Real-Time Normalization and Feature Extraction of 3D Face Data Using Curvature Characteristics

Tae Kyun Kim, Seok Cheol Kee and Sang Ryong Kim
Human Computer Interaction Lab.
Samsung Advanced Institute of Technology, KOREA
E-mail taekyun@sait.samsung.co.kr

Abstract

3D data has many advantages over image data. It is robust to illumination change and does not have a scaling problem caused by distance of an object. Also it can be viewed at various angles. Nowadays with advance of 3D capturing tools, laser scanners and high-speed stereo machines, interests in 3D data processing have been increased. But the number of 3D face recognition approaching is still little. The method of normalization and real-time feature extraction of 3D face data(range data) is presented in this paper.

The step of normalization of range data is performed first using the symmetry of the defined facial section pattern and characteristics of changes of the pattern according to head rotations. Normalization of the data for head rotations can not only give strong constraints on the positions of facial features but also reduce the dimension of parameters used in the deformable template matching which is done at the next step.

Facial features are found in a range image, which is obtained by projection of the normalized range data, using the deformable templates of eyes, nose and mouth. For reliable feature detection surface curvatures, which can represent a local surface shape, are used in this step. We define the energy functions of each template and the conditions of major control points using curvature information. Finally the facial features are positioned in three-dimensional space by back-mapping to the original range data. The back-mapping is the inverse process of getting the facial range image.

1 Introduction

Normalization and feature extraction of facial data is basic technology for human recognition and animation that are being given much attention in the field of human-computer interaction. Although many researches based on images are ongoing, approaches using 3D data have been increased to overcome shortcomings of image based approaches as 3D becomes a major concern and 3D input devices are getting popular. For the study of reliable face recognition[8,9,11,12] and automatic building of an animation face[1,2,3], the method of normalization and feature detection of facial range data is presented.

For reliable 3D face recognition, normalization of input data should be done especially considering 3D rotations. 3D rotations were not compensated in the previous approaches to 3D face recognition [2,3,8,9,11,12], which are some for face classification and others for facial feature extraction. And they do not seem to be robust to even a small rotation. First we tried to estimate rotation by finding more than three feature points in a plane and calculating normal vectors of the plane. But it was difficult to estimate correctly because of the smoothness of facial surface and complex rotations. Therefore we turned to the approach using global characteristics of facial range data. The efficient energy functions being related with the symmetry of data was defined and the range data was rotated to find the pose that minimizes the energy. The compensation for both in-plane and horizontal rotations is followed by the compensation for vertical rotation.

Let me explain the previous feature extraction methods more carefully. The method of finding features one by one for whole range image using only one-side variation of depth [1,2], probably involve large errors for non-uniform and noisy range data. Real 3D face data is non-uniform and noisy because specular reflection can occur or 3D data can be purposely compressed. For more reliable detection positions of features need to be restricted. In this paper, the positions of eye, nose and mouth are initialized by the defined feature line model of normalized data. And the features are grouped with the deformable templates so that probability of large errors is minimized.

In the research[3], LOG edge filter for a range image was used to detect feature contours, but it is not sufficient to find contours accurately because variations of the range data in eye contours and outlines of mouth are not as large as those of facial image. Surface curvatures were used here to define the conditions of feature points in a range image. More features could be found reliably using this information.

The proposed method largely consists of two steps, normalization of data and feature extraction of the normalized data. The normalization step is explained in Section 2 and the extraction step is explained in Section 3. In Section 2 a specific section pattern and a feature line model are defined for the normalization of range data, and the method to compensate in-plane and horizontal rotations of a head is introduced using the energy function which has minimum value when a face is frontal. And the method to compensate vertical rotation is explained using the defined feature line model from the pre-compensated data about in-plane and horizontal rotations. The concept of surface curvatures and the initial locating of the deformable templates are explained in Section 3. And the methods to extract feature points of eye, nose and mouth from the normalized range data are described. Finally, other passive feature points are found and all detected feature points are mapped to 3D space. The result of robust feature extraction about both various face shapes and rotations is showed in Section 4.

2 Normalization of Facial Range Data

2.1 Section Pattern and Feature Line Model for Normalization about Head Rotation

Generally a human head is rotated being accompanied with complex combinations of 3D rotations(in-plane, horizontal and vertical), so data can be normalized when all of the effects due to each independent rotation and mutual interactions are analyzed. Through inspecting change of data according to 3D rotation, the following section pattern and the feature line model are defined.

