Proceedings of The 2016 IEEE International
Conference on Real-time Computing and Robotics
June 6-9, 2016, Angkor Wat, Cambodia

Real-time Target Tracking and Positioning on FPGA

Chen Chen, Yunhui Liu, *Fellow, IEEE* , Congyi Lyu, Weiguo Zhou, Jianqing Peng, Xin Jiang, Peng Li

***Abstract*—Recently, FPGA becomes more important and popular in embedded systems due to its parallelizable inner structures. In this paper, a pipelined FPGA architecture will be present which will be used on a mobile robot. It implements algorithms about target recognition，tracking, positioning based on monocular vision , etc. The hardware platform for the design is the development board by Xilinx. The system uses monocular camera to measure the orientation of target in real time. It acquires the video stream from CMOS camera. The system realizes target tracking by extracting the color of the target as the feature, filtering，detecting feature of the target and then displaying on the screen. The system realizes target positioning according to the centroid of the tracked target, then computing the orientation which can be used for the intelligent robot vision.**

I. Introduction

Mobile Robot monocular vision system is a hot topic in the field of mobile robots [1-2]. Like humans, vision is one of the most important ways of accessing to information, it can provide like relative position and other information which can be used for the robot. Comparing to the panorama and binocular vision sensor, monocular vision sensor has a simple structure, an easy calibration and flexible application. Therefore, the study of mobile robot vision system is significant [3-4]. Consequently, researching on the tracking and positioning on target is a subject of importance.

There are several approaches on real-time image tracking. Some special features or matching template can be used to track objects. As study in [5], the target can be tracked based on LBP. In [6-8], The Sum of Absolute Differences method is used for matching. The proposed algorithm is implemented on FPGA. The algorithm is accelerated with effective usage of Block RAMS distributed on FPGA. The proposed algorithm became fast enough for real-time object tracking applications. Another approach is to distinguish objects and background from images. As study in [9-12], a block-based motion estimation method was used to track a moving object. By comparing between the images and the template, we can obtain the information of the moving object. This method is often used when the background is unchanged [13-15]. By comparing input image with the template, the different area will be referred to be the moving objects. But when the background is changeable, it will be matched by mistake. In study [16-17], an adaptive block matching algorithm was proposed for tracking. However, in [18], they need complex computation so they usually implemented only by PC [19].

This paper is mainly about target recognition and positioning based on monocular vision mobile robot system [20-21]. It includes target recognition, tracking and monocular vision positioning algorithms. The main contents of this paper are: target recognition, target tracking and monocular positioning.

For the target recognition module, this paper uses color as the feature to be recognized and then transfer RGB to HSV because HSV is more robust to lightness than RGB [22], the values of HSV are used to recognize the target and then the target parts are processed into binary image to calculate the centroid of the target easily. When the centroid of the target is calculated, PID feedback was used to control the motor to make it track in real-time [23]. This method is very effective for a single color target to be recognized and scenes without too much similar objectives to interference with.

In order to achieve visual orientation for the mobile robot with single camera using monocular vision technology, a spatial point positioning model is established. In [24], the model based on the camera perspective model of holes, two dimensional coordinates of images are mapped to three dimensional coordinates in the same system. The system of the robot is according to the geometrical relationships to realize the target positioning based on monocular vision.

II. Overview of the System

This design of video image processing system is based on Xilinx Zynq-7000 All Programmable SoC. The platform of this processing system combines FPGA and ARM efficiently. The system program can be divided into three modules, namely a target recognition module, target tracking module and monocular positioning module. The basic steps video image processing in this system can be divided into data acquisition, format conversion, data processing and data displaying. The following sections will describe the design of each module in FPGA development environment in detail. The block diagram of the system is shown in Figure 1.

III. TARGET RECOGNITION MODULE

Since the OV7725 CMOS Sensor original output as Bayer array, it will reduce the quality of the image. Therefore, we will transfer from Bayer to RGB888 directly to accelerate the subsequent color image processing system.

