Design of a Virtual Badminton Teaching System

Yingna Feng

Jiangxi Science and Technology Normal University,
Nanchang, 330013, China

Abstract
Due to the wide application in circles such as national defense, industrial production, transportation and PE, the computer vision technology has become more and more important. Combined with technical requirements and specifications of the "virtual badminton", this paper studies and implements the "virtual badminton" system based on the information technology. We can get the three dimensional space information of the target object based on the single target features of the "Virtual Badminton" system and real-time requirements with the help of the stereo visual algorithm. As for the errors in the three dimensional data, we use the curve fitting method to carry out the smooth processing on the data, and get virtual effect through the three dimensional reconstruction. The experimental result has shown the efficiency of algorithm.
Key words: INFORMATION TECHNOLOGY; MOTION DETECTION; ROBUSTNESS

Introduction

As an emerging discipline, the computer vision has been developed very fast in recent years [1]. This discipline originates from the photogrammetry and the computer technology was introduced when measuring and analyzing the images. Through digital processing of images, and then we can put the algorithm into a program. After we complete the quantitative analysis and the calculation of the scene structure, as well as the statistical pattern recognition on the other hand, our initial work is finished based on the analysis and identification of the two-dimensional images. After years of development under the framework of the Malta theory, the computer vision has made a large number of achievements. However, with the increasingly deepening of the computer vision researches, people have found the limitations of Malta vision theory, that is, we have met troubles in trying to figure out the accurate three dimensional geometric descriptions from the scenery images or image sequences and quantitatively determine the nature of the objects in the scenes. The perception is an action rather than just a passive response, and the computer vision should also develop towards the initiative vision.

Studies on mobile robots and other relative applications were conducted from the mid of 1980 until the later time. Due to its close relationship with the vision studies, the space geometry methods and physics knowledge were widely used for stereoscopic vision studies. Besides, the active vision research method, distance sensor and the fusion technology ware introduced in this period. Because this method can obtain the depth map directly or get the depth map by a moving object, many ill posed problems were turned into good ones.

With the development of technology and the extensive application of the computer vision, the emphasis of computer vision research has changed from studying static image to studying the dynamic image sequence. The detection of moving objects has also made a rapid development. Until now, the motion target detection technology has been widely used in the national defense, transportation, industrial production, security and other fields, and has been developing and updating gradually. As early as the mid of 1970, the method of inter frame difference classification has been applied to the moving target detection in the static scenes.

Most of the early detection of moving targets was based on the time domain and the algorithm of frame checking threshold was first proposed at that time. This algorithm is generally applicable to the stationary observer. The current method includes the background subtraction method, frame difference method and the optical flow method. Background subtraction method is the most common method in current motion detection. First of all, it captures the image as the background image and then uses the current image and the background image to detect the moving object. This method can almost provide a complete characteristic data base. The only shortcoming lies in that, it is easy to be influenced by the light in the scene and other interferences.

So after the research, the relative researchers have improved this method, which includes the method proposed in the literature [2], that is, we use the minimum strength value, the maximum strength value and the maximum time score to make statistics and establish the model in each pixel of the scene, and make the periodic updating on the background; In the literature [3], we take advantages of the background model adaptation method in the Carl full filter so as to adapt to the light In the literature [4], we use the self-adaptive mixed Gaussian background model to make models, and used the online estimated update model method to process the shut light changes, and dealt with the effects of the background interference caused by the chaotic. The Aiming at pixel-level processing, the frame difference method deals primarily with the adjacent difference of two or three frames, so as to find a detection method for the moving targets in the image. In the literature [5], we adopt the two-frame difference method to detect the moving target in the actual video, which is used for target tracking and classification; the improved method is the three frame difference method. In the literature [6], we use the three frame difference method combined with the self-adaptive background cancellation method to detect the moving targets from the background quickly and effectively. This method has a good effect on the dynamic targets, but it still cannot detect the static targets. The optical flow method refers to the motion speed of the model in the image. It is a kind of two dimensional instantaneous velocity fields, in which the two dimensional velocity vector is the image plane of the projection of the three dimensional velocity field [7]. This method is based on the model parameters. Under the conditions of a given flow, we use the flow field parameters as a feature. For the multiple independently moving objects, each of the optical flow vectors corresponds to a single opaque dimensional projection of the rigid body movement, mapping for each different series of movements that is used for the parameter as described in [8]. The "virtual badminton" system is based on the model of the Competition of the International Badminton in the Hawk-Eye system. At the same time, it is also an exhibition item in the permanent exhibition of the China Science and Technology Museum.

