Target Detection and Tracking System Based on FPGA

Xiuli Yu¹, Bingqi Li¹, Mingshuai Dong¹* and Weimin Yin¹

¹Automation school, Beijing University of Posts and Telecommunications, Beijing, China

E-mail: dongmingshuai@bupt.edu.cn

Abstract. The parallel processing mode of Field Programmable Gate Array (FPGA) and the ability to perform multiple tasks repeatedly meet the real-time and flexible requirements of target detection and tracking. The FPGA of Cyclone IV of ALTERA company is adopted as the hardware processor for the system platform to filter and preprocess color conversion of the collected image. The edge detection method of Sobel operator is adopted to realize the convolution and binarization processing of the third-order matrix of pixel points and the operator template, and the inter-frame difference algorithm is applied to the moving target. At the same time, the tracking algorithm based on centroid feature matching can realize the real-time tracking of target position. The system uses the characteristics of high real-time, low power consumption, parallel processing and strong expansibility of the FPGA, transplants machine vision algorithm into the FPGA, reduces the delay of image processing, and completes the hardware implementation of the target detection and tracking algorithm. Finally, the experiment verifies the real-time performance and effectiveness of the algorithm.

1. Introduction

With the development of integrated circuits, the application of integrated image processing chips, FPGA and Digital Signal Processing (DSP) processors become more and more extensive. Among them, the design cost of special integrated image processing chips is relatively high, and the flexibility, universality and portability of the chip are unsatisfactory. The calculation speed of DSP processors is very fast, which is suitable for completing the calculation of complex algorithms. But, the limitations of hardware optimization prevent the system from being widely used. On the contrary, because of its special design structure, FPGA can process data in pipelined and parallel ways⁴⁻⁵, has the characteristics of reconfiguration, which enables it to complete various situations, and has better compatibility and flexibility. Therefore, in the work of target detection and tracking which requires high computation speed and strict real-time performance, the advantage of FPGA is obvious⁶. Using FPGA processor instead of computer to process target detection and tracking can make the integrated system have the characteristics of small size, low cost, low energy consumption, fast speed, high flexibility, good real-time performance and good compatibility. Therefore, the research on image processing and target tracking based on FPGA has certain economic interests and broad application prospects.

This paper presents a target detection and tracking system based on FPGA. In this system, an image acquisition and display platform is built with FPGA as the hardware processor. The collected image data is stored in the memory. On this basis, moving objects in the scene are detected, and the features of moving objects are extracted by image segmentation. Finally, the tracking effect of the target is achieved by controlling the rotation of the platform and the movement of the car.
2. Overview of Overall System Functions
The system uses the FPGA as the processor, and designs the functional modules of the system, including image acquisition module, image preprocessing module, SDRAM data storage module, image algorithm processing module, VGA display driver module, motion control module, configuration module and mobile car based on STM32 controller, as shown in figure 1.

Figure 1 Architecture of System Platform Modules
The system platform uses the pipeline image processing architecture of parallel processing to reduce the delay of image processing. The dual-core architecture processor of FPGA + ARM is adopted. The parallel algorithm is implemented by FPGA. ARM can undertake some tasks of serial data processing and some control tasks. CMOS monocular industrial camera OV7725 is used for image acquisition. SDRAM memory for data storage and VGA display with 640*480 resolution and 60Hz refresh rate for image display. And the common interface and transmission protocol of visual system are used, such as USB, CAMERALINK, serial port, LVDS interface, I2C interface and so on. The communication between each functional module depends on the instantiation of the functional module and signal reference of the top-level file. Finally, the whole hardware platform is built.

2.1. VGA Display Driver Module
This module is used to drive VGA display. According to the protocol analysis of VGA, combined with the development mode and working principle of the FPGA, the VGA driver is decomposed into the structure shown in figure 2.

Figure 2 VGA driver structure partition
The system_ctrl_pll module belongs to the global clock management module, which is responsible for the input and output of the whole system’s global clock; the lcd_display module belongs to the image display module, which receives the data information of the pixels, outputs the displayed data, and finally completes the image production task; the lcd_driver module belongs to the VGA controller. The driver module is responsible for receiving display data, outputting RGB display data, and timing control and display control.

