3D FACE RECONSTRUCTION FROM UNCALIBRATED IMAGE SETS

Alexandros Moskofidis[*], *Nikos Nikolaidis**

{amoskofi,nikolaid}@aiia.csd.auth.gr

[*]Department of Informatics, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

ABSTRACT

In this paper, we present a framework for reconstructing the 3D model of a human face from a set of uncalibrated facial images in different poses. The resulting 3D model of the face can be used in applications like face recognition and face verification. Our approach comprises two steps: in the first step we utilize a 3D reconstruction method to calculate the 3D feature coordinates of some salient feature points of the face, marked manually on the input images, whereas in the second step we use a mass springs finite elements method (FEM) to deform a generic face model, based on the cloud of points produced from the first step. We further enhance the resulting 3D model by projecting it into the input images and manually refining its node coordinates.

1. INTRODUCTION

The task of reconstructing an object in 3D space from its images (projections) is one of the most demanding in computer vision. In the past years the biggest attention was given to the calibrated reconstruction case (i.e. the case where the position of the camera relative to the object and the camera intrinsic parameters are known beforehand) whereas nowadays researchers try to tackle the uncalibrated reconstruction problem, where the input images are taken with a camera at random position and orientation with respect to the human face.

It is well known [1] that utilizing the epipolar geometry one can yield depth estimates for an object just from two images of it. Unfortunately, the obtained coordinates do not lie on the Euclidean space [2], which makes this representation not very useful. In order to upgrade the representation, extra information is required. This extra information can be obtained either from the camera position or from the camera intrinsic parameters. The latter can be calculated either from the use of special calibration patterns or from the images of our input set. The procedure of utilizing the images that we have in order to calculate the camera intrinsic parameters is called self calibration as opposed to calibration where some specific calibration patterns are used in order to calculate the camera calibration matrix.

There are numerous approaches to the uncalibrated 3D reconstruction problem in literature, the more characteristic of which are the work of Faugeras [3], Beardsley et al [4], Hartley [5] and Pollefeys [2], who wrote an excellent tutorial on the subject.

Our approach utilizes the 3D reconstruction algorithm presented by Pollefeys in [2] in order to calculate the 3D coordinates of some salient feature points of the face based on a small number of facial images where feature points are manually marked. We have chosen to use this approach because of its flexibility, due to the fact that the input images can be taken with an off the self camera placed at random positions. The intrinsic camera parameters can be calculated from the input image set.

We further incorporate a generic face model (the Candide face model) and deform it, using a finite element method (FEM), based on the point cloud obtained from the first step. On top of that, to further improve our resulting 3D model, we reproject it back to the initial images and fine tune it manually using an interface that was developed especially for this purpose. The resulting face model can be used along with the corresponding texture in biometric applications such as face recognition and face verification.

The rest of this paper is organized as follows. Part 2 describes in brief the first part of the proposed methodology, which is the 3D reconstruction of a set of salient features of the human face. In section 3 we describe the incorporation and the deformation of a generic head model (Candide head model) whereas in part 4 we provide some experimental results. In part 5 future directions are described and conclusions follow in part 6.

2. 3D RECONSTRUCTION

As already mentioned, we have used the algorithm proposed by Pollefeys in [2] in order to calculate the 3D coordinates of some salient features of the face. We will briefly explain the steps of algorithm for the sake of completeness of this paper. Readers interested in obtaining additional information can consult [2].

For our camera we have adopted the ideal pinhole – perspective camera model [6] where no radial distortion is present. In such a camera, the projection of an object point on an image plane is described by the following equation

$$\boldsymbol{m} = \boldsymbol{PM} \quad (1)$$

where $\boldsymbol{m} = [x, y, 1]^T$ are the point coordinates on the image plane, $\boldsymbol{P}$ is the 3x4 projection matrix and $\boldsymbol{M} = [X, Y, Z, 1]^T$ are the object point coordinates in 3D space. Note that we use the homogenous coordinates where the '=' sign indicates an equality up to an non-zero scale factor.