In addition, it is also an opportunity for the computer vision study, especially for the binocular stereo vision. Computer vision itself also has a lot of significances, which is one of the hottest topics in current researches. After the 1990s, the rapid development of popular science has been achieved all over the country. The teenagers' popular science education has been included as an important item. When we introduce the popular science knowledge to the masses of teenagers through the popular science exhibition item, we can better guide the healthy development for the young people, lead them to explore science and technology, develop healthy interests and hobbies, and lay a good foundation for the future development of science and technology in China.

Consequently, with regard to the system requirements, this paper demarcates the camera by Tsai two-step method. In this way, we obtain the internal and external parameters of two cameras, the position and pose transformation relationship between the two cameras and the relationship between the camera coordinate system and the world coordinate system, laying foundation to obtain three-dimensional information. When positioning the targeted objects, we adopt the method with the combination of background difference, feature point statistics and open operation in the mathematical morphology. On these grounds, we can steadily and accurately detect the centroid in targeted objects. It has good robustness and lack of sensitivity to the light changes. The simulation experience can be conducted after obtaining the two-dimensional centroid in targeted objects.

Analysis and Design of the "Virtual Badminton" System

The "Virtual Badminton" system is a participatory interactive exhibition. Through the vision subsystem, we can track the moving trajectory of the badminton, and save the sports points in the air. The audience can see the badminton's first landing point from multiple perspectives by determining whether the first landing point is in the disputed area and whether we need three-dimensional reconstruction.

Based on the above analysis and introduction, we can know that the system is a visual system based on computer vision. The key issue of this system is that the stereo vision system should have the ability to capture the space point of targeted objects at running time accurately and preserve them for the three dimensional reconstruction. Of all the points in the space, we should put the first landing point at the top rank since its accuracy rate is the highest. Consequently, we have designed the monocular vision system for this specific purpose, which can help to determine the exact position of the first landing point.

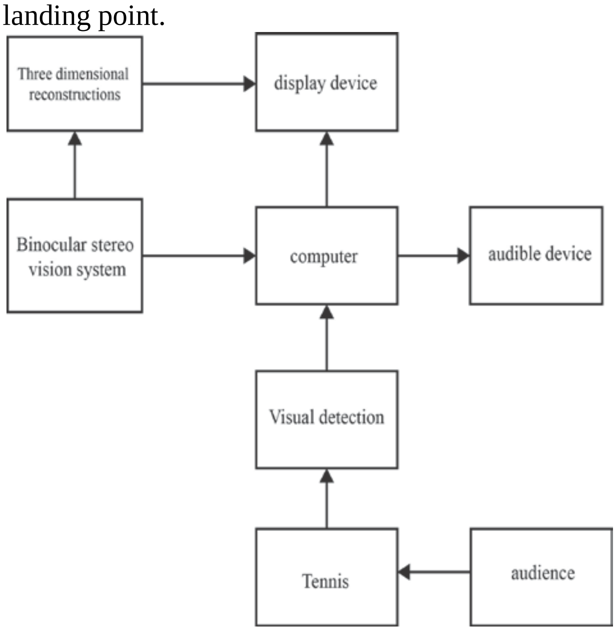

Figure 1. Overall Chart of the Virtual Badminton System

Two-Dimensional Detection and Location of the Moving Targets

In the "Virtual Badminton" system, the purpose of the detection for the moving targets is to find out whether there are any ball motions in the ground. If not, we can start the system introduction after a certain time; if any, we should save the captured images and then start the appropriate handler. The purpose of the "badminton" target detection is to detect the moving badminton from an image. Once the system has detected a moving object and saved the image, the next step is to find the target in the image and make positioning. In this experiment, we have used two methods: the background difference method and the extraction method based on the gray feature so as to find the target in the image.

After the foregoing theoretical analysis and comparison and the related hardware configuration, we select the background difference method to conduct the motion detection in "Virtual Badminton" system, companied by the background updates, feature point statistics and other methods, which make this method in the system more stable and accurate. As for the environmental effect on the system, we have adopted the updating method for the background image in order to eliminate the negative effects. The background checking method in the system can be divided into the following steps: open the camera to collect images and initialize the related parameters and set a separate logo bit Flags to detect whether there are any moving objects. We can initialize it as the team LSE; we have set a counting machine, which is used to make statistics about the number of the features point’s changes. Its initial value is set as 0; we have also set two closed values, one is used to make segmentation. When the absolute value of the pixel number is greater than the threshold value, the counter should be added by 1. The other threshold value is used to make comparison with the counter. Only when the value of the counter is greater than that of the threshold, we can regard that there are moving objects in the scene, and then set the value of the Flags as TRUE.

