Research on real-time tracking of table tennis ball based on machine learning with low-speed camera

Yun-Feng Ji\(^a\), Jian-Wei Zhang\(^b\), Zhi-hao Shi\(^a\), Mei-Han Liu\(^a\) and Jie Ren\(^a\)

\(^a\)China Table Tennis College, Shanghai University of Sport, Shanghai, People’s Republic of China; \(^b\)Department of Informatics, Institute of Technical Aspects of Multimodal Systems, University of Hamburg, Hamburg, Germany

**ABSTRACT**

This paper proposes a novel method to track table tennis ball in real time by a low-speed camera instead of a high-speed one. Several difficult problems are solved for practical applications, such as environmental interference, smear in low-speed video images and slow processing speed. In view of these difficulties, the VOCUS system is used to segment images and mark the three significant regions based on three contrast colour channels. These regions are utilized for image matching using the LGP\(^+\)adaboost algorithm. As a strong classifier based on machine learning, adaboost algorithm can recognize the features of smear balls with different shapes. Therefore, the region that is most similar to smear ball from the three significant regions is regarded as a target. Afterwards, through the moving ROI area algorithm, the identification time is greatly shortened in real-time video tracking. Finally, the feasibility of the algorithm is examined by experiments.

**1. Introduction**

With the development of artificial intelligence, more and more intelligent technologies are used in sports industry, such as wearable sensors, live video capture, technical and tactical analysis systems and so on. Among them, the live video capture technology is widely used in ball games, such as: ‘Hawkeye’ system in tennis, the ‘gate line technology’ in soccer and ‘Omni-player tracking system’ in basketball. However, in table tennis, there exists no intelligent real-time video capture system because of the prominent tracking difficulties and challenges caused by the small size, light weight, fast speed and strong rotation of the ping-pong ball.

In order to settle these problems a lot of researches have been done. Zhang proposed a colour segmentation method for special yellow table tennis ball tracking (Zhang, Wei, Yu, & Zhong, 2011). Stauffer presented a contour-based tracking algorithm by calculating the displacement vector flow field, and using iterative operation to detect the motion area in the scene (Horn & Schunck, 1993; Ince & Konrad, 2008; Stauffer & Grimson, 1999). Lampert used the colour difference between table tennis and the background to divide images, and marked the region with high similarity of desired colour as the target area (Lampert & Peters, 2012). Zhang used the frame difference method to obtain moving objects and recognize ping-pong ball base on its characters (Zhang, 2010; Zhang & Xu, 2009; Zhang, Xu, & Tan, 2010) and Ji expanded this method into match pictures and videos (Ji et al., 2016a, 2016b).

However, all the above tracking methods require sufficient light, plain background or high camera speed which are hard to realize in practical application. Especially in real-time tracking system, the main problem is the low efficiency in transmission and identification caused by the large amount of data from high-speed camera. Using low-speed camera instead can largely increase the efficiency, however smear and pixel distortion may appear because of the low shutter speed, which leads to the deformation of graphic and the difficulty of identification.

In terms of these difficulties, this paper proposed a novel method to track table tennis ball in real time with low-speed camera based on machine learning. The ping-pong ball is fast recognized from complex pictures by the combination of image segmentation and matching, afterwards ROI (region of interest) prediction is utilized to increase the real-time tracking efficiency in the video. The main contributions of our work are as the following:

1. An image segmentation method based on the eyeball gaze model is applied into the table tennis identification process for the first time, which overcomes pixel distortion and improves the identification efficiency and accuracy.
(2) An image matching based on machine learning is used in the identification of table tennis ball. Through this method, smear balls with different shapes under different shooting conditions can be recognized and the accuracy of recognition in different scenes can be improved through enlarging the training dataset.

(3) An algorithm is proposed to calculate the center of the real ball based on the smear ball and the direction of the ball’s movement.

(4) A tracking method based on ROI prediction is proposed to reduce the computation time in real-time video tracking.

