International Journal of Innovative
Computing, Information and Control
Volume 8, Number 1(B), January 2012

 ISSN 1349-4198
pp. 807–825

3D FACIAL SURFACE RECONSTRUCTION USING INTEGRATED ORTHOGRAPHIC MODELS TO APPROXIMATE PERSPECTIVE PROJECTION MODEL

Jin-Yi Wu and Jenn-Jier James Lien

Department of Computer Science and Information Engineering
National Cheng Kung University
No. 1, University Road, Tainan City 70101, Taiwan
{ curtis; jjlien }@csie.ncku.edu.tw

Received October 2010; revised May 2011

Abstract. *This study develops a 3D facial reconstruction system in which the perspective projection model is approximated by applying a factorization method to the piecewise orthographic projection model. The proposed system comprises five modules. The first and second modules reconstruct the 3D facial surface using a factorization method based on an orthographic projection model. However, the facial video is taken based on the perspective projection model rather than an orthographic projection model. Thus, to compensate for the difference between the two models, the third module is developed to approximate the perspective projection model by dividing the 3D facial surface into small groups and then reconstructs each group in orthographic projection module. These reconstructed results are then integrated to form a complete 3D facial surface, which is almost as accurate as the reconstruction result using a perspective projection model. The fourth module implements a novel smoothing process for the 3D facial surface by interpolating additional vertices from the vectors of the existing 3D vertices. Finally, the fifth module utilizes a new solution to overcome the missing point problem, which is caused by occlusion at high pan rotation angles, commonly arising in 3D reconstruction applications. The experimental results show that the proposed system achieves a promising result within a relatively short time.*

Keywords: 3D reconstruction, Shape from motion, Factorization, Orthographic projection, Perspective projection

1. **Introduction.** 3D structure modeling, in which 3D information is acquired from 2D data, provides the means to analyze the reconstructed 3D object under any view angle, to realize its 3D geometrical structure, and to render a more realistic and livelier view than that obtained in the 2D image plane. Accordingly, 3D modeling has widespread applications in many areas, such as medical diagnosis, robotic vision, navigation, building construction and simulation, and human-computer interaction. Of the various classes of 3D structure modeling, 3D facial reconstruction has attracted particular attention in recent years for applications such as human animation in the entertainment industry and user-friendly human-computer interaction. In [1], a method was developed for constructing 3D facial models by means of laser scanners. The reconstructed models were highly accurate. However, the computation process was time-consuming and the scanning equipment was extremely expensive. As the cost of digital cameras continues to fall, an increasing number of researchers have proposed methods for recovering 3D information from 2D photographic images using computer vision or machine learning techniques. Most existing 3D facial reconstruction algorithms are based on a single image [2] or a set of images taken from known (or unknown) camera viewpoints [3]. However, when using a

single image, insufficient information is available to permit an accurate 3D reconstruction. Similarly, when using multiple images, it is necessary to solve the problems of identifying the corresponding points in neighboring images, overcoming the missing point problem caused by occlusion, and estimating the extrinsic parameters of the camera.

In the Bundle Adjustment (BA) approach [4-7], the 3D reconstruction problem is treated as a minimization problem in which the objective is to minimize the distance between the actual 2D feature point locations and the estimated locations obtained by projecting the corresponding 3D point locations to the 2D plane. By using the Levenberg-Marquardt (LM) algorithm, they can optimize both the parameters of the 3D camera motion and the shape of the 3D object. However, when utilizing the LM algorithm in BA-based 3D reconstruction methods, it is necessary to manipulate an extremely large inverse Hessian matrix, and thus the solution procedure is very time-consuming. Therefore, the authors in [7] proposed a local optimization process to accelerate the solution process.

3D shape reconstruction is commonly performed using a class of systems referred to as "Shape from X", where X denotes a physical property such as shading, focus or motion. Shape from shading (SfS) methods [2,8-11] enable the reconstruction of 3D shape provided that the scene or object satisfies the Lambertian reflection model. However, in practice, not all scenes or objects satisfy this constraint. Accordingly, various researchers have turned to develop Bidirectional Reflection Distribution Functions (BRDFs) as a more general representation of the reflection model for 3D reconstruction purposes [12-14].

In shape from focus (SfF) methods, a sequence of images containing a focused region of the imaged object and an unfocused background is taken while the focus setting is progressively changed in incremental steps. SfF methods comprise two basic functions, namely a focus measurement function (e.g., Sum Modified Laplacian (SML) [15], Cross Sum Modified Laplacian (XSML) [16], moment of DCT coefficient [17], or ratio of wavelet coefficient [18]) and a smoothing function used to reduce the error caused by a false-determined focus value (e.g., Gaussian distribution [16] or Bezier-surface [17,19]). Since SfF methods utilize the focus value to estimate the 3D information, the depth of field must be very small. Accordingly, in implementing SfF methods, special image capture devices such as microscopes [15,17-19] or a holographic camera along with a CCD sensor [16] are required.

Shape from motion (SfM) methods solve the 3D reconstruction problem using a factorization approach [3,20-24]. By introducing the rank constraint and factorizing by singular value decomposition (SVD), factorization approaches based on orthographic [3], weak-perspective [20], or para-perspective [21] projection models factorize the location matrix of the 2D feature points in different image frames to the 3D rotation matrix of the 2D image frame and the 3D shape matrix of the object, respectively. In the method proposed in [22], both the 3D rotation information of the object and the 3D shape of the object were recovered by means of a perspective projection model. However, SfM methods generally acquire a less dense depth map than SfS or SfF methods since it is difficult to find the corresponding feature points with dense 2D feature points.

The literature contains many other techniques for solving the 3D reconstruction problem besides those described above. For example, in [25], the 3D object's shape is reconstructed using a Bayesian MAP technique, while in [26], 3D reconstruction is achieved using a graph cut method. Both methods yield convincing results. However, their respective models are more complicated than those used in the methods described above.

All of the methods described above are intended for the reconstruction of static or rigid objects. However, in recent years, various algorithms have been proposed for the 3D reconstruction of non-rigid objects over a sequence of images [27-29]. Generally speaking,

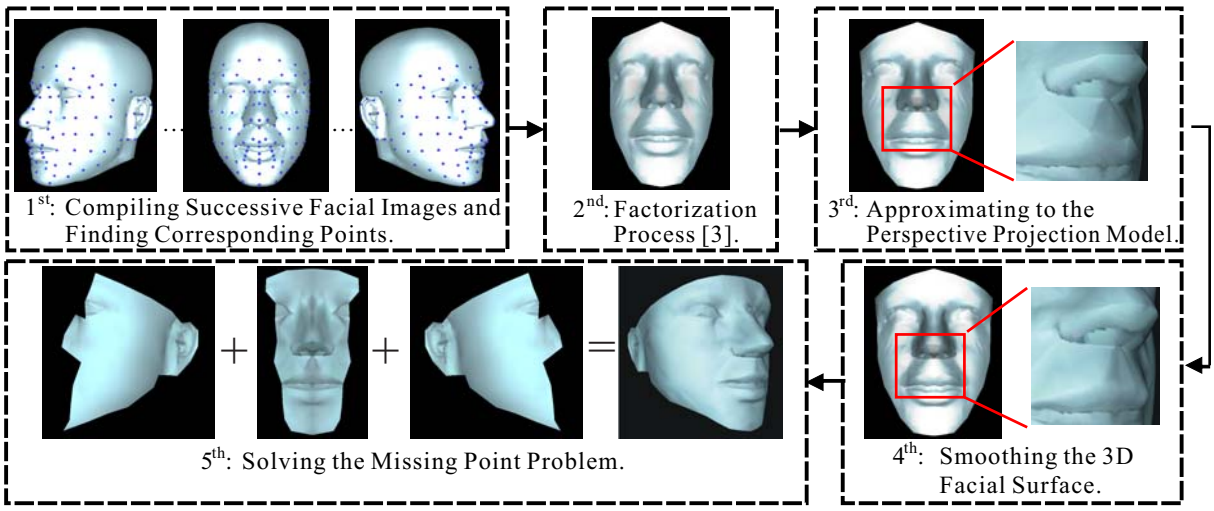

FIGURE 1. Workflow of the proposed 3D facial reconstruction system

such algorithms model the 3D shape as a linear weighted combination of a set of shape vectors, such that different shapes can be modeled by assigning different weights.

The present study proposes a method for the efficient reconstruction of the 3D human face using a web camera and a personal computer. In performing the reconstruction process, an assumption is made that the head rotates in the pan-rotation direction without non-rigid motions, such as facial expression. This assumption is consistent with the general environment setting in multi-frame 3D reconstruction algorithms [3-5,21], and is much simpler than that in [16]. Since the proposed method is implemented using a web camera, a SfF reconstruction approach cannot be applied because the very small depth of field required by such approaches cannot be achieved. Accordingly, a SfM factorization-based method is used since such methods are not only more straightforward to implement than BA approaches, but also achieve a more accurate reconstruction result than SfS approaches. There are three contributions in the proposed approach. The first one is that a divide and counter reconstruction method is proposed by dividing the 3D facial surface into small groups and then reconstruct each group by using factorization approach based on orthographic projection module. Then, the reconstruction results of all groups are integrated to form a complete 3D facial surface, which is almost as accurate as the reconstruction result using a perspective projection model. Second, a novel smoothing method is proposed for giving the surface of the reconstructed face a denser, smoother and more realistic appearance by interpolating additional 3D vertices on the facial surface from the vectors of the existing 3D vertices. Finally, a new divide and conquer solution is proposed to overcome the missing point problem, which is caused by occlusion at high pan rotation angles, commonly arising in 3D reconstruction applications. The image sequence is divided into 3 segments, and there is no occlusion problem within each segment. We can reconstruct each segment respectively and then integrate these reconstruction results together to form a complete 3D model. This article is organized in the following way. In Section 1, the related 3D reconstruction methods are given from the literature. And in Section 2, the proposed 3D reconstruction method is described in detail. Section 3 presents the result compare with different existing approaches. Section 4 is the conclusion of this study.

