A MODEL BASED ENERGY MINIMIZATION METHOD FOR 3D FACE RECONSTRUCTION

Davide Onofrio, Stefano Tubaro
{d.onofrio, tubaro}@elet.polimi.it

Dipartimento di Elettronica e Informazione - Politecnico di Milano – Milano - ITALY

ABSTRACT

In the paper we present a model based method for generating 3D face models from multiple images by means of an energy minimization algorithm. The energy function takes into account of: i) how well the luminance profiles are transferred (through the 3D model) from one image to the others; ii) how smooth are the reconstructed surfaces; iii) how it is congruent with an adapted face template mode (Candide model). It is important to notice that with the proposed method for each considered face two different 3D models are reconstructed: one with high resolution and one with low resolution obtained a reshaped Candide model. The first model can be used, for example, in 3D face analysis/recognition systems, while the second can be more useful for searching a specific face model in a large database.

1. INTRODUCTION

Face 3D reconstruction is an extensively studied topic of research mainly because the wide range of application ranging from recognition, identification to animation and video coding (see MPEG4-SNHC) and so on. Depending on the kind of application several 3D reconstruction algorithms are available; the choice depends on the accuracy with which it is necessary to build the 3D model. For model-based video coding simple models are normally preferable especially if low bit-rate communication channels are considered. For recognition and identification applications accurate models can guarantee higher performance than simpler ones, although simpler models are preferable to perform fast comparison between subjects. One possible strategy for the use of human-face 3D models for recognition applications is the creation, for each person under analysis, of two models: the first very accurate, but "heavy" from the data point of view; the second, simpler, defined by a limited number of parameters. In this paper we present an approach suitable to create, from a triplet of images, both these two models.

Traditional 3D reconstruction methods are based on inference of the 3D structure of an object or a scene from its 2D projections. Matching or establishing correspondences between point locations in images acquired from multiple views is the key problem in multi-view as well as in stereo image analysis. We used for the computation of correspondences an optical flow based approach that uses the brightness constancy assumption to find a transformation that maps corresponding points from one image to the others [1]. In order to acquire views of the 3D scene we used a trinocular calibrated camera set [2], the known camera configuration can provide a powerful epipolar geometry constraint for matching. As described in [3, 4] the brightness constraint can be used in energy minimization methods. In our case the energy is composed of two terms one that accounts for the correlation between areas of the images (the brightness constraint), and the other one for the smoothness of the reconstructed 3D surface. Establishing a balance between the two terms is a significant problem: if the correlation term exceeds the smooth term a very sparse surface is obtained, vice versa a flat surface is obtained, very far from having a "face structure". To overcome this difficulty we added an energy term to the total energy to be minimized: this last term takes into account for a template model of the object we want to reconstruct. In particular we adopted a simple face model: the Candide face model. The total energy expression can be minimized in an iterative fashion with approximated but fast methods (as for example belief propagation methods [4]), further, at every iteration the template model can be adjusted in order to adapt better to the clouds of available 3D points. Proceeding in this way, even if the balance between smooth term and the correlation term is not perfect, at the end of the reconstruction process we obtain a surface that has a face shape, moreover we obtain the parameter set that adapt in the best way the Candide model to the estimated 3D cloud of points representing the imaged face.

0-7803-9332-5/05/$20.00 ©2005 IEEE

2. A SIMPLE ENERGY MINIMIZATION METHOD FOR FACE RECONSTRUCTION

The experimental set-up is composed of a trinocular calibrated camera system [2] set as shown in Fig.1. The triplets of images are acquired synchronously. In Fig.2 it is illustrated how the proposed system works.

