Gesture Segmentation and Positioning based on Improved Depth Information

Haibo Lin[a], Shengbin Wang[b] and Yi Zhang[c]

Engineering Research and Development Center for Information Accessibility, Chongqing University of Posts and Telecommunication, Chongqing 400065, China.

[a]844038217@qq.com, [b]1357503397@qq.com, [c]1905613719@qq.com

Abstract

Aiming at the problem that segmented gesture by Kinect depth information usually contains wrist data, which easily causes subsequent false gesture recognition, a gesture segmentation and positioning algorithm based on improved depth information was proposed. Firstly, the gesture binary image was detected based on depth information threshold limit in experimental space. Secondly, according to characteristics of common gestures, accurate gesture was segmented by gesture endpoint detection and variable threshold algorithm. In order to obtain stable segmentation results, morphological processing of segmented gesture was conducted. Lastly, the gesture positioning algorithm was proposed based on the method of combining gesture gravity center coordinates and maximum inscribed circle center coordinates. The experimental results show that the proposed gesture segmentation method has better accuracy and stability than the existing algorithm. The combined gesture positioning is more stable than gesture gravity center positioning and skeletal data positioning of Kinect Software Development Kit (SDK) and it has no singular points.

Keywords

Gesture Segmentation, Gesture Positioning, Variable Threshold, Maximum Inscribed Circle.

1. Introduction

With the development of computer technology, gesture recognition based on machine vision as a very promising human-computer interaction technology has always been a worthy research topic[1]. In the gesture recognition process, gesture segmentation and positioning lay the foundation for the gesture trajectory extraction and gesture recognition. Skin color segmentation based on depth information is new[2], firstly,it needs get Kinect depth information and uses the depth threshold to split gesture[3-4], but the skin color detection traversals the entire frame of the image in long time and it is not real-time. Different skin color and skin-like objects can result in the interference of gesture segmentation with the wrist [5-6]. Literature [7-9] connect depth information and color image to split gestures, but it can not avoid the impact of light,which reduces the robustness of gesture segmentation. Literature [10-11] use depth information to split the gestures and wrist, then it uses the distance conversion operation in the two parts to find their own center point of distance gray image and connects the two points to take the vertical bisector, the method accurately splits the gesture and wrist in complex computing.

Therefore, a gesture segmentation and positioning algorithm based on improved depth information was proposed and accurate gesture was segmented by gesture endpoint detection and variable threshold algorithm, and the robustness is good, which reduces the computational complexity. The gesture positioning algorithm was proposed based on the method of combining gesture gravity center coordinates and maximum inscribed circle center coordinates, which increases the stability of the wheelchair control.

2. Gesture segmentation algorithm of improved depth information

2.1 Gesture detection based on depth information threshold limit

In gesture recognition process, accurate gesture detection is critical. The gesture is facing the Kinect camera and limits the motor up to about 10 °. According to the depth threshold gesture segmentation method in South China University of Technology[3], the depth information acquisition range is limited to the distance[d_1 - d_2] between the Kinect camera and the gesture, but can not meet the precise and safe control of wheelchair, in order to get more stable and accurate results,then $d_1 = 250\text{mm}$, $d_2 = 450\text{mm}$. To eliminate the interference of the complex environment outside experimental space and obtain better segmentation result, gesture experimental space is set as shown in Figure 1, the coordinate system $\{V\}$ is the gesture basis coordinate system.

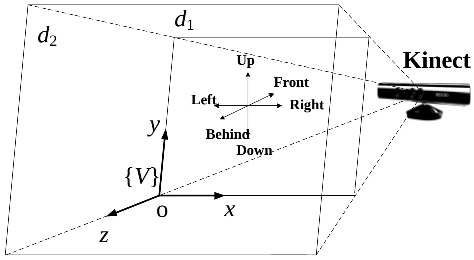


Fig. 1 Gesture experimental space setting

According to the existing information based on depth threshold gesture detection[3-4, 12], RGB image, depth shaded image (the pixel points in the depth [d_1 - d_2] is linearizated to the gray value [1-254], the gray is 0 and 255 in the depth less than d_1 and more than d_2) and gesture binary image (make the binarization for the depth shaded image, gesture gray value is 255, the rest is 0) is easy to obtain, the result is shown in Figure 2. Matrix $\boldsymbol{G}(i, j)$ is the pixel gray value of the depth shaded image in Figure 2 (b), Binary matrix $\boldsymbol{H}_1$ is gesture binary image in Figure 2 (c), the gesture value is 1, the rest is 0. $\boldsymbol{G}$ and $\boldsymbol{H}_1$ are both $m \times n$ matrix, m and n is pixel high and pixel wide.

