

Research Article

A Novel Music Player Controlling Design Based on Gesture Recognition

Yi Liu, Chenglei Dong, Min Zhang, Xinxin Chen and Yanjun Li

School of Information and Electrical Engineering, Zhejiang University City College, Hangzhou, China

Abstract: This study has proposed a novel music player controlling method based on gesture recognition, which translating the gesture interacting signal into a controlling instruction for the music player. Firstly, utilizing the laptop's webcam to capture the image information, and then employing the image processing method to tackle it. The skin color detection was used to obtain the information of gesture candidates, and a background subtraction was introduced to eliminate the distributing information. Moreover, in order to ensure the rapid and effective implementation of music player proposed here, the barycenter of a gesture was calculated as one Recognized Reference Information (RRI); a ratio between the gesture's width and height was also selected as the other RRI; the comparison of these two RRI values was utilized to obtain a pattern signal of the gesture which was corresponding to the controlling instruction for the Music Player. Eventually, a Music Player was programmed and the pattern signal generated by gesture recognition was used to control the Music Player as willing to realize four basic functions: "play", "pause", "previous" and "next". A series of tests using our gesture recognition based Music Player was conducted under the condition with different kinds of complex backgrounds, and the results showed the satisfactory performance of our interactive designing.

Keywords: Background subtraction, gesture recognition, music player, skin color detection, recognized reference information

INTRODUCTION

During the past decades, Human-Computer Interaction Techniques represented by multi-touch technology have aroused many attentions of scholars in the world, and have been ceaselessly popular in our daily life. However, the current Contact Interactions based products are no longer satisfied by customers. They prefer to have behavior recognition system that possessing beyond performance on handling comfort and convenience. Recently, a gesture recognition as a novel Human-Computer Interaction Technique-has been widely studied for its incomparable advantage: diversity, ambiguity, and difference in time and space. And it is also an interactive technology related to other areas such as image processing, pattern recognition and computer vision, etc.

At present, many meaningful research results relevant to gesture recognition have been proposed. Tan *et al.* (2008) proposed a gesture based control of mobile robots. A consumer electronics control system based on hand gesture moment invariants was put forward by Premaratne and Nguyen (2007); it satisfies the demand of real-time responses and is highly accurate in recognition towards various gestures. Zhang *et al.* (2011) represented a simple but efficient intelligent wheelchair control system based on gesture

information. Yang *et al.* (2012) proposed a hand posture recognition method based on average neighborhood margin maximization (Yang *et al.*, 2012), and this method recognition rate is above 90%.

In this study, a novel music player in view of gesture information recognition is proposed. To begin with, utilizing the webcam of laptop to capture the image signal, and then using the skin color detection to obtain the gesture image. Moreover, according to the morphology and geometry features of gestures, the barycenter of a gesture is calculated; at the same time, the ratio between the gesture's width and height is also computed; then, the comparing of these computation results are employed so as to acquire the gesture interacting signal. Eventually, the signals are programmed to control the Music Player realizing basic operations without any touching actions, such as play music, pause music, play previous song and play next song.

GESTURE INTERACTING BASED MUSIC PLAYER CONTROLLING DESIGN

The designing of a gesture interaction based Music Player system primarily consists of two parts: generating the gesture interaction signal, and controlling the Music

Corresponding Author: Yanjun Li, Key Laboratory of Intelligent System, Zhejiang University City College, Hangzhou, 310015, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

