Large-Screen Interactive Technology With 3D Sensor Based on Clustering Filtering Method and Unscented Kalman Filter

LEI YU, (Member, IEEE), AND SHUMIN FEI

1School of Mechanical and Electric Engineering, Soochow University, Suzhou, China
2Jiangsu Smart Factory Engineering Research Center, Huaiyin Institute of Technology, Huai’an, China
3School of Automation, Southeast University, Nanjing, China

Corresponding author: Lei Yu (slender2008@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 61873176.

ABSTRACT In this paper, a large screen interactive 3D sensor technology with clustering filtering method and unscented Kalman filter has been addressed. The surface and interactive area with the screen is mapped together using image processing and machine vision processing of 3D sensor. Then, the predictable and accurate multi-point interaction can be obtained. Also some operations can be achieved such as zooming, panning, selecting and moving image components, etc. Vision method, infrared induction, dynamic capture, image processing and other mixed technologies have been simultaneously utilized to recognize human gestures and simulate mouse event which can give participants a fully interactive experience. Experimental results show that this large-screen interactive technology with 3D sensor can achieve accurate multi-point interaction and gesture control, and have good human-machine interaction effect.

INDEX TERMS 3D sensor, multi-point interactive, gesture control, human-computer interaction, unscented Kalman filter.

I. INTRODUCTION Interactive technology has been widely used in our life. Touch-screen technology is regarded as a node on the way leading to the ideal human-computer interaction. Apple’s iPhone as the representative product, touch screen technology has brought us a new and simple-to-use human interface, and it also brings fantastic control experience. Scientists predict that in the near future, as long as human beings through words, gestures, facial expressions and other input, the machine will know what to do, which is the ideal state of human-computer interaction [1]–[12]. Analyzing hand gestures is a comprehensive task involving motion modeling, motion analysis, pattern recognition, machine learning, and even psycholinguistic studies [5]. Like “Avatar” movie, as soon as the human press the start button, a three-dimensional holographic user interface can be presented before our eyes.

When press a random button, the program opens. When swipe our gesture randomly, data moves with our wish. Capturing and recognizing gestures, expressions and vocal inputs led to the evolution of human-computer interfaces giving rise to the development of the so-called natural user interfaces. These recent developments are investigated by Lamberti et al. In [7], where a generic framework for a seamless integration of multi-modal natural interaction capabilities is illustrated. As we know, 3D sensing sensors are often employed to capture images and video to realize human-computer interaction for intelligent systems [10], [13]–[15], [24]. 3D sensing devices are used as a tool for inferring human’s hands position, poses and gestures. The information from the sensors can translated into suitable commands for the control of virtually every kind of digital system. One of the main feature of sensing device-based human-computer interaction systems is the ability of tracking human gestures and poses in the field of 3D sensing sensor—Kinect [13]–[15]. Nowadays, new types of human-computer interaction based on Kinect have been focused on improving interactive feeling and experience.

Research direction of this paper is an important recent field of human-computer interaction, in accordance with the tendency of screen interactive field’s developments to the large screen interactive fields. In this paper, a large-screen interactive technology with clustering filtering method and
unscented Kalman filter based on 3D sensor has been presented. Also some operations can be achieved such as zooming, panning, selecting and moving image components, etc. Simulating mouse event can bring participants to enjoy a fully interactive experience, and the intuitive, three-dimensional, real-time multiplayer interactive way translates a new interactive feeling and experience.

II. LARGE-SCREEN INTERACTIVE TECHNOLOGY BASED ON THE 3D SENSOR

A. WORKING PRINCIPLE

The architecture of the large screen interactive system is designed in this paper (see Fig.1) which consists of hardware and software components. The hardware mainly includes the 3D Sensing sensor(such as Kinect), industrial machines, power supplies, electronic dogs, mosaic panel and other components. The software includes the interface software client software and other software. The Working principle of the interactive system is as follows: Through the 3D sensor image processing, machine vision processing, as well as flat-screen displays and interactive coverage area of the map together, the surface and interactive area with the screen is mapped. Then, the predictable and accurate multi-point interaction can be obtained. After gesture classification, we simulate the mouse and keyboard event and then show the interactive images by the projector. This proposed interactive system has a good human-machine interaction effect in Fig.2.

