Table Tennis Identification and Tracking Model of the Edge Information Optimizing Canny Algorithm

Changhong Wu, Haiyan Geng

College of Sports, Langfang teachers college,
Langfang 065000, Hebei, China

Guangwei Sun

Langfang Radio and TV University,
Langfang 065000, Hebei, China

Abstract
As the Canny algorithm has the problems that the recognition accuracy is poor and the anti-noise ability is not strong in the table tennis identification and tracking model, this paper proposes an improved Canny algorithm based on edge information optimization. Firstly, it compares the Sobel operator with Canny operator and finds that the edge information of Canny operator has higher signal frequency, which will lose some detail information causing the edge unclear. Therefore, the OTSU method, which adaptively extract high and low threshold according to the characteristics of gradient magnitude, is adopted to avoid the manual effect, which can make Canny algorithm more stable and effective. Finally, the particle swarm optimization method is introduced, taking the mean value and standard deviation functions of local images as objective functions to optimize the edge information of Canny algorithm. Simulation experiment shows that the improved Canny algorithm has better performance than standard algorithm in table tennis identification and tracking and it has a higher identification precision.
Key words: CANNY ALGORITHM, TABLE TENNIS IDENTIFICAITON, OTSU, LOCAL IMAGE, EDGE INFORMATION OPTIMIZATION

1. Introduction

Playing table tennis seems a simple task for humans, but for robot, it is absolutely a challengeable job currently in related technologies [1]. Especially in this project, it requires breakthrough in stereo vision, rapid and flexible arm based on visual servo, integrated joint design and dynamic equilibrium control so that humanoid robot can play the table tennis with human beings [2]. Domestic and foreign researchers have been attracted by these challenges and carried out a lot of studies.

The vision system, as the basis of the task of table tennis robot, has been paid more attention by researchers. In terms of hardware structure, Andersson R.L. realized the man-machine rally by

taking the stereo vision system composed of four cameras. But with the increasing of the cameras, image processing tasks of the computer become more expensive, which contradicts the real-time requirement of visual system. Therefore, the researchers generally adopted two cameras to form the binocular vision system [3]. Acosta L. measured the three-dimensional position of table tennis with single camera, but the work environment was too strict to be appropriate in the standard table tennis competition [4]. Zhang et al have designed a high-speed vision system based on smart camera which can obtain good real-time performance in the identification and tracking of table tennis, but the system for 8-bit grayscale image acquired less information, limited in the design of image processing algorithms and increased the reliable recognition difficulty of table tennis under complex dynamic environment [5]. The main shape information of ball is its edge information. There are many edge extraction methods in image processing, such as Prewitt operator [6], Sobel operator [7], Laplacian operator [8] and Canny operator [9], where Canny operator can quickly extract the single edge information of the image. Because of the small imaging area of table tennis in the image, how to improve the detection accuracy is an important problem for the edge detection technologies. The most influential circle detection method using edge information is the Circle Hough transform (CHT) algorithm which precisely searches the circle with a given radius in an image, but when the radius of the circle is unknown, the algorithm needs to establish a costly three-dimensional parameter space including the radius of the circle. In order to reduce the computational time and memory and improve the accuracy of detection, many improved methods are generated [10]. D. Scaramuzza et al made improvement in edge chain code by determining the edge direction with the vicinity point of each edge point rather than directly using the gradient direction of the edge, which solved the problem that the lines obtained from the edge points on the same circle cannot intersect on a point [11]. T. Atherton et al proposed that the CHT algorithm and its improved method is equivalent in the form to convolve with a constant-scale kernel operator in edge image. Orazio T.D. improved it by introducing the normalization factor into the convolution kernel and ensuring the most complete circle from the extreme points [12]. It was applied in rapid and robust football detection from football match video image, and processed the ball shadow problems in football detection.

In light of the defects of standard Canny algorithm in table tennis identification and tracking, this paper put forward a table tennis identification and tracking model based on optimized Canny algorithm by edge information, and conducted experiments on real data to verify its effectiveness.

2. Identification model analysis based on Canny algorithm

Canny operator is a multi-stage optimization operator including filtering, enhancement and detection with good edge detection performance. The specific mathematic description is written below.

Firstly, the two-dimensional Gaussian function is set,

$$G(x,y)=\frac{1}{2\pi\sigma^2}\exp\left[\frac{-(x^2+y^2)}{2\sigma^2}\right] \quad (1)$$

Then, the first derivative of Gaussian function $G(x,y)$ in a direction p is took,

$$G_p=\frac{\partial G(x,y)}{\partial p}=p\cdot\nabla G(x,y) \quad (2)$$

$$p=\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \quad (3)$$

$$\nabla G(x,y)=\begin{bmatrix}\frac{\partial G}{\partial x}\\ \frac{\partial G}{\partial y}\end{bmatrix} \quad (4)$$

where p is the direction vector; $\nabla G(x,y)$ is the gradient vector.

