MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

13—12

Improvement of 3D Caricature by using Gray Scale Image in combination with Range Image

Takayuki Fujiwara [1]
School of Computer and Cognitive Sciences
Chukyo University

Hiroyasu Koshimizu [2]
School of Computer and Cognitive Sciences
Chukyo University

Abstract

We proposed a method of 3D caricature generation which is based on the automatic extraction of the facial parts for the 3D facial image. This method is likely to suffer sometimes fatal degradations in the feature extraction from a variation of the head pose (roll, pitch and yaw). Therefore, we propose a method of facial image modification by rotation (roll) which is based on the irises position extracted by Hough transform of the circle from gray scale image. We improved the facial feature extraction rate and the quality of the caricature generated by the improved 3D face data.

1 Introduction

We proposed a method of 3D facial caricature generation which is based on the automatic extraction of the facial parts for the 3D facial image [1]. This method is likely to suffer sometimes fatal degradations in the feature extraction from a variation of the head pose (roll, pitch and yaw). This system uses 3D range finder for measuring 3D facial images which are the range images and the gray scale images. Our method to generate the caricature is based on the processing of the range image, and the gray scale image is only mapped to the surface of the 3D face data. Therefore, we propose a calibration method of facial image (range image and gray scale image) by modifying the rotation (roll) on the basis of the irises position extracted by Hough transform of the circle from the gray scale image.

In section 2, the principle of the PICASSO system is summarized. In section 3, the details of a method to extract the feature points is presented. In section 4, a method to estimate the pose of the head is proposed. This method utilizes 2D gray images together with 3D facial data. Some experimental results are shown in sections 5 and 6.

2 "Mean Face" assumption and PICASSO

We are developing a facial caricaturing system PICASSO which extracts a facial characteristics of the facial image and deforms these characteristics. As for the basic principle of PICASSO, the facial caricature Q_{2D} could be generated by comparing the input face P_{2D} with the mean face S_{2D} that is defined by averaging input faces. We call this idea the "Mean face assumption" for facial caricaturing. The individuality features can be expressed by the vector ($P_{2D} - S_{2D}$), and the deformation parameter b. The facial data is measured by the range finder (VOXELAN). The difference between the input data P_{3D} and "the mean 3D face" S_{3D} can be utilized to generate the 3D facial caricature Q_{3D} as given in the eq. (1). This general idea of 3D caricaturing is shown in Fig. 1.

$$Q_{3D} = P_{3D} + b(P_{3D} - S_{3D}) \qquad (1)$$

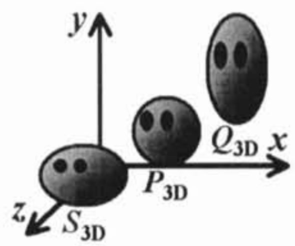


Fig. 1 General idea of the 3D facial caricaturing

3 Automatic detection of facial feature points

Feature points are extracted from the range image. As shown in Fig. 2, the nose vertex is extracted by investigating the value of neighborhood gradient to the candidates which are extracted as the maximum peaks from the range image. Some local feature points (peaks) are found by searching above (2 points) and below (5 points) from the nose vertex. An example is shown in Fig. 3.

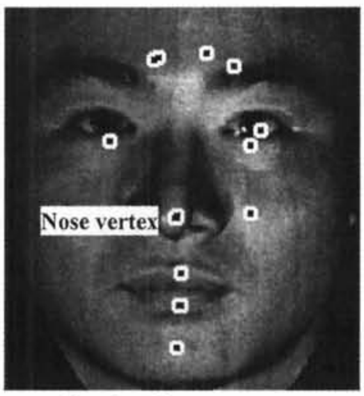


Fig. 2 Nose vertex

Fig. 3 Local features

3.1 Definition of each facial parts region

Each facial parts region is defined by the local peaks and the edge points of the face which are found by scanning the range image in the horizontal direction from the local peaks. These edge points are detected by thresholding from the gradient image of the depth data. For the nose region,

[1] Address: 101 Tokodachi, Kaizu-cho, Toyota City, Aichi, 470-0393, JAPAN. E-mail: fuji@koshi-lab.sccs.chukyo-u.ac.jp
[2] Address: 101 Tokodachi, Kaizu-cho, Toyota City, Aichi, 470-0393, JAPAN. E-mail: hiroyasu@sccs.chukyo-u.ac.jp

the coordinates of the top and bottom boundaries are defined by y_u^1 and y_b^1, respectively. The nose region is defined by eq.(2), and an example is shown in Fig. 4. For the mouth region, the top and bottom are defined by y_b^1 and y_b^5. The mouth region is defined by eq.(3), and an example is shown also in Fig. 4. For the eye region, the top and bottom are y_u^2 and $y_u^{0.5}$. The eye region is defined by eq.(4), and an example is shown also in Fig. 4. The facial feature points are generated by using these regions.

