Improving 3D Face Details based on Normal Map of Hetero-source Images

Chang Yang, Jiansheng Chen, Nan Su and Guangda Su
Department of Electronic Engineering, Tsinghua University
Beijing, 100084, China
yangchang11@mails.tsinghua.edu.cn, jschenthu@tsinghua.edu.cn
v377026@sina.com, susu@tsinghua.edu.cn

Abstract

For each person, there exist large unstructured photo collections in personal photo albums. We call these photos Hetero-source images, which imply abundant shape and texture information of the specific face. In this paper, we propose a novel 3D face modeling method combining the normal map of Hetero-source images with the fitting result based on a single image to achieve more accurate 3D shape estimates. Based on recent research showing that the set of images of convex Lambertian surfaces under general illumination can be well approximated using low-order spherical harmonics, we first incorporate spherical harmonics into the 3D morphable model to initialize the 3D shape. The fitting result, however, suffers from model dominance and lacks of fine details. The normal map inferred by Hetero-source image shading constraints allows the possibility of improving local details and challenging the model dominance. We estimate the normal map which contains more accurate orientation information in an alternating optimization way and apply it to improve the preliminary 3D surface. Experimental results on both synthetic and real world data demonstrate that our method could be used to capture discriminating facial features and outperforms the single image fitting result in accuracy.

1. Introduction

3D face modeling provides solutions for a wide range of applications as pose/illumination invariant face recognition, public security and human computer interactions. It has received a continual development over the past two decades and a considerable body of researchers is still working on the challenging parts of this task [1][6] [9][8][13].

Statistical model based 3D reconstruction methods have achieved great success in solving the problem of recovering the shape and texture of unseen subjects. Blanz and Vetter propose a generative 3D morphable model (3DMM) with potential of recovering highly satisfactory face models [5]. They minimize the texture discrepancy between the input and synthesized images to get the model parameters based on the Phong reflectance model. For the cases roughly lying within the linear subspace spanned using the training data, the method performs relatively well. To achieve better performance, Romdhani adds image features as edges and specular highlights to formulate a smooth cost function [17]. Brian et al. [1] develop a 3DMM based stereo system, which uses stereo information of multiple images taken simultaneously to improve accuracy and robustness. However, the increase of accuracy yields a more computationally expensive search. Moreover, those measures could not essentially solve the problem of model dominance and the reconstruction quality is entirely dependent on whether a similar subject is included in the training set [13].

Photometric stereo based methods [7] focus on image shading cues and explores images themselves. For instance, Georghiades et al. [6] propose an illumination cone based method which recovers accurate 3D face using 7 well illuminated and registered images. For unknown illuminations, there exists a generalized bas-relief (GBR) ambiguity [3] requiring the orientations for key positions to be set forehead. Some researchers are committed to resolving this ambiguity by iteratively estimating light positions [10] or using template model as constraints [9]. Despite the GBR ambiguity, photometric stereo provides high-frequency information that 3DMM based methods lack.

For such algorithms, images were acquired with special instruments or under controlled conditions. What we take into consideration is recovering 3D face from large unstructured image collections. We call them Hetero-source images, which show extreme variations in pose, illumination and facial expression. On the other hand, those images involve abundant information which may contribute to 3D modeling if used properly. Unfortunately, traditional reconstruction methods as illumination cone [6] cannot be directly applied to the problem. To our knowledge, only [9] attempts to address such a question, which recovers a model that is locally consistent with the photo collection. In this

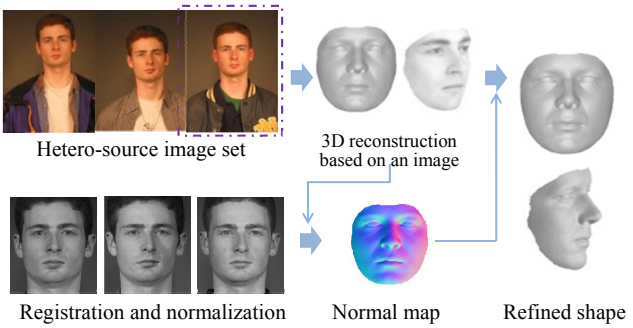


Figure 1. Overview of our approach, which consists of three major parts: 3D face reconstruction, normal map estimation and shape improvement.

