

# Local Spherical Harmonics for Facial Shape and Albedo Estimation

SUWOONG HEO<sup>ID</sup>, HYEWON SONG<sup>ID</sup>, JIWOO KANG<sup>ID</sup>,  
AND SANGHOON LEE<sup>ID</sup>, (Senior Member, IEEE)

Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Sanghoon Lee (slee@yonsei.ac.kr)

This research project was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2020.

**ABSTRACT** In this paper, we present a novel facial albedo and 3D shape recovery method with a local spherical harmonic illumination model. From a face in an image, the proposed method can produce a high-quality 3D shape and albedo using a novel parameterization of local illuminations. Because a facial shape is partially convex, a single spherical harmonics is generally used for the illumination of a face within a constrained illumination environment. However, when a facial image is captured in an unconstrained scene, the illumination is inappropriately estimated due to the presence of shadow and specular reflections. To address this issue, we propose a novel local spherical harmonic illumination model for representing the illumination of a face. Unlike the existing parameterization of local illumination, our local spherical harmonic illumination model utilizes a smooth weight function for the seamless representation of natural illumination. Therefore, the albedo and shape information in an image can be precisely estimated using the first-order spherical harmonics. For accurate estimation of albedo, we also utilize facial albedo statistics to prevent the estimated albedo from becoming biased toward input image. Furthermore, we developed an accurate and reliable 3D shape reconstruction method from a normal map based on tetrahedron-based deformation. Comparing to the Laplacian deformation based method, our method is applicable to any mesh regardless of its structure. Through rigorous experiments, we demonstrate that the proposed local spherical harmonic illumination model is effective in estimating the complex illumination and can recover a high-quality facial albedo and 3D shape.

**INDEX TERMS** Albedo estimation, face reconstruction, shape from shading, spherical harmonic illumination, tetrahedron-based deformation.

## I. INTRODUCTION

The analysis of facial geometry and appearance is a classical problem and its applications are related to many computer vision and graphics tasks such as face recognition [1], pose estimation [2]–[4], and facial animation [5]. 3D face reconstruction, which is the process of inferring the 3D geometry of a human face from 2D images, is the most very fundamental core that powers those applications.

Among many studies of 3D face reconstruction, estimating 3D geometry of a human face with a single image is particularly challenging because of the loss of depth information caused by camera projection. Therefore, many of the recent studies in the computer vision field have focused on approximating the 3D geometry in a single image using prior

knowledge regarding the target object [6]. A well-known prior model for face reconstruction is the 3D Morphable Model (3DMM), which is a statistical model of facial shape and albedo [7]. Several studies have attempted to reconstruct the 3D face model from a single image by optimizing the parameters of 3DMM [8]–[10] or utilizing the deep-learning method [11], [12]. However, such methods can only reconstruct the smooth parts of facial geometry and fail to reconstruct the high-frequency details such as beard, pore, and wrinkles. In recent years, several methods have been proposed to recover high-frequency details from the smooth geometry by using the shape from shading (SfS) technique [13], [14]. This technique is widely adopted for computing the 3D geometry of highly curved regions based on the illumination and reflectance model for shading variation [15], [16]. By assuming that a face can be represented by pure diffuse

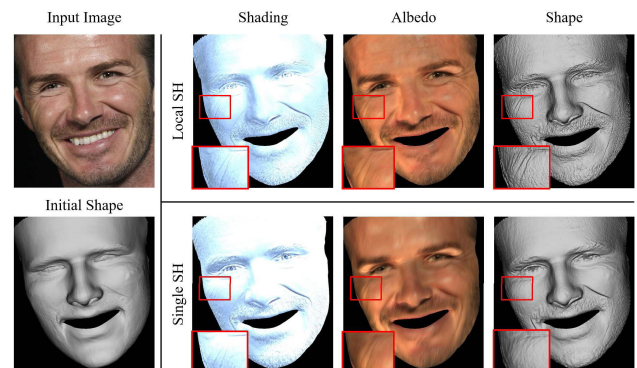
The associate editor coordinating the review of this manuscript and approving it for publication was Songwen Pei<sup>ID</sup>.

reflectance model, most of studies have focused on Lambertian reflectance in conjunction with second-order spherical harmonics (SH) for illumination model [9], [17], [18]. This approach is effective for images taken under controlled illumination conditions. However, for the face images with complex illumination, the SfS process with single second-order SH may fail to reconstruct locations containing effects that cannot be modeled accurately.

Although the second-order SH is proven to be robust for Lambertian diffuse surfaces [19], facial images in natural scenes contain various lighting effects such as cast shadow, specular, and indirect illumination. For specular illumination, one way to model such effects is to adopt a non-Lambertian reflectance model such as the dichromatic model, with a higher-order SH basis [20]. Although such a model is useful for estimating surface normals of a reflective object with a strictly constrained setup [21], the estimation for a single image is difficult because it is a highly under-determined problem. Additionally, to model cast shadows and inter-reflection, a computationally expensive algorithm such as ray-tracing [22], is required. In early works on facial shape recovery, several researchers attempted to reduce the complex lighting effects by using an intensity map called corrective field [17], [23]. This map is introduced to reduce the shading error on a wide shadow region on a face, but not small shadow regions, such as a nasal cavity. Moreover, shading properties are not fully considered, resulting in the filtering shadow regions during the process of recovering albedo and shape. Recently, a global and local SH model for scene depth and reflectance estimation for single image is proposed [24]. Local SH captures partial illumination that compensates for residual discrepancies between image and global illumination. In [24], the effective range of each local SH was defined on a particular rectangular grid. This strategy produces fine approximations of the illumination variation in a scene with a set of small objects. However, for an image with spatially varying appearances, this method can produce aliasing artifacts near the boundary of the grid.

Motivated by these previous works, we address the problem of complex illumination for facial 3D shape and albedo recovery from a single image using a novel local SH illumination model. Because the illumination on a face varies depending on facial structure and lighting environments, local SH illumination with a rectangular grid fails to model regions with complex illumination effects. Rather than using a rectangular grid, we propose a more flexible local SH illumination model by clustering the pixels based on its luminance cues. Based on these clustered pixels, the shape and local contribution of each local SH illumination is computed using a local weight function that accounts for smooth variations around the cluster. The aliasing artifact caused by rectangular grids can be resolved because of the property of signal interpolation between local weight functions. On the leverage of the proposed illumination model, we propose an optimization framework for recovering facial albedo and 3D shape information from a single image. By using a prior face model initialized

by the existing fitting method, we obtain illumination, albedo, normal, and 3D shape through an alternating optimization scheme. We observe that the proposed local SH illumination is able to model complex illumination efficiently by using a low order basis. Therefore, by using the first-order basis, we can linearize each problem while retaining the advantage of local SH illumination. To this end, we develop a simple but robust method for reconstructing a 3D facial shape by using tetrahedron-based deformation technique with triangle normals. Compared to the traditional approach using surface Laplacian, we show that this technique can reliably reconstruct a 3D shape regardless of a mesh structure of the prior 3D shape. An example of the proposed method compared to the single SH is presented in Fig. 1.



**FIGURE 1.** An example result of facial albedo and shape recovery. For shading variations because of the self-shadow in the red boxes, the proposed local SH reconstructs details of the facial shape, albedo as well as shading.

Our main contributions can be summarized as follows:

- We propose a robust and efficient optimization framework for estimating facial albedo and 3D shape from single image based on a simple but effective local SH illumination model for natural scenes.
- A local SH illumination model is constructed upon the smooth and flexible weight function based on superpixel clustering. This function smoothly interpolates the local illuminations of its surrounding regions.
- Unlike the traditional SfS method which recovers a 3D shape by optimizing the heightmap or surface Laplacian from computed normal map, we adopt a tetrahedron based deformation method that can be efficiently solved by linear optimization.

The remainder of this paper is organized as follows. In Section II, several studies related to the proposed method are briefly reviewed. In Section III, we introduce the proposed local SH illumination model. We then present an efficient framework to recover facial shape and albedo in Section IV. Following the comprehensive experiments in Section V, the final conclusion is summarized in Section VI.

