K-059

Interpretation of Emotional Gestures by Considering Hands Positions and Face Features

李 皆† ロクマン ジュアンダ† 今井 順一† 金子 正秀†
Jie Li Juanda Lokman Jun-ichi Imai Masahide Kaneko

1. Introduction

People tend to have special facial actions and special hand gestures which imply the state of feelings in our daily life. This paper mainly discusses the method to interpret emotional gestures by considering relative positions of hands and face. Based on an preparatory investigation, this research predefines certain kinds of emotional gestures which are thinking with left hand on the left cheek or on the left forehead, thinking with right hand on the right cheek or on the right forehead, yawning with hand covering mouth which implies sleepiness, crying with hands under the eyes and showing signs of worry with hands on the head. Samples of those input pictures are shown in Fig. 2(a).

Occlusion between hands or between hand and face occur in those predefined gestures, therefore colored gloves are used to make it easy to track and distinguish between hands region and face region. Although each of the recognition of facial expressions and that of hand gestures has been studied intensively, the interpretation of relative position between hands and face to represent emotional conditions has not been thoroughly studied.

2. Interpretation of Emotional Gestures

2.1 Face and Hand Extraction

The system extracts the face region using its skin color. Skin-color extraction method detecting human faces with varying illumination condition and presence of complex background has been proposed by Quan Huynh-Thu in [1]. The skin-color extraction relies upon a Gaussian Mixture Model (GMM), and efficient skin-color segmentation is achieved by computing an optimal threshold value for each of the Gaussian components of the mixture model. Hands with colored gloves are detected by using the H and S value of the glove color. Fig. 2(b) shows the results of skin-color and glove-color extraction.

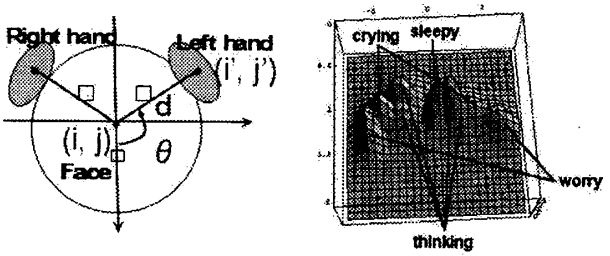

Fig. 1: (a) The relative positions of hands and face when worry. (b) 2D Gaussian distribution result.

2.2 Relative Position of Face and Hands

The center of gravity (i, j) of skin-color blob is calculated as origin of the coordinate of each frame, and the center of gravity (i', j') of glove-color is also calculated. Face area is divided into two parts using a vertical line passing the origin. As for natural emotional gestures, hand in one part seldom crosses this line to the other part. Right hand is supposed to be on the right side of the line, and vice versa. By transforming into polar coordinate, the distance d and angle θ show relative positions between face and hand.

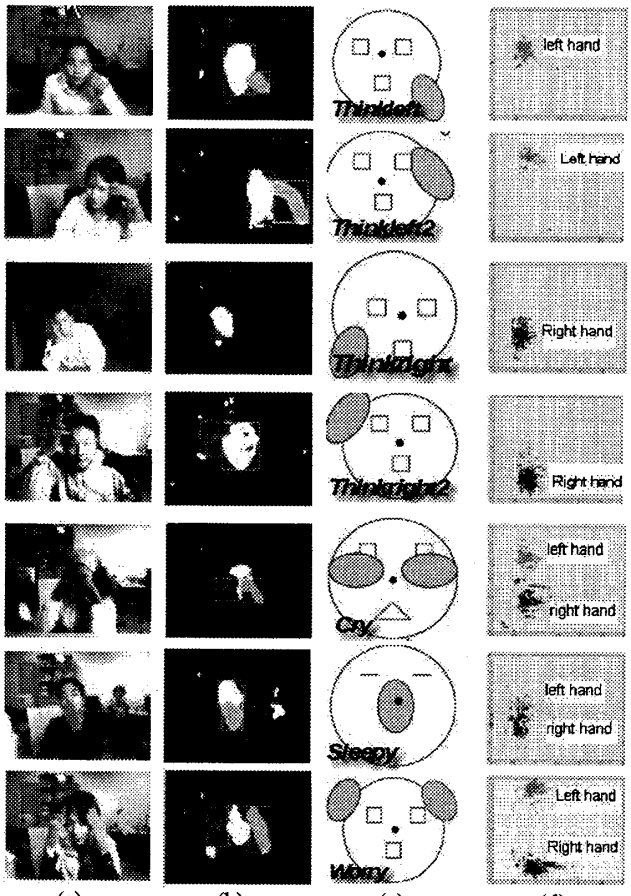

Fig. 2: (a) input images, (b) hand and face detection result, (c) models showing the relative position of hands and face, (d) distribution of center of gravity for hand

The case of worry is shown in Fig. 1(a) and all the models are shown in Fig. 2(c). Distance d is divided by the width of face contour as a normalized distance (d'), because distance d and width of face blob has a linear characteristic while the distance between human and camera is changing. The width of the face blob is used as a reference for it has a relatively stable value while facing the camera compared to height of face blob in the experiment. d' is used as a value to define the near face area and not near face area. Distribution of (d', θ) of the right hand and the left hand is shown in Fig. 2(d).

