

# Assuring liveness in biometric identity authentication by real-time face tracking

J. Bigun, H. Fronthaler, and K. Kollreider  
Halmstad University, SE-30118, Sweden

**Abstract**—A system that combines real-time face tracking as well as the localization of facial landmarks in order to improve the authenticity of fingerprint recognition is introduced. The intended purpose of this application is to assist in securing public areas and individuals, in addition to enforce that the collected sensor data in a multi modal person authentication system originate from present persons, i.e. the system is not under a so called play back attack. Facial features are extracted with the help of Gabor filters and classified by SVM experts. For real-time performance, selected points from a retinotopic grid are used to form regional face models. Additionally only a subset of the Gabor decomposition is used for different face regions. The second modality presented is texture-based fingerprint recognition, exploiting linear symmetry. Experimental results on the proposed system are presented.

**Keywords**— face tracking; Multi modal person authentication; Biometric identity authentication; Support Vector Machines; Gabor decomposition; Log-polar sampling; Fingerprint recognition.

## I. INTRODUCTION

Face localization or face tracking has gained increased attention [11], [13], along with biometrics in general. A class of powerful face recognition features are Gabor filters [6], [12], [18] which have impulse responses resembling those of simple cells, [8], [17], in visual cortex. A sparse log-polar retinotopic grid can be used, [18]–[21], to sample the Gabor responses at sparse points of an image to accelerate the feature extraction. Jumping to assumed points of interest, [22], instead of exhaustive search, is typical to biological vision systems from which our Retinotopic Vision approach is inspired. Additional time-savings enabling real-time performance in face tracking are obtained if a motion detection stage excludes motionless parts from the extraction [19]. An additional performance factor is to employ only specific frequency/orientation channels for modelling a facial region or landmark. Besides considering the speed issue, also, recognition rates can be raised by reducing the amount and adapting the range of frequencies for the decomposition. That the usage of all frequency channels is not the best strategy, is also confirmed by other authors, e.g. [7].

The benefits of taking multiple biometrics into account are being confirmed continuously in literature [1], [9]. A recent development is to use quality signals to improve multimodal performance in biometrics, [3]. An authentication in shifting environmental conditions involving several modalities in the decision process is shown to be more robust than any single-source authentication. Additionally, combing several traits decreases the success probability of a willful forging. The

system proposed here combines real-time face tracking and facial landmark localization as a prerequisite for fingerprint recognition.

This paper is structured as follows: In the next chapter, some application scenarios will be introduced. Chapter three and four will detail the applied face-tracking and facial landmark localization systems, whereas chapter five outlines the main points of the employed fingerprint recognition system. Chapter six and seven are devoted to the experiments and their subsequent discussion.

## II. APPLICATION SCENARIOS

Person - and especially face tracking is beneficial in various situations where the gap between an advanced security level and minimum client interference has to be bridged.

An example scenario would be a doctor trying to enter the surgery room. In that case one or more cameras could track an approaching person and the machine supervisor would try to determine whether the person is authorized to enter the restricted area. At best the doctor can pass without any further system interaction. In case no clear decision can be made based on the response of the first assessment, a second machine expert, which is specialized in another recognition trait, can be included in the process. For this purpose a fingerprint sensor, either stationary or embedded in a mobile device carried by the client, is applied to decide on access authorization. This second modality is activated depending on the response of the first one. The benefit of this process is the enhanced authenticity of the fingerprint, since it implies a high degree of certainty that the fingerprint was taken from a real person, because the fingerprint signal channel is opened only if the face channel signals an alive person. This aspect, directly mapped to a surveillance application would imply that a forged fingerprint, whether accepted or not, could be connected to the impostor's face and length of body parts that can be obtained from a designated camera system.

A related system could also be adapted in the event, where people should be protected from entering certain premises. An example for this would be an area that is only innocuous for humans when protective clothing is worn. In this situation, the proposed system could verify the concomitance of such clothing.

## III. TRACKING CONCEPT

The proposed face-tracking system measures image properties of a face [18], [19] non-uniformly via a retinotopic

grid. This sparse log-polar arrangement of points is shown in figure 1 and 4. Generally, only the features extracted at those points are available for processing. As indicated, not all points are used here to achieve real-time performance. Instead, the relevant points are dynamically selected.

While detecting and tracking a face, the system undergoes two different states, which are defined here next.

*Outer state:* The system is expecting a face coming from outside the grid as shown on the left in figure 1.

