Pattern Recognition 44 (2011) 1470–1486

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A SfM-based 3D face reconstruction method robust to self-occlusion by using a shape conversion matrix

Sung Joo Lee [a], Kang Ryoung Park [b], Jaihie Kim [a,*]

[a] School of Electrical and Electronic Engineering, Yonsei University, Biometrics Engineering Research Center, Seoul, South Korea
[b] Division of Electronics and Electrical Engineering, Dongguk University, Biometrics Engineering Research Center, South Korea

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form
18 October 2010
Accepted 16 November 2010
Available online 27 November 2010

Keywords:
Structure from motion (SfM)
3D face reconstruction
Self-occlusion
Shape conversion matrix

ABSTRACT

This paper presents a 3D face reconstruction method using multiple 2D face images. Structure from motion (SfM) methods, which have been widely used to reconstruct 3D faces, are vulnerable to point correspondence errors caused by self-occlusion. In order to solve this problem, we propose a shape conversion matrix (SCM) which estimates the ground-truth 2D facial feature points (FFPs) from the observed 2D FFPs corrupted by self-occlusion errors. To make the SCM, the training observed 2D FFPs and ground-truth 2D FFPs are collected by using 3D face scans. An observed shape model and a ground-truth shape model are then built to represent the observed 2D FFPs and the ground-truth 2D FFPs, respectively. Finally, the observed shape model parameter is converted to the ground truth shape model parameter via the SCM. By using the SCM, the true locations of the self-occluded FFPs are estimated exactly with simple matrix multiplications. As a result, SfM-based 3D face reconstruction methods combined with the proposed SCM become more robust against point correspondence errors caused by self-occlusion, and the computational cost is significantly reduced. In experiments, the reconstructed 3D facial shape is quantitatively compared with the 3D facial shape obtained from a 3D scanner, and the results show that SfM-based 3D face reconstruction methods with the proposed SCM show a higher accuracy and a faster processing time than SfM-based 3D face reconstruction methods without the SCM.

1. Introduction

Modeling 3D faces is a very useful technology in computer vision for its various applications, including pose-invariant face recognition [1,2], age-invariant face recognition [3], 3D person-specific game and movie character generation [4,5], teleconferencing, surgical simulation, etc. One approach used to model a 3D face requires special hardware or multi-cameras, such as a 3D laser scanner, stereo cameras, or a structured light [6]. However, applications using this approach are restrictive because of high costs and camera calibration. As an alternative, 3D face reconstruction methods by using an image sequence have been researched intensively. These methods can be categorized into model-based and SfM-based methods.

Model-based methods build 3D morphable face models offline to represent facial shape and texture, illumination, and camera geometry, with a large number of model parameters [1,7]. When 2D facial images are inputted, the methods find the model parameters of the 3D morphable face model iteratively in order to minimize the texture residual between the 2D facial image synthesized from the model parameters and the inputted 2D facial image. If the optimal model parameters are found, a detailed 3D facial shape can then be reconstructed from these model parameters. However, these methods require a high computational costs and possibly fall into a local minima because the number of model parameters is considerably large [8,9]. If they fall into a local minima, the reconstructed 3D face tends to be close to the mean face because the parameter optimization begins at the mean face [8,9]. In order to reduce the computational complexities of the original 3D morphable model, simplified morphable models representing only facial shape have been proposed in [3,10–12].

SfM-based methods estimate a 3D facial shape and a projection matrix from the corresponding 2D facial feature points (FFPs) of multiple facial images [8,9,13–16]. The basic idea of SfM is that the corresponding 2D FFPs can be factorized into the 3D facial shape and the projection matrix by using some reasonable constraints, such as the rank constraint [17]. Compared to model-based methods, these methods do not require the parameter optimization started from the mean of the 3D faces. As a result, a person-specific 3D facial shape can be reconstructed by using SfM [8,9].

SfM-based methods can be categorized into dense correspondence-based methods and sparse correspondence-based methods according to the density of the corresponding 2D FFPs. Dense correspondence-based methods find dense corresponding 2D FFPs,

* Corresponding author.
E-mail addresses: sungjoo@yonsei.ac.kr (S.J. Lee), parkgr@dongguk.edu (K.R. Park), jhkim@yonsei.ac.kr (J. Kim).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.11.012

as shown in Fig. 1(a), and reconstruct a dense 3D facial shape from them [8,9]. If dense corresponding 2D FFPs can be correctly found, these methods can reconstruct a detailed 3D facial shape. However, in general, it is difficult to find dense corresponding 2D FFPs since some facial regions, such as the cheek and the forehead, have no salient texture patterns. When point correspondence errors occur, the reconstructed 3D face appears rough. In order to obtain a smooth reconstructed 3D face, [8,9] used the mean 3D face to regularize the reconstructed 3D face.

Sparse correspondence-based methods find a predetermined number of sparse and salient corresponding 2D FFPs, as shown in Fig. 1(b), and reconstruct a sparse 3D facial shape from them [13–16]. A dense 3D mean face is then adapted to the reconstructed sparse 3D facial shape [13,14]. By using these methods, a smooth and dense 3D face can be reconstructed without dense corresponding 2D FFPs. In addition, in some applications, such as 3D game character generation, sparse corresponding 2D FFPs can be found with the assistance of the user. Therefore, these methods are very useful in such cases. The proposed 3D reconstruction method belongs to this category.

One of the fundamental problems of SfM-based methods is self-occlusion which means some facial parts occlude other facial parts when head rotation occurs. Fig. 2(b) shows the ground-truth 2D FFPs that do not include point correspondence errors, and the observed 2D FFPs when the head orientation is frontal. Fig. 2(a) and (c) show them when the head is significantly rotated. It can be seen from Fig. 2 that the error between the observed 2D FFPs and the ground truth ones increases as head rotation increases due to self-occlusion. This kind of correspondence error can degrade the performance of the SfM algorithms.

In order to solve this problem, previous SfM methods found the initial 3D facial shape based on all the observed 2D FFPs, and then reconstructed a detailed 3D facial shape iteratively by minimizing the shape residual of the visible 2D FFPs [18–20]. The matrix completion method estimated the true locations of the self-occluded FFPs by minimizing both the shape residual of the visible 2D FFPs and the nuclear norm of the estimated matrix [35,36]. However, these methods have the following limitations: Firstly, these methods are sensitive to point correspondence errors because of the reduced number of useful 2D FFPs. According to Szeliski and Kang [21], the SfM algorithms were less sensitive to point correspondence errors as the number of FFPs and the amount of object rotation increase if each FFP has an equivalent correspondence error. Unfortunately, in the face, it is difficult to obtain both a large number of FFPs and a sizable head rotation simultaneously, because more FFPs become self-occluded as head rotation increases. Secondly, these methods basically require a lot of iterations which lead to high computational costs. Finally, these methods require an additional self-occlusion detector in order to find the visible FFPs. As a result, performance is dependent on the self-occlusion detector.

In order to solve these problems, we propose a shape conversion matrix (SCM) to estimate the true locations of the self-occluded FFPs. In other words, the SCM transforms the observed facial shape to a converted facial shape that is closer to the ground truth facial shape shown in Fig. 2.

In order to make the SCM, observed 2D FFPs and ground-truth 2D FFPs were collected from subjects and two shape models were built to represent these 2D FFPs using principal component analysis (PCA). Then, two shape parameters were found by projecting the two types of 2D FFPs onto their corresponding shape models. The SCM was found to convert the observed shape parameter to the ground truth shape parameter by using a least square method. By using the SCM, the true locations of the self-occluded FFPs are estimated exactly with simple matrix multiplications. As a result, SfM-based 3D face reconstruction methods combined with the proposed SCM become more robust against correspondence errors with only a small additional computational cost. Furthermore, the proposed method does not require a self-occlusion detector to find the visible points because the self-occluded points are estimated from all of the observed FFPs by using the SCM. Table 1 shows a comparison between previous 3D face reconstruction methods and the proposed method.

The remainder of the paper is organized as follows: Section 2 describes the general procedures for sparse correspondence-based 3D face reconstruction and previous solutions for the self-occlusion problem. Section 3 presents the proposed method using the SCM. Section 4 lays out our experimental environment and the quantitative and qualitative results. Finally, conclusions are given in Section 5.

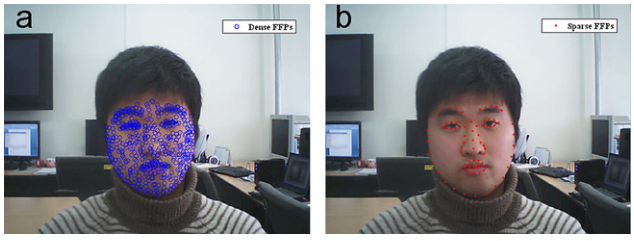

Fig. 1. Examples of 2D FFPs: (a) dense 2D FFPs and (b) sparse 2D FFPs.

2. Related works

2.1. General procedure for sparse correspondence-based 3D face reconstruction

As shown in Fig. 3, the general procedure for sparse correspondence-based 3D face reconstruction consists of sparse 2D FFPs extraction, 3D reconstruction of the sparse FFPs using SfM, 3D dense mean model adaptation, and texture mapping.

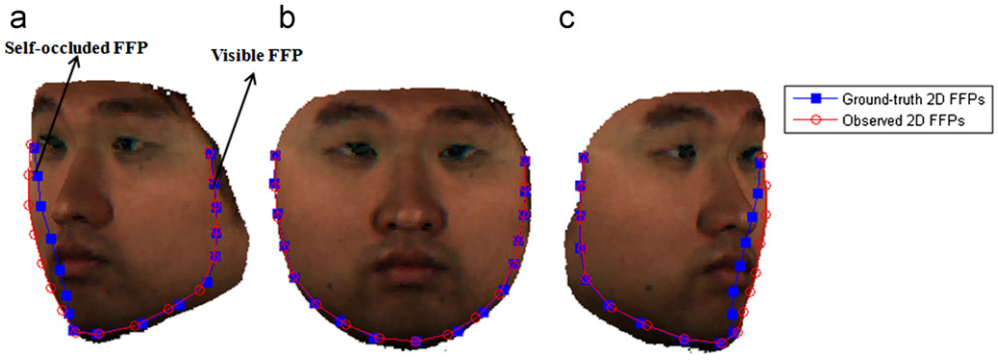

Fig. 2. 2D facial contour FFPs when head rotation is: (a) −45°, (b) 0°, and (c) 45°.

Table 1
Comparison of the proposed method to previous 3D face reconstruction methods.

