frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 13 April 2021
doi: 10.3389/fnbot.2021.652562

Geometrical Consistency Modeling on B-Spline Parameter Domain for 3D Face Reconstruction From Limited Number of Wild Images

Weilong Peng[1], *Yong Su*[2], *Keke Tang*[3]*, *Chao Xu*[4], *Zhiyong Feng*[4]* and *Meie Fang*[1]

[1] School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, China, [2] Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China, [3] Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China, [4] College of Intelligence and Computing, Tianjin University, Tianjin, China

A number of methods have been proposed for face reconstruction from single/multiple image(s). However, it is still a challenge to do reconstruction for limited number of wild images, in which there exists complex different imaging conditions, various face appearance, and limited number of high-quality images. And most current mesh model based methods cannot generate high-quality face model because of the local mapping deviation in geometric optics and distortion error brought by discrete differential operation. In this paper, accurate geometrical consistency modeling on B-spline parameter domain is proposed to reconstruct high-quality face surface from the various images. The modeling is completely consistent with the law of geometric optics, and B-spline reduces the distortion during surface deformation. In our method, 0th- and 1st-order consistency of stereo are formulated based on low-rank texture structures and local normals, respectively, to approach the pinpoint geometric modeling for face reconstruction. A practical solution combining the two consistency as well as an iterative algorithm is proposed to optimize high-detailed B-spline face effectively. Extensive empirical evaluations on synthetic data and unconstrained data are conducted, and the experimental results demonstrate the effectiveness of our method on challenging scenario, e.g., limited number of images with different head poses, illuminations, and expressions.

Keywords: 3D face modeling, B-spline, face reconstruction, geometrical consistency, parametric domain

OPEN ACCESS

Edited by:
Wei Shi,
Carleton University, Canada

Reviewed by:
Xiaofei Xie,
Nanyang Technological University, Singapore
Weisheng Li,
Chongqing University of Posts and Telecommunications, China
Min Meng,
Guangdong University of Technology, China

***Correspondence:**
Keke Tang
tangbohutbh@gmail.com
Zhiyong Feng
zyfeng@tju.edu.cn

Received: *12 January 2021*
Accepted: *15 March 2021*
Published: *13 April 2021*

Citation:
Peng W, Su Y, Tang K, Xu C, Feng Z and Fang M (2021) Geometrical Consistency Modeling on B-Spline Parameter Domain for 3D Face Reconstruction From Limited Number of Wild Images. Front. Neurorobot. 15:652562. doi: 10.3389/fnbot.2021.652562

1. INTRODUCTION

3D face has been extensively applied in the areas of face recognition (Artificial and Aryananda, 2002; Mian et al., 2006), expression recognition (Zhang et al., 2015). These face analysis technologies are of significance for human-robot cooperative tasks in a safe and intelligent state (Maejima et al., 2012). So 3D face reconstruction is a import topic, and it is meaningful to reconstruct specific 3D face from person-of-interest images under many challenge scenes. The images under challenge scene are also referred as images in the wild, having following characteristics: (1) significant changes in illuminations across time periods; (2) various face poses caused by different camera sensors and view points; (3) different appearances among different environment; (4) occlusions or redundant backgrounds. More seriously, only limited number of identity images are available under human-robot interaction, surveillance, and mobile shooting scenario as listed in **Figure 1**, sometimes.

As a whole, reconstruction technologies include single-image method, multiple images, and even unconstrained images based methods. Recent researches (Kemelmacher and Seitz, 2011; Roth et al., 2015, 2016) prove that good reconstruction depends on two aspects of efforts: (1) enough rich local information, e.g., normal, and (2) a good face prior, e.g., face template. Particularly, the latter is to find an embedding representation with good characteristic to register local information finely.

According to the template representation, these methods can be categorized into three classes: (i) methods without using template, e.g., integration (Kemelmacher and Seitz, 2011) and structure from motion (Koo and Lam, 2008), (ii) methods using a single discrete template, e.g., a reference face mesh (Roth et al., 2015), and (iii) methods using a statistic continuous template, e.g, T-splineMMs (Peng et al., 2017), or discrete template, e.g., 3DMMs (Piotraschke and Blanz, 2016; Roth et al., 2016). The methods with template always generate good global shape compared with those without template, and a statistic template contributes to a better personalization. Therefore, it is very significant to find a excellent template representation for face reconstruction. Mesh model is widely used due to its rapid computation and popularity in computer vision, but it is not well-compatible with geometric optics in vertex level, resulting in local mapping deviation of rays, seen in **Figure 1**. This makes local information not strictly registered physically. Additional, discretization of Laplace-Beltrami operation (LBO), i.e., cotangent scheme (Meyer et al., 2003), may bring a deformation distortion at local, which often happens when images are not enough for high-quality normal estimation. This distortion irregularly occurs at the edge and the location with large curvature changing, e.g., nose and mouth. Lastly the topology-fixed mesh also restricts an extended refinement. All above problem limits reconstruction precision of mesh.

To solve the existing issue in mesh template, we adopt classic B-spline embedding function (Piegl and Tiller, 1997) to register local information and reconstruct face. Firstly, B-spline surface is a parametric surface that can approximate the true shape of an object with fewer parameters (control points) than mesh. It contributes to correct rays in geometric optics, that makes local information, i.e., texture, feature points and normals, accurately registered. Secondly, we use 2nd-order partial derivative operator w.r.t. parameters as the local deformation constraint to reduce the deformation distortion. Lastly, B-spline surface also can be used to generate mesh in any precision or be extended for further refinement. The three characteristics of B-spline face show great advantages over a mesh template based method. Given a collection of images, we use B-spline embedding function as 3D face representation and model 0th- and 1st-order consistency of reconstruction in the parameter domain, which makes BP imaging rays completely compatible with geometric optics. The 0th-order consistency model guarantees that the images are well-registered to surface even if the face images has occlusion or expression; And the 1st-order consistency model guarantees that the surface normals is consistent to the normals estimated from images. Both qualitative and quantitative experiments are conducted and compared with other methods.

In a nutshell, there are two primary contributions:

1. Pinpoint geometrical consistency is modeled on B-spline embedding function for face reconstruction from multiple images, completely consistent with the law of geometric optics.
2. 0th- and 1st-order consistency conditions and its a practical solution is proposed to optimize B-spline face effectively, which is able to handle variations such as different poses, illuminations, and expressions with limited of number images.

In the following, we will first review related work in section 2. Section 3 provides a geometric modeling of multiple BP imaging in image-based stereo for our problem. We introduce the B-spline embedding and its brief representations in section 4 and present consistency modeling for B-spline face reconstruction in

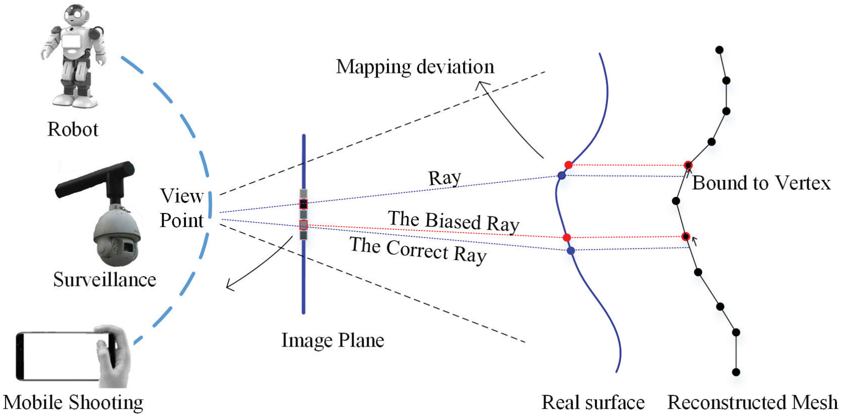

FIGURE 1 | Geometric optics of BP (i.e., back projection) imaging on two types of surfaces: the correct ray lines go through the blue points on the true shape, while the biased ones go through red points on the mesh shape because the cross point between a ray and mesh is bounded to vertex. The difference between red point and the blue point is referred to local mapping deviation.

section 5. In addition, a practical solution is proposed in section 6. We conduct experiment in section 7 and conclude in section 9.

2. RELATED WORK

2.1. 3D Face Required Scenes

With the development of robots and AIoT (Qiu et al., 2020), vision will play an very important role in safety (Khraisat et al., 2019; Li et al., 2019), scene and human understanding (Zhang et al., 2015; Meng et al., 2020). As a base technology, 3D face contributes to the scenes greatly. For example, to build humanoid robots that interact in a human-understanding manner, automatic face, and expression recognition is very import (Zhang et al., 2015). The recognition during real-life human robot interaction could still be challenging as a result of subject variations, illumination changes, various pose, background clutter, and occlusions (Mian et al., 2006). However, humanoid robot API of original version cannot always be able to handling such challenges. Optimal, robust, and accurate automatic face analysis is thus meaningful for the real-life applications since the performance of facial action and emotion recognition relies heavily on it. Many parametric approaches like 3DMMs (Blanz and Vetter, 1999; Blanz et al., 2004) and face alignment with 3D solution (Zhu et al., 2016) in the computer vision field have been proposed to estimate head pose, recognition identity, and expression from real-life images to benefit subsequent automatic facial behavior perception to address the above issues. Therefore, 3d face modeling in a humanoid robot view is of great significant to handling the challenging face analysis during interaction.

2.2. 2D Images Based Face Reconstruction

2D methods generally cover several kinds of fundamental methods including Structure from Motion (SFM) (Tomasi and Kanade, 1992), Shape from Shading (SFM) (Zhang et al., 1999), 3D Morphable Model (3DMM) (Blanz and Vetter, 1999; Blanz et al., 2004), and Deep learnings (Richardson et al., 2017; Deng et al., 2019). SFM methods compute the positions of surface points based on an assumption that there exists a coordinate transformation between the image coordinate system and the camera coordinate system. And SFS methods compute surface normals with an assumption that the subject surface is of Lambertian and under a relatively distant illumination. And the idea of 3DMM is that human faces are within a linear subspace, and that any novel face shape can be represented by a linear combination of shape eigenvectors deduced by PCA. SFS and SFM give the geometrical and physical descriptions of face shape and imaging, and 3DMM concentrates on the statistical explanation of 3D meshes or skeletons. Deep learning methods infer 3D face shape or texture (Lin et al., 2020) by statistically learning mapping between face images and their 3D shapes (Zhou et al., 2019). Being limited to data size, most of them relies 3DMM or PCA for synthesizing supplementary ground truths (Richardson et al., 2016) or as a priori (Tran et al., 2017; Gecer et al., 2019; Wu et al., 2019), resulting absence of shape detail. It's believed that face reconstruction is rather a geometrical optimization problem than a statistical problem, as 3DMM is more suitable to be an assistant of the geometrical method when building detailed shape, e.g., that by Yang et al. (2014).

