

Automatic Facial Feature Points Extraction and Expression Recognition Based on Video Database

Duo Feng^{1, a}, Shun Nishide^{1, b} and Fuji Ren^{1, c}

¹Faculty of Engineering, Tokushima University, 2-1 Minami-Josanjima, Tokushima 770-8506, Japan

^ac501437025@tokushima-u.ac.jp, ^bnishide@is.tokushima-u.ac.jp, ^cren@is.tokushima-u.ac.jp

Keywords: Expression Recognition, Support Vector Machine, Video Sequence Processing

Abstract. In this paper, we propose a facial expression recognition method using Support Vector Machine on mapped coordinate sequence features. The proposed system is almost fully automatic, in which landmark initialization is based on general knowledge with edge information, and missing information compensation is done by ASM. The geometric features of facial expressions were extracted from sequences of facial landmarks. Validation experiments were conducted using facial expression sequences extracted from a video based facial expression database.

1. Introduction

Facial expression recognition is a vital research field for human-machine interaction. Psychological research has shown that facial expressions are the most expressive means to display human emotions (facial expression 55%, vocal information 38%, and verbal information 7% [1]). Despite the activeness of work in scientific community for more than 20 years, development of an accurate facial expression recognition using machine is still a challenging issue. A summary of the history with current issues is shown in the meta-review on facial expression recognition [2].

An important issue on facial expression recognition is extraction of appropriate facial features. Two types of features are generally used, which are appearance-based features and geometry-based features [3]. Appearance-based features describe the texture of a face during an expression, whereas geometry-based features are vectors representing the shape and location of facial components, such as eyebrows, eyes, nose, and mouth. We utilize geometry-based features in our research.

Two types of models exist to track geometry-based features for facial expression recognition, active shape model (ASM) [4] and active appearance model (AAM) [5]. ASM is statistical model of the shape of objects, represented by a set of points, which iteratively deforms to fit to an example of the object in a new image. The shapes are constrained by PDM (point distribution model) Statistical Shape Model to deform only in ways seen in a training set of labeled examples. On the other hand, AAM uses shape and appearance to statistically model objects to match a new image. The two models require hundreds of images with corresponding coordinates, usually obtained manually, of landmarks in all images as training data. We utilize ASM for the feature tracking model.

In this paper, we propose a facial expression recognition method using Support Vector Machine (SVM) to classify sequential geometry-based features tracked by ASM. Occlusions of feature points are compensated by the ASM model. Normalization of feature points in the sequential data is done to deal with variations in location and inclination of the face. Facial expression recognition experiments were conducted with video sequences to validate the effectiveness of the method.

2. Automatic feature points extraction

In 1978, Ekman et al.[6] developed Facial Action Coding System(FACS), which measures facial expressions, to help psychologists with face behavior analysis. Based on the experiments using FACS, it has been found that eyebrows, eyes, nose, and mouth are the most related to changes in facial expression. Therefore, we create a 38-point model, shown in Fig. 1, to express the four main parts of the face. The location, change, and distance of facial parts can be extracted from the model.

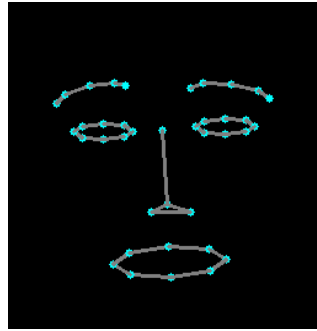


Fig.1. 38-point model of face

Figure 2 shows the overall process of landmark extraction using general knowledge of facial parts with edge information. The following processes are conducted.

- (1) Extract face region in image using Haar-like feature based detection method as in [7].
- (2) Estimate the area of facial parts based on general knowledge (e.g. eyebrows over eyes).
- (3) Calculate edge using contour detection from specific area of facial parts from Haar-like feature.
- (4) Modify edge information using edge detection and morphological transformation.
- (5) Reconstruct occluded area from relation of landmark location from ASM.

Contour detection cannot always detect the complete edge information. Fig. 3(a) shows an edge extraction example of eyes and eyebrows where portions of eyebrows are missing. Fig. 3(b) shows the result of edge detection and morphological transformation to connect the edges. By combining these methods, the complete facial parts can be extracted as in Fig. 3(c).

