Target Detection in NAO Robot Golfing

Erwei Su¹, Long Chen², Yingli Xu³ and Bin Hu⁴*

¹Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China
²School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
³Email: chenclong@yeah.net

Abstract. In robot golfing games, how to identify and track small balls is a critical step for scoring. Considering uneven illumination and intensity, irregular distribution, and blurred borders, an integrated detection method based on the random Hough transform and the Kalman filter is presented to improve the recognition accuracy, i.e., guaranteeing the detection accuracy and the stability in shooting. Experiment on the NAO robot is performed to show the effectiveness of the proposed detecting method. Both theoretical and experimental results suggest that the proposed recognition method can reduce the calculation time of the Hough transform and has a certain degree of robustness against uncertain environments.

1. Introduction

The robot’s perception of the environment comes from the mobile robot’s acquisition of environmental information, on basis of the embedded sensors, and extracts valid feature information in the environment for processing and understanding. Environmental perception is the premise basis for humanoid robots to have autonomous decision-making behaviour. Traditional robots use laser radar, ultrasonic waves, and cameras to obtain information about the surrounding environment to realize the perception and recognition of the environment. Based on the monocular vision [1] robot can perform small ball recognition. Because the colour space does not consider the impact of light intensity, the impact of light intensity is relatively large in the process of identifying the small ball, and the result of it will make the robot blind or cause a large error in object recognition. So in order to better adapt to different light intensities, use HSV color space. The corresponding colour space will not change drastically with the change of the light source intensity conditions, and the colour value of the target object will not show a large deviation, which weakens the effect of the lighting conditions on the robot vision system to a certain extent and enhances adaptive capabilities of robot vision systems. A colour watershed [2] image segmentation method has been proposed, which can extract the contours of objects through classification, but the relative error may be large due to uncertain factors.

The existing methods such as SIFT [3] and Camshift [4] are based on the Meanshift algorithm for the video sequence, the initial value of the search window is the result of the previous frame, but the ROI needs to be manually adjusted to obtain the colour characteristics of the target. Kalman Filter [5], and Particle Filter can be used to track moving objects. Camshift can be used to track moving objects, but this method cannot detect obstructing objects in the frame. Traditional template matching algorithms [6] cannot adapt to factors such as attitude changes and noise interference. Inter-frame difference method uses continuous frames in the video to detect and extract targets [7], which is not suitable for the moving target. In a static environment, [8,9] use the difference information of two
consecutive frames to obtain moving target information. The Kalman filter can be used to track moving objects and predict the next possible position of moving objects, but it is limited to the linearity of normal distribution of noise or stabilize the system. This paper adopts the method of Hough circle detection and feature fusion to identify the golf ball, by establishing a mathematical model and calculating the distance between the golf ball and the robot. With this method, NAO can detect the ball with a high accuracy. By adding a Kalman filter, the object position can be predicted such that NAO can shot the ball with more accuracy and robustness.

2. The Target Detection Algorithm

2.1. Image Preprocessing
The NAO robot has 15 joints, 25 degrees of freedom and flexible movement. NAO has two upper and lower cameras: one on the forehead, which is responsible for long-range horizontal scanning with a wide field of view, and the other is located at the NAO port, responsible for the short-term surroundings near the feet. For long-distance object visual recognition, the mouth camera can be used for accurate distance measurement and positioning over short distances. As the robot moves, the camera's horizontal and vertical angles change as the head rotates. With a resolution of 640 x 480, the system provides an image frame rate of 30 frames per second. In practice, the NAO robot is a monolithic camera model because the two cameras cannot operate at the same time, but can switch up and down at any time.

