IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 22, 2020, accepted July 3, 2020, date of publication July 10, 2020, date of current version August 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3008435

An Improved Moving Tracking Algorithm With Multiple Information Fusion Based on 3D Sensors

YIFAN FANG[1], LEI YU[1], (Member, IEEE), AND SHUMIN FEI[2]
[1]School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215000, China
[2]School of Automation, Southeast University, Nanjing 215096, China

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

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

ABSTRACT The continuously adaptive mean shift algorithm suffers from the tracking offset phenomenon while tracking targets with colors similar to that of the background. In this paper, to improve the performance of this algorithm, the depth information is combined with the back-projection color image and the information from the moving prediction algorithm. The target search window is predicted based on switching filtering algorithm with the Extended Kalman Filter (EKF) method. The moving tracking synthesis algorithm which used 3D sensors and combines color, depth and prediction information is used to solve the problems that the continuously adaptive mean shift algorithm encounters, namely disturbance and the tendency to enlarge the tracking window. The experimental results show that the proposed algorithm can accurately track a moving target in the presence of a complex background, and greatly improves the interference resistance and robustness of the system.

INDEX TERMS Continuously adaptive mean-shift, switching filtering algorithm, extended Kalman filter, moving tracking algorithm, 3D sensor.

I. INTRODUCTION

Moving tracking or motion tracking is a popular research topic in the field of computer vision. It has several applications in intelligent human-computer interaction, medical diagnosis, intelligent monitoring and military domains. To this purpose, several target tracking algorithms have been developed in engineering fields. Techniques such as the target tracking algorithm based on template matching, TLD (Tracking-Learning-Detection) target tracking algorithm, Mean Shift, Mode Seeking, and Clustering and continuous adaptive mean shift algorithm, have been developed and applied in the field of motion tracking [1]–[5]. The continuous adaptive mean shift algorithm (hereinafter referred to as Camshift) uses the color histogram back projection to obtain the probability distribution for tracking the movement of a target. The advantage of this algorithm is that it is less sensitive to the target deformation and can adaptively adjust the tracking window to continuously track a target with simple backgrounds. Due to its low complexity, small computation and good real-time performance, this algorithm is widely used for tracking moving targets. However, one of the shortcomings of this algorithm is that when the color or hue of the tracked target is similar to the background of the object, tracking errors are likely to occur. In such cases where tracking must be performed in the presence of a complex background or interference, the tracking based on the Camshift may be inaccurate or the target may be lost. In recent years, many researchers have made several improvements to the Camshift algorithm to improve its performance. For example, Hsia *et al.* proposed an indirect prediction Camshift algorithm based on Adaptive Search Pattern (ASP), which improves the accuracy and robustness of the target tracking task [6]. Camshift algorithm with the particle filter algorithm is presented to improve target tracking [7], [8]. Other researchers have combined the Camshift algorithm with particle filter, texture information and edge information to improve the robustness of the algorithm [9], [10].

Inspired by the improvements to the Camshift algorithm which aim to solve the problem of tracking errors in the presence of backgrounds with similar colors, an improved version of Camshift is proposed in this paper which performs

The associate editor coordinating the review of this manuscript and approving it for publication was Sunil Karamchandani.

synthesis of moving tracking with the help of a 3D sensor to combine the color, depth and prediction information. Firstly, a kinematic infrared (IR) camera is employed to obtain the depth information of the moving target. Then this depth information is fused with the Camshift color information. In addition, a simple switching mechanism is added to the system. The system automatically switches algorithms based on the characteristics of the image. The switching system is a concept in the field of control engineering, and 'switching' as a control idea comes from the switching servo system. In actual engineering control, due to various external intrinsic factors, the structure of the system changes during operation. Such a system can be regarded as a 'switching system' and modeled [11]. The application of switching control technology can not only solve the basic problems of the variable structure system, but also improve the transient performance of the system, which meets the needs of the development of intelligent control technology. Combining the switching filtering algorithm with the filtering algorithm can improve the efficiency of the entire interactive system [12], [13]. When the system detects that the image signal does not have much interference, it automatically switches to the adaptive median filtering algorithm to filter the image under the effect of the switching rule, which improves the system's operating efficiency [14]. Next, a switching filtering algorithm with the Extended Kalman Filter is used to predict the direction of the moving target. The location of the search window is dynamically adjusted to integrate the two algorithms. Experimental simulation results show that the algorithm proposed in this paper can accurately track moving targets under complex backgrounds. The algorithm also greatly improves the anti-interference ability and the robustness of the system.

This paper is organized as follows. Section II describes a fusion algorithm of 3D sensor depth information and color information. Section III presents an EKF predictive information fusion algorithm based on 3D sensor. To evaluate the proposed the performance of moving tracking technology with multiple information fusion, simulation results appear in Section IV, and the proposed method is respectively compared with the traditional Camshift algorithm and the original machine learning based RGB-D algorithm of the system. Finally, Section V gives the conclusions.