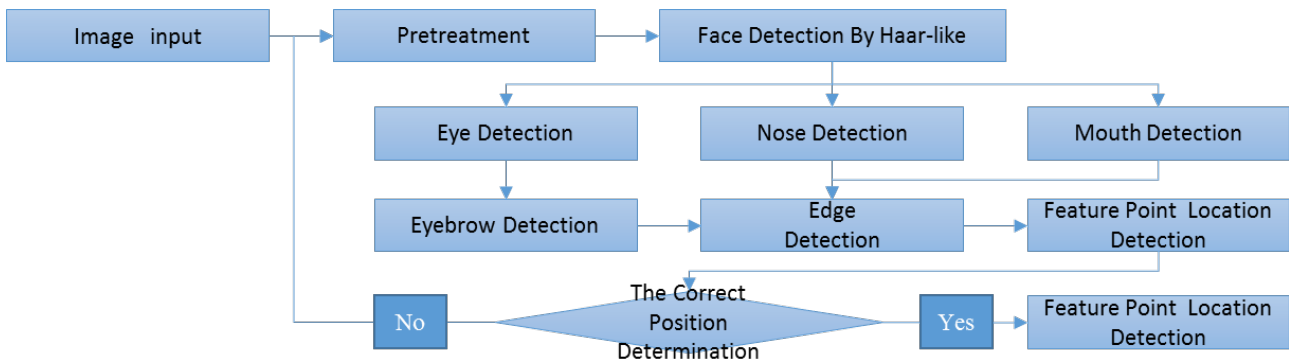


Fig.2. Overall process of landmark extraction by every frame



(a) Area of eyes and eyebrows (b) Edge detection and morphological transformation of eyebrows (c) Complete extraction result

Fig.3. An example of facial part detection

Some images contain occlusions where a portion of the facial part is concealed by another. Figure 4(a) shows an example where a portion of the eyebrow is occluded. We use the landmark location relationships obtained using 400+ images with good results as training data of ASM for inferring the missing landmark position. Figure 4(b) shows the fixed result using ASM.



Fig.4. Extraction example when facial parts are occluded

3. Facial expression recognition

3.1 Selection of representative frames in video

The proposed recognition method requires a same number of processing frames for each video sequence. Therefore, we use the standard deviation and inter-frame difference of facial feature points to select representative frames from the video sequence.

Using ASM, the set of calculated feature points of the f th frame of the video is calculated as

$$S_f = \{(x_1, y_1)_f, (x_2, y_2)_f, \dots, (x_{n-1}, y_{n-1})_f, (x_n, y_n)_f\}, \quad (1)$$

where S_f and $(x_i, y_i)_f$ are the tracking result and i th landmark coordinate position in the f th frame of the video, respectively, and n is the number of landmark in a frameset to 38 in our experiment. Denoting the i th landmark in the f th frame of expression sequence as P_{if} , (1) can be rewritten as

$$P_i = \{(x_i, y_i)_1, (x_i, y_i)_2, \dots, (x_i, y_i)_{N-1}, (x_i, y_i)_N\}, \quad (2)$$

$$S_f = \{P_{1f}, P_{2f}, \dots, P_{(n-1)f}, P_{nf}\}. \quad (3)$$

Focusing on the average feature value of the expressionless face \bar{P}_i , the difference between the feature in each frame from the expressionless face σS_f can be calculated as

$$\bar{P}_i = \overline{(x_i, y_i)} = \frac{(x_i, y_i)_1 + (x_i, y_i)_2 + \dots + (x_i, y_i)_{N-1} + (x_i, y_i)_N}{N}, \dots i = 1 \sim n, \quad (4)$$

$$\sigma S_f = \sum_{i=1}^n \text{dis}[\bar{P}_i, P_{if}], \quad (5)$$

where N is the number of the frames in the expressionless face. The difference between the current frame with the preceding frame, ΔS_f , can be calculated as

$$\Delta S_f = \sum_{i=1}^n \text{dis}[P_{if}, P_{i(f-1)}]. \quad (6)$$

The representative frames (10 frames in this paper) are obtained from σS_f and ΔS_f for processing.

3.2 Features generation

In the proposed approach, we focus on video based facial expression, where the expression changes from expressionless to a certain facial expression. While image based facial expression recognition cannot deal with small changes of facial expression, our method using video can use contextual information to focus on minor changes.

