

Facial Expression Recognition Based on the Two-Dimensional Structure of Affect

Young-suk Shin

Department of Information and telecommunication Engineering, Chosun University,
#375 Seosuk-dong, Dong-gu, Gwangju, 501-759, Korea
ysshin@mail.chosun.ac.kr

Abstract. We present an expression recognition system based on the two-dimensional structure of affect. The system is capable of identifying the various emotions using automated feature extraction. A method for extracting information about facial expressions from images is presented in three steps. In the first step, Gabor wavelet representation is constructed to provide edge extraction of major face components using the average value of the image's 2-D Gabor wavelet coefficient histogram. In the second step, sparse features of facial expression image are extracted using fuzzy C-means clustering(FCM) algorithm on neutral faces. In the third step, features of facial expressions are extracted using the Dynamic Linking Model(DLM) on expression images. The result of facial expression recognition is compared with dimensional values of internal states derived from semantic ratings of words related to emotion by experimental subjects. The two-dimensional structure of affect recognizes not only six facial expressions related to six basic emotions (happiness, sadness, surprise, anger, fear, disgust), but also expressions of various internal states.

1 Introduction

Face is an important social stimulus in human interactions. Specially, facial expression plays a major role in human communication. If a computer can understand emotions from human's facial expressions, it is possible to help humans in various situations dynamically.

Currently, most facial expression recognition systems use the six principle emotions of Ekman [1]. Ekman considers six basic emotions: happiness, surprise, fear, anger, disgust, sadness; and categorizes facial expressions with these six basic emotions. Facial action coding system is an analysis of facial muscle actions which make facial expressions based on the six basic emotions [2]. Most research on facial expression recognition includes studies using the basic emotions of Ekman[3, 4, 5, 6, 7], therefore these studies have limitations for recognition of natural facial expressions which consist of several other emotions and many combinations of emotions. Here we describe research extended on the dimension model of internal states for recognizing not only facial expressions of basic emotions but also expressions of various emotions.

Previous work on facial expression processing includes studies using representation based on optical flow from image sequences [3, 8, 9], principle components

analysis of single image [9,10], physically based models [11], and wavelets transformation[12]. These methods are similar in that they first extract some features from the images, then these features are used as inputs into a classification system.

The first stage detects the edges of major face components, using the average value of the image's 2-D Gabor wavelet coefficient histogram on all the images. Since Gabor vectors with neighboring pixels are highly correlated and redundant, it is sufficient to use sparse pixels on a face. The second stage, FCM clustering algorithm is used to select sparse pixels from edges of major facial components extracted previously from a neutral face of each expressor. The third stage is an application of the Dynamic Link Architecture [13] and is used here to detect sparse local features on expression images from preselected points in the neutral face. Lastly, from a multi-layer perceptron we recognize the facial expressions based on the two-dimensional structure of affect.

2 Database

The images used in this study were obtained from the Korean facial expression database for mapping of facial expressions into internal states [14]. This database consists of 500 facial expression images of males and females under well controlled lighting condition. Expressions were divided into two dimensions(pleasure-displeasure and arousal-sleep dimension) according to the study of internal states through the semantic analysis of words related with emotion by Kim et al. [15] using expressive 83 words.

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 person. The 11 expressions are happiness, surprise, sadness, disgust, fear, satisfaction, comfort, distress, tiredness, worry(including neutral face). A few of these are shown in Fig. 1. Our paper shows recognition of facial expressions using 2 dimension. The result of the dimension analysis of 44 emotion words related to internal emotion states is shown Fig. 2.



