Binocular Vision Detection and 3D Construction Based on Encoded Light

Hao Zhu[1,2](✉), Mulan Wang[2], and Kaiyun Xu[2]

[1] School of Communication, Nanjing Institute of Technology,
Nanjing 211167, China
zhuhao@njit.edu.cn

[2] Jiangsu Key Laboratory of Advanced Numerical Control Technology,
Nanjing 211167, China

Abstract. A non-contact 3D data acquisition system based on fringe encoded light is analyzed. In this paper, the characteristics of encoding light structure based on space-time domain are studied, and the encoding combination with fringe boundary features is designed. This paper discusses the hierarchical reconstruction technology based on the three levels of projective radiation metric, which is better adapted to the measurement environment of inaccurate or uncalibrated camera parameters calibration, and more convenient to introduce the optimization algorithm. On the other hand, the affine coordinate system and the projective coordinate system are more widely used than the Euclidean coordinate system. The active binocular measurement system is constructed to collect the spatial information of 3D objects and store it in the form of point cloud.

Keywords: Encoded light · 3D construction · Binocular vision

1 Introduction

3D reconstruction based on machine vision refers to the process of reconstructing 3D information from single view or multi view images. Its main task is to acquire the two-dimensional image of the target by non-contact way based on the camera, grating, projector and computer, and then analyze and extract the spatial information to obtain the three-dimensional coordinate information of the target and store it in point cloud or other formats. There are two kinds of methods to obtain coordinate information of three-dimensional objects by machine vision technology: passive three-dimensional sensing and active three-dimensional sensing. The advantage of passive sensing method is that the hardware structure is relatively simple and easy to implement, but the corresponding point matching

Supported by organization by Natural Science Foundation of Universities in Jiangsu Province (Grant No.: 16KJA460001), Creation Foundation Project of Nanjing Institute of Technology (Grant No.: ZKJ201611).

X. Sun et al. (Eds.): ICAIS 2020, CCIS 1254, pp. 607–617, 2020.
https://doi.org/10.1007/978-981-15-8101-4_54

algorithm is usually complex and the processing is long. On the other hand, the reflection of light on the measured object is larger in the passive sensing method. Based on the passive mode, the active three-dimensional sensing mode adds a structural light source to the target. Due to the depth change of the surface of the three-dimensional object in the direction of the projected light, the modulation of space or time will be produced to the projected structured light, forming distortion. The distortion of the two-dimensional image obtained from different angles is also different from each other. The three-dimensional data of the object can be extracted by demodulating the projection of the distortion. Relevant researches [1] shows that, using binocular vision technology can only reconstruct the target in projective space or affine space. Hartley [2] proposed that projective space, affine space and Euclidean space can use different transformation matrices to transform each other. Based on the above theory, this paper studies the hierarchical reconstruction technology composed of projective reconstruction, affine reconstruction and European reconstruction.

2 Mathematical Modeling of Vision Sensor

For a monocular vision system with only one projector and one camera, an equivalent camera can be virtual by rigid rotation and translation of the projector and camera [3]. But the precondition of the transformation is that the internal parameters of the projector and the camera are the same. In engineering practice, the above conditions are generally difficult to meet. Therefore, a binocular vision system consisting of one projector and two cameras is usually used for 3D detection. Generally speaking, the projector can be virtual as a pinhole image and the camera can be virtual as a linear camera, both of which meet the model described below. When the internal parameters of the two cameras are the same and the optical center is in the same horizontal plane, the height of the collected image is the same. The corresponding points can be determined by finding the feature points of the same height. Therefore, the binocular vision system can use two cameras of the same model, which are placed on a horizontal platform, and the projector is placed between the two cameras. The two cameras collect the projection image which is projected by the projector to the three-dimensional target at the same time (Fig. 1).

