

3D Face Modeling Based on Structured-Light Assisted Stereo Sensor

Boulbaba Ben Amor, Mohsen Ardabilian,
and Liming Chen

LIRIS Lab,
Lyon Research Center for Images and Intelligent Information Systems,
UMR 5205 CNRS, Centrale Lyon, France
{Boulbaba.Ben-amor, Mohsen.Ardabilian,
Liming.Chen}@ec-lyon.fr
<http://liris.cnrs.fr/>

Abstract. In this paper we present a 3D human face reconstruction framework based on stereo sensor coupled with a structured lighting source. Starting from two calibrated images, the active part (video projector) which project controlled lights, allows the operator to locate two sets of structured features with sub-pixel accuracy in both left and right images. Then, exploiting epipolar geometry improves the matching process by reducing its complexity from a bidirectional to a unidirectional search problem. Finally, we perform an adapted dynamic programming algorithm to obtain corresponding features in each conjugated scanline separately. Final three dimensional face models are achieved by a pipeline of four steps: (a) stereo triangulation, (b) data interpolation based on cubic spline models, (c) Delaunay triangulation-based meshing, and (d) texture mapping process.

1 Introduction

Over recent years, face detection [1], analysis, measurement and description have been applied widely in several applications [2] such as recognition, surveillance, multimedia document description, etc. Most face recognition technologies have two challenges: first, significant changes in lighting conditions cause system performance degradation. The alternative problem is their sensitivity to pose variations, indeed existing software compares probe image to gallery images taken at various angles. In our work, we propose to capture the 3D facial geometry in order to enhance recognition process. Having 3D images of faces we can generate a synthetic view at any angle and perform recognition algorithms.

In this paper we present a face reconstruction method based on a hybrid structured light and stereo sensor technique. This report is organized as follows: in section 2, we describe some existing methods for geometry recovering based on optical methods. Section 3 presents an overview of our system. A comprehensive description is given in sections 4, 5, 6, and 7. Finally, we present some experimental results and some future work in sections 8 and 9.

2 Related Work

In recent years, 3D human face acquisition technologies have made great progress. The successful solution of this problem has immense potential for applications in many domains. In this section we briefly review some approaches to recovering 3D facial geometry. Four potential optical methods are proposed: laser scanning, coded light range digitizers, silhouette-based methods, and multi-image/motion based approaches. Commercial 3D digitizers such as Cyberware [3] and Minolta [4] non-contact scanners are based on laser triangulation; Laser rays coming out of light source hit the object surface and are captured by a camera in a different angle using a rotating mirror. These devices take a short time to capture highly accurate reconstructions. However, they are expensive and the data are usually noisy requiring manual editing. In the case of multi-image based methods various approaches are developed. The classical ones use stereo sensors to acquire simultaneously two [5] or a set of images [6, 7], 3D information is found by triangulation; the intersection of optical rays going from projection centers joining corresponding features in images. The greatest challenge of these approaches is the matching process and the accuracy of the reconstructed models depends on precision of the matching. In the second category of these methods, the data source is a video sequence of face and Structure From Motion (SFM) algorithm is the most used approach. In [8] authors use SFM algorithms which are enhanced with a generic model as an initial solution. Here the obtained shape represents an approximation of the real face. Another solution is given by a structured light-based approach in which some kind of special lighting is directed to the object to be scanned. These active approaches help to solve the correspondence problem, which is a difficult task in passive methods. Depth information will be extracted from pattern deformation such as techniques presented in [9] and [10] but these devices are restrictive, so need to regroup a set of partial models. Other solutions have been developed such as silhouette extraction based methods [11], photometric methods [12] and face from orthogonal views methods [13] which produce an approximation of a real face.

