

ESTIMATION OF 3D MOTION AND STRUCTURE OF HUMAN FACES

Y. Xirouhakis, G. Votsis and A. Delopoulos
Computer Science Division,
Department of Electrical Engineering,
National Technical University of Athens,
Athens GR-15773, GREECE

1 Introduction

The extraction of motion and shape information of three dimensional objects from video sequences emerges in various applications especially within the framework of the MPEG-4 and MPEG-7 standards. Particular attention has been given to this problem within the scope of model-based coding and knowledge-based *3D* modeling. In this chapter, a novel algorithm is proposed for the *3D* reconstruction of a human face from *2D* projections. The obtained results can contribute to several fields with an emphasis on *3D* modeling and characterization of human faces.

The problem of modeling human faces from their projections has been tackled before by several authors such as [1, 2]. However, the accuracy of the obtained results is still far from the desirable. In [1] an algorithm is provided for the estimation of *3D* motion and structure from two orthographic projections (two frames) having available a relatively accurate initial guess of the object structure, for example a generic wireframe model for the case of human faces. In [2] a modification to the above algorithm is proposed in order to increase its tolerance to errors in the initial guess.

In the aforementioned approaches it is assumed that the human face is a generally rigid object. Local deformations of the face, such as mouth and eye movement, are generally tackled separately as local motions. This constraint is also set in this work, i.e. the human head is assumed to be globally rigid. Similarly to other approaches, it is assumed that there is no camera movement and that the video is obtained under orthographic projection. The latter, as it can be seen in [6], yields

satisfactory results and significantly reduces computational cost.

In the present work, the algorithm presented in [5] is utilized in order to obtain an accurate solution for the object's 3D motion and structure (without any initial guess for the object's 3D structure) employing three orthographic projections. The algorithm exploits the information provided by the fact that the object is a human head, in the sense that keypoints and keylines are extracted increasing the accuracy of the 2D motion estimates. As it will be quoted in the sequel, accuracy in the 2D motion estimates plays a significant role to the accuracy of depth estimates.

2 Overview

The first step of the presented algorithm is to obtain 2D motion estimates of the most characteristic face elements (iris and edges of the eyes, nasal points, lips edges, etc.). Reliable estimation of these feature correspondences forms a crucial presumption for 3D information extraction. In this work, reliable estimation of corresponding points is obtained through the use of adjustable template matching involving correlation coefficient calculations for 2D feature templates (see Section 3).

The second step uses these 2D point correspondences in order to yield estimates of the motion parameters corresponding to the actual 3D movement of the human face. These include the rotation angles and axes as well as the translation vectors that describe the movement of the object in 3D space (see Section 4).

The third step computes 3D shape information (depth and relative position) of the initial characteristic points. Computation is based on the 2D point correspondences obtained in the first step and the rotation/translation estimates of the second step. A variant of block-based 2D motion estimation algorithms that exploit the previously estimated motion parameters of the face are utilized for all points on the image (see Section 5). The estimates that these algorithm yield for the remaining - non characteristic - points of the face are next used for the accurate extraction of 3D position information of the human face.

Simulated experiments illustrating the performance of the proposed algorithms have been included.

3 Facial Feature Extraction

Research on face recognition has shown that some specific features of the face can be easily detected. Among these, we choose to extract those that can be considered as rigid points, throughout an arbitrary video sequence including the face. The relative solidity of the feature points is a crucial issue, since it affects the correct calculation of the rotation matrices. Such features can, for example,

be the eyes and nasal points. As it will be seen, the accuracy of computation of motion parameters depends upon the reliability of the extraction of a sufficient number of characteristic points (at least four) on the complex surface of the face.

Additionally, the above set is enriched by features that may deform along time, so as to achieve 3D shape recovery by attaining a dense wireframe mask of the human face. Thus, supplementary points may be placed, e.g. on the chin and cheek contours and the lips outline.

A variety of techniques have been proposed for extracting characteristic points in images containing faces. These techniques comprise simple or more elaborate tools such as curvature extrema, Fourier descriptors, generalized symmetry operators, the KL transform, the Hough transform, adjustable template matching, Gabor wavelet decomposition and knowledge-based vision systems [9]. In our approach, a few variations of template matching techniques are adopted.