*Inner state:* Though not exactly in the center, the face is assumed to be fully inside the coverage of the grid as indicated on the right in figure 1



Fig. 1. Outer state (left) and inner state(right)

Note that the actual coverage of the grid is larger than appears in figure 1 since feature extraction at these points includes a neighborhood of the grid points. Depending on the state, different face regions involving different grid points (hinted by crosses in figure 1) are most interesting. Here face regions are modelled with the help of Support Vectors, [5], classifying Gabor responses at corresponding points. In the operational phase, trained machine experts are consulted to decide, if necessary, the direction and distance to move a face towards the grid center is estimated. In practice this means that the pan-tilt mechanism is controlled by the camera. The components of the tracking system are sketched in figure 2:

Assume that the system is *outer state*. A certain set  $\{O_i\}$  of outermost grid points is considered for feature extraction. If a face approaches from the right hand-side (compare to figure 1), then the machine-expert for detecting any right hand-side facial region will respond (expert 4 in figure 2). As a consequence the camera will be moved towards the assumed face center, which is to the right direction at a certain angle  $\alpha$ . At the same time the system changes its state (to *inner state*). A different set  $\{I_i\}$ , comprising outermost as well as interior grid points, is now involved in the feature extraction. An expert assuming the face in the center of the grid will respond (expert 7), and in that case the current frame, referred to as prepared frame in the following, is stored for further processing. A small rectangular region around the center of the prepared frame is then supposed to contain both eyes. This may be evaluated by visual inspection afterwards should any verification and/or human intervention be necessary: If the rectangle contains both eyes (more than 50% of each eye), the system is assumed to have accomplished to track the face satisfyingly. The prepared frame is suitable for authentication. When the face leaves the center towards the, e.g. left, border of the grid, the expert detecting any right

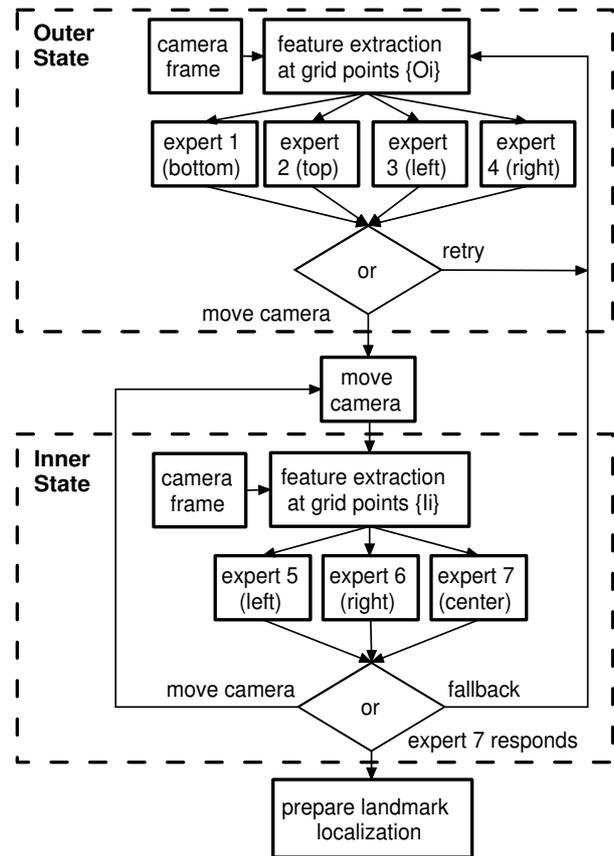


Fig. 2. Flow chart of the Tracking system

hand-side facial region in *inner state* will respond (expert 6). As a consequence the camera will move to the left with a certain angle  $\beta$ , to recenter the face. Again a frame is stored, provided that the responsible expert suggests to (expert 7). If the system loses the face from the center during *inner state*, it is designed to fall back to *outer state*.

In comparison to [19], the face is immediately centered after one expert is responding, while it can take several jumps in a system as proposed there. Furthermore, speed gains are achieved by picking out specific features at only some points of the grid for each expert. This is further detailed in the next chapter.

Further effort is put on facial landmark localization in the prepared frames. The landmarks to be detected are the left and the right eye. This is treated as a separate system working offline, nevertheless being very time-efficient, since the approximate position of the single eyes are known (rectangular areas half). Preferably the system controller switches to work on landmark localization straight after a single frame was stored by the tracking system, or it keeps the tracking ongoing while preserving several prepared frames for later processing. Both the tracking and the landmark localization systems can be distributed to different machines to maintain real-time performance. The schematics of the landmark localization system are displayed in figure 3:

