An Efficient Object Recognition System For Humanoid Robot Vision

Wei-Hsuan Chang, Chih-Hsien Hsia, Yi-Che Tai, Shih-Hung Chang, Fun Ye, and Jen-Shiun Chiang
Department of Electrical Engineering, Tamkang University, Taipei, Taiwan
E-mail: whchang@ee.tku.edu.tw; chhsia@ee.tku.edu.tw; choose29999@hotmail.com; 696450682@s96.tku.edu.tw; fyee@mail.tku.edu.tw; chiang@mail.tku.edu.tw

Abstract

The research of autonomous robots is one of the most important issues in recent years. In the numerous robot researches, the humanoid robot soccer competition is very popular. The robot soccer players rely on their vision systems very heavily when they are in the unpredictable and dynamic environments. This paper proposes a simple and fast real-time object recognition system for the RoboCup soccer humanoid league rules of the 2009 competition. This vision system can help the robot to collect various environment information as the terminal data to finish the functions of robot localization, robot tactic, barrier avoiding,..., etc. It can decrease the computing efforts by using our proposed approach, Adaptive Resolution Method (ARM), to recognize the critical objects in the contest field by object features which can be obtained easily. The experimental results indicate that the proposed approach can increase the real time and accurate recognition efficiency.

Keywords: Robot, RoboCup, Adaptive Resolution Method, Object Recognition, Real Time.

1. Introduction

RoboCup [1] is an international joint project to stimulate researches in the field of artificial intelligence, robotics, and related fields. According the rules of RoboCup for 2009 in the humanoid league of kid size [2], the competitions take place on a rectangular field of 600×400 cm2 area, which contains two goals and two landmark poles, as shown in Figure 1. A goal is placed in the middle of each goal line. One of the goals is colored yellow and the other is colored blue. The goal for the kid size field has 90cm of height for the crossbar, 40cm of height for the goal wall, 150cm of width for the goal wall, and 50cm of depth for the goal wall. The two landmark poles are placed at each side of the two intersection points between the touch line and the center line. The landmark pole is a cylinder and has a diameter of 20cm. It consists of three segments of 20cm height, stacked each other. The lowest and the highest segments are with the same color as the goal of its left side. The ball is the standard size orange tennis ball. All of the above objects are the most critical characteristics in the field, and they are also the key features which we have to pay attention to.

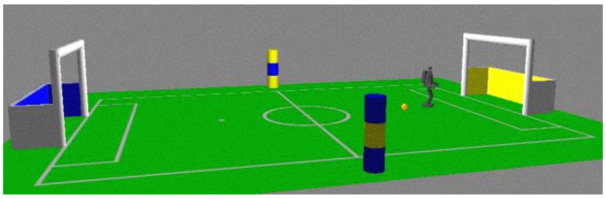


Figure 1. The field of the competitions [2].

The functions of humanoid robot vision system include image capturing, image analyses, and digital image processing by using visual sensors. For digital image processing, it is to transform the image into the analyzable digital pattern by digital signal processing. We can further use the image analysis technique to describe and recognize the image content for the robot vision. The robot vision system can use the environment information captured in front of the robot to recognize the image by means of the technique of human vision system. An object recognition algorithm is thus proposed to the humanoid robot soccer competition.

Generally speaking, object recognition uses object features to extract the object out of the picture frame, and thus color [10]-[11], shape [5]-[6], contour [7]-[9], texture, and sizes of object features are commonly used. Because the object color is distinctive in the contest field, we mainly choose the color information to determine the critical objects. Although this approach is simple, the real time efficiency is still low. Because there is a lot of information to be processed in every frame for real time consideration, Sugandi et al. [14] proposed a low resolution method to reduce the information. It can speed up the processing time, but the low resolution results in a shorter recognizable distance, and it may increase the false recognition rate. In order to improve the mentioned drawbacks, we propose a new approach, adaptive resolution method (ARM), to reduce the computation complexity and increase the accuracy rate.

The rest of this paper is organized as follows. Section 2 presents the related background such as the general color based object recognition method, low resolution method, and encountered problems. Section 3 describes

978-1-4244-5228-6/09/$26.00 c2009 IEEE

the proposed approach, ARM. The experimental results are shown in Section 4. Finally, the conclusions and future works are outlined in Section 5.