(a) RGB image (b) Depth shaded image (c) Gesture binary image

Fig. 2 Three gesture images

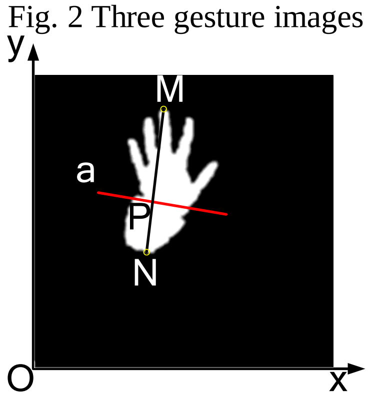


Fig. 3 Accurate gesture segmentation

2.2 Gesture segmentation based on variable thresholds

As shown in Fig. 2 (c), gestures based on the depth information threshold limit detection is inaccurate with wrist interference for the following gesture trajectory extraction and gesture recognition. Literature [10-11] split the gesture and wrist in complex computing by distance transformation. Accurate gesture was segmented by variable threshold algorithm, which is more simple and stable than the literature [10-11].

In the gesture experimental space, the general gesture characteristics are studied and the fingers are tilted upward or upward. In Figure 2 (c) gesture binary image the highest point and the lowest point are found quickly, as shown in Figure 3, line a is the any vertical line of line MN, P is the intersection. The required formula is set to:

$$\mathrm{MP} = \lambda \mathrm{MN}. \quad (1)$$

In Figure 3, the acquisition of M point and N point is susceptible to isolated points and holes noise, and the boundary of the segmented image is not smooth enough, resulting in the deviations of accurate acquisition of M point and N point. Therefore, the morphological processing method is used to eliminate the interference. The size of each connected area extracted from gesture binary image is calculated, and the smaller isolated points are removed by the threshold limit. The holes in the gesture image are eliminated by the closed operation in morphology and the accurate gesture binary image marked as $\boldsymbol{H}_2$ .is obtained.

Gesture random movement makes the wrist area size and gesture tilt constantly change and the appropriate variable threshold can ensure the accuracy of segmentation. Through a large number of experimental, the experimental space is set from depth d_1 to depth d_2, λ is got by the depth value, the experiment proved that the sub-quadratic function is shown as the following:

$$\lambda = f(\bar{g}_{\boldsymbol{H}_2}) = \begin{cases} c_1, & 1 \le \bar{g}_{\boldsymbol{H}_2} \le \bar{g}_1 \\ [\bar{g}_{\boldsymbol{H}_2}{}^2, \bar{g}_{\boldsymbol{H}_2}, 1][a,b,c]^{\mathrm{T}}, & \bar{g}_1 < \bar{g}_{\boldsymbol{H}_2} < \bar{g}_2 \\ c_2, & \bar{g}_2 \le \bar{g}_{\boldsymbol{H}_2} \le 254 \end{cases}. \quad (2)$$

According to the experimental, $[a,b,c]^{\mathrm{T}} = \begin{bmatrix} 0.00004093909191 \\ -0.00887636364636 \\ 0.97710000000000 \end{bmatrix}$, $c_1 = 0.5$, $c_2 = 1$, $\bar{g}_1 = 90$, $\bar{g}_2 = 234$. The gesture average depth value is:

$$\bar{g}_{\boldsymbol{H}_2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{G}(i,j)\boldsymbol{H}_2(i,j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{H}_2(i,j)}. \quad (3)$$

In formula (3), m and n is the depth image high and wide, the unit is pixel. $\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{G}(i,j)\boldsymbol{H}_2(i,j)$ is the multiplied two-type matrix. The gesture binary image is got to split the accurate gesture by the method of Figure 3 and the gesture binary image without wrist is matrix $\boldsymbol{H}_3$.

