Happy-Sad Expression Recognition Using Emotion Geometry Feature and Support Vector Machine

Linlu Wang, Xiaodong Gu*, Yuanyuan Wang, and Liming Zhang

Department of Electronic Engineering, Fudan University, Shanghai 200433, P.R. China
linlu_wang@yahoo.com.cn, guxiaodong@263.net, xdgu@fudan.edu.cn

Abstract. Currently human-computer interaction, especially emotional interaction, still lacks intuition. In health care, it is very important for the medical robot, who assumes the responsibility of taking care of patients, to understand the patient's feeling, such as happiness and sadness. We propose an approach to facial expression recognition for estimating patients' emotion. Two expressions (happiness and sadness) are classified in this paper. Our method uses a novel geometric feature parameter, which we call the Emotion Geometry Feature (EGF). The active shape model (ASM), which can be categorized mainly for non-rigid shapes, is used to locate Emotion Geometry Feature (EGF) points. Meanwhile, the Support Vector Machine (SVM) is used to do classification. Our method was tested on a Japanese Female Facial Expression (JAFFE) database. Experimental results, with the average recognition rate of 97.3%, show the efficiency of our method.

1 Introduction

During the past three decades, facial expression analysis has attracted more and more attention in the computer vision field for its many applications, such as human-machine interaction, image understanding, synthetic face animation [1], and web services. Facial expressions reflect not only emotions, but also other mental activities, social interaction and physiological signals [2]. Therefore, in health care, it is important for a medical robot that helps the patient, to recognize the patients' emotion through facial expression [3].

Even though much work has been done, facial expression recognition with a high accuracy is very difficult due to the non-rigidity and complexity of facial expression. The facial expressions were generally defined by psychologists as a set of six basic facial expressions [4], including anger, disgust, fear, happiness, sadness, and surprise. In this paper we mainly recognize the happy expression and sad expression.

Many research efforts have been put into facial expression recognition. A survey on the facial expression recognition can be found in [5] and [6]. In [6], Pantic and Rothkrantz surveyed the research work done in automating facial expression analysis. In [5], Fasel and Luettin introduced the most prominent automatic facial expression analysis methods and systems presented in the literature. In addition, they also discussed facial motion and deformation extraction approaches and classification

* Corresponding author.

M. Köppen et al. (Eds.): ICONIP 2008, Part II, LNCS 5507, pp. 535–542, 2009.

methods in [5]. According to the approach used for facial-expression feature extraction, methods about facial expression recognition can be distinguished as the feature-based method [7, 8] and the template-based method [9, 10].

Since the Facial Action Coding System (FACS), which is a system designed to describe changes in the facial expression in terms of observable activations of facial muscles [12], was developed by Ekman and Friesen [11] to code facial expressions by Action Units (AUs), several research efforts [12, 13] have been made to recognize facial expression based on the Facial Action Coding System (FACS). The Active Shape Model proposed by Cootes [14], is one of the sophisticated deformable template models to detect facial features [15]. However traditional facial features extracted by ASM cannot accurately represent facial expressions.

So, motivated by FACS and ASM, we propose a model of facial expression recognition for estimating patients' emotion based on EGF, which we introduce in this paper. Emotion Geometry Feature (EGF) concentrates on expression features, rather than facial features, to form a facial recognition system.

The rest of the paper is organized as follows. Section 2 introduces our facial expression recognition model. In this model, we introduce a novel expression feature recognizer, EGF. The Active Shape Model, which is used to locate Emotion Geometry Feature (EGF) points, is also described simply in Section 2. Using the Japanese Female Facial Expression (JAFFE) database as the test data, experimental results and analysis are shown in Section 3. Conclusions are drawn in Section 4.

2 Our Facial Expression Recognition Approach

Fig. 1 shows the flow diagram of the proposed model. Our model can be divided into three parts: one for the creation of The Active shape Model with a set of training images; two for the location of facial expression features; three for the extraction of EGF and the classification of two expressions.