3.1 MATCHING 2D FEATURES

Firstly, some characteristic 3D regions of a generic head model are isolated and projected onto the 2D space, in order to derive generic templates containing the features to be tracked. In fact, 3D generic templates are utilized in order to obtain a set of views (affine transformations) for each characteristic region of the face. This task aims at finding the region in the 1st frame that matches best each generic template. Since in the first frame we may not have a frontal view of the human face, one element from each set is appropriate. In this way, each characteristic region is derived in the first frame. In the remaining frames the characteristic regions can be obtained similarly. Equivalently, the generic templates can be substituted by the ones matched in the 1st frame, after they are subjected to a number of affine transformations forming new sets of characteristic templates.

Template matching is being performed with the use of correlation coefficient. Supposing that images are viewed as intensity matrices, the correlation coefficient between the prototype (matrix \mathbf{A}) and the repeatedly audited blocks of an appropriately selected area of the face (matrix \mathbf{B}) is computed as follows:

$$r = \frac{\sum_{n_1} \sum_{n_2} \mathbf{A}(n_1, n_2) \mathbf{B}(n_1, n_2)}{\sqrt{\sum_{n_1} \sum_{n_2} \mathbf{A}^2(n_1, n_2) \sum_{n_1} \sum_{n_2} \mathbf{B}^2(n_1, n_2)}} \quad (1)$$

The template of each group that achieves maximum correlation coefficient with its corresponding block of the image of interest, is the one used. If, however, the maximum correlation coefficient for a specific group doesn't exceed a confidence threshold, the specific feature is considered as not tracked. The ease for each feature's detection has been determined through years of research by computer engineers and neuro-physiologists [9]. The most reliable features are extracted

first. The knowledge of their position restricts the search area for the rest of the features, making their tracking faster and more confident.

Subsequently, hierarchical filtering through 'steerable filters' is used to detect facial keypoints. The term 'steerable filters' is used to describe the reconstruction of deformed filter kernels that are created by superimposing a small number of orthogonal basis functions, as described by the following formula:

$$\mathbf{F}_\alpha(\mathbf{x}) = \sum_{k=1}^N b_k(\alpha) \mathbf{A}_k(\mathbf{x}) \quad (2)$$

where \mathbf{F} denotes the kernel, \mathbf{A}_k denote the basis functions, b_k denote their weights and α is a general multi-deformation [10]. In this way, more complex filters are composed by simple ones, while filtering with more sophisticated filters is reduced to convolving the image with a small number of basis functions. Additionally, any number of basis functions reconstructs all deformed kernels, affecting only the reconstruction quality. In our case, low quality approximations are adequate, a fact that results in a small number of basis functions. For the detection of the iris, a model driven approach is used as in [10].

Alternatively, easier processing tools are being practiced, such as the integral projections towards vertical and horizontal orientation, as shown by the following relations respectively:

$$\mathbf{v}(x) = \sum_{y=y_1}^{y_2} \mathbf{I}(x, y) \quad (3)$$

$$\mathbf{h}(y) = \sum_{x=x_1}^{x_2} \mathbf{I}(x, y) \quad (4)$$

Simple image processing tools proved adequate for plain detection, while the more elaborate tool of hierarchical filtering is used for complex point extraction. Thus, characteristic face points are detected accurately within the tracked templates.

3.2 LINE MATCHING

Adjustable templates consisting of parabolas [11] are used for representing features, such as the chin contour. In this chapter, such features will be utilized in order to restrict the search area during the motion estimation procedure for all (non-characteristic) points on the face.

The area of the jaw that is visible in each frame can be automatically restricted, since the blocks already detected are enough to indicate its approximate position. By performing edge detection to this restricted area we find a set of points that are candidates for the description of the jaw line. Given these points, the estimation

of the jaw line is provided by its approximation by a parabola or a second order line. In the case of the second order line approximation, a line described by $y = \alpha x^2 + \beta x + \gamma$ is derived by computing the unknowns α, β, γ . Thus the candidate points located by edge detection techniques are approximated with the appropriate second order line. This is performed with a Least Squared distance approximation of that set. By defining the distance of each point from the curve as $d_i = y_i - \alpha x_i^2 - \beta x_i - \gamma$, and by minimizing the squared sum for all the candidate points, we get:

$$\theta_{LS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

where \mathbf{X} is the $N \times 3$ table containing in line i the vector $[x_i^2 \ x_i \ 1]$, \mathbf{y} is the $N \times 1$ vector containing in line i the element y_i , while θ_{LS} is the vector containing the factors α, β, γ .