Chen Chen is with the School of Mechanical Engineering and Automation, Harbin Institute of Technology Shenzhen Graduate School (e-mail: chenchen921103@163.com).

Yunhui Liu is with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong.

Congyi Lyu is with the Intelligent Robotics Institute, School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China. And the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong.

Weiguo Zhou, Jianqing Peng, Xin Jiang, Peng Li are all with School of Mechatronical Engineering, Harbin Institute of Technology Shenzhen graduate school, Shenzhen, China.

978-1-4673-8959-4/16/$31.00 © 2016 IEEE

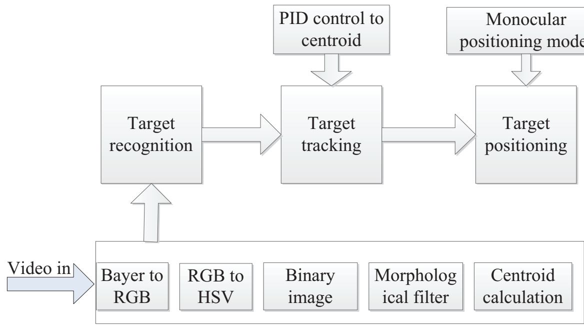

Figure 1. Block diagram of the system

In this paper, two shift register devices were used to build a line buffer. By this line buffer device the 3x3 window based filter algorithm can be established. In Figure 2, three lines of a window were stored in the three shift registers. By the shift controller, the data flow can shift sequentially in the line buffer.

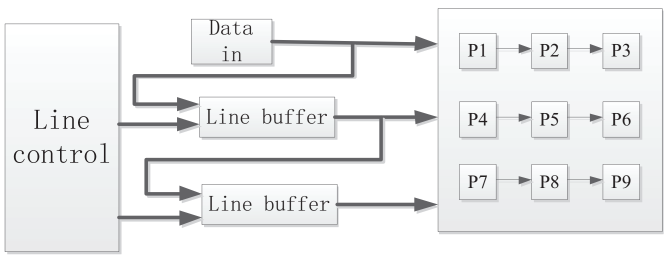

Figure 2. Diagram of line buffer

a. Bayer to RGB

Since the OV7725 CMOS Sensor output Bayer array originally, it will reduce the quality of the image. Therefore, the format of the image should be transferred from RAW to RGB888 directly for accelerating the subsequent color image processing system. The timing simulation diagram of transferring from Bayer to RGB is as Figure 3.

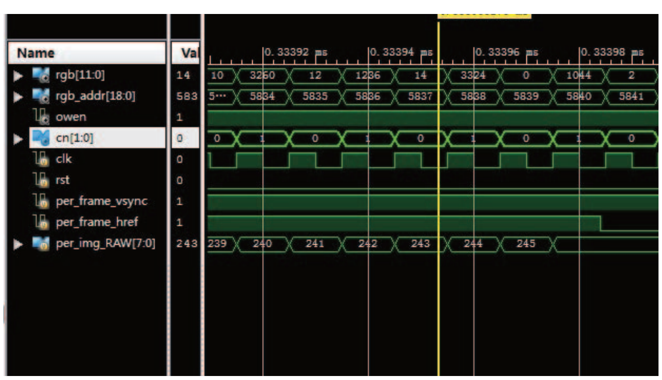

Figure 3. The timing simulation diagram of Bayer to RGB

b. RGB to HSV

The three components of RGB color space model are closely related to the brightness, which will be affected by light intensity easily. The same value of R, G, B will generate a large change in the color space when the light intensity changed, which is difficult to guarantee the stability of the experiment. We cannot judge a certain color through a single threshold value to identify the target object accurately. HSV color space model is a more intuitive color model which can be evolved from the RGB model. Because chrominance and luminance component are separated, color information will not be affected by luminance component V. The color information of the image is robust when the conditions of the ambient light are changing constantly. In view of this, HSV color space will be chosen as the feature to recognize the target. The timing simulation diagram of transferring from RGB to HSV is as Figure 4.

