Spotting Recognition of Head Gestures from Color Image Series

Haiyuan WU, Tadayoshi SHIOYAMA and Hirotomo KOBAYASHI
Dept. of Mechanical and System Engineering, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan.
e-mail: {wuh, shioyama}@ipc.kit.ac.jp

Abstract

This paper presents an approach for spotting recognition of human head gestures from color image series. First, we use an algorithm based on perceptually uniform color system to detect skin color region and hair color region of the face from input images. Then, the 3D pose of the head relative to the camera is estimated by calculating the primary moment and secondary moment of the skin color region and the hair color region. The standard patterns of each gesture were represented by the sequence of rotation angles in X , Y , Z axis, and the head area. Moreover, the human head gestures were recognized by using the CDP(Continuous Dynamic Programming) algorithm to compare input image series with the standard patterns. Extensive experiments show the effectiveness of this approach in the human head gestures recognition from live video sequence with different people even wearing glasses, with different head size, and with unknown complex background.

1 Introduction

Gestures are an important form of communication between people. We regard expressions of the face as one of the most natural forms of human expression and communication. For human gesture recognition, several studies have been reported recently. They have taken into account of the following features: a) time difference between sequential frames [1], b) silhouette of the target human [2], c) eigenvalue[3], and d) tracking the facial features [4]. But these features have following problems: 1) They depend on the target human, position, size and brightness. 2) They are not suitable with frame-wise calculation for real-time recognition. 3) They have instability depending on interpreting contents in each frame image. In this paper, we present an approach to recognize human head gestures, which can solve these problems.

2 Human head gesture

Figure 1. The head coordinate systems.

Figure 2. Recognisable gestures.

The 3D pose of the human head relative to the camera is represented by the rotation angles around X , Y , Z axes of head coordinate system, or its translation (See Fig.1). We define 10 kinds of head gestures(See Fig.2). They were defined through a questionnaire survey from many peoples with different nationalities. The defined gesture patterns are represented by the change of 3D pose of human head in time-sequence.

3 Extracting the head

The terms “skin color” and “hair color” are subjective concepts obtained by human’s observation through eyes. In order to obtain the results similar to human’s vision system, the color system used should agree with characteristic

of human's vision system. In our research, a uniform chromatic color system is used, which has been proposed by Farnsworth. Using the *Skin Color Distribution Model* and the *Hair Color Distribution Model* [5, 6], we can extract the skin and hair color region from color input images. Using the extracted color information, we can determine the head region according to the integral projection method [7].

4 Estimating the pose

Various kinds of information can be used to estimate the pose of the head in the image, such as facial features, shape connection about the skin and the hair region, etc. We use the information about the skin and hair color region to estimate the pose of head because they can be extracted from color images stably and both of them are not sensitive to the changes of facial expression, wearing glasses, and other local changes of facial feature. As a result, our approach of 3D head pose estimation is stable enough.

- The head rotation angle around Z axis: Since the shape of head is symmetrical, and it can be considered as an elliptical sphere approximately, we can infer the tilt of the head in the image θ_z by estimating the axis of least inertia of head region.

- The head rotation angle around Y and X axis: Since the face is approximately symmetric, we can estimate the horizontal rotation angle of the head θ_y by checking the difference S_x between the X -coordinates of the centers of the gravity of the head region and the one of the skin color region. We calculate the direction and approximate angle of the head as following:

$$\theta_y = \theta_{y0} + a_y S_x, \quad (1)$$

where, θ_{y0} is the horizontal rotation angle of the head when $S_x = 0$. And a_y is the gradient of the approximate straight line, which is obtained based on the investigation results from face image series of many men and women. Since the size of face is not fixed, the S_x needs to be normalized.

Figure 3. The centers of the hair, skin and face.

Similarly, we can calculate the direction and approximate angle of the head as following:

$$\theta_x = \begin{cases} \theta_{x1} + a_1 S_y & (S_y \leq S_{y0}), \\ \theta_{x2} + a_2 S_y & (S_y > S_{y0}), \end{cases} \quad (2)$$

where, θ_{x1} and θ_{x2} are the vertical rotation angles of the head when $S_y = 0$, respectively. And a_1 and a_2 are the gradients of the approximate straight line, respectively. The S_{y0} is the value of S_y when the head in frontal pose.

5 Head gesture recognition by CDP

In our approach, there are four features are extracted. Three of them are about angles, θ_x , θ_y , θ_z , another one is about head area a_h . This series of input feature vectors and the standard patterns corresponding to each gesture are matched in the CDP [1].

A standard pattern $P = \{p(\tau) | 1 \leq \tau \leq T\}$ is expressed as a series of the feature vectors $p(\tau) = \{p_k(\tau) | 1 \leq k \leq 4\} = \{\theta_x, \theta_y, \theta_z, a_h\}$.