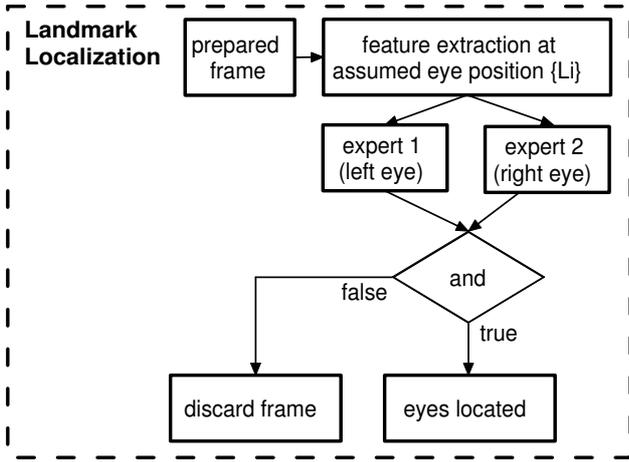


Fig. 3. Landmark localization system

Given a prepared frame, the system jumps to the assumed location, either of the right or left eye. The grid is randomly displaced in a restricted area around each eye's assumed position. Each time features are extracted at a different set  $\{L_i\}$  of the grid and classified by the according landmark expert (for example expert1 for the left eye). Initially the grid is placed into the center of one half the rectangle size, which is also the optimal position for that eye in a prepared frame. As more trials have to be made to locate the landmark more precisely, the considered area for displacing the grid randomly grows, even if the grid is constrained by one rectangle half. In this way the search for the landmark is fine tuned towards the expected position of an eye. Since the system should discriminate the two landmarks, but also the joint landmarks against other facial features, two machine experts are employed. Independently, the landmarks are looked for in a successive way, aborting each search after a maximum number of trials. Both experts must agree on successful landmark localization, to decide to keep or to drop the prepared frame. A kept frame can then be used for fast facial feature detection (e.g. mouth, geometry determination) and face authentication purposes in a system similar to [18]. For the purpose of this paper, the facial landmark localization system is employed in order to evaluate the accuracy of the tracking system by visually inspecting prepared frames. Furthermore, the actual presence of a face is confirmed.

#### IV. MODELLING THE FACE

##### A. Retinotopic sampling grid

The initial retinotopic grid applied here consists of 68+1 points distributed onto 4 circles, as displayed in figure 4. In the figure the contours of a properly centered face are indicated, too. For reading convenience, the grid points are additionally numbered in a spiral manner as they are stored into a 1-D data structure. The grids size is empirically determined by fitting a face's inter-eye distance to the second innermost circle. For this purpose the average distance between the eyes (pupils) of the faces used for training was applied, thus aligning the grid as mentioned. Having frames of the size 256x205/192

pixels in the training set, this results in radii of 10.7 and 60 pixels for the innermost and outermost circle, respectively. This arrangement allows breaking the grid into small point sets of special interest.

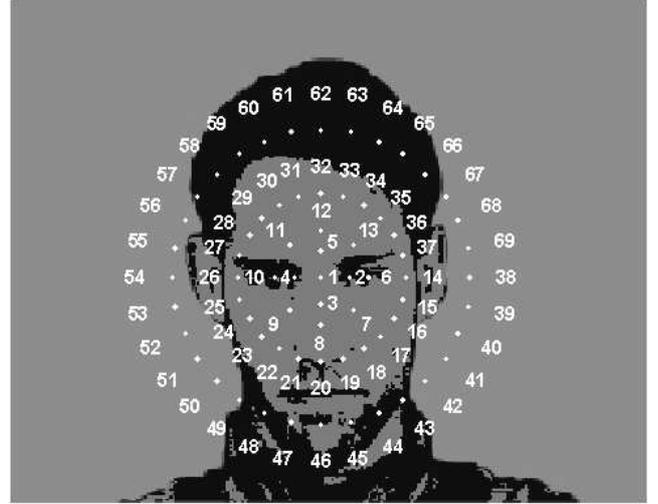


Fig. 4. The retinotopic grid with numbered points

##### B. Gabor features

At a grid point  $p$  retained for the feature extraction, Gabor feature vector  $\vec{x}$  is computed. This vector describes the neighborhood of the image  $IM$  at that point and consists of magnitude answers  $k_{\xi,\eta}$  of the Gabor filters  $f$ . A single Gabor filter, denoted as  $f_{m,n,\xi,\eta}$  is a 2-D complex valued filter corresponding to a certain frequency  $\xi$  and orientation  $\eta$  in the image neighborhood around  $p$ . The filter magnitude answer is formed by the absolute value of the scalar product of the image cut-out  $IM$  of size  $M \times N$  and the complex Gabor filter  $f$ . This is also shown in equation 1. The size  $M \times N$  of the considered neighborhood around  $p$  depends on the frequency  $\xi$ : A higher frequency implies a smaller neighborhood and vice-versa. Depending, which frequency and orientation channels are applied, the resulting feature vector  $\vec{x}$  consists of the single magnitude answers  $k_{\xi,\eta}$ . The dimension of  $x$  equals the product of the amount of applied frequency and orientation channels.

