Sparse Transfer for Facial Shape-from-Shading

Journal:	*IEEE Transactions on Human-Machine Systems*
Manuscript ID:	THMS-14-08-0291
Manuscript Type:	Regular Paper
Date Submitted by the Author:	14-Aug-2014
Complete List of Authors:	Hu, Jian-Fang; Sun Yat-Sen University, School of Mathematics and Computational Science Zheng, Wei-Shi; Sun Yat-Sen University, School of Information Science and Technology Xie, Xiaohua; Chinese Academy of Sciences, Shenzhen Institute of Advanced Technology (SIAT) Lai, Jianhuang; Sun Yat-Sen University, School of Information Science and Technology
Keywords:	Biometrics, face image processing, Visual processing, 3D face image processing

SCHOLARONE™
Manuscripts
Review

Sparse Transfer for Facial Shape-from-Shading

Jian-Fang Hu, Wei-Shi Zheng, Xiaohua Xie, and Jianhuang Lai

***Abstract*—We present a transfer approach to reconstruct a 3D face shape (depth map) from 2D images. First, some reference shapes are derived from sufficient images (more than 9) of source faces based on a complete non-Lambertian reflectance model, which allows the existence of cast shadow and specularity and makes the reference shapes inferred more accurately. Relying on the fact that human faces have similar 3D structures, the reference shapes are then used to guide the inference of the target shape even based on a few images(e.g., only two). We develop a sparse face shape adaptation method in order to adaptively and selectively transfer reliable face shapes to reconstruct the target's. During this process, neither any standard 3D shape (such as the scanned 3D face) nor explicit estimation of face albedo is required. The evaluation on three data sets including both synthetic and realistic ones has verified the effectiveness of the proposed method.**

***Index Terms*—shape recovery, Non-Lambertian model, sparse transfer face shape.**

I. Introduction

FACE shape is one of the key characteristics for describing a human face. The recovery of face shape is widely used in face recognition [10], [20], [14], face re-rendering [32], and face animation [27]. One approach to obtain face shape is using depth sensors such as the Microsoft Kinect and the $Cyberware^{TM}$ laser scanner. These devices tools often suffer from a low spatial resolution and a low depth precision. Moreover, some of them are too expensive. For these reasons, the depth-sensor-based method is not popular for precise face shape recovery. In contrast, thanks to convenient capture of high resolution data, the photograph-based method is more suitable for inferring the face depth map.

Recovering face shape from given images often referred to a complex optimization process which aims to minimize the error between the observed appearance and the reconstructed appearance by simultaneously estimating intrinsic components (shape and texture) and extrinsic parameters (pose, illumination, and camera properties) of the image face [40], [39]. Many researchers have proposed using salient structure or shading cues in still images to infer surface depths [31], [39], [10], [3], [17]. While salient structures are not easy to discover, the shading cue has been found more useful in recent researches. Most of previous shading-based work cover four key operations: (i) establishing an accurate reference face shape; (ii) modeling the reflection of light by the concerned object; (iii) estimating the lighting in target images; (iv) inferring the target shape restricted by the reference shape, so that the rendered shading fits the one hinted in target images. Furthermore, these methods require an albedo estimation or utilize the 3D scanners to aid the construction of reference shape [17], [12].

Compared with the 3D scanner based methods, obtaining a reference shape from 2D images is more convenient to track the correspondence between the 3D face shape and 2D face images. Meanwhile, it has been also suggested that it is more pragmatic to work with 2D images alone since the majority of face databases still remain 2D [20], [22]. In this work, we propose a new 3D face shape reconstruction model by transferring shape cues from source (non-target) 2D face images[1] via sparse modeling.

A. Related Works

i) Reflectance model. For methods that infer face shape from shading, one of the key challenges is how to accurately model the reflection of light by an object. The popular model is the bidirectional reflectance distribution functions (BRDF). There has been a tremendous effort to build parametric BRDF models [33], [20]. In all the proposed reflectance models, the Lambertian model, which assumes that the surface acts as a low-pass filter on the light signal, is widely used due to its simplicity [10], [3], [17]. In these methods, how to figure out the low-pass filter becomes the main challenge of the Lambertian based method. The problem may be addressed by mining a set of basis (i.e. subspaces) to represent the face-light interaction [10]. In [3] and [17], a spherical harmonic approximation of BRDF is presented. In [40], [39], [34], a linear or nonlinear approximation of the Lambertian rule is proposed. To capture the high-frequency lighting effects (e.g. cast shadow and specularity), a non-Lambertian model is introduced [20], [4], [5]. However, these non-Lambertian methods typically need a large mount of images to gain a reliable reconstruction. For instance, [20] uses at least 9 images.

ii) Use of 3D reference face shapes. Some researchers emphasize that depth information cannot be recovered from image intensity directly [2]. However, the analysis of a subject's shape can benefit the shape reconstruction for other subject. In other words, to get a reasonable face shape recovery, a reliable reference shape is always employed. Especially, 3D scanned faces are often used. For instance, [36], [8] incorporate a statistical model of face shape and traditional shape-from-shading algorithm together for obtaining a reliable face shape.

J. Hu is with the School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, China.
E-mail: hujianf@mail2.sysu.edu.cn
W. Zheng and J.H. Lai are with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China.
E-mail: wszheng@ieee.org and stsljh@mail.sysu.edu.cn
Xiaohua Xie is with Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences.
E-mail: xiaohua.xie@siat.ac.cn

[1]In this paper, we call the 2D images used to help reconstruction of target 3D face image as the source face images.

For Peer Review

Approaches	Albedo estimation	Reflectance	Transfer	Number of input images	Scanned 3D Shape	Other Assumption
Illumination cone[10]	Yes	Lambertian	No	≥ 7	No	No attached shadow, symmetric faces, ray tracing
Subspace [3]	No	Lambertian	No	1	No	Distant&isotropic lighting, 3D scans
Exemplar model [17], [16]	Yes	Lambertian	Yes	1	Yes	No attached shadow, manual alignment of 3D models and 2D images
Wild Face[18], [12]	Yes	Lambertian	Yes	1	Yes	Accurate alignment of 3D models and 2D images
Morphable[4], [5]	Yes	Phong [29]	Yes	1	Yes	Accurate alignment of 3D models and 2D images
Statistical [36]	Yes	Lambertian	Yes	3	Yes	Sufficient 3D aligned models
Nonlinear Mapping [24]	No	Lambertian	Yes	1	Yes	Large number of accurate aligned 3D face models
Tensor Spline [20]	No	Non-Lamb.	No	≥ 9	No	Point sources with known directions
Our method	No	Non-Lamb.	Yes	≥ 2	No	Point sources with known directions

TABLE I
REQUIREMENTS, ASSUMPTIONS OF CANDIDATE METHODS FOR RECOVERING FACE SHAPES [20]

However, they need to train a statistical model on relatively "clean" data. Therefore, they need a large number of accurate 3D face models as well as the point-wise correspondence between all the models. To address this limitation, [3], [17] present an exemplar based model for molding face shapes. They achieve object modeling by modifying the shape and albedo of a reference face, so that it can fit the input image well. In [18], [12], the authors also propose using varied reference model to minimize a reconstruction error. [24], [25], [26] aim to learn a mapping (linear or linear) between transformed image space and the reference face shape space to obtain their reconstruction. Among the existing works, 3D scanned faces are often used as reference shapes. However, obtaining a large set of precise shapes via 3D scanners and manually aligning these shapes are still an expensive. Recently, there exists work intending to accurately reconstruct a 3D face from large unstructured photo collections [20]. We are inspired by this and propose building reference shape using sufficient images.

iii) Application of transfer learning. Recent study on transfer learning show that knowledge gained from one learning task or domain can be used to improve the performance of another [30], [37], [28]. For instance, using the knowledge learned from unlabeled data to benefit their supervised classification and clustering tasks is considered in [30], [7]; methods for transferring the expression of one individual to animations of another subject are explored in [37], [38]. Meanwhile, it should be mentioned that the reference model based methods, including Exemplar model [17], [16], Wild Face[18], [12], Morphable[4], [5], Nonlinear Mapping [24] and Statistical [36], [8] in Table I, can also be viewed as instances of transfer learning as they utilize the scanned shape of the auxiliary person to help the reconstruction of the target subject. Our method is also a transfer based model that transfers the knowledge of reference samples for the shape recovery of target subject. However, different from theirs, our method is a 3D scanner free method as illustrated in Table I, and it would discover the transferred shape cues from a set of 2D non-target images by a sparse transfer modeling.

