Design of image recognition system based on FPGA

Chen Chen
School of Information Engineering
WHUT
Hongshan, Wuhan, Hubei, China
e-mail:916219053@qq.com

2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP) | 978-1-6654-7857-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSP54964.2022.9778604

***Abstract*—With the development of digital image processing technology, image recognition has a wide application prospect. This technology is widely used in military, industrial, people's livelihood and other fields. For example, face recognition system often used in our life is a typical application of image recognition technology. But the hardware realization of image recognition technology is not very mature. In this paper, an image recognition system based on FPGA is proposed, which uses the advantages of high concurrency and real time of FPGA to process the image quickly and realize the real time recognition of some simple objects.**

Keywords：FPGA; Image recognition;HDMI; Median filtering

I. INTRODUCTION

Image recognition system can be realized by software or hardware, in software, image recognition has been developed more mature, especially the development of neural network, so that image recognition in software can achieve a high recognition rate. But in the hardware realization of image recognition, technology is not very mature[1]. With the development of embedded technology, image recognition and tracking system is becoming more and more low cost, low power consumption, miniaturization, and hardware has the advantage of high speed, it has become a trend to use hardware to realize image recognition. This paper proposes an image recognition system based on FPGA. The camera is used to collect the image to be recognized, and the image data is sent to the FPGA, and the FPGA is used to tell the data, and according to the characteristics of the object to match, to achieve fast recognition of some simple objects.

II. OVERALL SCHEME DESIGN

This paper uses Altera EP4CE10F17C8N FPGA chip to realize the image recognition system. The system mainly includes four modules: image acquisition module, image pretreatment module, image recognition module, image output module. The image acquisition module is realized by OV5640 camera to collect images, and the collected images are stored in SDRAM for storage[2]. Image preprocessing module for image binarization, filtering and other image preprocessing operations, convenient for subsequent processing. Image recognition module for image segmentation and recognition according to the characteristics of the object, image output module includes the image output through HDMI, and the recognition results through RS232 serial port sent to the upper computer. The overall block diagram of the system is shown in Figure 1.

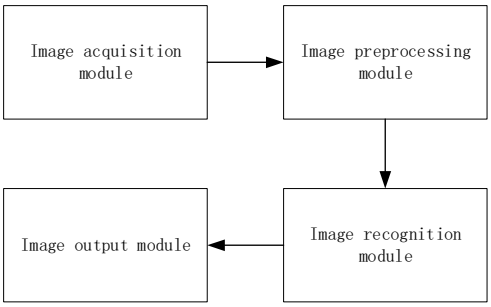

Figure 1 . Overall block diagram of the system

III. IMAGE ACQUISITION

A. Image Acquisition

This system uses OV5640 for image acquisition, OV5640 supports higher resolution, acquisition rate, and has higher image processing performance. Its image acquisition effect is better, and it is convenient for subsequent image processing. OV5640 uses IIC communication protocol to communicate with FPGA, and configures the register of the camera, which is convenient to use. In this design, we use a camera to collect images with a resolution of 640×480, which has a high recognition rate and not a very large amount of data and a faster processing speed

The clock frequency of FPGA collected by the system is 50MHZ, while the working frequency of the camera is 24MHZ. Therefore, clock frequency division must be carried out first when programming. The camera collects image data according to HREF and VSYNC signals. When HREF is high, pixel data is transmitted sequentially. At the end of each line of data transmission, HREF outputs a level jump signal to separate the current line and the next line of data. A frame of image is composed of N lines of data. When VSYNC is low power, the pixel data of each line is transmitted sequentially. After each frame of image is transmitted, VSYNC will output a level jump signal. The timing of camera acquisition is shown in Figure 2.

