Published in EUROCON 2003 (Computer as a Tool. The IEEE Region 8) 2, 365-369, 2003
which should be used for any reference to this work

Comparison of feature extraction techniques for face verification using elastic graph matching on low-power mobile devices

J.-L. Nagel, P. Stadelmann, M. Ansorge, and F. Pellandini
Institute of Microtechnology, University of Neuchâtel
Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland
email: jean-luc.nagel@unine.ch

***Abstract*—Biometric face verification using elastic graph matching is considered and two feature extraction techniques used as a pre-processing stage are compared. Low-complexity vs. performance is of particular interest for mobile devices. A modified technique based on a normalization of mathematical morphology feature is introduced and show a performance comparable to that of Gabor features at a much-reduced computational complexity.**

***Index Terms*--Biometrics, face verification, feature extraction, graph matching, low-power.**

I. Introduction

BIOMETRIC face verification increases the security of the access to a device by verifying physiological characteristics of the user whereas classical security techniques use either something that the user possesses (e.g. a key) or that the user knows (a password, or personal identification number, PIN). Alternatively at the same security level face-based access control can provide with more intuitive and user-friendly user interfaces than traditional password-based systems.

Performing face verification on a mobile device represents however a difficult challenge due to the unconstrained image acquisition conditions (viewpoint, illumination) and the power consumption limitation. Consequently robust yet power efficient supporting technologies need to be devised.

Seeking this goal the next section motivates the use of biometric identity verification on mobile devices and lists the most common categories of algorithm readily proposed. A provably efficient algorithm is then selected in Section III and two associated feature extraction techniques that show large differences in computational complexity are discussed in Section IV. Finally performances are measured on a standard database.

II. Biometrics for mobile devices

Convergence of personal digital assistants (PDA) and of mobile phones can definitely be observed, the latter containing more and more features of the first. The trend is naturally to have a single device with extended capabilities (wireless internet, phone, camera, GPS) resulting in an increased number of potentially sensitive and private data being stored on the device and transferred to remote locations. Thus the need for more effective security is clearly present. To achieve security and privacy of data during transmission all communications will be encrypted in third generation mobile phones. Similarly the subscriber account is protected by codes that are exchanged between the phone and the network. However none of these methods protects the access to the terminal itself.

According to a survey carried out by Clark [1] on user attitudes towards mobile phone security, around 80% of the current mobile phone users believe that enhanced security would be good or very good. The lack of convenience and of confidence in PIN codes is most often mentioned as an explanation of why the subscribers are not using them. Not surprisingly the leading biometric technology that those users would be ready to use is fingerprint (74% of positive responses). This can be explained by the fact that the vast majority of users already had some experience with this technology whereas it is generally not the case with face verification. However face verification is advantageous due to the direct presence of a digital image sensor in new generations of devices, thus sharing its cost between multiple applications. Finally the reluctance of users to disclose their biometric templates for everyday applications can be reduced by the fact that templates are stored locally and not on a remote server. This last point however involves that the matching of a test image against a reference template is done on the device itself. Since the image acquisition is carried out with the user holding the mobile device with the camera both the viewpoint and the lighting environment are unconstrained. The algorithm must therefore compensate for scale variations, rotations and homotheties. Furthermore features extracted from the image should be so-called illuminant invariant.

III. FACE VERIFICATION ALGORITHMS

Similarly to the holistic nature of face recognition by the brain – i.e. the face image being treated as a whole instead of independent facial parts [2] – template-based algorithms seem more robust to variation of pose and to occlusions than feature-based methods. Indeed in the latter the false detection of an important facial feature can lead to a false rejection of the client whereas in template-based approaches the importance of each facial feature is distributed over the whole template and a local discrepancy does not necessarily lead to a false rejection.

Compensation for viewpoint variations is usually handled with deformable models. True deformable approaches are indeed reported to perform better than rigid ones when faces are slightly turned away from the camera [3]. In this case most of the parts of the head still projects the same patterns with slightly different configuration properties. For larger rotations however (mainly out-of-plane rotations) the projection of facial parts produces completely different patterns, which can no longer be handled by simple deformations. This is especially true for salient facial features such as the nose.

Besides, some methods rely on general pattern extraction techniques – i.e. used in diverse pattern recognition problems up to a difference of parameters – while other rely on problem specific techniques that are learned from a training data set. The impact on the performance of one category over the other is not clear. However it can be assumed that the generalization efficiency of the trained patterns to a new data set is strongly dependent on the quality of this set. Thus in the case of small training sets untuned features are preferable.