ASM created by the training samples can search new facial expression feature points in the testing images. After the location of feature points we make efforts to analyze the difference between happiness expressions and sadness expressions and extract EGF including shape coordinates, the distance between upper lip and lower lip and the geometric angle between the line $\overline{\mathbf{O}_i\mathbf{A}_i}$ and the line $\overline{\mathbf{O}_i\mathbf{B}_i}$ (Fig. 3). Before applying SVM, it's very important to scale the input parameters [16]. An important objective is to avoid numerical difficulties during the calculation [17]. Ultimately classified facial expressions are obtained.

2.1 Overview of ASM

Cootes et al. [14] introduced the Active Shape Model (ASM), a method of fitting a set of local feature detectors to an object. The ASM procedure starts with a prior knowledge about the object shape, so that the model can extract the object shape in a new image by locating the outline of the object. So, we can divide the ASM procedure into two steps made up of modeling and matching.

A linear shape model can be generated by a set of manually labeled training image points, the formula is,

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi}_s \mathbf{b}_s, \tag{1}$$

where $\mathbf{x}$ is the synthesized shape, $\overline{\mathbf{x}}$ is the mean of shapes, and $\mathbf{b}_s$ is a set of shape model parameters, and $\mathbf{\Phi}_s$ is the matrix which consists of eigenvectors of the covariance matrix of the training set, and is an orthogonal matrix.

Virtually, the matching is to locate the outline points of the object by searching in the testing image by using the model built. In searching, we can obtain the optimal parameters for location and shape of the face by comparing the reference model from the training set to a new test image [24].

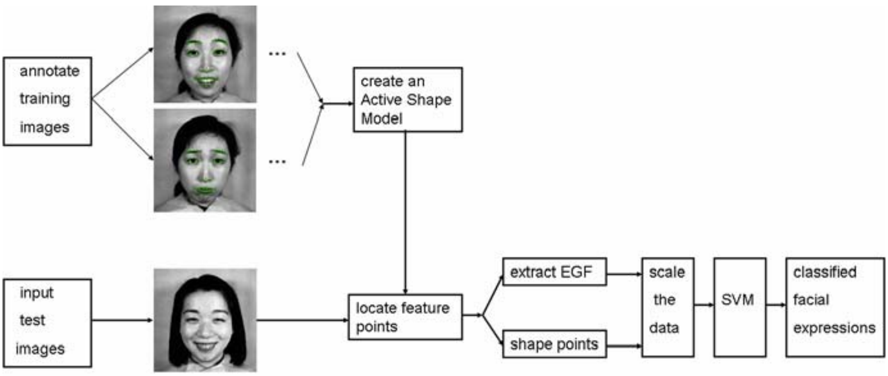

Fig. 1. The flow diagram of the proposed model

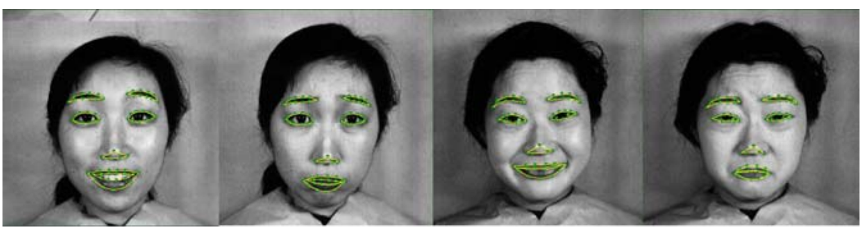

Fig. 2. The examples of labeled training data, and each of them is labeled with 66 points around the mouth, the nose tip, eyes, and eyebrows

2.2 Proposed Emotion Geometry Feature

In this section we compare happiness expressions with sadness expressions, and introduce the concept of the Emotion Geometry Feature (EGF) containing the angle and the distance, and the position of shape features. Here, the angle feature and the distance feature are the novel ones we mainly propose in terms of facial expression recognition in this paper, and the position information can be found in other papers.

