Facial Expression Recognition with Fuzzy C-Means Clustering Algorithm and Neural Network Based on Gabor Wavelets

Youngsuk Shin[1], Chansup Chung[2], Yillbyung Lee[3]

1. Graduate Program in Cognitive Science of Yonsei University
2. Department of Psychology in Yonsei University
3. Department of Computer Science & Industrial Systems Engineering in Yonsei University.

Abstract

This paper presents a facial expression recognition based on Gabor wavelets that uses a fuzzy C-means(FCM) clustering algorithm and neural network. Features of facial expressions are extracted to two steps. In the first step, Gabor wavelet representation can provide edges extraction of major face components using the average value of the image's 2-D Gabor wavelet coefficient histogram. In the next step, we extract sparse features of facial expressions from the extracted edge information using FCM clustering algorithm. The result of facial expression recognition is compared with dimensional values of internal states derived from semantic ratings of words related to emotion. The dimensional model can recognize not only six facial expressions related to Ekman's basic emotions, but also expressions of various internal states.

Keywords: facial expression recognition, fuzzy C-means algorithm, neural network, Gabor wavelets pleasure-displeasure(P-DP), arousal-sleep(A-S)

1. Introduction

The recognition of human facial expressions has been recently studied for advanced human interface. Most of works on facial expression recognition include studies using the six principle emotions of Ekman[1]. Ekman considers the six basic emotions: happiness, surprise, fear, anger, disgust, and sadness; and categorizes facial expressions with these six basic emotions. It limits the recognition of natural facial expressions which consist of several other emotions and many combinations of emotions.

In study of Kim Younga et al. [2], the dimension study about emotion through the semantic rating of emotion words which indicates two-dimensional structure: pleasure/displeasure, arousal/sleep. Each expressor of female and male posed 83 internal emotional state expressions when 83 words of emotion are presented.

Experimental subjects rated pictures for degrees of expression in each of the two dimensions on a nine point scale.

The images were labeled with a rating averaged over all subjects. These posed pictures may differ from the words of emotion between the posed expression and named expression since expressors depend strongly on the personal subjectivity. We thus picked up 44 internal state expressions being a high agreement of posed expression and named expression; and for our experiment used 11 expressions in a set of 44 internal state expressions from each of 6 people. The 11 expressions are happiness, surprise, sadness, disgust, fear, satisfaction, comfort, distress, tiredness, worry (including neutral face). Fig.1 shows examples of images used.

To extract features of major face components, the average value of the image's 2-D Gabor wavelet coefficient histogram on all the images was used.

Fig.1. Examples of images used.

Fuzzy C-means clustering algorithm and dynamic linking are proposed to extract sparse local features from edges on expression images extracted previously. This conclusion demonstrates recognition of facial expressions based on the two-dimensional structure of emotion using a neural network.

2. Gabor transformation

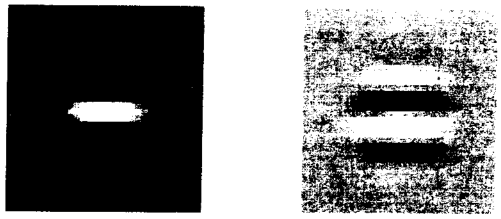
We use 205 images of females and males, each gray level image using 640 by 480 pixels and include face images almost in the frontal pose.

For feature extraction of major face components, an average value of 2-D Gabor coefficients histogram of the image is used. Each image, I was convolved with both even and odd Gabor kernels in an image $I(\vec{x})$ at position $\vec{x}_0$. In our implementation we chose a description of the complex-valued two-dimensional Gabor transform is given by Daugman[3].

$$\psi_{\vec{k}}(\vec{x}) = \frac{k^2}{\sigma^2}\exp(-\frac{k^2x^2}{2\sigma^2})[\exp(i\vec{k}\vec{x}) - \exp(-\frac{\sigma^2}{2})] \qquad (2.1)$$

The wave vector $\vec{k}$ determines the wavelength and orientation of the kernel $\psi_{\vec{k}}$. The parameter σ notes the width of the Gaussian window of the function.

To detect features of $\psi_{\vec{k}}$ major face components, we use a specific frequency band, a wave number k=0.78 and an angle of orientation, $\theta=0°$ and chose $\sigma=\pi$. The complex valued applied to each image combine an even and odd part (see Fig. 2).



