A Robust System for 3D Face Reconstruction using Multi-Stage Linear Approach

Divyapreethi.A.K. [1+] and A.Srinivasan [2]

[1] Department of Computer Science and Engineering, MNM Jain Engineering College, Thoraipakkam, Chennai.

[2] Head of the Department of Information Technology, MNM Jain Engineering College, Thoraipakkam, Chennai.

Abstract. Human head is three-dimensional object in 3D space with variations in position and in its structure. Consequently, 3D Face Modeling is largely acknowledged in Face Recognition application for Non-cooperative subjects. Structure from Motion (SfM) is a 3D Reconstruction technique to compute a 3D Facial shape by means of multiple 2D image sequence. Due to Self-Occlusion, this technique is susceptible to point correspondence error reducing its performance. To eliminate point correspondence error a Matrix called Shape Conversion Matrix (SCM) is used to obtain the true location of Self-Occluded Facial Feature Points (FFPs). In the proposed system, a new Structure from Motion method called Multi-stage linear approach is adapted. A novel Face alignment algorithm called Robust Alignment by Sparse and Low-rank decomposition (RASL) is incorporated with the system. More resourceful feature extraction techniques are amended. They are Gabor wavelet and Dual-tree Complex wavelet transform. A Generalized Polycube Trivariant Spline based 3D dense mean model adaptation is integrated. By applying these models, a proficient framework for robust 3D Face Reconstruction for self-occlusion is proposed.

Keywords: self-occlusion, 3d face reconstruction, structure from motion, 3d face modelling, multi-stage linear approach.

1. Introduction

Face Recognition is an application to biometric system for identification and verification of faces. A Face Recognition system is supposed to be able to identify or recognize a non-cooperative face in uncontrolled environment and an arbitrary situation without the notice of the subject. This generality of environment and situations, however, brought serious challenges to face recognition techniques, e.g., the appearances of a face, due to viewing condition changes may vary too much to tolerate or handle.

In the current trend, 3D face model is a very useful technology in computer vision for its various applications, including pose-invariant face recognition, illumination-invariant face recognition, age-invariant face recognition, 3D game and movie character generation, teleconferencing, surgical simulation, etc.

To model a 3D face requires special hardware or multi-cameras, such as a 3D laser scanner, stereo cameras, or a structured light. However, applications using this approach are restrictive because of high costs and camera calibration. As an alternative, 3D face reconstruction methods by using an image sequence have been researched intensively.

These methods can be categorized into two types.

- Model-based methods
- Structure from Motion-based methods

[+] Corresponding author. Tel.: +919962047166.
E-mail address: divyapreethiak@gmail.com.

1.1. Model-Based methods

Model-based methods build 3D face models to represent facial shape and texture, illumination, and camera geometry, with a large number of model parameters [1]. When 2D facial images are given as input, the methods find the model parameters of the 3D face model iteratively in order to minimize the texture residual between the 2D facial image synthesized from the model parameters and the input 2D facial image. If the optimal model parameters are found, a detailed 3D facial shape can then be reconstructed from these model parameters.

1.2. Structure from motion-based methods

Structure from Motion-based method estimate a 3D facial shape and a projection matrix from the corresponding 2D FFPs from multiple facial images [2]. The 2D FFPs from multiple images are factorized into 3D facial shape. It does not require parameter optimization like model-based method. SfM-based methods can be categorized into dense correspondence-based methods and sparse correspondence-based methods according to the density of the corresponding 2D FFPs.

- **Dense correspondence-based methods** find dense corresponding 2D FFPs, and reconstruct a dense 3D facial shape from them. If dense corresponding 2D FFPs can be correctly found, these methods can reconstruct a detailed 3D facial shape. However, in general, it is difficult to find dense corresponding 2D FFPs, since some facial regions, such as the cheek and the forehead have no salient texture patterns.
- **Sparse correspondence-based methods** find a predetermined number of sparse and salient corresponding 2D FFPs, and reconstruct a sparse 3D facial shape from them. A dense 3D mean face is then adapted to the reconstructed sparse 3D facial shape.

