FACIAL ACTION UNIT RECOGNITION AND INFERENCE FOR FACIAL EXPRESSION ANALYSIS

Yu Li Xue, Xia Mao and Qing Chang
School of Electronic and Information Engineering, Beihang University, Xueyuan Road 37, Haidian District, Beijing, China

Keywords: Facial Action Unit Recognition, Gabor Wavelet, Optical Flow, Support Vector Machine, Dynamic Bayesian Network.

Abstract: Human facial expression is extremely abundant, and can be described by numerous facial action units. Recognizing facial action units helps catching the inner emotion or intention of human. In this paper, we propose a novel method for facial action unit recognition and inference. We used Gabor wavelet and optical flow for feature extraction, and used support vector machine and dynamic bayesian network for classification and inference respectively. We combined the advantages of both global and local feature extraction, recognized the most discriminant AUs with multiple classifiers to achieve high recognition rate, and then inference the related AUs. Experiments were conducted on the Cohn-Kanade AU-Coded database. The results demonstrated that compared to early researches for facial action units recognition, our method is capable of recognizing more action units and achieved good performance.

1 INTRODUCTION

Human facial expression plays an important role in human daily communications. It is important to analyze facial expression in the fields of psychology and affective computing.

Many researchers have proposed methods for facial action unit recognition. Pantic et al. founded a facial expression recognition expert system including various methods to recognize 16 action units and 6 basic facial expressions in both frontal and profile views (Pantic and Rothkrantz, 2000). Tian et al. extracted both permanent and temporal facial features and recognized neutral expression, 6 upper facial action units and 10 lower facial action units (Tian et al., 2001). Kapoor et al. used the infrared camera to detect pupil, and then extracted parameters through principle component analysis, finally used support vector machine to recognize upper facial action units and combined action units (Kapoor et al., 2003). Tong et al. used dynamic bayesian network to present rigid and non-rigid facial movement and their temporal-spatial relationship (Tong et al., 2007; Tong et al., 2010). They obtained the facial action recognition result through facial movement measurement and probability inference, and achieved good recognition result to spontaneous facial expressions.

However, most of above researches have been implemented in controlled conditions and limited AUs have been recognized. Recognizing subtle facial action units in real life is still a challenge. This paper aims to propose a method to recognize and infer more action units of facial expression.

2 FACIAL REGION LOCATION

We used the method of haar-like wavelet feature extraction and AdaBoost classification (Viola and Jones, 2001) to detect face from image sequences. Then eyes were detected in the face using the same method. If the two eyes are not in a horizontal level, the face will be aligned using affine transformation.

Based on the eyes location, we obtain the local regions of face, such as nose region, above eyes region, below eyes region, and below nose region. The illustration of the facial regions is shown in figure 1.

3 GABOR WAVELET AND OPTICAL FLOW FEATURE EXTRACTION

To extract the Gabor features of image sequences for action unit recognition, firstly, the difference image

Li Xue Y., Mao X. and Chang Q..
FACIAL ACTION UNIT RECOGNITION AND INFERENCE FOR FACIAL EXPRESSION ANALYSIS.
DOI: 10.5220/0003834006940697
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 694-697
ISBN: 978-989-8565-03-7

(a) (b) (c) (d) (e)

Figure 1: Illustration of the facial regions (a) eyes detected (b) nose inferred (c) above eyes region (d) below eyes region (e) below nose region (copyright @Jeffrey Cohn).

is obtained by subtracting the neutral image from the apex image. Then, Gabor wavelet feature is obtained by convolving the difference image with a set of multiscale and multiorientation Gabor filters.

The whole normalized face region is convolved by a set of Gabor filters at two spatial frequences and four orientations. The Gabor wavelet coefficient is shown in equation 1.

$$J = \|J\| e^{j\phi} \tag{1}$$

where $\|J\|$ is the magnitude and ϕ is the phase. $\phi = \pi/4$, $\pi/2$, $3\pi/4$ and π.