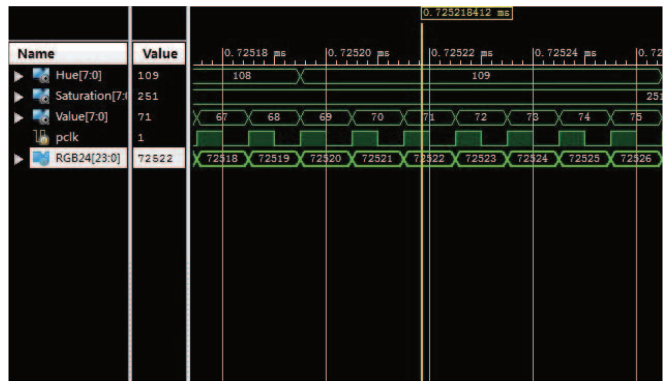

Figure 4. The timing simulation diagram of RGB to HSV

c. Color extraction and binary

Because the three components of the HSV are relatively independent, we can determine the threshold of the three components of HSV separately. By reading the values of image pixel and setting the threshold for comparison, the segmentation algorithm based on color threshold pixels will be applied. It determines whether meets the threshold requirement and the foreground area or the background area belongs to the target. If it belongs to the scope of the threshold of the target area, it will be set to 255, the remaining values are set to 0. The binary image can be obtained after the division. The Figure 5 is the binary image of the target recognition.

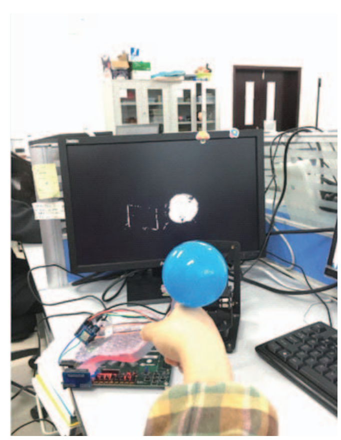

Figure 5. Binary image of the target recognition

d. Morphological filter

Morphological processing is based on the shape of a series of operations on binary images or gray morphological images. Erosion and dilation are conventional algorithms. After the image has transferred into binary image, there are still some holes and noise. It is not beneficial to the calculation of the target position subsequently.

Through erosion, the main object gets smaller and the small objects disappear, the fractured parts of the main object are removed and the object becomes smaller and the small (noisy) objects disappear. By taking the 3×3 block window for example, the equation(1) can be used. Because FPGA can parallel run, the equation (2) can used to carry on the erosion. The diagram of erosion effect is as Figure 6.

$$\begin{aligned} p = p11\,\&\,p12\,\&\,p13\,\&\,p21\,\&\,p22 \\ \&\,p23\,\&\,p31\,\&\,p32\,\&\,p33 \end{aligned} \tag{1}$$

$$\begin{cases} p1 = p11\&p12\&p13 \\ p2 = p21\&p22\&p23 \\ p3 = p31\&p32\&p33 \\ p \ = p1\&p2\&p3 \end{cases} \tag{2}$$

Where $p11$、$p12$、$p13$、$p21$、$p22$、$p23$、$p31$、$p32$、$p33$is the pixels of the 3×3block window.

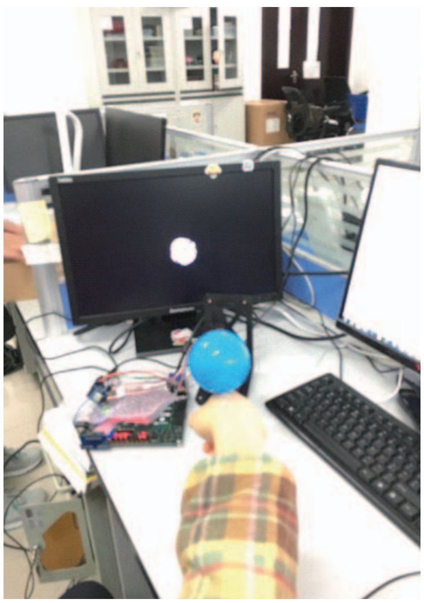

Figure 6. Diagram of the erosion effect

The term dilation refers to the fact that the object in the binary image is increased in size. In general, dilating an image results in objects in objects becoming bigger, small holes being filled, and objects being merged. By taking the 3×3 block window for example, the equation (3) can be used. Because FPGA can parallel run, the equation (4) can be used to carry on the erosion. The diagram of dilation effect is as Figure 7.

