

Embedded 3D Vision Measurement System Based on the Line Structured Light *

Zhao Wang, Xuewei Cao, Haitao Song, Han Xiao, Wenhao He, and Kui Yuan

Institution of Automation

Chinese Academy of Sciences

95 Zhongguancun East Road, Beijing, China

zhao.wang@ia.ac.cn

Abstract - In this paper a real-time embedded 3D vision measurement system (a vision sensor) is realized, based on the line structured light, mainly for applications like large object reconstruction or robot obstacle avoidance. The system mainly contains a sensor and an image acquisition and processing board developed by our research team. An FPGA chip and a DSP chip are embedded in the card as the major calculation units, which make real-time computation possible. The sensor consists of a high-speed camera, a laser projector and an optical filter. The whole recognition algorithm is divided into three parts: the camera calibration, the light plane calibration and the final 3D information extraction. The first two parts are achieved with Matlab before the first use of the sensor. The last part is the process of extracting 3D information for objects. In order to achieve real-time extraction, FPGA and DSP are used to realize certain algorithms separately. Experiments show the time of processing one image using the system proposed in this paper is 9.2ms, and the reconstruction error is less than 2mm when the distance between the sensor and the object is about one meter, which is enough for large object reconstruction and robot obstacle avoidance.

Index Terms - 3D vision measurement, embedded system, line structured light, vision sensor.

I. INTRODUCTION

Vision sensors are currently the main way for service robot achieving environmental perception. In recent years, with the continuous development of sensor, a variety of depth sensors appear. It can display the distance to the front object, so geometrical shape of the object can be obtained, which provides a new way to 3D object recognition and 3D environmental perception. Among them, 3D laser range finder, depth camera and 3D vision measurement are the most typical three kinds of depth information sensors. But the volume and power consumption of common laser range finder are very large, and ordinary depth camera has rather low resolution and high noise. In addition, laser range finder and depth camera with high precision are very expensive. Therefore, more and more studies focus on the 3D vision measurement technology. Among them, optical 3D measurement technology with its large range, non-contact, speed, moderate accuracy and other advantages has become the most promising 3D data acquisition method in engineering application [1].

In many of the optical 3D vision measurement methods, structured light 3D vision measurement method with its large range, large field of view, high precision, easy to extract the light bars, strong real-time, active control and other advantages has won widespread use in industrial application in recent years [2]. Depending on the light pattern, structured light 3D vision measurement method can be divided into point structured light method, line structured light method and surface structured light method [3]. Just as their name implies, the shape of laser projecting to the object surface is a point, line or surface. Point method is relatively simple from the light source manufacturing technology to the image processing algorithms, but once only a point is got from one image, so the measurement efficiency is rather low, and 3D scanning device is a must when the 3D data of an object surface is needed. On the contrary, the surface structured light method can get a surface without scanning, but it has matching problem, and this kind of laser is very expensive and the image processing algorithms are very complicated. Balancing these two methods, line structured light method attracted more and more attention. Compared with surface method, its image processing algorithm is much easier and matching problem does not exist. At the same time, the scanning time is much shorter than point method. So this method is chosen to realize 3D vision measurement in the paper.

There have been many studies [4, 9] focused on this field. However, they paid more attention to improve the reconstruction precision of small object. In this paper, we focus on the fast vision measurement for applications like large object reconstruction or robot obstacle avoidance, where need real-time measurement. Normally, general-purpose computer system is used to process the vision information. But the computation ability of the CPU is not enough to handle the algorithms of high complexity, and the low processing speed and large power consumption of general-purpose computer make it difficult to meet the requirements of industrial applications. As a customizable logic circuit, Field Programmable Gate Array (FPGA) can accelerate the algorithm by the parallel computation and pipelined structure. But for some complicated algorithms, the flexibility of FPGA is insufficient. As a microprocessor with a kind of special structure, Digital Signal Processor (DSP) can make up for the deficiency of FPGA by its flexible addressing mode. Besides, DSP also has powerful calculation ability. Therefore, using an FPGA chip and a DSP chip as the main computing elements,

* This work is supported by the National Natural Science Foundation of China Grant No. 61273360, and the National High Technology Research and Development Program of China Grant No. 2013AA040903.

our research team developed an intelligent image acquisition and processing board to realize the real-time computation for complex algorithms.