(a) (b)

Fig.2. (a) The real part: G_1

(b) The imaginary part : G_2

We use only the magnitudes because they represent local information of an image in a smoothly varying way. The computation proceeds as follows:

$$\begin{aligned} w_1 &= \sum G_1 I(\vec{x}) \\ w_2 &= \sum G_2 I(\vec{x}) \\ M &= (w_1^2 + w_2^2)^{1/2} \end{aligned} \qquad (2.2)$$

Fig.3 shows the result of the 2-D Gabor coefficients histogram using the magnitudes of Gabor coefficients from an expression image. This means these coefficients completely capture facial local feature points in special frequency and special orientation. Thus, we applied the average value of 2-D Gabor coefficient histogram to extract local facial feature points. The average value of Gabor coefficients histogram is controlled by optional value $\pm\alpha$ since experimental images may be a noise.

Fig.4 shows the result of image which applied an optional value to an average value of the Gabor coefficients histogram.

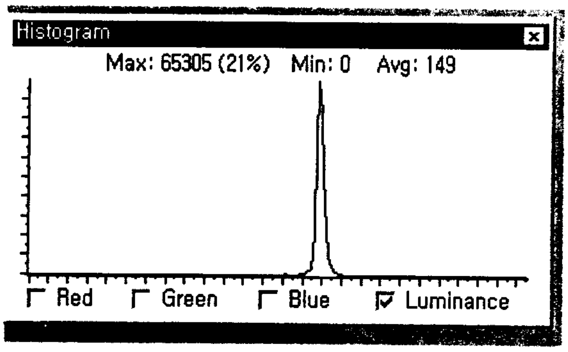


Fig.3. Histogram of image's 2-D Gabor coefficients

Fig.4. Edges of major face components

3. Sparse coding using fuzzy C-means clustering algorithm

Extracted feature points are similar to edges of major face components. Since Gabor vectors with neighboring pixels are highly correlated and redundant, it is sufficient to use sparse pixels on a face. Therefore, we pick out sparse feature points based on the Fuzzy C-means clustering(FCM) algorithm in feature points extracted from the image's 2-D Gabor coefficients histogram.

FCM is a data clustering algorithm in which each data point belongs to a cluster to a degree specified by a membership grade. Bezdek[4] proposed this algorithm in 1973. The potentiality of fuzzy clustering algorithms can be demonstrated by their application in clustering tasks which involve a large number of feature vectors of high dimensionality and a large number of clusters. Such an application is the codebook design required in image compression based on vector quantization, indicates that fuzzy C-means(also known as K-means) algorithm can design high quality codebooks, regardless of their initialization[5,6].

FCM algorithm applies to neutral face images that are used as a template to extract sparse feature points from edges of major face components on expression images (happiness, surprise, sadness, disgust, fear, satisfaction, comfort, distress, tiredness, worry).

FCM considers the set X formed by M feature vectors from an n-dimensional Euclidean space, $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$, $\mathbf{x}_i \in \Re^n \quad \forall i = 1,2,\ldots,M$. The clustering is based on the assignment of the feature vector $\mathbf{x}_i \in X$ into c clusters, which are represented by $\mathbf{c}_i \in \Re^n \quad \forall i = 1,2,\ldots,c$. The degree of the assignment of the feature vector $\mathbf{x}_i \in X$ into various clusters is measured by the membership function $u_{ij} \in [0,1]$, which satisfy the properties

$$\sum_{i=1}^{c} u_{ij} = 1, \ \forall_j = 1,\ldots,M \,. \tag{3.1}$$

Fuzzy C-means algorithm was developed by solving the minimization problem. The cost function for FCM is

$$J(U, c_1, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{M} u_{ij}^m d_{ij}^2 , \tag{3.2}$$

c_i is the cluster center of fuzzy group i ; $d_{ij} = \| \mathbf{c}_i - \mathbf{x}_j \|$ is the Euclidean distance between ith cluster center and jth data point ; and $m \in [1 < m, \infty]$ is a weighting exponent. The necessary conditions for Equation (3.2) to reach a minimum are

$$c_i = \frac{\sum_{j=1}^{M} u_{ij}^m X_j}{\sum_{j=1}^{M} u_{ij}^m} , \tag{3.3}$$

and

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} . \tag{3.4}$$

Fig.5 shows a result of the sparse coding with picture point of coefficient the most similar to average value of coefficients in each cluster by FCM algorithm. The number of clusters is decided in the range that can reflect the same topological relationship as major face components in human vision.