2. Background

2.1. Color Based Object Recognition Method

A good vision system is playing an important role for the humanoid robot soccer players. Many robot vision modules have provided some basic color information, and it can extract the object by selecting the color threshold. The flow chart of a traditional color recognition method is shown in Figure 2. The RGB [3] color model comes from the three additive primary colors, red, green, and blue. The main purpose of the RGB color model is for the sensing, representation, and display of images in electronic systems, such as televisions and computers, and it is the basic image information format. The RGB color model can describe all colors by the different proportion combinations. Because the RGB color model is not explicit, it can be easily influenced by the light illumination and make people select error threshold values.

An HSV [3] color model relates the representations of pixels in the RGB color space, which attempts to describe perceptual color relationships more accurately than RGB. Because the HSV color model describes the color and brightness component respectively, the HSV color model is not easily influenced by the light illumination. The HSV color model is therefore extensively used in the fields of color recognition. The HSV transform function is shown in equations (1), (2) and (3) as follows.

$$H=\begin{cases}\left(6+\frac{G\text{-}B}{C_{MAX}\text{-}C_{min}}\right)\times 60°, \text{ if } R=C_{MAX}\\ \left(2+\frac{B\text{-}R}{C_{MAX}\text{-}C_{min}}\right)\times 60°, \text{ if } G=C_{MAX}\\ \left(4+\frac{R\text{-}G}{C_{MAX}\text{-}C_{min}}\right)\times 60°, \text{ if } B=C_{MAX}\end{cases} \tag{1}$$

$$S=\frac{C_{MAX}\text{-}C_{min}}{C_{MAX}} \tag{2}$$

$$V=C_{MAX} \tag{3}$$

In (1), (2), and (3), H is hue, and its range is 0°~360°; S means saturation, and its range is 0~1; V represents value, and its range is 0~255. The RGB values are confined by (4):

$$C_{MAX}=MAX(R, G, B), C_{min} = min(R, G, B) \tag{4}$$

where C_{MAX} is the maximum value in the RGB color components, and C_{min} is the minimum value in the RGB color components. Hence, we can directly make use of H and S to describe a color range of high environmental tolerance. It can help us to obtain the foreground objects mask M(x, y) by the threshold value selection in (5).

$$M(x,y)=\begin{cases}1, \text{ foreground}\\ \text{if } T_L<H<T_H \cap S>Thd_S\\ \\ 0, \text{ background}\\ \text{otherwise}\end{cases} \tag{5}$$

$$T_L=Thd_H\text{-}R_H,\ T_H=Thd_H+R_H$$

where Thd_H, Thd_S and R_H are the threshold of hue, threshold of saturation, and the range of hue respectively by manual setting. The foreground object mask usually accompanies with the noise, and we can remove the noise by the simple morphological methods, such as dilation, erosion, opening, and closing. It needs to separate the objects by labeling when many objects with the same colors are existed in the frame. The following procedure is the flow for labeling :

Step 1: Scan the Threshold Image M(x, y).

Step 2: Give the value $Label^{i}_{color}$ to the connected component Q{n} of pixel(x, y).

Step 3: Give the same value $Label^{i}_{color}(x,y)$ to the connected component of Q{n}.

Step 4: Until no connected component can be found.

Step 5: Update $Label^{i}_{color}$, i = i+1. Then go to Step 1 and repeat Steps 2.~4.

Step 6: Completely scan the image.

By using the procedure mentioned above, the objects can be extracted. Although this method is simple, it is only suitable for low frame rate sequences. For a high resolution or noisy sequence, this approach may need very high computation complexity.

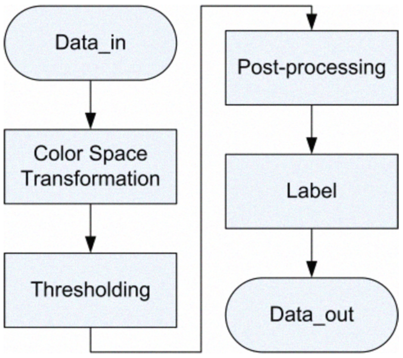


Figure 2. The flow chart of the traditional color recognition method.

