Facial Emotion Detection Considering Partial Occlusion of Face Using Bayesian Network

Yoshihiro Miyakoshi, and Shohei Kato
Dept. of Computer Science and Engineering
Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya 466–8555, Japan
Email: {miyakosi, shohey}@juno.ics.nitech.ac.jp

***Abstract*—Recently, robots that communicate with human have attracted much attention in the research field of robotics. In communication between human, almost all human recognize the subtleties of emotion in each other's facial expressions, voices, and motions. Robots can communicate more smoothly with human as they detect human emotions and respond with appropriate behaviors. Usually, almost all human express their own emotions with their facial expressions. In this paper, we propose an emotion detection system with facial features using a Bayesian network. In actual communication, it is possible that some parts of the face will be occluded by adornments such as glasses or a hat. In previous studies on facial recognition, these studies have been had the process to fill in the gaps of occluded features after capturing facial features from each image. However, not all occluded features can always be filled in the gaps accurately. Therefore, it is difficult for robots to detect emotions accurately in real-time communication. For this reason, we propose an emotion detection system taking into consideration partial occlusion of the face using causal relations between facial features. Bayesian network classifiers infer from the dependencies among the target attribute and explanatory variables. This characteristic of Bayesian network makes our proposed system can detect emotions without filling in the gaps of occluded features. In the experiments, the proposed system succeeded in detecting emotions with high recognition rates even though some facial features were occluded.**

***Keywords*—Facial Emotion Detection; Bayesian Network; Feature Selection; K2 Algorithm; Partial Occlusion of Face**

I. INTRODUCTION

Recently, diverse type of robots designed not only for industrial activities but also for communicating with human have been researched and developed. Therefore, opportunities for human who rarely use computers to have contact with robots are increasing. In addition, communication robots are aimed at accomplishing smooth communication with human. For these reasons, communication robots need more intuitive interaction systems.

It is reasonable to suppose that robots that give human a sense of familiarity can communicate more smoothly with human. According to reports in the field of human communication, human feel familiarity with one's counterpart as they synchronize with their nonverbal information, e.g., facial expressions and voices [1], [2]. Robots can communicate more smoothly with human as they detect human emotions and respond with appropriate behaviors. Usually, almost all human express their own emotions with their facial expressions. In this paper, we propose an emotion detection system with facial features using a Bayesian network.

In actual communication, it is possible that some parts of the face will be occluded by adornments such as glasses or a hat. In previous studies on facial recognition, these studies have been had the process to fill in the gaps of occluded features after capturing facial features from each image [3], [4]. However, not all occluded features can always be filled in the gaps accurately. Therefore, it is difficult for robots to detect emotions accurately in real-time communication. For this reason, we propose an emotion detection system considering partial occlusion of the face using causal relations between facial features. Bayesian network classifiers infer from the dependencies among the target attribute and explanatory variables. This characteristic of Bayesian networks makes our proposed system can detect emotions without filling in the gaps of occluded features.

II. EMOTIONS IN COMMUNICATION

According to reports in the field of evolutionary psychology, emotions have been selected in the evolution [5], [6]. These reports indicate that emotions are important for human life. Especially, in communication between human, almost all human recognize the subtleties of emotion each other. Moreover, in previous studies on human communication, the meaning of emotions have been debated for make smooth communication [7]. So, we consider that emotions are also important for make smooth communication between robots and human. For these reasons, communication robots need the emotion detection system to make emotional communication with human.

