Review

pISSN 2383-5389 / eISSN 2383-8116

Journal of International Society for Simulation Surgery 2014;1(2):57-61

Reconstructing 3-D Facial Shape Based on SR Imagine

Yu-Jin Hong[1], Jaewon Kim[2], Ig-Jae Kim[2]

[1]University of Science and Technology, HCI Robotics, Daejeon, Korea
[2]Korea Institute of Science and Technology, Imaging Media Center, Seoul, Korea

We present a robust 3D facial reconstruction method using a single image generated by face-specific super resolution technique. Based on the several consecutive frames with low resolution, we generate a single high resolution image and a three dimensional facial model based on it. To do this, we apply PME method to compute patch similarities for SR after two-phase warping according to facial attributes. Based on the SRI, we extract facial features automatically and reconstruct 3D facial model with basis which selected adaptively according to facial statistical data less than a few seconds. Thereby, we can provide the facial image of various points of view which cannot be given by a single point of view of a camera.

Key Words Face reconstruction · Super resolution · Warping · Face feature extraction · 3D reconstruction.

Received: November 27, 2014 / **Revised:** November 28, 2014 / **Accepted:** December 5, 2014
Address for correspondence: Ig-Jae Kim
Korea Institute of Science and Technology, Imaging Media Center, 5 Hwarang-ro 14-gil, Seongbuk-gu, Seoul 136-791, Korea
Tel: 82-2-958-5766, **Fax:** 82-2-958-5769, **E-mail:** drjay@kist.re.kr

Introduction

Although many recent techniques for automatic a 3D facial shape reconstruction from a captured image or the several frames of moving picture have been developed in this regard, the reconstruction result of a 3D facial shape in most of these techniques significantly depends on resolution of a corresponding image or moving picture. Thus, when a frame resolution of an image or moving picture is low, the reconstruction quality of a 3D facial model can be very poor, resulting in difficulties of using the 3D model for face recognition or an investigation of a suspect.

Typically, a reconstructed 3D facial model accurate enough to be used for face recognition or an investigation of a suspect may be obtained only if a human face in a close-up frontal position is generated with high resolution from a frame of a captured image or moving picture. Thus, the scope of using these techniques is very limited. In general, since it is difficult to acquire images of high quality by using cameras in access control devices, ATMs, CCTV monitoring devices, and so forth, the suspect's face constitutes a small percentage of the entire frame of an image or moving picture captured in a general state, a facial region usually has a low resolution, and thus, it is difficult to use captured images for an investigation of a suspect. In other words, the traditional studies of reconstructing 3D facial shape is only practical in police investigation under very specific environment condition, for example, the video is high quality and focused on the face.

In this paper, we propose a new method to overcome these previous limitations mentioned above. First, we apply the proposed SR technique on the sequential video frames which are selected manually including face area in consequence of the accuracy of motion estimation increases in case of moving objects. Second, we reconstruct a 3D facial model with the simple and fast recursive optimization technique based on adaptively selected 8 facial basis models according to facial statistical data. Thus, we can provide not only facial images of various points of view including a frontal view of image which is very useful for face recognition but cannot be provided by a single point of view camera if not captured in that view.

RELATED WORK

Super-Resolution: Previous works on SR can be roughly divided into four categories: interpolated-based (1), reconstruction based (2), classification-based (3) and learning-based (4, 5).

Among them, Liu et al. applied SR technique to enhance the quality of face image using a patch-based nonparametric Markov network to learn the relationship between local facial feature and global face. However, they generated SR image only for frontal view based on low resolution image with the same point of view (4).

3D Face Reconstruction: A number of researchers have described techniques for recovering 3D facial shape directly from images. Pighin et al. reconstructed the parametric 3D model with five view points of the faces based on feature points (6). Blanz and Vetter introduced the method of reconstructing 3D realistic face from image(s) and changed the facial appearance very naturally according to human 3D scans database (7). And with the auto generated 3D models, they exchanged face images across changes in viewpoint and illumination, for example, hairstyle (8). However, these previous techniques require a lot of time to build a 3D facial model with texture.

