ADVANCES IN STEREO VISION

Edited by **José R.A. Torreão**

INTECHWEB.ORG

Advances in Stereo Vision
Edited by José R.A. Torreão

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Davor Vidic
Technical Editor Teodora Smiljanic
Cover Designer Jan Hyrat
Image Copyright Saulius L, 2010. Used under license from Shutterstock.com

First published June, 2011
Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Advances in Stereo Vision, Edited by José R.A. Torreão
p. cm.
ISBN 978-953-307-837-3

Contents

Preface

Stereopsis is a vision process whose geometrical foundation has been known for a long time, ever since the experiments by Wheatstone, in the 19th century. Nevertheless, its inner workings in biological organisms, as well as its emulation by computer systems, have proven elusive, and stereo vision remains a very active and challenging area of research nowadays. In this volume we have attempted to present a limited but relevant sample of the work being carried out in stereo vision by researchers from around the world. We have chapters dealing with the implementation of stereo algorithms in dedicated hardware; with active stereo vision systems; with stereo based on omnidirectional images; with the application of stereo vision to robotic manipulation and to environment modeling; with the psychophysical aspects of stereo, and with the interface between biological and artificial stereo systems. Thus, we believe that we have covered significant aspects of stereopsis, both from the applied and from the theoretical standpoints.

We would like to thank all the authors who contributed to this project, and also the editorial staff at InTech, especially Mr. Vidic, for their continuous support.

José R.A. Torreão
Instituto de Computação
Universidade Federal Fluminense
Brazil

1

Active Stereo Vision for 3D Profile Measurement

Jing Xu[1], Qiang Yi[1], Chenglong Fu[1], Huabin Yin[2],
Zhengda Zhao[2] and Ken Chen[1]
[1]*Tsinghua University*
[2]*AVIC Chendu Aircraft Industrial(Group)Co., Ltd*
China

1. Introduction

Over the past decade, vision-based 3D sensing technology has been increasingly applied in manufacturing industries. The 3D shape of a part, which can be represented by using a point cloud, is usually required for two main purposes: reverse engineering or dimensional inspection. On the other hand, vision-based 3D sensing techniques can be divided into categories: passive stereo vision and active stereo vision.

Stereo vision based on no additional devices besides the cameras is known as passive stereo vision, which works in a similar way as the human eyes. In this case, the passive stereo vision can be very compact and low-cost without any extra components. The extensive application of the passive vision benefits from the epipolar geometry, first introduced in (Longuet, 1981). Epipolar geometry, which provides the geometric constraints between 2D image points in the two cameras relative to the same 3D points with the assumption that the cameras can be presented by using the pinhole model, has been utilized in camera calibration. However, it still has some drawbacks for industrial inspection. The first difficulty is the correspondence problem. In other words, determining the pixels of different views in terms of the same physic point of the inspected part is not a trivial step, especially for a texture-less object, such as a piece of white paper. Another problem is the sparse resolution of the reconstruction, usually with a small number of points. Furthermore, the inappropriate ambient light condition would also lead to the failure of the passive stereo vision.

In order to overcome the above drawbacks, active stereo vision, removing the ambiguity of the texture-less part with a special projection device, is commonly used when dense reconstructions are needed. For this technique, a special device (e.g. projector) is employed to emit special patterns onto the identified object, which will be detected by the camera.

In a word, compared with the passive strategy, the active one is advantageous for robust and accurate 3D scene reconstruction.

This chapter summarizes the coding strategy, 3D reconstruction, and sensor calibration for active stereo vision, as well as the specific application in manufacturing industry. Our contribution is to propose two pattern coding strategies and pixel-to-pixel calibration for accurate 3D reconstruction in industrial inspection.

2. Coding strategy

2.1 Related work

The key of the active stereo vision method is the encoding of the structured light pattern, used to establish the correspondence between the camera and the projector, since it would impact all the system performance, including measurement accuracy, the density of point cloud, perception speed and reliability.

This chapter will focus on the fast 3D profile management. For this purpose we only summarize the coding strategies with a single and a few patterns. A great variety of different patterns have been addressed during the past decades(Salvi et al., 2010), e.g., temporal-coding patterns, direct-coding patterns, and spatial-neighborhood patterns, among which the temporal-coding patterns are multi-shot and the other two patterns are one-shot.

