3D Face Reconstruction Methods for Unconstrained Images: A Survey

Dr. Shaimaa H. Shaker
University of Technology/ Department of Computer Sciences

Najlaa Abd Hamza
University of Baghdad/ College of Nursing

Abstract:

3D face reconstruction is an exceptionally difficult issue in computer vision. What's more, is utilized as an essential for numerous essential vision undertakings e.g., face recognition. The significance of 3D face reconstruction is that we can assess illumination and pose data by utilizing the 3D shape of the face. In this paper we survey four distinctive methodologies for 3D face reconstruction also a relevant literature on 3D face reconstruction approaches has been reviewed and evaluated, which addresses the problem of 3D face reconstruction for unconstrained image.

Keywords: 3D Reconstruction, Stereo-vision, Shape from shading, Shape from silhouette, Structure from motion.

المستخلص:

ان تقنية اعادة البناء (التكوين) للوجوه الثلاثية الابعاد هي مسالة غاية في الصعوبة في رؤية الكومبيوتر. وهي من الامور الاساسية والهامة للعديد من مشاريع الرؤية الأساسية على سبيل المثال، التعرف على الوجه. أهمية إعادة الإعمار الوجوه الثلاثية الابعاد هي أن نتمكن من تحديد الإضاءة وبيانات الوقفة من خلال الاستفادة من الشكل للوجه الثلاثي الابعاد. في هذا البحث ، تم استعراض اربع تقنيات لاعادة تكوين الوجوه ثلاثية الابعاد وكذلك الادبيات السابقة لطرق اعادة البناء تم استعراضها وتقيمها التي تعالج مشكلة اعادة البناء للوجوه ثلاثية الابعاد للصور غير المقيدة.

١. Introduction

Face reconstruction is the way toward making a 3D model of a face from 2D image(s)[1].It is important consideration in mind the end goal to create human face models looking like as genuine as could reasonably be expected. This procedure involves a change from two dimensional (2D) spaces to three dimensional (3D) spaces[2].

Face reconstruction is significant with applications in face recognition, video editing[3], virtual reality, animation[2], verification, expression recognition and facial animations[4] and that's only the tip of the iceberg. For example, exact face models have been appeared to fundamentally progress face recognition[3].

In every one of these applications, the reconstructed face should be compacted and precise, particularly around critical regions like nose, the mouth, and so on. Computer games can be appeared as a good illustration which needs accurate human face models. Individuals needed to utilize their imaginings greatly more than they utilize today while playing computer games, 10-15 years prior. In any case, now, the improvement presents an incredible virtual reality with the goal that individuals don't have to utilize their imaginings abilities much[٢].

Figure 1 show general points identified with the execution of 3D face reconstruction procedures Most of the 3d reconstruction algorithms share the same basic processing pipeline, and may be not running all the processing steps.

Figure 1: General steps of 3D face reconstruction [1].

The determination of key facial features in the input 2D face image(s) is an significant stage in most 3D face reconstruction procedures. Usually face feature determination has been utilized for[1]:

١. Introducing the situating of face models in 3D reconstruction methods based on model.

٢. Finding feature on sideways and forward images so it is conceivable to misshape general 3D face prototypes to accept the form of the specified face.

٣. Instating the procedure of point tracing in 3D reconstruction methods based on video.

٤. Creating point consistency in faces taken from unlike vantage point

Another important part of the reconstruction process is "data acquisition". It is gathering the 3D information about that object by using one of 3D reconstruction techniques (We show in section 2)

an significant task of 3D reconstruction methods is the face registration. It is bringing the whole default 3D model vertices as close as could be expected under the corresponding to the relating 3D coordinates of the feature points computed from images. The opposite is also potential, i.e., bringing the calculated 3D points close to the default 3D model. This corresponding involves rotating, translating and scaling of the one to be moved nearer to the other.

With respect the use of 3D in registration, approaches can be classified into two distinct collections:

١. **3D to 3D registration:** Corresponding is done among two 3D point clouds. One is a base mesh, and other is the 3D data produced in anyway after the data acquisition procedure.