After selecting the nodes(3D points) which are in the range of a constant depth difference(Z_{th}) to the highest point(which have max z coord.) of facial range data and projecting them into a XY plane, we obtain the section as shown in Figure 1. The ideal pattern for compensation of a rotation is not acquired using a fixed depth difference because of various face shapes. So we select a few node sets in small variations of the depth difference and calculate section areas of all the selected node sets. We get the section pattern whose area is the same as that of a reference. By minimizing the energy function E_p in Eq.(1) the Z_{th} and the section pattern P are obtained.

$$E_p = (\int P(x,y)ds - refarea)^2 \quad (1)$$

$$P(x,y) = \begin{cases} 1 & if\ Z(x,y) \geq Z(x_t,y_t) - Z_{th} \\ 0 & otherwise \end{cases} \quad (2)$$

($Z(x,y)$: z coord. of range data, $Z(x_t,y_t)$: maximum z coord., *refarea* and Z_{th} : constant)

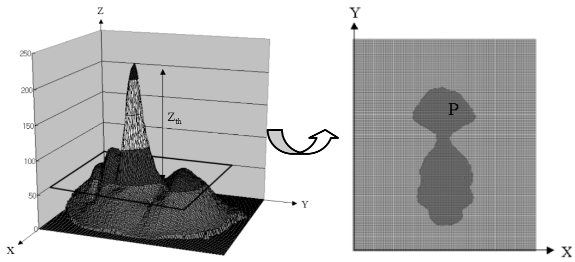

Figure 1. Section pattern from a facial range data. Z coord. of facial range data is exponentially transformed.

From the acquired pattern P the feature line model is defined in Figure 2. If in-plane and horizontal rotations were compensated, each feature line would put around the positions of eyes, eyebrows and a mouth center line.

Eyebrow line (y1) : the last local maxima
Eye line (y2) : the last local minima
Mouth center line (y3) : the first local minnima

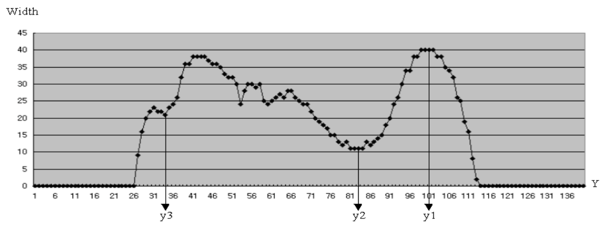

Figure 2. Graph of the pattern width.

The section pattern and the feature lines that are defined using depth characteristics of a face do not lose their peculiarity about various face shapes and noise. Figure 3 shows an example of the acquired section pattern and the feature lines.

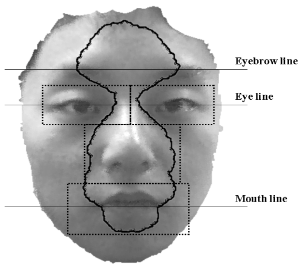

Figure 3. Feature line model from the pattern P

2.2 Compensation for In-plane and Horizontal Rotations

Energy function E_R is defined from the section pattern in Section 2.1. The E_R has a minimum value when a face is frontal with the pattern becoming symmetrical to Y and Z axes.

$$E_R = \int_y ((U(y) - x_t) - (x_t - L(y))^2 dy \quad (3)$$

$$U(y) = \max \arg .\, x \ of \ \{(x, y) \mid P(x, y) = 1\} \quad (4)$$

$$L(y) = \min \arg .\, x \ of \ \{(x, y) \mid P(x, y) = 1\} \quad (5)$$

(x_t : x coord. of a node that has maximum z coord.)

The E_R value depends on in-plane rotation(Z-axis) and horizontal rotation(Y-axis) largely but it doesn't change about vertical rotation(X-axis) of a head. Rotating facial range data with all possible combinations of in-plane and horizontal rotations, we calculate the E_R . By finding the rotation angles which make the function of E_R minimum in-plane and horizontal rotations of a head can be compensated. Figure 4 shows changes of the energy E_R for in-plane and horizontal rotations.

