FACE TRACKING ALGORITHM ROBUST TO POSE, ILLUMINATION AND FACE EXPRESSION CHANGES: A 3D PARAMETRIC MODEL APPROACH

Marco Anisetti, Valerio Bellandi
University of Milan - Department of Information Technology
via Bramante 65 - 26013, Crema (CR), Italy

Luigi Arnone, Fabrizio Beverina
STMicroelectronics - Advanced System Technology Group
via Olivetti 5 - 20041, Agrate Brianza, Italy

Keywords: Face tracking, expression changes, FACS, illumination changes.

Abstract: Considering the face as an object that moves through a scene, the posture related to the camera's point of view and the texture both may change the aspect of the object considerably. These changes are tightly coupled with the alterations in illumination conditions when the subject moves or even when some modifications happen in illumination conditions (light switched on or off etc.). This paper presents a method for tracking a face on a video sequence by recovering the full-motion and the expression deformations of the head using 3D expressive head model. Taking advantage from a 3D triangle based face model, we are able to deal with any kind of illumination changes and face expression movements. In this parametric model, any changes can be defined as a linear combination of a set of weighted basis that could easily be included in a minimization algorithm using a classical Newton optimization approach. The 3D model of the face is created using some characteristical face points given on the first frame. Using a gradient descent approach, the algorithm is able to extract simultaneously the parameters related to the face expression, the 3D posture and the virtual illumination conditions. The algorithm has been tested on Kanade-Cohn database (Kanade et al., 2000) for expression estimation and its precision has been compared with a standard multi-camera system for the 3D tracking (Elite2002 System) (Ferrigno and Pedotti, 1985). Regarding illumination tests, we use synthetic movie created using standard 3D-mesh animation tools and real experimental videos created in very extreme illumination condition. The results in all the cases are promising even with great head movements and changes in the expression and the illumination conditions. The proposed approach has a twofold application as a part of a facial expression analysis system and preprocessing for identification systems (expression, pose and illumination normalization).

1 INTRODUCTION

Three dimensional face tracking is an emerging and a crucial component of many systems in emotional expression analysis, lip reading, identity recognition, surveillance etc. However, this is still a challenging task because of the variability of the faces. This variability arises from the changes in the pose and facial expression deformations and from illumination modifications. Generally speaking, regarding light compensation, most of the methods proposed are established on a set of images in order to compensate the light effects. Some techniques can build geometrically the set of images (basis) (Ishiyama and Sakamoto, 2004), and others can derive them from a collection of pictures taken in different kind of illumination and then use an eigenfaces technique to create the basis (Dornaika and Ahlberg, 2003) ,(Cascia et al., 2000). Otherwise, in our proposed methods, the illuminance basis are created directly during the tracking by optimizing the parameters that control the effects of a light on a 3D face model. We developed an algorithm that can express all the effects due to both expression and illumination, on a 3D face model as a linear composition of a set of basis. One set can describe the changes in the expression (these are integrated in the 3D face model) and the other can deal with illumination changes taking advantage of the 3D face model. An inspiring work concerning illumination changes using basis approach is the one of Hager (Hager and Belhumeur, 1998) and (Eisert and Girod, 1997).They put the basis for illumination into an ef-

Anisetti M., Bellandi V., Arnone L. and Beverina F. (2006).
FACE TRACKING ALGORITHM ROBUST TO POSE, ILLUMINATION AND FACE EXPRESSION CHANGES: A 3D PARAMETRIC MODEL APPROACH.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 318-325
DOI: 10.5220/0001365103180325