2. **3D Facial Surface Reconstruction System.** As shown in Figure 1, the 3D facial surface reconstruction system proposed in this study consists of five modules. In the first module, a series of facial images are compiled as the face rotates in the pan-rotation

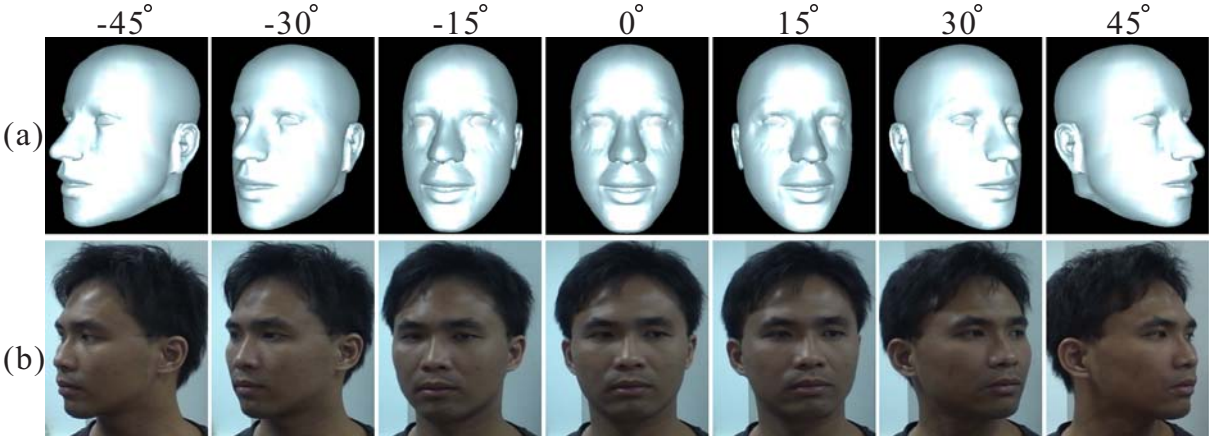

FIGURE 2. Rotation of (a) facial model and (b) real face in pan-rotation direction

direction. Certain feature points are automatically located in the first image frame by adopting the approach in [30] and are then tracked over the remaining frames using an optical flow technique [31]. In the second module, a factorization approach based on an orthographic projection model [3] is applied to reconstruct the 3D face. In the third module, the perspective projection model of the 3D face is approximated by integrating the piecewise orthographic projection models. In the fourth module, an interpolation-based smoothing method is applied to give the reconstructed 3D face a smoother and more realistic appearance. Finally, in the fifth module, a method aimed at resolving the missing point problem in 3D reconstruction applications is applied in order to reconstruct the complete 3D face.

2.1. **1st module: compiling successive facial images and finding corresponding points.** In the reconstruction method proposed in this study, a web camera is fixed in front of the subject at a distance about 40 cm and is used to capture a sequence of facial images as the subject gradually turns the head in the pan-rotation direction without non-rigid motions (see Figure 2). Note that the image sequence is intentionally captured by rotating the head in front of a stationary camera since this approach is more convenient than rotating the camera around the still head. And when the distance from the camera to the face of the user is around 40 to 45 cm, the face occupies around 3/4 of the image and is of a sufficient size to permit the accurate tracking of the feature points. Having acquired the sequence of facial images, various facial feature points are automatically located in the first frame of the sequence using the Direct Combined Model (DCM) method [30]. An optical flow technique [31] is then used to track the selected feature points over the remaining image frames in the sequence. In general, tracking errors may occur for feature points located in textureless regions of the face, such as the cheeks and chin. This problem is commonly resolved by increasing the size of the tracking window in order to include a sufficient amount of texture information, or by applying a manual correction procedure. While both methods successfully improve the quality of the reconstruction results, they also increase the amount of computing resources required and delay completion of the tracking procedure. A Delaunay Triangulation scheme [32,33] is then applied to construct 2D triangles from various feature points. The following factorization process, these triangles are used to construct corresponding 3D polygons, in which the position of each vertex m is defined as (x_m, y_m, z_m) in the 3D facial (i.e., 3D world) coordinate system.

2.2. **2nd module: factorization process based on orthographic projection model.** In the proposed system, the 3D face is reconstructed initially using the factorization approach based on the orthographic projection model [3]. The orthographic projection model is far simpler than the perspective projection model [20-22], and it involves first locating the centroid of the 3D feature points (or 3D shape) of the object and then specifying the 3D centroid point as the origin of the 3D world coordinate system. A point matrix W is then created comprising the locations of N corresponding 2D feature points over F facial image frames. Finally, the point matrix W is factorized into the rotation matrix R of the 2D image frames and the 3D shape S by applying the singular value decomposition (SVD) algorithm. The 3D shape S obtained by the factorization method [3] is referred to hereafter in this study as the *factorization result*.

2.3. **3rd module: approximating to perspective projection model.** As described above, the facial reconstruction method proposed in this study is based on the orthographic projection model. However, the web camera used to acquire the facial images is based on a perspective projection model. Therefore, if the orthographic model is to provide an accurate approximation of the perspective model, the ratio of the object depth, Δd, to the distance between the camera center and the object, d, must be very small, i.e., $\Delta d \ll d$ [34]. Clearly, if the value of Δd can be reduced (note that d is determined when the facial images are recorded), the ratio $\Delta d/d$ is also reduced. As a result, the reconstruction result obtained for the depth of each vertex on the facial surface, z_m, will be improved. Accordingly, in the third module of the proposed reconstruction system, the 3D facial shape S obtained by the factorization process is divided into n groups (or layers) $g_1, g_2, \ldots, g_n$ along the z-axis direction using a vector quantization (VQ) clustering method [35]. In other words, the 3D facial shape S is described by $S = g_1 \bigcup g_2 \bigcup \cdots \bigcup g_n$ and $|g_i \bigcap g_{i+1}| \geq 4$ for $1 \leq i \leq n-1$. In practice, a trade-off must be made between the number of groups or clusters and the accuracy of the 3D facial reconstruction results. Therefore, applying VQ by setting the requirement for minimum number of feature points within each group for grouping, we do not have to worry about how to determinate the number of the groups.

Each group, g_i, has its own coordinate system and consists of a minimum of 4 reconstructed 3D feature points. In general, registering two 3D coordinate systems requires the presence of at least 4 common points in the two coordinate systems in order to solve the 12-parameter rotation and translation transformation matrix, i.e., $|g_i \bigcap g_{i+1}| \geq 4$. Applying the factorization approach proposed in [3], the 3D reconstruction result g_i' for each group is constructed from a set of corresponding 2D projection points p_i. In other words, p_i is factorized into $r_i' g_i'$, where r_i' is the rotation matrix of group i. We say that the reconstruction results g_i' are reconstructed by piecewise orthographic projection model.

As stated above, each group g_i' has its own 3D coordinate system. However, to reconstruct a complete 3D face, it is necessary to register all of the groups into the same 3D coordinate system. The group registration procedure performed in the present study can be summarized as follows:

1. Neighboring groups g_i' and g_{i+1}' have overlapping regions $o_{i(i+1)}$ in g_i' and $o_{(i+1)i}$ in g_{i+1}', respectively. Therefore, the 12-parameter transformation matrix τ between $o_{i(i+1)}$ and $o_{(i+1)i}$ has the form

$$o_{i(i+1)} = \tau o_{(i+1)i}, \text{ where } \tau = \begin{bmatrix} R_{3\times 3} & T_{3\times 1} \\ 0_{1\times 3} & 1 \end{bmatrix} \tag{1}$$

2. Transform the coordinate of g_{i+1}' into that of g_i' as g_{i+1}'' using the transformation matrix, i.e., $g_{i+1}'' = \tau g_{i+1}'$.