This paper is organized as follows. Section II and Section III briefly discuss the whole hardware system and the fundamental of vision measurement based on the line structured light separately. The implementation scheme of the algorithm and detailed process of each step achieved in the intelligent image acquisition and processing board will be described in Section IV. Section V gives some experimental results of important steps. Section VI concludes the paper.

II. THE EMBEDDED 3D VISION MEASUREMENT SYSTEM

The vision measurement system (see Fig. 1) contains a sensor, an image acquisition and processing board (see Fig. 2), a scanning device and an upper computer. The sensor is constituted of a line laser projector and a high-speed camera, and it is installed on the scanning device to move continuously with the movement of an arm of the device, whose movement is controlled by a stepping motor. The image acquisition and processing board is the computing core. The image with line structured light is taken by the camera and is transferred into the intelligent image card, where operations like filtering, light knife center extraction and 3D information extraction are performed. Finally, the results are transferred to the upper computer by internet. At the same time, the upper computer controls the rotation of the stepper motor to control the movement of the scanning device.

The intelligent image acquisition and processing board (see the upper left of Fig. 1) is the core of the whole system, which is designed by WenHao He. The FPGA used on the card is Altera Cyclone III EP3C40F484, and the DSP on the card is TI's TMS320DM642. In addition, there are two 512K×16bits SRAM, two 4M×32bits SDRAM, one 32M×16bits DDR and a 4M×8bits Flash on the card.

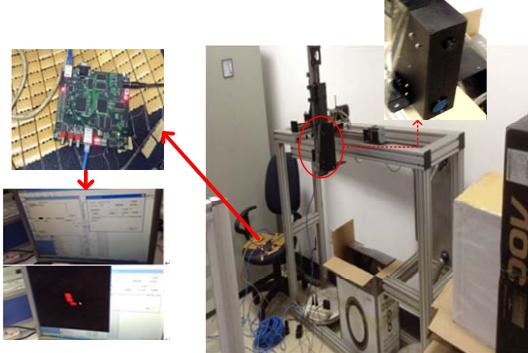


Fig. 1 A whole vision measurement system

The card also contains two ports: Gigabit Ethernet port and Fast Ethernet port. The former port is connected with the high-speed digital camera to achieve faster image acquisition, and the latter is connected with upper computer to output results. In addition, there are two analog video input ports and one analog video output port to input or output analog image data from ordinary CCD camera. On this board, FPGA is connected with DSP and two SRAMs, which makes DSP

convenient to access the memory resources in the FPGA and the SRAMs. What's more, DSP is also connected with a SDRAM and a FLASH, which makes the board have enough memory space. And FLASH can keep important data after power down of the board.

III. FUNDAMENTAL OF VISION MEASUREMENT BASED ON LINE STRUCTURED LIGHT

After introducing the hardware system, this part will briefly describe the fundamental of line structured light vision measurement. The basic idea of this algorithm (see Fig. 2) is that a light stripe is modulated by the depth or gap of the measured surface, which will lead to distortion or discontinuity of light stripe taken in a 2D image. And the degree of distortion is proportional to the depth, and discontinuity proves the existence of the physical gap in the surface of the object. All the points of the light stripe in the 3D space contain rich 3D information, which can be described by their 3D world coordinates (X_w, Y_w, Z_w) . And when the camera captures the image of the object with the deformed light stripe, the points in the 2D deformed light stripe contain their 2D image point coordinates (u, v) in 2D image plane coordinate system. The task of structured light vision measurement is to acquire the 3D information of the measured surface from the 2D deformed light stripe image [4]. This is to say, finding the correspondence between the two coordinates $((X_w, Y_w, Z_w) = f(u, v))$ is the key.

