A Laser Pointer Trajectory Recognition Method for Human-computer Interaction

Chong Li[1], Bin Xin[1], Qing Wang[*1]

[*1]School of Automation, Beijing Institute of Technology, Beijing 100081, China
E-mail: brucebin@bit.edu.cn

Abstract. In the mainstream teaching equipment, because the laser pointer is limited by distance and has a relatively single function, it cannot meet the needs of classroom teaching; in order to facilitate teaching, this paper designs a human-computer interaction oriented laser pointer trajectory recognition method, and proposes a hybrid algorithm that combines the traditional light spot recognition algorithm ,it can ensure the real-time update of the background image, at the same time accurately identify the tiny moving target; the improved BP neural network algorithm is proposed for trajectory recognition, which gives full play to the nonlinear fitting ability of the neural network. Compared with the previous trajectory recognition methods, the advantages and characteristics of this method are: the characteristics of the light spot and its background can be trained before use, which ensures the adaptability of the whole system; also, it can continuously identify different trajectories, after arranging and combining them, the laser pointer can be used to realize more complex functions, which is convenient for the speaker's operation. Therefore, it provides efficiency as well as scalability for us. Finally, after experimental verification, the accuracy rate of this method is as high as 95% when using an infrared laser pointer and normal light conditions, which proves the feasibility and robustness of the method.

Keywords: Human-computer interaction; Laser pointer; Trajectory recognition; Neural network

1. INTRODUCTION

Human-computer interaction is the study of the interaction between systems and users. Systems are various types of computer systems as well as corresponding software, and users refer to humans. The current mainstream human-computer interaction function is realized by some external devices and software, including keyboard, mouse, hard disks, floppy disks and other genre of devices, the corresponding software matching these commonly used devices act as the bridge of communication between user and system.

In recent years, laser pointers and electronic whiteboards have been widely applied in intelligent classroom in the study of human-computer interaction. At the beginning of the 21st century, there is little research on electronic whiteboard in China. Due to its high price and the uneven technical level of teachers at that time, very few people have paid attention to it. After 2004, with the breakthrough of electronic whiteboard technology and the reform of national curriculum emphasizing the interaction between teachers and students, electronic whiteboard has been widely introduced into modern teaching and become an important auxiliary tool for teachers to use. It has also led to extensive research. At present, the electronic whiteboard in the market is basically stagnant, due to the high cost and resources of research, and also become a constraint on its development.

The remote-control laser pointer has contributed its own indispensable power to the teaching of multimedia classrooms from beginning to end. Its development is roughly as follows: from the first generation of laser pointers that can only emit light spots to the second generation of laser pointers that can realize infrared remote control, which can realize the page turning function of PPT. The disadvantage is that it is less efficient if it is obstructed by objects. At present, the most widely used is the third-generation radio frequency remote control laser pointer, which realizes the corresponding functions through the well-developed radio frequency technology in recent years, eliminating the previous limitation of the need to align the direction, but the effect is better in the case of electromagnetic interference. As far as the current situation is concerned, the functions of remote-control laser pointer are relatively few and cannot be extended, which is not able to fully meet the needs of interactive classroom.

This paper takes human-computer interaction as the background, and makes full use of the convenience, portability and easy identification of laser pointers. It mainly exploits imgproc (image preprocessing), video (real-time image acquisition) and highgui (The realization of specific functions after recognition) algorithm modules in the computer vision library of OpenCV, then converts the video collected by the camera into operable and detectable image. In spot detection, a complex algorithm combining the

mainstream inter-frame difference method and background difference method is selected to detect and calculate the coordinates of spot in real time. Next, a series of discrete coordinate points are fitted into smooth curves by the least square method, forming a series of images of the training set and transmitting them to the BP neural network for training, which can not only ensure the accuracy of recognition, but also meet the rapidness of response time, finally, the output function is the given that matches the trajectory.

2. RELATED WORK

In terms of the extensible laser pointer, MAO Peng et al. used the camera to identify the Z-shaped broken line track, and used the pylon laser pointer built by the 2-DOF steering gear to make an attendance checking software based on Android[1]. On the basis of the original laser pointer, Liu Gang et al. transplanted the bright spot recognition algorithm to FPGA embedded platform, and established a set of auxiliary teaching system[2]. Based on the human-computer interaction laser pointer project, Zhang Qilin et al. developed the embedded hardware and upper computer software of the system and realized the function of multimedia content display[3]. A. Soetedjo et al. developed a cost-effective injection simulator using a laser pointer with high reliability in detection for indoor applications[4].

