VISION BASED REMOTE CONTROL SYSTEM BY USING HAND GESTURE AND VOICE RECOGNITION

Sanoofar. A.N
Department of Electrical and Electronics Engineering,NIET
sanoofar34@gmail.com

Abstract

In the present situation voice recognition and hand gesture are couple of emerging trend used to many applications such as control based works like systematic control, voice dialing, simple data entry ,television control and so many of controlling applications .But most of the existing works accompanied with lot of real time problems first of all the power consumption is not considerable, and among them the voice recognition is vastly altered by dialects ,accent ,physical state of user ,surrounding atmosphere and so many of this like problems ,but also in case of security based applications it is easily hacked by imitation and recording voice of sound. In this proposed model the automatic user state recognition is introduced as a supporting trend for hybridized voice recognition and hand gesture system to manage home system such as television, air conditioner, refrigerator, fan, light and so on. To achieving power management , less power consumption ultra sonic sensor as the operating element of automatic user state recognition, and the voice recognition is functioning by speech attention IC HM2007and microcontroller PIC. So that in this scheme the existed problems and power consumption can be effectively reduced Key words: AUSR, ultra sonic sensor, hand gesture, voice recognition, PIC, HM2007.

I.INTRODUCTION

The past few years, the home system that is the home appliances can be entirely changed in structure, size, operation features and everything. But the controlling method of they is not considerably changed, till now the majority of existing devices can be managed by the press type and remote control methods. This paper proposes an attractive solution for this problem.

When compared to the now existing and past controlling systems the hand gesture and voice recognition are future .But for considering a handicapped person the hand gesture is unimportant and in the situation of new users it is difficult to operate and utilize suddenly ,but the voice recognition can be utilize any user instantaneously ,there is no problem of know more about the controlling .but in the condition of noise surrounding or dump peoples it is also not have any importance .so that the voice recognition is possible where as the hand gesture is not possible and the hand gesture is possible where as the voice recognition is not possible. Hence the mutual interacting operation of hand gesture and voice recognition by hybridizing both can be acceptable for any situation. The power consumption of this system can be controlled by imposing automatic user state recognition (AUSR) with them.

In this integrated scheme, the AUSR effectively operated by comparatively low cost ultra sonic sensor and a depth camera. The ultra sonic sensor detects weather any user is there, if it is, then turn on the camera, it analyze movements of user other action, watching or controlling. If the user needs to controlling, recognize which way is selected to control, through hand or voice. If voice is detected then the voice recognition mode is turn on. The speech recognition IC HM2007 in voice recognition check the input voice is valid or not and process the voice by passes through a next microcontroller PIC.

If select the control way as hand then hand gesture mode is turn on and capturing the actions of hand. Through this, the proposed scheme effectively distorts computational cost as well as power consumption.

The remaining of the paper is arranged as follows. The related works can described as the next section 2, the section 3 shows implementation and hypothetical contents of the proposed model the sections 4 and 5 respectively indicates the experimental results and conclusion.

II. RELATED WORK

Based on hand gesture or voice recognition there some works exist. According to user interaction there are three variations of voice recognition user dependant, user independent and user adaptive and with respect to input type ,isolated word recognition (IWR), continues word recognition (CWR) and continues speech recognition.

The user dependant systems can access only by specific predefined users, the user must be introduced once to the system by a training period before initiate the control, This system having the benefits of easy to develop, cheaper to buy and accurate. If any operator can access the system except any training period is user independent voice recognition. Regardless of user dependant system it is user flexible, but less accuracy, complexity to develop and high expense to buy affects the possibilities of this model. The third variation speaker adaptive is the most emerging one it is a combined structure of user dependant and user independent models. Differently the introduction of user is done by a short training period at run time and it can be accessed by any type of users.