$$k_{(\xi,\eta)} = \left| \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} IM_{(m,n)} f_{(m,n,\xi,\eta)} \right| \quad (1)$$

The applied Gabor filter bank is designed in the so-called log-polar domain, a logarithmically scaled polar space where the Gabor filters are uniformly distributed 2-dimensional concentric Gaussian bells, [2]. The in this way determined Gabor filters efficiently cover a large area in the Fourier domain. The number of (radial or isotropic) frequency channels is 5 whereas 6 orientation channels are used.

##### C. Tracking models

For the tracking system, several face regions are modelled. Four models are constructed for a face coming from outside

(*outer state*) and three models for the face being centered or leaving from inside (*inner state*). As a result, state-dependent experts are employed in the operational phase. The different indices of the experts are illustrated in figure 2. Different models are used in outer and *inner state* in order to adapt best to the relative facial shape. In the following the models are referred to as *outer models* and *inner models* respectively.

*Outer models:* These models encode the necessary information to detect four facial regions (right/left hand-side, top, bottom) coming from outside of the grid. In the following they are referred to as right-outer model, bottom-outer model, etc. The left hand-side of figure 5 shows the grid points used for the outer models with an indicated face approaching from the right. Each model contains three grid points (numbers refer to the labels shown in figure 4): {14, 42, 66} for the right-outer model, {26, 50, 58} for the left-outer model, {20, 42, 50} for the top-outer model and {32, 58, 66} for the bottom-outer model. These twelve grid points constitute the point set  $\{O_i\}$ , mentioned earlier. Looking, e.g. at the indicated face on the left hand-side in figure 5, it is apparent that the points involved in the right-outer model are chosen to adapt best to the facial curvature. Likewise, e.g. the top-outer models points are bent downward. It is also worth mentioning that the outer models share four grid points. This means that in operational phase features extracted at those points can be used each twice, thus saving important processing time. The outer models encode features taken at low frequency channels. By this a larger area is covered. The final frequencies used for feature extraction are detailed further below.

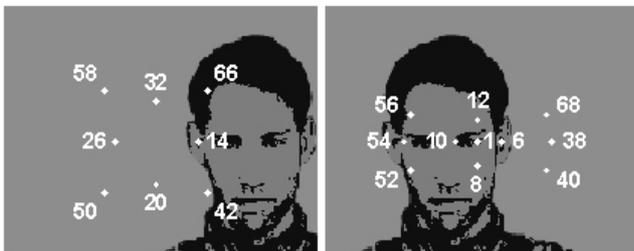


Fig. 5. Outer models (left) and inner models (right)

*Inner models:* The inner models encode the necessary information to recognize three facial regions (right/left hand-side, central) of a face centered to the grid or leaving the grid. In comparison to the outer models they are referred to as right-inner -, bottom-inner - and center model. The right hand-side in figure 5 shows the grid points used for the inner models with an indicated face leaving the center towards the left. The following grid points are used: {52, 54, 56} for the right-inner model, {38, 40, 68} for the left-inner model and {1, 6, 8, 10, 12} for the center model. These grid points together constitute the point set  $\{I_i\}$ . The indicated face on the right in figure 5 apparently suggests bending the sampling points for the right-inner and left-inner models in an opposite direction compared to the points used for corresponding outer models (note that the term right or left always denotes the facial region and not a direction). The center model uses five

points to increase the spatial precision of which, point 6 and 10 are furthermore likely to be directly at the pupils position of a centered face. The center model encodes higher frequency features, compared to all other tracking models, since it should focus on small areas of a face. The right-inner and left-inner models employ low frequency channels to cover larger facial areas.

#### D. Landmark models

The facial landmark localization system employs separate machine-experts, thus involving separate models. Figure 6 displays the used point set for the landmark models. The 15 points together constitute the point set  $\{L_i\}$ . Both of the models, named right-eye and left-eye model respectively, employ the same  $\{L_i\}$  for feature extraction. In figure 6 the right-eye model points are displaced to their position of interest. Both right-eye and left-eye model can be compared to the center model of the tracking system, but employing even more points at the inner grid circles. Point 14 and 26 are primarily used to discriminate the eye regions, thus low frequencies are applied for feature extraction there. At the remaining points, medium and high frequency channels are used to model the eye. The finally used frequencies are listed in the paragraph below.

