Review: Recent Structure-From-Motion Algorithms 3D Shape Reconstruction

K. Punnam Chandar
Dept. of E.C.E
University College of Engineering Khammam, INDIA

T. Satya Savithri
Dept. of E.C.E J.N.T.U Hyderabad, INDIA

Abstract
Existing face recognition systems are based on 2D facial images and exhibit well-known deficiencies. Accordingly, the face recognition research is gradually shifting from classical 2D to sophisticated 3D or hybrid 2D/3D. 3D shape reconstruction from multiview photographs and video sequences (2D images) is an active area of research which can fully leverage the potential of existing 2D image acquisition systems. Currently the 3D reconstruction algorithms may be grouped in to four categories. These are Shape-from-X, 3D morphable model (3DMM), structure from motion (SFM) and Learning based. In this paper, we introduce, discuss and analyze the recent SFM methods, depth estimation based on genetic algorithm (SM), constrained independent component analysis (cICA) and Non-linear Least Squares Modeling (NLS). We begin by introducing similarity transform, which forms the basis of SFM reconstruction techniques described here. This is followed by a review and comparison of the three methods. The characteristics of the three SFM methods are summarized in a table that should facilitate further research on this topic.

Keywords
Similarity Transform, 3D Reconstructio, Face Recognition, Structure From Motion, GA; cICA; NlS Optimization.

1. Introduction

In multimedia and surveillance applications, the acquisition of human face is from multiview photographs & video sequences where the data is 2D image frames. The face recognition and retrieval accuracy on this data is reduced. The key factors that reduce the accuracy of 2D face recognition algorithms are view point, expression, illumination and available small face region in a given frame when the person is not close to the camera. To overcome the above shortcomings, 3D face models are adopted where by the 3D face models are invariant to changes in view point, background cluster, illumination and occlusions [1-5]. With the superior performance of 3D models they are gaining importance in the field of security, 3D Virtual worlds, games, 3D simulation, educational software, research in psychology and graphics. K.W. Bowyer et al. presented a comprehensive review in [6] for 3D and multimodal 3D+2D face recognition.

Presently there are two main streams of reconstructing the 3D face models, one approach is to use 3D depth sensing cameras and the other is reconstructing the 3D face model from 2D images. The high cost of the 3D depth sensing cameras limit their deployment in multimedia and surveillance applications. The alternative is to develop algorithms to reconstruct the 3D face model from 2D images and is an active area of research. The reconstruction accuracy depends on the quality of the 2D image frame, due to the trade of between quality and other parameters, like acquisition speed & storage, the quality of the image frame is compromised. Therefore, when we reconstruct a 3D face model incorporating the available prior information the reconstruction accuracy of the algorithm can greatly alleviate the capabilities of existing 2D or 3D face recognition and can be a valuable tool that can be used in various multimedia and surveillance applications,.

The goal of the reconstruction algorithm is to derive the 3D shape information of the face from N-2D images ($N \geq 2$), one frontal view and others non-frontal view. The recovered shape can be expressed in depth Z(x, y), surface normal (nx, ny, nz), surface gradient (p, q), and surface slant, ϕ, and tilt, θ. The depth can be perceived either as relative distance from camera to surface points, or the relative surface height above the x-y plane.

During the past decade many algorithms have been developed, representative 3D shape reconstruction algorithms can be grouped into four categories, i) Shape-from-X [7-9], [27], ii) 3D morphable model [10-12], iii) structure from motion [16-21] and iv) learning [13-15].

i. **Shape-from-x** - algorithms extract the shape information, with an assumption of the image formation models like Lambertian model, specular model, and hybrid model. The solution of the shape extraction problem formulated with the aforesaid models is yet not simple, as the real image formation model is more complex in nature. To deal with this, additional constraints like, brightness constraint [22], smoothness constraint [23], intensity gradient constraint [24], integrability constraint and unit normal constraint are considered in sequence. The reconstructed shape with these constraints on a considered model exhibit a large variation with the ground truth, like, the nose will be imploded, and cheeks get exaggerated. To cope with this variation

Manuscript received April 5, 2016
Manuscript revised April 20, 2016

in reconstructed shape, extension algorithms use time consuming parameter search algorithms. In any case finding a unique solution or convergence to shape-from-x, is still a difficult problem [25].