![FIGURE 1. The entire design of the system architecture diagram.](image1)

![FIGURE 2. Large-screen interactive system based on the 3D sensor.](image2)

B. IMPLEMENTATION METHOD OF LARGE-SCREEN INTERACTIVE SYSTEM BASED ON 3D SENSOR

This design implementation of large-screen interactive system has four key steps. Firstly, we need to create interactive area. Then confirm the touch event by the particle analysis algorithm which points are true “touch”, which is the interference. After the gesture classification, the surface and interactive area with the screen is mapped together to simulate mouse and keyboard events. Also some operations can be achieved such as zooming, panning, selecting and moving image components, etc. So a good human-machine interaction effect can be achieved. Specific implementation steps of the method are as follows:

**Step 1 (Create the Interactive Area):** Firstly, create an interactive-field interactive platform (eg Kinect) through 3D sensors. Kinect installation location needs to make sure Kinect depth camera covering the entire range of interactive areas. Kinect is also needed to ensure that the installation will not be moved and interfered. Even slight vibration will cause interference on the depth values, and reduce the accuracy of interaction. If a wall as an interactive surface, the above rules still apply, but need to be very careful to confirm the installation location of Kinect. Kinect scanning field of view face interaction can not be totally obscured by a user standing. In short, the interactive wall is the best way that the entire interaction region should be covered.

**Step 2 (Determine the Touch Event):** The development of the interactive project is on the basis of Microsoft’s boiler-plate code SkeletalViewer on. Environment source code is written by C++ and the NET framework4.0. From the Kinect depth data obtained at the frame, the depth datas are stored in the depth buffer. Then, according to the mounting Kinect position we find the roughly interactive surface [14]. Then the next subsequent frame is obtained, the depth of each frame of data will be compared with each other to interact with the surface in accordance with the pre-obtained in order to confirm whether there is “touch” event. Particle analysis algorithm can identify which point is true “touch”, or which is the interference [15]. Finally, we set up a filter, judging these frames of the points one by one to represent a gesture which acts on behalf of a mouse or keyboard events.

**Step 3 (Touch Mode—Mouse Control):** System mission mode is the normal operating mode, which is used to capture real-time tracking touch gestures, and translate gestures into win7 system above the corresponding mouse and keyboard events. Define a depth value Dmax, and all the depth value exceeds the upper limit are deemed not to fall within touching range. Also define a depth value of the lower limit Dmin, and all values below the lower limit of the depth are considered meaningless touch value. Narrow region between Dmax and Dmin effective area is designed as a touch point. Left-click and right-click is corresponding with the mouse. The event of Right-click is based on a feedback of the time-related gesture.

**Step 4 (Touch Gesture—Control Mode):** For the system idle mode, gestures filter function and classification parameters need to be initialized and be ready to accept different particle information. When the values of particles are larger than 1, it can be conformed as an effective gesture. In addition, the entire analog I/O event is reset (the mouse button is released). We use a time counter, used to avoid a false gesture is reset to case0 (noise frame). If the time counter value over three (already more than three), we consider the previous gesture has ended. Every gesture (including mobile gestures, zoom gestures, hand gestures, etc.) is saved in the previous record of the gesture.
III. FILTERING METHOD

The traditional data filtering method is based on the arithmetic weighted averaging method. The detailed steps are in the following: Firstly, filter out distances values that first enter operating area and remove the first P points of entry (P = 2). Secondly, take the first point of the P+1 point as the first point of the weighted calculation and compare the rest point, and then remove the maximum and minimum of it. Thirdly, the remaining N-P-2 data is performed with an arithmetic weighted averaging Y. This filtering method is easy to eliminate the deviation of the sampling value caused by the pulse interference, with the characteristics of convenient calculation, fast speed, small storage capacity etc. Specific filtering algorithm is as follows:

\[ Y = \frac{1}{N-P-2} \sum_{k=P+2}^{N-1} X(k) \]  

(1)