2.2. Low Resolution Method

For overcoming the above-mentioned problems, several approaches of low resolution method were proposed [12][14]. The flow chart of a general low resolution method is shown in Figure 3. Several low resolution methods, such as the approach of applying 2-D Discrete Wavelet Transform (DWT) and the using of 2×2 average filter, were discussed. Cheng *et al.* [12] applied 2-D DWT for detecting and tracking moving objects and only the LL3-band image is used for detecting motion of the moving object. Because noises are preserved in high-frequency, it can reduce computing cost for post-processing by using the LL3-band image. This method can be used for coping with noise or fake motion effectively; however the conventional DWT scheme has the disadvantages of complicated calculation when an original image is decomposed into the LL-band image. Moreover if it uses an LL3-band image to deal with the fake motion, it may cause incomplete moving object detecting regions. Sugandi *et al.* [14] proposed a simple method by using the low resolution concept to deal with the fake motion such as moving leaves of trees. The low resolution image is generated by replacing each pixel value of an original image with the average value of its four neighbor pixels and itself as shown in Figure 4. It also provides a flexible multi-resolution image like the DWT. Nevertheless, the low resolution images generated by using the 2×2 average filter method are more blurred than that by using the DWT method, as shown in Figure 5. It may reduce the preciseness of post-processing (such as object detection, tracking, and object identification), because the post-processing depends on the correct location of the moving object detecting and accuracy moving object.

In order to detect and track the moving object more accurately, we propose a new approach, ARM, which is based on the 2-D integer symmetric mask-based discrete wavelet transform (SMDWT) [13]. It does not only retain the features of the flexibilities for multi-resolution, but also does not cause high computing cost when using it for finding different subband images. In addition, it preserves more image quality of the low resolution image than that of the low resolution method [14].

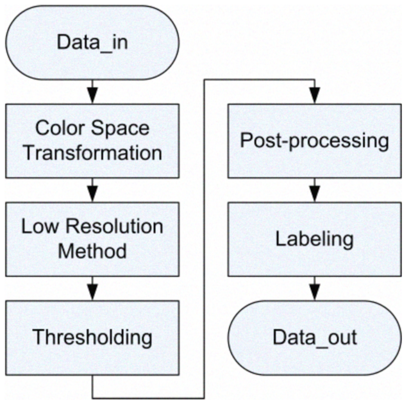


Figure 3. The flow chart of a general low resolution method.

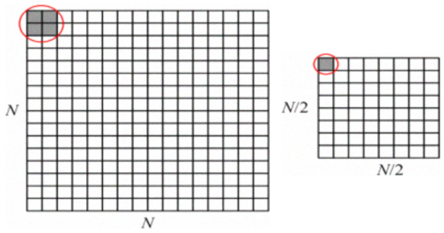
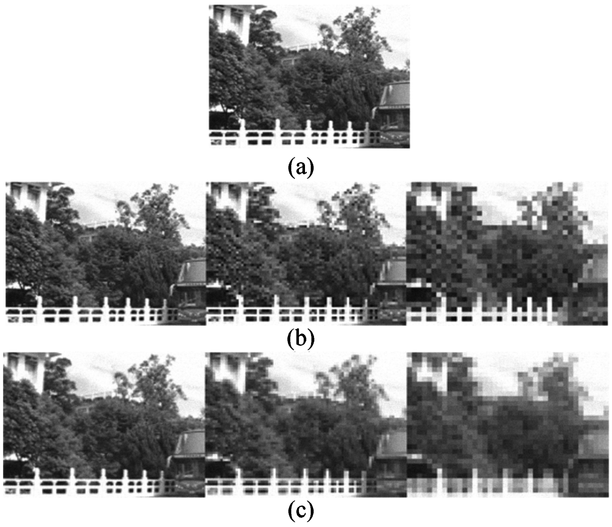


Figure 4. Diagram of the 2×2 average filter method.



(a)

(b)

(c)

Figure 5. Comparisons of low resolution images. (a) Original image (320×240). (b) Each subband image with DWT from left to right as 160×120, 80×60, and 40×30, respectively. (c) Each resolution image with the 2×2 average filter method from left to right as 160×120, 80×60, and 40×30, respectively.

