Human Emotions Estimation Using Combination of 3D Average Face and LUT-AdaBoost

Jinhui Chen[1,a)] Yasuo ARIKI[1,b)] Tetsuya TAKIGUCHI[1,c)]

1. Overviews

One of the most crucial techniques associated with Computer Vision is technology that deals with facial recognition, especially, the automatic estimation of human emotions. However, in real-time facial expression recognition, when a face turns sideways, the expressional feature extraction becomes difficult as the view of camera changes and recognition accuracy degrades significantly. Therefore, quite many conventional methods are proposed, which are based on static images or limited to situations in which the face is viewed from the front. In this paper, a method that uses Look-Up-Table (LUT) AdaBoost combining with the three-dimensional average face is proposed to solve the problem mentioned above. In order to evaluate the proposed method, the experiment compared with the conventional method was executed. The experiments show promising results and very good success rates. This paper covers several methods that can improve results by making the system more robust.

2. Introduction

Within the past decade or two, the need for human emotions estimation technology in various different areas keeps pushing the research forward every year, and lots of achievements have been obtained due to its great potentials in real-life applications, such as human-computer interaction (H-CI) [7], automatic recommendations analysis based on video contents [6] and M. S. Bartlett et al. [2] proposed Gabor feature based AdaSVM method to recognize expression, and obtained a good performance on Cohen Kanade expression database [5]. But lots of difficulties, there still exist, because of the variation and complexity of the facial expression across human population and even the same individual. So human emotion recognition is an interesting research topic, and many researchers attach great importance to this field.

However many of the conventional methods above focused on the single static picture, so that they are short of practical application. Under the condition of real-time recognition, when a face turns sideways, the expressional feature extraction becomes difficult and the recognition accuracy degrades significantly. Thus, many of the conventional approaches are limited to situations in which the face is viewed from the front.

Therefore, there is a great need to solve the problem that the face cannot be moved freely. Thus, in this paper, a method is proposed to estimate human emotions based on facial expression n a real-time event when the face moves freely. We use the LUT for AdaBoost to be trained on the multi-class features efficiently. Furthermore, we use three-dimensional average face, which is recovered from the original image based on Procrustes Analysis. With these approaches, we can solve the problem that many of the conventional facial expression recognition methods are limited to situations in which the face cannot be moved freely.

3. Proposed Method

Fig. 1 shows a processing flow of the proposed method. First, the facial area is detected based on AdaBoost, using Haar-like features on the input images. Next, if the feature of facial expressions can be extracted, the facial expression on this facial area will be estimated by LUT AdaBoost; otherwise, we use 3D average face to estimate the facial expressions, and the average face model is reconstructed based on Procrustes Analysis [3] to obtain the shape of face including effective expressional features.

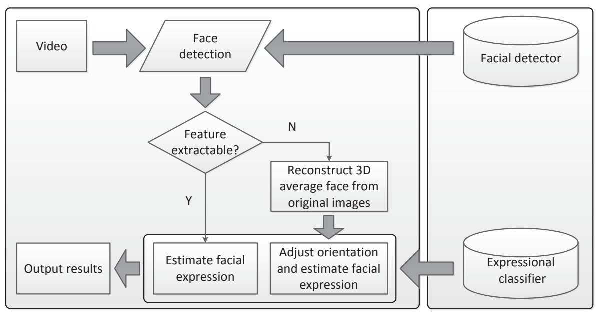

Fig. 1 Processing flow of the proposed method.

AdaBoost is a learning algorithm that selects a small number of weak classifiers from a large weak classifier pool or hypothesis space to construct a strong classifier. The two-class AdaBoost algorithm has been successfully used in face detection, gender classification, etc. However, many problems such as expression recognition are multi-class in nature.

1 Graduate School of System Informatics, Kobe University, Japan, 1-1 Rokkodai-cho, Nada Ward, Kobe 657-8501, Japan
a) ianchen@me.cs.scitec.kobe-u.ac.jp
b) ariki@kobe-u.ac.jp
c) takigu@kobe-u.ac.jp

For a multi-class, we use a real-valued 2D LUT type weak classifier, which was proposed by Y.Wang et al. [7]. But it is also limited to situations in which the face is viewed from the front. More over since their experiments were focused on the static pictures under the condition of real-time, the system does not have robustness against face direction changes. Therefor, we use three-dimensional average face to solve these problems.