Categorization			Strength	Weakness
Special hardware-based methods [6]			• Accurate	• High cost • Require calibration
Software-based methods	Model-based method [1,3,7–10]		• Require only single view • If optimal model parameters are found, detailed shape of cheek and forehead can be reconstructed	• Reconstructed results are rather close to the average 3D face when the parameter optimization falls into a local minima • High computational cost
	SfM-based methods	Dense correspondence [8,9]	• Can obtain person-specific 3D face • If dense correspondence is found without error, detailed shape of cheek and forehead can be reconstructed	• Vulnerable to self-occlusion • Unstable due to dense correspondence error • Require multi-view images
		Sparse correspondence [13–16]	• Can obtain person-specific 3D face • Relatively stable because only sparse correspondence is needed	• Vulnerable to self-occlusion • Reconstructed shapes of cheek and forehead are relatively coarse because these region do not have FFPs • Require multi-view images
		Sparse correspondence +SCM (Proposed)	• Can obtain person-specific 3D face • Relatively stable because only sparse correspondence is needed • **Robust to self-occlusion**	• Reconstructed shapes of cheek and forehead are relatively coarse because these regions do not have FFPs • Requires multi-view images

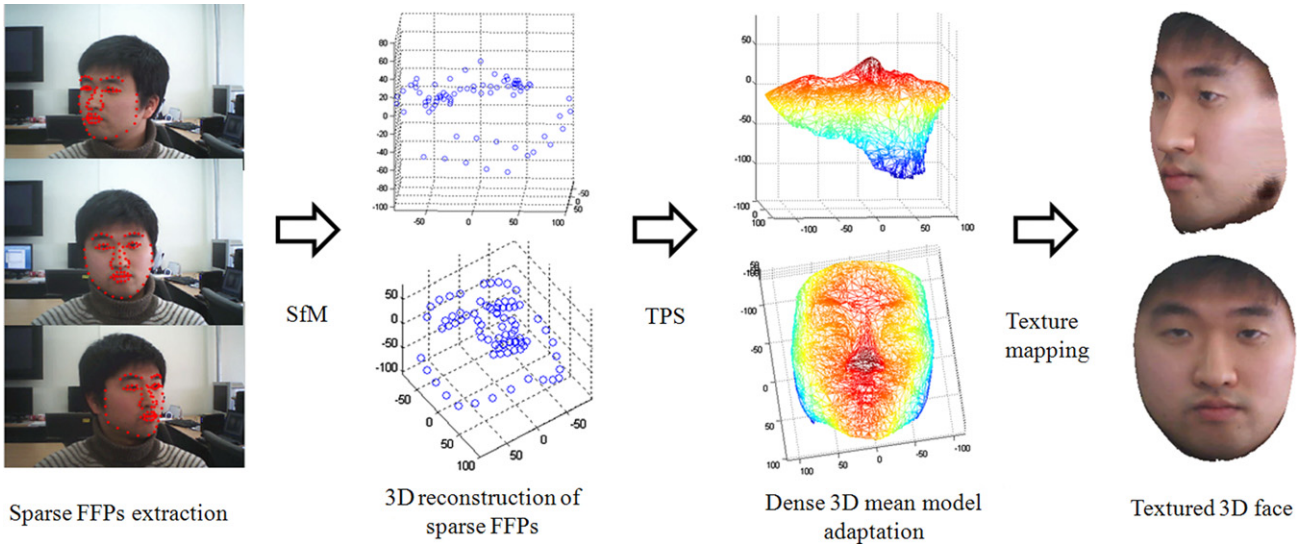

Fig. 3. General procedure for sparse correspondence-based 3D face reconstruction.

When F images, $\mathbf{I}_1,\mathbf{I}_2,\ldots,\mathbf{I}_F$, are captured, sparse correspondence-based methods find a predetermined number of sparse 2D FFPs such as the eyes, nose, and facial contour as shown in Fig. 3. These feature points can be localized by either manual annotation [13] or by an automatic feature detector such as an active appearance model [22–24], an active shape model [25], etc. If N points are localized, we can obtain the measurement matrix $\mathbf{W}$ as follows:

$$\mathbf{W}=\begin{bmatrix} x_{11} & \cdots & x_{1N} \\ y_{11} & \cdots & y_{1N} \\ \vdots & \ddots & \vdots \\ x_{F1} & \cdots & x_{FN} \\ y_{F1} & \cdots & y_{FN} \end{bmatrix} \quad (1)$$

where x_{fn} and y_{fn} refer to the x position of the nth feature point in the fth image and the y position, respectively.

In order to find the 3D sparse facial shape $\mathbf{S}$ composed of 3D FFPs, $\mathbf{W}$ is factorized by using the scaled-orthographic camera model as follows:

$$\mathbf{W}=\mathbf{PS}+\mathbf{t}^T\mathbf{1} \quad (2)$$

where $\mathbf{P}$ is the $2F\times 3$ projection matrix, $\mathbf{S}$ is the $3\times N$ 3D sparse shape, $\mathbf{t}$ is the $1\times 2F$ translation vector, and $\mathbf{1}$ is the $1\times N$ vector, the components of which are all 1. The scaled-orthographic camera model is a simplified camera model so that if the distance between the camera and the face, or the focal length of the camera is very small, then the reconstructed 3D face can be erroneous due to the perspective effect [38]. However, it is efficient and effective when the distance between the camera and the face, or the focal length of

the camera, is relatively large [17,37]. If we assume the origin of the world coordinates is placed at the centroid of the 3D object shape, the translation vector $\mathbf{t}$ can be described by

$$\mathbf{t}=\left[\frac{1}{N}\sum_{i=1}^{N}W_{1i}\cdots\frac{1}{N}\sum_{i=1}^{N}W_{2Fi}\right] \tag{3}$$

where W_{ab} is the (a, b) component in the matrix $\mathbf{W}$. Then the SfM algorithm [17,26] calculates the registered measurement matrix $\tilde{\mathbf{W}}$ by subtracting $\mathbf{t}$ from $\mathbf{W}$ as follows:

$$\tilde{\mathbf{W}}=\mathbf{W}-\mathbf{t}^T\mathbf{1}=\mathbf{PS} \tag{4}$$

From the given $\tilde{\mathbf{W}}$, the widely used SfM algorithm known as the factorization method finds $\mathbf{S}$ and $\mathbf{P}$ by using the rank constraint and singular value decomposition(SVD) [17,26]. The factorization method is a closed form solution, so that it requires very little computational costs. In addition, this method is proven to provide an optimal solution if we assume $\mathbf{W}$ has the equivalent Gaussian noise for each component of $\mathbf{W}$ [27].

After obtaining the reconstructed 3D sparse facial shape $\mathbf{S}$, a 3D dense mean face model is adapted to $\mathbf{S}$. A widely used adaptation method is the thin-plate spline (TPS) [28,29]. The TPS finds a nonlinear interpolation function f to convert the 3D sparse FFPs in the 3D dense mean face model into $\mathbf{S}$ as follows:

$$f(\mathbf{u})=\mathbf{t}+\mathbf{Ru}+\mathbf{D}^Ts(\mathbf{u})=\mathbf{S} \tag{5}$$

where $\mathbf{u}$ is $3\times N$ 3D sparse FFPs in the 3D dense mean face model. s is the Spline function, $\mathbf{R}$ is a rotation matrix, $\mathbf{D}$ is a deformation parameter matrix, and $\mathbf{t}$ is a translation vector. These parameters can be obtained by using the least square method [29]. The person-specific 3D dense shape can be reconstructed by converting other points in the 3D dense mean face model using the interpolation function f.

Finally, the textures in the captured images are mapped to the reconstructed 3D dense facial shape to make the textured 3D facial shape shown in Fig. 3.

2.2. Previous solutions

The factorization method shows a reliable 3D reconstruction result when all the FFPs are visible but its performance degrades seriously if errors caused by self-occlusion occur as it was explained in Section 1. If the self-occluded FFPs are excluded, missing components occur in the registered measurement matrix $\tilde{\mathbf{W}}$ and we cannot obtain the 3D sparse facial shape $\mathbf{S}$ by using the factorization method [26]. Although a solution for the missing components was proposed in [17], it cannot solve a generic missing data problem [18,20].

As an alternative, there have been approaches to find $\mathbf{S}$ from $\tilde{\mathbf{W}}$ having missing components by using iterative methods [18,19]. These methods minimized an objective function:

$$(\hat{\mathbf{P}},\hat{\mathbf{S}})=\mathrm{argmin}_{\mathbf{P},\mathbf{S}}||\mathbf{M}\odot(\tilde{\mathbf{W}}-\mathbf{PS})||_F \tag{6}$$

where $\mathbf{M}$ is the masking matrix, $\hat{\mathbf{P}}$ and $\hat{\mathbf{S}}$ are the estimated projection matrix and the 3D facial shape, respectively. The symbol $\odot$ refers to the Hardamard product and $\|\|_F$ denotes the Frobenius norm. The components of masking matrix $\mathbf{M}$ consist of 0 and 1, which correspond to self-occluded FFPs and visible FFPs, respectively. Marques et al. minimized (6) to find $\hat{\mathbf{P}}$ and $\hat{\mathbf{S}}$ [18]. They initialized the missing components with random numbers and found $\hat{\mathbf{P}}$ and $\hat{\mathbf{S}}$ by using a factorization method [18]. From $\hat{\mathbf{P}}$ and $\hat{\mathbf{S}}$, they estimated the missing components and updated the estimated result to $\tilde{\mathbf{W}}$ until there was a convergence. Wiberg also minimized (6) to find $\hat{\mathbf{P}}$ and $\hat{\mathbf{S}}$ [19]. His algorithm is called an alternation algorithm because he repeatedly found the projection matrix and the shape matrix from the fixed shape matrix and the projection matrix, respectively.

Some previous work [20] added regularized terms to (6) as follows:

$$(\hat{\mathbf{P}},\hat{\mathbf{S}})=\mathrm{argmin}_{\mathbf{P},\mathbf{S}}||\mathbf{M}\odot(\tilde{\mathbf{W}}-\mathbf{PS})||_F+\lambda_1||\mathbf{P}||_F+\lambda_2||\mathbf{S}||_F \tag{7}$$

where λ_1, λ_2 are the regularizing parameters. Buchanan et al. minimized (7) to find $\hat{\mathbf{P}}$ and $\hat{\mathbf{S}}$ by using a damped Newton algorithm [20].

The matrix completion method [35,36] minimize not only the shape residual of the visible FFPs but also the nuclear norm of matrix $\mathbf{X}$:

$$(\hat{\mathbf{X}})=\mathrm{argmin}_{\mathbf{X}}\frac{1}{2}||\mathbf{M}\odot(\tilde{\mathbf{W}}-\mathbf{X})||_F^2+\mu||\mathbf{X}||_* \tag{8}$$

where $||\mathbf{X}||_*$ is the nuclear norm of matrix $\mathbf{X}$ and μ is a weight. The nuclear-norm minimization shown in (8) is the tightest convex relaxation of the following NP-hard rank minimization problem:

$$\begin{aligned}&\text{minimize}\quad \mathrm{rank}(\mathbf{X})\\&\text{subject to}\quad \mathbf{M}\odot\tilde{\mathbf{W}}=\mathbf{M}\odot\mathbf{X}\end{aligned} \tag{9}$$

Therefore, if we minimize (8), we can find matrix $\mathbf{X}$ which has a minimized rank and whose visible components are close to those of the input matrix $\tilde{\mathbf{W}}$. From solution $\hat{\mathbf{X}}$, we can estimate missing components of matrix $\tilde{\mathbf{W}}$. In other words, we can estimate the real locations of the self-occluded FFPs. Note that the matrix completion method can estimate the missing components but it cannot decompose the estimated matrix $\hat{\mathbf{X}}$ to obtain the 3D facial shape. Therefore, previous SfMs [18–20] can be used to decompose the estimated matrix $\hat{\mathbf{X}}$, and to obtain the 3D facial shape $\hat{\mathbf{S}}$.

From (6) to (8), it is clear that the basic idea of the previous SfM methods and the matrix completion method is to exclude the erroneous self-occluded FFPs and to minimize the shape residual calculated with only visible FFPs by using the masking matrix $\mathbf{M}$. However, these methods based on the visible FFPs are sensitive to correspondence errors due to the reduced number of FFPs. In addition, to obtain $\mathbf{M}$, they require an additional occlusion detector to determine whether or not each FFP is occluded. If the occlusion detector makes an error, it inevitably degrades the reconstruction performance. Finally, the previous methods require a lot of iterations resulting in a long processing time.