$$\begin{aligned} p = p11|p12|p13|p21|p22| \\ p23|p31|p32|p33 \end{aligned} \tag{3}$$

$$\begin{cases} p1 = p11|p12|p13 \\ p2 = p21|p22|p23 \\ p3 = p31|p32|p33 \\ p \ = p1|p2|p3 \end{cases} \tag{4}$$

where $p11$、$p12$、$p13$、$p21$、$p22$、$p23$、$p31$、$p32$、$p33$is the pixels of the 3×3block window.

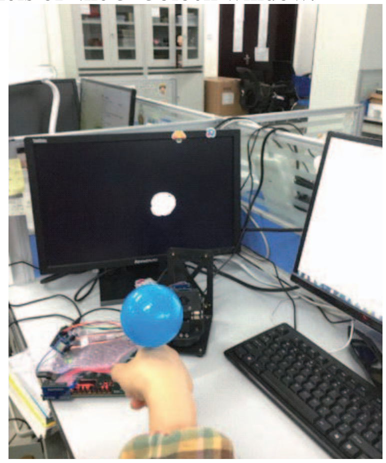

Figure 7. Diagram of the dilation effect

IV. TARGET TRACKING MODULE

Target tracking algorithm is implement based on object recognition module. Target tracking module includes the evaluation of the target centroid, reading steering angle position and PID closed-loop feedback control. After target was recognized, the binary image of the target is obtained through morphological filtering.

Since the output shaft and position feedback of potentiometer of the servo is connected to the helm. When the position feedback potentiometer follows rotating, a signal from the position feedback potentiometer will be led by reading the signal voltage, the current steering angle position will be known. Voltage potentiometer signal is analog which cannot be used directly by FPGA. The analog signal will be converted to digital signal by using the primitive of XADC. The primitive block diagram of XADC is Figure 8.The value of the voltage signal can be read. We can know the servo posture angle and track the target accurately according to feedback of the posture.

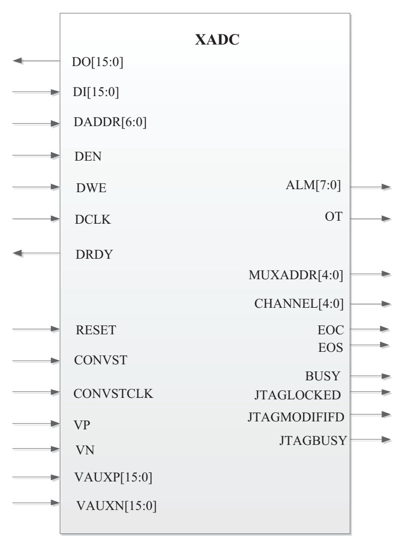

Figure 8. The primitive block diagram of XADC

V. MONOCULAR VISION LOCALIZATION MODEL MODULE

When the target has been tracked accurately, the position of the target in the Image Coordinate System will be known. Then the position of the target in space coordinate system by using localization model of monocular vision is calculated.

a. Localization model of monocular vision

When the height of the camera is fixed, the robot coordinate is taken as the actual three-dimensional coordinate system by using pinhole imaging principle. By converting between image coordinate system and the actual space coordinate system, the position of the target point can be sure. The condition is required as stated: the height of the target is lower than the height of the camera. The localization model of monocular visual spatial points is presented as Figure 9.