## II. RELATED WORKS

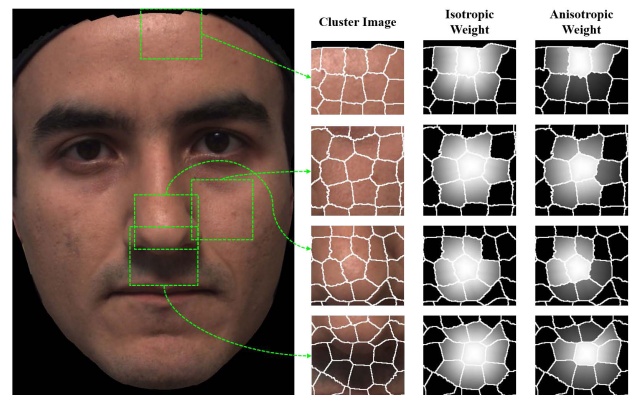
Because inferring facial albedo, and 3D geometry from a single image is an ill-posed problem, most studies have utilized prior models of a face such as the 3DMM [7]. By using the 3DMM, there are several approaches for reconstructing the 3D face model, which can be categorized into inverse rendering [8], SfS technique combined with 3DMM fitting [25], and end-to-end frameworks based on deep learning.

**Inverse rendering** aims to recover shape, illumination, and reflectance information from a single image or set of images. In particular, for 3D face reconstruction, the low-dimensional parameters for shape and albedo of the 3DMM are estimated based on a specific reflectance model. Blanz and Vetter [8] reconstructed a 3D face model using the Phong reflectance model with weak specular assumption. To incorporate complex illumination, a shading model represented by second-order SH was proposed [9]. Additionally, some recent works have attempted to merge the from-the-studio texture with in-the-wild texture [1], [10]. While the traditional 3DMM was constructed from 200 face scans, 9,663 distinct faces were used to generate the large-scale face model [26]. In [27], the linear representation of the 3DMM was transformed into a nonlinear representation using a deep neural network. However, this technique cannot reconstruct the fine details of facial geometry, such as deep wrinkles, because such models are constructed based upon the generalization of facial appearance.

**SfS**, which is a technique for inferring the normal map of a given image, has been widely adopted for enhancing high-frequency details with an initial estimate of a 3D model. After using various methods to compute an initial model, such as multi-view stereo [15], depth [16], or template-based methods [13], the surface normal, illumination, and albedo are inferred to reconstruct a point cloud and corresponding mesh. For human faces, SfS-based methods can successfully recover 3D geometry from a set of unconstrained [14] or studio level [13], [28] images. In [14], a 3D surface corresponding to computed point normals was generated by deforming a template shape using the Laplacian surface deformation technique [29]. Additionally, outlier pixels were filtered by using a structural similarity index (SSiM) [30] to avoid inaccurate albedo estimation. For single-image cases, the most common approach is to compute a depth map iteratively by updating albedo, illumination, and normal maps [25]. Biswas *et al.* [31] proposed the linear minimum mean squared error estimation for albedo in an image. Their reported results demonstrated that a linear least square process conjoint with error statistics computed by using 3DMM can generate an albedo that is visually similar to the true albedo. Another approach is to compute albedo by estimating coefficients for the principal components of a statistical albedo model constructed from hundreds of scan data [17]. To reduce the effects of complex illumination effects, such as cast shadows and specular highlights, [23] applied a corrective field to their illumination model. Most of the aforementioned works have

made efforts to attenuate the effects of facial regions with complex illumination effects.

**Deep-learning** based methods are the most preferred approach for end-to-end inference of 3D geometry from a single image [32]. [33] proposed a convolutional neural network (CNN) to decompose facial images into albedo, shading, and normal maps. Because of a lack of ground-truth data, their network was trained using both real and synthetic data generated by using the 3DMM. To compensate for the domain discrepancies between real and synthetic data, photometric rendering loss with single second-order SH was adopted. Although this method is robust to complex illumination effects because of the generalization capabilities of CNNs, the reconstructed facial albedo maps were biased toward the average albedo of the 3DMM. Furthermore, the reconstructed facial geometry lacked high-frequency details. To recover high-frequency details, serial architectures consisting of a coarse-to-fine network have been proposed [27]. Recently, a UV map, which is mapping of 3D mesh to the 2D image, was used for inferring 3D displacement from 2D image [34], [35]. The latest prominent results by using a UV map combined with a CNN was proposed in [36].

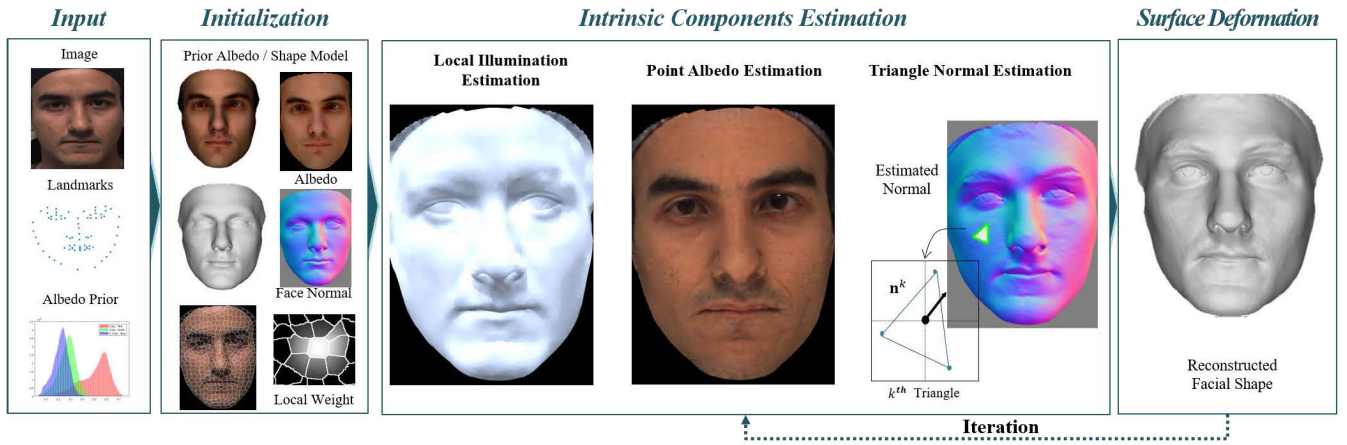


**FIGURE 2.** Example of local clustering and anisotropic weighting. By using the anisotropic weight  $\omega_j$ , each local SH can have a greater effect on its adjacent clusters, which have similar luminance distributions.

## III. LOCAL SH ILLUMINATION

Most methods for estimating illumination on a face assume that human skin acts as a diffuse reflector so that illumination can be modeled by single second-order SH. In [37], it was demonstrated that single second-order SH can capture the 99.2% of the illumination of a Lambertian object with distant lights. Since then, in many SfS methods for facial albedo and shape recovery, a set of coefficients of second-order SH has been used to represent illumination around a face. In such model, each pixel intensity  $I(x, y)$  in an image is expressed as the product of the facial albedo  $\rho(x, y)$  and illumination  $L(x, y)$ . We can further simplify this model by factoring the spectral power of  $\{R, G, B\}$  channels into the albedo, meaning a vector  $L(x, y)$  can be converted into a scalar  $L(x, y)$ . As a result, the pixel intensity of a face in an image can be defined





**FIGURE 3.** Overview of the proposed facial albedo and shape recovery method. For a given facial image, estimated landmarks, and the prior distribution of albedo modeled by a Gaussian mixture model (GMM), the proposed method estimates the intrinsic components of the image, including illumination, albedo, and normal maps, in sequence. Then, the facial shape is reconstructed using the computed normal map. This method iterates on intrinsic component estimation and surface decomposition until convergence is achieved.

as

$$\mathbf{I}(x, y) = \rho(x, y)L(x, y), \quad (1)$$

$$L(x, y) = \sum_{n=0}^N \sum_{m=-n}^n l_{nm} Y_{nm}(\mathbf{n}(x, y)), \quad (2)$$

where  $N$ ,  $l_{nm}$ , and  $\mathbf{n}(x, y)$  are the order of SH, illumination coefficients of the basis  $Y_{nm}$ , and surface normal projected onto the pixel position  $(x, y)$ , respectively. For a Lambertian object, such as a face, the global illumination is generally approximated by  $N = 2$ .