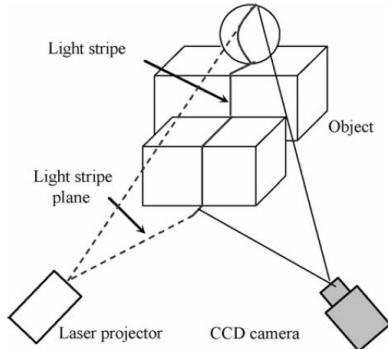


Fig. 2 Fundamental of vision measurement based on line structured light,
From Ref. [4]

In order to find the correspondence, a model based on the perspective projection transform is proposed, which involves the transformation between four coordinate systems: 3D world coordinate system, 3D camera coordinate system, 2D image coordinate system, and 2D image plane coordinate system. The purpose is finding the transformation from the image plane coordinate system (u, v) to world coordinate system (X_w, Y_w, Z_w) . Here we directly give the result. If you want to know more details, please refer to [4].

$$\begin{aligned}
Zc \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} 1/dx & \lambda & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (1) \\
&= \begin{bmatrix} fu & \lambda & u_0 \\ 0 & fv & v_0 \\ 0 & 0 & 1 \end{bmatrix} \times [R \ t] \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = K \times [R \ t] \times \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\
Z_w &= aX_w + bY_w + c \quad (2)
\end{aligned}$$

Eq. (1) describes the perspective projection model of camera and (2) describes the equation of the light stripe plane, both of which constitute a complete ideal mathematical model (lens distortion equations are omitted here) of the structured light stripe vision sensor. Where $fu = f/dx$, $fv = f/dy$, $\lambda = \lambda f$. K is camera intrinsic parameters matrix. $[R \ t]$ is camera extrinsic parameters matrix. a, b, c are the laser plane parameters. Once the model parameters are known, and the image plane coordinates (u, v) of any point on the light stripe taken by camera are read from image, then the world coordinates (X_w, Y_w, Z_w) of the point can be solved by (1) and (2). So the process of line structured light vision measurement proposed in the paper can be divided into two large parts. The first part is to estimate these parameters, which is called parameter calibration. The second part is to extract 3D information for objects. The following will detail these two large parts.

V. IMPLEMENTATION PROCEDURE OF LINE STRUCTURED LIGHT VISION MEASUREMENT SENSOR

As mentioned above, the implementation procedure consists of two large parts: parameter calibration and 3D information extraction. The first part needs to be completed before the first use of the sensor. And as long as the relative position of camera and the laser projector is fixed, the parameter of sensor will not change. So calibration process does not require to be repeated unless the application requirements are changed. However, this part is very important that can directly affect the precision of vision measurement. Therefore, this process mainly focuses on the improvement of accuracy but the speed, which is suitable to be implemented with software. On the contrary, the second part is actually a process of using the sensor to extract depth information, and this part pays more attention to the speed of extraction, especially for applications like large object reconstruction or robot navigation, so it will be realized with the embedded system in this paper.

Fig. 3 shows the sensor designed by this paper. The distance between the camera and projector is about 120mm. Although according to [3], the larger the distance is, the smaller measurement error of each direction is. But the applications here ask for speed and small volume more than precision, so the distance here is set much smaller than 500mm in [4]. But if you want higher precision, please firstly consider increasing the distance.