Optical flow has also been used to track the motion information of facial features in image sequence. Optical flow method assumes that the gray values in any image feature region do not change two consecutive frames, but only shift from one position to another. The calculation of optical flow is shown in equation 3.

$$I_x V_x + I_y V_y = -I_t \tag{2}$$

where V_x,V_y are the x and y components of the velocity or optical flow of $I(x,y,t)$ and I_x, I_y and I_t are the derivatives of the image at (x,y,t) in the corresponding directions.

4 SUPPORT VECTOR MACHINE FOR FACIAL ACTION UNIT RECOGNITION

We choose support vector machine (SVM) as the classifier for facial action unit recognition. Given l observations, each of which consists of a vector $\mathrm{x}_i \in R^n$, $i = 1, \cdots, l$ and related label y_i. The task of SVM is to study the mapping $\mathrm{x}_i \mapsto y_i$, the machine is defined by a list of possible mappings $\mathrm{x}_i \mapsto f(\mathrm{x}, \phi)$, where function $f(\mathrm{x}, \phi)$ can be obtained by parameter ϕ. Given ϕ, we can obtain a trained machine, the expectation of test error or expected risk of which is $R(\phi) = \int \frac{1}{2} |y - f(\mathrm{x}, \phi)| \, dP(\mathrm{x}, y)$, where $P(\mathrm{x}, y)$ is the unknown probability distribution. The experimental risk is the average error rate on the training set $R_{emp}(\phi) = \frac{1}{2l} \sum_{i=1}^{l} |y_i - f(\mathrm{x}_i, \phi)|$. Then the expected risk satisfied the equation 3.

$$R(\phi) \leq R_{emp}(\phi) + \sqrt{\frac{h(\log(2l/h) + 1) - \log(\eta/4)}{l}} \tag{3}$$

where h is the non-negative integer called VC (Vapnik Chervonenkis) dimension, which is the quantified measurement of the ability of learning machine. $0 \leq \eta \leq 1$, when no error $\eta = 0$, and when the worst situation $\eta = 1$. The right side of equation 3 is the risk boundary , and the lowest upper boundary can be obtained through choosing learning machine $f(\mathrm{x}, \phi)$.

5 DYNAMIC BAYESIAN NETWORKS FOR FACIAL ACTION UNIT INFERENCE

Bayesian Networks (BNs) are graph models for reasoning under uncertainty, where the nodes represent discrete or continuous variables, and the arcs represent the direct connections between them. The dependency is characterized by a conditional probability table (CPT) for each node. Dynamic bayesian networks (DBNs) is an extension of BNs to handle temporal models (Korb and Nicholson, 2004).

Firstly, we use a BN to model and learn relationships among AUs. Then, a DBN is made up of interconnected time slices of static BNs to model the dynamics in AU development and represent probabilistic relationships among AUs. Let θ_{ijk} indicate a probability parameter for a DBN with structure B_S, as seen in equation 4.

$$\theta_{ijk} = p\left(x_i^k | pa^j(X_i), B_S\right) \tag{4}$$

where i ranges over all the variables (nodes) in the DBN, j ranges over all the possible parent instantiations for variable X_i, k ranges over all the instantiations for X_i, x_i^k represents the kth state of variable X_i, and $pa^j(X_i)$ is the jth configuration of the parent nodes of X_i.

We learn the parameters of the DBN in order to infer each AU. The learning process maximizes the posterior distribution $p(\theta | D, B_S)$, given a database D and the structure B_S. Detailed description of DBN modelling for facial action unit inference can be seen in (Tong et al., 2007).

6 EXPERIMENTAL RESULTS

Experiments were conducted on the Cohn Kanade AU-Coded facial expression database (Kanade et al., 2000), which provides image sequences of facial expressions of 97 subjects. The facial expression images were coded into upper and lower AUs separately.