The above procedure is performed two times. After the first iteration and the determination of the unknown parameters, the points whose distance from the curve exceeds a threshold are discarded. In the second iteration, the calculations are performed using only the remaining points. Thus, the curve is now approximated precisely.

4 Computation of Motion Parameters

Earlier approaches in the problem of estimating 3D motion include [7, 8] in the case of perspective projections, and [4, 1, 2, 5] in the orthographic case. In this paper, the algorithm proposed in [5] is employed, since it provides an exact, computationally attractive solution even in the presence of noise. It should be mentioned that orthography poses no considerable limitation in the application of the proposed algorithm, as it is a reasonable approximation to perspective projection, when the object is relatively far away from the camera.

4.1 DEFINITIONS

The movement of a rigid object in 3D space can be seen as superposition of a 3D rotation and a 3D translation. Consequently, when a point (x, y, z) on the object moves to (x', y', z') , it holds:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{T} \quad (6)$$

where \mathbf{R}, \mathbf{T} are the rotation and translation matrices respectively. Generally:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad \mathbf{T} = [T_1 \ T_2 \ T_3]^T \quad (7)$$

Assuming a rotation about an arbitrary axis, matrix \mathbf{R} can be written in cartesian coordinates in terms of the rotation axis $\mathbf{n} = [n_1 \ n_2 \ n_3]^T$ and rotation angle a , as:

$$\mathbf{R} = \begin{bmatrix} n_1^2 + (1 - n_1^2) \cos a & n_1 n_2 (1 - \cos a) - n_3 \sin a & n_1 n_3 (1 - \cos a) + n_2 \sin a \\ n_1 n_2 (1 - \cos a) + n_3 \sin a & n_2^2 + (1 - n_2^2) \cos a & n_2 n_3 (1 - \cos a) - n_1 \sin a \\ n_1 n_3 (1 - \cos a) - n_2 \sin a & n_2 n_3 (1 - \cos a) + n_1 \sin a & n_3^2 + (1 - n_3^2) \cos a \end{bmatrix} \quad (8)$$

The computation of motion parameters involves the estimation of matrices \mathbf{R} , \mathbf{T} . In fact, as far as the rotation matrix \mathbf{R} is concerned, the estimation of the rotation axis \mathbf{n} and rotation angle a is necessary. It must be noticed that under orthography, the component T_3 of the translation matrix, which is perpendicular to the projection plane, cannot be estimated by any scheme. The computation of the structure parameters involves estimation of the z, z' coordinates for each point in (6), given the point correspondences (given in fact coordinates x', x, y', y) and having computed the motion parameters. However, absolute depth cannot be estimated by any scheme under orthographic projections, as in this case absolute depth information is lost due to the type of projection. In fact, relative point depth is estimated, providing 3D shape information.

4.2 REVIEW OF THE ALGORITHM

In this subsection, we will follow the notations used in [5] for compatibility matters.

Ullman in his classical work [3] proved that four point correspondences over three frames are sufficient to yield a unique solution to motion and structure up to a reflection. The latter is verified in all proposed algorithms [4, 5]. In this sense, a small set of at least four relatively exact point correspondences, as described in the previous section, are needed for computing the rotation and translation matrices.

According to Stokes assertion, any movement such as (6) can be described by a translation \mathbf{T} equal to the point correspondence of an arbitrarily chosen point P on the object, and a rotation \mathbf{R} about an appropriate axis and considering point P as the rotation origin. Consequently, by picking a point P and assigning P to the world origin $(0, 0, 0)$, translation \mathbf{T} is computed as the point correspondence of P . In this sense, the computation of \mathbf{T} is straightforward.