٢. **3D to 2D registration:** These methods exploit a 3D model which is used to fit 2D data.[3]

The goal of this paper is review different techniques for 3D reconstruction and for each type the most significant advantages and drawbacks are outlined, and suggest method for 3D reconstruction avoids the drawbacks and Benefit from the advantages of these 3D reconstruction methods.

In the rest of this paper, show a relevant literature on 3D face reconstruction approaches in section 2, in section 3 review four unlike approaches for 3D reconstruction, Section 4 show our proposed system, In section 5 illustrate the discussion, and, finally, references has been shown in section 6.

2. Literature Survey

In this section, a relevant literature on 3D face reconstruction methods has been reviewed and evaluated, which addresses the problem of 3D face reconstruction based on Unconstrained image.

The work presented in [11] include design a 3D face reconstruction method that takings as input either one single vantage or several unlike vantage, used sparse bundle adjustment (SBA) to reconstruct 3D landmarks from a set of 2D landmarks, The general 3D face model is warped by the reconstructed 3D landmarks and predefined 3D landmarks.

In [12] Presented 3D reconstruction approach which includes deform every image to an a general, frontal posture, improving an initial shape and lighting based on photometric stereo, and refining the model using limited vantage choice, repeat this whole system until joining.

The presented work in[2], include an iterative procedure to deform the general template face mesh based on the predictable 3D landmarks and the photometric stereo-based normal.

The work submitted in[13] proposed B-spline Shape from Motion and Shading (BsSfMS) to reconstruct constant B-spline surface for multi- vantage face images, face surface is seeming as a B-spline surface that can be reconstructed by improving B-spline regulator points, normals and 3D feature points calculated from shading and motion.

3. 3D Reconstruction Techniques:-

A 3D achievement categorization is given in Figure 2.

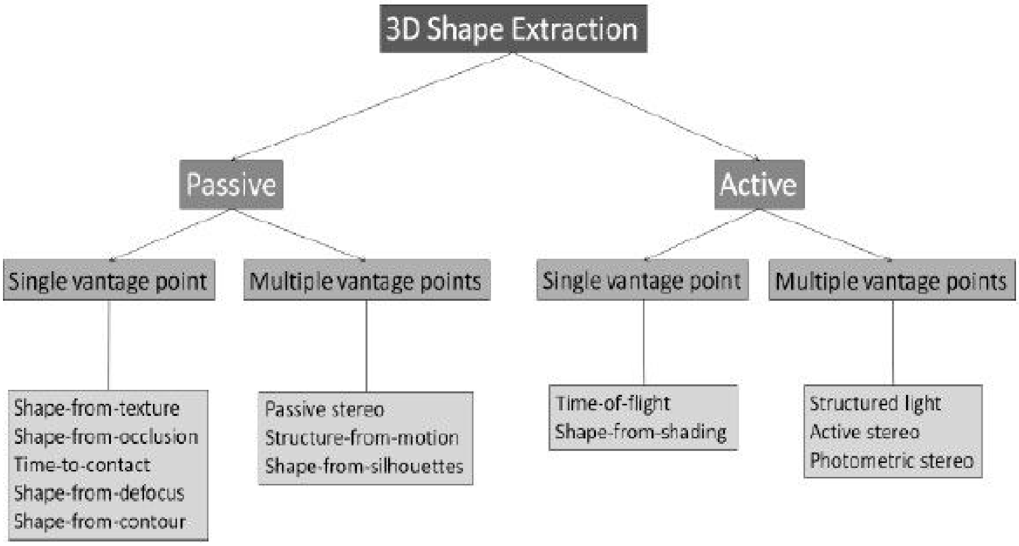

Fig. 2: Taxonomy of strategies for the extraction of data on 3D shape[5].

A first difference is among *active* and passive *techniques*, in active strategies the light sources are exceptionally controlled, as a major aspect of the methodology to land at the 3D data. In passive strategies, on the another aspect, light is not controlled or just as for image quality. Ordinarily passive techniques work with whichever sensible, surrounding light accessible. From a computational perspective, active techniques have a tendency to be less requesting, as the uncommon light is utilized to improve a portion of the means in the 3D catching procedure. Their appropriateness is limited to situations where the extraordinary lighting procedures can be connected.