3. The Proposed Method

3.1. Adaptive Resolution Method (ARM)

ARM takes the advantage of the information obtained from the motor to know the object distance and chooses the most proper resolution. The operation flow chart is shown in Figure 6. After HSV color transformation, ARM has two operation modes, Manual Mode and Tracking Mode. The Manual Mode can let us manually choose the resolution. The high resolution approach brings a better recognizable distance but with a slower running speed. On the other hand, the low resolution approach brings a lower recognizable distance but with a faster running speed. When we get the information from the motor angle, we can convert it as the "sel" signal through the adaptive selector to choose the appropriated resolution. The "sel" condition is shown in (6):

$$\begin{cases} \text{sel=0 (Do Nothing)} \quad , \text{if } D_{\text{ball}} > D_{\text{thd2}} \cup f_b = 1 \\ \text{sel=1 } \left(\text{1-Level DWT}\right), \text{if } D_{\text{thd1}} > D_{\text{ball}} \geq D_{\text{thd2}} \\ \text{sel=2 } \left(\text{2-Level DWT}\right), \text{if } D_{\text{ball}} \leq D_{\text{thd1}} \end{cases} \qquad (6)$$

The relationship between the resolution and the distance of the ball is described in Table 1. According to Table 1, we can conclude a distance equation as follows:

$$D_{\text{ball}} = H_{\text{cam}} \times \tan \theta_m \qquad (7)$$

where H_{cam} is the height of the camera place, and θ_m is the information of the motor angle. In (6), D_{thd1} and D_{thd2} are the threshold values for the recognizable distance and are set to 0.6 and 2.5, respectively. D_{ball} is the distance between the robot and ball that obtained from (7). In order to obtain more accurate D_{ball}, we have to keep the ball in the center of the frame to reach the function of ball tracking. If the ball disappears in the frame, the flag f_b is set to 1. At the same time, the frame changes into the original size to have a higher probability to find out the ball. Since the sizes of the other critical objects (such as goal and landmark) in the field are larger than the ball, they can be recognized easily. Figure 7 shows the results of different resolutions after the threshold processing.

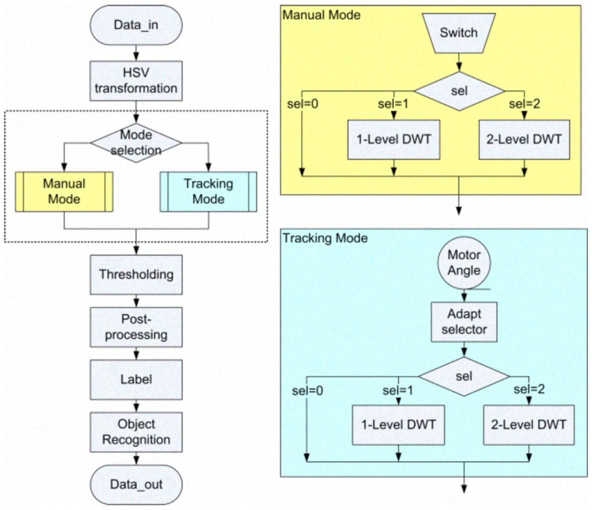


Figure 6. The flow chart of ARM.

Table 1. The relationship between the resolution and the distance of the ball.

Resolution	*Run time (sec)*	*Frame rate (fps)*	*Recognizable distance (m)*
320×240	0.072	13.8	3.8
160×120	0.041	24.4	2.6
80×60	0.034	29.4	0.8

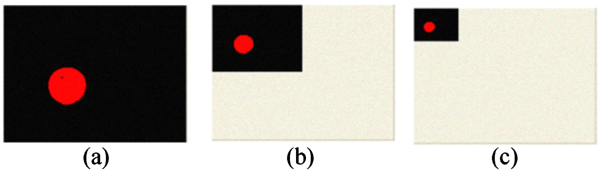


Figure 7. The results after the threshold-processing under the resolution (a) 320×240. (b) 160×120. (c) 80×60.

3.2. Sample Object Recognition Method

According to the above-mentioned color segmentation method, it can fast and easily extract the orange ball in the field, but it is not enough to recognize the goals and landmarks. The colors of the goals and landmarks are yellow and blue, and by color segmentation the extraction of goals and landmarks may not be correct as shown in Figure 8. Therefore we have to use more features and information to extract them. Since the contest field is not complicated, a simple recognition method can be used to reduce the computation complexity. The landmark is a cylinder with three colors. Let us look at one of the landmark with the upper and bottom layers in yellow, and the center layer in blue; this one is defined as the YBY-landmark. The color combinations of the other one are in contrast of the previous one, and the landmark is defined as the BYB-landmark. The labels of the BYB landmark can be calculated by (8). The YBY landmark is in the same manner as the BYB landmark.

