H-008

Improving a Bayesian Network Based Recognition of Spontaneous Facial Expressions of a Person who Watches Web News

- Utilizing Image Features for Blinks -

Chao Xu† Jun Ohya†

1． Abstract

Recently reading news through web news media is becoming popular. Most of web news is delivered together with a comment system, which asks the user to rate items such as “the news let you think”, and “boring”.

However, many users do not like manual operations for rating; therefore, not many users actually input the rates. One solution for this issue is to utilize results of recognizing facial expressions from the video sequence acquired by the camera that observes the user. So far, there are very many works on recognizing six fundamental expressions such as sad, surprise and happy, for example, by HMM (Hidden Markov Models) [1]. However, not many works dealt with recognizing expressions that could appear when the user watches web news; main difficulty in recognizing these expressions could be caused by the fact that these expressions are spontaneously generated. It is difficult to recognize the spontaneous expressions, which come with tiny movements hard to be detected. This paper proposes a method that aims at recognizing spontaneous expressions.

2 ． Basic Idea

A Bayesian network [2] is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). A Bayesian network can solve problems such as spontaneous behavior prediction of infants by gathering irregular behavior patterns. That is, a Bayesian network might work well for recognizing spontaneous expressions. ASM (Active Shape Model) [3] is a training set based method for locating featupoints in the face. The located feature points could be useful features for a facial expression recognition. This paper uses ASM as feature detector that provides a Bayesian network with parameter values based on the detected feature points.

3. Combing of Bayesian Network and ASM

Table 1 lists each spontaneous expression and its example of possible news that could let the viewer generate that spontaneous expression.

3.1 Feature Distances and Normalization

The process described in this chapter is shown in Fig.1, where this chapter deals with the training phase.

During the ASM process, as shown in Fig. 1, six distances (Distance A~F) between the 10 feature points are calculated as

† Waseda University, Graduate School of Global Information and Telecommunication Studies

fig.2.

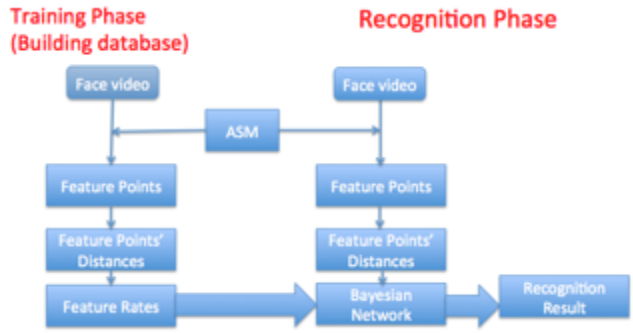

Fig 1 Basic Idea of This Reasearch

Table 1 Feature Web news of Spontaneous Expression

Feature web news	Spontaneous expression
Makes people think	People who acted as patriots to commit acts of vandalism need to be brought to justice
Being helpful	Evaluation of iPhone: AU vs. Softbank
Being interested	iPhone for free right now!
Feeling misery	Folk near Chernobyl still cloudy about health
Boring	Twitter of Big Ben: Bang, Bang, Bang…

In case the subject changes the distance between him/her and the computer screen, the distance between Point X and Point Y is normalized to 100 in each frame. Thus, for each distance of the six (fig.3.2), we can get the Absolute Distance and Relative Distance of each frame of feature spontaneous expression video as:

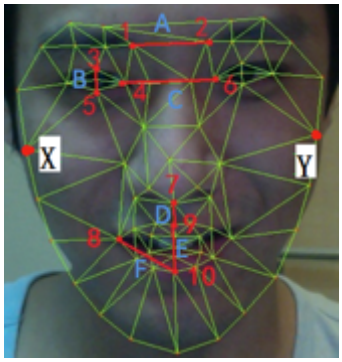

Fig. 2 Feature Points and Feature Distances

$$\text{Relative Distance}(A,B,C,D,E,F) = \frac{\text{Absolute Distance}(A,B,C,D,E,F)}{\text{Absolute Distance}(\text{Line}_{XY})} * 100 \quad (1)$$

Then, to simplify the data, from change in each Relative Distance in each frame of the video sequence, the Pre-Feature Change Value of each distance in one video can be obtained as follows.

$$\text{Pre} - \text{Feature Change Value} = \text{Max}[\text{Relative Distance}] - \text{Normalized Neutral Distance} \quad (2)$$

where Normalized Neutral Distance denote the normalized distance for neutral expression, Max[ ] indicates the operation for obtaining the maximum value in all the frames.

Furthermore, we define the Feature Change Value as a value that could have one of the following five types: -2<, -1, 0, 1, and

>2. The Feature Change Value is obtained from Pre-Feature Change Value for each type as follows.

$$\text{Feature Change Value} = \text{Count}(\text{Pre} - \text{Feature Change Value})$$

where Count counts the number of Pre-Feature Change Value for that type.

Each video can be represented by a Feature Change Value efficiently in a simple way.

