International Journal of Science and Engineering Applications
Volume 15-Issue 05, 52 – 56, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1505.1006

Design of a Table Tennis Motion Tracking System Based on Machine Vision

Yijun Liu
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: Table tennis is a globally popular sport that has developed rapidly and prosperously in recent years. Table tennis competitions have also become well-received large-scale sports events. Characterized by a fast rhythm, high technical requirements and excellent viewing value, table tennis is widely favored by the public. It not only occupies an important position in competitive sports events, but also serves as a common leisure activity for people of all age groups.

With the widespread promotion of table tennis competitions across the world, there is an increasing demand for improving the viewing quality and technical level of matches. In actual competitions, restricted by the small size and high moving speed of the table tennis ball, artificial referees are prone to judgment deviations and inaccuracies. To solve the above problem, this paper proposes a table tennis motion tracking system based on machine vision.

First, the imported table tennis match video is decomposed into sequential frames. Then, the RGB images after frame extraction are converted into HSV color space images. Combined with the orange color feature of the table tennis ball, irrelevant color interference information in the background is filtered out to obtain a single-target image containing only the table tennis ball. Next, hierarchical calculation is conducted on the processed images to acquire the center position of the ball and realize preliminary motion tracking. To optimize tracking precision, the Kalman filter algorithm is adopted to predict the subsequent movement position of the ball based on continuous coordinate data. Practical test results demonstrate that the system can stably track the moving table tennis ball and display the tracking and prediction results in the video in real time.

Keywords: object tracking; color detection; OpenCV; Kalman Filter; table tennis matches

1. INTRODUCTION

With the continuous advancement of artificial intelligence and computer technology, machine vision has evolved into an interdisciplinary discipline integrating computer science, artificial intelligence, image processing and other fields. It plays an increasingly vital role not only in cutting-edge national fields such as military industry and aerospace, but also in manufacturing, medical care and daily life scenarios[1]. As a core branch of machine vision, target detection and tracking technology has been widely applied in various industries, providing strong technical support for automation, intellectualization and high-efficiency operation[2].

For instance, autonomous driving, which is booming in the automotive industry, relies on target detection and tracking technology to identify and monitor vehicles, pedestrians, traffic signs and other key objects on roads, so as to improve driving safety and comfort. Similarly, forest fire alarm systems adopt relevant technologies to identify fire hazards rapidly and accurately with less manpower consumption, thereby reducing economic and ecological losses caused by forest fires[3]. It is evident that rational application of target detection and tracking technology can bring tremendous benefits to all walks of life.

In the sports industry, table tennis has gained widespread popularity and become a mainstream mass sport. Accordingly, combining target tracking technology with table tennis research has become an inevitable development trend. This paper proposes a machine vision-based table tennis tracking system to identify and locate the ball in real time, which is of great practical significance. Firstly, it can effectively improve the overall quality of table tennis matches. The system can capture the real-time position and motion trajectory of the ball quickly and accurately, providing objective data for referees and spectators to ensure impartial judgment. Secondly, it can analyze athletes' competition data including hitting speed and hitting angle, so as to deliver targeted training feedback and help athletes optimize technical movements and competitive performance. Thirdly, this research conforms to the intelligent development trend of sports technology, and provides new ideas and references for the innovative development of table tennis events and the sports industry[4].

Target recognition and tracking technology has been developed for decades, with a large number of algorithms emerging continuously. Selecting appropriate algorithms or multi-algorithm combination schemes to realize high-precision real-time detection and tracking of table tennis balls is the core research issue to be discussed in this paper.

2. A Target Tracking Technology Based on Color Detection

2.1 Introduction to Principles

2.1.1 Color Space Conversion

(1) Concepts of RGB and HSV Color Spaces

① RGB Color Space

The RGB color space is a color mode that generates a wide spectrum of diverse colors through the superposition of three primary colors: Red, Green and Blue. It is also known as the trichromatic color mode. Each of the red, green and blue channels can be divided into 256 gradient levels according to the spectral range. The superposition of the three channels can produce 256*256*256 distinct colors, exceeding 16 million in total. This quantity far surpasses the range of colors distinguishable by the human visual system and enables accurate representation of natural scenes. Hence, the RGB color space is also referred to as the natural color mode.

