Binocular Vision Three-dimensional Imaging Technology by Using Structural Light Projection

Xianmin Ma

Department of Information Engineering, Heilongjiang International University, Harbin, 150025, China

Abstract. SIFT matching algorithm is used to carry out the binocular three-dimensional imaging. Active projection is introduced to solve the problem of low feature quantity and poor matching results in the matching process. By means of projection random speckle, the matching feature is increased, and the matching quality is greatly improved. According to the train running part of the three-dimensional imaging experiment, achieved a good imaging result. Compared with the Fourier profilometry in the active three-dimensional imaging technology. The experimental results show that the structured light projection binocular three-dimensional imaging has a better effect.

1 Introduction

Three-dimensional optical measurement is divided into passive three-dimensional measurement technology and active three-dimensional measurement technology. The passive measurement technology does not rely on the structured light illumination in terms of the illumination mode, and it can use the captured two-dimensional digital image to restore the appearance of the object directly from one or more ready-made video recording systems [1-3]. The active measurement technology requires the illumination of structured light. From the digital image carrying the measuring object's three-dimensional topographic information, the measuring object's three-dimensional topography is obtained by some other algorithm [4-6].

Binocular vision is a passive three-dimensional imaging technology based on human eye imaging theory. It uses the imaging device to obtain two images of the measuring object from different positions based on the parallax principle, and calculates three-dimensional information through the relationship between corresponding points of images [7]. It has the advantages of high efficiency, suitable precision, simple system structure and low cost, so it is very suitable for online and non-contact product inspection and quality control systems at the manufacturing site. Since image acquisition is done instantaneously, it is an effective fast three-dimensional measurement method [8-9].

Railway plays a crucial role in China's economic and social development. In order to keep the running part of the train in good condition during running, it is necessary to conduct detailed inspection work on the running part before the train travels to meet the safe driving condition. In daily primary inspections, inspections are usually made visually or manually, which is time-consuming and labor-intensive. Besides, trains need to be in the high-speed train section or locale depots for a long time to complete inspections. Also, the numerous parts at the bottom and dark environment pose threats to the safety of maintenance personnel. Therefore, in order to save the inspection time and ensure the safety of inspection, an image inspection program should be designed so that the inspector may directly observe three-dimensional images of the bottom parts on the computer side.

A complete binocular stereo vision system consists of six parts: (1) camera calibration; (2) image acquisition; (3) feature extraction; (4) stereo matching; (5) depth calculation; (6) interpolation and reconstruction. For these six steps, domestic and overseas scholars have conducted a lot of research to improve the binocular vision measurement effect.

Based on SIFT binocular matching, the binocular three-dimensional imaging technology is studied in this paper. To handle the difficult matching problem of binocular vision, this paper greatly improves the matching accuracy by introducing the structured light projection, and greatly enhances the three-dimensional imaging effect. The binocular three-dimensional imaging method is used for the inspection of bottom parts and three-dimensional imaging verification is performed.

2 Binocular vision model

Binocular vision three-dimensional imaging mathematical model is as shown below. For convenience of mathematical derivation, two hypotheses are made here:

1. The lens imaging is not distorted;

2. The two cameras' imaging are line-aligned. Based on these two hypotheses, the mathematical relationship

Corresponding author: 1652402051@qq.com

between the depth of the spatial point $P(X,Y,Z)$ and the coordinate points $P_l(X_l,Y_l)$ and $P_r(X_r,Y_r)$ of the point in the left and right views is obtained.

The mathematical derivation is made based on the above two hypotheses below.

Assuming that $P(X,Y,Z)$ is a point to be measured in space, and its imaging points are $P_l(X_l,Y_l)$ and $P_r(X_r,Y_r)$ respectively in the left and right cameras. The main points of the two cameras are $C_l(C_{lx},C_{ly})$ and $C_r(C_{rx},C_{ry})$ respectively. Note that the main point is the intersection of the chief ray and the image plane, which is on the optical axis of the lens. Since the mechanical installation does not guarantee that the lens axis completely coincides with the normal of the CCD imaging center, the main imaging point does not completely coincide with the center of the image, that is, C_l and C_r are not the center points of the left and right views.