Notice: X (k) represents the real time data value of sensor acquisition and N represents the total number of data collected, generally N is greater than or equal to 8, and satisfied the condition:

\[ X(3) \leq X(4) \leq \cdots \leq X(k) \leq \cdots \leq X(N) \]  

(2)

In view of the real-time and accuracy requirements of 3D sensor tracking and localization, the adaptive threshold nearest neighbor clustering method is adopted in this paper. The point distance in the continuous point collection is relatively close, but the distance between the discontinuities is far away, and it is a collection of two different points. According to the characteristics of sequential scanning and spatial correlation of the data of 3D sensor, a simplified nearest neighbor clustering method is used [16]–[19]. If the distance between two points is less than the threshold value, the two points belong to the same point set. If they are larger than the threshold value, the two points belong to different point sets. At this point, the nearest neighbor clustering method is fast and effective. Specific procedures are as follows:

\( \text{i} \) Clustering initialization: \( i = 1, k = 1, C_k \) is point set clustering,

\( \text{ii} \) \( j = i, i = i + 1 \). The coordinates of two adjacent points are \( p_i = (x_i, y_i), p_j = (x_j, y_j) \). Where, \( x = l \cdot \cos \lambda, y = l \cdot \sin \lambda \). Then the distance between two adjacent points is \( R = \sqrt{(x_j-x_i)^2 + (y_j-y_i)^2} \).

\( \text{iii} \) If \( R < \delta_l \) is established, then \( p_i \in C_k \) and turn to (ii); If \( R > \delta_l \), then \( p_i \in C_k \) and \( k = k + 1, p_j \in C_k, i = i + 1 \), finally turn to (ii).

\( \text{iv} \) End of a frame 3D sensor data, turn to (i), then start next frame data.

The threshold selection is the key of the whole algorithm, which not only directly affects the clustering effect, but also affects the accuracy of the follow-up tracking and positioning. If the selected threshold is too small, the sensor data will be divided fragmented. Though the precision of judging will be higher, the real-time detection will discount [18]; On the contrary, if the threshold is too big, it is very difficult to detect small obstacles, the possibility of missing data will be increased [18]–[21].

Unscented Kalman filter(UKF) abandons the method of linearization of the nonlinear function, but to use a linear Kalman filter framework [20]–[25]. For one-step prediction equation, unscented transformation is used to solve nonlinear transformation of covariance and mean. Unscented Kalman filter can achieve great performance in bearings-only target tracking. Therefore, Unscented Kalman filter is used in this paper to realize further data filtering. Specific steps are as follows:

1) Using the formula (3) and (4) to obtain a set of sample points, and the weights can be calculated with equation(5):

\[
\begin{align*}
X^{(0)} & = \bar{X}, \\
X^{(i)} & = \bar{X} + (\sqrt{(n+\lambda)}P)_i, \quad i = 1 \sim n \\
X^{(i)} & = \bar{X} - (\sqrt{(n+\lambda)}P)_i, \quad i = n + 1 \sim 2n \\
w^{(0)} & = \lambda/\sqrt{n+\lambda} \\
w^{(i)} & = \frac{(1-a^2+\beta)}{n+\lambda} + \frac{\lambda}{\sqrt{2(n+\lambda)}}, \quad i = 1 \sim 2n
\end{align*}
\]

(3)

(4)

where, \( \lambda \) is the scaling parameter.

\[
X^{(i)}(k|k) = [\bar{X}(k|k)\hat{X}(k|k) + \sqrt{(n+\lambda)}P(k|k)] \\
\times [\bar{X}(k|k) - \sqrt{(n+\lambda)}P(k|k)] 
\]

(5)

2) one-step prediction of 2n + 1 Sigma point set can be obtained with equation(6):

\[
X^{(i)}(k+1|k) = f[k, X^{(i)}(k|k)] 
\]

(6)

3) One-step prediction and covariance matrix of the digital data can be calculated with equation(7) and (8):

\[
\hat{X}(k+1|k) = \sum_{i=0}^{2n} w^{(i)}(k+1|k) \\
= \sum_{i=0}^{2n} w^{(i)}(\hat{X}(k+1|k) - X^{(i)}(k+1|k)) + Q 
\]