According to hand recognition feature there are two, static hand and dynamic hand recognition and also recognition sensor base, recognition based on RGB camera and recognition based on depth camera. The static hand gesture taking input as the structure of the hand, including number of fingers, position and orientation. In this model it is necessary to keeping the hand some time statically for recognition happens, this makes this model not quite natural for control. Differently the other one dynamic hand gesture captures the path of hand movement as input, in this type provides the facility of control in a natural way. Presently RGB and depth camera are the two devices mostly used. The RGB camera recognizes the hand two dimensionally, now it is in affordable cost but light sensitiveness is the main problem facing this device. The depth camera is suitable for any environmental condition so that many researches are going through this device. In this proposed system, continues speech recognition and dynamic hand gesture are the methods adapted and along with them depth camera is used for user state recognition.

III. THE PROPOSED TOUCHLESS BASEDCONTROL SYSTEM

This proposed model aims that to control different home appliances in a touch less way by hand gesture and voice recognition. The figure 1 shows the hardware components of the scheme, consisting ultra sonic sensor, depth, camera voice recognition element and the air conditioner is taken as an example home appliance. For connecting the sensor to the device to be control and for processing the sensor data embedded software components can be used.

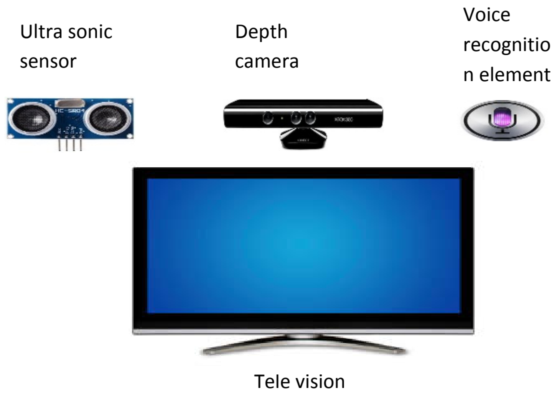

Figure 1: The proposed device control system including hardware components.

A. Overview of the Proposed System

This scheme hybridizes the hand gesture and voice recognition for make the system suitable for control the device at any situation, to minimize the computational cost as well as power consumption AUSR is associated. During the operation it involves several steps, first of all user state recognition consist of user state definition, initialization and transition The next step is controlling through either voice recognition or hand gesture or through both. The purpose of ultra sonic sensor is to detect the physical presence of user and if the user is there, it turns on the depth camera, the depth camera observe the movements of the user and update it.

While the user is trying to control the device, camera checks whether it is through hand or voice. If it is through hand then hand gesture mode is turn on else if it is through voice then voice recognition mode is turn on in order to control the device. Suppose there is no any user present or if the user does not trying to control the device then both the hand gesture and voice recognition are in sleeping mode.

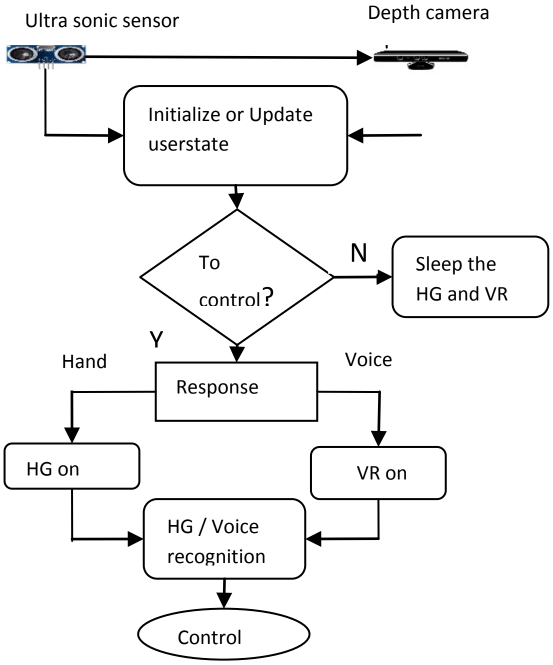

Figure2: Architecture of the proposed system.

