MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
8-12

An Artificial Emotion Imitator

Do Hyoung Kim, Kwang Ho An and Myung Jin Chung
Dept. of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
E-mail: {bborog, akh}@cheonji.kaist.ac.kr, mjchung@ee.kaist.ac.kr

Abstract—**In the last decade, face analysis, e.g. face recognition, face detection, face tracking, facial expression recognition, is a very lively and expanding research field. As computer animated agents and robots bring a social dimension to human computer interaction, interest in this research field is increasing rapidly. In this paper, we introduce an artificial emotion mimic system which can recognize human facial expressions and also imitate the recognized facial expression. And also we propose a classifier that is based on weak classifiers obtained by using modified rectangular features to recognize human facial expression in real-time. Next, we introduce our robot that is manipulated by a distributed control algorithm. Finally, experimental results of facial expression recognition and emotion expression are shown for the validity of our artificial emotion imitator.**

Keywords- facial expression recognition; AdaBoost; emotion expression; face detection; active vision

I. INTRODUCTION

The most effective way to recognize human's emotion is to notice human's facial expressions. The facial expressions have rich information about human's emotion or mood. For that reason, if computer animated agents or robots can automatically recognize the facial expressions, those artificial systems are easily able to understand or estimate human's emotion or mood. This recognition technique can be also used as a component of human-robot interaction (HRI).

Within the past decade, many researchers have been trying to automatically recognize facial expressions of a human being. Various pattern recognition methods have been used in order to recognize facial expressions [1] [2]. M.J. Lyons et al. revised PCA and LDA to analyze the expression training sets. C. Padgett et al. used a back-propagation neural network for emotion classification, whereas T. Otsuka used a hidden Markov model based method to recognize one of six facial expressions. N. Sebe et al. proposed an emotion recognition method using a Naive Bayes model [3]. More recently, M. S. Bartlett et al. proposed Gabor feature based AdaSVM method [4]. Y. Wang et al. proposed an expression classifier that is learned from boosting Haar feature based Look-Up-Table type weak classifiers [2].

Another area of research on HRI is to generate emotion of the robots. To interact socially with humans, a robot must be able to do more than simply gather information about its surroundings: it must be able to express its state or emotion, so that humans will believe that the robot has beliefs, desires, and intentions of its own [5]. Currently, there are many robots that can express their status, such as Kismet and Leonardo at MIT, the WE-4 at Waseda University, Saya at Tokyo University of Science, and Pearl at Carnegie Mellon University.

In this paper, we propose a simple and efficient method for automatic facial expression recognition using a small set of modified rectangular features selected and trained by AdaBoost learning algorithm. We also introduce our robot, called "Ulkni," which exploits human social tendencies to convey intentionality through motor actions.

The rest of this paper is organized as follows: in Section 2, our facial expression recognition method is presented; Section 3 describes how to make artificial emotion using our robot; in Section 4, we briefly introduce the system architecture and the motor control scheme of Ulkni; Section 5 present initial experiments to show the validity of the proposed system; and finally some conclusions and further works are presented.

II. FACIAL EXPRESSION RECOGNITION

Our emotion imitation system is composed of two modules, which are facial expression recognition and artificial emotion generation (see Fig. 1). The system firstly detects human's face in the image. The proposed facial expression recognition algorithm classifies the obtained face into one of P. Ekman's basic facial expressions that include neutral, happiness, sadness, anger, surprise, disgust and fear [6]. From the result of the recognition, our emotion imitation system knows user's emotion, and it copies the recognized emotion through the following procedures: artificial emotion generation, multiple motor control, and movements of robot's eyelid and mouse.

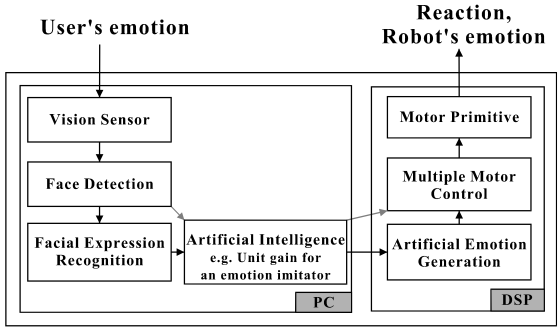

Figure 1. The whole system block diagram. The image pre-processing, face detecion and facial expression recognition algorithm run on a personal computer (PC) with a commercial microprocessor. In addition, the generation of robot's emotion and motor controller operate in a fixed point digital signal processor (DSP).