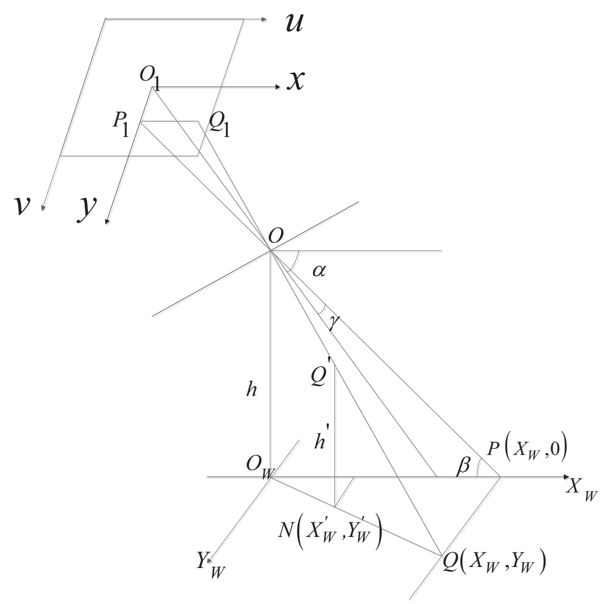

Figure 9. The localization model of monocular visual spatial points

The relationship diagram of angle is expressed as equation(5):

$$\begin{cases} \gamma = \arctan\left(\dfrac{x}{f}\right) \\ \beta = \alpha - \gamma \\ x_w = \dfrac{h}{\tan\beta} \end{cases} \tag{5}$$

Where α is the angle of pitch, *h* is the height from camera to the base, *f* is the focal length.

Obtained by the calculation to obtain the abscissa of point P is as equation(6):

$$O_wP = x_w = \frac{h}{tan\left(\alpha - arctan\dfrac{v-v_0}{f_x}\right)} \tag{6}$$

Where x_w is the abscissa of the object in the World Coordinate System, which equals to the O_WP.

Obtained by the calculation to obtain the ordinate of point P is presented as equation(7):

$$y_w = \frac{(u-u_0)\times d_y \times h}{\sin\left(\alpha - \arctan\dfrac{v-v_0}{f_x}\right)\sqrt{f^2+(v-v_0)^2\times d_x^2}} \tag{7}$$

Where u_0 and v_0 is the center coordinates of the camera in the image coordinate system, f_x is the unit length of the focal length in the *x* direction.

For a point in space coordinate plane projection in as the (x'_w, y'_w) , NQ' which is the height of the target, height must be known. Following a similar principle of similar triangles, as shown in equation(8):

$$\frac{NE}{PQ} = \frac{O_wE}{O_wP} = \frac{O_wN}{O_wQ} = \frac{h-h_1}{h} \tag{8}$$

Where the h_1 is the height from the object to the base.

So the coordinate of spatial point Q' is as shown in equation(9).

$$\begin{cases} X'_w = X_w\left(1-\dfrac{h_1}{h}\right) \\ Y'_w = Y_w\left(1-\dfrac{h_1}{h}\right) \end{cases} \tag{9}$$

b. Camera calibration and distortion correction

This experiment uses OmniVision's CMOS sensor. A checkerboard plane with the image resolution of 640 × 480 will be taken photos from different angles. Then the Matlab toolbox will be used for calibration. The diagram of parameter calibration is shown in Figure 10.

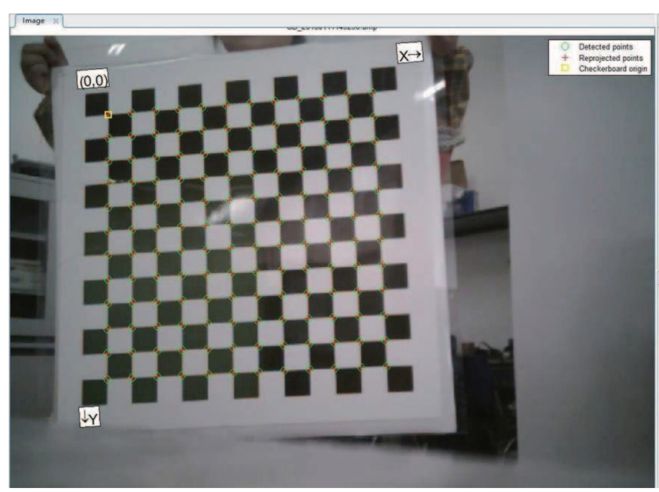

Figure 10. Diagram of parameter calibration

By using the Matlab toolbox to calculate intrinsic parameter, the intrinsic parameter matrix of the camera can be expressed in the equation(10):

$$M_3 = \begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 704.7820 & 0 & 277.0051 \\ 0 & 698.3087 & 252.8652 \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

Where (α_x, α_y) is the focal length in x and y axis, the (u_0, v_0) is principal point of the camera, s is the skew.