II. AN ALGORITHM FOR FUSING 3D DEPTH SENSOR INFORMATION WITH COLOR INFORMATION

The Camshift algorithm extracts the HUE component from the target HSV space as the eigenvalue, and distinguishes the target from the background based on the color difference [15]. When the colors or the hues of the target and the background are similar, it is difficult to distinguish between the two. The Camshift algorithm uses the color histogram to describe the target, which greatly weakens the anti-interference ability of the Camshift algorithm. Color images can contain a lot of noise due to variations in color and hue [16], while depth images are less sensitive to light, color, texture, etc. Therefore, depth images contain much less noise than color images and hence are more usable for object tracking. In this paper we aim to integrate the information from both depth and color to develop a multi-feature tracking algorithm with improved tracking capability. The improvement of this step of combining the two approaches is shown in Figure 1.

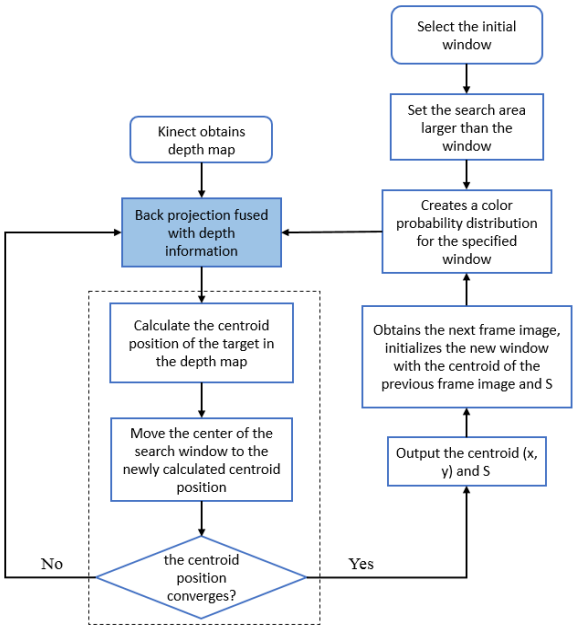


FIGURE 1. The algorithmic flow fusing depth information of kinect.

There are many devices and methods which can be used to obtain depth images. Among them, the Kinect with their optical coding technology provide a relatively inexpensive and reliable way to use off-the-shelf available components for our experiments [17], [18]. Kinect is RGB-D camera. It is possible to obtain a backward projection pattern of the color histogram and fuse it with the depth. The back projection of the depth information of the fused algorithm is based on the original back projection image, with the depth data being used as the weight for filtering. The specific process is explained as follows: Two images are acquired sequentially. Let the distance that the tracked target has moved between the two images be m. The distance from the center of the current target to the camera is d. Map the pixels whose value is in the range of $(d - m, d + m)$ in the current depth map to the back projection with a weighting factor of 1. That is, if the original value of a, after processing will become $a * 1$. Map the pixels whose value is in the range of $(d - 2m, d - m) \cup (d + m, d + 2m)$ in the current depth map to the back projection with a weighting factor of 0.8. Set the pixels whose value is in any other range to 0. For this, the corresponding weight coefficient is selected according to the size of the depth offset value, and the weight coefficient can be modified according to the actual situation. Now the Camshift algorithm is used to track the target on the colored back projection map of the fusion depth information.

As represented by the dashed box in the algorithm in Figure 1, let $I(x, y)$ be the pixels in the search window.

We use the Mean Shift iterative algorithm to obtain the zero-order and the first-order matrices [17]:

$$M_{00} = \sum_x \sum_y I(x, y),\ M_{10} = \sum_x \sum_y xI(x, y),$$
$$M_{01} = \sum_x \sum_y yI(x, y) \quad (1)$$

The centroid position of the search window is given by:

$$(x_c, y_c) = (\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}) \quad (2)$$

Next, we adjust the size of the search window according to the matrix M_{00} in equation (1), and reposition the center of the window to the centroid of the window. In the new image frame, the position and size of the new search window are again set based on the centroid position of the window and the M_{00} obtained from the previous image frame. This whole process is executed iteratively to achieve the goal of continuous tracking of the object through multiple frames.

When the gesture image is not very complicated, this algorithm can automatically switch to the weighted median filtering algorithm, which reduces the system operation amount and improves the system response speed.

Pixel p in image I shows a local window $R(p)$ centered on p and radius R. Different from the ordinary median filtering algorithm, for each pixel q belonging to $R(p)$, there is a weight coefficient w_{pq} based on the similarity of the corresponding feature image, as shown in equation 3:

$$w_{pq} = g(f(p), f(q)) \quad (3)$$

$f(p)$ and $f(q)$ are the eigenvalues of pixels p and q in the corresponding feature map. g is a weight function, the most used is the Gaussian function, which reflects the similarity of the pixels p and q.

Use $I(q)$ to represent the pixel value of pixel q, and the total number of pixels in the window is represented by n, then $n = (2r + 1) \times (2r + 1)$. Then the pixel value and weight value in the window constitute a pair sequence, that is, the sequence is sorted according to the value of $I(q)$. After sorting, the weighted weights are accumulated in sequence until the accumulated weight is greater than half of the weighted weight, and the corresponding $I(q)$ is used as the new pixel value of the center point of the local window.

$$p^* = \min k \quad \text{s.t.} \sum_{q=1}^{k} w_{pq} \geq \frac{1}{2} \sum_{q=1}^{n} w_{pq} \quad (4)$$

where, p^* is the filtered pixel point. Compared with ordinary median filtering, weighted median filtering has more feature maps and weight function terms during processing, making the filtering effect better.

