G-012

QUANTITATIVE ESTIMATION OF FACIAL PARALYSIS USING THREE-DIMENSIONAL LANDMARK MOVEMENTS

Truc Hung Ngo† Masataka Seo† Naoki Matsushiro‡ Yen-Wei Chen†*

1. Introduction

Facial paralysis is the loss of facial movement ability due to nerve damage. It usually occurs on just one side of the face. The quantitative evaluation of this condition is necessary so that appropriate treatment method can be applied. In Japan, the Yanagihara grading system (YGS) [1] is mostly used for the evaluation of this condition. When applying the YGS, patients are asked to perform ten facial expressions. Then, each expression is assigned a score depending on the seriousness of the condition. Ten facial expressions and respective assigned scores are presented in Table I.

The assignment of scores for facial paralysis from clinicians is based on clinical observation. It is subjective. Therefore, an objective and quantitative assessment of facial paralysis is desired.

This paper presents a technique for the objective evaluation of facial paralysis, in which features are extracted based on landmark's positions in three-dimensional space (3D-landmarks). The landmarks are initialized manually in the first frontal frame and are tracked in the subsequent frontal frames. Then, the landmark's positions are reconstructed in 3D-space using multiview images and a camera self-calibration technique. From the 3D-landmarks, the features are extracted in 3D-space and used for classification. These 3D-features may contain enhanced information such as depth information and, therefore, may help improve the accuracy rates of predicted score.

2. System Overview

The block diagram of our system is described in Fig. 1. For each expression, the first frontal frame is used for landmark initialization. The landmarks of facial structure are shown in Fig.2. These landmarks are tracked in the subsequent frontal frames using the Kanade-Lucas-Tomasi tracker [2] with a preprocessing of Gaussian filter and an intensity normalization. After that, the landmarks are reconstructed in 3D-space. The 3D reconstruction includes a camera self-calibration step (done only once for each subject) and a 3D reconstruction step using multiview geometry [3]. The camera self-calibration comprises a feature point (scale invariant feature transform, SIFT) detection step, a matching of the feature points of the multiview images, an estimate of fundamental matrix using the random sample consensus algorithm and the 8-points algorithm, and an inferring of camera projection matrix [3]. The 3D reconstruction step is performed by using the triangulation algorithm [3]. Next, the 3D-landmarks of facial structure are mirrored so that the paralysed sides of the subjects are on the same side. The normal side and paralysed side are detected based on the side with a greater movement of landmark. Then, an alignment and a normalization process are applied to 3D-landmarks. The alignment includes a rotation so that the inter-pupil line is made perpendicular to *y*-axis (vertical direction) and parallel to *x*-axis (horizontal direction). The normalization consists of a transform of 3D-landmarks so that the inter-pupil's distances of all subjects are the same. Finally, the 3D-features are extracted in the extreme state for classification. The classifier uses multi-class support vector machines based on the one-against-all technique. For each expression, only some related landmarks are used for extracting features. For EP1, the pair of landmarks 2-16 is used. Similarly, the pairs 5-19 and 7-21 are used for EP2, EP3, and EP4; the focused pair of landmarks for EP5 is 10-11; and the pairs 12-22, 13-23, and 14-24 are used for the remaining expressions. The pair 25-26 (inter-pupil) is used for alignment and normalization. Note that EP0 is not included in this study due to features being extracted based on the movements of landmarks.

† Ritsumeikan University
‡ Osaka Police Hospital
* chen@is.ritsumei.ac.jp

TABLE I. YANAGIHARA GRADING SYSTEM

De-note	Expression	Assigned score				
		Full palsy	Serious palsy	Moderate palsy	Slight palsy	Normal
EP0	Resting state	0	1	2	3	4
EP1	Raising eyebrows	0	1	2	3	4
EP2	Closing eyes gently	0	1	2	3	4
EP3	Closing eyes tightly	0	1	2	3	4
EP4	Closing paralysed eye	0	1	2	3	4
EP5	Screwing nose	0	1	2	3	4
EP6	Puffing out cheeks	0	1	2	3	4
EP7	Showing teeth	0	1	2	3	4
EP8	Whistle movement	0	1	2	3	4
EP9	Depressing under-lip	0	1	2	3	4

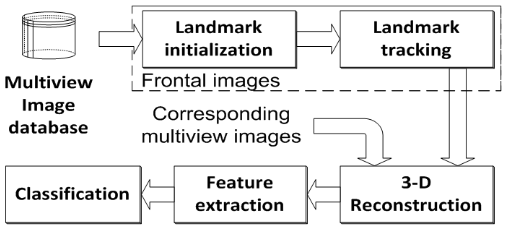

Fig. 1. Block diagram

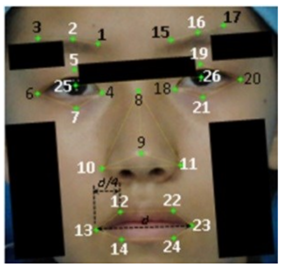

Fig. 2. Twenty-six landmarks of facial structure.