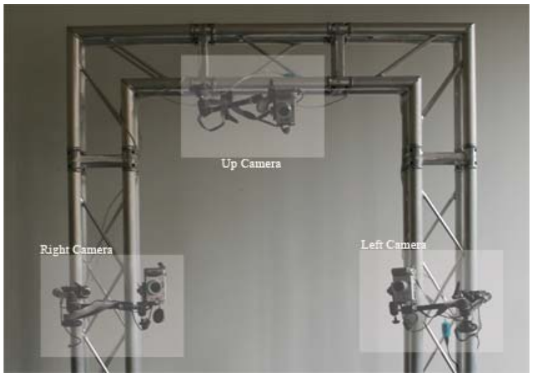

Figure 1. Experimental Set Up

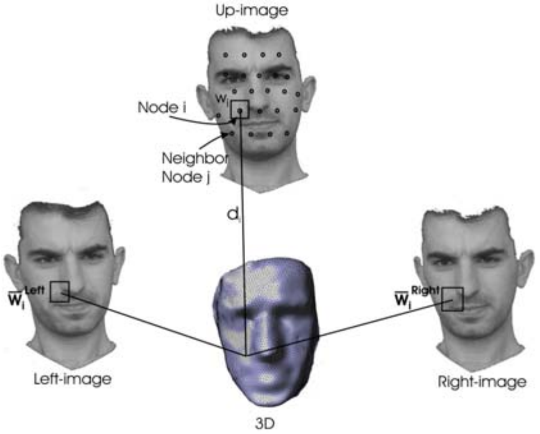

Figure 2. How the proposed system works.

In the UP-image we select an ensemble of points whose associated depths are computed. The depth image is modeled as a Markov Random Field (MRF), the energy associated to each point of the depth field is composed by two terms: one account for brightness constancy among correspondent patches, while the other (clique potential) describes how each node interacts with its neighborhoods. We find the depth map as the one that minimizes the following energy expression:

$$E(\{d\}) = \sum_i \min_{n\in\{Left,Right\}} \left\{ \iint_{W_i} (I^{UP}(w) - I^n(w + \widetilde{w}^n(d_i)))^2 dw \right\} + \lambda \cdot \sum_{(ij)} \psi_j \cdot \left| d_i - d_j \right|^2 \quad (1)$$

Where I^{Up} and I^n are respectively the intensity of the Up-image (taken as a reference) and the intensity on one of the other two images of the triplet, d_i is the depth associated with the point i considered in the Up-image, d_j is the depth associated with the neighbor of the point i, w is the patch window that surrounds the point i; $\widetilde{w}$ is the patch that surrounds the corresponding point of i in one of the other two images, ψ_j is a coefficient that depends on the luminance gradient in the Up-image, λ is a balancing term. A simulated annealing procedure could be applied in order to estimate the depth configuration that minimizes the above energy expression. A sub-optimal algorithm based on belief propagation is preferred to ensure that the final results are obtained in a fast way [4].

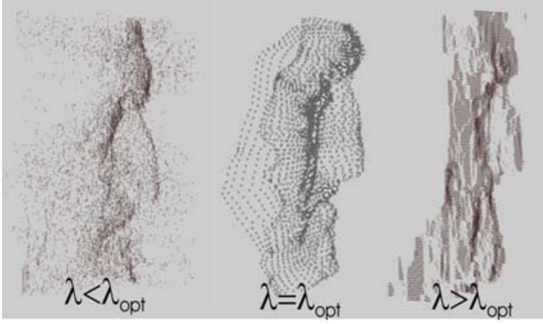

Figure 3. Examples of face profiles obtained with different λ values

As already stated, the choice of λ is critical: if it exceeds an optimal value the reconstructed surface is flattened otherwise it is scattered, an example is shown in the Fig.3. One method to avoid this is to add an additional constraint that takes into account for a template model of the object under reconstruction.

3. THE FACE TEMPLATE MODEL: CANDIDE

The original Candide face model, described in the report by Rydfalk [5], contained 75 vertices and 100 triangles; the model is widely used, since its simplicity makes it a good tool for image analysis tasks. Candide is a parameterized face mask defined as a wire-frame model (Fig.4) with a texture mapped onto its surfaces. The vertex coordinates can be seen as a 3N-dimensional vector $\widetilde{g}$ (where N is the number of vertices) containing the (x, y, z) coordinates of the vertices. The model is composed of Shape Units (SUs) that define a deformation of a standard face towards a specific face, and Animation Units (AUs) that define deformation of the specific face. The SUs thus describe the static shape of a face, they are invariant over time, but specific to each individual, while the AUs describe the dynamic shape.

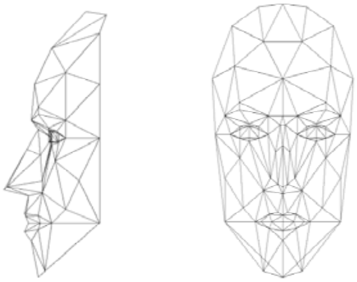

Figure 4. The Candide model

Candide can be controlled with the following expression:

$$g(\sigma,\alpha)=\tilde{g}+S\sigma+A\alpha \tag{2}$$

where the resulting vector g contains the new (x, y, z) vertex coordinates. The columns of S and A are the Shape and Animation Units respectively, and thus the vectors σ and α contain the shape and animation parameters. Since we also want to perform global motion, we need a few more parameters for rotation, scaling, and translation. Thus, we replace Eq.(2) with