Canny operator is established on the basis of two-dimensional convolution $\nabla G(x,y)\cdot f(x,y)$. Its edge intensity is determined by $\left|\nabla G(x,y)\cdot f(x,y)\right|$ and direction $p=\frac{\nabla G(x,y)\cdot f(x,y)}{\left|\nabla G(x,y)\cdot f(x,y)\right|}$. In order to improve the operation speed of Canny operator, the two-dimensional convolution model of $\nabla G(x,y)$ is spilt into two one-dimensional filter,

$$\frac{\partial G(x,y)}{\partial x}=kx\cdot\exp\left[\frac{-x^2}{2\sigma^2}\right]\exp\left[\frac{-y^2}{2\sigma^2}\right]=h_1(x)h_2(y) \quad (5)$$

$$\frac{\partial G(x,y)}{\partial y}=ky\cdot\exp\left[\frac{-y^2}{2\sigma^2}\right]\exp\left[\frac{-x^2}{2\sigma^2}\right]=h_1(y)h_2(y) \quad (6)$$

Here k is a constant.

$$\begin{cases} h_1(x) = \sqrt{k}x \cdot \exp\left[\dfrac{-x^2}{2\sigma^2}\right] \\ h_2(x) = \sqrt{k} \cdot \exp\left[\dfrac{-x^2}{2\sigma^2}\right] \\ h_1(y) = \sqrt{k}y \cdot \exp\left[\dfrac{-y^2}{2\sigma^2}\right] \\ h_2(y) = \sqrt{k} \cdot \exp\left[\dfrac{-y^2}{2\sigma^2}\right] \end{cases} \quad (7)$$

It is seen that

$$\begin{cases} h_1(x) = xh_2(x) \\ h_1(y) = xh_2(y) \end{cases} \quad (8)$$

These two templates are convoluted with image $f(x, y)$ respectively,

$$E_x = \frac{\partial G(x, y)}{\partial x} \cdot f(x, y) \quad (9)$$

$$E_y = \frac{\partial G(x, y)}{\partial y} \cdot f(x, y) \quad (10)$$

$$A(i, j) = \sqrt{E_x^2(i, j) + E_y^2(i, j)} \quad (11)$$

$$\alpha(i, j) = \arctan\frac{E_y(i, j)}{E_x(i, j)} \quad (12)$$

Where $A(i, j)$ indicates the edge intensity of point (i, j) in the image while $\alpha(i, j)$ is the direction of vertical edge.

This paper takes Sobel operator and Robert operator to compare with Canny operator. Taking a 256x256 table tennis video sequence for an example, their precision on table tennis identification and tracking is compared as follows.

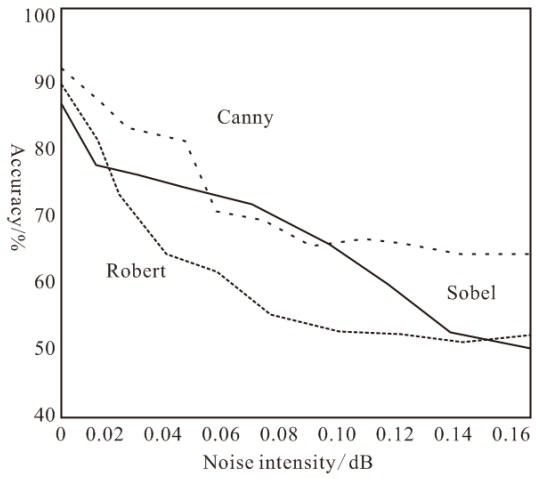

Figure 1. The comparison results of effects of noise on the accuracy

It is evident from the simulation results that Canny edge detection algorithm has better anti-noise ability and robustness. However, it still exists some problems: in the table tennis identification and tracking, image denoising is the first step and Canny operator removes the noises with Gaussian filter. The parameter controlling the degree of smoothness σ is manually determined and the higher σ is, the narrower the band is. At this state, high-frequency signal will be inhibited and fake edge point will be avoided. Due to the high frequency of signal in the edge, some detail information will be missed to make the image edge unclear. When σ is set lower, the band broadens to keep the detail information but cannot obtain ideal denoising effect. In practice, it is very hard to find accurate σ value to make this algorithm limited in table tennis identification and tracking.