Fig. 3 The sensor designed by us

A. Parameter calibration

According to (1) and (2), parameter calibration part can be divided into two separate steps: camera calibration and projector calibration. Camera calibration, i.e. estimating the intrinsic parameters and extrinsic parameters of the camera model, detailed stages of which have been described in [7] based on a camera calibration toolbox, which is fast, easy and accurate. So here we directly use it. The projector calibration, i.e. the calibration of the light stripe plane emitted from the projector. Although only three non-collinear control points on the light stripe plane are enough to calibrate a plane, often more points are necessary to improve the calibration accuracy [4]. There are many methods proposed to gain these points during this process [4, 8]. Among of these, the method proposed by Zhou et al in [4] is based on a freely moveable planar calibration target, and can obtain a large number of calibration points (located in the light line used to calibrate light plane) and high calibration precision without costly auxiliary equipment. So here light plane calibration is implemented based on this method. But due to the difference between application requirements and implementation platform, some parts of this method are changed. So some stages are needed to be discussed. Fig. 4 is the planar calibration target used in this paper, which contains 10×7 squares except the outermost squares, and the size of every square is $30 \times 30\text{mm}^2$.



Fig. 4 The planar calibration target

The basic stages of this method after improvement are as followed.

1) The sensor is fixed and the light is projected to the surface of front object.

2) Extract the center of the light knife and fit the light line under the image plane coordination system, as the green line shown in the Fig.5.

3) Fit the horizontal and vertical straight lines of the planar calibration target under the image plane coordination system, as the red lines shown in the Fig.5.

4) Compute the intersection points (called calibration points, see the black points in Fig. 5) of the light lines and the straight lines fitted in (3).

5) Compute the local world coordinates of these calibration points by cross-ratio described in [4].

6) Compute the camera coordinates of calibration points by the transformation from the camera to the local world coordination system, which is obtained by the calibration toolbox described in [7].

7) Fit the equation of the light plane under the camera coordination system. The result is shown in the right image of Fig.8).

8) Compute the global world coordinates of these calibration points by the transformation from the camera to the global world coordination system. The process of getting the transformation will be introduced later.

9) Continuously move the sensor with the scanning device, and repeat the steps from (2) to (8) for every image.

10) Combine all the global world coordinates obtained from above to reconstruct the surface of the object and to get the direction and distance from the object to the sensor.

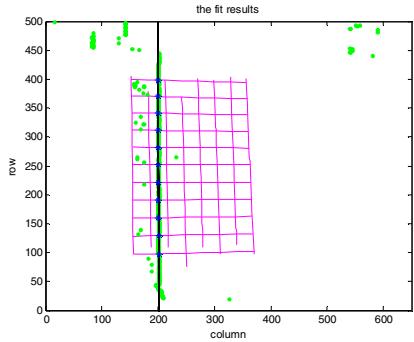


Fig. 5 The image of fitting lines and computing intersection points

Among of above steps, there are some keys which are different with [4] or even not mentioned in [4]. So here it is necessary to explain them in detail.

1) Make sure the distance between the sensor and the measured object. Depth information detection for service robot required that the distance between the object and the sensor cannot be too small, which is different with the applications like in [4]. What's more, the closer between the sensor and the object is, the higher the accuracy is (Ref. [3]). So the distance for 3D vision measurement is generally very close, like less than 20cm. However, the distance for depth information detection is based on the application environment or service robot arm length, which makes the distance unchangeable. In this paper, the distance is set to one meter, limited to the robot arm length.

2) Locate a global world coordinate system and compute the transformation between the camera coordinate system and the global world coordinate system. The global world coordinate system is set in Ref [4] to coincide with a local world coordinate system (located in the left upper of planar calibration target). The position of global world coordinate system is changeable with the change of the target position.

That's because that the location has little effect on the result of vision measurement in that paper. However, in this paper, the final results of detecting 3D information for robot navigation need can describe the direction and distance between the sensor and the front object. And the final information is obtained under the global world coordinate system. Therefore, the system in this paper cannot be located at any positions. It needs to be set close to the camera. Here the origin of this coordinate system is the origin of the camera coordinate system, and three positive axes point to right, front and top separately, which is matched with the direction of arms of scanning device (see Fig.1).