To build the Bayesian Network, a global conditional variable is needed in each links of nodes. Feature Rate is defined as the Feature Change Value per subject as follows.

$$\text{Feature Rate} = \frac{\text{Feature Change Value}}{\text{Number of subjects}} \quad (3)$$

The Feature Rates are used for building the Bayesian Network.

3.2 Eye State and Mouth State

Except the distance changing values, there might be other features such as eye and mouth states. If we combine these states, we might be able to improve the recognition performance.

This thesis deals with the human eye state; more specifically, open or close of eyes can be used for news classification framework based on a global scan and verification strategy. The method employs a cascade structure to organize a series of eye state classifiers trained by Adaboost. The experiments by Zhaorong et al. over a large dataset show that proposed system achieves a good performance [4]. Figure 3 shows some samples.

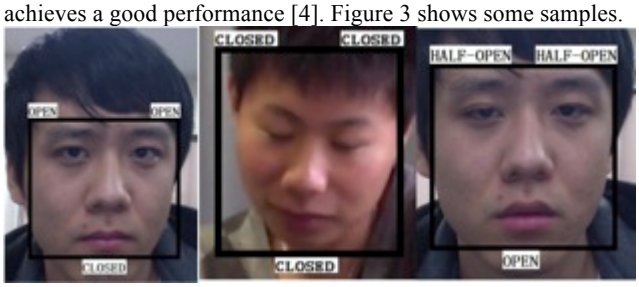

Fig. 3 Examples of Experiment

3.3 Bayesian Network Building

Figure 4 shows the proposed structure of the Bayesian Network. The root node is “Feature Distance”. From the Feature Distances, the six Feature Change Values are obtained; therefore, the “Feature Distance” node is the parent node of the six nodes for “Feature Change Value”. Since the recognition result is obtained from one of the “Feature Change Value” nodes, the “Recognition” node is the child node of “Feature Change Value” nodes.

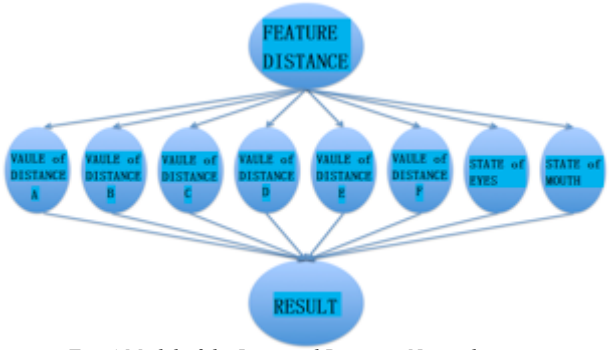

Fig. 4 Model of the Improved Bayesian Network

As described in Chapter 3, the final recognition result is obtained as the category (spontaneous expression) that maximizes the a-posteriori probability in condition of a-probability of the six “Feature Change Values”.

For example, if the viewer is feeling misery, the a-priori probabilities for the Feature Rate (<= -2) of distance A is 0.86, and for the Feature Rate (<= -1) of distance A is 0.14, thus the link from distance A to the “Result” node yields these two kinds of conditional variables, because the other three a-priori probabilities are zero in this particular example. The other links are set similarly. Two new nodes are added to the Bayesian Network we built. Figure 4 shows the specific structure of the new Bayesian Network.

4 . Experimental Results

Recognition results indicates that the ground truths for the videos are shown horizontally, and the category (expression) to be recognized are shown vertically. This experiment shows a 0.82 recognition result on average, which is a very promising result.

The results of recognizing the spontaneous facial expressions using the Bayesian Network with the eye state nodes are listed in Table 2. By adding the two state nodes of eyes and mouth, the result is improved to 0.86 on average, which is higher than the method with 6 distance nodes by 0.02.

Table 2 Improved Result by Eyes and Mouth State

		Feature videos of				
		Makes people think	Being helpful	Being interested	Feel Misery	Boring
Recognition result as	Makes people think	0.89	0.06	0.10	0.03	0.01
	Being helpful	0.01	0.84	0.05	0.12	0
	Being interested	0.02	0.06	0.79	0.06	0
	Feel Misery	0.06	0.03	0.05	0.78	0
	Boring	0.02	0.01	0.01	0.01	0.99

5 . Conclusion

This paper has proposed a method to solve the problem that web news comment system is barely voted because of its manual operation. As a result of conducting experiments using 17 subjects’ spontaneous expression videos, the recognition accuracy 84% are achieved using the non-eye-state model. Using “with eye-state” mode, 86% are achieved. These are considered to be promising results.

References

[1] Lawrence R. Rabiner, “*A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*”, Proceedings of the IEEE, 77 (2), p. 257–286, February 1989

[2] Motomura Yoichi, Nishida Yoshifumi, “*Graphical Modeling of Prior Probability in Bayesian Estimation for Behavior Understanding in Real Life*”, Information Processing Society of Japan, 2007

[3] T. F. Cootes, and C. J. Taylor, "*Active Shape Models - Smart Snakes*", British Machine Vision Conference, Springer-Verlag, pp.266-275, 1992.

[4] Li Zhaorong, Ai Haizhou, “*Real-time Robust Automatic Eye State Classification*”,TP391