3D FACE RECONSTRUCTION WITH A FOUR CAMERA ACQUISITION SYSTEM

Davide Onofrio, Stefano Tubaro**   *Antonio Rama[+], Francesc Tarres[+]*
{d.onofrio, tubaro}@elet.polimi.it   *{alrama, tarres}@gps.tsc.upc.es*

*Dipartimento di Elettronica e Informazione - Politecnico di Milano
[+]Department Teoria del Senyal i Comunicacions de la Universitat Politècnica de Catalunya

ABSTRACT

In this paper we present a method for 3D face reconstruction based on the use of a four camera acquisition system. Our method determines correspondences between surface patches on different views through a modeling of depth maps based on Markov Random Fields (MRFs). In order to reduce the occurrence of outliers the MRF-based modeling is bound to satisfy the epipolar constraint. We apply the belief propagation algorithm to the MRF model in order to perform a maximum-a-posteriori estimation of such correspondences. The use of a four camera system improves the accuracy of the obtained 3D model, in fact the occlusions are remarkably reduced, and moreover the extension of the reconstructed face is wider than in more traditional two or three camera systems, this is particularly useful in some 3D-based face recognition systems that work with only a limited regions of a face.

1. INTRODUCTION

The techniques for the creation of 3D models of human faces starting from one or more images have been largely considered by several research groups due to the wide range of possible applications: from person identification/recognition, virtual actor creation and animation, model based 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 models, although simpler models are preferable to perform fast comparison between subjects.

Recently different algorithms have been proposed for face recognition by means of only part of the face [1], nevertheless to verify the performance in terms of recognition rate of these algorithms a complete 3D face model should be used, models that can be generated with a binocular or trinocular systems only with some difficulties, for this reason we have decided to implement a 3D face reconstruction algorithm with four cameras.

In this paper we present an approach suitable to create, from four images these models.

Traditional 3D reconstruction methods are used to generate 3D models of objects or scenes from a set of 2D projections: matching or establishing correspondence between point locations in different images is the key problem.

We used for the computation of correspondences optical flow based approach that uses the brightness constancy assumption to find a transformation that maps corresponding points in multiple images into one another [2]. In order to acquire views of the 3D scene we used a four calibrated camera set [3], the known camera configuration can provide a powerful epipolar geometry constraint for matching. As described in [4, 5] 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, that is the brightness constraint, and the other for smoothness of the reconstructed 3D surface. One drawback is the difficulty of establishing a balance between the two terms: if the correlation term exceeds the smooth term a very sparse surface is obtained, vice versa a flat surface is obtained, very far from the right face structure.

The energy expression can be minimized in an iterative fashion with approximated but fast methods (as for example belief propagation methods [5]).

2. EXPERIMENTAL SET-UP

The experimental set-up is composed of a four calibrated camera system set as shown in Fig.1; the calibration procedure is a natural extension of [3]. The four images are acquired synchronously, with four Canon cameras G3, the resolution is about 1000x1000 pixels in the face region.

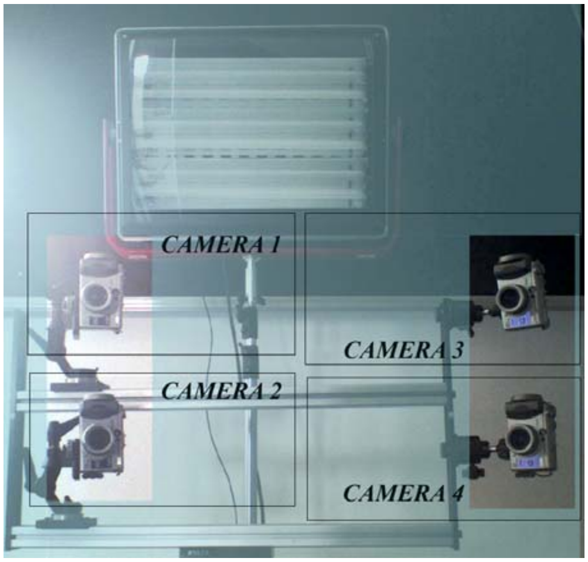

Figure 1. Experimental Set Up

In Fig.2 a schema of the method for 3D reconstruction is shown: the cameras are grouped in two pairs, each pair produces a depth map of the region of the face seen. Because the cameras are globally calibrated the two clouds of points generated don't need registration, indeed they are referred to the same world's reference system.