For the temporal-coding approach, a group of patterns are sequentially illuminated onto the measured surface. The codeword of each pixel is usually generated by its own intensity variance over time. Therefore, this approach is usually regarded as a pixel-independent and multiplexed approach. Because of the high accuracy and resolution performance, the temporal patterns are the most extensively employed method in optical metrology.

At present, the phase-shifting method (PSM), which is a typical example of the above temporal patterns, is the most commonly used pattern in 3D profile measurement for industrial quality inspection. The reason is that this method could reach pixel-level resolution with high density. Another benefit of this technique is its robustness to surface reflectivity and ambient light variations. For this technique, the minimum number of patterns required is three. Hence, a three-step phase shifting pattern is usually used, in which three sinusoidal patterns with $2\pi/3$ phase shifting relative to each other are utilized (Huang & Zhang, 2006).

However, the calculated phase distribution is constricted in the rage of $\left[-\pi\ +\pi\right)$ by means of anti-tangent function due to the periodic property of the sinusoidal waveform, which is named relative phase. Therefore, it is necessary to determine the order of phase shifting in the camera image plane to eliminate the ambiguity, in order to obtain the absolute phase, which refers to the continuous phase value relative to the standard phase.

The absolute phase φ_a is usually expressed using the relative phase φ_r as

$$\varphi_\mathrm{a} = \varphi_\mathrm{r} + 2k\pi \tag{1}$$

where k is the order of phase shifting. Furthermore, the relationship between the absolute phase φ_a and the relative phase φ_r can be demonstrated as in figure 1.

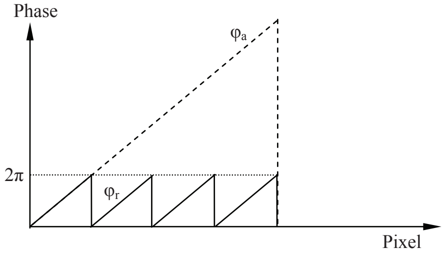


Fig. 1. The relationship between absolute phase and relative phase

To solve this problem, several unwrapping algorithms have been developed (Ghiglia & Pritt, 1998), among which a general unwrapping algorithm is to introduce a marker, i.e., a line in the perpendicular direction of the phase distribution. In this case, the absolute phase with respect to the reference marker can be obtained by using the assumption of continuity of the measured object. Several similar strategies have also been developed to solve this problem. It should be pointed out that the proposed algorithms can only be used for smooth surfaces with height variation no more than 2π within any adjacent pixels. Therefore, the 2π ambiguity problem will arise when measuring surfaces with abrupt steps, resulting in inaccuracy of the 3D measurement. Increasing the wavelength of phase shifting can solve this problem; however, the measurement accuracy will be affected and the system will be susceptible to noise.

One feasible solution is to take advantage of gray code and phase-shifting (GCPS) methods. The gray code is essentially a binary code in which only two intensity levels are used. Moreover, the constraint of Hamming distance is applied in the codeword formulation in the gray code method. Thus, this technique is robust to noise. The basic idea of the GCPS method is to divide the entire image plane into small patches by using the gray code method to remove the 2π discontinuities; and then determine the fine relative phase in each patch (a measured modulo 2π) by using the phase-shifting method. Thus, by integrating the gray code and phase-shifting methods, the GCPS method achieves high accuracy and removes the 2π ambiguity (Sansoni et al., 1999).

In addition to the GCPS method, an alternative way to resolve the above phase ambiguity problem is the multiple-wavelength phase-shifting method (Towers et al., 2005; Reich et al., 1997) as shown in figure 2. In this method, at least two different phase shifting patterns with wavelengths λ_a and λ_b are used to distinguish the phase shifting order by comparing the phase difference in an extended range with an equivalent wavelength λ_{ab}, which can be specified as:

$$\lambda_{\mathrm{ab}} = \frac{\lambda_{\mathrm{a}}\lambda_{\mathrm{b}}}{\lambda_{\mathrm{b}} - \lambda_{\mathrm{a}}} \tag{2}$$

The extended wavelength leads to a unique phase distribution for the entire image without loss of accuracy.

