Easy and Convincing Ear Modeling for Virtual Human

Hui Zhang[1] and In Kyu Park[2]

[1] Computing Laboratory, Samsung Advanced Institute of Technology,
YONGIN 449-712, KOREA
hui.zhang@samsung.com
[2] School of Information and Communication Engineering, INHA University,
INCHEON 402-751, KOREA
pik@ieee.org

Abstract. In this paper, we describe the human ear creation in a system for 3-D photorealistic face modeling from frontal and profile images taken by an uncalibrated handheld digital camera. The ear part is segmented in profile facial image with a boundary detection algorithm, and then mapped to the created face model. Our ear modeling procedure has several advantages. (1) A highly automatic detection algorithm is proposed to locate profile ear boundary accurately and robustly despite of a variety of illumination, hair shape, gender and race. (2) Deliberately designed shape deformation algorithm is proposed to stitch the ear and head smoothly. (3) Synthesized texture is utilized and merged with input images to compensate occluded ear-nearby areas and create realistic results. Experimental results show that the created ear is smooth and convincing, improving the face appearance dramatically.

1 Introduction

A large quantities of features in facial images under different view angles are analyzed for face detection, recognition and modeling [1][2][3][4]. However, ear analysis has been received little attention although it is significant feature, especially in profile image. Ears also affect the photorealistic appearance of a 3-D head model significantly.

In [5], some parts of profile ear boundary were used to create 3-D head models. However, only the boundary between hair and ear was detected by manual operation. In fact, the automatic and robust detection of complete ear boundary is quite difficult to achieve because: (1) the ear has large shape variation for different individuals; (2) the local contrast between ear and skin parts is very weak in many cases; (3) the appearance in nearby area varies dramatically because of hair and illumination.

The generation of ear shape and texture also needs special attention because: (1) the ear shape is too complex to be modeled accurately from image information with easy user interaction; (2) obvious occlusion occurs in ear-nearby area so that some compensation is needed to create convincing texture. However, existing face modeling system [6][7][8][9][10] paid little attention to these issues and resulted with ugly ears, which weaken the realism of the created models significantly.

In this paper, we propose algorithms for creating accurate ear models easily for the user of our face modeling system. The main contribution of our work includes:

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3331, pp. 394–401, 2004.

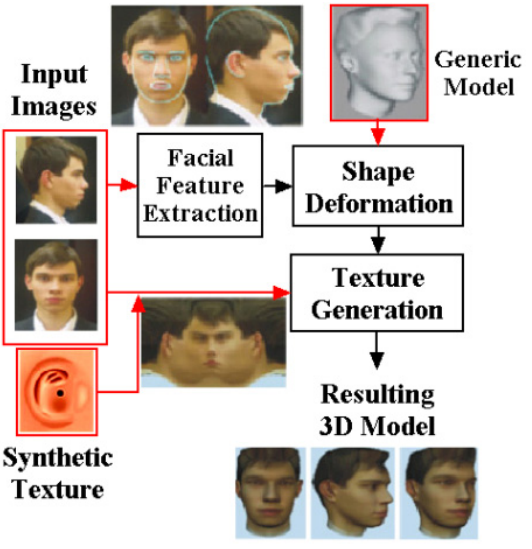


Fig. 1. Overview of the face modeling system.

- Presenting a robust and highly automatic algorithm to detect profile ear. This is achieved by exploring the different image information such as skin color and edge, and enforcing a priori knowledge about human ears with a template
- Presenting an automatic shape deformation algorithm to stitch ear and head smoothly
- Utilizing synthesized texture to compensate the occluded head areas, and combining it smoothly with the input image

The paper is organized as follows. In Section 2, the face modeling system is described briefly. Then Section 3, Section 4, and Section 5 are devoted to ear detection, ear shape deformation and ear texture generation, respectively. Finally, conclusive remarks are given in Section 6.

