High performance hardware architecture for depth measurement by using binocular-camera

GAO Jin-liang(高金良), WANG Peng(王 鹏), ZHANG Zhi-jie(张志杰)
(*Key Laboratory of Instrumentation Science & Dynamic Measurement* (*North University of China*), *Ministry of Education*, *Taiyuan 030051*, *China*)

Abstract: High performance hardware architecture for depth measurement by using binocular-camera is proposed. In the system, at first, video streams of the target are captured by left and right charge-coupled device(CCD) cameras to obtain an image including the target. Then, two different images with two different view points are obtained, and they are used in calculating the position deviation of the image's pixels based on triangular measurement. Finally, the three-dimensional coordinate of the object is reconstructed. All the video data is processed by using field-programmable gate array(FPGA) in real-time. Hardware implementation speeds up the performance and reduces the power, thus, this hardware architecture can be applied in the portable environment.

Key words: field-programmable gate array(FPGA); binocular-camera; Laplacian of Gaussian filtering; depth measurement

CLD number: TP216 **Document code**: A
Article ID: 1674-8042(2012)03-0211-04 **doi**: 10.3969/j.issn.1674-8042.2012.03.002

0 Introduction

As an important research branch in robot vision, the technology of depth measurement has been attractive[1]. It uses sensors to percept the information of scene depth, which can provide effective information for establishing scenes of the robot. Current general purpose microprocessors are too slow to perform stereo vision at video rate[2]. Generally speaking, it takes several seconds to execute a medium-sized stereo vision algorithm for a single pair of images on a 1 GHz general-purpose microprocessor. Designers have built custom-designed hardware systems to accelerate the performance of the vision systems to overcome this limitation. Hardware implementation allows one to exploit the parallelism that usually exists in image processing and vision algorithms[3]. Compared to other current technologies of depth measurement, the technology of depth measurement by using binocular-camera has higher measuring accuracy and better performance in real-time.

The system uses field-programmable gate array (FPGA) as the control chip and selects TVP5150 as the decode chip. This video-rate stereo machine is a special-purpose hardware that has been built with off-the-shelf components. All the algorithms of the system is designed and built in the FPGA according to the current performance. This system handles the distance range of 2 − 15 m by using 8-mm lenses. The stereo machine output a pair of intensity and depth images at 30 times per second.

1 Principle of depth measurement

1.1 Binocular vision ranging theory

In parallel optical-axis system, the binocular-camera changes the method for depth measurement of the mobile object in a 3-D environment into the method for two-dimension parallax image plane. Through it, the mapping relationship between the dots in the three-dimensional scene and dots in the two-dimensional scene can be established. The ideal projecting model is the traditional pinhole camera optical model.

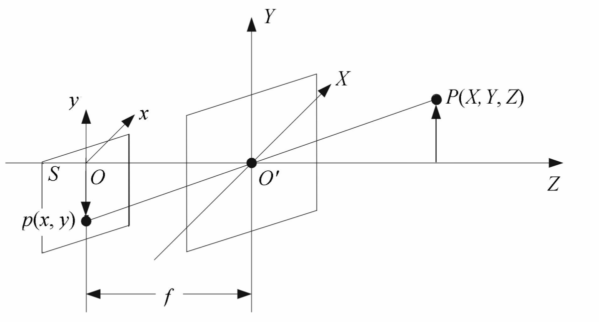

Fig.1 Model of pinhole imaging

※ **Received data**: 2012-02-18
Corresponding author: GAO Jin-liang(gaojie8563@163.com)

As shown in Fig. 1, S is a two-dimension image plane, O is the position of the pinhole. The point $p(x, y)$ on S plane is the projective point of another point $P(X, Y, Z)$ in 3-D. f is the focal length of the optical system. Z is the orient of the optical-axis.

In the traditional pinhole camera model, when the object distance is far greater than image distance, as shown in Fig. 2, the point O' is the viewpoint, f is the distance from the view plane to the point O.

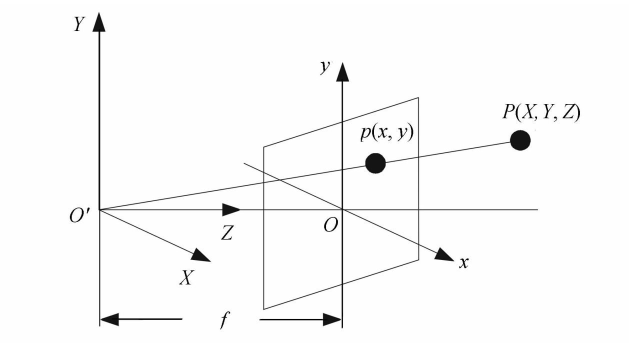

Fig. 2 Mode of human sight

Thus, according to these models, a conclusion can be drawn about relationship between the spatial point P and the planar point p. It can be described by Eq. (1)

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{1}$$

Binocular stereo vision obtains 3-D information by using trigonometry principle, which is based on principle of parallax. The image plane is established by the two cameras and the object is to be tested to form a triangle[4]. In the public view field, the values of objects in 3-D and 3-D coordinates of feature points can be obtained after the positional relation between the two cameras is known.