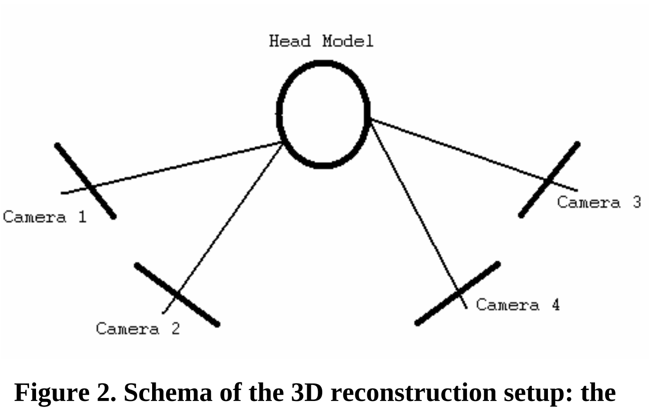

Figure 2. Schema of the 3D reconstruction setup: the four cameras are coupled in two pairs 1-2, 3-4.

3. FOUR CAMERAS 3D RECONSTRUCTION

As already stated in the previous section the four cameras are grouped in pairs: each pair generates a depth map of the face region seen, the depth image is modeled as a Markov random field (MRF) [5]. In order to accomplish the task of finding the two depth configurations we propose to calculate disparity fields by minimizing an energy expression, for each pair, involving a fidelity term and a smoothing term:

$$E_{(1,2)} = E_{data(1,2)} + E_{smooth(1,2)} \qquad (1)$$

$$E_{(3,4)} = E_{data(3,4)} + E_{smooth(3,4)} \qquad (2)$$

The two expressions can be processed separately to find the depth configuration that minimize the associated energy term.

Find the depth configuration, and not the disparity map, implies that the epipolar constraint is automatically satisfied.

The term $E_{data(x,y)}$ is very important: in theory it can generate, alone, a well reconstructed 3D surface, in practice some approximations are needed and the term $E_{smooth(x,y)}$ is necessary to avoid local minima as result of the minimization energy process.

We adopted for $E_{data(x,y)}$ the expression:

$$E_{data(x,y)}(\{d\}) = \sum_i \left\{ \iint_{W_i} (I^x(w) - I^y(w + \tilde{w}^y(d_i)))^2 \, dw \right\} \qquad (3)$$

Where I^x and I^y are respectively the intensity of the image acquired with camera x and that of the image acquired with camera y, d_i is the depth associated with the point *i* considered in the x-image, *w* is a patch window that surround the point *i*; $\tilde{w}$ is the patch that surround the corresponding point of *i* in the y image, the sum is over all the pixel in the face region.

As a first approximation we adopted the same size for the two corresponding patch *w* and $\tilde{w}$. In a second iteration of the algorithm, once we have an estimation of the reconstructed surface, we can use a second order approximation: at every point of the surface the tangent plane is calculated and an homography between *w* and $\tilde{w}$ established, this leads to a better correspondence between pixels in the different images and gives us a more precise surface reconstruction (Fig.3).