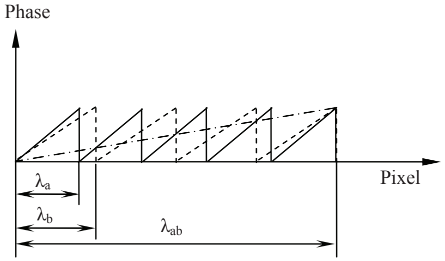


Fig. 2. Phase shifting with multiple-wavelength

However, both the GCPS method and the multiple-wavelength phase-shifting method require more structured light patterns, which will sacrifice the measurement speed. Meanwhile, these methods can only be used to measure stationary parts; otherwise, the sensor may capture non-corresponding pattern codes due to the displacement of the inspected target, resulting in inaccurate 3D shape measurement.

To reduce the number of patterns, a feasible solution is to integrate multiple phase shifting patterns into a single composite pattern for real-time measurement (Guan et al., 2003), at the expense of measurement accuracy. Another commonly used one-shot pattern strategy is based on the Fourier Transform Profilometry (FTP) (Takeda & Mutoh, 1983), in which a single pattern is projected and analyzed in the spatial frequency domain. It should be noted that the spectrum aliasing phenomena will affect the measurement accuracy. It should be mentioned that a common problem of the phase shifting methods is their susceptibility to sensor and environment noise.

In the direct-coding approach, to achieve a pixel level resolution, each pixel should have a unique color value in the same pattern. Thus, a great number of colors are required. Moreover, the captured color by the camera does not only depends on the color of the projected pattern, but also relies on the color of the scanned surface. Thus, this direct-coding technique is very susceptible to noise and ambient light and is inappropriate for quality inspection.

In the spatial-neighborhood approach, the codeword of the primitive is specified by its own value and the values of its adjacent primitives. Thus, this technique can be implemented in one-shot pattern for real-time 3D profile measurement. The most commonly used primitives are color and geometry. Some of the color-based patterns are colored slit pattern, colored stripe pattern, colored grid pattern, colored spot pattern, etc (Tehrani et al., 2008; Pages et al., 2004; Je et al., 2004; Salvi, 1998; Payeur, 2009). The codeword of the each primitive is usually formulated under the constraint of De Bruijn sequence(Pages et al., 2004), pseudorandom sequence(Payeur, 2009) or M-arrays(Salvi, 1998). As a well-known type of mathematical sequence, the De Bruijn sequence of order m with q different symbols is a circular sequence with the length of q^m, where each subsequence of length m exactly emerges once. Thus, each subsequence can be uniquely identified in the entire sequence. Similarly, a pseudorandom sequence is generated in the same way without the subsequence formed by 0. It is noted that both of the above two methods are one-dimension spatial coding approaches, whereas M-arrays are the two-dimension coding strategy. Assume that the total number of primitives in the pattern is m $\times$ n, and then the sub-window of u $\times$ v appears only once for M-arrays coding strategy. Examples of the geometry-based patterns are given in (Doignon, 2005).

Besides, the temporal monochromatic black/white stripe pattern is also usually adopted for high speed 3D shape measurement. The black/white pattern has the following advantages: first, the pattern identification is very easy and fast due to the simple image processing; second, the measurement is very reliable because of the robustness to the varied reflection property and ambient light. A temporal stripe coded pattern uses four different patterns for binary boundary code, and generates $2^8 = 256$ codes. Then, only 111 available codes are employed to avoid decode error(Rusinkiewicz, 2002).

Recently, several other coding strategies for real-time measurement have been reported. The black/white stripes combined with traversed color lines are used to form one-shot patterns. In this method, the epipolar geometry constraint is used to decode the intersections between the stripe boundaries and the color lines(Koninckx & Gool, 2006). A single stripe pattern is proposed to reconstruct the 3D human face. To clarify the index of each stripe, an algorithm based on the maximum spanning tree of a graph is used to identify the potential connectivity and the adjacency in recorded stripes(Brink et al., 2008).

For accurate, reliable and fast measurements of industrial parts (e.g., automotive parts), the projection pattern is supposed to meet the following requirements:
(a) high robustness to the reflectivity variance of the measured part;
(b) high consistence of the measurement performance;
(c) appropriate point cloud density to represent the 3D shape;
(d) accurate location of the primitives;

(e) rapid decoding capability.

Motivated by these facts, we developed two novel structured light patterns for rapid 3D measurement(Xu et al., 2010; 2011), inspired by previous research.