As shown in Fig. 3, it is the principle of binocular stereo vision, and b is the distance between projection centers of the two cameras. The two cameras photograph the same feature points of the object in 3-D at the same time. The image at point P can be obtained from the left camera and the right camera. It is assumed that the two cameras are installed in a plane. The relationship between coordinates Y is made in Eq. (2)

$$y_{\text{left}} = y_{\text{right}} = y. \tag{2}$$

In Fig. 3, the point P is the target, O_l and O_r are the optical centers of the two cameras which have the same focal length. p_l and p_r are projective points of the left and right cameras, d is the distance from the point P to the camera. According to Eq. (3)

$$d = \frac{bf}{m+n}, \tag{3}$$

the value of m plus n is the parallax of the left and right cameras, and thus, the problem of depth of 3-D is transferred into the problem of searching for the matching point.

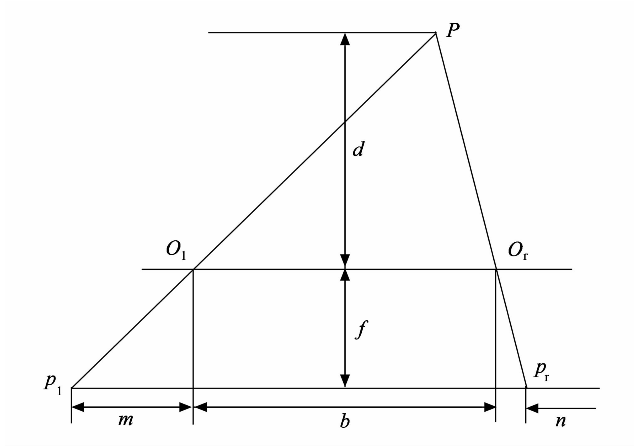

Fig. 3 Principle of distance measurement based on binocular camera

1.2 Simple space object matching

In the image matching process, taking a marker of space objects in the two images on the (CCD) image, the image point will be marked on the corresponding match points each other. The method is proposed to install a laser source in the cameras' system, and the laser marker is used to quickly resolve the target object at the same point among the two corresponding points about the CCD image matching problem, and thus the target object depth information can be easily calculated.

2 Design of stereo vision system

When a complex algorithm based on a reprogrammable hardware is implemented, the most important issue is that there is a fixed amount of hardware available. These hardware resources include on-FPGA and off-FPGA available memory, logic capacity, chip-to-chip communication bandwidth and memory access bandwidth. It is a systematical flow path as shown in Fig. 4.

The video interface unit which consists of TVP5150 chip receives the video signal from the cameras in the composite phase alternation line (PAL) format with an image size of 576×720 pixels[5]. However, the stereo vision system itself is capable of processing 30 frame per second if there is a way to simultaneously receive the two video-rate image streams from the video interface. The video interface unit sends two non-interlaced gray scale image streams to the data processing unit after frame buffering[6].

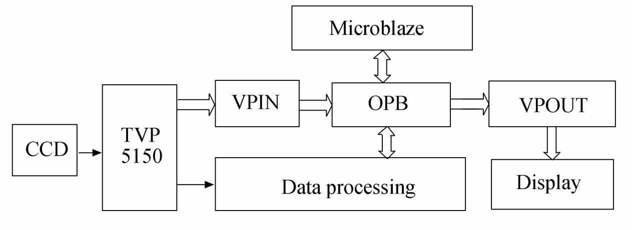

Fig. 4 Schematic diagram of the system

2.1 Laser marker acquisition and recognition

Data processing is to extract real-time video data processed to obtain the laser marker spot center position. The first is the Laplacian of Gaussian (LOG) filter modules.

The LOG filter unit can enhance the image features, and remove the effect of intensity variations among images due to the difference of camera gains, ambient light, etc. The LOG subsystem contains six processing channels, each of which can perform the LOG filtering on an image in real-time. The input image for each channel is read from the frame grabber and the output image is sent to the next subsystems.