The transformation from initial camera coordinate system to global world coordinate system can be obtained by the device shown in Fig.1. Firstly, fix the sensor an initial position on the device and set the planar target in a fixed position. Then compute the translation matrix ($t_0 = (x_0, y_0, z_0)^T$) of camera-local world coordinate system by toolbox. Thus we get the initial camera coordinates (x_0, y_0, z_0) of zero point (called P) in the local world coordinate system. Then move the sensor 50mm respectively along the x-axis, y-axis and z-axis of global world coordinate system, and respectively compute translation matrix to get the coordinates $((x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3))$ of point P after moved. It should be noted that every time the sensor is moved from its initial position. Suppose R is the rotation matrix of initial camera-global world coordinate system, then the equation (3) can be inferred.

Since the origins of coordinates are coincident of these two coordinate systems, so the translation matrix t of initial camera-global world coordinate system is $(0,0,0)^T$. When the sensor moves with the scanning device, the translation matrix t is changeable. Since the time of scanning 100mm is 1s, and the speed of extracting 3D information is 100 frames per second in this paper, thus, the distance of the device scanning during the time of processing one image is 1mm, and the translation distance of the sensor in the i-th second is i mm. For simplification, here we suppose the movement of the sensor is uniform and the direction is x-axis positive direction. So the translation matrix of i-th image is equation (4).

$$R \times \begin{bmatrix} x_0 - x_1 & x_0 - x_2 & x_0 - x_3 \\ y_0 - y_1 & y_0 - y_2 & y_0 - y_3 \\ z_0 - z_1 & z_0 - z_2 & z_0 - z_3 \end{bmatrix} = \begin{bmatrix} 50 & 0 & 0 \\ 0 & 50 & 0 \\ 0 & 0 & 50 \end{bmatrix} \quad (3)$$

$$t_i = (i, 0, 0)^T \quad (4)$$

3) Extract the light knife center based on the centroid method and the extreme method. The common methods of extracting knife center consist of threshold value method, edge method, midline method, extreme method and gray centroid method [1, 5]. The basic principle of the gray centroid method is that the centroid of gray distribution of cross section of light is as the center of light knife. This method is fast and has high accuracy. In addition, the extreme value method takes the position with gray maxima as the center of light knife. This

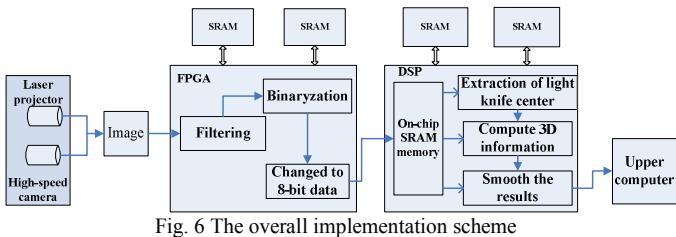
method is very simple and fast, but it only reached the pixel level accuracy and is easily subject to severe noise.

Since the low cost laser projector used in this paper leads to wide light bar in the image and edge flaws, and taking into account the simple implementation of hardware, a method combining these two methods is proposed here. Its detailed stages are as followed. Firstly, threshold the image by maxima of every row, i.e. set the maxima of grayscale of every row is 1, else is 0. Secondly, search the center of the longest all-1 section of every row as the centroid position. This method not only improves the preision to the sub pixel level and binaryzation makes the speed of reading image from FPGA much faster. What's more, this method generates less noise than only using one method.

4) RANSAC (Random Sample Consensus, Ref. [6]) is used to fit the light straight line to exclude outliers (see Fig. 5), which is generated by other bright points in the environment or reflection. From the figure 5, we can see that this method can effectively fit straight lines.

B. Extraction of depth information in hardware

As mentioned above, this part mainly focuses on the speed, and it is realized in the intelligent image acquisition and processing board, with the cooperation of DSP and FPGA. The implementation scheme is as Fig. 6.



The procedure of implementation with hardware can be described as follows.

1) An image with laser light stripe taken by high-speed camera is transferred into the FPGA by the Gigabit Ethernet port.

