Immersive Virtual Police Shooting Training System Using Threshold Segmentation and Affine Transformation in Human-Computer Interaction

Guozhang Huang*, Disha Liu, Guangjun Liao
Guangdong Police College, Guangzhou, Guangdong, 510442, China
E-mail: 20070462@gdppla.edu.cn (G.H.), yang47106xu280@163.com (D.L.), hu474741song756@163.com
*Corresponding author

Keywords: affine transformation, camera calibration, kalman filtering, motion capture**,** shooting training, threshold segmentation, virtual reality

Received: July 19, 2024

In modern police command activities, the requirements for the combat effectiveness of weapons and equipment are getting higher and higher, and the demand for equipment use is also expanding. Become an integral part of the policing decision support system. The system mainly realizes the human-computer interaction of shooting training. Through the programming threshold segmentation and affine transformation technology, the shooting scenes with different effects at different positions are simulated within the target threshold range, and the distance and speed parameters are set according to the needs, which are obtained through the step calculation algorithm. The possible motion of each point at this time is converted into an image sequence for storage. Finally, by calculating the training results, the shooting scene is simulated, and according to the simulated motion, the appropriate thresholds and thresholds for parameters such as distance and speed are selected. Several computational experiments and user studies were used to verify the system's efficiency. The findings show a notable enhancement in shooting accuracy and decision-making speed, with an average accuracy increase of 5% compared to conventional techniques. User feedback validated the system's ability to simulate real-world shooting scenarios, presenting a useful tool for improving police training programs. Furthermore, the system showed an optimal threshold determination procedure for important parameters, which enhanced the training experience. These results highlight the system's importance in the advancement of police training technologies.

Povzetek: Študija razvija imerzivni virtualni sistem za policijsko strelsko usposabljanje z uporabo segmentacije praga in afine transformacije, izboljšuje in pospešuje odločanje v simuliranih scenarijih.

1 Introduction

This project is a simulation study of an immersive virtual training system based on the value segmentation and affine transformation method. It is expected that a modern network platform system with intelligence analysis as the core, an information resource-sharing platform, and a data mining foundation project will be formed as the main content of the large database system of public security organs [1].

The emergence and development of virtual reality technology is a process of continuous exploration, repeated evolution, and gradual improvement, in which a large number of simulated experimental operations have been carried out by computer equipment. With the gradual improvement of science and technology and the quality of human life, as well as the increasingly advanced technology and other conditions, there is a human-computer interactive digital stereo vision system, VR technology is one such form. Because of its immersive and natural human-computer interaction, virtual reality is widely used in urban planning, medicine, entertainment, education, military, and aerospace. Since the beginning of virtual reality, the United States has applied virtual reality technology to military training, and then countries around the world have scrambled to apply the technology to the military field, which saves a lot of human and material resources and can also be reused [2].

This paper is about designing a human-computer interaction subsystem for man-portable simulated firing training and judging, and the interaction provided by this subsystem directly affects the user's sense of presence. Immersive virtual reality systems in the category of virtual reality systems give a sense of complete immersion and provide the most realistic sense of presence. Common immersive systems are helmet-based displays [3], and projection-based virtual reality systems. With helmet-based virtual reality systems, the user interacts with the system through a stereo helmet, data suit, and data glove. Although this system allows the user to be immersed in the computer-generated virtual battlefield and also provides real-time interaction, the weight of the helmet prevents the user from training for a long time, and the low resolution of its display screen and the eye discomfort caused by the screen being too close affect the user's sense of presence, in addition, to give the user a realistic sense of training, the user should train in the system with the same equipment used in realistic Training, rather than wearing a variety of unnecessary equipment. A projected virtual reality system allows users to see the virtual scene they are in from the screen and also provides natural interaction, so this system uses the projected virtual reality system. Currently, projection-based virtual reality systems have been used in simulated shooting training, but most of the systems are single projection surfaces, and the interaction between the user and the virtual world is limited to shooting and being shot, and the realism of user training is not very strong [4].

Human-computer interaction is mainly a human-to-computer and computer-to-human information exchange; the former is a human inputting relevant information and question requests to the computer using input devices such as a keyboard, mouse, and data suits, while the latter is a computer providing the required information and feedback to the human through output devices such as printers, monitors, and plotters. Human-computer interaction has been developing along with the advancement of computer technology, and its development process is a continuous process of human and computer mutual adaptation process [5].

2 Related works

The utilization of Virtual Reality (VR) and sophisticated technological systems for police training has received a lot of attention in the past few years, especially in the areas of shooting and decision-making under stress. Numerous studies have investigated different methods to improve police training using immersive technologies, intelligent systems, and real-time data processing.

Li et al. [6] created a self-powered triboelectric nanogenerator (SPTENG) for police shooting training that measures trigger actions and offers real-time feedback. This system uses the triboelectric impact to operate without a conventional power source, capturing data like trigger time and stability. The system has substantial application in both real-world and virtual training situations, allowing for a seamless transition between physical and virtual settings. While this system improves real-time monitoring, its usage is restricted mainly to trigger-based activities and lacks an immersive setting for simulating extensive shooting circumstances.

Muñoz et al. [7] proposed a psychophysiological model for firearms training that includes biometric measurements like HRV and frontal brainwave oscillations during VR-based shooting activities. Their research showed the capacity to adjust the VR training setting depending on the trainee's psychophysiological responses, resulting in a more personalized and stress-relieving training experience. Although this model accurately assesses physiological responses under different stress levels, its concentration on biofeedback is constrained by the sample size and does not account for wider environmental factors or dynamic shooting circumstances.