2. Related Works

Self-Occlusion is defined as occlusion of facial region by another facial region. The main cause of self-occlusion is due to different orientation of face image which act as input to surveillance system. Due to this, the performance of face modelling of the system is reduced. Unsung Park and Anil K. Jain [3] used the structure from motion for pose-invariant face recognition system in which the self-occlusion is considered noise in the facial feature point detection.

Manuel Marques and Joan Costeria [4] considered self-occlusion as missing data and proposed a new factorization method which computes the exact shape iteratively. The matrix completion method such as Damped Newton algorithm [5] and Fixed point and Bregman iterative method [6] estimate the self-occluded FFPs by minimizing the shape residual of the visible 2D FFPs. Nevertheless, they are affected by reduced number of useful 2D FFPs.

This paper is organised as follows. Section 3 explains the proposed system work flow which reconstructs 3D face, estimate SCM, eliminating the self-occlusion using SCM in the reconstructed 3D face and dense model adaptation on 3D face with texture mapping. Section 4 describes results and discussion of the proposed system.

3. Proposed System

3.1. 3D face reconstruction

A new proficient system for 3D face reconstruction is proposed and figure 1 describes the system development. Facial feature Points are the 2D points that indicate end point, corners and ridges of the face collectively representing the salient features like eyes, nose and facial contours. Given F images $I_1, I_2, \ldots \ldots \ldots, I_F$ are captured with different orientation for a Subject and their FFPs are extracted using Gabor Wavelet [7]. Gabor wavelet has ability in removing most of the variability in image due to variation in lighting and contrast. Furthermore they are robust against small shifts and deformations. The FFP are localized to compute a measurement matrix W.
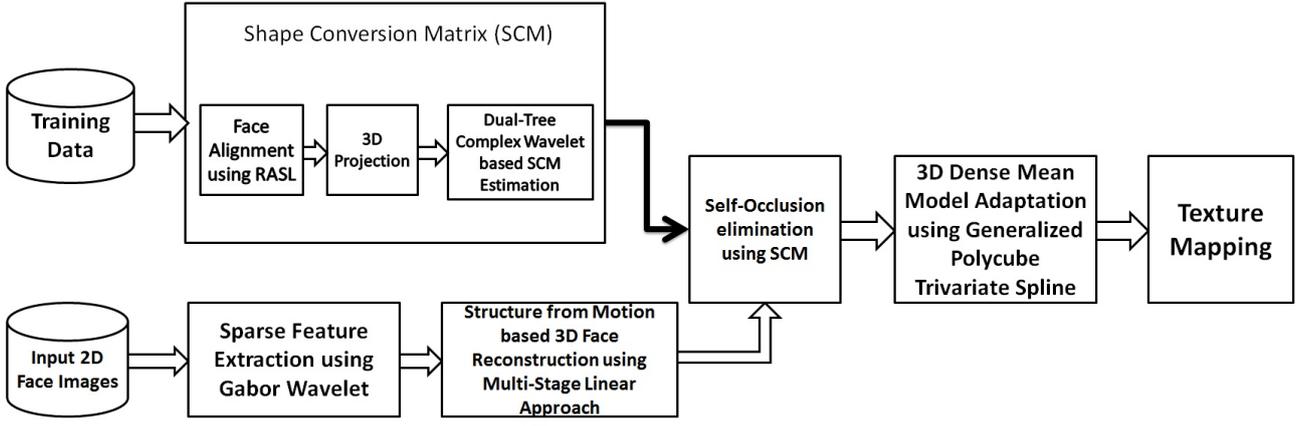
Fig. 1: System Architecture

Structure from Motion technique find the Facial shape **S** and Projection matrix **P** using the measurement matrix W. Multi-stage linear approach [8] estimate the 3D facial shape by, first performing pairwise reconstructions, then robustly aligning these in pairs, and finally aligning all of them globally by simultaneously estimating their unknown relative scales and translations is integrated in the system. The algorithm treats all input images equally. It efficiently copes with substantial outliers and fast. It is also easy to parallelize and removes the need for frequent bundle-adjustment.

