HandButton: Gesture Recognition of Transceiver-free Object by Using Wireless Networks

Weiling Zheng
College of Computer Science and Software
Shenzhen University
Shenzhen, China
zwlcle@hotmail.com

Dian Zhang*
College of Computer Science and Software
Shenzhen University
Shenzhen, China
serena.dian@gmail.com

***Abstract*—Traditional radio-based gesture recognition approaches usually require the target to carry a device (e.g., an EMG sensor or an accelerometer sensor). However, such requirement cannot be satisfied in many applications. For example, in smart home, users want to control the light on/off by hand gesture, without carrying any device. To overcome this drawback, in this paper, we propose two algorithms able to recognize the target gesture (mainly the human hand gesture) without carrying any device, based on just *Radio Signal Strength Indicator* (RSSI). Our platform utilizes only 6 telosB sensor nodes and is with a very easy deployment. Experiment results show that the successful recognition radio can reach around 80% in our system.**

***Keywords*—RSSI, wireless sensor networks, transceiver-free, gesture recognition**

I. Introduction

Gesture recognition usually aims to interpret human gesture, e.g., the body motion commonly originates from the hand or face. It attracts many researchers attention, since it is very useful in many applications to understand body language.

Traditional technologies usually utilize computer vision algorithms to interpret the body language [4] [5] [6] [7]. However, they usually cannot work in dark area due to the light requirement of the surrounding cameras. Thus, these technologies are limited to be applied in some specific applications, e.g., in the dark area. Although some other technologies are able to recognize the human gesture in dark area, most of them require the target to carry an EMG sensor or an accelerometer sensor [1] [2]. Thus, it is still a big limitation to interact with the environment. Considering the following scenario, when a person comes into a dark room, instead of trying to find the control button of the light in the dark, he/she wants to use his/her hand motion to control the light to turn on/off without carrying an additional device. Such similar requirements are very important in many applications (e.g., the smart home) to control any electronic device. Traditional technologies cannot satisfy such requirements.

In order to fill the gap between traditional technologies and real-world requirements, we propose *HandButton*, which is able to recognize human gesture without carrying any device in such scenario. The basic idea is to utilize the signal dynamic information (containing their occurrence time) caused by the human body (e.g., the hand) to detect the gesture. More detail, some wireless nodes are deployed in advance in a specific area (e.g., the door area). Each node acts as both transmitter and receiver. Thus, there are many wireless links among these nodes. When a target gesture is produced, the *Radio Signal Strength Indicator* (RSSI) value of some wireless links will change. We referred it as signal dynamics. The occurrence time of such signal dynamics are different for those wireless links. We utilize such difference to decide the target gesture.

$Figure.1$ gives an illustration of our idea, which contains six nodes marked A, B, C, D, E and F. If the number of deployed node may vary according to different applications. Node A, C and E are placed on the left side, while the other nodes are placed on the right side. Each node acts as both transmitter and receiver. When the target is motionless, as shown in the left part of $Fig.1$, the RSSI value of each wireless link will be stable. When the target performs a gesture, the RSSI value of some wireless links will change. For example, in the right part of $Fig.1$, when the target moves his/her hand from the top to down, the RSSI value of link AB usually will be first affected. Then it will be the other links, e.g., CD and EF. Our approach aims to recognize the target gesture without carrying any device in some specific scenarios, e.g., in the smart home to control any electronic device. It is easy to be extended and applied to other applications.

Our HandButton system has two algorithms to determine the target gesture. One is *Peak-Time*, the other is *Best-Fit*. The former one leverages the time with the largest signal dynamic value of each wireless link, to determine the target gesture. It is convenient for users to apply. The latter one utilizes localization algorithm to detect the target (e.g., the hand) trace, the target gesture is able to be derived. It has higher accuracy and can give more information

