A new approach for 3D face modeling using multi-view-stereo and ICP algorithms

Najlaa Abd Hamza[1], Ghassan H. Abdul-Majeed[2]
[1]College of Nursing, University of Baghdad
[2] College of Engineering ,University of Baghdad

ABSTRACT

This paper addresses the problem of facial reconstruction, in the form of a 3D model, from images taken with different viewpoints. The system consists of two parts: 3D dense reconstruction and 3D-3D registration. A Multi-View-Stereo algorithm is used in the 3D dense reconstruction, whereas the ICP algorithm is applied in the 3D-3D registration. The proposed system has been tested against data sets from different viewpoints. The results of the proposed system reveal its superiority over the existing models in terms of quality of reconstruction.

Keywords: Multi-View-Stereo , 3D-3D regisraion, ICP algorihm, 3D face modeling

Corresponding Author:

Najlaa Abd Hamza
College of Nursing, University of Baghdad, Iraq
E-mail: najlaaabd@conursing.uobaghdad.edu.iq

1. Introduction

Obtaining, modeling and producing realistic 3-D human faces in addition to their dynamics has appeared as an effective study issue in the limit range between the files of computer graphics and computer vision. 3D facial reconstruction is distinct as a procedure that recovers 3D information from 2D face images [1]. Acquiring a 3-D face surface model can be useful for various applications, for example, 3-D face recognition, 3D expression recognition. In spite of developing 3-D sensors for obtaining 3-D faces, precisely reproducing 3-D surface models from 2-D pictures is considered as an already existent problem of computer vision [2]. It is very significant to forensics; it has a role when investigating a crime. So it's the art of the reconstruction of what face might have looked like from a skull. Also, it is important in facial surgery to restore the form and function of the body.
3D acquirement approaches are generally characterized as passive and active reconstruction methods. In active methods, the sources of light are controlled, as main features of the procedure to reach the 3D information. In the passive methods, on the other hand, light is not controlled or just for quality of image. Generally, passive methods work with which ever sensible, surrounding light accessible. From a computational point of view, active techniques have a trend to be less demanding, as the special light is applied to recover a portion of the means in the 3D holding procedure. Their suitability is restricted to states where the special lighting measures can be workable[3]. Fig.1 displays a 3-D face modeling grouping related to reconstruction methods. Such taxonomy offers 2 classes: The first class goals 3-D static face modeling, whereas the methods of second class aim to get facial shapes in action (i.e., in 3D+t domain) [4].

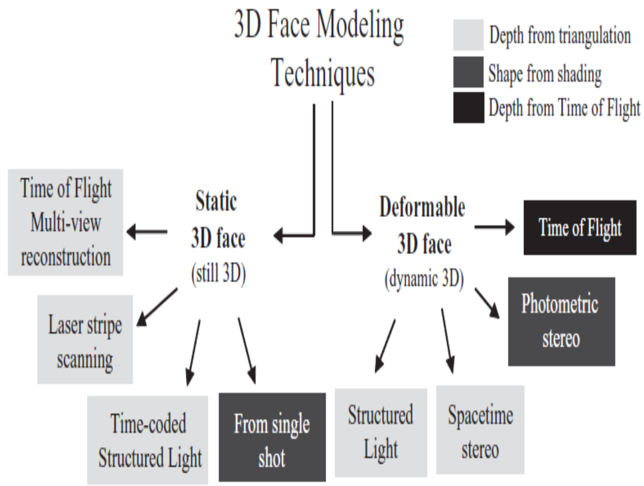

Figure1. Taxonomy for 3D face modeling methods [4]

The present work proposes a new system for creating a realistic 3D model of the face from multiple facial images (different in view points) using passive multi-view system (Multi-View Stereo, MVS). The output of MVS is combined with the general model using the algorithm of Iterative Closest Point (ICP) for the purpose of getting the final 3D face model.

2. Multi-view stereo technique

In the case when $\boldsymbol{q}$ is considered as an image of the same 3-D point and $\boldsymbol{P}$ is taken via a different camera from different angle of view, the 3-D coordinates related to $\boldsymbol{Y}$ could be extracted via calculating the intersection of 2 rays, $\mathbf{v_1}$ and $\mathbf{v_2}$, which begin from the centers of the camera crossing through $\boldsymbol{P}$ and $\boldsymbol{q}$ respectively. This is identified as the optical triangulation principle. $\boldsymbol{P}$ and q will be named corresponding or matching pixels due to the fact that they are images of the same 3-D point P.

