

Evaluation of Facial Paralysis Based on Spatial Features of Filtered Images

Truc Hung Ngo¹, Masataka Seo¹, Naoki Matsushiro², Yen-Wei Chen^{1*}

¹ Graduate School of Information Science and Engineering, Ritsumeikan University. Shiga, Japan.

² Osaka Police Hospital. Osaka, Japan.

* Corresponding author. email: chen@is.ritsumei.ac.jp

Manuscript submitted November 15, 2015; accepted January 9, 2016.

doi: 10.17706/ijbbb.2016.6.1.1-8

Abstract: Facial paralysis is a common clinical condition occurring in 30 to 40 patients per 100,000 people per year in Japan. A quantitative tool to support medical diagnostics is necessary. This paper presents a technique that we combined Gabor filters and wavelet decomposition to develop this tool. In our work, the Gabor filters and the wavelet decomposition are used as preprocessing steps to extract the feature. These features are used as the inputs of a multi-class support vector machines for quantitative evaluation of facial paralysis. Our method overcomes the drawbacks of the other techniques such as noisy removal and against variation of illumination. Experimental results show that our proposed method outperforms other conventional techniques testing on a dynamic facial expression image database.

Key words: Gabor filters, wavelet decomposition, quantitative analysis of facial paralysis.

1. Introduction

Facial paralysis is a medical condition where patients lose their facial movement ability. It is due to neural damage and usually occurs on only one side of the face. It is necessary to evaluate the degrees of condition so that the approximate treatment methods can be applied. To evaluate the degree of the facial paralysis, the patient is asked to perform various facial expressions, then the clinician assigns a score for each expression based on clinical observation. Such an evaluation is subjective and therefore highly undesirable in medical field. Hence, an objective quantitative assessment of facial paralysis is desired.

Lots of approaches have been proposed to objectively evaluate the degree of facial paralysis. Some typical methods include the methods of Watchman et al. [1] and He et al. [2]. The main principle is based on the measurement of asymmetry between two sides of the face. Watchman et al. measured the asymmetric features based on static images. Using static images is sensitive to noise and variation of illumination. He *et al.* used the multi-resolution local binary patterns (LBPs). The use of LBP provides a better tolerance against variation of illumination, but noise and redundant frequencies still have not been addressed.

In previous researches, we proposed a combining LBP with Gabor filters [3], multi-resolution analysis [4]. Using frequency techniques easily removes unnecessary frequencies. However, a combination of these frequency techniques is not mentioned for a complete system. In addition, the reference [4] did not solve problem of variation of illumination. This paper presents a technique which uses LBP transform to be against variation of illumination; then LBP images are filtered by Gabor filters and wavelet decomposition for feature extraction. These frequency techniques support each other to develop a complete system.

2. Triage-10 Points Clinical Diagnostics

Yanagihara grading system (YGS) [5] is accepted as a standard diagnostics of facial paralysis in Japan. However, there are some disadvantages too. Firstly, the performance of 10 expressions in the YGS is both physical and psychological burden for patients. Secondly, a research of Matsushiro [6] proved that there are some redundant expressions. Matsushiro proposed a simple and precise method, which was named a triage-10-points (T10P), for evaluation. The T10P requires the patient to perform 3 facial expressions, and each expression is assigned a score corresponding to one of five levels of severity, as shown in Table 1.

Table 1. Triage-10-Points Grading System

Denote	Expression	Evaluation of palsy				
		Full	Almost full	Moderate	Slight	Normal
EP1	Raise of eyebrows	0	1	1	1	2
EP2	Closure of eyes tightly	0	1	2	3	4
EP3	Toothy movement	0	1	2	3	4

3. Local Binary Patterns

LBP was first introduced in 1994 by Ojala et al. [7]. The most important property of the LBP is that it provides better tolerance against variation of illumination. LBP operator transforms a gray image into an image of integer labels. A pixel $I(x, y)$ is transformed into a label by Eq. (1),

$$LBP_{P,R}(x, y) = \sum_{p=0}^P S \left(I \left(x + R \cos \frac{2\pi p}{P}, y + R \sin \frac{2\pi p}{P} \right) - I(x, y) \right) 2^p, \quad (1)$$

where P is the number of circular neighbors, R is the circular radius with center (x, y) ; S is a step function.