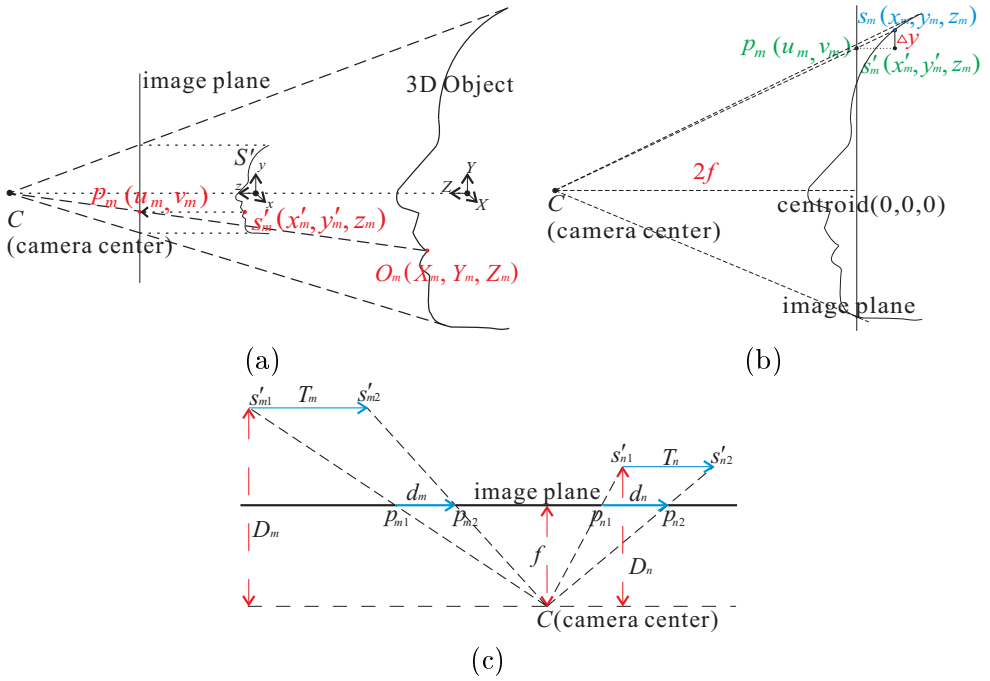

FIGURE 3. (a) Point p_m in the image plane represents the 2D projection point of both the 3D reconstructed point s'_m in the orthographic model and the 3D object point O_m in the perspective model; (b) difference Δy between the reconstruction results in the y-axis direction obtained from the orthographic and the perspective model, respectively; (c) determining the camera focal length f

3. Denote g''_{i+1} as g'_{i+1}, i.e., $g'_{i+1} = g''_{i+1}$.
4. Increase i by 1 for the next group and return to Step 1. Repeat Steps 1-4 until i equals n.

Once all n groups have been registered, the reconstructed 3D facial model (or facial shape) is described by $S' = g'_1 \bigcup g'_2 \bigcup \cdots \bigcup g'_n$. In other words, the reconstructed depth z_m of each vertex in the 3D S' is more accurate than that in the original model S. The VQ clustering procedure is then applied to the reconstructed 3D facial shape S' and the clustering results are compared with the previous results. If the samples in any cluster are changed, the clustering and registration procedures are applied once again to S'. The procedure is repeated iteratively in this way until no change in the samples in any of the clusters occurs.

Having improved the depth value z_m of each vertex in the 3D object, the accuracies of the x- and y-coordinate values of each vertex should also be improved. Analyzing the reconstructed 3D facial model S', it is found that the x- and y-axis components of the 3D coordinates $s'_m(x'_m, y'_m, z_m)$ of every vertex S'_m in S' are virtually identical to those of the corresponding 2D projection point $p_m(u_m, v_m)$, i.e., $x'_m = u_m$ and $y'_m = v_m$ (see Figure 3(a)). This finding is reasonable since the factorization approach is based on the orthographic projection model. As shown in Figure 3(b), $x'_m = u_m$, $y'_m = v_m$ holds for the perspective model when both the 3D object and the 2D image plane are located at a distance of twice the focal length f from the origin of the camera coordinate system [36].

As shown in Figure 3(c), the focal length f of the camera can be calculated from the reconstructed 3D coordinates of each vertex $s'_m(x'_m, y'_m, z_m)$ in S' and the rotation matrix R. Let s'_{m1} and s'_{n1} be two vertices in S' and let p_{m1} and p_{n1} be the projection points of s'_{m1} and s'_{n1}, respectively, in the 2D image plane. Furthermore, let T_m and T_n be

the translation vectors as vertices s'_{m1} and s'_{n1} in frame r move to vertices s'_{m2} and s'_{n2}, respectively, in frame $r+1$ in the 3D space. In addition, let d_m and d_n be the translation vectors as feature points p_{m1} and p_{n1} in frame r move to feature points p_{m2} and p_{n2} in frame $r+1$, respectively. Finally, let D_m and D_n be the depths of s'_{m1} and s'_{n1}, respectively, in the camera coordinate system. It can then be shown that

$$D_m = \frac{T_m}{d_m} f \text{ and } D_n = \frac{T_n}{d_n} f \tag{2}$$

In the orthographic projection model, the values of D_m and D_n are unknown. However, the relative depth between them can be calculated because the 3D position of each vertex in S' is known. Thus, the focal length f can be derived as

$$f = (D_m - D_n)\frac{d_m d_n}{T_m d_n - T_n d_m} \tag{3}$$

From Figure 3(b), it can be seen that the reconstructed vertex s'_m and the actual vertex s_m have an error Δy in the y-axis direction, i.e.,

$$\Delta y = y'_m \frac{z_m}{2f} \tag{4}$$

Therefore, the improved value y_m for vertex s_m is obtained as

$$y_m = y'_m + \Delta y = y'_m + \frac{y'_m z_m}{2f} \tag{5}$$

Similarly, the improved value x_m for vertex s_m is expressed as

$$x_m = x'_m + \Delta x = x'_m + \frac{x'_m z_m}{2f} \tag{6}$$

Therefore, the final result for the coordinates of vertex $s_m(x_m, y_m, z_m)$ is obtained by applying (5) and (6). The facial surface obtained from this improved reconstruction module is referred to hereafter in this study as the *improved result*.

2.4. **4th module: smoothing 3D facial surface.** In general, the higher the density of the 3D feature points in the facial image, the smoother the appearance of the reconstructed 3D facial shape. However, a high feature point density may cause tracking problems in that some points may be erroneously matched with one of their neighbors. Therefore, in the 3D facial reconstruction system proposed in this study, a small number of 2D feature points are used initially to create a draft 3D facial model containing only a limited number of polygon vertices. Each polygon of the mesh is a flat plane. However, the surface of the real face is curved. Therefore, a novel smoothing method is developed in which additional vertices are interpolated from the original vertices of the improved result.

In applying the smoothing process to the surface of the reconstructed 3D facial model, all of the flat planes on the 3D surface are transformed into curved planes by interpolating additional 3D vertices, e.g., vertices D, E, F and G in Figure 4(b.2). As a result, the surface takes on a more smooth and realistic appearance. Each polygon m in the 3D facial model S has a normal direction $\overrightarrow{N_m}$ and each of its vertices v is a common vertex of n neighboring polygons. Therefore, the normal direction $\overrightarrow{N_v}$ of each vertex v is the average normal direction of the n neighboring polygons. As shown in Figure 4(a.1), vertex B on edge $\overline{BC}$ has a normal direction $\overrightarrow{N_B}$; vertex C has a normal direction $\overrightarrow{N_C}$. Therefore, the vector $\overrightarrow{N_{PB}}$ lying on the plane formed by $\overrightarrow{N_B}$ and $\overrightarrow{BC}$ can be found by solving the

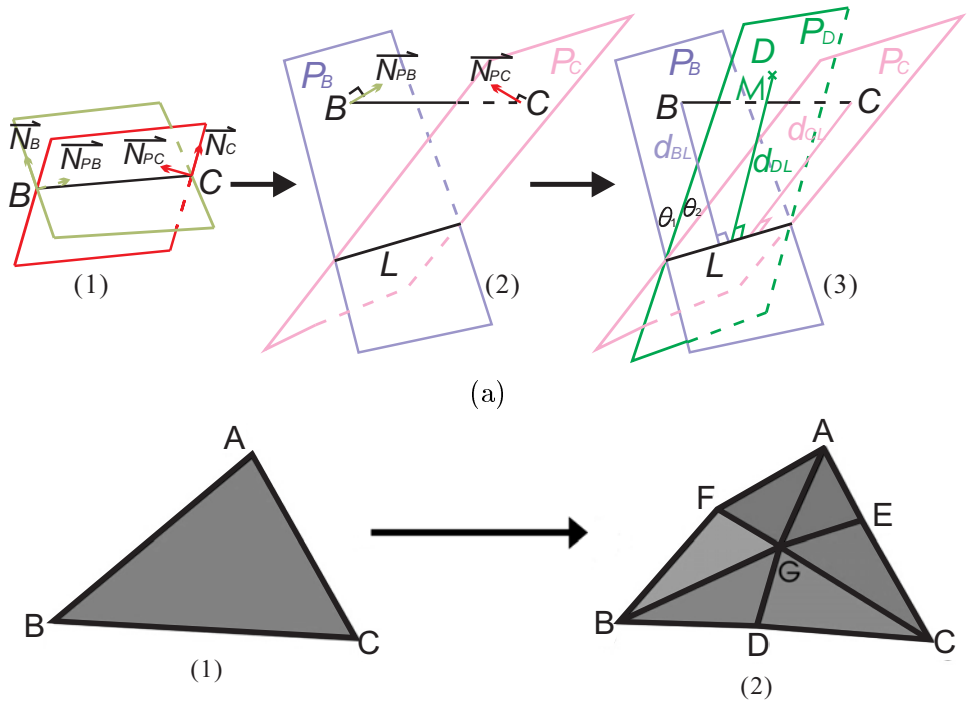

FIGURE 4. (a) Example of smoothing method and (b) smoothing result obtained for each polygon in 3D space

following equations:

$$\begin{cases} \overrightarrow{N_{PB}} \bullet \overrightarrow{N_B} = 0 \\ \overrightarrow{N_{PB}} \bullet \left(\overrightarrow{N_B} \times \overrightarrow{BC} \right) = 0 \end{cases} \tag{7}$$