A 3-D point P might be considered as the intersection of n (n >1) rays v_i crossing over the optical centers c_i of cameras $\{Ci\}$ where i= 1, . . . , n. Also, this could be donated to passive optical triangulation. Multi-View stereo reconstruction utilizes the principle of triangulation for extracting the depth map of an object from 2 or more projections.

As illustrated in Fig.2, Multi-View stereo algorithm defines the location of a point in space via the discovery of intersection related to rays v_i crossing via the center of projection c_i of the i_{th} camera and the projection of the point P in each image, pi.

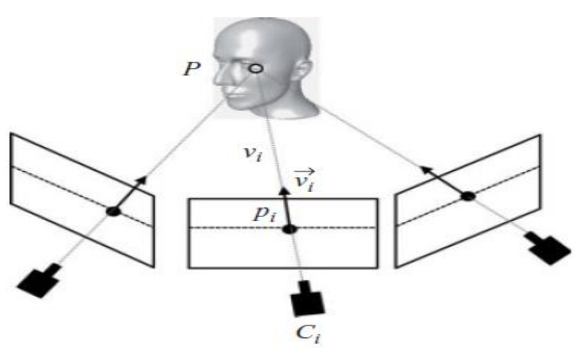

Figure 2. passive optical triangulation [4]

There are two goals of the multi-view stereo (MVS) algorithm. Initially, it allows supporting constraints on stereo corresponding, reject false corresponds, and raise the accuracy of good corresponds. Next, spatial arrangement of the cameras permits covering full face. The multi-view stereo technique creates dense 3D reconstruction of an object, by detecting correspondences between the images using the values of the camera parameters. These correspondences are triangulated to get 3D geometry. It was created as a natural enhancement to the two-view case. Rather than taking 2 pictures from 2 unlike vantage points, multi-view stereo could take the more vantage point for increasing robustness, e.g. to noise of image or texture of the surface[5].
Fig.3 shows an outline of a general MVS algorithm[5]. Different algorithms could utilize different executions regarding each of the main steps, yet, the general method is always the same[6]. The following steps are adopted in the present study:

- Capturing the images.
- Computing the parameters of the camera for each image.
- Reconstructing 3-D geometry of the object from the series of images using a triangulation process
- Optionally reconstruct the materials of the scene.

In this paper, firstly a set of real images from different viewpoints are calibrated and then entered into a robust, metric structure-from-motion (SfM) method [7][8]. This method depends on the SIFT feature detector [9] and produces intrinsic and extrinsic camera parameters (location, orientation) which represent the inputs to the MVS algorithm. The ICP algorithm is then used to perform registration between the results of MVS algorithm and the general model (natural model) to get the final model.

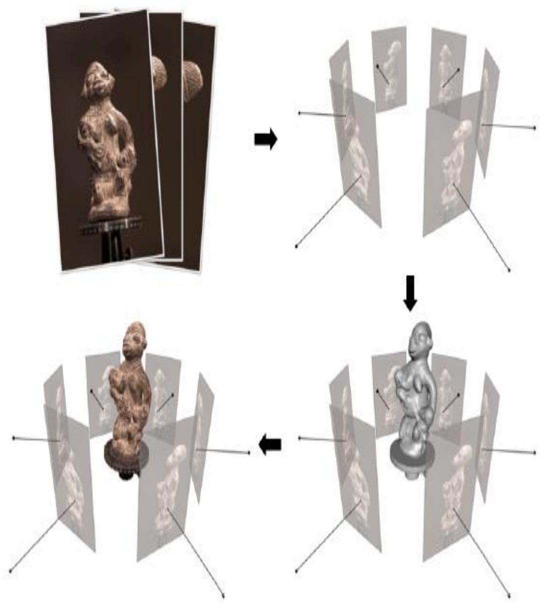

Figure 3**.** Multi-view stereo algorithm steps [5]

3. 3D-3D registration

This part brings the complete general 3D model (base mesh), which could be estimated under the matching to the relating 3D coordinates produced from 3D reconstruction procedure. The reverse is also possible, i.e., bringing the reconstructed 3D points near to the general 3D model. This matching includes translating, rotating, and scaling of the one to be moved closer to the other[10]. In the next subsection, the mathematical background regarding the Iterative Closest Point, ICP algorithm is going to be defined.

