Efficient and low-cost 2.5D and 3D face photography for recognition

Boulbaba Ben Amor, Mohsen Ardabilian, 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

Abstract

In this paper, we propose a complete 2.5D and 3D human face acquisition framework based on a stereo sensor coupled with a structured lighting source. We aim to develop an accurate and low-cost solution dedicated to the 3D model-based face recognition techniques (FRT). In our approach, we first calibrate the stereo sensor in order to extract its optical characteristics and geometrical parameters (the offline phase). Second, epipolar geometry coupled with a projection of special structured light on a face (the online capture phase), improves the resolution of the stereo matching problem, by transforming it into a one-dimensional search problem and a sub-pixel features matching. Next, we apply our adapted and optimized dynamic programming algorithm to pairs of features which are already located in each scanline. Finally, 3D information is found by computing the intersection of optical rays coming from the pair of matched features. The final face model is produced by a pipeline of four steps: (a) Spline-based interpolation, (b) Partial models' alignment then integration, (c) Mesh generation, and (d) Texture mapping. Furthermore, this paper presents an approach which evaluates the reconstruction techniques. We consider a scan from a laser scanner as "ground truth", then we compute spatial deviation between it and the homologue reconstructed model, based on the well-known Iterative Closest Point matching algorithm.[1]

1. Introduction and motivation

Over the past few decades, biometrics and particularly face detection, analysis, measurement and description have been applied widely to several applications such as recognition, video surveillance, access control, production of personal documents such as passport and national identity card, etc. As described in [1], most commercial face recognition technologies (FRT) suffer from two kinds of problems. The first one concerns inter-class similarity, for instance twins' classes, fathers and sons' classes. The second problem is the intra-class variations caused by significant changes in lighting conditions, pose variations (i.e. three-dimensional head orientation), and facial expressions.

Current state-of-the-art in face recognition is rich in developed works which aim to resolve problems regarding this challenge. One paradigm to recognition is the 3D model-based techniques [2, 3, 4] in which the researchers exploit, in addition to textural end silhouette information, the three-dimensional shape of the face in order to mitigate some limitations. In general, the 3D faces of interest are saved in a library during an offline phase. During the online recognition phase, a single captured model of the face, present in the scene, is matched with the model library in order to find the identity and the pose of the person. This task presents an active area of research and focuses many biometric applications; however the true handicap is the non-availability of an efficient and, at the same time, a low-cost solution to acquire the face models. It is, moreover, the main goal of this paper which proposes a novel efficient and low-cost 3D acquisition solution dedicated to person recognition and authentication systems.

The remainder of the paper is organized as follows: Section (2) presents the state-of-the-art in 3D acquisition techniques, particularly optical ones. Section (3) describes an overview of the proposed approach. In Section (4), we focus on the stereo matching strategy including structured light-based features localization and their matching. In Sections (5) and (6) we emphasize our pipeline of 2.5D and 3D face modeling. Section (7) presents a procedure which allows the measurement of the accuracy of our

[1] This study has been implemented on the Face-Checker platform at LIRIS Lab., Centrale Lyon.

reconstruction results. Finally, a conclusion and future work are presented in Section (8).

2. Related work on 3D face photography

Three-dimensional acquisition techniques have been applied in many research areas and industrial applications including robot vision, medical imaging, archeology, reverse engineering, and industrial quality control. Scanning human faces is notably one of the most important targets. A successful solution has immense potential applications for face animation in video games, face recognition and authentication, building statues of peoples, etc. In this section we briefly review some optical approaches focusing on recovering the three-dimensional object geometry particularly human faces. In this category of methods, four potential classes are proposed: laser scanning, coded light range digitizers, silhouette-based methods, and multi-image/motion based approaches.