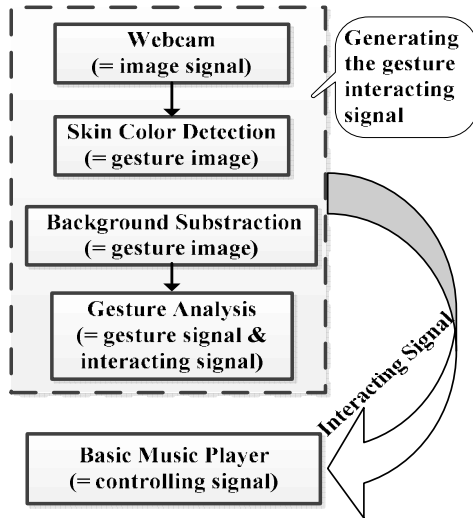


Fig. 1: The simple block diagram of music player based on gesture recognition

Player. The first part of generating the gesture interacting signal is mainly about dealing with the image information originated from webcam, by means of the skin color detection with the morphology and geometry analyses; and during this process, a background subtraction is introduced to eliminate the distributing information to get the gesture information. Furthermore, the other part principally refers to using the interacting signal to perform the Music Player realizing the four basic functions mentioned above. The simple block diagram of the gesture recognition based Music Player controlling design is shown in Fig. 1.

COLOR SPACE CONVERSION AND SKIN COLOR DETECTION

The color space of image information captured by the webcam of laptop is acquiescently RGB of computer system. However, in the RGB color space, the skin color pixels are readily influenced by the change of luminance resulting in the alteration of three components of RGB color space. This phenomenon will lead skin color candidates to deviating the main gathering area of the skin color and makes the skin based image recognition very difficult. Therefore, a color space conversion method (Wang *et al.*, 2012) is utilized here to tackle this problem. The following formula is used to transform an image components from the RGB color space into the *YCbCr* color space:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ -0.1687 & -0.3313 & 0.5000 \\ 0.5000 & -0.4817 & -0.0813 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

where, *Y* means the luminance in color space, *Cr* and *Cb* indicate the chromaticity in color space,

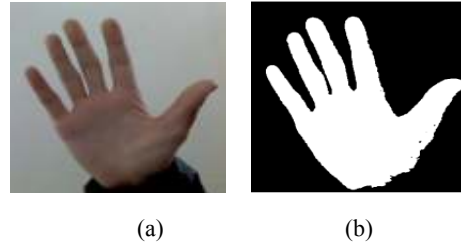


Fig. 2: A gesture image with simple background and processing result by the skin color detection

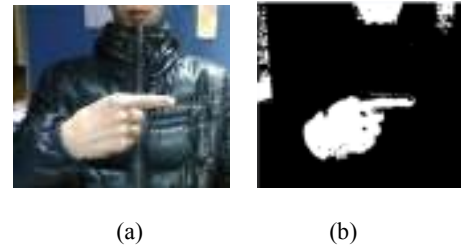


Fig. 3: A gesture image with complex background and processing result by the skin color detection

respectively. It is studied that the *Cb* and *Cr* components of *YCbCr* color space can effectively assemble the skin color in a closed region. Hereupon, this method is employed in our design to obtain the skin color candidates. Commonly, the thresholds of *Cr* and *Cb* can be used to divide the image information into skin color region and non skin color region; and a image pixel can be seen as the skin color candidate, when it lies in the region (Wang *et al.*, 2012): $77 \leq Cb \leq 127, 133 \leq Cr \leq 173$.

For simple background gesture image as shown in Fig. 2a, a gesture could be extracted completely by skin color detection, as shown in Fig. 2b. But for an image with more complicated environment, the extraction task is harder and will be dealt with and illustrated below in details.

Gesture acquisition: In order to generate gesture interacting signals to control Music Player, an effective method of gesture acquisition is indispensable. Commonly, a gesture image with complex background information are often presented with clothes, walls etc., such as displayed in Fig. 3a. And its processing result by skin color detection is showed in Fig. 3b.

As shown in Fig. 3, some distracting information such as curtain, face and sticky note are also selected by skin color detection as the gesture candidates because their color characteristics are similar to the gestures. Moreover, due to the influence of illumination intensity, the data in the binary image of the gesture is not complete, and this will disturb the calculation procedure.

Aiming at removing the distracting information assumed in Fig. 3, the method based on background subtraction is introduced.