Similarly, vector $\overrightarrow{N_{PC}}$ can be found by solving the following equations:

$$\begin{cases} \overrightarrow{N_{PC}} \bullet \overrightarrow{N_C} = 0 \\ \overrightarrow{N_{PC}} \bullet \left(\overrightarrow{N_C} \times \overrightarrow{CB} \right) = 0 \end{cases} \tag{8}$$

Having found vectors $\overrightarrow{N_{PB}}$ and $\overrightarrow{N_{PC}}$, the corresponding planes, i.e., P_B and P_C, respectively, can be determined. As shown in Figure 4(a.2), P_B is a plane with a normal direction $\overrightarrow{N_{PB}}$ and passing through B. Similarly, P_C is a plane with a normal direction $\overrightarrow{N_{PC}}$ and passing through C. Assume that planes P_B and P_C intersect at line L. Point M on edge $\overline{BC}$ and line L form a plane P_D with a normal direction $\overrightarrow{N_{PD}}$ (see Figure 4(a.3)). The vector $\overrightarrow{N_M}$, i.e., the normal direction vector of point M lying on plane P_D, can be found by solving the following equations:

$$\begin{cases} \overrightarrow{N_M} = \alpha \overrightarrow{N_B} + (1-\alpha) \overrightarrow{N_C} \\ \overrightarrow{N_M} \bullet \overrightarrow{N_{PD}} = 0 \end{cases} \tag{9}$$

where α is a scaling factor with a value in the range $0 \leq \alpha \leq 1$ and is used to control the relative weights of $\overrightarrow{N_B}$ and $\overrightarrow{N_C}$. Finally, the vertex D can be determined from

$$\begin{cases} \overrightarrow{MD} = \beta \overrightarrow{N_M} \\ d(D,L) = d(B,L) \dfrac{\theta_2}{\theta_1 + \theta_2} + d(C,L) \dfrac{\theta_1}{\theta_1 + \theta_2} \end{cases} \tag{10}$$

where β is a scaling factor with a value of $\beta \geq 0$ and $d(X, L)$ is the distance from X to L, where X denotes vertex B, C or D. Equations (7) to (10) can be used to obtain interpolated vertices E and F on edges $\overline{CA}$ and $\overline{AB}$, respectively, in a similar manner (see Figure 4(b.2)). Note that point M in Figure 4(a.3) represents the mid-point of the interpolated vertices D, E and F. To locate vertex G in Figure 4(b.2), lines AD, BE and CF are used to determine the interpolated vertices G_{AD}, G_{BE} and G_{CF}, respectively, by means of (7) to (10). Since G should be located at the centroid of the triangular polygon ABC, the M position of vertices G_{AD}, G_{BE} and G_{CF} should satisfy $d(X, Y) : d(M, Y) = 2 : 1$, where (X, Y) is (A, D), (B, E) or (C, F). Vertex G is then given by $G = (G_{AD} + G_{BE} + G_{CF})/3$.

Following the smoothing process, the original polygon m_{ABC} is transformed into six individual sub-polygons. These sub-polygons construct a curved surface rather than a flat plane, and hence the 3D surface has a smoother appearance. The outcome of the smoothing process is referred to hereafter in this study as the *smoothing result*.

2.5. **5th module: solving the missing point problem.** Some of the facial feature points become occluded at large values of the pan-rotation angle. As a result, some elements of the point matrix W are missing, and thus the factorization process fails. Although the authors in [23,24] proposed various algorithms to resolve this problem and acquired a good result, an alternative approach, which is more straightforward and easier, is presented here. From observation, it is found that when the pan rotation reaches an angle of approximately 10°-15° to the left or right, the feature points adjacent to the nose start to become occluded. And when the head is rotated to the left or right by an angle of more than about 30°, the ears of the face become clearly visible. Moreover, the proposed reconstruction algorithm is based on a small value of the relative depth Δd for each cluster, the projection model is close to the orthographic projection model, and thus a superior reconstruction result is obtained. To ensure that the overall depth of the 3D shape S reconstructed in the second module is minimized, it is important to choose an appropriate range of rotation angle for image sequence used to reconstruct the 3D shape. Accordingly, in the facial reconstruction method proposed in this study, the captured image sequence is divided into three segments, namely the left side-view V_1 (from $-45°$ to $-30°$), the frontal-view V_2 (from $-10°$ to $10°$) and the right side-view V_3 (from $30°$ to $45°$), as shown in Figure 5(a). The 3D rotation angle of the face is automatically estimated using the method proposed in [37,38]. The reconstruction modules described in Sections 2.1-2.3 are then used to reconstruct the 3D facial shape based on the images within segments V_1, V_2 and V_3, respectively. The corresponding reconstruction results are shown in Figures 5(b)-5(d), respectively. Side-views V_1 and V_3 are then registered to the frontal-view V_2 using the same registration procedure as that described in the third module (see Figure 5(e)).

However, the registration results obtained when aligning two different segments located at different depths in 3D space are sensitive to the depth variance of the overlapping regions. Specifically, the reconstruction error increases as the depth variance of the overlapping regions increases. Figures 5(f) and 5(g) show the improved reconstruction results (shown in red) obtained when using different overlapping regions to carry out the registration procedure. (Note that the white 3D shapes in the two figures indicate the original facial model). In Figure 5(f), the registration results (i.e., the red 3D shape) are based on the purple regions in Figures 5(b)-5(d), which have a small depth variance. Conversely, the registration results presented in Figure 5(g) are based on the cheek regions, which contain significant depth variances. Thus, comparing the two sets of registration results, it is apparent that Figure 5(g) represents a poorer reconstruction performance.

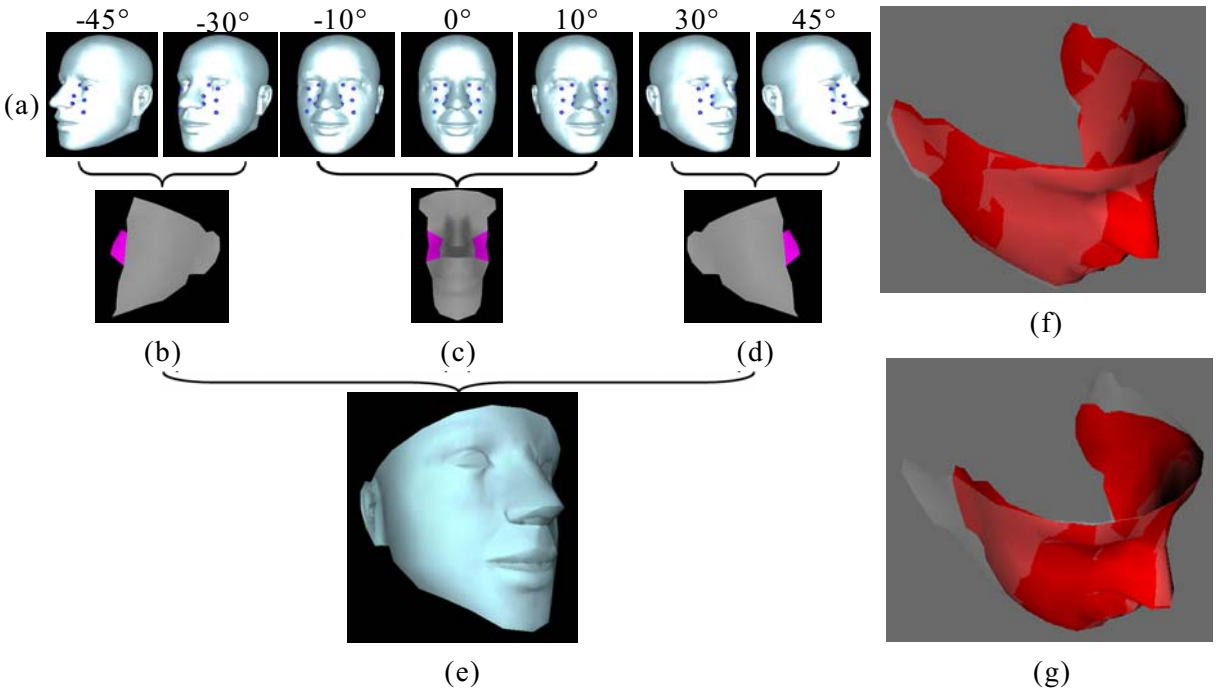

FIGURE 5. (a) Sequence of facial images under different pan-rotation angles; (b)-(d) improved results in which the purple regions denote overlapping regions (the dots shown in (a)); (e) registered result; (f) and (g) registration results obtained using different overlapping regions

3. **Experimental Results.** Seven sections present comparisons with other approaches and different evaluations of the reconstruction result. Section 3.1 is the comparison of reconstruction performance of different methods. Sections 3.2-3.4 are the performance evaluation of improved 3D reconstruction method, at different distances, for different 3D facial shapes, respectively. Section 3.5 evaluates the influence of tracking errors on the reconstruction result. Section 3.6 presents the smoothing result. And Section 3.7 presents the reconstruction results for actual human faces.