work presented herein, we propose a method sharing similar thoughts to [19], which combines multiview and photometric stereo for accurate 3D reconstruction. We build on our approach by fusing the 3DMM fitting algorithm based on a single image and the normal map of Hetero-source images. It maintains the high-frequency details and avoids low-frequency bias using normal and position information. As shown in Figure 1, the input of our method is a Hetero-source image set which consists of faces under variant illuminations. There exist no facial expression or pose variation among the set for illumination is our main concern. Additionally, it is an easy task to find sufficient faces of near frontal pose in the image set. The 3D shape is initialized by fitting the 3DMM to a single image. Then we adopt an iterative optimization method based on classic photometric stereo and initial shape prior to estimate the normal map, which is subsequently applied to improve the preliminary 3D geometry. In summary, our main contributions are as follows:

- exploring a new approach which takes advantage of a broad range of personal photo collections.
- combining the 3DMM fitting algorithm with the photometric stereo to achieve more reliable 3D models.

The rest of the paper is organized as follows. Section 2 reviews the modified LiST algorithm [16] and section 3 iteratively estimates the unknown illuminations and normal map of Hetero-source images. New normals are used to improve the preliminary 3D model based on the method in [12] (Section 4). Section 5 demonstrates experimental results on Basel Face Model (BFM) [14] and FRGC2.0 database [15]. The last section concludes our work.

2. 3D morphable model fitting

The 3D morphable model (3DMM) [5] is a widely used statistical model with potential to recover high quality 3D shape. In the first subsection, we will briefly describe the 3DMM and our modified LiST fitting algorithm based on Spherical Harmonic Lighting model [2]. In the following subsections, we detail the iterative optimization framework to recover model parameters from the input.

2.1. 3D morphable model

The idea of 3DMM is based on the assumption that human faces are within a linear subspace. Blanz and Vetter used 200 3D face meshes in full correspondence to establish eigenvectors of the subspace [5]. Any novel face shape and texture can be represented as

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{S}\alpha, \mathbf{t} = \bar{\mathbf{t}} + \mathbf{T}\beta \tag{1}$$

in which $\mathbf{S}$ and $\mathbf{T}$ denote the PCA matrices formed by stacking the scaled eigenvectors; vector α and β stand for the shape and texture coefficients. The common practice for fitting the model is to minimize the discrepancy between the input image I_{input} and the synthesized one I_{model} rendered with α, β and $\mathbf{p}$ ($\mathbf{p}$ is rendering parameter vector and consists of rotation, scale, translation and illumination) :

$$\min_{\alpha,\beta,\mathbf{p}} \delta_I = \sum_{x,y} ||I_{input}(x,y) - I_{model}(x,y)||^2 \tag{2}$$

It has been demonstrated that the set of images of a convex Lambertian object under a wide variety of lighting conditions can be well approximated by a low-order spherical harmonic basis [2]. We modify the LiST fitting algorithm [16] based on Phong Lighting model to build our LiST based on Spherical Harmonic Lighting model. Equations 3 4 5 summarize the process of transforming the shape and texture vectors, $\mathbf{s}$ and $\mathbf{t}$, into a gray level image using the first nine spherical harmonic basis:

$$\mathbf{s}_{3d} = R(\bar{\mathbf{s}} + \mathbf{S}\alpha) \Rightarrow \mathbf{n} = (n_x, n_y, n_z) \tag{3}$$

$$\mathbf{s}_{2d}^{model} = fP\mathbf{s}_{3d} + \mathbf{t}_{2d} \tag{4}$$

$$I_{model} \approx \rho .* \sum_{i=1}^{9} h_i(\mathbf{n}) l_i, \rho = \begin{bmatrix} 0.30 & 0 & 0 \\ 0 & 0.59 & 0 \\ 0 & 0 & 0.11 \end{bmatrix} (\bar{\mathbf{t}} + \mathbf{T}\beta) \tag{5}$$

