DISPARITY-BASED 3D FACE MODELING USING 3D DEFORMABLE FACIAL MASK FOR 3D FACE RECOGNITION

A-Nasser Ansari, Mohamed Abdel-Mottaleb, and Mohammad H. Mahoor

University of Miami, Department of Electrical & Computer Engineering
1251 Memorial Drive, Coral Gables, FL 33124
E-mails: a.alansari@umiami.edu, mottaleb@miami.edu, mmahoor@umsis.miami.edu

ABSTRACT

We present an automatic disparity-based approach for 3D face modeling, from two frontal and one profile view stereo images, for 3D face recognition applications. Once the images are captured, the algorithm starts by extracting selected 2D facial features from one of the frontal views and computes a dense disparity map from the two frontal images. Using the extracted 2D features plus their corresponding disparities in the disparity map, we compute their 3D coordinates. We next align a low resolution 3D mesh model to the 3D features, re-project it's vertices on the frontal 2D image and adjust its profile line vertices using the profile view. We increase the resolutions of the resulting 2D model only at its center region to obtain a facial mask model covering distinctive features of the face. The computation of the 2D vertices coordinates with their disparities results in a deformed 3D model mask specific to a give subject's face. Application of the model in 3D face recognition validates the algorithm and shows a high recognition rate.

1. INTRODUCTION

3D face modeling has been an active area in computer graphics and machine vision. The need for accurate face modeling for realistic facial animation, compression, and recognition has led researchers to explore 3D model-based algorithms. The approaches in the literature for 3D face modeling differ in many aspects depending on the application, computing efficiency, type of sensors, cost, and required accuracy. In [1], a 3D generic face deformation is proposed for videophone applications, where they use a single image to adapt the generic model to the face in the image. In [2,6,8], they adapt a generic face to facial features extracted from images taken by an expensive 3D scanner. In [4] stereo images were used to create 3D face models using differential geometry. Blanz et al. [8] develop a good system to create 3D face models from a single image. However, their system requires 3D scanner and initial manual intervention to create the 3D model. In addition their system is computationally more expensive. Medioni et al. [10] use stereo images based on internally and externally calibrated cameras. Their system differs from ours because they capture only two parallel frontal images and do not incorporate a profile view or a mesh model.

In this paper, we propose an automatic system for 3D face modeling using two frontal and one profile view images, taken by a calibrated stereo system. The system automatically extracts 57 facial features from one of the frontal views, the reference left image. Via stereo correspondences, a dense disparity map is computed from the two frontal view images. The map is then filtered and interpolated in areas of missing disparities. Using the 57 2D features points and their corresponding disparities, the system computes their 3D coordinates. It then rigidly aligns the vertices of a low resolution 3D mesh model to the 3D features, projects the entire vertices on the images for adjustment, and extracts 11 additional features along the profile line of the face. Next, the system increases the triangular resolution of the re-projected model, at the reference image, and obtains a high resolution mesh mask of the model only around areas of the eyes, eyebrows, nose, and mouth. Using the vertices locations of the model mask and their corresponding disparities, from the disparity map, the mask is deformed to the individual's face by computing their 3D coordinates using triangulations. Unlike the existing 3D algorithms that rely on range data, which can be accurate but suffer from the difficulty of extracting salient facial features [11-14], ours extracts a total of 68 facial features using 2D techniques because it is easier to extract corners and edges from the intensity frontal images than from 3D range images. In addition, some of these algorithms either extract or deform a model to local surface curvatures in 3D world plane. On the other hand, in our approach, we deform the mesh model mask using feature correspondences encoded in the disparity map, which are obtained from the 2D image plane. Moreover, we combine information from the disparity map and the 2D images in order to obtain more robust results than those obtained by using 3D data alone. We verify the uniqueness of the final deformed mask model to a given face by applying the model mask in 3D face recognition. The general objective of our research is to create a biometric system for recognizing frontal faces, with neutral expression, under cooperative environment, and controlled light variations.

This paper is organized as follows: Section 2 introduces our calibrated image acquisition system. Section 3 explains the computation of the disparity map. Section 4 briefly describes the facial feature extraction from one view image using an improved statistical method. Section 5 explains the use of the mesh model; its alignment and its local deformation. Section 6 presents the application of the deformed model in 3D face recognition. Conclusions and future work are given in section 7.