$$\begin{aligned}&\mathrm{Landmark}_{BYB}(x,y)=\mathrm{Label}_B^i(x,y)+\mathrm{Label}_B^j(x,y)+\mathrm{Label}_Y^k(x,y)\\ &\text{if}\left|\mathrm{Label}_B^j(X_{min})-\mathrm{Label}_B^i(X_{min})\right|<\alpha \qquad (8)\\ &\cap\left|\mathrm{Label}_B^j(X_{max})-\mathrm{Label}_B^i(X_{max})\right|<\alpha\\ &\cap\mathrm{Label}_B^i(Y_{max})<\mathrm{Label}_Y^k(Y_c)<\mathrm{Label}_B^j(Y_{min})\end{aligned}$$

where $\mathrm{Label}_B^i(x,y)$ is the pixel of the i-th blue component in a frame, X_{min} the minimum value for the object i at the x direction in the frame, X_{max} the maximum value, Y_{min} and Y_{max} the minimum value and the maximum value at y direction respectively, and Y_c the center point of the object at the vertical direction. The threshold value α is set as 15. The landmark is composed of two same color objects in the vertical line, and the center is in different color. If it can find an object with this feature, the system can treat this object as the landmark. Equation (9) is used to define the label of the ball.

$$\begin{aligned}&\mathrm{Ball}(x,y)=\mathrm{Label}_O^s(x,y) \qquad (9)\\ &\text{if the size of }\mathrm{Label}_O^s\text{ is maximum of }\mathrm{Label}_O\\ &\cap\beta_1<\frac{\mathrm{Label}_O^s(X_{max})-\mathrm{Label}_O^s(X_{min})}{\mathrm{Label}_O^s(Y_{max})-\mathrm{Label}_O^s(Y_{min})}<\beta_2\end{aligned}$$

where $\mathrm{Label}_O^s(x,y)$ is the pixel of the s-th orange component in a frame. Since the ball is very small in the picture frame, in order to avoid the noise, the ball is treated as the maximum orange object and with a shape ratio of height to width approximately equal to 1. Here β_1 and β_2 are set to 0.8 and 1.2, respectively. The goal recognition is defined in (10).

$$\begin{aligned}&\mathrm{Goal}_B(x,y)=\mathrm{Label}_B^m(x,y) \qquad (10)\\ &\text{if }\frac{\mathrm{Label}_B^m(X_{max})-\mathrm{Label}_B^m(X_{min})}{\mathrm{Label}_B^m(Y_{max})-\mathrm{Label}_B^m(Y_{min})}>\gamma\end{aligned}$$

where $\mathrm{Label}_B^m(x,y)$ is the pixel of the m-th blue component in a picture frame. Since the blue goal is composed of the blue object and the shape ratio of the height to the width is greater than 1.2, the parameter γ is set as 1.2. The yellow goal is in the same manner as the blue goal.

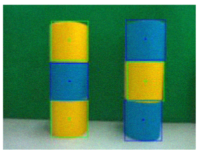


Figure 8. False segmentation of the landmark.

4. Experimental Results

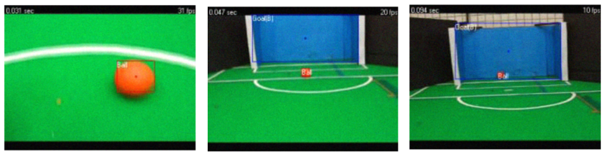
In this work, the environment information is extracted by the Logitech QuickCam Ultra Vision [16]. The resolution is 320×240 pixels, and the frame rate is 30 fps (frame per second). The Dynamixel RX-28 Motor [17] is used on the head of the robot as the motion device for horizontal and vertical direction movement. For simulation computer, the CPU is Intel Core 2 Duo CPU 2.1GHz, and the development tool is C^{++} Builder 6.

4.1. Room in and room out of the picture frame

In this experiment, the camera is set in the center of the field. The scene simulates that the robot kicks ball into the goal and the vision system will track the ball. The experimental results of the accuracy rate and average FPS (frame per second) under different resolutions and ARM are shown in Table 2. “False Positive” means the error misdiagnosis. “False Negative” means that it does not recognize the object. According to Table 2 we find that even though the 320×240 resolution had high accuracy rate, the process speed is slow. The 80×60 resolution has the highest processing speed, but it has low accuracy rate. By this approach, it gets high accuracy rate only when the object is close to the camera. On the other hand, the proposed ARM does not only have high accuracy rate, but also keep high processing speed. The result of ARM is shown in the Figure 9.

