ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Automatic User State Recognition For Hand Gesture Based Low-Cost Television Control System

M.Sudha Rani
M.Tech Student,
Malla Reddy Institute of Technology,
Hyderabad,TS India.

K.Santosh Kumar
Assistant Professor,
Dept of ECE,
Malla Reddy Institute of Technology,
Hyderabad,TS India.

ABSTRACT:

Hand gesture based television (TV) control is attracting more and more researchers. Most of existing works focus on the hand gesture recognition algorithm and the corresponding user interface, while the power consumption or computational cost is not considered carefully. In practice, keeping the camera device and the gesture recognition module running all the time often costs much energy. Till now, few methods have been reported to solve this problem. This paper proposes an automatic user state recognition scheme to recognize the TV user's state and activate the camera-based gesture recognition module only when the user is trying to control the TV. Specifically, the user's behavior active or not is detected by lowcost sensors, the user's gaze watching TV or not is tracked by the face-based view detection, and the user's state is then decided according to a finite-state machine composed of four states:Absent, OtherAction, Controlling, and Watching. The prototypes based on an ultrasonic distance sensor array, a red-green-blue (RGB) camera, and a depth camera are implemented and tested. The results show that the proposed scheme can effectively reduce the power consumption or computational cost of the original hand gesture based control schemes.

Index Terms:TV control, user state recognition, finite state machine, hand gesture.

INTRODUCTION:

Television (TV) is widely used all around the world. Till now, the TV display screen has been innovated for several generations while the TV controller keeps nearly unchanged during a long period.

Recently, with the enrichment of TV programs, more and more frequent controlling operations such as channel switching and program searching are required, which makes the traditional point-click-style hand-held TV controller inconvenient. Taking the channel switching for example, it requires the user to lower his/her head to first see and then press the small buttons. Next, he/she needs to look up at the TV screen to see whether the program is the expected one. Otherwise, he has to repeat this process to switch the channel. According to Fitts' Law [1], the interaction system's execution time depends on the easiness for the user to press the button. The frequent looking-up-and-down reduces the control speed and affects user experiences to a large extent.Recently, some new controlling apparatuses have appeared, e.g., the wearable or hand-held space controller. These kinds of controllers activate command signals by gesture recognitionbased on space sensors, e.g., the controller based on a space mouse [2], the one based on a data glove and a Heads up Display (HUD) [3], the wearable one like a wristwatch [4], the one based on a touch pad [5], the one with a low-cost data glove [6], and the ring-type controller [7]. Although these controllers improve user experiences in some extent, it requires the user to hold or wear a device which may be difficult to found sometimes in a big room. Additionally, some hand-free controlling means have also been studied, e.g., voice control [8][9]. With the speech recognition module, the TV can interpret the user's voice command and then response accordingly. However, there are still difficulties in voice control. Firstly, the speech recognition algorithm is not mature enough due to large number of vocabularies and various dialects. Secondly, the interference including the environmental noise and the voice coming from the TV's speaker will affect voice recognition to a large extent and are not easy to eliminate [8].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Additionally, the recognition module is always kept on in order to detect when and what kind of voice command is activated, which leads to much computational cost and power consumption.

Existing System:

Another kind of hand-free controlling is the hand gesture based TV control [10][11]. The detected and recognized hand gestures are used as the command signals for TV controlling. Some user interfaces, e.g., icon-based interface [10] or motion-based interface [11], are adjusted accordingly in order to support natural hand control. Free-hand control makes the controlling much easier for the user to master and use, and some new sensors capable of depth detection have been developed for more accurate hand gesture recognition [12][13], which makes it a potential future for TV control. However, similar with the speech recognition module in voice control, the hand gesture recognition module is kept running all the time in order to detect the activation gesture for launching the gesture control. Thus, the means to reduce the computational cost or power consumption should be considered carefully. Generally, with respect to the energy consumption, two aspects are considered: the sensor device, e.g., a red-green-blue (RGB) camera or a depth camera, and the gesture recognition module. Intuitively, it is not necessary to make these sensors always run especially when the user is not controlling the TV. Additionally, it is not reasonable for the user to turn on/off the gesture recognition module frequently and manually during TV watching.

Proposed System:

This paper proposes an automatic user state recognition scheme to recognize the TV user's state efficiently before activating or sleeping the camera device and hand gesture recognition module. The considered user states, including Absent, OtherAction, Controlling, and Watching, are initialized and updated according to the results of action detection and presentation detection based on low-cost sensors. Thus, the camera device and hand gesture recognition module are activated adaptively to reduce the system's computational cost and power consumption.