In 3D model, the geometry of a face is defined as a shape vector $S_{3D} = (x_1, y_1, z_1, \cdots x_n, y_n, z_n)^T \in \mathbf{R}^3$, which contains the x, y, and z-coordinates of n vertices. The mean shape s_0 and m shape variations s_i are then obtained, and a new shape S can be expressed as a linear combination of the mean shape s_0 and the shape variations s_i as follows:

$$S = s_0 + \sum_{i=1}^{m} \beta_i s_i \tag{1}$$

where $\beta = (\beta_1, \beta_2, \cdots, \beta_m)^T$ is the shape parameter and m is the dimension of the shape parameter which was determined to represent the shape of 3D face model. Given the input face image indicated as $S_{2D} = (x_1, y_1, \cdots, x_n, y_n) \in \mathbf{R}^2$, the shape parameter β needs to be determined such that it minimizes the shape residual between the projected 3D facial shape generated by the shape parameter and the input 2D facial shape. The optimal shape and pose parameters (β, R_θ, T) are obtained from

$$E_r = \|P(R_\theta S_{3D} + T) - S_{2D}\|^2 \tag{2}$$

where S_{3D} is a $3 \times n$ matrix that is reshaped from the $3\, n \times 1$ model shape vectors obtained using (1) and S_{2D} is $2 \times m$ matrix, P is a 2×3 orthographic projection matrix, T is a $3 \times n$ translation matrix consisting of n translation vectors $t = [t_x, t_y, t_z]^T$, and R_θ is a 3×3 rotation matrix where the yaw angle is θ. The average face creation is indicated as follows:

1. Initialization: set $\beta_0 = 0$ and $k = 1$.
2. Alignment: S_{2D} is aligned with the 2D shape obtained by projecting the frontal 3D shape (s_0) onto the $x - y$ plane.
3. Update R_θ and T with the fixed shape parameter by $min \ \ \|P(R_\theta S_{3D} + T) - S_{2D}\|^2$, and reconstruct $(S_{3D})_k$ using the shape parameter β_k.
4. Verify whether $E_r \leq \mu$ or $k > N$, if not, go to Step 3 and $k = k + 1$ (in Ref. [4], Gower suggested setting $\mu = 10^{-4}$).
5. Reconstruct S_{3D} using the final shape parameters.

4. Experiments

In order to evaluate the proposed method, we compared it among the conventional method which bases on LUT AdaBoost without three-dimensional average face, and the method based on K-means [1] (k=13). Fig. 2 shows the error rate of experimental results in the facial expression estimation. The estimation maximum error rate of proposed method is 13.60%, and the best result is 2.86%. On the other hand, the conventional method keeps the error rate approximately 28.69%, and the method of K-means is about 34.06%. Table 1 shows the average processing time per one image with the size of 640 pixels $\times$ 480 pixels by PC with CPU Core i3-330M 2.13 GHZ and 2.0 GB memory. The validity of the proposed method is confirmed by these experimental results.

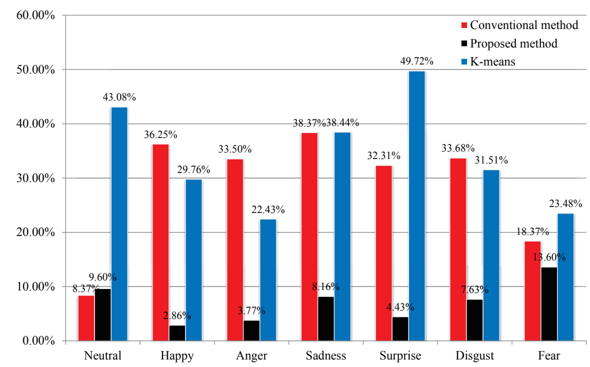

Fig. 2 Experimental results (error rate).

Table 1 Average processing time per one image.

Method	Proposed	Conventional	K-means
Time cost	19.36 ms	13.11 ms	31.70 ms

5. Conclusions

In this paper, we proposed the method combining 3D average face and ameliorated AdaBoost to recognize facial expressions against face direction changes. The experiments have shown that our approaches obtained more effective results and improved the exaction rate. Future work will considere a possible implementation in a real scenario. The research will be focussed on combining with the other techniques, just as speech recognition.

References

[1] Alldrin, N., Smith, A. and Turnbull, D.: Classifying facial expression with radial basis function networks, using gradient descent and K-means, *CSE253* (2003).

[2] Bartlett, M., Littlewort, G., Fasel, I. and Movellan, J.: Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction., *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, Vol. 5, IEEE, pp. 53–53 (2003).

[3] Goodall, C.: Procrustes methods in the statistical analysis of shape, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339 (1991).

[4] Gower, J.: Generalized procrustes analysis, *Psychometrika*, Vol. 40, No. 1, pp. 33–51 (1975).

[5] Kanade, T., Cohn, J. and Tian, Y.: Comprehensive database for facial expression analysis, *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, IEEE, pp. 46–53 (2000).

[6] Miyahara, M., Aoki, M., Takiguchi, T. and Ariki, Y.: Tagging video contents with positive/negative interest based on user's facial expression, *Proceedings of the 14th international conference on Advances in multimedia modeling*, Springer-Verlag, pp. 210–219 (2008).

[7] Wang, Y., Ai, H., Wu, B. and Huang, C.: Real time facial expression recognition with adaboost, *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, Vol. 3, IEEE, pp. 926–929 (2004).