3 Coding Optical Technology

3.1 Principle of Encoded Light

In the active 3D measurement system, the projector projects light with certain characteristics to the target, and analyzes the spatial position of the corresponding points according to the pattern characteristics on the target. This kind of light with certain characteristics is called encoded light. Griffin et al. [4] studied the theory of encoded light, and proposed four features that encoded light needs

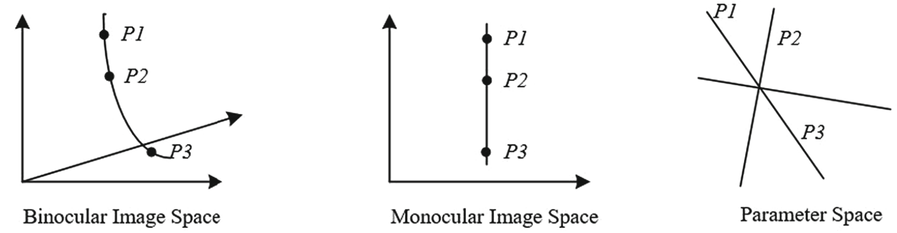

Fig. 1. Active binocular vision measurement system.

to meet: (1) The position of each point is determined by itself and four neighboring points around it. (2) The same code cannot correspond to more than two points. (3) All coding information is determined by a fixed base represented by symbols. (4) The size of coding matrix is required to reach the extreme value. After the above conditions are met, the corresponding relationship between the image feature points projected on the 3D object and the feature points on the projection template can be determined by decoding operation, so as to obtain the spatial position information of the data points.

According to the different coding forms, the common coding light can be divided into space-based coding and time-based coding. Space-based coding includes gray code grating coding, space phase coding, gray-scale coding and so on. In these models, the projection light encodes the spatial information with different stripe width, gray level and color information. Time-based coding refers to the acquisition of motion vectors in images at different times to obtain the shape information of objects, which is initially used to obtain the spatial position of moving objects. Because any measured surface can be regarded as the deformation of the reference plane over a period of time, the detection technology based on time coding is gradually applied to the field of 3D detection.

3.2 Coded Optical Design

In contrast, the advantage of space-based coding is that it is easy to extract features, the mathematical model is simple, and its disadvantage lies in its weak anti-interference ability to the outside world, especially not suitable for the information acquisition of moving objects. However, the anti-jamming ability based on time coding is strong, but how to determine the optical characteristics of coding is difficult to achieve. Based on the above characteristics, this section discusses a stripe encoded light model based on spatiotemporal coding.

Reference [3] studies how to design a coding structure to make it more suitable for the measurement of three-dimensional object space information. The basic characteristics of encoded optical structure are proposed:

For different sampling time, each pixel in the same position should form a feature vector; the same pixel is not associated with other pixels in at least one direction, to ensure that the coding structure has a certain degree of anti-interference.

For projection gray scale, only black and white colors are used to improve the calculation efficiency and anti-interference.

For an image with a given number of frames, there should be as much detection information as possible. When coding, the combination coding of current information should be used as much as possible to maximize the detection coding space.

For coding feature detection, fringe boundary feature is used as much as possible. On the one hand, for n stripes, there are 2n stripe boundaries, which can contain more information. On the other hand, the boundary information only has the jumping state but no width, which can avoid the stripe width corrosion and deformation caused by interference and is more in line with the ideal mathematical model.

Based on the above conditions, a coding combination with fringe boundary features can be designed, using '0' and '1' to represent the switching state of black-and-white stripes. The corresponding stripes in each frame constitute a binary coding state. With the change of time parameters, the change of projection pattern is represented by the switching of coding state. The principle of stripe coding is shown in Fig. 2.

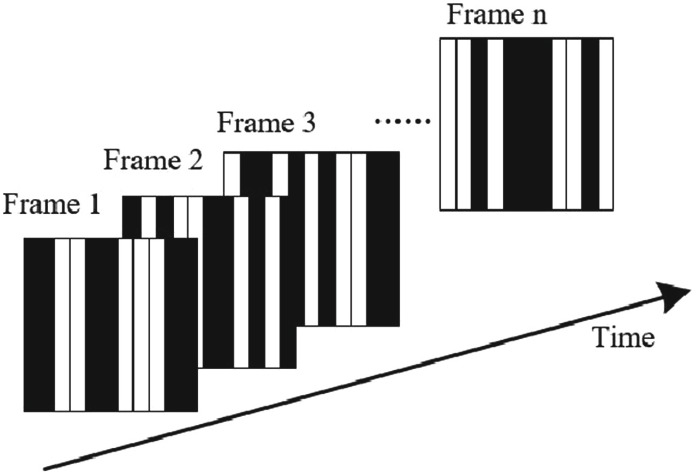

Fig. 2. Principle of stripe coding.