(7)

(8)

4) According to one-step prediction value, unscented transformation can be used to generate sigma point set;

5) Sigma point set generated in step4 can be inserted into the observation equation to achieve the prediction of observation, as shown in Equation (9):

\[
Z^{(i)}(k+1|k) = h[X^{(i)}(k+1|k)] 
\]

(9)

6) Predicted mean and covariance of the system can be calculated with predicted observation of sigma point set.
obtained in step 5 by weighted summation, as shown in Equations (10)-(12):

\[
\tilde{Z}(k+1|k) = \sum_{i=0}^{2n} w^{(i)} Z^{(i)}(k+1|k) 
\] (10)

\[
P_{\tilde{Z}Z}(k+1|k) = \sum_{i=0}^{2n} \left[ Z^{(i)}(k+1|k) - \tilde{Z}(k+1|k) \right]^T R + \sum_{i=0}^{2n} \left[ Z^{(i)}(k+1|k) - \tilde{Z}(k+1|k) \right] 
\] (11)

\[
P_{\tilde{Z}X}(k+1|k) = \sum_{i=0}^{2n} w^{(i)} [X^{(i)}(k+1|k) - \tilde{Z}(k+1|k)] 
\] (12)

⑦ Kalman gain array can be obtained with Equation (13):

\[
K(k+1) = P_{\tilde{Z}Z}(k+1|k) P_{\tilde{Z}Z}^{-1}(k+1) 
\] (13)

⑧ Status updates and covariance update of the system can be calculated with Equation (14) and (15).

\[
\hat{X}(k+1|k+1) = \hat{X}(k+1|k) + K(k+1)[Z(k+1) - \hat{Z}(k+1|k)] 
\] (14)

\[
P(k+1|k+1) = P(k+1|k) - K(k+1)P_{\tilde{Z}Z}(k+1|k) K^T(k+1) 
\] (15)

In this paper, the adaptive threshold nearest neighbor clustering algorithm can effectively remove the noise, run fast, and improve the interaction effect. The man-machine interaction system has been used to analyze the continuous and effective distance values. The coordinate points are calculated by the digital filtering method, which is the coordinates of the hand that is used to operate. Then the man-machine interactive control will be realized. At the same time, it can be used to count the effective coordinate points of the scanning area. The mean value coordinates will be calculated, when the operator leaves the scanning range. At the end of this set of data scanning, coordinate parameters reset.

IV. EXPERIMENT RESULTS

This paper has developed a set of interactive media solutions that shows the advantages of interactive equipment perfectly, which can realize the precise interaction and gesture control. Combining special high precision optical induction and photoelectric analog signal, it can be used in any environment to achieve a variety of interactive scene, and strive to become the leading mode of information dissemination in the future.

In this section, the installation position of the experimental system of the 3D sensor Kinect needs to ensure that the depth of the Kinect camera covers the entire interactive area, and ensure that, it will not be moved and interfered after installation. Based on the method of large screen interactive system mentioned above and filtering method with 3D sensor, this section conducted two experiments, named the single point filtering experiment and the multi point filtering experiment, comparing with the traditional filtering method and filtering method.

A. SINGLE POINT FILTERING EXPERIMENT

In Fig. 3, the red solid square represents the sensor, and the small red rectangle frame represents the effective operation area, only the point in which it can be effectively operated. Experiments are performed on the same object at the same position, and the only change place is the different filtering algorithm experiments. Fig. 3 (a) is the original image. Fig.3 (b) is the single point filtering effect by the traditional filtering method. Fig. 3 (c) is the single point filtering effect by the proposed filtering method.

![Figure 3](image-url)
From Fig.3, we can see that there are some isolated noise points before the single point filter, and the single point of the noise is large, and the accuracy is not well. After filtering with the traditional filtering method, the isolated noise points are basically filtered out, but the location is still not very accurate. Employing the integrated filter proposed in this paper, the single point is more accurate and the jitter noise is very small. The results show that the presented clustering algorithm is effective.