A. Relationships between emotions and facial expressions

Human express emotions in diverse ways, e.g., facial expressions, voices and body languages. However, voices contain verbal information, and the voice pitch is associated with a gender gap. Likewise, the meaning that is expressed through body language differs among cultures. Nevertheless, Paul Ekman and his colleagues found evidence to support universality in facial expressions [8]. Moreover, according to reports in the field of human emotions, when human mimic an emotional face, they then generate the same emotion [9]. These evidences indicate strong relationships between emotions and facial expressions. For these reasons, we focus on facial expressions to detect human emotions. In this paper, we use six different human emotions: happiness, anger, sadness, surprise, disgust, and fear as target attributes of the detection system. These emotions are represented as "universal facial expressions" by Ekman, et al. They studied facial expressions in different cultures, including preliterate cultures, and found much commonality in the expression and recognition of emotions on the face. For this reason, we believe that a system that detects these emotions is effective for people of different races and cultures.

III. Facial features for emotion detection

In this paper, we use magnitudes of some predefined motion of various facial features as explanatory variables of the detection system. Magnitudes of these motions are calculated by movements of facial features from a neutral face for each facial expression. Our proposed system is designed to detect human emotions for communication robots. We believe that it is the appropriate assumption that communication robots can get a neutral face of the interlocutor. The facial features are shown in Fig. 1, and are described in TABLE I. Each facial feature corresponds to a simple deformation on the face referring to Motion-Units [10].

IV. Bayesian network classifier for facial expression detection

Bayesian networks are powerful tools for knowledge representation and inference under conditions of uncertainty.

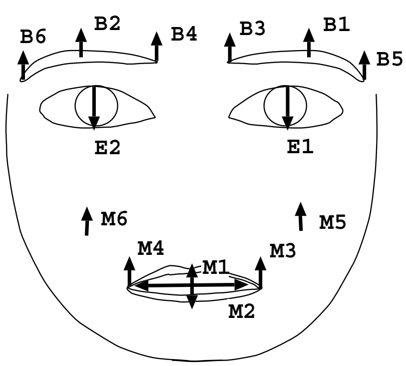


Fig. 1. Facial features.

TABLE I
Description of Facial Features

Index	Description
E1	Blinking of right eye
E2	Blinking of left eye
B1	Vertical movement of right brow
B2	Vertical movement of left brow
B3	Vertical movement of right brow left corner
B4	Vertical movement of left brow right corner
B5	Vertical movement of right brow right corner
B6	Vertical movement of left brow left corner
M1	Horizontal stretch of mouth height
M2	Vertical stretch of mouth width
M3	Horizontal movement of right mouth corner
M4	Horizontal movement of left mouth corner
M5	Lifting of right cheek
M6	Lifting of left cheek

According to reports on classification, Bayesian networks are also effective as classifiers [11].

A Bayesian network classifier represents the dependencies among the target attribute and explanatory variables in a directed acyclic graph. This graph is the structure of the Bayesian network. Many Bayesian network classifiers have been proposed in previous works, for example [11], [12], [13]. However, these Bayesian network classifiers almost no consider causal relations among explanatory variables. Therefore, we propose a Bayesian network classifier considering causal relations among explanatory variables. We construct the Bayesian network composed of the emotions and the facial features. The proposed system learns the structure of the Bayesian network in two phases: an internal phase and an external phase. Because of these phases, proposed system constructs a more robust Bayesian network classifier. Fig. 2 shows the process flow of learning the structure in the proposed system.

A. Gaussian distribution

Bayesian network classifiers calculate the maximum likelihood to classify the target attribute based on Bayes' theorem. The optimal classification rule under the maximum likelihood is expressed by the following equation.

$$\hat{c} = \arg\max_{c \in \{1,\ldots,|C|\}} P(X_1, \ldots, X_n | c; \Theta), \tag{1}$$

where, c, $X_1, \ldots, X_n$, and Θ each shows one of class label of the target attribute, explanatory variables, and the parameter set.