ii. **3D-Morphable-Model (3DMM)** - algorithms perform well in reconstructing the 3D shape. There are three crucial steps in reconstructing the 3D shape from a set of 3D laser-scanned heads. The first step is to align a training set of facial shapes and corresponding textures. Second step is to build the linear subspaces of 3D shapes and textures vectors, and finally in third 3D shape is reconstructed invariant to pose and illumination. A trained 3DMM represents a realistic human face as a convex combination of the linear subspace vectors built from the shape and texture. However, the reconstruction performance is achieved at the cost of the time-consuming alignment and fitting procedures [10], [11], [26].

iii. **Learning** – based algorithms exploit the common information shared by the 2D image subspaces and 3-D shape to recover the 3-D shape [15]. Thus the algorithms in this category require a coupled training set comprising of 2D and corresponding 3D faces. However, the reconstruction performance is affected by the illumination variation as these algorithms assume that the 2D and 3D faces are embedded in the corresponding linear subspaces [28], [29].

iv. **Structure from motion (SFM)** - algorithms recover the shape and motion parameters of 3D face from a 2D image sequence. 2D image are formed by projections from the 3D world. Structure from motion recovers the original 3D information by inverting the effect of the projection process. Two well-known projection models are perspective model and the orthographic model. Ullman [29] proved that four point correspondences over three views yield a unique solution to motion and structure. It is impossible to determine the motion and structure uniquely from two orthographic views no matter how many point correspondences one may have. Huang and Lee [30] and Hu and Ahuja [31] presented a linear algorithm to obtain the 3D motion and structure parameters. Shapiro et al. [32] considered the affine epipolar line properties and solved the affine epipolar line equation, and then determined all the unknown camera motion parameters. In [16] under orthographic projection, it is proved that observation matrix with rank-3 can be factorized into a shape matrix and a motion matrix using the singular value decomposition (SVD) technique. Xirouhakis and Delopouls [33] extracted the motion and shape parameters of a rigid 3D object by computing the rotation matrices via the eigenvalues and eigenvectors of appropriate defined 2x2 matrices, where the eigenvalues are the expression of four motion vectors in two successive transitions. In [19], a Gaussian prior is assumed for the shape coefficients, and the optimization is solved using the expectation – maximization (EM) algorithm. In [20], a kind of novel object independent shape basis-trajectory basis produced by the discrete cosine transform (DCT) was introduced to reduce the number of unknowns and to increase the stability of the estimation.

Among structure from motion methods, the spatial transformation approach is one important branch. Specifically, in [21] a similarity–transform [36] based 3D facial shape reconstruction algorithm is proposed to estimate the 3D structure of a human face, from a group of face images under different poses. The algorithm formulates a model and the parameter search is performed using time consuming genetic algorithm. In [34], a novel algorithm is proposed considering the observed 2D images as mixture signals and the depth information is recovered from a blind source perspective. In [35], a computationally efficient NLS-Model is presented based on the nonlinear least squares modeling of the similarity transform utilizing the prior available information. The beauty of the spatial transformation model is that they are sparse in nature, as they extract the depth information of only important features, and these models require, far smaller storage requirement than other techniques. These models find useful application in real-time applications.

This paper is about the comparison of 3D reconstruction performance analysis of the structure-from-motion algorithms based on the spatial transformation approach. We begin by introducing the similarity transform [36] which forms the foundation of all three of the reconstruction techniques described here. We have reviewed the recent depth estimation (shape reconstruction) algorithms Similarity Measure (SM) [21], Constrained Independent Component Analysis (cICA) [34] and Non-Linear Least Squares Modelling (NLS) [35] which fall into the category of SFM algorithms and compared them in terms of correlation coefficients of estimated and true depth values (mean and standard deviation), and timing (training time) in order to analyze the advantages and disadvantages of these approaches. A fair comparison for the reconstruction performance can be made as the database used in these techniques for experimentation is Bosporus 3D database, which provides the ground truth values for depth except [21]. For the SM algorithm performance can be analyzed based on face recognition results.