A similar vector series is also obtained from the input images. By denoting this feature vectors as $i_k(t)$ ($0 \leq t \leq \infty$), the local distance $d_k(t, \tau)$ of $i_k(t)$ and $p_k(\tau)$ is defined by the following formula:

$$d_k(t, \tau) = \|i_k(t) - p_k(\tau)\|; \quad 1 \leq k \leq 4. \quad (3)$$

A cumulative distance is denoted as $M_k(t, \tau)$. Initial condition:

$$M_k(-1, \tau) = M_k(0, \tau) = \infty \quad (1 \leq \tau \leq T). \quad (4)$$

Recurrence formula ($1 \leq t$):

$$M_k(t, 1) = 3d_k(t, 1), \quad (5)$$

$$M_k(t, 2) = \min \begin{cases} M_k(t-2, 1) + 2d_k(t-1, 2) + d_k(t, 2), \\ M_k(t-1, 1) + 3d_k(t, 2), \\ M_k(t, 1) + 3d_k(t, 2), \end{cases} \quad (6)$$

$$M(t, \tau) = \min \begin{cases} M(t-2, \tau-1) + 2d(t-1, \tau) + d(t, \tau), \\ M(t-1, \tau-1) + 3d(t, \tau), \\ M(t-1, \tau-2) + 3d(t, \tau-1) + 3d(t, \tau), \end{cases} \quad (3 \leq \tau \leq T). \quad (7)$$

Figure 4. The path of search by CDP matching.

Assume the motion ended at time t . Then, for $\tau = T$, $M_k(t, T)$ represents the cumulative distance to match the input time series best in the section $[1, T]$ of the standard pattern. $M_k(t, T)$ is normalized by

$$A_k(t) = \frac{1}{3T} M_k(t, T), \quad 1 \leq k \leq 4. \quad (8)$$

$$E(t) = \sum_{k=1}^4 \omega_k \cdot A_k(t) / \sum_{k=1}^4 \omega_k \quad (9)$$

where, $\{\omega_k \mid k \in \{1, 2, 3, 4\}\}$ are the weights. $E(t)$ is the output value of the CDP corresponding to the standard pattern at that time. Assume there are I standard patterns. The $E_l(t)$ ($l = 1, 2, \dots, I$) will be produced at time t . The ideal values of $E_l(t)$ for a gesture which corresponds to the input gesture, is expected to be a local minimum toward the end of the motion.

6 Experimental results

To make sure the effectiveness of the proposed method, we have implemented a human head gestures recognition system, and used it to test some real head motion image series. A PC (PentiumII-266MHz), a 3CCD video camera and a low-priced live image capture card are used. The input image series are full color image with 320×240 pixels. The main processing time:

Table 1. The main processing time.

processing contain	processing time(ms/frame)
Head extraction	21.4
Pose estimation	3.2
Gesture recognition	3.0
Otherwise	9.5

Table 2. The recognition result for same person.

Gesture name	Recognition rate(%)
Yes	100
No	95
Have doubts	100
Close observation	93
Surprise	100
Relax	100
Look up	87
Look down	97
Look right	100
Look left	100

- *Where the tester is the same person that created in the standard pattern:* The tester's head size is changed from 50×80 pixels to full size image. Each gesture was tested by over 50 sets of input image series. About 97.2% of human head gestures are recognized successfully, even wearing glasses, with different clothes, or with unknown background.

- *Where the experiments are conducted for different testers:* We use 7 men to test our system. Each gesture of each man was tested by 10 set of input image series. About 80% of human head gestures are recognized successfully.

7 Conclusions

This paper has described a new approach to recognize the human head from color image series. One of the contribution is the feature vectors which are obtained from the information of the skin color region and the hair color region. Since the color information can very robust and is easy to be extricated from color input image, our approach is very efficient and reliable. The experiment results show our approach can recognize successfully even for different people wearing glasses, with a wide change of head size, and unknown complex background. Merging this method with expression recognition system and/or voice recognition system, we can achieve a big progress in the man-machine communication toward the real application.

References

- [1] Takehashi, Seki, Kojima and Oka, "Spotting Recognition of Human Gestures from Time-Varying Images", ICAFG'96, pp. 318–322, 1996.
- [2] Wilson and Bobick, "Using Configuration states for the representation and recognition of gesture", ICCV'95, 1995.
- [3] Turk and Pentland, "Eigenfaces for recognition", Cognitive Neuroscience, 3:71-86, 1991.
- [4] Zelinsky, and Heinzmann, "Real-Time Visual Recognition of Facial Gestures for Human-Computer Interaction", ICAFG'96, pp. 351–356, 1996.
- [5] H. Wu, Q. Chen and M. Yachida, "An Fuzzy Theory Based Face Detector", ICPR'96, pp. 406–410, 1996.
- [6] Q. Chen, H. Wu and M. Yachida, "Face Detection by Fuzzy Pattern Matching", ICCV'95, pp. 591–596, 1995.
- [7] H. Wu, T. Yokoyama, Dadet and M. Yachida, "Face and Facial Feature Extraction from Color Image", ICAFG'96, pp.345-350, 1996.