There is the problem that how to model $P(X_1, \ldots, X_n | c; \Theta)$ that is the probability of explanatory variables given the class label c. In general Bayesian networks, the explanatory variables are treated as discrete, in which case the distributions are probability mass functions. The

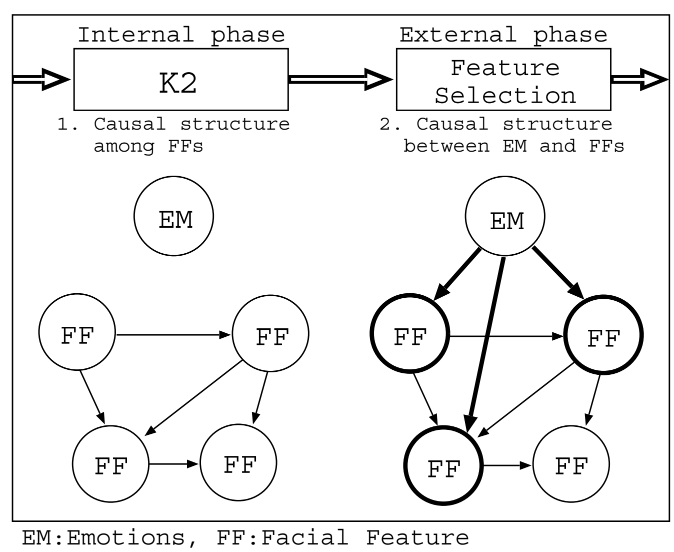


Fig. 2. Process flow of learning the structure of the Bayesian network.

explanatory variables can also be continuous as use the Gaussian distribution. The maximum likelihood can be used to obtain the estimate of the parameters (mean and variance). In this paper, we treat the facial features as continuous by using the Gaussian distribution.

B. Internal phase

In the internal phase, the system learns the structure among facial features using the K2 algorithm [14]. K2 is a greedy algorithm that learns the structure of the Bayesian network from given data. Therefore, the structure learned by using the K2 approximates real causal relations among facial features.

1) K2 algorithm: K2 algorithm is a Bayesian method for inducting a structure from given data. K2 attempts to select the network that maximizes the posterior probability of the network given the database of cases. K2 is a scoring-based learning algorithm. We use the Bayesian Information Criterion to determine the scores. When M, $\hat{\theta_M}$, and d each shows an explanatory variable, explanatory variables that have dependencies with M, and the number of $\hat{\theta_M}$; the $BIC(\hat{\theta_M}, d)$, which is the score of M, is expressed by the following equation.

$$BIC(\hat{\theta_M}, d) = \sum_{i=1}^{N} \log P(D_i) - \frac{d \log N}{2}, \tag{2}$$

where N, D_i, and $P(D_i)$ each shows the size of the database, a data, and a occurrence probability of D_i. K2 needs to be given a total order of the explanatory variables for learning the structure. In this paper, we arrange explanatory variables in descending order by the degree of separation of emotions. The ordering on the explanatory variables is shown by the following equation. The index in the following equation shows each facial feature (refer to TABLE I).

$$\begin{aligned} &B4 \prec M4 \prec M3 \prec E1 \prec B3 \prec B6 \prec M1 \\ &\prec B5 \prec M6 \prec M5 \prec M2 \prec B2 \prec B1 \prec E2. \end{aligned} \tag{3}$$

C. External phase

In the external phase, the system learns the structure between emotions and facial features with causal relations using feature selection. In previous studies on Bayesian network classifiers, the importance of selection to find a good subset of explanatory variables has been reported [15], [16], [17]. However, these studies selected features without considering causal relations among explanatory variables. We believe that proposed system can select the more appropriate subset of facial features for emotion detection because of using causal relations among facial features. In this paper, we use stepwise selection to find a good subset of facial features for emotion detection, and use the average recognition rate for all emotions as evaluation values.