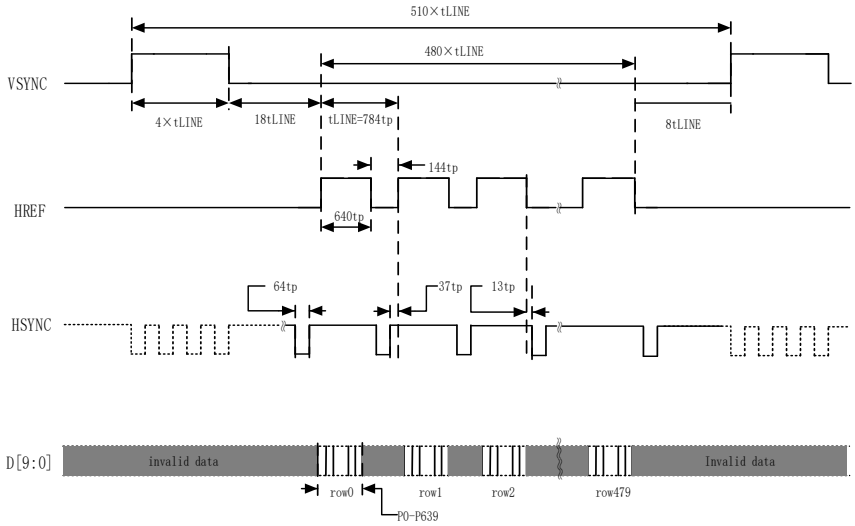

Figure 2. CMOS sequence diagram

B. Data Cache

This design uses SDRAM for data cache, SDRAM has the advantages of low price, high integration and fast reading and writing speed. Different from the usual asynchronous DRAM, SDRAM has a synchronous interface, and the clock frequency of its working clock is the same as that of the corresponding controller (CPU/ READ/write controller on FPGA), and the command sending and data transmission in SDRAM are based on this clock to realize the synchronous operation of instructions or data.

The system uses the SDRAM chip W9825G6KH-6 from WINBOND Company. The working frequency is up to 166MHZ, and the refresh frequency is up to every 8000/64ms. Its organizational structure is 32M×8bit. SDRAM chip is responsible for storing the image data collected by the camera, and output the image data for subsequent processing after caching.

IV. IMAGE PREPROCESSING

Image preprocessing includes converting RGB data collected by the camera into YCbCr format, and then performing median filtering. After filtering, image data will be binarized. After binarization, the image will be corroded and expanded to eliminate the noise at the boundary point =, so as to facilitate subsequent discrimination.

A. Change the RGB format to YCbCr format

Due to more interference in RGB space, sometimes the target image cannot be extracted, so we need to first convert RGB image into YCbCr image for processing[3]. YCbCr image format is a data format commonly used in image processing, in which Y component represents the brightness of the image, that is, the gray order value. If only Y component is displayed, RGB images can be turned into gray images. The Cb component reflects the difference between the blue part of the RGB input signal and the brightness value of the RGB signal, while the Cr component reflects the difference between the brightness value of the red part of the RGB input signal and the RGB signal. The conversion formula of RGB to YCbCr is as follows.

$$Y = 257 \times R + 0.564 \times G + 0.098 \times B + 16 \quad (1)$$

$$Cb = -0.148 \times R - 0.291 \times G + 0.439 \times B + 128 \quad (2)$$

$$Cr = 0.439 \times R - 0.368 \times G - 0.071 \times B + 128 \quad (3)$$

There are floating point numbers in the above conversion formula, but floating point numbers cannot be involved in the calculation in FPGA. Therefore, in the actual conversion, it is necessary to multiply all the coefficients on the right end of the formula by 256 and round it, and then move 8 bits to the right to complete the conversion. It is worth noting that the OV5640 camera is used here to collect RGB565 data format, while the capture formula is RGB888 format, so RGB565 needs to be converted to RGB888 in advance.

B. Median filtering

In digital image processing, there are noises in both gray image obtained directly and gray image converted from color image. Noise has great influence on image quality. Median filtering is a common filtering method, which can not only remove the isolated noise points, but also keep the edge characteristics of the image, so that the image will not produce significant blur.

Median filtering is a very common spatial filtering, which is a nonlinear smoothing technology. It takes every pixel and its neighborhood as a filtering template, calculates the median of all pixels, and uses it to replace the pixel value in the center of the template. The template of median filtering is shown in Figure 3.

