Facial identification problem: A tracking based approach

Marco Anisetti, Valerio Bellandi, Ernesto Damiani
Department of Information Technology
University of Milan
via Bramante, 65 - 26013, Crema (CR), Italy
Telephone: +390373898047, Fax: 0373899010
Email: {anisetti,bellandi,damiani}@dti.unimi.it

Fabrizio Beverina
AST Group, ST Microelectronics
via Olivetti, 5 - 20041,
Agrate Brianza (MI), Italy
Telephone: +390396037234
Email: fabrizio.beverina@st.com

Abstract

This paper presents a method for face identification using a query by example approach. Our technique is suitable for use within Ambient Security Environments and is robust across variations in pose, expression and illuminations conditions. To account for these variations, we use a face template matching algorithm based on a 3D head model created from a single frontal face image. Thanks to our tracking-based approach our algorithm is able to extract simultaneously all parameters related to the face expression and to the 3D posture. With these estimates, we are able to reconstruct a frontal, neutral and normalized image on which dissimilarity analysis for identification and anomalies detection is performed. Our tracking process combined with dissimilarity analysis was tested on Kanade-Cohn database [13] for expression independent identification and several other experimental databases for robustness.

1. Introduction

Face recognition has received significant attention as one of the most important applications in ambient security and environmental control, especially in the past few years. Although other very reliable methods of biometric personal identification exist, such as fingerprint analysis and retinal or iris scans, these methods rely on full, active cooperation of the participants. A personal identification system based on analysis of frontal or profile images of the face can be effective without the participant's cooperation or even knowledge. Also, other biometric methods may require special purpose hardware, while cameras are nowadays available in bundle with many electronic devices (cellular phones etc.). For this reason, unlike other biometric systems, facial recognition can be used for general surveillance, usually in combination with public video cameras. A well-known experiment was carried out in the US at 2001 and 2002 Super Bowls, where pictures were taken of every attendee as they entered the stadium and compared against a database of known offenders. In England, where public surveillance cameras are widespread, the town of Newham has also experimented with the technology. [10][12][17]. Although current face recognition systems have reached

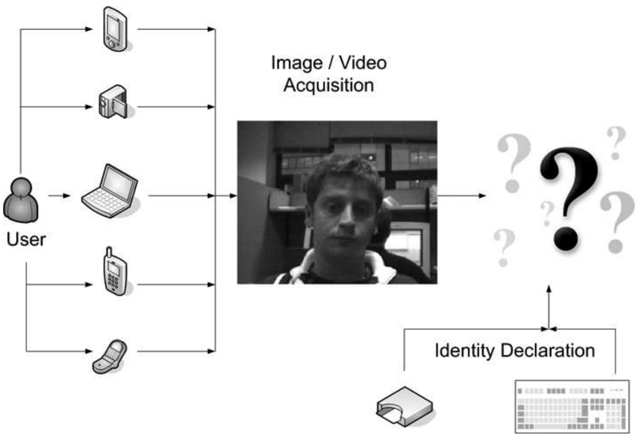


Figure 1. Identity authentication problem.