1). a and b in Fig. 3 illustrate that a mouth which is deformable in a happiness image is wider than one in a sadness images, but the nose tip is rigid in different circumstances. So we consider the nose tip, and two corners of the mouth as three vertices of a triangle. The coordinates of three vertices are,

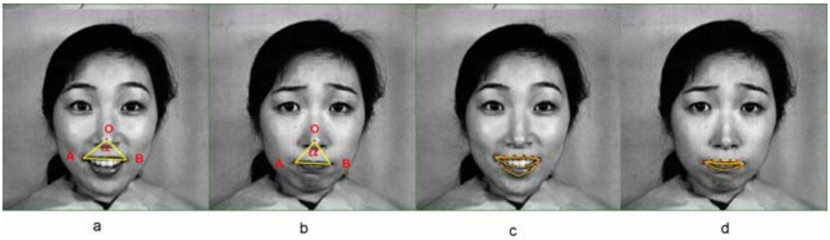

Fig. 3. Images a, b are labeled O, A, B on nose tip, and corners of mouth, respectively, and images c, d are labeled around the mouth. a and c are images with happiness expression, meanwhile, b and d are samples with sadness expression

$$\mathbf{O}_i = (x_i^0, y_i^0)^T, \mathbf{A}_i = (x_i^1, y_i^1)^T, \mathbf{B}_i = (x_i^2, y_i^2)^T, i = 1, 2, \cdots, N, \tag{2}$$

where N is the number of facial images, $x_i^j, y_i^j, j = 0,1,2,$ are the x, y coordinate of the three vertices in the ith images.

We can obtain EGF by the angle α_i between the line $\overrightarrow{\mathbf{O}_i\mathbf{A}_i}$ and the line $\overrightarrow{\mathbf{O}_i\mathbf{B}_i}$ to distinguish widths of the mouth with different expressions. And α_i can be calculated as,

$$\alpha_i = \arccos\left(\frac{\overrightarrow{\mathbf{O}_i\mathbf{A}_i} \bullet \overrightarrow{\mathbf{O}_i\mathbf{B}_i}}{\left|\overrightarrow{\mathbf{O}_i\mathbf{A}_i}\right|\left|\overrightarrow{\mathbf{O}_i\mathbf{B}_i}\right|}\right), \tag{3}$$

where α_i is angle between line $\overrightarrow{\mathbf{O}_i\mathbf{A}_i}$ and $\overrightarrow{\mathbf{O}_i\mathbf{B}_i}$, 'arccos' is the arc cosine transformation function, '$\bullet$' is the inner product of two vectors.

2). In Fig. 3, we obtain another EGF, which is denoted by the distance between upper lip and lower lip by four EGF points from images c and d. Because when you are happy or sad, your mouth is generally open or closed, respectively. The distance can be calculated by the difference of the coordinates of upper lip points and lower lip points.

3). Fig. 2 shows four examples of labeled training data, and each of them is labeled with 66 points around the mouth, the nose tip, eyes, and eyebrows. Using these 66 points can obtain a vector for planar image shapes,

$$\mathbf{x} = [x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n]^T, \tag{4}$$

where (xi, yi) represents the ith landmark point coordinate, n is the total number of points (n=66). After the the searching procedure using ASM, we can obtain a set of new shape positions represented by (4). As a position feature, the shape vector is also used as an input to SVM with other EGF.

2.3 Overview of Support Vector Machine

The Support Vector Machine is used to do classification of two expressions in this paper. Here we describe SVM simply.

The SVM algorithm [18] needs to solve the following optimization problem,

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i , \tag{5}$$

Subject to
$$y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+b) \geq 1-\xi_i ,$$
$$\xi_i \geq 0, i = 1,\cdots,l.$$

where C is the term which penalizes the training error, b is the bias for the SVM, ξ_i is the ith slack variable vector, $\mathbf{x}_i \in R^n, i = 1,\cdots,l$ is the ith training vector, $y_i \in \{-1,1\}$, is the ith class label, and ϕ is a nonlinear mapping function by which training vector $\mathbf{x}_i$ is mapped into a higher dimensional space.

Using a kernel function, SVM projects training samples onto a high dimensional feature space where these samples can be divided linearly.

The decision function is given as,

$$\mathrm{sgn}(\sum_{i=1}^{l} y_i\alpha_i K(\mathbf{x}_i,\mathbf{x})+b) , \tag{6}$$

here, α_i are the Lagrange multipliers of a dual optimization problem. Once (6) is obtained, classification of unseen test data is achieved.