ficient 2D tracking algorithm using parametric models. Contrarely to this approach that needs training for creating a good basis, our methods only use the information resulting from a 3D model and a fixed set of illumination sources. Using this illumination compensation based approach, we improved the quality of the 3D tracking also in realistic environment. Concerning 3D tracking in literature, there are several methods that like ours use a 3D template. They can be divided into two categories: one using a geometrical shape (plane or cylinder) and the other using a 3D head wire-mesh model (Anisetti et al., 2005). (Xiao et al., 2002) uses a cylinder to estimate the pose, and an Active Appearance Model (AAM) method is used to map the appearance head model to the face region. This method can handle limited head motion, because after a certain level of rotation the distortion due to the difference between the head and the cylinder are not negligible. The uses of 3D head model is a solution to improve the tracking, specially if the model can also morph according to certain expression parameters. Many recent works choose this solution (Tao and Huang, 1999), (Matthews et al., 2003) and (Dornaika and Ahlberg, 2004). (Tao and Huang, 1999) uses explanation-based facial motion tracking algorithm, based on a Piecewise Bezier Volume Deformation model (PBVD). In this way, they can (in two steps) track the posture and the head deformation. (Matthews et al., 2003), (Dornaika and Ahlberg, 2004), use another method that differs from our warping technique, because theirs is an affine piecewise instead of being based on expression shape basis. These control the position changes of every pixel of the images (not only the vertices) according to the expression parameters. This means that instead of the usual sequence of operations for every pixel (which is finding out which triangle it belongs to and performing for it the affine warp for that triangle) we do an all in one operation. This simplification is made thanks to the 3D face model definition for the face expression for every triangle, which simplifies the operation. Another difference, is that our methods do not need a training set (Matthews et al., 2003), (Dornaika and Ahlberg, 2004), to learn the parameters that control the shape modes of the face. Similarly to the (Cootes et al., 2000) approach, we use an optimization algorithm that matches shape and texture simultaneously. The difference is that we perform it inside the warping algorithm, considering at the same time the face movement parameters and the roto-translation ones. The novelty of our approach, regarding the expression inference, consists in the use of a set of basis that deal with the face deformation directly inside the warping transformation. Concerning the illumination compensation, we use a 3D model for a sort of illumination effect prediction. This implies that together with the warping parameters, we extract the action unit coding (FACS) and the position of the 3 virtual light sources, obtaining a greater precision in tracking. Many algorithms avoid issues related to the illumination estimation by updating the template frame after frame. As a result, any error in motion estimation is consequently propagated. Besides that, they cannot take into account sudden light variations like a light turning on abruptly. On this robust algorithm, we developed some interesting applications that use the normalized face as a final result of pose and expression inferences. For instance, we propose a system for classifying the emotional facial expressions in (Andreoni et al., 2004) or for identifying certification of a subject in (Damiani et al., 2005). In the following sections, we will explain all developed techniques in details.

2 A PARAMETRIZED 3D FACE MODEL

In literature there are many facial models both of 2D and 3D. Considering our objectives, we chose a 3D model for precision in the tracking estimation, for useful in illumination inference and for self occlusion monitoring. Furthermore, we need to adapt the 3D model to an image by aligning some fiducial points with the corresponding points on the 3D model. Therefore the 3D model must be morphable both for expression and for face adaptation on these fiducial points. Summarizing, we need a model with these characteristics:

1. Triangle based. This feature makes the model suited for the affine transformation, that, by definition, maps triangles into triangles. It then becomes useful in expression morphing;
2. Animation and shape parametrization. This makes it possible to describe shape and face expression (Animation Unit) by a simple linear formula:

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

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

Considering these characteristics, a 3D morphable model like (Blanz and Vetter, 2003) could be a optimal choice. Yet our aim is to use a free 3D model not dependent on training faces, and to demonstrate that with only few adaptations we can use a general purpose 3D face model. In our previous works (Bellandi et al., 2005) (Damiani et al., 2005) we chose

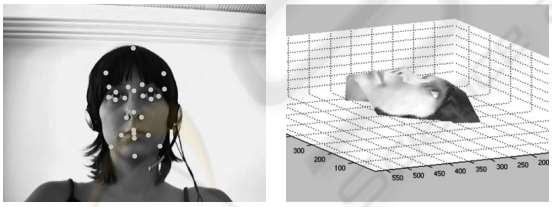
the Candide-3 model for our tests, obtaining good results. We also used the Candide-3 in this work for certain tests, but we decided to develop a more realistic model to deal better with expressions and especially with illumination changes.

2.1 3D Individual Shape Parametrization

Following these main characteristics, we can use the shape parameters of the 3D model σ, for computing a 3D individual shape model and texture on a single frame representing a frontal and neutral view of the subject. This represents the best fitting 3D model template on the subject's face and is tightly coupled to any different individuals. This fit is performed on the frontal frame manually, choosing some fiducial points (they could be selected by any automatic 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 correspondent points $\mathbf{p}$ on the model and the manually chosen ones. We define the error as follows:

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

We weight the error on the model's points according to the intrinsic insecurity of every point selection. 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. Figure(1) shows an example of the creation of an individual template.