Noise Processing

Noise processing in this system is based on the mathematical morphology. We have found that the binary image after the background differences contains the isolated noise points, which have a certain influence on the system. Therefore, we need to deal with the problem first.

Mathematical morphology is targeted on the morphological characteristics of images, which describes the basic features and structures of the images. Under normal circumstances, morphological image processing adopts the method of neighborhood structure elements. As for the adjacent elements in the pixel locations, we should make a specific logic operation and focus generally on the binary images as the processing objects.

Mathematical morphological operation is based on the erosion and dilation operations. Erosion means eliminating the border of the objects and making the border contract inward so as to eliminate the isolated noise points and separate those tiny connections. Its mathematical expression is as follows:

$$E = X \odot B = \left\{ (x,y) \middle| \hat{B}_{x,y} \subseteq X \right\} \quad (1)$$

X represents the image after the binaryzation. B represents the structural element used for corrosion. E represents the result arising from corrosion, which is also a binary image. Dilation and erosion act oppositely. It is a boundary expansion in the binary objects, and the expansion can fill the empty holes left by the image segmentation. Its mathematical expression is as follows:

$$D = X \oplus B = \left\{ (x,y) \middle| \left[\hat{B}_{x,y} \cap X \subseteq X \right] \right\}$$

X is an image after the binaryzation, and B represents the structural element used for expansion. We can use the structure element B to make the expansion processing with the image X, and reserve the results in D. Although the effect of erosion and dilation is just the opposite, yet it is not an inverse operation. Both of them are irreversible. The opening and closing operation are created in accordance with its irreversibility.

The opening operation is to corrode the image first to remove the isolated noise points. Then we carry out the expansion operation again so as to fill the holes left by the corrosion. This method has a lot of functions, including eliminating the small goals, separating the objects at the slender points, and smoothing the large borders without changing its area. Its mathematical expression is as follows:

$$X \circ B = (X \odot B) \oplus B \quad (2)$$

X is the image to be processed. B represents the structural element to be conducted the opening operation. The binary images in the system based on the background subtraction are accompanied by the noise points, so the open operation should be required to eliminate those isolated noise points and to make the images smoother. This method can effectively improve the segmentation effect and is conducive to the accurate positioning.

Feature Extraction

In practice, the grayscale image is very sensitive to the light and in many cases, a certain gray feature may introduce a great deal of noises. So at the beginning of the database design for the features of the gray values, we have taken the right mouse button event into account. We eliminate the gray-scale features liable to introduce too many noises from the feature database, which means assigning the values of the corresponding position in the database as 0.

The procedure in this section is as follows: First of all, read the images captured by the cameras at both sides, establish the database of the gray scale features through procedures, draw a rectangular area with the left mouse button events, which include the target areas in the image. After that, the procedure shall automatically make statistics on all the gray values in the area, and set the values of the gray feature in the corresponding position as 1, the other locations as 0. Finally, the image will be turned into a binary image.