Table 2. The experimental results of accuracy rate and average FPS under different resolutions and ARM.

Resolution	Total Frame	Object Frame	False Positive	False Negative	Accuracy Rate	Average FPS
320×240	124	87	0	3	96.55%	14.2
160×120	230	164	2	43	72.56%	24.3
80×60	242	181	1	110	38.67%	29.1
ARM	135	97	1	2	96.91%	20.5



(a) (b) (c)

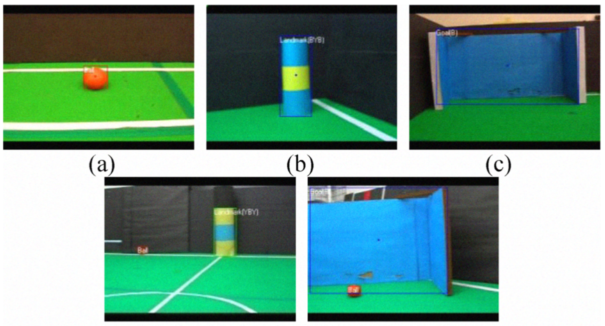
Figure 9. The result of ARM when the adaptive selector chooses (a) 80×60 resolution with level-2 SMDWT, (b) 160×120 resolution with level-1 SMDWT, (c) 320×240 resolution with the original input.

4.2. The function of object recognition

In this experiment, several scenes are simulated. Scene 1: it closes the ball slowly. Scene 2: the camera turns left to see the BYB landmark and keeps turning until the BYB landmark disappears. Scene 3: the camera turns up to see the goal and turns right and then turns left until the goal disappears. Scene 4: the YBY landmark is always in the frame and the ball enters from the bottom of the frame and then the camera turns left to see the similar color object. Scene 5: the camera turns left to see the YBY landmark, ball, and goal respectively. The experimental results of these scenes are shown in Table 3. According to the simulation results, our proposed method accommodates many kinds of scenes and has high accuracy rate of more than 94% on average. The experimental results are shown in Figure 10.

Table 3. The experimental results of the several kinds of scene simulation.

Scene	Total Frame	Object Frame	False Positive	False Negative	Accuracy Rate
(1) ball	691	691	0	7	98.99%
(2) landmark	290	191	3	21	87.43%
(3) goal	232	212	0	13	93.87%
(4) ball&landmark	753	753	0	18	97.61%
(5) ball&goal&landmark	616	616	12	68	87.01%
Total	2582	2463	15	127	94.23%



(a) (b) (c)

(d) (e)

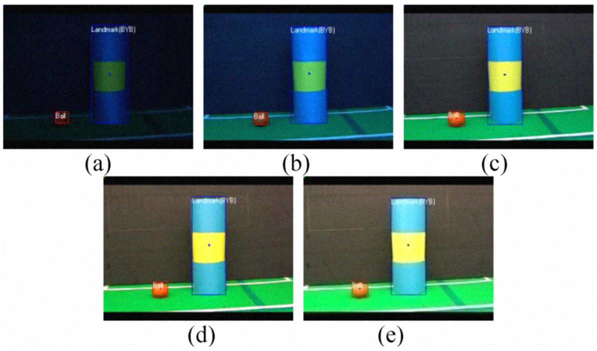
Figure 10. The experimental results of the several scenes: (a) Scene 1 at frame of 429. (b) Scene 2 at frame of 137. (c) Scene 3 at frame of 108. (d) Scene 4 at frame of 323. (e) Scene 5 at frame of 247.

4.3. The environmental tolerance

In this experiment, the object recognition is accomplished at the environment with different luminance. The luminance is set as 16 lux, 178 lux, 400 lux, 596 lux, and 893 lux, respectively. The camera took picture frames of the ball and landmark with a fixed time period in different luminance. The level-1 SMDWT is used in the manual mode. The recognition accuracy rates of the ball and BYB landmark in different luminance is shown in Table 4. According to Table 4, the proposed method is robust on the light luminance changing. The accuracy rate is more than 95% on average. It is more than 75% accuracy rate even though the luminance is very low (16 lux). The recognition results in different luminance are shown in Figure 11.

Table 4. The accuracy rate of the ball and BYB landmark recognition in different luminance.