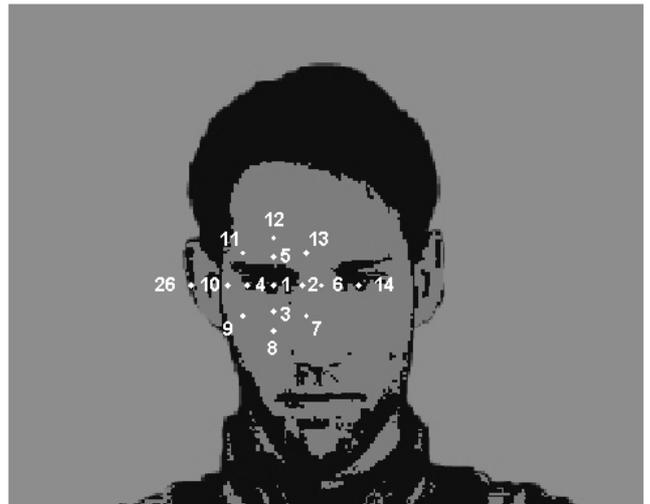


Fig. 6. Right-eye model, focusing on the corresponding eye

#### E. Training the experts

Face generalization is added to the system by training the used experts on a number of faces. For this purpose 100 frontal images from XM2VTS Database [15] are chosen at random. The images are downsized to a resolution of 256x205, which is nearest to the resolution of the 256x192 frames acquired from the camera. In addition 15 frontal camera images of 2 different persons extend the training set, which means that in total 130 frames are used for training purposes. Each expert of the tracking system and the landmark localization system is trained separately, whereas the training procedure can be described in a common way: The used grid points, employed

frequencies for the feature extraction and the facial regions to be focused on are defined by the respective models. While cycling the training frames, regions of interest are marked manually. Depending on the expert to be trained, the appropriate model parameters are applied to extract features. The features are labelled as positive class examples for the facial region to be trained. The same amount of negative examples for a facial region are extracted and labelled as negative class examples. Then non-linear SVM classifiers employing an RBF kernel are trained for each model. In order to determine the kernel parameters for optimal classification rates, cross validation is applied to each classifier. In this process also different packages of frequency channels are tried for each expert. In most cases (for all outer and inner models) the classification rate turns out to be higher when applying fewer rather than more frequency channels in the feature extraction process (compare line 5 and 6 in the table below). This is, considering the real-time aspect, a desirable outcome. For the landmark models, negative class examples for one eye include features extracted at the opposite eye. Table I lists the finally employed frequency channels (1 low .. 5 high) and achieved classification rates (in the cross validation tests) for all models:

Model	Frequency channel	Classification rate
Right-outer	1,2	0.98
Left-outer	1,2	0.99
Top-outer	1,2	0.93
Bottom-outer	1,2	0.88
Center	4,5	0.96
Center	3,4,5	0.95
Right-inner	1,2	0.89
Left-inner	1,2	0.92
Right-eye	1,2,3,4,5	1
Left-eye	1,2,3,4,5	1

TABLE I  
TABLE OF EMPLOYED FREQUENCY CHANNELS

The determination of the decision threshold for each expert was done by scanning pictures of the training set and calculating the analyzed machine expert's response at each pixel. Different thresholds were applied to the responses to study the performance of the models manually. The machine experts' threshold designs were based on these observations.

## V. FINGERPRINT ANALYSIS

In this section the basic principles of the fingerprint recognition process are described. The core point detection and the alignment of the fingerprints used for the experiments are based on [16], thus only briefly described in this section. The implemented recognition algorithm is a texture based approach, as the facial feature descriptors, which also utilize orientation based information.

### A. Extracting the core point

The core point of a fingerprint serves as reference point for the alignment described in 5.2. In order to extract this singular point a first order complex Gaussian filter is convolved with the orientation tensor of a fingerprint image. The position

of the core point is finally determined by means of the Gaussian Pyramid. The result of such a core point detection in fingerprints is shown in figure 7.



Fig. 7. Test (left) and reference fingerprint (right) with marked core point

### B. Translation and rotation of the fingerprint

The reference fingerprint is aligned with an arbitrary one by translation and rotation based on their core points. Tests in [16] have shown that the Gaussian pyramid levels 2 and 3 are most suitable for the extraction of the core rotation angle. After the coarse alignment process some fine tuning of rotation and translation is performed. In figure 8 the areas around the core point of the aligned test and reference fingerprints are displayed.

