A Neural Network Model for Classification of Facial Expressions Based on Dimension Model

Young-Suk Shin

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

Abstract. We present a new neural network model for classification of facial expressions based on dimension model that is illumination-invariant and without detectable cues such as a neutral expression. The neural network model on the two-dimensional structure of emotion have improved the limitation of expression recognition based on a small number of discrete categories of emotional expressions, lighting sensitivity, and dependence on cues such as a neutral expression.

1 Introduction

The work in facial expressions for human-computer intelligent interaction did not start until the 1990s. Models for recognizing facial expressions have traditionally operated on a short digital video sequence of the facial expression being made, such as neutral, then happy, then neutral[1,2,3]. All require the person's head to be easily found in the video. Therefore, continuous expression recognition such as a sequence of "happy, surprise, frown" was not handled well. And the expressions must either be manually separated, or interleaved with some reliably detectable cues such as a neutral expression, which has essentially zero motion energy.

In this paper, we present a new neural network model on the two dimensional structure of emotion for classification of facial expressions that is illumination-invariant and without detectable cues such as a neutral expression.

2 Facial Expression Representations for Invariant-Illumination and Neutral Expression

The face images used for this research were centered the face images with coordinates for eye and mouth locations, and then cropped and scaled to 20x20 pixels. The luminance was normalized in two steps. First, a "sphering" step prior to principal component analysis is performed. The rows of the images were concatenated to produce 1×400 dimensional vectors. The row means are subtracted from the dataset, X . Then X is passed through the zero-phase whitening filter, V , which is the inverse square root of the covariance matrix:

$$V = E\{XX^T\}^{-\frac{1}{2}}, W = XV \quad (1)$$

This indicates that the mean is set to zero and the variances are equalized as unit variances. Secondly, we subtract the local mean gray-scale value from the sphered each patch. From this process, W removes much of the variability due to lightening. Figure 1(a) shows the cropped images before normalizing. Figure 1(b) shows the cropped images after normalizing.

Fig. 1. (a) Images before normalizing. (b) Images after normalizing. (c) PCA representation only included the first 1 principle component. (d) PCA representation excluded the first 1 principle component

In a task such as facial expression recognition, the first 1 or 2 principal components of PCA do not address the high-order dependencies of the facial expression images, that is to say, it just displays the neutral face. Figure 1(c) shows PCA representation that included the first 1 principle component. But selecting intermediate ranges of components that excluded the first 1 or 2 principle components of PCA did address well the changes in facial expression (Figure 1(d)).

Therefore, to extract information of facial expression regardless of neutral expression, we employed the 200 PCA coefficients, P_n , excluded the first 1 principle component of PCA of the face images. The principal component representation of the set of images in W in Equation(1) based on P_n is defined as $Y_n = W * P_n$. The approximation of W is obtained as:

$$\overline{W} = Y_n * P_n^T. \quad (2)$$

The columns of $\overline{W}$ contains the representational codes for the training images (Figure 1(d)). The representational code for the test images was found by $\overline{W}_{test} = Y_{test} * P_n^T$. Best performance for facial expression recognition was obtained using 200 principal components excluded the first 1 principle component.

3 Recognition

The face images used for this research were a subset of the Korean facial expression database[4]. The data set contained 500 images, 3 females and 3 males, each image

Fig. 2. A rating result of facial expression recognition in Pleasure-Displeasure dimension (Str.,strangeness;Des.,despair;Del.,delight;Pro.,proud;Sle.,sleepiness;Bor.,boredom; Con.,confusion;Lik.,likable;Ted.,tedious;Int.,intricacy;Reg.,regret;Lon.,loneliness; Cry.,crying;War.,warmness;Hap.,happiness; Dep.,depression)

Fig. 3. A rating result of facial expression recognition in Arousal-Sleep dimension

using 640 by 480 pixels. 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 Younga et al. [5] using 83 expressive words. The system for facial expression recognition uses a three-layer neural network. The first layer contained the representational codes derived in Equation (2). The second layer was 30 hidden units and the third layer was two output nodes to recognize the two dimensions: Pleasure-Displeasure and Arousal-Sleep.

Training applies an error back propagation algorithm. The activation function of hidden units uses the sigmoid function. 500 images for training and 66 images excluded from the training set for testing are used. The 66 images for test include 11 expression images of each six people. The first test verifies with the 500 images trained already. Recognition result produced by 500 images trained previously showed 100% recognition rates. The rating result of facial expressions derived from 9 point scale on two dimension for degrees of expression by subjects was compared with experimental results of a neural network(NN).

Figure 2 and 3 show the correlation of the expression recognition between human and NN in each of the two dimensions. The statistical significance of the similarity for expression recognition between human and NN on each of the two dimensions was tested by Person correlation analysis. The correlation in the Pleasure-Displeasure dimension between human and NN showed 0.77 at the 0.01 level and 0.51 at the 0.01 level in the Arousal-Sleep dimension.

4 Conclusion

Our results allowed us to extend the range of emotion recognition and to recognize on dimension model of emotion with illumination-invariant without detectable cues such as a neutral expression. We propose that the inference of emotional states within a subject from facial expressions may depends more on the Pleasure-Displeasure dimension than Arousal-Sleep dimension. It may be analyzed that the perception of Pleasure-Displeasure dimension may be needed for the survival of the species and the immediate and appropriate response to emotionally salient, while the Arousal-Sleep dimension may be needed for relatively detailed cognitive ability for the personal internal states.

Acknowledgements. This study was supported by research funds from Chosun University, 2004.

References

1. Oliver, N. Pentland, A., Berard, F.: LAFTER:a real-time face and lips tracker with facial expression recognition. *Pattern Recognition* **33** (2000) 1369-1382
2. Cohen, I., Sebe, N., Garg, A., Chen, L. S., Huang, T. S.: Facial expression recognition from video sequence. *Proc. Int'l Conf. Multimedia and Exp(ICME)* (2002) 121-124
3. Cohen, I. :Semisupervised learning of classifiers with application to human-computer interaction. PhD thesis, Univ. of Illinois at Urbana-Champaign (2003)
4. Saebum, B., Jaehyun, H., Chansub, C.: Facial expression database for mapping facial expression onto internal state. '97 Emotion Conference of Korea, (1997) 215-219
5. Younga, K., Jinkwan, K., Sukyung, P., Kyungja, O., Chansub, 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