Fig. 1. Examples from the facial expression database

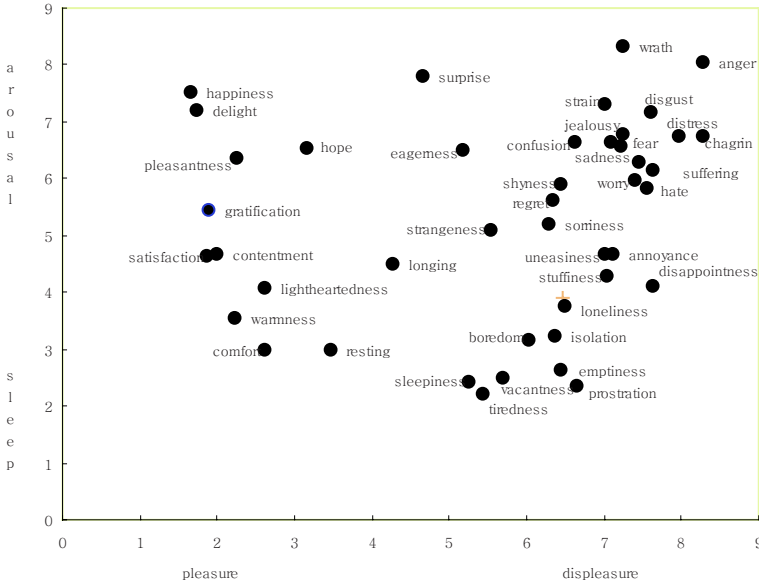


Fig. 2. The two-dimensional structure of affect

3 Automatic Feature Extraction

To extract information of facial expression, we use 287 images of 6 person, each image using 640 by 480 pixels included face images almost in the frontal pose. Original images have been rescaled and cropped such that the eyes are roughly at the same position with a distance of 60 pixels in the final image. In this section, we present the process of an automatic feature extraction. The process consists of three steps.

3.1 Preprocessing with Gabor Wavelets

For edges of major facial components, an average value of the image’s 2-D Gabor wavelet coefficient histogram is used. Each image was convolved with both even and odd Gabor kernels in an image $I(\vec{x})$ around a given pixel $\vec{x} = (x, y)$. The general form of two dimensional Gabor wavelets function is given by Daugman [16].

$$\psi_{\vec{k}}(\vec{x}) = \frac{k^2}{\sigma^2} \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right) [\exp(i\vec{k} \cdot \vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right)] \tag{1}$$

The wave vector \vec{k} of length $k \equiv \|\vec{k}\|$ defines the spatial wavelength and at the same time controls the width of the Gaussian window. The parameter σ denotes the width of the Gaussian window relative to the wavelength corresponding to k . To

detect features of major face components, we use a specific frequency band, a wave number, $k=0.78$, and 5 distinct orientations in 22.5° steps between 0 and π , and chose $\sigma = \pi$. The complex valued ψ_k applied to each image combines an even and odd part. We use only the magnitudes because they represent local information of an image in a smoothly varying way. Let G be the set of Gabor functions to be applied to I . G is G_1, G_2 . The computation proceeds as follows:

$$\begin{aligned} \omega_1 &= \sum \sum G_1 I, & \omega_2 &= \sum \sum G_2 I \\ \bar{\omega} &= \sqrt{(\omega_1^2 + \omega_2^2)}. \end{aligned} \tag{2}$$

Fig. 3(a) 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 local facial 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. 3(b) shows the resulting image which applied an optional value to an average value of the Gabor coefficients histogram.

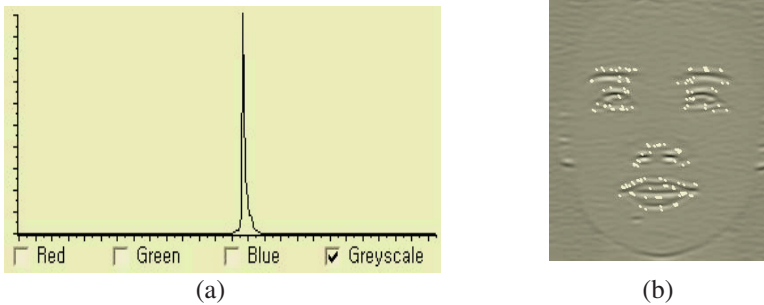


Fig. 3. (a) 2-D Gabor coefficient histogram. (b) Extracted edges of major face components

3.2 Sparse Feature Points Extraction Using FCM Clustering Algorithm on Neutral Face

Extracted feature points are similar to edges of major facial components. Since Gabor vectors with neighboring pixels are highly correlated and redundant, it is sufficient to use sparse pixels on a face. We thus pick out sparse feature points based on the FCM clustering algorithm in edges extracted from the 2-D Gabor wavelet coefficient histogram. FCM algorithm applies to neutral facial images that are used as a template to extract sparse feature points from edges of major facial components on expression images.