Here R denotes the rigid rotation transformation and P represents the orthogonal projection. After the projection a scaling factor f and a 2D translation $\mathbf{t}_{2d}$ are applied. ρ is the albedo vector, which is approximated by transforming the R, G, B color values of model texture $\mathbf{t}$ to gray value. $h_i(\mathbf{n})$ and l_i ($i = 1, 2, \cdots, 9$) represent the 9 spherical harmonics and their coefficients respectively. Furthermore, the first nine harmonic image set Y has the form of $(\rho .* h(\mathbf{n}))^T$ where the operator $.*$ denotes the component-wise product:

$$\begin{aligned} Y &= (\rho .* h(\mathbf{n}))^T \\ &= (\rho .* (\mathbf{1}, \mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z, \mathbf{n}_{xy}, \mathbf{n}_{xz}, \mathbf{n}_{yz}, \mathbf{n}_{x^2} - \mathbf{n}_{y^2}, 3\mathbf{n}_{z^2} - \mathbf{1}))^T \end{aligned} \tag{6}$$

where $\mathbf{n}_x$, $\mathbf{n}_y$ and $\mathbf{n}_z$ are the components of all vertex surface normals and $\mathbf{n}_{xy}$ is defined as $\mathbf{n}_x . * \mathbf{n}_y$ (similarly for $\mathbf{n}_{x^2}$, $\mathbf{n}_{y^2}$, $\mathbf{n}_{z^2}$, $\mathbf{n}_{xz}$, $\mathbf{n}_{yz}$).

Directly minimizing the cost function defined with intensity distance as Eq. 2 amounts to solving a highly complicated and nonlinear optimization problem. The LiST fitting algorithm approximates the non-linear optimization problem using two linear error parts: the shape and texture error functions linearly depending on the shape and texture parameters respectively [16]. The method alternately updates one parameter while maintaining the others constant in an iterative way. Similarly, we build our Spherical Harmonic Lighting model based LiST algorithm as follows.

2.2. Shape fitting

Instead of recovering shape parameters from δ_I, updating α from shape errors is adopted [16]. However, as the image does not contain any shape information, the shape error $\delta \mathbf{s}_{2d}$ is estimated by applying an optical flow algorithm [20] between the input and synthesized images. We denote the shape recovered by optical flow on the input by $\mathbf{s}_{2d}^{img}$.

Rotation R, Translation $\mathbf{t}_{2d}$ and Scale f parameters update. To account for variations in pose, we first calculate pose parameters using automatically located feature points of the input [11] and corresponding points on the 3DMM. Let q and Q denote the point sets on the image and the model respectively. We attempt to seek the R, f and $\mathbf{t}_{2d}$ such that

$$\min_{R,s,\mathbf{t}_{2d}} ||(fRQ + \mathbf{t}_{2d}) - q||^2 \tag{7}$$

In our work, we adopt the Least-Squares Rigid Motion Using SVD [18] to recover the pose parameters and set them as start values. In subsequent iterations, the minimization is performed on recovered shape $\mathbf{s}_{2d}^{img}$ and corresponding model points to avoid the feature point location errors.

Shape Parameters α Update. Given the newly updated pose parameters, shape error $\delta \mathbf{s}_{2d}$ is equal to the difference between the shape $\mathbf{s}_{2d}^{img}$ and the model shape $\mathbf{s}_{2d}^{model}$. While rotation and scale parameters remain constant, the relation between $\delta \mathbf{s}_{2d}$ and α is linear as

$$\mathbf{s}_{2d}^{img} - \mathbf{s}_{2d}^{model} = \delta \mathbf{s}_{2d} = fPR\mathbf{S}\delta\alpha = A\delta\alpha \tag{8}$$

By solving the over-constrained linear system of equations, shape parameter vector α is updated in a single step.

2.3. Texture fitting

Illumination Coefficients $\boldsymbol{l}$ update. We first estimate the illumination $\boldsymbol{l}$ by finding the best coefficients that fit the current model to the input:

$$\mathbf{min} \sum ||\boldsymbol{I_{input}} - \boldsymbol{\rho} . * \sum \boldsymbol{h_i}(\mathbf{n}) \boldsymbol{l_i}||^2 \tag{9}$$

Since all the other parameters are kept constant, this is a highly over-constrained linear system with only 9 unknowns. It has been demonstrated in research [8] that the error of recovering lighting by even using the 3D face of a different individual is sufficiently small.