* Dian Zhang is the corresponding author

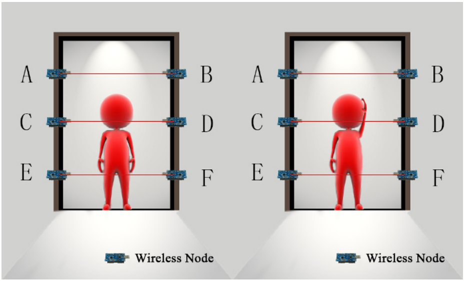


Fig. 1: An example of gesture recognition

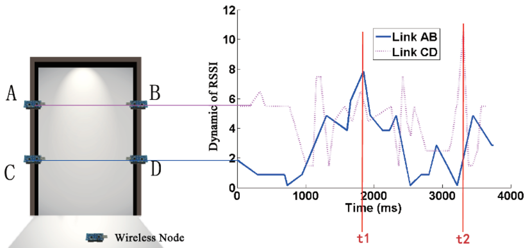


Fig. 2: An example of Peak-Time algorithm

about the gesture. Our experiments are based on 6 telosB sensors [3]. Experimental results show that the successful recognition radio can reach around 80% and the latency is only about $0.4s$, which show great potential prospects in future applications.

The main contributions of this paper are as follows. First, we are able to recognize target gesture without carrying any device, and it can be applied in dark area. It breaks the environment limitation of tradition technologies. Second, we propose Hand-Button, which contains two recognizing algorithms. Third, we conduct real experiments and comprehensive analysis in this area. At last, the cost is very low and the deployment is easy.

The rest of the paper is organized as follows. In the next section, we will discuss the related work about traditional technologies. In the following, we will introduce our two recognizing algorithms. $Section.4$ will show the experiment setup and results. Finally, we will conclude the work in the last section. Some possible future work directions are also listed in the last section.

II. Related Work

Basically, there are two kinds of gesture recognition technologies: Non-radio based technologies and Radio based technologies.

A. Non-radio Based Technologies

In Non-radio based technologies, video technologies are very popular. Video technologies [4] [5] use image processing algorithms to recognize the target gesture. In the computer vision area, the gesture, especially the extremities and head gesture can be even recognized in 3-D by using multiple cameras [6]. The gesture can also be recognized by using RGB-D data from a Kinect sensor [7], it creates the skeletal model from the depth data and the extracts the frame-level features from RGB frames, and then the depth data should be classified by multiple extreme learning machines. Although these technologies have high accuracy, they cannot be applied in the dark area. Thus, the application scenario is limited. New technologies, such as Kinect, have made a breakthrough in darkness. However, they should take a lot of training to recognize a gesture. On the other hand, the high cost prohibits the widespread use of these equipments. Moreover, they have major concerns on privacy and energy consumption.

B. Radio Based Technologies

Traditional radio-based technologies mainly use EMG sensor or Accelerometer sensor in recognition [8]. The gesture can be recognized based on the input signals from accelerometers [9]. EMG sensors measure the electrical activity produced by the muscles to recognize the gesture [10]. However, these technologies all require the target to carry a device (e.g., an EMG sensor or an accelerometer sensor).

There are some other technologies using WiFi [11], which doesn't require the target to carry a device. Most WiFi technologies use MIMO [12] technologies and USRP [13] to estimate the human gesture. However, they require devices with multiple antennas and should get physical layer information. It is a limitation to most common device. Our algorithms have no such requirements and they can be widely for almost all the wireless device.

III. Methodology

In this section, we will introduce two algorithms to recognize target gesture: Peak-Time and Best-Fit. The basic idea behind the two algorithms is to detect when the wireless link signal will be affected by the target gesture, then derive the gesture direction from the time difference between links. The wireless signal is measured by RSSI information. If in a *static environment* where no object moves around and the target object is motionless, the RSSI value of each wireless will be stable. If in a *dynamic environment* where the target performs some gesture, the RSSI value of some wireless link will change (referred as *RSSI dynamics*). Those changed wireless links plus their change time will be leveraged to derive the gesture.

A. Peak-Time Algorithm

This algorithm is able to be applied in different node topologies and recognize simple target gestures. In this algorithm, we may recognize gesture G_1 (from top to down) and G_2 (from down to top). Without loss the generality, we introduce this algorithm based on the grid setting of the nodes.