Fuzzy C-means clustering [17] is a data clustering algorithm in which each data point belongs to a cluster to a degree specified by a membership grade. The potential-

ity of fuzzy clustering algorithms can be demonstrated by their application in clustering tasks which involve a large number of feature vectors of high dimension and a large number of clusters. Such an application is the codebook design required in image compression based on vector quantity, regardless of their initialization [18].

FCM consider the set X formed by N feature vectors from an N dimensional Euclidean space, $\mathbf{x}_i \in \mathfrak{R}^n \quad \forall i = 1, 2, \dots, N$, $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$. The clustering is based on the assignment of the feature vector $\mathbf{x}_i \in X$ into c clusters, which are represented by $\mathbf{c}_i \in \mathfrak{R}^n, \quad \forall i = 1, 2, \dots, c$. 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

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, N. \tag{3}$$

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

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^N u_{ij}^m d_{ij}^2. \tag{4}$$

c_i is the cluster center of fuzzy group i ; $d_{ij} = \|\mathbf{c}_i - \mathbf{x}_j\|$ is the Euclidean distance between i th cluster center and j th data point ; and $m \in [1 < m, \infty]$ is a weighting exponent. The necessary conditions for Equation (4) to reach a minimum are

$$\mathbf{c}_i = \frac{\sum_{j=1}^N u_{ij}^m X_j}{\sum_{j=1}^N u_{ij}^m} \tag{5}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}. \tag{6}$$

We determined sparse feature points using the following steps: **Step1.** Initialize the membership matrix U with random values between 0 and 1 such that the constraints in Equation (3) are satisfied. **Step2.** Calculate c fuzzy cluster centers ($\mathbf{c}_i, i = 1, 2, \dots, c$) using Equation (5). **Step3.** Compute the cost function according to Equation (4), and stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold. **Step4.** Compute a new U using Equation (6), then go to Step2. Fig. 4(a) shows a result that extracted sparse pixel points by FCM algorithm: $c=60, m=2$. The number of clusters is decided in the range that can reflect the same topological relationship as major face components in human vision.

3.3 Sparse Feature Points Extraction Using DLM on Expression Image

After extracting the sparse feature points on neutral faces, which are used as a template to extract sparse feature points from edges on the 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 expression domain contain the jets of the Gabor transformation we defined in Equation (1) and Equation (2). 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 refers to \vec{J}_i^E . The entire wavelet family for our study consists of two frequency bands, the wave number $k = \left\| \vec{k} \right\| = (\pi/4, \pi/8)$ using inverse pixels and seven different orientations from 0° to 180° , differing in 30° steps.

The linking procedure is performed under the constraint that the matching points found in the expression face have approximately the same topological relations as the preselected points in the neutral image. From the assumption, before starting a dynamic linking we apply FCM algorithm to edges of major face components on the expression images in the expression domain with the same cluster number applied for clustering edges of neutral facial image. A coarse-to-fine matching process performed to locate facial features on an expression image as follows.

1. A matching point should be chosen in the neutral face and then computed in the Euclidean distance between the preselected point in neutral face and the centroid point of each cluster clustered in the expression image in $\vec{\Delta}_{ij}^{NE} = \vec{x}_i - \vec{x}_j$. Thus, we expect to speed up mapping feature points between neutral face and expression face.
2. After selecting a cluster with the centroid point of a minimum Euclidean distance in an expression image, again we compute the Euclidean distance between all the points in the cluster selected in an expression image and the preselected point in neutral face. Here, we choose the candidate points in the cluster selected in an expression image with the minimum Euclidean distance or more the minimum Euclidean distance calculated in the step 1.
3. Among the candidate points in the cluster selected in an expression image should be chosen as a final point. We calculate a distance value between an average (\mathcal{E}) of the Gabor transformation's jets of all the points(n) in the cluster including the candidate points in an expression image, $\mathcal{E} = \frac{1}{n} \sum_{i=1}^n \vec{J}_i^E$, and the Gabor

transformation's jet of each candidate point in an expression image, \vec{J}_i^E . Thus, a candidate point with a minimum distance value on the expression image is chosen as a final corresponding feature point against a preselected point in the neutral face. From the steps 2 and 3, we can reflect the local topology of non-rigid distortion of the face caused from changing expression. For the resulting the algorithm mentioned above see Fig. 4(b).