a certain level of maturity (for a complete overview, see [20]), their success has been thwarted by the constraints imposed by many real applications. For example, recognition of face images acquired in an outdoor environment with changes in illumination and/or pose remains a largely unsolved problem, as well as face recognition through expression changes. In other words, current systems are still far away from the capability of the human perception system. The human brain is highly adapted for recognizing faces, especially when compensating for changes in lighting and angle is required. Not surprisingly, early approaches to face-recognition have shown high rates of both "false positives" (wrongly matching innocent people with photos in the database) and ”false negatives” (not catching people even when their photo is in the database). Unfortunately, unlike other biometrics, faces do not stay the same over time, and early systems were easily tripped up by changes in facial hair, body weight or by simple disguises. Recently, variations in pose and illumination, seen by many as the main challenges for face recognition, have been widely investigated The goal of these approaches is separating the two main characteristics of faces: shape and color. Indeed posture (estimated with a 2D or 3D face model) and shape must be kept separated for obtaining a precise shape deformation parameter; the same can be said of illumination and texture. Morph-based approaches like [4] are in principle an effective way of enforcing such separation; however, they rely on initial information (like illumination condition and posture estimates) that in some real applications are not so easy to get. Also, morphing requires a large training set of 3D faces for producing a 3D Morphable Model. Our work is aimed at defining a *lightweight* technique, suitable to certify the identity of a subject that accepts to be identified for accessing certain services (Fig.(1)). Our identification service is based on a video stream acquired from a camera (e.g. a cellular camera, a web-cam or a set of surveillance cameras). Illumination conditions, posture and expression of the subject may be very different and *a priori* unknown. Our approach is based on face tracking as described by a 3D expression model, without requiring any frequency image intensity or neural network analysis. Our 3D expression model also tackles the problem of face expression changes, that, at the best of our knowledge, has never been taken into account in previous identification work. Expression-deformable tracking promises to be an important first step toward developing a face expression certified protocol (e.g., one capable of providing evidence that someone is about to fall asleep), dramatically enlarging the scope of the entire identification process. Our basic idea is using a template tracking process. Thanks to expression and posture parameters estimates we are able to reconstruct normalized frontal views of faces and then use some easy dissimilarity analysis to provide an acceptable degree of identification quality. Our intent is to demonstrate that the tracking approach can give good results in identification-related applications, even in the case of non-cooperative users. In the following sections we first describe all components of tracking process. We start from a short introduction to the template tracking problem; than we illustrate the 3D expression model and template construction, the tracking process for extraction of posture and our expression and dissimilarity analysis for validating the tracking results. Concluding, we present our identification architecture based on tracking and some experimental results on Cohn-Kanade database.

2. Face tracker

The extrapolation of the head posture from a monocular video sequence is the first step for the analysis and the recognition of face expression. Available methods for inferring head posture can be classified as 2D (image based) methods and 3D (model based) ones. Among the latter, two subcategories exist: the approaches relying on a geometrical shape (plane or cylinder) and the ones that use a 3D head wire-mesh model [2] . Xiao et al. [19] use a cylinder to estimate the pose beside other interesting method use 2D approach with an Active Appearance Model (AAM) [18] to map the head model to the face region. Cylinder based method can only handle limited head motion, because after a certain level of rotation the distortion due to the difference between the head and the cylinder becomes not negligible. 3D head models can substantially improve tracking only if the model can also morph according to some expression parameters. Many recent works chose this solution [15, 11, 7]. Tao et al. [11] use an explanation-based facial motion tracking algorithm based on a Piecewise Bezier Volume Deformation model (PBVD); in this way they can (in two steps) track both posture and head deformation. Matthews and Dornika [15, 7] use another method that controls position changes of every pixel of the image (not only the vertices) according to the expression parameters. Our approach is different, inasmuch instead of carrying out the usual sequence of operation for every pixel (namely, finding out which triangle intersects the pixel and do an affine warp for that triangle), we do all in one operation. This simplification is done thanks to the Candide-3 [1] definition of the transformation between every triangle. Also, our method does not need a training set [15, 7] to learn the parameters that control the shape of the face. Similarly to the Cootes approach [6], we use an optimization algorithm that matches shape and texture simultaneously. Optimization is carried out within the warping algorithm, considering simultaneously the face movement parameters and the roto-translation ones.

2.1 A parameterized 3D face model and template creation

Candide-3 is a parameterized face model. It consists in a 3D wire-frame of 113 vertices and 184 triangles. It has some advantages that make it preferable to other models used by face animators:

1. It is triangle based. This feature makes it very suitable for the affine transform, that, by definition, maps triangles into triangles;

2. The face shape and expression can be described by a

simple linear formula:

$$\mathbf{g} = \overline{\mathbf{g}} + \mathbf{S}\sigma + \mathbf{A}\alpha \quad (1)$$

where the resulting vector is $\mathbf{g}$, $\overline{\mathbf{g}}$ are the standard shape model vertices coordinates, $\mathbf{S}$ and $\mathbf{A}$ are the Shape and Animation Units. The Shape Units are controlled by the σ parameter and are used to determine the head specific shape. The animation units are controlled by the α parameter.

