

Evaluation of Facial Paralysis Degrees Using Multi-Resolution Analysis

T.H. Ngo*

Graduate School of Inf. Sci. and Eng.
Ritsumeikan University
Japan

M. Seo

Graduate School of Inf. Sci. and Eng., Ritsumeikan
University
Japan

Y.W. Chen

Graduate School of Inf. Sci. and Eng.
Ritsumeikan University
Japan

College of Computer Science and Information Technology
Central South Univ. of Forestry and Technology
Hunan, China

N. Matsushiro

Osaka Police Hospital
Japan

Abstract-Facial paralysis is a common clinical condition occurring in the rate 20 to 25 patients per 100,000 people per year. An objective quantitative tool to support medical diagnostics is necessary. This paper proposes a robust method that decomposes the images into multi frequencies-space domain, and then features are extracted for classification using a support vector machine (SVM). The method analyses the images in frequency-space domain, so it overcomes the problems of other techniques such as the change of illumination, noise and redundant frequencies. Experiments show that our proposed method outperforms other techniques testing on a dynamic facial expression image database.

Keywords- facial paralysis; multi-resolution analysis; wavelet decomposition

I. INTRODUCTION

Facial paralysis is a medical condition where the patient loses his or her facial movement ability. It is due to neural damage and usually occurs on only one side of the face. The patient can recover; however, serious sequelae may remain if not treated early and properly. For effective treatment, it is necessary to evaluate the degrees of condition so that apply treatment methods accordingly [1]. For the evaluation, the patient is asked to perform various facial expressions. Each expression is scored based on clinical observation. For example, the Yanagihara grading system (YGS) [2], the method is largely used in Japan for diagnosing, requires the patient to perform 10 facial expressions such as “at rest”, “raise of eyebrows”, “closure of eyes gently”, “nose screw up” expressions...; and then assigns score (points — pts) 0 for full paralysis, 2 for partial paralysis, or 4 for normal for each expression based on clinical observation (Detail as in reference [2]). Such an evaluation is subjective. The given score may differ to such a large extent that it leads to different decisions in the treatment methods [2]. This is undesirable problem. From the requirement in medical area, it is necessary to develop an objective quantitative assessment tool for facial paralysis.

In recent years, many approaches have been developed to objectively evaluate the degrees of facial paralysis. The main principle of those methods is based on the measurement of bilateral asymmetry between two sides of the face using intensity of pixel [3] or local binary patterns (LBP) [4] during an expression. Using the intensity of pixel is sensitive to illumination, shadow and noise. The use of LBP provides a better tolerance against illumination changes, but noise and redundant information in image such as unnecessary frequencies still have not been addressed completely. A technique based on frequency-space analysis can reduce these problems.

This paper presents a technique that uses the wavelet decomposition to analyse the images in frequency-space domain for discriminant feature extraction. It can remove effectively the noise and redundant information contained in the images. Therefore, it overcomes the problems of the other methods.

II. MULTI-RESOLUTION ANALYSIS OF IMAGE

In image processing, using the Mallat method [5], the basic level of wavelet transform decomposes an image into 4 sub bands: the approximation sub band (LL_1), the horizontal detail sub band (LH_1), the vertical detail sub band (HL_1), and the diagonal detail sub band (HH_1). Then the approximation sub band LL_1 in the basic level is used for decomposition in the second level; the approximation LL_2 in the second level is used for the next level decomposition and so on. The dimensions of sub band equals a half of the input. Fig. 1 shows an example of wavelet decomposition into 2 levels of an input image.

III. SYSTEM OVERVIEW

A. Block Diagram

The block diagram of our system is shown in fig. 2. The first frame of each expression is used as the reference frame for the structured construction of local regions. The face is

firstly detected, normalized by the inter-pupil distance, and rotated so that the inter-pupil line is made perpendicular to the vertical face midline. In the second step, the local regions or the regions of interest (ROIs) of the face such as eyebrow, nasal, and mouth regions are detected. Each ROI is divided into 2 equal regions, one on the left and one on the right side, vertically symmetric each other as in fig. 3. Then, each ROI of each frame is normalized in intensity, transformed into scale-space domain using wavelet decomposition. Next, the features are calculated based on the coefficients of the wavelet decomposition and are used as the inputs of an SVM for training and testing.