2.3. Shape in Shading and Structure in Motion

SFS has been widely used for reconstruction, e.g., single-view reconstruction (Kemelmacher Shlizerman and Basri, 2011), multiple frontal images based reconstruction (Wang et al., 2003), and unconstrained image based reconstruction (Kemelmacher and Seitz, 2011; Roth et al., 2015). As single-view is ill posed (Prados and Faugeras, 2005), a reference is always needed (Kemelmacher Shlizerman and Basri, 2011). For unconstrained images, photometric stereo is applied to obtain accurate normals locally (Kemelmacher and Seitz, 2011; Roth et al., 2015). SFM uses multiple frame or images to recover sparse 3D structure of feature points of an object (Tomasi and Kanade, 1992). Spatial-transformation approach (Sun et al., 2013) only estimates the depth of facial points. Bundle adjustment (Agarwal et al., 2011) fits the large scale rigid object reconstruction, but it cannot generate the dense model of non-rigid face. Incremental SFM (Gonzalez-Mora et al., 2010) is proposed to build a generic 3D face model for non-rigid face. The work by Roth et al. (2015) optimizes the local information with normals from shading, based on a 3D feature points-driven global warping. Therefore, shading and motion are important and very distinct geometric information of face, and they enhance the reconstruction when being combined. In our method, 0th- and 1st-order consistency of stereo is modeled to integrate the advantages of both shading and motion information.

2.4. Facial Surface Modeling

Surface modeling is dependent on the data input (point cloud, noise, outlier, etc), output (point cloud, mesh, skeleton), and types of shape (man-made shape, organic shape). Point cloud, skeleton, and mesh grid are the widely used man-made shape type for face reconstruction. Lu et al. (2016) present an a stepwise tracking method approach to reconstruct 3D B-spline space curves from planar orthogonal views through minimizing the energy function with weight values. Spatial transformation method (Sun et al., 2013) estimates positions of sparse facial feature points. Bundle adjustment builds the dense point cloud for large scale rigid object with a great number of images (Agarwal et al., 2011). Heo and Savvides (2009) reconstruct face dense mesh based on skeleton and 3DMM. Kemelmacher and Seitz (2011) apply integration of normals to get discrete surface points, which may produce incredible depth when the recovered normals are unreliable. Roth et al. (2015) reconstruct face mesh based on Laplace mesh editing, which may produce local mesh distortion after several iterations of local optimization. In work of mesh reconstruction, surface-smoothness priors is also needed to guarantee the smoothness of discrete mesh based on point cloud, e.g., radial basis function (Carr et al., 2001) and Poisson surface reconstruction (Kazhdan et al., 2006). Due to the fact that the point cloud and 3D mesh are discontinuous geometric shape, they cannot approximate the true shape of a face of arbitrary precision. There have been works of fitting B-splines to noisy 3D data, like Hoch et al. (1998). B-spline face model is

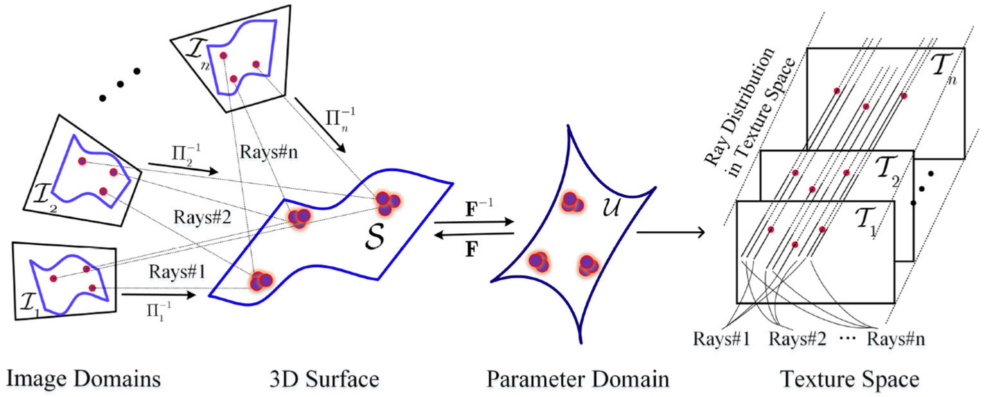

FIGURE 2 | Geometric optics of multiple BP imaging.

a continuous free-form surface that can be reconstructed from images directly, instead of intermediate point data, but it is not a detailed model by only using structure optimization (Peng et al., 2016). Because B-spline surface is a special case of NURBS (Non-Uniform Rational B-Spline) (Piegl and Tiller, 1997), it can also be imported to 3D modeling software like Rhino3D for further editing, analysis, and transformation conveniently by adjusting the B-spline control points. It can also be converted into mesh model with any precision according to appropriate parameter interval, conveniently, which is meaningful for a system with limited memory.

3. GEOMETRIC MODELING

Our problem modeling is illustrated in **Figure 2**. The domain of input image I_i from a camera is $\mathcal{I}_i \subset \mathbb{R}^2, i = 1, 2, \ldots, n$. Π^{-1} denotes the inverse operator of Π. The camera operator $\Pi_i \in C^{\infty}(\mathbb{R}^3, \mathbb{R}^2)$ map a point $P \in \mathcal{S}$ to $p = \Pi_i(P) \in \mathcal{I}_i$ using weak perspective projection, $i = 1, 2, \ldots, n$. And Π_i^{-1} determines the ray cluster Rays#i of BP imaging from $\mathcal{I}_i, i = 1, 2, \ldots, n$. Let s_i, R_i, and t_i denote scale, rotation, and translation parameter in projection Π_i. The ith projection operation is simply

$$\Pi_i(P) \triangleq s_i \cdot R_{i,[1,2]} \cdot P + t_i. \tag{1}$$

$R_{i,[1,2]}$ expresses the first two rows of R_i.

Let $\mathcal{U} \subset \mathbb{R}^2$ denote the parameter domain of human face surface. A certain embedding $\mathbf{F} \in C^1(\mathcal{U}, \mathbb{R}^3)$ maps a point $\mathbf{u} \in \mathcal{U}$ to the 3D point $P \in \mathcal{S}$. $\mathbf{F}^{-1}$ denote the inverse operator of $\mathbf{F}$. It is thus clear that different embedding F determine different face shapes. According to the geometric optics of BP imaging, a image point $p \in \mathcal{I}_i$ is back projected onto a point $\mathbf{u} = \tau_i(p) \in \mathcal{U}$ *via* the operator

$$\tau_i \triangleq \mathbf{F}^{-1} \circ \Pi_i^{-1}. \tag{2}$$

Therefore, an image I_i in the i-th view is mapped to surface S, and then is mapped to texture space by

$$\mathcal{T}_i \triangleq I_i \circ \tau_i^{-1}, \tag{3}$$

where we define

$$(I \circ \tau^{-1})(\mathbf{u}) \triangleq I(\Pi(\mathbf{F}(\mathbf{u}))), \quad \mathit{for} \ \ \mathbf{u} \in \mathcal{U}. \tag{4}$$

In fact, $\tau_i, i = 1, 2, \ldots, n$ generate discrete and inconsistent rays mapping in texture space because of the discrete and different images domains, as well as the noises, seen in **Figure 2**.

3.1. 0th- and 1st-Order Consistency

Generally, the problem is how to determine $\mathbf{F}$ according to from multiple images. If all images are the captures of a same $\mathcal{S}$, all $\{\mathcal{T}_i\}_{i=1:n}$ in texture space are hoped to be highly consistent in the geometry.

First, that satisfies

$$< \hat{\mathbf{F}}, \{\hat{\Pi}_i\} >= \underset{\mathbf{F},\{\Pi_i\}}{\arg\min} \operatorname{rank}([\operatorname{vec}(\mathcal{T}_1), \operatorname{vec}(\mathcal{T}_2), \ldots, \operatorname{vec}(\mathcal{T}_n)]), \tag{5}$$

with $\mathcal{T}_i = (I_i \circ \tau_i^{-1})^{\#}, i = 1, 2, \ldots, n$. And $(\cdot)^{\#}$ is a composition operator of fitting and sampling, to handle the inconsistency. It firstly fits a texture function based on the discrete texture and parameters mapped from one image, and then samples texture intensity values at unified parameter points $\{\mathbf{u}_j\}_{j=1:N_p}$.

Second, it satisfies

$$\begin{cases} \dfrac{\frac{\partial \mathbf{F}}{\partial u} \times \frac{\partial \mathbf{F}}{\partial v}}{\left\| \frac{\partial \mathbf{F}}{\partial u} \times \frac{\partial \mathbf{F}}{\partial v} \right\|} = \boldsymbol{n}, \\ \rho_j \boldsymbol{n}_j \cdot \boldsymbol{l}_i = \mathcal{T}_i|_{\mathbf{u}_j}. \end{cases} \tag{6}$$

which describes the equivalence relation between normal $\boldsymbol{n}$ and 1st-order partial derivative in the first formulation, and the equivalence relation among albedo ρ, normal $\boldsymbol{n}$, light direction $\boldsymbol{l}$, and image intensity $\mathcal{T}$ in the second. This follows a linear photometric model, as seen in **Figure 3**.

We refer to Equations (5) and (6) as 0th- and 1st- order consistence equations in 3D surface reconstruction respectively. Generally, researchers solve any one of the two consistence problem to reconstruct 3D surface, classically, by multi-view stereo (Seitz et al., 2006) for 0th-order consistence problem, or by photometric stereo (Barsky and Petrou, 2003) for the 1st-order one.

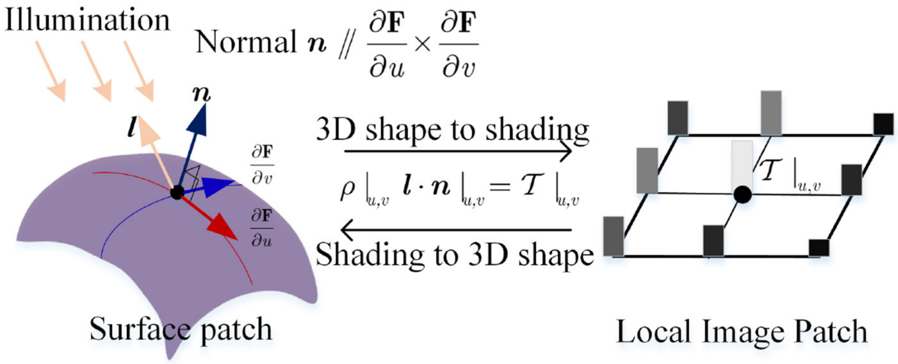
Illumination
Normal $\boldsymbol{n} \mathbin{/\!/} \frac{\partial \mathbf{F}}{\partial u} \times \frac{\partial \mathbf{F}}{\partial v}$
$\boldsymbol{n}$
$\boldsymbol{l}$
$\frac{\partial \mathbf{F}}{\partial v}$
$\frac{\partial \mathbf{F}}{\partial u}$
3D shape to shading
$\rho \big|_{u,v} \ \boldsymbol{l} \cdot \boldsymbol{n} \big|_{u,v} = \mathcal{T} \big|_{u,v}$
Shading to 3D shape
$\mathcal{T} \big|_{u,v}$
Surface patch
Local Image Patch

FIGURE 3 | The consistency mapping equivalence between embedding **F** and the image intensity.