The paper is organized as follows: Section 2 introduces the method of recognition and tracking. Section 2.1 and section 2.2 introduces the method of image segmentation, preprocessing and matching. Section 2.3 describes a way to locate the true ball center of a smear ball. Section 2.4 describes how to predict ROI area in video. In section 3, the feasibility of the algorithms is verified by the experiments. Section 4 discusses and summarizes the research questions of this article, and puts forward the research plan for the future work.

2. Methods

Generally, humans tend to focus on several special areas in certain scene consciously or unconsciously during observing. This phenomenon is called the human visual attention system (Olshausen, Anderson, & Van Essen, 1993; VOCUS, F. T. S. 2005). As to table tennis sport, ping-pong ball is the main focus of players’ visual attention system. Therefore, the VOCUS (Visual Object Detection with a Computational attention System) system, which mimics the human visual attention system, is utilized in image segmentation in this paper and several significant areas are marked. Therefore, based on characters of these areas, ping-pong ball is recognized efficiently through image matching. Specific to images with smear of low-speed camera investigated in this paper, a recognition method based on machine learning is adopted to improve the accuracy. Moreover, during video tracking, the ROI prediction method is used to obtain new significant areas instead after ping-pong ball is recognized by the VOCUS system, which saves large amount of time for image segmentation.

2.1. Image segmentation based on VOCUS system

During the image processing process, the segmentation of the image makes the image more effectively processed and identified. At present, iNVT (iLab Neuromorphic Vision C++Toolkit) system (Bruce & Tsotsos, 2009; Frintrop, Werner, & Martin Garcia, 2015) and VOCUS system (Klein & Frintrop, 2011; Borji, Sihite, & Itti, 2013; Itti, Koch, & Niebur, 1998) are the two most popular image segmentation methods based on human visual attention system.

As show in Table 1, the iNVT takes the direction target into consideration, which limits the segmentation ability of table tennis ball because of the uncertainty of ball’s direction. Meanwhile, one colour channel information is utilized in iNVT system, which means low accuracy of recognition for images with pixel distortion. In contrary, the VOCUS system mainly considers the colour information of three channels during segmentation. Without limitation of direction and with more information from all colour channel, the VOCUS system has a wide applicability and is more suitable for the problem in this study.

The key theory of the VOCUS system is based on the research of Hurvich. He proposed that there are three opposing colour channels in human being’s visual system: black and white, red and green, blue and yellow, through the study of human visual attention. Human being could achieve the observation of the scene (Hurvich & Jameson, 1957) through these three colour channels. In the VOCUS system, a picture is divided into three colour channels to identify and segmentation of the pictures could be obtained through filtering, differential, homing, fusion and other operations. The specific flow chart is shown in Figure 1.

The segmentation process is divided into following steps and shown in Table 2:

Step 1: The colour information is used as a linear filter split the input image into three channels based on colour information The channels are: black / white, red / green, blue / yellow, and the thresholds of the filters in turn are: $BW = \frac{R+G+B}{3}, RG = R - G, BY = B - Y$;

Step 2: The image Gaussian pyramid algorithm is used to blurred image in different scales. In this experiment, five sets of pyramid scale images.

Step 3: The Center-surround difference and normalization of images are performed in different scales, and the specific method is the DoG (Difference-of-Gaussian) filtering algorithm.

### Table 1. Differences between iNVT and VOCUS.

| Features                  | iNVT                                      | VOCUS                                      |
|---------------------------|-------------------------------------------|--------------------------------------------|
| Pyramid structure         | Colour(1 channel),orientation             | Colour(3 channels)                         |
|                           | One pyramid, One scale per layer          | Twin pyramid, Multiple scales per layer     |
| Feature fusion            | Down-sampling, Weighting by uniqueness    | Up-sampling, Arithmetic mean               |
|                           | Fuse colour and orientation               | Fuse 3 channels                           |