#### A. LOCAL SH MODEL

Although the aforementioned model is effective for modeling the global illumination of a Lambertian object, it cannot model cast shadows, self-shadows, and specular reflections, which appear in local areas of an object in natural scenes. To model such illumination effects, a more moderate model is required because of the non-linear nature of shadows and high-frequency characteristics of specular reflections. To address this issue, we propose a simple and efficient representation for complex illumination effects using the local SH illumination model. As demonstrated in [24], the global-local SH representation for estimating local illumination is effective for the robust estimation of illumination in an image. The authors of the aforementioned paper divided an image into rectangles so the illumination errors in each region could be compensated independently. However, because of the absence of smoothness constraint on border pixels, this approach yields aliasing artifacts. Furthermore, simply minimizing the illumination gradient between the adjacent regions would result in flat local illumination. To resolve this issue, an illumination effect on  $(x, y)$  is expressed as the sum of global illumination and the weighted contributions of

adjacent local SH as follows:

$$L(x, y) = L_G(x, y) + \sum_{k \in \mathcal{C}} w_k(x, y) L_k(x, y), \quad (3)$$

where  $L_G(x, y)$ ,  $L_k(x, y)$ ,  $\mathcal{C}$ , and  $w_k(x, y)$  represent the global illumination,  $k^{th}$  local SH illumination, a set of clusters, and their contributions to  $(x, y)$ , respectively. Note that each illumination follows the SH lighting formula in (2).

Unlike in [24], we compute each region for local SH using a pixel clustering method, such as simple linear iterative clustering (SLIC) [38], in the luminance domain. As shown in Fig. 2, regions with shadows, such as the fourth row, and specular effects, such as the first and third rows, can be effectively segmented. Compared to using a rectangular grid, this approach enables more intuitive construction of local SH illumination. For each cluster region, we assign SH illumination that affects not only the area inside the region but also nearby clusters with a certain weight which is detailed in the following subsection.

#### B. ANISOTROPIC WEIGHTING

For an image clustered based on luminance cues, the contributions of illumination for the pixel location  $(x, y)$  from the  $k^{th}$  cluster is expressed by a weighted value for the local SH as defined in (3). The weight should be monotonically decreasing function with distance from the centroid of the cluster. Since each cluster is not necessarily formed with ellipsoidal shape, simply using the distance between the centroid and a pixel location is not appropriate. Therefore, we define the weight  $w_k(x, y)$  in terms of a Gaussian with the distance function of each cluster. The distance function  $\phi_k(x, y)$  computed for the cluster region  $\Omega_k$  is expressed by the Euclidean distance  $d((x, y), \partial\Omega_k)$  between the pixel

location  $(x, y)$  and boundary  $\partial\Omega_k$  of the  $k^{th}$  cluster as follows:

$$\phi_k(x, y) = \begin{cases} -d((x, y), \partial\Omega_k) & (x, y) \in \Omega_k \\ d((x, y), \partial\Omega_k) & (x, y) \notin \Omega_k, \end{cases} \quad (4)$$

$$d((x, y), \partial\Omega_k) = \min_{(x', y') \in \partial\Omega_k} \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x' \\ y' \end{pmatrix} \right\|_2. \quad (5)$$

Then, the distance function  $\phi_k(x, y)$  is converted to the zero-centered distance function  $\phi'_k(x, y)$  by subtracting minimum value of  $\phi_k(x, y)$ . In addition to the distance function, we compute a local smoothness value  $\omega_j$ , which is defined as the Chi-square distance of the luminance distribution between the  $k^{th}$  cluster and its adjacent clusters  $\{j | j \in \text{adj}(k)\}$ . The weight  $\omega_j$  determines the width and smoothness of the weight function in the direction of the  $j^{th}$  adjacent cluster. This prevents  $w_k(x, y)$  from affecting regions with distinctive illumination effects. As shown in Fig. 2, the adjacent cluster with a small Chi-square distance has a wider and higher weight value. Based on this distance function and local smoothness, the anisotropic weight function of the  $k^{th}$  cluster for the pixel  $(x, y)$  for local SH is defined as

$$w_k(x, y) = \omega_j e^{-(\phi'_k((x, y)))^2 / 2\sigma_k^2}, \quad (6)$$

where  $\sigma_k$  is a control parameter that determines the radius of the weight. To determine  $\sigma_k$  automatically, we adopt the concept of the full width at half maximum. Let  $\alpha$  be a desired boundary value proportional to the maximum weight value of the cluster. Then,  $\sigma_k$  is computed as follows:

$$\sigma_k = \left( \frac{1}{|\partial\Omega|} \sum_{(x, y) \in \partial\Omega} \phi'_k(x, y) \right) / \sqrt{2 \ln \alpha}. \quad (7)$$

As shown in Fig. 2, excluding the region with flat illumination in the second row, the computed weights are consistent with the Gaussian and zero-centered distance function  $\phi'_k(x, y)$  for adjacent regions with similar illumination effects, but fall off significantly for other regions. For all experiments in this paper, we empirically set  $\alpha = 1.5$ .

#### IV. ALBEDO AND SHAPE RECOVERY

Because directly estimating facial shape and albedo from a single image is infeasible, we utilize the prior face information by fitting a 3DMM. In particular, the pose, expression, and identity parameters of the 3DMM are estimated from a given image with landmark points [39]. A weak perspective projection is assumed for the 3DMM fitting process. To estimate the prior albedo, we choose initial illumination coefficients as uniform white illumination by setting the first coefficient of global SH as  $2\sqrt{\pi}$  and setting all other coefficients to zero, as described in [20]. Because the mesh resolution of the model is insufficient for recovering facial details, we apply the Loop subdivision [40] to the initial shape model following initialization.

Based on the face region  $\Omega$  identified by projecting an initial face model, the estimation of illumination, albedo, and

normal from the image proceeds by minimizing the following data term:

$$E_I = \sum_{(x, y) \in \Omega} \|\mathbf{I}(x, y) - \boldsymbol{\rho}(x, y)L(x, y)\|^2. \quad (8)$$

Because jointly optimizing the albedo, illumination coefficients, and surface normal in (8) is infeasible, we separate this problem into estimating global-local illumination coefficients  $L_k, L_G$ , albedo  $\boldsymbol{\rho}$ , and surface normal  $\mathbf{n}(x, y)$  iteratively. Each component is estimated by minimizing the (8) combining with proper regularization terms. At the end of each iteration, the vertex displacement with the computed normal is applied by using the tetrahedron-based mesh deformation technique [41]. To improve computational efficiency, we set  $N = 1$  so that each optimization can be conducted using a linear least-squares solver, such as bi-conjugate gradient stabilized solver [42]. Although higher-orders are desirable for the accurate estimation of shape and albedo, we found that our model can accurately reconstruct facial details even using the first-order SH. An overview of the facial shape and albedo recovery process is presented in Fig. 3.

#### A. ILLUMINATION ESTIMATION

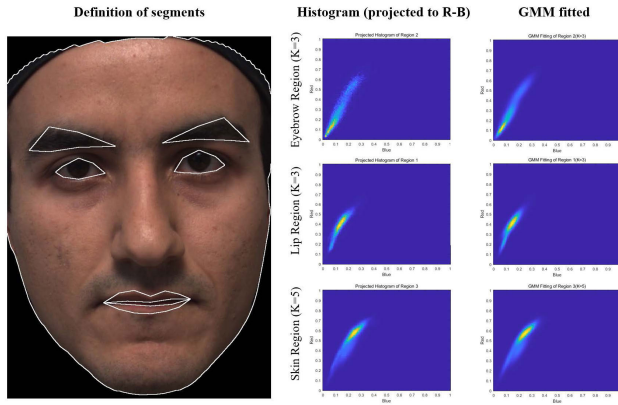
For a given  $\boldsymbol{\rho}(x, y)$  and  $\mathbf{n}(x, y)$  computed from the initialization or previous iteration, the illumination coefficients, including global and local SH, are estimated jointly. In conjunction with  $E_I$  from (8), the objective function  $E_L$  for the illumination estimation step is defined as

$$E_L = E_I + \lambda_{L_c} \sum_{k \in C} \|\mathbf{l}_k\|^2 + \lambda_{L_s} \sum_{(x, y) \in \Omega} \nabla L(x, y)^2, \quad (9)$$

where  $\lambda_{L_c}$ ,  $\lambda_{L_s}$ , and  $\mathbf{l}_k$  are the balancing weights for the regularization and smoothness terms, as well as a vector containing the illumination coefficients of (2) for the  $k^{th}$  cluster, respectively. Because a large value for the coefficients of local SH would affect the stability of the normal estimation step, the coefficient regularization term constrains the coefficients for local SH illumination to be small. Smoothness term is introduced for enforcing the estimated illumination smooth in the spatial domain. The illumination of the input image is estimated by minimizing (9) with respect to the SH coefficients of  $L_k$  and  $L_G$  simultaneously.