3. Edge information optimization of Canny algorithm

3.1 Double threshold optimization based on Otsu method

Canny edge detection algorithm uses double threshold to detect and connect edge and the setting of double threshold has a great impact on the detection result. Therefore, this paper uses Otsu method adaptively to set higher or lower threshold. The specific method is described as follows.

The histogram of gradient magnitude is made with the magnitude quantization as L, namely $G = 1, 2, ..., L$. The pixel number in level i is represented as f_i, and the total pixel number is denoted as N, namely

$$\begin{aligned} N &= f_1 + f_2 + ... + f_L \\ &= \sum_{i=1}^{L} f_i \end{aligned} \quad (13)$$

Here, P_i is the probability of level i, namely $P_i = f_i / N$, and then $P_i \geq 0$, $\sum_{i=1}^{L} P_i = 1$. The histogram is classified into two classes: C_0 and C_1, with threshold t, namely $C_0 = \{1, 2, ..., t\}$ and $C_1 = \{t+1, t+2, ..., L\}$. The probability distribution of these two classes is,

$$\begin{cases} \omega_0 = \sum_{i=1}^{t} P_i = \omega(t) \\ \omega_1 = \sum_{i=t+1}^{L} P_i = 1 - \omega(t) \end{cases} \quad (14)$$

The mean grey values are,

$$\begin{cases} u_0 = \sum_{i=1}^{t} \frac{iP_i}{\omega_0} \\ u_1 = \sum_{i=t+1}^{L} \frac{iP_i}{\omega_1} = \frac{u_T - u(t)}{1-\omega(t)} \end{cases} \quad (15)$$

where $u(t) = \sum_{i=1}^{t} iP_i$ and $u_T = u(L) = \sum_{i=1}^{L} iP_i$.

For arbitrary t, the equation below is true.

$$\begin{cases} \omega_0 + \omega_1 = 1 \\ \omega_0 u_0 + \omega_1 u_1 = u_T \end{cases} \quad (16)$$

The variance of these two classes is,

$$\begin{cases} \sigma_0^2 = \sum_{i=1}^{t} (i-u_0)^2 \frac{P_i}{\omega_0} \\ \sigma_1^2 = \sum_{i=t+1}^{L} (i-u_1)^2 \frac{P_i}{\omega_1} \end{cases} \quad (17)$$

To evaluate the performance of threshold, the evaluation functions are introduced.

$$\begin{cases} \lambda = \frac{\sigma_B^2}{\sigma_w^2} \\ \kappa = \frac{\sigma_T^2}{\sigma_w^2} \\ \eta = \frac{\sigma_B^2}{\sigma_T^2} \end{cases} \quad (18)$$

where σ_w^2, σ_T^2 and σ_B^2 represents the inner class variance, variance between classes and total variance, described respectively as,

$$\begin{cases} \sigma_w^2 = \omega_0\sigma_0^2 + \omega_1\sigma_1^2 \\ \sigma_B^2 = \omega_0\sigma_0^2 + \omega_1\sigma_1^2 - u_T^2 \\ \sigma_T^2 = \sum_{i=1}^{L} (i-u_T)^2 P_i \end{cases} \quad (19)$$

Then the optimization objective turns to select the optimal threshold t, to maximize λ, κ and η. Because $\sigma_T^2 = \sigma_w^2 + \sigma_B^2$, the relationship among λ, κ and η is,

$$\begin{cases} \kappa = \lambda + 1 \\ \eta = \frac{\lambda}{\lambda+1} \\ \kappa = \frac{1}{1-\eta} \end{cases} \quad (20)$$

From this expression, it is known their monotonicity is consistent. σ_T^2 is a known constant, independent of t, therefore $\eta(t) = \sigma_B^2(t)/\sigma_T^2$ is the simplest in three evaluation functions. Then interclass variance $\sigma_B^2(t)$ can be used as the evaluation function of the classification performance.

The optimal threshold t^* is,

$$\sigma_B^2(t^*) = \max_{1<t<L}\{\sigma_B^2(t)\} \quad (21)$$

The abovementioned maximum interclass variance can only obtain single threshold, but Canny algorithm adaptively gets the high and low threshold. Therefore, double threshold should be extended to it with following procedure.

Threshold T_l and T_h divide the histogram of gradient magnitude into three classes, $C_0 = \{1, 2, ..., T_l\}$, $C_1 = \{T_l + 1, t_1 + 2, ..., T_h\}$ and $C_2 = \{T_h + 1, t_2 + 2, ..., L\}$.