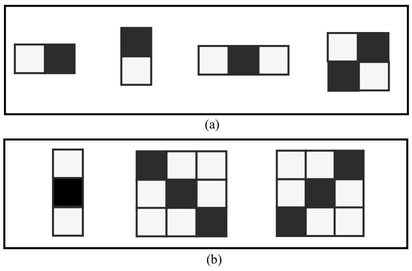

Figure 2. Modified rectangular features. In (a), four types of rectangular features which are proposed by P. Viola are shown.[7] Two additional types to focus on diagonal structures and a type for the vertical structure in the face image are shown in (b).

P. Viola et al. proposed some rectangular features to detect human faces in real-time [8] [9]. We use these simple and efficient rectangular features for real-time facial expression recognition. In addition, we propose some effective rectangular features to focus on diagonal structures in the face images. The value of rectangular features can be computed easily using integral image representation. A set of effective rectangular features are selected and trained by AdaBoost learning algorithm. In order to classify various facial expressions in real-time, we firstly use a small set of selected features to recognize initial facial expression and all selected features are used to determine the final facial expression.

A. *Modified rectangular features*

Our facial expression recognition procedure classifies facial images based on the value of simple features. The simple features used are reminiscent of Haar basis functions which have been used by Papageorgiou et al [7]. We use four kinds of features to recognize facial expressions. The value of a two-rectangle feature is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent. A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle. A four-rectangle feature computes the difference between diagonal pairs of rectangles. Finally to focus on diagonal structures in the face image we created another type of rectangular feature. These new features are illustrated in Fig. 2. The sum of the pixels in the black region is subtracted from the sum in the white region.

B. *Integral Image*

Rectangular features can be computed very rapidly using an integral image. The integral image at location (x, y) contains the sum of the pixels above and to the left of (x, y). Therefore we can define an integral image like equation (1).

$$ii(x,y) = \sum_{x' \le x,\, y' \le y} i(x', y') \quad , \tag{1}$$

where $ii(x, y)$ is the integral image and $i(x, y)$ is the original image. The integral image can be computed easily as follows:

$$\begin{aligned} s(x,y) &= s(x, y-1) + i(x,y) \\ ii(x,y) &= ii(x-1, y) + s(x,y) \end{aligned} \tag{2}$$

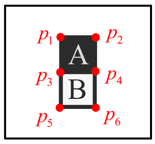

Figure 3. Two-rectangle feature value computation example. Two-rectangle feature value is the difference between the sum within A and the sum within B.

By using the integral image, any rectangle sum can be computed in four pixel references. For example, a two-rectangle feature value shown in Fig. 3 can be easily computed in six pixel references by the following manners.

Two-rectangle feature value

$$\begin{aligned} &= \sum_{(x,y)\in B} i(x,y) - \sum_{(x,y)\in A} i(x,y) \\ &= (ii(p_6) + ii(p_3) - ii(p_4) - ii(p_5)) \\ &\quad - (ii(p_4) + ii(p_1) - ii(p_2) - ii(p_3)) \\ &= ii(p_6) + 2ii(p_3) - ii(p_5) - 2ii(p_4) - ii(p_1) + ii(p_2) \end{aligned} \tag{3}$$

C. *AdaBoost learning algorithm*

AdaBoost is a simple learning algorithm that selects a small set of weak classifiers from a large number of potential features. So in our system, a variant of AdaBoost is used both to select a small set of features and train the classifier. Our boosting algorithm is basically same to P. Viola's boosting algorithm [8]. The boosting algorithm for learning a strong classifier is shown below. From this procedure, T weak classifiers are constructed each using a single feature. The final strong classifier is a weighted linear combination of the T weak classifiers.

- Given sample images (x_1, y_1), ..., (x_n, y_n) where $y_i = 0, 1$ for negative and positive samples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, ..., T$,

  1) Normalize the weights so that w_t is a probability distribution.

  $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$

  2) For each feature j, train a weak classifier h_j.

  $$h_j(x) = \begin{cases} 1 & , if\ p_j f_j(x) < p_j \theta_j \\ 0 & , \quad otherwise \end{cases}.$$

  A weak classifier $h_j(x)$ consists of a feature value f_i, a threshold value Θ_j and a parity p_j indicating the direction of the inequality sign. And the error is computed with respect to w_t,

  $$\varepsilon_j = \sum_i w_i \left| h_j(x_i) - y_i \right| .$$

  3) Choose the classifier ht with the lowest error ε_t.

  4) Update the weights.

  $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$, where $e_i = 0$ if sample xi is classified correctly, $e_i = 1$ otherwise.