Frintrop, Werner, & Martin Garcia, (2015) and VOCUS system (Klein & Frintrop, 2011; Borji, Sihite, & Itti, 2013; Itti, Koch, & Niebur, 1998) are the two most popular image segmentation methods based on human visual attention system.
### Table 2. The steps of image segmentation.

| Step | Process |
|------|---------|
| 1    | Decomposing the input image into three opposite colour channels |
| 2    | Using the image pyramid algorithm to process image at different scales |
| 3    | Center-surround difference and normalization of images |
| 4    | Multi-scale feature fusion operation |
| 5    | Generate a saliency map |
| 6    | Mark the salient area |

Step 4: The multi-scale feature fusion operation is performed on the feature map under the three channels to generate a set of conspicuity maps;

Step 5: Using linear fusion algorithm to change the conspicuity maps under three channels to one saliency map;

Step 6: Three most significant areas are marked with the red boxes.

### 2.2. Offline machine learning and image matching

General studies of table tennis ball tracking mainly extract the characteristics of the moving target based on the optical flow algorithm, but smear balls caused by low-speed camera may lead to the uncertainty of the characteristics and inaccuracy of recognition. Therefore, a novel machine learning algorithm for image matching is introduced in this paper to solve this problem. Meanwhile, the offline process of machine training and learning occupies no online computing time and has advantages in real-time tracking.

The matching of images is generally divided into two steps: image preprocessing and feature matching. At present, the commonly used methods of preprocessing are: LBP (Local Binary Patterns), LGP (Local Gradient Patterns) and HoG (Histogram of Oriented Gradient).
Generally, LBP and LGP algorithm is sensitive to the local texture features, while HoG algorithm maintains good invariance for optical geometry and image deformation. These preprocessing method can extract the characteristics of the image. As to training method, SVM (Support Vector Machine) and Adaboost algorithm are most widely used to classify the image features. The SVM algorithm can use the kernel function to fit the maximum interval hyperplane in the high-dimensional feature space, which is suitable for the classification of nonlinear data sets. Adaboost algorithm is the accumulation of several weak classifiers, and it has good adaptability to the classification of unknown data.

Since no machine learning algorithms for table tennis recognition has been researched before, we use these methods to design several experiments in turn and compare the accuracy and the calculation time of image matching in this paper. According to our experiment data, LGP adaboost algorithm shows the best performance and is discussed in detail as follows. The result of experiments is shown in section 3.

2.2.1. LGP algorithm

LGP is a feature extraction method based on local gradient of image and widely used in the preprocessing of image recognition. Generally, local gradient of each pixel is calculated on account of pixel block with certain scale except for edge pixels. As to a gray scale image matrix, for each of the $3 \times 3$ sub-matrices (denoted as matrix A), the pixel values of the center point are represented by $i_c$ and the pixel values of the surrounding points are represented by $i_n$, $n = 0, \ldots, 7$, which is defined in the order of the clockwise direction from the upper left corner. The gradient for each surrounding point is defined as follows:

$$g_n = |i_n - i_c|, \quad n = 0, \ldots, 7$$

(1)

The definition of average gradient of $3 \times 3$ sub-matrix is:

$$\bar{g} = \frac{1}{8} \sum_{n=0}^{7} g_n$$

(2)

For each point $(x_c, y_c)$ of the image, its LGP value is expressed as follows:

$$\text{LGP}(x_c, y_c) = \sum_{n=0}^{7} s(g_n - \bar{g})2^n$$

(3)

The function $s$ is defined as:

$$s(x) = \begin{cases} 
0, & x < 0 \\
1, & \text{others}
\end{cases}$$

(4)

LGP operation transforms the original matrix A to the substitution matrix A with the converted value according to Equation (3) and the particular calculating procedure is shown in Figure 2.