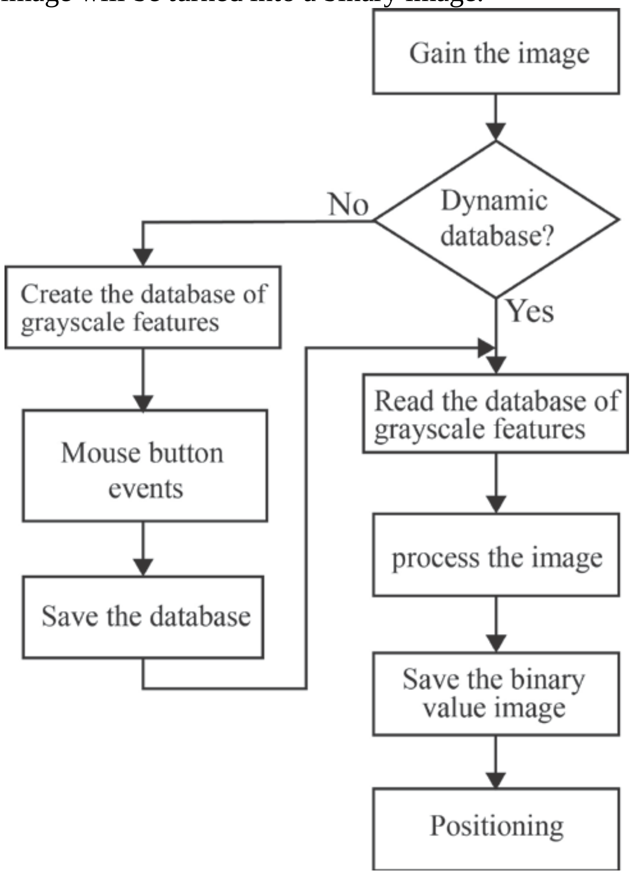

Figure 2. Flow Chart Based on the Grayscale Features

Positioning of the Badminton Target

After the target is extracted from the background image, we can conduct the target positioning by finding the centroid location of the target for the later calculation of the three dimensional space coordinates in the stereo vision system. The positioning point method is described as follows: While making the background difference of the images, we should establish a labeled area with same size as the image, and initialize the value as FALSE. If the absolute value of the pixel points difference is greater than some other value, we shall set the corresponding pixel value as 255, and set the value of the corresponding place in the region as TRUE, which means that we regard the points as the target objects.

Acquisition and Reconstruction of the Three-dimensional Information in the System Target

The stereo vision system composed by two parallel cameras is the simplest binocular stereo vision system. Its image-forming principle is as follows: The distance between the two cameras' projection centers is called the baseline distance, which represented as B. As for the object P in the space, the coordinate of the corresponding image on the cameras is $Pleft\left(Xleft, Yleft\right)$' $\Pr ight\left(Xright, Yright\right)$.respectively. Because the two cameras are in the same plane, the feature point $P\left(X_w, Y_w, Z_w\right)$ corresponding to the image in the left and right cameras has the same coordinate, that is $Yleft = Yright = Y$. Based on the principle of triangulation, its relationship can be obtained as follows:

$$\begin{cases} X_{left} = f\dfrac{x_w}{z_w} \\ X_{right} = f\dfrac{\left(x_w - B\right)}{z_w} \\ Y = f\dfrac{y_w}{z_w} \end{cases} \quad (3)$$

Usually, the two cameras are not placed in parallel; there is always a certain angle between the optical axes of the two cameras. The schematic diagram of normal binocular vision is shown as follows. If we change the target object's three dimensional coordinate $P\left(X_w, Y_w, Z_w\right)$ into the camera's three dimensional coordinate $P\left(X_c, Y_c, Z_c\right)$, the transform is as follows:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \quad (4)$$

R is the rotation matrix, and T is the translation vector. In this system, we should firstly obtain the centroid of the target point in the two dimensional image. Through the transform relationship between the centroid and the two cameras, we can obtain the three dimensional coordinate P(xc, yc, zc) under a certain camera coordinate system. Then we can obtain the three dimensional coordinate $P\left(X_w, Y_w, Z_w\right)$ in world coordinate system with the combination of the formula. After we have get the left camera coordinate system in the three dimensional space points, we can calculate the three dimensional points of the target in the world coordinate system. At this step, we need to enter the transformation relationship between the coordinate system of the left camera and the world coordinate system; the three dimensional coordinate we obtain from the first step of the left camera coordinate system; as well as the intrinsic and extrinsic parameters of the camera. In this step, we mainly used the Halcon function affine mat3d point 3d(). The output value is the value of the three dimensional information in the world coordinate system. In the "Virtual Badminton" system, we have set the origin of the coordinate system at the left corner on the white line of the world. The right side is the positive direction of X-axis, while the back side of it is the positive direction of the Y axis, and the direction to the field is the positive direction of the Z-axis. The unit of the data we get is "m". After we have gained the data, we converted the unit to "mm".

Through the introduction of principles stated above, we have gained some relative knowledge about the binocular stereo vision system and the transformation relationship between the various coordinate systems. Through the realization of the technology in the above two chapters, we have gotten both the internal and external parameters of the cameras, the distortion coefficients, as well as the centroid of target object in the two-dimensional image. The next step is to obtain the three dimensional coordinate of the object through the corresponding calculation. In this section, we have firstly studied the relevant theories on stereo matching, combined with the characteristics of this system to provide the method applied in the system. At the same time, the author has also discussed the relative knowledge about the curve fitting, so as to smooth the badminton movement curve.

Experiment and Evaluation

As can be seen from the data in table 1, we can get the simulation curve as shown in Figure 3, where the 50th point is the lowest one. If the Z value we have obtained in the original data is less than 60, then we would not match the points whose Z value is 60. Otherwise, we can match the corresponding point whose Z value is 60.