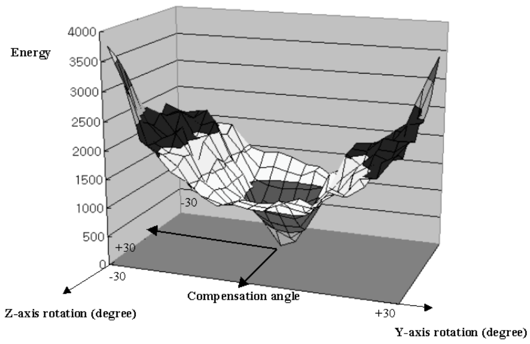

Figure 4. E_R for in-plane and horizontal rotations

The steepest descent method can be used to find the global minimum point instead of searching whole range of angles because the global minimum is clearly distinguishable as shown in Figure 4.

$$\frac{\partial \theta}{\partial t} = -\frac{\partial E_R}{\partial \theta} \qquad (6)$$

(θ : in-plane rotation angle or horizontal rotation angle)

2.3 Compensation for Vertical Rotation

After in-plane and horizontal rotations are compensated in Section 2.2, the feature lines defined in Section 2.1 are located around the positions of eye, eyebrow lines and a mouth center line. These characteristics do not depend on vertical rotation of a face. Height interval(along Y-axis) between two lines changes like Figure 5 for the change of a vertical rotation angle. That is, the interval becomes the maximum when a face is frontal and it decreases when the face is upward or downward. By finding an angle that makes the interval of lines maximum, vertical rotation can be normalized.

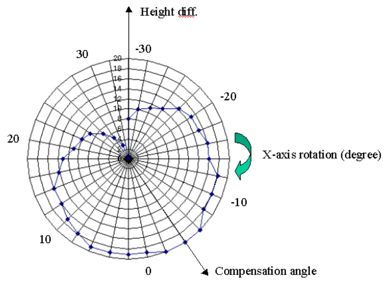

Figure 5. Interval between two feature lines for vertical rotation angle

3 Feature Extraction of Normalized Facial Range Data

3.1 Principle Curvatures and Initial Positioning of Deformable Templates

A range image is obtained from the normalized facial range data and facial feature points are found in the range image. The change of brightness in outlines of eyes and a mouth is not large in a range image compared with an image of CCD camera. Edge information obtained by general image processing techniques is not sufficient to detect facial feature points in a range image. Therefore the proposed method uses curvature information which describes characteristics of a surface.

Principle curvature(k_1,k_2) is the maximum or the minimum of curvatures of normal sections. Using these parameters shapes and slope speeds of a surface can be calculated. Figure 6 shows the segmentation results of facial range images using the principle curvatures k_1, k_2.

Figure 6. Segmentation results of facial range images based on principle curvatures (k_1,k_2)

(black: $k_1 > 0.5$, gray: $k_1 < 0$ and $-0.3 < k_2 < 0$, white: $k_1 < 0$ and , $k_2 < -0.3$)

Positions of important regions around eyes, nose and mouth are initialized using the feature line model used to compensate rotations and feature points are extracted in those regions. The Initial positions are illustrated in Figure 3. Four rectangular regions are obtained using the feature line model and general geometry of a face. In the constrained regions all of the feature points on the deformable templates are found by using energy minimization techniques with principle curvatures or searching the most reliable point first. The proposed initialization method is robust to non-uniform and noisy data. Also grouping of feature points with deformable templates makes the detection algorithm more robust.

3.2 Mouth Filter

The filter to find feature points of mouth uses the deformable template defined in Figure 7. Assuming an inner contour of lip is a part of an ellipse, the deformable template consists of six control parameters, Cx, Cy, a, L, B and T. Four parameters(Cx, Cy, a , L) can be found using the energy function E_{inner} in Eq.(7).

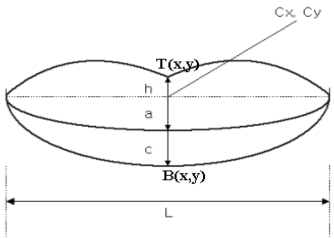

Figure 7. Deformable template for mouth

$$E_{inner} = -\frac{1}{|R_i|}\int_{R_i} \psi_i(x,y)ds \qquad (7)$$

$$\psi_i(x,y) = \begin{cases} 1 & if\ k_1(x,y) \geq 1 \text{ and } 0 \leq k_2(x,y) \leq 0.1 \\ 0 & otherwise \end{cases} \qquad (8)$$

(k_1, k_2 : maximum and minimum principle curvatures, R_i: inner contour)

$\psi_i(x,y)$ means the region which is a concave surface(k_1>0, k_2>0) and whose maximum slope is fast and minimum slope is slow. The center position of an inner ellipse is restricted using the mouth center line and the control parameters are found to minimize the energy function E_{inner}.