According to the method mentioned above, we convert each pixel value of each image in the training dataset into LGP value. The number of positive samples and negative samples are settled as $N_f$ and $N_{nf}$ separately. We use $G_{f_m}$ to represent the LGP feature pictures of the positive samples, and use $G_{nf_n}$ to represent the LGP feature pictures of the negative samples, where, $m = 1, \ldots N_f$, $n = 1 \ldots N_{nf}$.

2.2.2. Adaboost algorithm

Adaboost algorithm is used as the main algorithm for training. Adaboost algorithm is a learning process that a strong classifier is composed of several weak classifiers according to certain weight. In every process, it updates the weight of each training sample, which can increase the weights of the positive samples and reduce the weights of the negative samples. So it will lead to better distinguish between positive and negative samples. A four-stage cascade of classifiers is used in this paper, and the number of feature points in each cascade is selected: as:

$$F_1 = 26, F_2 = 60, F_3 = 120, F_4 = 300.$$
(a) Generate the weight from the training LGP feature images as:

\[
\begin{align*}
W_f^f(X, r) &= 0 \\
W_f^{nf}(X, r) &= 0
\end{align*}
\]  

(5)

For positive samples, set \( \gamma = G_m^f(X) \), it can obtain

\[
W_f^f(X, \gamma) = W_f^f(X, \gamma) + W_f^f(m)
\]  

(6)

For negative samples, set \( \gamma = G_n^{nf}(X) \), it can obtain

\[
W_f^{nf}(X, \gamma) = W_f^{nf}(X, \gamma) + W_f^{nf}(n)
\]  

(7)

(b) Compute the error \( e_t(X) \) to search for the best feature point as:

\[
e_t(X) = \sum_{r=0}^{255} \min\{W_f^f(X, r), W_f^{nf}(X, r)\}
\]  

(8)

(c) Select the best feature point \( X_t \) as:

\[
X_t = \min_X e_t(X)
\]

\[
P_{t+1} = P_t \cup X_t
\]  

(9)

(d) Select the type of weak classifier at the selected feature point \( X_t \) according to the sum of weight as:

\[
z_t(r) = \begin{cases} 0, W_f^f(X, r) > W_f^{nf}(X, r) \\ 1, W_f^f(X, r) \leq W_f^{nf}(X, r) \end{cases}
\]  

(10)

Update the weak classifier at the selected feature point \( X_t \) as:

\[
h_X(r) = h_X(r) + \alpha_t z_t(r)
\]  

(11)

where \( \alpha_t = \ln(1 - e_t)/e_t \).

(e) Update the weights of the training positive and negative LGP feature images as:

\[
\begin{align*}
w_{t+1}^f(m) &= w_{t+1}^f(m) \cdot e^{-m}, z_t(G_m^f(X_t)) = 0 \\
w_{t+1}^{nf}(n) &= w_{t+1}^{nf}(n) \cdot e^{-m}, z_t(G_n^{nf}(X_t)) = 1
\end{align*}
\]  

(12)

Normalization: \( w_{t+1}^f(m) = \frac{w_{t+1}^f(m)}{S_{t+1}} \) and \( w_{t+1}^{nf}(n) = \frac{w_{t+1}^{nf}(n)}{S_{t+1}} \),

where \( S_{t+1} = \sum_{m=1}^{N_f} w_{t+1}^f(m) + \sum_{n=1}^{N_{nf}} w_{t+1}^{nf}(n) \).

(f) When the number of feature point reaches the number of feature points in each cascade, output the corresponding feature point \( X \) and the weak classifier \( h_X(r), X \in P_{t+1} \), as the four-stage cascade of classifiers. When the number of feature points reaches the number of \( F_1 \), end the cycle.

Step 3: The final strong classifier is the sum of weak classifiers as:

\[
H^f = \sum_{X \in P_{t+1}} h_X^f(G(X))
\]  

(13)

where \( j \) represents \( j \)-stage cascade of classifiers, \( p^f \) represents the set of selected feature points and \( h_X^f() \) represents weak classifier of the feature point \( X \) in the \( j \)-stage cascade of classifiers.