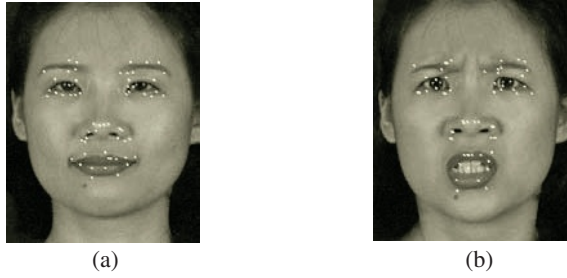


Fig. 4. (a) Sparse pixel points extracted with FCM algorithm on neutral face. (b) Sparse pixel points extracted with DLM on expression image

4 Recognition and Discussion

The system for facial expression recognition uses a three-layer neural network. The first layer is the distance values from each feature point on a neutral face to each feature point on an expression face which are normalized by size from 0 to 1. The second layer is 240 hidden units and the third layer is two output nodes to recognize the two dimensions: pleasure-displeasure and arousal-sleep.

Training applies error back propagation algorithm which is well known to the pattern recognition field. The activation function of hidden units uses the sigmoid function. 250 images for training and 37 images excluded from the training set for testing are used. The first test verifies with the 250 images trained already. Recognition result produced by 250 images trained previously showed 100% recognition rates. The rating result of facial expressions derived from the semantic rating of emotion words by subjects is compared with experimental results of a neural network (NN). The similarity of recognition result between human and NN is computed in Equation (7). The dimension values of human and NN in each two dimension are given as vectors of \vec{H} and \vec{N} .

$$S(\vec{H}, \vec{N}) = \frac{\vec{H} \cdot \vec{N}}{\|\vec{H}\| \|\vec{N}\|} \min\left(\frac{\|\vec{H}\|}{\|\vec{N}\|}, \frac{\|\vec{N}\|}{\|\vec{H}\|}\right) \quad (7)$$

Table 1 describes a degree of similarity of expression recognition between human and NN on the two-dimensional structure of affect. In Table 1, the result of expression recognition of NN is matched to the most nearest emotion word in 44 emotion words related to internal emotion states. The result of expression recognition of NN looks very similar to the result of expression recognition of human. The displeasure emotions of high level arousal dimension seems like fairly generalized and stabled emotions in expression space. Such expressions are surprise, disgust, fear, distress, and worry. It seems that the quantity of physical changes can be detected easily. High level arousal dimension in the pleasure dimension also seems like a important component to discriminate the expression images. For instance, satisfaction, tiredness, and comfort show low level arousal dimension (sleep dimension) in the pleasure dimension, while happiness shows high level arousal dimension in the pleasure dimension.

Table 1. The result data of expression recognition between human and NN(Pleasure-Displeasure: P – D, Arousal-Sleep: A – S)

Emotion words	Human(Mean)		Neural Network		Similarity	Recognition on Neural Network
	P – D	A – S	P – D	A – S		
happiness	1.65	7.53	3.88	3.44	0.54	lightheartedness
			4.92	4.6	0.71	boredom
			2.86	5.86	0.82	pleasantness
			1.31	5.69	0.75	gratification
			4.43	4.8	0.73	Longing
satisfaction	1.85	4.65	1.49	6.07	0.79	pleasantness
			2.14	4.96	0.92	contentment
			6.32	5.9	0.52	shyness
comfort	2.61	2.98	5.0	5.7	0.52	strangeness
			3.65	3.64	0.77	lightheartedness
sadness	7.22	6.57	7.07	5.23	0.89	shyness
			3.7	6.37	0.72	hope
			6.62	7.12	0.91	surprise
tiredness	5.44	2.2	7.94	6.29	0.56	strain
			4.06	4.05	0.90	sleepiness
			4.39	4.28	0.89	longing
			4.8	5.09	0.76	strangeness
			6.39	5.65	0.65	uneasiness
worry	7.4	5.96	6.89	6.09	0.97	confusion
surprise	4.65	7.8	7.39	6.84	0.94	strain
			4.55	8.29	0.95	surprise
			4.61	7.67	0.98	surprise
disgust	7.93	6.74	4.65	5.60	0.79	hope
			6.35	3.42	0.68	isolation
			7.33	6.14	0.91	hate
			7.68	6.03	0.98	distress
			6.05	6.72	0.86	surprise
Fear	7.25	6.77	6.75	4.49	0.80	sadness
			6.43	5.21	0.83	stiffness
			6.68	7.97	0.94	disgust
			7.30	7.96	0.91	chagrin
			5.91	4.17	0.72	isolation
distress	7.46	6.29	7.48	7.16	0.94	disgust
			4.28	5.81	0.72	hope
			4.77	4.97	0.70	boredom
			5.60	4.11	0.71	boredom
			5.81	5.05	0.79	strangeness