$$g(\sigma,\alpha)=RD(\tilde{g}+S\sigma+A\alpha)+t \tag{3}$$

where $R = R(r_x, r_y, r_z)$ is a rotation matrix, $D = D(d_x, d_y, d_z)$ is the scale in the three dimensions, and $t = t(t_x, t_y t_z)$ the translation vector. The geometry of our model is thus parameterized by the parameter vector:

$$p=[\nu,\sigma,\alpha]=\lfloor r_x,r_y,r_z,d_x,d_y,d_z,t_x,t_y,t_z,\sigma,\alpha\rfloor$$

where ν is the vector of global motion parameters.

For our purposes only the shape parameters σ, are considered. With the 12 SUs of the model is possible to reshape Candide to at least the most common head shapes. The included SUs are listed in Table 1.

	SHAPE UNIT
0	Head height
1	Eyebrows, vertical position
2	Eyes, vertical position
3	Eyes, width
4	Eyes, height
5	Eye separation distance
6	Cheeks z
7	Nose z-extension
8	Nose vertical position
9	Nose, pointing up
10	Mouth vertical position
11	Mouth width

Table 1. Shape Units of the Candide model

4. AN IMPROVED ALGORITHM FOR 3D FACE RECONSTRUCTION

To account for the Candide model into our reconstruction algorithm, we add the following term to the total energy expression in Eq.(1):

$$\gamma\cdot\sum_i dist\,(d_i,g(p))$$

That is the sum of the (Euclidean) distances between points of the reconstructed surface and the Candide model. The energy expression becomes therefore:

$$E(\{d\})=\sum_i \min_{n\in\{Left,Right\}}\left\{\iint_{W_i}(I^{UP}(w)-I^n(w+\tilde{w}^n(d_i)))^2dw\right\}+ \tag{4}$$
$$\lambda\cdot\sum_{(ij)}\psi_j\cdot\left|d_i-d_j\right|^2+\gamma\cdot\sum_i dist(d_i,g(p))$$

The reconstruction process is summarized in Fig.5.

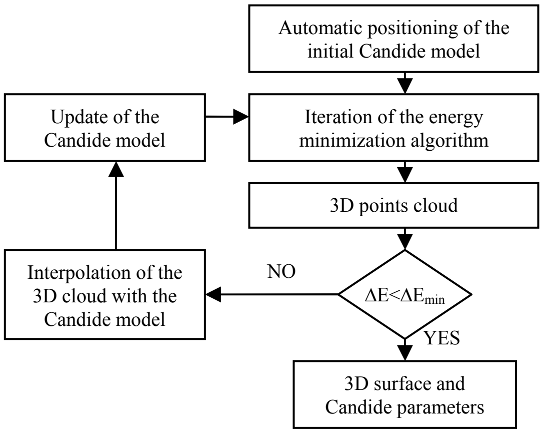
Automatic positioning of the initial Candide model
Update of the Candide model
Iteration of the energy minimization algorithm
3D points cloud
Interpolation of the 3D cloud with the Candide model
NO
ΔE<ΔE_min
YES
3D surface and Candide parameters

Figure 5. Diagram of the reconstruction process

In the initialization phase, face features (eyes, nose, mouth) are detected on the available calibrated images [6] and the initial Candide model, represented by the vector parameter $\boldsymbol{p}^{(0)}$, $g(\boldsymbol{p}^{(0)})$, is automatically positioned and adapted. Then iterations of the energy minimization algorithm are performed: at iteration k we obtain a cloud of 3D points $s^{(k)}(\{d_i\})$ that approximates the real face. This is used to update the model $g(\boldsymbol{p}^{(k)})$, searching for the new $\boldsymbol{p}^{(k)}$ that minimizes:

$$p^{(k)}=\arg\min_p\left[dist(s^{(k)}(\{d_i\}),g(p))\right] \tag{5}$$

As the iterations proceed we obtain better Candide approximations of the real face and we can progressively increase the value of γ in Eq.(4), ascribing more importance to the adherence of the estimated 3D points to the actual version of the Candide model. The iterations stop when the energy reaches a minimum E_{min}. As s a final result we obtain a 3D point cloud and the vector of parameters $\boldsymbol{p}$, for the best Candide model.

5. RESULTS

We have tested the method described with several triplets of face images. The triplets of images used for reconstruction are taken with three Canon cameras G3, the resolution is about 1000x1000 pxls in the face image area. Fig.6 shows some results of the proposed algorithm.