3.2. Embedding F

There are several types of representation for embedding **F**, such as discrete mesh and C^2 parametric surface. In fact the representation type of **F** also affects the reconstruction effect. Intuitively for mesh, on one hand there exists mapping deviation of rays from image points to vertices of mesh, which contributes to inaccurate texture charts $\{\mathcal{T}_i\}_{i=1:n}$ and affects the accuracy of reconstruction. On the other, discrete differential operator, i.e., LBO (Meyer et al., 2003), brings potential distortion error when there exists obtuse triangles in the mesh caused by error local normal. Additionally, the precision of mesh also limit the detail of reconstruction.

We consider to apply C^2 parametric surface as the representation of face. Generally, B-spline surface is recommended because of its advantages of good locality over other types of surfaces such as polynomial surface and Bessel surface. By B-spline surface, it doesn't exist mapping deviation in geometric optics, and it avoids the potential distortion brought by discrete differential operator. Therefore, accurate and continuous back projection texture charts $\{\mathcal{T}_i\}_{i=1:n}$ can be generated based on Equations (2), (3), and (5). Then accurate reconstruction can be implemented based on Equation (6). What's more, the precision can be enhanced for high-detailed reconstruction by inserting control points.

4. B-SPLINE FACE EMBEDDING F, AND THE 0TH-, 1ST-, 2ND–ORDER REPRESENTATION

The human face is assumed to be a uniform B-spline surface $\mathcal{S}$ of degree 4 × 4, with $\mathbf{B} = \{\boldsymbol{b}_{mn}\}_{M\times N}$ as its control points. In parameter domain $\mathcal{U}$, knots $U = \{u_m\}_{m=1}^{M+4}$ and $V = \{v_n\}_{n=1}^{N+4}$ split uv parameter plane into uniform grid. Let $\mathbf{u}$ denote parameter point (u, v). The surface function is

$$\mathbf{F}(\mathbf{u}) = \sum_{m=1}^{M}\sum_{n=1}^{N} R_{m,n}(\mathbf{u})\boldsymbol{b}_{mn},$$

with $R_{m,n}(\mathbf{u}) = N_{m,4}(u) \cdot N_{n,4}(v)$ and

$$\begin{cases} N_{i,1}(w) = \begin{cases} 1 \ \ u_i \le w < u_{i+1}, \\ 0 \ \ otherwise, \end{cases} \\ N_{i,j}(w) = \frac{(w-u_i)\cdot N_{i,j-1}(w)}{u_{i+j-1}-u_i} + \frac{(u_{i+j}-w)\cdot N_{i+1,j-1}(w)}{u_{i+j}-u_{i+1}}, (j = 4, 3, 2). \end{cases}$$

F is C^2, meaning that it can approximate the true shape in arbitrary uv precision with deterministic k-ordered partial derivative $\frac{\partial^k \mathbf{F}}{\partial u^k}$ and $\frac{\partial^k \mathbf{F}}{\partial v^k}$, $k = 1, 2$, and $\frac{\partial^2 \mathbf{F}}{\partial u \partial v}$.

4.1. 0th-Order Representation

We give a more brief formulation of 0th-order representation as follows:

$$\mathbf{F}|_{\mathbf{u}} = \mathbf{T}|_{\mathbf{u}} \cdot \mathbf{b}, \tag{7}$$

where **b** denotes a $3MN \times 1$ vector storing B-spline control points, and $\mathbf{T}|_{\mathbf{u}}$ denotes a sparse $3 \times 3MN$ matrix stacking the 0th-order coefficients at parameter $\mathbf{u} \in \mathcal{U}$.

In fact, we needn't consider all 3D points mapping to 2D images when estimating a operator Π. Instead, we only consider f landmark points on human face as shown in **Figure 4**, and their brief formulation is

$$\mathbf{F}|_{\mathbf{u}(l_i)} = \mathbf{T}|_{\mathbf{u}(l_i)} \cdot \boldsymbol{b}, i = 1, 2, \ldots, f, \tag{8}$$

where $\mathbf{u}(l_i)$ is the parameter point of the i-th feature point, $i = 1, 2, \ldots, f$. The landmarks cover a sparse structure of face.

4.2. 1st-Order Representation

The 1st-order partial derivatives of **F** w.r.t u and v are

$$\mathbf{F}'_u(\mathbf{u}) = \sum_{m=1}^{M}\sum_{n=1}^{N} N'_{m,4}(u) \cdot N_{n,4}(v)\boldsymbol{b}_{mn}$$
$$= \sum_{m=1}^{M}\sum_{n=1}^{N} (\frac{4}{u_{m+4}-u_i}N'_{m,3}(u) - \frac{4}{u_{m+5}-u_{m+1}}N'_{m+1,3}(u)) \cdot N_{n,4}(v)\boldsymbol{b}_{mn}$$

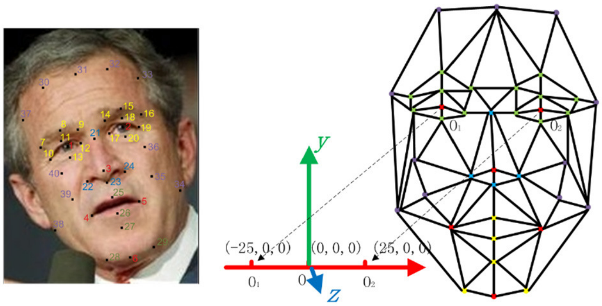

FIGURE 4 | Face structure defined by 40 feature points: the left side shows the point positions in a face image; the right side shows the structure topology with eye center points of O_1 (−25, 0, 0) and O_2 (25, 0, 0) in 3*D* space, which looks like a frontal 2D face structure from the direction of normal (0, 0, 1). (The face image used in the figure comes from LFW database[1]).

and

$$\mathbf{F}'_v(\mathbf{u}) = \sum_{m=1}^{M}\sum_{n=1}^{N} N_{m,4}(u)\cdot N'_{n,4}(v)\boldsymbol{b}_{mn}$$
$$= \sum_{m=1}^{M}\sum_{n=1}^{N} N_{m,4}(u)\cdot\left(\frac{4}{u_{n+4}-u_n}N'_{n,3}(v) - \frac{4}{u_{n+5}-u_{n+1}}N'_{n+1,3}(v)\right)\boldsymbol{b}_{mn}$$

respectively.

Similarly, we give a more brief formulation of 1st-order partial derivative as follows:

$$\begin{cases} \mathbf{F}'_u|_{\mathbf{u}} = \mathbf{T}_1|_{\mathbf{u}}\cdot\mathbf{b} \\ \mathbf{F}'_v|_{\mathbf{u}} = \mathbf{T}_2|_{\mathbf{u}}\cdot\mathbf{b} \end{cases}, \tag{9}$$

where $\mathbf{T}_1|_{\mathbf{u}}$ and $\mathbf{T}_2|_{\mathbf{u}}$ denote the matrixes stacking the 1st-order coefficients w.r.t u and v, respectively.

Therefore, the surface normal vector at $\mathbf{u}$ can be computed by the cross product

$$\boldsymbol{n}|_{\mathbf{u}} = \frac{\boldsymbol{F}'_u|_{\mathbf{u}}\times\boldsymbol{F}'_v|_{\mathbf{u}}}{\left\|\boldsymbol{F}'_u|_{\mathbf{u}}\times\boldsymbol{F}'_v|_{\mathbf{u}}\right\|} = s|_{\mathbf{u}}\cdot\boldsymbol{F}'_u|_{\mathbf{u}}\times\boldsymbol{F}'_v|_{\mathbf{u}}, \tag{10}$$

which is the key information for detailed reconstruction using photometric stereo method.

4.3. 2nd-Order Representation

And similarly, the 2nd-order partial derivatives w.r.t. u and v, respectively are

$$\begin{cases} \mathbf{F}''_{uu}|_{\mathbf{u}} = \mathbf{T}_{11}|_{\mathbf{u}}\cdot\mathbf{b} \\ \mathbf{F}''_{vv}|_{\mathbf{u}} = \mathbf{T}_{22}|_{\mathbf{u}}\cdot\mathbf{b} \end{cases}, \tag{11}$$

where $\mathbf{T}_{11}|_{\mathbf{u}}$ and $\mathbf{T}_{22}|_{\mathbf{u}}$ denote the matrixes stacking the 2nd-order coefficients w.r.t u and v, respectively. The 2nd-order information can be used for smooth control during optimization.

Based on face surface embedded with B-spline function, we present the pinpoint 0th- and 1st-order geometric consistency conditions in the following section.

5. CONSISTENCY MODELING IN B-SPLINE FACE RECONSTRUCTION

Reconstruction problem is to compute $\mathbf{F}$ by solving 0th-order consistence of Equation (5) or 1st-order consistence of Equation (6). Generally, two consistency conditions are combined for face reconstruction considering that estimating abundant consistent points in images is limited and that the estimated normals are unfaithful. Furthermore, how to obtain the accurate registration of 0th- and 1st-order information is the most important to high-detailed B-spline reconstruction.

The well-registered textures are low-rank structures of the back projection texture charts. But in practice, they can be easily violated due to the presence of partial occlusions or expressions in the images captured. Since these errors typically affect only a small fraction of all pixels in an chart, they can be modeled as sparse errors whose nonzero entries can have arbitrarily large magnitude.

5.1. Modeling Occlusion and Expression Corruptions in 0th-Order Consistence

Let e_i represent the error corresponding to image I_i such that the back projection texture charts $\mathcal{T}_i = (I_i \circ \tau_i^{-1})^{\#} - e_i = \mathcal{T}_i^e - e_i, i = 1, 2, \ldots, n$ are well registered to the surface $\mathbf{F}$, and free of any corruptions or expressions. Also combining with 0th-order representation of B-spline face in Equation (7), the formulation (5) can be modified as follows:

$$\begin{aligned} <\hat{\mathbf{b}}, \{\hat{\Pi}_i\}, \hat{\mathbf{D}}, \hat{\mathbf{E}}> = \arg \lim_{\mathbf{b},\{\Pi_i\},\mathbf{D},\mathbf{E}} \|\mathbf{D}\|_* + \eta\,\|\mathbf{E}\|_1\,, \\ s.t. \left\|\mathbf{D}^e - \mathbf{D} - \mathbf{E}\right\|_F \le \varepsilon. \end{aligned} \tag{12}$$

where $\mathbf{D}^e = \left[\mathrm{vec}(\mathcal{T}_1^e), \mathrm{vec}(\mathcal{T}_2^e), \ldots, \mathrm{vec}(\mathcal{T}_n^e)\right]$ and $\mathbf{E} = [\mathrm{vec}(e_1), \mathrm{vec}(e_2), \ldots, \mathrm{vec}(e_n)]$.

However, the solution $\hat{\mathbf{b}}$ of face surface $\mathcal{S}$ is not unique if all images are in similar views. And the reconstruction is not high-detailed even if we can make a unique solution by applying a prior face template. So we also need to model high details in 1st-order consistence.

5.2. Modeling High Details in 1st-Order Consistence

The resolution of reconstruction is determined by the density of correctly estimated normals. To enhance the resolution of B-spline surface, we use operator $(\cdot)^{\#}$ to sample N_p dense parameter points $\{\mathbf{u}_j\}_{j=1:N_p}$ on the domain $\mathcal{U}$ for the problem of Equation (6).