Suppose the gesture recognition area is a door area, as shown in $Fig.2$. In this example, we have

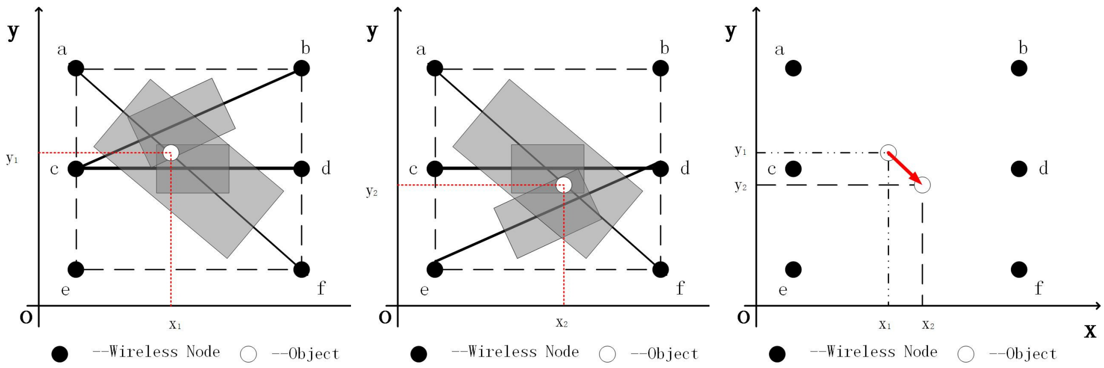


Fig. 3: An example of Best-Fit algorithm

4 nodes deployed in this area. Each node will act as both transmitter and receiver. Therefore, in total we have 6 wireless links (we regard the symmetric links as one link). In this algorithm, we only utilize the horizontal wireless link of each pair of nodes (in the example, they are link AB and CD). For each horizontal wireless link, when the target is motionless, the RSSI value of each horizontal wireless link is stable. When the target gesture occurs (e.g., a human gesture from down to up), link CD usually will be affected first. Its RSSI value will change first. As shown in the right part of $Fig.2$, the RSSI value of link CD changes dramatically. The change is the largest in time t_1 (referred to *Peak-Time*). The RSSI value of link AB changes dramatically with the largest in time t_2. Also we may find that, the time t_1 is earlier than time t_2. We may conclude that the gesture is G_2 (down to up). In order to get higher accuracy, we may deploy more nodes for redundancy. How many number nodes should be used in our system will be discussed in the later section.

In general, for each horizontal wireless link ij between node i and j, its RSSI value in static environment is r_{ij}. Suppose

$$\bar{R}_{ij} = \frac{\sum_1^N r_{ij}}{N} \tag{1}$$

where $\bar{R_{ij}}$ is the average RSSI value of N RSSI static value obtained from the horizontal wireless link ij.

In dynamic environment, when the target gesture occurs, the RSSI dynamic value of each horizontal wireless link ij between the node i and j is D_{ij}. It is noted as d_{ij}, which is the RSSI difference between static environment and dynamic environment.

$$d_{ij} = \left| D_{ij} - \bar{R}_{ij} \right| \tag{2}$$

For each horizontal wireless link ij between node i and j, we will calculate its peak time T_{ij} when the value of d_{ij} is the largest. It can be expressed as

$$T_{ij} = T_{\max(d_{ij})} \tag{3}$$

Therefore, the target gesture can be derived by all the peak time of all the horizontal wireless links. Among the whole set, for each two horizontal wireless links ij between node i and j and link uv between node u and v, suppose their Y axis-coordinates are y_{ij} and y_{uv}. Their corresponding peak times are T_{ij} and T_{uv}, respectively. We will decide the gesture G for link ij and uv as follows.

$$G = \begin{cases} G_1 & if\ (y_{ij} > y_{uv}\ and\ T_{ij} > T_{uv}) \\ & or\ (y_{ij} < y_{uv}\ and\ T_{ij} < T_{uv}), \\ G_2 & if\ (y_{ij} > y_{uv}\ and\ T_{ij} < T_{uv}) \\ & or\ (y_{ij} < y_{uv}\ and\ T_{ij} > T_{uv}), \\ F & if\ (T_{ij} = T_{uv}). \end{cases} \tag{4}$$

where F means the failure of the gesture recognition.