Upper and lower contours are also elliptical lines and defined using two control points (T,B) and the determined inner contour parameters. See the reference [4] for detailed line equations. Top point(T) and bottom point(B) are defined using curvature information like followings.

$$\begin{aligned} T(x,y) &= \{(x,y) \mid \partial Z(x,y)/\partial y\ is\ \min, x = Cx, Cy \leq y < Cy + y_1\} \\ B(x,y) &= \{(x,y) \mid \partial Z(x,y)/\partial y\ is\ \max, x = Cx, Cy - y_2 \leq y < Cy\} \end{aligned} \qquad (9)$$

(y_1,y_2: constant)

3.3 Nose Filter

The filter for detecting feature points of nose uses the deformable templates defined in Figure 8. The bottom contour of nose consists of the center half ellipse defined with Cx, Cy, L, h and two straight lines connected to C1, C2 and two curves connected to C3, C4. Four parameters, Cx, Cy, L and h are calculated by finding a bottom point and two nostril points of nose. C1,C2,C3 and C4 are directly searched using curvature changes and relative geometry similarly to Eq. (9).

C5 is the highest point of the normalized facial data. From the C5 to a brow, points with a slow curvature change are found as nose ridge points. The angle of a nose ridge line, θ , is the average of angles between the ridge points and C5. Ec is an intersection point of the nose ridge line and the eye line.

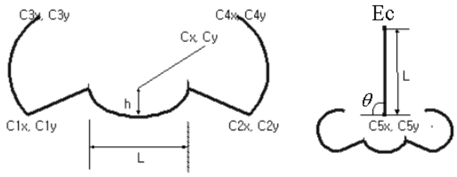

Figure 8. Deformable template for nose

3.4 Eye Filter

The ellipse parameters defined in Figure 9 are initialized using the eye line and the point Ec defined in Section 3.3. It is possible because sizes and intervals of two eyes of normal people are not largely different and a scale of range data is given. But it is still difficult to find eye contours because depth variations in eyes are relatively small compared with other regions in a face. Using the energy functions in Eq.(10) deformable template matching is performed. $E_{surface}$ means the area of a concave region inside an eye ellipse and E_{line} represents the length of an eye contour that is located in a convex region.

$$E_{eye} = E_{surface} + E_{line} \qquad (10)$$

$$E_{surface} = -\frac{c_1}{|R_{eye}|}\int_{R_{eye}} \psi_s(x,y)dA \qquad (11)$$

$$E_{line} = -\frac{c_2}{|\partial R_{eye}|}\int_{\partial R_{eye}} \psi_e(x,y)ds \qquad (12)$$

$$\psi_s(x,y) = \begin{cases} 1 & if\ 0.5 < k_1(x,y) \\ 0 & otherwise \end{cases} \qquad (13)$$

$$\psi_e(x,y) = \begin{cases} 1 & if\ k_1(x,y) < 0\ and\ k_2(x,y) \leq -0.3 \\ 0 & otherwise \end{cases} \qquad (14)$$

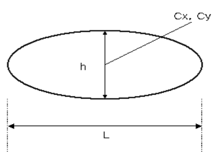

Figure 9. Deformable template for eye

3.5 Other Passive Features and Back-mapping

Other passive points are located using the major feature points which are on the contours of eyes, nose and mouth, the boundary of a face detected by boundary following algorithm in a range image and the feature line model. Passive points are intersections of horizontal and vertical lines passing through the major feature points, the feature lines and the boundary of a face.

Feature points in 3D space can be finally found by back-mapping of all feature points in a range image to original facial range data. The back-mapping is the inverse operation of linear scaling that is executed when projecting range data to a range image. Linear interpolation algorithm was executed concurrently with the inverse operation.

4 Result

To verify the proposed algorithm, feature points were extracted for various 3D face data. 119 feature points(Eye(44), Nose(16), Mouth(24), passive(35)) were found for each face. In this experiment facial range data were obtained using the 3 dimensional scanning equipment, RealScan 3D by Real 3D Inc.. About 20,000~50,000 vertex were obtained for not only small rotated faces but also largely rotated faces. Some of data were intentionally compressed using the RealScan 3D software.

