



# Design and Implementation of Stereo Vision System Based on FPGA

Qian Wang<sup>(✉)</sup>, Xin Gu, Hua Wang, and Guowei Yao

Research and Development Center,  
China Academy of Launch Vehicle Technology, Beijing, China  
aoranqian@126.com, nync396@126.com,  
e\_visawang@aliyun.com, 75312309@qq.com

**Abstract.** By comparing the different methods of vehicle ranging, the stereo vision system is selected for obstacle imaging and ranging. In response to low data transmission rate and low real-time performance and low resolution of the traditional ranging system design method, the stereo vision system is designed with integrated multi-core ARM and FPGA SOC. The system adopts image acquisition, distortion correction, target detection and tracking the image of the obstacle in front of the vehicle to compute the obstacle distance information. The delay and the power consumption of the system is measured by experiments, and it indicates that the system is suitable for vehicle-assisted driving and other image processing technologies field.

**Keywords:** Obstacle detection · Range finder · Stereo vision · Image match · System design

## 1 Introduction

With the rapid increase in car ownership, traffic safety has become an important issue for the public [1]. Therefore, the automobile active collision avoidance system based on the intelligent traffic system has become a research hotspot. The active collision avoidance system obtains the distance information of pedestrians and vehicles by sensor technology and transmits it to the drivers to realize the assisted driving of the vehicle. And the distance information also can assist the future car auto-driving collision avoidance system to make a decision.

This paper describes the design and implementation method of stereo vision system based on FPGA, the distance information can be generated by comparing binocular detection images, and the front obstacle distance image can also be output. At the same time, the paper provides the actual engineering indexes such as system delay and power consumption.

## 2 Obstacle Detection Method

At present, millimeter wave radar, laser ranging and depth vision imaging are normally used to detect obstacles. The following is a brief introduction.

## 2.1 Monocular Vision Measurement

Monocular vision measurement is a kind of passive ranging method, that needs to obtain the position information in the three-dimensional space from the two-dimensional image information. The characteristics of it is small size, low power consumption and good stability. This method usually detects the vehicle distance by mapping the width of the vehicle in the camera. If the road width has changed or no driveway, there may be a mistake or the distance can not be detected [2].

## 2.2 Millimeter Wave Radar Ranging

Millimeter-wave radar works in the millimeter-wave band. The 60 GHz, 120 GHz and 180 GHz bands are generally used on Car-borne radar. The ranging methods mainly include continuous wave radar ranging and pulsed radar ranging, which principle are the same. The distance is mainly calculated by measuring the time difference between transmission and reception [3]. The characteristic of this method is high resolution, frequency bandwidth, and can distinguish multiple small targets at the same time, with imaging capability, but in rain and fog weather, millimeter wave loss is serious, which is not conducive to imaging.

## 2.3 Laser Ranging

Laser distance measurement also belongs to radar distance measurement, which working principle is similar to the millimeter wave distance measurement but using light source to emit. Due to the narrow beam, short wavelength, and high resolution of the laser, a variety of information such as the emission characteristics, distance, and velocity of the target can be obtained [4]. However, weak global ability, low accuracy, and missed detection are weaknesses of this method.

## 2.4 Stereo Vision Ranging

Binocular stereo vision is based on the parallax principle and uses an imaging device to obtain two images of a measured object from different positions. According to calculating the positional deviation between the corresponding points of the image to obtain three-dimensional geometric information of an object. In actual application, the camera captures the image first, then uses a series of algorithms in the processor to find the same two-dimensional pixel points in two images at different viewing angles, and calculates three-dimensional information of the object according to the information such as the focal length and baseline distance of the camera [5].

This measurement method has the advantages of high efficiency, suitable precision, simple system structure, low cost. For the measurement of moving objects, the stereo vision method is a more effective measurement method because the image acquisition is completed instantaneously.

Comparing and analyzing the different ranging methods, and considering the factors such as the reliability and safety of the application in the automobile driving, this paper chooses to use FPGA to realize the Stereo Vision system.

The advantages and disadvantages of vehicle-mounted ranging methods are shown in Table 1 [5].

**Table 1.** Comparison of different ranging methods

	Stereo vision ranging	Monocular vision measurement	Laser ranging	Millimeter wave radar ranging
100 m ranging	○	×	◎	◎
View angle	◎	◎	△	△
Range accuracy	○	×	◎	◎
Depth accuracy	◎	○	△	△
Lane detection	◎	○	×	×
Rain, snow	○	○	○	◎
Fog	△	△	△	◎
Night	○	○	◎	◎
Target dependencies	○	○	△	△
Safety	◎	◎	△	○
Cost	△	◎	○	△

注: ◎Good ○Better △General ×Poor

### 3 System Design and Implementation

#### 3.1 Hardware Components

Visual system includes binocular camera and data processing board binocular camera is used for taking image, data processing board is used for processing a series of algorithms.