III. AN EKF PREDICTIVE INFORMATION FUSION ALGORITHM WITH SWITCHING FILTER

The default switching filtering algorithm does not have the movement prediction module and the mechanism for updating the target feature. If the background color is close to the color of the tracked target, there can be occlusion or other types of interferences, and the calculated target area will be expanded and the search window will be enlarged. Therefore, in the process of target tracking, the search algorithm employed needs to estimate the future position of the target and narrow the search range. Common predictive algorithms used in these scenarios are Kalman Filter and the Extended Kalman Filter [19], [20]. The Extended Kalman Filter (EKF) is studied to apply the Kalman Filter theory to the non-linear field. In case of tracking a moving object, the tracking system model is generally non-linear in nature. Inspired by the switching system, this paper proposes a switching filter algorithm with the Extended Kalman filter to predict the position of the moving target. The switching system process is shown in Figure 2.

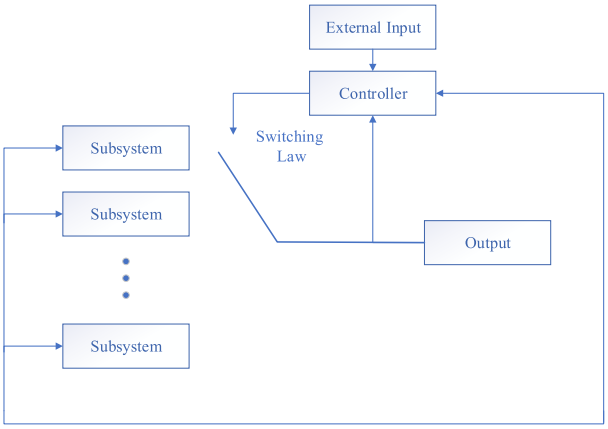


FIGURE 2. Flow chart of switching filtering system.

In general, a switching system consists of several subsystems and a switching rule. The switching rule can be likened to a switching device, which determines which subsystem the system will switch to at each moment, and only one subsystem can be activated at each time. Its mathematical model is:

$$\delta x(t) = f_\sigma(x(t), u(t), v(t)) \quad (5)$$
$$y(t) = g_\sigma(x(t), w(t)) \quad (6)$$

where, $x(t) \in R^s$ is the state vector of the system, $u(t) \in R^m$ is the input vector of the system, $y(t) \in R^l$ is the measurement output, $v(t) \in R^s$ and $w(t) \in R^q$ are external signal disturbances, σ is a segmented continuous signal function, the value range is N={1,...,N}, N is the number of subsystems. $f_i, i \in N$ is a vector field, $g_i, i \in N$ is a vector value function, and for a continuous time system, δ represents a differential operation $\delta x(t) = \dot{x}(t)$. For discrete time systems, δ represents the forward differential operation $x(t) = x(t + 1)$.

Each subsystem model can be expressed as:

$$\delta x(t) = f_j(x(t), u(t), v(t)), x(t_0) = x_0 \quad (7)$$
$$y(t) = g_j(x(t), w(t)), j \in N \quad (8)$$

The meaning of the switching signal $\sigma_0(t)$ is:

$$\sigma_0(t) = x_0; (i_0, t_0), (i_1, t_1), \ldots, (i_j, t_j), \ldots, \quad |t_j \in M, j \in N \tag{9}$$

where, e is the initial state of the system, t_0 is the initial time of the system. When $t_j \leq t < t_{j+1}$, the i_jth subsystem of the system is activated, and the trajectory of the system is generated by the i_jth subsystem. Assume that the system state does not jump at the moment of switching, and that there is a finite number of transitions in the interval [0, t], defining $T_{\min} = \min\{T = t_k - t_{k-1}\}$ as the minimum time interval of the i_jth subsystem. At a given time, it is assumed that any switching occurs at time $M^{def} = \{t_1, t_2, \ldots, t_j\}$.Since the time interval between successive switchings cannot be less than $T_{\min}$, and for any fixed switching law, finite time escaping is unlikely to occur, that is, under given initial conditions, the system cannot deviate from steady state for a limited time. That is, the system is stable.

In Figure 3, the process flow which demonstrates the idea of using EKF to improve the switching filtering algorithm is visualized [21], [22].

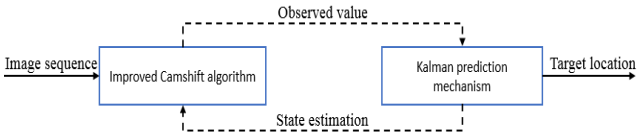


FIGURE 3. The idea of using EKF prediction mechanism.

Assuming that both the system noise and observation noise are white noise, the system state equation and the measurement equation are chosen based on EKF as follows:

$$X(k) = F(k|k-1)X(k-1) + G(k-1)W(k-1) \tag{10}$$

$$Z(k) = h(X(k)) + V(k) \tag{11}$$

where $X(k) \in R^{n\times 1}$ is the n-dimensional target state vector and $Z(k) \in R^{m\times 1}$ is the m-dimensional measurement vector. $F(k|k-1) \in R^{n\times m}$ and $G(k) \in R^{n\times p}$ are the state transition matrix and the input matrix, respectively. $W(k) \in R^{p\times l}$ and $V(k) \in R^{m\times l}$ are the state noise and the measurement noise respectively. $h(\cdot)$ is the observation function which can be expressed as follows:

$$h(X(k)) = \begin{bmatrix} r(k) \\ \theta(k) \end{bmatrix} = \begin{bmatrix} \sqrt{x^2(k) + y^2(k)} \\ \arctan \frac{y(k)}{x(k)} \end{bmatrix} \tag{12}$$

where $r(k)$ is the distance between the benchmark reference image (0, 0) and the target centroid. The points are expressed in the polar coordinate, where $\theta(k)$ is the azimuth in the polar coordinate. We use the Taylor series to expand $h(X(k))$ for the predictive value of $\hat{X}(k|k-1)$, and omit any component greater than the second higher order and subsequently linearize it to get the following:

$$H(k) = \frac{\partial h}{\partial X}|_{x=\hat{x}(k|k-1)} = \begin{bmatrix} \frac{\hat{x}(k|k-1)}{\hat{r}} & \frac{\hat{y}(k|k-1)}{\hat{r}} & 0 & 0 \\ -\frac{\hat{y}(k|k-1)}{\hat{r}} & \frac{x(k|k-1)}{\hat{r}} & 0 & 0 \end{bmatrix} \tag{13}$$

$$\hat{r} = \sqrt{[\hat{x}(k|k-1)]^2 + [\hat{y}(k|k-1)]^2} \tag{14}$$

The extended Kalman Filter is used recursively to achieve a state optimization. The recursive formulas used are shown as follows. It consists of a prediction equation expressed as:

$$\hat{X}(k|k-1) = F(k|k-1)\hat{X}(k-1) \tag{15}$$

The camera on the Microsoft Kinect can record up to 30 images per second. Since the time interval between adjacent shots is short, it can be assumed that the two frames are uniformly moving. The time interval of two adjacent images is taken as the sampling time interval Δt, and the position and velocity of the target are taken as the movement parameters of the target.

The movement state vector and the observed state vector of the target are defined in the following:

$$X(k) = [x(k), y(k), v_x(k), v_y(k)]^T \tag{16}$$

$$z(k) = [x(k), y(k)] \tag{17}$$

where $x(k)$ and $y(k)$ are the positions of the target's centroid on the axis, and $v_x(k)$ and $v_y(k)$ are the velocities of the target movement.

The switching filtering algorithm combined with the EKF realization process is as follows:

1) Select the size and position of the initial window as input to the switching filtering algorithm and calculate the target centroid as the observed value of the EKF $Z(k)$.
2) At time k in the tracking process, the optimal estimate $\hat{X}(k-1|k-1)$ of the target position at the last time (k-1) is substituted in the state equation (12) to obtain the optimal prediction estimate $\hat{X}(k|k-1)$ of $X(k)$.
3) $Z(k)$ is used to correct the estimated estimate $Z(k)$, that is, $Z(k)$ is brought into state filter formula to obtain the optimal estimate of $\hat{X}(k|k)$.

To estimate the target position at the next moment (time k+1), the input of the switching filtering algorithm uses the target centroid position component of $\hat{X}(k|k)$.

The flowchart for the final algorithm for the synthesis of the switching moving tracking algorithm which combines the depth information and the EKF prediction mechanism is shown as Figure 4.

IV. EXPERIMENT RESULTS

The algorithm of the improved system is employed by using the human-computer interaction platform set up by the research group (as shown in Figure 5). The platform was programmed using the OpenCV library and the Visual C++ 6.0 framework. It integrates: a) real-time data acquisition, b) filtering, c) display, d) algorithm simulation and e) control.

Figure 6 shows the overview of CPU and GPU steps of multi-information fusion CamShift. The CPU starts the tracking process, reads each frame of the image, and sets the initial search window range of the target area. During the operation, the input image needs to be converted into the

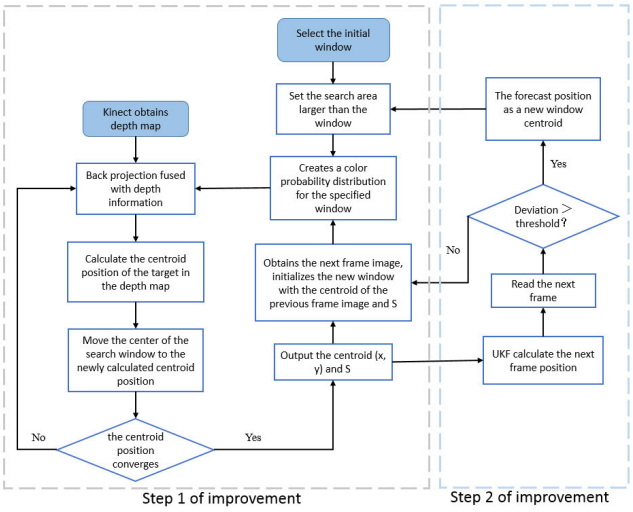


FIGURE 4. The algorithmic flowchart for 3D sensor moving tracking.