OVERVIEW

Our process consists of two stages. In super resolution stage, the several consecutive frames, chosen by a user, from the objective video in low resolution which is including the face area that we want to see in 3D model are processed to generate a SR image based on Probabilistic Motion Estimation (PME) method (10). In reconstruction stage, we first extract 23 facial features to align between the generated SR image and 3D generic model (Fig. 4). In the alignment process, we calculate TRS (translate, rotate, scale) matrix by minimizing the error of distance between feature points from SRI and the projected points from our 3D generic model. After alignment, we deform the shape of given 3D generic model to specific 3D model based on input image by linear combination with eight blendshape basis which are chosen adaptively by considering facial parameters. The rest of this paper is organized as follows. Section 4 formulates SR problem in a compact scenario and briefly introduces two-phase warping technique and elaborates PME method for SR in our scheme. Detail explanation of 3D face reconstruction process is given in Sec. 5.

SUPER-RESOLUTION

Input data for our SR method is low-resolution video frames, which is not necessarily consecutive, from CCTV or any kind of security camera devices. The input video frames are chosen by a user with a criterion, the inclusion of a target person's whole face. Our SR algorithm consists of three stapes. The first step is the detection of facial feature points in the input frames by STASM (9). Next step is warping process for input frames toward a reference frame using facial feature points. This is a critical process to achieve a good SR result.

Typically, a human face shows huge movements at video shooting and consequently imaging factors like zoom, translation and rotation of the face varies significantly in a capture video. Although PME method has strength in the usage of pixels uncovered by explicit motion estimation, pixels in a patch with significantly different zoom and rotation regarding a reference patch are still in low utilization. Accordingly, input images in such variation can produce a better SR result with compensat-

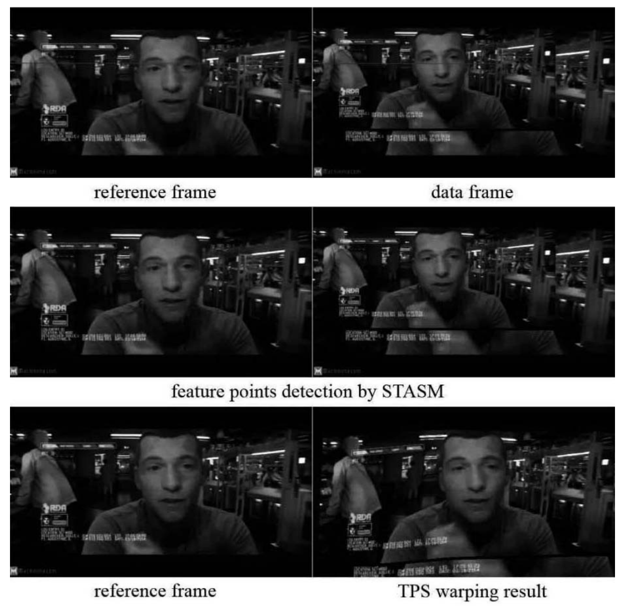

Fig. 1. Zoom compensation by TPS warping.

Fig. 2. Compensation of lip shape by PA.

ing the variation by warping. We apply two warping methods for each input frame, TPS (Thin Plate Spline) and PA (Piecewise Affine) warping. First, TPS is performed to compensate overall zoom and rotation as shown in Fig. 1. Then, PA compensates local shapes such as eyes, a nose, a lip, and so on (See Fig. 2).

The final step of our SR algorithm is applying PME to the warped images. Briefly, PME method computes patch similarities S_n between the target /P_0 and other frames /$P_{1\text{-}n}$ as follows. P_0 and $P_{1\text{-}n}$ are patches in a reference and input frames, respectively.

$$S_n = \sum_{k=1}^{m} (I_{p_0}^{k} - I_{p_n}^{k})^2 \tag{1}$$

Weights for each patch, W_n, is defined in Eq.2 regarding S_n in Eq.1 and a super-resolution image is produced by weighted average of patches in Eq.3.

$$W_n = e^{-\frac{S_n}{2\sigma^2}} \tag{2}$$

$$I_s = \frac{\sum_{j=1}^{n} I_{p_0} + I_{p_j} w_j}{\sum_{j=1}^{n} 1 + W_j} \tag{3}$$

Fig. 6 illustrates the input 117*80 in low resolution (top) and output 329*218 in SR image (middle) respectively.