Table 1. Data after Fitting

points		point coordinate(mm)	
	X	Y	Z
1	450.419	924.613	1046.713
2	449.012	908.329	1031.964
…	…	…	…
…	…	…	…
…	…	…	…
45	387.101	191.826	169.943
46	385.694	175.542	145.508
47	384.287	159.258	120.858
48	382.879	142.973	95.993
49	381.473	126.689	70.912
50	380.864	119.648	60
51	370.601	76.057	169.373

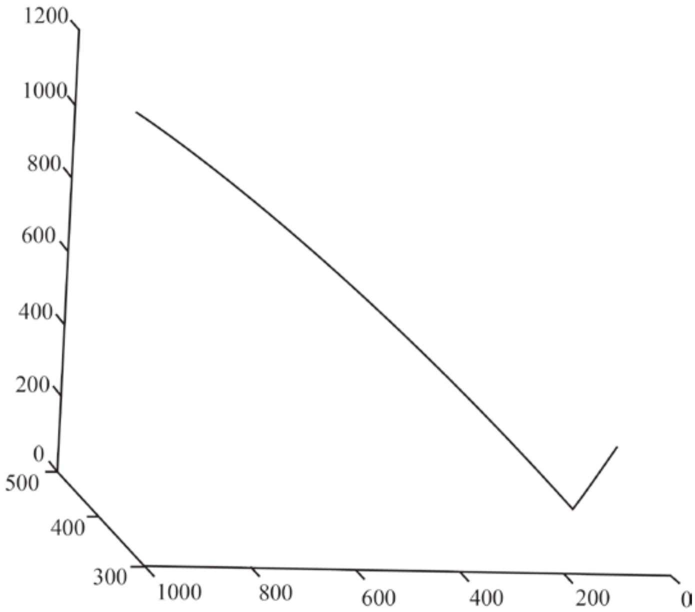

Figure 3. Simulation Curve after Fitting

In the "Virtual Badminton" system, we adopt the background difference method to detect the moving targets, make target positioning and conduct the image smoothing with the help of the open operation in the mathematical morphology. The results are shown in Figure 4. In Figure 4, (al) and (bl) are background images shot by leaf and right cameras, (a2) and (b2) are images with moving targets shot by leaf and right cameras, (a3) and (b3) are images after the segmentation positioning. The crossing center in the white area is the two dimensional centroid of the object obtained by the system. Each camera can capture 8 to 10 images when positioning before the first landing and we should handle these images. A group of images is shown in Figure 4.

The three dimensional reconstruction in the "Virtual Badminton" system is a playback system as a validation. When the audience throws the shuttlecock to the negotiating area, the system will automatically replay the reconstruction results to allow the audience to clearly see the first landing position of the badminton. Therefore, in this module, we have a high requirement for the art design, which needs to be produced by the professional animators. Consequently, we do not describe it in details here. Figure 5 is a screenshot of the three-dimensional reconstruction effect design. After the validation of a number of data, it turns out that the three dimensional reconstruction has a good effect.

A group of data can be obtained by the foregoing steps. The storing format of each line is as follows: the x-axis and y-axis of the target centroid in the image shot by the left camera; the x-axis and y-axis of the target centroid in the image shot by right camera. There are totally four data in each line and the corresponding data of images involved in this paper are shown in Table 2.

