Development of a Humanoid Robot for Emotion Recognition

YOUNG-GEUN SHIN[1], SANG-SUNG PARK[1], JUNG-NYUN KIM[1], DONG-SIK JANG[1],
[1] Industrial Systems and Information Engineering, Korea University
1, 5-ka, Anam-Dong, Sungbuk-Ku, Seoul 136-701, SOUTH KOREA

Abstract: This paper deals with a humanoid robot system of recognizing human's expression from a detected human's face and then showing human's emotion. A face detection method is as follows. First, change RGB color space to CIElab color space. Second, extract skin candidate territory. Third, detect a face through facial geometrical interrelation by face filter. Then, the position of eyes, a nose and a mouth which are used as the preliminary data of expression recognition are distinguished using geometrical features. Usually, when a person shows expression, he uses eyebrows, eyes and a mouth. In this paper, the change of eyebrows and a mouth are used as feature values of expression recognition. Feature values which are gained in this way are sent to a robot through serial communication. Then the robot operates a motor that is installed and shows human's emotion. The result of experiments on 60 persons shows 77.9% accuracy.

Key-Words: Emotion Recognition, Robot, CIElab, Face Detection, Expression, Filter

1 Introduction

Recently, as an increasing of the interest for robots, we can see a robot around us commonly. Like cleaning robots, security robots and service robots, Robots have a close relation with human in various fields. In the beginning, a robot is mainly developed to increase productivity in industry. But, recently the development of robots that can be utilized in daily life has grown rapidly. Especially, a human-friendly robot which recognizes human's emotion has become an important object to study. An emotion recognition robot need face detection technique and emotion recognition technique. Representative studies of face detection are by conversion of a color scene [1, 2], and using PCA (Principal Component Analysis) [3], and using LDA (Linear Discriminant Analysis) [4]. These detecting methods are coming off well about face images that see front side. But, it has defect that is hard to detect a face about contorted face images. We can divide the methods to recognize expression into the using mechanical device and using computer vision technology. In the first method, the recognition of facial muscle motion by establishing the device on user's face will be the easiest way. But, this method can give users inconvenience and activity imitations. The second method can be divided again into static expression recognition and dynamic expression recognition [5]. The expression recognition of static image has defect that is hard to find out the accurate position and shape from actual images regardless of skin color and lighting. And the expression recognition of dynamic image has a limitation of reliability, realtime responsibility, and environmental toughness etc. Therefore, it has some restrictions to use [6]. In this paper, we suggest a face and object detection algorithm using geometrical features of face and based on CIElab color space in the image sequence to solve difficulty detecting face in contorted image. In addition, to complement the defects that dynamic recognition and static recognition have, we suggest an algorithm that recognizes the variation and features of movement in window region. Emotion recognition window (ERW) can track down and recognize the change of movement. Decision of emotion is attained by selecting emotion which has the nearest distance with emotion cluster defined in the experiment. This paper is organized as follows. In section 2, we suggest the method of face detection. And our suggested method of emotion recognition is explained in section 3. Then section 4 shows Composition of recognition system and section 5 shows experiments, analysis of result. Finally concluding remarks are given in section 6.

2 Tracking Facial Components

Complexion samples used in this paper to make candidate area of skin color are 20 Korean men images found in the internet. Data space of RGB that is extracted in 20 sample face images is converted into CIElab area [7], and we decide skin candidate area, and then detect face through facial geometrical interrelation by face filter.

After detecting face, detect the territory of eyes, eyebrows, nose, and mouth using data of detected face. Position of eyes in this paper is the most important factor to recognize emotion, because position of nose, mouth and eyebrows can be detected via the position of eyes.

2.1 Detecting the area of eyes

Process that detects eyes is as follows.

Step1. Normalize after converting data of detected face area to gray level image.

Step2. Search central point of image using equation (1) in normalized image.

$$C_{(x,y)} = (imgwidth, imgheight) / 2 \tag{1}$$

Central coordinate of image is denoted by $C_{(x,y)}$. The symbol *imgwidth* represents width of image and *imgheight* represents height of image.

Step3. For the amount of operation and rate of emotion recognition, we assume that human's expression changes from side to side uniformly and use only left area of image as data of expression recognition.