The first one, X-point pattern, is a one-shot pattern based on geometrical feature and neighboring information. The primitive of this pattern is the corner of the black/white chessboard pattern. Compared with traditional geometric primitives, such as disc and stripe, the primitive of the X-point pattern is more reliable and accurate in detecting the primitive's location. The value of the primitive is represented by the direction of the X-point. This X-point pattern can be used for real-time 3D shape measurement thanks to its one-shot nature.

The second one, the two-level binary pattern strategy, makes use of both the temporal and spatial coding to reduce the number of required patterns. In this method, the value of the stripe boundary (primitive) is determined by the intensity variance in time domain. Then, the codeword of the primitive is calculated by using its own value and the values of the neighboring primitives in space domain. This is the reason why this method is termed as "two-level pattern" in this chapter.

2.2 X-point pattern coding strategy

The X-point pattern is based on the black/white binary pattern, through which the system robustness can be enhanced by removing the influence of the color property of the inspected parts. Only the geometrical feature can be used to distinguish different primitives in the pattern when using this method. The concept of the X-point method is derived from the chessboard, which is usually used in camera calibration due to the accurate positioning of corner points. Thus, the X-point method is very accurate for 3D measurement. The value of the primitive is represented by its orientation. As shown in figure 3, the corresponding values of the four primitives are denoted as 0, 1, 2, and 3, respectively. The angle between the orientation of the primitive and the horizontal line are 0, 45, 90, and 135 degrees, respectively.

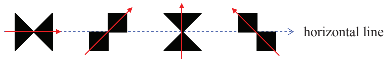


Fig. 3. The primitive design

Apparently, these four primitives are inadequate to remove the ambiguity in the pattern. To solve this problem, the neighboring coding information should be integrated to obtain much more amount of codeword. A straight-forward solution is to use both the value of a primitive and those of its eight neighboring primitives, as shown in figure 4. In this case, the pattern is able to recognize the $4^9 = 262,144$ unique primitives. Therefore, the maximum-allowable number of points in the proposed pattern is 262,144 in theory.

Another benefit of the X-point method is to decrease occlusion influences. As shown in figure 5, a primitive located on the edge of the inspected part usually leads to partial loss of the geometrical shape. However, it is evident that the primitive can still be detected by using the proposed method, resulting in improved system performance.

2.3 Two-level binary pattern coding strategy

Similarly, the two-level coding strategy is also based on the black/white pattern, to improve reliability. Furthermore, the pattern is a three-step pattern, i.e. , the number of patterns is three for 3D profile measurement. In this approach, the codeword of the primitive is determined by

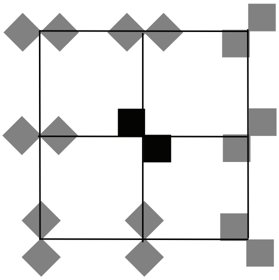


Fig. 4. The codeword based on 8 neighbor

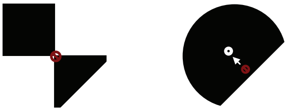


Fig. 5. An example of occlusion

its own value and those of the neighboring primitives. The method to generate the boundary value (represented by intensity variation in time domain) of two adjacent stripes is presented as follows.

In theory, the maximum possible number of intensity variance for each stripe over time is eight. In this chapter, the values are represented by 000, 001, 010, 011, 100, 101, 110, and 111, respectively. The value 001 means that the intensity of the stripe is switched in the order of white, black, and black over time. In this chapter, the values 000 and 111 are discarded to remove the influence of the reflectivity of the inspected part. In other words, the intensity of the stripe should change at least once during the measurement. Therefore, six remaining values 001, 010, 011, 100, 101, and 110 are used for coding the stripe. In order to achieve sub-pixel accuracy of stripe boundary detection, the location is specified by using the inverse intensity stripe, as shown in figure 6. A, B, C and D represent the intensity values of the successive pixels n and $n+1$ around the stripe boundary. The accurate location of a stripe boundary can be obtained as:

$$P = P_{n+1} - \frac{D - B}{(A + D) - (B + C)} \tag{3}$$

It is clear that the error is constrained within one pixel by using the above approach.