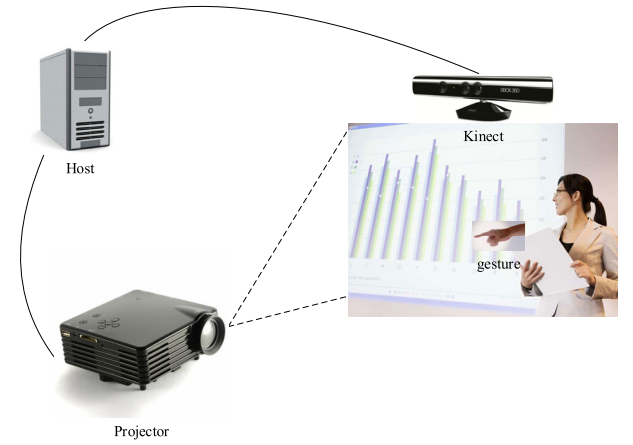


FIGURE 5. Human-computer interaction platform.

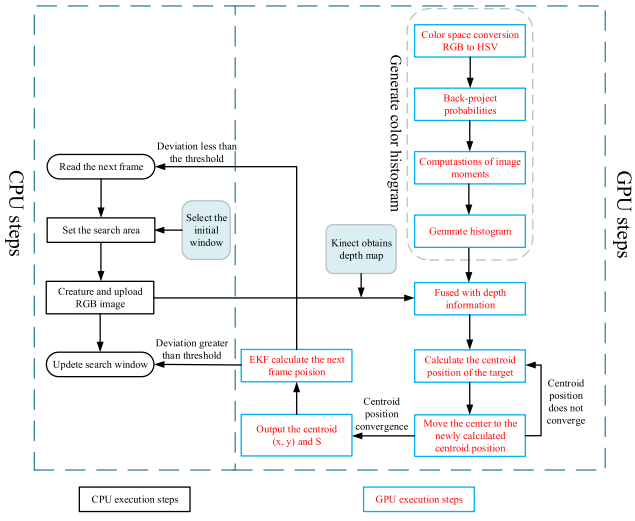


FIGURE 6. Overview of CPU and GPU steps of proposed method.

HSV color space. This work is completed by the GPU. Then the original input image is converted into a color probability distribution image based on the obtained color histogram (i.e., "back projection"). The time-consuming operations such as depth information and color information fusion, Mesnshift iteration, and EKF prediction of the next frame position are all performed by the GPU. The CPU computes the updated position and size [21]. If the deviation between the predicted position of the EKF and the actual position of the next frame is greater than the threshold, the size of the search window is updated; otherwise, it indicates that the tracking is accurate and the next frame of image is Read.

The Kinect sensor is placed above the projection surface and a projector is placed in front of it. The wall is used as an interactive projection plane, which enables the user to perform interactive operations Figure 6. Self-developed human-computer interaction platform based on the somatosensory perception. The monitor can also be connected directly to the host using a liquid crystal display panel, eliminating the need for a projector. Both displays are useful in the following experiments. The targets were people wearing black clothes and the disruptors were black boxes. They were both located in the monitoring area below the Kinect sensor. The comparison charts of the tracking effect are given. The experimental setting under which the user interacts with the system is shown in Figure 7.

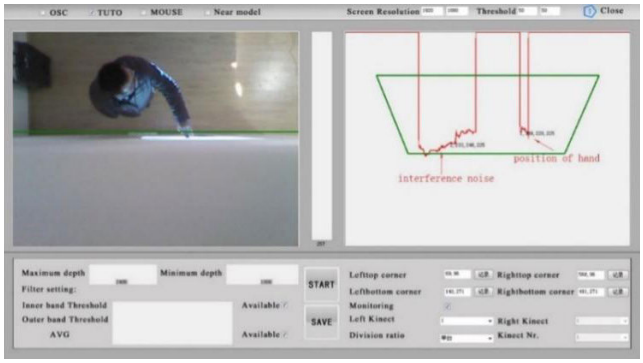


FIGURE 7. Self-developed human-computer interaction platform.

A. COMPARISION OF SINGLE EXPERIMENTAL RESULTS

The tracking experiment was conducted in an environment where indoor light and sunlight mixed. First, the default algorithm based on the color histogram is used to track and intercept the 13, 23, 33, 43, 53 frames of the tracking video. The effect is shown in Figure 8.a. As shown in the figure, at the beginning, the target body was within the tracking window. In the 33rd frame as the person approached the box similar in color to the subject, the tracking window was offset and could no longer follow the target. Next, use the algorithm proposed in this paper and then perform tracking experiments to track the target in the same scenario. As with the default algorithm, screenshots of the tracking situation at different distances from the target to the interferer are extracted, which are also 13, 23, 33, 43, and 53 frames. The comprehensive algorithm was used to combine the depth information and the forecast information proposed in this paper with the intercept.