3. Proposed method

3.1. Overall procedure for the proposed 3D face reconstruction method

The overall procedure for the proposed 3D face reconstruction method is described in Fig. 4. The proposed shape conversion matrix (SCM) is added to the previous SfM-based 3D reconstruction

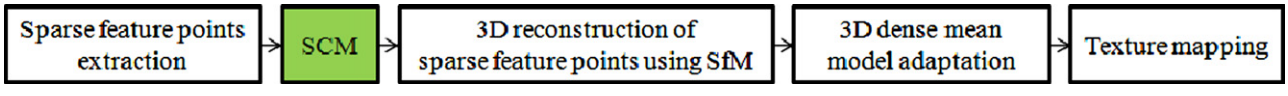

Fig. 4. Overall procedure for the proposed 3D face reconstruction method using SCM.

methods shown in Fig. 4. The SCM converts an observed 2D facial shape to a converted 2D facial shape that is closer to the ground-truth 2D facial shape, as shown in Fig. 5. By using the SCM, the proposed method can estimate the true locations of self-occluded 2D FFPs without the iterations required for minimizing the shape residual of visible FFPs, but with simple matrix multiplications.

3.2. Training procedure for making a SCM

The procedure used to obtain a SCM is shown in Fig. 6. In order to make a SCM, we need training ground-truth FFPs and observed FFPs from many training face images with different views. In order to obtain such training data, 3D face scans from different training subjects are obtained using a 3D scanner [30] and their 3D FFPs are obtained manually. A 3D face scan consists of over 0.1 million vertices, meshes, and texture coordinates. Let a 3D facial shape consisting of 3D FFPs in the kth face scan be $\mathbf{S}_k^{3D}$, then $\mathbf{S}_k^{3D}$ can be written as

$$\mathbf{S}_k^{3D} = \begin{bmatrix} x_{k1} & \cdots & x_{kN} \\ y_{k1} & \cdots & y_{kN} \\ z_{k1} & \cdots & z_{kN} \end{bmatrix} \tag{10}$$

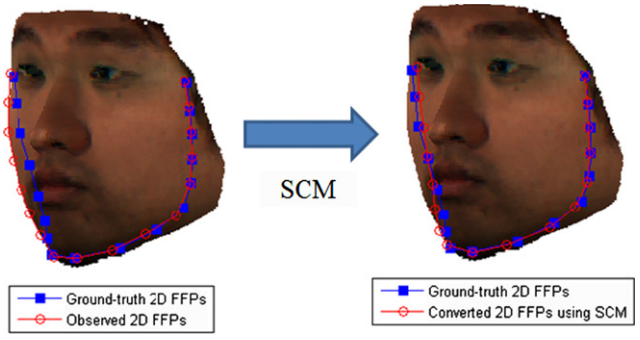

Fig. 5. Converted 2D FFPs using the SCM.

where N is the number of FFPs. x_{nk}, y_{nk}, and z_{nk} refer to x, y, and z position of nth FFP in the kth 3D face scan, respectively. Normally, the 3D face scans from different subjects are not well aligned. Consequently, a 3D alignment procedure is required to obtain aligned 3D face scans and their 3D facial shapes. The 3D alignment procedures are:

1. Choose a $\mathbf{S}_1^{3D}$ as a reference 3D facial shape.
2. Find the best matched similarity transforms to align the 3D facial shapes to the reference 3D facial shape by using the Procrustes analysis [31].
3. Select a new reference 3D facial shape as the mean of the aligned 3D facial shapes.
4. Repeat 2 and 3 until the average shape residual between the 3D facial shapes and the reference 3D facial shape is less than a predetermined threshold.
5. Align the 3D face scans by using the final similarity transforms obtained in Step 2.

The training ground-truth 2D facial shapes are then obtained by rotating the aligned 3D facial shapes and projecting them onto the 2D image plane. Likewise, different projected face images are obtained by rotating the aligned 3D face scans and projecting them, as shown in Fig. 6. From a projected 2D face image, its observed facial shape is obtained by manually annotating the FFPs. Specifically, $\mathbf{S}_k^{3D}$ is rotated and projected in order to obtain the ground-truth 2D facial shape $\mathbf{G}^{2D}$ consisting of $2 \times N$ FFPs.

$$\mathbf{G}^{2D} = s_I \mathbf{P}_I \mathbf{R}_C (\mathbf{R}_\theta \mathbf{S}_k^{3D} + \mathbf{t}_C) \tag{11}$$

where $\mathbf{R}_\theta$ is the 3×3 rotation matrix that rotates $\mathbf{S}_k^{3D}$, $\mathbf{R}_C$ and $\mathbf{t}_C$ are the rotation and translation matrices, respectively, that align the world coordinate to the camera coordinate. s_I and $\mathbf{P}_I$ are the scaling factor and the orthographic projection, respectively, that map a point in the camera coordinate to a point in the image coordinate. A virtual camera can be generated by using $\mathbf{R}_C$, $\mathbf{t}_C$, s_I and $\mathbf{P}_I$. These parameters are determined experimentally to obtain a proper image size. $\mathbf{R}_\theta$ determines the head rotation and we consider

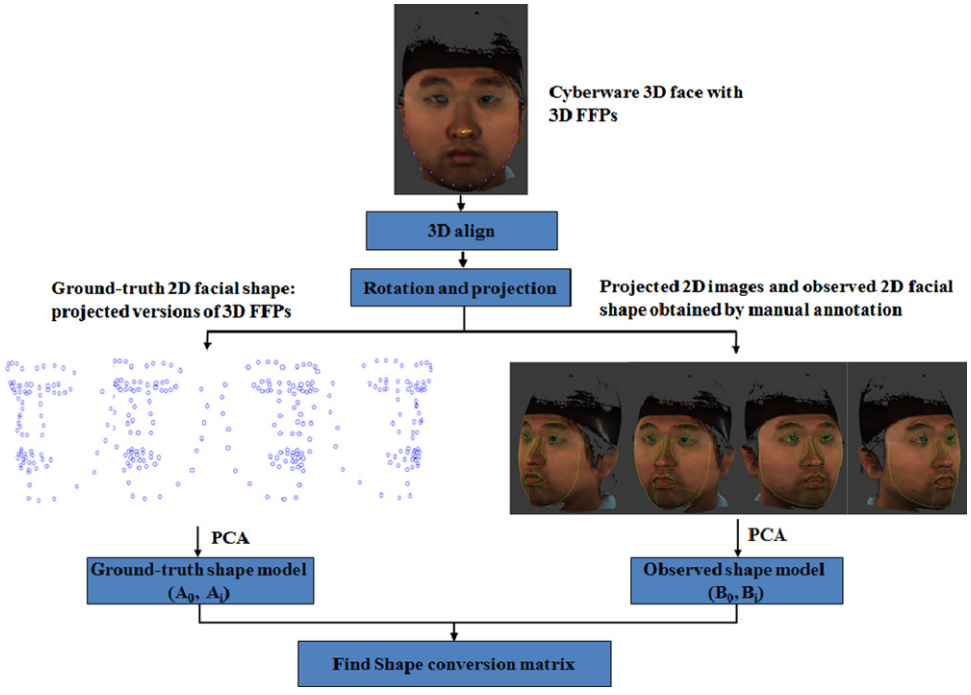

Fig. 6. Overall procedure for obtaining the SCM.

seven different yaws ($\theta = -45°, -30°, -15°, 0°, 15°, 30°$, and $45°$) in this paper. Therefore, seven different $\mathbf{R}_\theta$ are used to make the ground-truth 2D facial shapes in the 7 different view images. These procedures are repeated for all the aligned 3D facial shapes. The projected face images are obtained by rotating the aligned 3D face scans and projecting them in the same way by using $\mathbf{R}_\theta$, $\mathbf{R}_C$, $\mathbf{t}_C$, s_I, and $\mathbf{P}_I$. The observed facial shapes are then obtained by manual annotation in the projected face 2D images.

After obtaining the training ground-truth and the observed 2D facial shapes, the ground-truth shape model and the observed shape model were built from the ground-truth 2D facial shapes and the observed facial shapes, respectively. Specifically, let a ground-truth 2D facial shape of the vth projected image be $\mathbf{G}_v^{2D}$ and its observed facial shape is $\mathbf{O}_v^{2D}$. $\mathbf{G}_v^{2D}$ and $\mathbf{O}_v^{2D}$ are a $2N \times 1$ matrix composed of the x and y positions of N FFPs. $\mathbf{G}_v^{2D}$ is then represented by the ground-truth shape model that is the linear combination of the mean $\mathbf{A}_0$ and the shape variation $\mathbf{A}_i$ obtained by using the training ground-truth 2D shapes and the principal component analysis (PCA) as

$$\mathbf{G}_v^{2D} = \mathbf{A}_0 + \sum_{i=1}^{n} \alpha_{vi}\mathbf{A}_i \tag{12}$$

where $\boldsymbol{\alpha}_v = [\alpha_{v1}, \ldots, \alpha_{vn}]^T$ is the ground-truth shape parameter of a vth projected image and n is the dimension of the parameter. Note that the ground-truth 2D facial shapes include not only the frontal facial shape but also the rotated facial shapes. In addition, these shapes are the projected version of the 3D facial shapes. Therefore, a rotated 2D facial shape which is free from self-occlusion errors can be obtained by adjusting the ground-truth shape parameter $\boldsymbol{\alpha}_v$. Likewise, the observed facial shape $\mathbf{O}_v^{2D}$ can be represented by the observed shape model that is the linear combination of the mean $\mathbf{B}_0$ and the shape variation $\mathbf{B}_i$ obtained by using the training observed 2D facial shapes and the PCA as follows:

$$\mathbf{O}_v^{2D} = \mathbf{B}_0 + \sum_{i=1}^{m} \beta_{vi}\mathbf{B}_i \tag{13}$$

where $\boldsymbol{\beta}_v = [\beta_{v1}, \ldots, \beta_{vm}]^T$ is the observed shape parameter of a vth projected image and m is the dimension of the parameter. The training observed facial shapes are obtained from the projected face images. Therefore, these shapes include the correspondence errors caused by self-occlusion. Consequently, a rotated 2D facial shape free from self-occlusion errors cannot be obtained by adjusting the observed shape parameter $\boldsymbol{\beta}_v$. Figs. 7 and 8 show the mean facial shape, the first shape and the second shape variations of the ground-truth shape model and those of the observed shape model, respectively. Note that the first shape variations of both shape models represent the head rotation and the second shape variations of both shape models representing the facial width variations.

Finally, a SCM was found to transform an observed shape parameter to the corresponding ground-truth shape parameter. In order to obtain the relationship between the observed shape parameters and the ground-truth shape parameters, the distribution of the first and second components of both shape parameters are plotted, as shown in Fig. 9.

As shown in Fig. 9, the dominant components of the observed shape parameters were linearly correlated with those of the ground-truth shape parameters. Therefore, we can approximate the linear relationships between the observed and the ground-truth shape parameters. These linear relationships can now be defined by the SCM $\mathbf{C}$, which converts an observed shape parameter to the corresponding ground-truth shape parameter by

$$\boldsymbol{\alpha}_v = \mathbf{C}\boldsymbol{\beta}_v \tag{14}$$

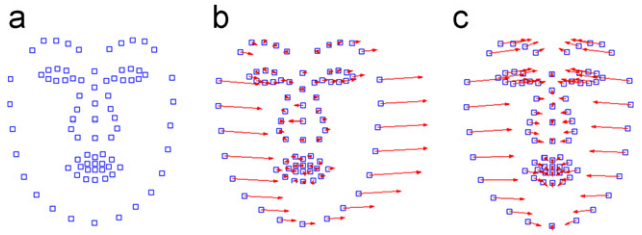

Fig. 7. Ground-truth shape model: (a) the mean face ($\mathbf{A}_0$), (b) the first shape variation ($\mathbf{A}_1$), and (c) the second shape variation ($\mathbf{A}_2$). Arrows represent the direction of the variation as a corresponding shape parameter varies toward a positive direction.