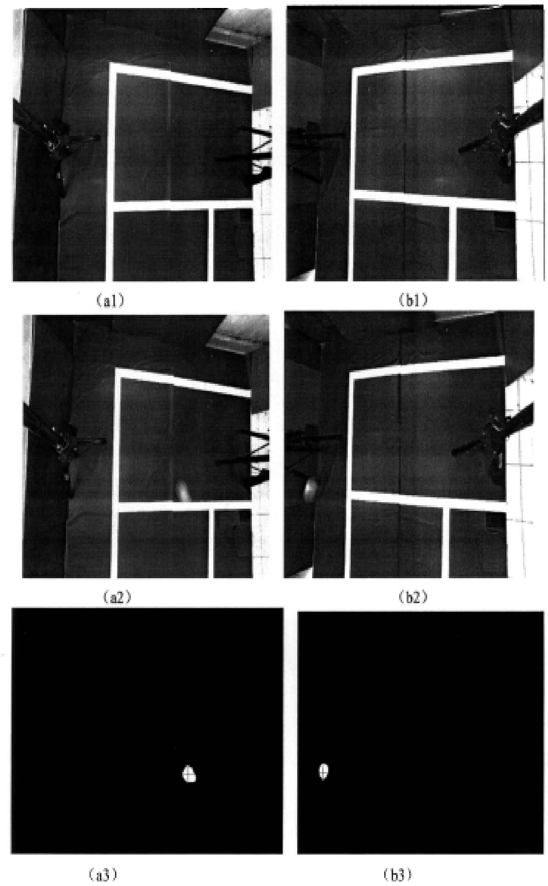

Figure 4. Result of Positioning Experience

Table 2. Target Centroid

Point	Centroid in the left image		Centroid in the right image	
	X-axis	Y-axis	X-axis	Y-axis
1	668	572	800	800
2	642	516	72	459
3	617	461	90	421
4	608	433	95	397
5	552	353	144	324
6	533	324	161	298
7	516	298	177	275
8	499	274	195	256
9	468	234	224	218
10	440	203	251	191
11	416	179	276	169
12	410	155	281	159
13	413	140	268	130
14	421	121	256	114
15	432	87	248	99

The first of above data includes point 800, while the image size is 780 x 582, which means that there's no object in the image and the tennis has no access to the vision scope of the right camera. Consequently, such data would be omitted in subsequent calculation of three-dimensional point and no error would appear.

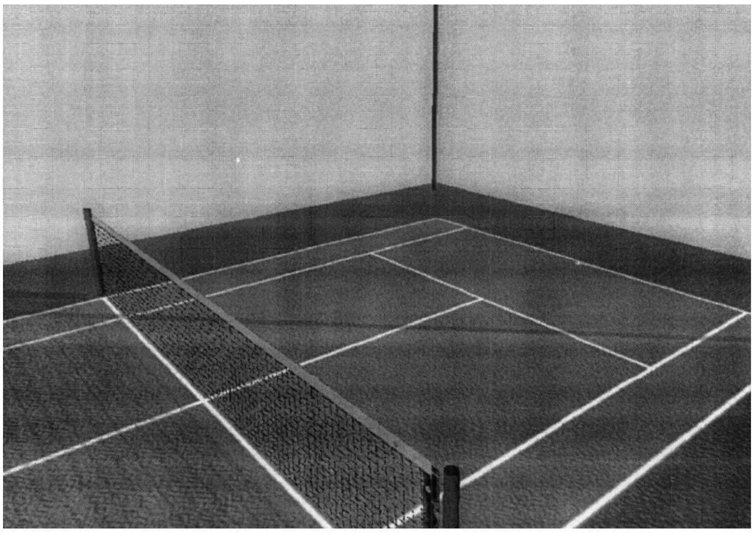

Figure 5. Effect Picture of Three-dimensional Reconstruction

Conclusions

The computer vision technology has gone through decades of development. It has been widely used in the national defense and security, industrial production, traffic monitoring, as well as people's daily life and other areas. Combined with the "Virtual Badminton" system, this paper has introduced the current research status of binocular stereo vision at home and abroad, and discussed about the camera calibration, moving target detection and positioning, stereo matching, three-dimensional reconstruction and other techniques.

References

1. Zheng N. (2012) Computer Vision and Pattern Recognition [M], National Defence Industry Press.
2. Haritaoglu I., Harwood D., Davis L. (2014) W4: Real-time Surveillance of People and Their Activities. *IEEE Trans Pattern Analysis and Machine Intelligence*, 22(8), p.p.809-830.
3. Karmann K., Brandt A. (2013) Moving Object Recognition Using an Adaptive Background Memory. *Proc. of Int. Conf. on Time-varying Image Processing and Moving Object Recognition*, p.p. 245-258.
4. Stauffer C., Crimson W. (2008) Adaptive Background Mixture Models for Real-time Tracking. *Proc. of Int. Conf. on IEEE Computer Vision and Pattern Recognition*, p.p.246-25.
5. Lipton A., Fujiyoshi H., Patil R. (2012) Moving Target Classification and Tracking from Real-time Video. *Proc. of Int. Conf. on IEEE Workshop on Applications of Computer Vision*, Princeton, USA, p.p.8-14.
6. Collins R. (2013) *A System for Video Surveillance and Monitoring: final report*, Carnegie Mellon University: Pittsburgh.
7. Silverstone C., Zhang X. (2011) Moving Targets Real-time Detecting Algorithm Based on Sequence Image. *Optical Technology*, 10(02), p.p.215-217.
8. Cai H. (2014) *Moving Object Detection Algorithm in Video and Application*. Zhejiang University: Hangzhou.