2 Overview of the Face Modeling System

We have proposed a system focusing on creating a highly automated procedure of a high-quality facial model generation from frontal and profile images without imposing strict conditions on picture taking conditions [11], thus to make photorealistic human head models very easily for a common PC user. As shown in Fig. 1, it takes two photos of a person (frontal and profile) and a generic head model as the inputs to produce a textured VRML model of the person's head. Two types of data are carefully combined: frontal and profile facial features. The recognition part of the system extracts the facial features robustly. A few algorithms have been developed to detect individual facial parts including eye, nose, mouth, ear, chin, and cheek. The generic head model is deformed to coincide with the detected facial features by employing Radial-Basis Functions (RBF)

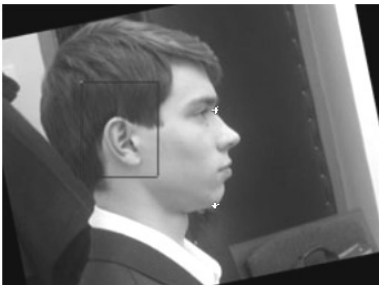


Fig. 2. Scale and orientation normalization for ear detection. Normalized image, calibration points and ear search area are shown.

interpolation [12]. The model texture is created by blending frontal and profile face photos, together with some synthesized texture.

3 Ear Detection in the Profile Image

The ear detection can be divided into two steps: profile image pre-processing, and boundary detection. They will be described in following subsections.

3.1 Profile Image Pre-processing

In this step some normalization operation is performed to facilitate the further ear detection. Two fiducial points, the nose bridge top and the chin point, are detected in profile image, and utilized as calibration points to determine the normalization. The original image is rotated to make the segment connecting them to be vertical, then scaled to make the distance between them to be a predefined value. In this stage, we also define a rectangle as the search area for the ear by statistical analysis on the relative positions between ears and the calibration points in our test images. In Fig. 2, the normalized image, the calibration points (the two white crossing points), and the search area (the rectangle with black border line) are all shown.

These two points are selected because they can be robustly detected, and they are distant enough so that the normalization is less sensitive to the detection errors. In order to detect them automatically, we first rely on skin color classification results to detect profile facial curves. Afterwards the local curvature properties of profile curve pixels are analyzed to locate profile fiducial feature points such as nose tip, nose bridge top, mouth and chin points. The algorithm is similar to that in [13].

3.2 Ear Boundary Detection

Since the local low-level clues are usually weak and erroneous in the area nearby the ear, a curve template, which represents the priori knowledge about human ears, is utilized to detect the ear boundary. The template is first matched with ear image to be translated to an initial position, and then refined by deform it to match the accurate ear boundary.

(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Ear initialization results with skin color boundary. The translated template (white smooth curve), the skin color boundary (long, irregular-shaped grey curve), and the matched partial segments (dark partial segments) are shown.

Note that the purpose of previous pre-processing step is to compensate different scale and orientation for template matching. In our implementation, the template is a 5-degree polynomial, as shown in Fig. 3 (the smooth curve with the same shape on all photos).

The 'skin color boundary' (the boundary of the face region detected with skin color classification) is used for ear initialization because it coincides with the ear boundary at some partial segment in most cases, as shown in Fig. 3. Then the problem is to find the corresponding partial segment between the template and the skin color boundary inside the search area. After the scale and orientation normalization, it can be solved with a simple curve-matching algorithm based on the similarity of curve gradient. In more detail, the two curves are preprocessed to be 4-connected, avoiding local duplication. We denote the resultant point sets as $\{q_i \in \mathbf{R}^2|_{1\le i\le N}\}$ for the template, and $\{p_j \in \mathbf{R}^2|_{1\le j\le M}\}$ for the skin color boundary. Next, two displacement arrays are constructed as $\{VQ_s = q_{a(s+1)} - q_{as}\}$ and $\{VP_t = p_{a(t+1)} - p_{at}\}$, where a is a coefficient for sampling step. Now we evaluate $l(s,t)$ as the maximum integer l that satisfies

$$\sum_{m=1}^{l} \|VQ_{s+m} - VP_{t+m}\| \le \delta, \tag{1}$$

where δ is a threshold to measure the similarity of the tangential direction at $q_{a\cdot s}$ and $p_{a\cdot t}$. The position (s_m, t_m) where $l(s,t)$ is maximum is found to give the match result as $\{q_i|_{as_m\le i\le a(s_m+l_m)}\}$ and $\{p_j|_{at_m\le j\le a(t_m+l_m)}\}$. Finally, the template is translated based on the partial segment match. Such a simple method performs very well in our experiments, as shown in Fig. 3.