B. Overview of the Proposed Method

We aim to transfer the shape cues from 2D source (non-target) face images to recover 3D target face shape. More specifically, we construct reference shapes using sufficient face images (more than 9 images for each source face). These reference shapes are then applied to assist the recovery of target face shape. During the reconstruction process, a target-special reference shape is implicitly produced by a sparse linear combination of all reference shapes. Our approach essentially transfers a face shape constraint, which is obtained by a reliable task (using sufficient images), to facilitate the solving of a seriously ill-posed problem (using a few target images). The flow chart of our method is presented in Fig. 1. During this process, neither any standard 3D shape nor explicit albedo estimation is required. A comparison between our method and the state-of-the-art methods is presented in Table I. Key characteristics of the proposed approach are explained below.

i) Reconstructing reference face shapes from 2D images. The reference shapes are expected to be reconstructed as accurately as possible. For this goal, besides using adequate face images (e.g., more than 9 images for each face), a complete reflectance model is employed for modeling the intensity variation. This reflectance model, which is based on the mathematical framework presented in [20], is beyond the Lambertian surface assumption and allows the existence of cast shadow and specularity.

ii) Inferring face shape in surface normal field. In our approach, the face shape is represented by surface normals rather than depth map. Figure 1 shows that face images of different people share a similar surface normal field (i.e. gradient field), which makes the transfer approach feasible. The advantage of using surface normals is two-folds. Firstly, it allows us to design a computationally efficient algorithm to recover target shape without explicitly providing an accurate definition of the boundary conditions. Secondly, it has been known that the surface normal is invariant to the captured angle and position, thus easily adapted to varying face poses [41].

iii) Sparse modeling for transferring the reference shape as the reconstruction constraint. Aided by reference shapes, we develop a sparse face shape adaptation method for inferring the target face shape. On one hand, the target shape should not deviate the reference shape too much. On the other hand, the synthesized 2D images of target face would fall close the observed target images as much as possible. Therefore, the reference shape severs as a template to our reconstructed shape, and the input target images are used as a guide to mold

Unedited Draft

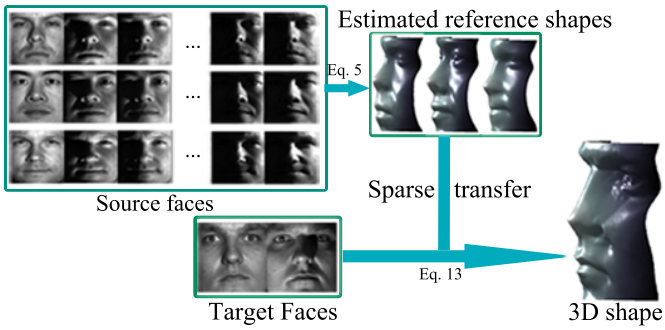


Fig. 1. An illustration of our method.

the reference shapes to reach a desired reconstruction, where a sparse modeling is used to make a selective transfer. In other words, our method intends to modify the reference shape to fit the target face images.

iv) Using different reflectance models in different tasks. In our approach, a non-Lambertian model is used for the reconstruction of reference shape while a Lambertian mode is used for the reconstruction of target shape. This ensures an accuracy reconstruction of reference shape, and enables the inference of target shape using even a small set of images (e.g., using only two images).

The remainder of this paper is organized as follows: In Section II we describe the proposed method in details. Section III presents the experimental results and discussion. Section IV gives the conclusion.

II. Our Approach

Our goal is to recover the 3D shape of a target face using a small number of 2D target face images, even in the case that only two images are available. Our method aims to select reference face shapes from non-target people to facilitate the recovery. For this objective, reliable reference shapes are first reconstructed from sufficient source images. Based on the reference shapes, the target shape can be generated by the proposed sparse transfer model, where a few or even only two garget images are required. Our method works under the fact that different faces hold similar shapes (represented by surface normals), which provides the premise of using a transfer modeling. Furthermore, our method does not require any estimation on albedo or light intensity. Details are presented as follows.

A. Reference Shape Recovery

In this section, We describe how to estimate the shape of a source face from multiple 2D images. To obtain an accurate estimation, we have proposed using a non-lambertian reflectance model to represent the interaction between light and surface, which allows the existence of cast shadow and specularity.

1) Non-Lambertian reflectance model: We develop a non-Lambertian reflectance model based on the mathematical framework proposed by Kumar et al. [20].

Consider the intensity of a face E as a spherical function with respect to an unit vector $\mathbf{s} = (s_1, s_2, s_3)$ that indicates the light direction. Such a spherical function in $\Re^3$ can be expressed using an n-th order Cartesian tensor [15], i.e.,

$$E = \sum_{k+l+m=n} T_{klm} s_1^k s_2^l s_3^m \tag{1}$$

where T_{klm} is the real-value tensor coefficient, and k, l, and m are no-negative integers encoding the information of the surface normal, the incident light intensity, and the albedo of the corresponding point, respectively.

The Lambertian reflectance model can be seen as a special case of Eq. (1) with $n = 1$, while the tensor model employed in [20] keeps the third order terms. In our method, the zero, first, and second order terms are all considered, and thus the reflectance model can be formulated as follows

$$E = \sum_{1 \le k+l+m \le 2} T_{klm} s_1^k s_2^l s_3^m \tag{2}$$

The Eq. (2) can also be rewritten in a matrix form, i.e.,

$$E = \bar{\mathbf{s}}^T \cdot U \cdot \bar{\mathbf{s}} \tag{3}$$

where $\bar{\mathbf{s}} = (1, s_1, s_2, s_3)^T$, and

$$U = \begin{pmatrix} 0 & T_{100}/2 & T_{010}/2 & T_{001}/2 \\ T_{100}/2 & T_{200} & T_{110}/2 & T_{101}/2 \\ T_{010}/2 & T_{110}/2 & T_{020} & T_{011}/2 \\ T_{001}/2 & T_{101}/2 & T_{011}/2 & T_{002} \end{pmatrix} \tag{4}$$

Our reflectance model, i.e., Eq. (3), consists of both linear and non-linear terms. A combination is quite necessary , because it provides a framework to capture the complete intensity variation caused by surface-lighting interaction. More specifically, the linear terms mainly capture the low-frequency information and the nonlinear terms would exhibit the high-frequency component such as cast shadow and specularity.

For each face to be represented by our reflectance model, totally nine (2^3+1) parameters need to be estimated. Intuitively, in order to obtain an estimation of the parameters, we need at least 9 uncorrelated equations. In practice, we use at least 9 images for each source subject to gain a robust parameter estimation.

2) Surface Normal inference for Reference Shape: Suppose J images $\{I_j\}_{j=1,2...J}$ of a source face are used and the light direction $\mathbf{s}$ of each image has been known. We can estimate the unknown T_{klm} in the reflectance model (2) by solving a linear system of J linear equations. After that, the surface normal can be estimated by simply calculating maximum of the following reflectance model

$$\tilde{\mathbf{n}} = (\tilde{n}_1, \tilde{n}_2, \tilde{n}_3) = \arg \max_{n_1,n_2,n_3} \sum_{1 \le k+l+m \le 2} T_{klm} n_1^k n_2^l n_3^m \tag{5}$$

This is based on the observation that the apparent bidirectional reflectance distribution function (ABRDF) typically reaches its maximum when the light direction is the same as the surface normal direction. When sufficient source images are used, they likely capture the light direction identical (or

For Peer Review

rather similar) to the surface normal direction for each point. The experimental results later also support this observation. To derive the surface normal, we adopt the standard projected gradient descent method with the initialization of all-one vector [6], [1].