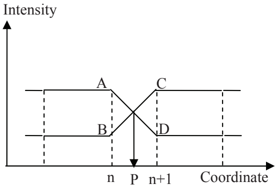


Fig. 6. The edge detection with inverse intensity

The above stripe boundary detection strategy imports another constraint for the configuration of adjacent stripes, where the intensity is supposed to vary twice in the space domain. In this case, assuming that one stripe is 001, the next stripe can only be selected from 010, 100, 110. Thus, the possible number for the arrangement of connected stripes is 3 × 6 = 18. The potential stripe boundaries are listed in figure 7

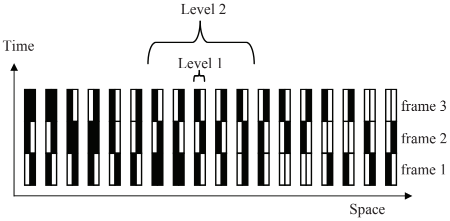
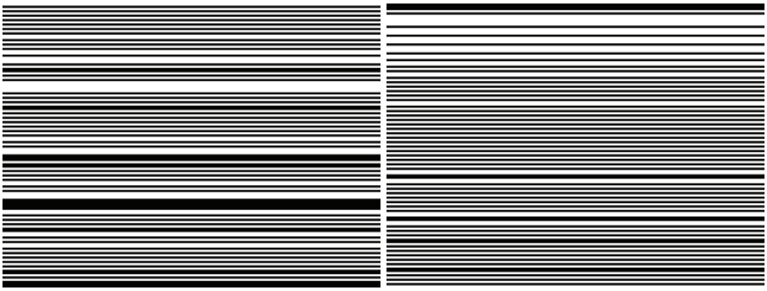
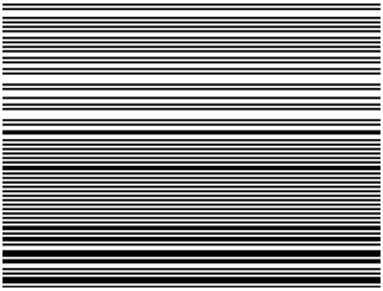


Fig. 7. The potential stripe boundaries



(a) Pattern 1 (b) Pattern 2



(c) Pattern 3

Fig. 8. The three two-level patterns

The second step is to form the codeword for each stripe boundary by using a series of successive stripes in space. Thus, the two-level pattern is essentially a sequence of stripe patterns. The codeword of each stripe boundary is determined by a successive subsequence. To uniquely locate each stripe boundary in a single pattern, the subsequence length n should be specified. Without loss of generality, we assume that the possible number of the first stripe is 6 while the possible number of the second one is 3. Similarly, there are 3 options for each stripe in the remaining adjacent stripes. Therefore, the number of unique subsequence is $6 \times 3^{n-1}$. For instance, if the length of the subsequence is 4, then 162 unique subsequences can be formulated under the above constraints. The subsequence can be generated by using

Fleury's algorithm, which is described in detail in (Xu et al., 2011). 128 unique subsequences are selected to form the illuminated patterns for 3D profile measurement, The pattern resolution is 768 × 1024 pixels, where the width of each strip is 6 pixels as shown in figure 8.

3. Phase-height mapping strategy

The phase-height mapping approach is the critical step of active vision, which converts the phase distribution in the camera image plane to the corresponding coordinate of the inspected object. The existing methods for transforming the phase to coordinate can be categorized into two types: absolute height method and relative height method.

In a stereo vision sensor, the projector is considered as an inverse camera, thus, both the camera and the projector can be represented by the pinhole model. When the distortion of the lens is ignored, the relationship between a point of the scanned object in the world frame and the corresponding pixels in the camera or projector image plane can be uniformly expressed:

$$sI = A \left[R \ t \right] X \tag{4}$$

where $I = \left[r \ c \ 1 \right]^T$ is the homogeneous coordinate of any arbitrary pixel in the image frame of the camera or projector; $X = \left[x \ y \ z \ 1 \right]^{\mathrm{T}}$ is the homogeneous coordinate of the corresponding point in the world frame; s is a scale factor; $\left[R \ t \right]$ is the extrinsic parameters representing the 3 × 3 rotation matrix and 3 × 1 translation vector from the world frame to the image frame; A is the intrinsic parameters matrix which is written as:

$$A = \begin{bmatrix} \alpha \ \gamma \ r_0 \\ 0 \ \beta \ c_0 \\ 0 \ 0 \ 1 \end{bmatrix} \tag{5}$$