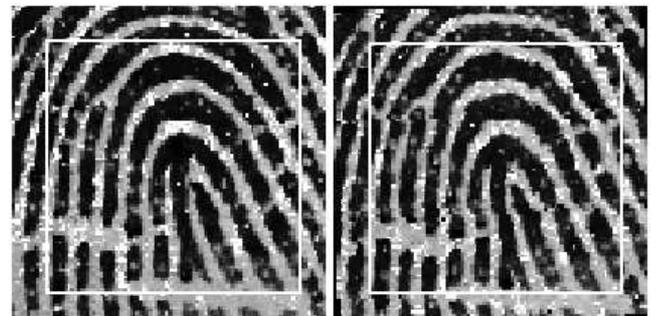


Fig. 8. Aligned test and reference fingerprints' core neighborhood

### C. Recognition algorithm

The similarity measurement is calculated by means of the second order complex moments  $I_{20}$  and  $I_{11}$ . Their detailed derivations can be found in [4], [10].

$$I_{20} = \iint_{-\infty}^{\infty} ((D_x + \imath D_y) f(x, y))^2 dx dy \quad (2)$$

$$I_{11} = \iint_{-\infty}^{\infty} |(D_x + \imath D_y) f(x, y)|^2 dx dy \quad (3)$$

The local neighborhood of a fingerprint consists of linear patterns, represented by its ridges and valleys. Only at minutiae

points, where ridges are either merged or broken, the linear symmetry is very small. The complex moments expressed in the equations 2 and 3 provide information about this lack of linear symmetry. In a perfect linear symmetry  $I_{11}$  and  $I_{20}$  are equal; however the smaller  $I_{20}$  compared to  $I_{11}$ , the lower the linear symmetry. The actual matching process between the two fingerprints is performed by means of the Schwartz inequality. In this case the similarity between the orientations of two local neighborhoods is calculated. The closer the resulting scalar of the quotient in equation 4 is to 1, the higher the similarity between the two fingerprints.

$$\frac{\left| \sum_x \sum_y LS_{\text{test}} \cdot LS_{\text{ref}} \right|}{\|LS_{\text{test}}\| \cdot \|LS_{\text{ref}}\|} \leq 1 \quad (4)$$

In the formula above  $LS$  denotes the quotient  $\frac{I_{20}}{I_{11}}$ . By dividing  $I_{20}$  through  $I_{11}$  unreliable orientations are attenuated, whereas others are preserved. As indicated by a square in figure 8 only a certain area (81x81 pixels) around the core points are regarded for the similarity measurement. Furthermore the test - and the reference fingerprint are illustrated visualizing the modified moment  $LS$ .

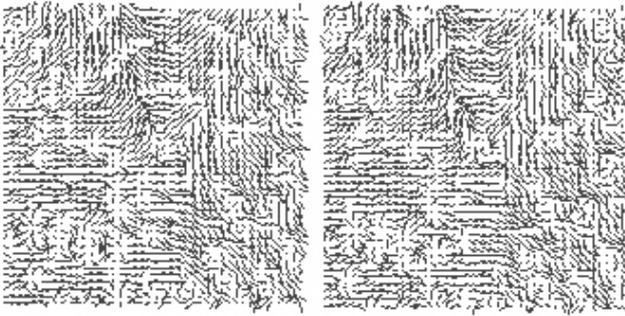


Fig. 9. Modified moment  $\frac{I_{20}}{I_{11}}$  of test and reference fingerprint

## VI. EXPERIMENTAL RESULTS

For the implementation and testing of the algorithms an IEEE1394 camera (Sony XCD-X700) and an USB fingerprint sensor (Fingerprint cards FPC1010) is used. The camera is mounted on a pan-tilt unit (Directed Perception PTU-46-17.5). The fingerprint sensor is only used for testing purposes, whereas all experimental results documented in this paper are based on tests employing the FVC2000 fingerprint database [14]. All camera images are downsized to a resolution of 256x192 pixels.

### A. Real-time face tracking

Additionally, the system performance is tested by tracking the face of 10 different test persons moving freely in front of the camera at a distance of about 1 to 2 meters. From each person 30 prepared frames are automatically acquired. Figure 10 shows one of these frames for each test person. A rectangle is drawn around the areas where the people's eyes are supposed to be in, to facilitate the visual evaluation by humans.



Fig. 10. Successfully prepared frames of 10 test persons

This assumed eye position rectangle comprises approximately 4% of the whole frame area.

The results of the visual inspection are as follows: 290 prepared frames (96,66%) out of 300 contain both eyes and another 8 (2,66%) contain at least one eye. Only in 2 cases (0,66%) the eyes are slightly outside the rectangular area. Figure 11 shows three such erroneous eye region detections.