We illustrate the reconstructed surface normal vectors of 38 subjects on extended Yale B database in Figure 2. For each subject, we have used 64 images for the reconstruction. From this figure, we can observe that faces of different subjects typically hold similar structure for surface normal distribution, which enables the development of transfer models for inferring a target face shape (surface normals) from source face images.

B. Target shape inference based on a single reference face shape

Given K images of target face and a reference shape (represented by the surface normal vector $\widetilde{\mathbf{n}}$) which has been reconstructed from sufficient source face images, our goal is to recover the target shape $\mathbf{n}$ using a transfer model. In most practical cases, only a small number of target images are available, and therefore it is a seriously ill-posed problem to solve the non-Lambertian mode. Instead, we employ the Lambertian model to capture the observed intensity variation on target face images. Though the Lambertian rule is more simpler than the non-Lambertian one, the experiment will show that our reconstructed result is still satisfactory. The main reason is that the reconstructed result is also controlled by the reference shape which has been built accurately based on a non-Lambertian model (Eq.(2)). According to the Lambertian rule, the task is to infer the surface normal $\mathbf{n}$ satisfying the following equations:

$$E_k(x,y) = I\rho(x,y)\mathbf{n}(x,y)^T\mathbf{s_k}, k = 1,2...K \tag{6}$$

where I and $\mathbf{s_k}$ indicate the illumination intensity and direction of light source, respectively. $\mathbf{n}(\mathbf{x},\mathbf{y})$ refers to the face surface normal and $\rho(x,y)$ indicates the albedo at point (x,y). For each pair of images (k_1,k_2), we can get the following equations:

$$\frac{E_{k_1}(x,y)}{E_{k_2}(x,y)} = \frac{\mathbf{n}(x,y)^T\mathbf{s}_{k_1}}{\mathbf{n}(x,y)^T\mathbf{s}_{k_2}}, k_1,k_2 = 1,2,...K \tag{7}$$

The above equations can be reformed as

$$E_{k_1}(x,y)\mathbf{n}(x,y)^T\mathbf{s_{k_2}} = E_{k_2}(x,y)\mathbf{n}(x,y)^T\mathbf{s_{k_1}}, \quad k_1,k_2 = 1,2,...K \tag{8}$$

Based on the above equation set, the surface normal $\mathbf{n}$ can be inferred without explicit estimation of albedo $\rho(x,y)$ and illumination intensity I. However, it is hard to find a solution that strictly fits all the equations. Hence, we introduce a slack variable $\varepsilon_{k_1k_2}$, that is

$$E_{k_1}(x,y)\mathbf{n}(x,y)^T\mathbf{s_{k_2}} = E_{k_2}(x,y)\mathbf{n}(x,y)^T\mathbf{s_{k_1}} + \varepsilon_{k_1k_2}, \quad k_1,k_2 = 1,2,...K \tag{9}$$

The slack variables actually refer to the reconstruction error of $\mathbf{n}$.

As stated before, we can use the pre-learned reference shape $\widetilde{\mathbf{n}}$ to guide the reconstruction by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{n}} \int_\Omega (\mathbf{n}(x,y) - \widetilde{\mathbf{n}}(x,y))^T(\mathbf{n}(x,y) - \widetilde{\mathbf{n}}(x,y)) \\ + \lambda \sum_{k_1,k_2} \varepsilon^2_{k_1k_2} dxdy \\ s.t. \mathbf{n}(x,y)^T\mathbf{n}(x,y) = 1, \forall(x,y) \in \Omega \end{aligned} \tag{10}$$

In the above objective function, the first term is to restrict our reconstruction close to the reference shape, namely shape deviation, and the second term measures the reconstruction error relevant to target images. λ is a parameter to control the trade-off between the reconstruction error and the shape deviation. The larger the λ is, the larger the shape deviation is allowed when constructing the target. Furthermore, according to Eq.(9), we have

$$\begin{aligned} \varepsilon^2_{k_1k_2} &= (E_{k_1}(x,y)\mathbf{n}(x,y)^T\mathbf{s_{k_2}} - E_{k_2}(x,y)\mathbf{n}(x,y)^T\mathbf{s_{k_1}})^2 \\ &= \mathbf{n}^T[(E_{k_1}\mathbf{s}_{k_2} - E_{k_2}\mathbf{s}_{k_2})(E_{k_1}\mathbf{s}_{k_2} - E_{k_2}\mathbf{s}_{k_1})^T]\mathbf{n} \end{aligned} \tag{11}$$

By defining $\mathbf{A} = \sum_{k_1,k_2}(E_{k_1}\mathbf{s}_{k_2} - E_{k_2}\mathbf{s}_{k_1})(E_{k_1}\mathbf{s}_{k_2} - E_{k_2}\mathbf{s}_{k_1})^T$, which is a positive definite matrix, the optimization system (10) can be represented as

$$\begin{aligned} \min_{\mathbf{n}} \int_\Omega \|\mathbf{n}(x,y) - \widetilde{\mathbf{n}}\|^2 + \lambda\mathbf{n}(x,y)^T\mathbf{A}\mathbf{n}(x,y)dxdy \\ s.t. \mathbf{n}(x,y)^T\mathbf{n}(x,y) = 1, \forall(x,y) \in \Omega \end{aligned} \tag{12}$$

Here, Ω indicates the face region to be reconstructed. Note that the optimization problem (12) is convex with respect to $\mathbf{n}$ and is a standard least squares problem with quadratic constrains. This constrained optimization problem can be solved using the gradient descent with iterative projection [6], [1]. Especially, the estimation of $\mathbf{n}$ can be iteratively updated and then be projected to the norm ball until it converges to a local minimal. Given the t-th estimator $\mathbf{n}(\mathbf{t})$ and the corresponding gradient $\mathbf{g}(t) = 2(\mathbb{I} + \lambda\mathbf{A})\mathbf{n} - 2\widetilde{\mathbf{n}}$, the new updated point can be determined by $\mathbf{n}(t+1) = cos(\tau)\mathbf{n}(t) - sin(\tau)\mathbf{p}(t)$, where the projection $\mathbf{p}(t)$ is given by $\frac{\mathbf{g}(t)-\mathbf{g}(t)^T\mathbf{n}(t)\mathbf{n}(t)}{\|\mathbf{g}(t)-\mathbf{g}(t)^T\mathbf{n}(t)\mathbf{n}(t)\|}$. Here τ is the iteration step size and an optimal step size is computed by a line search method in each iteration.

C. Target shape inference based on multiple reference face shapes

Sometimes there exists large difference between the given reference face shape and the target face shape. In this case, the shape recovery may be not accurately guided by a single source face shape. Moreover, it is sometimes hard to manually select the best reference shape for transferring. To overcome this limitation, we further extend our method to the case of multiple reference face shapes.