Then the well-registered and dense texture are obtained by

$$\mathcal{T}_i|_{\mathbf{u}_j} = \hat{\mathbf{D}}_{ji}, \tag{13}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, N_p$.

According to Lambertian illumination model seen in Equation (6), dense normals $\boldsymbol{n}_j$ as well as light $\boldsymbol{l}_i$ can be computed from the shading (intensity) of charts $\mathcal{T}_i$ by SVD method.

Finally, the high detailed reconstruction must satisfy

$$\min_{\mathbf{F}} \sum_{j=1}^{N_p} \left\|\boldsymbol{n}_j - s|_{\mathbf{u}_j}\mathbf{F}'_u|_{\mathbf{u}_j}\times\mathbf{F}'_v|_{\mathbf{u}_j}\right\|_2^2. \tag{14}$$

By putting Equation (9) into Equation (14), we get

$$\min_{\mathbf{b}} \sum_{j=1}^{N_p} \left\| \boldsymbol{n}_j - s|_{\mathbf{u}_j} (\mathbf{T}_1|_{\mathbf{u}_j} \cdot \mathbf{b}) \times (\mathbf{T}_2|_{\mathbf{u}_j} \cdot \mathbf{b}) \right\|_2^2. \tag{15}$$

Conditions of both Equations (6) and (15) have to be considered for a good reconstruction, which is very difficult. Therefore, we propose a practical solution that combining both 0th- and 1st-order consistence.

6. PRACTICAL SOLUTION COMBINING 0TH- AND 1ST-ORDER CONSISTENCE

The problems of both 0th-order consistence and 1st-order consistence are difficult to solve. For , Jacobian matrices w.r.t. $\{\tau_i^{-1}\}_{i=1:n}$ have to be computed, which is computing-expensive. And the solution of Equation (15) is not unique, either. Therefore, we aim to find a practical solution to handle both two consistence conditions in this section. We first define the subproblem for each condition, and then provide a iterative algorithm.

6.1. 0th-Order Solution

In Equation (6), three kind of parameters including camera parameters $\{\Pi_i\}_{i=1:n}$, surface parameters $\mathbf{F}$ (or $\mathbf{b}$), and texture parameters $\{\mathcal{T}_i\}_{i=1:n}$ (or $\mathbf{D}$) need to be computed, but they are difficult to be solved simultaneously. We adopt to optimize them by turns, instead.

6.1.1. Estimating Π_i

According to linear transformation from 3D to 2D in Equation (1), we can estimate scale s_i, rotation R_i and translationt_i of landmarks for each image $I_i, i = 1, 2, \ldots, n$ based on the and SVD method (Kemelmacher and Seitz, 2011). The image landmarks are detected by a state-of-art detector (Burgos-Artizzu et al., 2013) that has a similar high performance to human. And the 3D landmarks are defined on a B-spline face template with control point parameter $\mathbf{b}_0$, according to Equation (8).

6.1.2. Estimating $\boldsymbol{b}$

Let $\mathbf{f}$ denote a $2nf \times 1$ vector stacking f landmarks of n images, and $\mathbf{P}$ denote a $2nf \times 3f$ projection matrix stacking n views of parameters $s_i R_{i,[1,2]}$, and $\mathbf{t}$ denote a $2nf \times 1$ vector stacking f translation. The update of $\mathbf{b}$ can be implemented by solving:

$$\min_{\boldsymbol{b}} \left\| \mathbf{f} - \mathbf{t} - \mathbf{P} \cdot \mathbf{T}^{\#l} \mathbf{b} \right\|_2^2 + \zeta \left\| (\mathbf{T}_{11}^{\#} + \mathbf{T}_{22}^{\#})(\mathbf{b} - \mathbf{b}_0) \right\|_2^2 \tag{16}$$

where the first and the second are 0th- and 2nd-order item, respectively, and ζ is used to balance them. Operator $(\cdot)^{\#l}$ is a sampling operator that selects B-spline coefficients of landmarks at parameters $\{\mathbf{u}(l_i)\}_{i=1:f}$, and $(\cdot)^{\#}$ selects B-spline coefficients at $\{\mathbf{u}_j\}_{i=1:N_p}$. In fact, $\mathbf{T}^{\#l}$ is a $3f \times 3MN$ matrix that stacks $\mathbf{T}|_{\mathbf{u}(l_i)}, i = 1, 2, \ldots, f$, and $\mathbf{T}_{11}^{\#}$ (or $\mathbf{T}_{22}^{\#}$) is a $3f \times 3MN$ matrix that stacks $\mathbf{T}_{11}|_{\mathbf{u}_j}$ (or $\mathbf{T}_{22}|_{\mathbf{u}_j}$), $j = 1, 2, \ldots, N_p$.

The second item also work as a regularization measuring the distance of local information between faces $\mathbf{b}$ and $\mathbf{b}_0$. It helps eliminate affect of geometric rotation brought by 0st-order warping, and guarantee a smoothness changing during optimization. Particularly, ζ cannot be too small, otherwise a fast changing may bring a local optimal.

6.1.3. Estimating $\mathcal{T}_i$

τ_i^{-1} and τ_i is determined by Equation (2) when Π_i and $\mathbf{b}$ is known. Then texture chart with noise is obtained by applying consistent parameter sampling $\mathcal{T}_i^e = (I_i \circ \tau_i^{-1})^{\#}$. Let $\mathbf{D}^e = [\mathrm{vec}(\mathcal{T}_1^e), \mathrm{vec}(\mathcal{T}_2^e), \ldots, \mathrm{vec}(\mathcal{T}_n^e)]$. The update of texture charts is to minimize the following formulation

$$\begin{aligned} < \hat{\mathbf{D}}, \hat{\mathbf{E}} >= \arg \lim_{\mathbf{D},\mathbf{E}} \|\mathbf{D}\|_* + \eta \|\mathbf{E}\|_1 , \\ s.t. \|\mathbf{D}^e - \mathbf{D} - \mathbf{E}\|_F \leq \varepsilon. \end{aligned} \tag{17}$$

which can be solved by Robust PCA (Bhardwaj and Raman, 2016). And let $\mathcal{T}_i|_{\mathbf{u}_j} = \hat{\mathbf{D}}_{ji}$, for $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, N_p$.

6.2. 1st-Order Solution

Firstly, texture charts based photometric stereo method is used to estimate the local normals. Secondly, a normals driven optimization strategy is proposed to optimize the B-spline face.

6.2.1. Estimating $\boldsymbol{n}_j$

According to Photometric stereo, the shape of each point can be solved by the observed variation in shading of the images. Data of n texture charts are input into $M_{n \times N_p}$ for estimating the initial shape $\tilde{S}$ and lighting $\tilde{L}$ by factorizing $M = LS\ via$ SVD (Yuille et al., 1999). $\tilde{L} = U\sqrt{\Sigma}$ and $\tilde{S} = \sqrt{\Sigma} V^T$, where $M = U \Sigma V^T$. To approach the true normal information, we estimate the shape S and ambiguity A by following the work of Kemelmacher and Seitz (2011). Lastly, the normal at j-th point is $\boldsymbol{n}_j = S_j^T$, where S_j is the j-th row of S.

6.2.2. Estimating b

We normalize $\boldsymbol{n}_j$ and stack them into a $3N_p \times 1$ vector $\mathbf{h}$. Equation (15) can be rewritten as

$$\mathbf{O}_1 = \min_{\mathbf{b}} \left\| \mathbf{h} - \Lambda|_{\mathbf{b}} \cdot ((\mathbf{T}_1^{\#} \mathbf{b}) \otimes ((\mathbf{T}_2^{\#} \mathbf{b})) \right\|_2^2 ,$$

where Λ is a $3N_p \times 3N_p$ diagonal matrix that stores $3N_p$ reciprocals of lengths of the normals $\{\boldsymbol{n}_j\}_{j=1:N_p}$; and $(\cdot)^{\#}$ is a selection operator that selects $3N_p$ rows of 1st-order coefficients at parameter $\{\mathbf{u}_j\}_{j=1:N_p}$; and $\mathbf{b}_0$ represent the control points of a B-spline template face. Particularly, symbol $\otimes$ denotes a composite operator of cross product, which makes $\boldsymbol{w} \otimes \boldsymbol{v} = [\boldsymbol{w}_1 \times \boldsymbol{v}_1; \boldsymbol{w}_2 \times \boldsymbol{v}_2; \ldots; \boldsymbol{w}_{N_p} \times \boldsymbol{v}_{N_p}]$, where $\boldsymbol{w}$ and $\boldsymbol{v}$ are $3N_p \times 1$ vectors containing N_p normals.

However, there exists two issues: (1) the low-dimension $\boldsymbol{h}$ may not guarantee an unique solution of high-dimension $\boldsymbol{b}$; and (2) the system is not simply linear, which is difficult to be solved. Therefore, a frontal constraint based on template $\mathbf{b}_0$ is applied to make a unique solution; And a strategy of approximating to linearization is also proposed to make a linear solution.

6.2.2.1. Frontal Constraint

The frontal constraint is a distance measurement condition between surface $\mathcal{S}$ and template w.r.t. x- and y-component:

$$\mathbf{O}_2 = \left\| \mathbf{T}^{\#xy}(\mathbf{b} - \mathbf{b}_0) \right\|_2^2 < \epsilon,$$

where the matrix $\mathbf{T}^{\#xy}$ stacks 0th-order coefficients at parameter $\{\mathbf{u}_j\}_{j=1:N_p}$ corresponding to x- and y- components. Operator $(\cdot)^{\#sxy}$ also sets the coefficients corresponding to z- components to zeros.

Particularly, the first item $\boldsymbol{O}_1$ is not a simple linear form, for which an approximating to linearization is proposed.

6.2.2.2. Approximating to Linearization

According to the characteristics of the cross-product $\otimes$, the first item in $\boldsymbol{O}_1$ can be rewritten as a linear-like formulation:

$$\|\mathbf{h} - \boldsymbol{L}|_{\mathbf{b}} \cdot \mathbf{b}\|_2^2 \quad \text{or} \quad \|\mathbf{h} - \boldsymbol{R}|_{\mathbf{b}} \cdot \mathbf{b}\|_2^2,$$

where $\begin{cases} \boldsymbol{L}|_{\mathbf{b}} = \Lambda|_{\mathbf{b}} \cdot \left[\mathbf{T}_1^{\#}\mathbf{b}\right]_{\otimes} \cdot \mathbf{T}_2^{\#} \\ \boldsymbol{R}|_{\mathbf{b}} = -\Lambda|_{\mathbf{b}} \cdot \left[\mathbf{T}_2^{\#}\mathbf{b}\right]_{\otimes} \cdot \mathbf{T}_1^{(sn)}. \end{cases}$

Particularly, the operation $[\cdot]_{\otimes}$ makes a $3N_p \times 1$ vector $\boldsymbol{w} = \left[\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_{N_p}^T\right]^T$ become a $3N_p \times 3N_p$ sparse matrix $[\boldsymbol{w}]_{\otimes} = diag([\boldsymbol{w}_1]_{\times}, [\boldsymbol{w}_2]_{\times}, \ldots, \left[\boldsymbol{w}_{N_p}\right]_{\times})$, where $[\boldsymbol{w}_i]_{\times} = [0, -\boldsymbol{w}_i^z, \boldsymbol{w}_i^y; \boldsymbol{w}_i^z, 0, -\boldsymbol{w}_i^x; -\boldsymbol{w}_i^y, \boldsymbol{w}_i^x, 0], i = 1, 2, \ldots, N_p$.