2. Similarity Transform

A. Similarity Transform 2D

The 2D similarity transform measures the similarity and affine distance between two images or point sets P and Q and is defined in Eq (1).

$$D_{\min}^2 = \min_{s,R_{2x2}} \left\| Q - sR_{2x2}P \right\|^2 \quad (1)$$

Where s is a scalar and R is a 2x2 rotation matrix

$$R = \begin{pmatrix} \cos(\mu) & \sin(\mu) \\ -\sin(\mu) & \cos(\mu) \end{pmatrix}$$

Differentiating the equation (1) with respect to μ and s, and equating partial derivatives to zero, we get[1]

$$s = \sqrt{rt^2[\mathrm{P'(Q)^T}] + tr^2[\mathrm{P'(Q')^T}]}$$

$$\tan\mu = \frac{rt[\mathrm{P'(Q')^T}]}{tr[\mathrm{P'(Q')^T}]}$$

B. 3D Similarity Transform

3D Similarity transform gives the 3D to 2D transformation via given rotation matrix and scale. Under orthographic projection 3D face model is projected to the corresponding 2D face based on similarity transform 3D and is given below:

$$p_i = s_i * R_{i2x3} * C + T_i \quad \text{for } i = 1,2,3,4,5\ldots N \quad (2)$$

where N is the number of non-frontal-view face images, s_i, T_i and R_i denote the scaling factor, the translation Matrix and the rotation matrix between the frontal view image and the i^{th} non-frontal-view face image, respectively. R_i can be specified as three successive rotations around the x-, y-, and z-axes, by angles ϕ_i, Ψ_i, θ_i, respectively, and can be written as the product of these three rotations as follows:

$$R_i = \begin{bmatrix} \cos\phi_i & \sin\phi_i & 0 \\ -\sin\phi_i & \cos\phi_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\Psi_i & 0 & -\sin\Psi_i \\ 0 & 1 & 0 \\ \sin\Psi_i & 0 & \cos\Psi_i \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_i & \sin\theta_i \\ 0 & -\sin\theta_i & \cos\theta_i \end{bmatrix} = \begin{bmatrix} r_{i_{11}} & r_{i_{12}} & r_{i_{13}} \\ r_{i_{21}} & r_{i_{22}} & r_{i_{23}} \\ r_{i_{31}} & r_{i_{32}} & r_{i_{33}} \end{bmatrix} \quad (3)$$

R_{i2x3} contains the first two rows of the 3x3 rotation matrix R_i. Let n be the number of feature points in a face image. The matrix $C = [X_C, Y_C, Z_C]^T$ is a 3 x n matrix, which represents the 3D coordinates in the adapted face model. X_C, Y_C and Z_C are three nx1 matrices, which are the x-, y-, z-co-ordinates, respectively, of the feature points in the adapted face model. X_C and Y_C are measured from the image being adapted. Marked features of sample image form Bosphorus Database is shown in Fig.1, while Z_C is initially set at the default depth values of the CANDIDE 3D model with a particular scale according to the size of the face image.

The CANDIDE 3D face model is a parameterized face mask specifically developed for the model –based coding of human faces [39]. During the past several decades, candied has been a popular face model used in different face-related applications, because of its simplicity and public availability [21, 40]. The third version of the CANDIDE model, called CANDIDE-3, is composed of 113 vertices and 168 triangular surfaces, as shown in Fig.2. Each vertex is represented by its 3-D coordinates.

p_i is a 2 x n matrix which represents the 2D coordinates of the feature points in the i^{th} non-frontal-view face images. Also, the first row and the second row of p_i represent the x- and y- co-ordinates, respectively.

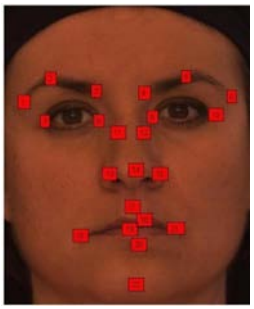

Fig.1 Positions of the 22 features marked on the Face.

Fig.2 Candide-3D Model of Generic Human Face.