Suppose $\{\widetilde{\mathbf{n}}_s\}_{s=1,2...S}$ are the surface normal vectors of S source faces. We extend the reconstruction model (12) to a

For Peer Review

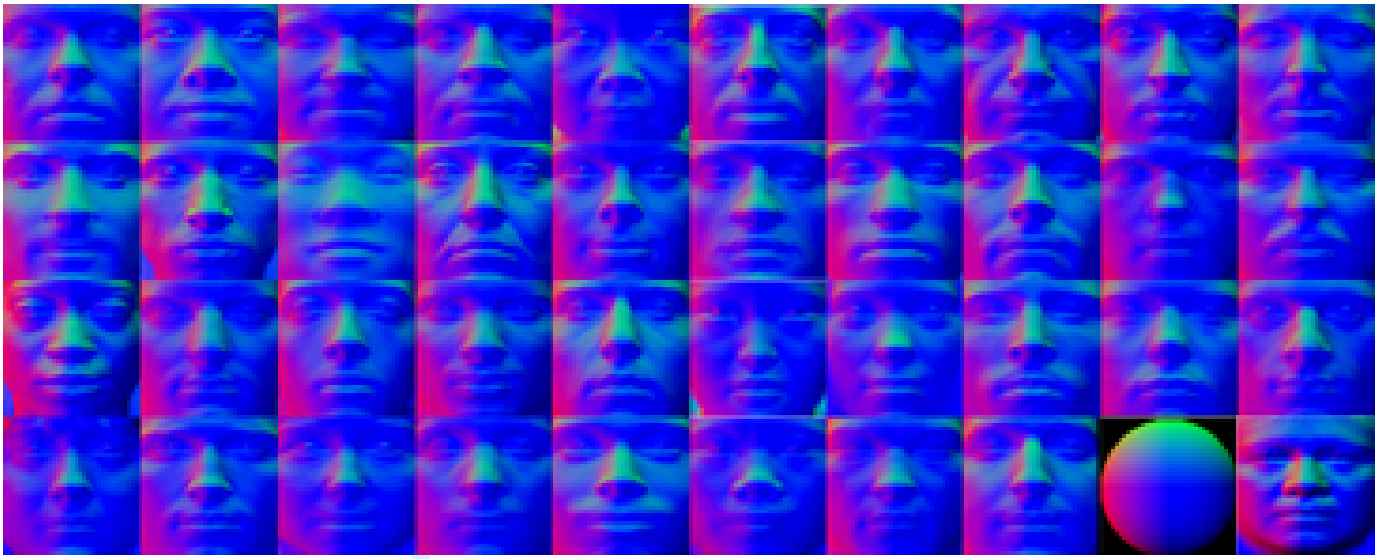


Fig. 2. Visualization of surface normal vectors calculated in Extended Yale B set. Each block shows the surface normal of one subject's face. For reference, we also show the surface normal vectors of sphere and a scanned face of the Texas 3DFRD database [11] in the last two blocks. This figure is suggested to visualize in color.

sparse face shape adaptation modeling as follows:

$$\min_{\mathbf{n},\boldsymbol{\beta}} \int_{\Omega} (\|\mathbf{n} - \sum_{s=1,2...S} \beta_{\mathbf{s}} \widetilde{\mathbf{n}}_{\mathbf{s}}\|^2 + \lambda \mathbf{n}^T \mathbf{A} \mathbf{n}) dxdy + \gamma \|\boldsymbol{\beta}\|_{\ell_1}$$
$$s.t. \mathbf{n}(x,y)^T \mathbf{n}(x,y) = 1, \forall (x,y) \in \Omega \tag{13}$$

In this model, the reference shape is treated as a combination of source shapes, with $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_S)$ as combination coefficients. The last term is a sparse constraint which essentially selects less reference shapes that fit the target shape best to combine, instead of using a simple weighted average on all reference shapes. γ is a regularization factor, where a larger γ would lead to a selection of less reference shapes for target shape reconstruction. In our experiments, we fix the key parameters λ as 1 and γ as 3000 in experiments unless otherwise stated. The influence of γ would be evaluated in the experiments.

The optimization system (13) can be solved by computing the following two steps alternatively until convergence to a local minima.

STEP 1, For fixed coefficients $\boldsymbol{\beta}$, optimize the surface normal $\mathbf{n}$:

$$\min_{\mathbf{n}} \int_{\Omega} \|\mathbf{n} - \sum_{s=1,2...S} \beta_{\mathbf{s}} \widetilde{\mathbf{n}}_{\mathbf{s}}\|^2 + \lambda \mathbf{n}^T \mathbf{A} \mathbf{n} dxdy$$
$$s.t. \mathbf{n}^T \mathbf{n} = 1, \forall (x,y) \in \Omega \tag{14}$$

STEP 2, For fixed surface normal $\mathbf{n}$, optimize the coefficients $\boldsymbol{\beta}$:

$$\min_{\boldsymbol{\beta}} \int_{\Omega} \|\mathbf{n} - \sum_{s=1,2...S} \beta_{\mathbf{s}} \widetilde{\mathbf{n}}_{\mathbf{s}}\|^2 dxdy + \gamma \|\boldsymbol{\beta}\|_{\ell_1} \tag{15}$$

The problem in (14) is a standard instance of single reference shape based optimization as formulated in (12). It can be solved by the projected gradient descent method described in previous section. Criterion (15) is a convex L_1-regularized least squares problem and can be efficiently solved by a gradient descend method [21]. Specially, the gradient $\mathbf{g} = (g_1, g_2, ..., g_s, ..., g_S)$ is determined by $g_s = 2\int_{\Omega}(\mathbf{n} - \sum_{s=1,2...S} \beta_{\mathbf{s}} \widetilde{\mathbf{n}}_{\mathbf{s}}) \widetilde{\mathbf{n}}_s dxdy + \frac{\beta_s}{|\beta_s| + \epsilon}$. Here, a small positive constant ϵ is added to avoid a zero denominator.

D. Depth recovery

Once the surface normal vectors have been calculated, we can use one of the standard shape-from-gradient methods to recover the depth map [13], [19], [9]. We adopt the method proposed in [9], which enforces the integrability of the reconstruction implicitly by projecting the possibly non-integrable surface slope estimates onto the nearest integrable surface slopes. This approach is appropriate for integrating noisy slope estimations and can enhance our reconstruction result.

It is worthy noting that the surface normal $\mathbf{n}(x,y)$ at point (x,y) can also be expressed as $(z_x, z_y, -1)/\sqrt{z_x^2 + z_y^2 + 1}$, where z_x and z_y indicate the partial derivative of the depth map z with respect to x and y directions, respectively. This results in the following equations

$$z_x = -\frac{n_1}{n_3}, z_y = -\frac{n_2}{n_3}, \mathbf{n} = (n_1, n_2, n_3)^T \tag{16}$$

In order to obtain a smooth and integrable depth map $\widetilde{z}$, the possibly non-integrable estimated surface slopes (z_x, z_y) are projected onto a set of integrable surface slopes $(\widetilde{z}_x, \widetilde{z}_y)$ where $\frac{\partial}{\partial x}\widetilde{z}_y = \frac{\partial}{\partial y}\widetilde{z}_x$, by minimizing the following distance measure:

$$d(\widetilde{z}, z) = \int_{\Omega} \mid F(\widetilde{z}_x) - F(z_x) \mid^2 + \mid F(\widetilde{z}_y) - F(z_y) \mid^2 dxdy$$

Here, $F(.)$ refers to a fourier transformation operator. Note that the derivatives of the fourier basis functions possess the

For Peer Review

Synthesized images Yale B CMU-PIE

Fig. 3. Some examples from the synthesized, Extended Yale B, and CMU-PIE dataset under different lighting conditions.

following useful properties

$$F(\widetilde{z}_x) = iw_x F(\widetilde{z}), F(\widetilde{z}_y) = iw_y F(\widetilde{z}) \tag{17}$$

where (w_x, w_y) is a two-dimensional index for Fourier transformation. Then by substituting Eq. (17) into the distance measure $d(\widetilde{z}, z)$ and optimizing it, we can get

$$F(\widetilde{z}) = \frac{-iw_x F(z_x) - iw_y F(z_y)}{w_x^2 + w_y^2}. \tag{18}$$

$F(\widetilde{z})$ are the Fourier coefficients of the constrained surface slopes indicated by $\widetilde{z}$. Then it is straightforward to obtain the integrable surface depth map $\widetilde{z}$ by performing the inverse 2D Fourier Transform on the coefficients $F(\widetilde{z})$.

It is worthy pointing out that the expression in Eq.(18) is valid except at the point $(w_x, w_y) = (0, 0)$, which means that the average value of $\widetilde{z}$ cannot be recovered directly. However, it does not affect the reconstruction of face shape.

III. Experimental results and analysis

We have conducted experiments on three data sets including one synthetic set and two real face sets to evaluate the performance of the proposed method. The synthetic set consists of images of ellipsoid and sphere under varying illumination captured from a frontal view. The face data sets include the images from Extended Yale B database [23] and CMU-PIE database [35]. Some examples of these data sets are shown in Figure 3.