The estimation of matrix \mathbf{R} is achieved through the use of the so-called 2x2 matrices $\tilde{\mathbf{K}}$ which can be expressed in terms of the differential reference and motion vectors for every three points as,

$$\Delta \mathbf{v} = \tilde{\mathbf{K}} \Delta \mathbf{r} \quad (9)$$

where $\Delta \mathbf{r} = [\mathbf{r}_3 - \mathbf{r}_1 \ \mathbf{r}_2 - \mathbf{r}_1]$, and $\Delta \mathbf{v} = [\mathbf{v}_3 - \mathbf{v}_1 \ \mathbf{v}_2 - \mathbf{v}_1]$. Vectors $\mathbf{v}_i, \mathbf{r}_i$ are the motion and the reference vectors respectively for a point on 1st frame.

Similarly, matrix \mathbf{K} can be used, defined as:

$$\Delta \mathbf{r}' = \mathbf{K} \Delta \mathbf{r} \quad (10)$$

where $\Delta \mathbf{r}' = [\mathbf{r}'_3 - \mathbf{r}'_1 \quad \mathbf{r}'_2 - \mathbf{r}'_1]$. Vector \mathbf{r}'_i is the reference vector of the corresponding point on the 2nd frame. It can be seen that matrices \mathbf{K} , $\tilde{\mathbf{K}}$ are the same for any three points on the same 3D planar surface. The rotation matrices can then be obtained using some manipulations, in terms of eigenvalues and eigenvectors, as it will be quoted in the sequel. In order to simplify the following equations, let matrix \mathbf{L} be the adjoint matrix:

$$\mathbf{L} = \text{adj}(\tilde{\mathbf{K}} + \mathbf{I}) = \text{adj}\mathbf{K} \quad (11)$$

where \mathbf{I} denotes the 2x2 identity matrix. Since \mathbf{L} also characterizes the projection of a 3D planar surface to the projection plane, let \mathbf{L}_{Rj} be the \mathbf{L} -matrix that corresponds to rotation \mathbf{R} of the plane j . In accordance to Ullman's assertion, four \mathbf{L} -matrices contain sufficient information, in order to estimate the motion and structure parameters. In fact, employing four point correspondences over 3 frames, two \mathbf{L} -matrices are derived corresponding to movement frame 1-2 and another two corresponding to movement from frame 1-3. Without loss of generality, let for simplicity matrix \mathbf{R} correspond to movement frame 1-2 and matrix \mathbf{S} correspond to movement frame 1-3. Using more than four \mathbf{L} -matrices, a Weighted Least Squares scheme is provided in [5] to improve the estimation of the motion parameters.

Let the rotation matrix \mathbf{R} be written as:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 \\ \mathbf{r}_2^T & r_{33} \end{bmatrix} \quad (12)$$

where $\mathbf{r}_0 = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$, $\mathbf{r}_1 = [r_{13} \quad r_{23}]^T$ and $\mathbf{r}_2 = [r_{31} \quad r_{32}]^T$.

Let matrix $\Delta \mathbf{L}_{Rk}$ be the difference between two arbitrary \mathbf{L} -matrices corresponding to the same rotation.

$$\Delta \mathbf{L}_{Rk} = \mathbf{L}_{Ri} - \mathbf{L}_{Rj} \quad (13)$$

Then, the unknown vector \mathbf{r}_1 is computed within a scalar ambiguity ρ as the unit-norm eigenvector of the summation of matrices:

$$\mathbf{Y}_R = \frac{1}{M} \sum_{k=1}^M \Delta \mathbf{L}_{Rk}^T \Delta \mathbf{L}_{Rk} \quad (14)$$

corresponding to its smallest eigenvalue. In this sense, let $\mathbf{r}_1 = \rho \mathbf{c}_1$. Subsequently, \mathbf{r}_2 is estimated within the same scalar ambiguity ρ as

$$\mathbf{r}_2 = -\frac{1}{N} \sum_{j=1}^N \mathbf{L}_{Rj} \mathbf{r}_1 = -\frac{1}{N} \sum_{j=1}^N \rho \mathbf{L}_{Rj} \mathbf{c}_1 = \rho \mathbf{c}_2 \quad (15)$$