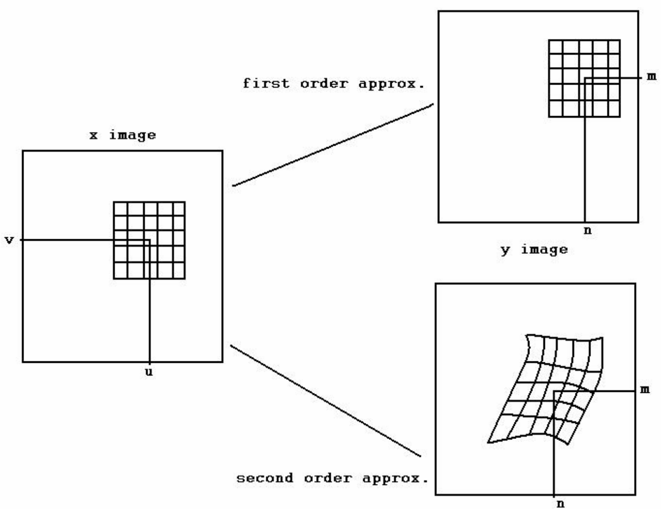

Figure 3. To a small rectangular region in the x image corresponds in the y image a rectangle of the same size if we do a first approximation, or a distorted rectangle at second order

The $E_{smooth(x,y)}$ has the expression:

$$E_{smooth\ (x,y)}(\{d\}) = \lambda \cdot \sum_{(ij)} \psi_j \cdot \left| d_i - d_j \right|^2 \qquad (4)$$

Where ψ_j is a coefficient that depends on the luminance gradient in the x image (taken as a reference), λ is a balancing term, d_j is the depth associated with the neighbor of the point *i*

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 fast results.[5]

At the end of the reconstruction procedure we have verified the first approximation adopted on assuming corresponding patches of the same shape and size: about the 30% of the pixels of the x image have been incorrectly matched to pixels in corresponding patch of the y image, this has been corrected in the second iteration where tangent plane approximations have been used.

The two cloud of points, one for each pair of cameras, generated at the end of this step overlap in proximity of the frontal region of the face, in that region the approximation of same size same shape for the patch is quite invalid because the tangent plane at the surface to be reconstructed is not parallel to the image plane of the cameras. Many outliers are generated in the frontal region and therefore a step for outlier reduction must be executed.

4. OUTLIERS REDUCTION

We find local outliers by iteratively fitting quadric patches around every data point.

We use all the points in a spherical neighborhood of $P_0=(x_0, y_0, z_0)$ and estimate the orientation of the local tangent plane. Given this orientation, we define a reference frame whose origin is P_0 itself and whose z axis is perpendicular to the plane; we fit a quadric of the form [6]:

$$z = quad(x, y) = ax^2 + bxy + cy^2 + dx + ey + f \qquad (5)$$

by minimizing a least square criterion

$$\delta = \sum_i w_i (z_i - quad(x_i, y_i))^2 \qquad (6)$$

where the (x_i, y_i, z_i) 1<i<n are the n neighbors of P_0 and the w_i are associated weights.

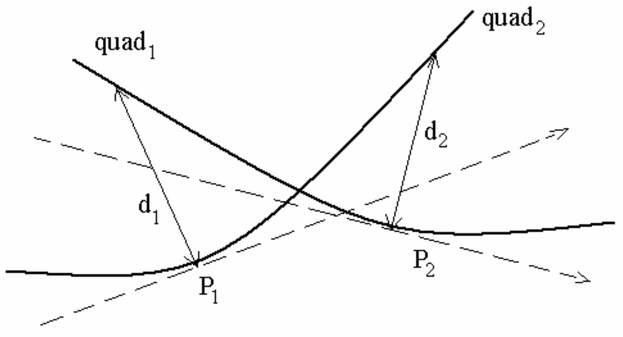

Figure 4. The distance between two points is taken to be the maximum value of the distance of one point to the local surface corresponding to the other, represented by the arrows. The dotted lines represent the axes of the reference frames in which the computations are performed.

To deal with outliers, we define a metric d_{quad} that measures whether or not two points appear to belong to the same surface. We take d_{quad} to be:

$$d_{quad}(P_1, P_2, quad_1, quad_2) = \max(dist_1, dist_2) \qquad (7)$$

Where

$$dist_1 = abs(z_1 - quad_2(x_1, y_1)) \qquad (8)$$

Expressed in the reference frame of $quad_2$

$$dist_2 = abs(z_2 - quad_1(x_2, y_2)) \qquad (9)$$

Expressed in the reference frame of $quad_1$ (Fig.4).