##### (1) Binocular camera

The system uses the D5M binocular digital camera from You-Your company to obtain left and right visual images.

The output of this camera is 500M pixels. The system design around the camera spacing can be placed as 6 cm, 10 cm, 15 cm and 20 cm.

##### (2) Visual image

Traditional imaging processing systems are mostly implemented in the form of FPGA +DSP. However, due to the complexity of interface design for data interaction between chips and insufficient data transmission bandwidth, problems such as insufficient imaging resolution and poor real-time performance are caused. In response to these problems, the FPGA of the cyclone system which is produced by Altera Corporation is

used in this system. It's a kind of medium scale logic processor and will not increase the R&D cost. It integrates multi-core Cortex-A9 processing Device, FPGA (referring to the internal logic part) communicates with ARM via AXI interface, communication clock is 400 MHz, bit width is 32 bit/64 bit/128 bit optional, FPGA can also read ARM external plug DDR directly at 500 MB rate, and FPGA integrates DDR, etc. Universal peripheral interface to meet application requirements (Fig. 1).

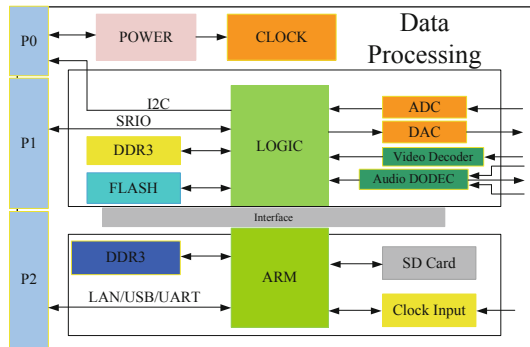


Fig. 1. The diagram of visual image processing board

The structure of the visual image processing board is shown in Fig. 2. The board is integrated into the VPX processing platform as a 3U board. Therefore, the external interface complies with the VPX standard. There are three groups of standard connectors: P0, P1, and P2. The board can be interpolated with other internal platforms. The board performs SRIO serial data transmission, which is beneficial for function expansion.

In the image processing board, DDR3 and FLASH are externally connected to the FPGA to receive frame data. At the same time, an analog interface is integrated to perform analog-to-digital conversion. The FPGA also has external audio and video interfaces and can directly process audio and video signals. The DDR3 which connects to ARM is used to storage the image and distance processing data, and the external SD CARD is used for program self-loading.

### 3.2 Image Processing Software

In the stereo vision system, the core is Cyclone V5 series SOC which includes two parts, FPGA and ARM. The FPGA in the SOC complete the image data acquisition and transmission, and performs image distortion correction, stereo visual imaging and

target detection; ARM run Linux system and completes the target tracking, transfer the image data through Gigabit Ethernet to the host computer to display. The image processing software module is divided as shown below.

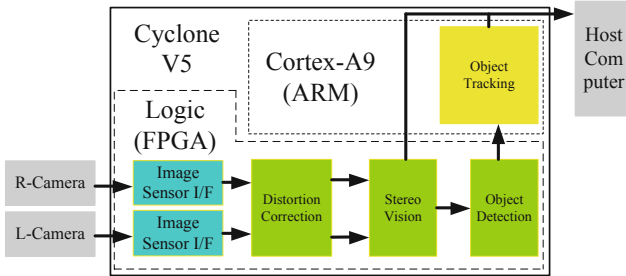


Fig. 2. The diagram of stereo vision system algorithm flow

(1) Image Distortion Correction

The fish-eye image correction algorithm is used which based on positioning retrieval. The standard coordinate system OXY is established according to the spherical coordinate model  $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$  in this fish-eye imaging principle. Assume that the fish-eye lens is located at the coordinate origin O, and its shooting direction is along the OZ axis, the fish-eye image falls on the OXY plane after shooting.