3. Improved joint gesture positioning algorithm

3.1 Gesture center of gravity positioning algorithm

Gesture positioning is the basis for controlling the wheelchairs stably and safely. The real-time gesture positioning is to extract the gesture trajectory[13]. The gesture trajectory is the important feature of dynamic gesture. In the experimental space, $\{V\}$ is the coordinate system, (x_g, y_g, z_g) is set as the gesture center of gravity coordinatesm and the gesture average depth shaded value is z_g. Gesture positioning formula is shown:

$$(x_g, y_g, z_g) = \left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} j\boldsymbol{H}_3(i,j)}{S_{\boldsymbol{H}_3}}, m-1-\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} i\boldsymbol{H}_3(i,j)}{S_{\boldsymbol{H}_3}}, \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{G}(i,j)\boldsymbol{H}_3(i,j)}{S_{\boldsymbol{H}_3}}\right), S_{\boldsymbol{H}_3} = \sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{H}_3(i,j). \quad (4)$$

$\boldsymbol{G}(i,j)\boldsymbol{H}_3(i,j)$ is the multiplied two-type matrix; the unit of x_g, y_g is pixel; the value of z_g is [0-255], the unit is the gray value. The size of the gesture is $S_{\boldsymbol{H}_3}$, the unit is pixel, that is, (x_g, y_g, z_g) is gesture positioning coordinates.

3.2 Maximum inscribed circle positioning algorithm

The maximum inscribed circle method is used to positioning the gesture based on the gesture shape feature. The gesture pixel point set is *T*, the gesture edge pixel point set is *E*. $E \subset T$, Point $t \in T$ is the pixel point within the gesture. For any $t_i \in T$,

$$d_i = \min_j(\sqrt{(\mathrm{x}_{t_i} - \mathrm{x}_{e_j})^2 + (\mathrm{y}_{t_i} - \mathrm{y}_{e_j})^2})\,. \tag{5}$$

$\mathrm{x}_{t_i}, \mathrm{y}_{t_i}$ and $\mathrm{x}_{e_j}, \mathrm{y}_{e_j}$ is the vertical and horizontal coordinates of the point t_i and e_j, d_i means the minimum distance from point t_i to the edge of the gesture. For all d_i, define the gesture radius:

$$R = \max_i(d_i)\,. \tag{6}$$

The largest inscribed circle center is point t_i and radius R is obtained by t_i, the corresponding x_{t_i} and y_{t_i} is the center of circle coordinates. The radius is determined by making inscribed circle with formulac (5) and (6), the calculation flow chart is shown in Figure 4.

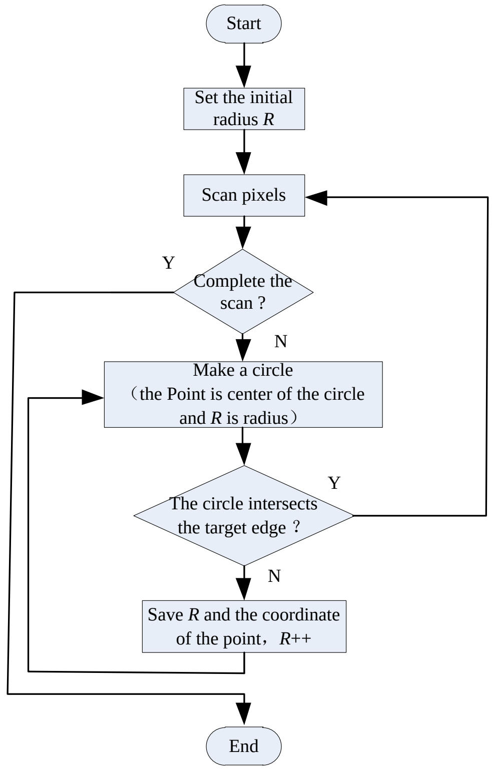


Fig. 4 positioning procedure of maximum inscribed circle algorithm

Figure 4 defines an initial radius, which is the minimum radius allowed. If the circle can not be made with the radius, it means that a sufficiently large part can not be found to lock target, so it is not a target gesture. Figure 5 intuitively shows that the center and radius of the gesture are obtained by the maximum inscribed circle method. For the target gesture, The circle with point A as the circle center is the largest circle, so point A is the largest inscribed circle center, the circle radius is the gesture radius.

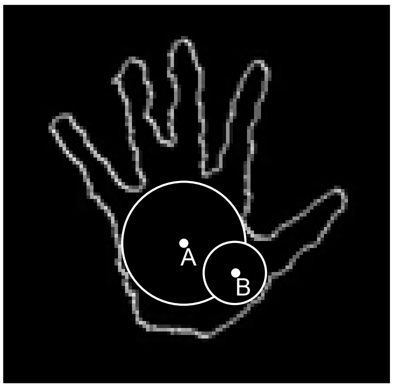


Fig. 5 Schematic diagram of maximum inscribed circle algorithm

Figure 6 shows that the gesture center and radius are obtained by using the maximum inscribed circle method. The four candidate targets are marked as different colors, A and B meets the condition of the area size, after the elimination process of internal point, the target area size of C and D is small and the resolution is low, so C and D can not be identified. In order to cut the cost, C and D are ignored. Since in target B area circle can not be made with the initial radius, which means target B is not gesture. The target A meets all the conditions and target A is processed by the improved method, which proved that gesture positioning is accurated.