The selection of a suitable kernel function is very important for the classification of SVM. The basic kernel is given as follows,

1). Radial Basis Function (RBF) kernel: $K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\gamma\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2), \gamma > 0.$

2). 'd' degree polynomial kernel: $K(\mathbf{x}_i,\mathbf{x}_j) = (\gamma\mathbf{x}_i{}^{\mathrm{T}}\mathbf{x}_j + r)^d, \gamma > 0.$

3). Sigmoid kernel: $K(\mathbf{x}_i,\mathbf{x}_j) = \tanh(\gamma\mathbf{x}_i{}^{\mathrm{T}}\mathbf{x}_j + r).$

3 Experimental Results and Discussions

In experiments, our method was applied to a Japanese Female Facial Expression (JAFFE) [19] database to evaluate the performance of classification of facial expressions. The database contains 213 images of 7 facial expressions including 6 basic facial expressions and 1 neutral posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects [19].

Because we concentrate on the classification of happiness expression and sadness expression, 60 images including happiness and sadness of the JAFFE database are selected for training and testing. Thus, of the 60 images, 30 are happiness images and another 30 are sadness images, and 10 Japanese females are included for both expressions.

The Active Shape Model is generated by 55 images randomly selected from 60 images. The EGF acquired from the shape coordinates of facial expression feature points obtained from the fitting of ASM are used as an input to SVM. An RBF kernel is our

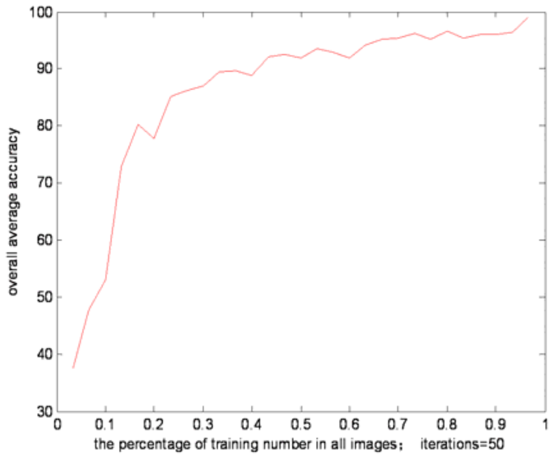

Fig. 4. The x-axis represents the number of training samples used as a ratio of the 60 total available samples for both expressions, and the y-axis represents the overall average accuracy for 50 iterations

first choice for the SVM system, and we use the cross-validation method to find the best parameters C and γ.

In the training and testing of SVM, we randomly select training samples from the set of 30 for each expression and vary the number of training samples from 1 to 29 for each expression. So the total ordinal training samples are 2, 4, … , 58. The testing is done on the remaining unused samples. The procedure is repeated for 50 iterations for each training sample size. We obtain the overall mean accuracy for each respective training sample size and plot the mean accuracy as a function of training sample size in Fig. 4. And this figure shows that the more training samples that are used, the higher the accuracy that can be obtained. From the figure, if we select 80% or even more as the training sets, the overall average accuracy is very high and does not have large fluctuation. So, we see the 80% as a very important turning point.

In experiments, ultimately 80% of samples for each class are used for training data, while the remaining samples form the test data. The overall mean accuracy using our proposed method is 97.3%. The happiness and sadness recognition rates are 98.5% and 96.1% respectively. Actually, 97.3% is one half of the sum of 98.5% and 96.1%.

In our experiments, we compare the classification performances of the proposed approach with other methods on the JAFFE database. For happiness expression recognition, the average accuracies in the papers [20], [21], [22] and [23] (six basic expressions are considered in these papers) are 50%, 50%, 30% and 70%, respectively, and the accuracy using our method is 98.5%. Meanwhile, for sadness expression recognition, the mean recognition rates in the above papers are 70%, 60%, 90%, and 60% respectively, and the recognition rate based on our method is 96.1%. In the paper [20], [21], their methods are based on the person-similarity weighted expression feature, but the similarities of persons obtained by face recognition algorithm are rough, thus the “true expression” feature can not be estimated accurately [20]. Nevertheless we extract distinguishable features, EGFs we propose, in terms of expression recognition in this paper. So we can get a relatively better result than theirs.