In the field of image processing, Xin Wang et al. introduced the bihaar wavelet transform in the wavelet domain, which is suitable for local multiscale analysis of moving Windows[5]. L. M. G. Fonseca et al. used image processing techniques to solve a problem in remote sensing application and provided an urban application case[6]. J. Zheng et al. introduced a new collaborative image processing algorithm to enhance the contour and surface details of the scene by combining multi-light images[7]. E. S. Dunstone derived and demonstrated a technique for rapidly performing affine image processing operations in a scale-independent manner using IAN networks[8]. J. Im et al. proposed a ghost artifact removal algorithm for obtaining high dynamic range (HDR) images without artifacts in the presence of camera motion[9].

In terms of track recognition, L. Lin integrated kalman filter prediction method into CamShift algorithm for target tracking, which improved target tracking speed and effectively solved the problem of fast target tracking in the case of interference[10]. Yan-hua Liang et al. proposed a SIFT target recognition algorithm based on color guidance and a target tracking algorithm based on compressed sensing theory, which can be applied to panoramic visual target recognition and tracking[11]. In the digital video surveillance system, Xiao-Yun Xiong uses image recognition technology to track suspicious moving targets and alarm during forest fire prevention, and to identify forest fires and alarm during forest fire monitoring[12].

In the development of human-computer interaction, Y. Liu et al. proposed the design research method of human-computer interaction interface based on design psychology, and developed a set of human-computer interaction interface based on design psychology suitable for the new era[13]. X. Zhao designed a smart home power management human-computer interaction system using r-GOMS model, which meets the requirements and principles of human-computer interaction design[14]. Y. Tao et al. realized a complete human-computer interaction system based on hand-eye positioning and simple gesture recognition based on Kinect, which can quickly locate hand-eye position through depth image and color image with strong real-time performance[15]. S. M. Sali Shajideen et al. proposed a single-camera gesture estimation algorithm for gesture tracking from a distance in 2D space and developed a real-time human-computer interaction system[16]. Y. Xu et al. introduced the design and development of a uav suitable for indoor navigation and an innovative human-computer interaction interface, which realized the effective and efficient control of uav[17].

3. OVERVIEW

In this paper, the basic idea of the human-computer interaction method for laser pointer trajectory recognition proposed is as follows: First, the camera is used to shoot a large screen in real time, then the specific trajectory drawn by the person holding the laser pen is transmitted to the computer, thus the computer can obtain a series of projected points by using image processing technology. Finally, after trajectory fitting and recognition algorithm, the output function will be given. The configuration diagram and flow chart of the system are shown in Figure 1 and Figure 2.

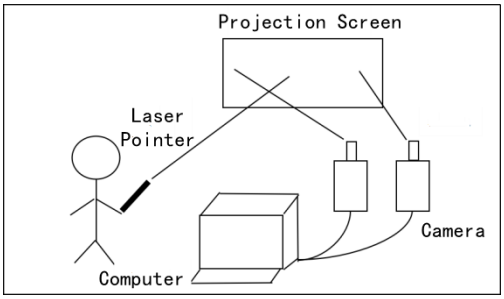


Fig. 1 System configuration diagram.

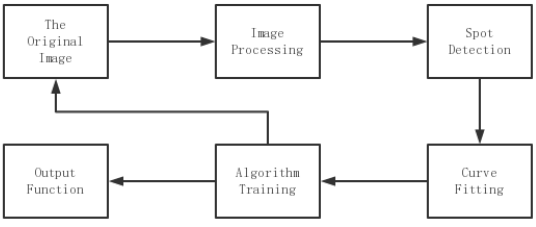


Fig. 2 Flowchart of the system.

Chapter 4 describes how the original image is processed into usable images. Chapter 5 introduces the algorithm and process of laser spot recognition used in

this paper. Chapter 6 shows the method of fitting coordinate of light spot into curve and the whole process of trajectory recognition. Chapter 7 presents a physical demonstration of the system's inputs and outputs, chapter 8 summarizes the content of the full text.

4. IMAGE PROCESSING

4.1. Gaussian Filtering

Gaussian filter is a kind of linear smoothing filter which is widely used in image processing. The basic idea is that in the process of image processing, the pixel values of the whole target image are superimposed and the average value is obtained. For the value of each specific pixel point, its own value and the value of other adjacent pixels are also superimposed and averaged to determine the value.