3 Proposed Approach: Overview

Our approach is based on a binocular sensor, the minimal system for performing optical triangulation, coupled with a video projector which helps to resolve the matching process with a sub-pixel precision. Having a pair of points, each of them in a different image, which are projections of the same 3D point, this spatial point can be reconstructed using the cameras' light-rays intersection. But, the real challenge is how to establish corresponding pairs of points. Figure 1 illustrates the overview of our system: in order to find camera parameters and the relationship between them, we calibrate the binocular sensor. After the calibration process, the correspondence problem, initially a two-directional search problem becomes a one-dimensional search problem, by applying image rectification to the pair of images. In the image acquisition process we take two images from each camera corresponding to normal and inverse light pattern projections. Consequently, the stripe intersections improve the edge detection process with a sub-pixel precision. Then an adapted dynamic programming algorithm is applied to perform matching feature sequences in left and right epipolar lines separately.

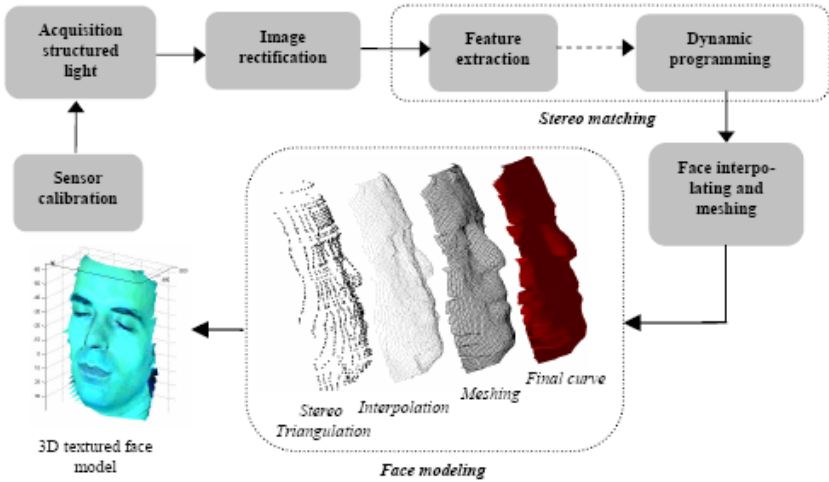


Fig. 1. Pipeline of 3d face modeling proposed approach

The final stage is to find a triangulation result for each pair of matching points. In order to obtain the 3D face model, we proceed by filtering, interpolating, meshing points and finally mapping texture onto a shape of face.

4 Calibration Step

This section briefly describes camera model geometry, stereo sensor geometry and the calibration processes which include computation of both intrinsic and extrinsic parameters. The main idea in camera calibration theory is to find a relationship between the 3D points of the scene and their 2D projecting point in the plane image. Here, the spatial point passes by different transformations in the camera acquisition process. Otherwise, calibration process computes changes from the initial 3D point P_w to the 2D image point P_i . First, the space point P_w is projected on an image plane in P_u by perspective projection. Second, the lens distortion changes the point position from P_u to P_d . Finally, P_d coordinates must be expressed in the image coordinate system to obtain P_i . The camera calibration technique used in our approach is described by Zhang in [14] where a planner pattern is used. Author computes a *closed-form solution* for initialisation, followed by a nonlinear optimization technique based on the maximum likelihood criterion which includes lens distortion.

After calibrations of each camera separately, we must compute the rigid transformation $A=(R,T)$ between them. Having this calibration data, it makes it possible to compute epipolar geometry and perform a rectification process in order to reduce the complexity of the correspondence problem. Indeed, given a feature point m in the left image, the corresponding feature point must lie on the corresponding epipolar line. In a standard stereo setup, conjugated epipolar lines are collinear and parallel to one of the image axis, usually the horizontal one. If the pair of images is taken with a general stereo configuration, an operation known as rectification can be applied to bring the two retinal planes to be coplanar to a common plane in space. The standard rectifica-

tion consists of selecting a plane parallel with the baseline ($O_L O_R$). The two images are then re-projected onto this plane. The new images satisfy the standard stereo setup. After image rectification the epipolar lines become collinear in the rectified cameras and the correspondence finding is limited to a conjugate scanline.