If the number of horizontal wireless links n is more than two, we will have C_n^2 pair of horizontal wireless links (for example, if there are 3 horizontal wireless links, then we will have 3 pair of those links). Applying $Equ.4$ to these C_n^2 pair of horizontal wireless links, suppose there are l number of gesture decision falls into gesture G_1, while k number of gesture decision falls into gesture G_2, here $l + k = C_n^2$. By using the $Equ.4$, we will have the final gesture decision G_{final} as

$$G_{final} = \begin{cases} G_1 & if\ l > k, \\ G_2 & if\ l < k, \\ F & if\ l = k. \end{cases} \tag{5}$$

B. Best-Fit Algorithm

Previous Peak-Time algorithm only can decide a simple target gesture. If users want to get a more accurate target gesture, they may use the Best-Fit algorithm. The basic idea behind this algorithm is to utilize localization algorithm to get the trace of target gesture. Then the gesture behavior is able to be decided.

According to the model of our previous research, for each wireless link, if the target is closer to the center of the link, the RSSI value will change more. Therefore, for each wireless link, once its RSSI value changes, we will estimate the possible target area for such link, which is able to be presented by a rectangle. The length and the width of the rectangle can be obtained from our previous model [14].

Fig. 4: Experimental environment

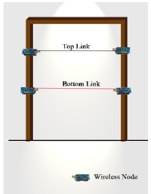


Fig. 5: The example of node deployment

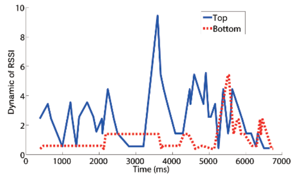


Fig. 6: The example of a figure caption (when node distance is 40cm)

The basic idea of our Best-Fit algorithm is illustrated in $Fig.3$. As $Fig.3$ shown, we have deployed six nodes in a door area. Each node will act as both transmitter and receiver. Therefore, in total we have 15 wireless links (we regard the symmetric links as one link). For each wireless link, we measure the RSSI dynamic value, which is the RSSI difference between static environment and dynamic environment, as introduced before. If the RSSI dynamic value is larger than zero, we will estimate the possible target area for this link, as the rectangle for link cd, af and cb, colored in grey. A larger RSSI dynamic value will cause a smaller rectangle area but with a large weight. These rectangles may be overlap as shown in $Fig.3$.

In the following, every fixed interval (in our experiment this value is $200ms$), we calculate the intersection points of the rectangles. The estimated target position is the weight average position of these intersection points. As shown in $Fig.3$, position (x_1, y_1) is the first estimate result. Position (x_2, y_2) is the second estimate result. So we may conclude that the target trace is from (x_1, y_1) to (x_2, y_2). Since $y_1 > y_2$, we may decide the gesture belongs to G_1. Moreover, it can give more information.

In general, suppose we have m nodes. Therefore, we have C_m^2 wireless links. As mentioned before, the RSSI dynamic value of each wireless link d_{ij} can be calculated from $Equ.1$ and $Equ.2$ (here we consider all the wireless links, not just the horizontal links). Each d_{ij} creates a rectangle area, in which the object body is likely to reside. For each rectangle area of link ij, its corresponding rectangle area is identified as A_{ij}. The weight value w_{ij} of A_{ij} is calculated by the follow formula

$$w_{ij} = d_{ij} + \sum_{u,v=1}^{m} \frac{A_{ij} \bigcap A_{uv}}{A_{uv}} \times d_{uv}(u < v) \quad (6)$$

here A_{uv} is the area of each rectangle overlapping with A_{ij}.