4. Gabor Filters

Gabor filter was introduced by Denis Gabor in 1946 [8]. In image processing, Gabor filters represents the best compromise between spatial and frequency localization. A 2D Gabor filter is obtained from the modulation of a 2D Gaussian function by a 2D complex sinusoidal function, and can be defined by Eq. (2),

$$g_{\sigma_x, \sigma_y, f, \theta}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \cdot \exp \left[-\frac{1}{2} \left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2} \right) \right] \exp(j2\pi f x'), \quad (2)$$

where σ_x and σ_y are the standard deviations of the Gaussian, $x'=x.\cos\theta+y.\sin\theta$ and $y'=-x.\sin\theta+y.\cos\theta$ are the coordinates rotated by the angle θ , f is frequency. The output of Gabor filter for an input image is the convolution between that image and the Gabor function. Fig. 1 shows a Gabor filter and filtered images.

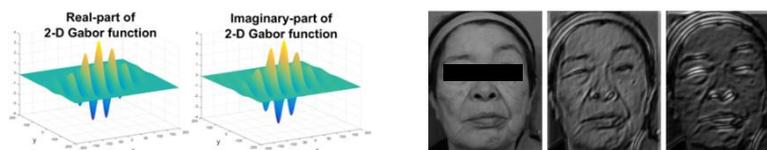


Fig. 1. A Gabor filter and images filtered by Gabor filters. The images, respectively from left to right, are the input image; the image filtered with $\sigma_x=6, \sigma_y=4, f=0.06, \theta=30^\circ$ and with $\sigma_x=8, \sigma_y=4, f=0.05, \theta=30^\circ$.

5. Wavelet Decomposition

Wavelet is a wave - like oscillation with amplitude beginning at zero, increasing, and then decreasing back to zero. Wavelet transforms convert a signal into a series of wavelets. In image processing, wavelet transforms [9] decompose an image into 4 sub-bands: the approximation sub-band (LL), the horizontal detail sub-band (LH), the vertical detail sub-band (HL), and the diagonal detail sub-band (HH). In addition, the image can be decomposed into several levels. The image of approximate sub-band is used for the next level of decomposition. Assuming that in the basic level, the image is decomposed into 4 sub-bands: LL1, LH1, HL1 and HH1. Then, the second level of wavelet transform uses the image LL1 for decomposition. The dimension of sub-band equals a half of the dimension of input image. Fig. 2 shows an example of the wavelet decomposition of an input image.

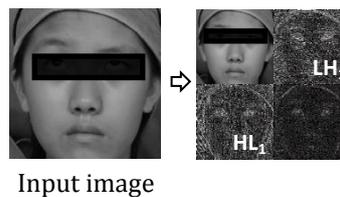


Fig. 2. Wavelet transforms decompose an input image into 4 sub-bands.

6. System Overview

6.1. Block Diagram

The block diagram of the system is shown in Fig. 3. The first frame of each expression is used as a reference frame for normalization and local region construction. The remaining frames are processed the same way as in the reference frame. In the first frame, the face is detected and constructed corresponding local regions of interest (cROIs), as shown in Fig. 3. Because our research focuses on the expressions of T10P, only the local regions of the eyebrows, eyes and mouth are extracted. Then, the pairs of cROIs are transformed by LBP operator, filtered by Gabor filters or wavelet decomposition, and then used for feature extraction. Finally, the extracted features are used as the inputs of a multi-class support vector machines (MC-SVMs) for classification.

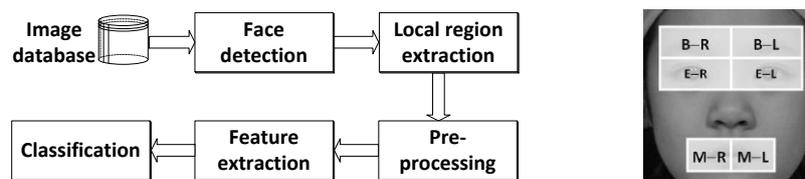


Fig. 3. Block diagram of our system and cROIs. B–eyebrow, E–eye, M–mouth, L–left, R–right.