In the above equations notice that N, M are the numbers of $\mathbf{L}, \mathbf{\Delta L}$ matrices involved in the computations respectively. Generally $M \cong \frac{N}{2}$. Let also λ_R denote the larger eigenvalue of matrix \mathbf{Y}_R . Notice as well that $r_{33}^2 = 1 - \rho^2$, since $\text{norm}(\mathbf{c}_1) = \text{norm}(\mathbf{c}_2) = 1$. Similar equations hold for rotation matrix \mathbf{S} . Let similarly λ_S denote the respective eigenvalue and $\mathbf{s}_1 = \sigma \mathbf{d}_1, \mathbf{s}_2 = \sigma \mathbf{d}_2$. It can be proved that the ratio $\frac{\sigma}{\rho}$ can be estimated up to a sign ambiguity as

$$w^2 = \left(\frac{\sigma}{\rho}\right)^2 = \frac{\lambda_S}{\lambda_R} \quad (16)$$

where one of the solutions can be easily rejected. Finally, r_{33} and s_{33} are computed from:

$$\begin{bmatrix} w\mathbf{Jc}_2 & -\mathbf{Jd}_2 \end{bmatrix} \begin{bmatrix} r_{33} \\ s_{33} \end{bmatrix} = -\frac{1}{N} \sum_{j=1}^N \begin{bmatrix} \mathbf{L}_{Sj} & \mathbf{L}_{Rj} \end{bmatrix} \begin{bmatrix} \mathbf{Jd}_1 \\ -w\mathbf{Jc}_1 \end{bmatrix}. \quad (17)$$

where $\mathbf{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. The rotation matrices \mathbf{R}, \mathbf{S} are estimated on the basis of $\mathbf{r}_1, \mathbf{r}_2, r_{33}, \mathbf{s}_1, \mathbf{s}_2, s_{33}$. Thus all the motion parameters are obtained.

4.3 ESTIMATION IN THE PRESENCE OF NOISE

In order to tackle the problem of noisy motion vector estimates, the proposed algorithm is extended in [5] to employ a large number of projected points. However, in this work only a small number of facial points is used. Moreover, these point correspondences are given great confidence, so the employment of the solution in noisy environments is not necessary. In fact, in order to take into consideration the respective confidence with which each facial characteristic is estimated, confidence weights are incorporated in the sums of (14) and (15).

5 Computation of Structure Parameters

Given the rotation matrices, the relative depth of the points on each of the three frames, can be estimated employing Equation (6). The z, z' coordinates are subsequently computed. In fact, it is possible to recover depth only for points whose correspondences are known, while accuracy in depth estimation, in general, requires accuracy in the estimation of all respective point correspondences. Thus, in order to recover depth for all points on the human face, a block-based motion estimation scheme is performed for all points on the images. To achieve faster and more accurate motion estimation, we have to exploit all information gained in the first step, which involves relatively precise detection of certain facial keypoints and keylines. For example, to confine the area of the cheeks -which is generally smooth, and is therefore very sensitive as far as motion estimation is concerned-

we used the limits of the neighbouring blocks tracked during the first step of the algorithm. At the same time, the rotation matrices obtained in the second step yield an a-priori estimation of the dense motion field and impose a constraint to the orientation of motion vector estimates. The acquired motion field, although not perfectly accurate, is appropriate for 3D reconstruction of the human face. In order to improve the reconstructed mask, the latter can then be compared to a generic wireframe model.

6 Simulations

The model depicted in the figures is the model 'Conor' found in Headus (<http://www.headus.com.au/vrml/conor10.wrl>). It is a scanned face of a child mapped with its texture. The model itself was not available, so it was rotated on-line in the vrml environment and grabbed. For this reason, the following results are not arithmetically verified. However, the efficiency of the algorithm is visually apparent.

In Figures 1(a) – (c), the three input frames are depicted with the tracked templates indicated. As it was mentioned in Section 3, 3D templates are employed for the production of sets of affine-transformed 2D templates. Such a generic template (eye template) is depicted in Figure 2(d). Localization of specific points on the human face with the use of hierarchical filtering and simple tools is performed. The derived point correspondences between frames 1 and 2 are depicted in Figure 1(d).

Using the aforementioned methodology for the computation of rotation parameters, we obtain angles $a = 37.8^\circ$, $b = 39.7^\circ$ and rotation axes $\mathbf{n}_a = [0.54 \ 0.84 \ 0.07]^T$ and $\mathbf{n}_b = [0.27 \ 0.96 \ 0.03]^T$ respectively. The obtained motion parameters were experimentally verified subjecting a generic wireframe model to the same rotations.