A. Synthetic shape recovery

Besides the human face, some other category of shapes may also be reconstructed by our approach as long as a similar reference shape is available. In order to give a quantitative evaluation for our method, we first conducted an experiment on synthetic data. We firstly formed some standard sphere models and an ellipsoid model. For each model, we synthesized a set of images of frontal view but under different lightings. Each sphere image set were used to reconstruct the corresponding sphere shape. All the reconstructed spheres are regarded as reference shapes and used to aid the reconstruction of an ellipsoid.

Data synthesis. Specially, each image is synthesized using the Phong reflectance model [29]

$$\begin{aligned} I(x,y) &= max(0, I_1\rho_1\mathbf{n}^T\mathbf{s} + I_2(\mathbf{n}^T\frac{\mathbf{s}+\mathbf{v}}{\|\mathbf{s}+\mathbf{v}\|})^{30}) + I_3 \\ \mathbf{n} &= (\frac{x-x_c}{a^2}, \frac{y-y_c}{b^2}, \frac{z-z_c}{c^2}) \\ \rho_1 &= 1 - \frac{|z-z_c|}{18c} \end{aligned} \tag{19}$$

where I_1,I_2, and I_3 indicate the Lambertian reflection magnitude, specular reflection magnitude, and environment illumination, respectively. In this experiment, we set $I_1 = 220$ and $I_2 = 30$, and let I_3 be a random value ranging from 3 to 8. a, b, and c are three key parameters to control the shape of the generated ellipsoid. Especially, the shape is a sphere when $a = b = c$. Parameters (x_c, y_c, z_c) indicate the location of the center of sphere or ellipsoid. The second term, i.e. $I_2(\mathbf{n}^T\frac{\mathbf{s}+\mathbf{v}}{\|\mathbf{s}+\mathbf{v}\|})^{30}$, in the max function is used to model the specular reflection.

In practice, we generated three sets of sphere images with radius being 40,50 and 60, respectively. Each image set contains 64 images captured from frontal view under different lighting directions identical to the light parameters in Yale B database. The pixel values of each image were scaled into a range of $[0, 255]$. Some samples can be seen in Figure 3. Three image sets were used to reconstruct the corresponding three spheres as reference shapes.

Similarly, we further generated 64 images based on an ellipsoid model with shape parameters $a = 100$, $b = 80$, and $c = 40$. Our experiment is to infer the ellipsoid shape from some of these ellipsoid images, aided by the reference sphere shapes.

Compared methods. We would like to assess how much benefit the proposed transfer model can bring as compared to the traditional learning methods. Therefore, we compared our method with two baselines without using any reference shapes. The first one is to directly recover target shape from images using Lambertian model and the second one is using our proposed non-Lambertian model. Both these two methods do not use a reference shape.

Since the standard form of the target shape has been known, we can calculate the square error between the standard shape (namely ground truth) and the reconstructed shape to measure the reconstruction performance. Specifically, both depth and surface normal vector errors are reported. The surface normal vector error is measured by the mean angle between the reconstructed surface normal and ground-truth surface normal, while the depth error is defined by the mean deviation of difference between the estimated depth map and the ground-truth depth map.

For each method, we randomly selected different numbers of target images used for reconstruction. For a certain number, 50-fold random test were conducted and the average reconstruction error was calculated.

Result. The average reconstruction errors (including depth errors and surface normal vector errors) under different numbers of input target images are shown in Figure 4. As shown, both the depth error and surface normal vector error for all the

For Peer Review

methods reduce as the number of input image K increases. In most cases, the non-Lambertian model based method achieves much lower reconstruction error compared to the Lambertian model based method. The proposed transfer model obtains the lowest reconstruction error among all the comparisons. It is worthy mentioning that the transfer model with 4 input images can perform comparable with the non-lambertian model based method with 10 input images. This demonstrates that the proposed transfer model can significantly reduce the number of input target images.

B. Real face recovery

Datasets. We further evaluate our method on two real face databases: Extended Yale B [23] and CMU Pose, Illumination, and Expression (CMU-PIE) [35]. For these two face databases, only the images captured from the frontal view were selected in our experiment. There are total 2432 frontal face images of 38 subjects under 64 illumination conditions in extended Yale B database, and 1428 frontal face images of 68 individuals under 21 lighting directions in CMU-PIE database.

Compared methods. Since the ground-truth shapes are not available in these database sets, we just used the method in Section II-A to reconstruct the shape from all the face images of each subject. These reconstructed shapes were then regarded as the ground-truth shapes to evaluate the reconstruction performance. When taking one subject as the target face to be reconstructed, the ground-truth shapes of rest subjects in the same database were used as reference shapes to guide the reconstruction. For a comparison, the Direct Lambertian baseline method specified in the last section was also implemented. Since the Direct Lambertian baseline method can only obtain a quite reliable shape with at least 4 input images, we randomly selected 4 target images as input of the Direct Lambertian comparison and our method to make a fair comparison. We also compared our method with tensor spline method introduced in [20]. According to the authors' instruction, the model parameters have to be estimated using 10 input images.

In order to quantitatively measure the difference between two shapes, i.e., h_1 and h_2, an error map was calculated by

$$Error(x,y) = [h_1(x,y) - h_2(x,y)]^2 \tag{20}$$

Based on this error map, we could further calculate the following Mean Squared Error (MSE) and the error variance as evaluation indexes to measure the similarity of given shapes as follows:

$$\begin{aligned} MSE =& \sqrt{\int_\Omega Error(x,y)dxdy} \\ Var =& \frac{1}{|\Omega|}\int_\Omega (Error(x,y) - \frac{MSE^2}{|\Omega|})^2 dxdy \end{aligned} \tag{21}$$

Results. The shape recovery results as well as the reconstruction errors are illustrated in Figure 5. As shown, our method gets better reconstruction result than the other methods in terms of both visual quality and quantitative error. Especially, other methods often fail to reconstruct the regions with drastic variation in depth, such as the canthus and the corner of mouth. We can also observe that our method with 4 input images can even obtain better results on CMU-PIE set than the tensor spline method with 10 input images. The main reason is that there are likely cast shadow and specular reflection appearing in these regions, so that their intensity variation is hard to capture using the Lambertian reflectance model or using a non-Lambertian reflectance model with a small number of images. However, guided by a reliable reference shape, the solution space could be constrained into a reasonable range, producing a more convincing reconstruction result. In the last column of the figure, the coefficients for combination of source face images are shown. For each reconstruction, as expected, only a small set of reference shapes (no more than 7) had been selected to assist the reconstruction of target shape.

In Figure 6, we show the reconstruction error (depth error) for each of the 38 subjects in Extended Yale B and 68 subjects in CMU-PIE data sets, respectively. We can observe that in most cases, the reconstruction error of our method is smaller, as compared to the method without reference shape transfer. The comparison is more notable especially for some extreme cases such as the face with awesome beard (see the face images shown in Figure 6(b)). We illustrate the reconstruction results of one extreme example in Figure 6(c). As shown, if we directly reconstruct the shape using 4 input target face images, the reconstruction quality is quite poor. However, with the help of reference face shape, the reconstruction would be greatly improved. It also can be seen that the reconstruction errors in Extended Yale B are quite lower than the ones in CMU-PIE set. It may be because there are more images per source face for obtaining a convincing reference shape in Extended Yale B set.

As stated in previous discussions, our method can obtain a convincing reconstruction result with only two target input images. We show one example of the reconstruction results using two input images in Figure 7. To further look insight into the reconstruction results, we display the face images after rendering (with texture mapping) based on the reconstructed shape under five different views in Figure 7(c). By observing the face images after rendering under different views, we can find that the reconstructed shape is very reasonable. This also implies that the proposed model can be further used to address the problem of the posture and lighting variations in face recognition by automatically generating images of face under varied postures and lighting conditions.