We choose the top K rectangles with the largest weight values (in our experiment we choose this values as 5) and calculate their intersection points. Then we calculate their average position of these intersection points as the target location (x_s, y_s). Such location estimate will be repeated every fixed interval (in our system this value is $200ms$). For each two temporally adjacent estimated target location (x_s, y_s) and (x_{s+1}, y_{s+1}), we may decide the gesture G as follows.

$$G = \begin{cases} G_1 & if\ y_{s+1} - y_s < 0, \\ G_2 & if\ y_{s+1} - y_s > 0, \\ F & if\ y_{s+1} - y_s = 0. \end{cases} \quad (7)$$

If the number of gesture decision p is more than two. Suppose there are r number of gesture decision falls into gesture G_1, while q number of gesture decision falls into gesture G_2 , here $r + q = p$, and the final gesture decision G_{final} will be generated by following formula.

$$G_{final} = \begin{cases} G_1 & if\ r > q, \\ G_2 & if\ r < q, \\ F & if\ r = q. \end{cases} \quad (8)$$

This Best-Fit algorithm can not only give the gesture decision, but also can give more information about the target trace.

IV. Performance Evaluation

A. Experiment Set Up

We run the experiments in our lab, which area is 80 square meters. The wireless nodes we use are telosB sensor nodes with Chipcon CC2420 radio chips [15]. TelosB is composed of the MSP430 (the MSP430F1611) microcontroller and the CC2420 radio chip. The microcontroller of this mote operates at $4.15MHz$ and has a $10kBytes$ internal RAM and a $48kBytes$ program Flash memory. The working band stands on $2.4GHz$. The sensors are deployed in a regular door area as shown in $Fig.4$. The width and the height of the door area are about $110.2cm$, and $213.8cm$, respectively. In detail, the sensors are deployed on either side of the door with fixed distance between them. Here we only consider the nodes deployment in a 2D area. We program all the sensor nodes to broad cast beacons periodically with the same interval. Each node broadcasts beacon messages periodically and listens to the beacons from its neighbors as well. The transmission power is defaulted at $0dBm$.

B. Infrastructure

In general, our HandButton system consists of two phases. First, in the initialization phase, each node builds a static table to store the static RSSI values for all its neighbors after receiving a few numbers of beacons (5 in our experiment). The average RSSI values of each pair of two nodes will

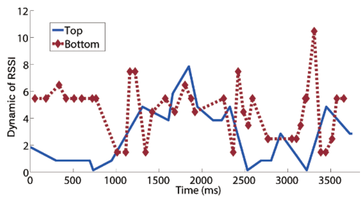


Fig. 7: The example of figure caption (with four nodes)

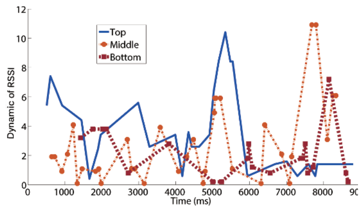


Fig. 8: The example of figure caption (with six nodes)

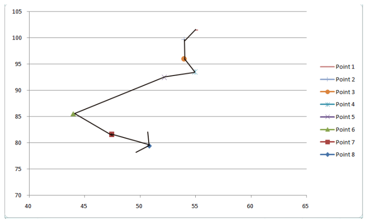


Fig. 9: The example of gesture trajectory

be adopted as the benchmark to estimate the posture. The initialization phase has to be carried out in the static environment. After initializing all the nodes, the system enters the recognition phase. Each node measures the RSSI dynamic value (compared with the static value) caused by the target gesture and reported back to the sink node. After receive the RSSI dynamic value and its occurrence time, we can estimate target gesture by leveraging our proposed algorithms.

C. The Impact of Node Distance

TABLE I: COMPARISON OF DIFFERENT NODE DISTANCE

Node Distance	30cm	40cm	50cm	60cm
Accuracy (%)	82.35	86.20	72.73	72.73

In this experiment, we will investigate how the node distance will impact the successful recognition radio. Since the sensors are deployed in the door area to recognize the body motion and the width of the door is fixed, we only test the vertical node distance along each side of the door. Therefore, we adjust the perpendicular distances of each node from $30cm$ to $60cm$, with each step separating by $10cm$. We leverage 4 sensors in the door area, with two on each side as shown in $Fig.5$. The line between the upper two sensors is the top link. The line between the lower two sensors is the bottom link. We test human gesture from top to down. For each node distance, we tested 15 rounds. The experiment result is shown in $Table.I$. We find that, when node distance is set to $40cm$, the result is the best, whose successful recognition ratio is 86.2%. $Figure.6$ is one of the results when node distance is $40cm$.