In the sequel a block matching technique is employed in order to obtain the final structure parameters. Significant help for this scheme is provided by the previously described line-matching scheme as well as by the tracked templates and the 3D motion parameters. Results of the line-matching scheme for the jaw line are depicted in Figure 2(a). The restricted search area for the cheek points is depicted in Figure 2(b). The motion field derived through the motion estimation procedure between frames 1 and 2 is illustrated in Figure 2(c), where only an indicative set of motion vectors is depicted. The reconstructed human mask is depicted in Figure 2(e). Since the motion parameters were estimated in accurate fashion, the reconstructed model resembles well the human face apart from an amount of noise. This noise in depth coordinates is due to the noise in the motion field estimates. Using some heuristical smoothing techniques, the model is slightly improved in Figure 2(f). Study and development of more systematic 3D smoothing techniques

is currently in progress.

7 Conclusions - Future Work

In the current work, the problem of 3D motion and structure extraction is treated in the case of human faces. This is achieved using a generalized algorithm for the estimation of 3D motion and structure parameters of rigid objects under orthography. In the particular case of human faces, the algorithm performance is aided by facial feature extraction schemes that employ adjustable template matching techniques. 2D point correspondences of reliably tracked facial feature points is applied to determine the rotation matrices, whereas the obtained 2D structure information of the human face is used for the extraction of 3D structure through a variant block-based motion estimation technique. Relevant experimental results were presented using views of a head model.

Study and development of systematic 3D smoothing and texture-mapping techniques is currently in progress in order to obtain realistic presentations of the obtained results.

References

- [1] K.Aizawa, H.Harasima, and T.Saito, "Model-based analysis-synthesis image coding (MBASIC) system for a person's face," *Signal Proc.: Image Comm.*, no.1, pp.139-152, 1989.
- [2] G.Bozdagi, A.M.Tekalp, and L.Onural, "An Improvement to MBASIC Algorithm for 3-D Motion and Depth Estimation," *IEEE Trans. Im. Proc.*, vol.3, no.5, pp.711-716, Sept. 1994.
- [3] S.Ullman, "The Interpretation of Visual Motion," Cambridge, MA, MIT Press, 1979.
- [4] T.S.Huang and C.H.Lee, "Motion and structure from orthographic projections," *IEEE Trans. PAMI*, vol.11, pp.536-540, May 1989.
- [5] A.Delopoulos and Y.Xirouhakis, "Robust Estimation of Motion and Shape based on Orthographic Projections of Rigid Objects," *accepted in Tenth IMDSP Workshop (IEEE)*, Alpbach Austria, July 1989.
- [6] A.M.Tekalp, "Digital Video Processing," *Prentice Hall*, 1995.
- [7] R.Y.Tsai and T.S.Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Trans. PAMI*, vol.6, pp. 13-27, Jan. 1984.