This study is a new approach of human’s emotion processing, it is interesting to note in this context that machine vision may represent various emotions similar to human with the combination of each dimension in the internal emotion states. To future study we are planning to recognize the expressions with person independent and a wider range of emotions in much larger database than present system.

References

1. Ekman, P.: Universal and cultural difference in facial expressions of emotions. In: J. K. Cole(Ed.), Nebraska symposium on motivation, Lincoln: University of Nebraska Press, (1972) 207-283
2. Ekman, P., Friesen, W.V.: Facial action coding system. Consulting Psychologists, Palo Alto, CA., (1977)
3. Mase, K.: Recognition of facial expression from optical flow. IEICE Transactions, E 74, **10** (1991) 3473-3483
4. Essa, I., Pentland, A.:Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, **19** (1997) 757-763
5. Lien, J.: Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity. Ph.D. Thesis, Carnegie Mellon University, (1998)
6. Oliver, N. Pentland, A., Berard, F.: LAFTER:a real-time face and lips tracker with facial expression recognition. Pattern Recognition **33** (2000) 1369-1382
7. Cohen, I., Sebe, N., Garg, A., Chen, L. S., Huang, T. S.: Facial expression recognition from video sequence:temporal and static modeling. Computer Vision and Image Understanding , In Press (2003)
8. Yacoob, Y., Davis, L.S.: Recognizing human facial expression from long image sequences using optical flow. IEEE Trans. Pattern Anal. Machine Intell. 18(6) (1996) 636-642
9. Bartlett, M., Viola, P., Sejnowski, T., Larsen, J., Hager, J., Ekman, P.: Classfying Facial Action. In: Advances in Neural Information Processing Systems 8. D. Touretzky et al. editors, MIT Press, Cambridge, MA (1996)
10. Padgett, C., Cottrell, G.: Identifying emotion in static face images. In Proceeding of the 2nd Joint Symposium on Neural Computation, 5 (1995) 91-101
11. Essa, I. Pentland, A. : Facial Expression Recognition using Visually Extracted Facial Action Parameters. Proceedings of the International Workshop on Automatic Face and Gesture Recognition (1995) 35-40
12. Lyons, M., Akamatsu, S.:Coding facial expressions with Gabor wavelets. Proceeding of the Third International Conference on Automatic Face and Gesture Recognition, (1998) 200-205
13. von der Malsburg, C.: Nervous structure with dynamical links. Ber. Bunsenges. Phy.Chem, **89** (1985) 703-710
14. Bahn, S., Hahn, J. and Chung, C.: Facial expression database for mapping facial expression onto internal state. '97 Emotion Conference of Korea, (1997) 215-219
15. Kim, Y., Kim, J., O, S., O, K., Chung, C.: The study of dimension of internal states through word analysis about emotion. Korean Journal of the Science of Emotion and Sensibility, **1** (1998) 145-152
16. Daugman, J: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America **2** (1985) 1160-1169
17. Bezdek, J.C.: Fuzzy mathematics in pattern classification. Ph.D. thesis, Applied Math. Center, Cornell University, Ithaca (1973)
18. Karayiannis, N.B., Pai, P.-I.: Fuzzy vector quantization algorithms and their application in image compression. IEEE Transactions on Image Processing, (1995)