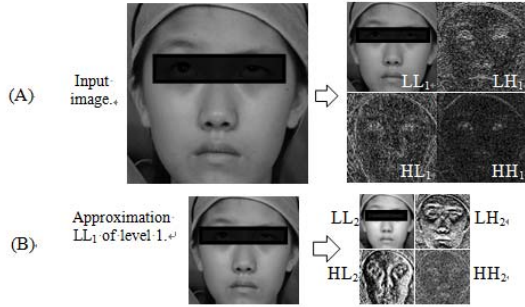


FIGURE I. AN EXAMPLE OF WAVELET DECOMPOSITION. (A) RESULTS OF DECOMPOSITION AT LEVEL 1. (B) RESULTS OF DECOMPOSITION AT LEVEL 2. IN THE FIGURE, THE VALUES OF APPROXIMATIONS LL ARE SCALED INTO IMAGE SCALE FOR DISPLAY.

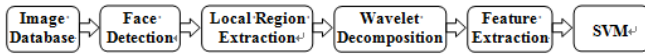


FIGURE II. BLOCK DIAGRAM OF OUR SYSTEM.

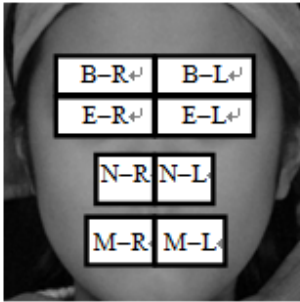


FIGURE III. STRUCTURE OF DETECTED ROIS. THE WHITE BACKGROUND REGIONS ARE THE ROIS. B — EYEBROW REGION, E — EYE REGION, N — NASAL REGION, M — MOUTH REGION, L — LEFT SIDE, R — RIGHT SIDE.

B. Feature Extraction

The facial paralysis degrees are measured by the asymmetric motion between two sides of the face. In addition, the motion ability of each side is also measured to solve the problem of natural asymmetry. Therefore, we define two kinds of feature: The asymmetric feature and the motion feature. In our system, we decompose the images to level 2 using wavelet

decomposition, and only approximation images (LLs) are used to calculate the features.

1) *Asymmetric feature.* In each frame, the asymmetric feature is measured by the correlation coefficient between an ROI on the left and corresponding ROI on the right side of the face. Because these two ROIs are vertically symmetric, one of them needs to flip horizontally before performance of the correlation calculation.

Consider two approximation images, one on the left denoted by I_L and one on the right denoted by I_R , at a decomposition level, the asymmetric feature $\rho_{L,R}$ is calculated as eqn (1).

$$\rho_{L,R} = \text{corr}(I_L(x, y), I_R(x, y)) \quad (1)$$

Where corr stands for correlation coefficient. Fig. 4 shows the calculation performance of the asymmetric feature between the B-R and the B-L regions at decomposition of level 1.

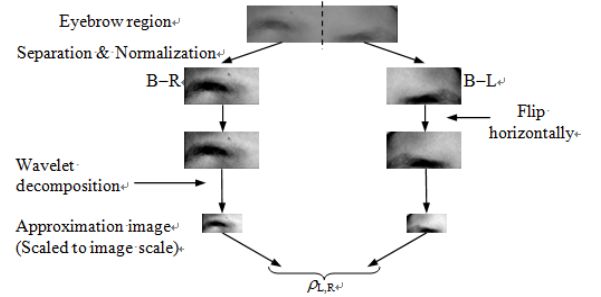


FIGURE IV. ILLUSTRATION OF THE ASYMMETRIC FEATURE CALCULATION. THE APPROXIMATION IMAGES ARE OBTAINED BY THE WAVELET DECOMPOSITION AT LEVEL 1.