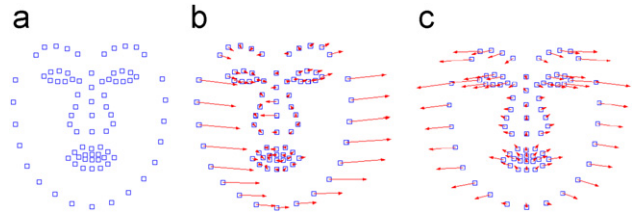

Fig. 8. Observed shape model: (a) the mean face ($\mathbf{B}_0$), (b) the first shape variation ($\mathbf{B}_1$), and (c) the second shape variation ($\mathbf{B}_2$). Arrows represent the direction of the variation as a corresponding shape parameter varies toward a positive direction.

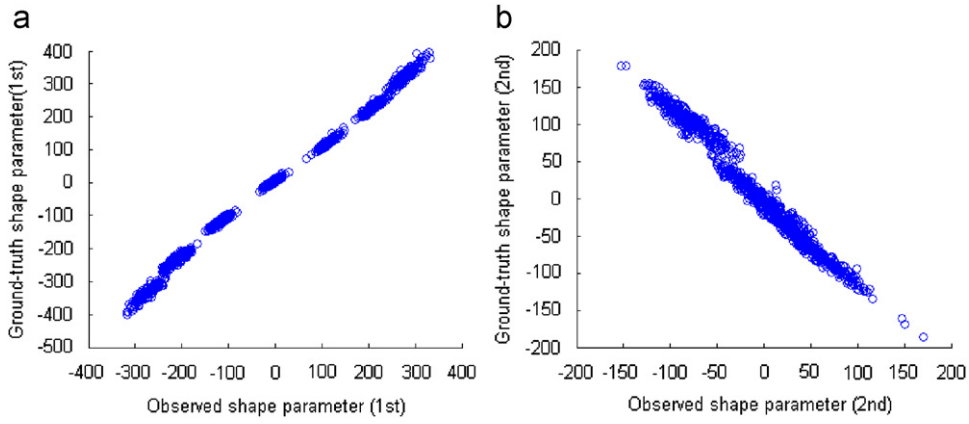

Fig. 9. Distribution of first and second components of the observed shape parameters and the ground-truth shape parameters: (a) the first component and (b) the second component.

Table 2
Comparison of the proposed method to previous methods.

	Previous SfM methods	Matrix completion+previous SfM methods	Proposed method (SCM+previous SfM methods)
Basic Idea	• Use only visible FFPs to reconstruct the projection matrix and 3D facial shape • Do not use statistical shape model	• Use visible FFPs and nuclear norm minimization to estimate the expected locations of self-occluded FFPs. • Use all FFPs to reconstruct the projection matrix and 3D facial shape • Do not use statistical shape model	• Uses statistical shape model to estimate the expected locations of self-occluded FFPs • Uses all FFPs to reconstruct the projection matrix and 3D facial shape
Strength	• If there are very small point observation errors, accurate 3D facial shape can be obtained • Do not require training data	• If there are very small point observation errors, accurate 3D facial shape can be obtained • Do not require training data	• Relatively robust to point observation errors by using all FFPs and statistical shape model • Does not need occlusion detectors • Fast processing time
Weakness	• Sensitive to observation errors because of reduced FFPs • Require occlusion detectors and the reconstruction performance is dependent on them • High computational cost caused by many iteration	• Sensitive to observation errors because of reduced FFPs • Require occlusion detectors and the reconstruction performance is dependent on them • High computational cost caused by many iteration	• Needs 3D face scans and manual annotation to obtain training data

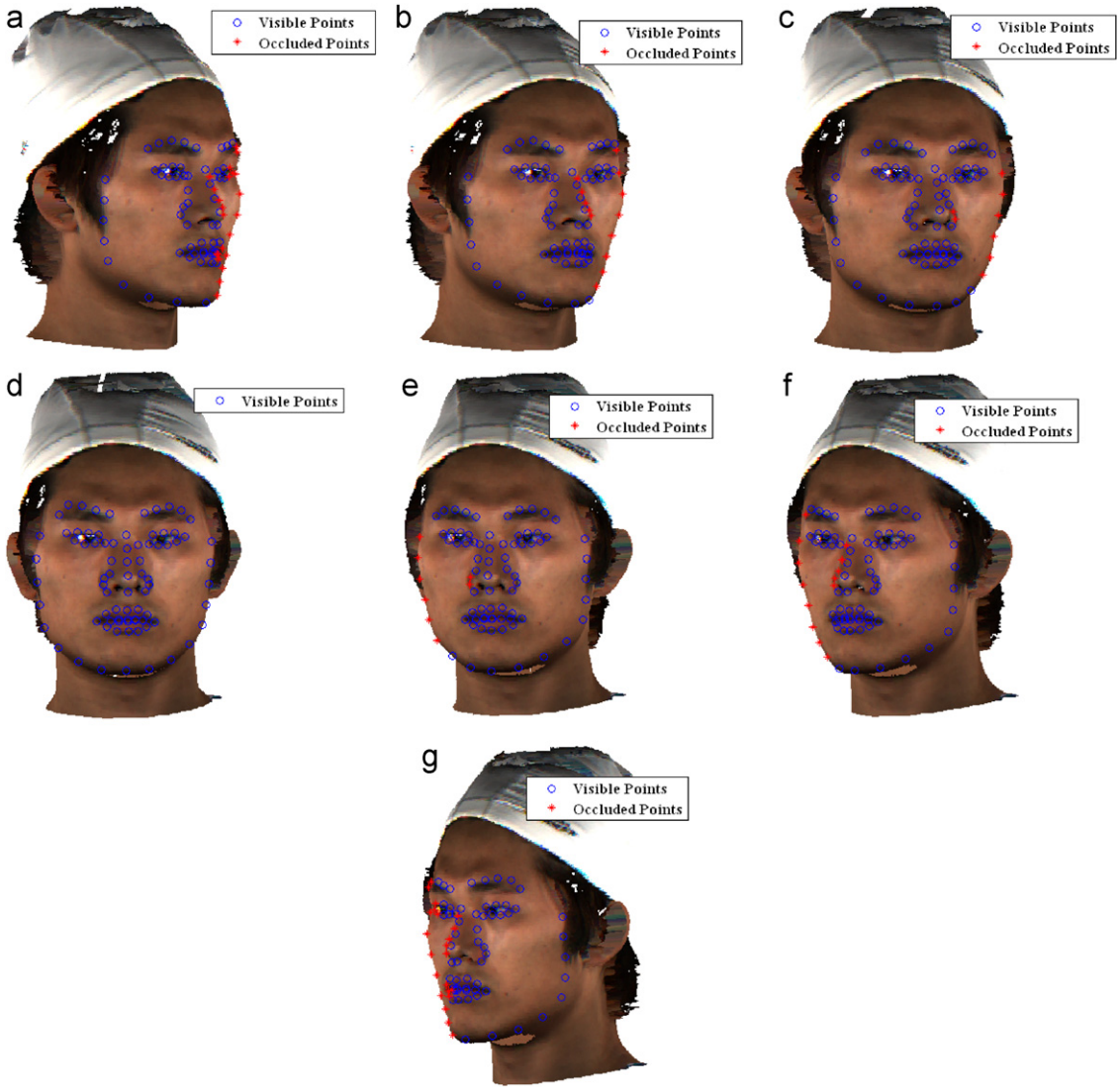

Fig. 10. Visible and self-occluded FFPs in the projected 2D facial images: (a) at $-45°$, (b) at $-30°$, (c) at $-15°$, (d) at $0°$, (e) at $15°$, (f) at $30°$, and (g) at $45°$.

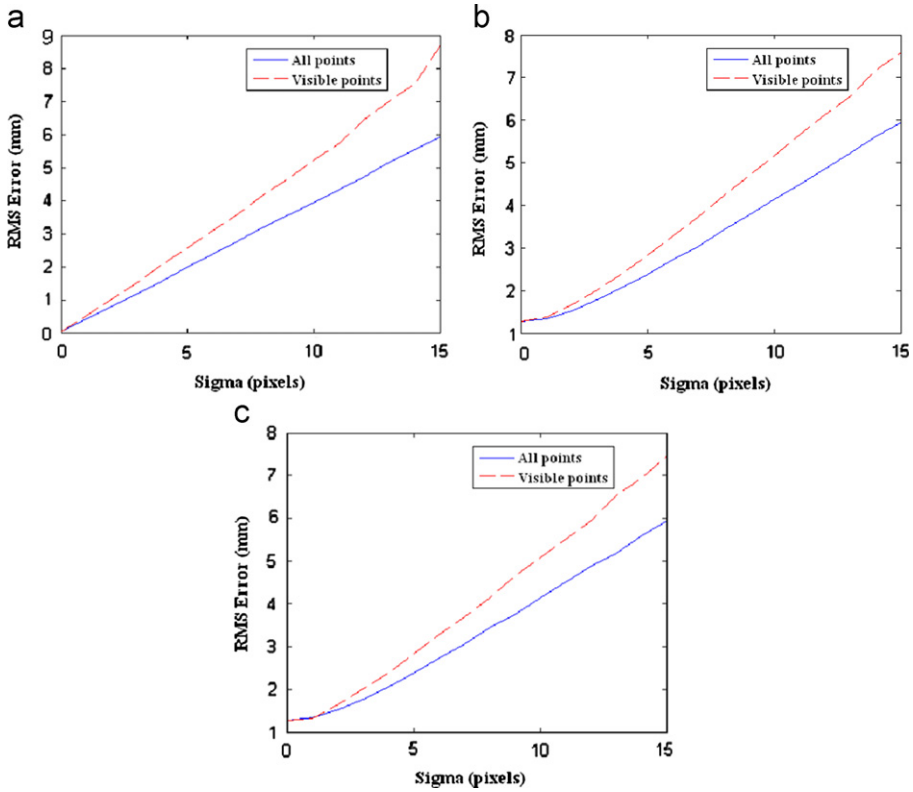

Fig. 11. RMS error between the reconstructed 3D shape and the ground-truth 3D shape when all 2D FFPs and visible 2D FFPs were used with: (a) SfM method 1 [18], (b) SfM method 2 [19], and (c) SfM method 3 [20].

The SCM $\mathbf{C}$ can then be found by using the least square method:

$$\mathbf{C} = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_V]([\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_V]^T[\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_V])^{-1}[\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_V]^T \quad (15)$$

where V is the total number of training projected images. For example, if we have 150 3D face scans and 7 rotated images per each 3D face scan, V is $150 \times 7 = 1050$.

3.3. 3D reconstruction using the SCM (testing)

In order to reconstruct a 3D facial shape, the 2D observed facial shapes obtained from multiple 2D face images are needed. Let a 2D observed facial shape be $\mathbf{O}^{2D}$, then $\mathbf{O}^{2D}$ is projected to the observed shape model in order to find the observed shape parameter $\boldsymbol{\beta}$. $\boldsymbol{\beta}$ is then converted to shape parameter $\boldsymbol{\alpha}$ by multiplying the SCM, as in (14). From the converted shape parameter $\boldsymbol{\alpha}$, a 2D converted facial shape is reconstructed by using (12). For multiple images, these procedures are repeated to find the 2D converted facial shapes. Finally, all SfM methods in Section 2.2 are used to reconstruct the 3D facial shape.