The RGB color space can be intuitively visualized as a cube established in a three-dimensional Cartesian coordinate system, as illustrated in Figure 2-1. The origin of the coordinate system

International Journal of Science and Engineering Applications
Volume 15-Issue 05, 52 – 56, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1505.1006

represents pure black, while the three coordinate axes correspond to red, green and blue respectively. The side length of the cube is defined as the parameter value of each color channel, which is divided into 256 units. Accordingly, the value range of the three primary color channels is normalized to 0–255.

Figure 2-1 RGB color space

② HSV Color Space

The HSV color space describes colors from a human visual perception perspective, which characterizes each color through three independent dimensions: Hue (H), Saturation (S), and Value (V). Hue determines the fundamental category of a color, such as red, yellow, green and other basic color tones. Saturation indicates the chroma purity of a color; a higher saturation corresponds to a more vivid color, whereas a lower saturation results in a dull and grayish visual effect. Value reflects the brightness of a color, with higher values representing lighter colors and lower values representing darker colors [16].

The HSV color space is geometrically modeled as a cone, commonly termed the hexagonal cone model, as shown in Figure 2-2. Hue is defined as the rotation angle around the V-axis, with a value range of 0-360°. Saturation is measured by the radial distance from a given point to the center of its corresponding circular cross-section, ranging from 0 to 1. Value is determined by the vertical distance from the circular cross-section to the apex of the cone, also normalized within the range of 0 to 1.

Figure 2-2 HSV color space

(2) Comparative Analysis of Advantages and Disadvantages

The RGB color space features wide application scenarios and intuitive comprehension, and it is the standard color mode for display devices. Nevertheless, it has prominent limitations in image processing. Slight variations in visual characteristics, especially changes in illumination and brightness, will cause

synchronous fluctuations in all three RGB channels, which restricts its applicability in professional image processing tasks.

In contrast, although the HSV color space cannot directly match the original visual signals captured by human eyes, it independently decouples hue, chroma and brightness. It delivers continuous and stable representation of color variations, making it more suitable for image processing, color segmentation and visual recognition tasks.

Taking the table tennis ball tracking task in this research as an example: the standard color of a table tennis ball is orange, yet variable ambient illumination may cause color deviation, presenting as cadmium orange, orange-yellow or yellow tones. When adopting the RGB color space, minor continuous color shifts of the target ball will lead to drastic fluctuations in the three RGB channel values. In the HSV color space, such subtle color changes are only reflected in slight continuous variations of hue or value. By setting reasonable threshold ranges for HSV components, the interference of complex illumination on target tracking can be effectively suppressed. Therefore, the original RGB video frames are converted to the HSV color space as a preprocessing step prior to target color tracking.

(3) Color Space Conversion Algorithm

Let (r, g, b) denote the normalized channel values of an RGB image, where all components are real numbers constrained within the range of 0 to 1. Define $\max$ as the maximum value among r, g and b, and $\min$ as the minimum value of the three channels. The conversion calculation is implemented to derive the HSV components (h, s, v). In this formulation, hue h is expressed in angular units, while saturation s and value v adopt normalized dimensionless values.

Among them, the value of V is always equal to the maximum value, and the value of S is 0 when the maximum value is 0; otherwise, it is calculated in accordance with Equation (2-1).

$$S = 1 - \min / \max$$

The calculation of the value of H is more complex, and the specific rules are shown in Equation (2-2):

$$h = \begin{cases} 0^\circ & \max = \min \\ 60^\circ \bullet (g-b)/(max-min) + 0^\circ & \max = r \cap g \geq b \\ 60^\circ \bullet (g-b)/(max-min) + 360^\circ & \max = r \cap g < b \\ 60^\circ \bullet (b-r)/(max-min) + 120^\circ & \max = g \\ 60^\circ \bullet (r-g)/(max-min) + 240^\circ & \max = b \end{cases}$$

2.1.2 Channel Value Setting

To detect the real-time position of the orange table tennis ball, we need to set a relatively large channel value for orange and small values for other colors, so as to disable the color channels other than orange. In this way, all colors except orange in the frame can be filtered out, enabling us to continuously track the position of the orange table tennis ball.