Using the shape parameters σ, we compute a 3D template of shape and texture on a single frame representing a frontal and neutral view of the subject. This template is the best fitting 3-D model on the frontal face view. On the frontal frame, 27 points are selected manually (they could be selected by any selection process) corresponding to some relevant features (eyebrows, eye, nose, mouth). Then, with a constrained optimization algorithm, we compute the model's shape parameters in order to minimize the error between the 24 points $\mathbf{p}$ on the model and the ones chosen manually. We define the error as follows:

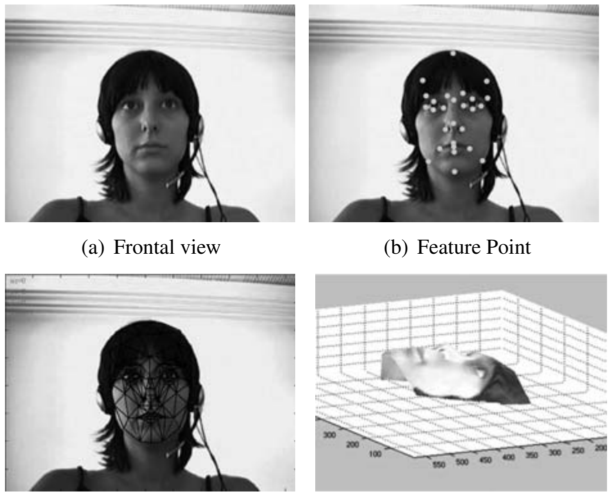
$$(\sigma, \mathbf{t}, \mathbf{R}) = \| \mathbf{w}(\mathbf{g}(\sigma, \mathbf{t}, \mathbf{R}) - \mathbf{p}) \|_2 \quad (2)$$

We weight the error on the model's points according to the intrinsic insecurity on the selection; usually the selection of the eyebrows' vertices is more difficult because of the hair, and the same applies to the center of the nose that cannot be determined precisely due to the light reflecting on it. Regarding the algorithm to find the parameters, we chose the Recursive Quadratic Programming (RQP) [16], an iterative method for constrained optimization, known to be very efficient and accurate. This method showed a great robustness to the noisy coordinates of the picked points, managing every time to reach a 3D template good enough for the tracking. This 3D template coupled with texture was stored attached to the identity of the subject. The system uses a stored 3D template for tracking process (Fig.(2) shows an example of the creation of an individual template). During the tracking process we deduce parameter α, while shape parameters σ remains constant because it represents an intrinsic property of the subject. In order to do this, we must find a set of bases that can express lineary the changes of the expression shape of our 3D template:

$$\mathbf{T} = \mathbf{T_0} + \sum_{i=1}^{n} \alpha_i \mathbf{B_i} \quad (3)$$

We know that the final position of the vertices of every triangle is:

$$\mathbf{V_f} = \mathbf{V_i} + \sum_{j=1}^{N} \alpha_j \mathbf{A}_j \quad (4)$$



(a) Frontal view (b) Feature Point

(c) Face mask (d) Template 3D

Figure 2. Example of 3D template extraction process.

$\mathbf{V_f}$ is the matrix containing the coordinate of the vertices in the final position, where α_j are the parameters that take in account every single movements of the N possible, $\mathbf{A}_j$ is the matrix that contains the component for the j-expression. A is a sparse matrix where nonzero values on a column j are the ones related to the vertices interested by that j-movement. Considering 3 vertices of the k_{th} triangle, we can find the transformation that brings a point belonging to the first triangle in a point on the second in this way:

$$\begin{bmatrix} V_{fk1} \\ V_{fk1} \\ V_{fk1} \end{bmatrix} = \begin{bmatrix} V_{ik1} \\ V_{ik1} \\ V_{ik1} \end{bmatrix} \mathbf{M_k} \quad (5)$$