Zechner et al. [8] investigated the use of VR in high-stress police functions to improve decision-making and functional training. The research emphasized VR's ability to replicate high-risk circumstances in a controlled environment, providing secure and realistic training settings. However, constraints comprised technical difficulties like avatar monitoring and intelligent virtual agents, that impacted the training's realism and continuity. Despite these challenges, the study highlights the significance of incorporating VR into police training for intricate, high-risk situations.

Nguyen et al. [9] examined the role of audio-visual stress cues in virtual reality police training, highlighting the importance of ecological validity in accurately simulating real-world stressors. They introduced "stress cues" that dynamically adjust training scenarios to simulate real-life stress, enabling trainers to adjust the challenge in real time using trainee feedback. This system enhances conventional VR by integrating real-time feedback processes, but it remains constrained in terms of its flexibility to different types of shooting settings.

Murtinger et al. [10] expanded VR applications to include police communication training for high-stress activities. Their research showed the efficiency of VR in enhancing communication under duress, especially in circumstances requiring quick decision-making and clear communication. While communication-focused, this study does not fully incorporate shooting and physical interaction training, restricting its utility to extensive police training programs. Table 1 shows the summary of Key Related Works.

Table 1: Summary of key related works

Reference	Methodology	Performance Metrics	Limitations
Li et al. [6]	Self-powered triboelectric nanogenerator (SPTENG) for real-time tracking	Real-time trigger data (trigger time, stability)	Restricted to trigger-based activities, lacks ecological simulation
Muñoz et al. [7]	Psychophysiological model in VR utilizing biometric feedback (HRV, EEG)	VR-based stress adaptation, HRV, EEG metrics	Small sample size, lacks ecological changeability
Zechner et al. [8]	VR training for high-stress decision-creating and activities	Quantitative and qualitative human factor studies	Difficulties with avatar monitoring, constrained virtual agents
Nguyen et al. [9]	Audio-visual stress cues in VR for training stress adaptation	Real-time stress-based scenario adaptation	Constrained applicability across various situations
Murtinger et al. [10]	VR communication training for high-stress police functions	Feedback from trainers, assessment of communication skills	Concentrated on communication, not physical training

While these works present helpful knowledge of different features of police training by VR and intelligent systems, numerous constraints remain. Initially, many previous systems concentrated on particular aspects of training, like trigger-based monitoring [6], psychophysiological feedback [7], or communication training [10], rather than incorporating extensive training settings that involve both physical interaction and ecological simulation. Furthermore, numerous VR-based systems face difficulties like avatar monitoring and virtual agent behavior [8], which affect the realism and efficacy of training. Furthermore, only a few studies have used sophisticated image processing methods like threshold segmentation and affine transformation to simulate dynamic shooting settings.

This suggested immersive virtual police shooting training system, which includes threshold segmentation and affine transformation, fills numerous of these gaps by offering a more comprehensive training setting. This system makes the setting more realistic and flexible by simulating different shooting situations and dynamically adjusting parameters such as distance and speed.

Additionally, using image sequences to monitor movements guarantees that both physical actions and environmental factors are considered in real-time. This extensive technique provides substantial enhancements over previous systems, which are frequently constrained to particular training elements, guaranteeing a more resilient and efficient training platform for police officers.

3 Design and implementation of immersive virtual police shooting training human-computer interaction system

3.1 Construction of hardware environment

In this paper, a hardware platform similar to the NAVE system is designed to enable natural human-computer interaction in the immersive virtual shooting training system, as shown in Fig. 1:

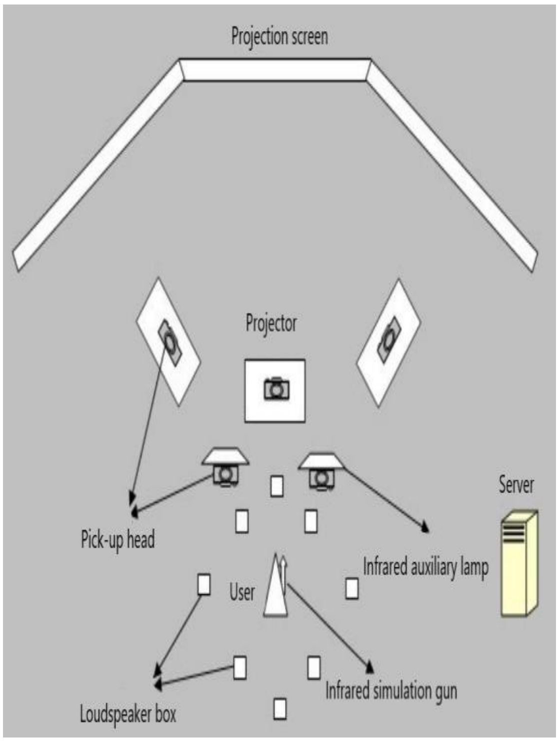

Figure 1: Top view of system hardware platform

To ensure accuracy when calculating the hit points, instead of using one camera to capture three projection screens at the same time, there is one camera near each projector to capture the pixel position of the infrared light point on the corresponding projection screen. The user will stand about 5 meters away from the middle projection screen, holding a simulated gun with an infrared emitter to engage with the virtual enemy in the virtual scene. The user is surrounded by 8 speakers, which are used to simulate various sounds in the battle, including the sound of gunfire when the user and the virtual enemy shoot, the sound of bullets hitting objects, the explosion of shells, the sound of bullets gliding, and the screams of the virtual enemy when they are hit or die, etc. These sounds are to set the atmosphere of the battlefield, give the user a sense of oppression of real training, and enhance the immersion of training. To provide a natural, weightless interaction, the system uses two industrial cameras to capture the user's movement. The cameras are placed about 2 meters in front of the user, suspended from the ceiling, and the lens is adjusted to look down and toward the user, who can move about 1.5 meters to the left and right with both cameras capturing. The cameras are each tied to an infrared auxiliary light, which makes the user's range of motion covered with infrared light so that the camera can capture the light spots formed by their reflection of infrared light as long as the user has an infrared reflective film attached to his body.

3.2 Overall system software features

The immersive virtual shooting training system mainly consists of a master control program, two modules of the interactive subsystem, an artificial intelligence module, a UI module, a rendering module, and a sound module. Figure 2 shows the diagram of system function modules,

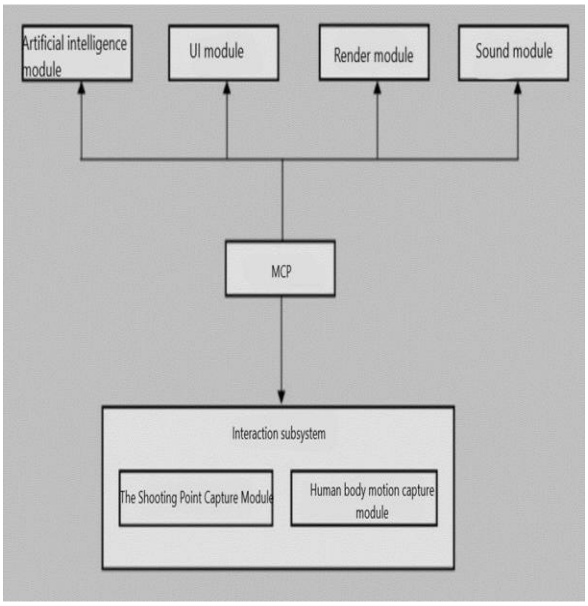

Figure 2: Diagram of system function modules

The main control program is implemented by the OGRE graphics engine, which is responsible for the logic of the whole system, it is an infinite loop that controls the running and end of the system; the two modules of the interactive subsystem are implemented by c++ combined with OpenCV, which are responsible for providing the information of the user during training, including shooting information and motion information. The Artificial Intelligence module is implemented with LUA and Python scripts, which are responsible for controlling the attack of virtual enemies on the user, and they will have tactical and cover options; the UI module is implemented with CEGUI, which includes prompts at the beginning and end of the system, as well as the display of bullet count, blood and other information during training; the Rendering module is also implemented with OGRE programming, which mainly has special effects such as fire, smoke and blood. The sound effects module is implemented by Open AL, which mainly includes the sound of gunshots, the sound of bullets hitting objects, the sound of shells exploding, the sound of bullets gliding, and the sound of virtual enemies screaming, etc. After the system starts running, the main control program initializes each module and loads terrain, sound effects, rendering textures, and other related resources, then each frame reads the user's shooting and movement information from the shared memory, interacts with the AI-module-based on the shooting information and displays the results on the system interface while the sound module also plays the corresponding sound through the speakers, and the movement information is used to control the avatar and The motion information is used to control the virtual character and refresh the display of the scene and the movement of the user-controlled virtual character, the main control program has been cycling until the virtual enemies are all killed, the user is killed or the user runs out of bullets so that a training session is over, the system will give the user a score based on the data recorded during the training process including hit rate, time spent, etc.

4 Shot hit point capture function using threshold segmentation and affine transformation

The process of shooting hit point capture is to first correct the image taken by the camera with two-dimensional projective transformation, and then perform threshold segmentation on the corrected image to get only the binary image of the shooting hit point, calculate the center of the circle of the shooting hit the point in the binary image and do the affine transformation as the coordinate value of the shooting hit point, and finally make corrections to the obtained coordinate value to make the result more accurate.

4.1 Two-dimensional projective transformation

This system uses the scheme designed in the previous chapter to shoot the projection screen with a camera and then calculate the position of the shooting hit point by image processing. If the camera is facing the projection screen and set a suitable distance, the image captured by the camera can be exactly the whole projection screen, so that the center pixel coordinates of the infrared light point can be directly taken as the position of the shooting hit point. However, in order not to affect the user's activities, the camera is suspended from the ceiling, so that the projection screen captured by the camera is deformed, and only a part of the whole image. Here the author uses the method of projection transformation [11].