6.1 Facial Action Unit Recognition

The faces in neutral and apex facial expression images were detected and resized to 48*48. For the Gabor feature extraction, we obtained 8 different Gabor features and down sampled to a vector of 4608 dimensions. For the optical flow feature extraction, we obtained the optical flow in x level and y level of each pixel, and combined to a vector of 4608 dimensions. Support vector machine classification was conducted using 5-fold cross validation.

The upper and lower action unit recognition results using Gabor feature extraction are shown in table 1 and table 2 respectively. We can see that the classifiers reached high recognition rates, with little decrease when the number of classes increases. Note that although AU9 (Nose Wrinkler) is introduced in the lower face action units in FACS, we code it with the upper face action units as the main feature of it is around the root of nose located in the upper face. Furthermore, we recognized some AU not recognized in early researches, such as AU43.

Table 1: Upper action unit recognition result using Gabor wavelet feature extraction.

N.	action unit category	RR
2	1+2+5;6	94.41%
3	1+2+5;6;4+7	92.76%
5	1+2+5;6;4+7;4+6+7+9;4	90.56%
7	1+2+5;6;4+7;4+6+7+9;4; 4+6+7+9+43;4+7+9	89.79%

note:N.= number of classes; RR=recognition rate.

Table 2: Lower action unit recognition result using Gabor wavelet feature extraction.

N.	action unit category	RR
2	11+12+25;25+27	96.62%
4	11+12+25;15;25;25+27	94.49%
6	11+12+25;15;25;25+27;23;20+25	92.23%
8	11+12+25;15;25;25+27; 23;20+25;17;11+20+25	91.38%
10	11+12+25;15;25;25+27;23; 20+25;17;11+20+25;11+12;15+17	89.94%
12	11+12+25;15;25;25+27;23; 20+25;17;11+20+25;11+12; 15+17;10+11;11+15	89.13%

note:N.= number of classes; RR=recognition rate.

We also use the local feature for action units recognition. We compared the global and local feature extraction using both Gabor wavelet and optical flow feature extraction. The above eyes upper action unit recognition results using Gabor wavelet and optical flow feature extraction are shown in table 3 and table 4 respectively. We can see that, for AU1+2+5 and AU4+7 classification, local optical flow feature extraction in the above eyes region can achieve best result; for AU1+2+5, AU4+7 and AU4 classification, global Gabor wavelet feature extraction can achieve best result.

Table 3: Global and local upper action unit recognition result using Gabor wavelet feature extraction.

N.	action unit category	RR (global)	RR (local)
2	1+2+5;4+7	91.18%	86.76%
3	1+2+5;4+7;4	91.95%	80.54%

note:N.= number of classes; RR=recognition rate.

Table 4: Global and local upper action unit recognition result using optical flow feature extraction.

N.	action unit category	RR (global)	RR (local)
2	1+2+5;4+7	88.24%	97.79%
3	1+2+5;4+7;4	81.88%	89.26%

note:N.= number of classes; RR=recognition rate.

However, the lower action unit recognition results using optical flow is not better than using Gabor wavelet in our experiments.

We also trained and recognized the upper and lower action units on hemifaces using optical flow. The results are shown in table 5 and table 6 respectively. We can see that the recognition rate of upper action units on hemifaces is higher than on full face. So we can recognize upper action units on hemifaces to improve the recognition rate.

For the lower action unit recognition, the recognition rate on hemifaces is higher than on full face when the number of classes is small; when then umber of classes is large, the recognition rate on hemifaces is lower than on full face.

Table 5: Upper action unit recognition results on hemifaces using optical flow.

N.	RR	RR (LH)	RR (RH)
2	88.24%	94.12%	94.85%
3	85.03%	90.42%	91.62%
4	77.78%	83.33%	85.56%
5	74.21%	79.47%	80.53%
6	70.85%	74.37%	77.39%

note:N.= number of classes; RR=recognition rate; LH=left hemiface; RH=right hemiface.

Table 6: Lower action unit recognition results on hemifaces using optical flow.