The proposed method estimates the true locations of the self-occluded points by using the SCM. As a result, all components of masking matrix $\mathbf{M}$ are filled with 1 and the objective function is changed from (6) to (16)

$$(\hat{\mathbf{P}},\hat{\mathbf{S}}) = \text{argmin}_{\mathbf{P},\mathbf{S}} ||(\tilde{\mathbf{W}} - \mathbf{PS})||_F \quad (16)$$

Therefore, the proposed method does not need to find $\mathbf{M}$.

In addition, the proposed method can obtain an optimal initial $\hat{\mathbf{S}}$ and $\hat{\mathbf{P}}$ for (16) without requiring a high computational cost by using the closed-form algorithm (the factorization methods of [18,26]) because all 2D FFPs are available. Furthermore, unlike the matrix completion method, the self-occluded FFPs can be estimated by using simple matrix multiplications. Therefore, the proposed method can estimate the 3D facial shape faster than previous methods. Table 2 shows a comparison between the proposed method and previous methods.

4. Experimental results

4.1. Experimental environment

In order to find the SCM, 150 3D face scans from 150 subjects were obtained by using a 3D scanner [30]. The age of subjects ranges from 10s to 60s. Therefore, the database includes the effect of age variations. These 3D face scans consisted of over 0.1 million vertices, meshes, and texture coordinates. From these 3D face scans, 80 3D FFPs were annotated manually and these points were used as the 3D ground-truth facial shape. The ground-truth 2D facial shapes were obtained for 7 different views; $-45°$, $-30°$, $-15°$, $0°$, $15°$, $30°$, and $45°$. Therefore, the total number of the ground-truth 2D facial shapes was 150 subjects $\times$ 7 views $= 1050$. The resolution of the projected 2D facial images was 1200×900 and the average facial width of the projected 2D facial images was approximately 300 pixels. From the projected 2D images, two different types of the observed 2D facial shapes were obtained by manual annotation and automatic annotation using an active appearance model (AAM) [22–24]. The AAM used in this paper is based on the simultaneous inverse compositional algorithm, because this algorithm showed a relatively good performance for unseen data [23,32]. The SCM was built as described in Section 3.2 by using the 2D ground-truth and the manually observed facial shapes. The dimensions of both the ground-truth and the observed shape parameters were determined to represent 99.9% of the shape variation [23,24]. The training and the testing facial shapes were divided using the leave-one-out method. In other words, the facial shapes of a subject were used for testing the 3D reconstruction performance and the facial shapes of the other 149 subjects were used for training the SCM and this procedure was repeated for all subjects.

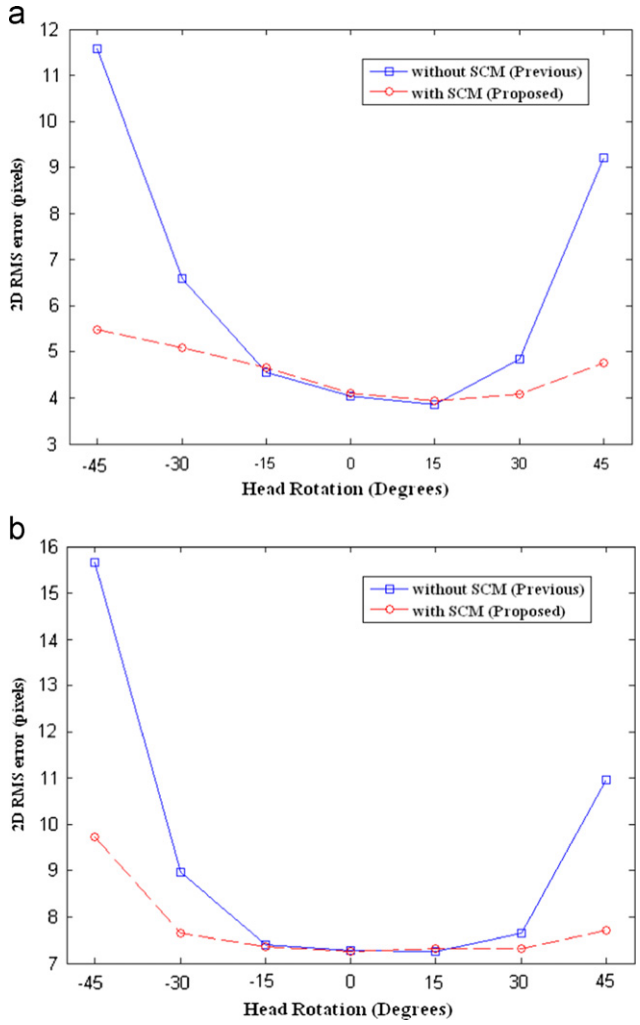

Fig. 12. 2D RMS errors according to head rotation: (a) manually annotated FFPs and (b) automatically annotated FFPs.

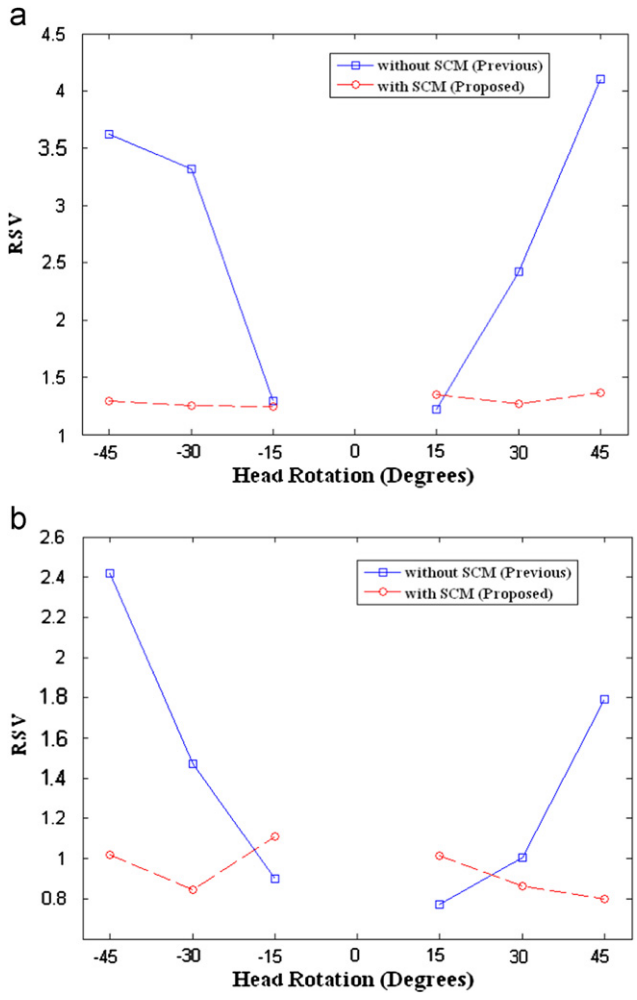

Fig. 13. RSV values according to head rotation: (a) manually annotated FFPs and (b) automatically annotated FFPs. Note that there is no RSV value at 0° because there are no self-occluded points at 0°.

In addition, the previous SfM methods in Section 2.2 require the masking matrix **M**. This matrix was defined by manually discriminating the self-occluded FFPs from the visible FFPs according to the head rotation angles. Fig. 10 shows the manually discriminated visible and self-occluded FFPs in the projected 2D facial images, according to head rotation angles.

4.2. Noise sensitivity test using ideal data

In order to compare the noise sensitivity of the SfM methods using all FFPs to that of SfM methods using only visible FFPs, 3D RMS errors were measured when the equivalent Gaussian noise was added to each ground-truth 2D FFP. SfM methods [18–20] were used to reconstruct the 3D FFPs and the standard deviation of Gaussian noise was increased from 0 pixels to 15 pixels. The results are shown in Fig. 11. From Fig. 11, it is clear that SfM methods using all 2D FFPs are relatively insensitive to Gaussian noise compared to those using only the visible part of the 2D FFPs. Therefore, if we can estimate the self-occluded FFPs with accuracy comparable to the visible FFPs, using all FFPs can make the SfM methods more robust against point correspondence errors.

4.3. Performance of estimating self-occluded facial feature points using SCM

In the second experiment, the performance of estimating the locations of the self-occluded 2D FFPs by using the SCM was evaluated quantitatively. To achieve this goal, the 2D RMS errors between the ground-truth 2D FFPs and the observed 2D FFPs without shape conversion were compared with the 2D RMS errors between the ground-truth 2D FFPs and the converted 2D FFPs. The observed 2D FFPs were obtained through manual annotation and automatic annotation using AAM. These 2D FFPs were obtained from the projected 2D images having 7 different views, as shown in Fig. 10. In addition, the ratio of the 2D RMS errors at the self-occluded 2D FFPs to the 2D RMS errors at the visible 2D FFPs that are located in a bilateral position to the self-occluded 2D FFPs (RSV) was measured in order to verify that the SCM can reduce the 2D RMS errors at the self-occluded 2D FFPs comparably to those at the visible 2D FFPs. All visible FFPs were not used to measure the RSV because the RMS error depends on each FFP. For example, the 2D RMS errors of the FFPs at the facial contour are usually higher than those at the eyes because the facial contour does not have a salient texture pattern. In short, the RSV can be written as

$$\mathrm{RSV} = \frac{\text{2D RMS error at self-occluded FFPs}}{\text{2D RMS error at visible FFPs located in a bilateral position to the self-occluded FFPs}}$$

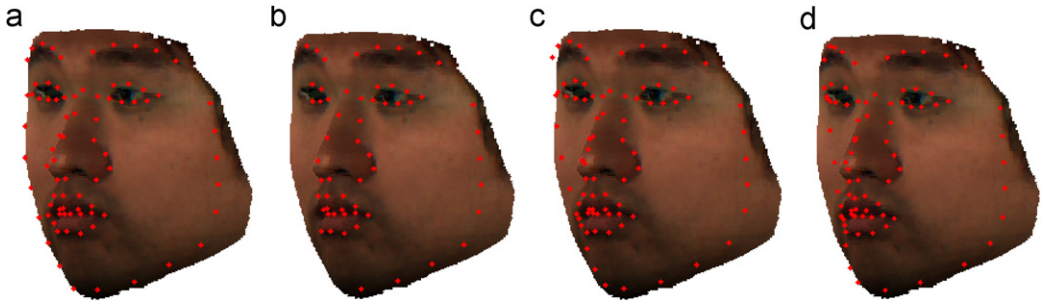

Fig. 14. Four kinds of input FFPs: (a) input FFPs1 (the observed 2D FFPs), (b) input FFPs2 (the visible part of the observed 2D FFPs), (c) input FFPs3 (all the FFPs estimated by the matrix completion method), and (d) input FFPs4 (the converted 2D FFPs using the SCM).

Table 3
Average 3D RMS errors calculated with the manually annotated FFPs.

Reconstruction method	Input feature points			
	Input FFPs1 (all points without SCM [15,16]) (mm)	Input FFPs2 (visible points [18–20]) (mm)	Input FFPs3 (all points with matrix completion [35,36]) (mm)	Input FFPs4 (all points with SCM-proposed) (mm)
SfM Method1[18]	2.47	1.88	1.99	1.65
SfM Method2 [19]	2.77	2.18	2.20	1.90
SfM Method3 [20]	2.77	2.27	2.20	1.90
AVG.	2.67	2.11	2.13	1.82

Table 4
Average processing time calculated with the manually annotated FFPs.