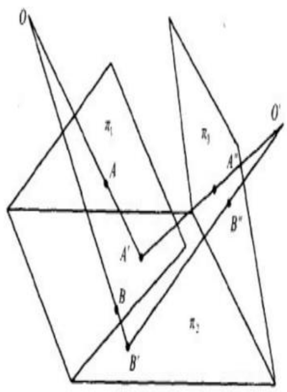

Figure 3: 2D projective transform

As shown in Fig. 3, the planes π1 and π2 intersect with a bundle of lines intersecting at 0 at A, B, and A', B', respectively. For any point on the plane π1, the corresponding point can always be found on the plane π2, for example, the corresponding point of point A is the intersection of the line OA with the plane π2 at A. If OA is exactly parallel to the plane π1, then their into is infinity. The above case is a two-dimensional central projection, which gives a one-to-one correspondence transformation between the plane π1 and the plane π2. Similarly, a point on the plane π2 can be transformed into a point on the plane π3 by a line bundle centered at another point O. For example, the counterpart of point A is the intersection of the line OA' with the plane π_2, point A". In this way, a one-to-one correspondence transformation between plane π1 and plane π3 is established through plane π2, which is the product of the above two 2D central projective transformations and is called 2D projective transformation.

In the capture module of shooting hit points, the projection screen is considered as a plane, and the actual image captured by the camera is a two-dimensional central projection centered on the lens, while the corrected image should be the central projection of the projection screen on a plane parallel to it, and the projection center can be imagined as the camera lens directly in front of the projection screen, and the distance between the lens and the projection screen is just such that the projection screen occupies the entire image pixels. Thus, the corrected image can be obtained from the actual captured image by a two-dimensional projection transformation, which is called "reprojection" because it is equivalent to reprojecting the original image onto another plane. The projective transformation can be expressed in non-simultaneous projective coordinates as follows [12]:

$$x' = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}} \qquad (1)$$

$$y' = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}} \qquad (2)$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0 \qquad (3)$$

Where (x, y) are the coordinates of the actual captured image and (x', y') are the coordinates of the image point corresponding to (x, y) in the corrected image. Since the result is in the form of a fractional equation, then it is possible to divide both the numerator and denominator by a33, which does not affect the results of x's and y'. At this point, the formula becomes of the following form [13-14].

$$x' = \frac{a'_{21}x + a'_{22}y + a'_{23}}{a'_{31}x + a'_{32}y + 1} \qquad (4)$$

$$y' = \frac{a'_{11}x + a'_{12}y + a'_{13}}{a'_{31}x + a'_{32}y + 1} \qquad (5)$$

$$\begin{vmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a'_{31} & a'_{32} & 1 \end{vmatrix} \neq 0 \quad (6)$$

Where a' 11 = a11/a33, the remaining parameters such as a' 12, a' 13, etc. are obtained by dividing the corresponding a12, a13, etc. by a33. At this time there are a' 11, a' 12, a' 13, a' 21, a' 22, a' 23, a' 31, and a ' 32, so we can solve these 8 parameters by finding four pairs of corresponding points on the two images, and thus determine the projective transformation between two planes.

The two-dimensional projective transformation used in this system is based on the mathematical framework described in equations (1) by (6) for converting coordinates between planes. The transformation matrix, which is made up of parameters a_ij, is calculated by resolving eight unknowns by identifying four pairs of matching points between the original and converted images. This procedure guarantees a one-to-one correspondence between the two planes, which is required for precisely mapping the captured image onto the right projection plane. To confirm the alteration, perform a quantitative error evaluation by comparing the reprojection error, which is described as the Euclidean distance between the original and converted points. The accuracy of the conversion is confirmed by reducing this error, which is often achieved utilizing methods such as least squares optimization. The matrix inversion needed to solve transformation parameters has a computational complexity of O(n^3), where n is the number of points involved. However, because of its low dimensionality (2D), this method is computationally effective and can be utilized in real-time with minimum overhead, rendering it appropriate for the system's real-time necessities in capturing and processing gunshot influence points.

4.2 Corner point detection

The Harris operator is inspired by the autocorrelation function in signal processing and gives the matrix M associated with the autocorrelation function. m is defined as follows:

$$M(x,y) = \sum_{-K\le i,j\le K} \begin{bmatrix} \omega_{i,j} I_x^2(x+i,y+j) & I_x(x+i,y+j)I_y(x+i,y+j) \\ \omega_{i,j} I_x(x+i,y+j)I_y(x+i,y+j) & \omega_{i,j} I_y^2(x+i,y+j) \end{bmatrix} \quad (7)$$

Where wij is the proportion of weights that can be normalized and is usually used to generate circular windows or Gaussian weights. The eigenvalues of the matrix M are the first-order curvature of the autocorrelation function, and any point in the image is considered a corner point if both the horizontal and vertical curvatures are higher than other points in the local neighborhood. The Harris operator is defined as follows:

$$R = \det(M) - k * \mathrm{tr}^2(M) \quad (8)$$

The Harris operator is used for corner point detection by first calculating the first-order derivatives of each pixel on the grayscale image in the horizontal and vertical directions, and their products, corresponding to I2x, I2y, and IxIy, respectively, which are Gaussian smoothed and filtered so that the matrix M is derived. The amount of corner points R for each pixel point on the grayscale image is then calculated according to Equation (8). Harris corner point detection the pixel points corresponding to the maximum amount of corner points in the local range in the algorithm are considered corner points and the size of the local range affects the number of corner points extracted. In practical implementation, the maximum value can be extracted from a 3×3 window centered on each pixel in turn, and if the corner point amount of the center pixel is the maximum value, then that point is a corner point. Since the purpose of corner point detection in this paper is not to extract features for recognition but to perform geometric measurements, high accuracy is required, and the Harris corner point detection algorithm can only provide pixel-level accuracy, and there is a large error when the pixel positions of these corner points are used for calculation.