6.2. Feature Extraction

The severity of facial paralysis is measured by the asymmetry between two sides of the face. In signal processing, the correlation is used to measure the similarity of two random variables. Similarly, the correlation represents the similarity between two images. Therefore, we can use the correlation coefficient of cROIs as extracted features. We define two kinds of feature: an asymmetric feature and a motion feature.

6.2.1. Asymmetric Feature

The correlation between cROIs measures the similarity of these two regions, one side has to be mirrored horizontally. We define this as the asymmetric feature. Consider a pair of cROIs at a frame t . Assuming that

I_L and I_R are images of cROIs after preprocessing. The asymmetric feature AS_t , is calculated by Eq. (3),

$$AS_t = \text{corr}(I_L(x, y), I_R(x, y)), \quad (3)$$

where corr stands for correlation coefficient. The procedures of the asymmetric feature extraction using LBP image filtered by Gabor filters or filtered by wavelet decomposition are illustrated in Fig. 4(a). For each pre-processed image, because we focus on only 1 pair of cROIs, there is only 1 feature to be extracted.

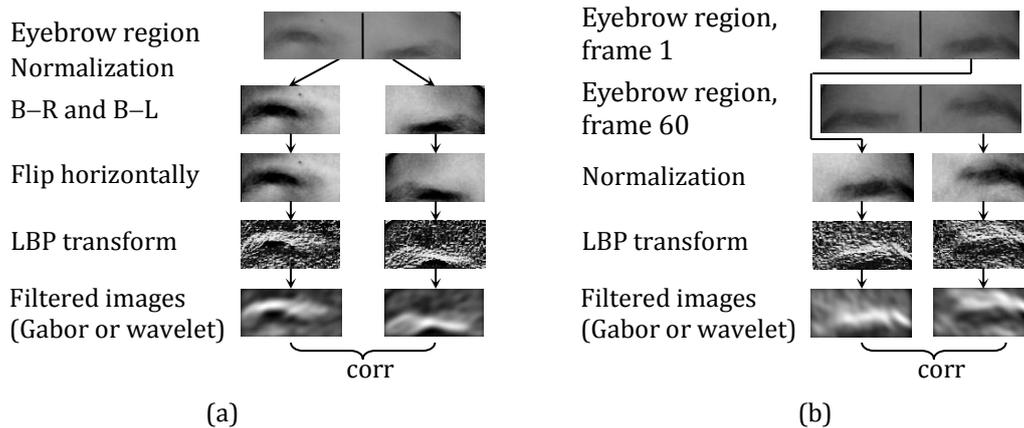


Fig. 4. Illustrations of feature extraction. (a) is the procedures of asymmetric feature extraction. (b) is the procedures of motion feature extraction.

6.2.2. Motion feature

During the expression period, the texture in an ROI is various. The variation of texture relates to the movement ability of muscles. Hence, the correlation between an image in frame-1 and an image in frame- t of one ROI represents the movement ability of muscles. We define this as the motion feature. Assuming that I_1 and I_t are images after the preprocessing step of an ROI in frame-1 and frame- t , respectively; the motion feature, denoted by MT_t , is calculated by Eq. (4).

$$MT_t = \text{corr}(I_t(x, y), I_1(x, y)). \quad (4)$$

The procedures of the motion feature extraction using LBP image filtered by Gabor filters or filtered by wavelet decomposition are illustrated in Fig. 4(b). For each pre-processed image, because there are 2 separate ROIs (left and right side), there are 2 motion features to be obtained.

