

# Facial Expression Recognition Based on Dimension Model Using Sparse Coding

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 dimension model of internal states that is capable of identifying the various emotions using automated feature extraction. Feature vectors for facial expressions are extracted from a hybrid approach using fuzzy c-mean clustering algorithm and dynamic linking based on Gabor wavelet representation. 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 dimensional model recognizes not only six facial expressions related to six basic emotions (happiness, sadness, surprise, angry, fear, disgust), but also expressions of various internal states. In this paper, with dimension model we have improved the limitation of expression recognition based on basic emotions, and have extracted features automatically with a new approach using FCM algorithm and the dynamic linking model.

## 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. Most research on facial expression recognition includes studies using the basic emotions of Ekman[2, 3, 4, 5], 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 [6, 7], principle components analysis of single image [7,8], physically based models [9], and wavelets transformation[10]. 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.

In this paper, we use a hybrid approach for automatic feature extraction. 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. 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 [11]. This capability is used here to detect sparse local features on expression images from preselected points in the neutral face. Finally, we show the recognition of facial expressions based on the dimension model of internal states using a multi-layer perceptron.

## 2 Facial Expression Database Based on Internal States

The images used in this study were obtained from the Korean facial expression database for mapping of facial expressions into internal states [12]. 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. [13] using expressive 83 words. For experiment we 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. The result of the dimension analysis of 44 emotion words related to internal emotion states is shown in Fig. 2. The dimension model expresses a degree of expression in each of the two dimensions on a nine point scale. Our paper shows the recognition of facial expressions on dimension model.

## 3 Sparse Coding of Facial Expressions

To extract information of facial expression, we use 287 images of facial expressions, 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. For edges of major facial components, an average value of the image's 2-D Gabor wavelet coefficient histogram is used. The general form of two dimensional Gabor wavelets function is given by Daugman [14]. 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  $\sigma$  denotes the width of the Gaussian window relative to the wavelength corresponding to  $k$ .

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



Fig. 1. Examples from the facial expression database

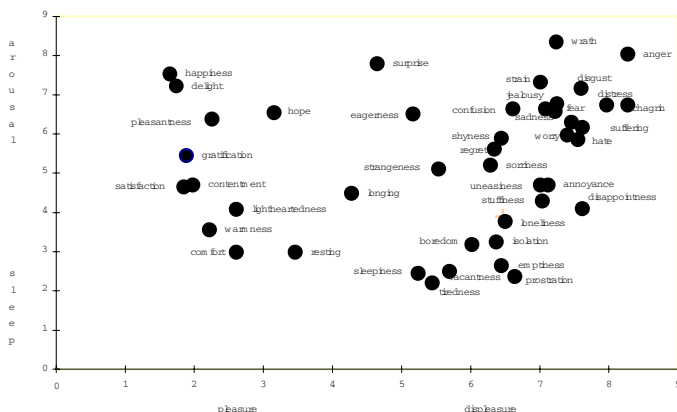
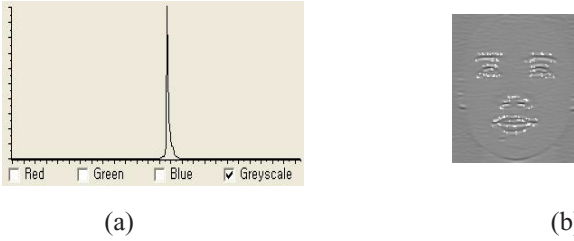


Fig. 2. Dimension Model: dimension analysis of 44 emotion words

To detect features of major face components, we use a specific frequency band, a wave number,  $k=0.78$ , and 5 distinct orientations in  $22.5^\circ$  steps between 0 and  $\pi$ , and chose  $\sigma=\pi$ . The complex valued  $\psi_k^-$  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 function  $\psi_k^-$  to be applied to  $I$ .  $G$  is  $G_1, G_2$ . The computation proceeds as follows:

$$\omega_1 = \sum \sum G_1 I, \quad \omega_2 = \sum \sum G_2 I, \quad \sigma = \sqrt{(\omega_1^2 + \omega_2^2)}$$

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

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 is used as a template to extract sparse feature points from edges of major facial components on expression images.

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 dimension and a large number of clusters[15]. Fuzzy C-means clustering [16] is a data clustering algorithm in which each data point belongs to a cluster to a degree specified by a membership grade. 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  $\sum_{i=1}^c u_{ij} = 1, \forall_j = 1, \dots, N$ . The cost function for FCM is  $J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^N u_{ij}^m d_{ij}^2$ .