Where r_0 and c_0 are the coordinates of the principle point; α and β are the focal length along two image axes of the image plane; γ is the skew parameter of the two image axes. Further, Eq.(4) can be represented by using the perspective projection matrix:

$$s \begin{bmatrix} r \\ c \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} \ m_{12} \ m_{13} \ m_{14} \\ m_{21} \ m_{22} \ m_{23} \ m_{24} \\ m_{31} \ m_{32} \ m_{33} \ m_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{6}$$

where $\begin{bmatrix} m_{11} \ m_{12} \ m_{13} \ m_{14} \\ m_{21} \ m_{22} \ m_{23} \ m_{24} \\ m_{31} \ m_{32} \ m_{33} \ m_{34} \end{bmatrix} = A \left[R \ t \right] = M$ is is the perspective projection matrix, which is utilized to map a 3D point in the word frame to a 2D point in the image plane.

Next, we eliminate the homogeneous scale s in Eq. (6) and obtain the general formula for both the camera and projector as:

$$\begin{cases} r = \frac{m_{11}x + m_{12}y + m_{13}z + m_{14}}{m_{31}x + m_{32}y + m_{33}z + m_{34}} \\ c = \frac{m_{21}x + m_{22}y + m_{23}z + m_{24}}{m_{31}x + m_{32}y + m_{33}z + m_{34}} \end{cases} \tag{7}$$

Without loss of generality, any pixel in the camera image plane generates a viewing ray-line through the optical center of the camera in the world frame. Then, we can obtain the viewing ray-line equation by using the following liner equation:

$$\begin{cases} r^c = \frac{m^c_{11}x+m^c_{12}y+m^c_{13}z+m^c_{14}}{m^c_{31}x+m^c_{32}y+m^c_{33}z+m^c_{34}} \\ c^c = \frac{m^c_{21}x+m^c_{22}y+m^c_{23}z+m^c_{24}}{m^c_{31}x+m^c_{32}y+m^c_{33}z+m^c_{34}} \end{cases} \tag{8}$$

Moreover, we can get the absolute phase distribution when the stripe pattern (i.e., two-level binary pattern) or the phase shifting pattern is adopted. In this case, the corresponding pixel in the projector image plane must lie on a line with the same phase value. To be specific, we assume that the line is a horizontal line with coordinate c_p, thus, the line forms a projecting ray-plane through the optical center of the projector in the world frame, intersecting the scanned surface.

$$c^p = \frac{m^p_{21}x + m^p_{22}y + m^p_{23}z + m^p_{24}}{m^p_{31}x + m^p_{32}y + m^p_{33}z + m^p_{34}} \tag{9}$$

Obviously, the scanned surface point is the intersection of the ray-plane and the ray-line as shown in figure 9 by using linear equations bringing the Eq.(8) and Eq.(9) together. So this method is referred to as the plane-line mapping approach in this chapter.

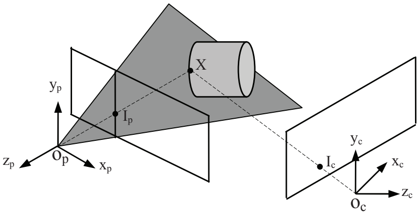


Fig. 9. The intersection of the plane and line

Actually, the projector is regarded as an inverse camera, since it projects images instead of capturing them. Consequently, both the camera and the projector have the same mathematic model such that the epipolar constraint is also satisfied by the projector and the camera. As shown in figure 10, point X is a measured point of the distorted stripe boundary on the inspected part. Point I_c is the projection of X in the camera image plane; while point I_p is the corresponding point of X in the projector image plane. Thus, the point I_p is restricted to lie on the epipolar line l_p due to the constraint of the epipolar geometry in stereo vision:

$$l_p = F \cdot I_c \tag{10}$$

where F is the fundamental matrix determined by calibration.

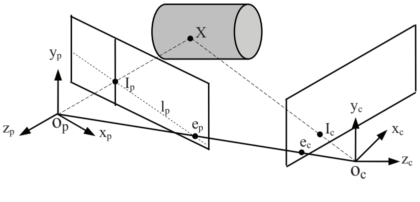


Fig. 10. The intersection of the line and line

When the stripe pattern or the phase shifting pattern is used, the corresponding pixel I_p for pixel I_c in the camera plane is the intersection the epipolar line l_p and the line with the phase equal to that of I_c in the projector plane.