The techniques in which special lights are used in order to extract dimensional information, called active methods, consist of two categories: First, 3D non-contact scanners such as Minolta [6] which are based on laser triangulation. Here, laser rays coming out of a light source hit the object and are captured by a camera in different angles using a rotating mirror. These devices take a short time to capture highly accurate reconstructions. However, they are expensive and their outputs are usually noisy requiring manual editing. In addition, it suffers from the surface's reflection which generates small unreal variations in the surface of the scanned object. The second solution is a structured-light based approach in which special light is directed onto the object to be scanned, such as the techniques presented in [13] and [14]. This process helps to solve the correspondence problem, which is a difficult task in passive methods. In the case of projecting one pattern of light on the measuring scene, depth information will be extracted by analyzing the pattern deformations. In the other case where a set of light patterns is projected, the extraction of the codeword assigned to each pixel in the images allows the formation of the range image. The major limitation of the active techniques is the devices' restriction, so need to acquire and register a set of partial models in order to build the entire models.

In the second category of approaches, known as passive methods, photogrammetry is the most active research area in which various algorithms are proposed. In the classical multi-images based sensors which acquire simultaneously two [7, 10] or a set of images [8, 9], 3D information which is ambiguous in only one optical ray, can be found by triangulation (i.e. the intersection of multiple optical rays going from projection centers of the cameras and passing through corresponding features in the images). The hardest problem regarding these approaches is the matching problem and the accuracy of the reconstructed models depends on the precision of this process. In the second sub-category of passive methods, the data source is a video sequence and the *Structure From Motion* algorithm is the most used approach to estimate image disparity. For instance, the authors in [11] and [12] use SFM algorithms which are enhanced with a generic model as an initial solution. In their papers, they present as results an approximate face model.

There are also other solutions in literature such as silhouette extraction-based methods [15], photometric methods [16] and face from orthogonal views methods [17] which produces an approximation of a real object. A combination of some methods can be significant such as work presented in [18], in which the authors combine a shape from a silhouette-based technique which provides a coarse 3D initial model with a multi-stereo carving technique. This technique is used only for reconstructing static objects.

3. Overview of the proposed approach

The proposed approach uses one binocular sensor assisted by a structured lighting source. This helps to resolve the matching process with sub-pixel precision then finds 3D dimensional points by optical triangulation. In this paper we briefly describe the basic approach used for range image formation which is already presented in [5], and we emphasize additional steps which focus on producing the entire face model and the accuracy evaluation method and results.

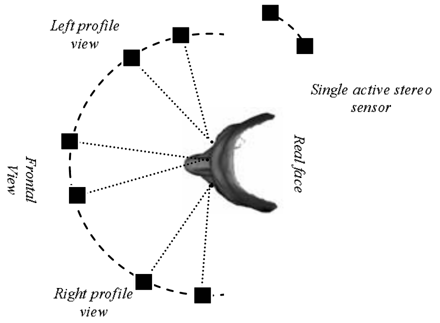


Figure1. Principal of view changing for complete 3D face acquisition

Figure 1 illustrates the basic idea of changing the sensor's viewpoint (or head orientation) in order to acquire different parts of the face to be scanned. After acquisition of the range images by applying the photography process three times in respectively, frontal view, left profile view and right profile view, we aim to register the obtained partial models using the Iterative Closet Point algorithm (as a fine alignment) assisted by a manual initialization (as a coarse alignment). The rigid transformation which is the result of the previous stage is applied to partial models in order to blend them considering the frontal model as the reference view. Finally, a step of texture mapping onto the reconstructed shape achieves the acquisition process and adds the true appearance to the face.

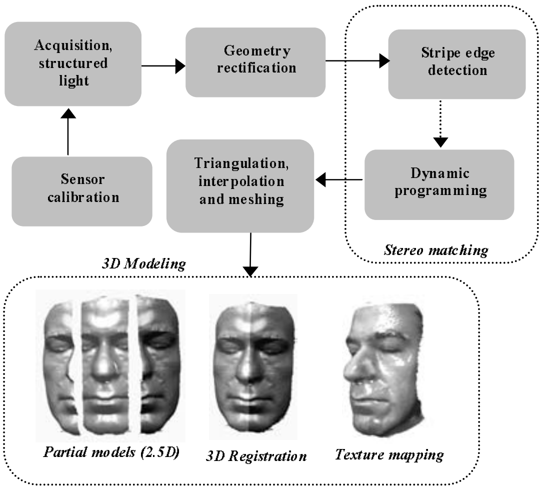


Figure2. Pipeline of our approach for 3D face reconstruction

Figure 2 describes the basic and the additional processes in our framework of 2.5D then 3D human face photography. The sections below describe major stages and contributions in this work; in other words the Spline-based face modeling and the novel step to perform sub-pixel features localization and matching. In the end, the paper addresses the depth accuracy measuring problem and a novel technique is proposed where the "ground truth" data are acquired by a 3D laser-based scanner.