(a) Feature Point (b) Template 3D

Figure 1: Example of 3D template extraction process.

Obviously, following this shape minimization we do not deal with the third dimension of the face, but we only consider the 2D deformations as the best fit shape. Nevertheless, using this approximate shape parameterized process, the tracking still produces a good result thanks to the error correction phase explained in the following sections. In order to test the difference between this rough shape model and a more realistic one, we use a real 3D mesh of a subject's face extracted by a stereo camera system.

2.2 3D Model Morphing for Expression Recognition

Another important feature of the 3D model is its morphability. During the tracking process, for expression inference, we deduce expression parameters α, while shape parameters σ remains constant because they represent an intrinsic property of the subject. In order to do this, we must find a set of basis B that can express linearly the changes of the expression shapes of our 3D template $\mathbf{T}$ starting from neutral template $\mathbf{T_0}$ obtaining one formula such as followed:

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

We know that the vertices final position of every triangle of 3D model are:

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

where $\mathbf{V_f}$ is the matrix containing the coordinates of the vertices in the final position, where α_j are the parameters that take into account every single movement of the N possible ones, $\mathbf{A}_j$ is the matrix that contains the components for the j-expression. $\mathbf{A}$ 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 matrix M_k that brings a point belonging to the first triangle into a point on the second in this way:

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

From which it follows that the transformation for each triangle can be written

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

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

derived from the template T_0 in this way:

$$T_{ik} = T_{0k}\mathbf{M}_k \quad (7)$$

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

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

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

2.3 3D Illumination Basis

We created a set of basis in order to compensate the effects of the light changes. Since the light sources are "distant", all the points on a face will be at the same orientation according to the light-source direction. This means we will have the same light intensity reflected back to the viewer from all the points of the face. We can then compute the intensity of that face, depending on the intensity and direction of the light source, the intensity of ambient light, and store that into a matrix. For the Lambertian case, there is no problem with multiple (distant) light sources it is just a matter of adding up their individual contributions. It has been demonstrated that in case of Lambertian surfaces, in absence of self shadowing, given 3 basis of 3 linear independent light source direction, one can reconstruct the image of the surface under a new light direction, by the linear combination of the 3 basis. In fact the irradiance at a point x can be given by

$$I = \alpha \mathbf{nL} \quad (10)$$

where **n** is the normal vector to the surface,α the albedo coefficient and **L** is the power and the direction of incident light rays. So, firstly we calculate the direction of the normal vector of the normal direction of each face. Then we create a matrix in which we associate at every point of the template its normal vector. Naming B_x, B_y and B_z the vectors of each direction component than the new template T can be written starting from the previous template T_0 as:

$$T = T_0 + \sum_{l=1}^{3} \lambda_l (B_x \cos(\theta_l + rot_x) + + B_y \cos(\phi_l + rot_y) + B_z \cos(\psi_l + rot_z)) \quad (11)$$

where θ_l, ϕ_l, ψ_l are the directions of every $l^t h$ light and rot_x, rot_y, rot_z are the estimations of the rotation of the template. The λ_l parameters are the intensity of each light, and the parameters that will be estimated by the algorithm. Furthermore, this set of normal vector basis are useful in the managing of the hidden triangle in the tracking algorithm. Thanks to that we can know which is the facets that is not visible anymore and we do not use it in the tracking algorithm.

