verification step [29], the localization accuracy has to be below $d_{eye} = 0.05$. In the evaluation, we treat localization with d_{eye} above 0.05 as unsuccessful. The advocated method was assessed and compared on the XM2VTS [30], BANCA [31], and BioID http://www.bioid.com face databases. These databases were specifically designed to capture faces in realistic authentication conditions and a number of methods have been evaluated on these data sets [1], [2]. For XM2VTS and BANCA, precise verification protocols exist which give an opportunity to evaluate the overall performance of the whole face authentication system. Although other publicly available face databases exist (like CMU, FERRET, etc.), evaluation on them is beyond the scope of this paper, due to the different purpose of the databases (nonfrontal faces, difficult acquisition conditions, multiple faces in the scene, poor resolution unsuitable for verification and recognition etc.). For training and parameter tuning, a separate set of 1,000 images from the worldmodel part of the BANCA database was used.

Evaluation of the XM2VTS database—frontal set. The d_{eye} measurements were collected for three methods—two variants of the proposed method and the reference method of Jesorsky et al. [7]. The two variants differ in the way the final face location hypothesis is selected. The simplest method (denoted as "linear SVM"), selects the most promising hypothesis by linear SVM. The second, "nonlinear SVM" uses a third degree polynomial kernel SVM in higher resolution to select the most promising hypothesis from the 30 best candidates, generated by the linear SVM. Fig. 2 shows that the GMM "nonlinear SVM" variant gives a similar
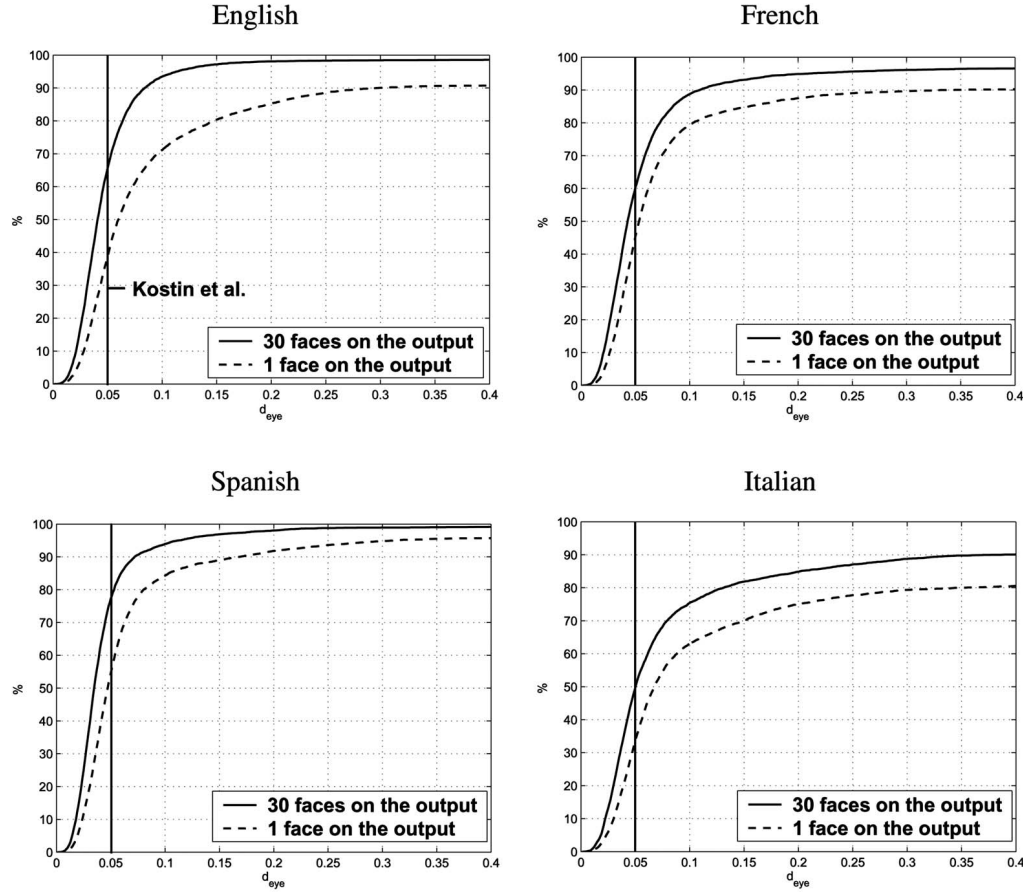
Fig. 3. Cumulative histograms of d_{eye} on the BANCA database.