RELATED WORK:

There exist some works on hand gesture based TV control. With respect to the gesture's features, there are two types, i.e., static gesture recognition and dynamic gesture recognition.

With respect to the recognition sensors, there are two types, i.e., the recognition based on RGB camera and the one based on depth camera. Static hand gestures are identified by the shape of hand, including position, orientation and number of fingers. A straightforward static gesture control method is hand-based mouse that changes hand gestures into mouse manipulations [10]. Specifically, the palm is detected and tracked for implementing the mouse moving, the change from palm to fist is detected for mouse clicking, and the fist is detected and tracked for mouse dragging. Some other controlling methods use a specific number of fingers and their angles [14], a predefined sequence of gesture state transitions [15], or the index finger and thumb [16][17]. Generally, during static gesture controlling, hand poses have to keep static for a certain period of time. Although the correct recognition rate can be increased, the controlling action is not natural enough. Differently, the dynamic hand gesture methods, which recognizing hand motion patterns as command signals, seem more natural for TV control [18]. For example, the gesture drawing based on motion tracking is used for command signals [11]. In this scheme, hand gestures are recognized by capturing the motion path when the user draws different symbols in the air. These gestures are used to interact with the TV. For dynamic hand gesture recognition, it is important to define a suitable set of motion patterns easy to remember and recognize. Currently, two kinds of devices are popular: the RGB camera and the depth camera.

Especially, the latter provides three-dimension information of hand or finger position, and could thus improve the hand motion recognition. With a RGB camera, the two-dimension motion information of hand is decided [18][19]. Generally, due to the camera's high resolution, the color information can be used to accelerate the detection of hand regions. This kind of camera is now inexpensive and acceptable by TV manufactures or consumers. However, the camera is often sensitive to light conditions in room, which degrades hand recognition. The depth camera is now attracting more and more researchers and engineers, although the available resolution of the depth map is still low [12][20][21]. Now, three kinds of depth camera are often used: the camera array for stereo vision, the camera based on time of flight (TOF), and the one based on structured light. For example, various motions are recognized from the spatiotemporal trajectory features composed of horizontal, vertical, temporal, and depth information captured by a TOF camera [12].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

In another scheme, the TOF camera is combined with RGB camera toimprove the hand gesture recognition [20]. The depth camera based on TOF or structured light is often robust against light changes in room, and is thus suitable for various environments. As can be seen, with the improvement of sensing devices and recognition algorithms, the free-hand TV control scheme can provide higher recognition rate and more natural user experiences. Nevertheless, the system's energy consumption is a concern and still not considered carefully, e.g., the computational cost on continuous gesture recognition, or power consumption on continuous device running. Some novel methods are expected to activate or sleep the devices and recognition modules adaptively according to the user state. This paper will propose a solution for this problem.

THE PROPOSED HAND GESTURE BASED LOW-COST TV CONTROL SYSTEM:

The proposed TV control system, targeted for low-cost hand gesture control in an automatic manner, composed of both hardware and embedded software components. Fig. 1 shows the hardware components, including the TV, set-top box, and the other sensors connected to the TV system, including the RGB or depth camera module and ultrasonic distance array module. The embedded software components which connect the sensors with the TV, process the sensor data, and drive the TV control process will be presented as follows.

Fig. 1. The proposed TV control system's hardware components include the TV, set-top box, and some sensors, including the RGB or depth camera module and ultrasonic distance array module.

Overview of the Proposed Control System:

Compared with traditional TV control systems based on hand gestures, the proposed system introduces a new scheme, named Automatic User State Recognition (AUSR). This scheme consists of several steps, e.g., the user state initialization or updating, the device activation or sleeping, the recognition module activation or sleeping, and the user input notification, as shown in Fig. 2. Based on the ultrasonic distance array and camera module, the user's action is detected and user state is initialized or updated automatically. The user state tells whether the user is controlling the TV or not, and will be defined in the following section. If the user is controlling the TV, the camera device and gesture recognition module will be turned on, and the user will be informed to begin hand gestures by e.g., displaying a message on the screen. Otherwise, the camera device or gesture recognition module is slept in order to save energy. When informed to begin, the user makes control gestures, which will be recognized by the system and then drive the TV. In the following content, the AUSR, including user state definition, initialization and transition based on automatic action detection and presentation detection, will be introduced in detail, and the performance evaluation on the implemented hand gesture based TV control systems will be presented.