Reconstruction method	Input feature points			
	Input FFPs1 (all points without SCM [15,16]) (ms)	Input FFPs2 (visible points [18–20]) (ms)	Input FFPs3 (all points with matrix completion [35,36]) (ms)	Input FFPs4 (all points with SCM-proposed) (ms)
SfM Method1 [18]	18	1793	2990	24
SfM Method2 [19]	4	125	2964	16
SfM Method3 [20]	7	5815	2977	18

Table 5
Average 3D RMS errors calculated with the automatically annotated FFPs.

Reconstruction method	Input feature points			
	Input FFPs1 (all points without SCM [15,16]) (mm)	Input FFPs2 (visible points [18–20]) (mm)	Input FFPs3 (all points with matrix completion [35,36]) (mm)	Input FFPs4 (all points with SCM-proposed) (mm)
SfM Method1 [18]	4.37	3.93	3.97	3.60
SfM Method2 [19]	4.11	3.64	3.65	3.24
SfM Method3 [20]	4.11	3.67	3.64	3.24
AVG.	4.20	3.75	3.75	3.36

Table 6
Average processing time calculated with automatically annotated FFPs.

Reconstruction method	Input feature points			
	Input FFPs1 (all points without SCM [15,16]) (ms)	Input FFPs2 (visible points [18–20]) (ms)	Input FFPs3 (all points with matrix completion [35,36]) (ms)	Input FFPs4 (all points with SCM-proposed) (ms)
SfM Method1 [18]	27	2375	2979	33
SfM Method2 [19]	4	124	2987	16
SfM Method3 [20]	6	4487	2965	19

Fig. 12(a) and (b) show the average 2D RMS error for each view when using the manually annotated 2D FFPs and the automatically annotated 2D FFPs, respectively. Fig. 13(a) and (b) show the average RSV for each view when using the manually annotated 2D FFPs and the automatically annotated 2D FFPs, respectively.

From Figs. 12 and 13, it is clear that (1) the 2D RMS errors in highly rotated facial images (-45°, -30°, 30°, and 45°) were reduced when the shape converted 2D FFPs were used no matter what types of the observed 2D FFPs were used, (2) the RSV values in highly rotated facial images were significantly reduced when the shape converted 2D FFPs were used no matter what types of the observed 2D FFPs were used, (3) the 2D RMS errors and RSV values of both 2D FFPs in the slightly rotated facial images (-15°, 0°, and 15°) were similar because the amount of self-occlusion is very low in these images and the 2D RMS errors at the visible FFPs of the automatically annotated FFPs increased in these images because there are no salient texture patterns in the facial contour. As a result, the RSV values without the SCM in -15° and 15° were under 1, which means the visible FFPs have larger RMS errors than the self-occluded FFPs. The SCM transforms the observed FFPs to the converted FFPs relatively close to the ground-truth FFPs. Therefore, in the -15° and 15°, the SCM reduce the RMS errors at the visible FFPs which have larger RMS errors than the self-occluded FFPs. Consequently, the RSV values with the SCM were larger than those without the SCM at these angles. (4) The RSV values were close to 1 when the SCM was used. In other words, the expected locations of the self-occluded 2D FFPs were found with accuracy comparable to the visible 2D FFPs. Therefore, using both visible and estimated self-occluded FFPs was helpful to reconstruct the 3D facial shape reliably, because the SfM showed robustness against the correspondence errors when all the FFPs were used, as explained in Section 4.2.

4.4. Quantitative 3D reconstruction results

In the third experiment, we evaluated the accuracy and processing time of the previous SfM-based 3D reconstruction methods, the previous SfM-based 3D reconstruction methods with the matrix

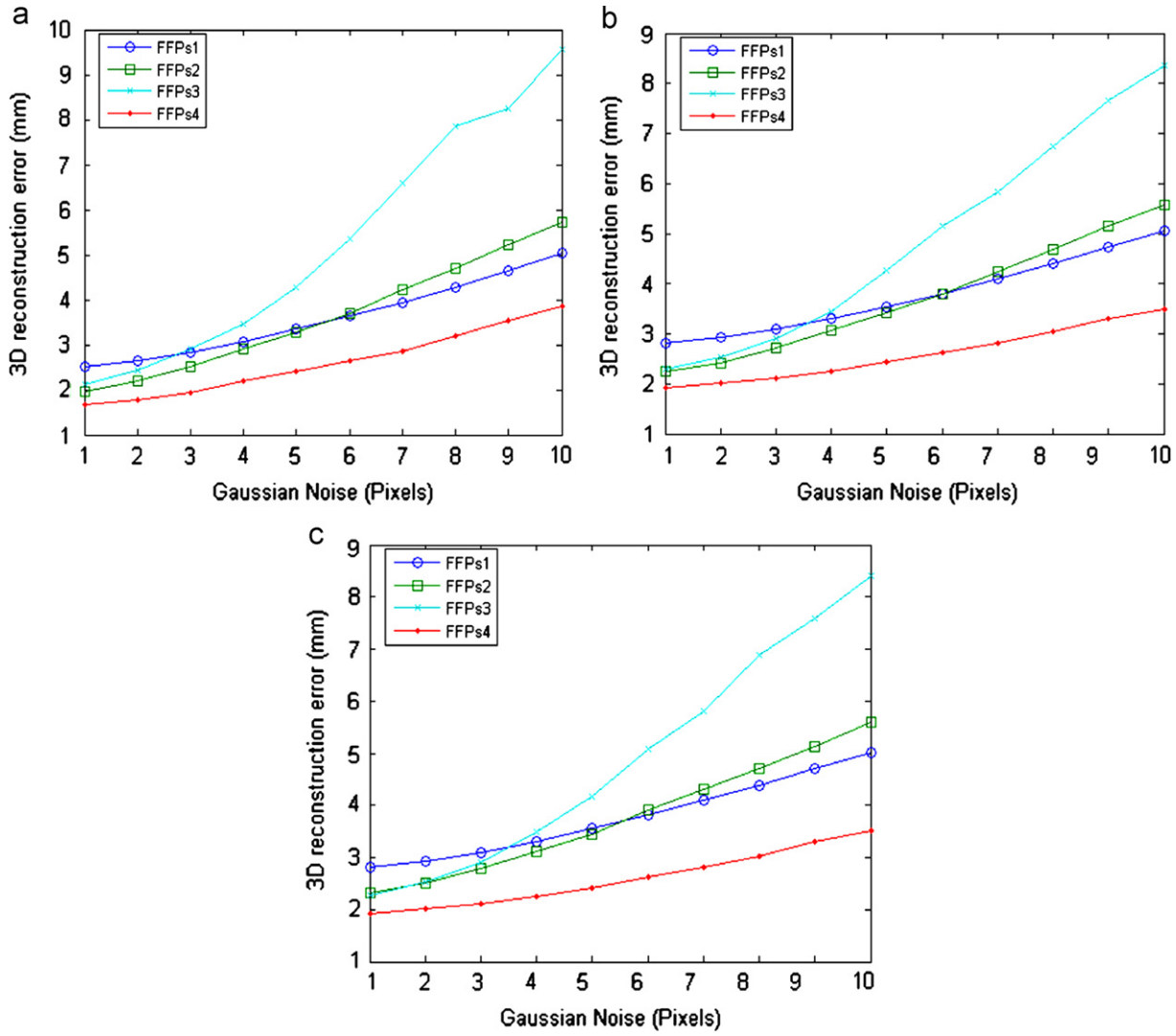

Fig. 15. Average 3D RMS errors with manually annotated 2D FFPs corrupted with increasing Gaussian noise: (a) with SfM 1, (b) with SfM 2, and (c) with SfM 3.

completion method, and the previous SfM-based 3D reconstruction methods with the proposed SCM. From the observed 2D FFPs, four kinds of input FFPs were found to reconstruct the 3D FFPs. The first input FFPs (FFPs1) is all of the FFPs without shape conversion, as shown in Fig. 14(a) [15,16]. These input points were obtained by manual annotation and automatic annotation using AAM. The second input FFPs (FFPs2) is the visible FFPs as shown in Fig. 14(b) [18–20]. The third input FFPs (FFPs3) is all the FFPs using the matrix completion method as shown in Fig. 14(c). The fourth input FFPs (FFPs4), the one proposed in this paper, is all the FFPs using the SCM, as shown in Fig. 14(d). SfM methods 1 [18], 2 [19], and 3 [20] were used for reconstruction of the 3D FFPs. The accuracy of the reconstructed 3D FFPs was measured by the 3D RMS errors. To measure the 3D RMS errors in an mm scale, the reconstructed 3D FFPs were aligned to the ground-truth 3D FFPs by using a similarity transformation that consists of scaling, rotation, and translation [31]. Note that the similarity transformation does not change the intrinsic structure of the reconstructed 3D FFPs. The 3D RMS errors were then measured between the aligned reconstructed 3D FFPs and the ground-truth 3D FFPs.

Table 3 shows the average 3D RMS errors for each of the input points and the reconstruction methods when manually annotated 2D FFPs were used. From Table 3, it is clear that (1) self-occlusion is a cruel problem in the 3D reconstruction using SfM because all the SfM methods showed larger 3D RMS errors with input FFPs1 than with input FFPs2, input FFPs3, and FFPs4, (2) the proposed method outperforms the previous methods because all the SfM methods showed a better accuracy when they used the proposed input FFPs4 than with the previously used input FFPs2 and FFPs3. Note that the previous methods (FFPs2 and 3) require occlusion detectors to determine the visible FFPs. In this paper, the visible FFPs were detected manually so that there were no errors in finding the visible FFPs. However, in a practical case, the visible FFP detection may have some errors. As a result, the 3D RMS errors using the input FFPs2 and 3 can be increased in a practical case.

Table 4 shows the processing time needed for each of the input points and the reconstruction methods when the manually annotated 2D FFPs were used with an Intel core quad CPU 2.33 GHz with 3 GB RAM. From Table 4, it is clear that the processing time of SfMs can be reduced significantly when all of the FFPs are used because the processing time with input FFPs1 and FFPs4 were far shorter than that with input FFPs2. This result occurred since the closed-form algorithms (the factorization methods [18,26]) give an optimal initial facial shape when all the 2D FFPs are available. Consequently, the SfM methods [18–20] do not require many iterations in this case. In addition, the processing time to estimate the self-occluded FFPs by using the matrix completion method was

Fig. 16. Qualitative 3D reconstruction results using the manually annotated FFPs; the first to fifth columns show the frontal view of the ground-truth, reconstructed 3D face using input FFPs1, input FFPs2, input FFPs3, and input FFPs4 (proposed). The last five columns show their profile views. The red lines show the ground-truth facial shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

far greater than those by using the SCM because the SCM do not require iterations but require only simple matrix multiplications. As a result, the processing time with input FFPs3 was far greater than those with input FFPs1 and 4, although input FFPs3 consists of all the FFPs. SfM method 3 took much processing time as this method calculates the Hessian matrix at every iteration [19].