6.2.3. Feature selection

Consider an expression and a parameter set of either the Gabor filter, or a level of wavelet decomposition. For a frame at time t , there are 1 asymmetric feature and 2 motion features to be obtained. In our experiments, we choose only the features in resting state and extreme state of expression for classification. The selection is performed as follows: Firstly, the motion feature in normal side is found based on greater change of feature values among two motion features. Secondly, the minimum value of this feature is detected. Next, the extreme time-point is looked up when the value of motion feature just falls below a threshold to the minimum (threshold value = 0.015). The extreme time-point is used for selecting one asymmetric feature and two motion features. These features represent the extreme state of an expression.

In addition, the maximum value of the asymmetric feature represents the feature of rest state. This feature is also selected. Therefore, for one parameter set of Gabor filter, or at a level of wavelet decomposition, there are 4 feature values to be selected. For n parameter sets or n levels of wavelet decomposition, there are $n \times 4$ features to be obtained and used for classification.

7. Experiments and Discussions

The dynamic facial expression image database [10] was used in our experiments. There were 85 subjects, including 75 patients and 10 healthy volunteers. In each training and testing run, 80% random samples were used for training, and the remaining 20% samples were used for testing. The recognition rates are the average results of 1,000 repeated turns. Recognition rate is ratio between number of samples with accurate recognition scores and total number of samples. A sample is considered as an accurate recognition if there is no difference between the score given by our system and the score given by clinicians.

For each expression we adjusted and combined several parameter sets for Gabor filters and used approximation images of wavelet decomposition for feature extraction. From experimental results, we see that the combination of parameter sets of $\theta = \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$, ($\sigma_x = \sigma_y = 8.3$, $f = 0.022$) for raising eyebrows, ($\sigma_x = \sigma_y = 4.4$, $f = 0.045$) for closing eyes tightly and ($\sigma_x = \sigma_y = 4.0$, $f = 0.05$) for toothy movement expression gave the best recognition rates. Also, the combining approximation images of level 1 and level 2 of wavelet decomposition gave the highest results. Table 2 presents the recognition rates of using of LBP images filtered by Gabor filters (GBLBP), LBP images decomposed by wavelet transform (WLLBP). Our results are also compared with other conventional methods such as method using intensity of pixel (IP) [1],[3], LBP images (LBP) [2],[3],[7], intensity image filtered by Gabor filters (GBIP) [3] and wavelet decomposition of intensity images (WLIP) [4]. From the result table, we can see that the average recognition rates of GBLBP and WLLBP are better than the conventional methods. We can conclude that the features extracted from filtered images are more suitable for facial paralysis evaluation. GBLBP and WLLBP are also better than GBIP and WLIP, respectively. This may be that LBP helps reduce the influence of variation of illumination. Therefore, the use of GBLBP or WLLBP is better against noise and variation of illumination. However, GBLBP shows better result than WLLBP for raising eyebrows expression, while WLLBP shows superior results to GBLBP for closing eyes tightly and toothy movement expression. Therefore, a combining these two methods to develop the tool for quantitative analysis of facial paralysis may improve overall result.

Table 2. Recognition Rates, Average Errors and the Rates of Disagreement ≤ 1

Method	Recognition rate (%)			AvEr (pts)			%disagreement ≤ 1 (%)		
	EP1	EP2	EP3	EP1	EP2	EP3	EP1	EP2	EP3
IP	67.3	45.8	42.4	0.33	0.98	0.89	99.9	78.2	81.0
LBP	72.2	48.8	46.7	0.29	0.87	0.77	99.9	73.8	82.0
GBIP	76.2	50.1	53.7	0.26	0.82	0.71	99.9	85.7	81.7
WLIP	75.0	52.6	58.1	0.27	0.69	0.64	99.9	89.1	84.6
GBLBP	78.4	51.8	55.2	0.22	0.71	0.67	99.0	84.9	82.8
WLLBP	75.5	54.5	59.3	0.25	0.62	0.60	99.9	91.3	86.2

