Training of Feature Extractor via New Cluster Validity – Application to Adaptive Facial Expression Recognition

Sang Wan Lee[1], Dae-Jin Kim[2], Yong Soo Kim[3], and Zeungnam Bien[1]

[1] Department of Electrical Engineering and Computer Science, KAIST, Daejeon, 305-701, Korea
bigbean@ctrsys.kaist.ac.kr, zbien@ee.kaist.ac.kr
http://ctrgate.kaist.ac.kr/~bigbean

[2] Human-friendly Welfare Robot System Engineering Research Center, KAIST, Daejeon, 305-701, Korea
djkim@ctrsys.kaist.ac.kr

[3] Division of Computer Engineering, Daejeon University, Daejeon, 300-716, Korea
kystj@dju.ac.kr

Abstract. A lot of researches on classifiers, which can perform well with a given set of feature vectors, have been done. However, researches on feature vectors, which extract better feature vectors automatically, have not been done very much. We face two problems when we consider feature extraction process. One is how we can make a good feature extractor, and the other is what more separable features are. In this paper, we solved these two problems by proposing feature extractor-training methodology that uses new cluster validity as an objective function. By combining feature extractor to Fuzzy Neural Network Model, we achieve on-line adaptation capability as well as optimized feature extraction. The result shows recognition rate of 97% when on-line adaptation is being done.

1 Introduction

Facial Expression has various characteristics such as interconnection among components, vagueness, and subjectivity[1]. To consider these characteristics, researchers have used various classification techniques[1-5]. However, most of previous researches are focused on classification only, not feature extractor, so the complexity of interconnection increases as using more features. In this paper, to solve this problem and to exclude heuristics in selecting feature, we propose training method of feature extractor based on Levenberg-Marquardt concept.

So then, what is separable feature? And how can we make feature more separable? Various researches related with this question have been studied on the category of Fuzzy C-Means algorithm. To improve Bezdek[6] and Xie and Beni[7]'s concepts, Kim developed inter-cluster proximity[8]. Though it can evaluate validity of clusters, it cannot measure separability when there is no overlap.

Then the question, what are more better features guaranteeing separability, still remains unanswered. Therefore, to guarantee the separability of class vector from feature extractor, we propose novel cluster validity as an objective function for training. Also, we introduce adaptation for personalized facial expression recognition, specially using Fuzzy Neural Network Model[9] performing on-line adaptation. In section 2,

R. Khosla et al. (Eds.): KES 2005, LNAI 3684, pp. 542–548, 2005.

we introduce overall structure of Facial Expression Recognition System. In section 3, new cluster validity for training is introduced, and then training method of feature extractor is explained in section 4. The superiority of on-line adaptation capability is shown in section 5.

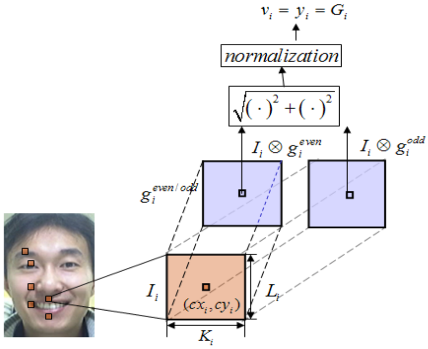


Fig. 1. Feature Extraction Process for given 6-facial feature points

2 Adaptive Facial Expression Recognition System

2.1 Feature Extraction Process via Gabor Wavelets [10]

Face and its each region are found by Haar-classifier using openCV library[11]. Because we can approach to each facial point in this manner, we focus on facial expression recognition process in this paper, rather than face detection process. In facial expression recognition process, the first step is a feature extracting layer which has a set of Gabor filters and normalization process[10](Fig. 1). Each feature vector, as follows, can show the rate of specific frequency and angle that sub image contains.

$$\vec{v} = [v_1, v_2, v_3, v_4, v_5, v_6]^T \tag{1}$$

2.2 Fuzzy Neural Network Model[9]

To organize adaptive classifier, output of feature extractor is applied to Fuzzy Neural Network Model[9] working in the unsupervised manner. In this algorithm, after performing initial iteration with training data, the centers of each cluster are converged, and then perform an on-line adaptation.