The two-dimensional Gaussian distribution is shown in the figure:

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

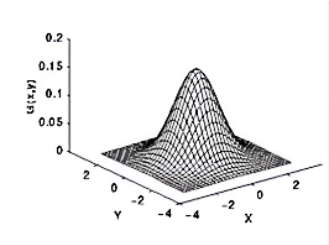


Fig. 3 Two-dimensional Gaussian distribution

When the computer receives the image from the camera, it carries out weighted average of the whole image, and selects the size of the filtering kernel as 3×3 according to the characteristics of the image, while the standard deviation of the convolution kernel in the vertical direction is calculated by the function itself to obtain an appropriate value. The image obtained by Gaussian filtering will greatly reduce the background noise and show the identified subject object more clearly.

4.2. Threshold Segmentation

The basic principle of threshold segmentation is to roughly divide the pixels of the target image into several different categories by setting a series of special thresholds. Its calculation formula is as follows:

$$R(x,y)=\begin{cases}255 & f(x,y)>T\\ 0 & others\end{cases} \tag{2}$$

Where $f(x,y)$ means gray value of target midpoint pixel; T means the original threshold; $R(x,y)$ means the pixel value of the point after binarization (255 or 0).

4.3. Morphological Operation

The word morphology usually refers to the branch of biology that deals with the morphology and structure of animals and plants. Because most of the objects in image processing are digital images, it refers to mathematical morphology here. The basic operations include: binary corrosion expansion, binary open and close operation, top hat transformation, particle analysis, etc.

In short, morphological operation is a series of image processing operations based on shape. This paper briefly introduces the morphological operations used.

(1) Inflation. Expansion is the operation of finding a local maximum. Mathematically, dilation and corrosion are the convolution of a target image with a known kernel. The mathematical expression is:

$$dst(x,y)=\max src(x+x',y+y') \tag{3}$$

Among them $(x',y'):element(x',y')\neq0$

(2) Corrosion. Expansion and corrosion are a pair of opposite operations, the opposite of the maximum is the minimum. That's what corrosion means. The functional prototype is the same for both. The mathematical expression is the same as inflation:

$$dst(x,y)=\min src(x+x',y+y') \tag{4}$$

Among them $(x',y'):element(x',y')\neq0$

(3) Binary open operation. A is set as the target object, B is the structural element, and the opening operation of B on A is defined as:

$$A\circ B=(A\Theta B)\oplus B \tag{5}$$

(4) Binary closed operation. The closed operation is just the opposite to the open operation. The expansion operation is carried out first, and then the corrosion operation is carried out. Its definition is as follows:

$$A\bullet B=[(A\oplus(-B)]\Theta(-B) \tag{6}$$

According to the characteristics of the laser pointer's light spot, gray-scale transformation and threshold segmentation are performed on the Gaussian filtered image. The color image is first transformed into grayscale image, which avoids a lot of time consumption caused by processing color information, and then transformed into binary image, which greatly reduces the amount of information of the image and speeds up the processing speed. For some images of burrs and holes, we shall use morphological operation processing, including a combination of opening operation and closing operation, to effectively remove the burr and fill the holes, at the same time improve the stability of the system for subsequent processing.

5. SPOT RECOGNITION

Background difference method is one of the most commonly used methods in motion detection, its basic idea is to use the current real-time image and fixed background image to do effective difference, so as to detect moving objects. Its advantage is that it can provide complete image feature data. However, its disadvantages are also obvious, it is sensitive to changes in dynamic environment. Therefore, the background difference method is not effective in extracting moving objects with large background changes, and its formula is as follows:

$$I_k(x,y)=\begin{cases}1 & Foreground \quad if\,|(f_k(x,y-B_k(x,y))\geq T| \\ 0 & Background \quad if\,|(f_k(x,y-B_k(x,y))\leq T|\end{cases} \tag{7}$$

Among them $f_k(x,y)$ means the current k frame image, $B_k(x,y)$ means current background, T means threshold for binarization, and $I_k(x,y)$ means the target image after binarization.