3.1. Iterative closest point (ICP) method

The algorithm of ICP is broadly applied for corresponding, matching, and registering 3-D surfaces. The basic idea of ICP structure is founded on estimating the rotation and translation (*R*, *T*) in an iterative way, that minimizes mean square error E_{ICP} between 2 mean point cloud sets (P,G)[11]. In many applications of 3D surfaces registration, ICP is generally utilized due to the fact that it has these benefits[12]:

- It could get precise registration result.
- It could be utilized with 3D point sets, implicit surfaces, parametric curves, and other surface representation forms. That is, the procedure is independent of surface representation.
- There is not high importance in doing the extraction of features and segmentation to the treated surface.
- The procedure's convergence could be assured under the condition of good initial value .

Theoretically, there are four stages that construct and compose the ICP technique, which are basically characterized by[13]:

- Finding the nearest neighbors between two 3D data point sets (P,G)
- Apply rotation and translation of one of 3D data point sets (R, T)
- Compute Error Function E_{ICP} (P,G)
- Minimization of error metric by terminating the iteration (steps 1, 2 and 3) when the value of the error function below threshold.

The standard ICP chooses all points in the point set for calculating the equivalent points, and since the number of the point set, utilized for registration, is extremely high, then the computing time will be extremely long. Thus, in later study, various approaches are suggested for decreasing the number of the point set elements, that is, to sample the matching point set. In this work, we suggest to choose only the main feature in both data points (eyes and nose) instead of choosing all the points to decrease the computing time and get an accurate result of registration. In addition, we suggest to use the point to point distance matching (as in standard ICP) by choosing the first point in the model point and compare it with all points in the second data point (brute Force match).

The approaches of minimizing the error function have the singular value decomposition approach, the orthogonal matrix approach, the quaternion approach and the dual quaternion approach[12]. The singular value decomposition approach has been used in the present study. This implementation of this approach is simple and the obtained result have extra accuracy. There are 2 sets of points: the model points (the space time faces neutral face model is adopted in this study) and the points that have been collected (we choose the points resulted from the 3D dense reconstruction). The model points are distinct as $\{P_i, i= 1, 2, ...,N\}$, the data points are defined as $\{G_i, i= 1, 2, ...,N\}$. Assuming the rotation matrix is R, and the translation matrix is T, the corresponding points P_i gained by the transformation of points Pi, have the following equation:

$$\acute{P_\iota} = RP_i + T + N_i \qquad (1)$$

The center of model points and data points can be calculated by:

$$p = \frac{1}{N}\sum_{i=1}^{N} p_i \qquad (2)$$

$$\acute{p} = \frac{1}{N}\sum_{i=1}^{N} G_i \qquad (3)$$

The stages of the SVD process can be summarized as follows:

1. Compute q_i and $\acute{q_\iota}$ by p and p

$$q_i = P_i - p \qquad (4)$$

$$\acute{q_\iota} = Q_i - \acute{p} \qquad (5)$$

2. Compute matrix H

$$H = \sum_{i=1}^{N} q_i \acute{q}_i^t \qquad (6)$$

3. Compute SVD decomposition of the matrix H

$$H = U\Delta V^t \qquad (7)$$

4. Compute the rotation matrix R

$$R = VU^t \qquad (8)$$

5. Calculate the translation matrix T

$$T = \acute{P} - R * p \qquad (9)$$

4. System overview

The proposed system uses three algorithms, as shown in Fig.4, namely SfM for producing the sparse 3D points and parameters of the cameras, MVS for creating dense reconstruction, and ICP for merging the MVS geometry with a global model (natural model), that corrects the errors in the 3D face points estimate by registration (fitting) general model (nature) according to the feature properties in 3D points (eyes, mouth and nose). In the following, we give a brief overview of the process:-