**B. MULTI-POINT FILTERING EXPERIMENT**

Also, in Fig.4, the red solid square represents the sensor, and the small red rectangle box represents the effective operating area. The point in the red rectangle can be only effective. The experiment for the same object is used to operate at the same position, only change for the different experimental filtering algorithms. Fig.4 (a) is the original image. Fig.4 (b) is the multi point filtering effect by the traditional filtering method, and Fig.4 (c) shows the effect of multi point filtering by the proposed filtering method in the paper.

From Fig.4, before multi-point filtering, there are a lot of noises and the jitter noise is very large. After filtering by the traditional filtering method, the isolated noise points around the multi points can be eliminated, and the denoising effect can be improved. By using the proposed method in this paper, we can remove the noise from the interaction region, retain the useful signal, and improve the signal-noise ratio, so that the multi points can be located precisely. Fig.5 denotes the interactive picture of single point in the actual situation, and Fig.6 is the actual effect diagram of large screen and multi points, analyzed by method mentioned above. It supported gesture control and reached a relatively ideal interactive effect in accuracy, stability and anti disturbance and other aspects. In summary, a large number of experiments demonstrate that the proposed method is effective for the large screen human-computer interaction system based on 3D sensor, and can filter out the noise points and have good multi point interaction effect.

**FIGURE 4.** (a) The original image. (b) the multi-point filtering effect by the traditional filtering method. (c) the effect of multi-point filtering by the proposed filtering method.

**FIGURE 5.** Interactive picture of single point in the actual situation.

**FIGURE 6.** Interactive picture of multi-point in the actual situation.
V. CONCLUSION
This paper has presented a large screen interactive technology system based on clustering filtering method and unscented Kalman filter, using the 3D Sensing sensor as a man-machine interactive sensor. By image processing and machine vision processing of 3D sensor, as well as mapping the display plane and the interaction range to the screen, predictable and accurate multi point interaction can be achieved. Also, some wonderful operations such as zoom, translation, image components selection and mobile can be realized, and design collocation, stability and versatility can also be ensured the best. Research and development of this large screen interactive technology system based on clustering filtering method and unscented Kalman filter method can have the natural interactive effect, infectious human-computer interaction surface and plentiful interactive contents. The proposed interactive system has greater economic benefits and can promote the significance, which can be widely used in entertainment games, digital signage, and other air control, etc.

ACKNOWLEDGMENT
The work is supported by the National Natural Science Foundation of China (No.61873176). Natural Science Foundation of Jiangsu Province, China (BK20181433); The open fund for the Jiangsu Smart Factory Engineering Research Center. The authors would like to thank the referees for their constructive comments.