If $\mathbf{b}$ is a known parameter, e.g., as $\mathbf{b}_0$, for $\boldsymbol{L}|_{\mathbf{b}}$, the minimization of $\left\|\mathbf{h} - \boldsymbol{L}|_{\mathbf{b}_0} \cdot \mathbf{b}\right\|$ will be a linear system. That is also true for $\boldsymbol{R}|_{\mathbf{b}}$.

In fact, we can use formulation $\left\|\mathbf{h} - \boldsymbol{L}|_{\mathbf{b}_0} \cdot \mathbf{b}\right\|$ to optimize the control points in parameter space of v by fixing u, and use $\left\|\mathbf{h} - \boldsymbol{R}|_{\mathbf{b}_0} \cdot \mathbf{b}\right\|$ to optimize in parameter space of u by fixing v.

Algorithm 1: Iterative Algorithm for B-spline Face Optimization

```
Input: Face images $\{I_i\}_{i=1:n}$, B-spline template face $\mathbf{b}_0$, and
    landmark parameters $\{\mathbf{u}(l_i)\}_{i=1:f}$ in domain $\mathcal{U}$.
Detect facial landmark points of images
while b is not converged do
    do // LOOP1: 0th-order consistence
        Estimate camera parameter $\{\Pi_i\}_{i=1:n}$ according to landmarks.
        Estimate b via Equ(16), and update $\mathbf{b}_0$ with b.
        // Obtain well-registered texture
        Register images to texture space by $\{I_i \circ \tau_i^{-1}\}_{i=1:n}$, and
        build $\mathbf{D}^e$ based on unified parameter $\{\mathbf{u}_j\}_{j=1:N_p}$.
        Solve Equ(17) to obtain $\hat{\mathbf{D}}$.
    while $\|\mathbf{D}\|_* + \eta\,\|\mathbf{E}\|_1$ is not converged // LOOP1 END
    Extract texture charts $\{\mathcal{T}_i\}_{i=1:n}$ from $\hat{\mathbf{D}}$.
    while $\mathbf{b}_0$ is not converged // LOOP2: 1st-order consistence do
        Estimate normals $\{\boldsymbol{n}_j\}$ from $\{\mathcal{T}_i\}$.
        Estimate b via Equation (18.a), and update $\mathbf{b}_0$ with b.
        Estimate b via Equation (18.b), and update $\mathbf{b}_0$ with b.
    end while
end while
Output: Solution of B-spline objective face b.
```

A practical skill is to optimize the control points on u and v parameter spaces by turns. The two iteration items are rewritten as

$$\begin{cases} \left\|\mathbf{h} - \boldsymbol{L}|_{\mathbf{b}_0} \cdot \mathbf{b}\right\|_2^2 + \lambda \left\|\Lambda_1|_{\mathbf{b}_0} \cdot \mathbf{T}_1^{\#} \cdot (\mathbf{b} - \mathbf{b}_0)\right\|_2^2, \\ \left\|\mathbf{h} - \boldsymbol{R}|_{\mathbf{b}_0} \cdot \mathbf{b}\right\|_2^2 + \lambda \left\|\Lambda_2|_{\mathbf{b}_0} \cdot \mathbf{T}_2^{\#} \cdot (\mathbf{b} - \mathbf{b}_0)\right\|_2^2. \end{cases}$$

where the second term for each formulation is unit tangent vector constraint on the fixed the directions. $\Lambda_1|_{\mathbf{b}_0}$ (or $\Lambda_2|_{\mathbf{b}_0}$) is a $3N_p \times 3N_p$ diagonal matrix that stores $3N_p$ reciprocals of lengths of tangent vector $\frac{\partial \mathbf{F}}{\partial u}$ (or $\frac{\partial \mathbf{F}}{\partial v}$) at $\{\mathbf{u}_j\}_{j=1:N_p}$. During this procedure $\mathbf{b}_0$ is updated step-by-step. As shown in **Figure 5**, two partial derivatives $\frac{\partial \mathbf{F}}{\partial v}$ and $\frac{\partial \mathbf{F}}{\partial u}$ at (u, v) are updated until $\frac{\partial \mathbf{F}}{\partial v} \times \frac{\partial \mathbf{F}}{\partial u}$ converges to $\boldsymbol{n}$.

By integrating with $\mathbf{O}_2$, the final formulation of optimization consists of two items as follows:

$$\begin{cases} \min\limits_{\mathbf{b}} \left\| \begin{bmatrix} \mathbf{h} \\ \mathbf{T}^{\#xy}\mathbf{b}_0 \end{bmatrix} - \begin{bmatrix} \boldsymbol{L}|_{\mathbf{b}_0} \\ \mathbf{T}^{\#xy} \end{bmatrix} \mathbf{b} \right\|_2^2 + \lambda \left\|\Lambda_1|_{\mathbf{b}_0} \cdot \mathbf{T}_1^{\#}(\mathbf{b} - \mathbf{b}_0)\right\|_2^2, (a) \\ \min\limits_{\mathbf{b}} \left\| \begin{bmatrix} \mathbf{h} \\ \mathbf{T}^{\#xy}\mathbf{b}_0 \end{bmatrix} - \begin{bmatrix} \boldsymbol{R}|_{b_0} \\ \mathbf{T}^{\#xy} \end{bmatrix} \mathbf{b} \right\|_2^2 + \lambda \left\|\Lambda_2|_{\mathbf{b}_0} \cdot \mathbf{T}_2^{\#}(\mathbf{b} - \mathbf{b}_0)\right\|_2^2. (b) \end{cases} \tag{18}$$

The $\mathbf{b}_0$ is initialized by value of $\mathbf{b}_0$. Then we can solve $\mathbf{b}$ and update $\mathbf{b}_0$ orderly by minimizing (a) and (b) in Equation (18) iteratively until convergence.

6.3. Algorithm

An iterative algorithm is presented for this practical solution in Algorithm 1. Processes of 0th-order consistence and 1st-order consistence are separately conducted in the inner loop. And the outer loop guarantees a global convergence on two consistence problem.

6.3.1. Computational Complexity

The computation in above Algorithm 1 involves linear least square for solving Equations (16), (18.a), and (18.b), SVD for estimating camera parameter, and Robust PCA for Equation (17). In detail, the computational complexity for solving Equation (16) is $\mathrm{O}(n^2 f^2 MN)$, and that of both Equations (18.a) and (18.b) are $\mathrm{O}(N_p^2 MN)$. The computational complexity of robust PCA comes to be $\mathrm{O}(N_p^2 k)$, where k is the rank constraint. By assuming $N_p > M > N >> f > n$, computational complexity of the other parts can be negligible. In addition, we need considering the number of iteration for total computation of Algorithm 1.

7. EXPERIMENT

In this section experiments are presented to verify our automatic free-form surface modeling method. We first describe the pipeline to prepare a collection of face images of a person for B-spline face reconstruction. And then we demonstrate the quantitative and qualitative comparisons with recent baseline methods on projected standard images from ground truth 3D data (Zhang et al., 2017) with various expressions, illuminations and poses. Finally, we conduct challenging reconstructions and comparison based on real unconstrained data taken from the challenging Labeled Faces in Wild (LFW) database[1] (Huang et al., 2007).

[1] http://vis-www.cs.umass.edu/lfw/

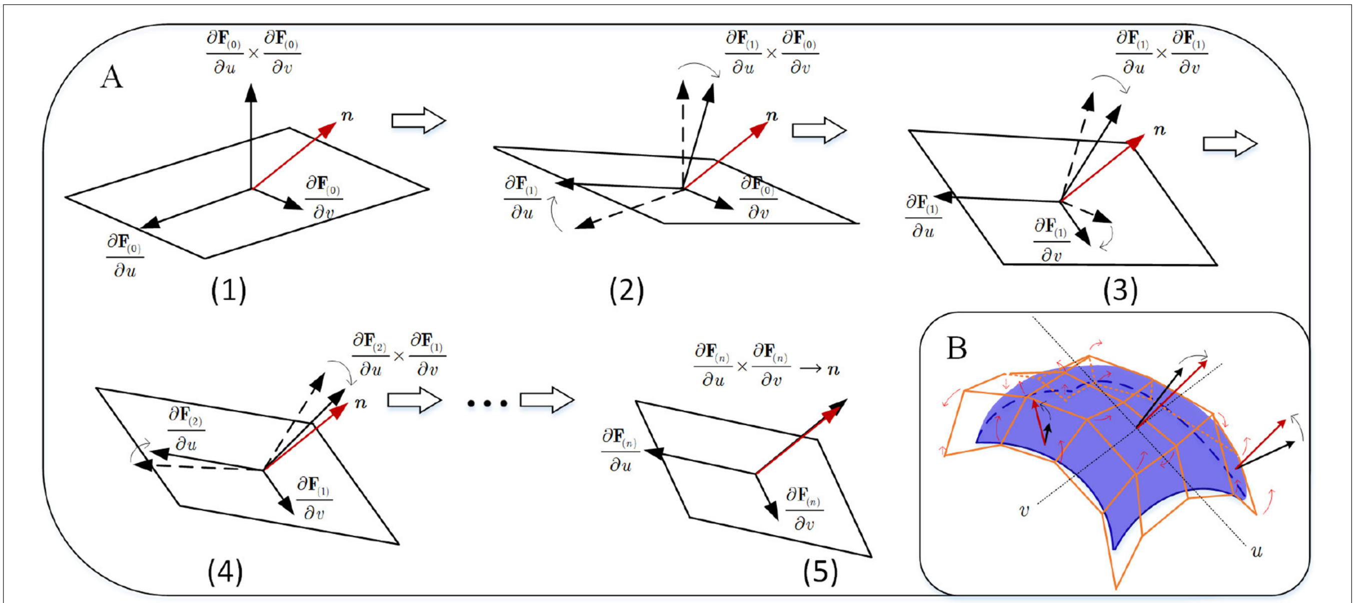
A
$\frac{\partial \mathbf{F}_{(0)}}{\partial u} \times \frac{\partial \mathbf{F}_{(0)}}{\partial v}$
n
$\frac{\partial \mathbf{F}_{(0)}}{\partial v}$
$\frac{\partial \mathbf{F}_{(0)}}{\partial u}$
(1)
$\frac{\partial \mathbf{F}_{(1)}}{\partial u} \times \frac{\partial \mathbf{F}_{(0)}}{\partial v}$
n
$\frac{\partial \mathbf{F}_{(1)}}{\partial u}$
$\frac{\partial \mathbf{F}_{(0)}}{\partial v}$
(2)
$\frac{\partial \mathbf{F}_{(1)}}{\partial u} \times \frac{\partial \mathbf{F}_{(1)}}{\partial v}$
n
$\frac{\partial \mathbf{F}_{(1)}}{\partial u}$
$\frac{\partial \mathbf{F}_{(1)}}{\partial v}$
(3)
$\frac{\partial \mathbf{F}_{(2)}}{\partial u} \times \frac{\partial \mathbf{F}_{(1)}}{\partial v}$
n
$\frac{\partial \mathbf{F}_{(2)}}{\partial u}$
$\frac{\partial \mathbf{F}_{(1)}}{\partial v}$
(4)
$\frac{\partial \mathbf{F}_{(n)}}{\partial u} \times \frac{\partial \mathbf{F}_{(n)}}{\partial v} \rightarrow n$
$\frac{\partial \mathbf{F}_{(n)}}{\partial u}$
$\frac{\partial \mathbf{F}_{(n)}}{\partial v}$
(5)
B
v
u

FIGURE 5 | Iterative adjustment on two partial derivatives: Process (1) to (2) adjusts $\frac{\partial \boldsymbol{F}}{\partial u}$ by fixing $\frac{\partial \boldsymbol{F}}{\partial v}$, and process (3) to (4) adjusts $\frac{\partial \boldsymbol{F}}{\partial v}$ by fixing $\frac{\partial \boldsymbol{F}}{\partial u}$, … until that $\frac{\partial \boldsymbol{F}}{\partial u} \times \frac{\partial \boldsymbol{F}}{\partial v}$ is infinitely close to objective $\boldsymbol{n}$; Process A implements a practically and iteratively linear handle for B-spline surface adjustment in B.