4. Stereo matching strategy

The key step in the stereo-based approaches is the matching problem which consists of finding the correspondence features between left and right images. In our basic approach, after sensor calibration, we rectify images by applying epipolar geometry transformation. This operation reduces the complexity of the correspondence problem from a bi-dimensional to a one-dimensional search problem. In fact, matched points have necessarily the same Y-coordinate, in rectified images. But, the problem is not yet totally resolved and we must sweep for each feature point, in the left image, the conjugated scanline in the right image in order to find the corresponding feature. To achieve this, we propose to match only a set of features which are discriminated by projecting, successively, negative and positive patterns of light on the face. This process allows us to localize some feature points with sub-pixel precision, and so improves precision of the matching and the 3D point's localization stages, see figure 3.

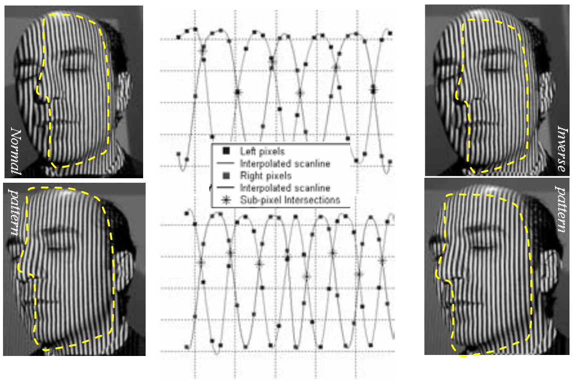


Figure3. Acquisition and sub-pixel stripe edge detection processes

The second step in our matching strategy is to join the sets of features which are detected in conjugated rectified scanlines. We are choosing to perform the principal of dynamic programming which is a very useful optimization technique for sequence matching and alignment. It aims to solve an N-stage decision process as N single-stage processes.

The DP-based matching approach allows us to find the optimal solution for each pair of conjugated scanlines separately by computing similarity matrix between features. The monotonic ordering constraint lets the global cost function be determined as the minimum cost path through a disparity space image. Here, the cost of the optimal path is the sum of the costs of the partial paths obtained recursively (1). We define cost function as a matrix where lines and columns are indexed by left and right features for each scanline (figure 4 (a)), as done by Ohta and Kanade, in [20], on natural edges.

$$\sigma(\Phi^*_{j,i}) = \begin{cases} 0, \text{ if } \ j=0 \ \text{ and } \ i=0; \\ \max\begin{Bmatrix} \sigma(\Phi^*_{j-1,i-1}) + \text{score}(q_j, e_i) \\ \sigma(\Phi^*_{j-1,i}) + \text{occ} \\ \sigma(\Phi^*_{j,i-1}) + \text{occ} \end{Bmatrix}, \\ . \quad \text{otherwise} \end{cases} \qquad (1)$$

Occlusions, which are the main complications in the matching problem, are modeled by assigning a group of pixels in one image to a single pixel in a second image and penalizing the solution by an occlusion cost *occ.* The term, *score(q$_i$,e$_i$)*, is a correlation measurement between features q_i and e_i, respectively in left and right images (2).

$$\mathrm{score}(q_i, e_i) = \frac{\sum\sum(N(e_i) - \bar{e}_i)((N(q_i) - \bar{q}_i)}{\sqrt{(\sum\sum(N(e_i) - \bar{e}_i)^2)(\sum\sum((N(q_i) - \bar{q}_i)^2)}} \quad (2)$$

The optimized DP algorithm calculates only the diagonal terms of the similarity matrix.