Table 1 shows estimation results of in-plane and horizontal rotations of the obtained 3D face data. The data were artificially rotated about X-axis to show the independence of the algorithm in Section 2.2 to vertical rotation.

Table 1. Estimation of in-plane and horizontal rotations

True angle		Estimated angle (horizontal, in-plane)				
Man1	4,0	4,0	4,0	3,0	3,1	4,0
Man2	-2,0	-5,3	-4,0	-2,-1	-1,2	0,1
Man3	-2,1	-9,0	-3,0	-3,0	-2,1	0,1
Man4	1,0	1,-1	0,1	1,0	3,0	3,0
Man5	-2,2	3,-10	-2,1	-2,1	-2,2	-1,2
Wom1	-6,-1	-6,-2	-6,-1	-5,-1	-8,-2	-8,-1
Wom2	4,-10	4,-11	3,-12	4,-10	3,-9	4,-10
Wom3	13,1	13,1	14,1	14,1	13,-1	13,3
Vertical rotation		-20	-10	0	10	20

(unit : degree)

Average error is 0.9 for the estimation of in-plane rotation and 1.1 for the estimation of horizontal rotation. When vertical rotation is negatively large, the ideal section pattern can not be acquired and errors occur. After compensation of the data about in-plane and horizontal rotations, vertical rotation was estimated.

Table 2. Estimation of vertical rotaion

	True	Estimated		True	Estimated
Man1	-3	-2	Man5	3	2
Man2	-5	-6	Wom1	-3	-3
Man3	-10	-10	Wom2	7	5
Man4	1	0	Wom3	-1	-1

(unit : degree)

Average error is 0.75 for vertical rotation. In the proposed normalization step accuracy is under 2 degrees and execution time for global search is 3(s) for Pentium III PC. To extract feature points exactly from these rotated data the steps of normalization were essential.

From the normalized data feature points were detected in 3D space. Figure 10 shows examples of extraction results. The points on the outlines of eyes and mouth were also located exactly. Our algorithm could find feature points reliably in the range of 30 degree rotations for each axis. Average errors in feature location are shown in Table 3.

Table 3. Average errors in feature location

			Eye(8)	Nose(7)	Mouth(7)
Avg. Error (mm)	Slightly rotated faces	Man1	2.65	0.95	0.95
		Man2	0.836	0.74	0.246
		Man3	0.81	0.62	0.246
		Man4	0.836	0.708	0.74
		Man5	0.56	0.62	0.352
		Wom1	1.19	1.695	0.634
	Largely rotated faces	Wom2	0.277	0.955	0.779
		Wom3	0.62	2.79	1.448

True positions of features are marked manually in xy and yz planes and errors were calculated only for selected major control points. Average execution time of feature detection algorithm is 200~300(ms) for Pentium III PC.

5 Conclusion

In this paper, a new method of facial feature extraction in three dimensional space was proposed. The proposed method is robust to 3D head rotation and spatially non-uniform data. It is possible to compensate a 3D rotated face through the proposed normalization step and facial features can be detected accurately in 3D sapce. Also contours of eyes and mouth can be extracted using the defined deformable templates with curvature information reliably. The facial features found by the proposed algorithm can be used in automatic building of an animation face and in the robust 3D face recognition.

Future work will be study of more accurate and robust detection of features. For better performance the number of parameters of deformable templates can be increased or the techniques such as active shape model[7] and active contours[5,6] can be applied.

Also a concrete method for automatic building of an animation face will be studied. It is possible by finding all moving contours and a muscle model.

References

[1] AT&T, *System and apparatus for customizing a computer animation wireframe*, CA2236235 patent, 1998.

[2] T. Fujiwara, "On the detection of feature points of 3D facial image and its application to 3D facial caricature", International Conference on 3-D digital Imaging and Modeling, pages 490-496, 1999.

[3] Y. Lee, "Realistic modeling for facial animation", Proceeding of SIGGRAPH, 1995.

[4] A.L. Yuille, P.W. Hallinan and D.S. Cohen, "Feature extraction from faces using deformable templates", International Journal of Computer Vision, 1992.

[5] L.D. Cohen and I. Cohen, "Finite element methods for active contour models and balloons for 2-D and 3-D images", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 15, pages 1131-1147, 1993.

[6] T. Yokoyama, Y. Yagi and M. Yachida, "Active contour model for extracting human faces", International Conference on Pattern Recognition, volume 1, pages 673-676, 1998.