#### B. ALBEDO ESTIMATION

After computing the global and local SH illumination coefficients, the albedo  $\boldsymbol{\rho}(x, y)$  is computed. We incorporate additional regularization energy to enforce the smoothness of the albedo according to the uniform Laplacian [44] of mesh information of the face model. Additionally, facial albedo priors from the MERL/ETH Skin Reflectance Database [43] are adopted to prevent the computed albedo from overfitted to the input image. Unlike the definition of facial segments in this database, we merge some facial regions. Although predefined segments can be obtained using facial landmarks, this would produce an undesirable discontinuity between contiguous segments. Therefore, we utilize the definition of



**FIGURE 4.** Definition of facial segments. We divide a face into four regions: lip, eyebrows, eyes, and facial skin. The corresponding histogram in the  $(R, G, B)$  domain and its fitted GMM are visualized by marginalizing it along  $G$  axis. Because there were no statistics for the eye region in [43], we do not use the region for (12). Regarding the other segments, the number of mixtures for each GMM is set to  $K = 3$  for the lip and eyebrows and  $K = 5$  for the facial skin region.

segments for salient areas, including the eyebrows and lip, while merging other facial skin regions into a single segment. For each region, a prior distribution model of facial albedo is computed by fitting the measured data to a GMM. The definitions of each region and the computed GMM with the number of modes are presented in Fig. 4. Finally, the objective function  $E_\rho$  for albedo estimation is defined as follows:

$$E_\rho = E_I + \lambda_p E_p + \lambda_s E_s. \quad (10)$$

A smoothness term  $E_s$  is introduced to obtain smooth albedo as follows:

$$E_s = \sum_i \sum_{j \in \mathcal{N}(i)} \frac{1}{d_i} \|\rho_j - \rho_i\|^2, \quad (11)$$

where  $\mathcal{N}(i)$  and  $d_i$  represent the local connectivity around the  $i^{\text{th}}$  vertex and the normalizer, which represents the number of vertices connected to the  $i^{\text{th}}$  vertex, respectively. The albedo prior term  $E_p$ , which is used to make the estimated albedo to have higher likelihood in terms of GMM, is defined as

$$E_p = \sum_i -\log p(\rho_i), \quad (12)$$

where  $p(\rho_i)$  is GMM computed by applying the expectation-maximization method to the database.

Because (12) is non-convex, we use an iterative method to relax the problem, as described in [45]. In each iteration, we compute the weighted log-likelihood of the computed albedo for the  $i^{\text{th}}$  vertex with its adjacent vertices and derive the mixture  $k^*$  with the highest likelihood. By using the chosen mixture  $k^*$  with its mean  $\mu_{k^*}$  and variance  $\Sigma_{k^*}$ , the equation can be converted into a linear form as follows:

$$E_p = \sum_i (\rho_i - \mu_{k^*})^T \Sigma_{k^*}^{-1} (\rho_i - \mu_{k^*}). \quad (13)$$

In our experiments, five iterations were sufficient to achieve convergence. It should be noted that (11) is similar to the albedo regularization term proposed in [25], except that regularization is applied in the vertex domain. Because the albedo is estimated on the entire vertex, our formula can generate smooth albedo across an entire face model. This is particularly beneficial when visible triangles are updated during the iteration by z-buffering.

### C. NORMAL ESTIMATION

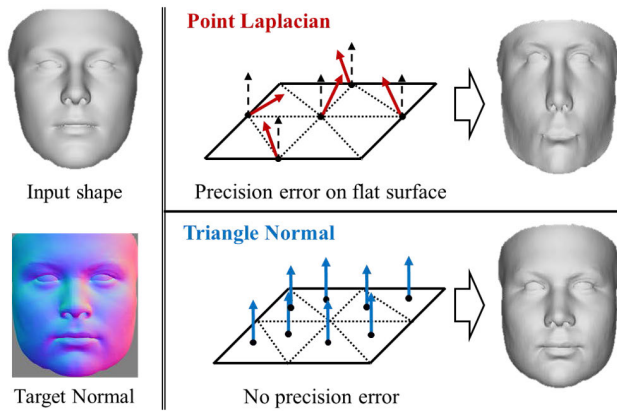
After both the global and local SH illumination coefficients and albedo are estimated, we can compute the triangle normal  $\mathbf{n}(x, y)$ . Because solely minimizing  $E_I$  is an under-constrained problem, the normal stabilization constraint defined as the difference between the estimated normal  $\mathbf{n}(x, y)$  in the previous iteration and current normal  $\mathbf{n}(x, y)$  is applied to the objective function. Additionally, the smoothness of the estimated normals is penalized by adding a gradient term. The normal at each location  $(x, y)$  is then computed using weights  $\lambda_{n_o}$  and  $\lambda_{n_s}$  by minimizing the following energy formula:

$$E_n = E_I + \lambda_{n_o} \sum_{(x,y) \in \Omega} \|\mathbf{n}(x, y) - \mathbf{n}^o(x, y)\|^2 + \lambda_{n_s} \sum_{(x,y) \in \Omega} \|\nabla \mathbf{n}(x, y)\|^2. \quad (14)$$

Note that, a mapping between the triangle of shape model to the location  $(x, y)$  at each iteration is computed by using z-buffering. Since we use only first-order SH, minimizing (14) is a linear least-squares problem that can be solved efficiently.

### D. 3D SHAPE RECOVERY

In most SfS methods for single-view facial shape recovery, the surface is reconstructed by computing a heightmap for the computed normal map. Such methods compute a depth value for each pixel location with assistance from a prior face model [17], [25]. However, these methods take a lot of time to converge, because these are formulated in a non-linear least-square problem. Roth *et al.* [14] proposed a linear SfS method based on mesh deformation using a surface Laplacian. Although this method was developed for unconstrained images, it is applicable to single-image face reconstruction with an appropriate regularizer. However, it requires suitable re-meshing prior to optimization because the surface Laplacian can be computed incorrectly around flat surfaces due to the precision error of the processor. Since the surface Laplacian is computed as the sum of the differences between a vertex and its adjacent vertices, if the vertex lies on the flat surface, the computed direction of Laplacian should be a zero-valued vector. In practice, due to limited precision, the Laplacian around a flat surface is a vector that has a magnitude larger than zero in a random direction. Therefore, the computed shape by using the surface Laplacian would be severely distorted.



**FIGURE 5.** Comparison of our deformation method with [14]. Without adapting the mesh structure, tetrahedron-based deformation produces accurate results, whereas the output of the Laplacian-based method is severely distorted.