Suppose the point q on the OXY plane is the imaging of an imaging point q' in the imaging visual field region. When the distance (l) between the imaging point (q) and the coordinate origin (o) is less than or equal to the quarter of the spherical radius (r) of the spherical coordinate model in the imaging principle of the fish-eye lens, the distance between the fish-eye unfolded image point (q') and the unfolded image center position (o')

$$l' = \frac{l}{\sqrt{1 - (\frac{l}{r})^2}} \tag{1}$$

If the distance (l) between the imaging point (q) and the coordinate origin (o) is greater than or equal to one fourth of the spherical radius (r) of the spherical coordinate model in the imaging principle of the fish-eye lens, the distance between fish-eye expansion image point (q') and the expanded image Center position (o') are obtained from Expression (2).

$$l' = \frac{l - m}{\sqrt{1 - (\frac{l-m}{r})^2}} + n \tag{2}$$

$l'$  is the distance between the unfolded image point  $q'$  and the unfolded image center position  $o'$ , and  $l$  is the distance between the projected imaging point  $q$  and the projection coordinate origin  $o$ ,  $r$  is the spherical radius of the fish-eye imaging spherical coordinate model,  $m$  and  $n$  are correct-parameters. Finally, the position of the image point  $q'$  in the unfolded image is calculated by Eq. 3 and the angle  $\theta$  which is formed by the imaging point  $q$  in the OXY plane.

$$\begin{cases} q'_x = l' \times ctg\theta \\ q'_y = l' \times tg\theta \end{cases} \tag{3}$$

**(2) Stereo Vision**

In this paper, SAD (Sum of absolute difference) algorithm is used to perform block matching on the images captured by the left and right cameras ( $4 \times 4$  image blocks are used) to obtain the parallax of the two images. The SAD method is calculated as follows:

$$\min_{d=d_{min}}^{d_{max}} \sum_{i=-w}^w \sum_{j=-w}^w |I_{right}(x+i)(y+j) - I_{left}(x+i+d)(y+j)| \tag{4}$$

$D_{max}$  and  $D_{min}$  are the maximum disparity and the minimum disparity value,  $d$  is the disparity value and  $w$  is the matching template size,  $I_{right}$  and  $I_{left}$  are the gray values of the points to be matched on the right and left two images.

- target tracking

Target tracking Using 3D-coordinate data and motion-vector data, the algorithm predicts the object's location in the next following frame, and then the object data in the first frame will be linked with the actual object data in the next frame. Vehicle A and vehicle B are moving in a frame at the same time, but vehicle C is tracking each of them separately by calculating their direction and speed.

**(3) Object detection**

- Object detection

It is a moving object detection method under a typical dynamic background.

- Retrieves the current frame image  $f_{cur}(x, y)$  captured by the right camera and the previous frame image  $f_{pre}(x, y)$  by Fourier transform to obtain  $F_{cur}(u, v)$  and  $F_{pre}(u, v)$  according to FFT;
- Estimate the global motion parameter;
- Perform inverse Fourier transform to calculate rotation angle  $\theta_0$  and scaling parameter  $\sigma$ ;
- On the basis of the previous step, calculate the translation  $(x_0, y_0)$ ;

- Based on the estimated global motion variables  $\sigma$ ,  $\theta_0$ , and  $(x_0, y_0)$ ,  $f_{cur}(x, y)$  is used as a benchmark to perform background matching on  $f_{pre}(x, y)$  by bilinear interpolation.
- The difference between  $pre(x, y)$  and  $f_{cur}(x, y)$  is calculated to get  $d(x, y)$ , of two value image and the morphological image processing is used to extract the moving target
- Cycle through the steps above.

- Distance calculation

Two cameras of the same mode with a focal length of  $f$  and an optical center distance of  $b$  are used to calculate distance. The optical axes of the two cameras are parallel. Using the target detection to extract the target area, select the matching point pair  $(x_l, y_l)$ ,  $(x_r, y_r)$  with the best matching effect in the target area, and use the following formula to calculate the target distance:

$$R = \frac{bf}{(x_l - x_r)d_x} \quad (5)$$

$x_l - x_r$  is called parallax, which means the position difference of the target on the left and right images captured by the camera. This value is obtained by matching the image feature points and finding the corresponding matching points on the left and right images;  $d_x$  represents the physical length of the unit pixel in the horizontal direction.