† 電気通信大学, UEC

2.3 Rotation of Face

Rotation of face is a big problem when considering the relative position of face and hands. An ellipse fit is given to face blob. A rotation of the head/face (Θ) is obtained in each frame. For each person and each gesture, the rotation angle distribution is different. Positive Θ represents the rotation to left, while negative represents the rotation to right.

2.4 Classification

Normalized distance d' was used to classify the near face area and not near face area and θ' is added with Θ. Distribution of data (d', θ') obtained in the experiment is modeled by 2D Gaussian distribution for each hand. 2D Gaussian distribution shown below is used to model and classify the near face area.

$$p(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu))} \quad (1)$$

Where μ is the vector of mean, Σ represents the covariance of the distribution. The 2D Gaussian distributions of predefined gestures in the near face area are used to classify the input gestures. 2D Gaussian distributions of all the predefined gestures in the near face area are shown in Fig. 1(b).

2.5 Face Features

(a) (b)

Fig. 3: (a) Mouth absence, (b) Mouth presence

Sleepy and thinking are two states which can sometimes confuse human as shown in Fig. 3(a) and (b). To cope with this problem, mouth and eyes are detected by using the contrast filter,

$$s = ie^{-(i/i_0)^4} \quad (2)$$

where s and i denote the output and input gray level respectively and i_0 is a constant.

3. Experiment

For the skin-color based approach, people are required to wear long-sleeved clothes of which colors are different from the skin in this system. Twenty samples from participants were taken for the research. Predefined gestures were performed in different situations, distance between human and camera and rotation of face varied from person to person.

Sample images of the input sequence are shown in Fig. 4(a). We suppose that there is only one hand in the near face area for some emotional gestures implying thinking and sleepiness, which only one 2D Gaussian distribution is given to these gestures. This is also one of the features used for classification. Hands distributions of predefined gestures which have the highest probability are calculated and detection results are given to each frame as shown in Fig. 4(b). Detection rate of single frames is about 85% by only using the relative position information. By studying a sequence of gestures, short sequence which corresponds to gestures' changing is discarded in the decision making.

Our system gives the interpretation results in real time by capturing 9 frames per second, which is implemented on a Pentium Celeron 1.0 GHz processor and 512 MB memory.

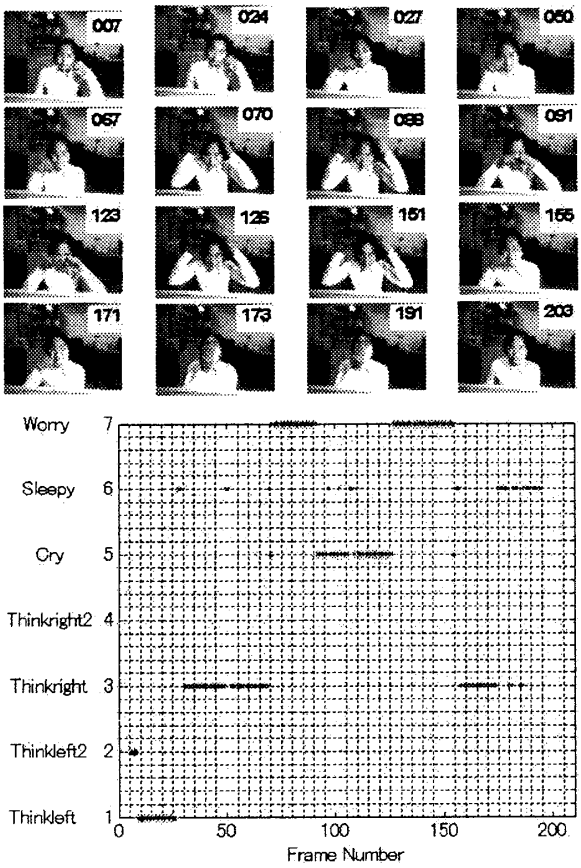

Fig. 4: (a) Sample images in the input sequence, (b) Detection result of each frame

Detection rate of the case of sleepiness and thinking, or one part of crying is insufficient when only using the relative position information. Robust interpretation of these cases can be given by using the presence of mouth or not.

4. Conclusion and Future Work

The relative position between hands and face is used to interpret some predefined emotional gestures and also the face features. This research gives visual support to understand human's feeling for robot which is an emotional companion of human. The proposed system gives proper interpretation according to the emotions in real time.

It is impractical for human to wear color gloves in human daily lives or office environment. Tracking hands during this overlapping duration remains a major issue in the vision-based approach in the future. Detailed information of face features, such as open or closed eyes and mouth, need to be studied to enhance this system.

Reference

[1] Quan Huynh-Thu, Mitsuhiko Meguro, and Masahide Kaneko, "Skin-color Extraction in Images with Complex Background and Varying Illumination," IEEE WACV2002, pp.280-285, USA, 2002.12.