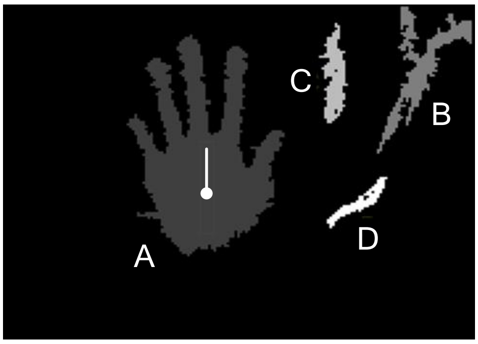


Fig. 6 Experimental results of maximum inscribed circle algorithm

3.3 Joint gesture center of gravity and maximum inscribed circle center positioning algorithm

Nikhil used radiation method to accurately positioning gesture[14], which is not simple and produces much redundant data. Firstly, when the gesture is binarized, the accuracy of the gesture center of gravity is reduced when the arm appears in the target video, and the maximum inscribed circle positioning method based on the gesture shape feature positioning gesture with fixed-point operations, which expands the application range of the algorithm, but the maximum inscribed circle positioning method only obtains the circle center two-dimensional coordinate and lacks the average depth gray value (the vertical coordinate) of the gesture area, in order to make up the deficiency of the two positioning methods, Combining the gesture center of gravity and the maximum inner circle center to get gesture positioning result. The two-dimensional coordinates of the center of gravity is the two-dimensional coordinates of the gesture positioning, gesture vertical coordinates is z_g, which gets the accurate joint positioning results, including the gesture radius R. The method is important in the gesture trajectory extraction, gesture tracking and gesture recognition. Joint gesture positioning flow chart is shown in Figure 7.

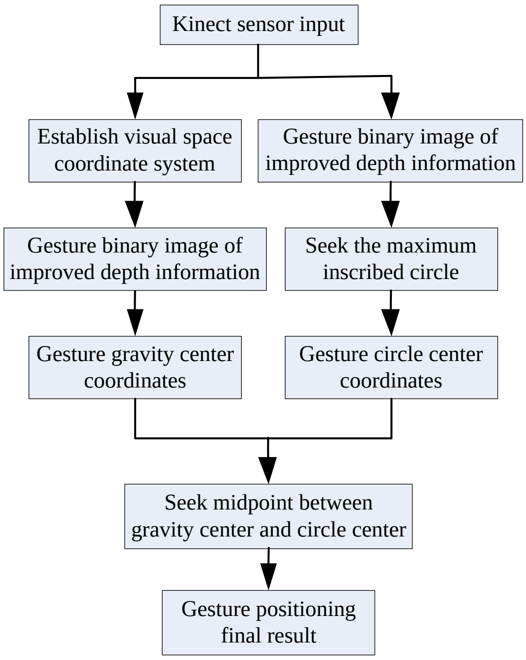


Fig. 7 Combined gesture positioning procedure

4. Experimental comparison and analysis

4.1 Gesture segmentation experiment

In order to verify the validity of the algorithm, experiment is conducted. Microsoft company Kinect (XBOX 360) depth camera is running for the visual information collection by the software Kinect for Windows Software Development Kit-v1.7 and Kinect driver software, Microsoft Visual Studio 2010 development environment and Matlab2012a software in the CPU for 3.30 GHz is used in the computer with memory capacity on a 3.19 GB. In gesture segmentation process, the depth data format for the depth image is 480 pixel high and 640 pixel wide.