3 3D MOTION, EXPRESSION, AND ILLUMINATION RECOVERY

Our goal is to obtain posture estimation parameters, the AUV deformation parameters and illumination condition parameter in one minimization process between two frames using morphing and illumination basis technique explain in previous session. First of all, for clearness, we describe the steepest descent algorithm for 3D posture and morphing estimation. Based on the idea that 2D face template $T_i(x)$, (extracted by projecting 3D Template T on image plane) appears 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 movement parameters and x are pixel coordinate from image plain, we can obtain the movement and expression parameter p by minimization of the function (12); in fact if $T_i(x)$ is the template at time t with the correct pose and the 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 $t+1$ is obtain by minimization of sum of squared error between $T(x)$ and $I(W(x;p))$:

$$\left(\sum_x [I(W(x;p)) - T(x)]^2 \right) \quad (12)$$

For this minimization we use an approach like (Lucas and Kanade, 1981) with forward additive implementation that assumes that current estimate of p is known and iteratively solves for increments to the parameters Δp. Equation (12) after some well known passage becomes:

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

$$H = \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T \left[\nabla I \frac{\partial W}{\partial p} \right] \quad (13)$$

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

cremental warp parameters. Because we need 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: $X(t + 1) = (X(t) + \sum_{i=1}^{m}(\alpha_i \cdot B_i))$. Where α_i and B_i follows expression based representation described in the previous section and the matrix M follows Bregler (Bregler and Malik, 1998) and the twist representation by (Murray et al., 1992). With these matrix the motion parameters p becomes $(\omega_x, \omega_y, \omega_z, t_x, t_y, t_z, \alpha_1, \ldots, \alpha_m)$ presented in equation , α_i and B_i follows expression based representation described in the previous section.With this consideration the warping $W(x; p)$ in (12) becomes:

$$W(x; p) = M(X + \sum_{i=1}^{m}(\alpha_i \cdot B_i)) \quad (14)$$

In situation of perspective projection, assuming the camera projection matrix depends only on the focal length f_L, the image plane coordinate vector x is obtain with:

$$\mathbf{x(t+1)} = \begin{bmatrix} x - y\omega_z + z\omega_y + t_x + B_x \\ x\omega_z + y - z\omega_x + t_y + B_y \end{bmatrix} \cdot \frac{f_L}{-x\omega_y + y\omega_x + z + t_z + B_z}(t) \quad (15)$$

where:

$$\begin{aligned} B_x &= \textstyle\sum_{i=1}^{m}(\alpha_i(a_i - b_i\omega_z + c_i\omega_y)) \\ B_y &= \textstyle\sum_{i=1}^{m}(\alpha_i(a_i\omega_z + b_i - c_i\omega_x)) \\ B_z &= \textstyle\sum_{i=1}^{m}(\alpha_i(-a_i\omega_y + b_i\omega_x + c_i)) \end{aligned} \quad (16)$$

This function maps the 3D motion and morphing in image plane. Following the Lucas-Kanade algorithm the Jacobian matrix $\frac{\partial_W}{\partial p}$ at $p = 0$ becomes:

$$\begin{bmatrix} -xy & (x^2+z^2) & -yz & z & 0 & -x & +DB_x \\ -(y^2+z^2) & xy & xz & 0 & z & -y & +DB_y \end{bmatrix} \cdot \frac{f_L}{z^2}(t) \quad (17)$$

where:

$$\begin{aligned} DB_x &= (a_1 \cdot z - c_1 \cdot x) + \cdots + (a_m \cdot z - c_m \cdot x) \\ DB_y &= (b_1 \cdot z - c_1 \cdot y) + \cdots + (b_m \cdot z - c_m \cdot y) \end{aligned} \quad (18)$$

Using a forward additive parameter estimate approach, we are able to obtain the correct 3D motion posture and morphing parameters of the template between two frames in one minimization phase. In order to obtain more robustness for global and local illumination changes, we also introduce in our minimization algorithm another five additional parameters using linear appearance variations. If we consider the image template $T(x)$ as:

$$T(x) + \sum_{i=1}^{5} \lambda_i B_i(x) \quad (19)$$

where $B_i i = 1, \ldots 5$ is a set of known appearance variation images and $\lambda_i, i = 1, \ldots 5$ are the appearance parameters. Global illumination changes can be modelled as an arbitrary change in gain and bias between the template and the input image by setting B_1 to be the T template and B_2 to be the unitary "all in one" image. For lateral illumination we use the other illumination basis $B_i i = 3, \ldots 5$ explained in the previous section. Using the equations 19 instead of the $T(x)$ in 12 we obtain the following equation that we should minimize:

$$min\left(\sum_{x}[I(W(x; p)) - T(x) - \sum_{i=1}^{5}\lambda_i A_i(x)]^2\right) \quad (20)$$

In accordance to the linear appearance variations technique, this can be minimized using the steepest descent approach. In Figure (2) there are some examples of tracking experiments with illumination changes in realistic environment with a standard lowquality webcam. We demonstrated the improvement in accuracy of posture and expression estimation with this one step minimization process and the quality of 3D model AUV tracking in the result section.