Texture Parameters $\boldsymbol{\beta}$ Update. At this stage, shape and rendering parameters are maintained constant. Equation 2 can be simplified as

$$\boldsymbol{\delta I} = \left[\begin{smallmatrix} \mathbf{0.30} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.59} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0.11} \end{smallmatrix} \right] \mathbf{T} . * \sum_{\boldsymbol{i=1}}^{\mathbf{9}} \boldsymbol{h_i}(\mathbf{n}) \boldsymbol{l_i} \boldsymbol{\delta\beta} = \boldsymbol{B\delta\beta} \tag{10}$$

Similar to the update of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ is acquired in a single step by calculating the pseudo-inverse of $\boldsymbol{B}$. In practice, however, with the existence of noise and nonlinearity, minimizing Eq. 8 and Eq. 10 by directly calculating pseudo-inverse may cause over-fitting and lead to a result far from the real face [4]. We instead turn to solving the following L2 constrained Least Squares problem, in which the tradeoff between fitting quality and plausibility can be controlled by a constant $\boldsymbol{\eta}$ [16].

$$\begin{aligned} \min_{\boldsymbol{\delta\alpha}} (||\boldsymbol{\delta \mathbf{s}_{2d}} - \boldsymbol{A\delta\alpha}||^2 + \boldsymbol{\eta}||\boldsymbol{\alpha_{cur}} + \boldsymbol{\delta\alpha}||^2) \\ \min_{\boldsymbol{\delta\beta}} (||\boldsymbol{\delta I} - \boldsymbol{B\delta\beta}||^2 + \boldsymbol{\eta}||\boldsymbol{\beta_{cur}} + \boldsymbol{\delta\beta}||^2) \end{aligned} \tag{11}$$

The whole iteration procedure starts from the pose parameters update, and ends when the image intensity discrepancy $\boldsymbol{\delta I}$ obtained after this iteration is less than a threshold. To make the process less computationally expensive, the algorithm randomly performs on a subset of all visible vertices. Compared to the original LiST in [16], our method automatically initialize the pose parameters without requiring manual interactions. Moreover, the optimization steps are all linear benefited from the Spherical Harmonic Lighting model. The resulting shape and texture vector are used as initial values of our surface improvement algorithm. On average, the reconstruction gets converged within 6 iterations and costs about 40s on a 2.5GHz Intel Core i5-3210M CPU.

3. Normal map

At this stage, we have obtained a 3D shape which is globally accurate but lacks of fine local details due to the problem of model dominance [13]. In this section, we will present how the surface orientation inferred using Heterosource image shading cues can be used to generate a more precise 3D model. Generally, this technique offers the possibility of increasing the accuracy of current fitting algorithms since it is independent from the single image based reconstruction.

Consider an image set consisting of $\boldsymbol{n}$ images taken with different cameras and under different illuminations. In the

first step, we attempt to recover illuminations using normals and albedo initialized by the 3DMM fitting result. In consideration of the linear relation between normal components and spherical harmonics, we take the first-order approximation to formulate the over-constrained linear least squares optimization with only four unknowns for each illumination:

$$\min \sum ||\boldsymbol{I}_{n\times m} - \boldsymbol{L}_{n\times 4}\boldsymbol{Y}_{4\times m}|| \quad (12)$$

in which $\boldsymbol{I}_{n\times m}$ stands for $\boldsymbol{n}$ images with $\boldsymbol{m}$ valid pixels each; $\boldsymbol{L}_{n\times 4}$ represents the $\boldsymbol{n}$ illumination conditions and $\boldsymbol{Y}_{4\times m}$ stands for the first 4D spherical harmonic space.

Given the newly updated illumination coefficients, we now attempt to recover the spherical harmonic basis and albedo. For every visible vertex $\boldsymbol{j}$ we choose different subsets by evaluating how well the $\boldsymbol{n}$ images fit the current shape, which works as reference [9]. For point $\boldsymbol{j}$, the distance between the input and synthesized image intensities $||\boldsymbol{I}_j - \boldsymbol{L}_{n\times 4}\boldsymbol{Y}_j||$ are calculated where $\boldsymbol{I}_j$ is a vector denoting the intensities of a pixel in all images (column of $\boldsymbol{I}_{n\times m}$) and $\boldsymbol{Y}_j$ represents the harmonic images (column of $\boldsymbol{Y}_{4\times m}$). Those $\boldsymbol{k}$ pixels whose distances are less than a threshold are chosen as the subset. In this way, our method can suppress effects of cast shadows and noise to some extent. The final error function of two terms is defined using the selected image subset as