The probability distribution of these three classes is,

$$\begin{cases} \omega_0 = \sum_{i=1}^{T_l} P_i \\ \omega_1 = \sum_{i=T_l+1}^{T_h} P_i \\ \omega_2 = \sum_{i=T_h+1}^{L} P_i \end{cases} \quad (22)$$

Each mean grey value is,

$$\begin{cases} u_0 = \sum_{i=1}^{T_l} \frac{iP_i}{\omega_0} \\ u_1 = \sum_{i=T_l+1}^{T_h} \frac{iP_i}{\omega_1} \\ u_2 = \sum_{i=T_h+1}^{L} \frac{iP_i}{\omega_2} \end{cases} \quad (23)$$

And the total mean grey value is,

$$u_T = \sum_{i=1}^{L} iP_i \quad (24)$$

For arbitrary T_l and T_h, the following expression is true,

$$\begin{cases} \omega_0 + \omega_1 + \omega_2 = 1 \\ \omega_0 u_0 + \omega_1 u_1 + \omega_2 u_2 = u_T \end{cases} \quad (25)$$

The interclass variance $\sigma_B^2(T_l, T_h)$ is,

$$\sigma_B^2(t_1, t_2) = \omega_0 u_0^2 + \omega_1 u_1^2 + \omega_2 u_2^2 - u_T^2 \quad (26)$$

The optimal threshold (T_l^*, T_h^*) is,

$$\sigma_B^2(T_l^*, T_h^*) = \max_{1<t_1<t_2<L}\{\sigma_B^2(T_l, T_h)\} \quad (27)$$

The OTSU method adaptively extract the high threshold and low threshold according to the characteristic of gradient magnitude without any manual settings. It avoids the manual effect and has simple and stable calculation.

3.2. Edge pixel optimization based on PSO algorithm

In most images, the mean value and standard deviation of corresponding pixels are usually calculated. Because Particle Swarm Optimization (PSO) algorithm searches the global optimum, this paper takes the mean value and standard deviation of pixel of local image as objective functions, namely fitness function.

The mean value of image: the grey sum of all pixel over the number of pixel, a reflection of the central tendency of grey value. $\bar{D}(i, j)$ is the mean value of pixel in 3×3 subimage.

$$\bar{D}(i, j) = \frac{\sum_{x=-1}^{1}\sum_{y=-1}^{1} D(i+x, j+y)}{9} \quad (28)$$

The variance of image is obtained by the minus of the arithmetic mean from all numerical value followed by sum of squares and mean value. This variance represents the mutual discrete degree of the observed values. If the object is an image, then its variance is the grey value of each pixel minus the sum of square of mean grey value over the number of total pixels. This variance can embody the discrete degree between pixels and the fluctuating degree across the image.

Here, for image D, the standard deviation of pixel of each 3×3 subimages are calculated. The standard deviation of 3×3 subimage centering (i, j) is,

$$\delta_{i,j} = \sqrt{\frac{\sum_{x=-1}^{1}\sum_{y=-1}^{1} [D(i+x, j+y) - \bar{D}(i, j)]^2}{9}} \quad (29)$$

When the standard deviation of the pixel is smaller, the fluctuation of pixel in this area is less and the mean value is more close to the real pixel of test image. The specific steps are written as follows.

(1)To initialize the location and velocity of each particle and determine the required parameters for PSO algorithm, including population size, weight, learning factor and fitness value.

(2) To calculate the objective function value, setting the mean value f_1 and standard deviation f_2 as the objective functions or fitness functions.

(3)To calculate the single fitness value $f_1(X_1(k))$ and $f_2(X_2(k))$ of objective functions f_1 and f_2, and single extreme $Xb_1(k)$, $Xb_2(k)$ and global extreme value Pp_1 and Pp_2. Then the vector mean value of global extreme value is $gBest = (Pp_1 + Pp_2)/2$ and the distance of global optimum is $dgBest = abs(Pp_1 + Pp_2)$, and the distance between each particle is $dpBest = abs(Xb_1(k) + Xb_2(k))$. Here, the optimal threshold is the distance between each particle $dpBest$, and the lowest threshold is set half of the highest threshold to extract the edge information.

(4) To obtain the fitness value by calculating the standard deviation and variance of detected image. The fitness value of two functions are introduced into $dgBest = abs(Pp_1 + Pp_2)$ respectively. The obtained $dgBest$ is compared with the distance between each particle $dpBest$. If better, the $dgBest$ will replace corresponding threshold of $dpBest$.

(5)To update the velocity and location of particles in images to be detected.

(6)To end up the procedure when the maximum iteration number is reached, otherwise go back to step (2).