Expression sequences vary based on individuality and the magnitude of expressions. Therefore, we implement a normalization method to reduce such influences. In addition, features of each independent frame are processed to represent the role of the frame in the overall sequence.

Assuming that the expressionless part exists in the beginning of the sequence, we calculate the coordinate sequence of each feature point P_i , the average value \bar{P}_i , and the deviation σP_i of each expressionless feature points. P_i and \bar{P}_i are calculated as (2) and (4). The σP_i is calculated as

$$\sigma P_i = \frac{[\bar{P}_i - P_{i1}]^2 + [\bar{P}_i - P_{i2}]^2 + \dots + [\bar{P}_i - P_{iN}]^2}{N}, \quad (7)$$

where N is the number of processing frames in the expressionless part.

Normalizing the average value \bar{S} to 0, the deviation σS_f to 1, the coordinate sequence P_i is changed to a normalized value sequence P'_i , given as

$$P'_i = (P_i - \bar{P}_i) * (1/\sigma P_i), \quad (8)$$

where P'_{if} is the normalized value of the i th feature point of the f th frame in the sequence.

The features of each frame compose a feature pool for the normalized value S'_f .

$$S'_f = \{P'_{1f}, P'_{2f}, \dots, P'_{(n-1)f}, P'_{nf}\}. \quad (9)$$

The same calculation process is conducted for extracting feature points of other facial expressions.

The normalized features in each frame shown by S'_f not only normalize individuality of facial expressions, but also express the change of feature points in the entire sequence, which enables expression recognition using a single frame.

4. Experiment

4.1 Database description

As we focus on Ekman's six facial expression model [8], we utilize a facial expression video database [9] for our experiment, composed of six basic facial expression classes (anger, disgust, fear, happiness, sadness, and surprise) for 40 people in the 20's to 40's (26 men and 14 women). The database consists of three sets of data per facial expression (in different magnitudes) per person, for a total of 18 data per person. The duration of the video from expressionless to a certain expression was five to seven seconds, incorporating the expressionless face as the onset to peak formation of a certain expression. The videos were digitized to 720×480 pixel arrays. 364 videos were selected from the database for the experiment.

4.2 Facial expression recognition using SVM

SVM is a linear classification method, which searches for a separating hyperplane with the widest margin between two different data categories. We utilize a publically available implementation of SVM, libsvm [10], using radial basic function (RBF) kernel, and grid search strategy [11] for optimal parameter selection.

Two experiments were conducted to evaluate the model. The first experiment uses the coordinate sequence of landmarks S_f , calculated as (3), as feature vectors for classification of facial expressions using SVM. The second experiment uses normalized coordinate sequence S'_f , calculated as (9), as feature vectors to evaluate the effects of normalization. Since S_f and S'_f are coordinate sequences, complex numbers can be used to represent coordinate information, changing (2) to

$$P_i = \{(x_i + y_i \cdot i)_1, (x_i + y_i \cdot i)_2, \dots, (x_i + y_i \cdot i)_{N-1}, (x_i + y_i \cdot i)_N\}, \quad (10)$$

Note that there is no change in calculation equation of S_f , P'_i and S'_f from (3), (8) and (9). The dimensionality of feature vectors in the two experiments for SVM classification was $n=38$. Tables 1 and 2 show the results of the two experiments on facial expression recognition.

The average recognition accuracy was 82.83% and 89.12%, for the first and second experiments, respectively. It is notable that there is an improvement of 6.29 points for facial expression recognition using SVM with normalized coordinate sequence S'_f , over facial expression recognition using coordinate sequence of landmarks S_f .

Table 1. Facial expression recognition for first experiment using S_f as features. (in percentages)

	Anger	Fear	Happiness	Surprise	Sadness	Disgust
Anger	71.92	0.2	24.68	0.56	1.25	1.39
Fear	0.18	77.58	18.71	1.87	0.2	1.46
Happiness	0.16	0.16	98.09	0.39	0.74	0.46
Surprise	0	0.85	16.35	81.13	1.33	0.34
Sadness	0.12	0	13.71	0.76	84.46	0.95
Disgust	0.61	0.45	14.54	0.15	0.48	83.77

Table 2. Facial expression recognition for second experiment using S'_f as features. (in percentages)

	Anger	Fear	Happiness	Surprise	Sadness	Disgust
Anger	83.37	0.6	2.81	2.29	8.69	2.25
Fear	0.51	87.38	2.26	1.45	6.98	1.42
Happiness	0.75	0.56	90.85	1.76	4.82	1.25
Surprise	0.15	0.98	2.81	88.96	6.09	1
Sadness	0.60	0.92	2.13	1.11	94.16	1.08
Disgust	0.76	0.17	2.04	0.61	6.45	89.98

4.3 Comparison with State-of-the-Art Methods

The recognition accuracy achieved by the video based facial expression database is comparable with the best accuracy in other works.