Specific to the blurring of images caused by low-speed camera, amount of smear ball’s images are added into the dataset of our research. Therefore, after training with adaboost algorithm, the strong classifier shows a great recognition ability to smear balls.

### 2.3. Real core detection and planning ROI area

#### 2.3.1. Real core detection

As one of the main reason of inaccuracy during recognition, smear ball is caused by the relative movement between ping-pong ball and camera during the exposure time. During filming process, because the exposure time of camera is very long and the speed of ping-pong ball is very fast, the image formed on the chip is changing during the exposure process. Therefore, the final formation of the picture is the image of the superposition within a continuous changing space. For a picture of a smear ball, the smear is the overlap image of the table tennis ball’s trajectory and the actual location of the ball is in the forefront of movement direction. In account of the analysis above, a novel algorithm is proposed in this paper to locate the ping-pong ball’s real center.

Generally, since the shape of a smear ball is elliptical, the elliptic formulation is utilized to fit the boundary points. The shape fitting is based on the least squares method and points of target boundary are collected by the contour search method (Xiao, Zhao, & Chen, 2009, November). The actual location of ping-pong ball is the center of the forefront circle target in the smear ball. As shown in Figure 3, the ellipse is the fitting ellipse for a ping-pong ball and the circle stands for the boundary of the actual ping-pong ball.

The coordinates of the ellipse’s center is \((x_1, y_1)\), \( l \) represents the long axis of the ellipse; \( s \) represents the minor axis of the ellipse and \( \alpha \) represents the angle between the major axis of the ellipse and the \( x \) axis. The values of these parameters could be obtained after the elliptic equation is fitted by the least squares method. As can be
seen in Figure 3, the short axis of the ellipse approximates the radius of the actual ball, which means the distance between the smear ball’s center and the actual ball’s center is the difference between the long axis \( l \) and the minor axis \( s \). The Calculation formulas are as follows:

\[
\begin{align*}
x_2 &= x_1 + (l - s) \cos \alpha \\
y_2 &= y_1 + (l - s) \sin \alpha
\end{align*}
\]

(14)

2.3.2. Prediction of the moving ROI area

Video tracking is a complex issue because of the light, the background and other interference problems. Since video is a playing sequence consist of multi-frame pictures, detecting for full images constantly during video tracking leads to huge amount of computation, large amount of calculation time and other problems. However, video is a continuously image sequence, in which two continuous images are strongly related to each other.

According to this characteristic, this paper proposes a method by predicting the possible location area of ping-pong ball in next frame to realize real-time video tracking. This prediction is based on the ball’s location in current frame and the possible location area moves with the movement of ping-pong ball, which is called moving ROI area. As to video to be tracked, each frame is detected by the image segmentation and matching algorithm discussed above until ping-pong ball is recognized for the first time. Afterwards, a square moving ROI area is located based on the ping-pong ball as the search area in next frame instead of full image. The center of this area is settled at the ball’s center in current frame, and the side length of this square is 2.4 times the product of the ball’s highest speed and the time interval between neighboring two frames. Therefore, because of the speed limitation, ping-pong ball is unable to escape from the moving ROI area in the next frame. In addition, while the center of this area moves with ping-pong ball, the target ball can always be detected in the moving ROI area. This method not only reduces the identification time, but also increases the identification accuracy.

3. Experiment and results

According to the discussion above, the algorithms in this paper is tested for ping-pong identification and tracking in the video shot in the real scene. The frame rate of the camera is 30fps. The experimental results are divided into three parts: the image segmentation results, the image matching and ball’s true center detection results, and the ROI setting results.

3.1. Result of image segmentation

According to the method described in this paper, images are segmented based on the VOCUS system, and the most significant three regions are marked in the split image with a red box as shown in Figure 4.