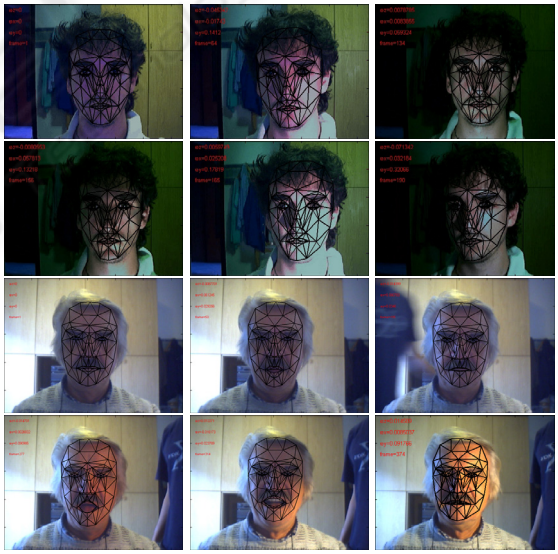


Figure 2: Example of tracking during extreme illumination condition variation and with variation in pose and expression. In black the tracking mask.

4 TEMPLATE MANAGEMENT

Recovery head position, AUV face morphing with different illumination environment, as explained before, is a difficult task for a 3D algorithm because the face has a complex 3D mesh that produces,

when moving, frequent illumination changes and self-occlusions, and because of the difference between the real face and the adapted 3D model. For that reason, to avoid these problems we have developed some techniques that we generally call "Template management". One of these techniques concerns the selection of the significative part of the 3D Template mask for performance improvement. In this context, we perform this optimisation by tresholding the gradient images in order to reduce the number of points in the face template. By our experience, the quality of our tracking does not decrease much by reducing the number of points and on the other hand the speed of the algorithm increase of an magnitude order.

4.1 Dissimilarity Analysis

Our principal goal is to estimate posture, expression and illumination parameters of the subject, in order to reconstruct the normalized neutral frontal view of the face (reported face). To analyze the quality of our tracking process, we have developed a technique called "dissimilarity analysis" that works on the retreived face. In fact, if the estimated posture, illumination and morphing parameters are correct, the frontal image will be consistent with the face template(Figure(3)). With this approach, good tracking occurs in good normalization reconstruction. If there is any approximation error in the posture evaluation, the normalized face suffers of distortion effect depending on the amount of the error. For this reason, by analyzing the retrieved images we can estimate the quality of the tracking that we define the "dissimilarity level". This analysis is performed by a 2D tracking algorithm (with an inverse compositional implementation) of some fiducial points on the normalized face (an extended set is represented as "+" in Figure(3). If the tracker of this fiducial points shows that there is a translation bigger than a threshold, we will label this tracking with low confidential level or with high dissimilarity. Dissimilarity is useful when we do not have the correct posture or expression, but we need to know how accurate our approximations are.

Figure 3: Example of posture, morphing estimation and frontal view normalization.

4.2 Dynamic template update and mosaic technique

After the head pose, the AUV morphing, the illumination parameter is obtained by the 3D Motion technique. We update our 3D template with the one recovered from the current frame if the confidence level of dissimilarity is enough. Otherwise, the template is not modified. With this template updating strategy, we can track the head and the face movements and partially deal with drift problems derived from the dynamic update. To maintain a good performance, we continue to update every parameter except the texture, so that we do not introduce errors in the template image that is the main responsible for the drift effect. Therefore,this solution is not enough for a long time tracker. For that reason, in the previous work, we had introduced a technique called "mosaic template" (Anisetti et al., 2005). This technique consists in creating and dynamically updating a collection of templates according to the position the subject is in. Practically speaking, it is the same as storing some head poses and the relative templates. When the estimated head pose is close to the one of that registered template, we use it for correcting the drifting problem by re-alignment. In this way, the drifting effect is strongly limited without many correction steps. This also permits to adapt the correction to the dynamic changes of the environment and of the subject himself (Figure(4)). During the posture correction phases in mosaic techniques, expression parameter are also re-corrected.