The system created by Kotsia has shown by far the best performance achieving 99.7% recognition rate [12]. However, landmark initialization in [12] is done manually, and the number of landmarks is larger than the number of landmarks in our method. In addition, our method is almost fully automatic, and data processing is simple and efficient.

The system created by Ghimire achieved 97.35% recognition rate, using multi-class AdaBoost with DTW similarity between feature vectors and SVM on boosted features [13]. In their method, 800 dimensional features were used to achieve the best accuracy. In our method, only 38 dimensional features are used for the facial expression recognition model.

Therefore, the achievement of 89.12% recognition rate using normalized coordinate feature sequences, is sufficiently comparable to other works.

5. Conclusion

In this paper, we proposed a facial expression recognition method by SVM, using normalized coordinate sequence feature recognition. The proposed method is almost fully automatic, in which landmark initialization is done based on general knowledge and edge information of facial parts, and fixed missing information is compensated by ASM. Geometric feature sequences were extracted from tracking results of facial landmark coordinates. Facial expression recognition was done using representative frames selected using standard deviation of frames and inter-frame difference. Experiments have shown that usage of normalized feature points could achieve a recognition rate of 89.12%, 6.29 points higher than that of normal feature points.

Although the experiments showed effectiveness of the method to other works, several issues remain to be dealt with. First, recognition rate of facial expression degrades when the change of facial expression is small. Second, an improvement of the model is required to achieve a higher recognition rate by including appearance features and adding constraints in the feature vectors. As real-time facial expression recognition is a demanding need for conversation with robot, we plan to implement our model to online applications. We believe that our model will contribute to active studies on human-computer interaction, and ultimately, to realize a human-robot coexisting society.

Acknowledgment

This research has been partially supported by JSPS KAKENHI Grant Number 15H01712.

References

- [1] Mehrabian, A. Communication without words, *Psychol. Today* 1968, 2, 53–56.
- [2] Valstar, M.F.; Mehu, M.; Jiang, B.; Pantic, M.; Scherer, K. Meta-analysis of the first facial expression recognition challenge, *IEEE Trans. Syst. Man. Cybern. B Cybern.* 2012, 42, 966–979.
- [3] Y. Tian, T. Kanade, J. Cohn, *Handbook of Face Recognition*, Springer, 2005.
- [4] T. Cootes, C. Taylor, D. Cooper and J. Graham, *Active Shape Models-Their Training and*

- Application, *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.
- [5] G. Edwards, C. Taylor and T. Cootes, Interpreting face images using active appearance models, *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.
- [6] P. Ekman and W.V. Friesen., *Facial Action Coding System: Investigator's Guide*, Consulting Psychologists Press, Palo Alto, CA, 1978.
- [7] Viola, P.; Jones, M.J. Robust real-time face detection, *Int. J. Comput. Vision* 2004, 57, 137–154.
- [8] Ekman, P. Strong evidence of universal in facial expressions: A reply to Russell's mistaken critique, *Psychol. Bull.* 1994, 115, 268–287.
- [9] T. Kuroda, M. Jia, F. Ren and S. Kuroiwa, *The Construction of The Facial Expression Video Database*, 2006 International Conference on Communication Technology, 2006.
- [10] Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines, 2001. Available online:<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>. [Accessed: 29- Nov- 2015].
- [11] Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. *A Practical Guide to Support Vector Classification*; Technical Report, Department of Computer Science, National Taiwan University, Taiwan, 2010.
- [12] I. Kotsia and I. Pitas, Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines, *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172-187, 2007.
- [13] D. Ghimire and J. Lee, Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines, *Sensors*, vol. 13, no. 6, pp. 7714-7734, 2013.