To address this issue, we propose a tetrahedron-based deformation technique with triangle normals for the SfS process. Once the normal at  $(x, y)$  is computed as in the previous subsection, the 3D facial shape  $\mathbf{X}$  is recovered by computing the rotation of triangles of current facial shape  $\mathbf{X}^c$ . For each triangle, the target rotation matrix  $\mathbf{R}^j \in SO(3)$  between the current normal  $\mathbf{n}^c$  and target normal  $\mathbf{n}^j$  of the  $j^{\text{th}}$  triangle is computed. The target normal can be computed easily by back-projecting the 3D facial shape onto an image. The rotation between two vectors can be expressed by a tetrahedron representation [41] with vertices  $[\mathbf{x}_1^j, \mathbf{x}_2^j, \mathbf{x}_3^j]$  defined as

$$\mathbf{V} = [x_2^j - x_1^j, x_3^j - x_1^j, x_4^j - x_1^j], \quad (15)$$

where  $\mathbf{x}_4^j = \mathbf{x}_1 + \mathbf{n}^j$  is an additional vertex introduced to define the tetrahedron. By using (15), the rotation matrix  $\hat{\mathbf{R}}^j$  between the current  $\mathbf{V}^{jc}$  and target  $(\mathbf{V}^j)^{-1}$  for the  $j^{\text{th}}$  triangle can be computed as

$$\hat{\mathbf{R}}^j = \mathbf{V}^{jc} (\mathbf{V}^j)^{-1}. \quad (16)$$

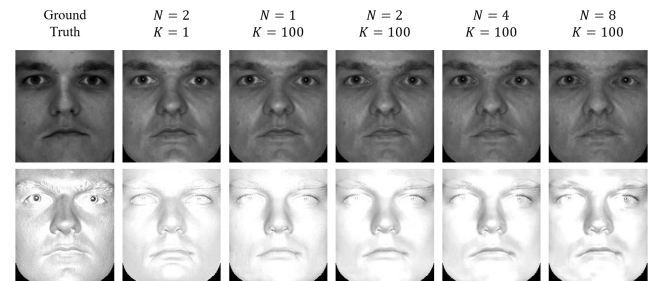
By using this representation, surface deformation can be performed by minimizing the following objective function  $E_d$ :

$$E_d = \sum_k \|\mathbf{R}^k - \hat{\mathbf{R}}^k\|_F^2 + \sum_{i \in \mathcal{B}} \|\mathbf{x}_i^c - \mathbf{x}_i\|_2^2, \quad (17)$$

where  $\mathcal{B}$  is the set of boundary vertices of the 3D facial shape. Because the first term in (17) can be converted into a system of linear equations, we can solve this minimization problem using the bi-conjugate gradient stabilized solver [42]. Fig. 5 presents an example of recovering 3D shape for input shape and target normal map by using the method proposed in [14] without re-meshing and tetrahedron-based method. Compared to the method proposed in [14], which stretches or shrinks flat regions, such as the cheek and forehead, our method reliably and accurately recovers the target shape from the given normal map.

## V. EXPERIMENTS

In this section, we present comprehensive experimental results. As an initial face model, we adopted the Basel face model [46], which is a widely used statistical parametric model for the shape and albedo of a face. The initial model with its projection parameters, including scale, translation, and rotation, was estimated using the 3DMM fitting method described in [4] with detected landmarks from [39]. Our method was tested using several public datasets and high-resolution photos from the internet. In all experiments, the hyperparameters  $\lambda_{L_c}$ ,  $\lambda_{L_s}$ ,  $\lambda_p$ ,  $\lambda_s$ ,  $\lambda_{n_o}$ , and  $\lambda_{n_s}$  are set to 1, 2, 1, 5, 0.1, and 0.1 respectively.



**FIGURE 6.** Estimated results of facial albedo with a fixed number of clusters ( $K = 100$ ) and increasing  $N$ .

**TABLE 1.** Estimation errors in terms of RS-MSE, A-SSiM, and S-SSiM. The error between  $K = 1$  and others are distinguishable, while the errors within  $K = 100$  exhibit no significant difference with orders.

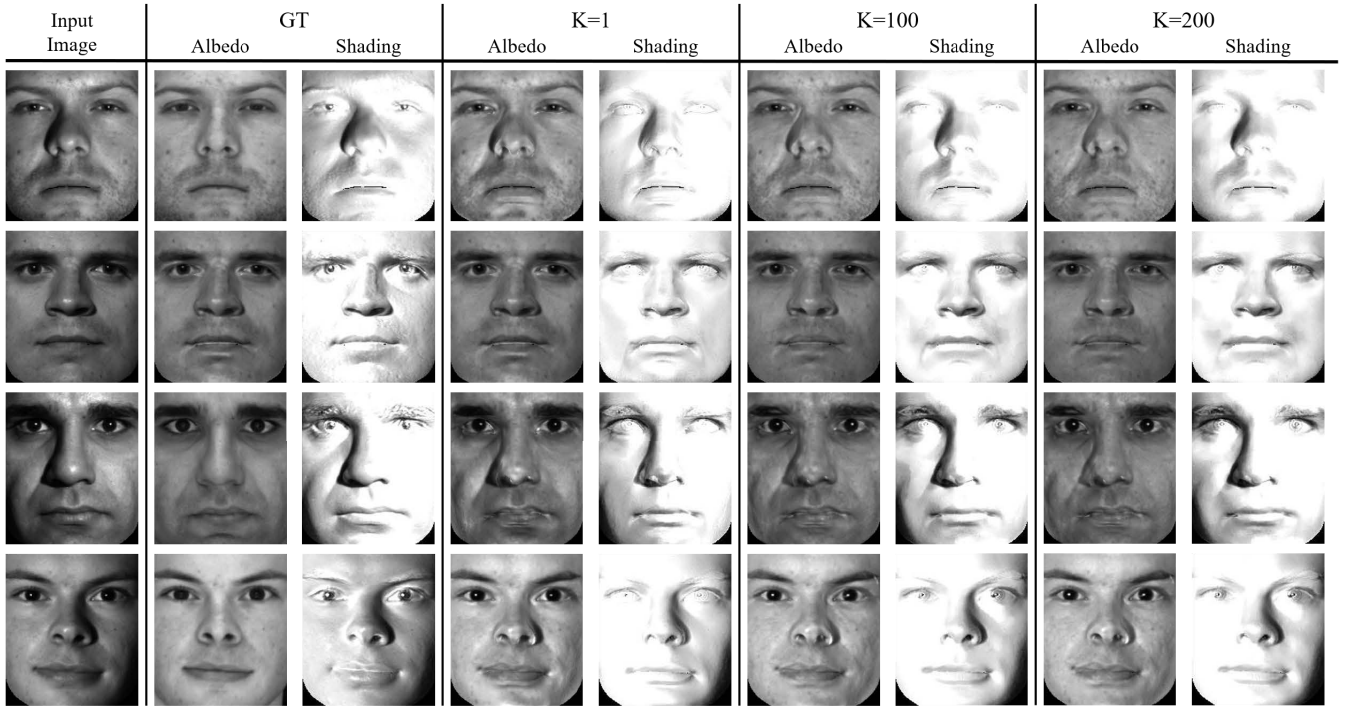
	N=2 (K=1)	N=1 (K=100)	N=2 (K=100)	N=4 (K=100)	N=8 (K=100)
RS-MSE	0.077	0.072	0.072	0.073	0.074
A-SSiM	0.859	0.870	0.868	0.867	0.863
S-SSiM	0.670	0.702	0.704	0.709	0.710

### A. ALBEDO RECOVERY

To demonstrate the effectiveness of the proposed lighting model, we tested our albedo recovery method, which was described in Section IV-B, on the YaleB face database [47]. The YaleB face database contains images of faces captured under different lighting conditions from a fixed viewing angle. For each subject, we computed a ground truth albedo using a generic photometric stereo technique based on the given ambient intensity image and direction of a light source for each image.

For the quantitative measure of estimated albedo and shading, we adopted three different error metrics. Reflectance/shading mean squared error (RS-MSE) is an error metric introduced in [48] that measures locally scale-invariant error for both reflectance and shading with values between zero and one. For the estimated gray albedo  $\rho$ , illumination (shading)  $L$ , and corresponding ground truth  $\rho^*$ ,  $L^*$ , the error is computed by summing the scale-invariant





**FIGURE 7.** Example results for albedo and shading estimation with ( $K = 100, 200$ ) or without ( $K = 1$ ) using the local SH lighting method. Effects such as self-shadow and specular highlights on nose-tip are effectively suppressed with  $K = 200$ .

MSE  $e(x, x^*)$  for all local windows  $w$  of size  $20 \times 20$  with strides of 10 as follows:

$$\text{RS-MSE}(\rho, L) = \frac{1}{2} \left( \frac{e(\rho, \rho^*)}{e(\rho, 0)} + \frac{e(L, L^*)}{e(L, 0)} \right), \quad (18)$$

$$e(x, x^*) = \sum_{w \in W} \min_a \|ax_w - x_w^*\|^2. \quad (19)$$

We also computed the SSiM [30], which is a measurement of the perceptual quality error between two images. This metric is sensitive to the local distortion of reconstructed albedo and shading. A high value of the SSiM indicates that a reconstructed signal is perceptually similar to the ground truth. Note that, the resulting SSiM values were calculated by averaging the SSiM values for all pixels in an image. The SSiM values for the estimated albedo and shading are denoted as A-SSiM and S-SSiM, respectively.