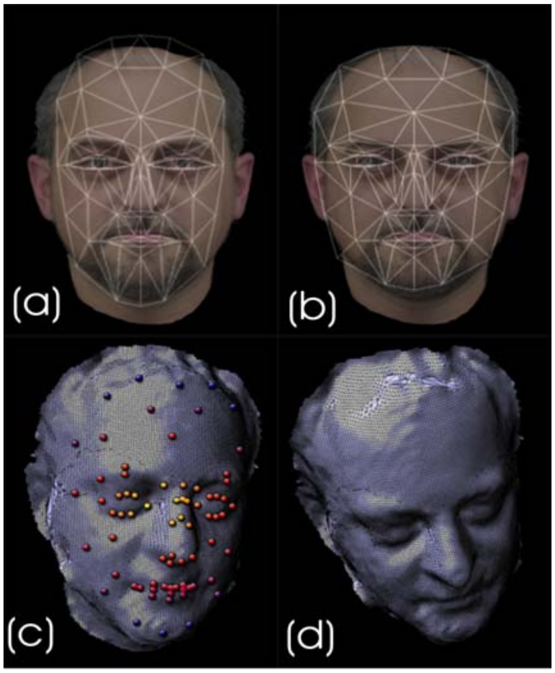

Figure 6. (a) A view of the initial Candide model is shown superimposed to the up-image; (b) (c) last iteration: the Candide model is adapted to the face; (d) the obtained 3D face model.

To minimize the energy term we used belief propagation algorithm as in [4], the addition of the energy term that deals with the Candide model increases the computation time of a factor 1.5 with respect to the original algorithm: 12 seconds instead of 8 seconds on Pentium IV 1,4 GHz. Results highlight that the reconstruction outliers are reduced thanks to the Candide driven energy term as shown in Table 2, furthermore this term permits to reach convergence of the process in less iterations.

	Time (s)	N° Iterations	N° Outliers(*)
With Candide	12.52	3	2%
Without Candide	8.35	5	12%

(*) the percentage of outliers is with respect to the total number of points reconstructed, are considered outliers those points with estimated depth 1% far off from the right depth.

Table 2. Results presented with respect to the original algorithm as in [4].

6. CONCLUSIONS

The presented algorithm generates as output a cloud of 3D points (and its corresponding wireframe) and a template model both representing 3D models of the face imaged in the considered images. The proposed approach is based on the minimization of an energy function composed by three terms. The first two are those normally used in this type of application while the third is related to how the estimated 3D points agree with an adapted Candide face model. Simulation results point out that this reduces the risk of obtaining too scattered or smoothed surfaces.

The algorithm obtains a 3D point cloud representing the face surface and a vector of parameters for the interpolating template face model, which can be used as a first, very fast, comparison term in recognition systems [7]. As a future work we want to finalize the algorithm to acquire data for the 3D recognition system presented in conjunction with Universitat Politecnica de Catalunya.

7. ACKNOWLEDGMENTS

The authors wish to acknowledge the support provided by the European Network of Excellence VISNET.

8. REFERENCES

[1] F. Pedersini, P. Pigazzini, A. Sarti, S. Tubaro, “3D Area Matching with Arbitrary Multiview Geometry,” *EURASIP Signal Processing: Image Communication - Special Issue on 3D Video Technology*, Elsevier, vol. 14, N. 1-2, pp.71-94, October 1998.

[2] F. Pedersini, A. Sarti, S. Tubaro, “Multicamera Systems: Calibration and Applications,” *IEEE Signal Processing Magazine, Special Issue on Stereo and 3D Imaging*, vol. 16, N. 3, pp. 55-65, May 1999.

[3] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.

[4] D. Onofrio, A. Sarti, S. Tubaro, “Area Matching Based On Belief Propagation With Applications To Face Modeling,”, *ICIP04*, Singapore 2004.

[5] M. Rydfalk, “CANDIDE, a parameterized face”, *Report No. LiTH-ISY-I-866, Dept. of Electrical Engineering*, Linköping University, Sweden, 1987.

[6] P. Viola, M. Jones, “Robust real-time Object Detection”, *II International Workshop on Statistical and Computational Theories of Vision-Modeling, Computing and Sampling*,Vancouver, Canada, 2001.

[7] A-Nasser Ansari, Mohamed Abdel-Mottaleb, “*3-D Face Modeling Using Two Views and a Generic Face Model with Application to 3-D Face Recognition*”, Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS’03), Miami, Florida, USA, 2003.