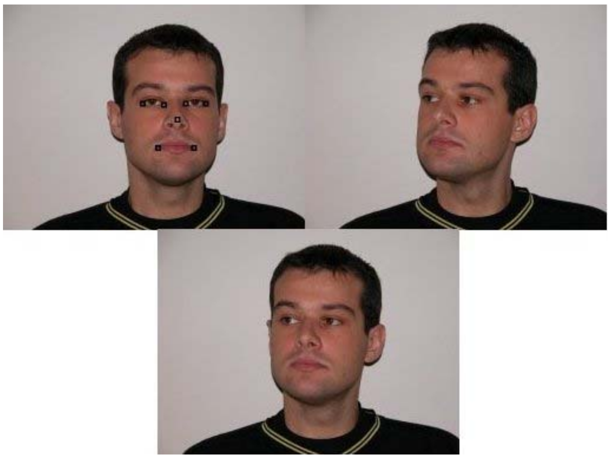

Figure 1 : The input images

At the first step we manually select some salient feature points of the face in the input images and define their correspondences (figure 1). The coordinates of these feature points over the input images constitute the input to the 3D reconstruction algorithm. It has to be noted that we have used some easily recognizable and distinct feature points of the face such as the corners of the eyes, the corners of the mouth and the tip of the nose. Unfortunately it is very difficult to define a big number of feature points on the human face due to its lack of texture and characteristic points that can be uniquely identified over a number of images.

What comes next is the calculation of the Fundamental Matrix [1]. The calculation of the Fundamental Matrix is based on the first two images of the set. Those two images must be selected efficiently so that they correspond to viewpoints that are as far apart as possible but in the same time have all the feature points visible on both of them. The overall performance of the algorithm relies heavily on the efficient selection of these first two frames.

After the calculation of the Fundamental Matrix it is possible to obtain a reference frame which will eventually help us get an initial estimate of the depth for the selected feature points. Unfortunately, this representation does not lie in the metric space and thus additional procedures should be followed in order to upgrade it to metric.

Next the rest of the images of the input set are incorporated in the algorithm and the projection matrices that describe the projection of the face in each image of the set are evaluated.

In the subsequent step the algorithm performs an optimization which is based on all the images of the input set and thus refines the representation. This is called Bundle Adjustment [7] and it is the most computationally intensive part of the algorithm.

Finally the algorithm uses a self calibration technique in order to calculate the camera intrinsic parameters. These parameters are subsequently used to upgrade the representation to the metric space and yield the final cloud of points.

3. GENERIC MODEL DEFORMATION

The next part of the proposed approach deals with the incorporation of a generic face model, namely the Candide face model, into the reconstruction procedure.

The Candide face model has been developed by the Linkoping University [8] and in its current version has 104 nodes, distributed all around the human face and 184 triangles that connect those nodes creating a wire frame. The nodes of the model correspond to characteristic points of the human face e.g. nose tip, outline of the eyes, mouth etc. The feature points selected on the facial images are described in the previous section and should correspond to Candide nodes. A procedure for defining the correspondences between the 3D reconstruction of the selected feature points and the Candide model nodes was followed.

3.1 FEM Deformation

A mass spring finite element method was employed to deform the generic Candide model. The deformation process incorporates a list of pivotal points (our 3D reconstructed points from the first part of the algorithm), the Candide model and a list which contains the correspondences between the pivotal points and the Candide nodes, and produces a deformed model.

The FEM deformation can be outlined as follows: at first the Candide model undergoes global rotation translation and scaling so that it is roughly aligned with the cloud of 3D points. In order to determine the scale factor the mean distances between the two corners of the eyes and the two corners of the mouth were evaluated both in the point cloud and the Candide model and their ratio was used as the scale factor. Then the model was translated so that the center of mass of the point cloud coincides with the center of mass of the corresponding model nodes.

Furthermore the Candide model has to be appropriately rotated. To achieve this, a triangle whose vertices are the outer tips of both eyes and the tip of the nose was defined. The same triangle was defined for the corresponding nodes of the Candide model and the model was rotated so that the outwards pointing normal vectors of the two triangles are aligned. The deformation process moves the corresponding nodes of the Candide model so that they coincide with the points of the cloud and deforms the rest of the nodes. As it is obvious from the latter, the pivotal points must spawn the entire face, otherwise the deformation process will produce poor results.