- The final strong classifier is

$$H(x)=\begin{cases}1 & , if \sum_{t=1}^{T}\alpha_t h_t(x) \geq 0.5\sum_{t=1}^{T}\alpha_t \\ 0 & , \quad otherwise\end{cases}, \quad \alpha_t=\log\frac{1}{\beta_t}.$$

D. Expression estimator

Our method to recognize a certain facial expression from multiple expressions is to divide the space of expressions into various classes and train different classifiers for each expression class. In order to avoid the computational expense, we use a two-stage approach. Firstly, we estimate the expression of the face and then evaluate only the classifier trained on that expression. The expression estimator is a multi-class classifier.

III. Emotion Expression

A. Definition of Facial Expression

Our emotion generation method is based on the facial action coding system (FACS) proposed by P. Ekman [6]. He defined 6 basic facial expressions, which are happiness, sadness, surprise, fear, disgust, and anger. In our system, we consider 7 facial expressions including neutral clues to make 6 basic facial expressions are shown in Table 1.

B. Generation of Artificial Facial Expression

To make an artificial facial expression for our robot, we let our robot make some movements, e.g. wrinkles, but the robot can't do this because it doesn't have any actuator to move the desired facial components. Therefore we should adapt the clues of each facial expression to our robot. We choose what our robot can express directly or indirectly, and apply the selected clues to our robot.

TABLE I. Clues of Human's Facial Expression.

Emotion	Clues
Happiness	- Corners of lips are drawn back and up. - The mouth may or may not be parted, with teeth exposed or not. - A wrinkle runs down from the nose to the outer edge beyond the lip corners. - The cheeks are raised. - The lower eyelid shows wrinkles below it.
Sadness	- The inner corners of the eyebrows are drawn up. - The skin below the eyebrow is triangulated, with the inner corner up. - The upper eyelid inner corner is raised. - The corners of the lips are down or the lip is trembling.
Surprise	- The brows are raised, so that they are curved and high. - The skin below the brow is stretched. - Horizontal wrinkles go across the forehead. - The eyelids are opened - The jaw drops open so that the lips and teeth are parted, but there is no tension or stretching of the mouth
Anger	- Vertical lines appear between the brows - The lower lid is tensed and may or may not be raised. - The upper lid is tense and may or may not be lowered by the action of the brow. - The eyes have a hard stare and may have a bulging appearance. - The lips are pressed firmly together, with the corners straight or down.
Disgust	- The upper lip is raised. - The lower lip is also raised and pushed up to the upper lip. - The nose is wrinkled. - The cheeks are raised. - Lines show below the lower lid, and the lid is pushed up but not tense. - The brow is lowered, lowering the upper lid.
Fear	- The brows are raised and drawn together - The wrinkles in the forehead are in the center, not across the entire forehead - The upper eyelid is raised, exposing sclera, and the lower eyelid is tensed and drawn up. - The mouth is open and the lips are either tensed slightly and drawn back or stretched and drawn back.

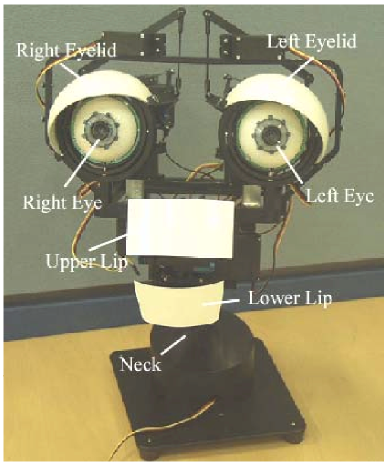

Figure 4. Ulkni's mechanisms. The system has 12 degrees of freedom (DOF). The eyes and the neck can pan and tilt independently. The eyelids also have two DOF to roll and to blink. The lips can tilt independently.

IV. System Description

We have designed our system architectures to meet the challenges of real-time visual-signal processing (about 30Hz) and real-time position control of all actuators (1KHz) with minimal latencies. Ulkni's vision system is built around a 2.6 GHz commercial PC. Ulkni's motivational and behavioral systems run on a TMS32LF2407A processor and 12 internal position controllers of commercial RC servos. The cameras in the eyes are connected to the PC by the IEEE 1394a interface, and all position commands of actuators are sent from PC. Ulkni has 12 DOF to control its gaze direction, two DOF for its neck, four DOF for its eyes, and six DOF for other expressive facial components, which are the eyelids and lips. According to [5], the positions of the system's eyes and neck are important for expressive posture, as well as for gazing toward an object of its attention (see [10] in detail).

V. Experimental Results

A. Facial Expression Recognition

A training image set consists of 1065 facial expression images with the JAFFE(Japanese Female Facial Expression) database and its modifications, e.g. mirror reflection, in-plane rotation and brightening. Using 7 rectangle feature types shown in Fig. 2, the boosting process was conducted to select 60 weak classifiers for each facial expression. And the images for testing the recognition rate consist of a total of 407 frontal face images chosen from the AR face database, PICS database and Ekman's face database [6][11].