The points in the physical world are projected onto the camera, which can be expressed by the following Equation (1).

$$a = QA \tag{1}$$

Where,

$$a=\begin{bmatrix} X' \\ Y' \\ Z' \\ w \end{bmatrix}, Q=\begin{bmatrix} 1 & 0 & 0 & -c_{lx} \\ 0 & 1 & 0 & -c_{ly} \\ 0 & 0 & 0 & f \\ 0 & 0 & -\frac{1}{T} & \frac{c_{lx}-c_{rx}}{T} \end{bmatrix}, A=\begin{bmatrix} x_l \\ y_l \\ x_l-x_r \\ 1 \end{bmatrix}$$

Then:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{w}\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}$$

It should be noted that the actual situation is that the horizontal center distance between the two cameras should be written as T_x. Here is the simplified derivation under the assumption that the X coordinate of the world coordinate system is taken on OO', so $T_x = T$.

x_l, x_r are the horizontal pixel positions of the P point in the imaging view of the right and left cameras, and the difference $x_l - x_r$ is the so-called parallax. Due to the diversity and complexity of the measuring object, the solution to parallax is often the key point in the whole 3D imaging.

In combination with the camera distortion, according to the pinhole camera model, the relationship between the object's world coordinate point and the corresponding pixel on the computer image is as Equation (2)

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & \gamma' & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} R & T \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = K\begin{bmatrix} R & T \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{2}$$

Where, K : camera internal reference matrix

$R = [r_1 \quad r_2 \quad r_3]$: Rotation matrix of camera:

T : translation vector of camera

In the Equation (2), the internal reference matrix contains the camera's focal length, distortion and other information, while the binocular camera's rotation, translation, and other information are contained in the external reference matrix. By integrating the internal reference matrix with the external reference matrix, the author may yield the following Equation (3):

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{3}$$

Where, M is the projection matrix (also called homography matrix).

After the equation is expanded, three linear equations can be obtained. After the scale factor s is removed, two relations are left and there are three unknowns, so at least two cameras are required to complete the depth measurement.

3 SIFT feature matching algorithm

Scale Invariant Feature Transform (SIFT) was patented by a Canadian professor David G. Lowe. SIFT feature remains invariant to rotation, scale and brightness changes, and it is a very stable local feature. The algorithm first establishes a scale-space description of the image and identifies potential scales and selects invariant extreme points through Gaussian differential functions. According to its local gradient direction, a feature description vector is established, which is a feature detection description method with scaling, rotation and affine invariance.

After convolving the image I with the two-dimensional Gaussian function G of different Gaussian kernels, the scale space G at different scales is obtained, and the Gaussian difference image $D(x,y,\sigma)$ is obtained after the two are subtracted. Each of these pixels is compared with its upper layer, lower layer and neighbors for a total of 26 pixels, and the maximum or minimum value is used as a candidate feature point. In order to enhance the matching stability, it is also necessary to remove the low-contrast extreme points and the unstable edge response points in the candidate feature points, so as to accurately locate the extreme points to obtain the local feature points.

Taking each local feature point as a center, a 16×16 window is taken in its neighborhood and is divided into

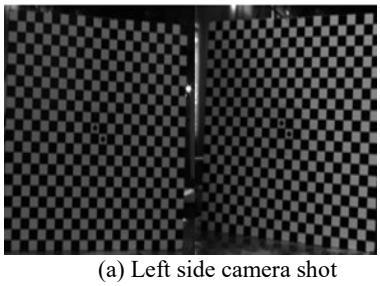
4×4 pixel blocks. A gradient histogram of 8 directions is calculated on each pixel block, which belongs to 8 intervals. The value of each interval is a Gaussian weighted cumulative value of gradient amplitude. Therefore, each 4×4 block of pixels can be represented by a description vector of 8 dimensions. In this way, a SIFT feature vector of 128 is generated for each local feature point. In addition, the feature vector length is normalized to further remove the effects of illumination.

The matching of feature vectors is to measure the similarity of SIFT feature vectors by calculating the Euclidean distance $U_{ab} = \sum (a_i - b_i)^2, i \in (1,2,,128)$ of the local feature points in the two images to be matched, that is, to find the nearest local feature point of the first image in another image adjacent. A ratio threshold R is set and $0 < R \le 1$. It is determined to be a correct matches when $\min < R$, where min is the nearest neighbor distance.