The transformation for each triangle can be written as follows:

$$\begin{aligned} \mathbf{M_k} &= \begin{bmatrix} V_{ik1} \\ V_{ik1} \\ V_{ik1} \end{bmatrix}^{-1} \begin{bmatrix} V_{fk1} \\ V_{fk1} \\ V_{fk1} \end{bmatrix} \\ &= \begin{bmatrix} V_{ik1} \\ V_{ik1} \\ V_{ik1} \end{bmatrix}^{-1} \left(\begin{bmatrix} V_{ik1} \\ V_{ik1} \\ V_{ik1} \end{bmatrix} + \sum_{j=1}^{N} \alpha_{jk} \mathbf{A}_{jk} \right) \\ &= \mathbf{I} + \begin{bmatrix} V_{ik1} \\ V_{ik1} \\ V_{ik1} \end{bmatrix}^{-1} \left(\sum_{j=1}^{N} \alpha_j \mathbf{A}_{jk} \right) \\ &= \mathbf{I} + \sum_{j=1}^{N} \alpha_j \tilde{\mathbf{B}}_{jk} \end{aligned} \quad (6)$$

Where $\mathbf{A}_{jk}$ are the matrix of the displacement related to the j-expression for the k^{th} triple of points, and $\tilde{\mathbf{B}}_{jk}$ is the transformation matrix for the j-expression and k triple. In this way the i_{th} point of the template, according to the triangle it belongs to, can be derived from the template T_0 in this way:

$$T_{ik} = T_{0k}\mathbf{M}_k \tag{7}$$

$$= T_{0ik} + \sum_{j=1}^{N} \alpha_j T_{0ik}\tilde{\mathbf{B}}_{jk} \tag{8}$$

$$= T_{0ik} + \sum_{j=1}^{N} \alpha_j \mathbf{B}_{jk} \tag{9}$$

In this way we obtain the set of bases for a formula (3). In fact points $\mathbf{B}_{jk}$ depend only on the definition of the Action Units [8] and from the initial template. Now we can write any expression deformation as a weighed sum of a set of fixed bases. This characteristic of the model will be exploited in the tracking phase, making the algorithm able to estimate the set of α directly.

2.2 3D Motion and Expression Recovery

Our goal is to obtain posture estimation parameters and Action Units deformation parameters in one minimization process between two frames using the morphing basis described in previous session. We rely on the idea that face 2D template $T_i(x)$, (extracted by projecting 3D Template T on image plane) will appear in next frame $I(x)$ albeit warped by $W(x;p)$, where $p = (p_1, \ldots, p_n, \alpha_1, \ldots, \alpha_m)$ is vector of parameters for 3D face model with m Candide-3 animation units parameters and x are pixel coordinate from image plain. Thanks to this assumption, we can obtain the movement and expression parameter p by minimizing function (10). If $T_t(x)$ is the template at time t with correct pose and expression p and $I(x)$ is the frame at time $t+1$, assuming that the illumination condition does not change much, the next correct pose and expression p at time $t+1$ can be obtained by minimizing of sum of the square errors between $T(x)$ and $I(W(x;p))$:

$$\left(\sum_x [I(W(x;p)) - T(x)]^2\right) \tag{10}$$

For this minimization we use the Lucas-Kanade approach [14] with forward additive implementation that assumes that current estimate of p is known and iteratively solves for increments to the parameters Δp. Equation (10) after some straightforward passages, becomes:

$$\Delta p = H^{-1} \sum_x \left[\nabla I \frac{\partial W}{\partial p}\right]^T [T(x) - W(x;p)]$$

$$H = \sum_x \left[\nabla I \frac{\partial W}{\partial p}\right]^T \left[\nabla I \frac{\partial W}{\partial p}\right] \tag{11}$$

with $\nabla I$is the gradient of the image I evaluated at $W(x;p)$, $\frac{\partial W}{\partial p}$ is Jacobian of warp evaluated in p and Δp is the update of the warp parameters.

To recover the 3D posture and expression morphing parameter, we consider that the motion of head point $X = [x, y, z, 1]^T$ between time t and $t+1$ is: $X(t+1) = M \cdot X(t)$ and expression morphing of the same point is: $X(t+1) = (X(t) + \sum_{i=1}^{m}(\alpha_i \cdot B_i))$. Here, α_i and B_i follows expression based representation described in the previous section and the matrix M follows Bregler's definition and the twist representation by Murrey [5]. Using these matrixes, the motion parameters p become $(\omega_x, \omega_y, \omega_z, t_x, t_y, t_z, \alpha_1, \ldots, \alpha_m)$ while

the warping $W(x;p)$ in (10) becomes:

$$W(x;p) = M(X + \sum_{i=1}^{m}(\alpha_i \cdot B_i)) \tag{12}$$

In a situation of perspective projection, assuming the camera projection matrix depends only on the focal length f_L, the image plane coordinate vector x can be obtained as follows:

$$\mathbf{x(t+1)} = \begin{bmatrix} x - y\omega_z + z\omega_y + t_x + B_x \\ x\omega_z + y - z\omega_x + t_y + B_y \end{bmatrix} \cdot \frac{f_L}{-x\omega_y + y\omega_x + z + t_z + B_z}(t) \tag{13}$$

where:

$$\begin{array}{l} B_x = \sum_{i=1}^{m}(\alpha_i(a_i - b_i\omega_z + c_i\omega_y)) \\ B_y = \sum_{i=1}^{m}(\alpha_i(a_i\omega_z + b_i - c_i\omega_x)) \\ B_z = \sum_{i=1}^{m}(\alpha_i(-a_i\omega_y + b_i\omega_x + c_i)) \end{array} \tag{14}$$

where the parameters a_i, b_i, c_i are translation vectors due to the morphing basis.

This function maps the 3D motion and morphing in the image coordinates.

Using a forward additive parameter estimation algorithm we are now able to obtain the correct posture 3D motion and morphing parameter of template between two frames in a single minimization phase (some example of expression tracking in Fig.(10)).

In order to increase the robustness towards global illumination changes we also introduce in our estimation parameter other two additional parameters using a linear appearance variations technique. We consider the image template $T(x)$ as:

$$T(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \tag{15}$$

Figure 3. Example of posture and morphing estimation on Cohn-Kanade DataBase.

where A_i with $i = 1, \ldots m$ is a set of known appearance variation images and λ_i with $i = 1, \ldots m$ are the appearance parameter. Global illumination change can be modeled as an arbitrary change in gain and bias between template and the input image by setting A_1 to be the T template and A_2 to be the unitary image that takes into account the template completely illuminated. Substituting inside the equation 10 the template calculated in the equation (15) we obtain the following equation to be minimized:

$$\sum_x [I(W(x;p)) - T(x) - \sum_{i=1}^{2} \lambda_i A_i(x)]^2 \qquad (16)$$

Thanks to linear appearance variations techniques, this function can be minimized using a Lucas-Kanade like approach. Figure (4) shows an example of illumination change. Lateral illumination change can be managed easily using linear appearance variation techniques with some specific illumination base. All tracking process include occlusion management. We consider two types of occlusion, self-occlusion (posture occlusion), and occlusion by an out-face object. Because of this ”external factor” some pixels in the face template should contribute less (or nothing) to the motion estimation than the others. To perform this we apply a well known, and commonly used technique called iteratively re-weighted last squares (IRLS) [3] already used by Xiao [19].

2.3 Dissimilarity analysis

When that tracking process converges to some posture and parameter estimate for morphing, it is possible to create a normalized neutral frontal view of the subject's face.

Figure 4. Tracking 3D with global illumination robustness technique.

In fact the intent of face tracking is to estimate posture and expression parameter of the subject for the reconstruction of the normalized neutral frontal face, where localization of interesting facial features are easier to do. Using this idea, we use the reconstructed frontal view to evaluate the quality of identification. This quality evaluation is our ”dissimilarity analysis”. To build the normalized face we back-project each frame on the 3D model using the estimated posture and morphing parameter. There, we warp the model on the frontal view. If the estimated posture and morphing parameter is correct, the frontal image will be consistent with the face (Fig.(6)).

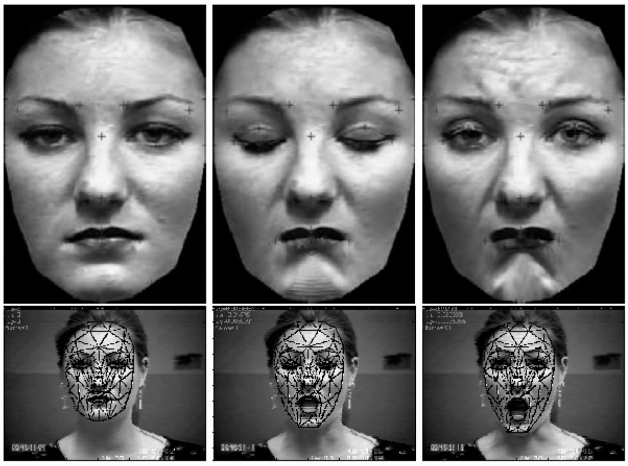


Figure 5. Example of posture and morphing estimation and frontal view normalization on Cohn-Kanade DataBase.