Because the camera has distortion, we have to rectify the parameters. The Matlab will be used to calculate the correction coefficient to correct distortion. Then the distortion coefficient matrix can be calculated as equation:

$$\begin{aligned}&[k_1, k_2, v_1, v_2] \\ &= [0.0419 \quad 0.4475 \quad -0.0041 \quad -0.0355]\end{aligned} \quad (11)$$

Where $[k_1, k_2]$ is the radial distortion and the $[v_1, v_2]$ is the tangential distortion.

c. The extrinsic model

The height h from the center of the camera to the base plane can be achieved easily and the pitch angle α can be achieved by the feedback angle of the motor. Then the extrinsic parameter of the localization model can be known.

d. CORDIC algorithm

CORDIC is mainly used in trigonometric, hyperbolic, exponential and logarithmic calculation. The algorithm uses add and shift instead of multiplication, rotation of the orientation vector calculations, trigonometric functions, multiplication, square root, inverse trigonometric, exponential or other functions. Taking the sine function for example, the simulation timing diagram of CORDIC is presented as Figure 11.

Figure 11. The simulation timing diagram of CORDIC

e. Measurement experiment

According to the positioning model based on monocular, the Modelsim simulator can be used to simulate positioning algorithm, when the height of the target to the base and the height of the camera to the base is known, the simulation timing diagram can be shown in Figure 12.

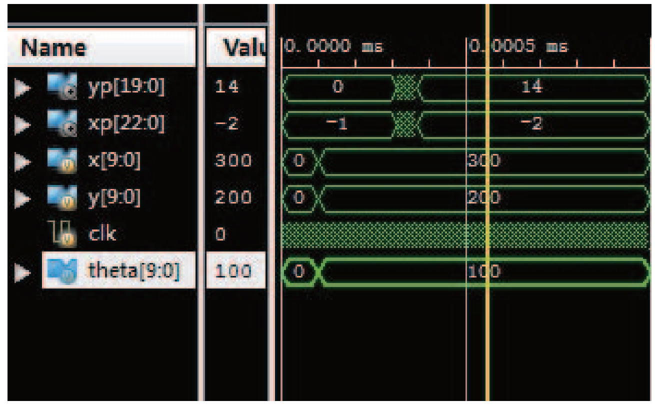

Figure 12. The simulation timing diagram of positioning algorithm

VI. CONCLUSION

This paper presents a real time target tracking and positioning system based on FPGA. The color of the target is used as the main feature to track the target and the monocular vision positioning model is established to locate the target.

For the future research, the tracking algorithm will be improved to be more stable. More accurate and robust algorithms will be developed to track some sophisticated target.

ACKNOWLEDGMENT

The subject was supported by Shenzhen Peacock Plan Team grant (KQTD20140630150243062), Shenzhen Fundamental Research grant (JCYJ20140417172417120), Shenzhen Fundamental Research grant (JCYJ20140417172417145) and Shenzhen Key Laboratory grant (ZDSYS20140508161825 065).

REFERENCES

[1] Vu A, Ramanandan A Chen, "Real-time Computer Vision Navigation System for Lane-Level Vehicle Navigation," *IEEE Intelligent Transportation Systems,* 2012, 13(2):899-913.

[2] Jani Boutellier, Ismo Lundbom, "Application-specific instruction processor for extracting local binary patterns," *IEEE Design and Architectures for Signal and Image Processing*. pp.23-25,October 2012.

[3] H. Aktas, R. Sever, "A two stage template matching algorithm and its Implementation on FPGA," *IEEE Signal Processing and Communications Applications Conference*, vol.34, pp. 2214–2217,May 2016.