Fig. 4: The background image and processing result by the skin color detection



Fig. 5: The result of Fig. 3b subtracts the common part of Fig. 3b and Fig. 4b

First of all, considering that the region of disturbing information will mainly keep stable as long as the camera's position do not change dramatically in a short interval. Therefore, for the convenience of computer calculation for the binary values, we firstly define the Binary Images of gesture with distracting information as BIG_{di}, and define the Binary Images of distracting information (background image) as BI_{di}; then let Brigade subtract the common part of Brigade and Bide; thus the binary image only with gesture are obtained.

For instance, the Fig. 4a is regarded as the background of Fig. 3a, which is captured by a webcam before the appearance of gesture in Fig. 3. And its processing result by skin color detection is presented in Fig. 4b.

The Fig. 5 is the result of binary processing of Fig. 3b subtracting the common part of Fig. 3b and 4b. Subsequently, the minimum rectangular area covered the whole gesture is obtained after eliminating the white area less than certain setting threshold, which is provided for the later gesture analysis to get the controlling signal.

GENERATION OF GESTURE INTERACTING SIGNAL

In the area of engineering applications, the desired systems should be timely and efficient so as to meet the user's practical demand. Our aim in this study is to realizing that employing the gesture information to instruct the Music Player performing four basic functions. As presented in Fig. 6a to d represent gestures of achieving the four basic functions, respectively.

- In order to ensure the rapid and effective implementation of Music Player proposed here, the barycenter of a gesture image represented b

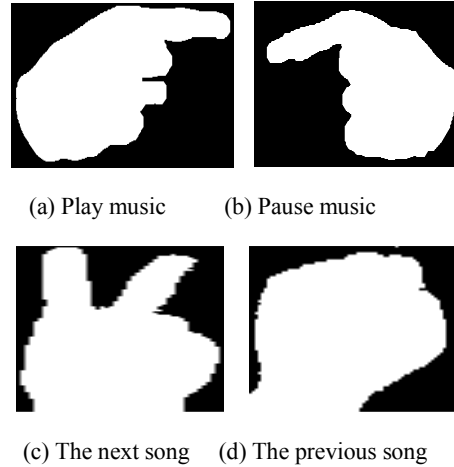


Fig. 6: Gestures of achieving the four basic functions of music player

(x_c, y_c) is calculated as one Recognized Reference Information (RRI); a ratio α between the gesture's width G_w and height G_h is also selected as the other RRI; the comparison of these two RRI values is utilized to obtain a pattern signal of the gesture which is corresponding to the controlling instruction for the Music Player. The equations of ratio α and the barycenter of a gesture (x_c, y_c) are represented in the Eq. (2) and (3), respectively.

$$\alpha = \frac{G_w}{G_h} \quad (2)$$

$$\begin{cases} x_c = \sum_{i=1}^n \sum_{j=1}^m i \cdot \delta(i, j) / \sum_{i=1}^n \sum_{j=1}^m \delta(i, j) \\ y_c = \sum_{i=1}^n \sum_{j=1}^m j \cdot \delta(i, j) / \sum_{i=1}^n \sum_{j=1}^m \delta(i, j) \end{cases} \quad (3)$$

where, n and m indicate the height and width of the minimum rectangular region covering the gesturing, respectively. $\delta(i, j)$ denotes the value of the gesture's binary image at point (i, j). Based on Eq. (3), the state variables of gesture signals are defined by g_{cx} and g_{cy} and are determined as follows:

$$g_{cx} = \begin{cases} \{Up\} & x_c < m/2 \\ \{Down\} & x_c > m/2 \end{cases} \quad (4)$$

$$g_{cy} = \begin{cases} \{Left\} & y_c < n/2 \\ \{Right\} & y_c > n/2 \end{cases} \quad (5)$$

Obviously, there are four states which are respectively representing the directions of a gesture inclining to left, right, up or down.