2.2. RAM Driver Module
SDRAM chip is selected to design the RAM drive module to store the collected image data, and the ping-pong operation of the memory is used to solve the frame interleaving problem. The principle is shown in figure 3.
In the Sdram_Control_2Port file, two FIFO memories are instantiated. Video stream data is input. Two consecutive frames of images are not written into the same BANK through the gated module, and then the image data stored in the two FIFOs are output sequentially through the output gated module, that is, \(0+1\)B is achieved through the buffer. The incomplete *ping-pong* operation of ANK and \(2+3\) Bank data. After testing, the function works well.

Figure 3 The schematic diagram of table tennis operation principle

2.3. Camera Driver Module

This module is used to drive the camera to collect image data, capture the light source data through the photosensitive sensor, and then convert the captured analog signal into digital signal through AD conversion processing for subsequent modules to calculate and store.

The board selected in this article has reserved the interface of the CMOS camera module, and the interface follows the interface definition of all official CMOS modules. The camera model selected OmniVision's OV7725 Sensor as the image acquisition source[4].

Figure 6 Structure flow chart of video acquisition and display system based on OV7725 camera

3. Hardware Implementation of Vision Algorithms

When designing a complete vision system, image processing algorithms should be divided into three categories: bottom, middle and high-level algorithms according to their different functions and tasks. The following figure is the flow chart of the system algorithm processing.

Figure 7 Flow chart of system algorithm
3.1. Image Preprocessing Algorithms

Image preprocessing algorithm is to process the image at the pixel level, and does not segment and recognize the image. Its main purpose is to process the complex and changeable moving image locally or globally, so that the processed image is suitable for the processing of higher-level algorithm. Its main processing method is convolution calculation with images and templates, so as to change the structure of the pixel data. Considering the logic realization of the algorithm, the preprocessing algorithm is very suitable for the front-end devices which can be directly connected with the image sensor, just like the FPGA we used. In the preprocessing part, the original image is transformed into RGB image, RGB image into YCbCr image, YCbCr image into gray image, median filter and Sobel edge detection are completed.

3.1.1. Convert Original Image into Gray Image. In the process of image acquisition by camera, the output data of OV7725 sensor is Raw data (Raw RGB), which is the original image file. It includes image data acquired from scanners, digital cameras and so on. After color interpolation, it becomes RGB image. Then the RGB image is transformed into YCbCr image by the conversion formula between RGB and YCbCr, as shown in (1).

\[
\begin{align*}
C_b &= 0.568(B - Y) + 128 = -0.172R - 0.339G + 0.511B + 128 \\
C_r &= 0.713(R - Y) + 128 = 0.511R - 0.428G - 0.083B + 128 \\
Y &= 0.299R + 0.587G + 0.114B
\end{align*}
\] (1)

Because VHDL language is needed for hardware transplantation, but VHDL can not perform floating-point operations, so the equivalent transformation of formula is carried out by expanding the value by 256 times and then moving the value to the right by 8 Bit bit, as shown in formula (2):

\[
\begin{align*}
Y &= ((77R + 150G + 29B) \gg 8) \\
C_b &= ((-43R - 85G + 128B) \gg 8) + 128 \\
C_r &= ((128R - 107G - 21B + 32768) \gg 8) + 128
\end{align*}
\] (2)

Then the algorithm formula hardware is transplanted to the FPGA, and the formula is combined by the adder, shift register and multiplier in the FPGA. If the above formula is used directly, when the simultaneous expression is less than 0, the operation will be wrong. Therefore, before transplanting the hardware of the algorithm, the operation will be split and transformed. First, 128 is extracted into brackets, and the formula (3):

\[
\begin{align*}
Y &= (77R + 150G + 29B) \gg 8 \\
C_b &= (-43R - 85G + 128B + 32768) \gg 8 \\
C_r &= (128R - 107G - 21B + 32768) \gg 8
\end{align*}
\] (3)

Because R, G and B are all positive numbers, the value of Y must be greater than 0, and the calculated Cb and Cr must be greater than 0, so they can be calculated directly without spillover.