In addition, to estimate the discrepancy of score between clinician's score and system's score for an input sample, we calculate the average error of the scores (AvEr). The AvEr is calculated by Eq. (5),

$$AvEr = \sum_{i=0}^n i \times \%disagreement(i) \tag{5}$$

where disagreement is the difference of clinician’s score and system’s score. Assuming that for a sample, the given score by clinician is $S_{clinician}$ and the given score by our system is S_{system} , then the level of disagreement of that sample is calculated by Eq. (6),

$$disagreement = abs(S_{clinician} - S_{system}) \tag{6}$$

where $abs(.)$ is absolute operator. The $\%disagreement(i)$ in Eq. (5) is ratio between number of samples with disagreement at level i and total number of samples; n is number of disagreement levels ($n=2$ for EP1, $n=4$ for EP2 and EP3). Because the resolution of score of T10P method is higher than that of YGS as presented in [5-6], the evaluations with average errors smaller than 1 (pt) can be acceptable. Similarly, the evaluations with disagreement smaller than or equal to 1 (pt) are also accepted. From the result table, we can see that our methods satisfy these conditions. Fig. 5(a) and Fig. 5(b) show graphic comparisons of recognition rates and average errors between our proposed methods and the other conventional ones. The figure highlights that our proposed methods are superior to the others.

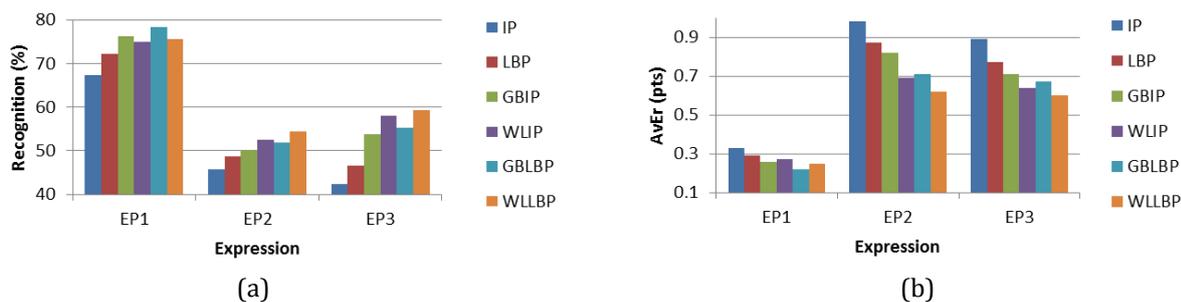


Fig. 5. Graphic comparisons of recognition rates and average errors.

8. Conclusions

We have proposed the use of combining LBP images with Gabor filters and the combining LBP images with wavelet decomposition as preprocessing steps for feature extraction for evaluation of facial paralysis. While the other methods used intensity images, we focused on filtered images. The GBLBP is superior to the other methods for the “raise of eyebrows” expression, while the WLLBP is better than the other methods for the “closure of eyes tightly” and the “toothy movement” expressions. Therefore, an approximate selection between GBLBP and WLLBP depending on the expression will help achieve the highest recognition rates. There are two important conclusions in this research: Firstly, we have shown that the use of frequency techniques is more effective; and secondly, combining with LBP images to be against variation of illumination helps obtain superior results.

Acknowledgements

This work is supported 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.