If the pose of the face model and the depths of the feature points fit the i^{th} non-frontal-view face images, the following equation will be a minimum:

$$D_{\min} = \min_{s_i, R_{i2x3}} \left\| p_i - s_i R_{i2x3} C - T_i \right\|_2^2 \quad (4)$$

3. Selected Structure from Motion Algorithms

C. Depth Estimation Based on Genetic Algorithm (SM)

In this algorithm three or more face images of the same subject are used to construct a 3D face model. One of them is a frontal view, while the other images are under limited arbitrary poses. To recover the 3D face structure, the 2D frontal-view face image is adapted to the candide model. Then, the pose and the feature-point depths of the candide model are adjusted to fit the poses of the respective 2D non-frontal-view face images in such a way that the feature-point distance between the projected 3D model and the 2D face images under different poses is minimized under the similarity transform.

In this algorithm all the point sets (marked n features as shown in Fig.1) to be compared are translated to their respective centroids so that the centroids become the origin of the coordinate system, and their first moments are zero. Let $M = [X_M, Y_M, Z_M]^T$ be a 3xn matrix which represents the centered model point set. Similarly, suppose that q_i denote a 2xn matrix which represents the centered point sets of the i^{th} image. In other words, q_i and M are the centered point set of p_i and C, respectively, then the translation term T_i can be omitted and the Eq. (4) becomes

$$D_{\min} = \frac{1}{N}\sum \left\| q_i - s_i R_{i2x3} M \right\|^2 \qquad (5)$$

To accomplish the goal of optimal alignment i.e, finding the best pose minimizing eq (5) is a computationally intensive task. The authors of SM algorithm employed Genetic Algorithm to evolve the solution from large searching space. The evolution of the solution vector by the GA depends on how efficiently the chromosome is defined and on the genetic operator's parameters.

1. Chromosome

The chromosome designed for the GA should be able to represent the solution effectively, and its length should be as short as possible. The number of elements in the chromosome is 3N. The chromosome structure is shown below.

Θ_1	Ψ_1	Φ_1	Θ_2	Ψ_2	Φ_2	…	θ_N	Ψ_N	Φ_N

2. The genetic operators.

The genetic operators are selection, crossover, and mutation. These operations are performed to search the optimal poses of the face images and the optimal depths of the face model. Rank selection method is used to select two chromosomes to perform crossover and/or mutation. After selecting two chromosomes, two crossover points are selected randomly. The values between these two crossover points in the two chromosomes are exchanged to form a pair of new offspring.

Mutation is intended to prevent all the solutions in a population falling into a local minimum by exploiting new candidates randomly. The N elements in each chromosome are randomly selected and replace by N randomly generated numbers, where N is the number of non-frontal face images. After estimating the pose values the z-coordinates in M are calculated by applying partial differentiation to eq. (5) and N different z-coordinates are obtained using the equation below:

$$\boldsymbol{Z}_{\boldsymbol{M}}^{\boldsymbol{T}} = \frac{\sum_{i=1}^{n} \boldsymbol{s}_i.\boldsymbol{r}1_i + \boldsymbol{s}_i^2.\boldsymbol{r}_{i_{33}}.\boldsymbol{r}2_i.\boldsymbol{M}_{xy}}{\sum_{i=1}^{n} \boldsymbol{s}_i^2 - \boldsymbol{r}1_i.\boldsymbol{r}1_i^T} \qquad (6)$$

To the best of our knowledge SM algorithm is the first published structure from motion depth estimation algorithm based on Similarity Measurement. The SM algorithm does not require any prior knowledge of camera calibration, and has no limitation on the possible poses or the scale of the face images. In addition the method has been verified that it can be extended to face recognition to alleviate the effect of pose variations. It is reported that the maximum runtime required to generate a face model is about 50 Sec. with a Pentium IV computer system with 2.3 GHz and 512 MB RAM. Unfortunately, the genetic algorithm (GA) used to estimate the depth usually encounters a heavy computational burden. Moreover, how to design a reasonable chromosome, how to make a feasible gene operation scheme, and how to adjust the parameters remain difficult problems.