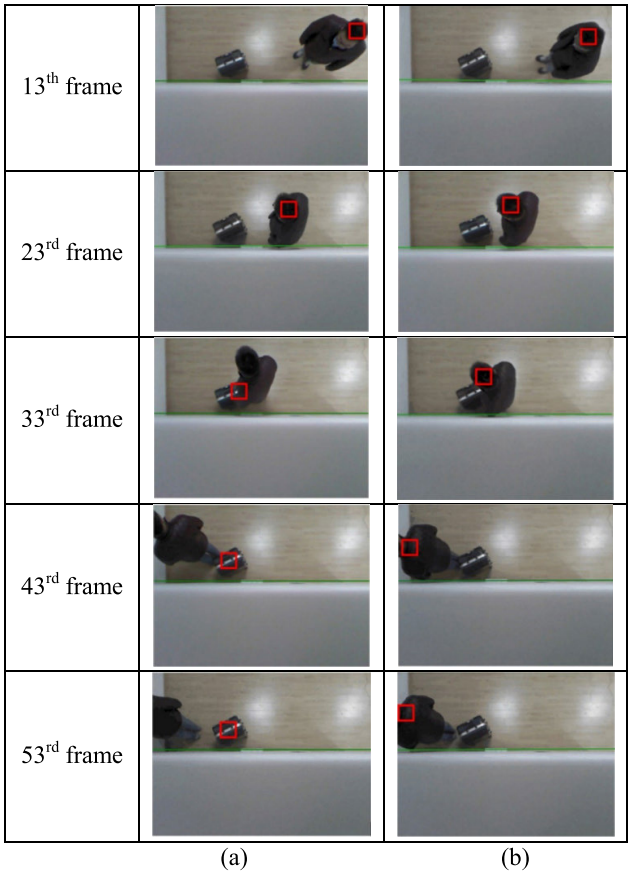


FIGURE 8. Comparison of tracking results (a) Tracking results using the default algorithm (b) Results of tracking after switching filter algorithm.

The results are shown in Figure 8.b. When the user's head and clothes are all black, the system can accurately locate the target on the head. When the person approached the interfering object of a similar color, the tracking window did not shift, and the algorithm was able to successfully track the target body throughout the whole process. The results of this single experiment demonstrate that the algorithm proposed is more robust than the original tracking algorithm.

B. COMPARISION OF NUMEROUS EXPERIMENTAL RESULTS

To ensure the reliability of the experimental results, the tracking experiments were repeated one hundred times. All the experiments were done with both the original algorithm and the improved algorithm under the same experimental conditions for each.

The statistical results for the successful tracking are shown in Table 1. Statistics show that the original algorithm's success rate of tracking the target in each frame of the image is 37%, the success rate of the improved algorithm is 92%, and the success rate of the comprehensive filtering algorithm is 96%. The image processing delays of the original algorithm and the algorithm proposed in this paper are 87ms and 94ms. The algorithm in this paper has little effect on the delay. As a comparison, the average delay of the comprehensive filtering algorithm is as high as 121ms, which is difficult to meet the real-time requirements. At the same time, in terms of tracking accuracy, the root mean square error of the error of the filtering algorithm and the reference trajectory in this paper is only 0.405. Obviously, the algorithm in this paper can accurately reflect the real gesture trajectory. By comparing the two experimental results we successfully demonstrate that the integrated tracking algorithm proposed in this paper that uses a fusion of color, depth and prediction information can highly improve the accuracy of tracking.

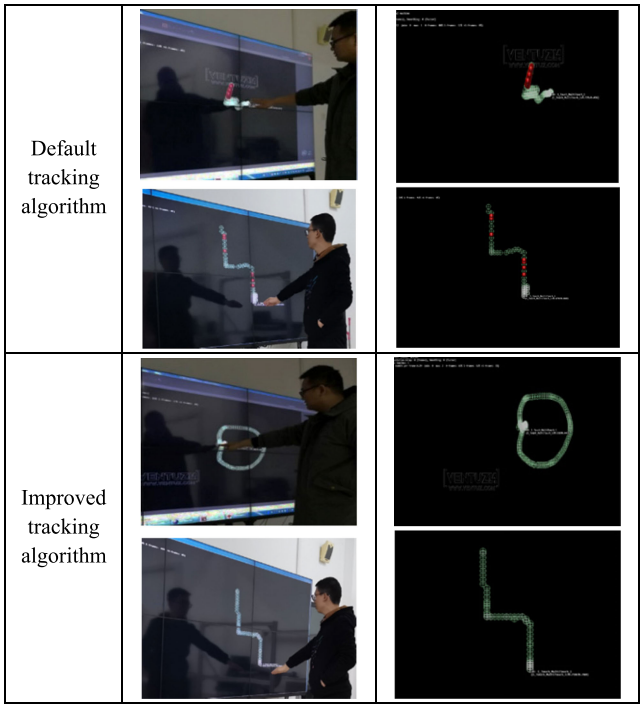
TABLE 1. Comparison of tracking success rates between two algorithms.

Tracking method	Number of experiments	Number of successes	Rate of success	Average processing time
Camshift	100	37	37%	87ms
Improved algorithm proposed in this paper	100	92	92%	94ms
Comprehensive filtering algorithm	100	96	96%	121ms

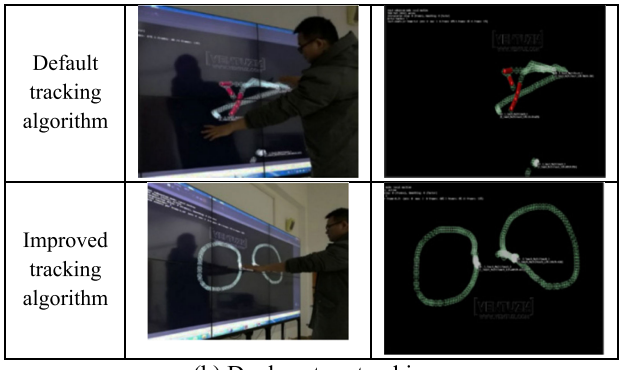
C. COMPARISION OF FILTERING ALGORITHM AND MACHINE LEARNING BASED RGB-D TRACKING ALGORITHM

The comparison between the filtering algorithm in the previous section and the switching filtering algorithm that only contains color information does not seem to explain the superiority of the improved algorithm [22]–[25]. This section compares the effects of the proposed multi-information fusion algorithm with the Kinect SDK's original machine-based learning RGB-D tracking algorithm based on the "Kinect-based large-screen interactive projection system" constructed by the research group. The interactive projection system can also perform gesture interaction. Next, a series of gesture tracking contrast experiments are performed in a dark natural light environment, and the tracking trajectory can be directly observed through the VENTUZ interactive design software demonstration interface.