B. User State Definition

It is the detection of action of user in the region where the device to be control exist, weather it is absent or other action or watching or controlling. If there is no any user exist in region then it is the state of absent, other action means that the user exist but not considering the device, do some other work. The state of controlling is the user trying to control the device either through hand or voice, suppose the selected way is hand then hand gesture is active otherwise if it is through voice then voice recognition module is turn on.

The letter X denotes that the motion of user is there or not, if X=1 then there is a motion of user in the device region otherwise X=0. Similarly the letter Y=1 shows user watching the device, if it is not then Y=0.

The figure 3 draws that the above mentioned states and transition between the states.

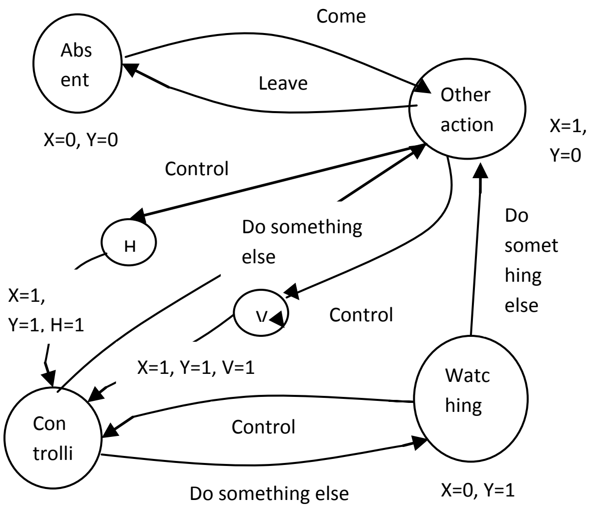

Figure 3: Various states and transition.

Initially the user state is absent X=0, Y=0, if any user comes to the region the user state changes from absent to other action X=1, Y=0.The user state changes from other action to controlling, if the existed user watch and try to control the device by hand X=1, Y=1, H=1 or voice X=1, Y=1, V=1. After that, suppose the user intentionally keeping silence and watch the device then the state turns to watching. Differently, if the user not needs to control anymore and do some other work then the user state transfer controlling to other action.

C. Automatic User State Recognition

The identification of weather the user is there in the device region, if the user is in there then found what is the present state of him/her it is the automatic user state recognition (AUSR). The presence of user is captured by ultra sonic sensor, what is it may be A=0 or A=1.

Suppose the action is detected (A=1) the sensor turn on the camera device, it will detect what is the present state of user may be B=0 or B=1. By the information's from ultra sonic sensor and camera (from A and B) decide the available condition of user, that means find out the present state of user.

The figure 4 highlights the action detection and presentation detection indetail next.

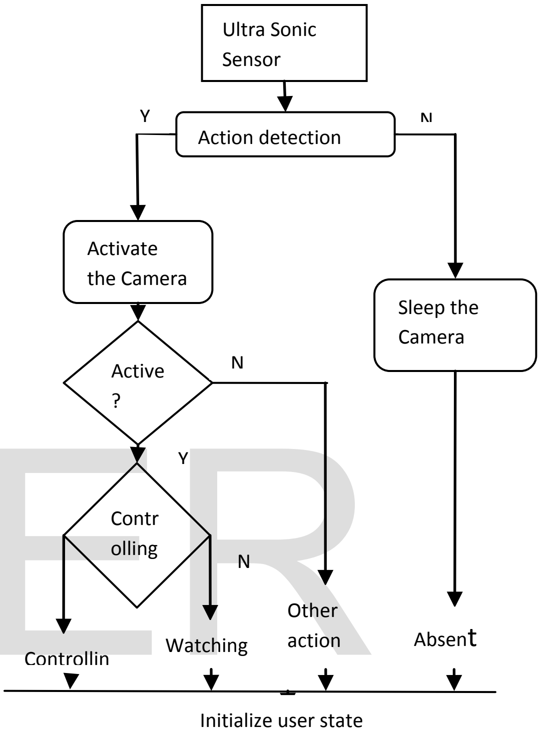

Figure 4: Flow diagram of automatic user state recognition.