[7] T.F. Cootes and C.J. Taylor, *Statistical models of appearance for computer vision*, Tech. Report of University of Manchester.

[8] T. Nagamine, T. Uemura and I. Masuda, "3D facial image analysis for human identification", International Conference on Pattern Recognition, pages 324-327, 1992.

[9] H.T. Tanaka, M. Ikeda and H. Chiaki, "Curvature-based face surface recognition using spherical correlation", IEEE International Conference on Automatic Face and Gesture Recognition , pages 372-377, 1998.

[10] M.J.T.Reinders and B.Sankur, "Transformation of a general 3D facial model to an actual scene face", International Conference on Pattern Recognition, volume 3, pages 75-78, 1992.

[11] C. Beumier and M. Acheroy, "Automatic 3D face authentication", Image and Vision Computing, 2000.

[12] C.S. Chua, F. Han and Y.K. Ho, "3D human face recognition using point signature", IEEE International Conference on Automatic Face and Gesture Recognition, pages 233-238, 2000.

[13] P.W. Hallinan, *Two- and three- dimensional patterns of the face*, pages 201-215, A K Peters. Ltd. 1999.

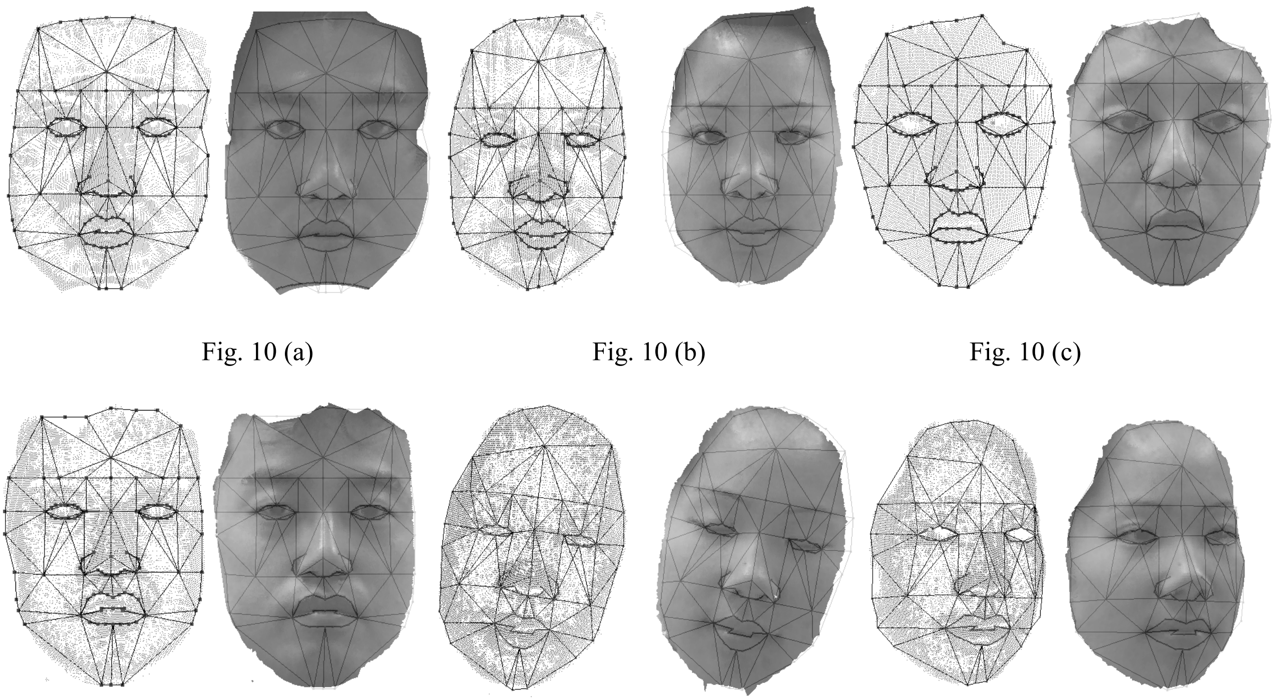

Fig. 10 (a) Fig. 10 (b) Fig. 10 (c)

Fig. 10 (d) Fig. 10 (e) Fig. 10 (f)

Figure 10. Some results of feature detection for various facial range data.
Left images are projected 3D point data and texture-mapped 3D data are on the right in each figure.