C. More Evaluations

1) Why Transfer Learning Benefits the Reconstruction: A key idea behind our approach is that all faces share similar 3D structures. Therefore the other subject's shape can serve to guide the reconstruction of the target face's shape. This is also indicated by the first term in the objective function of Eq. (13). Since the reference shape has been estimated using enough source images, it would provide a strong prior for our reconstruction and restrict the search of solution space. Meanwhile, the term $\lambda \mathbf{n}^T \mathbf{A} \mathbf{n}$ captures some individual details

For Peer Review

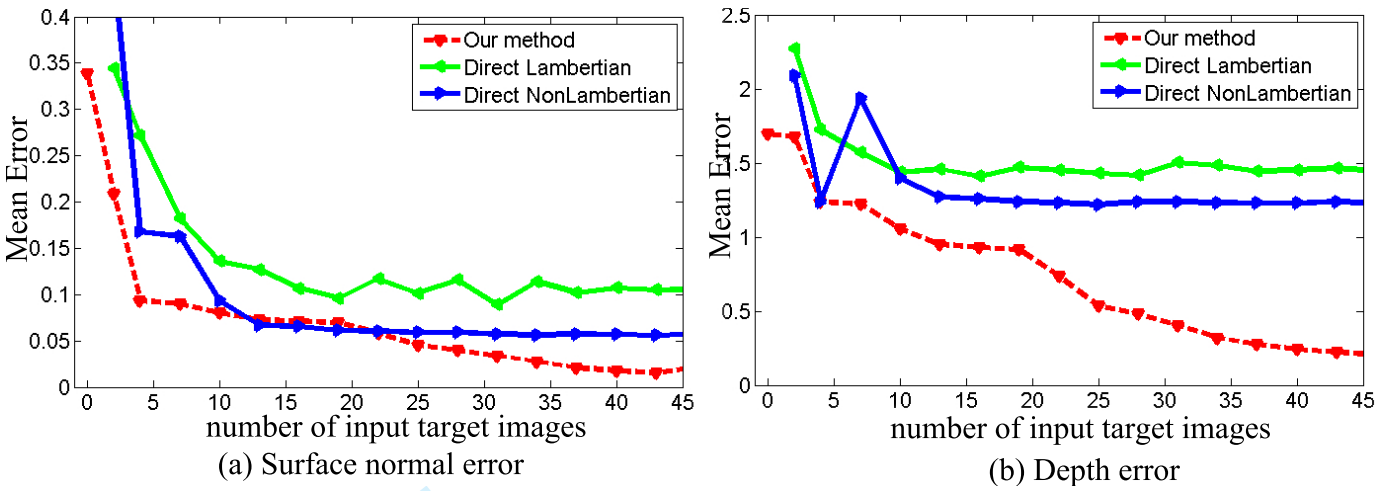


Fig. 4. Reconstruction errors using different methods on synthesized data.

of the target face and makes the reconstruction fit the target face better.

2) Convergence of the Method: Problem (13) is a convex L_1-regularized least squares problem. By using a standard gradient descent algorithm, the optimization is guaranteed to converge to a local minima in a few iterations (about 4 iterations in our experiment). One example of reconstruction iteration is shown in Figure 8, where the reconstructed shape at each iteration as well as the deviation map between two neighboring iterations are displayed. As shown, the deviation turns to decrease as more iterations go.

3) Influence of the number of input target images: We study the influence of the number of the input target face images on the performance of our method. Figure 9 shows the reconstruction results using 2 to 64 images, respectively. For a certain number, the depth error was calculated. As shown, a small number of input images leads to a larger reconstruction error generally. This is desirable, because the less the target images are used, the small the term $\mathbf{n}^T\mathbf{A}\mathbf{n}$ in our objective function (13) becomes. And then the less face details would be captured and so the larger depth error would be obtained. It is worthy pointing out that the reconstruction with less than 6 input images also achieves a small error and a good reconstruction.

4) The effect of weight γ: Here, we evaluate the effect of the weight γ in our objective function Eq. (13). We show some reconstruction results and the mean depth errors in Fig. 10 and Fig. 11, respectively. As expected, a γ that is too large or too small will lead to an inferior reconstruction. Specially, if we discard the sparse constrain by setting $\gamma = 0$, the shape near the nose and forehead would be less accurate. However, according to our observation, the reconstruction is more appealing when γ is between 2000 and 4000.

5) Using inferred reference shapes vs. using scanned reference shapes: Our transfer method makes use of reference face shapes learned from source face images. However, the method can also utilize the scanned face images as reference shapes. In this section, we investigate the performance difference between them. For each target face from the Yale B database, in the first approach, the shapes of the rest 37 subjects are employed to infer the reference shapes. In the second approach, we selected 50 scanned face shapes from the Texas 3D Face Recognition database (Texas 3DFRD) [11] as reference shapes. For each scanned shape, the face surface normal vector $\mathbf{n}$ is computed from depth map Z (ranging from 0 to 1) by

$$\mathbf{n}(x,y) = \frac{(rZ_x, rZ_y, 1)^T}{\sqrt{1 + r^2 Z_x^2 + r^2 Z_x^2}}$$

where Z_x and Z_y indicate the partial derivatives for x and y, respectively, and r is a constant to uniform the pixel distance and the depth value. In our experiment, r was set as 100. One example of the computed surface normal vector of scanned face is shown in the last block of Figure 2. For recovering the target face shape, four images per target face were used as input. From the mean square error point of view, using estimated reference shapes is 338.12 and in comparison using scanned reference shapes is 341.86. The reconstruction results of 3D face shapes relying on different kinds of reference shapes are shown in Figure 12. As shown, learning reference shapes from related source 2D images achieves a comparable result with the method that employs scanned 3D face images under the proposed sparse transfer model.

IV. Conclusion and Future Work

We have proposed reconstructing a target face shape from a few face images, aided by a set of reference shapes which are learned from sufficient face images based on a complete reflectance model. In particular, a sparse face shape adaptation method is proposed to perform a selective transfer. A key characteristic of our technique is that it takes advantage of the knowledge learnt from a reliable task, to guide the solving for a non-reliable problem, without using any 3D scanning.

One limitation of our method is that it is based on the point-source lighting and the light direction of face images need to be provided. However, the point lighting model is not available for the face images captured in the wild. In the