5. 2.5D face model reconstruction

For any viewpoint and after feature matching process, we triangulate by finding intersection points, in space, of the optical rays each one coming from the camera's projection center. Once this is done, we mesh the obtained points in order to connect them and to build a 3D surface. Before that, we interpolate between the triangulated points by applying the cubic spline models in order to increase the points' resolution.

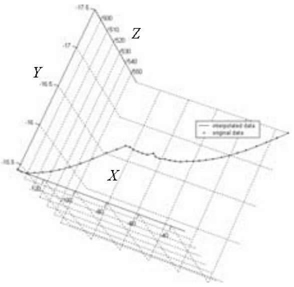


Figure5. Optical triangulation and cubic-spline based interpolation for one reconstructed scanline

Figure 5 represents reconstructed features in pair of conjugated scanlines and the interpolated points. This improves, as illustrated, the model's resolution and draws the partial face model.

In this field, many data interpolation methods exist in literature such as, Linear, Polynomial, Lagrange, Hermit, Spline, etc. In our approach, we use cubic spline functions [19] which are very popular interpolation models. These functions are 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 (3). In our case, cubic spline interpolation is significantly better than others for relatively smooth data such as faces, as presented in figure 6.

$$S_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i)^1 + d_i \quad (3)$$
$$\text{for} \quad i = 1,\ldots,n-1$$

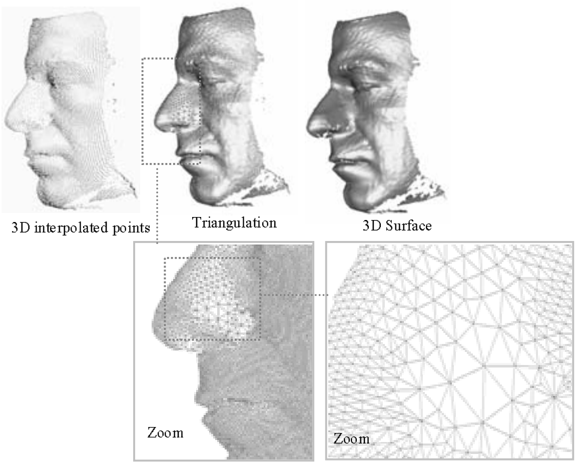


Figure6. Interpolation and mesh processes

Having the 3D interpolated point set *(S={p$_1$, p$_2$,..., p$_n$})*, we generate one mesh between them in order to obtain a coherent shape. For that, we consider the Delaunay triangulation/Voronoi diagram duality based 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 sites can be obtained as follows: given a set *S* of *n* distinct points in R^2, the Voronoi diagram is the partition of R^2 into *n* polyhedral regions *Vo(p), p* ∈ *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:

$\mathrm{Vo}(p) = \{x \in R^2 / \mathrm{dist}(x, p) \leq \mathrm{dist}(x, q) \forall q \in S - p\}$,where dist is the Euclidean distance function.

The convex hull *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 *conv(S)* into the Delaunay cells of Voronoi vertices together with their faces.

This mesh generation step based on Delaunay triangulation, applied to the projection in (x, y) plan of the space points, gives a 2.5D face model as shown in figure 6. It remains to align reconstructed partial models, merge them, and then map texture in order to achieve the reconstruction process. The following sections describe these additional stapes for entire face modeling.

6. 3D face modeling from 2.5D models

In order to perform face recognition from any viewpoint, complete 3D model representations are indispensable in the model gallery. In this section we focus, consequently on building an entire three-dimensional face model from partial scans. First, three scans, from respectively three viewpoints, are provided by the basic approach. Second, a complete 3D model is built by aligning and integrating these obtained partial models. The first task is to register these 2.5D models in order to align and transform them for the merging step. Our registration algorithm is based on the well-known Iterative Closet Point (ICP) algorithm [21] which is an iterative procedure minimizing the mean square error (MSE) between points in one view and the closest points, respectively, in the other. At each iteration of the algorithm, the geometric transformation that best aligns the two partial models is calculated. Intuitively, starting from the two sets of points $P = \{p_i\}$, as a reference data, and $X = \{y_i\}$, as a test data, the goal is to find the rigid transformation (R,t) which minimizes the distance between X and P. ICP consists of determining for each point p_i of the reference set P the nearest point in the second set X within the meaning of the Euclidean distance. The rigid transformation (R,T) minimizing a least square criterion (4) is calculated and applied to the each point of P.