2. STEREO SYSTEM

We use consumer type cameras and arrange them in parallel/orthogonal stereo configuration as shown in Fig.1. Unlike our previous work [18,19], where we used one frontal and one profile views, this system simultaneously captures two frontal and

one profile view images of the face. Instead of using trifocal tensor, we treat the two left and right frontal cameras as one stereo system and the left and profile cameras as another one. In either configuration, we have the origin in the left camera reference frame. We use the procedures given in [7] to find the extrinsic and intrinsic cameras parameters. In addition, to minimize the Euclidian distance errors between the observed image points and the re-projected image points over the three cameras, we refine the calibrated external parameters using bundle adjustment and finally perform image rectification on the frontal view images.
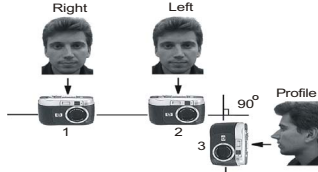

Fig. 1 Image acquisition set-up.

3. COMPUTATION OF THE DISPARITY MAP

The disparity map is computed from correspondences in the two rectified frontal images. We use a modified version of the progressive stereo matching algorithm in [17]. The core idea of [17] is to treat the disparity d at image location (u,v), as a surface in a volume of an algebraic function, and to extract the surface for seed voxels (a voxel is a 3D pixel), which have high probability of being correct matches. The function used is the normalized cross-correlation with a window. Two pixels indicate a correct match when their cross-correlation is close to one. Matching two pixels is therefore equivalent to extracting a maximal surface from the volume. To compute the disparity map, the algorithm undergoes two steps: seed voxel selection and disparity surface tracing to find the maximal voxel with high correlation. This algorithm requires accurate initial seeds and does not produce accurate disparity map. Our modification to the algorithm comes in the initial seed voxels and final interpolation of the disparity map. In addition, we use fully calibrated parameters and enforce guided matching restricted about the epipolar lines using a computed fundamental matrix F. For each pair of corresponding image coordinates $p_1 = (u_1,v_1)$ and $p_2 = (u_2,v_2)$, the fundamental matrix satisfies $p^t{}_1 F p_2 = 0$. To find accurate and populated initial seeds, we start by performing background segmentation in the original un-rectified left and right images in order to speed up subsequent steps. See Fig.2.a. The use of the undistorted rectified images will be explained subsequently. Next, we extract random corners in the segmented images [15]. This produces a large number of corners as shown by the red points in Fig.2.b. Then we perform correlation matching for each detected corner and select the ones with highest correlation match. These points are indicated by the green color in Fig.2.b. As shown in Fig.2.c, we further refine the corresponding corners and select the ones within sub-pixel distance from their epipolar lines. For that we use the Sampson's distance function. Given the fundamental matrix and a corresponding pair, Sampson distance returns the distance, which is a first order approximation to the geometric distance, called the gradient distance. For a paired correspondence, it's given in the discrete form as

$$Sd = \frac{\left| [u_2\ v_2\ 1]^t \cdot F \cdot [u_1\ v_1\ 1] \right|^2}{[Fu_1]^2 + [Fv_1]^2 + [F^t u_2]^2 + [F^t v_2]^2} \quad (1)$$

In our algorithm, we discard any correspondence with distance $Sd > 0.5$ pixels.
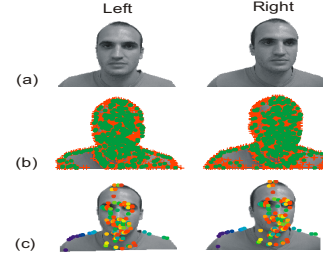

Fig. 2 (a) Background segmented images, (b) red, random and green corresponding corners detection, (c) shown with common colors, refined corresponding corners using Sampson's distance.

Finally we convert both, the detected corners and the face images to their rectified image coordinates. The above procedure finds accurate initial seeds, which are used as inputs, along with the rectified images, to the approach given in [17]. Fig.3.a shows the result of the subject's face given in Fig.2.a using [17]. Unfortunately, some holes exist either due to occlusions or failure of the algorithm to detect correspondences, especially at areas of reflection such as the forehead and the tip of the nose. Also, overshoots which exist around the boundary edges. For our application, we are only interested in the interior and un-occluded region of the face.
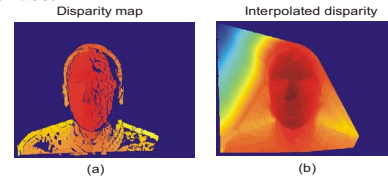

Fig. 3 (a) Computed and (b) interpolated disparity maps.