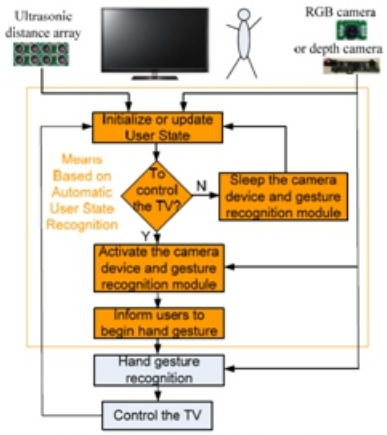

Fig. 2. The proposed TV control system's software components include user state initialization or updating, camera device and hand gesture recognition module activation or sleeping, user input informing, etc. The user's action is detected in order to decide the intention to control TV.

User State Definition:

According to user's actions in front of TV, the user state is classified into four types, i.e., Absent, OtherAction, Controlling, and Watching. Absent means no user is sitting in front of the TV although it may be open. OtherAction means that some user is sitting in front of the TV and doing something, but he is not watching or controlling the TV. Controlling means that the user is controlling the TV, e.g., changing the channel or adjusting the volume. Watching means the user is watching TV. Based on Finite-State Machine (FSM), these states are denoted by (A=0,B=0), (A=1,B=0), (A=1,B=1), and (A=0,B=1), respectively. Here, if the user's intentional motion in front of the TV is detected, then A=1, otherwise A=0.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

If the user is looking at the TV screen, then B=1, otherwise B=0.The user state will be changed from Absent to OtherAction if the user comes in front of the TV, from OtherAction to Controlling if the user makes hand gestures to control the TV, from Controlling to Watching if the user keeps silent and stays to watch TV, from Watching to Controlling if the user makes hand gestures to control the TV, from Watching or Controlling to OtherAction if the user does something else in front of TV, and from OtherAction to Absent if the user leaves the TV region. In the proposed scheme, these conditions for state transition are detected automatically based on the ultrasonic sensor array and the RGB or depth camera.

User State Transition:

In the proposed scheme, the user state is updated automatically according to the following steps.Firstly, the action detection is done to decide whether the user is moving intentionally. If no movement is detected, the inactive time will be counted and compared with the threshold Tm. If the inactive time is bigger than Tm, then the user state will be changed, otherwise not. Here, Tm is used to decide the user state transition when the user does not move during this period. For example, the user state may be changed from Controlling to Watching, if the user watches TV without movements during the Tm period after several controlling actions. If the intentional movement is detected, the presentation detection will be followed with respect to the original user state, and the user state will be updated accordingly. Specially, the user state may be forced to be initialized in order to avoid deadlocks caused by unpredicted environment changes.

USER STATE TRANSITION

The Algorithm of User State Transition

```
If (the intentional motion is detected)
    If (A=1 and B=1)
      Detect hand gestures
    Else if (A=0 and B=0)
      Set (A=1,B=0)
    Else if (A=1 and B=0)
      if (the face watching TV is detected)
        Set (A=1,B=1)
    Else
      if (the face watching TV is detected)
        Set (A=1,B=1)
Else if (the intentional motion is not detected for a long time)
    If (A=1,B=0)
      Set (A=0,B=0)
    Else if(A=1,B=1)
      Set (A=0,B=1)
```

EXPERIMENTAL RESULTS AND DISCUSSIONS:

The TV control prototypes are composed of a LCD TV (Liquid-crystal-display televisions), an ultrasonic sensor array, a RGB or depth camera, and the low-cost computing module embedded in a Personal Computer (PC).

In detail, the TV screen is 42-inch (0.93-meter height, and 0.52- meter width) with the optimal controlling distance ranging from 1.0 meters to 3.5 meters, and thus the face library is composed of faces with the angle (α=25°, β=15°, Φ =30 °). The ultrasonic sensor array is composed of four distance sensors, each of which has the sensing angle of θ=15° and maximal sensing distance D=4.5m. At the distance of 3.5 meters, the sensing square's size S=2.7m. The RGB camera outputs the picture with resolution 1024x768, and depth camera with resolution 640x480. In experiments, the low cost means based on various hand gesture recognition algorithms will be evaluated and discussed, including the study on user behavior, the user state detection accuracy, and the energy consumption with respect to different gesture algorithms.

CONCLUSIONS:

In this paper, an automatic user state recognition scheme is proposed to reduce free-hand TV control's computational cost and power consumption. In this scheme, the low-cost ultrasonic distance sensor array is constructed for detecting the user's intentional actions, the face detection method based on restricted face library is proposed to decide whether the user ispresent and actively watching TV, and the TV user's states are defined according to the finite-state machine (FSM). Based on the results of action detection and presentation detection, the user state is initialized and updated automatically.