3.2 Manual Refinement

After the application of the deformation we obtain a model that fits the individual's face depicted in the input set of images. Unfortunately, due to limitations on the 3D reconstruction algorithm, the deformation process and to errors in the selection of the feature points coordinates, the output model may not be ideal, in the sense that some nodes may not have the correct position in 3D space. Therefore a manual refinement procedure is adopted.

According to this procedure, we reproject the deformed face model in every image of the input set and manually change the location of certain model nodes in the 2D domain. In order to return to the 3D domain from the manually refined projections, a triangulation process is used [6]. This was facilitated from the fact that the projection matrices for each frame were available from the 3D reconstruction algorithm.

In order to be able to use the triangulation method to estimate the 3D coordinates of a model's node we must specify manually the new positions of the nodes in two frames. By so doing we can yield new, improved coordinates, in the 3D space. When the manual adjustment of the selected nodes is finished the deformation process is applied once again but this time with an extended set of pivotal points – the initial cloud of points produced from the 3D reconstruction algorithm along with the additional 3D coordinates of the points that have been manually refined.

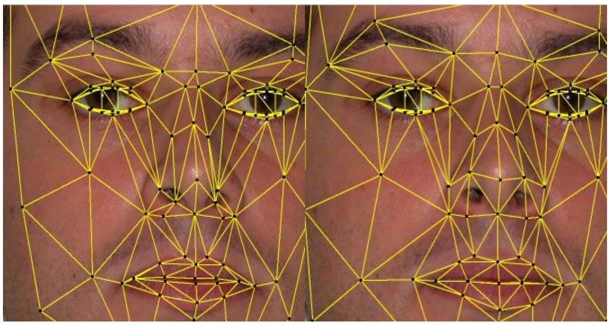

Figure 2 : The manual refinement procedure

The manual refinement procedure is presented in figure 2 which depicts the projection of the deformed model into an image of the input set prior and after the manual refinement. It is evident that with the manual refinement the generic model can fit more efficiently to the individual's face.

4. EXPERIMENTAL RESULTS

For our experiments we have used a minimal set of 3 images of human faces in different positions and we have selected and matched manually the feature points across those images.

Feature points were selected on the two corners of the mouth, the two corners of the eyes and on the tip of the nose, as shown on figure 1.

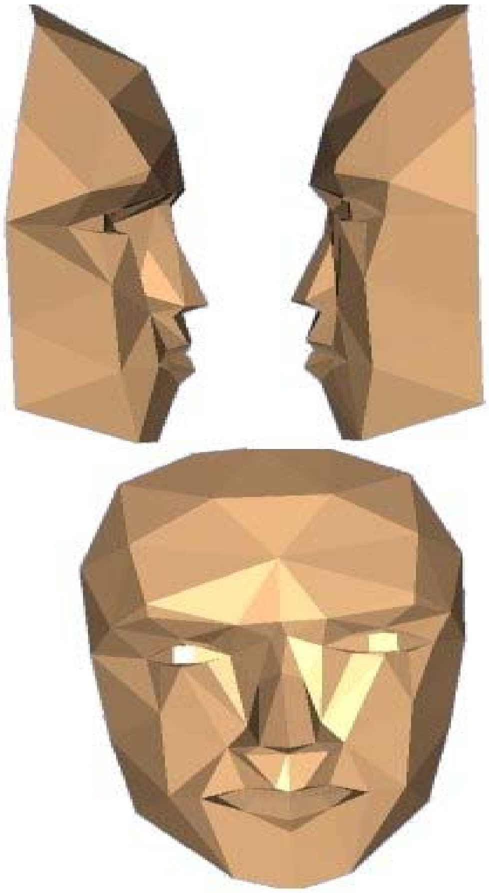

Figure 3 : The deformed Candide Model

The deformed Candide model derived from the facial images of figure 1, after applying limited manual refinement is presented in figure 3. Obviously the deformed model is not perfect which can be attributed to the errors in the feature selection process as well as to the limited resolution of the Candide model.