Based on the initialized ear template and the matched segment on the image ear boundary, we perform contour following to deform the template to match with the whole ear boundary in the image. In more detail, we approximate the template with line segments, using an adaptive polyline fitting algorithm. Then, the line segment that has its vertex on the ear boundary is selected as the starting position of contour following. We denote this vertex as $Cont_n$, and the next vertex along the polyline as $Cont_{next}$. Afterwards the segment is rotated to a new position that gives the best match evaluation, which is defined by combining local edge strength across the ear boundary with the segment similarity along it. All the following segments after $Cont_n$ along the polyline are rotated with respective to $Cont_n$, as illustrated in Fig. 4. Finally letting $n = next$, we perform this operation iteratively to deform the whole template. This procedure

is employed twice for both directions of the polyline. The fully automatic boundary detection performs very well for significant variety of illumination, hair shape, gender, and race, as shown in Fig. 5. Even in the only one failure case (Fig. 5 (l)), the initialization and contour following are successful, but the determination of ear bottom point fails due to the earring.

In order to have a complete ear we also need to determine another half ear boundary, *i.e.* the open part of the detected curve. The completing curve part is detected semi-automatically. The user needs to manually specify the top and bottom points of internal ear boundary. Then an automatic algorithm is developed to connect them with a curve, by solving a minimal-path graph optimization problem, following the idea in [14]. Finally this inside curve is connected with outside ear boundary to form a complete ear contour. The gap between them is automatically filled by linear interpolation of ear vertices in the generic model.

4 Ear Shape Deformation

The ear shape can be determined automatically during the shape deformation stage shown in Fig. 1. That is, ear vertices are deformed together with other head vertices (Please refer to [11] for more details about head shape deformation). However, this usually gives un-satisfactory results. In fact, ear vertices are too far away from the head center, and the RBF interpolation determined by all facial feature points (eyes, nose, mouth, ears and profile fiducial points) sometime generates strange shapes for them due to extrapolation. Furthermore, the ear shape is so complex that its smoothness after RBF deformation is not ensured, and it is also difficult to be modeled accurately from image information. In our system a fixed ear shape from the generic model is used in created individual models. It is scaled to fit the specific ear size, and combined smoothly with the created heads. Such a simple strategy avoids ugly ear shape.

A key issue of this strategy is the smooth shape combination between the scaled generic ear and the deformed head model. This is solved with RBF data interpolation. We manually specify two boundaries on the ear patch. The outside boundary is the blending boundary between the deformed head model and the ear patch, while the inside boundary separates the ear patch into a facial part and a stitching-out one. The idea is to deform the facial-part ear patch to generate seamless effect across the blending boundary, while

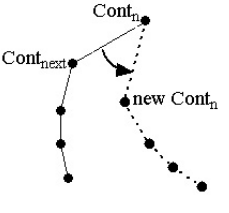


Fig. 4. The template segments before (solid line) and after (dashed line) rotation, during ear contour following.

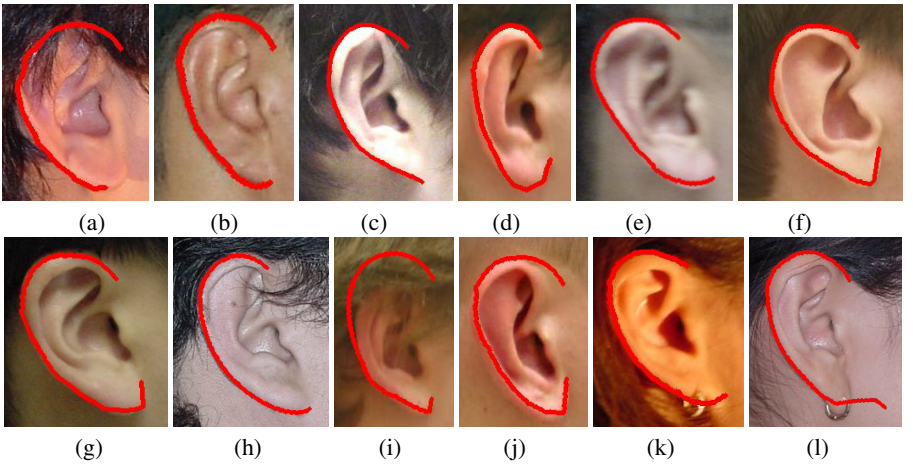
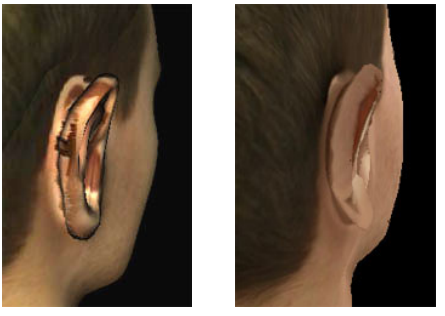


Fig. 5. Profile ear detection results.