Similarly, if the two-dimension coding pattern (i.e, X-point pattern), providing the location in the two axes of projector image plane, is adopted, then the pixel I_p can be directly obtained without the help of epipolar geometry.

Once the corresponding pixels are determined, we can get the a projecting ray-line through the optical center of the projector as:

$$\begin{cases} r^c = \frac{m^c_{11}x+m^c_{12}y+m^c_{13}z+m^c_{14}}{m^c_{31}x+m^c_{32}y+m^c_{33}z+m^c_{34}} \\ c^c = \frac{m^c_{21}x+m^c_{22}y+m^c_{23}z+m^c_{24}}{m^c_{31}x+m^c_{32}y+m^c_{33}z+m^c_{34}} \end{cases} \tag{11}$$

In this case, the calculation of the 3D surface point is converted to a line-line intersection problem. Combining Eq.(8) and Eq.(11), we can get the coordinate of the surface point. The equation can be solved by using the least-squares method, however, due to errors in the pinhole model of the camera and projector, the two ray-lines will not intersect in the 3D space. A better way is to compute the closest approach of the two skew lines, i.e., the shortest line segment connecting them. If the length of this segment is less than the threshold, we assign the midpoints as the intersection of the two lines; if it is larger than the threshold, we assume that there are some mistakes for the correspondence. The elaborated description for this method can be found in (Shapiro, 2001).

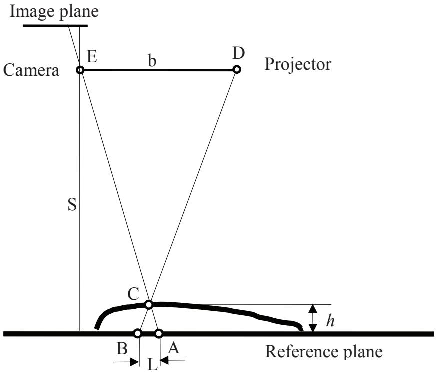


Fig. 11. The relative height caculation

Instead of measuring the absolute coordinate, the relative height variation is more emphasized in quality inspections. So another method is to obtain the relative height with respect to the reference plane using the triangular similarity method. As shown in figure 11, from the similar triangles ΔABC and ΔCDE , the relative height h from the surface to the reference plane can be calculated by

$$h = \frac{L \times S}{b + L} \tag{12}$$

where b denotes the baseline distance between the optical centers of the camera and projector; S is the standoff distance between the reference plane and the optical center of the camera; L is the distance of two corresponding pixels A and B. Eq.(12) can be further rewritten by using the pixel coordinate as:

$$h = \frac{res \times m \times S}{b + res \times m} \tag{13}$$

in which *res* is the resolution of the camera with a unit of mm/pixel and m signifies the number of pixels from A to B.
Furthermore, a simplified calculation of the relative height can be directly expressed as the production of a coefficient and the phase when the stand off distance S is much larger than the relative height h. (Su et al., 1992)