Four 7×7 convolvers are used for each channel. By loading arbitrary 7×7 coefficients, a large class of filtering operations can be carried out. For instance, a LOG filtering is achieved by loading a Gaussian mask into the first three convolvers and a Laplacian filter into the final one. The maximum size of the LOG filter becomes 25×25 pixels by this cascading technique. The LOG subsystem also has a multi-resolution capability which produces an image pyramid by repeatedly shrinking the images[7], then selects the image frame deposit center as the image origin, separately calculates it, and realizes laser marker spot center positioning.

2.2 Object depth information calculation

The "left eye" and "right-eye" at the point p of x and y coordinate are shown as

$$\begin{cases} p_{\text{left}} = (x_{\text{left}}, y_{\text{left}}), \\ p_{\text{right}} = (x_{\text{right}}, y_{\text{right}}). \end{cases} \tag{4}$$

In Eq. (4), when calibration of the two camera images is in the same plane, the characteristic point p of the y-coordinate of the image coordinates are the same as

$$y_{\text{left}} = y_{\text{right}}. \tag{5}$$

So, the depth information d is calculated by CCD calibration of the parameters as

$$d = \frac{f \times b}{(x_{\text{left}} - x_{\text{right}}) \times \partial x}, \tag{6}$$

where ∂x is the component of CCD.

2.3 Processing result display

For the coordinates receiving laser marker spot to the center or the parameters of depth information being calculated can display on the digital tube, the digital tube real-time display of the binary data must be converted into decimal number before, and then sent to the digital pipe display. 8-bit seven sections digital tube can be used as 2-D coordinates of the display elements. Here is the coordinates of the measurement of the laser marker spot in a pixel surface. As shown in Fig. 5, the left three digital tubes show the three decimal X coordinates, and the right three ones show the three decimal Y coordinates.

Fig. 5 Laser point coordinates

3 Experimental results

The proposed real-time stereo vision system is designed and coded using VHDL and implemented using a Virtex-4 FPGA from Xilinx.

First, a camera calibration model is established by utilizing basic geometry model of stereovision. Based on the camera calibration model, the calibration experiment is designed, which can work out the unknown camera parameters through the target coordinates, target image and the calibration model set up above. In the experiment, for the sake of testing the precision of the system, laser is used as the target to calculate the distance from the laser point to the camera, which can be compared with the actual distance.

Table 1 Experimental data

Actual distance/mm	Calculative distance/mm	Range error/%
2 250	2 284	1.51
5 730	5 865	2.36
9 080	9 208	1.40

In Table 1, the depth measurements obtained by the system are compared with the actual distance.

4 Conclusion

This paper describes high performance hardware architecture for the depth measurement by using binocular-camera on FPGA. The system uses laser positioning technology, CCD cameras and TVP5150 decoding chips as key components, togethering with FPGA, to compose visual sensor. The CCD camera acquires the laser marker spot which irradiates the object surface to be tested, TVP5150 receives the signals of the video camera and turns them into digital signals and send them into the FPGA, and the FPGA finishes image processing to determine the coordinates of the spot pixel, When the CCD parameters calculate the depth information, the CCD parameters must be calibrated. The flare coordinate information or depth information in the digital tube can be displayed. The whole system of data processing parts of the structure rely on the FPGA's powerful logic gate resources integrated in chips internally, and power consumption and stability of the system have been greatly improved.

References

[1] Azevedo C, Poignet P, Espiau B. Moving horizon control for biped robots without reference trajectory. Proc. of IEEE International Conference on Robotics and Automation, IEEE Computer Society, Washington DC,2002, 3: 2762-2767.

[2] Burt P J, Adelson E H. The Laplacian pyramid as a compact image code. IEEE Trans. on Communication, 1983, 31(4): 532-540.

[3] HU Tao, HUANG May. A New stereo matching algorithm for binocular vision. International Journal of Hybrid Information Technology, 2010, 3(1): 1-8.

[4] Kanade T, Narayanan P J, Rande P W. Virtualized reality: concepts and early results. Proc. of IEEE Workshop on the Representation on Visual Scene, Boston, 1995: 69-76.

[5] Baker H H, Binford T O. Depth from edges and intensity based stereo. Proc. of the 7th International Joint Conference on Artificial Intellegence. Vancouver, 1981.

[6] Diaz J, Ros E, Pelayo F, et al. FPGA-based real-time optical-flow system. IEEE Trans. on Circuits System for Video Technology, 2006, 16(2): 274-279.

[7] Carneiro G, Jepson A D. Local phase-based features. Proc. of the 2002 European Conference on Computer Vision, Copenhagen, Denmark, 2002, 2350: 282-296.