3.1. **Comparison of reconstruction performance of different methods.** To evaluate the performance of the reconstruction modules (i.e., Modules 1-3), a small box was designed and then reconstructed using values of $\Delta d/d$ equal to 1/3 and 1/7, respectively (see Figures 6(a)-6(d) and 6(e)-6(h)). In both cases, the box contained 101 feature points (see Figures 6(b) and 6(f)). In Figures 6(c), 6(g), 6(d) and 6(h), the red boxes indicate the reconstruction results, while the white boxes indicate the original box. Comparing Figures 6(c) and 6(g), it is observed that the reconstruction results obtained using the larger $\Delta d/d$ ratio of 1/3 are distorted more significantly than those obtained using a smaller ratio of 1/7. In other words, a closer agreement is obtained between the orthographic projection model and the perspective projection model as the value of $\Delta d/d$ is reduced. Figures 6(d) and 6(h) present the improved results.

From observation, the box will be out of the field of view of the camera when $\Delta d/d$ exceeds a value of 1/3. Furthermore, the size of the box is too small to track the feature points correctly when $\Delta d/d$ has a value of less than 1/7. Thus, in evaluating the improved results in the proposed method, the reconstruction error was compared with that of the orthographic factorization method (Orthographic FM) [3], weak-perspective factorization method (Weak-pers. FM) [20], and papa-perspective factorization method (Para-pers. FM) [21] for $\Delta d/d$ ratios of 1/3, 1/5 and 1/7. The corresponding results are presented

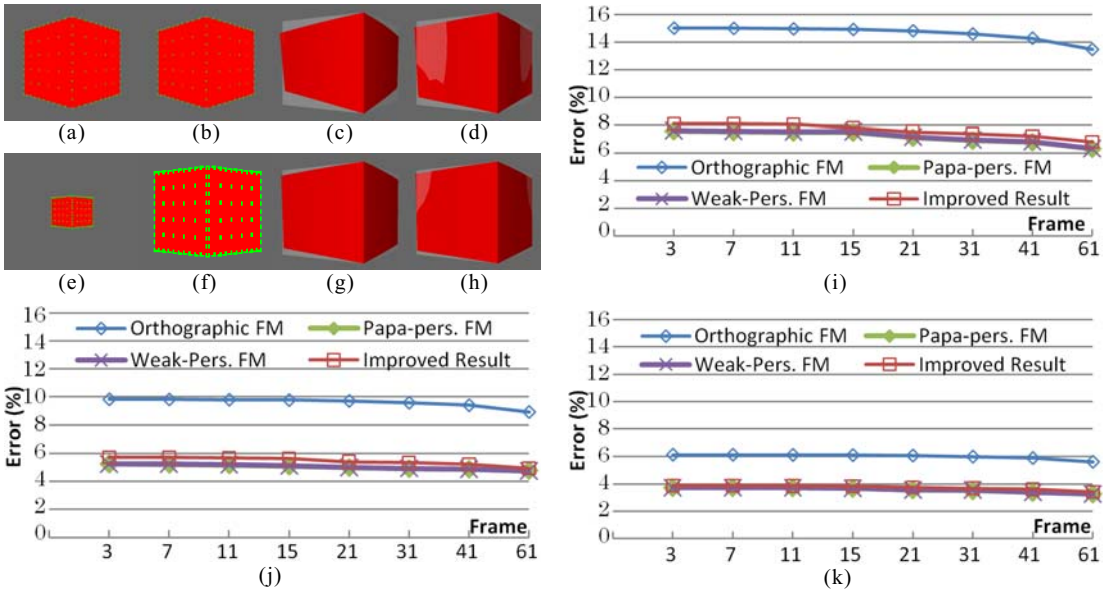

FIGURE 6. (a)-(d) and (e)-(h) Reconstruction results obtained when using $\Delta d/d$ ratios of 1/3 and 1/7, respectively. (a) and (e) Original images, (b) and (f) close-up view of boxes, (c) and (g) factorization results, and (d) and (h) improved results. (i)-(k) are the estimation errors of reconstructed boxes given $\Delta d/d$ equal to 1/3, 1/5 and 1/7, respectively.

TABLE 1. Average improvement for different $\Delta d/d$ ratios

Ratio ($\Delta d/d$)	1/3	1/4	1/5	1/6	1/7
Improvement (%)	43.3	41.5	40.0	38.6	37.8

in Figures 6(i)-6(k), respectively. The estimation error is defined here as the average Euclidean distance between all the reconstructed feature points and the original feature points divided by the average Euclidean distance between all the feature points and the origin of the original box. And the result of Orthographic FM equals to the factorization result in this study. It is seen that the error in the improved result is only about 0.1%-0.5% higher than that obtained from the other two perspective factorization methods. In other words, the proposed method yields a good approximation of the perspective projection model. Table 1 summarizes the average improvement percentage obtained by the proposed method at each value of $\Delta d/d$.

The performance of the proposed 3D reconstruction method was further evaluated by comparing the reconstruction error with that obtained from various existing methods. In performing the experiments, the 3D vase model used in the SfS methods presented in [2,9] was taken as the ground truth. Several feature points were manually selected on the 3D vase model and were recorded as it was rotated at a rate of one degree per frame from $-10°$ to $10°$ in the pan-rotation direction. The vase was then reconstructed by the proposed method. Table 2 compares the 3D reconstruction errors of the proposed method ($\Delta d/d = 1/5$) in the z-axis direction with those obtained from the SfS methods presented in [2,9]. The computational costs of SfS methods in [2] are less and those in [9] are more than the proposed method. However, the results presented in Table 3 show that the proposed method yields the most accurate reconstruction results of all the considered methods. Note that the BRDF method [13] uses a more sophisticated reflectance model, and is therefore better suited to the 3D reconstruction of synthetic or real objects. However, it has a high computation complexity, i.e., $O(N^3)$, where N is the number of reconstructed points. As a result, reconstructing the 3D vase considered in the present evaluation trials

TABLE 2. Comparison of 3D reconstruction errors (units) in z-axis direction of SfS methods in [2,9] and proposed method ($\Delta d/d = 1/5$)

Method	z-axis Error Value (units)
Zheng and Chellappa [2]	9.69
Lee and Kuo [2]	9.39
Kimmel and Sethian [9]	7.60
Tankus, Sochen and Yeshurun [9]	5.03
Proposed Method	1.85

TABLE 3. Comparison of 3D reconstruction errors (%) of BA methods [5] and proposed method by adding Gaussian noises with different standard deviations ($\Delta d/d = 1/5$)

Standard Deviation	0.5	1.0	1.5
Model-Based BA Method [5]	1.22%	1.57%	1.83%
Classical BA Method [5]	3.70%	6.00%	6.13%
Proposed Method	5.92%	6.03%	6.60%

requires a time of around tens of minutes. Recently, the BRDF work in [14] improves the computational time to minutes. The computational complexity of the proposed shape from motion (SfM) method is dominated by the SVD process [39], i.e., $O(N^3)$. However, the proposed method reconstructs the 3D model in just seconds since it reconstructs only some of the feature points, while the reconstruction using BRDF relies on dense points in order to acquire a good result. As a consequence, the computational time of BRDF is much longer than the proposed method.

Table 3 compares the reconstruction errors of the proposed method ($\Delta d/d = 1/5$) with that of two reconstruction methods based on the BA approach presented in [5] by adding Gaussian noise with standard deviations of 0.5, 1.0 and 1.5, respectively. The results show that the two BA-based methods yield a better 3D reconstruction performance than the proposed method for all values of the standard deviation. However, the computational time of the proposed 3D reconstruction method (i.e., a few seconds) is much faster than that of the two BA methods (i.e., several minutes). In performing the reconstruction process, BA methods solve a set of complex differential equations iteratively in an attempt to minimize the distances between the actual locations of the 2D image feature points and the locations of these points when projected from the corresponding 3D feature points onto the 2D image. Thus, the computational complexity is equal to $O(Mf(M + 2f)^2)$ [6], where M is the number of parameters to be solved and f is the number of frames. Conversely, the method proposed in this study uses a relatively simple method to find the relationship between the actual locations of the 2D points and the locations of the 2D points projected from the 3D geometric points. Consequently, although the reconstruction accuracy of the proposed approach is around 4.5% lower than that of the two BA-based approaches, the computational time is reduced by a factor of more than 100.