To assess the effects of local SH with varying orders, we compared estimation results for albedo. We set the initial number of clusters to  $K = 100$  for SLIC segmentation in the luminance domain. For orders of  $N = 1, 2, 4, 8$ , the results for RS-MSE, A-SSiM, and S-SSiM are listed in Table 1. Compared to the generic representation of illumination with  $N = 2$  and  $K = 1$ , the proposed representation produces better results in terms of the error metrics. However, for RS-MSE and A-SSiM, error increases with an order. Because the proposed method estimates illumination in local regions, rather than an entire image, the results with higher-order SH could be biased to the appearance of local clusters, resulting in distortion of the recovered albedo. For visual compari-

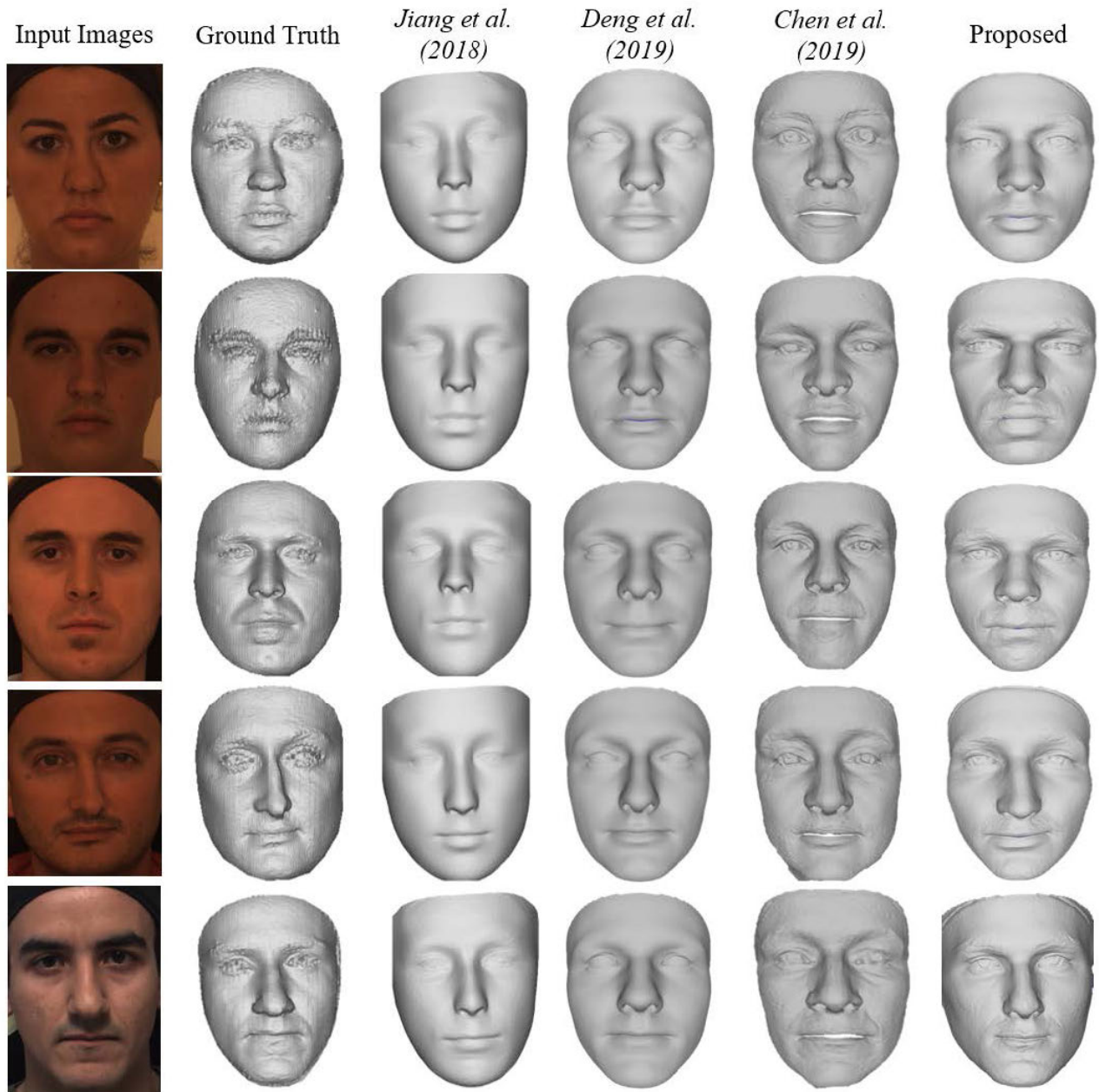
son, estimation results for albedo and shading are presented in Fig. 6. Note that  $K = 1$  corresponds to the model without local SH (i.e., global illumination only). One can see that there are no visible differences between the estimated albedo with varying orders, except for the result of global illumination with  $K = 1$ . Although there are no significant differences with increasing order, the specular region around the nose tip is diminished with the proposed method. These results serve as experimental proof for the validity of using first-order SH for local illumination, as discussed in Section IV.

**TABLE 2.** Average estimation errors for the YaleB face database. The means and standard deviations of the error values are presented.

	RS-MSE	S-SSiM	A-SSiM
$K = 1$	0.119 ( $\pm 0.025$ )	0.762 ( $\pm 0.064$ )	0.603 ( $\pm 0.049$ )
$K = 20$	0.114 ( $\pm 0.024$ )	0.772 ( $\pm 0.063$ )	0.632 ( $\pm 0.044$ )
$K = 50$	0.110 ( $\pm 0.022$ )	0.776 ( $\pm 0.062$ )	0.650 ( $\pm 0.040$ )
$K = 100$	0.108 ( $\pm 0.022$ )	0.779 ( $\pm 0.061$ )	0.660 ( $\pm 0.040$ )
$K = 150$	0.107 ( $\pm 0.022$ )	0.781 ( $\pm 0.0599$ )	0.666 ( $\pm 0.039$ )
$K = 200$	<b>0.106</b> ( $\pm 0.022$ )	<b>0.783</b> ( $\pm 0.060$ )	<b>0.670</b> ( $\pm 0.040$ )

Based on the observation that local SH of the first order provides sufficient quality for the estimated albedo,