References

- [1] Wachtman, G. S., Liu, Y., Zhao, T., J. Cohn, K. Schmidt, T. C. Henkelmann, J. M. VanSwearingen, & Manders, E. K. (2002). Measurement of asymmetry in persons with facial paralysis. *Combined Annual Conf. of the H. I. Robert and Ohio Valley societies of Plastic and Reconstructive Surgeon*.
- [2] He, S., Soraghan, J. J., O'Reilly, B. F., & Xing, D. (2009). Quantitative analysis of facial paralysis using local binary pattern in biomedical videos. *IEEE Trans. on Biomedical Engineering*, 56(7), 1864-1870.
- [3] Ngo, T. H., Seo, M., Chen, Y. W., & Matsushiro, N. (2014). Quantitative assessment of facial paralysis using local binary patterns and Gabor filters. *Proc. of the 5th International Symposium on Information and Communication Technology* (pp. 155-161).
- [4] Ngo, T. H., Seo, M., Chen, Y. W., & Matsushiro, N. (2015). Evaluation of Facial Paralysis Degrees Using Multi-Resolution Analysis. *Int. Conf. on Comp. Infor. Systems and Industrial Applications* (pp. 648-651).
- [5] Yanagihara, N., Nishimura, H., Hazama, H., Tamaki, H., Tomita, H., Okuda, Y., et al. (1977). On standardized documentation of facial palsy. *Nippon Jibiinkoka Gakkai Kaiho*, 80(8), 799-805 [In Japanese].
- [6] Matsushiro, N., Hakata, M., & Kojima, H. (2014). A simple and precise assessment method for facial paralysis: Triage-10-points grading system. *The 37th Annual Meeting of the Japan Society of Facial Nerve Research* [in Japanese].
- [7] Ojala, T., Pietikäinen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, vol. 1 (pp. 582-585).
- [8] Gabor, D. (1946). Theory of communication. *J. the Institute of Electrical Engineers*, 93, 429-457.
- [9] Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *Pattern Analysis and Machine Intelligent, IEEE Transaction on*, 11(7), 674-693.
- [10] Kihara, Y., Duan, G., Nishida, T., Matsushiro, N., & Chen, Y. W. (2011). A dynamic facial expression database for quantitative analysis of facial paralysis. *Computer Sciences and Convergence Information Technology (ICCIT), the 6th International conference on* (pp. 949-952).



Truc Hung Ngo received a B.E degree in 1998 from Can Tho Univ., Vietnam, a M.E. degree in 2009 from Ritsumeikan Univ., Japan. From 2013 to present, he has been pursuing a PhD course in Ritsumeikan Univ. His researches focus on signal processing, image processing, computer vision, machine learning and intelligent systems.



Masataka Seo received the B.E. degree in 2007, the M.E. degree in 2010, and the D.E. degree in 2012, from Ritsumeikan Univ., Shiga, Japan. He was a research fellow with the Japan Society for the Promotion of Science, from 2010 to 2013. From April 2014 to December 2014, he was a research associate with the Department of Science and Technology, Univ. of the Kanseigakuin, Hyogo, Japan. He is currently a lecturer with the college of Information Science and Engineering, Ritsumeikan Univ., Shiga, Japan. His research interests include facial image analysis, pattern recognition and computational intelligence.



Naoki Matsushiro received his B.E. and M.D. degree in 1998 and Ph.D. degree in 2003, both from Osaka Univ., Osaka, Japan. From April 2003 to March 2005, he was a fellow with Department of Otolaryngology, Osaka General Medical Center, Osaka, Japan. From April 2005 to June 2009, he was an associate director with Department of Otolaryngology, Osaka Rosai Hospital, Osaka, Japan. From July 2013 to present he is a chief director with Department of Otolaryngology, Facial Nerve Disorder, Ear and Skull Base Surgery, and Hearing Loss Center, Osaka Police Hospital, Osaka, Japan. He is also an associate professor with the Graduate School of Medicine, Osaka Univ., Osaka, Japan. His research interests include treatments and analysis of facial paralysis, lateral skull base surgery, and ear surgery.



Yen Wei Chen received a B.E. degree in 1985 from Kobe Univ., Japan, a M.E. degree in 1987 and a D.E. degree in 1990, both from Osaka Univ., Osaka, Japan. From 1991 to 1994, he was a research fellow with the Institute of Laser Technology, Osaka. From October 1994 to March 2004, he was an associate professor and a professor with the Department of Electrical and Electronic Engineering, Univ. of the Ryukyus, Okinawa, Japan. He is currently a professor with the Graduate School of Information Science and Engineering, Ritsumeikan Univ., Japan. He is also a chair professor with the College of Computer Science and Technology, Zhejiang Univ., China. He was a visiting professor with the Oxford Univ., UK, 2003. He is an overseas assessor of Chinese Academy of Science and Technology. He is an associate editor of *International Journal of Image and Graphics (IJIG)* and editorial board members of the *International Journal of Knowledge based and Intelligent Engineering Systems* and the *International Journal of Information*. His research interests include pattern recognition, image processing and machine learning. He has published more than 200 research papers in these fields.