3.2. **Performance evaluation of improved 3D reconstruction method.** As described in the Introduction section, the 3D facial reconstruction method proposed in this study is implemented using a web camera and a personal computer. In general, the depth of the frontal face from the tip of the nose to the cheek is around 7 to 8 cm for most individuals. In other words, in the present study, the object depth Δd is 7 to 8 cm. When the distance from the camera to the face of the user is around 40 to 45 cm, the face occupies around 3/4 of the image and is of a sufficient size to permit the accurate

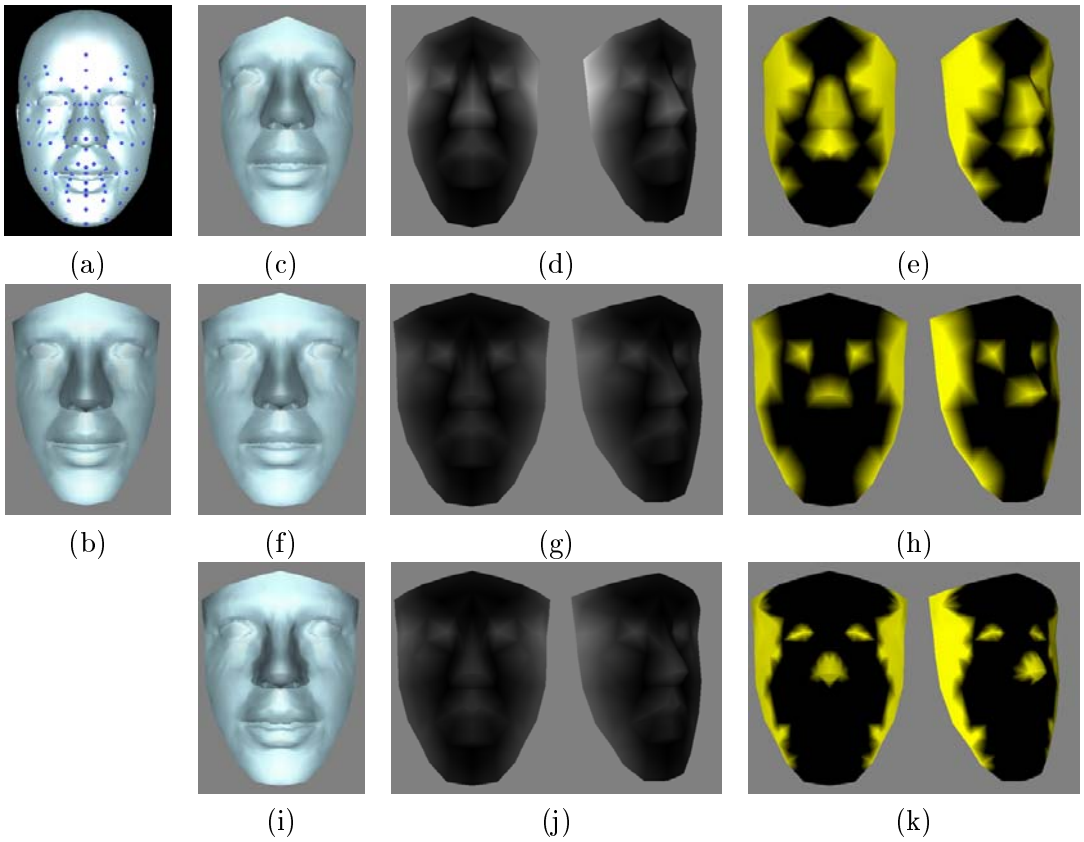

FIGURE 7. (a) 3D talking head model with 90 feature points. (b) Original facial model without feature points. (c), (f) and (i) Factorization result, improved result and smoothing result, respectively. (d), (g) and (j) Estimation errors of (c), (f) and (i), respectively. (e), (h) and (k) Binary images corresponding to (d), (g) and (j), respectively.

tracking of the feature points. Accordingly, the $\Delta d/d$ ratio in the proposed reconstruction system is around 1/5. The performance of the proposed 3D facial reconstruction system was further evaluated using Microsoft's 3D "Talking Head" facial model to provide the ground truth for the 3D reconstruction results. As shown in Figure 7(a), 90 3D feature points and contour points were assigned to the face in accordance with the contours of the facial features (i.e., the eyebrows, eyes, nose, mouth, chin and cheeks). The model was then rotated from left ($-15°$) to right ($15°$) in the pan-rotation direction. Figure 7(b) shows the original facial model with texture but no feature points.

In evaluating the performance of the proposed facial reconstruction system, the reconstruction error of the improved result was estimated in the same way as that described in Section 3.1. Furthermore, to evaluate the smoothing performance of the proposed system, the coordinates of the smoothed 3D facial shape (495 vertices) and the original 3D facial model (dense vertices) were aligned to the origin $O(0,0,0)$. The reconstruction error of the smoothing result was then defined as the distance between v, i.e., a vertex on the smoothed 3D shape, and the vertex on the original 3D model having the shortest distance to line vO of all the vertices on the original 3D model. The average reconstruction error of the smoothing results was then calculated in the same manner as that in Section 3.1.

Figures 7(c), 7(f) and 7(i) show the factorization results, improved results and smoothing results obtained from the 2^{nd}, 3^{rd} and 4^{th} modules in the proposed reconstruction system, respectively. Comparing the original facial model shown in Figure 7(b) with the factorization result presented in Figure 7(c), it is apparent that the reconstructed face

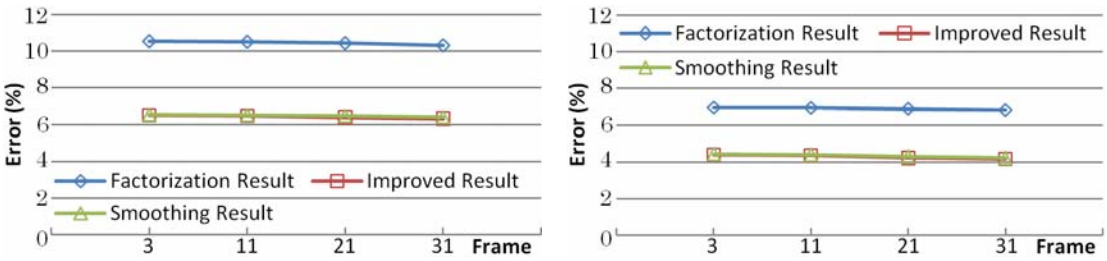

FIGURE 8. Comparison of reconstruction errors of different reconstruction results: (a) at distance d of 35 units ($\Delta d/d \approx 1/5$) and (b) at distance d of 50 units ($\Delta d/d \approx 1/7$)

is much narrower in the x-axis direction with reconstruction error about 3%. However, the reconstruction errors of the 3D facial shapes presented in Figures 7(f) and 7(i) are less than 1% in both the x- and the y-axis directions. Figures 7(d), 7(g) and 7(j) show the estimation errors of Figures 7(c), 7(f) and 7(i), respectively, using a gray value representation, in which the brightness increases with an increasing estimation error. Figures 7(e), 7(h) and 7(k) present binary images corresponding to Figures 7(d), 7(g) and 7(j), respectively. Note that in these binary images, the bright pixels indicate a reconstruction error of 3% or more relative to the original facial model. It can be seen that the improved reconstruction module successfully reduces the reconstruction error (see Figure 7(h)).

3.3. **Performance evaluation at different distances.** From observation, the head falls beyond the field of view when $\Delta d/d$ is larger than 1/3, i.e., d is less than 25 units (centimeters in our case). Moreover, the face is too small to track the feature points reliably when $\Delta d/d$ is less than 1/7 (d is larger than 50 units). Figures 8(a) and 8(b) present the reconstruction errors of the factorization result, the improved result and the smoothing result at values of d equal to 35 units ($\Delta d/d$ is close to 1/5) and 50 units ($\Delta d/d$ is close to 1/7), respectively. The average reconstruction errors of Orthographic FM [3] are found to be 10.5% and 7.0% for values of d equal to 35 units and 50 units, respectively. Meanwhile, the average reconstruction errors of the improved results are approximately 6.5% and 4.1%, respectively. In other words, for a given value of $\Delta d/d$, the degree of reconstruction improvement obtained by the proposed method for the 3D facial model is very similar to that obtained for the simple box model (see Figures 6(j) and 6(k)). Thus, the versatility and effectiveness of the proposed system are confirmed.

3.4. **Performance evaluation for different 3D facial shapes.** Figure 9 presents 10 additional face models used to evaluate the reconstruction performance of the 2$^{\text{nd}}$ (factorization) module, the 3$^{\text{rd}}$ (improved reconstruction) module and the 4$^{\text{th}}$ (smoothing) module. In performing the reconstruction experiments, the reconstruction error was evaluated given a $\Delta d/d$ ratio equal to 1/5. Moreover, the number of groups in the 3$^{\text{rd}}$ module was assigned a value between 10 and 12 by applying the VQ process. The reconstruction errors of the factorization result, the improved result and the smoothing result are shown in Table 4. It can be seen that the reconstruction errors of the improved result and the smoothing result in the x-, y- and z-axis directions are around 78%, 70% and 17%, respectively, better than those of the factorization result. In other words, the improved reconstruction module yields a significant reduction in the reconstruction error in the x- and y-axis directions. However, a less obvious improvement is obtained in the z-axis (i.e., depth) direction. This result arises because the ratio of $\Delta d/d$ used in the layered reconstruction approach is insufficiently small. Nonetheless, the improved reconstruction

FIGURE 9. 10 additional face models used to estimate reconstruction errors

TABLE 4. Reconstruction error (%) of each axis with $\Delta d/d = 1/5$ (factorization result [3]/improved result/smoothing result)

Angles	**+1° to +1°**	**−5° to +5°**	**−10° to +10°**	**−15° to +15°**
***x*-axis**	2.60/0.74/0.92	2.57/0.70/0.88	2.54/0.63/0.83	2.47/0.57/0.77
***y*-axis**	3.49/1.15/1.21	3.49/1.13/1.20	3.47/0.93/1.08	3.46/0.92/1.02
***z*-axis**	7.49/6.19/6.27	7.47/6.17 /6.24	7.40/6.14/6.20	7.29/6.05/6.15

module still provides a more accurate reconstruction result than the factorization module. The results presented in Table 4 show that the average reconstruction error of the smoothing result is only slightly larger than that of the improved result. In other words, the reconstruction errors of the additional vertices interpolated in the smoothing process are almost equal to those of the original vertices in the improved result. Therefore, the ability of the smoothing process to generate a convincing result is confirmed.