V. Related study

Cohen, et al. studied an emotion detection system with facial features using a Bayesian network [18]. Their system learns the structure of the Bayesian network using Tree-Augmented-Naive Bayes (TAN) [12]. TAN forms strong relationships between emotions and facial features. On the other hand, relationships among facial features become a tree structure. So, a facial feature has at most one parent. When some parts of the face are occluded, it is difficult for TAN to correctly detect emotions because it does not have sufficiently relationship to make up for the occluded features. In contrast, the structure of our proposed system forms strong relationships among facial features. We believe that the proposed system can handle the occlusion of facial features robustly by using causal relations between facial features. We discuss here the effectiveness of the proposed system by comparing the recognition rates between the proposed Bayesian network classifier (Proposed method) and the TAN classifier (Conventional method).

A. Tree-Augmented-Naive Bayes

The algorithm for learning TAN first learns a tree structure over explanatory variables, using mutual information tests conditioned on the target attribute. It then adds a link from the target attribute to each explanatory variable. When X_i and X_j, which $i \neq j$, and C each shows an explanatory variable, and the target attribute, then $I_P(X_i; X_j|C)$ that the average mutual information conditioned on a target attribute is shown by the following equation.

$$I_P(X_i; X_j|C) = \sum_{X_i, X_j, C} P(x_i, x_j, c) \log \frac{P(x_i, x_j|c)}{P(x_i|c)P(x_j|c)}. \tag{4}$$

When using the Gaussian distribution for explanatory variables, Equation (4) is replaced by the following equation.

$$I_P(X_i; X_j|C) = -\frac{1}{2} \sum_{c=1}^{|C|} P(C = c) \log(\frac{1}{1 - \rho^2_{(ij)|c}}), \tag{5}$$

where $\rho_{(ij)|c}$ is the correlation coefficient between X_i and X_j given the class label c.

VI. Experiments

To verify the effectiveness of the proposed method, we use the Japanese Female Facial Expression (JAFFE) database collected by Lyons, et al., [19]. JAFFE is composed from facial images that are acted by ten Japanese female subjects, and posed giving three or four examples of each emotion as well as a neutral face, for a total of two hundred thirteen images (one hundred eighty three images are emotional faces and thirty images are neutral faces). We calculate magnitudes of the motion of facial features from the neutral face for each facial expression. We use one hundred eighty three data for the experiments, with at least three data for each emotion for each subject. In order to normalize each image, we set the distance between the inner corners of one's eyes to 30 pixels.

In the experiment, we conduct ten-fold cross validation: nine subjects are used for training data, and another one subject is used for test data in each validation. In a validation, we learn the structures of the Bayesian network using proposed method, and conventional method, respectively. With the learned structures, we then conduct following two experiments, Ex.1 for detecting facial emotions with all of the facial features, and Ex.2 for detecting facial emotions with facial features that are occluded in some way. Moreover, in Ex.2, we also detect the occluded facial features.

A. Learning the structure of the Bayesian network

Fig. 3 and 4 each shows an example of the structure of the Bayesian network using the proposed method and the conventional method respectively. In the figures, the bold arrowed lines show the causal relations between the emotions and facial features, and the thin arrowed lines show the causal relations among facial features. TABLE II lists the selected rates of facial features by stepwise selection over cross validation. Facial features getting high selected rates are E1, E2, B1, B2, B4, M1, M2, and M4. These facial features indicated a strong relationship with emotions in Ekman's report. Therefore, we consider that the proposed method is able to find a good subset of facial features for emotion detection. In terms of relations among facial features, we can see that the proposed method is able to form strong relationships among facial features compared with the conventional method. Moreover, as a result of using K2, facial features are linked by the strength of the relationship between them. We suggest that the proposed method is able to approximate real causal relations among facial features.

B. Ex.1: Experiment on emotion detection

We conduct an experiment on emotion detection with the learned structure of the Bayesian network using the proposed method , and the conventional method, respectively.

TABLE III lists the results of emotion detection using the proposed method. The columns show the induced emotions and the rows show the detected emotions. TABLE IV shows the comparison of recognition rates between the proposed and conventional methods. The total is the average recognition rate for all emotions.