A second difference is between the sum of view points from where the scene is watched or potentially lit up. With *single-vantage methods* the method workings from a single view point, With *multi-vantage systems*, several views and/or controlled lighting source points are involved[5].

In this section, we survey four distinctive methodologies for 3D face reconstruction. A synopsis of properties of various techniques is appeared in Table 1.

Table 1. The compare of various 3D reconstruction strategies [6].

3D Reconstruction strategies	Input	Output
Stereo-vision	Single	dense
Shape from shading	Single	dense
Shape from silhouette	multiple	dense
Structure from motion	multiple	spare

3.1. Stereo-vision Technique:-

Stereo-vision might be defined as the technique for computing the depth map from two images acquired from two camera frameworks. It is trademark for stereo-vision that the separation of the cameras is moderately little and consistent. It is accepted that, the image gaining happens in similar time for both cameras. One can recognize the subsequent stereo-vision arrangements[7]:

- side-camera standard (parallel optical axes).
- side-camera with crossing optical axes.
- axis-motion.

The rule behind 3D reconstruction based on stereo is basic: specified the two projections of similar point on the world onto the two images, its 3D position is found as the crossing point of the two projection rays. The state is represented in Figure 3. Repeating such procedure for a few points yields the 3D shape and structure of the object in the scene. Take note of that this structure referred to as triangulation involves the calculations of the rays and, consequently, total information of the cameras: their (relative) locations and orientations. The procedure to decide these parameters is called (camera) calibration.

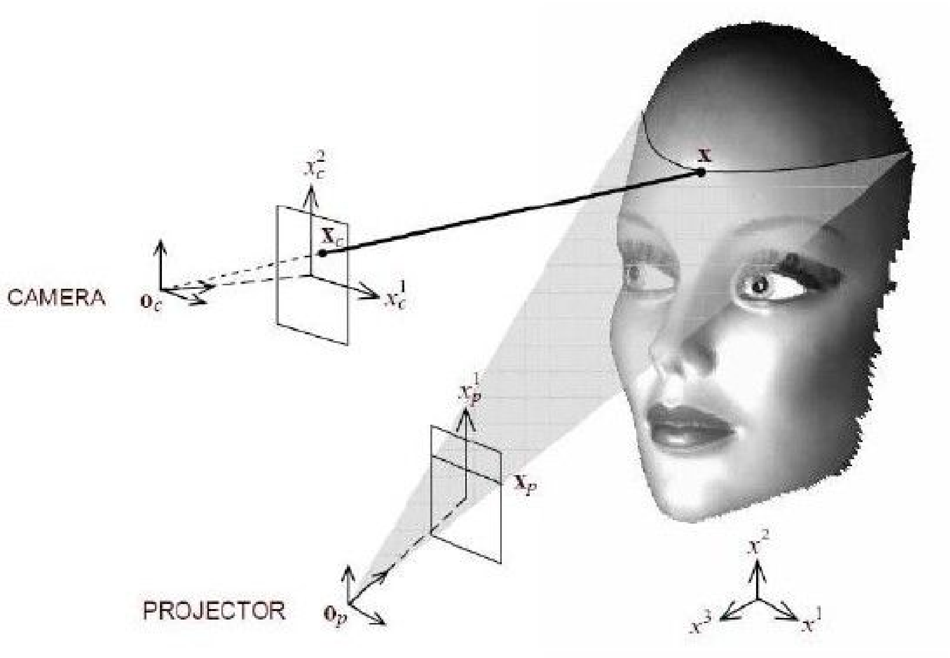

Figure 3: stereo-based 3D reconstruction[8]

In addition, to achieve this triangulation procedure, one needs methods for solving of the correspondence issue. i.e., discovery the point in the second image that relates to a particular point in the main image, or the other way around. Correspondence seeking really is the inflexible portion of stereo, and some would normally need to resolve it for several points[5].