The NAO robot looks around and recognizes the surrounding environment and objects through monocular vision and obtains the image in the video stream input of in the RGB colour space format. However, the RGB colour model is an uneven visual colour space [10]. For objects with the same colour attributes, the distribution of colour values measured under different light intensities or different light sources is very discrete, making the RGB correlations increase with each other, and the band will also be correspondingly wider. The proportion of the entire space occupied has also increased a lot, making the threshold range difficult to determine. When the light is not strong enough or becomes dark, it will cause the robot to fail. The corresponding HSV colour space will not change drastically with the change of the light source intensity conditions, and the colour value of the target object will not show a large deviation. To a certain extent, the impact of the lighting conditions on the robot vision system is reduced, and the adaptive ability of the robot vision system. HSV colour space has a prominent advantage in image processing, it separates the brightness, hue and saturation attributes of the colour, and can reduce the impact of lighting changes [11,12]. Robots use a limited range of precision to store RGB values, but image processing takes a lot of time, which constrains accuracy. First convert the colour space, the conversion formula is as follows:

\[ V = \max (R, G, B) \]

\[ S = \begin{cases} 
\frac{V - \min(R, G, B)}{V} & \text{if } V \neq 0, \\
0 & \text{otherwise}
\end{cases} \]

\[ H = \begin{cases} 
\frac{60(G - B)}{V - \min(R, G, B)} & \text{if } V = R, \\
\frac{120 + 60(B - R)}{V - \min(R, G, B)} & \text{if } V = G, \\
\frac{240 + 60(R - G)}{V - \min(R, G, B)} & \text{if } V = B
\end{cases} \]
The change of these two properties can be controlled by proportion. Due to the characteristics of the RGB model and the HSV model, singular points and unstable points appear in the H component converted from RGB to HSV. The existence of these singular points and unstable points brings inconvenience to subsequent image processing, and often needs to be combined for colour compensation.

![Image of colour segmentation and grayscale processing](image)

**Figure 1.** Image processing.

Figure 1 shows the red threshold of the target red ball to determine the red pixels in the threshold range in the photo taken by the robot. As the NAO robot approaches the target, more and more red pixels will appear in the field of view. When the threshold value of red pixels in the field of view is met, recognition is started. If the area block satisfies the outline as a circle, the target red ball is considered to be detected, otherwise the next red area block to be determined is selected. In such cases, if the robot finds a red target area on the court, and the outline of this area is circular, the target red ball is said to be recognized.

### 2.2. Algorithm Flow

Target detection is an important step before you hit the golf ball accurately. The ultimate goal is to accurately extract the 3 contour features of the target, so as to locate the position of the target ball. Target feature extraction is a key technique in target recognition and has a decisive influence on the final effect of recognition. Since most local features need to ensure brightness, proportion, translation and rotation have certain invariance, the extraction of local features depends more on the specific problem and the knowledge of the corresponding field.

The specific process of red ball recognition is as follows:

**Input:** Video stream input from robot camera  
**Output:** Coordinates of center of the target ball  
1) NAO robot captures images through camera.  
2) Median filtering of the images.  
3) Colour space conversion: Convert RGB colour space to HSV colour space, set the optimal threshold to segment binary images with the same colour characteristics as the target.  
4) Red ball detection: According to the actual size of the ball, set the radius of the circle, use the Hough circle function in OpenCV to detect the target red ball, and extract the outline of the small ball.  
5) Return the coordinates of the red ball in the picture.

### 2.3. Target Detection

In order to reduce noise, the median filtering and graying processing have been performed before the Hough circle transform detects the sphere. Hough transform is a method for describing the shape of the region boundary. It transforms the image space to the parameter space, and describes the curve in the image with some parameter form that most boundary points meet. By setting each edge point to provide the accumulator with a circle of radius R, the peak points corresponding to the accumulated circles at the positions where the centers of the original circles overlap are output to the accumulator space, and the required parameters are obtained. Based on efficiency considerations, Hough circle detection in OpenCV is implemented based on image gradients. The edges are first detected, the possible circle centers are found, and the optimal radius is calculated from the candidate circle centers. Hough circle detection is not based on the results of binary images or edge detection, but instead finds
candidate areas based on the gradient of grayscale images, and then implements Hough circle
detection based on the candidate areas. This will greatly reduce the amount of calculation and improve
the execution of the program Speed and performance. In the separation, the value of the red channel is
enhanced and the result of other spaces is reduced, which can make the separation result better. In
addition, adding Gaussian filtering to blur the local information is beneficial to the detection of the
Hough circle.