D. Voice Recognition

The voice recognition module can be activated when the user try to control the device through voice (X=1, Y=1, V=1). Here the contributed voice recognition is user dependant continues speech recognition type. The following figure 6 points the architecture of the proposed voice recognition module.

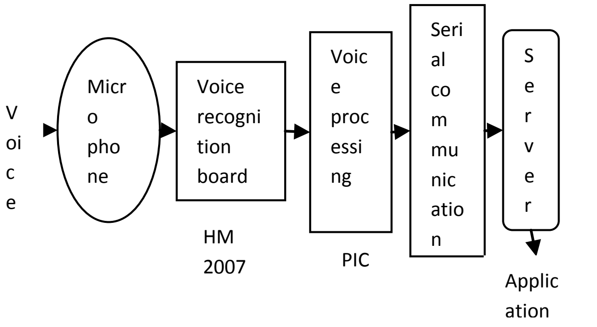

Figure 5: Functional view of voice recognition.

E. Hand Gesture Recognition

When the user is trying to control the device through hand (W=1, Y=1, H=1) the hand gesture recognition mode is turn on. Before beginning the hand gesture the system ask permission to user for start the recognition and if the user needs to start, then only the gesture module is active, in all other cases it will be in sleeping mode. Once the hand gesture recognition is active it check whether the user is frequently active and trying to control the device by hand in a time period T. Throughout this time period the gesture is in active mode, if the user is not active it count the inactive time suppose this time exceed T, automatically the hand gesture recognition mode is turns to sleeping mode, this model helps to reduces the power consumption as well as computational cost in a very often manner.

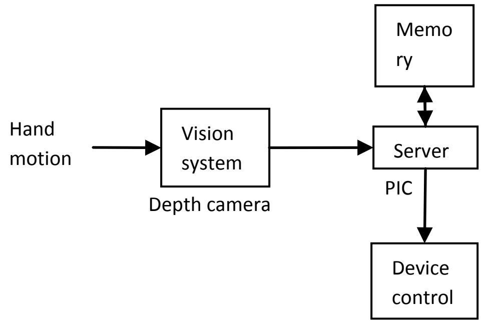

Figure 6: Functional view of hand gesture

F. User State Transition

The next possibility of the proposed scheme is user state transition and automatic updating of user state. Initially the ultrasonic sensor check the presence of user, if not any user is detected set the user state as absent suppose the presence of user is identified it turn on the depth camera. The camera device observe that the user is active or not that means consider the device or do some other work, if the user is not active count the inactive time and if exceeds threshold time TV, the user state transfer to other action and update it. When the user is trying to control by only hand, count the inactive time of voice recognition t and if this inactive time is greater than Tv the updated user state changes to control by hand, in a same way suppose the user is control only by hand count inactive time of hand gesture recognition and update corresponding user state. In other cases the user stop tries to control the device and only need to watching, user state is changes from controlling to watching.

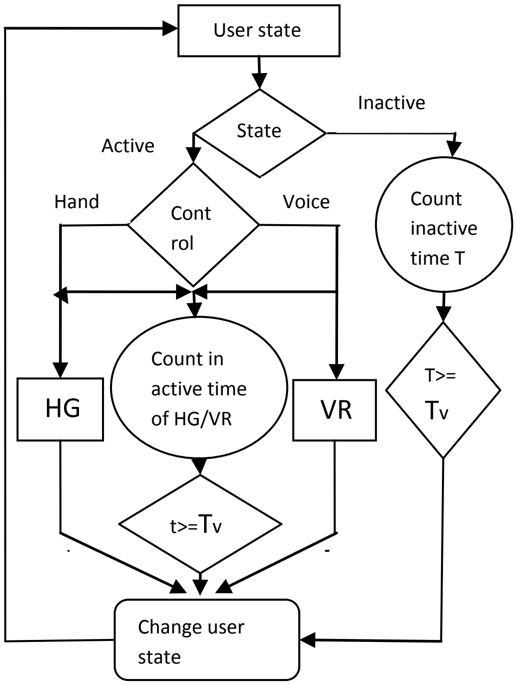

Figure 7: Flow diagram of user state transition.