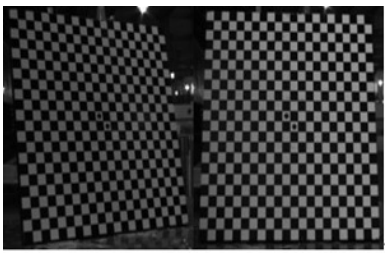
4 Experimental results and analysis

4.1 Binocular three-dimensional Imaging Experiment

Before performing three-dimensional imaging, the binocular camera needs to be calibrated, i.e., the homography matrix of the camera should be calculated. This paper adopts the Zhang Zhengyou [10] camera calibration method. Calibration checkerboard is shown in Figure1.



(a) Left side camera shot



(b) Right side camera shot

Figure 1. Camera Calibration Checkerboard

The following camera parameters are obtained after calculation, it is shown in Equation (4).

$$H_1 = \begin{vmatrix} 31.9 & 1.05 & 941.76 \\ -5.58 & 39.31 & 626.26 \\ 0 & 0 & 1 \end{vmatrix}$$

$$H_2 = \begin{vmatrix} 36.54 & -0.55 & 1261.3 \\ -1.23 & 39.66 & 603.33 \\ 0 & 0 & 1 \end{vmatrix} \quad (4)$$

After the camera parameters are calibrated, the two images acquired by the binocular camera are as shown in the Figure2.

(a) Left side camera shot

(b) Right side camera shot

Figure 2. Image pair of the measuring parts

The Figure3 shows the train's bottom parts, which are featured by dark environment, a large number of flat areas, and fewer extractable feature points. Images are binocularly matched directly:

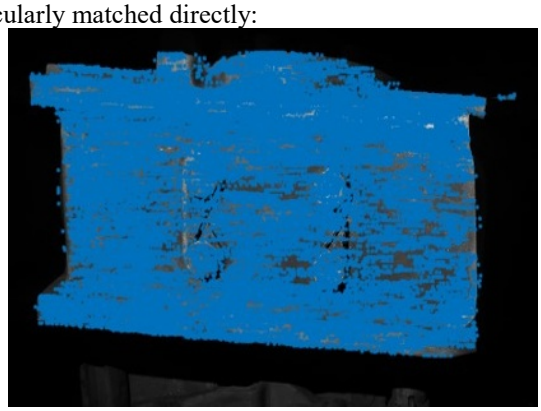


Figure 3. Binocular matching results

According to the matching results, it can be seen that there are a large number of unmatched gaps in the planar area, failing to meet the standard of 3D imaging. In order to enhance the matching effect, the feature points are increased, and the structured light projection is cited here. The projected image is a random speckle, it is shown in Figure 4.

(a) Left side camera shot

(b) Right side camera shot

Figure 4. Image pair of speckle projection parts

Compared with the matching results of the images not projected, it can be seen that the images with random speckles are almost completely matched, and the camera calibration parameters can be used for three-dimensional reconstruction. The results are shown in the Figure5.

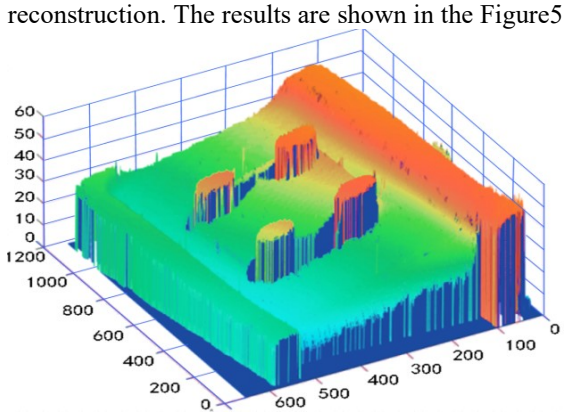
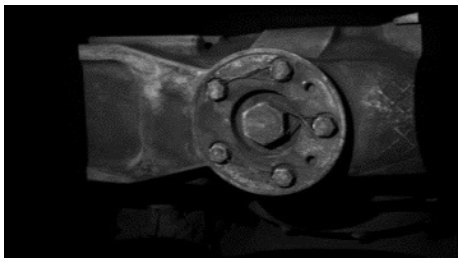