To overcome the holes in the disparity we apply cubic surface interpolation function S at locations (u,v) over the disparity of Fig.3.a and obtain the interpolated disparity $d' = S(u,v)$. The result of the interpolation is given in Fig.3.b, which indicates obvious improvement compared to Fig.3.a. Fig.4 shows a comparative example of testing the disparity map with facial features correspondences using both disparities given by Fig.3.
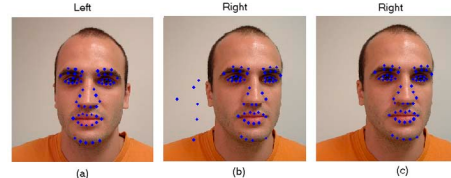

Fig. 4 (a) Features extraction, left view, (b) feature correspondence in right view to features in (a) using disparity of Fig.3.a and (c) using interpolated disparity of Fig.3.b.

Using the algorithm described in the next section, we obtain the facial features given by Fig.4.a. Using these features plus the disparity map of Fig.3.a, we compute the locations of the corresponding features in the right view. We can see from Fig.4.b that due to the holes in the disparity, many features have the wrong correspondences located away form their correct positions. On the other hand, using the interpolated disparity of Fig.3.b with the extracted features of Fig.4.a gives the correct correspondences as shown in Fig.4.c. As a result 3D coordinates can accurately be computed.

4. FACIAL FEATURE EXTRACTIONS

982

We automatically extract important facial features around the eyes, eyebrow, mouth, nose, and chin, using our improved version of Active Shape Model, ASM [20]. The standard ASM developed by Cootes et al. [16] suffers from poor initialization and modeling the local structure of the facial features in the images. In the original version of ASM, the initial feature points are obtained from the mean shape, which is derived from training data, and its accuracy depends on the size of the training set. In addition, the local structure of the feature points is represented by the change in the intensity values of the pixels along a profile line (i.e. edge location) that runs through the feature points. The core of our improvement relies on (a) initialize the ASM using the centers of the mouth and eyes using our algorithm in [5], (b) incorporating RGB color information in the feature matching process, and (c) unlike the standard ASM, which uses Euclidean transformation, we use 2D affine transformation which better aligns the extracted facial features to the shape model. Fig.4.a shows one example for facial feature extraction for one subject out of our captured database. Using these features and their corresponding disparities, we compute their 3D coordinates using stereo triangulation. These 3D coordinates are used in aligning and deforming a face mesh model as we describe next.

5. 3D FACE MODELING

This section deals with molding the human face using its extracted features and interpolated disparities. The idea is to align a 3D model to the computed 3D features in 3D, re-project its vertices on the frontal 2D image, obtain a high resolution 2D model mask at distinctive areas of the face using triangles sub-division, and compute the 3D vertices coordinates of the mask using their locations in the image plus their corresponding disparities. Fig.5.a shows our neutral 3D model with a total of 109 feature vertices and 188 mesh faces. Fig.5.b shows the 2D model mask that is generated from an aligned and deformed version of the model as we explain next. We point out that the neutral model is designed such that the left and right sides of the jaw fall within but not on the edges of the face boundary. This approach avoids inaccurate correspondences that might exist at the edges of the disparity map.

5.1 Global alignment

In the global alignment step, we rigidly align the 3D model using the 3D feature points, P_I, obtained from the images, and their corresponding feature vertices, P_M, in the model. Subscripts I and M indicate image features and model vertices, respectively. To achieve this goal, the model must be rotated, translated, and scaled. Eq.2 gives the sum squared error between P_I and P_M in terms of scale S, rotation R, and translation T for N points.

$$Min \ E(S,R,T) = \sum_{j=1}^{N} \left\| P_{I_j} - P_{M_j} \right\|^2 \quad (2)$$

Fig.5.c-d demonstrates an example of the aligned 3D model to the calculated 3D features and its projected vertices on the image.
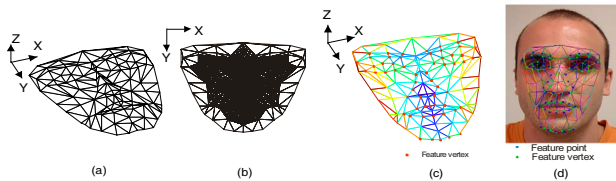

Fig. 5 (a) 3D model (b) 2D model mask (c) aligned 3D model vertices to extracted features, (d) projected model on the image.