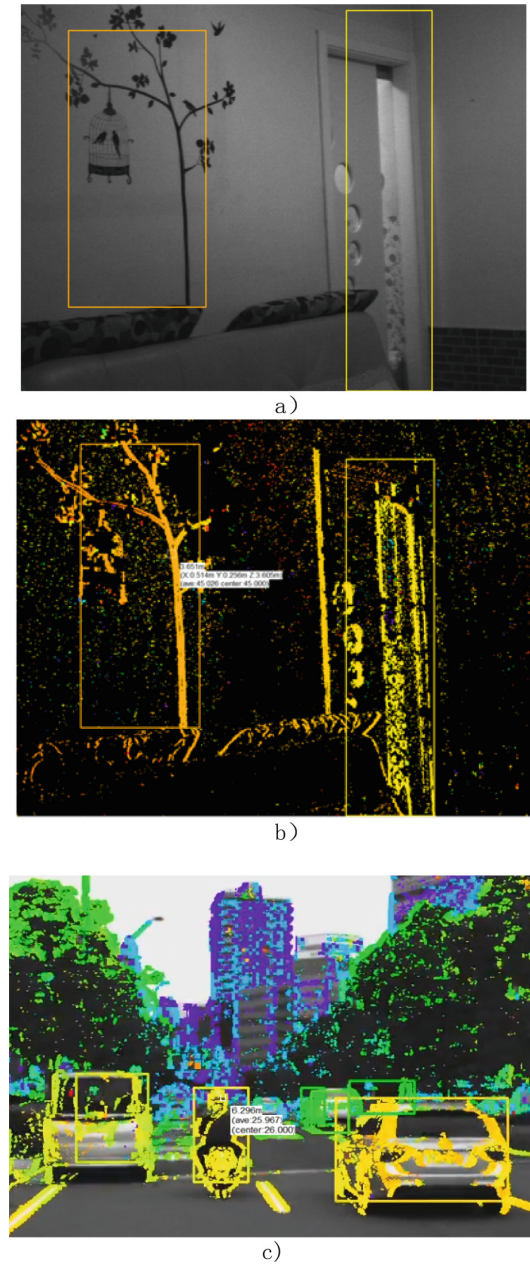
- target tracking

Target tracking Using 3D-coordinate data and motion-vector data, the algorithm predicts the object's location in the next following frame, and then the object data in the first frame will be linked with the actual object data in the next frame.

## 4 Experimental Result

The stereo vision system designed in this paper is used for target detection and tracking experiments. The camera's baseline are 10 cm and 20 cm and the camera's field of view was  $42^\circ$ , the pixel was  $1280 * 720$ , and the frame rate was 30fps.

The video output of the vision system is shown in Fig. 3a–c, where the image output from Fig. A and Fig. B is the same static scene, Fig. A is the indoor image directly output by the camera without performing the stereo imaging algorithm, and the outline of the door frame and wall decoration can be recognized; Fig. B is the implementation of the algorithm, which marks target in the field of view by color. As the distance grows, the color becomes colder, and putting the mouse on the recognized target image, the distance of the target can be displayed.



**Fig. 3.** The output figure of stereo vision system (Color figure online)

Figure c shows the moving target output image. Due to the target position is changing, optical camera image is not given. From the figures, the system can identify people, vehicles, and buildings in front of the visual field and output the distance of specific target.

Through experiments, it can be seen that the processor FPGA consumes 23K of ALM, about 32K LE resources, and RAM usage is about 4 Mbit. The entire software code runs about 50 lines of delay, about 2–3 ms under 720p frame rate. The final test results show that when the baseline is 10 cm, the maximum perceived distance is about 160 m, the error is about 2.3%; with the baseline of the camera being 20 cm, the maximum perceived distance is about 330 m, and the distance error is about 1.1%.

Then using FPGA to achieve stereo vision system design, hardware development costs are low, the system output delay is small, the system's power consumption can also be controlled at about 1.5 W, does have the conditions for large-scale engineering applications.

## 5 Conclusion

Compared with monocular imaging, stereo vision system collects a lot of target contour, behavior and distance information, and with the development cost reduction, system power consumption reduction and frame processing capability, depth vision system is more and more widely used in the car's driver assistance system. However, the application of in-depth visual technology is not limited to this technology. The technology can also be applied to taxis to report real-time road conditions for road navigation. It can even be applied to aerospace assistance for spacecraft docking and parking. With the development of stereo visual technology, the application potential of this technology in various fields such as industrial inspection, virtual reality, etc. This technology will continue to emerge and drive the climax of technological change [6].

## References

1. Lu, H.: Research on obstacle detection and obstacle avoid ancestrai. DongHua University, pp. 1–2 (2012)
2. Liu, Q., Pan, M., Li, Y.-w.: Design of vehicle monocular ranging system based on FPGA. *Chin. J. Liq. Cryst. Displays* **29**(3), 422–428 (2014)
3. Liao, S.-j., Liu, R., Cui, D.-Q.: Research on vehicle collision Prevention system based on Millimeter Wave Radar ranging, *Technology and Market* (2010)
4. Hua, Z.: Head detection and counting method based on stereo vision. Hunan University, pp. 16–17 (2015). 20150411
5. Saneyoshi, K.: Stereo vision system on automobile for collision avoidance. In: MVA 2011 IAPR Conference on Machine Vision Applications, Nara, Japan, 13–15 June 2011, p. 399 (2011)
6. Wang, X.-j., Zhou, C.-y., Zou, Z.-R.: Overview of stereoscopic vision. *Comput. Knowl. Technol.* (2005)