11	12	13
21	P	23
31	32	33

Figure 3. Filter template

On how to obtain the median filter median, the order of the sorting method to achieve. Such as bubble sort method, quick sort method. But none of these methods are suitable for implementation in Verilog. Here we use the pipelining method to get the median value, and the processing speed is also faster. The flow diagram of sorting algorithm is shown in Figure 4

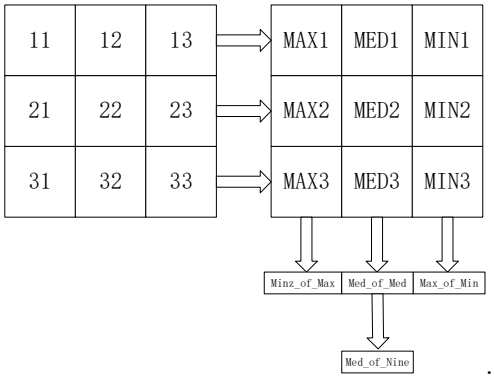

Figure 4 . Sorting algorithm flow chart

C. Image binarization

In image recognition, binarization is a commonly used method of image data processing. After binarization, the amount of data in the image is greatly reduced, and the contour of the image is clearer, and the image features are better extracted. In this design, we carry out image binarization according to the YCbCr format converted from RGB. First, the threshold range is obtained according to the color of the target object. Then, according to the threshold range, the pixel value within the threshold range is set to 1, and the pixel value outside the threshold range is set to 0, and the binarization image can be obtained. The effect is shown in Figure 5.

Figure 5. Binarization figure

V. IMAGE RECOGNITION

A. Connected domain algorithm

To recognize the object captured by the camera, the object should be separated from the background first. There are many methods to segment objects, but there are not many methods to segment objects in Verilog. At first, we used vertical projection method to segment objects, but in multi-object recognition, if the projection of two objects overlaps in the X and Y directions, there will be problems in object segmentation. Therefore, we use the connected domain method to separate objects. Here we adopt eight connected domain algorithm, that is, to a pixels to the left, upper left, upper, upper right, judge whether the pixel values agree with the pixels, if consistent criterion of the pixels and the pixels in the same tag, without pixel values agree with the pixel values, add a label on the pixel[4].

A connected domain is a pixel set composed of adjacent pixels with the same pixel value. Therefore, we can search for connected regions in the image through these two conditions. For each found connected domain, we assign a unique Label to it to distinguish it from other connected domains[5].

Through the connected domain algorithm, we can not only separate different objects, but also get the area and perimeter of the object, so as to facilitate the subsequent image recognition.

B. Matching algorithm

For image recognition, we mainly use three features of the image, that is, the color, perimeter and area of the image.

The color values were obtained earlier in the image preprocessing section. The area and perimeter of the image after binarization have also been obtained in the previous connected domain algorithm. Now we only need to use these two features for object recognition.

When using the feature of the ratio of perimeter to area, we use cosine distance algorithm for similarity detection, and compare the target object with the data stored in the data set. Cosine distance, also known as cosine similarity, is a measure of the magnitude of the difference between two individuals using the cosine of the Angle between two vectors in the vector space. The closer the cosine value is to 1, the closer the Angle is to 0, that is, the more similar the two vectors are. This is the cosine similarity. When the Angle between the two vectors is very small, it can be said that the a and b vectors have high similarity

Cosine similarity formula is derived from the law of cosines, in non-right triangles, the law of cosines is

$$\cos(\theta) = \frac{a^2 + b^2 - c^2}{2ab} \tag{4}$$

The triangle represented by vectors is shown in Figure 6.

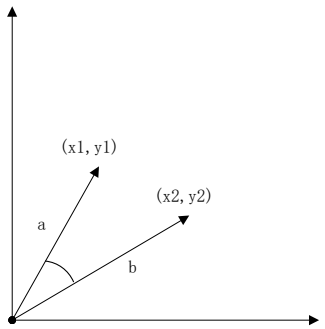

Figure 6. a and b vectors

The a and b vectors here are two-dimensional. When the vector is extended to n dimensions, the above method still holds, and the calculation formula is shown as (5)

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} \times \sqrt{\sum_{i=1}^{n}(y_i)^2}} \quad (5)$$

In this design, we adopt a 2-dimensional formula, and regard the perimeter and area as x and y respectively to form two-dimensional vectors (x, y). The most similar object is the one with the largest cosine.