Tables 5 and 6 show the average 3D RMS errors and the processing time for each of the input points and the reconstruction methods when the automatically annotated 2D FFPs were used. As can be seen in Tables 5 and 6, the average 3D RMS errors with the automatically annotated FFPs were larger than those with the manually annotated FFPs because the automatically annotated FFPs contained more 2D point correspondence errors. However, all of the SfM methods found more accurate 3D FFPs with the proposed input FFPs4 than with the previously used input FFPs, and the processing time was reduced when the input FFPs1 and 4 were used. SfM method 1 shows the best performance in Table 3 but shows the worst performance in Table 5. The main difference between SfM method 1 and SfM methods 2 and 3 is that only SfM method 1 used an orthogonal constraint to find the projection matrix. This constraint reduces the 3D RMS errors when the 2D correspondence errors are small but as the 2D correspondence errors increase, the constraint increases the

Fig. 17. Qualitative 3D reconstruction results using the automatically annotated FFPs; the first to fifth columns show the frontal view of the ground-truth, reconstructed 3D face using input FFPs1, input FFPs2, input FFPs3, and input FFPs4 (proposed). The last five columns show their profile views. The red lines show the ground-truth facial shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3D RMS errors. Fig. 11 also support this statement. In Fig. 11, SfM method 1 showed the best performance when the Gaussian noise was small, but its performance became more degraded as the Gaussian noise increased as compared to the other SfM methods.

Current automatic FFPs detectors are not always perfect if there are variations caused by lighting, expression, aging, and so on. In order to evaluate noise robustness, we measured the average 3D RMS error with manually annotated 2D FFPs corrupted with increasing Gaussian noise whose standard deviations range from 1 to 10 pixels. According to a previous work on an automatic FFP detector, active appearance models (AAMs), the 2D RMS errors of AAMs were reported to be from 2.5% to 3% of the facial width when there were severe facial variations caused by expression and illumination changes. In our database, the average of the facial width is about 300 pixels. Therefore, we set the range of the standard deviations from 1 to 10 pixels. The average 3D RMS error with the noisy 2D FFPs is shown in Fig. 15. As shown in Fig. 15, the proposed method (FFPs4) is more robust to 2D FFPs localization errors than previous methods (FFPs1, 2, and 3) when three different structure-from-motion SfM algorithms were used. Note that the methods based on the minimization of the visible shape residual

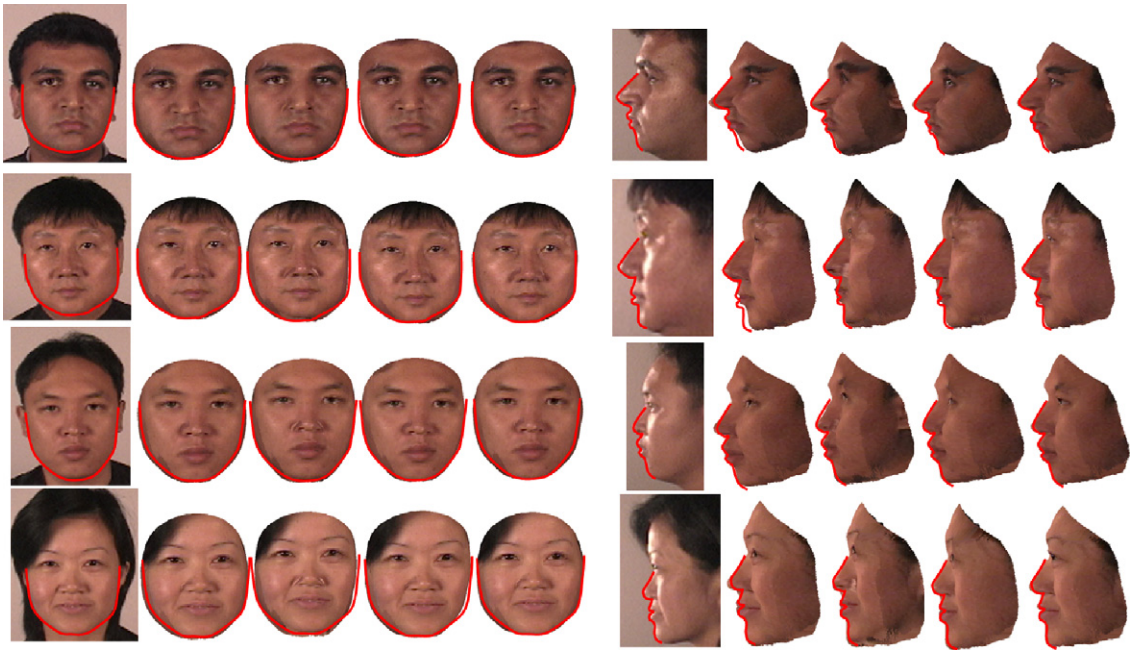

Fig. 18. Qualitative 3D reconstruction results using the FacePix database and the manually annotated FFPs; the first to fifth columns show the frontal view of the ground-truth, reconstructed 3D face using input FFPs1, input FFPs2, input FFPs3, and input FFPs4 (proposed). The last five columns show their profile views. The red lines show the ground-truth facial shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

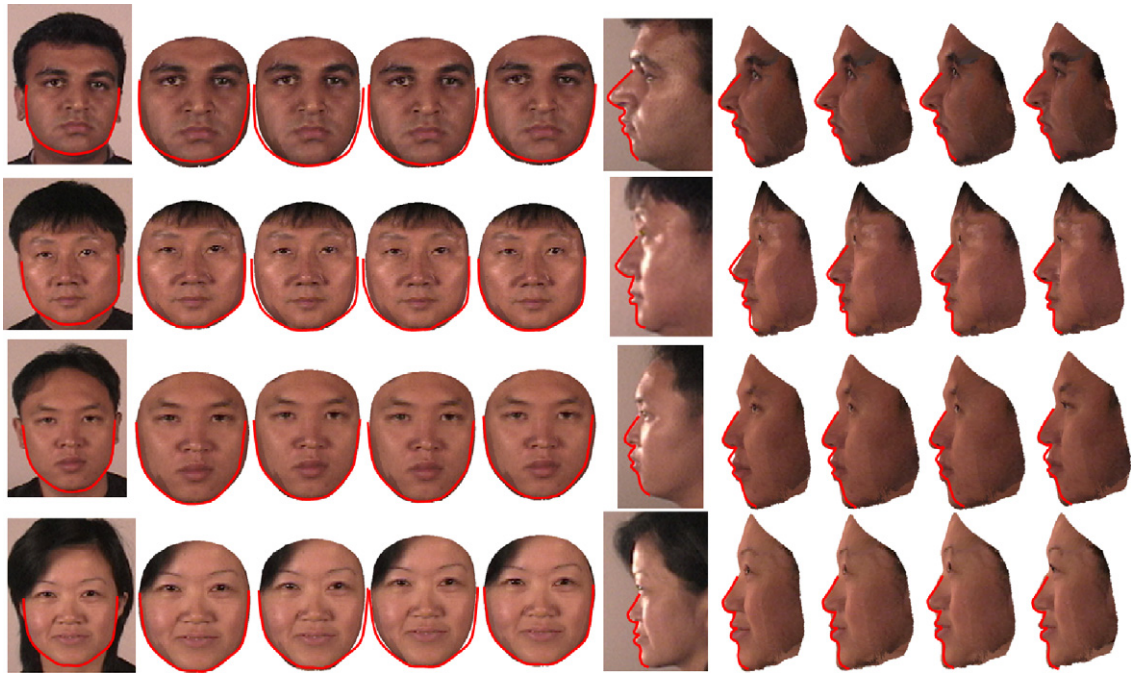

Fig. 19. Qualitative 3D reconstruction results using the FacePix database and the automatically annotated FFPs; the first to fifth columns show the frontal view of the ground-truth, reconstructed 3D face using input FFPs1, input FFPs2, input FFPs3, and input FFPs4 (proposed). The last five columns show their profile views. The red lines show the ground-truth facial shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(FFPs2 and FFPs3) were more sensitive to the Gaussian noise in 2D FPPs than the proposed SCM (FFPs4). In addition, the FFPs4 outperformed the FFPs1 because the correspondence errors caused by self-occlusion in the FFPs1 were reduced by the SCM in the FFPs4.

In short, it is clear that all SfM methods found more accurate 3D FFPs with the proposed input FFPs4 than with previously used input FFPs, regardless of whether manual or automatic FFP annotations were used. The proposed method was more robust to FFPs localization errors than previous methods and the processing time of the SfM methods was significantly reduced when all the 2D FFPs were used instead of only the visible 2D FFPs. Therefore, by using the proposed input FFPs4, the SfM methods can reduce both the 3D RMS errors and the processing time.

4.5. Qualitative 3D reconstruction results

Fig. 16 shows the qualitative 3D reconstruction results using the manually annotated FFPs with SfM method 1. Fig. 17 shows the results using the automatically annotated FFPs and SfM method 2. SfM method 1 was used for the manually annotated FFPs because it shows the best performance in terms of the RMS errors. SfM method 2 was used for the automatically annotated FFPs for the same reason. From Figs. 16 and 17, it is clear that (1) the reconstructed results with input FFPs1 were the worst among the four input FFPs because the self-occlusion problem is not solved in this FFPs. Specifically, the reconstructed facial contour was rather wide compared to the ground-truth (refer to the first row of Fig. 16), the facial profile shapes had large differences compared to the ground-truth, and the eye regions were flatter than the ground-truth. (2) The reconstructed results using input FFPs2 and 3 were worse than those with the proposed input FFPs4. The reconstructed facial profile shapes using input FFPs2 and 3 were improved compared with those with input FFPs1 but the results had larger errors than the reconstructed facial profile shapes with the proposed input FFPs4. In addition, the reconstructed eye regions with input FFPs2 and 3 were flatter than the ground-truth. This problem was solved by using the proposed input FFPs4. Therefore, more realistic and exact 3D reconstructed faces were obtained by using the proposed FFPs4. (3) The correspondence errors degrade the performance of the SfM methods because the reconstructed 3D faces with the manually annotated FFPs were closer to the ground-truth faces than those with the automatically annotated FFPs. For instance, the reconstructed sixth subject in Fig. 16 was relatively close to the ground-truth but it showed a far wider facial width in Fig. 17.

One common limitation of sparse correspondence-based 3D face reconstruction methods is that they cannot reconstruct a detailed 3D shape of a facial region that contains no FFPs. For example, these methods cannot reconstruct the detailed shapes of cheeks because there are no FFPs in the cheek area (refer to the sixth row of Fig. 16).

The qualitative 3D reconstruction results in real face images are shown in Figs. 18 and 19. Fig. 18 shows the reconstructed results using the manually annotated FFPs with SfM method 1, and Fig. 19 shows those using the automatically annotated FFPs with SfM method 2. The FacePix(30) database was used, which includes face images taken from 30 subjects, 181 pose angles, and 181 illumination angles [33,34]. From this database, face images taken from seven different views (-45°, -30°, -15°, 0°, 15°, 30°, and 45°) were used for the 3D face reconstruction. From Figs. 18 and 19, it was found that the reconstructed facial shapes using the proposed input FFPs4 are also more accurate in the real image database than the other two methods.

4.6. Experimental results with various expressions

Finally, an additional database including 3 different facial expressions (neutral, smile, surprise, as shown in Fig. 20) from 20 subjects was collected, and the average 3D RMS errors were measured in order to evaluate the performance of the proposed method and the previous methods under facial expressions. The SCMs were made by using the leave-one-out methodology. Tables 7 and 8 show the average 3D RMS errors when manually annotated FFPs and automatically annotated FFPs were used, respectively. From the tables, we can find that the proposed method (FFPs4) showed the lowest average 3D RMS error among the four different FFPs. By comparing the average 3D RMS errors of three expressions in Table 8, we can see that the neutral case (no expression) (4.36 mm) was larger than smile (3.95 mm), but it is almost the same as the surprise case (4.48 mm). Therefore, the proposed method can be used with different facial expressions.