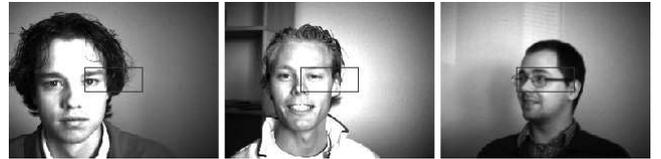


Fig. 11. Some tracking errors

The proposed system is indifferent to gender, hair and skin colour of the test person. Also wearers of glasses did not present difficulties. Additionally illumination changes have little influence on the system. Two out of the ten test persons were included in the training set of the SVM and, due to this, the corresponding prepared frames were free of errors. Nevertheless the system has some discrimination errors. In 8 cases only one eye is inside the small rectangle. The major source of this error is long curly hair, which is considered as the second eye in the worst case.

### B. Facial landmark localization

The facial landmark localization process should confirm the results of the visual inspection performed in 6.1. For this purpose all 300 prepared frames are analyzed by the facial landmark localization system. Figure 12 outlines the located eyes of four prepared frames.

The following results are obtained by this process: in 291 prepared frames out of 300 (97%) the rectangular area contains

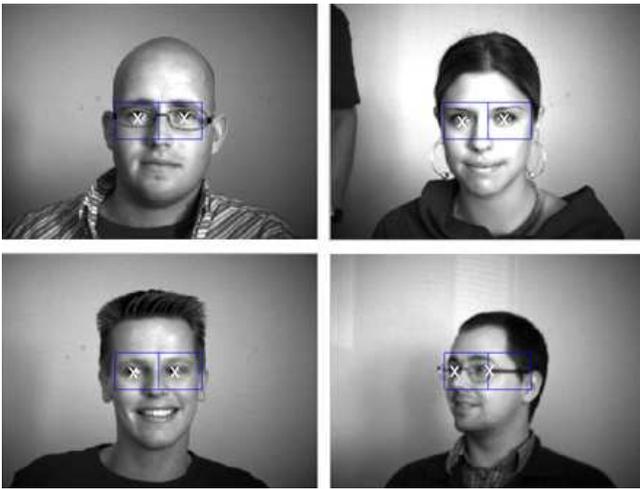


Fig. 12. Facial landmark detection in 4 prepared frames

both eyes and in another 7 (2,33%) at least one eye is inside. The the landmark localization system found two eyes in a prepared frame in which the human inspection determined only one. This frame is displayed at the bottom right in figure 12. Due to the fact that this frame is not very suitable for further processing steps, e.g., face authentication, the frame can be regarded as a discrimination error of the landmark localization system. Nevertheless this is the only case in this test that showed disagreement with the manual inspection.

### C. Fingerprint recognition

The tests of the fingerprint recognition algorithm are carried out on the subset DB2 of the FVC2000 fingerprint database, which contains 8 fingerprints of 100 different persons. Among these, 50 persons having a core point are chosen for the actual recognition test. In order to calculate the FRR one fingerprint of each test candidate serves as the reference and the remaining seven are used as genuine fingerprints. In the FAR calculation one fingerprint is regarded as client and all other fingerprints of the remaining persons (492) served as imposters. The same procedure is repeated for all 50 persons. The resulting FRR and the FAR curves are shown in figure 13. As indicated an EER of 4% is achieved.

## VII. CONCLUSION

The main novelties of the system proposed in this paper are the usage of sampling points and frequency channels employed in the face-tracking part, which enable its real-time performance. Low error rates are achieved in both the face tracking and the facial landmark localization systems. They are robust to typical sources of errors like glasses or illumination changes, considering the rather small training set. The facial landmark localization system turns out to be an attractive alternative to the human inspection. The slight problems dealing with long curly hair could be counteracted by increasing the decision threshold of the expert in charge of the central facial region, risking at the same time weaker tracking performance. One way to solve this problem is to

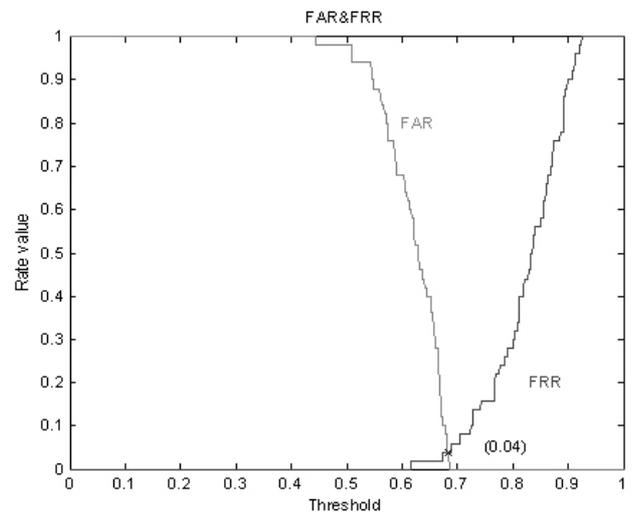


Fig. 13. Curve displaying FRR, FAR and EER of the fingerprint recognition system

bootstrap, e.g. to supplement the training with people who are not satisfyingly tracked by the system. This has not been done currently due to lack of data.