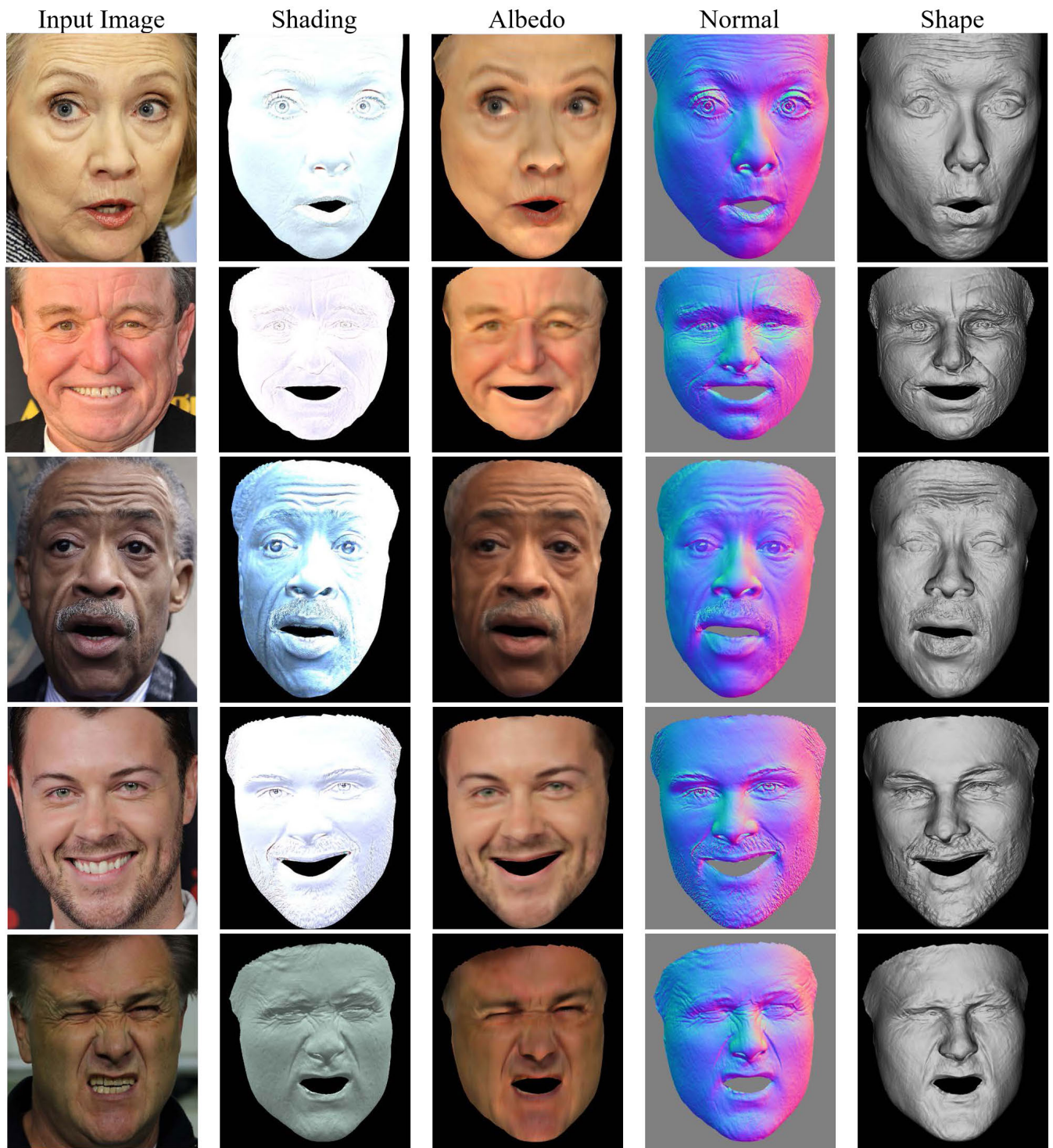




**FIGURE 8.** Comparison of results for 3D facial shape recovery from the Bosphorus database. We compare our proposed method to the methods proposed by Jiang et al. [17], Deng et al. [49], and Chen et al. [36]. Note that ground truth was constructed by meshing the noisy depth map provided in the database.

we present some example results with  $N = 1$  in Fig. 7. For albedo estimation with single SH with  $K = 1$ , the results reveal albedo highlights stronger than the true albedo around the specular regions, such as the nose tip and forehead. In contrast, our method can capture the specular regions more accurately, so specular highlights are suppressed compared to the single SH case. Moreover, regions with self-shadow can be effectively handled by our method. Except for the regions which are not under-saturated, the proposed method can recover the albedo accurately in softly shadowed regions. Even the hard shadow regions can be

approximated using the albedo prior constraint introduced in (13). The averages and standard deviations of RS-MSE, A-SSiM, and S-SSiM are listed in Table 2. As the number of divisions increases, better estimation results for albedo and shading are obtained. It seems that the increment of each SSiM value or decrement of RS-MSE with respect to increasing  $K$  saturated after  $K = 100$ . We found that the optimal number of clusters depends on the resolution of an input image. For our experiments on HD images, local SH with  $K = 300$  are sufficient for producing acceptable results.



**FIGURE 9.** Results of the proposed albedo and shape recovery method for faces under natural illumination. Because the subject in the last row was captured in an indoor environment, the estimated shading is darker than that in the other images.

### B. 3D SHAPE RECOVERY

We tested the entire proposed process from illumination estimation to 3D face shape recovery on images captured under constrained illumination and natural images with  $N = 1$

and the number of clusters  $K = 300$ . For images with constrained illumination, we sampled several images from the Bosphorus database [50], which provides structured-light scanned 3D point clouds of the faces of 105 subjects, as well



as corresponding RGB images. Note that, for the SfS-based method without any metric information, such as the true depth information of several points, error estimation using MSE is unconvincing. Furthermore, the scanned point clouds in the database are relatively noisy, making the quantitative measurement of real data impractical. Therefore, we present qualitative results for shape recovery with high-frequency details and exclude quantitative measures.

Fig. 8 presents comparative results for several images from the Bosphorus database for various methods. All of the compared methods are constructed using a single SH for modeling the illumination. The method from [17] is an SfS method that optimizes the height (depth) map using second-order SH. Because this method uses albedo statistics from 3DMM texture directly, the results exhibit over-smoothed geometry. It seems that the smoothness constraint is over-weighted because of the inaccurate estimation of albedo. The other two results were generated using pre-trained deep-learning models. While the method from [49] estimates the shape of a face by inferring the shape parameters of a 3DMM, a detailed facial shape with a UV displacement map is generated in [36]. Although our method is based on iterative linear optimization, the quality of the recovered 3D shapes is comparable to that of the latest methods based on deep-learning.

We conducted further experiments on the photos taken under natural illumination. The resulting shading, albedo, and normal, as well as the recovered 3D shape of faces, are presented in Fig. 9. One can see that the proposed local SH model can effectively handle natural illumination. Therefore, plausible results for albedo and normal, as well as detailed facial geometry, such as wrinkles around the eye and forehead or beards, can be obtained.

## VI. CONCLUSION

In this paper, we presented a novel 3D facial shape recovery framework for a single image with a novel parameterization of local SH illumination. The proposed local SH illumination model with an anisotropic weighting strategy allows facial albedo and 3D shape recovery to enhance the details of a face while preventing noisy deformation. Additionally, shape recovery using tetrahedron-based deformation facilitates reliable and accurate reconstruction for SfS with a prior face model. Experimental results demonstrated that our method is capable of recovering plausible facial albedo and 3D shape with enhanced details compared to recent state-of-the-art methods.

Nevertheless, the proposed local SH method is proven to be robust to complex illumination in natural scenes, identifying local regions by performing clustering based on pixel luminance may yield poor results. Because each local region is computed based on its luminance discrepancies, the resulting segmentation can produce regions containing rough geometry. Therefore, the geometry of the face should be incorporated in determining local clusters. This can be accomplished by incorporating the geometrical information of the initial shape using a graph-based optimization technique. Also, our

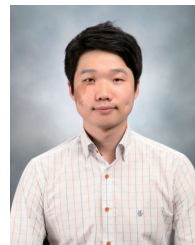
method can be combined with a statistical framework for physically based facial skin reflectance, such as the framework described in [51]. In future work, the proposed local SH model could be extended by accounting for these considerations.