IV. IMPLEMENTATION DETAILS AND DISCUSSION

The proposed system is practically implemented in a way of, use 42 inch LCD TV (Liquid crystal display television) as a device to be control and the ultrasonic sensor is taken as a combination of four distance sensors, here the sensing distance is ranges from 1m to 3.5m and sensing square is 2.7m. Next the depth camera of giving output resolutions 640X480 of picture is used and for voice recognition module the MEMS microphone of having application range 3m square is adapted. The entire system is accompanied with a single server so that the voice recognition and hand gesture modules have a common server PIC and the application such as TV control is interfaced with this server.

Figure 8: Implementation of TV control system

A. Implementation of Automatic User State Recognition

The correct action detection rate (Cadr) of implemented and tested automatic user state recognition depends on four parameters i.e., sampling interval N, length of action K so that NxK is the time duration in which the user's action can be identified. The next parameter is R threshold value of amplitude of motion, it is a tool for identify intentional user action and the rest one is continues action threshold Q, it is used to decide whether the sequence of action is enough to make an effective action.

Experiment 1:

In the first experiment action detection period NxK is tested, the other parameters are set as N=0.2, Q=K/2 and R=0.1. The result shows that the correct action detection rate is nearly 100% while the action decision period is 1.5m to 4m.

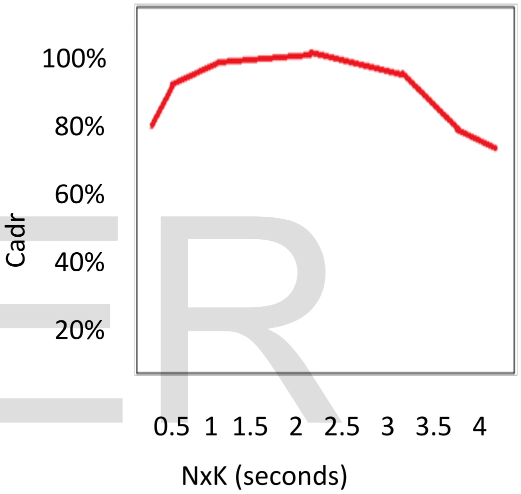

Figure 9a: The relation of action detection period and correct action detection rate.

Experiment 2

Next the threshold value R of motion is tested by setting NxK=2.4, Q=K/2 and N=0.2. It gives a result that the correct action detection rate is approximately 100% when the R value is ranges between 0.07 to 0.17 meters.

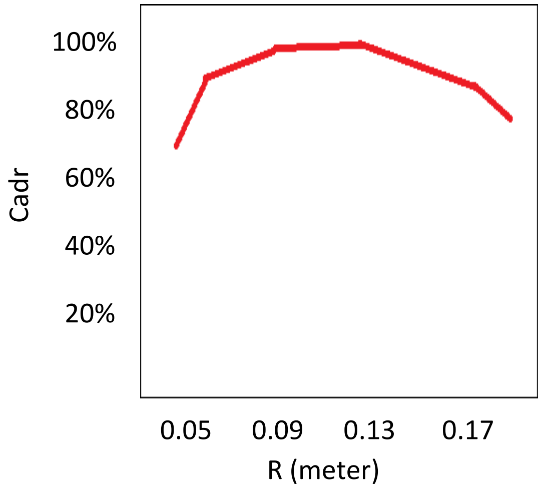

Figure 9b: The relation of amplitude motion and correct action detection rate.