For Peer Review

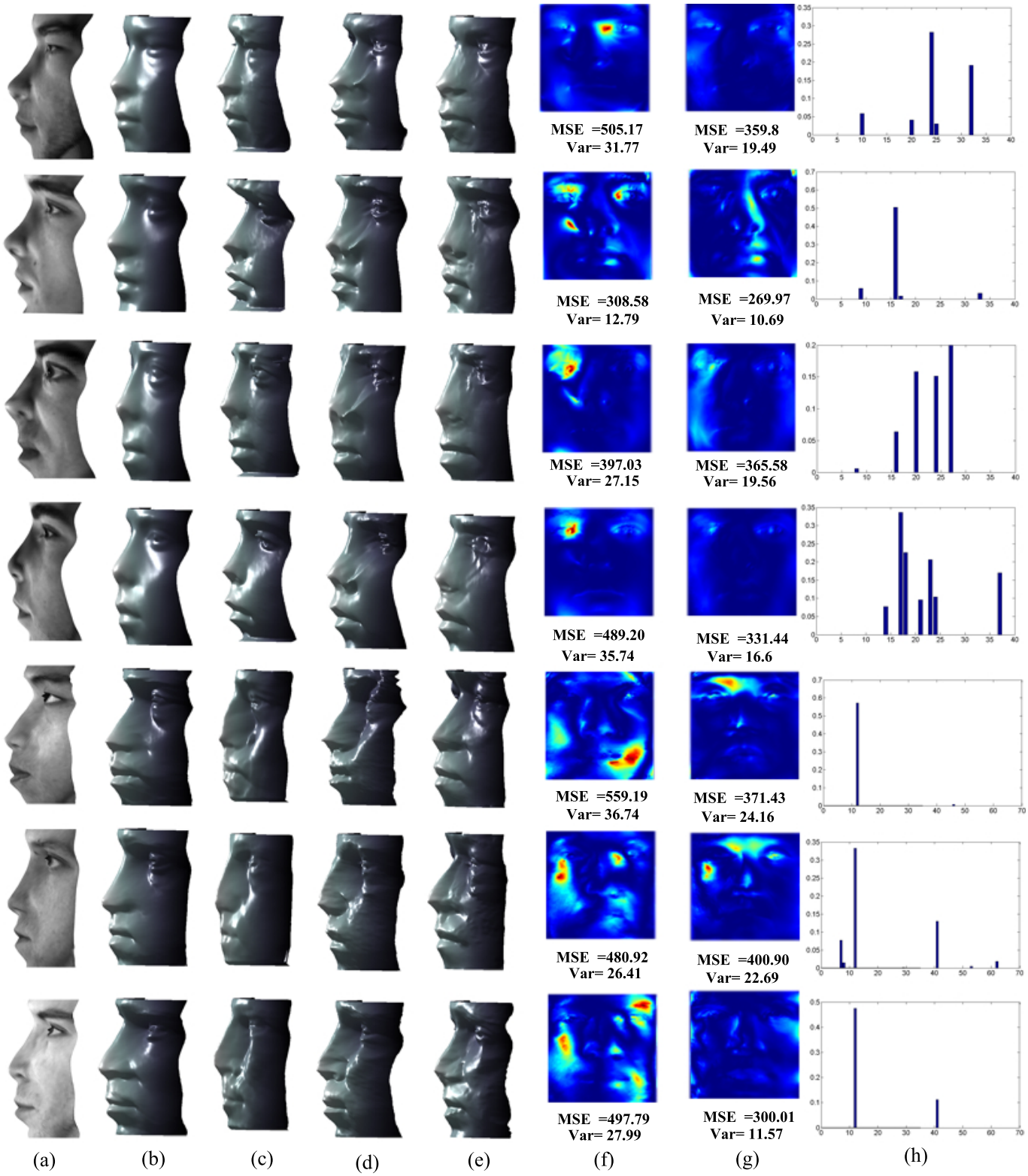


Fig. 5. Example results on Extended Yale B and CMU-PIE data sets. Each row is the result of one subject, the first four rows come from Yale B data set and the rest three ones from CMU-PIE data set. (a) The rendered face image based on the face shape reconstructed by our transfer method; (b) ground truth face shape; (c) face shape estimated by tensor spline model [18] with 10 input target images; (d) face shape recovered from 4 input target images by Lambertian model; (e) face shape estimated by our method with 4 input target images; (f) and (g) reconstruction error maps and the corresponding depth errors (MSE) as well as its variance (Var) of the direct Lambertian method and our method, respectively; (h) the sparse coefficients (β_s in our objective function (13)) for the combination of source faces.

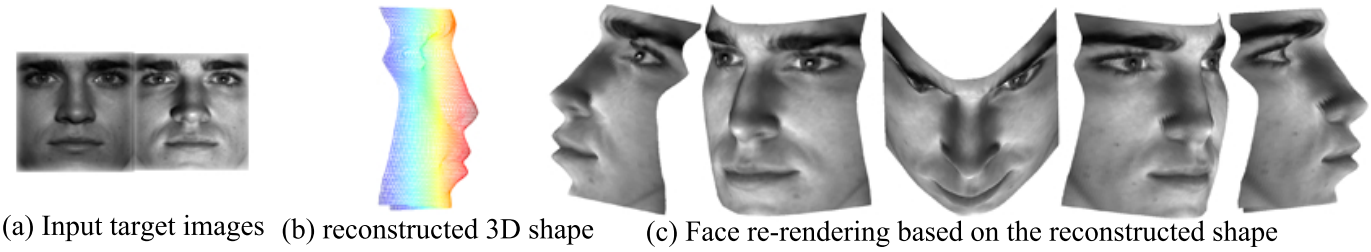


Fig. 7. 3D face shape reconstruction using two target images. For better visualization, facial textures have been mapped on the face shape.

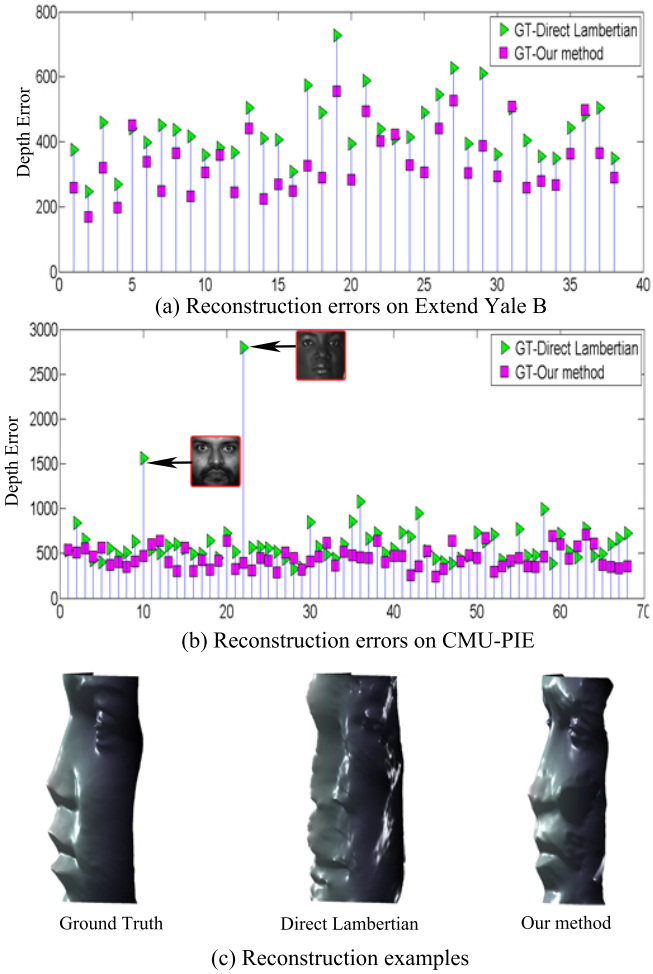


Fig. 6. Reconstruction errors of our method and direct Lambertian method for different subjects on Yale B and CMU-PIE data sets, respectively. Each face image in (b) indicates one of the input images for the corresponding face to be reconstructed. For these two extreme sample face images, the reconstruction errors of different methods are quite different. The reconstruction results of one extreme sample are shown in (c).

future, we would like to extend our method to work in a more complicated lighting environment.

Acknowledgment

The authors would like to thank Dr. Jun-Yong Zhu for helpful discussion on this project. This research was supported by Natural Science Foundation Of China (No. 61102111, No.61173084)

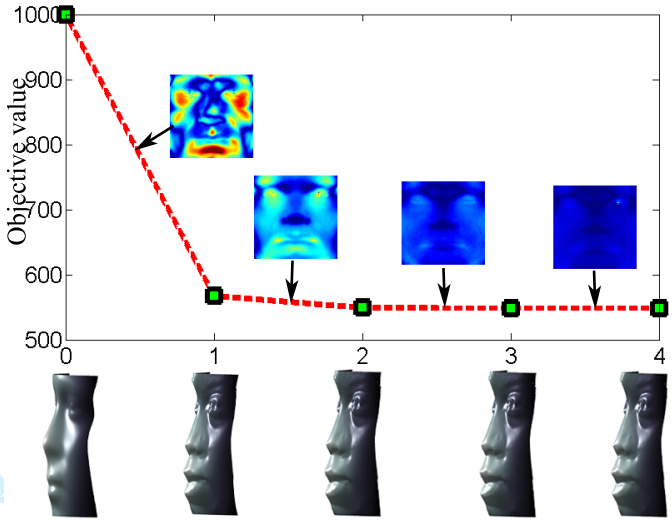


Fig. 8. Convergence of the proposed method. The reconstructed shape at each iteration as well as the deviation map between two neighboring iterations are displayed. A brighter color indicates a larger deviation value. The X-axis and Y-axis indicate the number of iterations and the objective function value Eq. (13), respectively.