$$e(R,t) = \frac{1}{N}\sum_{i=1}^{N}\|(Rp_i + t) - y_i\|^2 \quad (4)$$

This procedure is alternated and iterated until convergence (i.e. stability of the minimal error). Indeed, total transformation (R,t) is updated in an incremental way as follows: for each iteration k of the algorithm, $R=R_kR$ and $t=t+t_k$, The criterion to be minimized in the iteration k becomes (5):

$$e(R_k, t_k) = \frac{1}{N}\sum_{i=1}^{N}\|(R_k(Rp_i + t) + t_k - y_i\|^2 \quad (5)$$

The ICP algorithm presented above always converges monotonically to a local minimum [21]. But, we can hope for a convergence to global minimum if initialization is good. For this reason, we perform a coarse registration procedure, before the fine one. The coarse alignment consists of finding correspondences between distinctive features that may be present in the overlapping area. The goal of this initialization is to find a set of approximate registration transformations. Figure 7 illustrates different stages of the registration process. For the three 2.5D scans we perform, two by two, the registration procedure (the frontal scan is considered as the reference data). We present also the statistical errors in each registration step. Firstly, the registration of left profile scan and frontal scan outputs as mean error equal to 0.49 mm (0.3%) and standard deviation equal to 0.31 mm. Then, the registration of right scan and the frontal scan presents 0.43 mm (0.27%) as mean error and 0.28 mm as standard deviation. The color maps of error deviations presented in figure7 show the spatial distances. Once the alignment is done, a mesh integrating procedure is performed in order to build a unique mesh of the 3D face model. We can show, in figure 7, that the final model presents a coherent 3D shape.

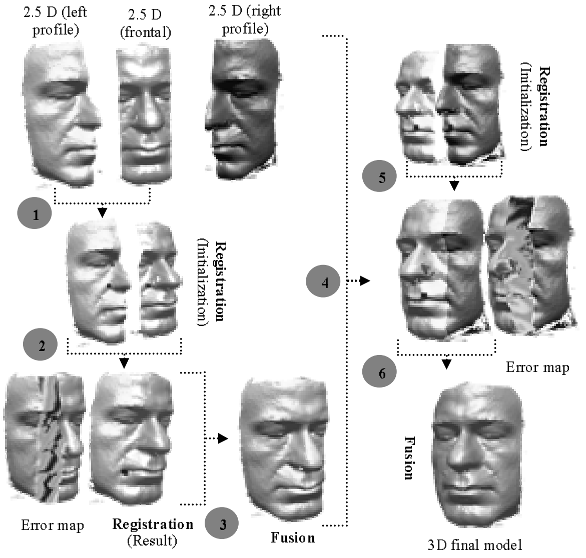
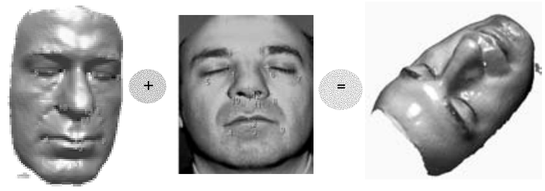


Figure7. Registration and fusion processes



3D Shape of face *Texture of face* *Textures face (VRML Model)*

Figure8. Texture mapping into 3D face shape

Having a 3D face shape, it remains to give realism by mapping the appearance image which is acquired at the same time with the frontal range image. Indeed, 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. This procedure is performed by warping texture image to a 3D surface grid using interpolation. In order to guarantee an excellent warping, we manually select some feature points in the texture image and corresponding feature points in the 3D face shape to give pasted points. Figure 8 illustrates this process:

7. Evaluation and Accuracy measurement

The main object of our approach is to develop a 2.5D and 3D face photography technique which is, at the same time, accurate and solves the high-cost problem of the 3D acquisition techniques. In this section, we detail the sources of error in our 3D face measurement technique. Then, we present some evaluation results obtained by computation of a global spatial deviation between reference data (from a scanner) and measured data (from our sensor).