3.5. **Performance evaluation by given different noise additions to tracking results.** The effect of tracking errors on the accuracy of the reconstruction results was evaluated by adding white Gaussian noise with various standard deviations to the tracking results obtained in the first module of the proposed system. Figures 10(b)-10(d) show the improved results obtained give the addition of white Gaussian noise with standard deviations of 1, 3 and 5, respectively. The corresponding estimation errors of the factorization result and the improved result are shown in Figures 10(e)-10(g), respectively. It can be seen that the reconstruction performance of both methods is degraded as the standard deviation of the Gaussian noise increases. In addition, Figures 10(b)-10(d) show that the reconstruction error in the z-axis direction (i.e., the depth direction) increases more significantly with an increasing standard deviation than the errors in the x- or y-axis directions.

Comparing the reconstruction errors shown in Figures 10(e)-10(g) with those shown in Figure 8(a) for the same $\Delta d/d$ ratio of 1/5 but no Gaussian noise addition, it is observed that the reconstruction error is sensitive to errors in the tracking results. However, unlike the estimation errors shown in Figure 8, when tracking errors are taken into consideration, the reconstruction error reduces significantly as the number of frames considered in the reconstruction process increases. This result is reasonable since the amount of feature point information included in the point matrix W increases with an increasing number of frames, and thus the factorization procedure is more robust toward the effects of noise interference.

3.6. **Evaluation of smoothing results.** Figure 11 shows the improved results and smoothing results for two facial models. Note that the images in the left-hand column show the original facial models, while those in the middle column show the improved results, and those in the right-hand column show the smoothing results. In addition, the images in the third row of Figure 11 present enlarged views of the nose and philtrum regions of the facial models, while those in the fourth row show enlarged views of the forehead region. It is seen that the boundaries of the facial features such as the forehead, nose, mouth and philtrum in the images in the right-hand column are the smoothest

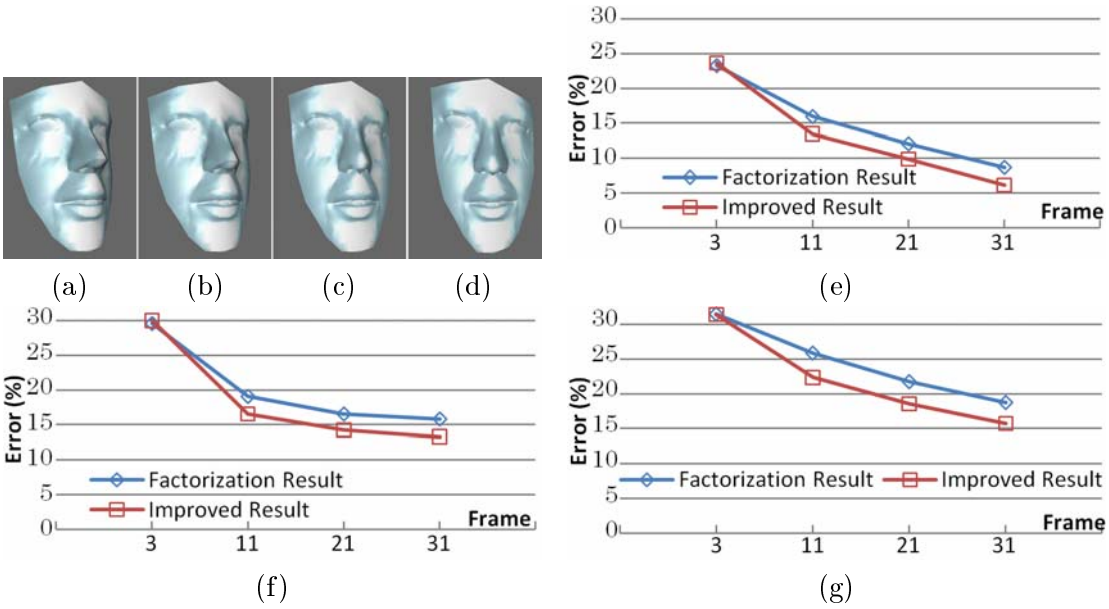

FIGURE 10. (a) Reconstruction result obtained using the improved method in absence of noise; (b)-(d) improved results obtained when white Gaussian noise is added to the tracking results with standard deviations of 1, 3 and 5, respectively; (e)-(g) reconstruction errors given white Gaussian noise addition with standard deviations of 1, 3

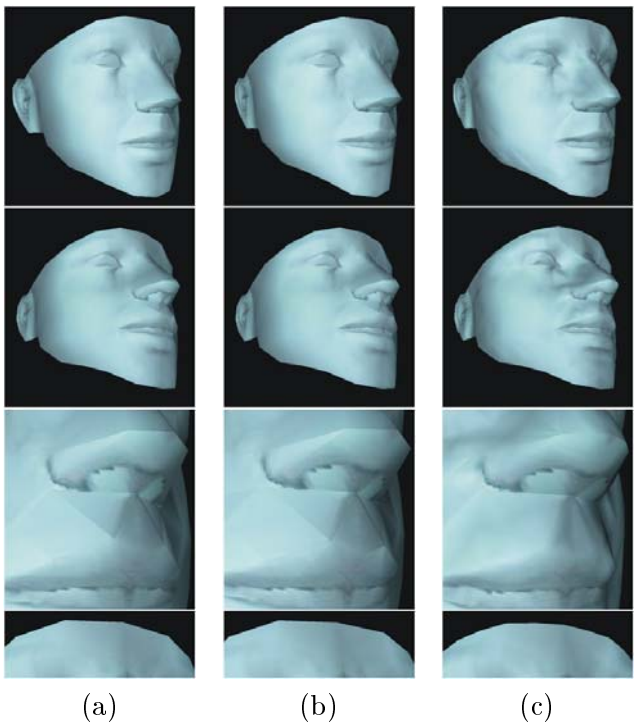

FIGURE 11. (a) Original facial model, (b) improved result and (c) smoothing

of all the images shown in the figure. In other words, the effectiveness of the proposed smoothing method is confirmed.

3.7. **3D reconstruction results for actual human faces.** Figure 12 presents the 3D facial reconstruction results obtained using the proposed method for three image sequences in which the head rotates from $-45°$ to $45°$ in the pan-rotation direction. And each image

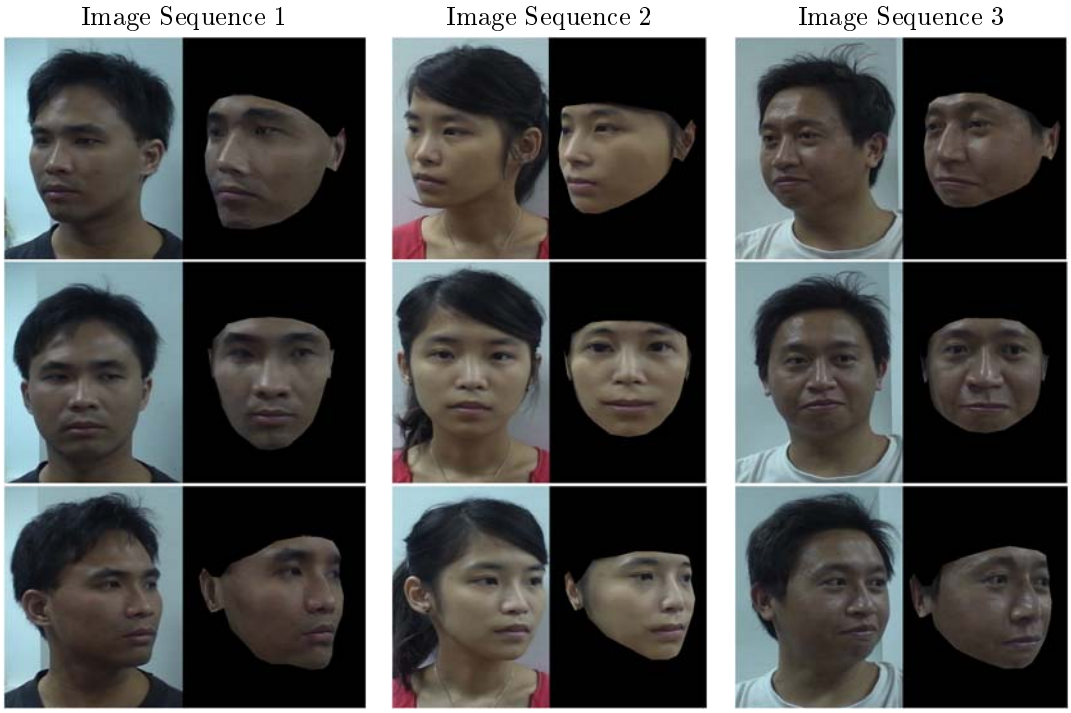

FIGURE 12. 3D facial reconstruction results. Images in odd columns are original facial images while those in even columns are corresponding 3D reconstruction results.

has a resolution of 720×480 (width×height) pixels. In performing the reconstruction process, 139 feature points were automatically located on each face using the method presented in [30] and the Lucas and Kanade optical flow technique [31] was then applied to track each feature point over the sequence of video frames. Some tracking errors were found to occur in the textureless regions of the face such as the forehead and cheek. Thus, the corresponding feature points were manually refined prior to the factorization process.