Also the test results of the fingerprint recognition seem to be at a satisfying level, regarding the difficulty level of the used fingerprint database.

## ACKNOWLEDGEMENT

We would like to thank Lorenz Gruebler and Situ Xin, both Halmstad University students, for their help in implementing the platform used in this study.

## REFERENCES

- [1] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multi modal person authentication systems by bayesian statistics. In J. Bigun, G. Chollet, and G. Borgefors, editors, *Audio and Video based Person Authentication - AVBPA97*, pages 291–300. Springer, 1997.
- [2] J. Bigun. Speed, frequency, and orientation tuned 3-d gabor filter banks and their design. In *Proc. International Conference on Pattern Recognition, ICPR, Jerusalem*, pages C–184–187. IEEE Computer Society, 1994.
- [3] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Multimodal biometric authentication using quality signals in mobile communications. In *Proc. of 12'th Int. conf. on image analysis and processing, Mantova, Italy*, pages 2–11. IEEE Computer Society Press, Piscataway, NJ, september 17-19 2003.
- [4] J. Bigun and G. H. Granlund. Optimal orientation detection of linear symmetry. In *First International Conference on Computer Vision, ICCV, June 8–11, London*, pages 433–438. IEEE Computer Society Press, Washington, DC., 1987.
- [5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.
- [6] B. Duc, S. Fischer, and J. Bigun. Face authentication with Gabor information on deformable graphs. *IEEE Trans. on Image Processing*, 8(4):504–516, 1999.
- [7] Ian R Fasel, M. S Bartlett, and J. R. Movellan. A comparison of gabor methods for automatic detection of facial landmarks. In *International conference on Automatic Face and Gesture Recognition*, pages 242–248, May 2002.
- [8] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J. physiology (London)*, 148, 1959.
- [9] A. Jain, L. Hong, and Y. Kulkarni. A multimodal biometric system using fingerprint, face and speech. In *Audio and Video based Person Authentication - AVBPA99*, pages 182–187, 1999.

- [10] M. Kass and A. Witkin. Analyzing oriented patterns. *Computer Vision, Graphics, and Image Processing*, 37:362–385, 1987.
- [11] C. Kotropoulos and I. Pitas. A rule based face detection in frontal views. In *ICASP*, volume IV, pages 2537–2540, 1997.
- [12] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Hertz, and W. Konen. Distortion invariant object recognition in the dynamic link architectures. *IEEE Trans. on Computers*, 42(3):300–311, March 1993.
- [13] D. Maio and D. Maltoni. Real-time face location on gray-scale static images. *Pattern Recognition*, 33(9):1525–1539, 2000.
- [14] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A. K. Jain. Fvc 2000: Fingerprint verification competition. *IEEE-PAMI*, 24(3):402–412, february 2002.
- [15] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In *Audio and Video based Person Authentication - AVBPA99*, pages 72–77. University of Maryland, 1999.
- [16] K. Nilsson and J. Bigun. Localization of corresponding points in fingerprints by complex filtering. *Pattern Recognition Letters*, 24:2135–2144, 2003.
- [17] G. A. Orban. *Neuronal operations in the visual cortex. studies of brain functions*. Springer, 1984.
- [18] F. Smeraldi and J. Bigun. Retinal vision applied to facial features detection and face authentication. *Pattern Recognition Letters*, 23:463–475, 2002.
- [19] F. Smeraldi, O. Carmona, and J. Bigun. Real-time head tracking by saccadic exploration and gabor decomposition. In A. T. Almeida and H. Araujo, editors, *Proc. the 5th International Workshop on Advanced Motion Control*, volume IEEE Cat. Num. 98TH8354, pages 684–687. IEEE Service Center, 445 Hoes Lane, P.O. Box 1331 Piscataway, NJ 08855-1331, USA, 1998.
- [20] B. Takacs and H. Wechsler. Face localization using a dynamic model of retinal feature extraction. In M. Bichsel, editor, *Proc. the International workshop on automatic face and gesture recognition, Zurich*, pages 243–247. Multimedia Laboratory; Univ. of Zurich; Winterthurers. 190 CH-8057 Zurich, June 26-28 1995.
- [21] M. Tistarelli and G. Sandini. On the advantages of log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Transactions on Pattern Analysis and Machine Recognition*, 15(4):401–410, 1993.
- [22] A. L. Yarbus. *Eye movements*. Plenum, New York, 1967.