We use voting mechanism in the recognition that conform to a certain characteristics of a ticket, the data set corresponds to the highest votes object that goal, first we extract the target image color threshold range, if the color of the image of an object conforms to the data set threshold, a ticket to the template to write, and then compare similarity, Vote for the data template with the highest similarity to the target object. The data template with the highest number of votes is the target object.

VI. IMAGE OUTPUT

A. HDMI System

HDMI is a new generation of multimedia interface standard, namely high-definition multimedia port[6]. It can transmit video and audio simultaneously, simplifying the interface and wiring of the device; At the same time, it provides higher data transmission bandwidth and can transmit uncompressed digital audio and high resolution video signals. HDMI interface because the interface volume is small, anti-interference ability, good compatibility and other advantages, so we use HDMI to display.

B. RS232 serial port

RS232 is a type of UART, there is no clock line, only two data lines, respectively Rx and Tx, both of which are 1bit wide. Rx is the line that receives data and TX is the line that sends data. We send the data to the upper computer through RS232 serial port. Serial port structure block diagram is shown in Figure 7.

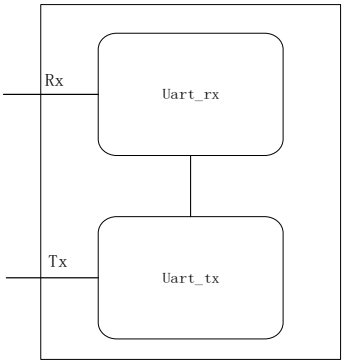

Figure 7. Serial port loopback diagram

VII. EXPERIMENT AND ANALYSIS

A. Test plan

The system uses Quartus II software of Altera to describe Verilog language and deploy the code into FPGA.

The test uses fruit identification to verify the availability of the system. Before the test, the characteristic information of different fruits is stored in FPGA, and then the fruits are placed under the camera of the system, and the serial port is opened to record the information sent by FPGA to the serial port.

B. The test results

When the apple is placed under the camera, the system successfully frames the apple and displays it on the HDMI display screen, as shown in Figure 8. At the same time, the serial port receives the recognition result sent by FPGA, including the recognized color and type. As shown in Figure 8, the system test passed.

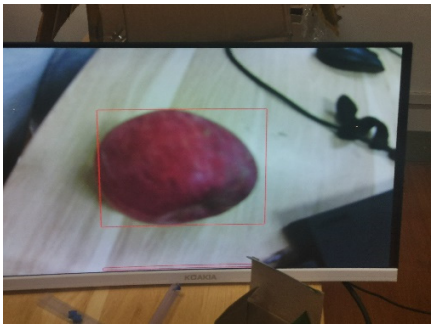

Figure 8. identify the effect drawing

VIII. CONCLUSION

This design has successfully realized the identification of the target object, and the identification information is sent to the PC end through the serial port, and some simple objects can be accurately identified. At the same time, the system's characteristic data entry is simple, and the voting mechanism is used to facilitate the subsequent function expansion and improve the accuracy, has a certain practical value.

REFERENCES

[1] Zhao Yili. Fast image blending for high-quality panoramic images on mobile phones[J]. Multimedia Tools and Applications,2020,80(1):

[2] Wan Quan and Li Shao fu. Fusion method based on FPGA image overlay and cross-screen stitching[J]. CHINESE JOURNAL OF LIQUID CRYSTALS AND DISPLAYS, 2020, 35(10) : 1066-1072.

[3] Anonymous. National Instruments; National Instruments Announces Module for Camera Link FPGA Image Processing[J]. Computer Weekly News, 2010,

[4] Zhang Ziyuan et al. Recognition Method of Digital Meter Readings in Substation Based on Connected Domain Analysis Algorithm[J]. Actuators, 2021, 10(8) : 170-170.

[5] Zhang Ziyuan et al. M. I. Faisal. Study of simply connected domain and its geometric properties[J]. Journal of Taibah University for Science, 2019, 13(1) : 993-997.

[6] Cao, Min. Face recognition robot system based on intelligent machine vision image recognition[J]. International Journal of System Assurance Engineering and Management, 2021: 1-10.