In TABLE IV, the total recognition rate of the proposed method outperforms that of the conventional method. The recognition rate for disgust using the conventional method is below 50%. On the other hand, all of the recognition rates with the proposed method are above 55%. This result is due to the difference in the structure between emotions and facial features. The conventional method overfits the training data because it forms a relationship between emotions and all facial features. In comparison, the proposed method selects a good subset of facial features for emotion detection. The proposed method obtain high recognition rates for total recognition, and especially happiness and surprise are relatively high. In terms of emotions with low recognition rates, angry, sadness, and disgust are mistaken for each other. These mistakes occur caused by the similarity of the facial expressions among these emotions, e.g., brows and corners of one's mouth are lowered. In addition, according to Ekman's report, it is even difficult for human to distinguish between angry and disgust. For these reasons, we suggest that the proposed method achieves a human-like emotion detection system.

Fig. 5 shows occurrence probabilities of each emotion for each facial feature selected by the proposed method over all cross validations. It seems that detecting emotions with only one facial feature is difficult, however almost all emotions can be detected by inferring from all selected features. We can confirm that the proposed method is able to select facial features that have effect to detect emotions from Fig. 5.

C. Ex.2: Experiment on emotion detection considering facial occlusion

In actual communication, not all facial features can always be captured accurately. We conduct an experiment when some facial features are occluded in order to verify the effectiveness for the robustness about facial occlusion of the proposed method. Emotion detection systems using Bayesian networks can detect emotions without filling in the gaps of occluded features. In addition, these systems can also detect occluded features by using relationships among facial features. So, we first detect emotions considering facial occlusion, and then detect occluded features. Parts of occluded facial features are eyes (E1, E2), brows (B1, B2, B3, B4, B5, B6), and mouth (M1, M2, M3, M4, M5, M6). These parts can easily be occluded by adornments such as glasses, a hat, or a surgical mask. In this study, we assumed that two or more occlusions of facial parts would not occur at the same time. We conduct ten-fold cross validation the same way as in VI-B.

TABLES IV, V, and VI list the recognition rates of emotion detection for the partial occlusion. TABLE VIII(a), (b), and (c) give the average absolute errors for actual measurement values based on estimated values inferred from relationships among

TABLE II
SELECTED RATES OF FACIAL FEATURES

Facial Feature	E1	E2	B1	B2	B3	B4	B5
Rates (%)	**100**	**90**	**100**	**80**	70	**100**	20

Facial Feature	B6	M1	M2	M3	M4	M5	M6
Rates (%)	60	**80**	**80**	40	**90**	60	40

TABLE III
CONFUSION MATRIX BY PROPOSED METHOD

Emotion	Happiness	Anger	Sadness	Surprise	Disgust	Fear
Happiness	**25**	1	0	4	0	2
Anger	2	**19**	0	0	7	2
Sadness	0	3	**19**	0	2	6
Surprise	3	0	0	**27**	0	0
Disgust	0	6	5	1	**16**	1
Fear	1	0	4	3	1	**23**

TABLE IV
RECOGNITION RATES

	Happiness	Anger	Sadness	Surprise	Disgust	Fear	Total
Proposed (%)	78.1	63.3	**63.3**	**90.0**	**55.2**	**71.9**	**70.3**
Conventional (%)	**84.4**	**73.3**	53.3	86.7	48.3	59.4	67.6