RECONSTRUCTION

To reconstruct a precise 3D facial model fast, we prepare a

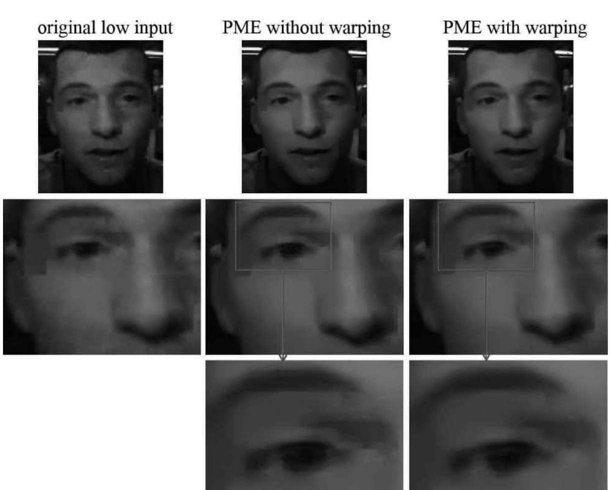

Fig. 3. Super-resolution result for 20 input frames based on PME method. PME with warping process in the third column show better quality than PME without warping process in the second column.

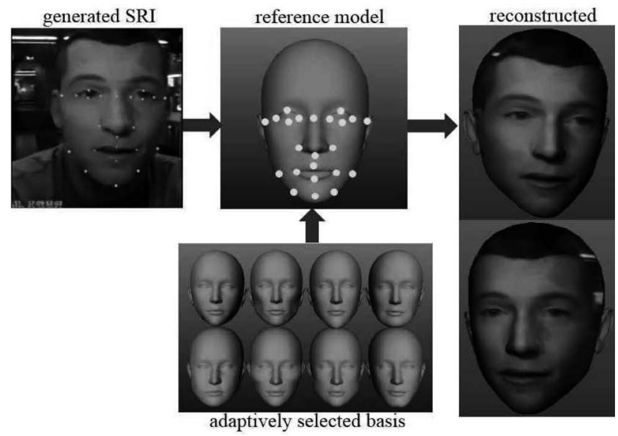

Fig. 4. 3D reconstruction flow based on a SRI and a 3D generic model with 8 blendshape basis which are selected adaptively. 23 yellow points on the generic model indicate corresponding features between 3D model and SRI. Right most figure shows the final reconstructed results.

single 3D reference face model with 3,500 vertices and a set of basis models for linear blending (11). For more precise reconstruction, we built a database of 3D facial scan data with HDI Advance (12) that consists of 100 heads of 15-73 year-old male and female. Fig. 4 illustrates the flow of reconstructing 3D face based on a SRI generated from the previous section.

In Fig. 4, we show a single 3D generic model, the eight base models selected adaptively among our scanned database according to the facial parameters extracted from SRI such as the distance between the left and right corners of the mouth, the distance between the left and right eyebrow and so on. Corresponding feature points between the generic model and the input image are shown in the figure. Once we set the models, which include a 3D reference model and 8 basis elements (models), and input image, we try to minimize the difference between the projected features from 3D reference model and the facial features from given image to find the unknown 3D transformation parameters, such as translation, rotation, and scaling, and estimated parameters for shape deformation (See Fig. 4). For finding the 3D transformation parameters, $\dot{X}_k$, to align between the 3D reference model with given face image, we can formulate the following equation.

$$\hat{X}_{k+1} = \hat{X}_k + K_k(\hat{z}_k - H\hat{P}_k) \tag{4}$$

where, $\tilde{X}_k$ is the vector which includes translation, rotation, and scaling parameters of 3D reference model at kth iteration. Z_k is 2D facial feature points from given image and H is projection matrix which projects our 3D reference model to 2D image plane. $\dot{P}_k$ is 3D facial features of the reference model. K_k is update gain. After enough iterations, we can finally get 3D transformation parameters to align the 3D reference model with 2D input image precisely. Once we get the parameters, we can utilize 2D facial feature points to deform global shape of the reference model based on 8 adaptively selected base models. When the weights are given, we deform the 3D reference model to the individual facial model by

$$v = v^0 + \sum_{i=1}^{n} \omega_i B_i \tag{5}$$

where, B_i is adaptively selected ith basis model. n is the number of basis elements. v^0 is 3D reference facial model and v is deformed 3D facial model according to the input image. ω_i is weights for linear combinations of the basis elements. We find the *n*-dimensional weight vector by minimizing

$$E = \sum_{i=1}^{N} \left| \hat{z}_j - H \sum_{i=1}^{n} \omega_i B_{i_j} \right| \tag{6}$$