*D. Depth estimation based on Constrained ICA Model **(cICA)***

3D depth estimation algorithm based on Constrained Independent Component Analysis , introduced Sun, Zhanli, and Kin-Man Lam is an efficient 3D structure estimation algorithm from 2D images. The algorithm is formulated from 2D feature points marked manually/automatic. Denote $\{(q_{xi}, q_{yi})\}_{i=1}^{n}$, n feature points of a non-frontal-view 2-D face q and (M_{xi}, M_{yi}, M_{zi}) represent the i^{th} feature point of a frontal-view 3-D face model **M**. The rotation matrix **R** for **q** is given in Eq. (3). Then the rotation and translation process for mapping the frontal-view face image to the non-frontal-view face image is given by

$$\begin{pmatrix} q_{x_i} \\ q_{y_i} \end{pmatrix} = k\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{pmatrix}\begin{pmatrix} M_{x_i} \\ M_{y_i} \\ M_{z_i} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}_{i=1,\ldots,p} \qquad (7)$$

Where k is the scale factor and (t_x, t_y) are the translation along x and y axes respectively. The matrix form of (7) can be written as follows:

$$q = kR_{2x3}M + t \quad (8)$$

where q is a 2xp matrix such that each column represents the (x, y) co-ordinates $(q_{xi}, q_{yi})^T$ of one feature point, M is a matrix such that each column represents (x, y, z) the co-ordinates $(M_x, M_y, M_z)^T$ of one feature point, and t is a 2xp matrix such that all columns are $(t_x, t_y)^T$.

In terms of the shape-alignment approach [21], the translation term t can be eliminated if both q and M are centered at the origin. Then

$$q = kR_{2x3}M \quad (9)$$

Denote

$$A = kR_{2x3} \quad (10)$$

Equation (9) can be then be written as

$$q = AM \quad (11)$$

It can be seen from (11) that A can be viewed as a mixing matrix and q as a mixture of M. Assuming that the distributions of the variables M_x, M_y and M_z are non-Gaussian, the 3-D structure estimation problem can is a Blind Source Separation (BSS) problem. Under the linear mixture process BSS can be solved using Independent Component Analysis (ICA). The unknown source signals in M=[M_x, M_y, M_z] can be recovered via the ICA algorithm by maximizing the non-Gaussian distribution. There is only one unknown signal in M i.e., M_z the depth information. Therefore, we need only extract M_z by means of the corresponding available reference signal. Based on the above considerations, the authors used constrained Independent Component Analysis (cICA), to estimate the 3-D structure.

Denote y as the estimated signal of M_z, then

$$y = wq \quad (12)$$

where w is the unmixing matrix.

In the cICA algorithm, the negentopy J(y) is used as a contrast function, and the cICA is formulated as a constrained optimization problem as follows:

Min J(y)

s.t. g(y: w) ≤ 0 and h(y: w) = 0.

Here the functions g(y, w) and h(y, w) represent the inequality and equality constraints, respectively. The inequality constraints are the closeness measurements of the estimated output and their corresponding references, and the equality constraints are adopted to eliminate the correlation relationship between any of the two different output components [37]. We can obtain the source signal y by optimizing the objective function (12). The optimization problem in (12) can be solved using Lagrange multipliers as follows:

$$L(w, \mu, \lambda) = J(y)+g(y,w, \mu) + h(y,w, \lambda) \quad (13)$$

Where μ and λ denote the Lagrange multipliers, and g(y, w, μ) and h(y, w, λ) are the terms corresponding to the inequality and equality constraints respectively. In each iteration, the changing of the multipliers μ and λ is given by

$$\Delta \mu = \max\{- \mu, \eta g(y,w)\} \quad (14)$$

and

$$\Delta \lambda = \gamma h(y, w)\} \quad (15)$$

Where η and γ are the learning rates. In [37] the gradient of L with respect to w is given as follows:

$$\Delta_w L = E\{J'(y)x^T\} + \mu\Delta_w g(y) + 4\lambda(E\{y^2\}-1)E(yx^T) \quad (16)$$

Where $g(y) = E\{(y-r)^2\}$, and $\Delta_w g(y)$ is the derivative of g(y) with respect to w. The Newton-like learning rule of w can be given by [37], [38].

$$\Delta w = - \eta(\Delta_w^2 L)^{-1}\Delta_w L \quad (17)$$

The initial unmixing matrix is $W_0 = q^{\dagger}r$. Where $q^{\dagger}$ is the Moore-Penrose generalized inverse of q, r is the reference signal derived from the candied 3D model.