Luminance (lux)	*Total Frame*	*Object Frame*	*False Positive*	*False Negative*	*Accuracy Rate*
16	205	205	0	47	77.07%
178	387	387	0	9	97.17%
400	214	214	1	5	97.20%
596	296	296	0	2	99.32%
893	350	350	0	3	99.14%



(a) (b) (c)

(d) (e)

Figure 11. The results of the ball and BYB landmark recognition in the different luminance at (a) 16 lux, (b) 178 lux, (c) 400 lux, (d) 596 lux, (e) 893 lux.

5. Conclusion

An outstanding humanoid robot soccer player must have a powerful object recognition system to fulfill the functions of robot localization, robot tactic, and barrier avoiding. In this paper we propose an HSV color based object segmentation method to accomplish the object recognition. The object recognition approach uses the proposed adaptive resolution method (ARM) and sample object recognition method, and it can recognize objects more robustly. The experimental results indicate that the proposed method is not only simple and fast, but also achieves high accuracy and efficiency with the functions of object recognition and tracking.

Because the error recognition happens when the objects occlude and the objects are just entering or exiting the frame, in the future we will solve this kind of problems to increase the accuracy rate and add more powerful functions, such as line detection and the opponent or partner robots recognition. Our future work aims to find a way to build a "real" humanoid robot vision system.

Acknowledgement

This work was supported by the National Science Council of Taiwan, R.O.C. under grant number NSC 98-2218-E-032-003.

References

[1] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "Robocup: The robot world cup initiative," *IJCAI-95 Workshop on Entertainment and AI/ALife*, pp. 19-24, 1995.

[2] *RoboCup Soccer Humanoid League Rules and Setup for the 2009 competition.* http://www.robocup2009.org/153-0-rules.

[3] R. C. Gonazlez and R. E. woods, *Digital Image Processing*, 2nd edition, Addison-Wesley, 1992.

[5] F. Chaumette, "Visual servoing using image features defined on geometrical primitives," *IEEE Conference on Decision and Control*, pp. 3782-3787, Dec. 1994.

[6] J.-H. Jean and R.-Y. Wu, "Adaptive visual tracking of moving objects modeled with unknown parameterized shape contour," *IEEE Conference Networking, Sensing and Control*, pp. 76-*81, March 2004.*

[7] S. J. Sun, D. R. Haynor, Y. M. Kim, "Semiautomatic video object segmentation using V snakes," *IEEE Transactions on Circuits System Video Technol.* vol. 13, no. 1, pp.75-82, Jan. 2003.

[8] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision.*, vol. 1, pp.321–331, Jan. 1988.

[9] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, Nov. 1986.

[10] N. Herodotou, K.N Plataniotis, and A.N. Venetsanopoulos, "A color segmentation scheme for object-based video coding," *IEEE Symposium on Advances in Digital Filtering and Signal Processing*, pp.25-29, June 1998.

[11] O. Ikeda, "Segmentation of faces in video footage using HSV color for face detection and image retrieval," *International Conference on Image Processing*, vol. 3, pp. 913-6, Sep. 2003.

[12] F.-H. Cheng and Y.-L. Chen, "Real time multiple objects tracking and identification based on discrete wavelet transform," *Pattern Recognition*, vol. 39, no. 3, pp. 1126-1139, Jun. 2006.

[13] C.-H. Hsia, J.-M. Guo, and J.-S. Chiang, "Improved low-complexity algorithm for 2-D integer lifting-based discrete wavelet transform using symmetric mask-based scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 8, pp. 1-7, Aug. 2009.

[14] B. Sugandi , H. Kim, J. K. Tan, and S. Ishikawa, "Tracking of moving objects by using a low resolution image," *International Conference on Innovative Computing, Information and Control,* pp. 408-408, Sep. 2007.

[15] C.-H. Hsia, J.-M. Guo, J.-S. Chiang, and C.-H. Lin, "A novel fast algorithm based on SMDWT for image applications," *IEEE International Symposium on Circuits and Systems,* pp. 762-765, May 2009.

[16] The specification of Logitech QuickCam Ultra Vision. http://www.logitech.com/index.cfm/webcam_communications/webcams/devices/238&cl=tw,zh.

[17] The specification of Dynamixel RX-28 Motor. http://www.crustcrawler.com/motors/RX28/docs/RX28_Manual.pdf.