Different from the range of H, S, V values in the HSV color space described above, here we use the cvtColor function in OpenCV with the parameter set to CV_BGR2HSV. Consequently, the ranges of our H, S, V values are changed to [0, 180], [0, 255], and [0, 255], instead of the previous [0, 360], [0, 1], and [0, 1]. The approximate HSV value intervals for each color are shown in the following table (Table 2-1):

Table 2-1 HSV Values of Various Colors

	Black	White	Red	Orange	Yellow	Green
h_{min}	0	0	156	11	26	35
h_{max}	180	180	180	25	34	75
s_{min}	0	0	43	43	43	43
s_{max}	255	30	255	255	255	255
v_{min}	0	211	46	46	46	46
v_{max}	46	255	255	255	255	255

Considering the influence of indoor lighting on the color of the table tennis ball and combined with practical conditions, the HSV value ranges in this experiment are set as h_{min}=11 ,h_{max}=179 ,s_{min}=128 ,s_{max}=255 ,v_{min}=90,v_{max}=255.

2.1.3 Target Contour Detection

We have now obtained the approximate shape of the target table tennis ball, and the next step is to acquire the coordinates of its contour. We perform hierarchical processing on the obtained shape, calculate the area enclosed by each contour layer, and sort the obtained areas in descending order. The contour with the largest area is the outermost contour of the object. Finally, we fit a bounding rectangle to the obtained outermost contour and

International Journal of Science and Engineering Applications
Volume 15-Issue 05, 52 – 56, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1505.1006

return the length, width, and coordinates of one corner of the rectangle, thereby obtaining the coordinates of the entire object contour.

Since the table tennis ball is a regular sphere, we can obtain its center coordinates by summing the coordinates of the four corners of the bounding rectangle in pairs and dividing by 2. At this point, a hollow circular frame centered on the sphere's center coordinates can perfectly represent the position of the table tennis ball in the frame (Figure 2-3).

Figure 2-3 Position of the Table Tennis Ball in the Frame

2.2 Optimization of Actual Tracking Effect Based on Kalman Filter

2.2.1 Introduction to Principles

The Kalman filter is an optimal estimation algorithm adopted for the optimized estimation of target quantities of interest. Based on the assumptions of linear system model and Gaussian noise, it fuses system measurement data and state prediction results to estimate the true state of the system [17].

Taking a moving trolley on the road as an example, its speed and position are defined as corresponding state variables, and the position at the next moment can be expressed by Equation (2-3):

When the trolley accelerates or decelerates, the control quantity of motion is defined as the control variable. The corresponding relational expressions are derived as Equation (2-4) and Equation (2-5):

The above formulas are rewritten in matrix form to characterize the current state of the trolley, as shown in Equation (2-6):

Let the corresponding state parameters be substituted into the formula, and the equation can be converted into Equation (2-7):

Thus, the state prediction formula of the Kalman filter is obtained. In the formula, the state estimation represents the posterior estimated state of the trolley; the state transition matrix describes the inference logic for updating the current state based on the previous state estimation; the control matrix quantifies the influence of motion control variables on the real-time state of the trolley.

In actual measurement and system operation, measurement errors and process errors are inevitable, which are collectively denoted by e. The error covariance matrix $\boldsymbol{P}$ is defined as Equation (3-8): Subsequently, the error covariance matrix $\boldsymbol{P}$ is transmitted iteratively at different time steps. Meanwhile, the process noise $\boldsymbol{Q}$ of the prediction model is taken into consideration. Combined with the basic properties of the covariance matrix shown in Equation (3-9), the iterative update formula of the covariance matrix is obtained as Equation (3-10): This constitutes the second core formula of the Kalman filter, which characterizes the transmission law of state uncertainty in continuous motion.