The location of the gesture tracking in each frame of the image is indicated by a circle. Green circles indicate that the tracking can be performed normally, and red exclamation marks indicate that the frame has failed to track the position of the gesture. The trajectory is shown in Figure 9. In Figure 9(a), when there is no EKF algorithm, the trajectory tracking of a single gesture is lost, and the tracking accuracy is not satisfactory. After the EKF algorithm is involved, the gesture trajectory is greatly stabilized, and the gesture position can be tracked well. The multi-point operation in Figure 9(b) also shows that EKF algorithm has greatly improved the stability of gesture tracking. From the experimental results, we can see that both the single gesture tracking and dual gesture tracking, the system-based tracking algorithm based



(a) Single gesture



(b) Dual gesture tracking

FIGURE 9. Comparison of improved algorithm and machine learning based RGB-D tracking algorithm.

on machine learning will have errors, and the two-hand tracking error is more obvious. Based on the improved RGB-D tracking algorithm with EKF observer, the tracking position is accurate.

At the same time, in order to verify the universal adaptability of the algorithm, we performed a gesture interaction experiment on another self-developed platform. We apply the proposed algorithm to this new system. The interface diagram of the interactive system is shown in Figure 10. After the algorithm in this article is processed, gestures can also be displayed at the corresponding positions in the interactive frame. Based on this, VENTUZ is still used to show the trajectory of gestures, as shown in Figure 11. Users can perform multi-point operations smoothly. This fully proves the advantages of the algorithm in gesture trajectory tracking.

Remark 1: Compared with the traditional filtering method, we can make the observations that under the improved

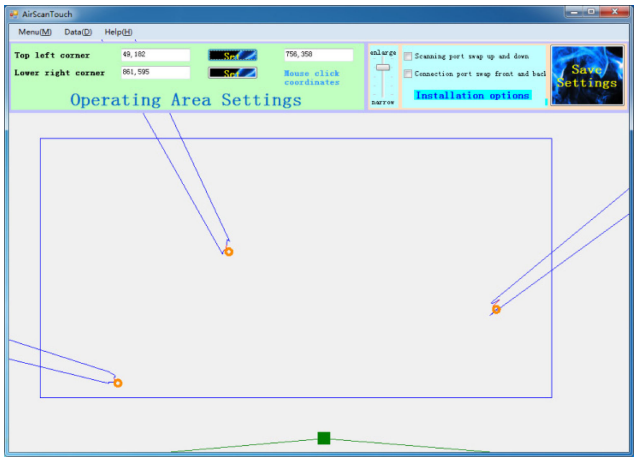


FIGURE 10. Schematic diagram of another interactive interface for multi-point operation.

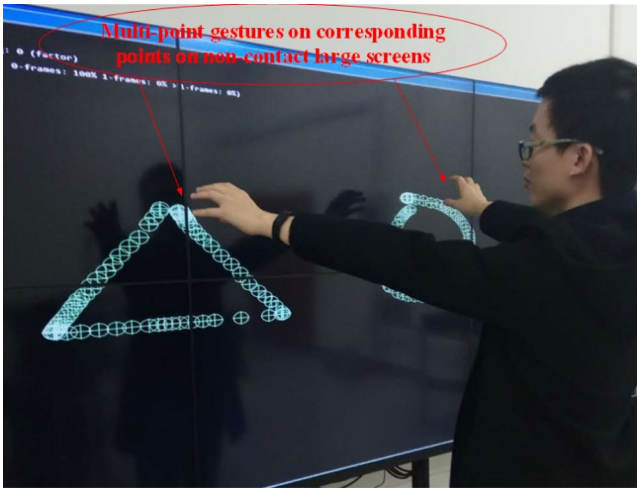


FIGURE 11. Multi-point operation trajectory diagram.

moving tracking synthesis algorithm with 3D sensor in this paper: (1) We can successfully solve the problem where the continuously adaptive mean shift algorithm encounters disturbance and tends to enlarge the tracking window; and that (2) We can robustly overcome the influences of external disturbance on the system. Therefore, we can state that the presented improved switching filtering algorithm is effective and has better performance than the original machine learning based RGB-D tracking algorithm method of Kinect SDK.

V. CONCLUSION

In this paper, an improved moving tracking algorithm utilizing multi-feature fusion with switching filter is proposed. First, the depth and color information are combined to generate a weighted original background color projection. Next, a switching filter algorithm is performed on the resulting image. Then the prediction information is added to the algorithm and the direction of movement is predicted using the extended Kalman filter. The proposed algorithm effectively solves the problem caused due to the tracking error in cases when the target and the background color are close.

The algorithm enlarges the interference search window and improves the accuracy of tracking. The experimental results verify the effectiveness of the improved moving tracking algorithm. The results demonstrate that the algorithm proposed in this paper greatly improves the system's robustness and ability to resist interference.