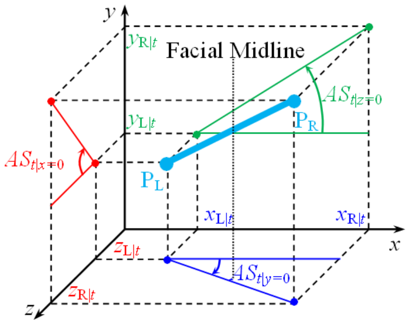

Fig. 3. A pair of corresponding landmarks and its asymmetric features extracted on 3 coordinate planes.

3. Feature Extraction

We define two kinds of features: an asymmetric feature and a movement feature.

The asymmetric features are the angles created by a pair of corresponding landmarks. Consider a pair of corresponding landmarks (P_L and P_R) as shown in Fig. 3. The asymmetric features extracted on 3 coordinate planes x=0, y=0 and z=0; denoted $AS_{t|x}$=0, $AS_{t|y}$=0 and $AS_{t|z}$=0; are calculated by Eq. (1), Eq. (2) and Eq. (3), respectively,

$$AS_{t|x=0} = \begin{cases} \arctan\left(\dfrac{y_{R|t} - y_{L|t}}{z_{R|t} - z_{L|t}}\right) & \text{if } |z_{R|t} - z_{L|t}| \geq \varepsilon \\ \arctan\left(\dfrac{y_{R|t} - y_{L|t}}{\varepsilon}\right) & \text{Otherwise} \end{cases} \qquad (1)$$

$$AS_{t|y=0} = \arctan\left(\frac{z_{R|t} - z_{L|t}}{x_{R|t} - x_{L|t}}\right), \qquad (2)$$

$$AS_{t|z=0} = \arctan\left(\frac{y_{R|t} - y_{L|t}}{x_{R|t} - x_{L|t}}\right), \qquad (3)$$

where ϵ is a threshold to remove the problem of tiny denominator.

The movement feature is the distance of a 3D-landmark moving from a location in frame 1 to a new location in frame t. The movement feature MD_t is calculated by Eq. (4),

$$MD_t = \sqrt{(x_t - x_1)^2 + (y_t - y_1)^2 + (z_t - z_1)^2} \qquad (4)$$

where (x_1, y_1, z_1) and (x_t, y_t, z_t) are the coordinates of the 3D-landmark in frame 1 and in frame t, respectively.

4. Experimental Results

A dynamic facial expression database [4] was used in our experiments. There were 83 samples together with their correct judgment of score. We tested on 1,000 repetitions and got average results. For each run, 17 random samples (20%) were used for testing. Four expressions including EP1, EP3, EP5 and EP7 were selected in our experiments. Table II presents the accuracy rates of predicted scores. These rates are graphically comparison in Fig. 4. The PI–B, LBP–B, Gabor–B, Tracking–2D and Tracking–3D are methods which use intensity images, local binary pattern images, images filtered by Gabor filters, movement of 2D-landmarks and movement of 3D-landmarks (our method) respectively.

TABLE II. ACCURACY RATES OF PREDICTED SCORES.

Method	Accuracy rates (%)			
	EP1	EP3	EP5	EP7
PI–B	50.7	48.2	48.1	39.2
LBP–B	58.3	48.9	41.8	40.1
Gabor–B	62.4	53.1	50.5	54.5
Tracking–2D	69.4	62.1	57.3	70.6
Tracking–3D	70.9	63.3	58.2	73.5

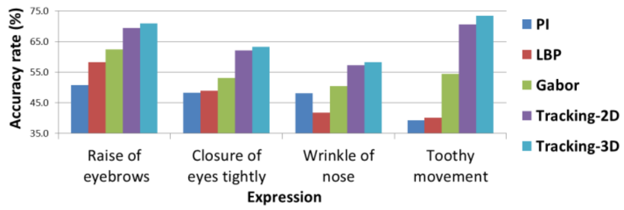

Fig. 4. Graphic comparison of the accuracy rates.

From the result table and the graph, it can be seen that the use of features extracted from the movement of 3D-landmarks is better than the use of features extracted from the other methods. The graph highlights that our method is better than the others.

5. Conclusion

We have presented a method which extracts features from 3D-landmarks for the quantitative assessment of facial paralysis. In our method, the landmarks are tracked in each frame. Then, they are reconstructed in 3D-space using multiview images. Because these features contain enhanced information in 3D-space, they help improve the accuracy rates of predicted scores. For overall evaluation, experimental results showed that our proposed method is superior to the others.

Acknowledgement

This work is supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 15H01130, No. 16H01436, No. 26330212, in part by the MEXT Support Program for the Strategic Research Foundation at Private Universities (2013-2017), in part by the R-GIRO Research Fund from Ritsumeikan University, and in part by the Recruitment Program of Global Experts (HAIOU Program) from Zhejiang Province, China.

Reference

[1] N. Yanagihara et al., “On standardized documentation of facial palsy,” Nippon Jibiinkoka Gakkai Kaiho, vol. 80, no. 8, pp. 799–805, 1977.
[2] Shi and C. Tomasi, “Good Features to Track,” Computer Vision and Pattern Recognition, Proceedings of IEEE Computer Society Conference, pp. 593–600, 1994.
[3] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.
[4] Y. Kihara, G. Duan, T. Nishida, N. Matsushiro and Y. W. Chen, “A Dynamic Facial Expression Database For Quantitative Analysis of Facial Paralysis,” The 6th Int. Conf. on Comp. Sci. and Convergence Info. Tech., pp. 949–952, Nov. 2011.