3.2. Shape from shading Technique

Shape-from-shading (SFS) manages the retrieval of 3d shape to single monocular image by using a progressive variety of shading in the image. Artists have since quite a while ago misused lighting and shading to pass on striking dreams of profundity in sketches. To solve the SFS issue, it is critical to concentrate how the images are formed. A straightforward model of image development is the Lambertian model, in which the grey level at a pixel in the image relies on upon the light basis direction and the surface normal. In SFS, known a gray level image, the goal is to get well the light birthplace and the surface form at each pixel in the image. SFS methods can be separated into four gatherings: minimization methods, propagation methods, local methods and linear methods. Minimization methods get the result by reducing an energy function. Propagation methodologies proliferate the shape data from an collection of surface points (e.g., singular points) to the entire image. Local methodologies determine shape in view of the suspicion of surface type. Linear methodologies calculate the result in view of the linearization of the reflectance map[9].

3.3. Shape from silhouette Technique

The goal of shape from silhouette technique is to products a 3D points for a surface(object) by utilize many surface boundary derived from various surface video or a images sequence. A whole set of silhouettes taken from unlike angles offers specifics connected to the geometrical construction of a surface therefore it is potential to produce a 3D surface from silhouettes. Figure 4 shows classic face silhouettes take out from a sequence of image. Shape from shading strategies have been utilized either independently or by mix together with other 3D reconstruction techniques like shape from motion or stereo. In the most straightforward formula a 3D reconstruction technique based on silhouette holds the subsequent stages:

١. Catch images of the similar surface from unlike vantage point.

٢. Utilize image handling procedures for deriving surface boundary in every image.

٣. Warp a deformable model unto the separated silhouettes equal the form of the model as perceived from unalike viewpoints.

٤. Accumulate texture data of the sharing images and produce the texture for the 3D surface[1].

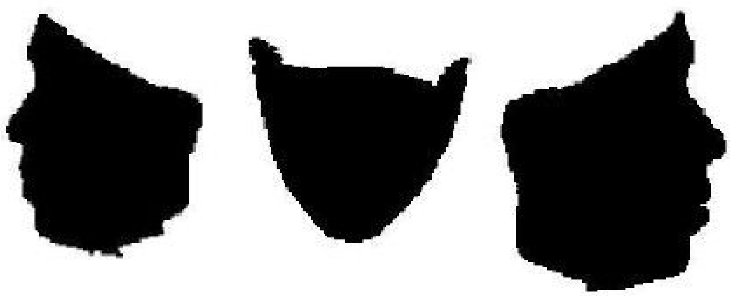

Fig 4: Cases of silhouettes of face derived from a sequence of image[1].

3.4. structure-from-motion technique:-

The objective of structure-from-motion (SfM) approaches is to products a 3D reconstruction for a surface (object) via take more than two image, while moving the camera.

SfM procedures takings as input a series of images and yield two things: the camera factors of each image, and a set of 3D points visible in the images which are regularly determined as tracks. A track is characterized as the 3D coordinates of a reconstructed 3D point and the list of corresponding 2D coordinates in a subset of the input images. Most of the present SfM algorithms have the similar simple handling pipeline (See Figure 5):

- Identify 2D features in each input image.
- correspond 2D features among images.
- Build 2D tracks after the corresponds.
- Explain the SfM model from the 2D tracks.
- Enhance the SfM model by bundle adjustment[10]

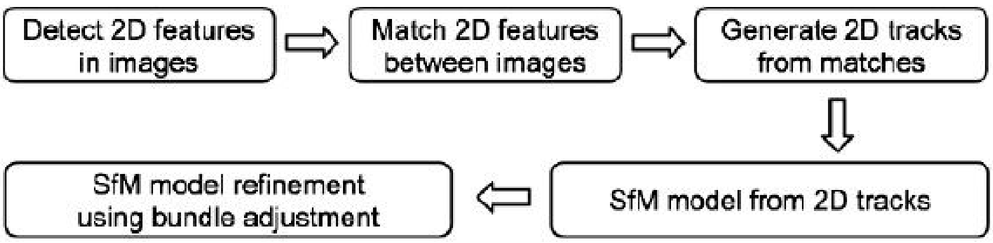

Figure5: Principle phases of a general SfM pipeline[10].