4 Conclusions

Experimental results based on the Japanese Female Facial Expression (JAFFE) database show that the approach to facial expression recognition for estimating patients' emotion we proposed in this paper, can classify happiness and sadness expressions with high accuracy. In this approach, we introduce a novel geometry feature, Emotion Geometry Feature (EGF), and Active Shape Model is used to extract the Emotion Geometry Feature points, and SVM ultimately classifies happiness and sadness expressions. We plan to extend happiness expression and sadness expression to six basic expressions further by adding more expression features to EGF.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 60671062 and 60571052) and National Basic Research Program of China (2005CB724303), and Shanghai Leading Academic Discipline Project (No. B112).

References

1. Aleksic, P.S., Katsaggelos, A.K.: Automatic facial expression recognition using facial animation parameters and multiStream HMMs. IEEE Trans. Inf. Forensics Secur. (2006)
2. Fasel, B., Luettin, J.: Automatic facial expression analysis: A survey. Pattern Recognition 36, 259–275 (2003)
3. Gholam Hosseini, H., Krechowec, Z.: Facial Expression Analysis for Estimating Patient's Emotional States in RPMS. In: Proc. of the 26th Annual International Conference of the IEEE, San Francisco, CA, USA, pp. 1517–1520 (2004)
4. Ekman, P., Friesen, W.V.: Emotion in the Human Face. Prentice-Hall, Englewood Cliffs (1975)
5. Fasel, B., Luettin, J.: Automatic Facial Expression analysis: a survey. Pattern Recognition 36, 259–275 (2003)
6. Pantic, M., Rothkrantz, L.L.M.: Automatic analysis of facial expressions: The state of the art. IEEE Trans. Pattern Ana. and Machine Intelligence 22, 1424–1455 (2000)
7. Guo, G., Dyer, C.R.: Learning from examples in the small sample case: Face expression recognition. IEEE Trans. Syst., Man, Cybern. B, Cybern. 35, 477–488 (2005)
8. Ma, L., Khorasani, K.: Facial expression recognition using constructive feedforward neural networks. IEEE Trans. Syst., Man, Cybern. B, Cybern. 34, 1588–1595 (2004)
9. Essa, I.A., Pentland, A.P.: Facial expression recognition using a dynamic model and motion energy. In: Int. Conf. Computer Vision, Cambrdige, MA (1995)
10. Bartlett, M.S., Littlewort, G., Braathen, B., Sejnowski, T.J., Movellan, J.R.: An approach to automatic analysis of spontaneous facial expressions. In: 5th IEEE Int. Conf. Automatic Face and Gesture Recognition, Washington, DC (2002)
11. Ekman, P., Friesen, W.V.: The facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, Inc., San Francisco (1978)

12. Pantic, M., Rothkrantz, L.J.M.: Facial action recognition for facial expression analysis from static face images. IEEE Trans. Sys., Man, Cybern. B: Cybern 34, 1449–1461 (2004)
13. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Analysis and Machine Intelligence 27, 699–714 (2005)
14. Cootes, T.: Introduction to active shape models, technical report (1998)
15. Kwo, W., Kin, L., Kit, N.: An accurate active shape model for facial feature extraction. In: Proc. Int. Sym. Int. Multimedia, Video and Speech, pp. 109–112 (2004)
16. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001), http://www.csie.ntu.edu.tw/~cjlin/libsvm
17. Hsu, C.-W., Chang, C.-C., Lin, C.-J.: A Practical Guide to Support Vector Classification (2008)
18. Vapnik, V.: Statistical learning theory. Wiley, New York (1998)
19. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: 3rd International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
20. Tan, H., Zhang, Y.-J.: Person-Similarity Weighted Feature for Expression Recognition. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part II. LNCS, vol. 4844, pp. 712–721. Springer, Heidelberg (2007)
21. Tan, H., Zhang, Y.: Person-Independent Expression Recognition Based on Person Similarity Weighted Distance. Jounal of Electronics and Information Technology 29, 455–459 (2007)
22. Wang, H., Ahuja, N.: Facial Expression Decomposition. In: ICCV, pp. 958–965 (2003)
23. Tian, Y., Kanade, T., Cohn, J.: Recognizing Action Units for Facial Expression Analysis. IEEE Trans. On PAMI 23, 97–115 (2001)
24. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models- their training and application. Computer Vision and Image Understanding 61, 38–59 (1995)