Fig.5. Sparse coding using FCM algorithm

After extracting the sparse feature points on neutral faces, extracted feature points are used as a template to extract sparse feature points from features on expression images extracted previously since each neutral face plays a standard role to decide the degree of expression change against an expression image.

To match point to point feature points on an expression face against each feature point on a neutral face, it consists of two different domains, which are called the neutral domain(N) and the expression domain(E).

The neutral domain and expression domain contain the jets of the Gabor transformation we defined in (2.1). The Gabor jet refers to the set of Gabor magnitudes obtained by sampling the image at the point $\vec{x}_i$ with sampling functions of all sizes(frequencies) and orientations, and refer to $\vec{J}(\vec{x}_i)$.

The linking procedure is performed under the constraint that the matching points found in the expression face have approximately the same topological relations as preselected points in the neutral image.

A match point should be chosen in the neutral face and then compute the Euclidean distance between the preselected point in neutral face and each point in the expression image. We compute the Euclidean distance, $\vec{\Delta}_{ij}^{NE} = \vec{x}_i^{N} - \vec{x}_j^{E}$, with i of preselected point in neutral face and j of center point in each cluster by FCM on an expression image. This evaluates the quality of local topological preservation.

The dynamical linking of points is guided by a function S which determines the similarity between neutral face jet, $\vec{J}_i^{N}$ and expression image jet, $\vec{J}_i^{E}$. Our entire wavelet family consists of 2 frequency bands, the wave number $k = \|\vec{k}\| = (\pi/4, \pi/8)$ using inverse pixels and angles at intervals of $\pi/6$ from 0 to π.

$$S(\vec{J}_i^{N}, \vec{J}_i^{E}) = \frac{\vec{J}_i^{N} \cdot \vec{J}_i^{E}}{\|\vec{J}_i^{N}\| \|\vec{J}_i^{E}\|} \tag{3.5}$$

Evaluation of match quality is done with following stages:

1) A match point should be chosen in the neutral face. Continue until a match point of the neutral face arrives at the final point.

2) Calculate the Euclidean distance and then if $\vec{\Delta}_{ij}^{NE}$ arrives at the minimum value, we choose to a point on an expression face having approximately the same topological relations as preselected point in the neutral image.

3) For much more correct matching, compute the similarity between a neutral face jet and an expression image jet using Equation (3.5).

4) If S arrives at the maximum value, the point on an expression face will be accepted as the corresponding point of preselected point in the neutral image. Once selected clusters in an expression image are excluded from next matching. Go to step 1.

We can arrive at a fast matching to points on expression images because of considering only points in the selected cluster on the expression image. (see Fig.6)

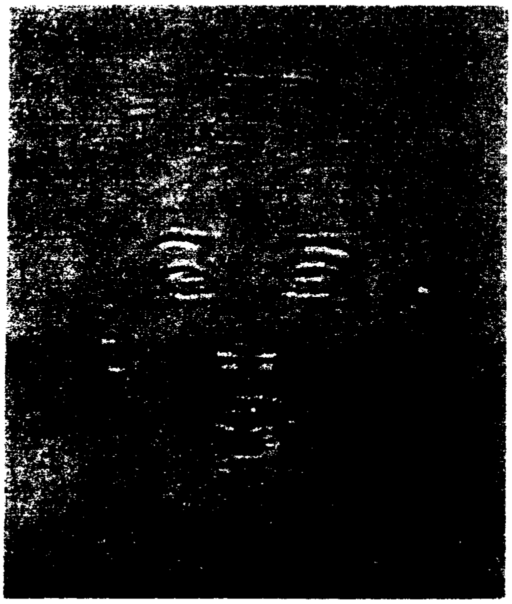


Fig.6. Sparse pixel points extracted on an expression image.