The above coding structure shall meet the following conditions:

Every coding stripe has at least one jump between the images. If there is no jump all the time, it is considered to be caused by the inherent texture of the object.

If the color of adjacent stripes is the same in a certain frame, the fringe boundary is considered invisible.

Stripe coding should be conducive to amplifying the error information in detection, so as to discover and discard the frame in time.

The jump structure in fringe coding should be distributed reasonably, which is convenient for the realization of projection and acquisition.

In reference [3], a comprehensive coding rule based on algebra, graph theory and Hadamard matrix is introduced.

The first step is to use graph theory to set the center core of coding, that is, the center stripe and the nearby coding.

The second step is to generate the initial coding matrix by Hadamard algorithm.

The third step is to make use of different columns in the commutative algebra commutative coding matrix to make the matrix meet the design requirements of stripe coding.

After the projection of the fringe on the three-dimensional measurement target is obtained, the pixels in the same position in the multi frame image are encoded according to the time sequence, the corresponding fringe boundary position and the jump situation are calculated, the projection data points are matched with the corresponding fringe coding points, and then the depth information of the three-dimensional object surface can be solved by the reconstruction model in the following paper.

4 3D Layered Reconstruction Technology

4.1 A Subsection Sample

The technology of three-dimensional layered reconstruction is gradually developed in the relevant theories put forward by Faugeras [1] and others in the 1990s. Different from the traditional 3D reconstruction technology, which directly establishes the 3D data structure of the target in the European coordinate system, the 3D hierarchical reconstruction establishes a reconstruction framework, which reconstructs the geometric model of the target step by step. The advantage of this technology is that it can better adapt to the measurement environment where the camera parameters are inaccurate or uncalibrated, and it is easier to introduce the optimization algorithm. Compared with the target Euclidean coordinate system of 3D reconstruction, affine coordinate system and projective coordinate system are more widely used. Therefore, Faugeras [1] divides hierarchical reconstruction technology into three steps: projective reconstruction, affine reconstruction and metric reconstruction.

The results show that there is a set of non singular transformations between the reconstructed projective matrices in the above three steps. When the transformation is affine transformation, affine reconstruction can be obtained, and when the transformation is Euclidean transformation, metric reconstruction can be formed.

4.2 Implementation of Projective Reconstruction

Set the corresponding point set in the two images as x_j and ${x_j}'$. Existence matrix F for any existence j:

$$x_j{}'Fx_j = 0 \tag{1}$$

This matrix is called the basic matrix between two images. Let $(P_1, P_1{}')$ and $(P_2, P_2{}')$ be two cameras, $(P_1, P_1{}', \{X_1\})$ and $(P_2, P_2{}', \{X_2\})$ are two reconstructions from x_j to $x_j{}'$, then there is a non singular matrix

$$P_2 = P_1 H, P_2{}' = P_1{}' H \tag{2}$$

And for each j:

$$X_{2j} = H^{-1} X_{1j} \tag{3}$$

Therefore, there are the following relationships:

$$P_2 X_{2j} = P_2(H^{-1} X_{1j}) = P_1 H H^{-1} X_{1j} = P_1 X_{1j} = x_j \tag{4}$$

Equation 4 shows that both point $H^{-1}X_{1j}$ and point X_{2j} are mapped to the same point x_j, and all pass through the corresponding center line of the camera P_2. In the same way, it can be deduced that the above two points are also on the center line corresponding to the camera $P_2{}'$.

For multiple images, if the camera takes m images of n target points from the perspective m, where $x_j{}'$ is the image point of the j' point under the i-th camera P^i, then for all the images, there are the following equations:

$$W = \begin{bmatrix} \lambda_1^1 x_1^1 & \ldots & \lambda_n^1 x_n^1 \\ \ldots & & \ldots \\ \lambda_1^m x_1^m & \ldots & \lambda_n^m x_n^m \end{bmatrix} = \begin{bmatrix} P^1 \\ P^2 \\ \ldots \\ P^m \end{bmatrix} \begin{bmatrix} X_1 & X_2 & X_n \end{bmatrix} = \tilde{P}\tilde{X} \tag{5}$$

Where W is the measurement matrix and λ_j^i is the projection depth.