2) After average filtering and threshold processing for the image by the maximum of every row in the FPGA, it is saved in the SRAM directly connected with FPGA with the format of 8-bit, which means a data with 8-bit represents eight adjacent pixels, and the order of pixels in the image is on the contrary with the order of the bits. This method saves many memories.

3) DSP read the 8-bit image data from the SRAM to its memory and transfer it into the SDRAM connected with the DSP. It should be noted that two memory spaces in the SDRAM are needed to save adjacent two images to avoid one image being washed by the next image.

4) In the DSP, light knife center of this image based on the centroid method is extracted and 3D information under the camera coordinate system is computed by equation (1) and (2).

5) Smooth the camera coordinates by averaging adjacent three pixels in the light line.

6) Save the results obtained from stage (5) into the upper computer, and convert the camera coordinates into the global world coordinates by the transformation between the camera and the global world coordinate system obtained by calibration.

7) Scan the surface of the object with the movement of scanning device to get 3D information of more points.

8) Combine all 3D global world coordinates to get the depth information of the object from the sensor and its 3D information of this surface.

9) Show the reconstruction image of the object with VC or use the depth information to navigate for service robot.

V. EXPERIMENTS AND RESULTS

A structured light vision sensor (see Fig. 3) was designed in this paper. It contains a high-speed camera, a laser projector and an optical filter. The camera is acA640-100gm/gc, using GigE interface standard, whose resolution is 640*480. It is the progressive scan and can achieve real-time acquisition of 100 frames per second. The wave length of the laser projector is 850nm and the power consumption is 200 mW. The wave length of the optical filter is about 850nm to filter the environmental light and keep the light stripe emitted by laser projector.

The device used for the 3D information extraction in this paper is shown in Fig 1. The left bottom displays the reconstruction result for the object in the front of sensor.

As already discussed, the algorithm proposed in this paper can be divided into three parts: camera calibration, light plane calibration and 3D information extraction. Now we will separately show some results of every part.

Camera calibration was done with a calibration toolbox described in [7]. Fig. 7 shows the images used to calibrate camera intrinsic parameters. More than twenty images are used to improve the precision of calibration. The results of calibration under the initial camera coordinate system are shown in Table I.

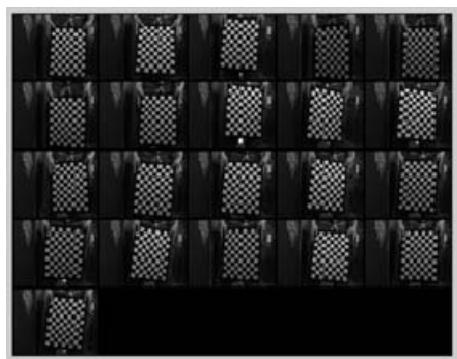


Fig. 7 Images used to calibrate camera parameters

Fig. 8 left shows the images used to calibrate light plane parameters. 11 calibration points (see Fig. 5) are extracted from every image. All the points are used to estimate the light plane with RANSAC. The right image is the result of fitting light plane. The equation of light plane under the initial camera coordinate system is shown in Table I.

The procedure of calculating the transformation between the camera coordinate system and global world coordinate system has been described in the previous section, so here we only give the result of rotation matrix R and translation matrix t in Table I.

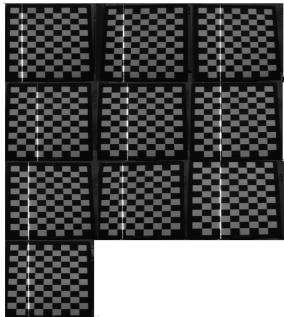


Fig. 8 Images used to calibrate light plane parameters and the results

TABLE I
RESULTS OF CALIBRATION

Camera calibration	Intrinsic matrix K	$\begin{bmatrix} 1066.72 & 0 & 293.4 \\ 0 & 1066.7 & 257.74 \\ 0 & 0 & 1 \end{bmatrix}$
	Extrinsic matrix [R t]	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
Light plane calibration	Equation of light plane	$Z_c = 608.53X_c - 4.73Y_c - 0.09$
Transformation from initial camera to global world coordination system	Rotation matrix R	$\begin{bmatrix} 1.0046 & -0.0591 & -0.0532 \\ -0.0682 & -0.9941 & -0.0048 \\ 0.1830 & -0.1541 & 1.1087 \end{bmatrix}$
	Translation matrix t	$t_i = (i, 0, 0)^T$