5.2 local model adjustment / incorporation of the profile view

The global alignment rigidly aligns the entire 109 vertices of the model based on the subset of 57 extracted features. Here, we loosen the rigidity requirement at the facial feature components by replacing the coordinates of the vertices of the eyes, eyebrows, mouth, nose, and chin with the coordinates extracted from the image. In addition, there are 11 vertices that lie vertically along the middle of the face through the nose. Their initial positions are obtained after the global alignment, yet they do not reflect actual extracted positions and may not have the correct correspondences in the other views. Moreover, they are hard to extract from the frontal image because they lack texture information and intensity characteristics. In our algorithm, we extract the 11 features from the profile edge line obtained from the profile image. Fig.6.a shows the edge image obtained using edge detection and linking. Fig.6.b shows, in blue, the 11 vertices of the frontal image projected on the profile view. From the figure, we can see that the locations of the blue feature points in the profile view need adjustment. We adjust their locations by computing the intersections of their epipolar lines, from the frontal image, with the profile edge line. This establishes correct correspondence at the selected features. Fig.6.b shows, in green, the new adjusted positions. Finally we modify the disparity map, only at these 11 locations, such that they project on their new locations on the profile edge. Fig.6.c-d show the 3D model and its projected vertices on the image with all the 68 adjusted vertices locations, as explained in this section, which can be compared with Fig.5.c-d.
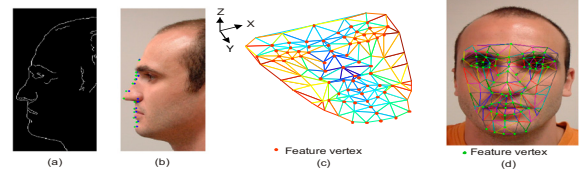

Fig. 6 (a) Profile edge image (b) original, in blue, and adjusted, in green, vertices (c) adjusted 3D model (d) projected model.

5.3 Face model mask and deformation

The objective of this section is to create a high resolution facial model mask and use it to capture the shape variations of a subject's face encoded in the dense disparity map. We increase the resolution of the model of Fig.6.d at the area of the face that has more discriminatory power among individuals (see mask in Fig.5.b). Each triangle within the mask area is divided in a manner similar to Fig.7.a. New vertices and triangles are generated based on the locations of the aligned vertices in the image plane. The result is given by the mask model superimposed on the image in Fig.7.b, after 1:16 triangle subdivisions. For clarity, we show in Fig.7.c the model without the image. Furthermore, every feature vertex superimposed on the image has a disparity at its location which can be easily extracted from the disparity map of Fig.3.b.
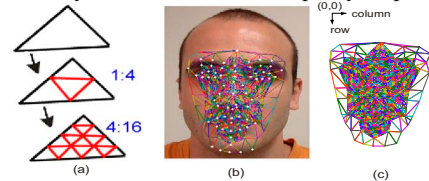

Fig. 7 (a) Triangle subdivisions, (b) 1:16 mask subdivision of Fig.6.d, the extracted features shown in white, (c) same as (b) without the image.

Combining all the vertices locations from Fig.7.b with their corresponding disparities from Fig.3.b and using stereo vision, we can compute accurately their positions in 3D. The final result is a deformed 3D model specific to a given face. The deformation process is a result of vertices correspondences in the disparity and 3D triangulations. Fig.8.a shows two views of the deformed 3D models of the subject in Fig.7.b. Due to the viewing angle in the figure, some triangles within the mask area in the frontal view of Fig.8.a may appear empty. In reality they are not. We show in Fig.8.b a zoomed version of the mask model alone.
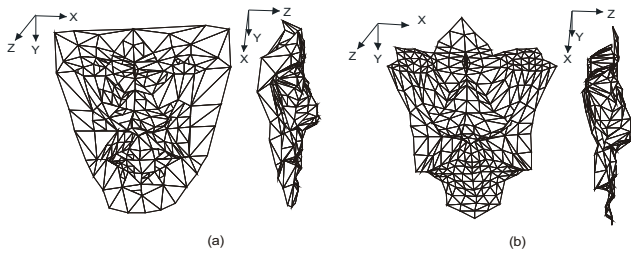

Fig. 8 (a) Deformed 3D model mask (b) zoomed view of the mask alone of the model in (a).