Equation 5 shows that the measurement matrix W can be decomposed into a matrix $\tilde{P}$ representing camera motion and a matrix $\tilde{X}$ representing the shape of space objects. Therefore, as long as the photographing depth λ_j^i can be estimated by some method, three-dimensional space points can be calculated by Eq. 5.

Equation 1, as long as the corresponding image points are known, the camera matrix is not needed to solve F. In reference [5], an 8-point algorithm is provided, which can solve the above basic matrix linearly based on 8 matching points. 8-point algorithm is a commonly used method to solve the basic matrix. The disadvantage of the 8-point algorithm is that on the one hand, when the selection point changes, the difference of the basic matrix is large, on the other hand, the data calculation of the algorithm is large. Based on the above shortcomings, a basic matrix estimation.

The basic idea of Hough transform is to use the duality of point and line to map the curve in the space to a point in the parameter space, so as to transform the problem of detecting the shape of space into the problem of peak statistics. Considering that Hough transform can transform the collinear points in rectangular coordinate system into a group of curves intersecting at one point in parameter domain, and the spatial points encoded by the same fringe boundary are also mapped into a group of collinear points on monocular imaging surface, so we can first transform the fringe encoded feature points in binocular vision system into collinear points in monocular vision system, and then transform Hough into a group of curves. When there is no error in the conversion process, the above curves should intersect at one point. If there is an error, there will be multiple intersections for the above curve group. The above process is shown in Fig. 3.

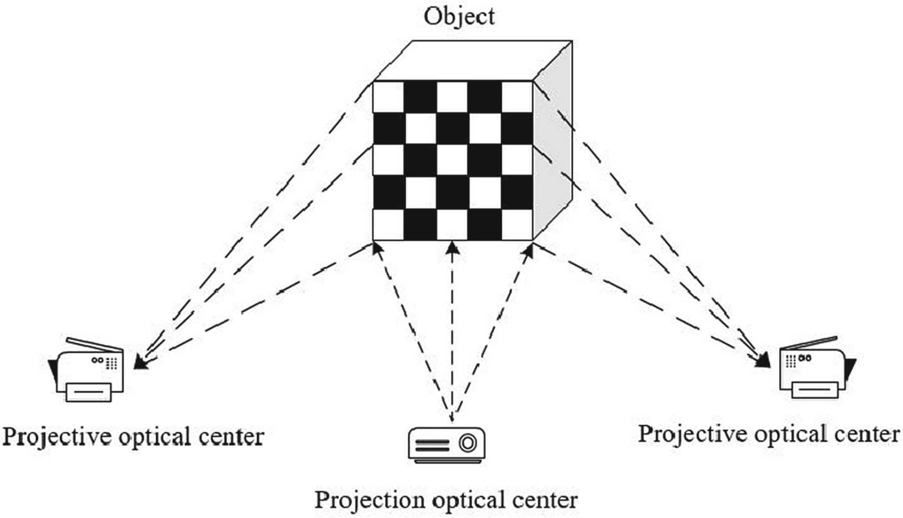
Object
Projective optical center
Projective optical center
Projection optical center

Fig. 3. Monocular imaging transformation and Hough transformation.

As shown in Fig. 3, after Hough transformation, if there are multiple intersections in the curve group, and the intersections are distributed in the circle with radius r, the radius is defined as Hough radius, which is recorded as R_{Hough}. In order to simplify the calculation, the maximum distance between two intersections can be taken as Hough radius.

According to the above definition, there is a correlation between Hough radius and basic matrix estimation. If the true value of the basic matrix is F and the estimated value is $\hat{F}$, there is the following relationship:

$$\lim_{F \to \hat{F}} R_{Hough} = 0 \tag{6}$$

Therefore, Hough radius can be used as the evaluation function of the basic matrix. When the noise distribution of image acquisition is isotropic zero mean Gaussian distribution and independent distribution, the 8-point algorithm can be used as the initial value of L-M estimation, and the Hough radius as the L-M optimization factor, so as to solve the optimal solution in the sense of maximum likelihood.

$$\min_{p \in P} \sum R^{p}_{Hough} \tag{7}$$

After estimating the basic matrix F, a group of projection matrices can be constructed according to the geometric model of trigonometry.