The nose vertex is the origin, and the range image is scanned to find the contour of the face region along x-axis, y-axis, $y=x$, $y=-x$, $y=2x$, $y=-2x$, $y=1/2$ and $y=-1/2$. Because the position of the upward point of the eye region and the top of $x=y$ and $x=-y$ are very close, they are not scanned. And, because there was a neck, 3 points of the bottoms (y-axis, $y=2x$ and $y=-2x$) were extracted by thresholding. An example is shown in Fig. 5.

$$\{(x,y) \mid x_{(L,n)} \le x \le x_{(R,n)}, y_u^{(1)} \le y \le y_b^{(1)}\} \quad (2)$$

$$\{(x,y) \mid x_{(L,m)} \le x \le x_{(R,m)}, y_b^{(1)} \le y \le y_b^{(5)}\} \quad (3)$$

$$\{(x,y) \mid x_{(L,e)} \le x \le x_{(R,e)}, y_u^{(2)} \le y \le y_u^{(0.5)}\} \quad (4)$$

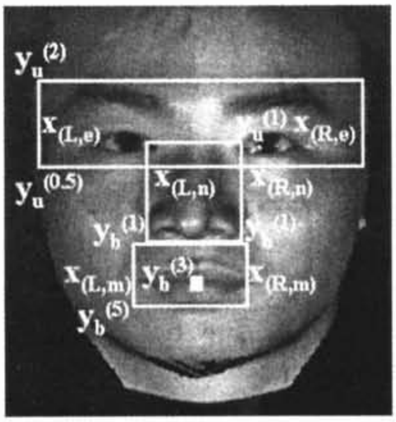


Fig. 4 Facial parts regions

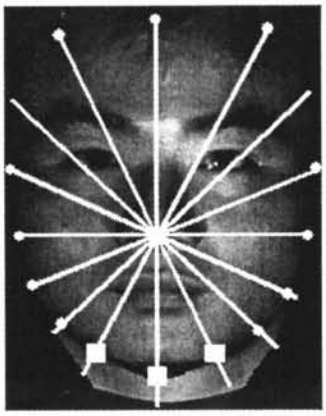


Fig. 5 Example of the face contour

3.3 Generation of the triangular patches

The wire-frame model of the 3D face was constructed by using 68 triangular patches, which are defined on 39 feature points including 8 points at the corners of the image. Fig. 6 shows an example of the generated triangular patches.

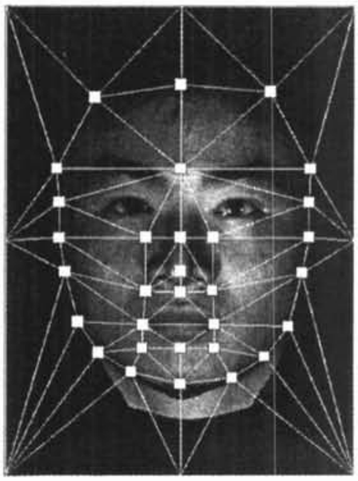


Fig. 6 Example of the triangular patches

4 Modification by the Roll rotation

We propose a method for the extraction of the roll axis of the head. Fig. 7 shows the variation of the head pose (roll, pitch and yaw). The roll angle θ of the head pose was extracted by the following method based on a pair of irises because the contour and the shape of the face [2][3][4][5] are not always symmetric in general.

A pair of irises was extracted by Hough transform for circle [6] from the gray images that are measured by the range finder together with the range images. The head axis for the roll parameter is defined by the normal of the line passing through a pair of irises. As shown in Fig. 8, the nose vertex and other feature points are detected by scanning the range image along the extracted head roll axis. The roll angle θ was given by

$$\theta = \tan^{-1}\{(y_2 - y_1)/(x_2 - x_1)\} \quad (5)$$

where (x_1, y_1) and (x_2, y_2) are the center of the irises. Fig. 8(c) shows the facial image that is modified by the roll rotation.