3 Proposed Cluster Validity for Training of Feature Extractor

In this section, we will introduce three kinds of criterions and total cluster validity consisted of them. To get more separable feature sets through the training process in section 4, proposed total cluster validity is used as an objective function, and so it has the form that is easy to differentiate.

Definition 1. Cluster Validity Based on Distance
Basically in this paper, conventional distance-based cluster validity, namely within and between class scatterness, are used. They are as follows.

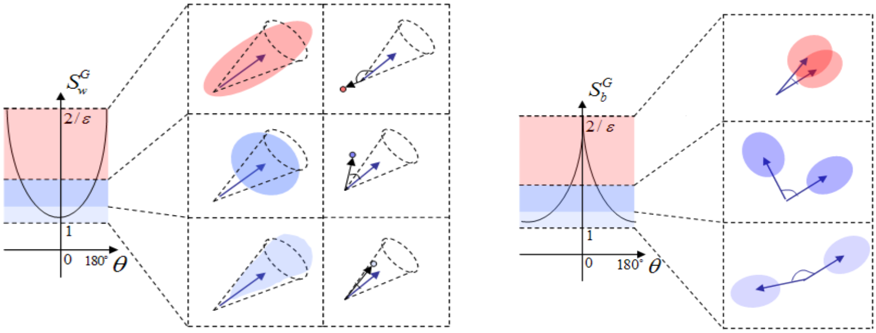


Fig. 2. Within-class guide $s_w^{G\,(j)}(n)$ about n-th point of class j and Between-class guide $s_b^{G\ (j,k)}$ about class j and k are shown. Both reflect separability of inner-class, and between classes

$$S_w = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{N_j}\sum_{n=1}^{N_j}\left\{\sum_{i=1}^{6}(v_i^{(j)}(n)-M_i^{(j)})^2\right\} \tag{2}$$

$$S_b = \frac{1}{J}\sum_{j=1}^{J}\left\{\sum_{i=1}^{6}(M_i^{(j)}-M_{oi})^2\right\} \tag{3}$$

where $\bar{M}^{(j)} = \sum_{n=1}^{N_j}\frac{\bar{v}^{(j)}}{N_j}$, $\bar{M}_o = \frac{1}{J}\sum_{j=1}^{J}\sum_{n=1}^{N_j}\frac{\bar{v}^{(j)}}{N_j}$ is center of each and total class. N, N_j, J is number of training data, number of data per class, and number of class, respectively.

Definition 2. Proposed Cluster Validity Based on Angle
Because the scatterness in definition 1 considers trace of matrix only, additional criterion which considers scattering direction is needed. The most well-known criterion to fulfill this need is Mahalanobis distance and first term of Battacharyya distance. To apply these characteristics, we introduce a minimization problem based on vector concept. Minimizing cost function means that data are pulled from center of total class so that they can locate around center of their class.

$$\text{minimize}\ \ \angle\left(\bar{v}^{(j)}(n)-\bar{M}_o\right)\left(\bar{M}^{(j)}-\bar{M}_o\right) \equiv \text{minimize}\ \left|\cos^{-1}\left(\frac{\left(\bar{v}^{(j)}(n)-\bar{M}_o\right)\cdot\left(\bar{M}^{(j)}-\bar{M}_o\right)}{\left|\bar{v}^{(j)}(n)-\bar{M}_o\right|\left|\bar{M}^{(j)}-\bar{M}_o\right|}\right)\right| \tag{4}$$

Because cosine function is concave in $-180^\circ \sim 180^\circ$, we simplify above problem.

$$\text{minimize}\ \ 2/\left\{\varepsilon+1+\frac{\left(\bar{v}^{(j)}(n)-\bar{M}_o\right)\cdot\left(\bar{M}^{(j)}-\bar{M}_o\right)}{\left|\bar{v}^{(j)}(n)-\bar{M}_o\right|\left|\bar{M}^{(j)}-\bar{M}_o\right|}\right\} \tag{5}$$

where $\varepsilon\ (0<\varepsilon\le 1)$ means control parameter of concavity. From above, we define cost function to be minimized as a guidance of within-class.