performance to the baseline method of Jesorsky at $d_{eye} \leq 0.05$. The "nonlinear SVM" reduced the number of false negatives by 32.2 percent in comparison with the "linear-SVM." The curve denoted as "30 faces" shows d_{eye} distance of the hypothesis nearest to ground truth among the 30 candidates. It should also be noted that multiple localization hypotheses on the output do not present a problem in personal authentication scenarios, where person-specific information can help with the selection of the fittest hypothesis. The "nonlinear SVM" variant was also compared with the method of Kostin and Kittler [32]. Kostin's detector is a two-stage process. First, a holistic, sliding-window search produces a rectangular box where an SVM-based eye-detector is applied. At $d_{eye} \leq 0.05$ "nonlinear SVM" had a success rate of 78 percent compared to 30 percent of Kostin's detector.

Evaluation on the BioID database. The database consists of 1,521 gray-scale images. It is regarded as a difficult data set, which reflects more realistic conditions than the XM2VTS database. The same variants ("linear SVM," "nonlinear SVM," "Jesorsky") as in the XM2VTS case were evaluated, see Fig. 2b. The "nonlinear SVM" variant outperforms Jesorsky's method by 12 percent. Compared with Kostin detector at $d_{eye} \leq 0.05$, "nonlinear SVM" had a success rate of 51 percent compared to 20 percent of Kostin's detector.

Evaluation on the BANCA database. The BANCA database is a challenging data set designed for person authentication experiments. It was captured in four European locations in two modalities (face and voice). The subjects were recorded in three different scenarios, controlled, degraded and adverse over 12 different sessions spanning three months [31]. In total, 208 people were captured, half men and half women. Fig. 3 depicts the "nonlinear SVM" variant results on the English, French, Spanish, and Italian

parts of the database (denoted as "1 face on the output"). At $d_{eye} \leq 0.05$ on the English part, the "nonlinear SVM" had a success rate of 39 percent compared to 29 percent of Kostin's detector. No other face detection results on the BANCA have been reported yet, but the database and the verification protocol are publicly available.

Feature detector evaluation. We measured the accuracy of the Gabor-based feature detectors in order to assess their contribution to the overall performance. For training, faces were normalized to an intereye distance of 40 pixels (i.e., scale and orientation was removed). We define a successful feature localization as the situation when among all the detected features in the image there exists at least one with $d_F \leq 0.05$, where d_F is defined in a similar manner as d_{eye}:

$$d_F = \frac{||(x_F, y_F) - (x_G, y_G)||}{||C_l - C_r||} \qquad (2)$$

(x_F, y_F) are the coordinates of the detected feature F, (x_G, y_G) the ground truth coordinates of the feature F, and C_l, C_r the ground truth coordinates of the left and right eye centers. Tables 1 and 2 depict several performance measurements on all three databases introduced earlier. It is clear that, if we decided to use only the eye-centers, we would always localize significantly fewer faces than with the method requiring any triplet of features—the difference is visible by comparing the entries in Table 2. If we assume that every localized triplet would lead to a successful localization, then by using our method, the total performance on XM2VTS would be 88.3 percent, i.e., the improvement achieved would be 13.8 percent. In realistic databases, our method gains even bigger performance boost (the most on the BANCA database). Please note that we

TABLE 1
Performance of Each Feature Detector (in %), Feature Matrix with Five Orientations and
Four Scales, GMM = Gaussian Mixture Model, SCC = Subcluster Classifier