We consider the 2D location of n feature points in N frames and we stack the 2D points to a matrix W:

$$W = \begin{pmatrix} u_1^0 & u_2^0 & \cdots & u_n^0 \\ v_1^0 & v_2^0 & \cdots & v_n^0 \\ \vdots & \vdots & \vdots & \vdots \\ u_1^N & u_2^N & \cdots & u_n^N \\ v_1^N & v_2^N & \cdots & v_n^N \end{pmatrix} \quad \ldots\ldots\ldots (1)$$

The W matrix can be factorized to $M \in \mathrm{R}^{2(N+1)\times 3(\bar{m}+1)}$ matrix which contains scaled projection matrices and $B \in R^{3(\bar{m}+1)\times n}$ which contains

3D shape bases, $W = MB =$

$$\begin{pmatrix} P^0 & \bar{\alpha}_1^0 P^0 & \dots & \bar{\alpha}_{\bar{m}}^0 P^0 \\ P^1 & \bar{\alpha}_1^1 P^1 & \dots & \bar{\alpha}_{\bar{m}}^1 P^1 \\ \vdots & \vdots & \vdots & \vdots \\ P^N & \bar{\alpha}_1^N P^N & \dots & \bar{\alpha}_{\bar{m}}^N P^N \end{pmatrix} \begin{pmatrix} \bar{s}_0 \\ \vdots \\ \bar{s}_{\bar{m}} \end{pmatrix} \quad \dots\dots(2)$$

Where P^i are camera projection matrices with weak perspective model assumption, $\{\bar{S}_\iota\}_{i=}^{\bar{m}}$, are estimated 3D bases of face and $\bar{\alpha}_\iota$ are coefficients of estimated 3D shape bases. The factorization can be achieved with the

Singular Value Decomposition (SVD) on W and after calculating the corrective matrix for imposing the rotation and the basis constraints, M and B matrixes can be determined uniquely. The matrix B which holds predictable 3D shape bases can be used for estimating the 3D shape of the objective[6].

4. The Proposed System of 3D Faces reconstruction

Figure (6) shows the proposed system for 3d face reconstruction for unconstrained image.

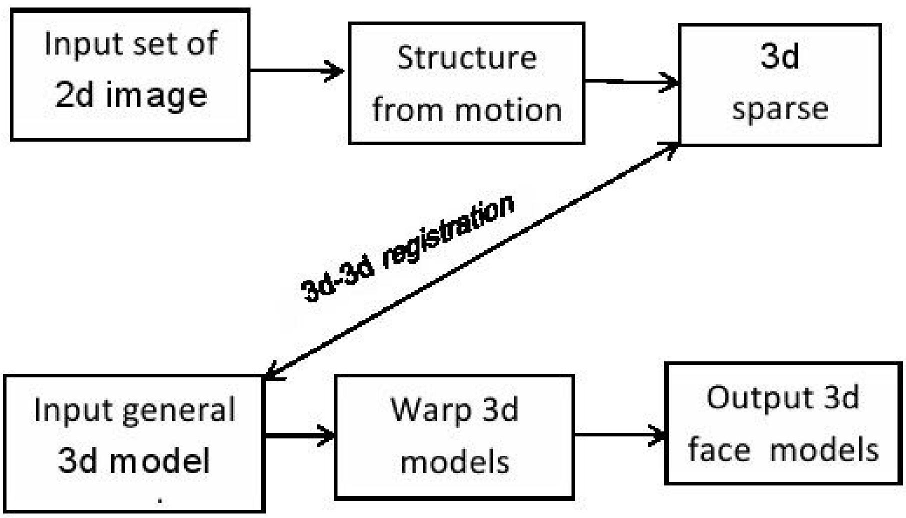

Figure 6: The proposed System.

As shown in the figure(6) The input to procedure is an "unconstrained" group of face images taken under a various variety of positions, expressions, and illuminations, without information about cameras or timing. The production of our procedure is a true 3D face surface model characterized as triangulated surface with albedo data or texture information.

Our procedure dependents on improvement structure from motion to take different image (different pose, expression, and illuminations) in the corresponding steps (see figure 5) to result 3d point.

5. Discussion

In this paper, we reviewed four different 3D face reconstruction methods. The shape from shading and stereo-vision approaches clearly estimation the dense 3D shape of the face from a single image but the structure from motion and the Shape from silhouette methods estimation the 3D shape of the face from multiple face images.