The now famous work of Turk and Pentland [4] on Eigenfaces is based on an earlier work by Kirby and Sirovich [5] who used the Karhunen-Loéve (KL) transform to characterize human faces. The baseline KL transform is clearly problem-tuned and do not allow deformations unless they are present in the training set. A solution to view-point dependence is the removal of shape information prior to the eigenface calculation by using several control points related to fiducial landmarks in order to warp the image face to a standard face [6]. Although handling efficiently viewpoint variations this solution has two major drawbacks: the warping requires a costly interpolation and the method falls back in the feature-based category due to the use of fiducial landmarks.

Lades [7],[8] introduced the "Dynamic Link Architecture" method later called "Elastic (Bunch) Graph Matching" (EGM). In EGM a face is holistically represented by labeled graphs – i.e. nodes connected by edges – containing simultaneously global (i.e. topographical) and local information. The local template – extracted with untuned pattern recognition techniques such as Gabor filters – associated to a node is not related to a particular fiducial

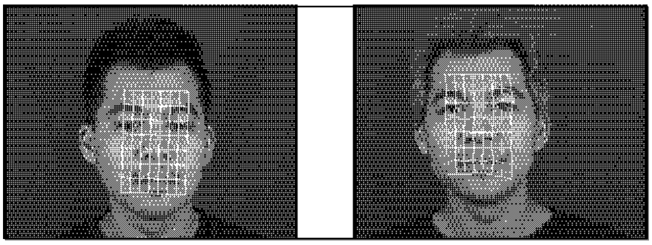

Fig. 1. Matching of a test image against a reference. Left: rigid reference graph laid on the inner part of the face; right: best matching test graph minimizing the feature distance and a deformation penalty. The feature distance leads to an acceptance of the subject.

point but is sampled on a regular grid. When a novel view is presented to the system a correspondence (i.e. an isomorphism) is searched between the two images allowing elastic deformations of the graph. A distance is minimized, which is obtained by computing a metric between local reference and test features nodes with the addition of a penalty accounting for the graph deformations.

Being holistic, untuned and handling intrinsically deformations this method thus fulfils all the requirements for mobile applications. The robustness to lighting variations is however dependent on the quality of the features used as local templates. A trade-off between efficiency and computational complexity needs to be performed as will be seen in the next section.

IV. GABOR VS. MORPHOLOGICAL FEATURES

Information used at each node of the attributed graph may be extracted from the image by various pattern analysis techniques such as local statistics of a region, edge/contour extraction, moments, transform features, texture or shape. Lades used Gabor filtered images due to its similarity with cortical cells [9] and its extended use in texture analysis. Kotropoulos, Tefas and Pitas [10] alternatively used mathematical morphology as a shape extraction tool for it is much less computationally expensive. These two approaches are compared here simultaneously with the introduction of modified morphological features that are shown to perform almost as well as Gabor features at a much-reduced calculation cost.

A. *Gabor filterbanks*

Basically a Gabor filter is a sinusoidal plane wave of given frequency and orientation modulated by a gaussian envelope. Gabor filters are particularly interesting for time-frequency signal analysis because they present an interesting compromise between spatial (or time) and frequency extension. Although many authors agree that indeed Gabor filters are performing admirably for image analysis they don't all agree on the form to use and how to handle the images resulting from filtering. A special case of complex Gabor filter is obtained with the major axis of the gaussian having the same orientation as the sine plane wave:

$$f(x,y) = \exp\left\{-\frac{1}{2}\left(\frac{x_r^2}{s_x^2}+\frac{y_r^2}{s_y^2}\right)\right\}\exp\left(i2pw_r x_r\right) \quad (1)$$

with x_r and y_r being the rotated coordinates.

$$\begin{pmatrix} x_r \\ y_r \end{pmatrix} = \begin{pmatrix} \cos f & \sin f \\ -\sin f & \cos f \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \quad (2)$$

The Gabor filter above is thus a gaussian shaped band-pass filter with minor axis oriented at an angle f, centered at the radial frequency w_r, and s_x, s_y being space constants of the Gaussian envelope along x_r and y_r respectively. Combining multiple gaussian filters with shifted frequency responses and octave band structure forms a filterbank that possibly uniformly covers the spatial-frequency domain. This allows any finite-dimensional function to be expressed as a sum of shifted Gabor functions (Gabor expansion).

It is sometimes proposed [11],[12] to use the magnitude of the image filtered with the real version of the filter. However it can be observed that the complex part of the filter being actually the Hilbert transform of the real part, the magnitude of the complex response will give the instantaneous envelope of the filtered signal, whereas the magnitude of the cosine filtered image will lead to a rectified signal that is only a weak approximation of the envelope. Consequently the latter solution would need a smoothing although requiring half of the filtering. A negative impact on the convergence of the matching algorithm might be observed when using undulating responses.

