2011년도 대한전기학회 하계학술대회 논문집 2011. 7. 20 - 22

Hand gesture recognition for player control

시란얀, 김진규, 염동회, 주영훈
군산대학교 제어로봇공학과

Hand gesture recognition for player control

Lan Yan Shi, Jin Gyu Kim, Dong Hae Yeom, Young Hoon Joo
Kunsan National University

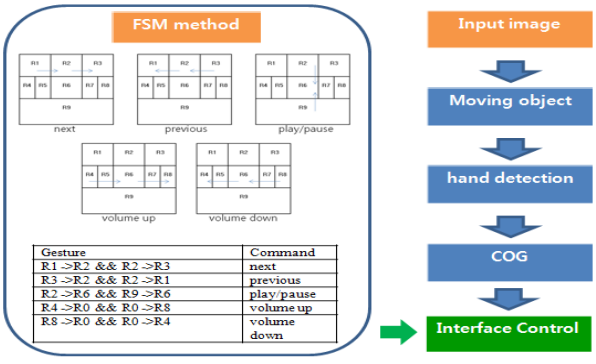
Abstract - Hand gesture recognition has been widely used in virtual reality and HCI (Human-Computer-Interaction) system, which is challenging and interesting subject in the vision based area. The existing approaches for vision-driven interactive user interfaces resort to technologies such as head tracking, face and facial expression recognition, eye tracking and gesture recognition. The purpose of this paper is to combine the finite state machine (FSM) and the gesture recognition method, in other to control Windows Media Player, such as: play/pause, next, pervious, and volume up/down.

1. Introduction

To recognize the hand gesture, it is needed to find the information contained in the input images. The important clues in the information can be made use of to control machines or robots. The gesture analysis consists of some processes such as: hand detection, hand tracking, gesture recognition and so on. Once a particular gesture is recognized, it needs to be mapped to a corresponding robot action. A user can assign the various commands to hand gestures and control intelligent robots using them [1]. The outline of the paper as follows: In Section 2, we use skin colors to detect the hand region and the FSM method to recognize the gesture command. Section 3 presents the experimental results and, Section 4 concludes with some remarks and gives an outlook on future work.

2. Gesture Recognition

The preprocessing of the gesture recognition includes the smoothing and the binarization of the input images. Fig. 1 shows the block diagram of the proposed method.



〈Figure.1〉 Block diagram

2.1 Hand detection

In this paper, we use the skin color method to detect the skin region, label the hand region and find the center of gravity (COG).

2.1.1 Skin color method

In the computer vision, there are various color models such as RGB, YCbCr, YUV, HSV and so on, which have been used to extract the skin region. In this paper, we use RGB and YCbCr color model to detect the skin region. Eq. (1) shows the relationship of RGB and YCbCr [4].

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.449 \\ 0.449 & -0.418 & -0.0813 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.344 & -0.714 \\ 1 & 1.772 & 0 \end{bmatrix} * \begin{bmatrix} Y \\ Cb-128 \\ Cr-128 \end{bmatrix}$$

Applying the threshold processing to the input image of Fig. 2, we can get the result image of Fig. 3.

〈Figure.2〉 input image

〈Figure.3〉 result image

2.1.2 Labeling and COG

After detecting the skin region, we find the hand by using the labeling and COG. Among the targets found with the breadth iteration, the biggest skin region is recognized as the hand region by the labeling. The COG algorithm is shown in (2).

$$P_{COG}(x) = \frac{Px_2 + Px_1}{2} \quad (2)$$

$$P_{COG}(y) = \frac{Py_2 + Py_1}{2}$$

〈Figure.4〉 Labeling

〈Figure.5〉 COG

In Fig. 4, $P(x_1, y_1)$ and $P(x_2, y_2)$ are two corner point on the hand region. As shown in Fig. 4 and 5, the red rectangle is

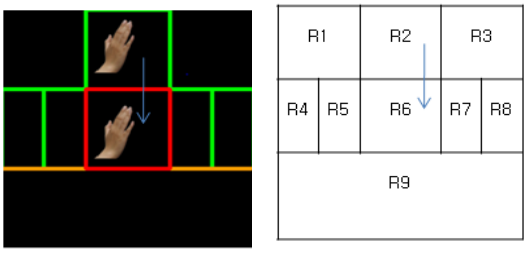
the hand region, and the blue X is the COG of the hand region.

2.2 FSM for gesture recognition

We use the simple and practical rules of the FSM method to recognize the meaning of the hand gesture.