## REFERENCES

- [1] G. Hu, F. Yan, C. Chan, W. Deng, W. Christmas, J. Kittler, and N. M. Robertson, "Face recognition using a unified 3D morphable model," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 73–89.
- [2] J. Park, S. Heo, K. Lee, H. Song, and S. Lee, "Robust facial pose estimation using landmark selection method for binocular stereo vision," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 186–190.
- [3] T. Choi, J. Kang, H. Song, and S. Lee, "Fitting facial models to spatial points: Blendshape approaches and benchmark," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2650–2654.
- [4] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 787–796.
- [5] C. Cao, Q. Hou, and K. Zhou, "Displaced dynamic expression regression for real-time facial tracking and animation," *ACM Trans. Graph.*, vol. 33, no. 4, p. 43, 2014.
- [6] S. Choi, A.-D. Nguyen, J. Kim, S. Ahn, and S. Lee, "Point cloud deformation for single image 3d reconstruction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2379–2383.
- [7] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Techn.* Reading, MA, USA: Addison-Wesley, 1999, pp. 187–194.
- [8] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
- [9] O. Aldrian and W. A. P. Smith, "Inverse rendering of faces with a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1080–1093, May 2013.
- [10] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou, "3D face morphable models 'in-the-wild,'" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5464–5473.
- [11] P. Dou, S. K. Shah, and I. A. Kakadiaris, "End-to-end 3D face reconstruction with deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 21–26.
- [12] X. Zhu, X. Liu, Z. Lei, and S. Z. Li, "Face alignment in full pose range: A 3D total solution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 78–92, Jan. 2019.
- [13] X. Cao, Z. Chen, A. Chen, X. Chen, C. Wang, and J. Yu, "Sparse photometric 3D face reconstruction guided by morphable models," 2017, *arXiv:1711.10870*. [Online]. Available: <http://arxiv.org/abs/1711.10870>
- [14] J. Roth, Y. Tong, and X. Liu, "Adaptive 3D face reconstruction from unconstrained photo collections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4197–4206.
- [15] F. Langguth, K. Sunkavalli, S. Hadap, and M. Goesele, "Shading-aware multi-view stereo," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 469–485.
- [16] M. Zollhöfer, A. Dai, M. Innmann, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, "Shading-based refinement on volumetric signed distance functions," *ACM Trans. Graph.*, vol. 34, no. 4, p. 96, 2015.
- [17] L. Jiang, J. Zhang, B. Deng, H. Li, and L. Liu, "3D face reconstruction with geometry details from a single image," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4756–4770, Oct. 2018.
- [18] X. Zeng, X. Peng, and Y. Qiao, "DF2Net: A dense-fine-finer network for detailed 3D face reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 2315–2324.
- [19] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.
- [20] C. Li, S. Lin, K. Zhou, and K. Ikeuchi, "Specular highlight removal in facial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3107–3116.
- [21] B. Tunwattapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec, "Acquiring reflectance and shape from continuous spherical harmonic illumination," *ACM Trans. Graph.*, vol. 32, no. 4, p. 109, 2013.



- [22] T. Whitted, "An improved illumination model for shaded display," in *Proc. ACM Siggraph Courses*, 2005, p. 4.
- [23] A. E. Ichim, S. Bouaziz, and M. Pauly, "Dynamic 3D avatar creation from hand-held video input," *ACM Trans. Graph.*, vol. 34, no. 4, p. 45, 2015.
- [24] H. Zhou, X. Yu, and D. Jacobs, "GLOSH: Global-local spherical harmonics for intrinsic image decomposition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 7820–7829.
- [25] I. Kemelmacher-Shlizerman and R. Basri, "3D face reconstruction from a single image using a single reference face shape," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 394–405, Feb. 2011.
- [26] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3D morphable model learnt from 10,000 faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5543–5552.
- [27] L. Tran and X. Liu, "Nonlinear 3D face morphable model," 2018, *arXiv:1804.03786*. [Online]. Available: <http://arxiv.org/abs/1804.03786>
- [28] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec, "Multiview face capture using polarized spherical gradient illumination," *ACM Trans. Graph.*, vol. 30, no. 6, p. 129, 2011.
- [29] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel, "Laplacian surface editing," in *Proc. Eurograph./ACM SIGGRAPH Symp. Geometry Process.*, 2004, pp. 175–184.
- [30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [31] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 884–899, May 2009.
- [32] D. Nguyen, S. Choi, W. Kim, and S. Lee, "GraphX-convolution for point cloud deformation in 2D-to-3D conversion," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 8628–8637.
- [33] S. Sengupta, A. Kanazawa, C. D. Castillo, and D. W. Jacobs, "SfSNet: Learning shape, reflectance and illuminance of faces 'in the wild,'" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6296–6305.
- [34] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3D face reconstruction and dense alignment with position map regression network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 534–551.
- [35] J. Kang, S. Lee, and S. Lee, "UV completion with self-referenced discrimination," in *Eurographics—Short Papers*, A. Wilkie and F. Banterle, Eds. Darmstadt, Germany: The Eurographics Association, 2020.
- [36] A. Chen, Z. Chen, G. Zhang, K. Mitchell, and J. Yu, "Photo-realistic facial details synthesis from single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 9429–9439.
- [37] D. Frolova, D. Simakov, and R. Basri, "Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting," in *Proc. Eur. Conf. Comput. Vis.* Prague, Czech Republic: Springer, 2004, pp. 574–587.
- [38] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [39] H. Song, J. Kang, and S. Lee, "ConcatNet: A deep architecture of concatenation-assisted network for dense facial landmark alignment," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2371–2375.
- [40] C. Loop, "Smooth subdivision surfaces based on triangles," M.S. thesis, Dept. Math., Univ. Utah, Salt Lake City, UT, USA, 1987.
- [41] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 399–405, 2004.
- [42] H. A. van der Vorst and A. Henk, "Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems," *Stat. Comput.*, vol. 13, no. 2, pp. 631–644, Mar. 1992.
- [43] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross, "Analysis of human faces using a measurement-based skin reflectance model," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1013–1024, Jul. 2006.
- [44] A. Nealen, T. Igarashi, O. Sorkine, and M. Alexa, "Laplacian mesh optimization," in *Proc. 4th Int. Conf. Comput. Graph. Interact. Techn. Australasia Southeast Asia (GRAPHITE)*, 2006, pp. 381–389.
- [45] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 479–486.
- [46] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 296–301.
- [47] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
- [48] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2335–2342.
- [49] Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong, "Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2019. [Online]. Available: [https://openaccess.thecvf.com/content\\_CVPRW\\_2019/html/AMFG/Deng\\_Accurate\\_3D\\_Face\\_Reconstruction\\_With\\_Weakly-Supervised\\_Learning\\_From\\_Single\\_Image\\_CVPRW\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPRW_2019/html/AMFG/Deng_Accurate_3D_Face_Reconstruction_With_Weakly-Supervised_Learning_From_Single_Image_CVPRW_2019_paper.html)
- [50] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and B. L. Akarun, *Bosphorus Database for 3D Face Analysis*. Roskilde, Denmark: Springer, 2008, pp. 47–56.
- [51] S. Yamaguchi, S. Saito, K. Nagano, Y. Zhao, W. Chen, K. Olszewski, S. Morishima, and H. Li, "High-fidelity facial reflectance and geometry inference from an unconstrained image," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–14, Aug. 2018.



**SUWOONG HEO** received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering. His main research interests include computer vision, graphics, and machine learning.



**HYEWON SONG** received the B.S. degree from Hongik University, Seoul, South Korea, in 2015. She is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul. Her main research interests include computer vision and machine learning.



**JIWOO KANG** received the B.S., M.S., and Ph.D. degrees from Yonsei University, Seoul, South Korea, in 2011 and 2019, respectively, all in electrical and electronic engineering. He is currently working with the Multidimensional Insight Laboratory, Yonsei University, as a Postdoctorate Researcher of Electrical and Electronic Engineering. His research interests include computer graphics, computer vision, and image analysis.



**SANGHOON LEE** (Senior Member, IEEE) received the B.S. degree from Yonsei University, Seoul, South Korea, in 1989, the M.S. degree from the Korea Advanced Institute of Science and Technology, South Korea, in 1991, and the Ph.D. degree from the University of Texas at Austin, Austin, TX, USA, in 2000. From 1991 to 1996, he worked with Korea Telecom, South Korea. From 1999 to 2002, he worked with Lucent Technologies, Murray Hill, NJ, USA. In 2003, he joined the EE Department,

Yonsei University, as a Faculty Member, where he is currently serving as a full-time Professor. His current research interests include image/video processing, computer vision, and graphics. He is a Board of Governors Member of APSIPA. He received the 2015 Yonsei Academic Award from

Yonsei University, the 2012 Special Service Award from the IEEE Broadcast Technology Society, and the 2013 Special Service Award from the IEEE Signal Processing Society. He was the General Chair of the 2013 IEEE IVMSWP Workshop and has served as a steering committee member for the IEEE and APSIPA conferences. He has also served as the Chair of the IEEE P3333.1 Quality Assessment Working Group since 2011, and worked as the IVM Technical Committee Chair of APSIPA from 2018 to 2019. He served on the IEEE IVMSWP Technical Committee from 2014 to 2019, and the IEEE MMSP Technical Committee since 2016. He has served as the Editor for the *Journal of Communications and Networks* from 2009 to 2015 and as a special issue Guest Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING in 2013. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014, and served as an Associate Editor from 2014 to 2018. He is also a Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS.

• • •