Step4. Find accumulation value of gray level through scanning separated data of the top part of left side to Y-axis direction by line. Human's eye area is closest from facial center to the direction of the top part of left side and less value of gray level than any other area. So if accumulation value of gray level is in the range of fixed thresholds resulted from repeated experiment, select the area as the candidate area of eyes. The accumulation value of gray level of selected area is evaluated by using equation (2).

$$accum(y) = \sum_{x=0}^{W/2} G(x,y) \qquad y = 0,1,\ldots,H/2 \tag{2}$$

Symbol *H* represents the height of image and *W* represents the width of image. And the gray level of *(x,y)* is denoted by *G(x,y).*

Step5. Y axis bottom's coordinate of the square of eye area is obtained by using *accum (y)*. This value is less than the accumulation value of gray level of other line. After obtaining the bottom’s coordinate if *accum(y)* through scanning is bigger than any fixed value, stop scanning. At this time, y axis coordinate value is the top coordinate value of eye area.

Step6. In y axis coordinate section obtained in step5, we can obtain the coordinate of the eye area square by scanning each line to x axis direction.

This area is the final eye area. Figure 1 shows a process of scanning for searching eye area.

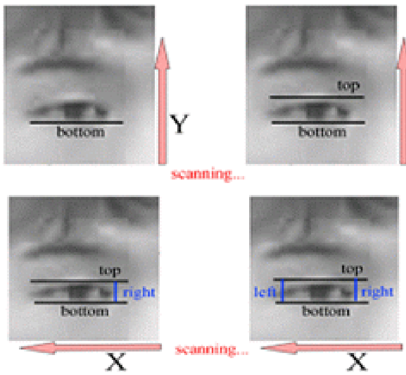

Fig 1. Process of scanning for searching eye area

2.2 Detecting the area of nose

Nose is not used in recognition of expression. But the area of nose is used to detect the area of mouth. An algorithm to detect nose is as follows.

Step1. Convert the image of gray level into the image of binary level and scan in lower direction from the center point of eye area.

Step2. Verify whether the values of detected area are less than threshold.

Step3. If a target area is verified, verify whether candidate area is one or more.

Step4. If it is one, the candidate area is selected as nose. Otherwise repeat Step1.

2.3 Detecting the area of mouth

An algorithm to detect the area of mouth is as follows.

Step1. Scan in lower direction from the center point of detected nose area.

Step2. If a value of target area is less than threshold, Select as mouth area after verifying a width ratio of target area. If a value of target area is more than threshold, decide after next scanning. If the area is not detected, repeat Step 1.

3 Emotion Recognition

3.1 Emotion Recognition Window

Eyebrows and mouth are the feature point of face used in the experiment for emotion recognition. We measured a variation by making Emotion Recognition Window (ERW) around eyebrows and mouth. Then we recognized 4 emotions on human's face by calculating dimension, distance and angle etc. which are basis of emotion recognition.

3.1.1 Decision of window size

Because of the operation speed like the other studied which use image sequence, it is right way to measure the amount of change in minimum area. We make ERW which uses geometrical data of the area of eyebrows and mouth in this paper. ERW are window of the part of eyebrows (ERW-E) and window of the part of mouth (ERW-M). In the ERW-E case, window size is decided by detected square of eye area in normalized image. The width of ERW-E is two times longer than the width of square of eye area based on the center coordinate of square. The reason why the size of window is two times longer than the square of eye area is because of the changing position. When the location of eyebrows is changed, the additional area will be added to original location. The result of experiment shows the area of eyebrows will be bigger two times than eye area. Detected eye area in normalized image is 30 × 15 Pixel in this paper. We decide the size of ERW-E is 60 × 30 Pixel considering the change of eyebrows. Divided window in this way is made into window like figure 2 has 20 × 10 pieces of block that do 3 × 3 Pixel as one block interiorly.

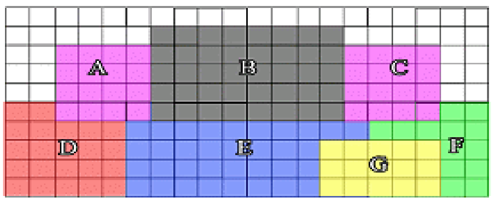

Fig 2. Emotion Recognition Window (ERW-E)

3.1.2 Extraction of emotion features

If change is sensed in image sequence, obtain three emotion features and decide emotion with the given data. Three emotion features are as follows.