Assuming that a satellite observes the motion state of the trolley in real time and feeds back observation values, the observation matrix is introduced to describe the mapping relationship between observed quantities and system state quantities. In addition, observation values are inevitably disturbed by observation noise v, whose covariance matrix is defined as $\boldsymbol{R}$. On this basis, the third core formula of the Kalman filter is established as Equation (3-11): The system state is updated through Equation (3-12), in which the residual error reflects the deviation between observation and prediction, and the Kalman gain is calculated by Equation (3-13):

$\tag{3-12}$

$\tag{3-13}$

Although the physical meaning of Kalman gain seems complicated, its essential principle is intuitive. For instance, two weighing scales with different precision parameters R_1 and R_2 are used to weigh the same object, obtaining measured values m and n respectively. To minimize the deviation between the final result and the true weight, the weighted average method shown in Equation (3-14) is adopted to calculate the optimized weight M:

By analogy with the above example, the process noise covariance $\boldsymbol{P}$ and observation noise covariance $\boldsymbol{R}$ in trolley motion correspond to the precision parameters of the two scales, and the Kalman gain acts as the weight coefficient. Similar to assigning weights according to equipment precision, the Kalman filter judges the reliability of predicted values and measured values through two covariance matrices, so as to adaptively solve the optimal Kalman gain.

Finally, the noise distribution parameters are corrected recursively by using the calculation results of the current time step, as shown in Equation (3-15). With the updated and optimized noise parameters adopted in the next iteration, the output results are closer to the real motion state. Through continuous recursive calculation, the deviation between predicted values and actual values is gradually reduced, which is the core advantage and essential principle of the Kalman filter algorithm [18].

2.2.2 Advantages of the Optimization Algorithm

In summary, the Kalman filter possesses multiple superior characteristics, including high estimation accuracy, excellent real-time performance, and the capability to predict the motion state of a target at the next moment merely relying on the state information of the previous frame [19, 20].

In this experiment, multiple optimization schemes such as color space conversion and adaptive threshold setting of channel parameters are adopted to suppress external interference in table tennis detection. Nevertheless, actual test scenarios still suffer from non-negligible interference, including uneven indoor illumination at specific viewing angles and temporary occlusion of the target. After introducing the Kalman filter for position prediction of the table tennis ball, the system can rely on the accurate historical state of the target in the previous frame to constrain the tracking result. Even under complex interference conditions, the detected position will not deviate severely from

International Journal of Science and Engineering Applications
Volume 15-Issue 05, 52 – 56, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1505.1006

the actual position, which effectively improves the robustness of target tracking.

3.1 Model Building and Training

The DBNet model was established using segmentation-based methods in deep learning. The basic principles and steps of DBNet are as follows:

1）Data Preprocessing: Before performing text detection on the nameplates of power equipment, data preprocessing is necessary. This includes resizing images, cleaning the data, applying image augmentation, and removing non-standard annotation boxes to expand the training dataset size.

2）Feature Extraction: The input images pass through the Backbone network, undergoing a convolution and downsampling operation, resulting in four feature maps of different sizes.

3）Feature Enhancement: The extracted feature maps are fed into the Feature Pyramid Network (FPN) structure. After cascading through the FPN network, a feature map one-fourth the size of the original image is obtained.

4）Text Position Prediction: The head network predicts the probability map and threshold map using the cascaded feature maps. An approximate binary mapping is computed from the probability and threshold maps.

5）Model Training and Testing: DBNet is utilized to train and test the dataset, evaluating its intelligent detection capabilities for text detection on power equipment nameplates.