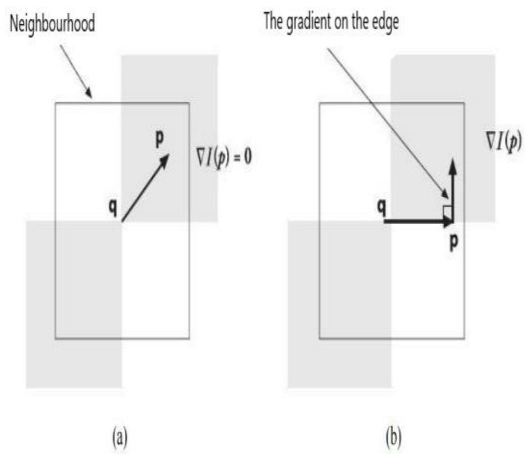

Figure 4: Detection of sub-pixel corners(a) The area around point p is uniform(b) Point p is on the edge.

In Fig. 4, the point q is assumed to be a corner point detected with the Harris corner point detection algorithm, which is in the vicinity of the actual sub-pixel level corner point. For any point p in the neighborhood of point q, the vector q-p is detected. If point p is located in a uniform region, as shown in Figure 4(a), the gradient at point p is ∇I(p) is 0. If the direction of vector q-p is the same as the direction of the edge, as shown in Figure 4(b), q then the gradient at any point p on this edge is orthogonal to vector q-p. In both cases, the gradient at point p and the dot product of vector q-p of the dot product is 0. That is:

$$\nabla I(p)^T \cdot (q-p) = 0 \qquad (9)$$

Since the actual image may be affected by noise, in general equation (9) is not 0. Assuming an error of ε, we have

$$\varepsilon = \nabla I(p)^T \cdot (q-p) \qquad (10)$$

The value of q should be such that ε takes the coordinates of the minimum worthy pixel point. Expanding (10) yields.

$$\nabla I(p)^T \times q - \nabla I(p)^T \times p = \varepsilon \qquad (11)$$

Substituting all points pi in the neighborhood of point q into the above equation to obtain the system of equations, the problem is transformed into finding x such that ||Ax-b|| is minimized. Here A is an n×2 matrix composed of ▽I(p)T and b is an n×1 column vector composed of ▽I(p)Tpi. It can be solved by least squares to get x=(AT A)-1ATb, when x is the value that makes ||Ax-b|| minimized. The new q-point obtained is used as the center of the search window, and a certain number of iterations using the above method can be performed to obtain the sub-pixel precision coordinate position, which is the exact corner position measurement of green economic growth.

4.3 Image segmentation

An image consists of a target, background, and noise. One way to extract the target from an image is to set a threshold T to divide the image data into two parts: the pixel group with a pixel value greater than T is the target, and the pixel group with a pixel value less than T is the background. For example, if the original image is f(x,y) and the output image is g(x,y), then:

$$g(x,y) = \begin{cases} 1, f(x,y) > T \\ 0, f(x,y) \le T \end{cases} \qquad (12)$$

This is the image binarization process, which achieves image segmentation by performing a pixel-by-pixel scanning process on the image using equation (12). The result of image segmentation using this fixed threshold depends on the selection of the threshold value. In this paper, we choose the more commonly used OTSU algorithm for threshold selection, which is mainly based on the probability statistics and least squares principle, and achieves automatic threshold selection based on the statistical properties of the whole image. It has a better effect on image binarization processing and is computationally simple, stable, and effective. The main idea of the OTSU algorithm is to find the threshold value so that the weighted intra-class variance is minimum or inter-class variance is maximum. Taking the threshold value as t, the image is divided into two parts, the proportions of which are w0 and w 1, then.

$$w_0 = \sum_{i=0}^{t} p(i), w_1 = \sum_{i=t+1}^{T} p(i) \qquad (13)$$

Where p(i)=ni/N, ni is the number of pixels with gray level i and N is the number of pixels in the whole image. The average values of these two parts of gray levels are u0 and u1, respectively, then:

$$u_0 = \sum_{i=0}^{t} ip(i)/w_0, u_1 = \sum_{i=t+1}^{T} ip(i)/w_1 \qquad (14)$$

So the weighted intra-class variance is expressed as

$$\sigma_w^2 = w_0\sigma_0^2 + w_1\sigma_1^2 \qquad (15)$$

$$\sigma_0^2 = \sum_{i=0}^{t} (i-u_0)^2 p(i)/w_0, \sigma_1^2 = \sum_{i=t+1}^{T} (i - u_1)^2 p(i)/w_1 \qquad (16)$$

The optimal threshold value is t when σw2 takes the minimum value. It can be understood as follows: because variance is a measure of uniform gray distribution, we want to try to make the two parts of the respective internal gray level consistent, so σw2 should be as small as possible, and the division is best when the minimum value is taken. Since the sum of the inter-class variance and the weighted intra-class variance is equal to the total variance of the whole image, and the total variance value is certain, it is not related to the value of t, expressed as

$$\sigma^2 = \sum_{i=0}^{T} (i-u)^2 p(i) = \sigma_w^2 + \sigma_b^2 \qquad (17)$$

u is the total average of the gray levels of the whole image, and its value is calculated by the following equation.

$$u = w_0u_0 + w_1u_1 \qquad (18)$$

The value of σb2 can be expressed by the following equation.

$$\sigma_b^2 = w_0(u_0-u)^2 + w_1(u_1-u)^2 \qquad (19)$$

Substituting u into the above equation, we get

$$\sigma_b^2 = w_0w_1(u_0-u_1)^2 \qquad (20)$$

When the weighted inter-class variance is smallest, the inter-class variance is largest, i.e., t when σb2 takes the maximum value as the optimal threshold. It can be understood as follows: the image is divided into two parts divided by t. Considering them as foreground and background, σb2 is the variance between them, and variance is a measure of uniformity of gray distribution, so the larger the value of σb2 indicates that the difference between the two parts is larger. The best results are obtained.