E. Depth estimation based on Non linear least-squares Model

Nonlinear least squares is the problem of finding a set of optimal values of the parameters x = (x_1, x_2, x_3,..,x_k), which minimize the square sum of nonlinear functions $f_i(x)$ (i=1,….,*l*)

$$\min_x \|f(x)\|_2^2 = \min_x (f_1^2(x) + f_2^2(x) + ... + f_l^2(x))$$

Where f(x) is a vector–valued function with component *i* of f(x) equals to $f_i(x)$. The shape features, which are represented by the (x, y) coordinates of the facial feature points, are used in the algorithm to estimate the corresponding depth values i.e., z. Assume that n feature points $\{(q_{xi}, q_{yi})\}_{i=1}^{n}$ are marked on the face images. (M_{xi}, M_{yi}, M_{zi}) represent the i^{th} feature point of a frontal-view 3D face model M, and (q_{xi}, q_{yi}) the i^{th} feature point of a non-frontal view 2D face q. the rotation matrix R for q is given in Eq. (3).

The distance between the feature points, q, of the 2D face image concerned and the corresponding points, M, of the 3D face model can be 3D similarity measurement and is given below:

$$d = \|q - sR_{2x3}M\|_2^2 \quad (18)$$

Where s, is the scale parameter. Denoting the vector x= (ϕ,Ψ,θ,s,M_{z1},…..M_{zn}) as the parameter vector, including both the pose parameters and the depth values of the feature points, the similarity measurement (q-sR_{2x3}M) in Eq. (18) can be rewritten as a vector function, as follows:

$f(x) = (f_1(x),…….f_n(x),f_{n+1}(x),……..f_{2n}(x))^T$

The Parameter ϕ, Ψ, θ, s and depth values M can be obtained by minimizing the distance d

$$\min_x \|f(x)\|_2^2 = \min_x \sum_{i=1}^{2n} f_i^2(x) \quad (19)$$

Therefore, the pose and shape estimation problem is formulated as a NLS model.

The authors of NLS-model incorporated facial symmetry information, optimization regularization term based on linear correlation, efficient model integration method to alleviate the sensitivities arising due to different poses to improve the depth estimation accuracy, different from SM algorithm & cICA.

In the NLS-Model when more than one non-frontal view face image of a subject are available, then the dimension of the objective function and the parameter number in eq. (19) increases linearly with N. This is one of the drawbacks of NLS-Model. To alleviate this increase in dimensionality the authors proposed model integration with a small sacrifice in depth estimation accuracy.

4. Comparison of the three "Structure from Motion" methods

With reference to SM, cICA, and NLS, we note that the discussed structure from motion algorithms share the following similarities.

1. The discussed three methods are based on Similarity Transform, i.e., Orthographic Projection from 3D space to 2D.
2. The three methods require at least one frontal view and one non-frontal view, but do not permit arbitrary facial poses.
3. The three algorithms require 22 feature point indices and initial candide depth values to initialize the 3D reconstruction algorithm.
4. Though different methods are used to estimate the 3D facial shape they are not robust to variation in pose of the sample selected.
5. Storage requirement of the discussed SFM methods is small compared to state-of-art Shape-from-x, 3DMM, Learning 3D shape reconstruction methods.
6. These methods can be used in offline face recognition with the constructed 3D face model of subject of interest.

5. Conclusion and Future Research

Although all the three reconstruction algorithms are based on similarity transform, the depth estimation procedures are different. To begin with SM algorithm employs Genetic Algorithm (GA) for estimating the depth values. Unfortunately, the GA encounters heavy computational burden and it requires many parameters fine tuning which is practically difficult. Different from SM algorithm, in cICA depth information of facial feature points are recovered assuming the orthographic projection as mixing process and the variables Mx, My, Mz as non-gaussian. In cICA Newton-like learning rule is used to estimate the depth values, which is also a time consuming procedure. Finally, in NLS-model, the objective function is minimized using non-linear least squares optimization. The computational complexity of this later method is superior to computationally burden GA & cICA. The authors of this method also provide quantitative analysis of reconstruction error accuracy. Table I. summarizes the qualitative differences between the three structure-from-motion methods. Comparison of the correlation coefficients of the estimated depth values and True depth values as given by the authors in their respective papers are tabulated in Table I. In general, the nonlinear least square modeling and optimization algorithm is fast and while the other two SM and cICA are time consuming and computationally burden.