2) *Motion feature.* The motion feature at a level of decomposition is measured by the correlation coefficient between an ROI in frame-1 and itself in another frame. For example, the motion feature of B-L region of frame-2 is calculated as eqn (2).

$$\rho_{1,2}^{B-L} = \text{corr}(I_2^{B-L}(x, y), I_1^{B-L}(x, y)) \quad (2)$$

Where I_2^{B-L} and I_1^{B-L} are approximations at considered level of B-L region in frame-2 and frame-1, respectively.

IV. EXPERIMENTAL RESULTS

In our experiments, the dynamic facial expression image database [6] that we have constructed in our previous research is used for training and testing. There are 85 subjects, including 75 volunteer patients and 10 healthy volunteers in this database. Each subject was taken 60 dynamic images for each facial expression with 1 second/image. Based on the suggestion by clinicians at Osaka Police Hospital, for more detail in the quantifying of the facial paralysis, each expression is assigned by one of five score levels (0, 1, 2, 3, or 4) instead of three levels in the YGS. The scores in our database were assigned by the clinicians there.

Three representative expressions of the YGS are chosen for our experiments: The raise of eyebrows, the closure of eyes

gently, and the nose screw-up expression. In each training and testing turn, 68 random samples (80% of number of samples) are used for training and the 17 remaining samples (20% of number of samples) are used for testing. The recognition rates are the average results of 1000 repeated turns.

Three methods have been developed in our work for the comparisons of recognition rate. The difference of these methods is that what kind of image is used for feature calculation. These three methods, named IP-B [3], LPB-B [4], and Wavelet, are based on the intensity of pixel, the LBP image, and the multiresolution wavelet decomposition, respectively. The third one is our proposed method. From experimental results that have highest recognition rates, we only decompose images into 2 levels, and then approximation images at level 1 and level 2 are used for feature calculation. The asymmetric features and the motion features are calculated as described in section 3.2. After that, these features are used as the inputs of the SVM for training and testing.

Consider at a level of decomposition k , we choose only four features that are used as the input of the SVM: two features are the asymmetric features and two more features are the minimum values of the motion features of each facial side. Therefore, the total of features is $4 \text{ (features/level)} \times 2 \text{ (levels)} = 8 \text{ (features)}$ that are used as the inputs of the SVM for training and testing.

Table 1 to table 3 present the recognition rates of the raise of eyebrows, the closure of eyes gently, and the nose screw-up expressions, respectively. Disagreement (%) is used to highlight the difference between the score of clinicians and the recognition score. The average score error (pts) that is calculated as in eqn (3) represents the average error point of each testing sample.

$$\text{Average score error} = \sum_{i=0}^4 i \times \text{disagreement}(i) \quad (3)$$

The tables of the results show that the wavelet decomposition method is better results than the other methods. Compare with the best cases of the other methods, respectively, our method improves the recognition rates 8.26%, 6.9%, and 4.55%; and reduces average score errors 0.27 pts, 0.02 pts, and 0.09 pts for the raise of eyebrows, the closure of eyes gently, and the nose screw-up expressions. Fig. 5 and fig. 6 present the graphic comparisons of the recognition rates and the average score errors, respectively. They show clearly that our proposed method is better than the other ones.

TABLE I. RECOGNITION RESULTS OF THE RAISE OF EYEBROWS EXPRESSION.

Disagreement (%)	0	1	2	3	4	≤ 1	Average score error
IP-B	49.15	30.65	15.32	4.00	0.88	79.80	0.77
LBP-B	56.15	25.06	12.00	6.21	0.58	81.21	0.70
Wavelet	64.41	28.77	6.43	0.35	0.04	93.2	0.43

TABLE II. RECOGNITION RESULTS OF THE CLOSURE OF EYES GENTLY EXPRESSION.

Disagreement (%)	0	1	2	3	4	≤ 1	Average score error
IP-B	50.28	28.23	11.49	8.74	1.26	78.51	0.82
LBP-B	48.85	21.16	15.52	10.56	3.91	70.01	1.00
Wavelet	57.18	18.42	12.99	9.94	1.92	75.6	0.81

TABLE III. RECOGNITION RESULTS OF THE NOSE SCREW-UP EXPRESSION.