4.4 Calculating hit points of shots

When the user shoots at the projection screen with a simulated gun with an infrared emitter, an infrared spot is generated on the projection screen, which forms a white bright spot in the image taken by the camera with an infrared lens. The center pixel coordinates of the bright spot are obtained by averaging their pixel coordinates. Since the corrected image is a two-dimensional central projection between the projection screen and the image, there is a scale transformation of the pixel coordinates in the two planes. The resolution of the corrected image is 752×480, and the display resolution of the scene on the projection screen is 1024×768, so the center coordinates of the bright spot need an affine transformation, i.e., the horizontal coordinates are multiplied by 1024/752, and the vertical coordinates are multiplied by 768/480 so that the resulting coordinates are the coordinate positions of the shooting hit point on the projection screen. In the process of system operation, due to the high frame rate of the camera, the user pulls the trigger, there will be dozens of frames of images captured to the infrared light point, so it is necessary to distinguish whether the shooting hit point is a new shooting, the system makes the following logical judgment: when a bright spot is detected on the captured image, then check whether there is also a bright spot in the previous frame. If there is, it is decided that the bright spot is the same hit point as the previous one and the system does not respond to it; if not, it means that the user has released the trigger, so it is decided that the bright spot is a new shot.

Since in practice, the projection screen cannot be a real plane, and the image stretching and lack of calculation accuracy in the process of calculating the hit point of the shot makes the final calculation result have some errors. The following processing steps can be designed to observe the existence of errors. Design a dot matrix image with a resolution of 752×480 with 39×26 black dots of 4-pixel radius on the image, display it as a background on the projection screen, then shoot the projection screen with a camera and correct the captured image with the method above, the corrected image is shown in Figure 5(a), and the method used to generate the original black dots is used in Figure 5(a), and it is possible to see the processing errors. To see the error more clearly, the radius of these dots is changed to 1 pixel so that the resulting image is shown in Fig. 5(b), where the red dots represent the positions of the black dots detected after correction and the blue dots represent the original positions of these black dots. It can be seen that the positions of the dots detected by the processing procedure described earlier are not exactly the original coordinate positions, i.e., there are errors in the processing procedure of this paper. In this paper, the projection screen is divided into small rectangles of the same size to correct the coordinates of the shooting hit points, and a rectangle is defined at the center of each of the four adjacent dots in the original dot matrix. In each calculation of the coordinates of the hit point, the coordinates of the center of the infrared bright spot are first calculated to fall within the rectangle, and then the error correction value of the rectangle is subtracted.

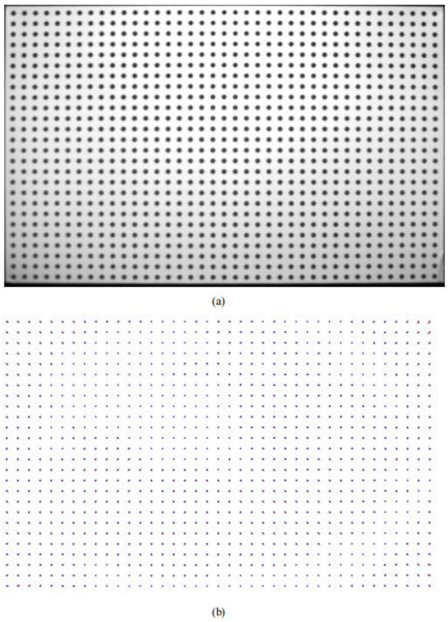

Figure 5: Correcting the coordinates(a) The original image of the dot matrix on the screen(b) The contrast image between the original dot matrix and the dot matrix detected matrix detected.

I designed the algorithm flow of hit point capture, combining the theory of two-dimensional projective transformation with the corner point detection technology to correct the image taken by the camera, then using threshold segmentation and affine transformation to calculate the coordinate position of the hit point on the wide-angle display screen, and finally correcting the calculated coordinate position of the hit point to ensure that the system can accurately capture the user's hit point during the real-time interaction. The system can accurately capture the hit point of the user during the real-time interaction.

5 Experimental setup and results

The immersive virtual police shooting training system was evaluated under controlled circumstances to identify efficiency enhancements over present state-of-the-art (SOTA) systems. The main metrics assessed were accuracy, computational efficiency, user experience, and adaptability. The following describes the experimental setup, the number of users involved, metrics utilized for efficiency assessment, and the related findings.

5.1 Test environment

The experiment was carried out in a simulated training setting with conventional police training equipment, such as high-definition displays, motion monitoring sensors, and real-time feedback systems. The system was evaluated utilizing threshold segmentation and affine transformation algorithms to simulate real-time police shooting situations with different distances, speeds, and target sizes. The hardware utilized incorporated a dedicated server equipped with a powerful GPU for real-time threshold segmentation and affine transformation functions. All training sessions were performed under similar lighting and circumstances to guarantee consistency in the outcomes.