A comparative test is taken for the segmentation results based on iNVT system. A total amount of 100 images taken with a low-speed camera of table tennis balls’ motion is selected, using both the iNVT and VOCUS system for image segmentation. The accuracy and average calculation time of each segmentation method are recorded. It is regard as a successful segmentation when
Table 3. Contrast between iNVT and VOCS.

|        | Accuracy (%) | Average calculation time (ms) |
|--------|--------------|------------------------------|
| iNVT   | 74           | 117.51                       |
| VOCS   | 87           | 91.83                        |

the ping-pong ball appears in one of the three most significant areas, and the accuracy is defined as the ratio of successful segmentations. The detailed results of the two segmentation image systems are shown in Table 3.

As can be seen from Table 3, the VOCS system is superior to the iNVT system on accuracy and average calculation time, which means the VOCS system is more appropriate for the real-time video tracking of ping-pong ball.

3.2. Result of image matching and real ball's center detecting

The accuracy of recognition raising with the increase of training samples is the greatest advantage of machine learning. Therefore, pre-collecting training samples under different scenarios improves the adaptability of the recognition algorithm for different scenes. Besides, since the training process of machine learning is offline, the increase of the training time caused by the gain of samples will not influence the online recognition time.

In this paper, the experiment is aimed at tracking ping-pong ball in videos shot by low-speed camera in realistic scene. Therefore, machine learning is utilized as the image matching method to reduce the interference caused by shape of smear balls and different scenes. This paper collects 100 pictures of smear balls with different motion directions and backgrounds similar to those in Figure 5(a) below as positive samples. Then 100 pictures similar to those in Figure 5(b) are also collected, which are randomly selected from different playing table tennis scenes as negative samples of training.

The detailed image matching method is separated into the following three steps:

(1) Preprocessing the image by Local Gradient Patterns (LGP);

(2) Training the adaboost level classifier by training samples;

(3) Adding significant areas from the VOCS system to the classifier for testing, and selecting the area with the highest value of similarity as the location of the table tennis ball.

The result is shown in Figure 6 and the location of ping-pong ball is marked with a red box in the figure.

After marking the significant area where the table tennis ball is located, the approximate position of the smear ball can be obtained using the ellipse fitting methods, and then the location of the real ball’s center is calculated according to the algorithm discussed in section 2.3. The real location of the table tennis ball is marked with a red circle in Figure 7.

In this experiment, six methods of image matching are compared as follows, including the image preprocessing methods and training methods. These different algorithms are LBP+SVM, LBP+adaboost, LGP+SVM, LGP+adaboost, HoG+SVM and HoG+adaboost. After training, 50 images are used to test the recognition accuracy and processing time of six different algorithms. Correctly identifying the ping-pong ball in a single image is recorded as an accurate result, and the ratio of the accurate results in the 50 images is the recognition accuracy of the algorithm. The processing time refers to the time required for processing a picture to complete the identification. The specific result is shown in Figure 8.

Figure 5. Part of the positive and negative samples of training.

Figure 6. Matching results.
3.3. Results of moving ROI area

Previous experiments focus on segmentation of images and identification of table tennis ball in images. However, the whole process costs more than 100 ms for one picture, which is unable to realize the real-time video tracking with a low-speed camera. The whole recognition time must be less than 33 ms since the frame rate of the camera used in this study is 30fps. Therefore, in video tracking, the moving ROI area algorithm is adopted to set the ROI area by predicting the possible location of table tennis ball in advance. The ROI area is used for recognition instead of the whole image, which reduces the amount of calculation and shortens the calculation time. Then the algorithm proposed in section 2.3 is utilized to calculate the real ball’s center. The experimental results are shown in Figure 9.

In Figure 9, the ROI area is marked with blue box and moves following the table tennis ball. After using the moving ROI area method, the average recognition time for a picture is reduced to 23 ms, which can basically realize the real-time tracking with a low-speed camera.