REFERENCES

[1] H. Wang, H. Li, J. Fang, and H. Wang, "Robust Gaussian Kalman filter with outlier detection," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1236–1240, Aug. 2018.
[2] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking learning detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Dec. 2012.
[3] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[4] R. Zhang, X. Feng, R. Zhang, S. Liu, and Y. Zheng, "Adaptive photoacoustic sSensing using matched filter," *IEEE Sensors Lett.*, vol. 1, no. 5, Oct. 2017, Art. no. 3501003.
[5] L. Dai, M. Yuan, and X. Zhang, "Speeding up the bilateral filter: A joint acceleration way," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2657–2672, Jun. 2016.
[6] 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.
[7] M. A. Zulkifley, M. M. Mustafa, and A. Hussain, "On improving CamShift performance through color constancy approach," in *Proc. Int. Conf. Comput. Inf. Sci.*, Jun. 2012, pp. 375–378.
[8] M. Yin, J. Zhang, H. Sun, and W. Gu, "Multi-cue-based CamShift guided particle filter tracking," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6313–6318, May 2011.
[9] S. Li, S. Zhao, B. Cheng, and J. Chen, "Accelerated particle filter for real-time visual tracking with decision fusion," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 1094–1098, Jul. 2018.
[10] G. Du and P. Zhang, "A novel human–manipulators interface using hybrid sensors with Kalman filter and particle filter," *Robot. Comput.-Integrated Manuf.*, vol. 38, pp. 93–101, Apr. 2016.
[11] L. Yu, J. Huang, and S. Fei, "Robust switching control of the direct-drive servo control systems based on disturbance observer for switching gain reduction," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 8, pp. 1366–1370, Aug. 2019.
[12] C. Bolle, A. Sridhar, H. Safar, D. Neilson, M. Cappuzzo, and M. Earnshaw, "A fast and low-loss electromagnetically tunable optical filter," *IEEE Photon. Technol. Lett.*, vol. 30, no. 9, pp. 837–840, May 1, 2018.
[13] C. Li, L. Yu, and S. Fei, "Large-scale, real-time 3D scene reconstruction using visual and IMU sensors," *IEEE Sensors J.*, vol. 20, no. 10, pp. 5597–5605, May 2020.
[14] J. Huang, X. Ma, H. Che, and Z. Han, "Further result on interval observer design for discrete-time switched systems and application to circuit systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, early access, Dec. 6, 2019, doi: 10.1109/TCSII.2019.2957945.
[15] K. Jung, "Hybrid approach of texture and connected component methods for text extraction in complex images," *Signal Process.*, vol. 41, no. 6, pp. 1033–1044, 2004.
[16] W. Li, S. Sun, Y. Jia, and J. Du, "Robust unscented Kalman filter with adaptation of process and measurement noise covariances," *Digit. Signal Process.*, vol. 48, pp. 93–103, Jan. 2016.
[17] B. Ghojogh, H. Mohammadzade, and M. Mokari, "Fisherposes for human action recognition using kinect sensor data," *IEEE Sensors J.*, vol. 18, no. 4, pp. 1612–1627, Feb. 2018.
[18] L. Yu, C. Li, and S. Fei, "Any-wall touch control system with switching filter based on 3-D sensor," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4697–4703, Jun. 2018.
[19] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 845–848, Dec. 2012.
[20] 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.
[21] R. Lamberti, Y. Petetin, F. Desbouvries, and F. Septier, "Semi-independent resampling for particle filtering," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 130–134, Jan. 2018.
[22] Y.-L. Hsu, J.-S. Wang, and C.-W. Chang, "A wearable inertial pedestrian navigation system with quaternion-based extended Kalman filter for pedestrian localization," *IEEE Sensors J.*, vol. 17, no. 10, pp. 3193–3206, May 2017.
[23] J. Yoo, J. H. Lee, and H. Jung, "A real-time face tracking algorithm using improved camshift with depth information," *J. Elect. Eng. Technol.*, vol. 12, no. 5, pp. 2067–2078, 2017.
[24] L. Yu and S. Fei, "Large-screen interactive technology with 3D sensor based on clustering filtering method and unscented Kalman filter," *IEEE Access*, vol. 8, pp. 8675–8680, 2020.
[25] 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.

YIFAN FANG received the B.S. degree from Changzhou University, Changzhou, China, in 2018. He is currently pursuing the M.S. degree with the School of Mechanical and Electrical Engineering, Soochow University, Suzhou, China. His current research interests include filter system intelligence control and switched systems.

LEI YU (Member, IEEE) was born in Xuancheng, China, in 1983. He received the Ph.D. degree from the College of Automation, Southeast University, Nanjing, China. He is currently a Professor with the School of Mechanical and Electrical Engineering, Soochow University, Suzhou, China. His main research interests include human–computer interaction, filter systems, and robust control.

SHUMIN FEI was born in 1961. He received the Ph.D. degree from Beihang University, Beijing, China, in 1995. From 1995 to 1997, he did his postdoctoral research at the Research Institute of Automation, Southeast University, Nanjing, China, where he is currently a Professor with the Research Institute of Automation. His research interests include analysis and synthesis of nonlinear systems, robust control, adaptive control, and analysis and synthesis of time-delay systems.