N.	RR	RR (LH)	RR (RH)
4	86.81%	89.83%	88.14%
6	78.72%	82.33%	79.51%
8	70.06%	72.0%	68.62%
10	67.44%	67.24%	64.66%
12	66.21%	65.76%	63.32%

note:N.= number of classes; RR=recognition rate; LH=left hemiface; RH=right hemiface.

Consequently, for the global feature extraction, Gabor wavelet feature extraction can reach better recognition rate than optical flow feature extraction; for the local feature extraction, optical flow can reach better recognition rate in the above eyes region. We combined both of these feature extraction methods to achieve best recognition result.

To get accurate recognition result, we first recognize the action units using the classifier of most classes, if the recognized action units combination is included in other classifiers, these classifiers will also be used to verify the recognition result.

6.2 Facial Action Unit Inference

Using the AU codes of 452 samples of facial expressions as the training data, we learned the BN of 22 action units, as seen in figure 2. Compared to the BN of 14 action units learned by (Tong et al., 2007), more nodes and links are learned in our research. This means that there are more complex relationships among the AUs. Based on the recognized AUs as evidence, the DBN can infer related AUs according to the corresponding predicted probabilities.

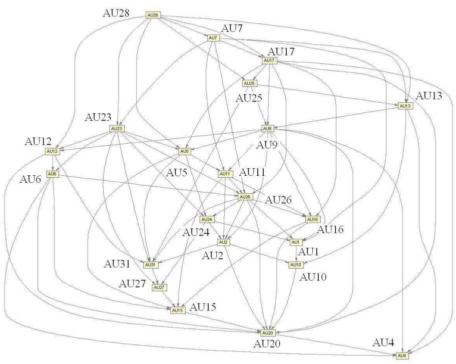

Figure 2: The learned BN of 22 action units.

Specially, the AUs we don't recognize directly because of lack of samples such as AU13(Sharp lip puller), AU16(Lower lip depress), AU24(Lip presser), AU26(Jaw drop) and AU31(Jaw clencher) can be inferred with their probabilities. For example, according to the CPTs of the learned DBN, $P(AU31=1|AU26=1)=0.7308$, means that when Jaw drop occurs, Jaw clencher will also occur with the probability of 0.7308.

7 CONCLUSIONS

Aiming to recognize facial action units efficiently, we analyze the Gabor wavelet and optical flow feature extraction in global and local facial regions, and use support vector machine and dynamic bayesian network for classification and inference respectively. The proposed method is capable of recognizing and inferring most action units in FACS, and can reach good performance.

ACKNOWLEDGEMENTS

This work is supported by National Nature Science Foundation of China (No.60873269, No.61103097), International Cooperation between China and Japan (No.2010DFA11990) and the Fundamental Research Funds for the Central Universities.

REFERENCES

Kanade, T., Cohn, J., and Tian, Y. L. (2000). Comprehensive database for facial expression analysis. In *Proc. of 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 46–53, Washington, DC, USA. IEEE Computer Society.

Kapoor, A., Qi, Y., and Picard, R. W. (2003). Fully automatic upper facial action recognition. In *Proc. of IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, pages 195–202.

Korb, K. B. and Nicholson, A. E. (2004). *Bayesian Artificial Intelligence*. A CRC Press Company.

Pantic, M. and Rothkrantz, L. J. M. (2000). Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 18(11):881–905.

Tian, Y. L., Kanade, T., and Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Trans. on PAMI*, 23(2):97–115.

Tong, Y., Chen, J. X., and Ji, Q. (2010). A unified probabilistic framework for spontaneous facial action modeling and understanding. *IEEE Trans. on PAMI*, 32(2):258–273.

Tong, Y., Liao, W. H., and Ji, Q. (2007). Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Trans. on PAMI*, 29(10):1–17.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc. of IEEE Conf. on CVPR*, pages 511–518, Washington, DC, USA. IEEE Computer Society.