After the above processing, the gray image is obtained, and the original image is compared with the gray image to obtain figure 8 and figure 9:

![Figure 8 Original image](image1)

![Figure 9 Gray level image](image2)

3.1.2. Median Filtering. Noise exists in the gray-scale image, so filtering can not only remove the solitary noise, but also preserve the edge characteristics of the image and avoid the obvious blurring of
In this paper, we use the fast sorting method as shown in figure 10 to get the median of the pixels. The steps are as follows:

The RTL principle of the above process is shown in figure 11. Input 9 pixels and output target median pixels:

Through the simulation test of Modelsim, figure 12 is obtained.

In the yellow box above, you can clearly see that the data is input into the shift register and generate a third-order matrix to verify the correctness of the method. The simulation test of median filter is also carried out by Modelsim, and the simulation figure is shown in figure 13.
The image after median filtering is shown in figure 14. The image is relatively dark because the maximum value is discarded.

3.1.3. Sobel Edge Detection Algorithms. Sobel operator is a discrete difference operator, which is used to calculate gradient approximation of image brightness function. Using Sobel operator at any pixel of the image, the corresponding gray gradient value of the point will be obtained. Sobel convolution operator contains a series of 3x3 matrices, which refer to transverse and longitudinal calculators respectively. By plane convolution of these two operators with the image to be processed, the gradient values of the pixels in the X and Y directions of the image can be obtained respectively\(^5\). Sobel edge detection algorithm can be carried out according to the following steps in figure 15:
Then, we use Modelsim to simulate and test the correctness of Sobel edge detection algorithm, as shown in figure 16.

Figure 16 Modelsim simulation of Sobel edge detection method

The logic circuit representation of the calculation formula of the Sobel edge detection algorithm above is shown in figure 17. According to the graph, the logic relationship of the hardware circuit of the algorithm in the FPGA can be clearly seen.

Figure 17 Logic circuit design of Sobel edge detection algorithm

The image processed according to Sobel edge detection algorithm is shown in figure 18.

Figure 18 Sobel Edge Detection Processed Image

3.2. Target Detection Algorithms

In this paper, we use the inter-frame difference algorithm to extract moving objects. The main idea is to analyze the images of different frames in succession in the video images with moving objects. The different parts of the image are the regions where the moving objects are located. By doing the difference operation on the images of different frames in succession, we can extract the moving objects. By establishing a mathematical model, the pixel values of two consecutive adjacent frames are calculated differentially and then binarized to extract moving objects in the image. If the video frame
image at any time \( t \) is expressed as \( f(x, y, t) \), then the video frame image at time \( t + 1 \) is expressed as \( f(x, y, T + 1) \). The calculation formula is shown in formula (4):

\[
 f_d(x, y; t, t + 1) = \begin{cases} 1, & |f(x, y, t) - f(x, y, t + 1)| > T \\ 0, & |f(x, y, t) - f(x, y, t + 1)| \leq T \end{cases}
\]  

(4)

In formula (4), \( T \) is a pre-set differential threshold. By comparing the results of the differential operation with the threshold, the binary processing is carried out. The region with 0 pixel value is identified as the background region, and the region with 1 pixel value is identified as the moving target region. That is to say, the threshold judgement is to identify the pixels with relatively large changes in the two frames as the target pixels, and the smaller changes as the background pixels without changes.

3.3. Target Tracking Algorithms

Target tracking algorithm is the process of segmenting and locating the determined moving target from the collected continuous multi-frame images, and finally obtaining the complete trajectory of the moving target or predicting the location of the target\(^{[6][7][8]}\).

Because the edge feature extraction method is used in the target detection algorithm, the tracking algorithm based on feature matching is adopted. In the previous stage, the edge information of the target extracted from the image is processed by the inter-frame difference method to track the target. That is to say, the centroid position is calculated by the edge information of the moving target\(^{[9]}\). The formula is as follows:

\[
\bar{x} = \frac{\sum x_i A_i}{\sum A_i} \quad (5)
\]

\[
\bar{y} = \frac{\sum y_i A_i}{\sum A_i} \quad (6)
\]

Among them, \( x_i \) is the abscissa of the pixels, \( y_i \) is the ordinate of the pixels, and \( A_i \) is the pixel of the pixels. The centroid coordinates of \((\bar{x}, \bar{y})\) type can be obtained by calculation. The image is processed by the inter-frame difference method and the moving object center extraction algorithm under the square area around the core (the area is increased for observation), as shown in figure 21.