A. An occupied area ratio of each window that is obtained in ERW-E, ERW-M
B. Distance between representative point in ERW-E and the square of eye area
C. Angle between the representative point in ERW-E and the topside of a square of eye area

The target data is binary level. Because 3 × 3 Pixel becomes one block, we must select mean value in each block. In case of mean value, if value which is divided sum of pixel value by 9 through equation (3) is larger than 0.5, do appropriate block value by 0 otherwise by 1.

$$B_{r\_avg} = \sum_{y=1}^{3}\sum_{x=1}^{3}(I(x,y))/9$$
$$B_{avg} = \begin{cases} 1 & if(B_{r\_avg} > 0.5) \\ 0 & otherwise \end{cases} \quad (3)$$

Sum of window weight in each emotion is always 1. The larger a positive value is, the more important a window is. Weight of obtained emotion feature using occupancy area ratio of window in ERW-E is shown in table 1. Weight is the data that is obtained in the experiment.

Region / Emotion	A	B	C	D	E	F	G	etc
Happiness	0.2	0.3	0.2	0.1	0.4	0.1	-0.1	-0.2
Surprise	0.3	0.4	0.3	0.1	0.3	-0.1	-0.2	-0.1
Sorrow	0.05	0.05	-0.1	0.1	0.3	0.5	0.4	-0.3
Angry	0.0	0.0	-0.1	0.1	0.2	0.5	0.6	-0.3

Table 1. Weight of area according to emotion of ERW-E

Ar, *Br*, *Cr*, *Dr*, *Er*, *Fr*, *Gr*, *etcr* are denoted by the binary occupancy ratio of each block. Then emotion feature can be obtained through equation (4).

$$F_w = A_r \times A + B_r \times B + C_r \times C + D_r \times D + E_r \times E + F_r \times F + G_r \times G + etc_r \times etc \quad (4)$$

In the case of ERW-M, measure amount of change of the left side area among whole window area. Emotion feature in ERW-M shows an organization ratio of binary area. Also, another emotion feature is the value which is expressed numerically both the distance and angle between a representative point of an eyebrow (Point1, Point2) and a square of eye area. It is shown in figure 3. Measure the distance between top points of left side, middle, right side of a square of eye area and three points which meet in vertical direction at those points. And measure the angle between the topside of a square of eye area and joining a top point of right side of a square of eye area to a top point of middle.

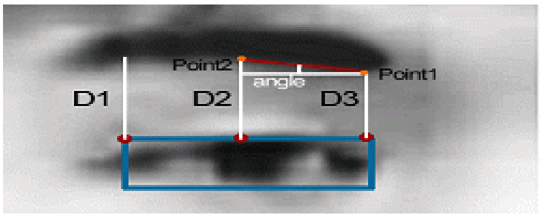

Fig 3. Distance and angle between an eyebrow and square of eye area

(x, y) is denoted by the coordinate of a left side point of a eye area square. Point1, Point2 are denoted by the point which meets an eyebrow. Angle is denoted by the angle between Point1 and Point2. Symbol A represents a width of eye area square. We can obtain the distance between a square of eye area and an eyebrow(D1, D2, D3 ; equation(5-7)) and the angle between the topside of a square of eye area and joining Point1 to Point2 (equation(8)) from following equations. (p_{x1}, p_{y1}) is denoted by coordinate of an eyebrow that meets with left apex of a square of eye area.

$$D1 = \sqrt{(x - p_{x1})^2 + (y - p_{y1})^2} \quad (5)$$

$$D2 = \sqrt{((x + \frac{A}{2}) - p_{x2})^2 + (y - p_{y2})^2} \quad (6)$$

$$D3 = \sqrt{((x + A) - p_{x3})^2 + (y - p_{y3})^2} \quad (7)$$

$$\cos\ \alpha = \frac{A/2}{\sqrt{(p_{x2} - p_{x3})^2 + (p_{y2} - p_{y3})^2}} \quad (8)$$

Points of an eyebrow can be obtained when occupancy ratio of window is calculated. Data (F_{m2}) is shown in equation (9).

$$F_{m2} = (\cos\ \alpha : \frac{D3}{D2}) \quad (9)$$

3.2 Decision of emotion

We can select emotion by matching emotion features (F_{m2}) that is obtained by distance and angle with an eyebrow and a square of eye area to emotion data which is obtained in the experiment. It is shown in figure 4. And select emotion by comparing emotion features (F_{m1}) that are component ratio of binary area in ERW-M with four emotion data that is mapped previously. Then we can select emotion by comparing F_{m2} with F_{m1}. The nearer F_w that is obtained by equation (4) is to 1, the more similar extracted data is to actual expression. If F_w is larger than threshold (0.9), extraction of emotion is done right. We decide emotion if emotion that is concluded by F_w equals emotion that is concluded by comparing F_{m1} with F_{m2}.