$$\min_{\boldsymbol{Y}_j} ||\boldsymbol{I}_{k\times 4} - \boldsymbol{L}_{k\times 4}\boldsymbol{Y}_j||^2 + \kappa||\boldsymbol{Y}_j - \boldsymbol{Y}_{jcur}||^2 \quad (13)$$

where the first term represents the lighting consistency relation and the second acts as a regularization avoiding large departures from the current value. The penalizing parameter $\boldsymbol{\kappa}$ is set as 1 and decreases by 0.2 for every iteration ($\boldsymbol{\kappa}$ is no smaller than 0). The solution to this minimization problem is

$$\boldsymbol{Y}_j = (\boldsymbol{L}_{k\times 4}^T\boldsymbol{L}_{k\times 4} + \kappa\boldsymbol{E})^{-1}(\boldsymbol{L}^T\boldsymbol{I}_{k\times 1} + \boldsymbol{Y}_{jcur}) \quad (14)$$

in which $\boldsymbol{E}$ stands for a 4×4 identity matrix. The first element of $\boldsymbol{Y}_j$ is the new albedo and the other three comprise the new normal vector $\mathbf{n_j}^*$, creating the normal map of Hetero-source images as Figure 3(a). Once the normal map is reconstructed, we repeat this procedure iteratively using the current albedo and spherical harmonics.

To review, in this section we build a two-step iterative optimization method to alternately recover illuminations by solving Eq. 12 and normal map by solving Eq. 14. In practice, two to four iterations prove a significant improvement for the normal orientation. The recovery of normal map gets converged within 4 iterations and costs about 20s on a 2.5GHz Intel Core i5-3210M CPU.

4. Surface Improvement

In this section, the normal map is applied to the initial shape for a more discriminating face model. Nehab [12] suggests that a triangular mesh can be manipulated to match the new normals by minimizing a cost function of two error terms as:

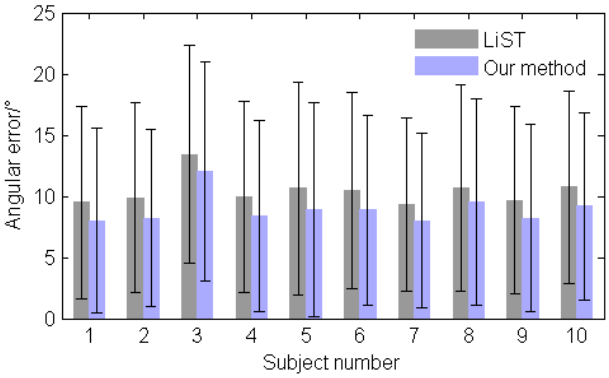


Figure 2. Mean and standard deviation of angular errors for 10 BFM subjects.

Figure 3. Normal map of Hetero-source images (a); Angular errors distribution presented by gray value for LiST result (b), and improved surface (c).

$$\boldsymbol{E}^p = \sum_j ||\boldsymbol{M}_j(\mathbf{s}_j^* - \mathbf{s}_j)||^2 \quad (15)$$

$$\boldsymbol{E}_n = \sum_j \sum_{u,w} ||[\mathbf{n}_j^* \cdot (\mathbf{s}_u^* - \mathbf{s}_w^*)]^2||^2 \quad (16)$$

The position error $\boldsymbol{E}^p$ encourages a solution close to the initial shape and the normal error $\boldsymbol{E}_n$ considers the difference between the new normals and its tangent space, in which $(\boldsymbol{u}, \boldsymbol{w})$ are vertices of edges surrounding the center vertex $\boldsymbol{j}$. The refined shape $\mathbf{s}^*$ is given by minimizing the final error function

$$\min_{\mathbf{s}^{new}} \boldsymbol{\lambda}\boldsymbol{E}^p + (\mathbf{1} - \boldsymbol{\lambda})\boldsymbol{E}^n \quad (17)$$

in which $\boldsymbol{\lambda} \in [\mathbf{0}, \mathbf{1}]$ controls how much effect the positions and normals have on the optimization. A value of 0.15 gives satisfactory results in our experiment. This optimization is formulated as a large over-constrained linear system and can be efficiently solved using sparse least squares. Generally, this step dealing with about 25000 vertices gets finished in 10s on a 2.5GHz Intel Core i5-3210M CPU.