Where, $\dot{Z}$ is facial features from 2D input image as described above. N is the number of feature points and n is the number of basis elements, B_{ij} is *j*th facial feature point on the ith basis model. H is already shown in equation 4.

We tested our algorithm with another 8 consecutive images and reconstructed 3D facial model as shown in Fig. 6.

We performed our reconstruction algorithm with Intel i7 single core 2.4 Ghz CPU and it takes about 2 seconds to get the final 3D model. For evaluating how precisely we can reconstruct 3D model, we measured L2 norm between the 3D feature point location on the 10 facial scan data (ground truth) and reconstructed 3D models.

Table 1 shows the average errors for each coordinate. In Fig. 5, we showed the comparison of reconstructed results with different basis elements. Leftmost face shows the ground truth data which we acquired with our face scanner, middle faces shows the result from reconstruction with predefined basis elements and right faces shows the result from reconstruction with adaptively selected basis elements. As shown the Fig. 5, the face with adaptively selected bases is much closer to the ground

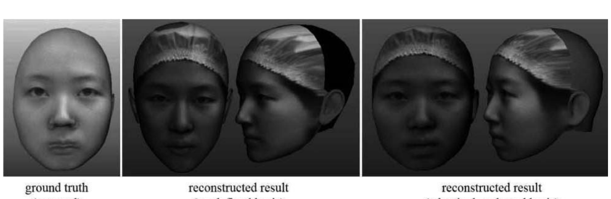

Fig. 5. This figure shows how closely we can reconstruct 3d model to the ground-truth. Ground-truth model (Left), the result of reconstructed with predefined blendshape basis (Center), Reconstructed result from adaptively selected basis (right).

Table 1. The reconstruction error

Coordinate	x	y	z
Average error (cm)	0.15227	0.004416	0.34874

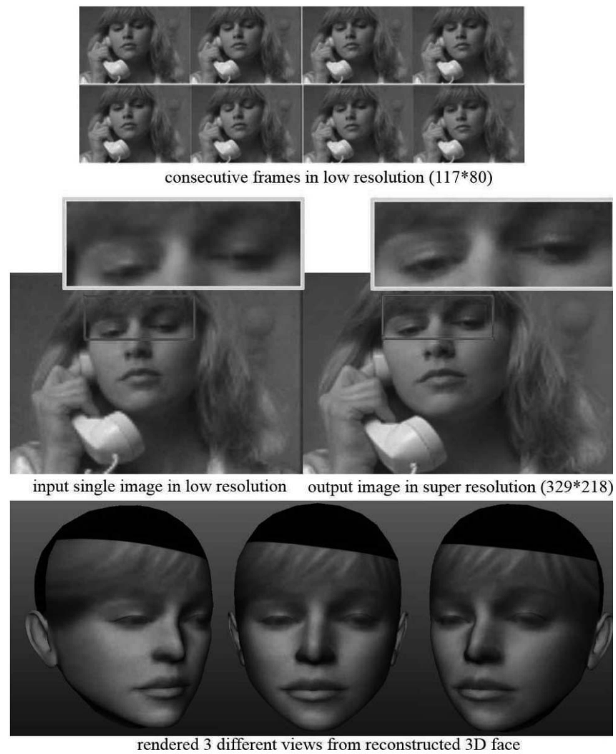

Fig. 6. Another reconstruction result with 8 consecutive low resolution frames. Input sequence (top), Generated SRI (middle), Rendered scenes from reconstructed model (bottom).