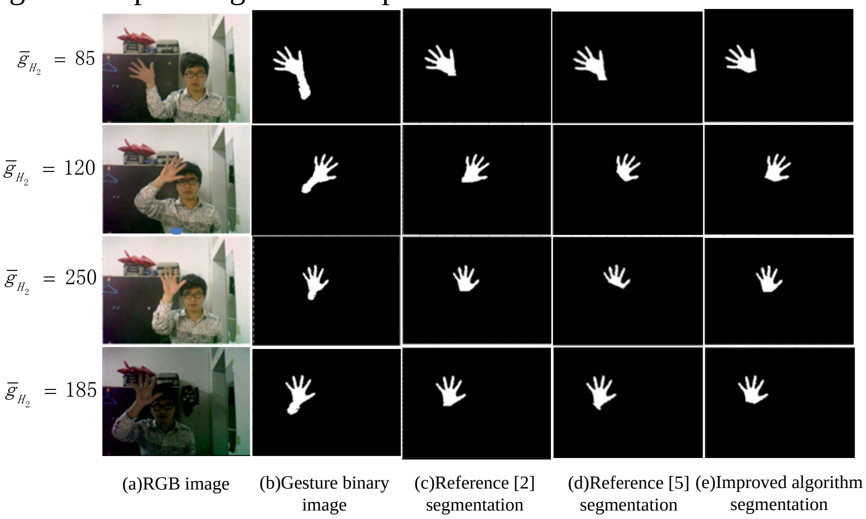


Fig. 8 Gesture segmentation comparison in different conditions

In the complex background condition contrast and analyze the experimental results, gesture segmentation results comparison is shown in Figure 8. Each line in the graph means the different average distance from the camera, the different gesture angles, and the different lighting conditions (tilted to the left, tilted to the right, vertical and dark environment vertical). By comparing the segmentation results, the gesture segmentation method can obtain the accurate gesture segmentation results in complex and variable conditions, which lays the foundation for the subsequent gesture trajectory extraction, gesture tracking and gesture recognition.

4.2 Gesture positioning experiment

In Kinect's own coordinate system, Kinect can provide human skeletal data format [13] (including 20 human body main skeleton point three-dimensional coordinates with gesture). Microsoft obtains the estimated depth information through a special algorithm for the whole body in the visual scene. The gesture trajectory information can be obtained by directly reading the gesture skeleton point, the frequency of the skeletal data and the depth data frame are both 30 Hz. Gesture in the Kinect scene do the spiral movement and the skeletal data positioning the gesture trajectory, but with the shelter of gesture and body, the trajectory is easy to jump [15], as shown in Figure 9, and the skeletal data positioning trajectory is easy to overlap, as shown in Figure 10 (a).

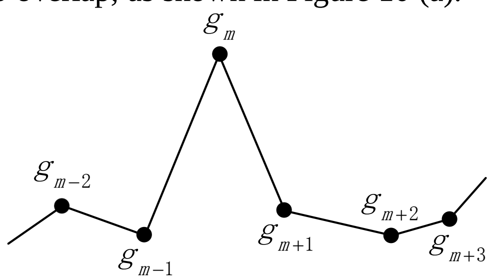
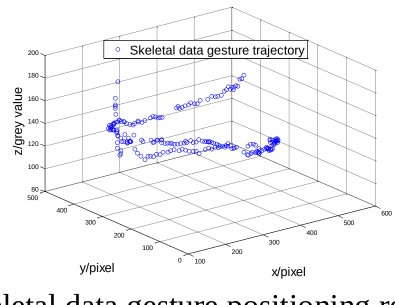
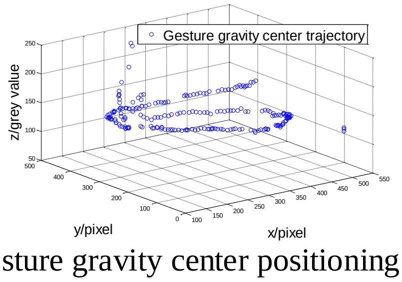


Fig. 9 Abnormal gesture trajectory

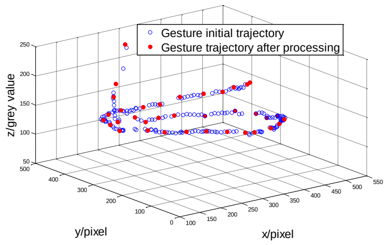
Figure 10 (b) shows that gesture center of gravity positioning trajectory has singular point interference. Jointing center of gravity and center of the maximum inscribed circle positioning algorithm is proposed. The midpoint of center of gravity and center of the maximum inscribed circle is selected as the gesture to capture. For the possible abnormal gestures trajectory, the spherical threshold limit algorithm can be used to deal with it, 1 second is divided into three time slices to calculate the average trajectory point, take the exception gesture spherical limit threshold $\tau_r = 100$ to get the initial gesture trajectory and the trajectory after processing, as shown in Figure 10 (c) which is more stable and reduces the trajectory redundancy. The unit of x, y axis in Figure 10 is pixel, z axis unit is gray value. Contrasting Figure 10 (a), 10 (b) and 10 (c) results, skeletal data gesture positioning is easy to jump and overlap, the gesture center of gravity positioning has abnormal trajectory. The gesture positioning results of the improved algorithm is more stable and it has certain resistance ability to the abnormal gesture trajectory.