Finally, yet importantly comparison of the SM, cICA, and NLS-Model is given in Table II. From the Table II it can be observed that all the methods are sample sensitive i.e., the depth estimation accuracy depends on the pose of the face image considered in the objective function, there is no formulated criterion for early stopping of optimization, other optimization regularization terms need to be identified and there should be quantitative analysis of the reconstruction accuracy. All these remain to be done in the future.

Table1.Comparision of Mean and Standard Deviation of the Correlation Coeffecients of discussed SFM Methods.

	μ (Mean)	σ (STD)
SM	0.4920	0.2620
cICA	0.8396	0.0631
NLS1_R_MI	0.9290	0.0313

References

[1] Yuille, Alan L., et al. "Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability." International Journal of Computer Vision 35.3 (1999): 203-222.

[2] Tsalakanidou, Filareti, Sotiris Malassiotis, and Michael G. Strintzis. "Face localization and authentication using color and depth images." Image Processing, IEEE Transactions on 14.2 (2005): 152-168.

[3] Wang, Yueming, Jianzhuang Liu, and Xiaoou Tang. "Robust 3D face recognition by local shape difference boosting." Pattern Analysis and Machine Intelligence, IEEE Transactions on 32.10 (2010): 1858-1870.

[4] Gupta, Shalini, Mia K. Markey, and Alan C. Bovik. "Anthropometric 3D face recognition." International Journal of Computer Vision 90.3 (2010): 331-349.

[5] Gonzalez-Mora, Jose, et al. "Bilinear active appearance models." (2007).
[6] Bowyer, Kevin W., Kyong Chang, and Patrick Flynn. "A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition." Computer vision and image understanding 101.1 (2006): 1-15.
[7] Thelen, Andrea, et al. "Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation." IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 18.1 (2009): 151-157.
[8] Castelán, Mario, and Edwin R. Hancock. "Acquiring height data from a single image of a face using local shape indicators." Computer Vision and Image Understanding 103.1 (2006): 64-79.
[9] Castelán, Mario, and Edwin R. Hancock. "A simple coupled statistical model for 3d face shape recovery." Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 1. IEEE, 2006.
[10] Jiang, Dalong, et al. "Efficient 3D reconstruction for face recognition." Pattern Recognition 38.6 (2005): 787-798.
[11] Romdhani, Sami, and Thomas Vetter. "Efficient, robust and accurate fitting of a 3D morphable model." Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003.
[12] Zhang, Chongzhen, and Fernand S. Cohen. "3-D face structure extraction and recognition from images using 3-D morphing and distance mapping." Image Processing, IEEE Transactions on 11.11 (2002): 1249-1259.
[13] Song, Mingli, et al. "Three-dimensional face reconstruction from a single image by a coupled RBF network." Image Processing, IEEE Transactions on 21.5 (2012): 2887-2897.
[14] Castelán, Mario, Gustavo A. Puerto-Souza, and Johan Van Horebeek. "Using subspace multiple linear regression for 3D face shape prediction from a single image." Advances in Visual Computing. Springer Berlin Heidelberg, 2009. 662-673.
[15] Li, Annan, et al. "Recovering 3D facial shape via coupled 2D/3D space learning." Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on. IEEE, 2008.
[16] Tomasi, Carlo, and Takeo Kanade. Shape and motion from image streams: a factorization method: full report on the orthographic case. Cornell University, 1992.
[17] Fortuna, Jeff, and Aleix M. Martinez. "Rigid structure from motion from a blind source separation perspective." International journal of computer vision 88.3 (2010): 404-424.
[18] Bregler, Christoph, Aaron Hertzmann, and Henning Biermann. "Recovering non-rigid 3D shape from image streams." Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. Vol. 2. IEEE, 2000.
[19] Torresani, Lorenzo, Aaron Hertzmann, and Christoph Bregler. "Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors." Pattern Analysis and Machine Intelligence, IEEE Transactions on 30.5 (2008): 878-892.
[20] Akhter, Ijaz, et al. "Trajectory space: A dual representation for nonrigid structure from motion." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.7 (2011): 1442-1456.
[21] Koo, Hei-Sheung, and Kin-Man Lam. "Recovering the 3D shape and poses of face images based on the similarity transform." Pattern Recognition Letters 29.6 (2008): 712-723.
[22] Lee, Kyoung Mu, and C-CJ Kuo. "Shape from shading with a linear triangular element surface model." Pattern Analysis and Machine Intelligence, IEEE Transactions on 15.8 (1993): 815-822.
[23] Horn, Berthold KP. "Height and gradient from shading." International journal of computer vision 5.1 (1990): 37-75.
[24] Zheng, Qinfen, and Rama Chellappa. "Estimation of illuminant direction, albedo, and shape from shading." Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on. IEEE, 1991.
[25] Zhang, Ruo, et al. "Shape-from-shading: a survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 21.8 (1999): 690-706.
[26] Blanz, Volker, and Thomas Vetter. "Face recognition based on fitting a 3D morphable model." Pattern Analysis and Machine Intelligence, IEEE Transactions on 25.9 (2003): 1063-1074.
[27] Castelán, Mario, and Johan Van Horebeek. "3D face shape approximation from intensities using Partial Least Squares." Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008.
[28] Nandy, Dibyendu, and Jezekiel Ben-Arie. "Shape from recognition: a novel approach for 3-D face shape recovery." Image Processing, IEEE Transactions on 10.2 (2001): 206-217.
[29] Ullman, Shimon. "The interpretation of visual motion." (1979).
[30] Huang, Thomas S., and C. H. Lee. "Motion and structure from orthographic projections." IEEE Transactions on Pattern Analysis and Machine Intelligence 11.5 (1989): 536-540.
[31] Hu, Xiaoping, and Narendra Ahuja. "Motion estimation under orthographic projection." Robotics and Automation, IEEE Transactions on 7.6 (1991): 848-853.
[32] Shapiro, Larry S., Andrew Zisserman, and Michael Brady. "3D motion recovery via affine epipolar geometry." International Journal of Computer Vision 16.2 (1995): 147-182.
[33] Xirouhakis, Yiannis, and Anastasios Delopoulos. "Least squares estimation of 3D shape and motion of rigid objects from their orthographic projections." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.4 (2000): 393-399.
[34] Sun, Zhanli, and Kin-Man Lam. "Depth estimation of face images based on the constrained ICA model." Information Forensics and Security, IEEE Transactions on 6.2 (2011): 360-370.
[35] Sun, Zhan-Li, Kin-Man Lam, and Qing-Wei Gao. "Depth estimation of face images using the nonlinear least-squares model." Image Processing, IEEE Transactions on 22.1 (2013): 17-30.
[36] Werman, Michael, and Daphna Weinshall. "Similarity and affine invariant distances between 2d point sets." Pattern