4. Calibration

The key procedure to guarantee accurate profile reconstruction of the inspected object is the proper calibration of the components of the active stereo vision, involving camera, projector, and system calibration(Li & Chen, 2003). The fundamental difference between the passive stereo vision and the active stereo vision is that one camera is replaced by a projector, leading to time-consuming and complicated calibration procedure for the reason that the projector cannot directly view the scene. To solve this problem, the projector is treated as an inverse camera. Then, we can calibrate the camera and the projector separately. We fist calibrate the camera and then determine the correspondence between the pixels in the projector and those in the calibration gauge using the camera(Zhang & Huang, 2006). In this case, the projector can be calibrated by using a similar technique as camera calibration. To be specific, the calibration of intrinsic and extrinsic parameters of both camera and projector can be implemented by using in the online Matalab toolbox.
Other methods involve neural networks, bundle adjustment, or absolute phase. However, these traditional calibration methods for active stereo vision treat both the camera and the projector as pin-hole models. A pinhole model is an ideal mathematical model where all the incident light rays go through a single point. However, a calibration residual error always exist when using a pinhole model, especially for affordable off-the-shelf equipment.
In this chapter, a pixel-to-pixel calibration concept has been adopted to improve system accuracy. For this technique, pixel-wise correspondence between the projector and the camera is established, instead of using the unique transformation matrix as in the approach mentioned above. Therefore, the significant merit is to improve the measurement accuracy because of the elimination of residual error of the sensor calibration. Additionally, another advantage is to avoid the projector calibration, which appears more tedious and complicated since the projector cannot view the calibration gauge in the scene.
From Eq. (13), the motivation of the active stereo vision sensor calibration is to obtain the parameters S, b and *res*. First, we will explain how to calibrate the parameter (S, b) for each couple of corresponding points in the pixel-to-pixel calibration approach.
In figure 12, the points D_i and E_i are corresponding pixels belonging to the same physical point C_i, where E_i, a virtual point, perhaps out of the image plane of the camera, is the intersection of two lines: the reflective ray of light C_iE_i , and the baseline D_iE_i parallel to the reference plane. In this case, a set of sensor parameter matrices $\left(b_{(i,j)}, S_{(i,j)}, i = 1, 2 \cdots m; j = 1, 2 \cdots n\right)$ are required to be calculated for each point on the projector image plane, where i, j are image coordinate indices of the correspondences in camera. A group of $L_{n(i,j)}$ can be calculated while the reference plane is moved to different heights for n times. Consequently, the sensor parameters $b_{(i,j)}$ and $S_{(i,j)}$ are computed using a linear least squares approach:

$$\begin{bmatrix} h_{1(i,j)} & -L_{1(i,j)} \\ \cdots & \cdots \\ h_{n(i,j)} & -L_{n(i,j)} \end{bmatrix} \cdot \begin{bmatrix} b_{(i,j)} \\ S_{(i,j)} \end{bmatrix} = \begin{bmatrix} -h_{1(i,j)} \times L_{1(i,j)} \\ \cdots \\ -h_{n(i,j)} \times L_{n(i,j)} \end{bmatrix} \tag{14}$$

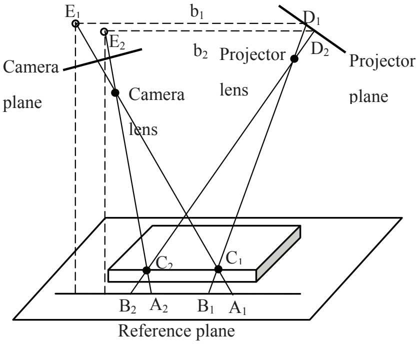


Fig. 12. Calibration of parameter (S, b)

So, Eqn.(12) can be rewritten as

$$h_{(i,j)} = \frac{L_{(i,j)} \times S_{(i,j)}}{b_{(i,j)} + L_{(i,j)}} \tag{15}$$

The remaining parameter *res* can be obtained by counting the pixels of a line with known length in the image. Next, the distance L can be further determined by using the calibrated parameter *res*.

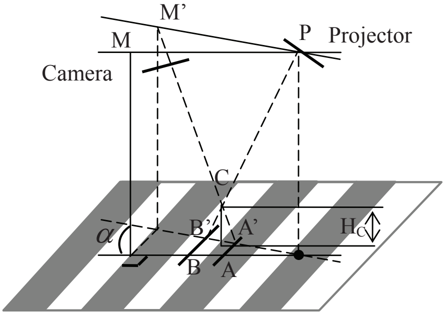


Fig. 13. Calibration of the offset angle α

The previous discussion focuses on the calibration the baseline distance b and standoff S provided that the L is accurately measured. However, inaccurate placement of the camera and projector will generate a system offset angle α, resulting in error of L, As shown in figure 13, point C is the point on the surface of the inspected object, with relative height H_C to the reference plane. AB and $A'B'$ are the projected lines of MP and $M'P$ on the reference plane, respectively. The calculated line AB is perpendicular to the stripe boundary. The angle between the calculated line AB and actual line $A'B'$ is called the offset angle α, which has to be calibrated. When using a pixel-to-pixel calibration method, M is assumed point corresponding to P. However actually M' is the real point corresponding to point P. $\Delta M'CP$ and $\Delta A'CB'$ are similar triangles and hence the distance L' between point A' and B' should be used to compute H_C. However, the patterns used are encoded along the image rows, which means codes are identical in one dimension. If α is not calibrated, instead of L', L will be