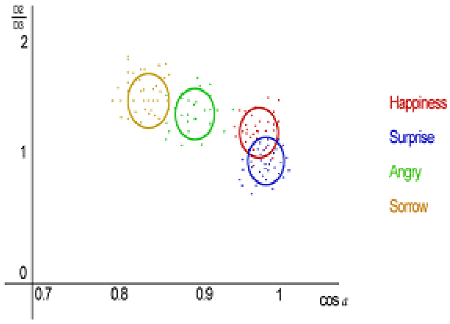

Fig 4. Ratio of distance and angle about each emotion

4 Composition of system

4.1 Diagram of system

Composition diagram of whole system that is suggested in this article is shown in figure 5.

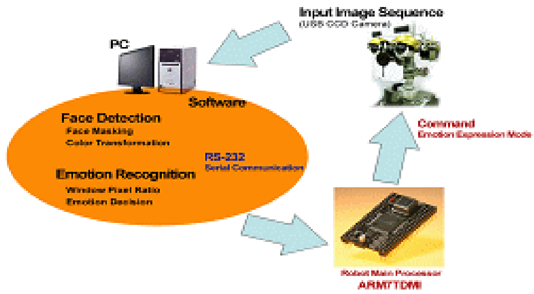

Fig 5. Diagram of whole system

Image sequence is gone through a process of face detection using CCD camera loaded in robot. After a process of face detection, recognize emotion using features of eyebrows and mouth. Then recognized emotion data is transmitted through RS232 serial communication to robot. The transmitted data is passed to main processor ARM board, and robot expresses human's emotion by running RC servo motor based on the transmitted data.

4.2 Hardware system

We make robot which can have more clear and natural look than human's expression by designing greatly eyelid, mouth and pupil for robot to do well facial representation in this paper. It is shown in figure 6.

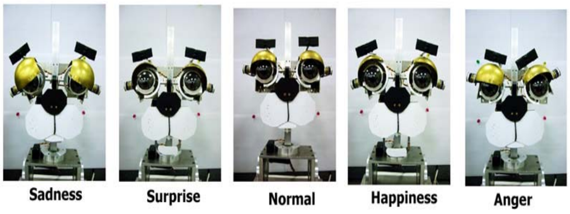

Fig 6. Motion that is expressed by robot about each emotion

4.3 Software system

As seen in figure 7, software processes emotion data and expresses emotion data on screen.

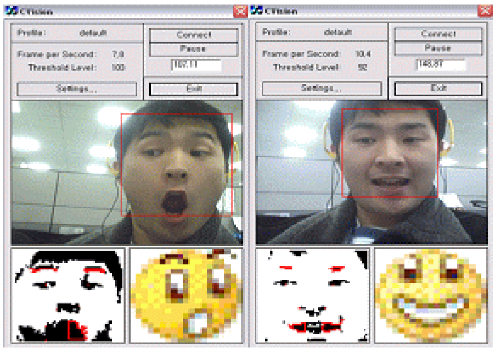

Fig 7. Recognition of surprise and happiness

5 Experiments and Analysis of Result

5.1 Experimental environment

These experiments were performed indoors, and camera is equipped on the part of robot head, we assume that only one reagent comes in camera. A fundamental process of experiment is shown in figure 8.

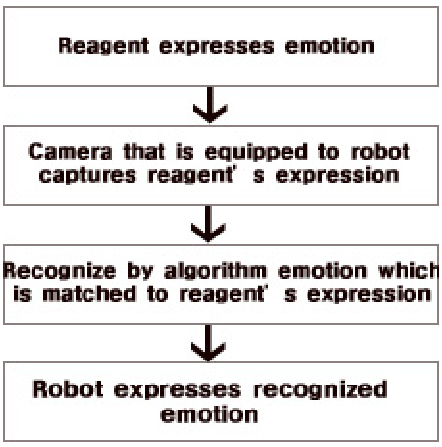

Fig 8. Diagram of process

5.2 Result and analysis

In this experiment, we performed about 4 emotions of surprise, happiness, angry, sorrow. After letting 30 reagents of each man and woman express emotions, measure whether failure or success by matching emotion which robot expresses to emotion which reagent expresses actually. Also, experiments about each emotion are performed 50 times. Experiment result for 4 emotions is as following.