3.2 Experimental Results and Analysis

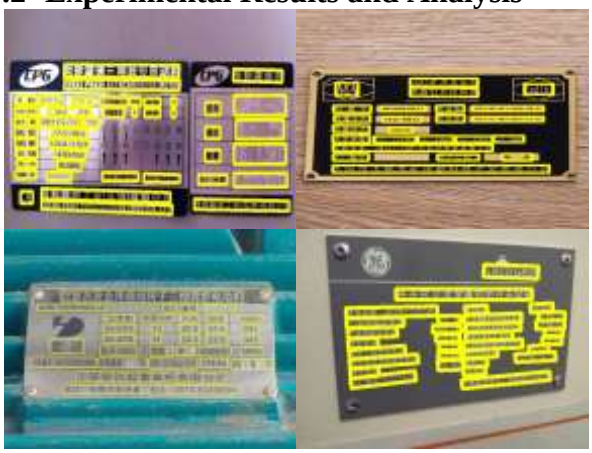

Figure.3 Detection Result

3. EXPERIMENT AND RESULT ANALYSIS

A comparison of the DBNet model with other models for text detection on power equipment nameplates is presented in Table 1.

From this table, it can be observed that DBNet outperforms other networks in overall performance due to its powerful feature extraction capabilities and the improved ability to generate threshold maps and binary maps at the detection stage. The DBNet-based text detection for power equipment nameplates excels in both accuracy and speed. In diversified scenario tests, this method accurately detects text regions and adapts well to complex backgrounds and embossed character features. Compared to traditional methods, DBNet shows significant improvements in detection efficiency and accuracy, providing reliable support for subsequent text recognition tasks.

4. CONCLUSION

In the field of text detection for power equipment nameplates, traditional methods primarily rely on manual efforts, which are time-consuming, labor-intensive, and prone to omissions, thereby causing inconvenience and risks during production or maintenance processes. To address these issues, this paper investigates text detection technology for power equipment nameplates based on deep learning. The DBNet algorithm is adopted to perform text detection on nameplates, and several attempts are made to improve detection performance.

DBNet is a segmentation-based text detection algorithm in deep learning, designed to detect text in images and annotate it in the form of bounding boxes. The algorithm introduces a differentiable binarization module, enabling the model to utilize an adaptive threshold map for binarization processing. This adaptive threshold map is incorporated into the loss calculation, which assists in optimizing the results during model training.

Experimental results show that this method not only significantly improves text detection performance but also simplifies the post-processing steps. Compared to other text detection models, DBNet demonstrates clear advantages in both effectiveness and performance. Generally, segmentation-based text detection methods require pixel-level prediction followed by post-processing algorithms to generate bounding boxes. However, post-processing algorithms are often complex and can lead to reduced computational speed. DBNet addresses this issue by integrating the binarization process into the training phase to enhance segmentation results, simplify post-processing, and maintain inference speed.

Table 1. Comparison Results

Models	P(%)	R(%)	F(%)
DBNet	80.2	72.4	76.1
PSNet	76.4	70.1	73.1
EAST	71.2	66.2	68.6

5. REFERENCES

[1] Wang Yifan, Wang Jiayu, Zhong Linlin, et al. Text Recognition of Electrical Equipment Nameplates Based on Deep Learning [J]. Electric Power Engineering Technology,

2022, 41(05): 210-218.

[2] Wang Jianxin, Wang Ziya, Tian Xuan. A Survey on Text Detection and Recognition in Natural Scenes Based on D eep Learning [J]. Journal of Software, 2020, 31(5): 1465-1496.

[3] Wang Daolei, Kang Bo, Zhu Rui. A Text Detection Meth od for Electrical Equipment Nameplates Based on Deep Learning [J]. Journal of Graphics, 2023, 44(04): 691-698.

[4] Liao M, Wan Z, Yao C, et al. Real-time scene text detecti on with differentiable binarization[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11474-11481.

[5] Zhang K, Sun M, Han T X, et al. Residual networks of re sidual networks: Multilevel residual networks[J]. IEEE T ransactions on Circuits and Systems for Video Technolog y, 2017, 28(6): 1303-1314.

[6] Xie J, Pang Y, Nie J, et al. Latent feature pyramid networ k for object detection[J]. IEEE Transactions on Multimed ia, 2022, 25: 2153-2163.