Fig. 6. Resulting face and ear texture.

keeping the shape of stitching-out ear part. Thus the combined model will be smooth around ear area.

In more detail, we first perform a global scale transform on the generic ear patch, to make the length of the blending boundary on it match with the counterpart on a deformed head. it is also translated so that the center of blending boundary vertices coincides to that of the deformed head model. Afterwards a RBF interpolation is performed for all ear-patch vertices, utilizing the two boundaries as constrained points. The blending boundary vertices are displaced to their corresponding positions on the deformed head, while the inside boundary vertices are remained at their original positions. Finally, the stitching out part can be scaled to match with the image ear size. This fully automatic algorithm creates smooth blending results in our experiment.



(a) (b)

Fig. 7. Improved appearance with the proposed algorithm.

(a) (b)

(c) (d)

Fig. 8. Face and ear models generated by the proposed modeling system.

5 Ear Texture Generation

Special attention on ear is paid during face texture generation. First, during creating texture coordinates for all model vertices, we unfolded ear patches to create a non-overlap 2-D plane. Second, synthetic ear texture, as shown in Fig. 1, is combined with the profile image to compensate the occluded ear area, utilizing a multi-resolution spline algorithm [15]. Note that synthetic texture is first color aligned to ensure the smooth combination with input images, with an algorithm introduced in [16]. Experiments show non-overlap, realistic texture is generated, as shown in Fig. 6. The appearance improvement with the customized ear modeling algorithms can be obviously observed in Fig. 7. Final textured individual head models are shown in Fig. 8. It is observed that the ear part shows quite natural 3-D shape and matches well with the overall head model.

6 Conclusion

The image-based ear modeling issues, including ear detection in profile image, ear shape deformation and ear texture generation, are described. The detection algorithm is highly automatic, and achieves robust and accurate results for a variety of illumination, hair shape, gender and race. Both smooth shape and non-overlap, realistic texture is created for human ears, based on the detected ear boundary. The ear modeling part is integrated into a face modeling system, and improves the appearance of created face models dramatically.

References

1. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. on Pattern Analysis and Machine Intelligence **24** (2002) 34–58
2. Hjelmas, E., Low, B.: Face detection: A survey. Computer Vision and Image Understanding, **83** (2001) 236–274
3. Brunelli, R., Poggio, T.: Face recognition: Features versus templates. IEEE Trans. on Pattern Analysis and Machine Intelligence **15** (2003) 1042–1052
4. Goto, T., Lee, W., Thalmann, N.: Facial feature extraction for quick 3d face modeling. Signal Processing: Image Communication **17** (2002) 243–259
5. Lee, W., Kalra, P., Thalmann, N.: Model based face reconstruction for animation. Proc. of Multimedia Modeling (1997) 323–338
6. Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., Salesin, D.: Perceptual 3d shape descriptor: Result of core experiment. Proc. of SIGGRAPH (1998) 75–84
7. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3-d faces. Proc. of SIGGRAPH (1999) 187–191
8. Shan, Y., Liu, Z., Zhang, Z.: Model-based bundle adjustment with application to face modeling. Proc. of IEEE International Conference on Computer Vision (2001)
9. Fua, P.: Regularized bundle-adjustment to model heads from image sequences without calibration data. International Journal of Computer Vision **38** (2000) 153–171
10. Sarris, N., Grammalidis, N., Strintzis, M.: Building three dimensional head models. Graphical Models **63** (2001) 333–368
11. Park, I.K., Zhang, H., Vezhnevets, V., Choh, H.K.: Image-based photorealistic 3-d face modeling. Proc. of Sixth IEEE International Conference on Automatic Face and Gesture Recognition (2004) 49–54
12. Carr, J., Beaton, R., Cherrie, J., Mitchell, T., Fright, W., McCallum, B., Evans., T.: Reconstruction and representation of 3-d objects with radial basis functions. Proc. of SIGGRAPH (2001)
13. Akimoto, T., Suenaga, Y., Wallace, R.: Automatic creation of 3d facial models. IEEE Computer Graphics and Applications **13** (1993) 16–22
14. Mortensen, E., Barrett, W.: Intelligent scissors for image composition. Prof. of SIGGRAPH (1995) 191–198
15. Burt, P., Adelson, E.: The earth mover's distance as a metric for image retrieval. ACM Trans. on Graphic **2** (1983) 217–236
16. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Computer Graphics and Applications **21** (2001) 34–41