Feature description	Feature label	XM2VTS SCC	XM2VTS GMM	BioID SCC	BioID GMM	Banca SCC	Banca GMM
Outer left eye corner	1	51.4	56.1	50.2	55.6	36.9	41.4
Left eye center	2	76.3	84.2	66.1	67.6	46.9	60.3
Inner left eye corner	3	71.3	70.9	43.7	51.2	38.1	44.5
Inner right eye corner	4	49.8	50.9	37.1	39.1	36.6	44.0
Right eye center	5	76.1	84.9	68.9	61.1	63.7	67.4
Outer right eye corner	6	60.0	64.2	50.5	54.8	28.2	34.3
Left nostril	7	72.2	70.4	48.4	29.5	41.3	54.8
Right nostril	8	76.8	75.5	49.4	34.5	61.2	63.6
Left mouth corner	9	34.3	54.2	40.7	40.0	45.9	49.6
Right mouth corner	10	47.2	45.8	46.4	48.7	58.5	61.8

actually reached the top theoretical performance in the case of XM2VTS database with 30 faces on the output (see Fig. 2).

5 CONCLUSIONS

We have presented a bottom-up algorithm that successfully localized faces using a single gray-scale image. We have assessed the localization performance of the proposed method on three benchmarking data sets. In Section 4, we demonstrated that, regarding accuracy, the advocated approach outperforms the baseline method of Jesorsky et al. [7] and also it is superior to a typical representative of the sliding window approach. In a coarse resolution, linear SVM-model cannot be relied upon if only one localization hypothesis on the output is needed. For such purposes, a third degree polynomial SVM trained in finer resolution dramatically improved the results. Other appearance models and classifiers can be exploited in the appearance test stage in the future, leaving room for further improvement. We also showed that feature detectors alone could not lead to a satisfactory performance. Nevertheless, when combined with the proposed constellation and appearance models, the performance boost achieved is dramatic. Our research implementation does not focus on speed (approximately 13 seconds/image on Pentium 4, 2.8GHz); nevertheless, we believe that real-time performance is achievable. The Gabor-bank feature detection is the critical part, however, the algorithm is easily parallelizable and dedicated parallel hardware is available.

ACKNOWLEDGMENTS

TABLE 2
Triplets and Eye Pair Detection Rates (in %), Feature Matrix
with Five Orientations and Four Scales, A = At Least One
Well-Posed Triplet Detected, B = Both Eye Centers Detected,
GMM = Gaussian Mixture Model, and SCC = Subcluster Classifier

	XM2VTS SCC	XM2VTS GMM	BioID SCC	BioID GMM	Banca SCC	Banca GMM
A	86.7	88.3	76.1	73.4	75.2	81.4
B	62.3	74.5	51.5	48.6	32.1	44.0