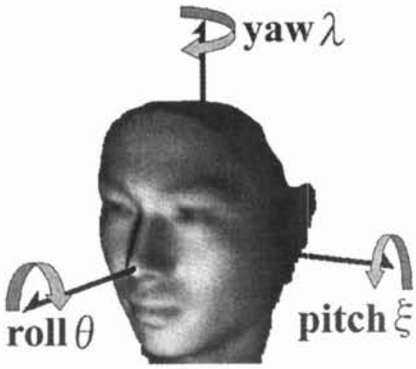
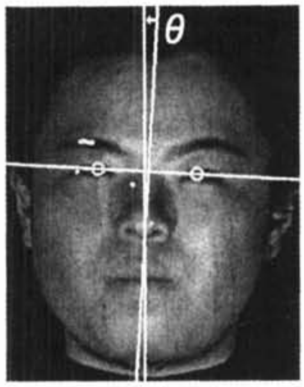


Fig. 7 Pose of the head



(a) Irises and the head axis

(b) Original image (c) Modified image

Fig. 8 Modification by the roll rotation

5 Experiments and Discussions

5.1 Extraction rate of the feature points

The improvement of the extraction success rate was evaluated by comparing the results of original images. Since the proposed method for the extraction of feature points is based on 8 local features including the nose vertex on the Y-axis, the improvement of the roll rotation of the input image given in section 4 could be effective in the feature extraction introduced in section 3. Table 1 shows the experimental results of 6 features among 8 local feature points, and Fig. 9 shows the corresponding positions of these 6 features. As known from these results, 10-20% local feature points were successfully improved.

Table 1 Accuracy of the extraction (22 images)

	Original image		Modified image		Improvement factor (%)
	Success	Successful rate	Success	Successful rate	
①	16	72.3%	16	72.3%	0
②	13	59.1%	19	86.3%	27.2
③	16	72.3%	17	77.3%	5
④	13	59.1%	13	59.1%	0
⑤	10	45.5%	16	72.3%	26.8
⑥	17	77.3%	20	90.9%	13.6
All	14.2	64.4%	16.8	76.5%	12.1

Fig. 9 Local features on Y-axis

5.2 Generation of the mean face

The mean faces are generated by using the triangular patches of many input faces, and the range image of the mean face is generated from the range images of the input faces by using triangular patches constructed on the feature points of the mean face. The triangular patches of the mean face (x_i^s, y_i^s) is given by

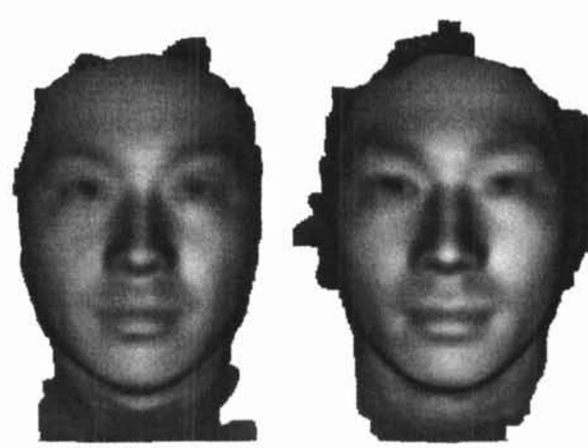
$$x_i^S = \frac{1}{N}\sum_{j=1}^{N} x_i^j, \quad y_i^{(S)} = \frac{1}{N}\sum_{j=1}^{N} y_i^j \tag{6}$$

where (x_i^j, y_i^j) is a node of the triangular patches of the input faces, and each facial parts region is corresponded to those of the mean face and of the input faces by Affine transform. $z^S{}_{(x,y)}$ is the range image of the mean face which is given by

$$z_{(x,y)}^S = \frac{1}{N}\sum_{j=1}^{N} z_{(x,y)}^j \tag{7}$$

where $z^j{}_{(x,y)}$ is the range images of the input faces.

Fig. 10 shows the results of the mean faces that are generated from original images of Fig. 10(a) and modified images of Fig. 10(b). The texture around both eye regions of Fig. 10(a) is sharper than Fig. 10(b) for example.



(a) Generation from original images (b) Generation from modified image

Fig. 10 3D mean faces

6 Generation of the 3D Caricature

The caricature is generated by using the triangular patches of the input face and the mean face, and the range image of the caricature is generated from the input face and the mean face by using the triangular patches of the caricature. The triangular patches of the caricature (x_i^Q, y_i^Q) is given by