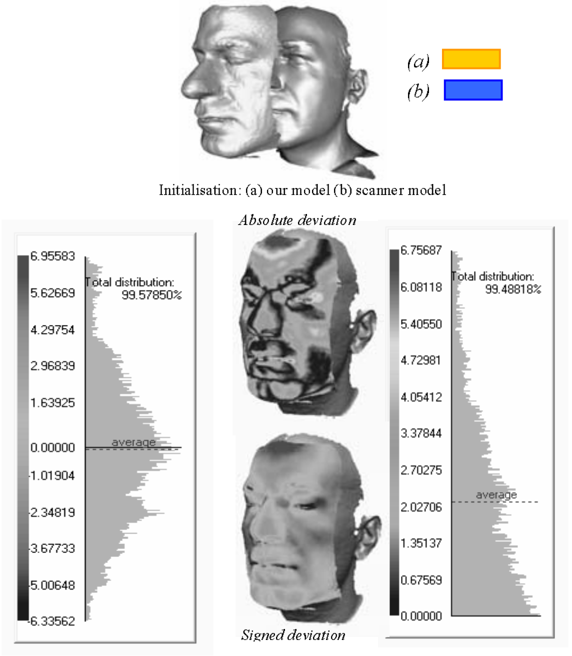
As described at the beginning of this report, the basic system in our platform is a binocular sensor, where 3D reconstruction goes through some stages which are calibration, matching and triangulation. As is known, the accuracy of this sensor depends on these stages particularly finding correspondences. For that, we are choosing to project a kind of special structured light on a face in order to distinguish, with sub-pixel precision, two sets of features in the stereo images. This process, coupled with the application of our dynamic programming algorithm on these located features and cubic-spline based interpolation, makes up our main contribution.

The accuracy of this kind of sensor depends also on two other important factors. The first one is the distance from the sensor to the object to be scanned (i.e. depth Z). The second source factor is the distance between the stereo cameras (i.e. baseline B).

$$Z = \frac{f.B}{d} \quad (a) \quad ; \quad \delta Z = -\frac{Z^2}{f.B}\delta d \quad (b) \qquad (6)$$

The equation (6(b)) is the derivation of the classical formula of depth from stereo (6(a)) after rectification. Here, f is the focal length and d is the disparity (displacement of correspondent points). Here, the depth error is proportional to the square of depth and inversely proportional to the baseline. Consequently, the nearer the object is, more accurate is the reconstruction and the shorter the baseline is, the less accurate is the triangulation. Currently, we are working to find the optimal parameters to improve our results.

In order to evaluate our acquisition technique, we consider models, which are reconstructed from a laser triangulation-based scanner, as reference data or "ground truth". The evaluation procedure consists of digitizing one face using our technique and acquiring the same face by the scanner both with neutral expressions. Then, we compute spatial deviation between the two sets of points. In this procedure we use, after manual coarse alignment, the Iterative Closest Point algorithm which minimizes the global distance between them. Figure 9 shows the two data sets and the two distribution curves: the signed and the absolute deviations.



Initialisation: (a) our model (b) scanner model

Deviation/Type	Mean error	Standard deviation
signed	*-0.086*	*2.11*
absolute	*2.66 (1.65%)*	*1.62*

Figure9. Experimental results and validation procedure

Figure 9 shows also mapped colors which represent error values. In the absolute deviation map, the blue color corresponds to the small errors and the red color represents the significant errors. In contrast, the signed deviation map represents maximum errors by saturated colors. We give also the mean error and the standard deviation values for each kind of deviation. Here, the mean depth error presented by our approach represents 1.65% *(i.e. Mean error/[Zmax-Zmin])* and standard deviation as 1.62 mm. The curve of the signed distribution has a unimodal silhouette centered at -0.086 mm and 2.11 mm as a standard deviation. In this measuring error procedure, we must take into account some considerations, especially errors from laser scanner acquisitions and ICP-based distance calculation.