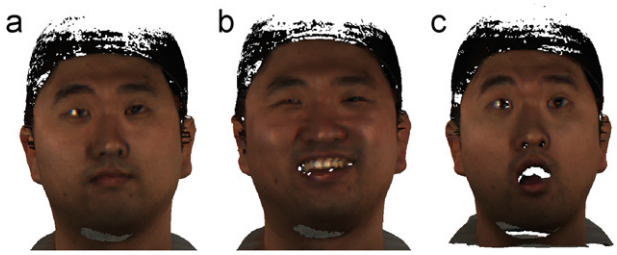

Fig. 20. Projected images with different expressions: (a) neutral, (b) smile, (c) surprise.

Table 7
Average 3D RMS errors under different expressions when manually annotated FFPs were used.

	FFPs1 (mm)	FFPs2 (mm)	FFPs3 (mm)	FFPs4 (mm)
Expression1 (Neutral)				
SfM1	2.56	2.25	2.39	2.18
SfM2	2.93	2.48	2.59	2.34
SfM3	2.93	2.70	2.56	2.34
AVG.	2.81	2.48	2.51	2.29
Expression2 (Smile)				
SfM1	3.02	2.55	2.61	2.26
SfM2	3.15	2.72	2.70	2.43
SfM3	3.15	2.76	2.69	2.43
AVG.	3.11	2.68	2.67	2.37
Expression3 (Surprise)				
SfM1	2.80	2.45	2.56	2.34
SfM2	3.58	3.15	3.24	2.95
SfM3	3.58	3.38	3.23	2.95
AVG.	3.32	2.99	3.01	2.75

Table 8
Average 3D RMS Errors under different expressions when automatically annotated FFPs were used.

	FFPs1 (mm)	FFPs2 (mm)	FFPs3 (mm)	FFPs4 (mm)
Expression1 (Neutral)				
SfM1	5.83	5.94	5.76	5.05
SfM2	4.75	4.41	4.51	4.01
SfM3	4.75	4.61	4.55	4.01
AVG.	5.11	4.99	4.94	4.36
Expression2 (Smile)				
SfM1	4.80	4.37	4.35	4.22
SfM2	4.44	4.07	4.10	3.81
SfM3	4.44	4.21	4.08	3.81
AVG.	4.56	4.22	4.18	3.95
Expression3 (Surprise)				
SfM1	6.24	5.66	5.12	5.03
SfM2	4.96	4.71	4.77	4.20
SfM3	4.96	4.90	4.74	4.20
AVG.	5.39	5.09	4.88	4.48

5. Conclusions

In this paper, a SfM-based 3D face reconstruction method robust to self-occlusion is proposed. The proposed method has three novelties over the previous methods. First, we propose the SCM that can estimate the true locations of the self-occluded FFPs with accuracy comparable to the visible FFPs and with simple matrix

multiplications. Second, by using the SCM, both the visible and the self-occluded FFPs were used to reconstruct the 3D facial shape. Consequently, the proposed 3D face reconstruction method becomes more robust against correspondence errors and finds a better initial 3D facial shape than SfM methods not using the SCM. Finally, the proposed method does not require an occlusion detection method to find the visible FFPs because the expected locations of the self-occluded FFPs are estimated from all the observed FFPs by using the SCM. By quantitative and qualitative experimental results, it was clear that SfM-based 3D face reconstruction methods with the proposed SCM show a higher accuracy and a faster convergence time than previous SfM-based 3D face reconstruction methods without the SCM. The average 3D RMS error between the reconstructed 3D face and the ground-truth 3D face obtained from the 3D scanner was less than 2 mm when the 2D FFPs were annotated manually and it was less 3.5 mm when the 2D FFPs were annotated automatically.

In this paper, multiple 2D facial images including the frontal face were used to reconstruct the 3D Face. In a future work, we will research 3D face reconstruction from 2D facial images that do not contain the frontal face but contain only side view images. In this case, the self-occluded FFPs may always be invisible and the proposed SCM would be useful to alleviate this problem. In addition, research about face recognition across poses by using the proposed 3D face reconstruction method will also be done as a future work.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) through the Biometrics Engineering Research Center (BERC) at Yonsei University (No. R112002105070030(2010)).

References

[1] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (9) (2003) 1063–1074.
[2] X. Zhang, Y. Gao, Face recognition across pose: a review, Pattern Recognition 42 (2009) 2876–2896.
[3] U. Park, Y. Tong, A.K. Jain, Age invariant face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (5) (2010) 947–954.
[4] Example of game character generation [Online] available at 〈http://fightnight.easports.com/featureFrame.action?id=Feature-TBA1&fType=video〉.
[5] A. Maejima, S. Wemler, T. Machida, M. Takebayashi, S. Morishima, Instant casting movie theater: the future cast system, IEICE Transactions on Information and Systems E91-D (4) (2008) 1135–1148.
[6] B. Gökberk, A.A. Salah, N. Alyüz, L. Akarun, 3D face recognition: technology and applications, in: M. Tistarelli, S.Z. Li, R. Chellappa (Eds.), Handbook of Remote Biometrics for Surveillance and Security, Springer, 2009, pp. 217–246.
[7] V. Blanz, T. Vetter, A morphable model for the synthesis of 3D faces, in: Proceedings of SIGGRAPH 99, Los Angeles, CA, 1999, pp. 187–194.
[8] D. Fidaleo, G. Medioni, Model-assisted 3D face reconstruction from video, Lecture Notes in Computer Science 4778 (2007) 124–138.
[9] A.K. Roy Chowdhury, R. Chellappa, Face reconstruction from monocular video using uncertainty analysis and a generic model, Computer Vision and Image Understanding 91 (2003) 188–213.
[10] Y. Shan, Z. Liu, Z. Zhang, Model-based bundle adjustment with application to face modeling, in: Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV), vol. 2, 2001, pp. 644–651.
[11] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, W. Gao, Efficient 3D reconstruction for face recognition, Pattern Recognition 38 (2005) 787–798.
[12] Z. Zhang, Z. Liu, D. Adler, M.F. Cohen, E. Hanson, Y. Shan, Robust and rapid generation of animated faces from video images: a model-based modeling approach, International Journal of Computer Vision 58 (2) (2004) 93–119.
[13] F. Pighin, R. Szeliski, D.H. Salesin, Modeling and animating realistic faces from images, International Journal of Computer Vision 50 (2) (2002) 143–169.
[14] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D.H. Salesin, Synthesizing realistic facial expressions from photographs, in: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '98. ACM, New York, NY, pp. 75–84.
[15] U. Park, A.K. Jain, A. Ross, Face recognition in video: adaptive fusion of multiple matchers, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[16] U. Park, A.K. Jain, 3D model-based face recognition in video, Lecture Notes in Computer Science 4642 (2007) 1085–1094.
[17] C. Tomasi, T. Kanade, Shape and motion from image streams under orthography: a factorization method, International Journal of Computer Vision 9 (2) (1992) 137–154.
[18] M. Marques, J. Costeira, Estimating 3D shape from degenerate sequences with missing data, Computer Vision and Image Understanding 113 (2009) 261–272.
[19] T. Wiberg, Computation of principal components when data are missing, in: Proceedings of the Symposium of Computational Statistics, 1976, pp. 229–326.
[20] A.M. Buchanan, A.W. Fitzgibbon, Damped Newton algorithms for matrix factorization with missing data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 316–322.
[21] R. Szeliski, S.B. Kang, Shape ambiguities in structure from motion, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (5) (1997) 506–512.
[22] S. Baker, R. Gross, I. Matthews, Lucas-Kanade 20 years on: a unifying framework: part 3, Technical Report CMU-RI-TR-03-35, Carnegie Mellon University Robotics Institute, 2003.
[23] R. Gross, I. Matthews, S. Baker, Generic vs. person specific active appearance models, Image and Vision Computing 23 (11) (2005) 1080–1093.
[24] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 681–685.
[25] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Computer Vision and Image Understanding 61 (1) (1995) 38–59.
[26] R. Hartley, A. Zisserman, N-view computational methods, in: Multiple View Geometry in Computer Vision, Cambridge, 2003, pp. 434–457.
[27] I.D. Reid, D.W. Murray, Active tracking of foveated feature clusters using affine structure, International Journal of Computer Vision 18 (1) (1996) 41–60.
[28] F.L. Bookstein, Principal warps: thin-plate splines and the decomposition of deformations, IEEE Transactions on Pattern Analysis Machine Intelligence 11 (6) (1989) 567–585.
[29] U. Park, A.K. Jain, 3D face reconstruction from stereo video, in: Proceedings of the Third Canadian Conference on Computer and Robot Vision, 2006, p. 41.
[30] 3D Scanner specification, available at 〈http://www.cyberware.com/products/pdf/headFace.pdf〉.
[31] A.-N. Ansari, M. Abdel-Mottaleb, 3D face modeling using two views and a generic face model with application to 3D face recognition, in: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03), 2003, pp. 37–44.
[32] S.J. Lee, K.R. Park, J. Kim, A comparative study of facial appearance modeling methods for active appearance models, Pattern Recognition Letters 30 (2009) 1335–1346.
[33] J.A. Black, M. Gargesha, K. Kahol, P. Kuchi, S. Panchanathan, A framework for performance evaluation of face recognition algorithms, in: Proceedings of the SPIE Internet Multimedia Management Systems, vol. 4862, 2002, pp. 163–174.
[34] G. Little, S. Krishna, J. Black, S. Panchanathan, A methodology for evaluating robustness of face recognition algorithms with respect to changes in pose and illumination angle, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), vol. 2, 2005, pp. 89–92.
[35] E.J. Candès, Y. Plan, Matrix completion with noise, Proceedings of IEEE 98 (6) (2010) 925–936.
[36] S. Ma, D. Goldfarb, L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization, Technical Report, 2008.
[37] J. Gonzalez-Mora, F. De la Torre, N. Guil, E.L. Zapata, Learning a generic 3D face model from 2D image databases using incremental structure-from-motion, Image and Vision Computing 28 (2010) 1117–1129.
[38] R. Hartley, A. Zisserman, Camera Models, Multiple View Geometry in Computer Vision, Cambridge, 2003, pp. 153–177.

Sung Joo Lee received his B.S. degree in electrical and electronic engineering and his M.S. degree in biometric engineering in 2004 and 2006. respectively, from Yonsei University, Seoul, Korea, where he is currently working toward his Ph.D. degree in electrical and electronic engineering. His current research interests include 3D face reconstruction, biometrics, pattern recognition, and computer vision.

Kang Ryoung Park received B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1994 and 1996, respectively. He also received the Ph.D. degree in computer vision at the Department of Electrical and Computer Engineering in Yonsei University in 2000. He was an assistant professor in the division of digital media technology at Sangmyung University from March 2003 to February 2008. He has been an associate professor in the Division of Electronics and Electrical Engineering at

Dongguk University from March 2008. He has been also a research member of Biometrics Engineering Research Center (BERC). His research interests include computer vision, image processing, and biometrics.

Jaihie Kim received the B.S. degree in electronic engineering at Yonsei University, Seoul, Korea, in 1979, and the M.S. degree in data structures and the Ph.D. degree in artificial intelligence at Case Western Reserve University, Cleveland, OH, in 1982 and 1984, respectively. Since 1984, he has been a professor in the School of Electrical and Electronic Engineering, Yonsei University. He is currently the Director of the Biometric Engineering Research Center in Korea. His research areas include biometrics, computer vision and pattern recognition. Prof. Kim is currently the Chairman of Korean Biometric Association.