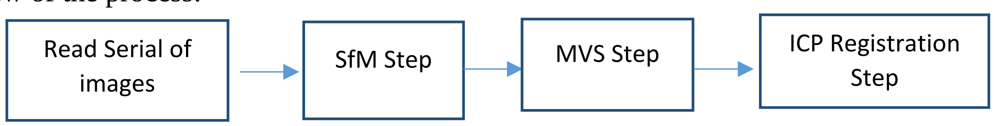

Figure 4. Steps of the Proposed System

4.1. First step: SfM algorithm

This step begins with reading a series of face images (different in view points) and produces the orientation and translation of the camera. These camera parameters represent the input to the second step (using MVS algorithm) and 3D face points which are rejected in the second step because of its sparse points. An illustration of the estimation camera parameters process is stated in the Algorithm 1.

Algorithm 1. Estimation the orientation and translation of the camera

```
Input:  sequence of images
Output: camera parameters
Begin
  //save view parameter for initial view
  Point= feature in first image
  Orientation=ones(3) // array of 3*3 of one value
  Location=[0 0 0]
  // save connection parameter between views
  ViewRelation=0     // the connection between views
  Correspond=0  //corresponding points between the two views
  Llocation=0 // location of the next camera according to the initial camera
  Lorientation=0 // orientation of the  next camera according to the  initial  camera
   For i=2 :no(image)
   Begin
SecondPoint=feature in second image
```

```
MatchPoint = matched Features between the two images
      //compute the camera parameters
        //Compute the essential matrix
        K1 = Matrix of Intrinsic paramter in   camera1
        K2 = Matrix of Extrinsic  paramterin  camera2.
        E = K2 *  fundamental matrix of camera  * K1'
     //Compute the Rotation matrix (R) and Translation vector (T)
        [U, D, V] = svd(E);
        W = [0 -1 0; 1 0 0; 0 0 1];
        Z = [0 1 0; -1 0 0; 0 0 0];
        // Rotation Matrix
        R = U * W * transpose(V)
        // Translation vector
        T = U * Z *  transpose (U)
        //add later view parameter
        SecondPoint=feature in the later image
        //add new connection
        Matches= matched points between the sequential images
        Update the L location according to the new view
        Update the L orientation  according to the new view
        orientation = L location * R
         location    = L orientation + T * Relative location
 end
End
```

4.2. Second step: MVS algorithm

In this step, the MVS algorithm is used to produce the 3D dense reconstruction, based on the series of images and the camera parameters. The Community Photo Collections Approach suggested by Goesele et al.[14], has been used in the present work. A detailed description, to get dense 3D point process, is given in the following Algorithm2.

Algorithm 2. 3D Dense Reconstruction

```
Input: matched points and view set camera parameters
Output:3D points
```

```
Begin
Calculate matched feature points between the images Mp
Read camera parameter by  using algorithm(1)
//Calculate camera projection matrix//
        for j = 1:NoView // number of View
        Begin
         O  = Orientation (j) // the orientation of camera
         L  = Location (j) // the location of camera
         I=  Intrinsicparamter(j)
             R=Transpose (O)
```

```
        T= -L * transpose (R )
      P = [R;T] * I
     end
for i = 1:no of matched feature points (Mp)
      begin
         for j = 1:no of View
        A(2*i-1,: )=MP(i,1)*P(3,: )-P(1,: )
        A(2*i,: )=MP(i,2)*P(3,: )-P(2,: )
      end
      apply SVD on A and choose the minimum singular value of A (X)
      point3d = transpose(X(1:3))
      End
```

4.3. Third step : 3D-3D registration

This step includes a registration method between the output of the second step and the general model (space time faces neutral face model[15]) to correct the error in the 3D points resulted from the MVS algorithm. For implementing this process, we suggest using the ICP algorithm. To reduce the processing time, the ICP will be appled to part of the 3D point (eyes and nose). The proposal mentioned in [16] has been adopted in this study to determine the important feature (eyes and nose) in 3D general model and 3D face Model. The registration step is described in Algorithm 3.