4. Facial expression recognition

Our system for facial expression recognition uses a 3-layer neural network. The first layer is 60 coordinate positions within the x-axis and y-axis positions derived with dynamic linking which are normalized by size in a standard 640 by 480 image. The second layer is 240 hidden units and the third layer is 2 output nodes to recognize the two dimensions. Training applies error back propagation algorithm[7] which is well known to the pattern recognition field. 180 images for training and 25 images excluded from the training set for testing are used. The first test verifies with the 180 images trained already. Recognition result produced by 180 image trained previously showed almost 100% recognition rates. The rating result of facial expressions derived from the semantic rating of emotion words by experimental subjects is compared with experimental results of a neural network(NN). The similarity of recognition result between human and NN is computed as follows.

$$S(\vec{H},\vec{N}) = \frac{\vec{H}\cdot\vec{N}}{\|\vec{H}\|\|\vec{N}\|}\min\left(\frac{\|\vec{H}\|}{\|\vec{N}\|},\frac{\|\vec{N}\|}{\|\vec{H}\|}\right)$$

Dimension values of human and NN on each two dimension are given as vectors of $\vec{H}$ and $\vec{N}$.

Tab.1 describes a degree of similarity of expression recognition between human(H) ard NN with 25 expression images for testing on two-dimensional structure of emotion. In Tab.1, the result of expression recognition of NN is matched to the most nearest emotion word in 44 emotion words related internal states, which looks very similar to the result of expression recognition of human. The result of the dimension analysis of 44 emotion word related internal states is shown Fig.7.

The pleasure-displeasure dimension and the arousal-sleep dimension of NN are quite a correspondence with the values in each dimension of human being. Above all, expression data of the high level of arousal and that of displeasure dimension have well recognized by neural network.

5. Discussion

This system demonstrates a facial expression recognition on the two-dimensional structure of emotion. Global structure of facial expression recognition is represented by matching topology between sparse points in neutral face of local template and points in an expression image.

The system can not be stated to generalize as a result for various expressions recognition, but we describe a study that implements a much more generalized condition than preceding studies.

In a pleasure-displeasure dimension, the degree of arousal can make an effect on discriminating from facial expressions and the combination in displeasure dimension with high level of arousal can be well recognized by neural network because that is possible to detect easily the quantity of physical changes: The degree of open state of mouth and eye [8]. Therefore, the relative importance of dimension can have an effect on facial expression recognition on the two-dimensional structure of emotion. This corresponds to the study of Kim Jinkwan et al.[9] recently.

The future work will deal with the facial expression recognition with many more various emotion words in a larger database.

References

[1] P. Ekman. Universal and cultural difference in facial expressions of emotions. In J. K. Cole(Ed), Nebraska symposium on motivation, pp.207-283. Lincoln: University of Nebraska Press, 1972.

[2] K. Younga. et al. The study of dimension of internal states through vocabulary analysis about emotion. '97 emotion conference of Korea, pp.209-219, 1997.

[3] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America,vol.2, No7,pp.1160-1169, 1985.

[4] J. C. Bezdek. Fuzzy mathematics in pattern classification. PhD thesis, Applied Math. Center, Cornell University, Ithaca, 1973.

[5] N. B. Karayiannis. MECA: Maximum Entropy Clusterings Algorithms. Systems, Neural Nets, and Computing Technical Report No. pp.94-106, University of Houston, June 1994.

[6] N. B. Karayiannis and P.-I Pai. Fuzzy vector quantization algorithms and their application in image compression. IEEE Transactions on Image Processing, in press, 1995.

[7] P. D. Wasserman. Neural computing theory and practice, Van Nosrand Reinhold, New York, pp.43-59,1989.

[8] H. Jaehyun and C. Chansup. Mappimg facial expression onto internal states. Korean Journal of the Science of Emotion and Sensibility, Vol.1. No.1. pp.41-58, 1998.

[9] K. Jinkwan. et al. Validating the stability of two-dimensional structure of emotion. Korean Journal of the Science of Emotion and Sensibility, Vol.2. No.1. pp.43-52, 1999.