3.2. Training phase for SCM estimation

The training phase estimates the Shape Conversion Matrix for self-occlusion elimination and figure 2 explains progression of SCM valuation. The training Face images are aligned using RASL: Robust Alignment by Sparse and Low-rank Decomposition [9]. RASL method seeks an optimal set of image domain transformations such that the matrix of transformed images can be decomposed as the sum of a sparse matrix of errors and a low-rank matrix of recovered aligned images. It doesn't require pre-filtering or matching and act directly on input images. It is tractable, scalable and simultaneously aligns hundreds of image in less time by solving a sequence of convex optimization problems.

The images are then projected into 2D plane with different orientations. The Facial features are extracted using Dual-tree Complex Wavelet Transform [10] for estimating Ground-truth and Observed facial feature points. Dual-tree Complex Wavelet Transform, novel wavelets transform exhibit approximate shift invariant property and improved angular resolution. DT-CWT employs two real wavelet transforms in parallel where the wavelets of one branch are the Hilbert transforms of the wavelet in the other and generate six sub-bands with complex coefficients in different directionalities. At that point, 2D Facial shapes are produced from Ground-truth and Observed feature points. From 2D Facial shape, 2D face models are computed for both Ground-truth and Observed facial shape. The dominant component of the observed shape parameters were linearly correlated with those of the ground-truth shape parameter. This association is used to obtain a discrepancy matrix called SCM. SCM is used to find the self-occluded FFPs in the reconstructed 3D face shape. The Input observed shape parameter is multiplied with Shape Conversion Matrix (SCM) [11] to obtain the true location of Facial Feature Points (FFPs).

$$\alpha_v = C\beta_v$$

Now C, the SCM is calculated as,

$$C = [\alpha_1 \dots \alpha_v]([\beta_1 \dots \beta_v]^T[\beta_1 \dots \beta_v])^{-1}[\beta_1 \dots \beta_v]^T$$

Where v is total number of training projected images, α and β are Ground-truth and Observed shape parameter respectively.
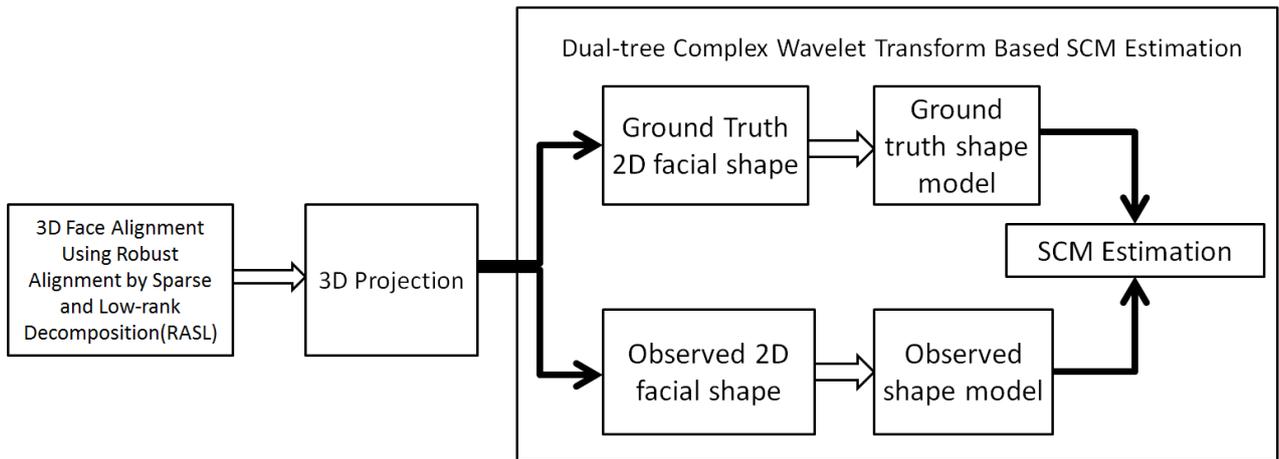
Fig. 2: SCM Estimation.