In this communication 18 complex filters corresponding to 6 orientations and 3 frequency bands were used similarly to the configuration described by Duc [13].

B. *Morphological features*

Mathematical morphology is a technique widely used in image processing for different purposes: segmentation, shape analysis, filtering, and texture analysis [14]. The two basic operations called erosion and dilation are used to define many other morphological operations (e.g. opening and closing). Given an object X and a structuring element B (both binary) the erosion and dilation are defined as:

$$E_B(x) = \{x | (B_X \subseteq X)\},\ D_B(x) = \{x | (B_X \cap X) \neq 0\} \quad (3)$$

where B_X denotes the structuring element translated at position x. This means that the eroded image is formed of all the points x such that B_X is included in X. Similarly the dilation is formed of all the points x such that B_X hits X i.e. that have a non-empty intersection.

These operations can be extended to grayscale images. An interesting special case occurs when using a flat binary structuring element to dilate and erode a grayscale image:

$$\begin{aligned} E_G(x) &= \min\{X(x+z), z \in G, (x+z) \in X\} \\ D_G(x) &= \max\{X(x-z), z \in G, (x-z) \in X\} \end{aligned} \quad (4)$$

where G is the support of the grayscale structuring element. This operation thus simply consists in finding the maximum/minimum on the support of the SE. A set of nine multiscale flat discs are used here as structuring elements:

$$G(z) = \begin{cases} 1, \|z\| < s \\ 0, else \end{cases} \quad (5)$$

where s is the radius of the n-th multiscale structuring element and is ranging from 1 to 9 pixels. The use of multiscale flat structuring elements further allows optimizing the computations by reusing results calculated at the previous level because it is entirely included in the new

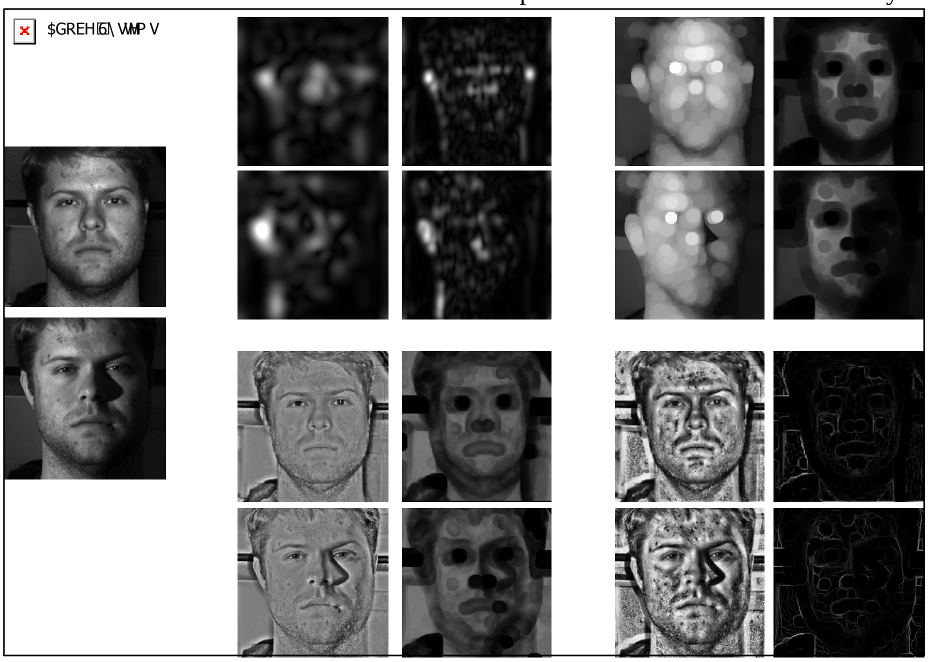

Fig. 2. Features extracted from two images taken from the Yale database [16] with frontal and side illumination. a) Original images; b) Gabor filtered images with two frequency bands; c) dilated and eroded images using a flat disc with $s = 9$; d) normalized morphology using (7); e) normalized morphology using (8).

one. These features are thus well suited for parallel hardware implementations [15]. These morphological features are grouped into a feature vector similarly to the Gabor features:

$$J(x,y) = \{D_9(x,y);...;\ D_1(x,y); I(x,y);\ E_1(x,y);...;\ E_9(x,y)\} \quad (6)$$

Compared to Gabor filtered images the images obtained by mathematical morphology are strongly illumination dependent due to the non-removal of the DC component. Thus normalization needs to be applied either to the original image or to the morphological features to create new illumination invariant features. Applying a model-independent normalization on the original image e.g. by considering a local neighborhood of $N \times N$ pixels and applying normalization based on statistics of this area imply the use of more data (N^2) than when performing the normalization using the statistics of the dilated and eroded values at a single point. Indeed these values account for an equivalent size of windows (typically $N = 19$) but with a reduced set of values.