3. Mathematical Models for Target Positioning

3.1. Distance Calculation between Robot and Ball
There are three coordinate systems in the Figure 2, which are the image coordinate system $UOV$, the
camera coordinate system with $O_3$ as the origin, and the world coordinate system $UOY$. We can see
that the points in world coordinates are proportional to the points in image coordinates through the
optical axis imaging. The proportion medium is the pixel point $O_i$ of the camera lens center on the
image and its actual point $M$ in world coordinates. It can solve the length of $O_3P$, Other recent
achievements in golf robot technology are improvements in the training system that enables people to
improve their swing skills. The manipulator controls the position and direction of the golf club’s grip,
the calculated parameters are shown in Equation (4).

![Figure 2. Coordinate system representation: $O_3$ is the location of NAO, and the ball is at P.](image)

The known quantities include the camera height $H$, the distance between the world coordinate
point corresponding to the image coordinate center and the camera on the $y$ axis $O_3M$, the image
coordinate $(ucenter, vcenter)$ of the lens center point, and the image coordinate $P_1$ of the measured
pixel point $(u, 0)$. The actual pixel length $x$, the actual pixel width $y$, and the camera focal length $f$.

\[
\begin{align*}
\alpha &= \arctan \left( \frac{H}{O_3M} \right) \\
\gamma &= \arctan \left( \frac{O_3P_1 \times y}{f} \right) \\
\beta &= \alpha - \gamma \\
O_3P &= \frac{H}{\tan \beta}
\end{align*}
\] (4)

3.2. Kalman Filter Tracking
The Kalman filter is used to predict the location of the target, and the target search area in the current frame is determined according to the prediction result. The state equation and measurement equation of the system are given by Equations (5) and (6).

\[ x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \]  

\[ z_k = Hx_k + V_k \]

where \( x_k = (x, y, v_x, v_y)^{-1} \), \( x, y \) are position variables, \( v_x, v_y \) are the speed variables. \( A \) represents an \( n \)-order state transition matrix of a continuous system. \( H \) is the parameter of the measurement system, which maps the real state space to the observation space.

In the above two formulas, \( x \) is the system state at time \( k \), \( z \) is the measured value at time \( k \). \( B_{u_{k-1}} \) is the input control model, and \( w_k \) and \( V_k \) represent process and measured noise, respectively. They are assumed to be Gaussian white noise, and their covariances are \( Q \) and \( R \) respectively (here we assume that they do not change with system state changes). The system state equation is a differential equation (such as an inertial navigation system). Figure 3 shows the difference between the estimate and the true value, where the estimate gets closer to the true value and eventually fluctuates near the real value.

**Figure 3.** The error between the estimated value and the true value.

The core points of Kalman filtering are state and covariance, as described by

\[
\begin{align*}
\dot{x}_{k|k-1} &= A\hat{x}_{k-1} + B_{u_{k-1}} + w_{k-1} \\
P_{k|k-1} &= AP_{k-1|k-1}A^T + Q_{k-1} \\
K_k &= P_{k|k-1}H_k^T (H_kP_{k|k-1}H_k^T + R_k)^{-1} \\
\hat{x}_k &= \dot{x}_{k|k-1} + K_k (z_k - H_k\hat{x}_{k|k-1}) \\
P_k &= (I - K_kH_k)P_{k|k-1}
\end{align*}
\]

(7)
where $\hat{x}_{k|k-1}$ represents the prior state estimation value at time $k$, which is an unreliable estimation made by the algorithm based on the result of the previous iteration, $\hat{x}_k$ and $\hat{x}_{k-1}$ respectively represent the posterior state estimate at time $k$ and $k-1$, which is the output to be The best estimate at the moment, this value is the result of Kalman filtering. $P_{k|k-1}$ represents the a priori estimated covariance at time k. When this covariance matrix determines the initial $P_0$, it can be derived later, and if the initial covariance matrix $P_0$ is not 0, its value pair. The filtering effect has little influence and can converge quickly. $P_k$ and $P_{k-1}$ respectively represent the covariance of the posterior estimation at time $k$ and $k-1$, which are one of the filtering results. $K_k$ represents Kalman gain, which is the intermediate result of filtering. $Q$ represents the covariance of process excitation noise, which is the error between the state transition matrix and the actual process.

The robot hitting the golf ball in a relatively short time can be considered as a linear model. According to the algorithm, the motion characteristics of the ball are obtained. In the tracking process, the Kalman filter completes the prediction and calculates the prior error covariance. Find the target location in the search area.