REFERENCES

[1] J. Matas, M. Hamouz, K. Jonsson, J. Kittler, Y. Li, C. Kotropoulos, A. Tefas, I. Pitas, T. Tan, H. Yan, F. Smeraldi, J. Bigun, N. Capdevielle, W. Gerstner, S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Comparison of Face Verification Results on the XM2VTS Database," *Proc. Int'l Conf. Pattern Recognition,* vol. 4, pp. 858-863, 2000.
[2] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vanderdorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, B. Kepenekci, F. Tek, G. Akar, F. Deravi, and N. Mavity, "Face Verification Competition on the XM2VTS Database," *Proc. Fourth Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 964-974, 2003.
[3] E. Hjelmås and B.K. Low, "Face Detection: A Survey," *Computer Vision and Image Understanding,* vol. 83, pp. 236-274, 2001.
[4] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 511-518, 2001.
[5] H. Rowley, S. Baluja, and T. Kanade, "Rotation Invariant Neural Network-Based Face Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 38-44, 1998.
[6] M.-H. Yang, D. Roth, and N. Ahuja, "A SNoW-Based Face Detector," *Advances in Neural Information Processing Systems,* vol. 12, pp. 855-861, 2000.
[7] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz, "Robust Face Detection Using the Hausdorff Distance," *Proc Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 90-95, 2001.
[8] K. Yow and R. Cipolla, "Feature-Based Human Face Detection," *Image and Vision Computing,* vol. 15, pp. 713-735, 1997.
[9] M. Weber, M. Welling, and P. Perona, "Unsupervised Learning of Models for Recognition," *Proc. Sixth European Conf. Computer Vision,* vol. 1, pp. 18-32, 2000.
[10] D. Cristinacce and T. Cootes, "Facial Feature Detection Using AdaBoost with Shape Constraints," *Proc. British Machine Vision Conf.,* vol. 1, pp. 213-240, 2003.
[11] M. Weber, W. Einhauser, M. Welling, and P. Perona, "Viewpoint-Invariant Learning and Detection of Human Heads," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 20-27, 2000.
[12] T. Cootes, D. Cooper, C. Taylor, and J. Graham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, no. 1, pp. 38-59, 1995.
[13] T. Cootes and C. Taylor, "Constrained Active Appearance Models," *Proc. Int'l Conf. Computer Vision,* vol. 1, pp. 748-754, 2001.
[14] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Computers,* vol. 42, no. 3, pp. 300-311, Mar. 1993.
[15] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.
[16] C. Kotropoulos and I. Pitas, "Face Authentication Based on Morphological Grid Matching," *Proc. IEEE Int'l Conf. Image Processing,* vol. 1, pp. 105-108, 1997.
[17] J. Matas, P. Bílek, M. Hamouz, and J. Kittler, "Discriminative Regions for Human Face Detection," *Proc. Asian Conf. Computer Vision,* pp. 604-609, Jan. 2002.
[18] M. Hamouz, J. Kittler, J. Matas, and P. Bílek, "Face Detection by Learned Affine Correspondences," *Proc. Joint IAPR Int'l Workshops Structural and Syntactic Pattern Recognition SSPR02 and SPR02,* pp. 566-575, Aug. 2002.
[19] G.H. Granlund, "In Search of a General Picture Processing Operator," *Computer Graphics and Image Processing,* vol. 8, pp. 155-173, 1978.
[20] T.S. Lee, "Image Representation Using 2D Gabor Wavelets," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 10, pp. 959-971, Oct. 1996.

[21] V. Kyrki, J.-K. Kamarainen, and H. Kälviäinen, "Simple Gabor Feature Space for Invariant Object Recognition," *Pattern Recognition Letters,* vol. 25, no. 3, pp. 311-318, 2004.

[22] J.-K. Kamarainen, V. Kyrki, H. Kälviäinen, M. Hamouz, and J. Kittler, "Invariant Gabor Features for Face Evidence Extraction," *Proc. MVA2002 IAPR Workshop Machine Vision Applications,* pp. 228-231, 2002.

[23] M. Hamouz, J. Kittler, J.-K. Kamarainen, and H. Kälviäinen, "Hypotheses-Driven Affine Invariant Localization of Faces in Verification Systems," *Proc. Fourth Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 276-286, 2003.

[24] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, and H. Kälviäinen, "Affine-Invariant Face Detection and Localization Using GMM-Based Feature Detector and Enhanced Appearance Model," *Proc. Sixth Int'l Conf. Face and Gesture Recognition,* pp. 67-72, 2004.

[25] M. Figueiredo and A. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 3, pp. 381-396, Mar. 2002.

[26] N.R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," *The Annals of Math. Statistics,* vol. 34, pp. 152-177, 1963.

[27] K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, pp. 39-50, Jan. 1998.

[28] M. Yang, D.J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 1, pp. 34-58, Jan. 2002.

[29] M. Sadeghi, J. Kittler, A. Kostin, and K. Messer, "A Comparative Study of Automatic Face Verification Algorithms on the BANCA Database," *Proc. Fourth Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 35-43, 2003.

[30] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," *Proc. Second Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 72-77, 1999.

[31] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran, "The BANCA Database and Evaluation Protocol," *Proc. Fourth Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* pp. 625-638, 2003.

[32] A. Kostin and J. Kittler, "Fast Face Detection and Eye Localization Using Support Vector Machines," *Proc. Sixth Int'l Conf. Pattern Recognition and Image Analysis: New Information Technologies,* pp. 371-375, 2002.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.