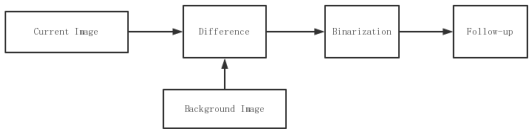


Fig. 4 Schematic diagram of background difference method

The principle of inter-frame difference method is to extract moving areas and objects by using the slight changes between adjacent frames of the target image. Its practices is to two different background but same time effectively the difference of adjacent frames, the image of a background part because there are few changes would be cut, therefore all basic moving targets as the change of time in a different position, and the grey value and the background is far, so the differential motion targets will emerge after the operation, You can roughly determine the position of the moving object in the image;The details are as follows:

$$I_k(x,y)=\begin{cases}1 & Foreground \quad if\,(f_k(x,y-f_{k-1}(x,y))\geq T \\ 0 & Background \quad if\,(f_k(x,y-f_{k-1}(x,y))\leq T\end{cases} \tag{8}$$

Among them, $f_k(x,y)$ represents the image of the current frame; $f_{k-1}(x,y)$ means the last image of the current frame; T means threshold of binarization; $I_k(x,y)$ means the image after binarization.

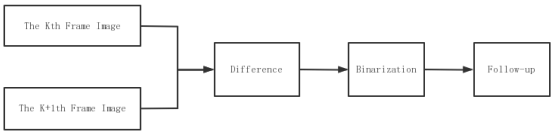


Fig. 5 Schematic diagram of difference method between frames

The moving target that needs to be detected in this paper is the coordinates of the light spot (infrared light) of the laser pointer's trajectory. Because the target light spot is a small object compared with the big screen, the tracking and recognition effect obtained by using a single algorithm cannot meet the expected requirements.

Therefore, the hybrid organic algorithm combining background difference method and frame difference method is adopted in this paper, and part of the content is optimized and improved on this basis, and the recognition and detection of moving light spot is well realized. The specific process is as follows:

(1) Using inter-frame difference method to judge the position of moving area and target light spot;

(2) Using background difference method to detect and identify real-time moving targets. After distinguishing the changing region and the non-changing region, the moving object is detected again according to the changing region and the background of the current frame;

(3) Perform threshold segmentation and morphological transformation on the difference image;

(4) Detect and extract real-time spot coordinates and store them in a two-dimensional array.

6. CURVE FITTING AND TRAJECTORY RECOGNITION

6.1. Curve Fitting

The least square method is a mathematical optimization method. It minimizes the sum of error squares to find the optimal matching function for a given data. By using the least square method, the unknown data can be easily obtained, and the sum of error squares between the obtained data and the actual data can be minimized, which makes the data approximately equivalent to the actual data. The least square polynomial curve fitting, according to a given m points, which does not require the curve to pass through these points perfectly, but an approximation of curve $y=f(x)$ to curve $y=\varphi(x)$.

The specific steps are as follows:

(1) Assign the resulting two-dimensional array to the point set in the form of a stack;

(2) Use constructors to calculate polynomial coefficient matrix A;

(3) Take x coordinate as independent variable, calculate corresponding Y according to coefficient matrix A;

(4) Stack the new coordinates into the new point set;

(5) Use the function of drawing polygons to draw the fitted curve.

6.2. Track Identification

6.2.1. SELECTION OF TRAJECTORY RECOGNITION ALGORITHM

The basic idea of motion trajectory recognition is to extract the real-time coordinate information of the trajectory of the moving object in the completion, and then use these coordinates as the breakthrough point to judge the corresponding direction vector and order of the corresponding judging conditions to realize the motion trajectory recognition. Artificial Neural Networks (ANN) is a Neural network used to simulate and abstract the characteristics of human brain Neural Networks. It is a kind of nonlinear information processing system which is composed of a large number of interlocking processing units. Its advantages and characteristics are as follow:

(1) Adaptive and self-organizing ability

Artificial neural network is a system with learning ability, which can develop knowledge beyond the designer's original knowledge level.

(2) Generalization ability

Generalization ability refers to the ability to predict and control untrained samples.

(3) Nonlinear mapping ability

Neural network does not need to have a deep understanding of the system, but at the same time can

realize the mapping relationship between input and output, greatly simplifying the difficulty of design.

(4) Highly parallel

Neural network is an abstract model based on human brain. Neural networks are also very parallel from the point of view of people doing things at the same time.

6.2.2.DESIGN OF BP NEURAL NETWORK

(1) Determine the number of network layers

The more layers of BP neural network, the smaller the error will be, and the complexity and response time will be correspondingly longer. According to its characteristics, if there is a hidden layer in the three-layer neural network, and the neuron of the hidden layer has an S-type function, it can fully approximate and approach an arbitrary class of nonlinear problems infinitely. Therefore, this paper considers the choice of three-layer network.