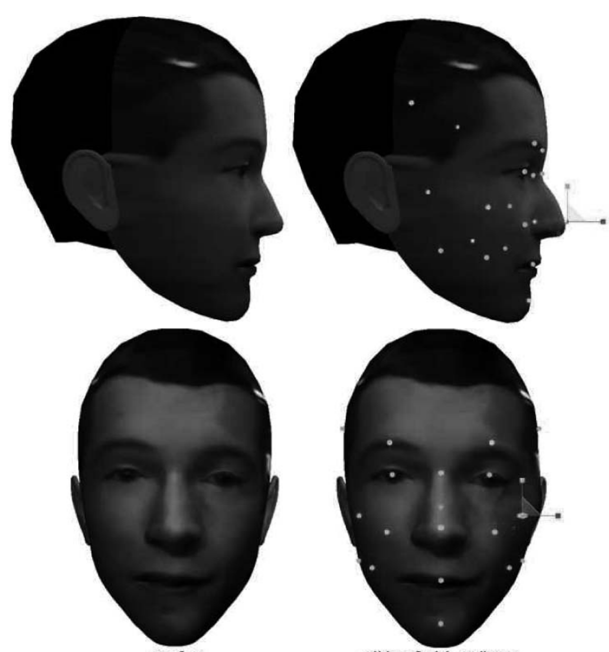

Fig. 7. Example of facial model deformation.

truth than the one with predefined (fixed) bases. Fig. 7 also shows the application of 3D editing facial features.

Once we built a 3D face model based on SR image, we provide interactive editing function to let users deform the face model as they want. For this, we predefined several joints in the face model and calculated weight values of nearby vertices from each joint in advance (joints are shown in Fig. 7 as green dots). Unlike previous control points based deformation, users can select any vertex on the face and then deform the shape by moving the selected vertex freely. In Fig. 7, we showed some results of deformation.

CONCLUSION

In this paper, we propose a robust 3D facial reconstruction method using a single image from PME based super-resolution method. Even though security cameras, such as

CCTV, are prevailing in these days, the usefulness of the images obtained by the camera would be limited at important cases, such as scene of crime, due to the low resolution of face area and the captured face angle. Our proposed method enhances the resolution of captured face images considering facial features and reconstructs 3D face model precisely for enabling us to predict the frontal face even though it is not a frontal view. For this, we apply face-specific SR method to increase the resolution and develop a fast and robust reconstruction method based on linear blending of adaptively selected facial basis elements according to the facial parameters. Hence, we get 3D reconstructed facial model and can predict the unseen part of the face.

Acknowledgements

This work was supported by KOCCA research project, "Development of recording and rendering methods based on multiple drones for performance and broadcasting".

REFERENCES

1. Xin L, Orchard MT. "New edge-directed interpolation," IEEE Trans. Image Processing, 2001:1521-1527
2. Morse BS, Schwartzwald D. "Image magnification using level-set reconstruction," IEEE Conference on Computer Vision and Pattern Recognition, 2001:333-340
3. Bouman CA, Atkins CB, Allebach JP. "Optimal image scaling using pixel classification," IEEE Conference on Image Processing, 2001: 864-867
4. Shum HY Liu C, Zhang CS. "A two-step approach to hallucinating faces: Global parametric model and local non-parametric model," IEEE Conference on Computer Vision and Pattern Recognition, 2001: 192-198
5. Yeung D, Chang H, Xiong Y. "Super-resolution through neighbor embedding," IEEE Conf. on Computer Vision and Pattern Recognition, 2004:275-282
6. Pighin F, Hecker J, Lischinski D, Szeliski R, Salesin DH. "Synthesizing realistic facial expressions from photographs," In Proceedings of SIGGRAPH, 1998:75-84
7. Blanz V, Vetter T. "A morphable model for the synthesis of 3d faces," In Proceedings of SIGGRAPH, 1999:187-194
8. Vetter T, Blanz V, Scherbaum K, Seidel HP. "Exchanging faces in images," 2004
9. Milborrow S, Nicolls F. "Locating facial features with an extended active shape model," October 2008.
10. Protter M, Elad M. "Super resolution with probabilistic motion estimation," IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 18, no. 8, 2009.
11. Kim IJ, Ko HS. "Intuitive quasi-eigen faces," ACM International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, December 2007.
12. "http://www.3d3solutions.com".