4. Simulation experiment

To verify the effectiveness of the improved algorithm, this paper conducted the simulation experiment. Firstly, taking a table tennis video sequence with 256x256 size for example, the identification and tracking on the table tennis was made with standard Canny algorithm and improved Canny algorithm. The comparison results are shown below.

Figure 2. The tracking results of standard algorithms.

Figure 3. The tracking results of improved algorithms.

Precision simulation experiment about these two algorithms was conducted with 10 table tennis video sequences, as shown in Figure 4.

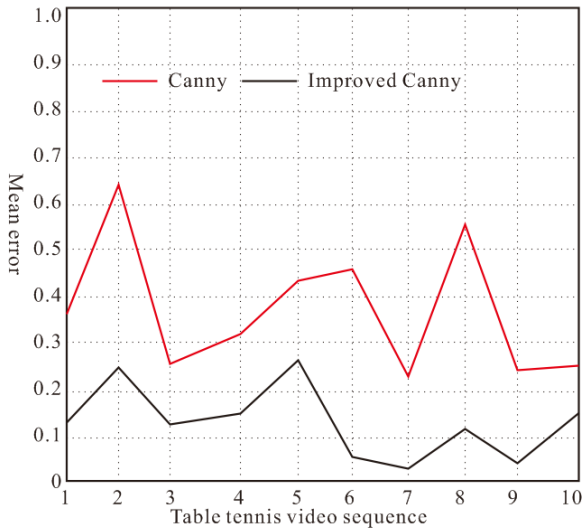

Figure 4. Comparative error analysis of two algorithms

From the results, the improved Canny algorithm presented in this paper has better performance than standard algorithm in table tennis identification and tracking because it optimizes the edge information of images. The simulation results prove that this improved algorithm has higher identification precision.

5. Conclusion

In the project of humanoid robot playing table tennis, the visual system of robot requires to solve the real-time problems like object identification, location in three-dimensional space and tracking, which is the foundation of robot execution. It has important theoretical and practical meaning in the development of visual measurement technology. For the existing defects in table tennis identification and tracking, this paper proposed a model based on Canny algorithm with improved edge information. The simulation results show that the improved Canny algorithm has better performance than standard algorithm in table tennis identification and tracking and has higher identification precision.

References

1. Chen G.D. (2014) An Vision-based Pose Measurement of Ping-pong Robot Body With Sub-Pixel Accuracy. *Control and Decision*, 27(10), p.p.1571-1574.
2. Zhang Y.H. (2014) Shadow Based Single Camera Vision System Calibration. *Journal of Image and Graphics*, 33(9), p.p.1895-1899.
3. Li X.H. (2014) Binocular Vision Measurement Applied in Analysis of Speed of Table-tennis. *Computer Science*, 35(3), p.p. 256-257.
4. Hu F.J., Zhao Y.W., Chen J. (2014) SIFT Feature Points Detection and Extraction of Three-Dimensional Point Cloud. *WIT Transactions on Information and Communication Technologies*, (60), p.p.603-611.
5. Bu J. (2011) A Real-time Billboard Trademark Recognition Algorithm in Sports Video. *Acta Automatica Sinica*, 37(4), p.p.418-426.
6. Song G. (2010) Video Classification Based on Region Features and HMM. *Journal of Southwest China Normal University(Natural Science)*, 35(2), p.p.180-184.
7. Zhu Y.Y. (2010) Analysis and Detection of Redundancy Data on Basketball Video. *Mini-micro Systems*, (9), p.p.1873-1876.
8. Lv, Z.H., Alaa H.W., Feng S.H., Li H.B., Réhman (2014) Multimodal Hand and Foot Gesture Interaction for Handheld Devices. *ACM Transactions on Multimedia Computing, Communications and Applications* 11(1s) P.P.10-10.
9. Hu F.J., Zhao Y.W., (2013) An Improved Method of Discrete Point Cloud Filtering based on Complex Environment. *International Journal of Applied Mathematics and Statistics*, 48(18), p.p. 514-522.
10. Xu H.N. (2009) Algorithm for Maneuvering Target Tracking in Sports Video Frequency Based on IMM. *Journal of Image and Graphics*,14(5), p.p.920-924.
11. Han B. (2013) Enhanced Sports Video Shot Transition Detection Based on a Unified Feature Model. *Video Engineering*, 31(8), p.p.76-79.

12. Guo H. (2014) Table Tennis Decision System Based on Double Channel Object Moving Detection. *Journal of Guangzhou Physical Education Institute*, 34(6), p.p. 67-70.