When d_{quad} is zero the two points belong to the same local surface, d_{quad} increases when their respective local surfaces become inconsistent. This distance, hence, can be used to discount outliers; moreover in the fitting quadrics the weighting factor w_i corresponding to the outlier can be reduced

5. RESULTS

Several face model were reconstructed with the proposed set-up, in Fig.5 results of the algorithm are shown.

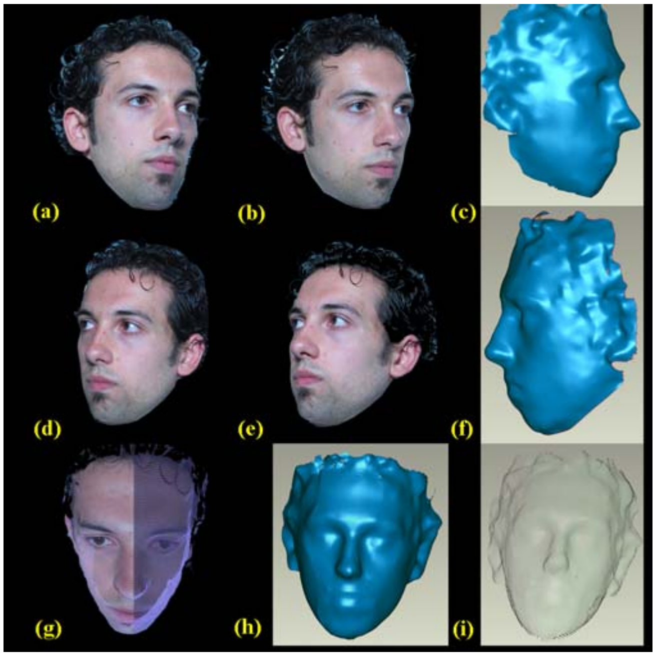

Figure 4. (a) (b) The views of the camera 3 and 4; (c) the reconstructed surface from camera 3,4; (d) (e) the views of camera 1 and 2; (f) the reconstructed region from camera 1,2; (g) (h) (i) the global reconstruction: the outliers elimination provides good results in the central region of the face.

We have compared these results and those obtained from a trinocular algorithm [7] we have implemented in the past: in the frontal region of the face, incline to the presence of outliers, there are no significant deviations for the two reconstructed surfaces.

The time required a complete 3D face reconstruction on a Pentium IV 3 GHz is 32 seconds.

6. CONCLUSIONS

In this paper we have presented an algorithm for a face 3D reconstruction based on energy minimization by means of a four camera system.

The four cameras are coupled in two pairs each generating a 3D point cloud representing the face surface seen by the pair. There is an overlapping region, the frontal region that can produce outliers, for this reason a step for outlier elimination has been adopted.

With this algorithm we intend to generate a 3D face database, the obtained model are wider than that obtained with traditional binocular or trinocular system, and can be used to assess performance of face recognition system that exploits recognition over partial regions of a face like Partial Principal Component Analysis [1].

7. ACKNOWLEDGEMENTS

This work has been done within the VISNET European network of excellence project.

7. REFERENCES

[1] T. Rama, F. Tarres, D. Onofrio, S. Tubaro, “Using Partial Information for Face Recognition and pose Estimation”, *ICME05*, Amsterdam 2005

[2] 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.

[3] 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.

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

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

[6] P. Fua, “A parallel stereo algorithm that produces dense depth maps and preserves image features”, *Machine Vision and Applications, 1991.*

[7] D. Onofrio, S.Tubaro, T. Rama, F. Tarres, “Model Based 3D Face Reconstruction by Means of an Energy Minimization Method”, *Icme05*, Amsterdam 2005.