We also find that, the nodes distance less than $30cm$ will not be benefit to the gesture recognition successful ratio. The reason may be the following. If the distance between the top link and bottom link is too close, the time difference between these two links caused by the body motion is not obvious. Thus, the gesture is more difficult to recognize. We also skip testing those node distances larger than $60cm$. The reason is that, in general human body motion, e.g., the hand gesture is unlikely to cover a very long distance. Therefore, in the following experiment, we will set the node distance default as $40cm$. We mainly compare the following results with the best setting in our experiment.

D. The Impact of Node Numbers

TABLE II: COMPARISON OF DIFFERENT NODE NUMBER

Node Number	Four Nodes	Six Nodes
Accuracy (%)	73.33	80.00

We totally tested averaged 20 rounds of human hand gesture. $Figure.7$ is one of the results when the node number is four. $Figure.8$ is one of the results when the node number is six. The top link, middle link and the bottom link represent the horizontal wireless link between the top two nodes, two middle nodes and the two bottom nodes, respectively.

Base on all the samples, as $Table.II$ shows, the successful recognition ratio with four nodes is 73.3% and 80% with six nodes, which is improve by 6.7%.Therefore, in the later experiments, we utilize this setting with 6 nodes in a 2D area. The reason why we do not use more nodes is that, since the vertical node distance we choose is $40cm$, totally the vertical distance this setting will cover is $80cm$. A typical human hand gesture will not be over this distance.

E. The Impact of Moving Speed

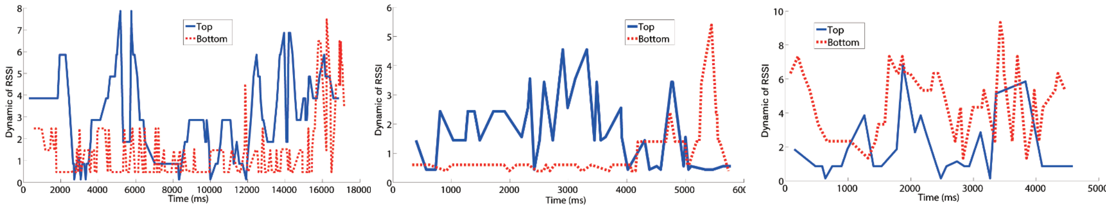
TABLE III: COMPARISON OF DIFFERENT MOVING SPEED

Moving Speed	Slow	Normal	Quick
Accuracy (%)	71.43	88.89	66.67

To learn how the target gesture will influence the successful recognition rate, we arrange a person to sweep his arm down through a fixed trace between the two sensor grids on the either side of the door. The vertical distance of the nodes are fixed at $40cm$ and in total we use 6 nodes, which has been introduced in the previous subsection. We test three kinds of moving speed: slow (about $2.27m/s$), normal (about $4.48m/s$), and fast (about $10.02m/s$). The examples of three moving speed is shown in $Fig.11$. The Experiment results are shown in $Table.IV$. We find that, when the gesture is performed at a normal speed, the successful recognition ratio is the best.

F. Algorithm Comparison

As introduced in the previous subsection, we choose 6 nodes and vertical node distance as $40cm$



(a) Slow (b) Normal (c) Quick

Fig. 10: The example of figure caption (slow, normal and quick)

TABLE IV: ALGORITHM COMPARISON

Algorithm	Peak-Time	Best-Fit
Accuracy (%)	75.56	76.67

in a 2D area as the final gesture recognition deployment. We compare our Peak-Time and Best-Fit algorithms, based on 55 rounds of different human hand gesture test. The experiment result is shown in $Table.V$. We can find that, the successful recognition ratio of the Peak-Time algorithm and Best-Fit algorithm are 75.56% and 76.67%, respectively. We may see that, Best-Fit algorithm performs the best. Moreover, the Best-Fit algorithm is able to find more complicate target gesture than Peak-Time algorithm. As shown in $Fig.9$, the estimate trace is able to give more information of the target gesture.