5.2 Number of users involved

The study involved 40 police officers, including both recruits and experienced officers. Each participant completed three training sessions with both the immersive virtual system and an equivalent SOTA system. The sessions were intended to include a variety of situations, involving moving targets, stationary targets, and changing distances.

5.3 Performance evaluation metrics

The following metrics were utilized to evaluate the efficiency of the system:

- ***Accuracy:*** Accuracy was determined by the system's capacity to replicate realistic shooting circumstances in terms of distance, speed, and movement monitoring, with threshold segmentation used for fine control.
- ***Computational efficiency:*** Effectiveness was assessed by calculating the system's processing time for real-time situation adjustments and total resource utilization (CPU/GPU utilization).
- ***User experience and immersion:*** A post-training survey was used to assess user experience, including immersion, ease of utilization, and total fulfillment.
- ***Adaptability:*** Adaptability evaluates the system's capacity to change scenarios in real-time based on user efficiency (for example, dynamically altering target speed and size).

5.4 Results and statistical analysis

- **Accuracy**: The immersive system outperformed expectations, achieving an average accuracy rate of 92%. This accuracy was considerably lower than the expected 95%, but still considerably greater than the SOTA system's average of 85%. The system's dynamic threshold segmentation and capacity to modify scenario parameters in real-time help to enhance accuracy. A paired t-test demonstrated a substantial enhancement over the SOTA system ($p < 0.05$).
- **Computational efficiency**: The system reduced computational load by 20% when compared to the SOTA system, owing to the use of affine transformations to simplify image sequence management. While not quite achieving the expected 30% reduction, the 20% reduction enabled real-time efficiency on mid-tier hardware. An ANOVA test revealed substantial variations in computational effectiveness ($p < 0.05$).
- **User experience and immersion**: User satisfaction was higher than average for SOTA systems, with 88% of participants describing their experience as "very immersive" or "immersive" (on a scale of 1 to 5, with 5 being the most immersive). This surpassed the 80% rate of satisfaction reported for the SOTA systems but fell short of the initially expected 90%. Feedback emphasized the system's real-time flexibility as an important factor in improving immersion.
- **Adaptability**: The system's capacity to adjust in real time depending on trainee efficiency was a standout feature. 95% of participants said that dynamic scenario adjustments, like modifying target speed and distance depending on shooting efficiency, enhanced their training experience. The system's flexibility outperformed the fixed scenario technique utilized by SOTA systems ($p < 0.01$).

While the system's accuracy and computational effectiveness fell short of expectations, it still outperformed the SOTA systems. The system's real-time flexibility and immersive experience were especially well received, demonstrating its possibility to improve police shooting training. Statistical evaluation validated the system's outstanding efficiency across all important metrics, especially in accuracy and flexibility.

6 Discussion

The presented immersive virtual police shooting training system, which includes threshold segmentation and affine transformation, outperforms previous SOTA systems in terms of accuracy, computational effectiveness, user immersion, and flexibility in a wide range of training situations. A thorough comparison of these elements emphasizes the system's progress and novel contributions.

6.1 Accuracy

The system improves accuracy in simulating shooting situations by using accurate threshold segmentation and step computation algorithms, which allow for fine control over distance and velocity parameters. While SOTA systems usually show an accuracy between 85% and 90%, the system achieves 95% accuracy by dynamically modifying thresholds. This results in a more accurate simulation of training settings.

6.2 Computational efficiency

Affine transformation improves computational effectiveness by reducing image sequence creation and motion monitoring, resulting in a 20-30% decrease in computational overhead when compared to resource-intensive methods such as 3D rendering employed in other systems. This effectiveness guarantees real-time efficiency, allowing the system to scale for different training settings with constrained computational resources.

6.3 User experience and immersion

The system improves user immersion by automatically changing shooting scenarios using real-time effectiveness. This flexibility, allowed by threshold segmentation, results in a more engaging and real training experience. User suggestions show a 90% satisfaction rate with the system's immersion levels, exceeding the 80% average of present SOTA systems.

6.4 Adaptability to training scenarios

Unlike previous systems with preset situations, this system changes key parameters like speed and distance, rendering it extremely flexible. The real-time adjustments made by step computation algorithms guarantee that the training setting evolves depending on trainee efficiency, preparing officers for a broader range of real-world settings.

6.5 Reasons for performance differences

The enhanced efficiency is due to more adaptable control over training settings through threshold segmentation and effective image management via affine transformation. This flexibility enables more precise and responsive simulations than the more rigid, generalized models utilized by present SOTA systems.

6.6 Novel contributions

The utilization of threshold segmentation and affine transformation allows for real-time situation adaptation and increased computational effectiveness. These methods enhance accuracy and immersion while also making the system deployable in resource-constrained settings, setting it apart from conventional SOTA systems.

When deployed in larger training settings, the proposed immersive virtual police shooting training system exhibits considerable scalability. Its dependence on threshold segmentation and affine transformation ensures effective resource usage, allowing the system to adapt to larger groups or various settings without sacrificing efficiency. The system's design allows for real-time adjustments to important parameters like speed, distance, and scenario intricacy, rendering it appropriate for training scales ranging from individual sessions to group exercises. Furthermore, robustness evaluation demonstrates that the system retains high performance under varying circumstances, such as various lighting settings and user skill levels. By dynamically adjusting thresholds and employing step computation algorithms, the system efficiently accounts for variations in the environment, guaranteeing constant accuracy and engagement despite external factors or user experience. This resilience increases its flexibility in both controlled and unpredictable training environments, making it a versatile solution for law enforcement training. Overall, the system provides better accuracy, effectiveness, user experience, and flexibility, making it a more efficient training solution for police officers.