Skeletal data gesture positioning result



(b) Gesture gravity center positioning result



(c) A the improved algorithm gesture positioning and result after processing

Fig. 10 Comparison of gesture positioning results

5. Conclusion

A gesture segmentation and positioning algorithm based on improved depth information was proposed. The gesture segmentation result is more accurate by introducing the variable threshold. The algorithm removes the wrist and do the morphological processing for gesture segmentation, which makes the gesture segmentation more accurate. In order to solve the instability problem of gesture positioning, the gesture positioning algorithm was proposed based on the method of combining gesture gravity center coordinates and maximum inscribed circle center coordinates to locate the gesture accurately and reliably and use the spherical threshold limit method to deal with abnormal gesture trajectory. The algorithm provides a strong guarantee for the follow-up high-precision gesture trajectory extraction, gesture tracking and gesture recognition. The experimental results proved the feasibility and validity of the improved algorithm.

Acknowledgements

Scientific and Technological Research Project Funds of Chongqing Municipal Education Commission (KJ130512)

References

[1] M. Jiang, K. Jiang, J. Kong, et al: Robust Hand Gesture Recognition Using Depth Image[J]. Journal of Computational Information Systems, Vol. 11 (2015) No.3, p.1093-1100.

[2] Y. Luo, C. He, Y. Wang, et al: A Novel Dynamic Hand Gesture Recognition Method Based on Depth Pre-segmentation Combining with Camshift Tracking Algorithm[J]. Semiconductor Optoelectronics, Vol. 36 (2015) No.1, p.155-159.

[3] J. Mai: Robotic Control Based on Natural Gesture (Ph.D., South China University of Technology, China 2013), p.12-14.

[4] Z. Li: Design of Remote Control of Manipulator Based on Gesture Recognition (Ph.D., Wuhan University of Technology, China 2014), p.22-29.

[5] W. Yanga, M. Wangc: Depth Information and Motion Context Feature Based Dynamic Hand Gesture Detection and Recognition[J]//Journal of Information and Computational Science, Vol. 12 (2015) No.10, p.3775-3782.

[6] T. Lee, T. Höllerer, AR. Handy: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking[C]// IEEE International Symposium on Wearable Computers. IEEE Computer Society (Boston, MA, USA, October 11-13, 2007). Vol. 16, p.83-90.

[7] Z. Ren, J. Yuan, J. Meng, et al: Robust Part-based Hand Gesture Recognition Using Kinect Sensor[J]. Multimedia, IEEE Transactions on, Vol. 15 (2013) No.5, p.1110-1120.

[8] S S. Rautaray, A. Agrawal: Real Time Hand Gesture Recognition System for Dynamic Applications[J]. International Journal of Ubicomp, Vol. 3 (2012) No.1, p.21.

[9] N H. Dardas, N D. Georganas: Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques[J]. IEEE Transactions on Instrumentation & Measurement, Vol. 60 (2011) No.11, p.3592-3607.

[10] S. Qin, X. Zhu, Y. Yang, et al: Real-time Hand Gesture Recognition From Depth Images Using Convex Shape Decomposition Method[J]. Journal of Signal Processing Systems, Vol. 74 (2014) No.1, p.47-58.
[11] C Y. Kao, C S. Fahn: A Human-machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion [J]. Procedia Engineering, Vol. 15 (2011) No.1, p.3739-3743.
[12] Z. Chen: Gesture Detection and Tracking Based on Depth Information (Ph.D., South China University of Technology, China 2012), p.20-29.
[13] S S. Rautaray, A. Agrawal: Vision Based Hand Gesture Recognition For Human Computer Interaction: a Survey[J]. Artificial Intelligence Review, Vol. 43 (2015) No.1, p.1-54.
[14] H. Wu, J. Wang: A Visual Attention-based Method to Address the Midas Touch Problem Existing in Gesture-based Interaction[J]. The Visual Computer, Vol. 32 (2016) No.1, p.123-136.
[15] S. Zhang, Y. Zhang, L. Kong: Self-occlusion Detection Approach Based on Depth Image[J]. Journal of Chinese Computer Systems, Vol. 31 (2010) No.5, p.964-968.