$$s_w^{G\,(j)}(n) \triangleq \left(2/\left\{\varepsilon+1+\frac{\left(\bar{v}^{(j)}(n)-\bar{M}_o\right)\cdot\left(\bar{M}^{(j)}-\bar{M}_o\right)}{R^{(j)}(n)\,R_M^{(j)}}\right\}\right), \qquad S_w^G \triangleq \frac{1}{J}\sum_{j=1}^{J}\frac{1}{N_j}\sum_{n=1}^{N_j}s_w^{G\,(j)}(n) \tag{6}$$

where j is class index and

$$R^{(j)}(n) \triangleq \left|\bar{v}^{(j)}(n)-\bar{M}_o\right|, \qquad R_M^{(j)} \triangleq \left|\bar{M}^{(j)}-\bar{M}_o\right| \tag{7}$$

Secondly, maximizing problem is introduced as guidance among classes. Maximization of angle makes classes avoid each other in viewpoint of direction.

$$\text{maximize } \angle\left(\bar{M}^{(j)}-\bar{M}_o\right)\left(\bar{M}^{(k)}-\bar{M}_o\right) \;\equiv\; \text{maximize } \left|\cos^{-1}\left(\frac{\left(\bar{M}^{(j)}-\bar{M}_o\right)\cdot\left(\bar{M}^{(k)}-\bar{M}_o\right)}{\left|\bar{M}^{(j)}-\bar{M}_o\right|\left|\bar{M}^{(k)}-\bar{M}_o\right|}\right)\right| \tag{8}$$

By similar simplification as the first case, we can convert Eq. (8) to a maximization problem. In this manner, we define cost function to be minimized as a guidance of between-class.

$$s_b^{G\ (j,k)} \triangleq \left(2\Big/\left\{\varepsilon+1-\frac{\left(\bar{M}^{(j)}-\bar{M}_o\right)\cdot\left(\bar{M}^{(k)}-\bar{M}_o\right)}{R_M^{(j)}\,R_M^{(k)}}\right\}\right), \qquad S_b^G \triangleq \frac{1}{{}_J C_2}\sum_{j=1}^{J}\sum_{k=1}^{J} s_b^{G\,(j,k)} \tag{9}$$

Definition 3. Additional Cluster Validity for Fuzzy Neural Network Model
To utilize FNNM[19], we need an additional condition, because FNNM performs vigilance test by fixed vigilance parameter. It means a variation of within-class scatterness. Therefore, it is similar to the covariance term of Battachayya distance.

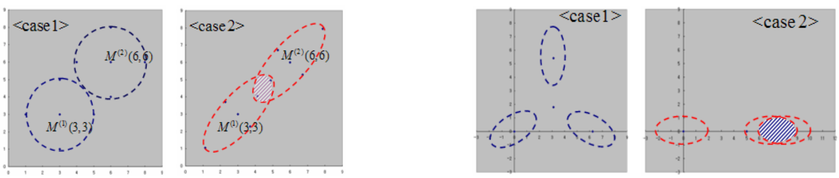
$$S_b^V \triangleq \frac{1}{J}\sum_{j=1}^{J}\left(s_w^{(j)}-S_w\right)^2 \tag{10}$$

Definition 4. Proposed Total Cluster Validity
Criterions introduced through Def. 1 to 3 constitute total cluster validity as follows.

$$S_{proposed} \triangleq \begin{pmatrix} S_b^{new} & 0 & 0 \\ 0 & S_w^{new} & 0 \\ 0 & 0 & \left(1+S_b^V\right) \end{pmatrix} \quad \text{where } S_b^{new}=\begin{pmatrix} S_b^{-1} & 0 \\ 0 & S_b^G \end{pmatrix},\ S_w^{new}=\begin{pmatrix} S_w & 0 \\ 0 & S_w^G \end{pmatrix} \tag{11}$$

To evaluate and differentiate Eq. (11), we use a determinant.