4. Conclusion and discussions

In this paper, a new method is proposed to realize real-time video tracking of table tennis ball with a low-speed camera. This method segments images based on VOCUS system, and then identifies the smear ball through image matching based on LGP+ adaboost algorithm. Finally, an algorithm is designed to calculate the coordinates of the real table tennis ball’s center, and the moving ROI area algorithm is used to shorten the identification time to achieve the requirement of real-time video tracking.

Although the goal of tracking table tennis by a low-speed camera could be achieved in this study through the integration and utilization of the experimental methods, but there are still some limitations and deficiencies:

(1) Image processing methods are not innovative enough, and other methods have not been compared with.
(2) The samples for machine learning is not large enough. This study mainly focuses on the experiment of one certain real scene.

(3) The identification problem with shelters has not been considered.

Considering the shortcomings of this method, it can be improved in the further study. The training samples of machine learning will be expanded so that the algorithm can be applied in different scenes. Meanwhile, the problems of tracking under complicated situations will be studied, such as the re-identification of the ball after the target’s lost produced by shelters or other issues.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This article is funded by the Graduate Study Abroad Program in Shanghai University of Sport [grant number stfx20170115] and Project of Shanghai Science and Technology Committee [grant number 15490503100].

**References**

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1), 55–69.

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3), 5–5.

Frintrop, S., Werner, T., & Martin Garcia, G. (2015). *Traditional saliency reloaded: A good old model in new shape*. Proceedings of the IEEE conference on computer vision and pattern recognition. (pp. 82–90).

Horn, B. K., & Schunck, B. G. (1981). “Determining optical flow”: a retrospective. *Artificial Intelligence*, 59(1-2), 81–87.

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, 64(6p1), 384–404.

Ince, S., & Konrad, J. (2008). Occlusion-aware optical flow estimation. *IEEE Transactions on Image Processing*, 17(8), 1443–1451.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259.

Ji, Y., Shi, Z., Wang, C., & Ren, J. (2016a). Table tennis service umpiring system based on video identification—for height and angle of the throwing ball’s problems. *Computer Engineering and Applications*, 52(10), 207–212. (Chinese).

Ji, Y., Zhu, L., & Shen, Y. (2016b). Table tennis recognition based on OpenCV. *Microcomputer Applications*, 32(4), 68–70. (Chinese).

Klein, D. A., & Frintrop, S. (2011, November). **Center-surround divergence of feature statistics for salient object detection.** Computer vision (ICCV), 2011 IEEE international conference on (pp. 2214–2219). IEEE.

Lampert, C. H., & Peters, J. (2012). Real-time detection of colored objects in multiple camera streams with off-the-shelf hardware components. *Journal of Real-Time Image Processing*, 7(1), 31–41.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11), 4700–4719.

Stauffer, C., & Grimson, W. E. L. (1999). **Adaptive background mixture models for real-time tracking.** Computer vision and pattern recognition, 1999. IEEE computer society conference on. (Vol. 2, pp. 246–252). IEEE.

VOCUS, F. T. S. (2005). *A visual attention system for object detection and goal directed search* (Doctoral dissertation). University of Bonn, Bown.

Xiao, P., Zhao, G., & Chen, Y. (2009, November). **An algorithm for ellipse detection based on geometry.** Pattern recognition, 2009. CCPR 2009. Chinese conference on (pp. 1–4). IEEE.

Zhang, Z. (2010). **Visual measurement and control of table tennis robot.** Beijing: Chinese Academy of Sciences. (Chinese).

Zhang, Y. H., Wei, W., Yu, D., & Zhong, C. W. (2011). A tracking and predicting scheme for ping pong robot. *Journal of Zhejiang University-Science C*, 12(2), 110–115.

Zhang, Z., & Xu, D. (2009). **High-Speed vision system based on camera and Its target tracking algorithm.** *Robot*, 31(3), 229–234,241. (Chinese).

Zhang, Z., Xu, D., & Tan, M. (2010). **Visual measurement and prediction of ball trajectory for table tennis robot.** *IEEE Transactions on Instrumentation and Measurement*, 59(12), 3195–3205.