5 Feature Extraction and Matching

The key step in the stereo approaches is the matching process which consists of finding the corresponding points between the left and right images. In our approach we propose to match a set of features which are detected by projecting negative and positive patterns of light on a subject. The projection of this kind of structured light helps to discriminate some feature points with sub-pixel precision and therefore increases precision of the matching process. The section below details the method used for stripe edge localization.

5.1 Stripe Edge Localization

The classical method for edge detection consists of detecting edges by simple binarization; i.e. with *pixel accuracy*. However, it is desirable to determine the stripe edge position with *subpixel accuracy*. This can be done by finding zero-crossings of the second derivative of the original image in the direction orthogonal to the stripes. Another method consists of locating intersections after projecting successively normal and inverse patterns.

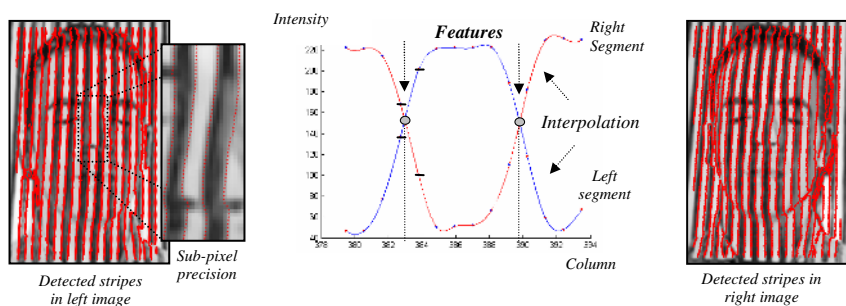


Fig. 2. Stripe boundary localization and results (detected features stereo images)

In figure 2 it has been shown how stripe edge can be detected by locating intersections of interpolated segments AB and EF . It shows also the results of stripe boundary detection in both left and right images.

5.2 Stereo Matching Based on Adapted Dynamic Programming Algorithm

Dynamic programming is a very useful technique for sequence matching and alignments. It solves an N -stage decision process as N single-stage processes. It reduces

the computational complexity to the logarithm of the combinatorial problem. In our approach we use this mathematical method to match left and right features. When images are rectified dynamic programming allows us to find the optimal solution for each scanline separately. The monotonic ordering constraint allows the global cost function to be determined as the minimum cost path through a disparity space image. The cost of the optimal path is the sum of the costs of the partial paths obtained recursively (1). Occlusions are modeled by assigning a group of pixels in one image to a single pixel in the other image and penalizing the solution by an occlusion cost occ . The $score(q_i, e_i)$ is a normalized correlation measurement between features q_i and e_i .

$$\sigma(\Phi_{j,i}^*) = \begin{cases} 0, & \text{if } j=0 \text{ and } i=0; \\ \max \left\{ \begin{array}{l} \sigma(\Phi_{j-1,i-1}^*) + score(q_j, e_i) \\ \sigma(\Phi_{j-1,i}^*) + occ \\ \sigma(\Phi_{j,i-1}^*) + occ \end{array} \right\}, & \\ \text{otherwise} & \end{cases} \tag{1}$$

We define cost function as a matrix where lines and columns are indexed by left and right features for each scanline. However the principal disadvantage of this method is the possibility that local error may be propagated along a scanline, corrupting other potentially good matches. In this stage of our work, we simply filter the final result and false matches are cancelled in order to solve this kind of problem.

6 Face Modeling

To obtain a 3D face model we firstly triangulate matched points by finding intersecting points in space of obtained optical rays. Secondly, we mesh points after interpolation based on cubic spline models. Finally we map texture onto the obtained shape in order to add realism. Detailed descriptions are given in the following paragraphs:

6.1 Stereo Triangulation

Once having obtained the correspondence pairs, the 3D face model, or to be more precise its depth, can be recovered. The depth is (approximately) inversely proportional to the disparity (2), which symbolizes the shift of the corresponding points between the two images. Consequently, assuming the rectification, the disparity between a matched pair of points from the left image I_L and the right image I_R , respectively $I_L(u, r)$ and $I_R(v, r)$ is defined by $d(u,r) = u - v$. Now, we concentrate on the recovery of the position of a single point, P , from its projections, p and q . The classical relationship between depth Z and disparity d is given by formula (2):

$$\frac{b + d(u,r)}{Z - f} = \frac{b}{Z} \tag{2}$$

The 3D coordinates of a point $P(X, Y, Z)$, which correspond to the matched pair $I_L(u, r)$ and $I_R(v, r)$, can be expressed as follows (3):

$$X = \frac{b \cdot u}{d(u,r)}; \quad Y = \frac{b \cdot r}{d(u,r)}; \quad Z = \frac{b \cdot f}{d(u,r)} \tag{3}$$

6.2 Model Interpolation and Meshing

Knowing a set of 3D points we interpolate in order to ameliorate the model's resolution and draw the face curve. Many interpolating data methods exist such as, linear, polynomial, Lagrange, Hermit, spline, etc. In our approach we use cubic spline functions [15] which are a very popular model for interpolation. The interpolating function is made up of a sequence of cubic polynomials across each interval of the data set curves that meet at the given data points with continuous first and second derivatives. Cubic spline interpolation is significantly better than some other interpolation methods for relatively smooth data such as faces.

Once having a 3D interpolated data set ($S=\{p_1, p_2, \dots, p_n\}$), we triangulate in order to obtain the meshing curve of the face. For that, we use the Delaunay triangulation and Voronoi diagram duality, approach amongst the most useful data structures of computational geometry. The main idea of this algorithm is based on the Voronoi diagram which partitions the plane into convex regions, one per point or site. Given the Voronoi diagram of a set of sites the Delaunay triangulation of those sites can be obtained as follows: Given a set S of n distinct points in R^2 Voronoi diagram is the partition of R^2 into n polyhedral regions $Vo(p)$, $p \in S$. Each region $Vo(p)$, called the *Voronoi cell* of p , is defined as the set of points in R^2 which are closer to p than to any other points in S , or more precisely, $Vo(p) = \{x \in R^2 / \text{dist}(x, p) \leq \text{dist}(x, q) \forall q \in S - p\}$, Where dist is the Euclidean distance. The convex hull $\text{conv}(nb(S, v))$ of the nearest neighbor set of a Voronoi vertex V is called the Delaunay cell of V . The Delaunay complex (or triangulation) of S is a partition of the convex hull $\text{conv}(S)$ into the Delaunay cells of Voronoi vertices together with their faces. This Delaunay triangulation applied to the obtained space points gives a 3D face model and the results are shown in figure 3.

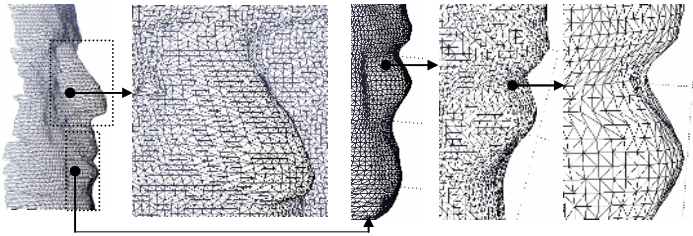


Fig. 3. Result of Delaunay triangulation process

6.3 Texture Mapping

In computer graphics, texture mapping refers to the technique where an image is pasted onto a three-dimensional surface. This technique can significantly increase the realism of a scene without increasing the complexity.

The final stage in our face reconstruction is to map texture onto the 3D shape in order to add realism to the face model. Currently, this step is performed by simply warping texture image to a 3D surface grid using interpolation. Figure 4 illustrates this process: first, we construct the texture image from the pair of left and right images then the warping procedure allows us to apply a texture onto the 3D shape with

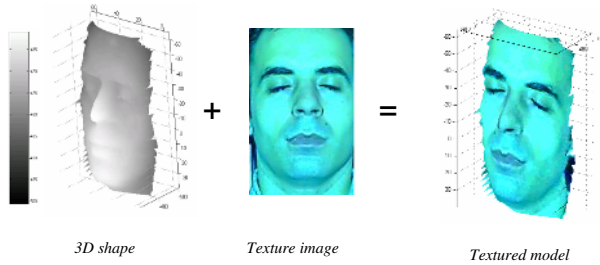


Fig. 4. Texture mapping stage

specific parameters. This method is not the best one, so in our future work we will ameliorate image texture and provide some corresponding feature points from 3D shape and image texture.