Fig. 7: The interface of music player based on gesture interacting

Here, a simple but efficient decision based on gesture information recognition to generate the Interacting Signal (IS) is performed, with regard to the relationship of two RRI values mentioned above. And the preferences of α could be redesigned at new situations:

- **Play music:** If $\alpha \in [0.25, 0.50]$ and $g_{cx} = \{down\}$, then generating IS = "Play": I.S.1.
- **Pause music:** If $\alpha \in [0.6, 1.0]$ and $g_{cx} = \{up\}$, then generating IS = "Pause": I.S.2.
- **Play next song:** If $\alpha \in [2.0, 3.0]$ and $g_{cy} = \{right\}$, then generating IS = "Next": I.S.3.
- **Play previous song:** If $\alpha \in [2.0, 3.0]$ and $g_{cy} = \{left\}$, then generating IS = "Previous": I.S.4.

MUSIC PLAYER BASED ON GESTURE INTERACTION

The primary feature of Music Player based on gesture interaction proposed in this study is that we can use the interacting signals originated from gesture recognition as willing to instruct the Music Player to complete its basic functions rather than utilizing traditional mouse and touching. In order to achieve this aim, a novel Music Player is required to develop based on MFC and OpenCV programming as shown in Fig. 7. And, in this Music Player, there are five function buttons: button of the "Gesture Recognition", button of "Play Previous" song, button of "Play/Pause" song,

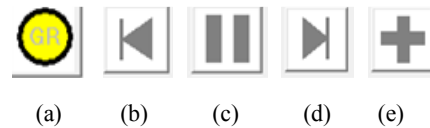


Fig. 8: The function button of music player based on gesture interaction

button of "Play Next" song and button of "Adding" song, respectively, as displayed in the Fig. 8a to e.

The mainly procedure to operate our Music Player is as follows. If the button of gesture recognition shown in Fig. 8a is pressed down, the webcam will be open and the task of gesture recognition including skin color detection, gesture acquisition and recognition will be automatically executed continuously until webcam is close. During the process of webcam working, gesture interacting signal I.S. is always detected. For instance, if interacting signal I.S.1 is produced, play music will be performed and corresponding button icon of "Play/Pause" will become yellow for a while. Certainly, it is likely to generate an unexpected interacting signal when withdrawing our gesture posed in webcam. Considering this circumstance, a feasible strategy is proposed in this study to overcome it, i.e., making the interacting signal invalid for a while after one gesture interacting is just completed.

CONCLUSION

This study presented a novel Music Player based on gesture recognition and provided a simple but efficient method to identify the gesture information.

The designing primarily consisted of two parts: generating the gesture interaction signal, and programming a Music Player controlled by the gesture signal. To generate the gesture interacting signals, some methods were adopted such as the skin color detection, as well as the gesture recognition upon the relationship between the ratio α and the barycenter position of gesture. Eventually, interacting signal is generated to guide our Music Player to implement four basic functions. Our gesture recognition based Music Player was tested under the condition with different kinds of complex backgrounds, and the results showed the satisfactory performance of our interactive designing. Furthermore, the influences resulting from the change of webcam's position and human being's posture will be thoroughly considered in our future work, to improve the adaptability and flexibility of our recognition method. Besides, a more effective skin color detection method will be deeply studied to overcome the interference of illumination changing.

REFERENCES

- Premaratne, P. and Q. Nguyen, 2007. Consumer electronics control system based on hand gesture moment invariants. *IET Comput. Vis.*, 1(1): 35-41.
- Tan, W.C., S.K. Sim, G.L.G. Seet and K.W. Ong, 2008. Gesture based control of mobile robots. *Proceedings of the IEEE Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA, 2008)*, pp: 20-25.
- Wang D., H. Shen and H.T. Lou, 2012. Study of a skin detection algorithm based on the H-CrCb color space. *Comput. Sci.*, 10(39): 223-225.
- Yang, Y., N.N. Cheng and Z. Meng-Meng, 2012. Research of hand posture recognition method based on average neighborhood margin maximization. *Comput. Eng.*, 9(38): 19-21.
- Zhang, Y., J. Zhang and Y. Luo, 2011. Intelligent wheelchair control system based on hand tracking. *J. Chongqing Univ., Posts Telecommun.*, 6(23): 741-745.