The performance of the 3D reconstruction algorithm was evaluated based on the reprojection error namely the Euclidean distance between the manually selected feature points and the projections of the derived 3D features. Results for the model presented in figure 3 can be seen in table 1.

Manually Selected Coordinates	Calculated Coordinates	Reprojection Error (pixels)
(1131,1151)	(1131,1151)	(0,0)
(1420,1164)	(1420,1164)	(0,0)
(1050,776)	(1051,775)	(-1,1)
(1221,786)	(1218,788)	(3,-2)
(1392,794)	(1395,795)	(-3,-1)
(1567,793)	(1566,792)	(1,1)
(1244,957)	(1244,957)	(0,0)

Table 1 : Reprojection error – image 3

One can observe that the reprojection error is very small and does not exceed 3 pixels for input images of dimensions 2560x1920. Similar results were obtained when the algorithm was applied to other image sets.

A number of provisions can be taken in order to make the algorithm more robust. The camera positions used for capturing the image should be sufficiently apart but at the same time care has to be taken in order to ensure that all feature points are visible in the first three images of the set. The most error prone part of the algorithm is the initial triangulation (depth estimation) where a small angle between the viewpoints used to acquire the two images can have a severe effect on the overall reconstruction.

Moreover through experimentation we have reached the conclusion that the quality of the results is mainly affected by the quality of the input features i.e. whether corresponding points selected on the images are indeed projections of the same points on the 3D space. Thus care should be taken in order to select these points as accurately as possible.

5. FUTURE WORK

Our work in the field of modeling a human face from a set of uncalibrated images is not complete yet. In the future we plan to incorporate some new techniques that will aim towards a more robust 3D reconstruction, namely a new method for bundle adjustment that besides the reprojection error will incorporate additional constraints derived from the geometry of the human face (e.g. the relative distances of the eyes, mouth etc).

Furthermore, we are experimenting with different techniques for the deformation of the generic Candide face model to be used as an alternative to the finite elements method.

6. CONCLUSIONS

In this paper we have presented a framework for the challenging task of reconstructing a face in three dimensions from a set of uncalibrated images.

In the first part of the proposed approach we used a 3D reconstruction algorithm proposed by Pollefeys [2] to calculate the 3D coordinates of some salient facial feature points manually selected on a set of images. At the second part of the algorithm, we use a generic face model in order to produce a more detailed representation of the object's face. This is substantially an interpolation process with all the advantages and disadvantages that this entails. In order to obtain a more detailed and accurate model of the individual's face an iterative manual refinement process is employed, which can improve the quality of the resulting face model. The experimental results prove that the proposed methodology can yield very satisfactory 3D reconstructions.

REFERENCES

1. Andrew Zisserman, Richard Hartley, "Multiple View Geometry in Computer Vision", Second Edition, Cambridge University Press 2004

2. Marc Pollefeys, "Tutorial on 3D Modelling from Figures", June 2000 (http://www.esat.kuleuven.ac.be/~pollefey/tutorial/)

3. Olivier Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig", ECCV '92, p. 563-578

4. Paul A. Beardsley, Andrew Zisserman, D.W. Murray, "Sequential Updating of Projective and Affine Structure from Motion", International Journal of Computer Vision, June/July 1997, p. 235-259

5. Richard Hartley, "Euclidean Reconstruction from Uncalibrated Views", Second joint European – US Workshop on Applications of Invariance in Computer Vision, 1993 p. 237-256

6. Emanuele Trucco, Alessandro Verri, "Introductory Techniques in 3-D Computer Vision" , Prentice Hall 1998

7. Bill Triggs, Philip McLauchlan, Richard Hartley, Andrew Fitzgibbon, "Bundle Adjustment – A Modern Synthesis", International Workshop on Vision Algorithms 1999 p.298-372

8. J. Ahlberg, CANDIDE-3 -- an updated parameterized face, Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.

The authors would like to thank George Moschos for his implementation of the Finite Elements Method library.