With this approach, good recognition occurs if normalization reconstruction was flawless. If any approximation error happens in the posture and expression evaluation (means that the template of the subject do not match), the normalized face suffers from a distortion effect, depending on the amount of the error. For this reason, we can estimate the quality of identification by quantifying this distortion on retrieved images. We can consider the probability P_s of correct identification based on dissimilarity value as follow:

$$P_s = \exp\left(-\left(\sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}}\right)\right) \qquad (17)$$

This analysis is performed by tracking n trusted points on the normalized face (an extended set of trusted points is represented as "+" in Fig(6)) obtain the d_i value of equation (17).

3 Cooperative Identity certification

Identity certification is based on idea that the subject wants to be identified, and that the identity of the subject is known to the identification system. The aim of tracking algorithms is to certify identity independently of illumination change, expression and posture. The identification process can be summarized as follows:

1. The user, using any devices with camera, declares identity (using a microcamera on top of the badge scanner).
2. This system retreves the corresponding 3D template (3D model + normalized facial texture) from database.
3. Face detection if performed on video streaming with initial posture estimation.
4. Tracking is started with posture estimation from the previous step.
5. If tracking dose not converge, identity matching is rejected.
6. If tracking occurs in convergence, the system produce a normalized face representation using estimated posture and expression parameter.
7. A dissimilarity measure on normalization face is computed to assess the quality of identification.

Note that, when identifications is successful, the subject template texture is gradually updated to prevent aging. From our experience of many experiments on face video sequences, obtained from different sources, tracking has a optimal results for subjects with very different 3D face shapes and very good robustness on expression changing and on illumination condition variations. When two subject have a similar face 3D shape, tracking alone can not provide a good identification response. For this reasons we improved the identification method with dissimilarity analysis. Another interesting use of the tracking capability of our algorithm is a certification protocol based on a fixed sequence of facial movements (Fig.(6)). We present some results of tracking algorithms and identity certification methodologies in the following section.

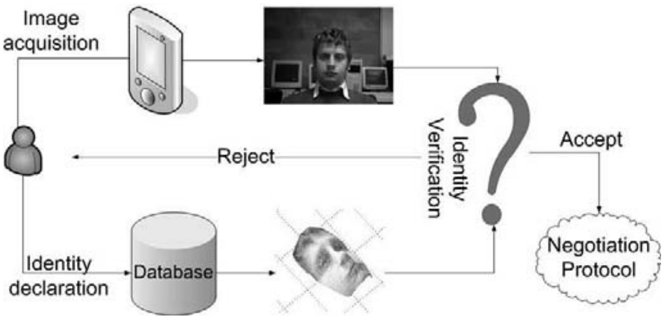


Figure 6. Identification process overview.

4 Experimental results

We carried out two experiments to evaluate the performance of our system. The first compared tracking quality in cases of morphing and non-morphing 3D tracking. The second dealt with the face identification process: for expression changes we used the Cohn-Kanade database, while for robustness testing we relied on our own laboratory's database, including 4 different subjects that move, change expression and talk, registered during ELITE2002 [9] tracking experimental session. This video database was recorded in Politecnico of Milan laboratory with a commercial webcam with a resolution of 640x480 pixel synchronized with ELITE2002 for real movement evaluation. The ELITE2002 system is an optoelectronic device able to track the three-dimensional coordinates of a number of reflecting markers placed non invasively on a helmet on the subject head and on the web-cam. This system, thanks to a set of 6 cameras, can perform the tracking of a point with the precision of the range of the 0.3 mm. Thanks to this high confidence in posture estimation we were able to compare our 3D tracking with real subject movement.