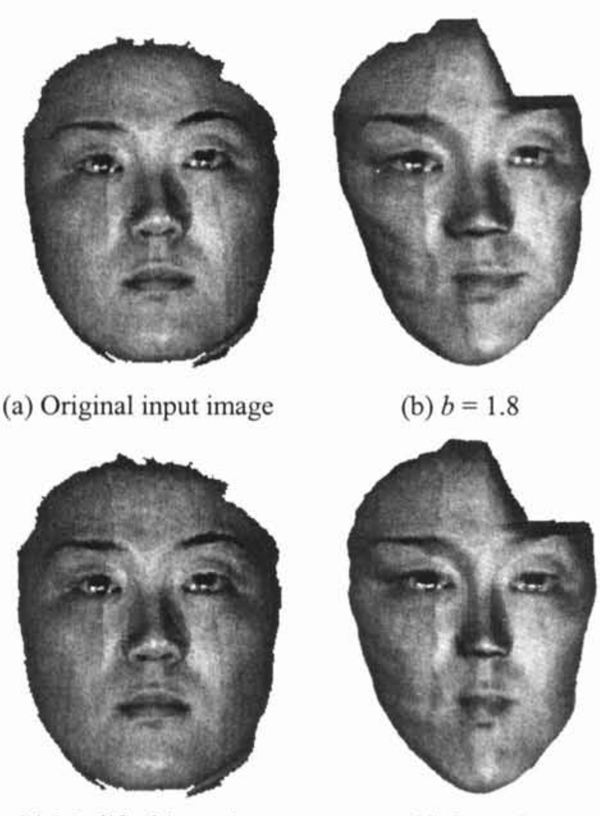
$$x_i^Q = x_i^P + b(x_i^P - x_i^S) \quad (8)$$

$$y_i^Q = y_i^P + b(y_i^P - y_i^S) \quad (9)$$

where b is exaggeration rate, and the range image of the caricature $z^Q_{(x,y)}$ is given by

$$z^Q_{(x,y)} = z^P_{(x,y)} + b(z^P_{(x,y)} - z^S_{(x,y)}). \quad (10)$$

Fig. 11 shows the results of the caricatures generated by the our new method. Fig. 11(b) is generated from the original images, and Fig. 11(d) is generated from modified images. The caricature of Fig. 11(b) leans to right side, and the mouth moves to left side because the inclination of the face is exaggerated. On the other hand, the result in Fig. 11(d) is a typical example of the improved caricature generated from the modified input face of Fig. 11(c), and each facial part of this caricature is properly exaggerated. Comparing with Fig. 11(b), it is clearly known that the shape of the nose and the size of the mouth are adequately exaggerated in Fig. 11(d). This is because the mean face in this procedure was made from many male input faces, and the features of the female nose and mouth are successfully exaggerated to be more feminine.



(a) Original input image (b) $b = 1.8$

(c) Modified input image (d) $b = 1.8$

Fig. 11 Result of the caricature generation

7 Conclusions

We have proposed a method for the estimation of the head pose in the roll rotation from the range images in combination with the gray images. Because our method used a pair of irises that are the most invariant part of the face, the caricatures were successfully improved. Our final goal however is to estimate other pitch and yaw of the head pose in order to realize the better human interface, and the recognition of the facial parts such as irises should be improved as the coming subjects of this research.

8 Acknowledgement

We would like to express many thanks for giving facial caricatures and helpful discussion to Mr. Katsuhiko Yoshida of the Design Dep. of SUNTORY and Mr. You Nakagawa of Tokyo Shinbun. This paper was partially supported by Ministry of Education, High-Tech Research Center Promotion, and IMS HUTOP Research Promotion.

References

[1] T. Fujiwara, M. Tominaga, K. Kato, K. Murakami and H. Koshimizu, "On the Detection of Feature Points of 3D Facial Image and Its Application to 3D Facial Caricature", Proc. MVA2000, No. 3-21, pp. 111-114 (2000)

[2] S. Katahara, and M. Aoki, "Face Parts Extraction Using Gradient Direction Symmetry", Proc. MIRU98, pp. II-411-416 (1998)

[3] Y. Furuichi, S. Ando, and M. Nakajima, "Human Verification Ushing 3-D Face Shape", CVIM 132-12, pp. 83-88 (2002)

[4] Y. Araki, N. Shimada and Y. Shirai, "Detection of Faces of Various Directions and Estimation of Face Direction in Complex Backgrounds", PRMU 2001-217, pp. 87-94 (2002)

[5] M. Yasumoto, H. Hongo, H. Watanabe, K. Yamamoto and H. Koshimizu, "Face Recognition and Face Direction Estimation by the Cooperation of Multiple Cameras", Trans Inst Electron Inform Communi Engnr Jpn, D-II, Vol.J84-D-II, No.8, pp. 1772-1780 (2001)

[6] T. Yamaguchi, M. Tominaga and H. Koshimizu, "Estimation and modeling of eye movement by facial image processing and its application", Proc. ISHF2001, PS22-7-2, pp. 298-303 (2001)