2.2.1 FSM method

The FSM is the most commonly used for the sequential components of digital systems [3]. Logic synthesis, which converts the symbolic description of the FSM to a hardware implementation. The existing methods start with FSM state minimization and state encoding in order to optimize design objectives such as area, delay, and testability[3]. In this paper, we divided the video window into several sections named R1 to R9 as shown in Fig. 6. A gesture whose location is the COG of the hand region is found in the section R2, which becomes the start section. Next, the hand moves to the section R6. And then, we get the gesture from R2 to R6, which is defined as a command, play/pause.



<Figure.6> Divided video window

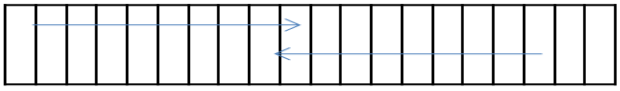
2.2.2 Definition of rules

If a gesture belongs to the defined rules, it is recognized as a meaningful command. If not, it is considered as meaningless, so the procedure turns back and waits for a new start. The Table 1 shows the five defined rules.

<Table 1> Rules

gesture	command
R2->R6 || R9->R6	play/pause
R1->R2 || R2->R3	previous
R3->R2 || R2->R1	next
R4->R0-R4 || R0-R8->R8	volume up
R8->R0-R8 || R0-R4->R4	volume down

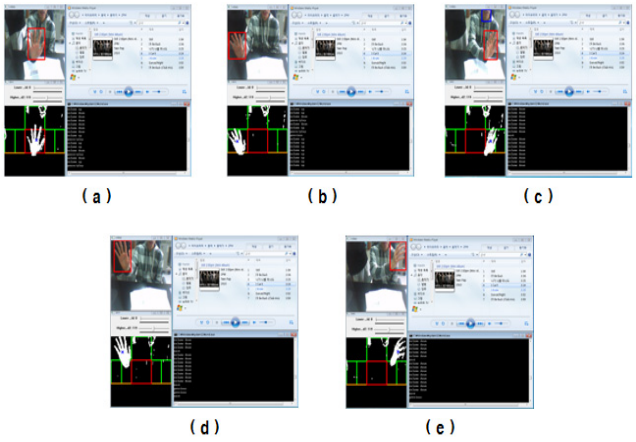
In Table 1, R0=R4+R5+R6+R7+R8. Because the volume control is a continuous process, we make the volume section larger and devide the section into 20 steps as shown in Fig. 7.



<Figure.7> volume control

3. Experimental Results

In this paper, we used Windows Media Player to have a test. This player has the keyboard short-cuts such as F8 meaning volume up, F9 volume down, SPACE play/pause, CTRL+F next, and CTRL+B previous. The output commands correspond to the keyboard short-cuts. The images used in this paper were 320x240 and 24-bit color images obtained from a web camera, with a frame rate of 17 frames/sec. The experiment was carried out in a usual office environment and the distance between camera and human was 0.4-0.6 meters. Fig.8-(a) shows play/pause control, Fig.8-(b) shows volume up control, Fig.8-(c) shows volume down control, Fig.8-(d) shows next control, and Fig.8-(e) shows previous control.



<Figure.8> Snap shots of experiments

The success rate of this experiment is shown in Table 2, which supports the reliability of this program.

<Table 2> Result

Gestures	Total text	Ture result	Rate
play/pauss	30	28	93.3%
next	30	26	86.7%
previous	30	28	93.3%
volume up	30	29	96.7%
volume down	30	30	100%

4. Conclusions

In this paper, we proposed an air interface for Window Media Player by means of the recognition of hand gestures and showed the applicability of the proposed program. But, this program has problems: If the distance between a user and a camera is too long, the success rate is deteriorated. And the brightness of the light affects the success rate of this program. We are devoting to the further research to cope with the problems.

감사의 글: 본 연구는 2011년 교육과학기술부의 재원으로 한국연구재단 지역혁신인력양성사업의 지원을 받아 수행된 연구임.

[References]

[1] M. K. Bhuyan, D. Ghosh, and P. K. Bora, "Threshold Finite State Machine for Vision Based Gesture Recognition," IEEE Indicon 2005 Conference, Page 379-382, Chennai, India, Dec. 2005.

[2] H. Hong and X. Zhu, "A Human Hand-image Detection Based on Skin-colour and Circular Degree," 2009 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing. Page 373-376, 2009.

[3] L. Yuan, G Qu, T. Villa, and A. Sangiovanni-Vincentelli, "FSM Re-Engineering and Its Application in Low Power State Encoding," Digital Object Identifier. Page 254-259. 2005.

[4] Y. Yang, P. Yuhua, and L. Zhaoguang, "A Fast Algorithm for YCbCr to RGB Conversion," IEEE Transactions on Consumer Electronics, Vol. 53, No. 4, Page 1490-1493. Nov. 2007.

[5] G. Bradski and A. Kaebler, "Leaning OpenCV".