4.3 Affine Reconstruction Design

According to the previous analysis, there is an affine transformation between affine reconstruction and Euclidean reconstruction, and the affine transformation remains unchanged at infinity. Therefore, if the coordinates of infinite plane in projective space can be determined, it can be transformed from projective reconstruction to affine reconstruction.

Figure 4 shows the establishment of an ideal binocular reconstruction model in affine space. C_1 and C_2 are two cameras with identical internal parameters in affine space. The optical axis and y axis of the two cameras are parallel and coincide with the x axis. P is the measured point. p_1 and p_2 are the projections of P on C_1 and C_2. L_1 and L_2 are the corresponding polar lines. Because L_1 and L_2 are parallel to each other, their poles e_1 and e_2 are at infinity. If the projective model can be transformed into the above model, then the infinite surface can be determined, thus the affine reconstruction can be completed.

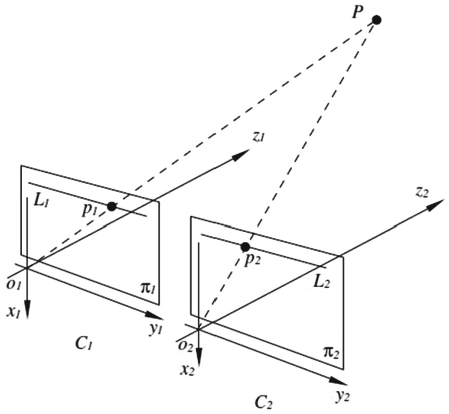

Fig. 4. An ideal binocular imaging model in affine space.

Since the poles e_1 of the camera C_1 are at infinity, it is advisable to set:

$$e_1 = \left[1, 1, 0\right]^T \tag{8}$$

According to the camera basic model, the basic matrix between C_1 and C_2 is

$$\bar{F}e_1 = 0 \tag{9}$$

Then

$$\bar{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{10}$$

The binocular correction algorithm proposed by Charles [7] provides a non singular transformation, which can transform the basic matrix F obtained from projective reconstruction into the matrix v in Eq. 10, so the poles in the original image can be projected to the infinite position, thus completing the affine transformation.

4.4 European Reconstruction Design

In the model shown in Fig. 4, let the coordinate of P point in C_1 coordinate system be (x, y, z), the translation distance between C_1 and C_2 be d, then the coordinate of P point in C_2 coordinate system is $(x - d, y, z)$. According to the central projection principle of the camera, we can know:

$$\begin{cases} x_1 - x_0 = f_x \frac{x}{z} \\ y_1 - y_0 = f_y \frac{y}{z} \\ x_2 - x_0 = f_x \frac{x-d}{z} \\ y_2 - y_0 = f_y \frac{y_1}{z_1} \end{cases} \tag{11}$$

Where, x_0, y_0, f_x, f_y are the internal parameter of the camera, $(x_1, y_1), (x_2, y_2)$ are the coordinates of p_1, p_2.

It can be seen from Fig. 4 that there is only a translation relationship between the coordinate system C_1 and the world coordinate system, so the coordinate system C_1 can be set as the world coordinate system, which can be obtained from Eq. 11

$$\begin{cases} x = \frac{d(x_1 - x_0)}{x_1 - x_2} \\ y = \frac{d f_x (y_1 - y_0)}{f_y (y_1 - y_2)} \\ z = \frac{d f_x}{x_1 - x_2} \end{cases} \tag{12}$$

It can be seen from Eq. 12 that when the camera internal parameters are determined, the three-dimensional coordinates of the space points P can be obtained from the coordinates of p_1, p_2 of P point in C_1 and C_2 coordinate systems.. The three-dimensional coordinates of all data points on the surface of the target can be obtained by traversing the image points in the two cameras and saved as a point cloud format.