7.1. Data Pipeline and Evaluation

7.1.1. Synthesized Data With Expression

The ground truth data are from the space-times faces (Zhang et al., 2017) which contains 3D face models with different expressions. We use the data because it is convenient to evaluate our method with ground truth. And different poses and illuminations can also be simulated by the spaces-times faces, seen in **Figure 6**. Images with various poses and illuminations are collected, and feature points manually labeled. The reconstruction is evaluated by the error to the ground truth model.

7.1.2. Real Data in the Wild

The wild data (Huang et al., 2007) has characteristics of subject variations, illumination changes, various pose, background clutter and occlusions. Images of each person are collected and input into a facial point detector (Burgos-Artizzu et al., 2013) that has a similar high performance to human, to find the 40 facial points shown in **Figure 4**. The initial B-spline template face is computed from a neutral model of space-time faces.

7.1.3. Comparison

To verify the accuracy of automatic surface reconstruction, discrete points are sampled from the generated continuous free-form shape, and are compared to the traditional discrete reconstructions, e.g., work by Kemelmacher and Seitz (2011) and Roth et al. (2015). For a memory-limited capture system, it is not available to collect thousands of images as what Kemelmacher and Seitz (2011) and Roth et al. (2015) have done, so we limit all the reconstructions to less than forty images. We also compare them with an end-to-end deep learning method by Sela et al. (2017) qualitatively. Deep learning methods rely training on a large amount of unconstrained data, so we just use the model provided by Sela et al. (2017) that have been training on unconstrained images, and test it on the images in the wild.

7.2. Synthesized Standard Images

We conduct five sessions of reconstructions: the first four are used to reconstruct expression *S1*, *S2*, *S3*, and *S4* by using their corresponding images, and the fifth session *S5* is based on images with different expressions. Each session contains 40 images with various illumination and different poses. Reconstruction results are compared with the re-implemented method Kemel_meth by Kemelmacher and Seitz (2011) and Roth_meth by Roth et al. (2015). Kemel_meth generates frontal face surface based on integration in image domain of size 120 × 110. We clip it according to the peripheral facial points and interpolate points to get more vertices. Roth_meth generates a face mesh based on a template with 23,725 vertices. In our method, control point grid of 102 × 77 is optimized for a B-spline face surface.

7.2.1. Quantitative Comparison

To compare the approaches numerically, we compute the shortest point-to-point distance from ground truth to reconstruction. Point clouds are sampled from B-spline face and aligned according to absolute orientation problem. As done in work of Roth et al. (2015), mean Euclidean distance (MED), and the root mean square (RMS) of the distances, after normalized by the eye-to-eye distance, are reported in **Table 1**. Particularly, evaluation of Roth_meth is based on surface clipped with same facial points like the other two methods by considering a fair comparison. In the table, the best results are highlighted in boldface, and the underlined result has no significant difference with the best. To our knowledge, Roth_meth is the state-of-art

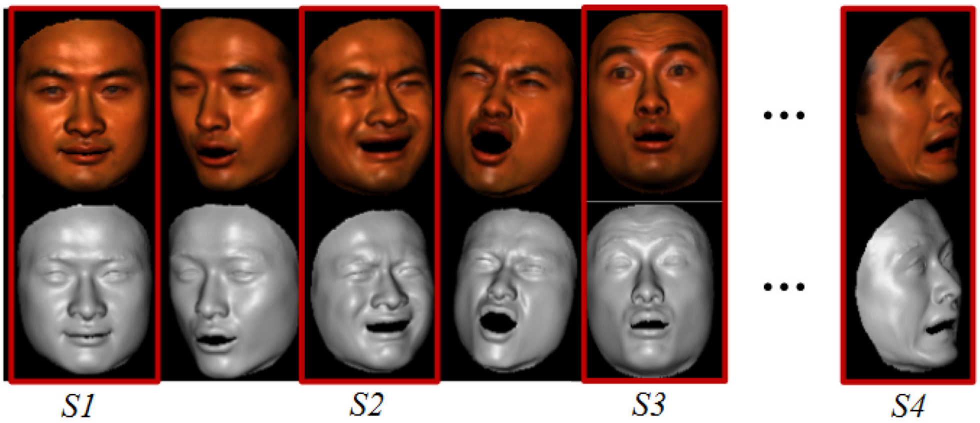

FIGURE 6 | Sample data simulated by the spaces-times faces (Zhang et al., 2017) : images and 3D model with various poses and illuminations are available; data of sample *S1*, *S2*, *S3*, and *S4* are used for evaluation.

TABLE 1 | Distances of the reconstruction to the ground truth.

Meth.	Index	S1	S2	S3	S4	S5
Kemel_meth	MED (%)	8.08	8.18	8.18	10.75	8.65
	RMS (%)	6.64	6.93	4.29	7.11	6.90
Roth_meth	MED (%)	**5.25**	7.06	5.43	6.63	**6.96**
	RMS (%)	**4.36**	5.79	4.54	4.42	**4.62**
Ours	MED (%)	6.31	**6.49**	**4.43**	**6.46**	6.98
	RMS (%)	4.10	**4.66**	**2.91**	**4.06**	4.34

The bold means the best value of MED and RMS, while the underline indicates the values next to the best.

method for face reconstruction from unconstrained images. Its re-implementation version is affected by the noisy normal estimation because of limited number images, showing results that are not good like as in its original paper. But it still performs good on all sessions. As a whole, results by both Roth_meth and our method have lower errors than Kemel_meth. On session S1 and S5, Roth_meth obtains the lowest mean error 5.21 and 6.96%, respectively. However, we obtains lower RMS 4.10 and 4.34% while its errors is quite close to the best especially on session S5. And on session S2, S3, and S4, our method obtains the best results, 6.49 ± 4.66, 4.43 ± 2.91, and 6.46 ± 4.06%. In contrast, the errors by Kemel_meth exceed 8%, and the RMS is also very large on every session. These numerical comparisons supply highly persuasive evidence that our B-spline method can build promising reconstructions based on face images.

7.2.2. Visual Comparison

The visual results in **Figure 7**. We show 3D models in mesh format for three methods on different sessions, and vertex numbers of models are also presented. It also demonstrates that our method has a promise performance by comparisons in the figure. An important fact is that Kemel (Kemelmacher and Seitz, 2011) cannot make a credible depth information and global shape, e.g., the global shape of reconstruction S2 and the mouse and nose of S3 are obviously incorrect, but our method solves global and local problem by optimization of 0th- and 1st-order consistency. And while Roth (Roth et al., 2015) generates more detailed information of an individual, it also produces distortion at the detailed shape, e.g., the eye of reconstruction S2 and the nose of reconstruction S3 and S4. In contrast, our method obtains realistic shape both globally and locally.

7.2.3. Characteristic Comparison

We give statistics of characteristics of the results generated by the three methods in **Table 2**, covering the global shape, local detail, credible depth, smoothness, distortion, and derivability. Depending on the quantitative and qualitative comparisons, we also give a rough rating. One star, two stars, and three stars represents bad, general, and good reconstruction respectively in the rating system. Both Roth_meth and our method obtain good scores on global shape, local detail, and credible depth. And both Kemel_meth and our method obtain a good score on smoothness. Because of the bad depth, Kemel_meth also gets bad score on global shape and distortion, and gets general scores on local detail. In addition, B-spline face model has better smoothness than the models by Kemel_meth and Roth_meth, because it is C^2 differentiable parametric surface while the other two are discrete model. Conclusively, 0th- and 1st-order consistency modeling using B-spline surface is efficient to reconstruct parametric surface of individual face.

7.3. Real Unconstrained Images

Our method is also tested based on real unconstrained data. Unconstrained data mean that the images are captured under uncertain condition, and the faces in the images are different in expression, pose and illumination condition. It is difficult to build the geometrical consistency for reconstruction using such data. Unlike the experiments in the work by Kemelmacher and Seitz (2011) using hundreds of images, we conduct reconstruction with limited number of images, because a large mount of face images for one person are not always available for small sample size tasks such as criminal investigation. In the experiment, uniformly 35 images are collected for each person from LFW database[1] covering different poses, illuminations and expressions.