Disagreement (%)	0	1	2	3	4	≤ 1	Average score error
IP-B	47.93	23.14	23.36	4.00	1.57	71.07	0.88
LBP-B	40.14	26.14	24.93	6.43	2.36	66.28	1.05
Wavelet	52.48	22.36	20.34	3.62	1.2	74.8	0.79

V. CONCLUSION

We have proposed a robust technique that is based on the wavelet decomposition of image to extract the discriminant features for objective measurement of the facial paralysis. Our method overcomes the problems of the other techniques such as noise, illumination, and redundant frequencies. Therefore, it improves the recognition rates. Experiments have showed that our method outperforms other techniques.

The limitation of our method is that we have not synchronized the start and the end frame in expression yet. Therefore, only 4 features in each decomposition level are represented as the input of the SVM for classification. Time synchronization and speed normalization are our future works. After these steps, we can choose more features during expression, so it promises to improve the recognition rates.

ACKNOWLEDGEMENTS

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. 2430076, 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 Otsuki Memorial Scholarship for Asian and African Countries.

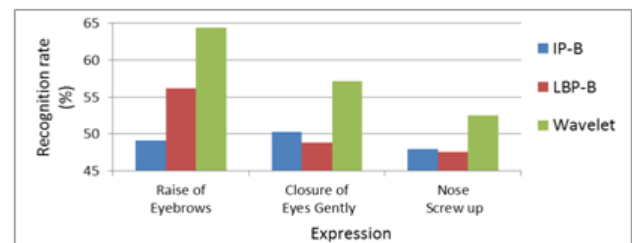


FIGURE V. COMPARISON OF RECOGNITION RATES BETWEEN THREE DIFFERENT METHODS.

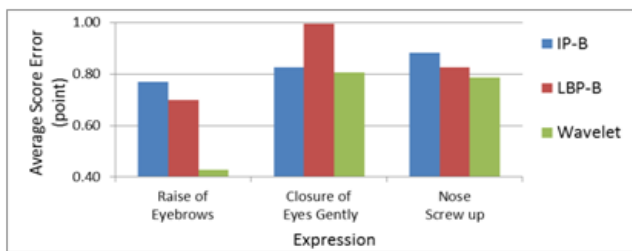


FIGURE VI. COMPARISON OF AVERAGE SCORE ERRORS BETWEEN THREE DIFFERENT METHODS.

REFERENCES

- [1] Shaw, M., Nazir, F., Bone, I., Bell's palsy: A Study of the Treatment Advice Given by Neurologists, *Journal of Neurology, Neurosurgery & Psychiatry*, 76(2), pp. 293–294, February 2005.
- [2] Matsushiro, N., Differences in the Evaluation of Facial Palsy (Yanagihara Grading System) among 52 ENT Doctors at Osaka University, *Journal of Facial Nerve Research*, 29, pp. 60–62, 2010.
- [3] Wachtman, G. S., Liu, Y., Zhao, T., Cohn, J., Schmidt, K., Henkelmann, T. C., VanSwearingen, J. M., and Manders, E. K., Measurement of Asymmetry in Persons with Facial Paralysis, *Combined Annual Conference of the Robert H. Ivy & Ohio Valley societies of Plastic and Reconstructive Surgeon*, 2002.
- [4] He, S., Soraghan, J. J., O'Reilly, B. F., and Xing, D., Quantitative Analysis of Facial Paralysis Using Local Binary Pattern in Biomedical Videos, *IEEE Transactions on Biomedical Engineering*, 56(7), pp. 1864–1870, 2009.
- [5] Mallat, S., A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, *Pattern Analysis and Machine Intelligent, IEEE Transaction on*, 11(7), pp. 674–693, July 1989.
- [6] Kihara, Y., Duan, G., Nishida, T., Matsushiro, N., and Chen, Y.W., A Dynamic Facial Expression Database For Quantitative Analysis of Facial Paralysis, *The 6th International conference on Computer Science and Convergence Information Technology*, pp. 949–952, Jeju Island, Korea, 29 November–01 December, 2011.