$c_i$  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  $J(U, c_1, \dots, c_c)$  to reach a minimum are

$$c_i = \frac{\sum_{j=1}^N u_{ij}^m X_j}{\sum_{j=1}^N u_{ij}^m} \text{ and } u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}.$$

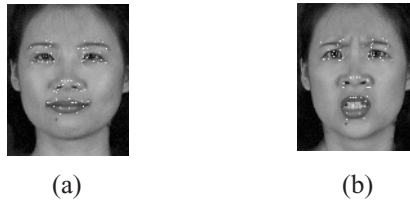
We determined sparse feature points using the following steps: **Step1.** Initialize the membership matrix  $U$  with random values between 0 and 1 such that the constraints in  $\sum_{i=1}^c u_{ij} = 1$  are satisfied. **Step2.** Calculate  $c$  fuzzy cluster centers ( $\mathbf{c}_i, i = 1, 2, \dots, c$ ) using  $c_i$ . **Step3.** Compute the cost function according to  $J(U, c_1, \dots, c_c)$ , 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  $u_{ij}$ , 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.

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. The Gabor jet  $\vec{J}(x_i)$  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. Sparse feature extraction using DLM on expression images is guided by a function  $S$  in  $S(\vec{J}_i^N, \vec{J}_i^E) = \vec{J}_i^N \cdot \vec{J}_i^E / \left\| \vec{J}_i^N \right\| \left\| \vec{J}_i^E \right\|$  which determines the similarity between neutral face jet,  $\vec{J}_i^N$  and expression image jet,  $\vec{J}_i^E$ . The entire wavelet family 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  $0^\circ$  to  $180^\circ$ , differing in  $30^\circ$  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. A match point should be chosen in the neutral face and then computed in the Euclidean distance between the preselected point in neutral face and each point in the expression image in  $\vec{\Delta}_{ij}^{NE} = \vec{x}_i^N - \vec{x}_j^E$ . This evaluates the quality of local topological preservation. The dynamic linking of selected points in the neutral face image to points in the expression image is formulated as an optimization problem. The cost function  $H$  in  $H = \vec{\Delta}_{ij}^{NE} + \sum S(\vec{J}_i^N, \vec{J}_j^E)$  to be optimized measures the quality of proposed point matches. We chose for cost function the special form. The feature on the expression images was accepted if the cost function  $H$  satisfies two conditions at the same time : (1) Reach to the minimum value. (2) Do not exceed a maximum distance value that the matching points found in the expression face have approximately the same topological relations as the preselected points in the neutral image(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 Facial Expression Recognition

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

$$s(\bar{H}, \bar{N}) = \frac{\bar{H} \cdot \bar{N}}{\sqrt{\bar{H} \cdot \bar{H}} \sqrt{\bar{N} \cdot \bar{N}}} \min \left( \frac{\bar{H}}{\bar{N}}, \frac{\bar{N}}{\bar{H}} \right)$$

ues of human and NN in each two dimension are given as vectors of  $\bar{H}$  and  $\bar{N}$ .

Table 1 describes a degree of similarity of expression recognition between human and NN on two-dimensional structure of emotion. 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(see Table 1).

### 5 Discussion and Conclusion

This paper presents an expression recognition system based on dimension model of internal states using sparse coding. Facial expression on dimension model includes Two dimensions which are pleasure to displeasure dimension and arousal to sleep dimension. The result of expression recognition of NN looks very similar to the result of expression recognition of human. Above all, the expression images of the high level of arousal and displeasure emotion have been most effectively recognized by neural network. In a pleasure-displeasure dimension, the degree of arousal could make an effect on discriminating facial expressions like happiness, satisfaction, and comfort. The combination in displeasure dimension with the high level of arousal dimension could be well recognized by neural network. Such expressions are fear, surprise, distress, worry, and disgust. These results appear to have an effect on physical changes between neutral pattern and expression pattern in major facial components.

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.

This study was supported by research funds from Chosun University, 2003.

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

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

## 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. Lien, J.: Automatic recognition of facial expressions using hidden Markov models and estimation of expression intensity. Ph.D. Thesis, Carnegie Mellon University, (1998)
3. Oliver, N. Pentland, A., Berard, F.: LAFTER:a real-time face and lips tracker with facial expression recognition. Pattern Recognition 33 (2000) 1369-1382
4. Tian, Y.L, Kanade, T., & Cohn, J. F.: Recognizing Action Units for Facial Expression Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), (2001) 97-116

5. 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)
6. 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
7. Bartlett, M., Viola, P., Sejnowski, T., Larsen, J., Hager, J., Ekman, P.: Classifying Facial Action. In: *Advances in Neural Information Processing Systems 8*. D. Touretzky et al. editors, MIT Press, Cambridge, MA (1996)
8. Padgett, C., Cottrell, G.: Identifying emotion in static face images. In *Proceeding of the 2<sup>nd</sup> Joint Symposium on Neural Computation*, 5 (1995) 91-101
9. 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
10. 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
11. von der Malsburg, C.: Nervous structure with dynamical links. *Ber. Bunsenges. Phy.Chem*, 89 (1985) 703-710
12. 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
13. 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
14. 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
15. Karayiannis, N.B., Pai, P.-I.: Fuzzy vector quantization algorithms and their application in image compression. *IEEE Transactions on Image Processing*, (1995)
16. Bezdek, J.C.: Fuzzy mathematics in pattern classification. Ph.D. thesis, Applied Math. Center, Cornell University, Ithaca (1973)