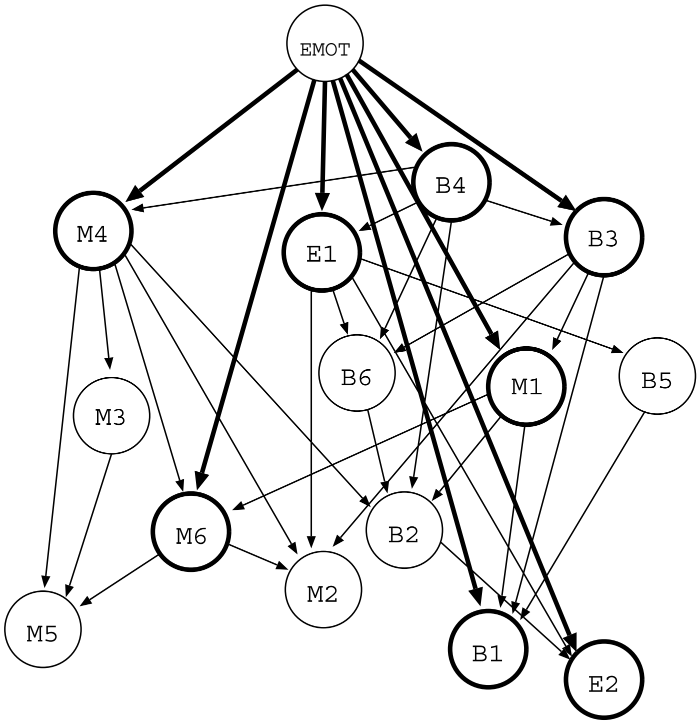


Fig. 3. Learned structure using conventional method.

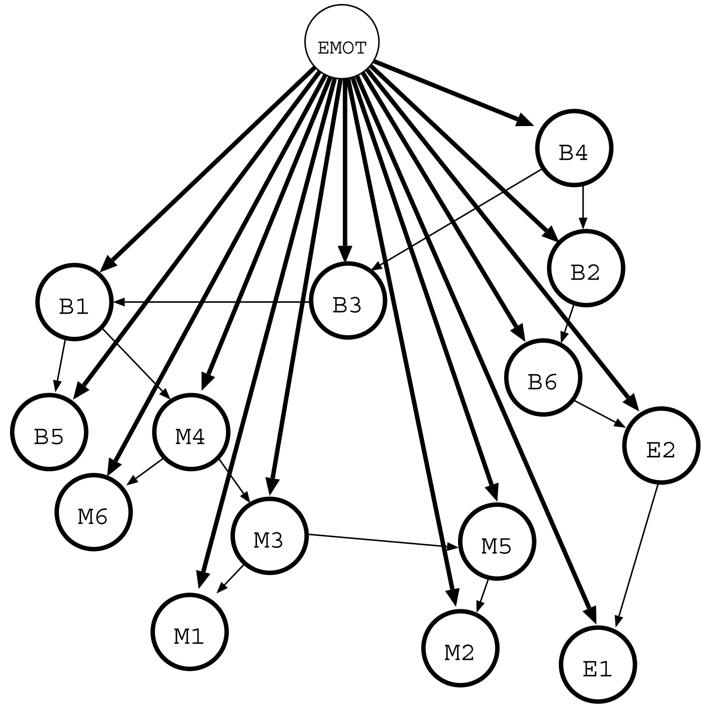
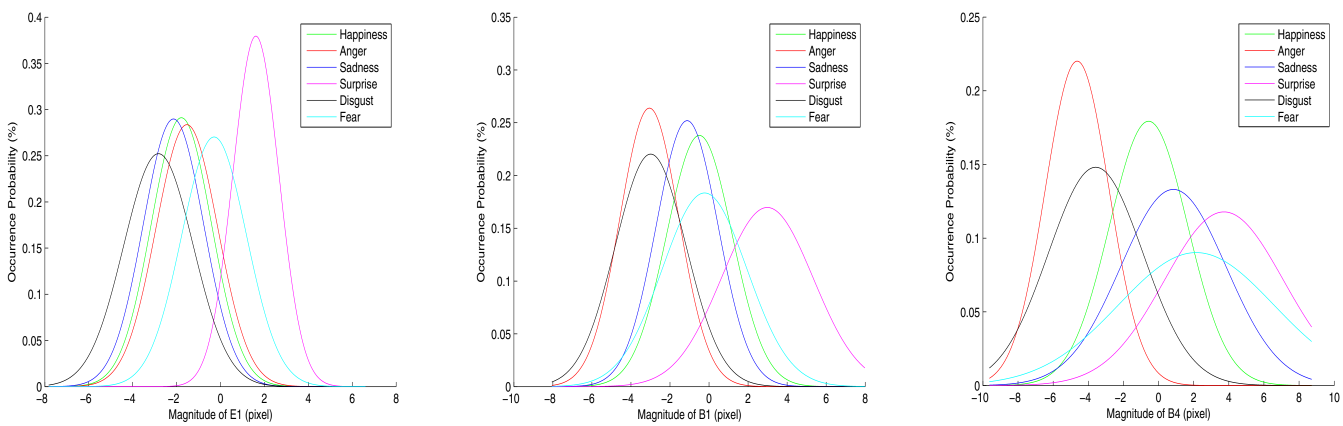