Algorithm 3. ***3D-3D Registration***
Input: 3D points, 3D Model Output: 3D Model
Begin **//Read the vertices of 3D Modeland 3D points** For i=1: M Read (Vi , Fi) End **// Extract the feature points from both 3D Model and 3D points** Calculate Gaussian and mean curvature (G,E) Extract the feature points depends on value of G,E (Fm, Fd) **// perform brute Force matching between the feature points** $D=(Fm-Fd)^2$ Choose min (D) // **Apply SVD to get rotation and translation matrices** Calculate the rotation matrix by using equation 8 (R) Calculate the translation matrix by using equation 9 (T) // **Transform feature points of Model depends on ICP result** NewFm=(R * Fm)+ T **End**

5. Results and discussion

In the steps mentioned in section 4, were implemented to generate the reconstructed model. Fig.5 shows the image sequence (differences in viewpoint) which represent the input data to the proposed system. Fig.6 demonstrates experimental results (sparse reconstruction) in the form of 3D point cloud and camera factors which result from SFM algorithm, takes only camera factors as input to the second step (MVS Algorithm) and reject the 3D point cloud, and Fig.7 demonstrates experimental results (dense reconstruction) in the form of 3D face model which results from MVS algorithm.

The resulting 3D face model (Fig.7) needs smoothing in some regions where there are errors and may have an outlier. These regions will be defined with the help of generic model (Fig.8). The final 3-D model is acquired after combining the generic model with the 3D face model utilizing the ICP algorithm which corrects the errors in the 3D face model estimate by deform general model (spacetime faces neutral face model) according to the feature properties in both 3D models (eyes and nose) to get final 3D model (Fig.9) which result from 3D-3D registration.

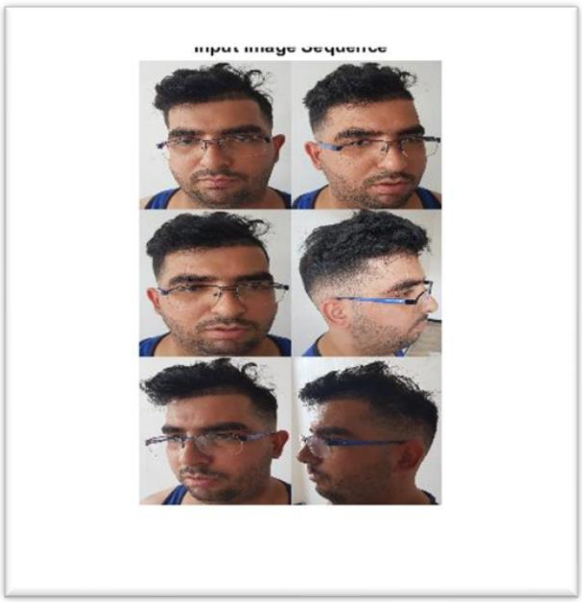

Figure5. Input image sequence

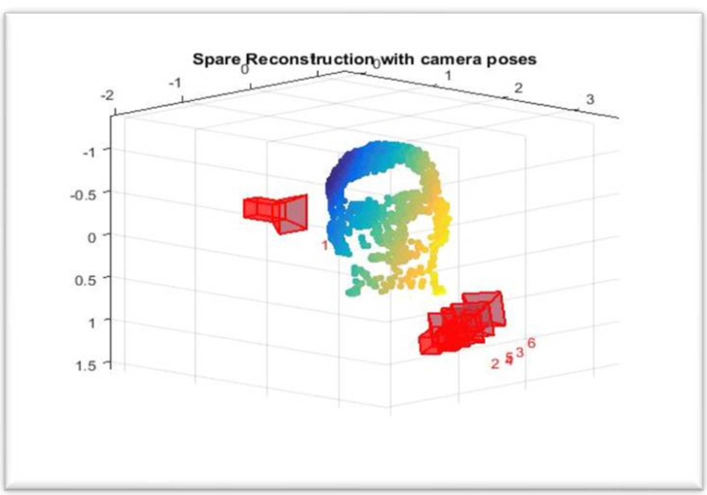

Figure 6. 3D point cloud with camera factors

Figure 7. Initial result 3D face model

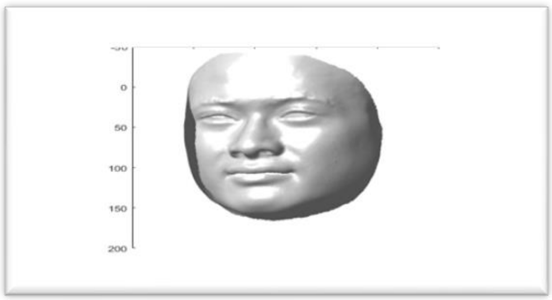

Figure 8. General Model

The generic model deforms according to only 3D feature, begins in the iterative procedure to deform the 3D model with 3D feature by finding the nearest neighbor between 3D general model (Fig.8) feature and 3D face model (Fig.7) to get the final 3D model which matches with the feature in 2D image as shown in Fig.9.