3.3. Dense model adaptation and texture mapping

3D Dense mean model is adapted on the sparse 3D facial shape to obtain a smooth and dense 3D face. Generalised Polycube Trivariate Spline [8] a global one-piece parametric representation for shape with nontrivial topology that prevents artefacts caused by stitching or trimming over shared boundaries of solid primitives is used. A Global trivariate T-splines can be constructed on Generalized PolyCube (GPC) and transition functions can be effectively computed using the GPC graph. The entire spline construction framework affords hierarchical refinement and level-of-detail control. The Texture in the input images is mapped to the reconstructed 3D dense facial shape to make the textured 3D facial shape.

4. Results and Discussion

The features from input 2D face images are extracted and aligned. A 3D Face model is reconstructed from the features of input 2D images. Using SCM, self-occlusion will be eliminated and a dense mean model will be adapted on 3D Face Model with the texture of input 2D images. The result of the system is compared with suitable databases based on Root Mean Square (RMS) error value obtained in both training and test data.

5. Conclusion

A new framework is proposed for a robust 3D Face Reconstruction for Self-Occlusion efficiently. The system estimate Shape Conversion Matrix (SCM) to obtain the true location of Self-Occluded Facial Feature Points (FFPs). The system incorporates original and efficient algorithms, thus a proficient system for eliminating point correspondence error caused by Self-Occlusion for robust 3D Face Reconstruction is proposed and the results will be compared with existing system.

6. References

[1] Volker Blanz, Thomas Vetter, "A Morphable model for the synthesis of 3D faces", SIGGRAPH conference proceeding, 1999.

[2] Unsang Park, Anil K. Jain, "3D Model-based face recognition in video", 2nd International conference on biometrics, 2007.

[3] Unsung Park, Anil K. Jain, "Face Recognition in video: Adaptive fusion of multiple matcher", proceeding in IEEE Computer society workshop on Biometrics, CVPRW 2007.

[4] Manuel Marques and Joao Costeria, "Estimating 3D Shape from degenerate sequence with missing data", Journal in Computer vision and image understanding, 2009.

[5] A.M.Buchanan and A.W.Fitzgibbon, "Damped newton algorithm for matrix factorization with missing data", CVPR 2005, Computer vision and pattern recognition, IEEE Computer society.

[6] Shiqian Ma, Donald Goldfarb, and Lifeng Chen, "Fixed point and Bregman iterative methods for matrix rank minimization", Journal in Mathematics programming: Serial A and B, 2011.

[7] LinLin Shen, Li Bai, Michael Fairhurst, "Gabor Wavelet and General Discriminant Analysis for face identification and verification", Image and vision computing 25 553-563, 2007.

[8] Sudipta N. Sinha, Drew Steedly, and Richard Szeliski, "A multi-stage linear approach to structure from motion", ECCV 2010 Workshop on Reconstruction and Modeling of Large-Scale 3D Virtual Environments.

[9] Yigang Peng, Arvind Ganesh, John Wright, Yi Ma, "RASL: Robust Alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images", Computer Vision and Pattern Recognition (CVPR), 2010.

[10] Turgay Celik, Huseyin Ozkaramanh, Hasan Demirel, "Facial feature extraction using complex dual-tree wavelet transform", Computer vision and image understanding 111 229-246, 2008.

[11] Sung Joo Lee, Kang Ryoung Park, Jaihie Kim, "A SfM-based 3D face reconstruction method robust to self-occlusion by using a Shape Conversion Matrix", Pattern recognition, 2011.

[12] Bo Li, Xin Li, Kexiang Wang, Hong Qin, "Generalized PolyCube Trivariate Splines", Shape Modeling International Conference (SMI), 2010.