Fig. 4. Learned structure using proposed method.



(a) Facial feature: E1

(b) Facial feature: B1

(c) Facial feature: B4

Fig. 5. Occurrence probabilities of each emotion for each facial feature.

facial features. The index in the tables indicates each facial feature (refer to TABLE I).

The total recognition rate of the proposed method outperform that of the conventional method for each occlusion. Recognition rates for some emotions using the conventional method show a sharp downturn. On the other hand, almost all recognition rates using the proposed method remain above 50%. In terms of errors, the proposed method outperform the conventional method for all facial features. The error for M1 with the conventional method is abnormally large. In contrast, the facial feature is in the same range as other facial features with the proposed method. Because of the abnormally large error, we consider that the recognition rates of the conventional method for occluded mouth are not obtained from the correct inference. Moreover, we can see from Fig. 4 and TABLE VIII that the errors of facial features that have only one relation between facial features, e.g., B5, M2, and M6 tend to be large compared with the proposed method. The errors of the proposed method are sufficiently small considering the fact that the distance between the inner corners of one's eyes is 30 pixels. We can suggest that the proposed method is more effective in recognizing emotions despite facial occlusion compared with the conventional method because of strong relationships among facial features.

TABLE V
RECOGNITION RATES FOR OCCLUDED EYES

	Happiness	Anger	Sadness	Surprise	Disgust	Fear	Total
Proposed (%)	78.1	70.0	66.7	83.3	48.3	56.3	**67.1**
Conventional (%)	68.8	73.3	53.3	83.3	41.4	56.3	62.7

TABLE VI
RECOGNITION RATES FOR OCCLUDED BROWS

	Happiness	Anger	Sadness	Surprise	Disgust	Fear	Total
Proposed (%)	71.9	56.7	50.0	70.0	34.5	53.1	**56.0**
Conventional (%)	71.9	26.7	53.3	70.0	13.8	65.6	50.2

TABLE VII
RECOGNITION RATES FOR OCCLUDED MOUTH

	Happiness	Anger	Sadness	Surprise	Disgust	Fear	Total
Proposed (%)	56.3	46.7	40.0	70.0	31.0	53.1	**49.5**
Conventional (%)	46.9	66.7	23.3	80.0	24.1	43.8	47.5

TABLE VIII
AVERAGE OF ABSOLUTE ERRORS

(a) Occluded Eyes

	E1	E2
Proposed ($pixels$)	1.3	1.4
Conventional ($pixels$)	5.6	3.6

(b) Occluded Brows

	B1	B2	B3	B4	B5	B6
Proposed ($pixels$)	1.6	1.8	2.5	2.7	1.4	1.7
Conventional ($pixels$)	1.8	2.0	2.8	3.1	3.1	2.3

(c) Occluded Mouth

	M1	M2	M3	M4	M5	M6
Proposed ($pixels$)	3.9	3.6	3.5	3.5	4.2	4.3
Conventional ($pixels$)	**314.2**	4.3	3.9	4.0	4.7	5.0

VII. CONCLUSION

We proposed an emotion detection system considering partial occlusion of the face using a Bayesian network. The proposed Bayesian network forms strong relationships among explanatory variables. In experiments, we confirmed high recognition rates for total recognition, and especially for happiness and surprise. Moreover, in the experiment on emotion detection considering facial occlusion, we confirmed that almost all recognition rates with the proposed method remained above 50%, and the errors were small compared with the conventional method. These results indicate that the propose method was more effective in recognizing emotions than the conventional method even if facial features were partially occluded.