In Fig. (8) we present an example of estimated tracking value for jaw rotation (the major rotation in the presented sequence) with and without morphing. It is clear that if the model can estimate morphing parameter, posture evaluating becomes more precise. In our experiments, the error of tracking with expression morphing is at maximum 2-3% compared with the ones without morphing tracking. This is an impressive results especially considering that even no-morphing tracking gives good results, with a maximum error of about 3 grades. Our tracking algorithms produce a estimated posture and expression parameters very close to the real ones (Fig. (7)).

These results also suggest that, tracking algorithms can indeed be used for identification because they do not introduce morphing variations that have a negative impact on identification and on any normalized frontal face reconstruction analysis. Some interesting experiments were fo-

Figure 7. Example of face tracking during ELITE experimental session.

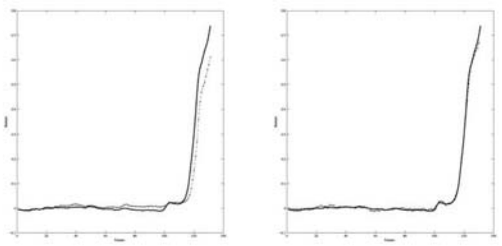
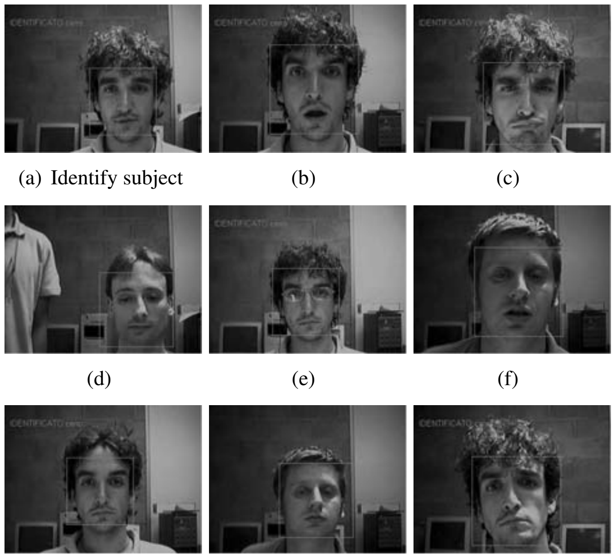


Figure 8. Comparison between Elite2002 (solid line) pose estimate and no morphing(left) and morphing (right) tracking posture estimate.

cused on changes to subjects' appearance (wearing glasses, changing hairstyle etc.). Thanks to linear appearance variation base A_1 in formula (15) we were able to deal with any moderate face appearance modification like, for instance, beard grows, without introducing mistakes. Figure (9) shows some examples of identification robustness w.r.t. appearance changing. Some frames captured by a videocamera are shown to underline the robustness of our identification process. Another experiment was focused on identify expression invariants. We used the Cohn-Kanade data base that contains many expression changes and many different subjects including some neutral frames from which we could construct our 3D template database. In the following tables (Tab.(1), Tab.(2)) we summarize our results. Based on intensity of expression we classify expressions from neutral to average, up to maximum intensity.



(a) Identify subject (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 9. Example of identification tests across expression and illuminations changing. Subfigure (a) represent the subject that wants to be identified. In Subfigure (d)(f)(h) false identity declarations without system accept response.

number of experiments	neutral	low	average	high
76	97%	95%	87%	78%

Table 1. Percentage of identifications in cases of expression variation.

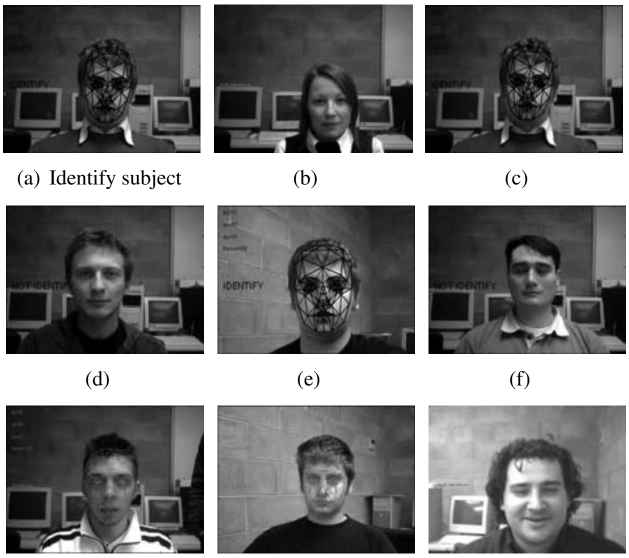
number of experiments	0-5	5-15	15-35	>35
46	98%	96%	88%	69%

Table 2. Percentage of identification in cases of pose variations.