$$\det(S_{proposed}) = S_b^{-1}S_b^G S_w S_w^G(1+\beta S_b^V) = \det(S_b^{new})\det(S_w^{new})(1+\beta S_b^V) \tag{12}$$



	Case1	Case2
ⓐ	0.96	0.96
ⓑ	1.07	4.47

	Case1	Case2
ⓐ	0.19	0.19
ⓑ	0.24	1.40

Fig. 3. Numerical Example ; Comparison between conventional cluster validity in Eq.(2,3) and proposed cluster validity in Eq. (11)

Now the training problem of feature extractor is determined to minimize Eq. (12). Figure 3 is a numerical example showing that the proposed method is superior to the conventional criterion as introduced in Eq. (2) and (3).

4 Training of Feature Extractor

We introduce training technique of feature extractor using proposed separability criterion in section 3. Basically, training methodology follows Levenberg – Marquardt(LM) concept[12,13]. First, partial derivatives of S_w, S_b are as follows.

$$\frac{\partial S_w}{\partial f_i} = \frac{2}{J}\sum_{j=1}^{J}\frac{1}{N_j}\sum_{n=1}^{N_j}(v_i^{(j)}(n) - M_i^{(j)})\left\{\frac{\partial v_i^{(j)}(n)}{\partial f_i} - \frac{\partial M_i^{(j)}}{\partial f_i}\right\} \tag{13}$$

$$\frac{\partial S_b}{\partial f_i} = \frac{1}{J}\sum_{j=1}^{J}2(M_i^{(j)} - M_{oi})\left\{\frac{\partial M_i^{(j)}}{\partial f_i} - \frac{\partial M_o}{\partial f_i}\right\} \tag{14}$$

where $\frac{\partial M_{oi}}{\partial f_i} = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{\partial M_i^{(j)}}{\partial f_i}\right)$, $\frac{\partial M_i^{(j)}}{\partial f_i} = \frac{1}{N_j}\sum_{n=1}^{N_j}\frac{\partial v_i^{(j)}{}_{(n)}}{\partial f_i}$

Secondly, partial derivative of S_w^G has a final numerical formula as follows.

$$\frac{\partial S_w^G}{\partial f_i} = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{N_j}\sum_{n=1}^{N_j}\frac{\partial s_w^{G(j)}(n)}{\partial f_i} \tag{15}$$

$$\begin{aligned}\frac{\partial s_w^{G(j)}(n)}{\partial f_i} = -\frac{\{s_w^{G(j)}(n)\}^2}{2R^{(j)}(n)R_M^{(j)}}\cdot&\left\{\frac{\left(v_i^{(j)}(n) - M_{oi}\right)}{\left(M_i^{(j)} - M_{oi}\right)}R_M^{(j)}\frac{\partial R_M^{(j)}}{\partial f_i} + \frac{\left(M_i^{(j)} - M_{oi}\right)}{\left(v_i^{(j)}(n) - M_{oi}\right)}R^{(j)}\frac{\partial R^{(j)}}{\partial f_i}\right.\\ &\left.+\left(\varepsilon + 1 - \frac{2}{s_w^{G(j)}(n)}\right)\left(R^{(j)}(n)\frac{\partial R_M^{(j)}}{\partial f_i} + R_M^{(j)}\frac{\partial R^{(j)}(n)}{\partial f_i}\right)\right\}\end{aligned} \tag{16}$$

where $\frac{\partial R^{(j)}(n)}{\partial f_i} = \frac{\left(v_i^{(j)}(n) - M_{oi}\right)}{R^{(j)}(n)}\left(\frac{\partial v_i^{(j)}(n)}{\partial f_i} - \frac{\partial M_{oi}}{\partial f_i}\right)$, $\frac{\partial R_M^{(j)}}{\partial f_i} = \frac{\left(M_i^{(j)} - M_{oi}\right)}{R^{(j)}(n)}\left(\frac{\partial M_i^{(j)}}{\partial f_i} - \frac{\partial M_{oi}}{\partial f_i}\right)$

Also, partial derivative of S_b^G is given by computation similar to the case above.