Fig. 9 shows some results of extracting 3D information for some objects, where the left is the reconstruction of a coffee machine surface, and the right is the reconstruction for a plane.

The time of processing one image is about 9.2ms on the 600MHz DSP, which means the processing speed can reach 100 frames per second. And the error of measure is less than 2 mm when the distance between the sensor and the object is about one meter. Although the error is larger than Ref. [4] mentioned, it is enough for reconstruction of large object or robot obstacle avoidance, and without regard to the speed of scanning, the speed of processing is more important in these applications.

If higher precision is needed, then you can try by increasing the distance between the laser projector and the camera, or by decreasing the distance between the sensor and the object, or by changing a longer lens or better laser projector.

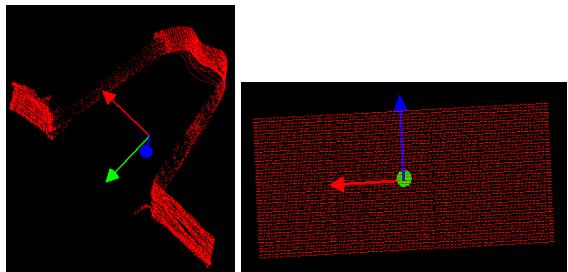


Fig. 9 Reconstruction results for some objects

VI. CONCLUSIONS

This paper designs an embedded 3D vision measurement system and implements the real-time extraction of 3D information for large object based on line structured light on a low-power embedded image processing board with an FPGA and a DSP as the major calculation units. Experiments show that the time of processing an image is about 9.2ms on the 600MHz DSP, fast enough for 3D information reconstruction and robot navigation. At the same time, the measurement error of the algorithm is less than 2mm.

The paper splits the algorithm into three steps: camera calibration, light plane calibration and 3D information extraction, and mainly describes the difference with Ref. [4] and the implementation details on embedded system. At last, some advice is proposed to further improve the precision, if it is necessary. In the future, the system will be used for experiments of service robots navigation and 3D object recognition.

ACKNOWLEDGMENT

The authors would like to thank all the authors of references and the anonymous reviewers for improving the quality of this paper.

REFERENCES

- [1] Qingyang Wu, "A study for key techniques of 3D-sensor based on line-structure light," *Sichuan University*, 2004.
- [2] Chengguo Zhang, "Study on data registration and reduction of reverse engineering," *Ocean University of China*, 2005.
- [3] G Zhang, *Vision Measurement*, Beijing: Science Press, 2008, pp.183-184.
- [4] Fuqiang Zhou, Guangjun Zhang, "Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations," *Image and Vision Computing*, vol. 23, no. 12, pp. 59-67, 2005.
- [5] Jingzhu Xu, "Research on methods and evaluation of stripe center extraction in structured light 3D measurement," *Nanjing University*, 2012.
- [6] David A. Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, Beijing: Publishing House of Electronics Industry, 2004, pp.294-296.
- [7] http://www.vision.caltech.edu/bouguet/calib_doc/htmls/example.html.
- [8] Zhenzhong Wei, Meng Xie, "Calibration method for line structured light vision sensor based on vanish points and lines," *International Conference on Pattern Recognition*, pp. 794-797, 2010.
- [9] J Espinal, M Ornelas, H.J. Puga, et al, "3D scanning system for digital dental based on line structured light sensor," *International Conference on System Science and Engineering*, pp. 353-357, 2012.