8. Conclusion

We have presented in this paper a complete low-cost and accurate solution for 2.5D and 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 pairs with sub-pixel precision. The global matching optimization is performed by a dynamic programming algorithm for each pair of the scanlines independently. Finally, depth is obtained by a pipeline of light-ray intersections identification, points' interpolation based on cubic spline models, points' meshing and texture mapping. This work presents a novel framework in which we associate existing techniques to new ones for face reconstruction. The main contributions are first in the stereo matching problem where feature extraction descends to less than the pixel unit. The second contribution concerns face modeling by using cubic spline interpolation in the X-direction where the Y-direction is densely acquired. This operation conserves details obtained from original points and ameliorates face resolution by the introduction of new points between original points.

9. References

[1] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E Tabassi, and J.M. Bone, "FRVT 2002: Evaluation Report",. March 2003

[2] V. Blanz, T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, September 2003, pp. 1063-1074

[3] A. Bronstein, M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition", Proc. Audio & Video-based Biometric Person Authentication (AVBPA), Lecture Notes in Comp. Science 2688, Springer, 2003, pp. 62-69

[4] X. Lu and A.K. Jain, "Integrating range and texture information for 3D face recognition", Proc. of WACV (Workshop on Applications of Computer Vision), pp. 156-163, Breckenridge, Colorado, January 2005.

[5] B. Ben Amor, M. Ardabilian, L. Chen, "3D Face Modeling Based on Structured-light Assisted Stereo Sensor". Proceeding of ICIAP 2005, Cagliari, Italia, 6-8 September 2005.

[6] Home page: www.minolta-3d.com

[7] E. Trucco and A. Verri, Introductory Techniques for 3D Computer Vision, ISBN 0-13-261108-2 Prentice Hall 1998.

[8] 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.

[9] T. Fujiwara, H. Koshimizu, K. Fujimura, H. Kihara, Y. Noguchi, N. Ishikawa , "3D Modeling System of Human Face and Full 3D Facial Caricaturing", Third International Conference on 3-D Digital Imaging and Modeling (3DIM '01) , Quebec City, Canada, May 28 - June 01, 2001.

[10] S. Lao, M. Kawade, Y. Sumi, F. Tomita: "Building 3D Facial Models and Detecting Face Pose in 3D Space". 3DIM 1999: Ottawa, Canada, pp. 390-404.

[11] 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.

[12] Zicheng Liu, Zhengyou Zhang, Chuck Jacobs, Michael Cohen., "Rapid Modeling of Animated Faces From Video", Journal of Visualization and Computer Animation, Vol 12, No.4 (Sep. 2001), Page 227-240.

[13] E. Garcia, J.-L. Dugelay, H. Delingette, "Low Cost 3D Face Acquisition and Modeling", ITCC, Las Vegas, Nevada, April 2001.

[14] L. Zhang, B. Curless, and S. M. Seitz. "Rapid shape acquisition using color structured light and multipass dynamic programming". In Int. Symposium on 3D Data Processing Visualization and Transmission, Padova, Italy, June 2002.

[15] Yang Liu, George Chen, Nelson Max, Christian Hofsetz, Peter McGuinness. "Visual Hull Rendering with Multi-view Stereo". Journal of WSCG. Feb. 2004.

[16] Chia-Yen Chen, Reinhard Klette and Chi-Fa Chen, "3D Reconstruction Using Shape from Photometric Stereo and Contours", October, 2003

[17] 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.

[18] C. Hernández Esteban and F. Schmitt, "Silhouette and Stereo Fusion for 3D Object Modeling", 3DIM 2003, 4th International Conference on 3D Digital Imaging and Modeling, Banff, Alberta, Canada, October 2003, pp. 46-53.

[19] De Boor, C., *A* Practical Guide to Splines, Springer-Verlag, 1978.

[20] Y. Ohta and T. Kanade, "Stereo intra- and inter-scanline search using dynamic programming", IEEE Trans. PAMI 7:139-154, 1985.

[21] P. Besel and N. Mckay: "A method for registration of 3D-shapes". IEEE trans. Pattern analysis and Machine intelligence, 14(2):239-256, 1992.