To statistically explain the effectiveness of our surface improvement algorithm, we test it on synthesized Hetero-source image sets of BFM out-of-sample subjects, one subject of which is shown in Figure 4. Angular errors of LiST

Figure 4. Synthesized Hetero-source images across varying illuminations of BFM subject.

fitting result and improved surface relative to the ground truth are analyzed with the mean and standard deviation in Figure 2. It can be observed that our method helps decrease the angular errors for all subjects. Figure 3 shows the distribution of angular errors, in which brighter pixels indicate larger errors. It is apparent that angular errors of LiST fitting result are larger in global compared with the improved surface. Moreover, the nose and mouth parts contain regions with maximum errors and need more correction, which is corresponding to our experimental results.

5. Experiments

We would like to evaluate our algorithm on both synthetic data generated from BFM [14] and real world data from FRGC2.0 [15]. The experiments provide both quantitative error measures with respect to the ground truth and visual evaluation. For experiments below, we first apply the modified LiST fitting algorithm to initialize the 3D shape and texture using a picture randomly selected from the Hetero-source image set. The fitting result also acts as the state-of-the-art comparison to our improved geometry.

To create Hetero-source image sets for BFM subjects, we randomly generated 10 renderings across varying illuminations based on Spherical Harmonic Lighting model and continually applied different nonlinear transformation to the synthesized images, e.g., histogram equalization, image intensity adjustment and logarithm transformation. One example is illustrated in Figure 4, some images of which show extreme lighting conditions.

To quantitatively measure the reconstruction error, we estimate the per-vertex average of distances between the recovered surface and ground truth using the inner points (neglecting neck and ears). It can be observed in Figure 5 that our algorithm improves the accuracy for all cases of two databases. Moreover, our algorithm performs much better than the LiST fitting algorithm in depicting face local fine details, which is corresponding to our original intention. Two of the BFM refined face shapes are presented in the first two rows of Figure 6. As shown in the figure, the corrected surfaces are capable of describing detailed and discriminating features such as dimples on the faces. Experimental results on real world data from FRGC2.0 in the last two rows of Figure 6 demonstrate more visual resemblance