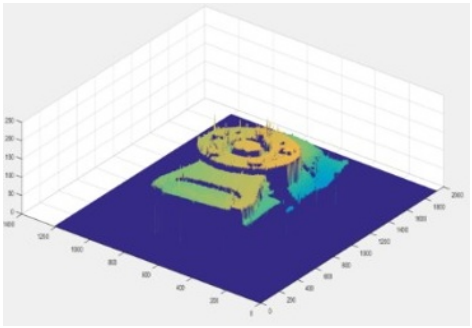
Figure 5. Three-dimensional shape restoration

4.2 Experimental Comparison with Fourier Profilometry

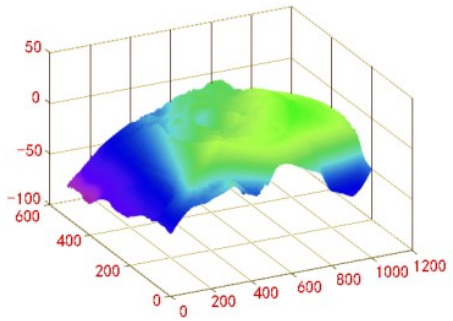
Fourier profilometry is an active three-dimensional measurement technology. The frequency spectrum of the raster fringe image is calculated by using a fast Fourier transform. After filtering, the phase value of the measuring object surface is extracted from the frequency spectrum, and then the three-dimensional shape data of the object is calculated. It is shown in Figure6.



(a) Two-dimensional image



(b) Binocular three-dimensional restoration image



(c) Fourier profilometry

Figure 6. Comparison experiment

Through the above experiments, it is concluded that the key factor of binocular three-dimensional imaging lies in the accuracy and redundancy of matching points. When the feature points are more and similar to the edges, the restoration results are better; while the planes with fewer feature points (continuous smooth surfaces, such as wheel treads), the restored three-dimensional images will be sparse and have certain noise points. These are caused by inaccurate matching. As for the Fourier profilometry, due to the periodic nature of the grating, it is inaccurate in dealing with phase cutoff regions (discontinuous portions in actual objects, such as cross-sections and fractures). This method has better restoration effect for continuous surfaces such as wheel treads, and when the components in the area are complex, it cannot produce accurate three-dimensional images. In general, binocular vision has superior imaging performance.

References

1. Guo Weiqing, Tang Yiping, Lu Shaohui, Chen Qi. A survey of 3D stereovision measurement and reconstruction based on mirror imaging technology. Computer science, 43 (09),1-10(2016)

2. Wei Shaopeng. 3D imaging technology based on the combination of depth camera and binocular vision. Zhejiang University, 2015
3. Wang Xiangjun, Yu Ya Nan. 3D visual testing of MAV aerodynamic shape based on stroboscopic imaging technology. nanotechnology and precision engineering, 9(06), 509-514(2011)
4. Li Keguo. Research on 3D facial imaging technology and algorithm based on binocular stereo vision. Dalian Polytechnic University, 2009
5. Liao Xiangrong, Chen Xiaoqing. Exploring new visual effects in animation from 3D virtual imaging technology. Journal of fine arts, (04), 64-65(2004)
6. Hu Xiaokun, Bi Yuan Wei. Four wheel alignment parameter modeling method based on 3D imaging technology. computer measurement and control, 22 (10), 3362-3364(2014)
7. Su Xian Yu, Zhang Qican, Chen Wenjing. Structured light 3D imaging technology. China laser, 41 (02), 9-18(2014)
8. Zahedi M, Salehi S M.License plate recognition system based on SIFT feature. Procedia Computer Science. 3, 998-1002(2011)
9. Belcher C,Du Y. Region-based SIFT approach to iris recognition .Optics&Lasers in Engineering. 47(1),139-147(2009)
10. Zheng you, Zhang. Flexible camera calibration by viewing a plane from unknown orientations. Proceedings of the 7th International Conference on Computer Vision, 666-673(1999)