6. APPLICATION TO 3D FACE RECOGNITION

From Fig.8, we can see that the model is successful in capturing the surface characteristics of the face. Due to limited space, our goal in this section is to briefly demonstrate the potential of the constructed 3D model mask in 3D face recognition. Face recognition has received great attention in the past few years. A recent literature survey of 3D face recognition is given in [9]. Because our algorithm uses calibrated orthogonal cameras for capturing images, we cannot experiment on public databases, since there is no public database captured with available internal and external stereo camera parameters and setup similar to Fig.1. In our work, we start by creating a database that contains a 3D face model for each person. Currently we have 3D models for 50 persons of various ethnic backgrounds and ages. For each person we capture two sets of two frontal and one profile views. One set is used for building the database model and the second is used for testing. For 50 people, we captured a total of 300 images. Given a test face, taken by the cameras setup of Fig. 1, we obtain the deformed 3D model as given by Fig.8. We then rigidly align the model of the test face with each face model in the database using the global alignment explained in section 5.1. Next, we calculate the weighted distances between each of the corresponding vertices in the test model and the corresponding vertices in the database models. The database model that has the largest number of vertices close to the test model is selected as the recognized face. We tested the algorithm with all the 50 faces and found that 49 people were recognized correctly, i.e. a recognition rate of 98 %.

7. CONCLUSION AND FUTURE WORKS

We presented a disparity based algorithm for 3D face modeling using 2D and 3D information. We rely on the 2D data for facial features extraction and disparity map computations. We use an a priori deformable mesh model as a tool for capturing the dense surface characteristics of a given face encoded in the disparity map and compute its 3D coordinates. The model is designed to emphasize the center area of the face using triangles subdivisions. An application to 3D face recognition is presented. The quality of the results is encouraging in both 3D face modeling and recognition. We are in the process of building a database of 3D models for more people to demonstrate larger scale recognition.

8. REFERENCES

[1] M. Reinders, P. Beek, B. Sankur, J. Lubbe, "Facial feature localization and adaptation of a generic face model for model-based coding," *SP: IC*, Vol.7, No.1, March 1995.

[2] R.L. Hsu and A.K. Jain, "Face modeling for recognition," *ICIP*, Greece, Oct., 2001.

[3] B. Guenter, C. Grimm, H. Malvar, and D. Wood, "Making faces," *Proc. SIGGRAPH,* 1998.

[4] R. Lengagne, P. Fua, and O. Monga, "3D face modeling from stereo and differential constraints," *ICAFGR*, 1998.

[5] R. Hsu, M. Abdel-Mottaleb, and A. Jain, "Face detection in color images," *IEEE Trans. on PAMI*, 2002.

[6] W. Lee and N. Thalmann, "Fast head modeling for animation," *Image and Vision Computing*, March 2000,

[7] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *ICCV*, Greece, Sep. 1999.

[8] V. Blanz and T. Vetter, "A morphable model for the syntheses of 3D faces," *Int. Conf. on CG*, 1999.

[9] K. Bowyer, K. Chang, and P. Flynn "A Survey of approaches to three-dimensional face recognition,"*ICPR*, Cambridge, England, August 2004.

[10] G. Medioni and R. Waupotitsch, "Face recognition and modeling in 3D," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 232-233, 2003.

[11] C. Hehser, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range images," *Int. Symposium on Signal Processing and its Application*, 2003.

[12] C. Chua, F. Han, and Y. Ho, "3D human face recognition using point signature," *Proc. of AFGR*, 2000.

[13] X. Lu, D. Colbry, and A. Jain," Three-dimensional model based face recognition," *Pattern Recognition* ,UK, 2004.

[14] Y. Lee, K. Park, J. Shim, and T. Yi, "3D face recognition using statistical multiple features for the local depth information," *Proc. of Vision Interface*, 2003.

[15] C. Harris and M. Stephens. "A combined corner and edge detector," *Proc. AVC Conference*, Manchester, 1988.

[16] A. Lanitis, C. Taylor, T. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE PAMI*, vol. 19, issue 7, pp.743–756, 1997.

[17] H. Zhang and S. Negahdaripour, "Fast and robust progressive stereo reconstruction via symmetry guided fusion," *Proc. Oceans '05 Europe*, Brest, France, June, 2005.

[18] A-Nasser Ansari, Mohamed Abdel-Mottaleb. "3-D face modeling using two orthogonal views and a generic face model," *Proc. IEEE Conf. ICME*, July, 2003.

[19] A-Nasser Ansari and Mohamed Abdel-Mottaleb, "Automatic facial feature extraction and 3D face modeling using two orthogonal views with application to 3D face recognition," *Pattern Recognition*, Vol. 38, Issue 12, pp. 2549-2563, Dec. 2005.

[20] Mohammad Mahoor and Mohamed Abdel-Mottaleb, "Facial feature extraction in color images using enhanced active shape model," *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, Southampton, UK, April 2006.