B. Implementation of Hand Gesture

The icon based hand gesture; motion based hand gesture and depth camera based hand gesture are the three modes of hand gesture recognition tested here. The icon based module is used to perform a mouse like operations i.e., clicking and double clicking but differently the motion based mode is helps to swift left, right, up and down by tracking hand and drawing path of hand. The depth camera based hand recognition operated by tracking deep information's and performs like motion based module but in addition to this helps to do forward and backward options. The figure illustrates the tested hand gesture recognition here.

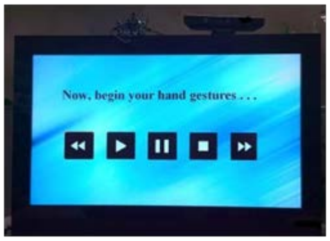

Figure 10: Information to the user while starting the hand gesture recognition.

C. Implementation of Voice Recognition

The implemented voice recognition module is user dependant and continues speech recognition type. So that initially the user's voice is familiarize to the system by a short training period, since it is user dependant and continues speech recognition type this introduced users can only access the system and while operating it allow user to give commands as a quite natural way.

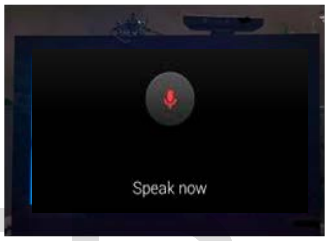

Figure 11: Information to the user while starting the voice recognition.

Here the correct voice detection rate (Cvdr) depends on three parameters user voice intensity I, external noise intensity E and distance between user and implemented system D. By keeping I and E static the D is tested, the result shows that the correct voice recognition rate obtained is nearly 100% when the D is at a range of 1 to 4 meters. In the second experiment user voice intensity I is tested by maintaining E and D as static value, here the correct voice recognition rate is approximately 100% while the voice intensity is ranges from 45 to 75 decibel. The third experiment is used to decide the noise intensity range; the correct voice detection rate is 100% when the E value is less than 20 decibel.

V. CONCLUSION

This paper proposes a touch less method for controlling the home devices i.e., voice recognition and hand gesture recognition in addition to this automatic user state recognition is accompanied. The automatic updating of user state is done by automatic user state recognition module, it consist of a ultra sonic sensor and depth camera and according to theseuser state the system decide whether and which control mode is activated and or not. Suppose the selected control way is through hand then hand gesture recognition module turn on or it is through voice then voice recognition module is turn on, a MEMS microphone, speech attention IC HM 2007 and PIC micro controller are the operating elements of voice recognition system.

In the implemented system TV is taken as a device to be control, the experimental results points that the automatic user state recognition helps to reducing the power consumption and computational cost effectively. In future makes that such a single system can be capable to control entire home devices by using ARM processor.

REFERENCES

[1] 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.

[2] 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, May2011.

[3] 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.

[4] 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, August2010.

[5] 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, [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.

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

[7] 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.

[8] 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, May2011.

[9] 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. 2011.

[10] A. Wilson, and N. Oliver, “GWindows: robust stereo vision for gesture based control of windows,” in *Proceeding of ICMI'03*, pp. 211-218, Nov. 2003.

[11] Y.-W. Bai, L.-S. Shen, Z.-H. Li, "Design and implementation of an embedded home surveillance system by use of multiple ultrasonic sensors," *IEEE Trans. on Consumer Electronics*, Vol. 56, No. 1, pp. 119-124, February 2010.

[12] Y.-W. Bai, L.-S. Shen, Z.-H. Li, "Design and implementation of an embedded home surveillance system by use of multiple ultrasonic sensors," *IEEE Trans. on Consumer Electronics*, Vol. 56, No. 1, pp. 119-124, February 2010.

[13] M. Takahashi, M. Fujii, M. Naemura, and S. Satoh, “Human gesture recognition using 3.5-dimensional trajectory features for hands-free user interface,” in *Proceeding of ARTEMIS'10*, pp. 3-8, Oct. 2010.

[14] X. Liu, and K. Fujimura, “Hand gsture recognition using depth data,” in *Proceedings of 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 529-534, May 2004.

[15] 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.