Figure(10) shows other identification tests with dissimilarity analysis evaluation.

5 Conclusion

Concluding we develop an promising face identification technique using an innovative one step tracking algorithm for facial posture and expression estimation. Our goal is to produce an identity certification based on cooperation of the user that declare her identity. This identity certification process, in case of correct identification, provide quality of identification for each frame of the captured video.



(a) Identify subject (b) (c)

(d) (e) (f)

(g) Wrong identification (h) (i)

Figure 10. Example of identification tests. The black tracking mask plot means tracking success, therefore high identify probability, while a gray one indicates lower probability. In Subfigure (g) we present a wrong tracking identification that will be afterwards discarded thanks to dissimilarity analysis.

The base idea of our approach is that if our precise template tracking algorithm do not converge than the identification process fails. With this technique we are able to prevent at last 21% of FAR (False acceptance rate) with maximum 2% of FRR (False Rejection Rate). To improve the FAR of our identification technique we use the dissimilarity analysis on normalized frontal face that quantify the imprecision of the tracking process as the probability of correct identification. This work was aimed at proving sensibility of this approach. Further investigation needs to be done on normalized faces to eliminate these ambiguous cases and improve FAR.

Acknowledgment

The authors would like to thank to the TBM Lab of Politecnico di Milano for the help during experiment sessions with Elite2002.

References

[1] J. Ahlberg. An active model for facial feature tracking. *EURASIP Journal on Applied Signal Processing*, 2002 (6)(6):566–571, 2002.

[2] M. Anisetti, V. Bellandi, and F. Beverina. Accurate 3d model based face tracking for facial expression recognition. In *Proceedings of the Fifth IASTED International Conference on Visualization, Imaging, and Image Processing*, Sept. 2005.

[3] M. Black. *Robust incrementel optical flow*. PhD thesis, Yale university, 1992.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1 – 12, Sep. 2003.

[5] C. Bregler and J. Malik. Tracking people with twists and exponential maps. *In CVPR98*, pages 8–15, 1998.

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance mode. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681 – 685, Jun. 200.

[7] F. Dornaika and J. Ahlberg. Fast and reliable active appearance model search for 3-d face tracking. *IEEE Transactions on Systems, Man and Cybernetics*, 34(4):1838 – 1853, Aug. 2004.

[8] P. Ekman and W. Friesen. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press*, 1978.

[9] G. Ferrigno and A. Pedotti. Elite: a digital dedicated hardware system for movement analysis via real-time tv signal processing. *IEEE Trans Biomed Eng.*, pages 943–50, 1985.

[10] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[11] T. H. Hai Tao. Explanation-based facial motion tracking using a piecewise bzier volume deformation model. *In CVPR99*, 1999.

[12] P. Hallinan. *A Deformable Model for the Recognition of Human Faces under Arbitrary Illumination*. PhD thesis, Harvard Univ. Cambridge, Mass., 1995.

[13] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. *Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pages 46–53, Grenoble, France 2000.

[14] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. Int. Joint Conf. Artificial Intelligence*, pages 674–679, 1981.

[15] I. Matthews and S. Baker. Active appearance model revisited. Technical Report CMU RI TR 03-02, The Robotic Institute, Carnegie Mellon University, 2002.

[16] M. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In G.A.Watson, editor, *Numerical Analysis*. Springer Verlag, 1978.

[17] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. 2001.

[18] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2004.

[19] J. Xiao, T. Kanade, and J. Cohn. Robust full-motion recovery of head by dynamic templates and re-registration techniques. *Proceeding of the fifth IEEE international Conference on automatic face and gesture recognition*, May,2002.

[20] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys,*, 35(4):399458, December 2003.