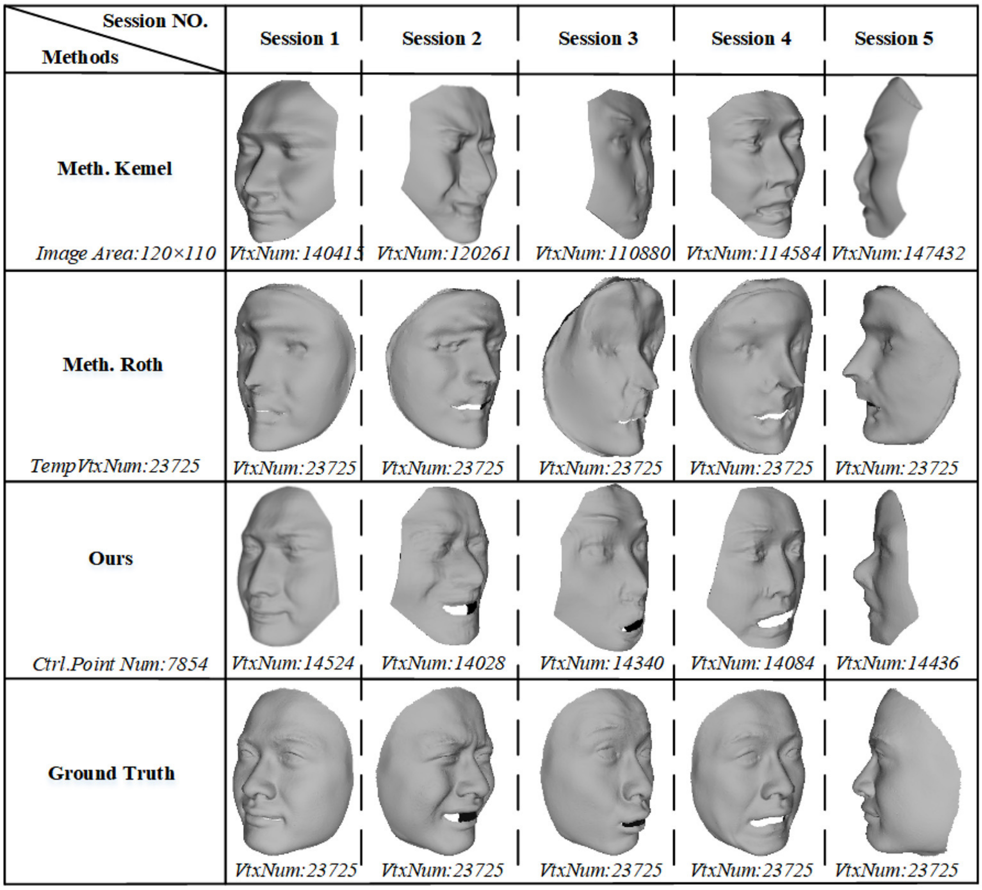

FIGURE 7 | Visual reconstructions and comparisons for session *S1*, *S2*, *S3*, *S4*, and *S5*: for each session of reconstructions, a column lists the 3D results of Kemelmacher and Seitz (2011), Roth et al. (2015), and us, as well as ground truth. (VrxNum means vertex number; TempVtxNum means vertex number of template; and Ctrl.Point Num means the control point number of B-spline face surface. Particularly, the vertices of B-spline face are points sampled from the reconstructed parametric surface).

TABLE 2 | A characteristics summarization of three methods by rough rating with number of ✰.

Characteristics	Kemel_meth	Roth_meth	Ours
Global shape	✰×1	✰×3	✰×3
Local detail	✰×2	✰×3	✰×3
Credible depth	✰×1	✰×3	✰×3
Smoothness	✰×3	✰×2	✰×3
No distortion	✰×1	✰×2	✰×3
C^2 differentiable	NO	NO	YES

Visual face reconstructions for Colin Powell, Donald Rumsfeld, George W. Bush, Hugo Chavez, and Gloria Macapagal Arroyo are compared with other two methods, as shown in **Figure 8**. Let *A* label the results generated by the reimplemented Kemel_meth, and let *B* label the results generated by the reimplemented Roth_meth, and let *C* label the method Seta_meth of deep learning by Sela et al. (2017) and let *D* label our results. Particularly, the input for Seta_meth is one image selected from the 35 images. Images in column 1, 5, and 8 are corresponding mean textures and two views of images respectively. By comparing these results, we observe some phenomena as follows:

(1) In frontal viewpoint, *A* and *D* show more vivid details than *B*, e.g., eyes and nose of Colin Powell. But in an other viewpoint, *D* shows more credible shape than *A*, e.g., the eyes and the forehead of Colin Powell, and the forehead and the mouth of Donald Rumsfeld.
(2) When the normals are incorrectly estimated from a limited number of images, e.g., for Gloria Macapagal Arroyo, *A* loses the local information completely, but *B*, *C*, and *D* still maintain general geometrical shape of face. For all methods, reconstructing nose is a challenge because the geometric curvature of the nose varies greatly. When the images are not enough, the noise could be amplified. So *B* shows bad results at nose being limited by number of input images.
(3) The input of *C* is a approximately frontal face image selected. As the model of *C* is learning on a set of 3D face data, it may not handle the uncertain noise and identity of inputs. So the

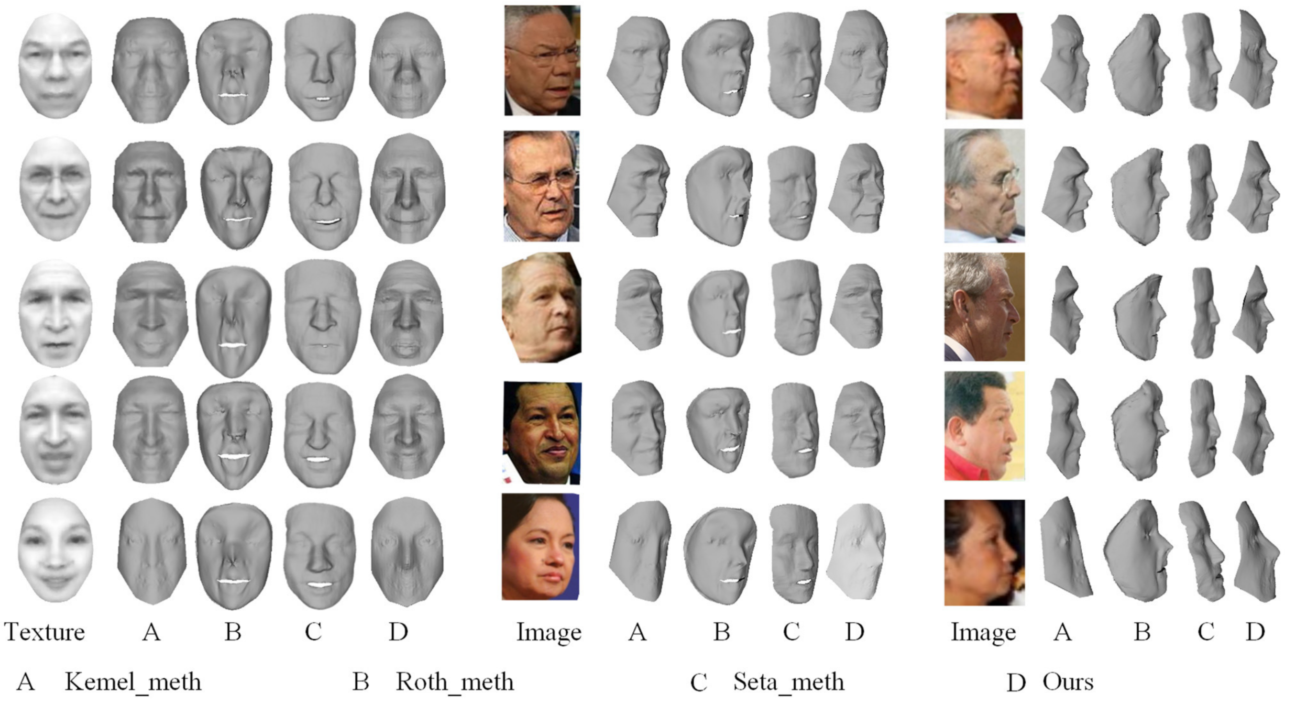

FIGURE 8 | Visual reconstructions and comparisons for Colin Powell, Donald Rumsfeld, George W. Bush, Hugo Chavez, and Gloria Macapagal Arroyo: (1) Images in column 1, 5, and 9 are corresponding mean textures and two views of images, respectively; (2) Columns labeled by *A* show the results generated by the reimplemented Kemel_meth, and columns labeled by *B* show the results generated by the reimplemented method Roth_meth, and columns labeled by *C* show the results generated by the Seta_meth, and columns labeled by *C* show our results. (The face images used in the figure come from LFW database[1]).

details in reconstruction by *C* don't look real, although their global shapes are stable and like human faces.

(4) By comparison, our method steadily produces better looking results than others from different viewpoints in the dataset. Clear and vivid details can be seen at key components such as eyes, nose and mouth, forehead, and cheek.

8. DISCUSSION

All the above experiments prove that our method can build pinpoint geometrical consistency on the limited number of real unconstrained data. Our method may not be best method in area of 3D reconstruction from multiple images, as the results in the original work by *B* looks better. It could deal with 3D reconstruction with limited number of images. Because we may not obtain large amount of images for reconstruction as done by Roth et al. (2015), for some condition restricted system. The shortcomings of *A* are mainly resulted from the inauthentic depth generated by integration method. And the bad results of *B* are caused by that the mesh template cannot build correct geometric consistency of number limited of unconstrained images and that the discrete differential operating on estimated noisy normal brings distortion errors. In contrast, we build pinpoint geometric consistency using B-spline surface. B-spline can smooth the noise in estimated normal better. So *D* can reconstruct correct face shape with little distortion, showing better result as a whole.

In the comparison, we don't consider other deep learning methods based methods appeared in recent years (Dou et al., 2017; Richardson et al., 2017; Lin et al., 2020; Sengupta et al., 2020; Shang et al., 2020). Because almost all recent works are focused on deep learning methods for single image based 3D face reconstruction (Dou et al., 2017; Richardson et al., 2017; Lin et al., 2020; Sengupta et al., 2020), as well as using a 3DMM model as prior. And the multi-view deep learning method only handle constrained face images (Shang et al., 2020). It means the deep learning methods can use a large amount of training data, and also a good prior. The input are different between these learning based methods and our method. So we conduct comparison with the classic optimization-based approaches for the sake of fairness. Nevertheless, we also select one representative method by Sela et al. (2017) to show result by deep learning as a reference in the comparison. It proves that if the test are not satisfactory to the prior and distribution of training data, it may obtain bad result.

9. CONCLUSIONS

This study set out to present high-detailed face reconstruction from multiple images based on pinpoint 0th- and 1st-order geometric consistence using B-spline embedding. Based on the good consistence modeling in geometric optics, the method works well for data with different poses and expressions in the wild. The key contribution of this study is that surface modeling adapts the correct rays in geometric optics by using B-spline embedding. This makes the high-detailed B-spline modeling from a number limited of face images captured under wild condition become reality. The method could also be applied to expression tracking and assisting face recognition in a monitoring or robot system.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

WP and ZF has contributed equally to the core idea as well as the experiment design and results analysis. YS, KT, and CX has provided assistance in experiments and analysis, under ZF's supervision. Besides, KT and MF provided the research group with financial support and experimental equipments. KT and ZF are supportive corresponding authors. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly supported by Science and Technology Program of Guangzhou, China (No. 202002030263), Shenzhen Science and Technology Foundation (JCYJ20170816093943197), Guangdong Basic and Applied Basic Research Foundation (2020A1515110997), National Natural Science Foundation of China (Nos. 61772164 and 62072126), and National Key R&D Program of China (No. 2019YFB1706003).

REFERENCES

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., et al. (2011). Building Rome in a day. *Commun. ACM* 54, 105–112. doi: 10.1145/2001269.2001293

Artificial, L. A., and Aryananda, L. (2002). "Recognizing and remembering individuals: Online and unsupervised face recognition for humanoid robot," in *Proc. of IROS* (Lausanne), 1202–1207.