We propose using the structure from motion method in our proposed system because it is take as input a set of images and produce the camera parameters of every image(output) but all another methods(Shape from silhouette, shape from shading and stereo-vision) need Compute camera parameters for each image in input step.

The stereo-vision method is a quick and general strategy to get data about the 3D structure of the scene. Unfortunately it has faults which restrict its variety of use. As a matter of first importance the necessity of image acquisition, utilizing an uncommon set of video cameras. Another important fault is the necessity of greatly dense scene texturing. At the point when the scene's shading is homogeneous with a little measure of trademark focuses then regardless of which calculation of coordinating images will be utilized, the disparity map's quality is generally low[7].

The benefits of the shape from shading method are that it neither necessities an aligned dense 3D shapes nor requirements to utilize the symmetric segments of the face[6].

Shape from silhouette strategies can be utilized as remain solitary or complimentary techniques for advancing the quality of reconstructed faces. Such strategies require availability of various image examples or video frames of a specific face. Some essential issues identified with the utilization of silhouette based techniques are (a) how to precisely extract the silhouette and (b) how to utilize data from numerous silhouette to create precisely the 3D structure of a face[1].

The main experience of structure from motion strategies can prompt the underlying impression that they are better as a greater number of information is accessible than help reconstruction. Since the face is captured from numerous different angles, it is sensible to infer that more exact 3D faces can be reconstructed. Capturing a human face in video represents the preferred standpoint that there is more opportunity, as the human does not need to pose for photographs, making such strategies more appropriate for surveillance frameworks[1].

We propose improvement the structure from motion algorithm to take different image (different pose, expression, and illuminations) in the corresponding steps to solve the problem of unconstrained image because the structure from motion algorithm(and another methods) take as input a set of constrained image.

References

[1] Georgios Stylianou and Andreas Lanitis, "**Image Based 3D Face Reconstruction: A Survey**", International Journal of Image and Graphics · April 2009.

[2] Joseph Roth, Yiying Tong, and Xiaoming Liu," **Unconstrained 3D Face Reconstruction**", IEEE Xplor,2015.

[3] Mehmet Dikmen, "**3d face reconstruction using stereo vision**" , M.sc Thesis, the graduate school of natural and applied sciences of middle east technical university, 2006.

[4] Joseph Roth, Yiying Tong, and Xiaoming Liu, " **Adaptive 3D Face Reconstruction from Unconstrained Photo Collections**", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[5] Theo Moons, Luc Gool, and Maarten Vergauwen, "3D Reconstruction from Multiple Images", Foundations and Trends in Computer Graphics and Vision, Vol. 4, No. 4 ,2008.

[6] Amin Jourabloo, " **A Survey of Different 3D Face Reconstruction Methods**", Technical Report Jan.15, 2016.

[7] Mariusz Siudak and Przemyslaw Rokita, " **A Survey of Passive 3D Reconstruction Methods on the Basis of More than One Image**", Machine Graphics & Vision vol. 24, no. 3/4, 2014.

[8] Oisn Mac Aodha, " **Reconstruction and Recognition of 3D Face Models**" MSc Thesis, University College London, 2008

[9] Ruo Z,hang, Ping-Sing Tsai, James Edwin Cryer and Mubarak Shas, " **Shape from Shading: A Survey**", Computer Vision Lab, School of Computer Science, University of Central Florida,2000.

[10] Yasutaka Furukawa and Carlos Hernandez," Multi-View Stereo: A Tutorial, Foundations and Trends in Computer Graphics and Vision, Volume 9, Issue 1-2, 2013.

[11] Jongmoo Choi, G´erard Medioni, Yuping Lin, "**3D Face Reconstruction Using A Single or Multiple Views**", International Conference on Pattern Recognition, 2010.

[12] Ira Kemelmacher-Shlizerman and Steven M. Seitz ,"**Face Reconstruction in the Wild**", University of Washington, 2011.

[13] Weilong Peng, Zhiyong Feng and Chao Xu, "**B-spline Shape from Motion & Shading: An Automatic Free-form Surface Modeling for Face Reconstruction**", arXiv:1601.05644v1 [cs.CV] 21 Jan 2016.