The future direction of this study will be follows.

- Increasing the recognition rate:
  In this paper, we learned the structure among facial features using the K2. In terms of relationships among facial features, they have been researched in the field of anatomy [20]. We will reform relationships learned by K2 based on researches on anatomy and form real causal relations. And also, we will confirm whether proposed structure can comply with a larger database.
- Communication between robots using our proposed system and human:
  Our proposed system is designed to detect human emotions for communication robots, and is aimed at accomplishing smooth communication with human. We will compose robots using our proposed system, and then confirm the usability of the robots for communication with human.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Education, Science, Sports and Culture, Grant–in–Aid for Scientific Research under grant #20700199.

REFERENCES

[1] J. L. Lakin and T. L. Chartrand, "Using nonconscious behavioral mimicry to create affiliation and rapport," *Psychological Science*, vol. 14, pp. 334–339, 2003.

[2] T. Hayashi, S. Kato, and H. Itoh, "A synchronous model of mental rhythm using paralanguage for communication robots," *Lecture Notes In Artificial Intelligence*, vol. 5925, pp. 376–388, 2009.

[3] A. M. Martinez and Y. Zhang, "Subset modelling of face localization error, occlusion, and expression," *Face Processing: Advance Modeling and Methods*, 2005.

[4] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Journal of Image and Visual Computing*, vol. 26, pp. 1052–1067, 2008.

[5] M. Ridley, *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*. Penguin Books, 1998.

[6] V. Johnston, *Why We Feel: The Science of Human Emotions*. Basic Books, 2000.

[7] P. A. Andersen and L. K. Guerrero, *Handbook of Communication and Emotion: Research, Theory, Applications, and Contexts*. Elsevier Science and Technology Books, 1997.

[8] P. Ekman, "Strong evidence for universals in facial expressions: a reply to russell's mistaken critique,," *Psychol Bull*, vol. 115, pp. 268–287, 1994.

[9] L. K. Bush, C. L. Barr, G. J. McHugo, and J. T. Lanzetta, "The effects of facial control and facial mimicry on subjective reactions to comedy routines," *Motivation and Emotion*, vol. 13, pp. 31–52, 1989.

[10] H. Tao and T. S. Huang, "Bõzier volume deformation model for facial animation and video tracking," *Lecture Notes In Computer Science*, vol. 1537, pp. 242–253, 1998.

[11] P. Langley, W. Iba, and K. Thompson, "An analysis of bayesian classifiers," *Proceedings of the tenth national conference on artificial intelligence*, pp. 223–228, 1992.

[12] N. Friedman, D. Gieger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, pp. 131–163, 1997.

[13] J. Cheng and R. Greiner, "Comparing bayesian network classifiers," *Proceedings of the Uncertainty in Artificial Intelligence*, pp. 101–108, 1999.

[14] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, pp. 309–347, 1992.

[15] P. Langley and S. Sage, "Induction of selective bayesian classifiers," *Proceedings of the Uncertainty in Artificial Intelligence*, pp. 399–406, 1994.

[16] M. Singh and G. M. Provan, "Efficient learning of selective bayesian network classifiers," *Proceedings of the International Conference on Machine Learning*, 1996.

[17] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273–324, 1997.

[18] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huanga, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, pp. 160–187, 2003.

[19] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," *Proc. the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200–205, 1998.

[20] R. C. Jacobs, *Royal's Guide to Muscular Anatomy*. Royal Jacobs Publishing Company, 2010.