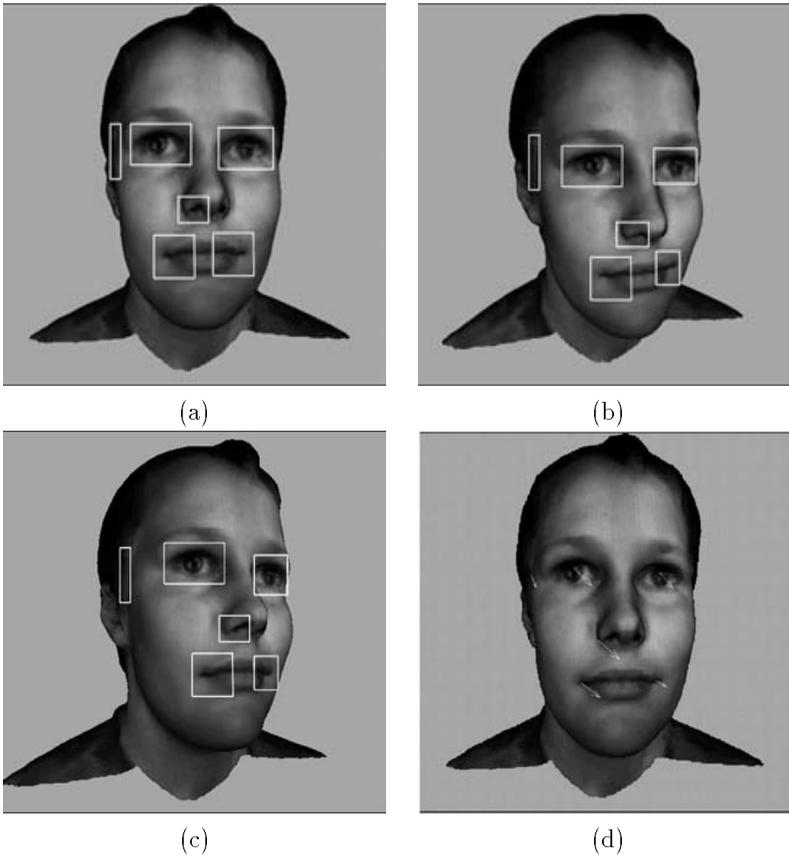
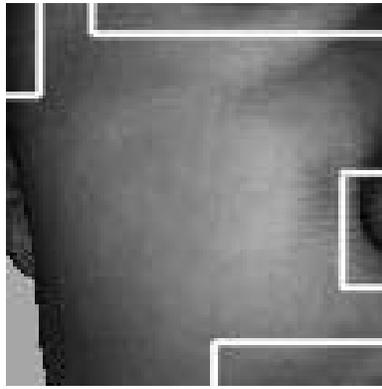


Figure 1: (a)Matched templates (frame 1), (b)Matched templates (frame 2), (c)Matched templates (frame 3), (d)Characteristic point correspondences

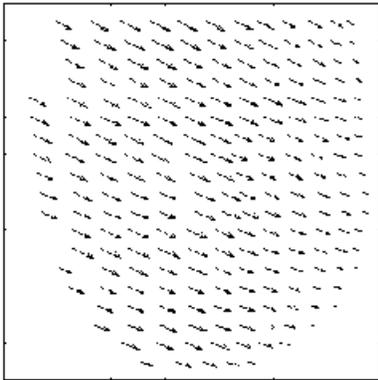
- [8] J.Weng, N.Ahuja and T.S.Huang, "Optimal Motion and Structure Estimation," *IEEE Trans. PAMI*, vol.15, no9, pp. 864-884, September 1993.
- [9] R.Chellappa, C.L.Wilson, and S.Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. of IEEE*, vol.83, no.5, pp.705-740, May 1995.
- [10] M.Michaelis, R.Herpers, L.Witta, and G.Sommer, "Hierarchical Filtering Scheme for the Detection of Facial Keypoints," *Proc. of ICASSP*, 1997.
- [11] M.Kampmann, "Estimation of the Chin and Cheek Contours for Precise Face Model Adaptation," *Proc. of ICIP*, 1997.



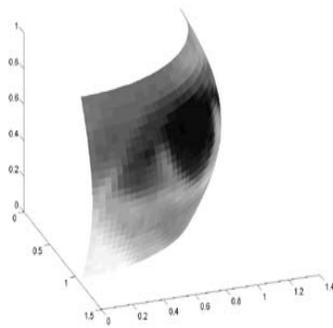
(a)



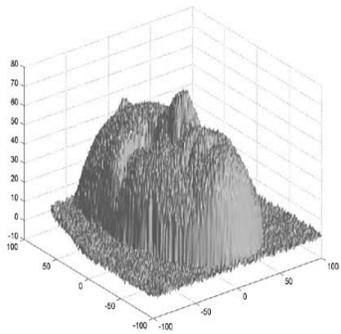
(b)



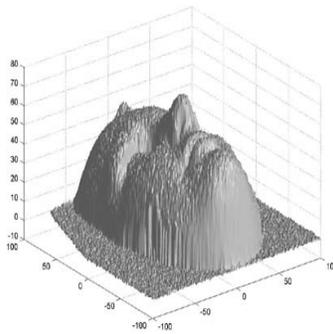
(c)



(d)



(e)



(f)

Figure 2: (a) Detected left chin line (frame 2), (b) Detected left cheek region (frame 2), (c) Indicative Motion Estimates, (d) 3D eye template, (e) 3D Reconstructed Human Face, (f) Reconstructed after 3D smoothing