A first normalization tested (7) was obtained by dividing each morphological value by the average of the 19 values. Another possibility (8) was the division of each value by the two extremes leading to only 17 normalized values.

$$J'_n(x,y) = J_n(x,y)/M(x,y), M(x,y) = 1/19 \sum_{n=1}^{19} J_n(x,y) \quad (7)$$

$$J''_n(x,y) = (J_n(x,y) - J_{19}(x,y))/(J_1(x,y) - J_{19}(x,y)) \quad (8)$$

Examples of images obtained with Gabor filters, baseline dilation and erosion, and with normalized morphology are shown in Figure 2 for two different images.

V. Results

Technology evaluation [17] was performed using the XM2VTS face database using the Lausanne protocol configuration II [18]. From the 4 training samples available only the first one was used to build a reference template. The 25×8 impostor evaluation samples were used for a-priori normalization of both genuine and impostor distances using the z-norm technique [19].

Results on the XM2VTS database (see Figure 3) show that the normalized morphological features perform better than the original morphological features, even though this database does not present large variations in illumination! This can be understood because even if no gradients are present the average illumination is still slightly varying from image to image. In addition we can observe that the normalization performed in (8) seems to be more effective than the normalization in (7).

Gabor features outperform morphological features with an equal-error-rate (EER) slightly above 6% (7.5% for the best normalized morphology features). For false-acceptance-rate (FAR) below 2% however the false-rejection-rate (FRR) is almost similar between both implementations. The difficulty for normalized

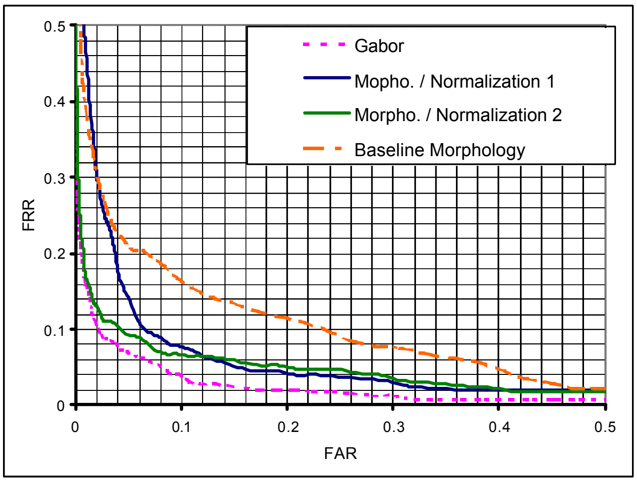

Fig. 3. Receiver Operating Curve (ROC)

morphological features to reach low levels of FRR can be explained by the convergence of the iterative matching algorithm, which is less smooth than with Gabor features. It means however that the morphological are very efficient at discriminating faces but that a more robust matching approach is necessary.

The normalized morphology features were especially designed for robust matching in varying illumination conditions (color and intensity). Because the XM2VTS does not allow testing directly this robustness scenario tests were performed by enrolling a volunteer and testing it in different lighting conditions. Only qualitative results were obtained showing interesting performance when changing the light direction and the type of illuminant (e.g. sun, artificial).

It can be assumed that features are extracted on the whole image since the scanning operation carried out during rigid matching [7] will reach a large number of the possible pixels. Consequently the Gabor filtering is efficiently done in the frequency domain. According to the configuration used this involves calculation of 18 complex FFTs, 18 pixel-by-pixel complex multiplications, and 18 complex inverse FFTs. The magnitude calculation needs also a consecutive post-processing. Comparatively the morphological features require only 361 maximum and minimum calculation per pixel using the structuring elements previously described. The normalization can be done as a post-processing requiring only 19 divisions per image position.

VI. Conclusion

In this communication the Elastic Graph Matching techniques was first identified as an efficient candidate for mobile applications of face verification. Additionally robust feature extraction was shown to be necessary while maintaining a low-complexity due to the reduced computational power available on the mobile device.

Two types of feature extraction techniques were compared and a novel technique based on mathematical morphology was shown to perform better than the standard approach. Moreover the best normalized morphological features reached performances close to those obtained with Gabor features although the complexity of the prior is much lower.