4. Experimental Results

The experiment was performed on the office floor using a laptop and a sixth-generation NAO robot. The experimental program is based on python2.7 and the open source library OpenCV. In the experiment, the NAO robot and the unprofessionally trained students will participate in the test, placing the golf ball in 5 different positions, and performing 10 tests each time. The NAO robot and the students do not hit the ball more than 1 times each time. The test ends when the golf ball does not hit the hole or cross the boundary. The prerequisite for the NAO robot to win the game is to accurately hit the ball into the hole.

| Distance/m | 0.6 | 1.2 | 1.8 | 2.4 | 3.0 | Average success rate |
|-----------|-----|-----|-----|-----|-----|----------------------|
| Student   | 9   | 7   | 6   | 4   | 4   | 60%                  |
| Nao       | 9   | 8   | 8   | 7   | 7   | 78%                  |

Figure 4 and Figure 5 show the whole process of the robot and hitting the ball and as can be seen from Table 1, the average success rate of the Nao robot is 18% higher than the average success rate of students. The accuracy rate of the students in close range is very high, but the accuracy in long range is low. The reason for the low accuracy rate is the uneven human force and the deviation of the sense of direction. The eyes, the ball, and the hole form a three-dimensional three-point line. There is human error, so the long-distance accuracy rate decreases. But the robot is positioned through accurate calculations, and the swing is stable, so the accuracy is high. So the average success rate of NAO robots is higher than humans.
5. Conclusion

In this paper, a golf ball recognition method based on colour threshold segmentation and Kirchhoff circle detection is proposed. NAO can realize the rapid detection of small balls, reducing the impact of different light on the experiment, ensuring accuracy and robustness. Considering the situation of losing the ball, the Kalman filter method is used to predict and track the ball. Experiments show this proposed algorithm improves the accuracy when NAO detects golf balls. Future work includes designing learning-based adaptive recognition methods.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 61976100 and 61672245 and by the HUST Independent Innovation Fund 5003187039.

References

[1] Zhang Z, Hou LW and Wang H S 2018 Research on robot golfing method based on feature fusion and kalman filter 2018 IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (Shenyang, China, 2018) pp 231-239
[2] Lezoray O and Cardot H 2002 Cooperation of color pixel classification schemes and color watershed: a study for microscopic images in IEEE Transactions on Image Processing (July 2002) 11(7) pp 783-789
[3] Lu Z P 2010 Fast image matching based on the simplified algorithm of SIFT Manufacturing Automation 32 133-5 (in Chinese)
[4] Michael B 2001 The Effects of Capture Conditions on the CAMSHIFT Face Tracker Report 2001-691-14
[5] Weng S, Kuo C M and Tu S K 2006 Video object tracking using adaptive kalman filter Journal of Visual Communication and Image Representation (S1190-1208) 17 1194-7
[6] Ru A, Tao X X and Wang H L 2010 Rotation and translation invariant template matching based on edge points for remotely sensed imagery from two sources The 15th National Academic Conference Proceedings about Image and Graphics 23 374-380
[7] Zheng B, Xu x Y, Dai Y P and Lu Y Y 2012 Object Tracking Algorithm Based on Combination of Dynamic Template Matching and Kalman Filter Intelligent Human-Machine Systems and Cybernetics (IHMSC) pp 136-139
[8] Chien S Y, Ma S Y and Chen L G 2002 Efficient moving object segmentation algorithm using background registration technique IEEE Transactions on Circuits and Systems for Video Technology 12(7) 577-586
[9] Lipton A J, Fujiyoshi H and Patil R S 1998 Moving target classification and tracking from real-time video Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201) (Princeton, NJ, USA) pp 8-14
[10] Sara I and Alfredo P. 2015 Salient feature based graph matching for person re-identification Pattern Recogn. 48(4) 1074–85
[11] Mohammad A S, Aini H and Noorita T et al 2012 Appearance-based methods in re-identification: A brief review 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications (CSPA) pp 404-408

[12] An L, Mehran K, Yang S F 2013 Reference-based person re-identification IEEE International Conference on Advanced Video and Signal Based Surveillance pp 244-9