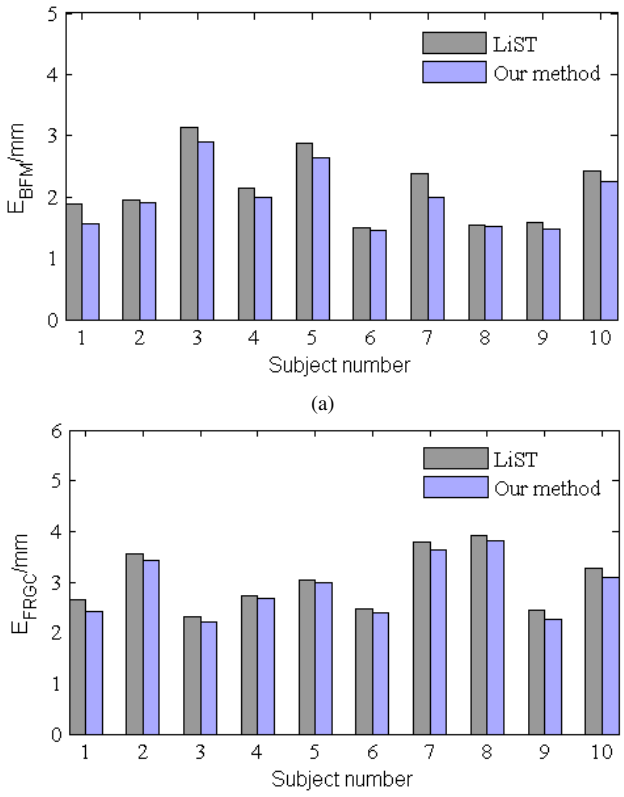


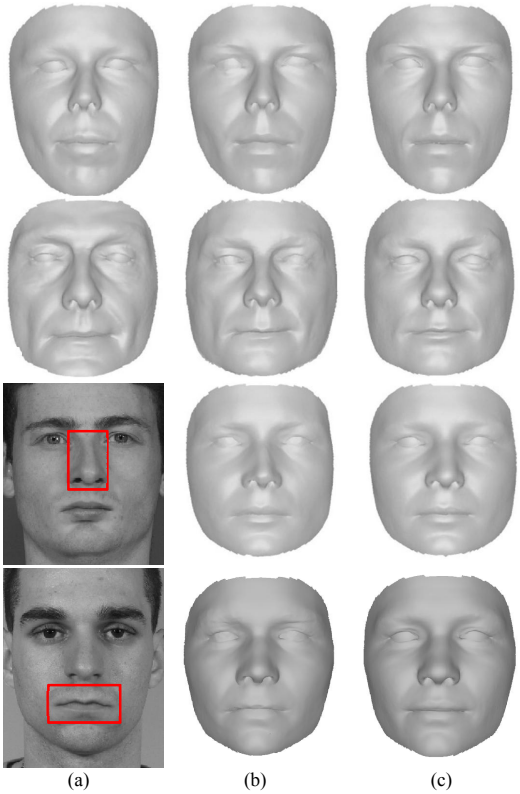
Figure 5. Reconstruction error of BFM (a) and FRGC2.0 (b). All distances between eye centers are normalized to 60 mm.

for the nose and mouth as emphasized in red rectangles.

6. Conclusion

In this paper, we creatively proposed a method capable of reconstructing high quality 3D face models from unstructured Hetero-source images. We began by fitting the 3DMM to an input image using the modified LiST algorithm. The resulting shape and texture are set as initial values for our two-step estimation of normal map inferred by image shading cues. By combining the 3DMM fitting result and the normal information, our method provides a route to improving the shape local details and reduce the problem of model dominance. Experimental results demonstrate the ability of our algorithm to capture discriminating facial features and improve quantitative accuracy.

Our work could be extended in a variety of ways. We would like to incorporate it with more other reconstruction methods since it is independent from the single image based recovery. And fortunately it adds only a little time burden because all the steps are linear. Additionally, the limitation of non-frontal cases could be solved by incorporating rotation transformation in the normal map optimization.



(a) (b) (c)

Figure 6. Ground truth or input faces (a); our improved shapes (b); LiST fitting shapes (c).

Acknowledgement

This work was supported by the National Natural Science Foundation of China (61101152) and the Tsinghua University Initiative Scientific Research Program.

References

[1] B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani, and T. Vetter. Reconstructing high quality face-surfaces using model based stereo. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[2] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(2):218–233, 2003.

[3] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *International journal of computer vision*, 35(1):33–44, 1999.

[4] V. Blanz, A. Mehl, T. Vetter, and H.-P. Seidel. A statistical method for robust 3d surface reconstruction from sparse data. In *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*, pages 293–300. IEEE, 2004.

[5] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.

[6] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):643–660, 2001.

[7] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(6):1060–1071, 2010.

[8] I. Kemelmacher-Shlizerman and R. Basri. 3d face reconstruction from a single image using a single reference face shape. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(2):394–405, 2011.

[9] I. Kemelmacher-Shlizerman and S. M. Seitz. Face reconstruction in the wild. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1746–1753. IEEE, 2011.

[10] T. Migita, S. Ogino, and T. Shakunaga. Direct bundle estimation for recovery of shape, reflectance property and light position. In *Computer Vision–ECCV 2008*, pages 412–425. Springer, 2008.

[11] S. Milborrow and F. Nicolls. Active Shape Models with SIFT Descriptors and MARS. *VISAPP*, 2014. `http://www.milbo.users.sonic.net/stasm`.

[12] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 536–543. ACM, 2005.

[13] A. Patel and W. A. Smith. Driving 3d morphable models using shading cues. *Pattern Recognition*, 45(5):1993–2004, 2012.

[14] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 296–301. IEEE, 2009.

[15] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005.

[16] S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *Computer Vision?aECCV 2002*, pages 3–19. Springer, 2002.

[17] S. Romdhani and T. Vetter. Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 986–993. IEEE, 2005.

[18] O. Sorkine. Least-squares rigid motion using svd. *Technical notes*, 120:3, 2009.

[19] C. Wu, Y. Liu, Q. Dai, and B. Wilburn. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *Visualization and Computer Graphics, IEEE Transactions on*, 17(8):1082–1095, 2011.

[20] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(9):1744–1757, 2012.