The supporting techniques presented here could be extended to other modalities requiring low-complexity. Namely face-based applications such as audio-visual speech recognition might benefit from this directly. Other applications include gaze tracking and hand tracking.

ACKNOWLEDGMENT

This work was part of the SmartPix and SmartCam projects supported by the Swiss Center for Electronics and Microtechnology, Inc. (CSEM), Neuchâtel and Zurich, Switzerland, under Grant OIH3.

REFERENCES

[1] N. L. Clarke, S. M. Furnell, P. M. Rodwell, and P. L. Reynolds, “Acceptance of subscriber authentication methods for mobile telephony devices”, *Computers & Security*, vol. 21, no.3, pp. 220-228, 2002.

[2] M. J. Farah, M. Drain, and J. N. Tanaka, “What Is ‘Special’ About Face Perception”, *Psychological Review*, nol. 105, no. 3, pp. 482-498, 1998.

[3] J. Zhang, Y. Yan and M. Lades, "Face Recognition: Eigenface, Elastic Matching, and Neural ets", *Proc. of the IEEE*, vol. 85, no. 9, pp. 1423-1435, September 1997.

[4] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, nol. 3, no. 1, 1991, pp. 71-86.

[5] M. Kirby and L. Sirovich, “Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces”, *IEEE PAMI*, vol. 12, no. 1, pp. 103-108, 1990.

[6] I. Craw, N. Costen, T. Kato, G. Robertson and S. Akamatsu, "Automatic Face Recognition: Combining Configuration and Texture", *Proc. of Int'l Workshop on Automatic Face- and Gesture-Recognition*, Zürich, Switzerland, June 26-28, 1995.

[7] M. Lades, J. C. Vorbrügen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Computers*, vol. 42, no. 3, March 1993, pp. 300-311.

[8] M. Lades, “Face Recognition Technology”, in *Handbook of Pattern Recognition and Computer Vision*, 2nd edition, Eds. C. H. Chen, L. F. Pau and P. S. P. Wang, World Scientific Publishing Company, 1998, pp. 667-683.

[9] D. A. Pollen, and S. F. Ronner, “Visual Cortical Neurons as Localized Spatial Frequency Filters”, *IEEE Systems, Man, and Cybernetics*, vol. 13, no. 5, pp. 907-915, September-October 1983.

[10] C. Kotropoulos, A. Tefas and I. Pitas, " Morphological Elastic Graph Matching applied to frontal face authentication under well-controlled and real conditions", Pattern Recognition, vol. 33, no. 12, pp. 1935-1947, 2000.

[11] A.K. Jain, and F. Farrokhnia, “Unsupervised Texture Segmentation Using Gabor Filters”, *Pattern Recognition*, vol. 24, no. 12, pp. 1167-1186, 1991.

[12] T. Randen, and J. H. Husøy, “Filtering for Texture Classification: A Comparative Study”, *IEEE PAMI*, vol. 21, no. 4, pp. 291-310, April 1999.

[13] B. Duc, S. Fischer, and J. Bigün, “Face Authentication with Gabor Information on Deformable Graphs”, *IEEE Image Processing*, vol. 8, no. 4, pp. 504-516, April 1999.

[14] J. Goutsias, L. Vincent, and D.S. Bloomberg, “Mathematical Morphology and its Applications to Signal Processing”, *Series "Computational Imaging and Vision"*, Kluwer, Dordrecht, 2000.

[15] P. Stadelmann, J.-L. Nagel, M. Ansorge, F. Pellandini, “A Multiscale Morphological Coprocessor for Low-Power Face Authentication”, *Proc. of XIth European Signal Processing Conference (EUSIPCO 2002)*, Toulouse, France, vol. 3, pp. 581-584, Sept. 03-06, 2002.

[16] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE PAMI, vol. 23, no. 6, pp. 643-660, 2001.

[17] P.J. Phillips, A. Martin, C.L. Wilson, and M. Przybocki, “An Introduction to Evaluating Biometric Systems”, *IEEE Computer*, pp. 56-63, February 2000.

[18] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, “XM2VTSDB: the Extended M2VTS database”, *in Proc. of the 2nd international conference on Audio- and Video-based Biometric Person Authentication (AVBPA'99)*, Washington DC, U.S.A., pages 72-77 1999.

[19] S. Bengio, J. Mariéthoz, and S. Marcel, “Evaluation of biometric technology on XM2VTS”, *Technical Report IDIAP-RR 01-21*, IDIAP,Martigny, Switzerland, 2001.