Emotion word	Human(H)		NN		Recognition on NN	Similarity (H & NN)
	P-DP	A-S	P-DP	A-S		
happiness	1.65	7.53	2.11	7.09	delight	0.95
			3.26	1.06	comfort	0.22
surprise	4.65	7.8	4.55	8.29	surprise	0.95
			4.61	7.67	surprise	0.98
sadness	7.22	6.57	5.48	7.74	confusion	0.94
			6.47	1.29	emptiness	0.58
disgust	7.93	6.74	3.89	2.63	resting	0.45
			6.65	5.23	emptiness	0.56
			6.41	7.1	hate	0.91
fear	7.25	6.77	7.68	6.03	distress	0.98
			2.21	1.79	comfort	0.29
			6.32	7.01	confusion	0.95
satisfaction	1.85	4.65	2.43	2.87	comfort	0.71
			4.33	2.87	sleepiness	0.79
			2.3	4.15	lightheartedness	0.94
comfort	2.61	2.98	1.94	1.63	comfort	0.63
			5.15	3.36	vacantness	0.62
			2.07	2.03	comfort	0.73
distress	7.46	6.29	3.89	2.63	resting	0.48
			4.28	5.81	hope	0.72
			7.66	7.39	chagrin	0.91
tiredness	5.44	2.2	6.64	5.66	shyness	0.64
			7.12	2.02	loneliness	0.79
worry	7.4	5.96	4.23	1.78	resting	0.46
			6.30	6.16	shyness	0.92

Tab.1. The result data of expression recognition between human and NN

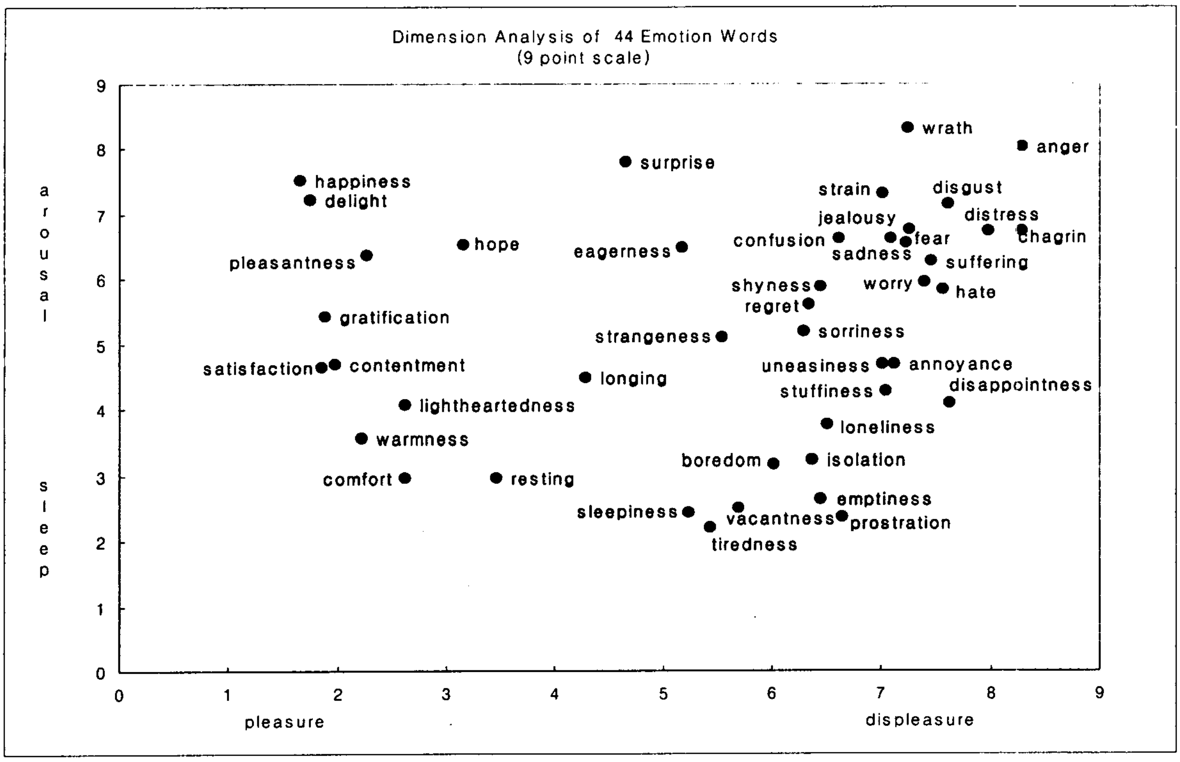


Fig. 7. Dimension model