3.4. Target Tracking System Experiments

The object detection and tracking system can detect moving objects and mark their centroid. After comparing with the coordinates of the image center, the direction of centroid movement can be judged and the car can track the centroid. The implementation process of the algorithm is to determine the position of the centroid coordinate of the moving object relative to the central coordinate of the VGA display screen, and control the signal 1 of left rotation or 0 of right rotation sent by the FPGA to
STM32. The signal is transmitted to the STM32 controller of the car through the serial port line for motion control.

In the motion control algorithm program of STM32, the signal value received by STM32 is judged. When it is 1, the left-wheel motor of the car is controlled to be stationary, the right-wheel rotates at a certain speed, and the rotation time is preset to achieve the effect of turning the car to the left. When it is 0, the right-wheel motor of the car is controlled to be stationary and the left-wheel rotates at a certain speed. Rotate and preset the rotation time to achieve the effect of turning the car to the right.

The flow chart of the tracking experiment is shown in figure 22.

![Flow chart of target centroid tracking experiment](image)

**Figure 22 Flow chart of target centroid tracking experiment**

Experiments show that the car can complete the expected rotation and move with the moving target. figure 23, 24 and 25 are the results of the target detection and tracking system controlling the car to track the moving target.

![Tracking Starting Point](image)

**Figure 23 Tracking Starting Point.**

![Tracking Left Rotation](image)

**Figure 24 Tracking Left Rotation.**

![Tracking Right Rotation](image)

**Figure 25 Tracking Right Rotation.**

4. Conclusion

The system platform built in this paper uses two FIFOs to design ping-pang storage, which solves the frame interleaving phenomenon when using a single SDRAM, and is verified by the color bar experiment. The formulas of the inter-frame difference algorithm are realized and transplanted by program, and the binary images of objects in static and moving states are obtained. The start-up program of the inter-frame difference algorithm is designed, and the control system switches between gray image and the image after the inter-frame difference processing. Selecting the edge characteristics of the moving object for tracking algorithm based on feature matching, that is, extracting the edge information of the moving object, calculating the centroid position of the moving object, and finally controlling the car to track the centroid position in real time. The centroid effect map of moving object is obtained by experiment, and the tracking of object by car is also realized.

The system platform uses the FPGA as the core of the hardware processor of image processing, which improves the speed and real-time of image processing, reduces the volume and energy consumption, and meets the requirements of the vision processing system towards the direction of miniaturization, integration, real-time and intelligent development.

References

[1] Wang Yandong. Design and implementation of stepper motor cooperative motion control system based on embedded real-time system [D]. Hebei University of Technology, 2016.
[2] Zhang Jin, Liu Haiyan, Wu Shihai. Application of FPGA in Moving Object Detection and Tracking [J]. Electronic Measurement Technology, 2010, 33 (12):64-67.

[3] Chen Quanjin. Research on hardware implementation of dynamic image detection and tracking algorithm based on video image [D]. University of Electronic Science and Technology of China, 2012.

[4] Zeng Yongxi, Huang Meihong, Chen Weihao. Video surveillance system based on FPGA + OV7725 [J]. Fujian Computer, 2015, 31 (09):8-9+39.

[5] Gonzalez R C, Woods R E. Digital Image Processing[M]. Addison-Wesley Longman Publishing Co., Inc., 2001.

[6] Xiao Dan. Intelligent video analysis system based on motion detection and tracking [D]. Tianjin University, 2012.

[7] Xiong Bei. Design of Bayer image color restoration based on FPGA [D]. Chengdu: Chengdu University of Technology, 2015.

[8] Li Tong. Research on multi-target tracking technology under intelligent video surveillance [D]. Hefei: University of Science and Technology of China, 2013.

[9] Jain R, Nagel H H. On the analysis of accumulative difference pictures from image sequences of real world scenes. [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1979, 1(2):206-214.