Figure 4: Example of 3D tracking with mosaic correction (right) during ELITE2002 experiment session.

4.3 Occlusion management

Another important problem about face tracking is the face occlusion. There are two types of occlusion: self-

occlusion (posture occlusion), and occlusion by object that do not belong to the face. Because of the presence of this "external factors", some pixels in the face template should contribute less (or not at all) to the motion estimation. To perform this, we apply a well known IRLS technique with a modified compensated approach used by (Xiao et al., 2002). Regarding self-occlusion, our occlusion manager determines the hidden facets by posture analysis. This techniques combined with mosaic and dissimilarity analysis, permits to prevent drift error and wrong mosaic template registration problems that may occurs with some other techniques presented in literature.

5 EXPERIMENT RESULTS

We have conducted three type of experiments for evaluating the precision of our system. For testing the tracking quality improvement in cases of morphing and no morphing 3D tracking. Secondly we tested our algorithms for extracting the features linked to the AU (Ekman and Friesen., 1978) on the Cohn-Kanade Data-Base. Finally we tested our illumination parameter estimation with synthetic model videos and in a realistic environment . For the first experimental evaluation, we used our data base including 10 different subjects that move, change expression and talk, registering during ELITE2002 tracking experimental session. For the real movement evaluation, the database was recorded in the Politecnico laboratory of Milan with a commercial web-cam at a resolution of 640x480 synchronized with ELITE2002. ELITE2002 system is an optoelectronic device able to track the three-dimensional coordinates of a number of reflecting markers that we placed on a helmet on the subject's head and on the web-cam. This system, thanks to a set of 6 cameras, can perform a tracking of a point with a range precision of 0.3 mm. Thanks to this high posture estimation confidence, we are able to compare our 3D tracking with real subject movement. Figure (5) shows an example of estimate tracking values for yaw rotation (the major rotation in the presented sequence) with and without morphing comparing with ELITE2002. It is clear that if the model could estimate morphing parameters, posture evaluation would become more precise. By our experiments, the error of tracking with expression morphing is at maximum 2-3% compared to the ones by no morphing tracking. This is an impressive result considering that the no morphing tracking with some occlusion technique improvement has good results: a maximum error of 5 grades comparing to ELITE2002. During these experiments, we also monitored the dissimilarity values that describe the quality of the tracking, obtaining that with morphing the dissimilarity values still

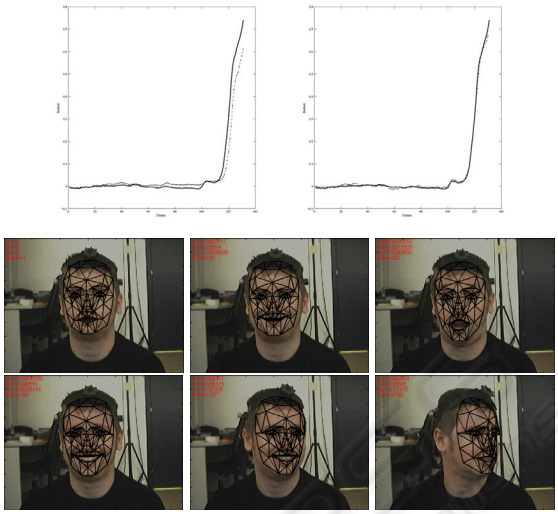


Figure 5: Comparison between ELITE2002 (solid line) pose estimate and no morphing(left) and morphing (right) tracking posture estimate. We also present some crop frame.

remain smaller than without the morphing. These results confirm the strong correlation between the quality of the tracking and the dissimilarity value and reinforce the improvement of the morphing for tracking purposes. Regarding the second type of experiments, the table(1) shows the link between the AUV of Candid-3 model extracted during the tracking and the AU of FACS codify used for comparing our results on Cohn-Kanade Data-Base. The same things could be done with the morphing basis of any 3D morphable face model. Our results using this data base

Table 1: Link between the AUV of Candide-3 model extracted directly during the tracking and the classical AU