5 Experiment and Analysis

The fringe encoded structured light shown in Fig. 5 is projected on the model to detect the 3D data of the model, and the results are represented by point cloud format. The test results of 3D data of the model are shown in Fig. 6. The left figure shows the measurement target and the right figure shows the measurement result. As shown in the figure, the number of three-dimensional points of the reversing lamp modelF is 27614. As can be seen from the figure, the 3D point cloud results can basically reflect the shape characteristics of the model.

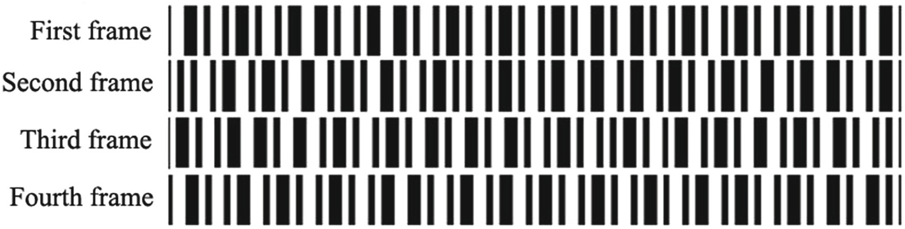

Fig. 5. Stripe encoded optical structure.

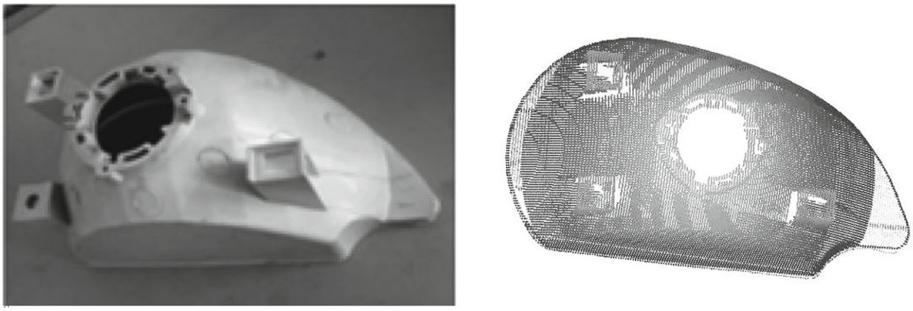

Fig. 6. Sampling experiment of reversing lamp model.

6 Conclusion

In this paper, the coding light detection technology based on machine vision is studied. The vision measurement technology based on encoded light technology is analyzed and discussed. This paper studies the construction of machine vision measurement system, and discusses the calibration technology of camera and projector. The mathematical model of active sensing binocular vision measurement system is obtained. This paper discusses the technology of structured light coding, analyzes the characteristics of structured light coding, and introduces the method of boundary stripe coding based on spatiotemporal coding. This paper analyzes the basic principle of three-dimensional layered technology, and realizes the measurement and description of the target three-dimensional point cloud model by using the layered method of projective reconstruction, affine reconstruction and metric reconstruction. The experimental results show that the proposed method is feasible.

References

1. Maybank, S.J., Faugeras, O.D.: A theory of self-calibration of a moving camera. Int. J. Comput. Vis. **8**(2), 123–151 (1992)
2. Hartley, R.I.: Euclidean reconstruction from uncalibrated views. In: Mundy, J.L., Zisserman, A., Forsyth, D. (eds.) AICV 1993. LNCS, vol. 825, pp. 235–256. Springer, Heidelberg (1994). https://doi.org/10.1007/3-540-58240-1_13
3. Ankang, Y.: 3D layered reconstruction technology based on structure encoded light. Southeast University, March 2009
4. Griffin, P., Narasimhan, L., Yee, S.: Generation of uniquely encoded light patterns for range data acquisition. Pattern Recogn. **25**(6), 609–616 (1992)
5. Luong, Q., Faugeras, O.: The fundamental matrix: theory, algorithms, and stability analysis. Int. J. Comput. Vis. **17**(1), 43–75 (1996)
6. Leavers, V.: Shape Detection in Computer Vision Using the Hough Transform. Springer, New York (1992). https://doi.org/10.1007/978-1-4471-1940-1
7. Loop, C., Zhang, Z.: Computing rectifying homographies for stereo vision. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1999, January 1999