$$\frac{\partial S_b^G}{\partial f_i} = \frac{1}{{}_JC_2}\sum_{j=1}^{J}\sum_{k=1}^{J}\frac{\partial s_b^{G(j,k)}(n)}{\partial f_i} = \frac{2}{J(J-1)}\sum_{j=1}^{J}\sum_{k=1}^{J}\frac{\partial s_b^{G(j,k)}(n)}{\partial f_i} \tag{17}$$

$$\begin{aligned}\frac{\partial s_b^{G(j,k)}}{\partial f_i} = \frac{\{s_b^{G(j,k)}\}^2}{2R_M^{(j)}R_M^{(k)}}\cdot&\left\{\frac{\left(M_i^{(j)} - M_{oi}\right)}{\left(M_i^{(k)} - M_{oi}\right)}R_M^{(k)}\frac{\partial R_M^{(k)}}{\partial f_i} + \frac{\left(M_i^{(k)} - M_{oi}\right)}{\left(M_i^{(j)} - M_{oi}\right)}R_M^{(j)}\frac{\partial R_M^{(j)}}{\partial f_i}\right.\\ &\left.-\left(\varepsilon + 1 - \frac{2}{s_b^{G(j,k)}}\right)\left(R_M^{(j)}\frac{\partial R_M^{(k)}}{\partial f_i} + R_M^{(k)}\frac{\partial R_M^{(j)}}{\partial f_i}\right)\right\}\end{aligned} \tag{18}$$

Also, Partial derivative of additional cluster validity S_b^V is given as follows.

$$\frac{\partial S_b^V}{\partial f_i} = \frac{2}{J}\sum_{j=1}^{J}(s_w^{(j)} - S_w)\cdot\left(\frac{\partial s_w^{(j)}}{\partial f_i} - \frac{1}{J}\sum_{j=1}^{J}\frac{\partial s_w^{(j)}}{\partial f_i}\right) \tag{19}$$

$$\frac{\partial s_w^{(j)}}{\partial f_i} = \frac{2}{N_j}\sum_{n=1}^{N_j}(v_i^{(j)}{}_{(n)} - M_i^{(j)})\left\{\frac{\partial v_i^{(j)}{}_{(n)}}{\partial f_i} - \frac{\partial M_i^{(j)}}{\partial f_i}\right\} \tag{20}$$

Finally, partial derivative of $\det(S_{proposed})$ is calculated using Eq. (13) to (20).

$$\frac{\partial \det(S_{proposed})}{\partial f_i} = \left\{\frac{\partial \det(S_b^{new})}{\partial f_i}\det(S_w^{new}) + \frac{\partial \det(S_w^{new})}{\partial f_i}\det(S_b^{new})\right\}(1 + \beta S_b^V) + \beta\frac{\partial S_b^V}{\partial f_i}\det(S_b^{new})\det(S_w^{new}) \tag{21}$$

where $\frac{\partial \det(S_b^{new})}{\partial f_i} = \frac{\partial S_b^G}{\partial f_i} S_b^{-1} - \frac{S_b^G}{S_b^2} \frac{\partial S_b}{\partial f_i}$, $\frac{\partial \det(S_w^{new})}{\partial f_i} = \frac{\partial S_w^G}{\partial f_i} S_w - \frac{\partial S_w}{\partial f_i} S_w^G$. And all calculation process about θ is just the same.

Additionally, by using technique introduced in [10], we can get partial derivatives of feature extraction process, $\partial v_i / \partial f_i$, $\partial v_i / \partial \theta_i$. Hereby all calculations for Levenberg – Marquardt algorithm are completed so as to find best cluster validity.

5 Results

We used 42 images of EKMAN DB and 45 JAFFE DB to get a general classifier. The learning process was done for three expressions - happy, sad, and angry. To show the system's adaptation capability, BSCL DB was used. This database, which consisted of 5 individuals and totally 600 images, was collected by PC-CAM in real circumstances during 6 days. First 60 data of each user were used as training data to organize personalized classifier. Then another 60 data, collected during 5 days after first data set had been collected, were used as on-line test data.