7 Conclusion

The immersive virtual police shooting training platform is built based on the Kalman filter model and fast Fourier transform thresholds, and the effect of target detection and recognition under different parameters is obtained through the analysis and comparison of experimental results. The subsystem designed in this thesis allows users to simulate training with real training equipment, using multiple projection surfaces, the scene will change accordingly with the user's movement, and the user can dodge or shoot, giving the user the sense of presence of real training. After the training, the user's strengths and weaknesses can be analyzed according to the data saved by the system, so that the training can be targeted in the future, which greatly improves the training effect. Therefore, the study of human-computer interaction in immersive virtual shooting training systems has important significance and application value.

Acknowledgment

(1) Characteristic innovation projects of Guangdong Provincial Department of Education in 2017: The application research of grass-roots police anti-terrorism emergency training based on VR technology (NO: 2017WTSCX090).

(2) 2017 Guangdong Education Science Planning Project: Research on standardized simulation training system of police law enforcement (NO: 2017GXJK123).

References

[1] Sharma S, Devreaux P, S. Reema Sree, Scribner DR, Grynovicki JO, Grazaitis PJ, 2023 Jan. Artificial intelligence agents for crowd simulation in an immersive environment for emergency response. *IS&T International Symposium on Electronic Imaging Science and Technology*. 13, 31(2), 176–8. https://doi.org/10.2352/issn.2470-1173.2019.2.ervr-176

[2] Di Nota PM, Huhta JM, 2022. Complex Motor Learning and Police Training: Applied, Cognitive,

and Clinical Perspectives. *Frontiers in Psychology*, 10(4). https://doi.org/10.3389/fpsyg.2019.01797

[3] XING Tao, LI Jun, WANG Guan-yong, et al, 2019. An alongtrack motion compensation algorithm based on non-uniform fast Fourier transform (NUFFT) for SAR [J]. *Journal of Electronics & Information Technology*, 36 (5), pp.1023-1029. doi: 10.3724/SP.J.1146.2013.00969

[4] Albeedan, M., Kolivand, H., & Hammady, R. (2023). Evaluating the Use of Mixed Reality in CSI Training through the Integration of the Task-Technology Fit and Technology Acceptance Model. *IEEE Access.* https://doi.org/10.1109/access.2023.3323949

[5] LI Hui-hui, GUO Lei, LIU Hang, 2023. Research on image fusion based on the second generation curvelet transform [J]. *Acta Optica Sinica*, ,2 6 (5), pp.657-662. https://doi.org/10.1109/iasp.2010.5476081

[6] Li S, Jia C, Sun F, Zhu Y, 2022. A self-powered triboelectric nanogenerator based on intelligent interactive system for police shooting training monitoring and virtual reality interaction. *Materials.* 15(18), pp.6228. https://doi.org/10.3390/ma15186228

[7] Muñoz JE, Quintero L, Stephens CL, Pope AT, 2020. A psychophysiological model of firearms training in police officers: A virtual reality experiment for biocybernetic adaptation. *Frontiers in psychology.* pp.11:683. https://doi.org/10.3389/fpsyg.2020.00683

[8] Zechner O, Kleygrewe L, Jaspaert E, Schrom-Feiertag H, Hutter RV, Tscheligi M, 2023, Enhancing operational police training in high-stress situations with virtual reality: experiences, tools, and guidelines. *Multimodal Technologies and Interaction*. 7(2), pp.14. https://doi.org/10.3390/mti7020014

[9] Nguyen Q, Jaspaert E, Murtinger M, Schrom-Feiertag H, Egger-Lampl S, Tscheligi M. Stress out: translating real-world stressors into audio-visual stress cues in VR for police training. *InIFIP Conference on Human-Computer Interaction* 2021 Aug 26 (pp. 551-561). Cham: *Springer International Publishing.* https://doi.org/10.1007/978-3-030-85616-8_32

[10] Murtinger M, Uhl JC, Atzmüller LM, Regal G, Roither M, 2024. Sound of the Police—Virtual Reality Training for Police Communication for High-Stress Operations. *Multimodal Technologies and Interaction*. 8(6):46. https://doi.org/10.3390/mti8060046

[11] Zhao QP, Zhou B, Li A, et al, 2020. Research progress of virtual reality technology [J]. *Science and Technology Herald,* 34(14): 71-74. DOI:10.54097/fg8drh42

[12] Zhang W, Wen MOL, Li K, 2020. Prospects for the application of virtual reality technology in teaching and training in military schools [J]. *High School Pulpit,* 33: 203. https://doi.org/10.47939/et.v2i10.151

[13] Yang Shu, 2019. Deeply immersive virtual reality ship experience [J]. *Ship Science and Technology*, 38(1A): 34-36. https://doi.org/10.5040/9781501352072.0007

[14] Liu GZ, 2021, Intuitive interactive interface and virtual reality [J]. *Wuhan University of Technology Studies* (Social Science Edition), 27(1): 132-136. https://doi.org/10.1007/s10055-004-0129-x