Figure 9**.** 3D face model, render face by final model and 3D face model With Texture

In this paper, two measures are presented: the reprojection error[12], the signal-to-noise ratio (SNR)[13] as shown in table1. Moreover, a comparison between the traditional reconstruction procedure (using only MVS Algorithm) and the proposed reconstructon method (combine the result of MVS algorithm with the generic model using ICP registration).

Table1. Value of reprojection error and SNR for both the traditional and proposed system methods

	The traditional method	**The proposed method**
Reprojection error	2.12	1.20
SNR	6.4	7.5

The goal in many face reconstruction methods is to make the error as small as possible and when the value of SNR is increased, it indicates that image accuracy is higher as shown in the above Table, we get reprojection error smaller than the traditional method and the value of SNR is bigger than the traditional method. The results of the proposed algorithm show it is good as far as the quality of reconstruction.

As shown in Fig.9, the proposed system gives 3D model more accurate and smooth from the traditional method. The 3D model which result from the traditional method need to be smoothed in some regions and not corresponding to 2D image features.

References

[1] E. Zheng, ”Toward 3d Reconstruction Of Static And Dynamic Objects”, Ph.D. Dissetration, University of North Carolina at Chapel Hill,2016.

[2] M. Piotraschke and V. Blanz,"Automated 3D Face Reconstruction from Multiple Images using Quality Measures", IEEE Explore,pp.3418-3427,2016.

[3] H. Najlaa Abd ,” 3D Face econstruction Method using A Hybrid Feature Detector ”, Ph.D Thesis, department of computer science, University of Technology, 2018.

[4] M. Daoudi, A. Srivastava and R.Veltkamp, "3D Face Modeling, Analysis And Recognition", Book Published in the John Wiley & Sons Ltd, 2013.

[5] Y. Furukawa and C. Hernández, "MultiView Stereo: A Tutorial”, Foundations and Trends in Computer Graphics and Vision , vol. 2, no. 1. 2007.

[6] S.Fuhrmann, F.LangguthandM.Goesele, “ MVE – A Multi-View Reconstruction Environment”, EUROGRAPHICS Workshops on Graphics and Cultural Heritage ,2014.

[7] M. Brown and D. G. Lowe,”Unsupervised 3D object recognitionand reconstruction in unordered datasets”, In Proc. 3DIM, pages 56–63, 2005.

[8] N. Snavely, S. M. Seitz, and R. Szeliski,”Photo tourism: exploringphoto collections in 3D”, In SIGGRAPH Conf. Proc., pages 835–846, 2006.

[9] S.Wua, A. Oerlemansb, E. M. akkerc and M. Lewc," Acomprehensive evaluation of local detectors and descriptors ', Elsevier journals, signal processing : Image communication , Vol.59,Nov. 2017, Pages: 150-167.

[10] M. Dikmen, "3D Face Reconstruction Using Stereo Vision" M.Sc thesis, Middle East Technical University, 2006.

[11] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2. pp. 239 –256, 1992.

[12] F. Wang and Z. Zhao, " A Survey of Iterative Closest Point Algorithm",IEEE , 2017.

[13] H. Kjer and J. Wilm, "Evaluation of surface registration algorithms for PET motion correction" Thesis,Technical University of Denmark, 2010.

[14] G. Michael, S. Noah, and S. Steven," Multi-View Stereo for Community Photo Collections, University of Washington, 2007.

[15] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz,"Spacetimefaces: high resolution capture for modeling and animation",ACM Trans. Graph., 23:548– 558, Aug. 2004.

[16] H. Shaimaa, H. Najlaa Abd," Surface Shape Descriptors on 3D Faces", Iraqi Journal of Science, 2017, Vol. 58, No.3C, pp: 1740 - 1750