4. **Conclusions and Future Work.** This study has developed a system for reconstructing 3D faces in which a factorization method based on the orthographic projection model is used to approximate the perspective projection model. The proposed system is straightforward and computationally efficient, and avoids the need to calibrate the camera in advance. The experimental results have shown that the proposed method provides a promising technique for reconstructing 3D faces within a matter of seconds. However, some manual refinements are required to compensate for the tracking errors which occur for feature points located in the textureless regions of the face. The accuracy of the reconstruction results is significantly dependent on the tracking performance. Therefore, the number of frames used in the reconstruction process requires careful consideration. Specifically, a greater number of frames increases the amount of feature point information available to the optimization process in the factorization method, and therefore improves the accuracy of the reconstruction results.

This study has also developed a novel smoothing method based on the linear interpolation of additional 3D vertices in order to improve the smoothness appearance and realism of the reconstructed facial surface. However, linear interpolation results in a poor smoothing performance for regions of the face with large curvatures, e.g., the surface of the nose. Accordingly, a future study will explore the feasibility of using a non-linear curvature function such as the Bezier Surface function to improve the smoothing result.

The missing point problem commonly occurs in 3D reconstruction applications. Accordingly, this study has proposed a new solution in which the facial image sequence is divided into three discrete segments in accordance with the rotation angle of the head. The improved reconstruction method is applied to each individual segment and a group registration procedure based on the overlapping regions of the separate segments is then conducted in order to construct the complete 3D facial model. The experimental results have shown that an appropriate choice of the overlapping regions is essential when creating the complete model. Specifically, the chosen regions should have a small depth variance (Δd) in order to optimize the reconstruction results.

Acknowledgment. This work was supported by the National Science Council, Taiwan, under Contract NSC-100-2628-E-006-013. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Y. C. Lee, D. Terzopoulos and K. Waters, Constructing physics-based facial models of individuals, *Graphics Interface*, pp.1-8, 1993.

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

[3] C. Tomasi and T. Kanade, Shape and motion from image streams under orthography: A factorization method, *International Journal of Computer Vision*, vol.9, no.2, pp.137-154, 1992.

[4] P. Fua, Regularized bundle-adjustment to model heads from image sequences without calibration data, *International Journal of Computer Vision*, vol.38, no.2, pp.153-171, 2000.

[5] Y. Shan, Z. Liu and Z. Zhang, Model-based bundle adjustment with application to face modeling, *Proc. of IEEE Conf. on Computer Vision*, pp.644-651, 2001.

[6] H. Shum, Q. Ke and Z. Zhang, Efficient bundle adjustment with virtual key frames: A hierarchical approach to multi-frame structure from motion, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.538-543, 1999.

[7] E. Mouragnon, F. Dekeyser, P. Sayd, M. Lhuillier and M. Dhome, Real time localization and 3d reconstruction, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol.1, pp.363-370, 2006.

[8] E. Coiras, Y. Petillot and D. M. Lane, Multiresolution 3-D reconstruction from side-scan sonar images, *IEEE Transactions on Image Processing*, vol.16, no.2, pp.382-390, 2007.

[9] A. Tankus, N. Sochen and Y. Yeshurun, A new perspective [on] shape-from-shading, *Proc. of IEEE International Conf. on Computer Vision*, vol.2, pp.862-869, 2003.

[10] F. Courteille, A. Crouzila, J. D. Duroua and P. Gurdjos, 3D-spline reconstruction using shape from shading: Spline from shading, *Image and Vision Computing*, vol.26, no.4, pp.466-479, 2008.

[11] A. Braquelaire and B. Kerautret, Reconstruction of Lambertian surfaces by discrete equal height contours and regions propagation, *Image and Vision Computing*, vol.23, no.2, pp.177-189, 2005.

[12] S. Soatto, A. Yezzi and H. Jin, Tales of shape and radiance in multi-view stereo, *Proc. of IEEE Conf. on Computer Vision*, pp.974-981, 2003.

[13] A. Hertzmann and S. M. Seitz, Example-based photometric stereo: Shape reconstruction with general, varying brdfs, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.8, pp.1254-1264, 2005.

[14] A.Ghosh, S. Achutha, W. Heidrich and M. O'Toole, BRDF acquisition with basis illumination, *Proc. of IEEE International Conf. on Computer Vision*, pp.1-8, 2007.

[15] S. Nayar and N. Yasuo, Shape from focus, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.16, no.8, pp.824-831, 1994.

[16] A. Thelen, S. Frey, S. Hirsch and P. Hering, Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation, *IEEE Trans. on Image Processing*, vol.18, no.1, pp.151-157, 2009.

[17] M. T. Mahmood, A. Khan and T. S. Choi, Approximating 3D shape through bezier curve and moments in discrete cosine transform, *International Journal of Innovative Computing, Information and Control*, vol.5, no.10(A), pp.2947-2958, 2009.

[18] M. T. Mahmood, S. Shim and T. S. Choi, Shape from focus using principal component analysis in discrete wavelet transform, *Optical Engineering*, vol.48, pp.057203:1-057203:9, 2009.

[19] M. S. Muhammad and T. S. Choi, A novel method for shape from focus in microscopy using Bezier surface approximation, *Microscopy Research and Technique*, vol.73, no.2, pp.140-151, 2010.
[20] J. L. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*, MIT Press, 1992.
[21] C. Poelman and T. Kanade, A paraperspective factorization method for shape and motion recovery, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.19, no.3, pp.206-218, 1997.
[22] B. Triggs, Factorization methods for projective structure and motion, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.845-851, 1996.
[23] W. K. Tang and Y. S. Hung, A subspace method for projective reconstruction from multiple images with missing data, *Image and Vision Computing*, vol.24, no.5, pp.515-524, 2006.
[24] A. Gruber and Y. Weiss, Multibody factorization with uncertainty and missing data using the EM algorithm, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol.1, pp.707-714, 2004.
[25] P. Gargallo and P. Sturm, Bayesian 3D modeling from images using multiple depth maps, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.885-891, 2005.
[26] G. Vogiatzis, P. H. S. Torr and R. Cipolla, Multi-view stereo via volumetric graph-cuts, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol.1, pp.391-398, 2005.
[27] V. Blanz and T. Vetter, A morphable model for the synthesis of 3D faces, *Proc. of the 26th Annual Conf. on Computer Graphics and Interactive Techniques*, pp.187-194, 1999.
[28] J. Xiao, J. X. Chai and T. Kanade, A closed-form solution to non-rigid shape and motion recovery, *European Conf. on Computer Vision*, pp.573-587, 2004.
[29] J. Xiao and T. Kanade, Uncalibrated perspective reconstruction of deformable structures, *Proc. of IEEE International Conf. on Computer Vision*, vol.2, pp.1075-1082, 2005.
[30] C. T. Twu and J. J. Lien, Automatic location of facial feature points and synthesis of facial sketches using direct combined model, *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.40, no.4, pp.1158-1169, 2010.
[31] B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, *International Joint Conf. on Artificial Intelligence*, vol.3, pp.674-679, 1981.
[32] R. A. Dwyer, A faster divide-and-conquer algorithm for constructing Delaunay triangulations, *Algorithmica*, vol.2, no.1, pp.137-151, 1987.
[33] R. Hassanpour and V. Atalay, Delaunay triangulation based 3d human face modeling from uncalibrated images, *Computer Vision and Pattern Recognition Workshop, Conf. on CVPRW'04*, pp.75-75, 2004.
[34] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2008.
[35] W. H. Equitz, A new vector quantization clustering algorithm, *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol.37, no.10, pp.1568-1575, 1989.
[36] E. Hecht, *Optics*, Addison Wesley, 2003.
[37] T. Horprasert, Y. Yacoob and L. S. Davis, Computing 3-D head orientation from a monocular image sequence, *IEEE International Conf. on Automatic Face and Gesture Recognition*, pp.242-247, 1996.
[38] T. H. Wang and J. J. Lien, Rigid and non-rigid motion separation using 3D model, *Computer Vision, Graphics, and Image Processing*, pp.A2-5, 2004.
[39] J. W. Demmel, *Applied Numerical Linear Algebra*, 1st Edition, Society for Industrial and Applied Mathematics, Philadelphia, 1997.