References

[1] E. A, A. T. A, and S. S. T., "The geometry of algorithms with orthogonality constraints," *SIAM journal on Matrix Analysis and Applications*, vol. 20(2), pp. 303–353, 1998.

[2] J. J. Atick, P. A. Griffin, and A. N. Redlich, "Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images," *Neural Computation*, vol. 8, no. 6, pp. 1321–1340, 1996.

[3] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

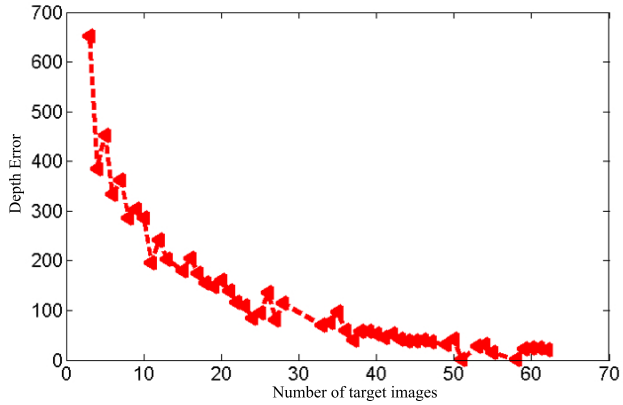
[4] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[5] ——, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[6] Y. Censor, *Parallel optimization: Theory, algorithms, and applications*. Oxford University Press, 1997.

[7] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Self-taught clustering," in *International conference on Machine learning (ICML) 2008*. ACM, pp. 200–207.

[8] R. Dovgard and R. Basri, "Statistical symmetric shape from shading for 3d structure recovery of faces," in *European Conference on Computer Vision (ECCV) 2004*. Springer, pp. 99–113.



(a) reconstruction errors

(b) reconstructions with 2, 6, 8, 20 and 64 target images, respectively.

Fig. 9. Influence of the input target image number.

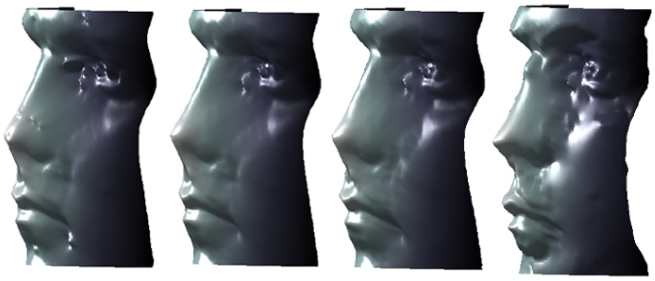


Fig. 10. Reconstruction by setting $\gamma = 0, 1000, 3000$ and 6000 (from left to right), respectively.

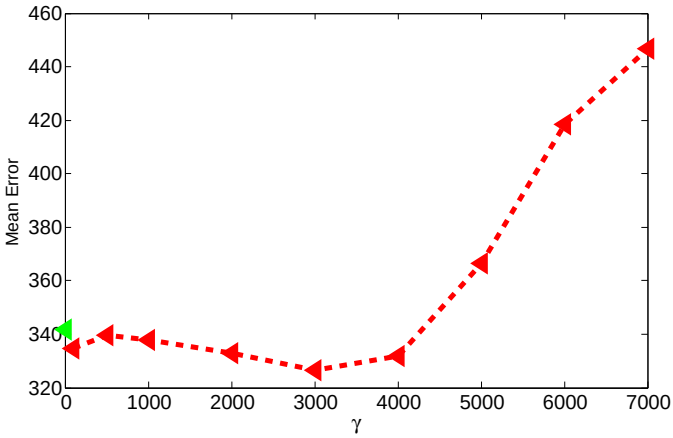


Fig. 11. Mean reconstruction errors with different γ. The green point indicates the error obtained by setting $\gamma = 0$.

Fig. 12. Comparison of 3D face shape reconstruction using different methods. The first row are the results using scanned face images as reference shapes, and the second row are the results using reference shapes inferred from 2D images by our model.

[9] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 439–451, 1988.

[10] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[11] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3d face recognition database," http://live.ece.utexas.edu/research/texas3dfr/index.htm.

[12] T. Hassner, "Viewing real-world faces in 3d," in *IEEE International Conference on Computer Vision (ICCV) 2013*.

[13] B. K. P. Horn, *Robot Vision: MIT ELECTRICAL ENGINEERING AND COMPUTER SCIENCE SE*. the MIT Press, 1986.

[14] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "A novel geometric facial representation based on multi-scale extended local binary patterns," in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011, pp. 1–7.

[15] H. Jeffreys, "cartesian tensor." Cambridge Univ. Press, 1931.

[16] I. Kemelmacher and R. Basri, "Molding face shapes by example," in *European Conference on Computer Vision (ECCV) 2006*. Springer, pp. 277–288.

[17] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 394–405, 2011.

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

[19] P. Kovesi, "Shapelets correlated with surface normals produce surfaces," in *IEEE International Conference on Computer Vision (ICCV) 2005.*, vol. 2, pp. 994–1001.

[20] R. Kumar, A. Barmpoutis, A. Banerjee, and B. C. Vemuri, "Non-lambertian reflectance modeling and shape recovery of faces using tensor splines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 533–567, 2011.

[21] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," in *Advances in neural information processing systems (NIPS 2006)*, pp. 801–808.

[22] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, "A bilinear illumination model for robust face recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV) 2005*, vol. 2, pp. 1177–1184.

[23] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[24] M. Lee and C.-H. Choi, "A robust real-time algorithm for facial shape recovery from a single image containing cast shadow under general, unknown lighting," *Pattern Recognition*, vol. 46, no. 1, pp. 38–44, 2013.

[25] ——, "Real-time facial shape recovery from a single image under general, unknown lighting by rank relaxation," *Computer Vision and Image Understanding*, vol. 120, pp. 59–69, 2014.

[26] Z. Lei, Q. Bai, R. He, and S. Z. Li, "Face shape recovery from a single image using cca mapping between tensor spaces," in *IEEE Conference*

on Computer Vision and Pattern Recognition (CVPR) 2008., pp. 1–7.

[27] H. Li, J. Yu, Y. Ye, and C. Bregler, "Realtime facial animation with on-the-fly correctives," *ACM Transactions on Graphics*, vol. 32, no. 4, July 2013.

[28] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[29] B. T. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, no. 6, pp. 311–317, 1975.

[30] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *Proceedings of the 24th international conference on Machine learning (ICML) 2007*. ACM, pp. 759–766.

[31] A. Saxena, M. Sun, and A. Y. Ng, "Make3d: Learning 3d scene structure from a single still image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 824–840, 2009.

[32] A. Shashua and T. R. Raviv, "The quotient image: Class based re-rendering and recognition with varying illuminations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129–159, 2001.

[33] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi, "A biquadratic reflectance model for radiometric image analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012*, pp. 230–237.

[34] I. Shimshoni, Y. Moses, and M. Lindenbaum, "Shape reconstruction of 3d bilaterally symmetric surfaces," *International Journal of Computer Vision*, vol. 39, no. 2, pp. 97–110, 2000.

[35] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 46–51.

[36] W. A. Smith and E. R. Hancock, "Recovering facial shape using a statistical model of surface normal direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1914–1930, 2006.

[37] D. Vlasic, M. Brand, H. Pfister, and J. Popović, "Face transfer with multilinear models," in *ACM Transactions on Graphics*, vol. 24, no. 3, 2005, pp. 426–433.

[38] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas, "Expression flow for 3d-aware face component transfer," in *ACM Transactions on Graphics)*, vol. 30, no. 4, 2011, p. 60.

[39] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690–706, 1999.

[40] W. Y. Zhao and R. Chellappa, "Symmetric shape-from-shading using self-ratio image," *International Journal of Computer Vision*, vol. 45, no. 1, pp. 55–75, 2001.

[41] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012*, pp. 2879–2886.

For Peer Review