In the process of initial face detection, the system selects some candidate regions from the entire input image area and then performs pattern classification in the candidate region. Once the face is detected, the system searches the face in the tracking window. When the face is detected, the system can deal with 20~25 image frames per second. Fig. 5 shows the results of face detection and Fig. 6

Figure 5. Multifle-face detection results. Our face detection algorithm also use modified rectangular features which have 7 types of rectangular feature.

Figure 6. Facial expression recognition results. These test images are made by just merging 6 facial expression images to one image.

Figure 7. Facial expression recognition results. This photo is captured at frame 0, 30, 60, 90, 120, 150, 180 and 210 respectively.

shows the results of our emotion recognition. To test the validity of the proposed algorithm, we make some test images which respectively include 6 facial expression images. Our system can recognize only one desired facial expression, "happiness," among all facial expressions. Fig. 7 shows the feasibility of our proposed recognition method in real-time sense. The initial face detection process takes about 250~300 milliseconds, whereas it takes 40~50 milliseconds to track the detected face region and to recognize which facial expression is in the detected face region (see Table II).

B. *Emotion Expression*

As mentioned above, Ulkni is composed of 12 RC servos, with four DOF to control its gaze direction, two DOF for its neck, and six DOF for its eyelids and lips. Therefore our system can make various facial expressions; neutral, anger, happiness, fear, sadness, and surprise, by using its eyelids and lips based on P. Ekman's observation shown Table I (see Fig. 8).

TABLE II. PROCESSING TIME OF IMPLEMENTED FACIAL EXPRESSION RECOGNITION SYSTEM.

	Average processing per image(320×240)
Initial face detection	250~300 ms
Face tracking & facial expression recognition	40 ~ 50 ms

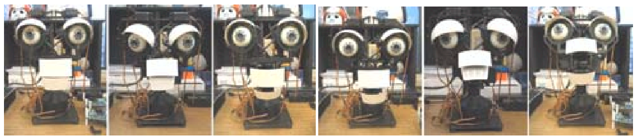

Figure 8. Ulkni's various facial expressions. There are 6 facial expressions which show his status; neutral, anger, happiness, fear, sadness, and surprise.

VI. CONCLUSION AND FURTHER RESEARCH

In this paper, our imitation system has been developed from two parts, facial expression recognition and emotion generation. And a real-time facial expression recognition algorithm is proposed using modified rectangular features selected by AdaBoost. We also developed an emotion imitation system for social interaction with humans. Therefore humans can easily perceive motor actions semantically and intuitively, regardless of what the robot intends. In conclusion, we introduced an artificial emotion imitation system using a robot head. This kind of imitation system can be used for showing the validity of real-time facial expression recognition algorithm and artificial emotion expression simultaneously.

Our research is focusing on an analysis of facial expression recognition algorithm, e.g. false rate and recognition rate. And also we consider more complex feature types suitable for recognizing facial expression.

REFERENCES

[1] M. Pantic and L. J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Transaction on Pattern Recognition and Machine Intellignece, vol. 21, no. 12, pp. 1357-1362, 1999.

[2] Y. Wang, H. AI, B. Wu, and C. Huang, "Real Time Facial Expression Recognition with Adaboost," Proceedings of the 17th International Conference on Pattern Recognition, 2004.

[3] N. Sebe, I. Cohen, A. Garg, M. Lew, and T. Huang, "Emotion Recognition Using a Cauchy Naive Bayes Classifier," Porceedings of the 16th International Conference on Pattern Recognition, vol. 1, pp. 17-20, 11-15 August 2002.

[4] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Developemnt and Applications to Human Computer Iinteraction," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, June 2003.

[5] C. Breazeal, "Designing Sociable Robots," MIT Press, 2002.

[6] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System," Research Nexus division of Network Information Research Corporation, 2002.

[7] C. Papageorgiou, M. Oren, and T. Poggio, "A general frame-work for object detection," Proceedings of International Conference on Computer Vision, 1998.

[8] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Kauai, HI, December 2001.

[9] M. Jones and P. Viola, "Fast Multi-view Face Detection", Technical Report TR2003-96, Mitsubishi Electric Research Laboratories, 2003.

[10] D. H. Kim, D. Kim, H. S. Hong and M. J. Chung, "An Image-Based Control Scheme for an Active Stereo Vision System," Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3375-3380, October 2004.

[11] A.M. Martinez and R. Benavente, "*The AR Face Database*," CVC Technical Report #24, June 1998.