[4] Q.Zhan, B.Huang, "Automatic Navigation for A Mobile Robot with Monocular Vision," *IEEE Automation and Mechatronics,* pp.1005-1010, September 2015.

[5] Y.P.Hsu, H.C.Miao, C.C.Tsai, "FPGA Implementation of Real-Time Image Tracking System," *IEEE SICE Annual Conference*, pp.18-21,August 2010.

[6] Hcon Soo Shin, Sung Min Kim, Joo Woo Kim, Ki Hwan Eom, "Novel object tracking method using the block-based motion estimation," *IEEE Transactions on International Symposium on Computer Science and its Applications*, pp.45-52,December 2007.

[7] Wu-Chih Hu, "Adaptive template block-based block matching for object tracking," *IEEE Transations on intelligent System Design and Applications*, pp.45-52,December 2007.

[8] Wei You, Hao Jiang, Ze-Nian Li, "Real-time multiple object tracking in smart environment,"*Proc. of IEEE Conf. on Robotics and Biomimetics,* pp.818-823,February 2009.

[9] S.A. Schneider, R.H. Cannon, "Object impedance control for cooperative manipulation: Theory and experimental results," *IEEE Transactions on Robotics and Automation,* 1992, 8 (3):383–394.

[10] D.Geronimo, A.M.Lopez, A.D.Sappa, T.Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance," *IEEE Transactions on Pattern Analysis and Machine Intelligende,*2010,32(7):1239-1258.

[11] Lu W, Wang T. "The method of Real-Time Distance Measurement Based on Monocular Vision," *Advanced Materials Research,* 2012,403:1451-1454.

[12] V Kruger. "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding* , 2006, 104(2-3):90-126.

[13] R.Poppe, "Vision-based human motion analysis :An overview," *Computer Vision* ,2007,108(1-2):4-18.

[14] Liu YJ, Luo X, Xuan YM, Chen WF, Fu XL, "Image Retargeting Quality Assessment Computer Graphics Forum," 2011,30(2):583-592.

[15] YoongKang Lim, Lindsay Kleeman , and Tom Drummond. "Algorithmic methodologies for fpga-based vision," *Machine Vision and Application*, 2013,24(6):1197-1211.

[16] J. Zhu, M.W. Ren, "Fast matching algorithm based on corner detection," *Journal of Nanjing University of Science and Technology,* vol. 35, no. 6, pp. 755–758, 2011.

[17] Y.J.Lee, B.D.Yim, and J.B.Song, "Mobile robot localization based on effective combination of vision and range sensors," *Int.J.Control,Autom.Syst.*,vol.7,no.1,pp.97-104,Feb.2009.

[18] Chen C.H., Cheng C.,Page D.,et al. "A moving object tracked by a mobile robot with real-time obstacles avoidance capacity," *Pattern Recognition,*2006,3:1091-1094.

[19] "Accelerating high-performance computing with FPGAs," *Altera FPGA solutions-White Paper*, pp.1-5,January 2010.

[20] M. Qasaimeh, A. Sagahyroon, "FPGA-Based Parallel Hardware Architecture for Real-Time Image Classification," *IEEE Transactions On Computational Imaging*, vol. 1, no. 1, pp, 56-70, March 2015.

[21] C.L.Sotiropoulou, L.Voudouris, "FPGA-based machine vision implementation for Lab-on-Chip flow detection," in Proc. *IEEE* Int.Symp. Circuits and System,2012, pp.2047-2050.

[22] Xu Min. "Research on Object Recognition and Location for Mobile Robot System Based on Monocular Vision," *A Dissertation Submitted for the Degree of Master,*June 2014.

[23] Tu Ya, Zhou Runjing, "Application of FPGA in direct current motor servo system," *IEEE Control Conference*. pp.16-18,July 2008.

[24] Wang Zhongcheng, Zhao Xuezeng, "Research on Space Target Orientation of NAO Based on Monocular Vision," *Machinery and Electronics.* 2015,09-0063-04.