	Angry	Happiness	Surprise	Sorrow
Man	70.8%	86.0%	83.4%	70.6%
Woman	70.6%	87.4%	85.0%	69.2%
Total	70.7%	86.7%	84.2%	69.9%

Table 2. Success ratio of emotion recognition

Because mouth and eyebrows change greatly compared with other area in case of expressions of surprise and happiness, expressions of surprise and happiness have high recognition ratio among 4 emotions like shown in table 2. Specially, because expression about these emotions was clear in woman's case, recognition ratio was better than in case of man. But in case of expressions of angry and sorrow, many failure factors are happened according as the action about some emotion varies each reagent. From the analysis of various failure factors, we knew that each reagent expresses a variety of facial looks at each time, even though same reagent expresses. To reduce this ratio of failure, an experiment that uses greater and various samples is needed. According to a study of emotion recognition using Hidden Markov Model(HMM) recently [8] among studies of emotion recognition, Single HMM shows 55 % as success ratio and Multilevel HMM shows 58 % in case of Person-Independent experiment intended for 5 reagents. This article comparing with existing studies shows 77.9 % as improved success ratio of emotion recognition.

6 Conclusions

Humanoid robot system that recognizes person's emotion is developed in this paper. We suggest a method of face detection, a method of emotion recognition and a methodology of robot system based on vision for this. Existing studies of emotion recognition are by using HMM algorithm and conceptual fuzzy set and etc. But these studies have problems with low reliability about ratio of recognition and construction of robot system. This paper shows greater recognition ratio of emotion than existing studies and realizes head of emotional intelligent robot. So, this can be applied to development of intelligent module that has the function of vision on movable robot that is studied briskly. But studies about these fields are needed a lot, because the objective information about expression is much insufficient as yet. Also an intelligence system that enables robots to accumulate intelligence and utilize it themselves is needed. For this, it will be performed to study about system that enable feedback using SVM [9], genetic algorithm etc.

Acknowledgement:
This work was supported by the Brain Korea 21 Project in 2006.

References:

[1] Y. Dai, Y. Nakano, Face-texture model-based on SGLD and its application in face detection in a color scene, *Pattern Recognition*, Vol. 29, No. 6, 1996, pp.1007-1017.

[2] H. Wu, Q. Chen and M.Yachida, Face Detection From Color Images Using Fuzzy Pattern Matching Method, *IEEE Tran. On Pattern Analysis and Machine Intelligence*, Vol.21, No.66, 1999, pp. 557-563.

[3] J. Zhang, Y. Yan, M. Lades, Face recognition: eigenface, elastic matching, and neural nets, *Proceedings of the IEEE*, Vol. 85, No. 9, 1997, pp. 1423-1435.

[4] W. Zhao, R. Chellappa, and N. Nandhakumar, Empirical performance analysis of linear discriminant classifiers, *In Proceedings of the 1998 Conference on Computer Vision and Pattern Recognition*, 1998, pp.164–169.

[5] Daewon Kim, *VR Tech Memo 98-04*, Virtual Reality Laboratory, Dept. of Computer Science, KAIST, 1998.

[6] Mark Rosenblum, Yaser Yacoob, and Larry S.Davis, Human Expression Recognition from Motion using a Radial Basis Function Network Architecture, *IEEE Transaction on Neural Network*, Vol. 7 , No.5, 1996, pp.1121-1138.

[7] Christine Connolly, Thomas Fliess, A Study of Efficiency and Accuracy in the Transformation from RGB to CIELAB Color Space, *IEEE Transactins on Image Processing*, Vol. 6, No.1. 1997, pp.1046-1048

[8] Ira Cohen, Ashutosh Garg, Thomas S.Huang, *Emotion Recognition from Facial Expressions using Multilevel HMM Neural Information Processing Systems*, Proc. NIPS Workshop on Affective Computing, Colorado. 2000.

[9] Christopher J.C. Burges, *A tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery, 2, 1998, pp.121-167