Generalization performance for EKMAN and JAFFE DB is 74% and 89% each. Personalization(Off-line adaptation) test of BSCL DB is successful as shown in second column of Table 2. In this table, clustering rate averages 98.7% after initial clustering of FNNM when learning of feature extractor is done, whereas 80.2 % without learning. From this result, we can conclude that performance is improved by minimizing proposed cluster validity.

Table 1. Left table shows initial clustering result (98.7%), and right table shows on-line adaptation test after initial clustering (97% for new data)

	FNNM only	**Training of Feature Extractor + FNNM**
USER#1	98.3 %	100.0 %
USER#2	76.7 %	100.0 %
USER#3	81.2 %	100.0 %
USER#4	61.7 %	98.3 %
USER#5	83.3 %	95.0 %

	No Adaptation	**On-line Adaptation**
USER#1	95.0 %	100.0 %
USER#2	91.7 %	96.7 %
USER#3	95.0 %	98.3 %
USER#4	98.3 %	98.3 %
USER#5	68.3 %	91.7 %

After organizing personalized GWNN, we continued to perform on-line adaptation with other 60 data during 5 days. Table 4 shows the results. The accuracy of 97.0% was obtained in average when performing on-line adaptation, whereas generalization performance of 89.7%, which does not utilize on-line adaptation.

6 Conclusion

Though previous researches considered various kinds of features and the interconnection among them by using various classification techniques, they have not been considered fundamental problem, feature extraction process itself. As the solution for this problem, a training method of feature extractor in Gabor Wavelet Neural Network for facial expression recognition has been proposed.

By using FNNM as an on-line adaptation, the system becomes to have capability of tracking user's change of facial expressions as time goes by without any supervisory manner. Therefore, the recognition rate can be improved when unlearned new user continues to use the system using adaptation process.

References

1. Park, G.-T.: A Study on Extraction of Emotion from Facial Image using Soft Computing Techniques. Ph.D Thesis, Dept. of Electrical Engineering and Computer Science, KAIST. (1998)
2. Kapoor, A., Yuan, Q., Picard, R.W.: Fully Automatic Upper Facial Action Recognition. IEEE International Analysis and Modeling of Faces and Gestures. (2003) 195 – 202.
3. Fasel, B.: Robust Facial Analysis using Convolutional Neural Networks. Proc. of the ICPR. (2002)
4. Guodong, G., Charles, R. D.: Simultaneuous Feature Selection and Classifier Training via Linear Programming: A Case Study for Face Expression Recognition. Proc. Of IEEE CVPR. (2003)
5. Ira Cohen et al.: Facial Expression Recognition From Video Sequences. ICME. (2002)
6. Bezdek, J.C.: Numerical taxonomy with fuzzy sets. J. Math. Biol vol 1. (1974) pp. 57-71.
7. Xuanli, L., Beni, G.: A Validity Measure for Fuzzy Clustering. IEEE Trans. on PAMI. Vol 13. (1991)
8. Kim, D.-W., Lee, K.H., Lee, D.H.: Fuzzy Cluster Validation Index based on Inter-Cluster Proximity. Pattern Recognition Letters. (2003)
9. Kim,Y.S., Ham,C.H., Baek, Y. S.: A Fuzzy Neural Network Model Solving the Underutilization Problem. Journal of Korea Fuzzy Logic and Intelligent Systems Society, Vol. 11. (2001) pp. 354-358.
10. Lee, S.W., Kim, D.-J., Kim, Y. S., Bien, Z.: An Adaptive Facial Expression Recognition System Using Fuzzy Neural Network Model and Q-learning, SCISISIS. Yokohama. Japan. (2004)
11. http://www.cs.bham.ac.uk/resources/courses/robotics/doc/opencvdocs/ref/OpenCVRef_Experimental.htm#decl_cvGetHaarClassifierCascadeScale
12. Ranganathan A.: The LM Algorithm, Report of BORG Lab. in Georgia Institute of Technology. (2004)
13. Edwin, K. P., Stanislaw H. Zak.: An Introduction to Optimization, John Wiley & Sons. (2001)