7 Experimental Results

Figure 4 shows one textured model result. Our results are obtained from two images having 640×480 pixels where faces occupy $1/6$ of the total surface and 16 stripes are projected onto the face subject. The presented model has 8081 vertices and 63948 triangles. However, these results can be enhanced by increasing the number of stripes and increasing photos' resolutions in order to cover more details of the face. Some parameters are adjustable such as *baseline*: the distance between the set of cameras and the distance from the subject to the sensor. At this moment, we are concentrating on computing the best set of parameters for reconstruction.

8 Conclusion and Future Work

We have presented in this paper a complete low-cost solution for 3D face acquisition using a stereo-structured-light coupled technique. The sensor is first calibrated and parameters are extracted especially baseline and focal length. Second, epipolar geometry is also computed in order to reduce the complexity of the correspondence search problem. Then, the projection of normal and inverse structured light provides a set of point pairs with sub-pixel precision. The global matching optimization is performed by a dynamic programming algorithm for each scanline independently. Finally, depth is obtained by light-ray intersections. The final face geometry is achieved by performing interpolation and meshing techniques. The next step in our work is to introduce, in addition to the intra-scanline solution, the inter-scanline consistency constraint in order to enhance the matching process and reduce false matching.

References

1. D. Tsishkou, L. Chen and E. Bovbel, Semi-Automatic Face Segmentation for Face Detection in Video, Proc. of MEDIANET'04, Tozeur, Tunisia, pp. 107-118.
2. Kun Peng, Liming Chen and Su Ruan, Face Image Encryption and Reconstruction for Smart Cards Using Fourier-Mellin Transform, Proc. of MEDIANET'04, Tunisia, pp 311-317.

3. Home page: www.cyberware.com
4. Home page: www.minolta-3d.com
5. E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*, ISBN 0-13-261108-2 Prentice Hall 1998.
6. D'Apuzzo N., Modeling human faces with multi-image photogrammetry. In: Corner, B.D., Pargas, R., Nurre, J.H. (Eds.), *Three-Dimensional Image Capture and Applications V*, Proc. of SPIE, Vol. 4661, San Jose, USA, 2002, pp. 191-197.
7. S. Lao, M. Kawade, Y. Sumi, F. Tomita: Building 3D Facial Models and Detecting Face Pose in 3D Space. *3DIM 1999*: 390-404.
8. A.R. Chowdhury, R. Chellappa, S. Krishnamurthy, and T. Vu, 3D Face Reconstruction from Video Using a Generic Model, *International Conference on Multimedia*, Switzerland, pp. I:449-452, 2002.
9. E. Garcia, J.-L. Dugelay, H. Delingette, *Low Cost 3D Face Acquisition and Modeling*, ITCC, Las Vegas, Nevada, April 2001.
10. L. Zhang, B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *Int. Symposium on 3D Data Processing Visualization and Transmission*, Padova, Italy, June 2002.
11. Yang Liu, George Chen, Nelson Max, Christian Hofsetz, Peter McGuinness. *Visual Hull Rendering with Multi-view Stereo*. *Journal of WSCG*. Feb. 2004.
12. Chia-Yen Chen, Reinhard Klette and Chi-Fa Chen, *3D Reconstruction Using Shape from Photometric Stereo and Contours*, October, 2003
13. Horace H S Ip and L.J. Yin, *Constructing a 3D Head Model from Two Orthogonal Views*, *The Visual Computer*, Vol 12, No. 5, pp. 254-266, 1996.
14. Z. Zhang, A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330-1334, 2000.
15. de Boor, C., *A Practical Guide to Splines*, Springer-Verlag, 1978.