REFERENCES
[1] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, “Human-computer interaction using eye-gaze input,” IEEE Trans. Syst., Man, Cybern., vol. 19, no. 6, pp. 1527–1534, Nov./Dec. 1989.
[2] L. A. Alexandre, “Gender recognition: A multiscale decision fusion approach,” Pattern Recognit. Lett., vol. 31, no. 11, pp. 1422–1427, Aug. 2010.
[3] M. Avlonitis, “On the problem of early detection of users interaction outbreaks via stochastic differential models,” Erg. Appl. Artif. Intell., vol. 51, pp. 92–96, May 2016.
[4] P. Viola and M. J. Jones, “Robust real-time face detection,” Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, May 2004.
[5] J. Shin and C. M. Kim, “Non-touch character input system based on hand tapping gestures using kinect sensor,” IEEE Access, vol. 5, pp. 10496–10505, 2017.
[6] S. L. Phung and A. Bouzerdoum, “A pyramidal neural network for visual pattern recognition,” IEEE Trans. Neural Netw., vol. 18, no. 2, pp. 329–343, Mar. 2007.
[7] J. S. Kim, S. J. Yun, D. J. Seol, H. J. Park, and Y. S. Kim, “An IR proximity-based 3D motion gesture sensor for low-power portable applications,” IEEE Sensors J., vol. 15, no. 12, pp. 7009–7016, Dec. 2015.
[8] H. Hasan and S. Abdul-Kareem, “Retraction note to: Human-computer interaction using vision-based hand gesture recognition systems: A survey,” Neural Comput. Appl., vol. 28, no. 4, p. 849, Apr. 2017.
[9] G. Paravati and V. Gatteschi, “Human–computer interaction in smart environments,” Sensors, vol. 15, no. 8, pp. 19487–19494, Aug. 2015.
[10] J. Palacios, C. Sagüés, E. Montijano, and S. Llorente, “Human–computer interaction based on hand gestures using RGB-D sensors,” Sensors, vol. 13, no. 9, pp. 11842–11860, Sep. 2013.
[11] W. Zhang, M. L. Smith, L. N. Smith, and A. Farooq, “Gender and gaze gesture recognition for human-computer interaction,” Comput. Vis. Image Understand., vol. 149, pp. 32–50, Aug. 2016.
[12] A. Pimenta, D. Carneiro, J. Neves, and P. Novais, “A neural network to classify fatigue from human-computer interaction,” Neurocomputing, vol. 172, pp. 413–426, Jan. 2016.
[13] T. Mallick, P. P. Das, and A. K. Majumdar, “Characterizations of noise in Kinect depth images: A review,” IEEE Sensors J., vol. 14, no. 6, pp. 1731–1740, Jun. 2014.
[14] G. Chen, J. Li, B. Wang, J. Zeng, G. Lu, and D. Zhang, “Reconstructing 3D human models with a Kinect,” Comput. Animation Virtual Worlds, vol. 27, no. 1, pp. 72–85, Jan. 2016.
[15] S.-C. Hsu, J.-Y. Huang, W.-C. Kao, and C.-L. Huang, “Human body motion parameters capturing using Kinect,” Mach. Vis. Appl., vol. 26, nos. 7–8, pp. 919–932, Nov. 2015.
[16] S. Mehrkanoon, O. M. Agudelo, and J. A. Suykens, “Incremental multi-class semi-supervised clustering regularized by Kalman filtering,” Neural Netw., vol. 71, pp. 88–104, Nov. 2015.
[17] W. Zhou and J. Hou, “A new adaptive robust unscented Kalman filter for improving the accuracy of target tracking,” IEEE Access, vol. 7, pp. 77476–77489, 2019.
[18] D.-S. Jang, S.-W. Jang, and H.-I. Choi, “Tracking a partially occluded target with a cluster of Kalman filters,” Int. J. Intell. Syst., vol. 17, no. 6, pp. 595–607, Jun. 2002.
[19] P. Zhang, W. Qi, and Z. Deng, “Hierarchical fusion robust Kalman filter for clustering sensor network time-varying systems with uncertain noise variances,” Int. J. Adapt. Control Signal Process., vol. 29, no. 1, pp. 99–122, Jan. 2015.
[20] W. Li, G. Wei, F. Han, and Y. Liu, “Weighted average consensus-based unscented Kalman filtering,” IEEE Trans. Cybern., vol. 46, no. 2, pp. 558–567, Feb. 2016.
[21] H. M. T. Menegar, J. Y. Ishihara, G. A. Borges, and A. N. Vargas, “A systematization of the unscented Kalman filter theory,” IEEE Trans. Autom. Control, vol. 60, no. 10, pp. 2583–2598, Oct. 2015.
[22] S. Ting Goh, S. A. Reza Zekavat, and K. Pahlavan, “DOA-based endoscopy capsule localization and orientation estimation via unscented Kalman filter,” IEEE Sensors J., vol. 14, no. 11, pp. 3819–3829, Nov. 2014.
[23] O. Straka, J. Duník, and M. Šimandl, “Unscented Kalman filter with advanced adaptation of scaling parameter,” Automatica, vol. 50, no. 10, pp. 2657–2664, Oct. 2014.
[24] W. Zhou and J. Hou, “A new adaptive high-order unscented Kalman filter for improving the accuracy and robustness of target tracking,” IEEE Access, vol. 7, pp. 118484–118497, 2019.
[25] A. Muruganathan, K. C. Tan, and P. Vadakkepat, “Evolutionary dynamic multiobjective optimization via Kalman filter prediction,” IEEE Trans. Cybern., vol. 46, no. 12, pp. 2862–2873, Dec. 2016.