(2) Number of nodes in input layer

The dimension of feature vector extracted from the object to be recognized is 8-dimensional direction vector, so there are 8.

(3) Number of nodes in the output layer

Since the system currently has 5 kinds of tracks to identify, 5 nodes are selected.

(4) Number of nodes of hidden layer

The experimental results show that the learning and information processing ability of neural network is weak if the number of hidden layer nodes is too small. On the contrary, too much will greatly increase the complexity of the system and lead to slow learning. Therefore, the selection of hidden layer neurons is very important. At present, the classic methods are:

Kolmogorov's theorem

The number of nodes in the hidden layer, where N is the number of nodes in the input layer, and I is a small increment, usually a constant 1. $s = 2n + I$

In this paper, Kolmogorov's theorem is adopted to determine the number of nodes in the hidden layer, so the number of neurons in the hidden layer is 17.

(5) Selection of activation function

Activation function Generally, s-type function is selected:

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{9}$$

Compared with the standard S-type function, the convergence time can be reduced by half.

6.2.3.BP NEURAL NETWORK LEARNING PROCESS

The specific process of neural network recognition is as follows:

(1) Establish ANN neural network model;

(2) Set the number of input layer, hidden layer and output layer;

(3) Set the algorithm as the back propagation algorithm and select parameters;

(4) Set the activation function as Sigmoid function;

(5) Set termination conditions and start training;

(6) The training set is obtained after the training;

(7) Use the training set as a reference to identify the real-time track and output the results.

7. EXPERIMENTAL RESULTS

In order to facilitate the speaker's operation and improve the accuracy of recognition training, five different kinds of gesture tracks are established in this paper, which are " | ", "/", "\", "∧" and "∨". The original image and the image after Gaussian filtering are shown in Figure 6, and the image after threshold segmentation is shown in Figure 7. The coordinates of spot recognition and their display are shown in Figure 8.The curve fitted by the least square method is shown in Figure 9.The corresponding functions of the final input track output are shown in Figure 10.

This method is based on Windows10 computer platform and supported by OpenCV technology and Visual Studio 2017. The recognition accuracy of five kinds of trajectory is more than 95%, and the recognition speed is about 0.5s. In the test process, when the trajectory drawn differs from the given gesture, as long as the outline or direction is similar, the system can output correct results, which greatly improves the robustness of the system. In addition, the speaker can use different types of gesture arrangement and combination, so as to achieve more complex functions. For example, in the class roll call, drawing a gesture A means opening the list of a designated directory, under this condition, drawing a gesture B is about to indicate that someone has arrived, then we can add the arrival marker after the person's name; drawing a gesture C means that someone is absent, then we can add the absent mark after the person's name; After using a variety of track arrangement and combination, the speaker can realize functions can be geometrically multiplied. Compared with most human-computer interaction methods, this method has obvious advantages, which can meet the various functional requirements and response time requirements of speakers, and ensure the rich scalability of the system.



Fig. 6 Original image and Gaussian filtered image

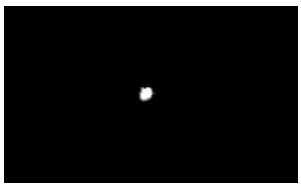


Fig. 7 Threshold segmentation image

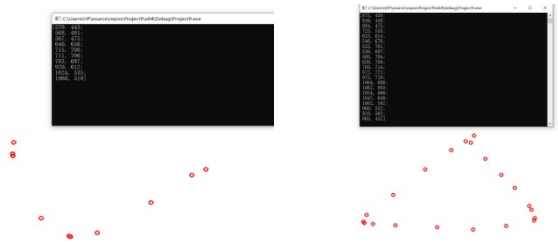


Fig. 8 Light spot recognition image

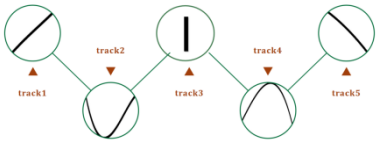


Fig. 9 Five curve-fitted trajectories



Fig. 10 Final output of track 5 and track 1

8. CONCLUSION

This paper mainly constructs a new type of laser pointer trajectory recognition method oriented to human -computer interaction. Its purpose is that the speakers can use the laser pointer in order to control the computer, making it perform necessary and complex function. Furthermore, the method successfully lets the speakers get rid of the constraints of the mouse and keyboard, then step off the stage and integrate into the audience. In addition, this method is not limited by distance, thus the speaker can function in most parts of the classroom. It is believed that in the future, this new human-computer interaction method can add its own richness to the multimedia teaching.