The camera device and gesture recognition module is activated or closed with respect to the recognized user state. For example, only when the user state is Controlling, the hand gesture recognition module is activated, otherwise not. The implemented prototype and experimental results show that the proposed scheme reduces existing gesture control systems' computational cost and power consumption greatly. This will improve the practical usability of the hand gesture based TV control. In the future, this scheme can be extended to some other touchless TV control systems, such as the voice-based control system and so on.

REFERENCES:

[1] I. S. Mackenzie and W. Buxton, "Extending fitts' law to twodimensionaltasks," in Proceedings of CHI'92, pp. 219-226, May 1992.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[2] S. K. Kim, G. H. Park, S. H. Yim, S. M. Choi, and S. J. Choi, "Gesture recognizing hand-held interface with vibrotactile feedback for 3D interaction," IEEE Trans. Consumer Electronics, vol. 55, no. 3, pp. 1169-1177, Aug. 2009.

[3] H. Heo, E. C. Lee, K. R. Park, C. J. Kim, and M. C. Whang, "A realistic game system using multi-modal user interfaces," IEEE Trans. Consumer Electronics, vol. 56, no. 3, pp. 1364-1372, Aug. 2010.

[4] D. W. Lee, J. M. Lim, S. W. John, I. Y. Cho, and C. H. Lee, "Actual remote control: a universal remote control using hand motions on a virtual menu," IEEE Trans. Consumer Electronics, vol. 55, no. 3, pp. 1439-1446, Aug. 2009.

[5] R. Aoki, M. Ihara, A. Maeda, M. Kobayashi, and S. Kagami, "Expanding kinds of gestures for hierarchical menu selection by unicursal gesture interface," IEEE Trans. Consumer Electronics, vol. 57, no. 2, pp. 731-737, May 2011.

[6] Y. M. Han, "A low-cost visual motion data glove as an input device to interpret human hand gestures," IEEE Trans. Consumer Electronics, vol. 56, no. 2, pp. 501-509, May 2010.

[7] L. C. Miranda, H. H. Hornung, M. C. .Baranauskas, "Adjustable interactive rings for iDTV," IEEE Trans. on Consumer Electronics, Vol.56, No. 3, pp. 1988-1996, August 2010.

[8] J.-S. Park, G.-J. Jang, J.-H. Kim, S.-H. Kim, "Acoustic interference cancellation for a voice-driven interface in smart TVs," IEEE Trans. On Consumer Electronics, Vol. 59, No. 1, pp. 244-249, February 2013.

[9] I. Papp, Z. Saric, N. Teslic, "Hands-free voice communication with TV," IEEE Trans. on Consumer Electronics, Vol. 57, No. 1, pp. 606-614, February 2011.

[10] W. T. Freeman and C. D. Weissman, "Television control by handgestures," in Proceeding of IEEE International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, pp. 179-183, June 1995.

[11] S. Jeong, J. Jin, T. Song, K. Kwon, and J. W. Jeon, "Single-Camera Dedicated Television Control System using Gesture Drawing," IEEE Trans. on Consumer Electronics, Vol. 58, No. 4, pp. 1129-1137, November 2012.

[12] M. Takahashi, M. Fujii, M. Naemura, and S. Satoh, "Human gesturerecognition using 3.5-dimensional trajectory features for hands-free user interface," in Proceeding of ARTEMIS'10, pp. 3-8, Oct. 2010.

[13] X. Liu, and K. Fujimura, "Hand gsture recognition using depth data," inProceedings of 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 529-534, May 2004.

[14] D. Ionescu, B. Ionescu, C. Gadea, and S. Islam, "An intelligent gesture interface for controlling TV sets and set-top boxes," in Proceeding of 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 159-164, May 2011.

[15] D. Lee, and Y. Park, "Vision-based remote control system by motion detection and open finger counting," IEEE Trans. Consumer Electronics, vol. 55, no. 4, pp. 2308-2313, Nov. 2009.

[16] S. Lenman, L. Bretzner, and B. Thuresson, "Using marking menus to develop command sets for computer vision based hand gesture interfaces," in Proceeding of NordiCHI'02, pp. 239-242, Oct. 2002.

[17] P. Premaratne, and Q. Nguyen, "Conumer electronics control system based on hand gesture moment invariants," IET Computer Vision, vol. 1,no. 1, pp. 35-41, March 2007.

[18] N. Henze, A. Locken, S. Boll, T. Hesselmann, and M. Pielot, "Free-hand gestures for music playback: deriving gestures with a user-centered process," Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia, no. 16, Dec. 2010.

[19] M. Chen, L. Mummert, P. Pillai, A. Hauptmann, and R. Sukthankar, "Controlling your TV with gestures," In Proceeding of International Conference on Multimedia Information Retrieval (MIR), pp. 405-408, March 2010.