G. Latency

The latency of gesture recognition system mainly depends on how much time for the data was collected. In our experiment, to avoid collision among 6 nodes, we set the beacon interval as $200ms$ to transmit a packet with $51bytes$. For Peak-Time algorithm, the system latency is $200ms$. But for Best-Fit Algorithm, we should estimate the target location over 2 times to decide the gesture. Therefore, the latency should be larger than $2 \times 200ms = 400ms$.

V. Conclusion And Future Work

In this paper, we have presented two algorithms for recognizing target object gesture behavior (mainly the human being) by using wireless sensor networks. At the same time, the target does not require to carry any device. The first Peak-Time algorithm uses the time difference of RSSI dynamics among difference links for recognition. It is easy to perform and with well successful recognition probability. The latter Best-Fit algorithm introduce localization algorithm, which has higher successful recognition probability and able to recognize more complicate gesture. We perform our algorithm in real environment. The experiment results show that we may recognize target gesture with the successful probability up to about 80%. As future work, we will try to recognize more different gestures. Furthermore, we may try to recognize gestures of multiple objects. More nodes deployment will be under consideration in our future work. We also may try to deploy the nodes in a 3D area.

Acknowledgment

This research was supported by The China NSFC Grant 61202377, 61170076 and U1301252.

References

[1] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," presented at IEEE Transactions on Systems, Man, and Cybernetics, Part A, 2011, pp.1064-1076.

[2] V. Kosmidou and L.J. Hadjileontiadis, "Sign language recognition using intrinsic-mode sample entropy on sEMG and accelerometer data," presented at IEEE Trans. Biomed. Engineering, 2009, pp.2879-2890.

[3] "XBOW Corporation," TelosB Mote Specifications, http://www.xbow.com/Products/pdofuctdetail-s.aspx?sid=252,2005.

[4] R.Z. Khan and N.A. Ibraheem, "Survey on gesture recognition for hand image postures," presented at Computer and Information Science, 2012, pp.110-121.

[5] Z. Ren, J. Meng, J. Yuan and Z. Zhang, "Robust hand gesture recognition with kinect sensor," in Proc. ACM Multimedia, 2011, pp.759-760.

[6] C. Tran and M.M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," presented at IEEE Trans. Industrial Informatics, 2012, pp.178-187.

[7] X. Chen and M. Koskela, "Online RGB-D gesture recognition with extreme learning machines," in Proc. ICMI, 2013, pp.467-474.

[8] O.D. Lara and M.A. Labrador, "A survey on human activity recognition using wearable sensors," presented at IEEE Communications Surveys and Tutorials, 2013, pp.1192-1209.

[9] R.Z. Xu, S.L. Zhou and W.J. Li, "MEMS accelerometer based nonspecific-user hand gesture recognition," IEEE Sensors Journal, vol.12, no.5 pp.1166-1173, May 2012.

[10] Z. Lu, X. Chen, Q. Li, X. Zhang and P. Zhou, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices," presented at IEEE T. Human-Machine Systems, 2014, pp.293-299.

[11] F. Adlib, Z. Kabelac and D. Katabi, "Multi-person motion tracking via RF body reflections," Computer Science and Artificial Intelligence Laboratory Technical Report, 2014.

[12] Q. Pu, S. Gupta, S. Gollakota and S. Patel, "Whole-home gesture recognition using wireless signals," in Proc. MOBICOM, 2013, pp.27-38.

[13] F. Adib and D. Katabi, "See through walls with WiFi!," in Proc. SIGCOMM, 2013, pp.75-86.

[14] D. Zhang, J. Ma, Q. Chen and L. Ni, "An RF-based system for tracking transceiver-free objects," Proc. Fifth Ann. IEEE Int'l Conf. Pervaise Computing and Comm. (PerCom '07), pp. 135-144, 2007.