Barsky, S., and Petrou, M. (2003). The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1239–1252. doi: 10.1109/TPAMI.2003.1233898

Bhardwaj, A., and Raman, S. (2016). Robust PCA-based solution to image composition using augmented lagrange multiplier (ALM). *Visual Comput.* 32, 591–600. doi: 10.1007/s00371-015-1075-1

Blanz, V., Mehl, A., Vetter, T., and Seidel, H. P. (2004). "A statistical method for robust 3D surface reconstruction from sparse data," in *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)* (Thessaloniki), 293–300. doi: 10.1109/TDPVT.2004.1335212

Blanz, V., and Vetter, T. (1999). "A morphable model for the synthesis of 3D faces," in *Proceedings of Conference on Computer Graphics and Interactive Techniques* (New York, NY), 187–194. doi: 10.1145/311535.311556

Burgos-Artizzu, X. P., Perona, P., and Dollár, P. (2013). "Robust face landmark estimation under occlusion," in *IEEE International Conference on Computer Vision (ICCV)*, (Sydney, VIC), 1513–1520. doi: 10.1109/ICCV.2013.191

Carr, J. C., Beatson, R. K., Cherrie, J. B., Mitchell, T. J., Fright, W. R., Mccallum, B. C., et al. (2001). Reconstruction and representation of 3D objects with radial basis functions. *ACM Siggraph* 67–76. doi: 10.1145/383259.383266

Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., and Tong, X. (2019). "Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set," in *IEEE Computer Vision and Pattern Recognition Workshops*. Long Beach, CA. doi: 10.1109/CVPRW.2019.00038

Dou, P., Shah, S. K., and Kakadiaris, I. A. (2017). "End-to-end 3D face reconstruction with deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI), 1503–1512. doi: 10.1109/CVPR.2017.164

Gecer, B., Ploumpis, S., Kotsia, I., and Zafeiriou, S. (2019). "Ganfit: generative adversarial network fitting for high fidelity 3D face reconstruction," in *IEEE Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA), 1155–C1164. doi: 10.1109/CVPR.2019.00125

Gonzalez-Mora, J., De la Torre, F., Guil, N., and Zapata, E. L. (2010). Learning a generic 3D face model from 2D image databases using incremental structure-from-motion. *Image Vis. Comput.* 28, 1117–1129. doi: 10.1016/j.imavis.2010.01.005

Heo, J., and Savvides, M. (2009). "In between 3D active appearance models and 3D morphable models," in *Computer Vision and Pattern Recognition* (Miami Beach, FL). doi: 10.1109/CVPRW.2009.5204300

Hoch, M., Fleischmann, G., and Girod, B. (1998). Modeling and animation of facial expressions based on b-splines. *Vis. Comput.* 11, 87–95. doi: 10.1007/BF01889979

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. Technical Report 07-49, University of Massachusetts, Amherst.

Kazhdan, M., Bolitho, M., and Hoppe, H. (2006). "Poisson surface reconstruction," in *Proceedings Symposium on Geometry Processing (SGP) 06* (Goslar), 32.

Kemelmacher Shlizerman, I. and Basri, R. (2011). 3d face reconstruction from a single image using a single reference face shape. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 394–405. doi: 10.1109/TPAMI.2010.63

Kemelmacher Shlizerman, I. and Seitz, S. M. (2011). "Face reconstruction in the wild," in *IEEE International Conference on Computer Vision (ICCV)* (Barcelona), 1746–1753. doi: 10.1109/ICCV.2011.6126439

Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J., and Alazab, A. (2019). A novel ensemble of hybrid intrusion detection system for detecting internet of things attacks. *Electronics* 8:1210. doi: 10.3390/electronics8111210

Koo, H.-S., and Lam, K.-M. (2008). Recovering the 3D shape and poses of face images based on the similarity transform. *Pattern Recogn. Lett.* 29, 712–723. doi: 10.1016/j.patrec.2007.11.018

Li, M., Sun, Y., Lu, H., Maharjan, S., and Tian, Z. (2019). Deep reinforcement learning for partially observable data poisoning attack in crowdsensing systems. *IEEE Intern. Things J.* 7, 6266–6278. doi: 10.1109/JIOT.2019.2962914

Lin, J., Yuan, Y., Shao, T., and Zhou, K. (2020). "Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks," in *2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5891–5900. doi: 10.1109/CVPR42600.2020.00593

Lu, Y., Yong, J. H., Shi, K. L., Song, H. C., and Ye, T. Y. (2016). 3D b-spline curve construction from orthogonal views with self-overlapping projection segments. *Comput. Graph.* 54, 18–27. doi: 10.1016/j.cag.2015.07.010

Maejima, A., Kuratate, T., Pierce, B., Morishima, S., and Cheng, G. (2012). "Automatic face replacement for humanoid robot with 3D face shaped display," in *2012 12th IEEE-RAS International Conference on Humanoid Robots* (Osaka), 469–474. doi: 10.1109/HUMANOIDS.2012.6651561

Meng, M., Lan, M., Yu, J., Wu, J., and Tao, D. (2020). Constrained discriminative projection learning for image classification. *IEEE Trans. Image Process.* 29, 186–198. doi: 10.1109/TIP.2019.2926774

Meyer, M., Desbrun, M., Schroder, P., and Barr, A. H. (2003). *Discrete Differential-Geometry Operators for Triangulated 2-Manifolds*. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-662-05105-4_2

Mian, A., Bennamoun, M., and Owens, R. (2006). "Automatic 3D face detection, normalization and recognition," in *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)* (Chapel Hill, NC) 735–742. doi: 10.1109/3DPVT.2006.32

Peng, W., Feng, Z., Xu, C., and Su, Y. (2017). "Parametric t-spline face morphable model for detailed fitting in shape subspace," in *IEEE Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI), 5515–5523. doi: 10.1109/CVPR.2017.585

Peng, W., Xu, C., and Feng, Z. (2016). 3D face modeling based on structure optimization and surface reconstruction with b-spline. *Neurocomputing* 179, 228–237. doi: 10.1016/j.neucom.2015.11.090

Piegl, L., and Tiller, W. (1997). *The Nurbs Book*. Monographs in Visual Communication. doi: 10.1007/978-3-642-59223-2

Piotraschke, M., and Blanz, V. (2016). "Automated 3D face reconstruction from multiple images using quality measures," in *Proc. IEEE Computer Vision and Pattern Recognition* (Las Vegas, NV), doi: 10.1109/CVPR.2016.372

Prados, E., and Faugeras, O. (2005). "Shape from shading: a well-posed problem?," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, *Vol. 2*, (San Diego, CA), 870–877.

Qiu, J., Tian, Z., Du, C., Zuo, Q., Su, S., and Fang, B. (2020). A survey on access control in the age of internet of things. *IEEE Intern. Things J.* 7, 4682–4696. doi: 10.1109/JIOT.2020.2969326

Richardson, E., Sela, M., and Kimmel, R. (2016). "3D face reconstruction by learning from synthetic data," in *International Conference on 3D Vision (3DV)* (Stanford, CA), 460–469. doi: 10.1109/3DV.2016.56

Richardson, E., Sela, M., Or-El, R., and Kimmel, R. (2017). "Learning detailed face reconstruction from a single image," in *IEEE Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI), 5553–5562. doi: 10.1109/CVPR.2017.589

Roth, J., Tong, Y., and Liu, X. (2015). "Unconstrained 3D face reconstruction," in *IEEE Computer Vision and Pattern Recognition (CVPR)* (Boston, MA). doi: 10.1109/CVPR.2015.7298876

Roth, J., Tong, Y., and Liu, X. (2016). "Adaptive 3D face reconstruction from unconstrained photo collections," in *Proc. IEEE Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2016.455

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (New York, NY), 519–528.

Sela, M., Richardson, E., and Kimmel, R. (2017). "Unrestricted facial geometry reconstruction using image-to-image translation," in *2017 IEEE International Conference on Computer Vision (ICCV)* (Venice). doi: 10.1109/ICCV.2017.175

Sengupta, S., Lichy, D., Kanazawa, A., Castillo, C. D., and Jacobs, D. W. (2020). SfSNet: Learning shape, reflectance and illuminance of faces in the wild. *IEEE Trans. Pattern Anal. Mach. Intell*. doi: 10.1109/TPAMI.2020.3046915

Shang, J., Shen, T., Li, S., Zhou, L., Zhen, M., Fang, T., et al. (2020). "Self-supervised monocular 3D face reconstruction by occlusionaware multi-view geometry consistency," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Glasgow). doi: 10.1007/978-3-030-58555-6_4

Sun, Z.-L., Lam, K.-M., and Gao, Q.-W. (2013). Depth estimation of face images using the nonlinear least-squares model. *IEEE Trans. Image Process*. 22, 17–30. doi: 10.1109/TIP.2012.2204269

Tomasi, C., and Kanade, T. (1992). Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vis.* 9, 137–154. doi: 10.1007/BF00129684

Tran, A. T., Hassner, T., Masi, I., and Medioni, G. (2017). "Regressing robust and discriminative 3D morphable models with a very deep neural network," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, (Honolulu, HI). doi: 10.1109/CVPR.2017.163

Wang, H., Wei, H., and Wang, Y. (2003). "Face representation under different illumination conditions," in *International Conference on Multimedia and Expo (ICME)* (Baltimore, MD), 285–288.

Wu, F., Bao, L., Chen, Y., Ling, Y., Song, Y., Li, S., et al. (2019). "MVF-net: multi-view 3D face morphable model regression," in *IEEE Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA), 959–968. doi: 10.1109/CVPR.2019.00105

Yang, C., Chen, J., Su, N., and Su, G. (2014). "Improving 3D face details based on normal map of hetero-source images," in *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)* (Columbus, OH), 9–14. doi: 10.1109/CVPRW.2014.7

Yuille, A. L., Snow, D., Epstein, R., and Belhumeur, P. N. (1999). Determining generative models of objects under varying illumination: shape and albedo from multiple images using SVD and integrability. *Int. J. Comput. Vis*. 35, 203–222. doi: 10.1023/A:1008180726317

Zhang, L., Mistry, K., Jiang, M., Chin Neoh, S., and Hossain, M. A. (2015). Adaptive facial point detection and emotion recognition for a humanoid robot. *Comput. Vis. Image Understand*. 140, 93–114. doi: 10.1016/j.cviu.2015.07.007

Zhang, L., Snavely, N., Curless, B., and Seitz, S. M. (2017). "Spacetime faces: high-resolution capture for modeling and animation," in *Data-Driven 3D Facial Animation* eds Deng, Z., and Neumann, U. (Los Angeles, CA: Springer), 248–276. doi: 10.1007/978-1-84628-907-1_13

Zhang, R., Tsai, P.-S., Cryer, J. E., and Shah, M. (1999). Shape-from-shading: a survey. *IEEE Trans. Pattern Anal. Mach. Intell*. 21, 690–706. doi: 10.1109/34.784284

Zhou, Y., Deng, J., Kotsia, I., and Zafeiriou, S. (2019). "Dense 3D face decoding over 2500 fps: joint texture shape convolutional mesh decoders," in *IEEE Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA), 1097–1106. doi: 10.1109/CVPR.2019.00119

Zhu, X., Lei, Z., Liu, X., Shi, H., and Li, S. Z. (2016). "Face alignment across large poses: a 3D solution," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV), 146–155. doi: 10.1109/CVPR.2016.23

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.