Acknowledgements

This work was supported in part by the National Outstanding Youth Talents Support Program 61822304, in part by the Basic Science Center Programs of NSFC under Grant 62088101, in part by Shanghai Municipal Science and Technology Major Project (2021SHZDZX0100), in part by Shanghai Municipal Commission of Science and Technology Project (19511132101), in part by the National Natural Science Fund of China under Grant 62003044.

References:

[1] MAO Peng, Shen Xinfeng, Li Wei, et al. For trajectory recognition based on image acquisition and tracking system design [J]. Journal of information technology, 2015 (5) : 85-87, living, DOI: 10.13274 / j.carol carroll nki HDZJ. 2015.05.022.

[2] Liu Gang. Laser Pointer Assisted Teaching System Based on FPGA [D]. Shanghai: Shanghai Jiao Tong University,2011.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[3] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[4] A. Soetedjo, E. Nurcahyo and Y. I. Nakhoda, "Development of a cost-effective shooting simulator using laser pointer," Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, 2011, pp. 1-5, doi: 10.1109/ICEEI.2011.6021835.

[5] Xin Wang, "Moving window-based double haar wavelet transform for image processing," in IEEE Transactions on Image Processing, vol. 15, no. 9, pp. 2771-2779, Sept. 2006, doi: 10.1109/TIP.2006.877316.

[6] L. M. G. Fonseca, L. M. Namikawa and E. F. Castejon, "Digital Image Processing in Remote Sensing," 2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing, 2009, pp. 59-71, doi: 10.1109 / SIBGRAPI - Tutorials. 2009.13.

[7] J. Zheng, Z. Li, S. Rahardja, S. Yao and W. Yao, "Collaborative image processing algorithm for detail refinement and enhancement via multi-light images," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 1382-1385, doi: 10.1109 / ICASSP. 2010.5495456.

[8] E. S. Dunstone, "Image processing using an image approximation neural network," 1 st Proceedings of International Conference on Image Processing, 1994, pp. 912-916, vol. 3, doi: 10.1109 / ICIP. 1994.413713.

[9] J. Im, S. Jang, S. Lee and J. Paik, "Geometrical transformation-based ghost artifacts removing for high dynamic range image," 18 th, 2011 IEEE International Conference on Image Processing, 2011, pp. 357-360, doi: 10.1109 / ICIP. 2011.6116490.

[10] L. Lin, "Research on Application of Image Recognition and Tracking Technology Based on Moving Objects," 2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA), 2020, pp. 222-225, doi: 10.1109 / ICSGEA51094.2020.00054.

[11] Y. -h. Liang and C. -t. Cai, "Robot target recognition and tracking based on panoramic vision," 2017 IEEE International Conference on Mechatronics and Automation (ICMA), 2017, pp. 131-135, doi: 10.1109 / ICMA. 2017.8015801.

[12] X. Xiong and B. Wang, "Study on forest fires recognition and moving target tracking in video surveillance system," Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, 2011, Pp. 2853-2856, doi: 10.1109 / EMEIT. 2011.6023697.

[13] Y. Liu, "Human-Computer Interface Design Based on Design Psychology," 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), 2020, pp. 5-9, doi: 10.1109 / ICHCI51889.2020.00009.

[14] X. Zhao, C. Zhou and W. Huang, "Smart home power management system design based on human-computer interaction model," Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, 2013, pp. 1247-1250, doi: 10.1109 / ICCSNT. 2013.6967328.

[15] Y. Tao, Z. Yao and B. Liang, "Human-Computer Interaction Using Fingertip Based on Kinect," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018, pp. 888-893, doi: 10.1109 / ITOEC. 2018.8740411.

[16] S. M. Sali Shajideen and V. H. Preetha, "Human-Computer Interaction System Using 2D and 3D Hand Gestures," 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), 2018, pp. 1-4, doi: 10.1109 / ICETIETR. 2018.8529064.

[17] Y. Xu and J. E. Meng, "Human-computer interaction in intelligent control of an unmanned aerial vehicle," 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI), 2016, pp. 42-46, doi: 10.1109 / ICICPI. 2016.7859670.