Analysis and Machine Intelligence, IEEE Transactions on 17.8 (1995): 810-814.
[37] Huang, De-Shuang, and Jian-Xun Mi. "A new constrained independent component analysis method." Neural Networks, IEEE Transactions on 18.5 (2007): 1532-1535.
[38] Lu, Wei, and Jagath C. Rajapakse. "ICA with reference." Neurocomputing 69.16 (2006): 2244-2257.
[39] Ahlberg, Jörgen. "An active model for facial feature tracking." EURASIP Journal on applied signal processing 2002.1 (2002): 566-571.
[40] Xie, Xudong, and Kin-Man Lam. "Elastic shape-texture matching for human face recognition." Pattern Recognition 41.1 (2008): 396-405.

Table 2.Comparision of Three Recent 3D Facial Shape Reconstruction Based on Structue from motion.

		3D Face Reconstruction – Structure-From-Motion Methods			
		SM	***cICA***	***NLS-Model***	***Remarks***
Min. No. of Input Images N		N=2	N=2	N=2	One Frontal One Non-Frontal
Initalizing	Facial Features	22 Facial Features	22 Facial Features	22 Facial Features	Fig.1
	Initial Depths	Candide Face Model	Candide Face Model	Candide Face Model	Fig.2
Optimization		GA	BSS	Non-linear Least Squares Optimization	
Texture Recovery		No	No	No	Separately Captured
Sample Sensitivity		Yes	Yes	Yes	
Efficiency Training Time.		~ 50Sec.	< 10Sec.	< 1Sec.	
Quantitative Analysis of Reconstruction Accuracy		No	No	Yes	
Face Recognition		Based On Reconstructed 3D Model	Based On Reconstructed 3D Model	Based On Reconstructed 3D Model	