AUV	AU	AUV	AU
1	10	7	42 43 44 45
2	25 26 27 17	8	7
3	20	9	9
4	4 1	10	23 24 28
5	12 13 15	11	5
6	2		

is very promising even using a simply fuzzy classifier. In the last type of experiments, we tested the quality of the tracking with illumination changes. To do this, we made synthetic cases where the subject and the light source rotate in different situations (Figure(6)). We obtain better quality using illumination techniques than using IRLS or other weight techniques. Further tests were done, performing the tracking in situations characterized by extreme and localized illumination conditions (Figure(2)) that thanks to illumination basis becomes trackable. Experiments were also made in realistic situations with different light sources together with expression changes (Figure(2)), and a

comparison with ELITE2002 system with changes in illumination. Sumarizing we obtain a better precision and an extended trackability in all cases of strong illumination changes.

Figure 6: Example of synthetic test for illumination change. (a) shows the tracked face, (b) and (c) shows face normalizate without and with illumination adjusting.

6 CONCLUSION

Concluding, we developed a robust expression analysis oriented face tracker with posture confidence evaluations, that makes the tracking good and very close to ELITE2002 estimation. The algorithm proposed is robust to face morphing and illumination changes in spite of the difference between the 3D face model and the real subject face. Our system performs good results thanks to the correction techniques like the mosaic ones and the dissimilarity analysis. We also showed that this method permits to extracted many measures linked to the AU that can be used for face expression detection.

REFERENCES

Andreoni, C., Anisetti, M., Apolloni, B., Bellandi, V., Balzarotti, S., Beverina, F., Campadelli, P., M.R.Ciceri, P.Colombo, F.Fumagalli, G.Palmas, and L.Piccini (2004). E(motional) learning. In *Technology Enhanced Learning 2004 (TEL04), Milan Italy*.

Anisetti, M., Bellandi, V., and Beverina, F. (Sept. 2005). Accurate 3d model based face tracking for facial expression recognition. In *Proc. of International Conference on Visualization, Imaging, and Image Processing (VIIPO5)*, pages 93 – 98.

Bellandi, V., Anisetti, M., and Beverina, F. (Sept. 2005). Upper-face expression features extraction system for video sequences. In *Proc. of International Conference on Visualization, Imaging, and Image Processing (VIIP05)*, pages 83–88.

Blanz, V. and Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063 – 1074.

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

Cascia, M. L., Scarloff, S., and Anthitsos, V. (2000). Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2000 (22)(4):322–336.

Cootes, T., Edwards, G., and Taylor, C. (Jun. 2000). Active appearance mode. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681 – 685.

Damiani, E., Anisetti, M., Bellandi, V., and Beverina, F. (2005). Facial identification problem: A tracking based approach. In *IEEE International Symposium on Signal-Image Technology and InternetBased Systems (IEEE SITIS'05)*.

Dornaika, F. and Ahlberg, J. (2003). Face and facial feature tracking using deformable models. *International Journal of Image and Graphics*.

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

Eisert, P. and Girod, B. ('July 1997). Model-based 3d-motion estimation with illumination compensation. In *Conference Publication*.

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

Ferrigno, G. and Pedotti, A. (1985). Elite: a digital dedicated hardware system for movement analysis via real-time tv signal processing. *IEEE Trans Biomed Eng.*, pages 943–950.

Hager, G. D. and Belhumeur, P. N. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1998 (20)(10):322–336.

Ishiyama, R. and Sakamoto, S. (2004). Fast and accurate facial pose estimation by aligning a 3d appearance model. In *Proc. of 17th international conference on pattern recognition (ICPR'04)*.

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

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

Matthews, I., Ishikawa, T., and Baker, S. (2003). The template update problem. In *Proc. of the British Machine Vision Conference*.

Murray, R., Li, Z., and Sastray (1992). *A mathematical introduction to robotic manipulation*. CRC press.

Tao, H. and Huang, T. (1999). Explanation-based facial motion tracking using a piecewise bier volume deformation model. *In CVPR99*.

Xiao, J., Kanade, T., and Cohn, J. (2002). Robust full-motion recovery of head by dynamic templates and re-registration techniques. *Proc. of Conference on automatic face and gesture recognition*.