

## Chapter 2

# Active Vision Sensors

This chapter presents the sensing fundamentals, measurement principles, and 3D reconstruction methods for active visual sensing. An idea of sensor reconfiguration and recalibration is also described which endows a robot with the ability of actively changing its sensing parameters according to practical scenes, targets, and purposes. These will be used in the next chapters in formulating the methods of sensor reconfiguration and sensor planning.

### 2.1 3D Visual Sensing by Machine Vision

Similar to human perception, machine vision perception is one of the most important ways for acquiring knowledge of the environment. The recovery of the 3D geometric information of the real world is a challenging problem in computer vision research. Active research in the field in the last 30 years has produced a huge variety of techniques for 3D sensing. In robotic applications, the 3D vision technology allows computers to measure the three-dimensional shape of objects or environments, without resorting to physically probing their surfaces.

#### 2.1.1 Passive Visual Sensing

One class of visual sensing methods is called passive visual sensing where no other device besides cameras is required. These methods were usually developed at the early stage of computer vision research. By passive, no energy is emitted for the sensing purpose and the images are the only input data. The sensing techniques were often supposed to reflect the way that human eyes work. The limited equipment cost constitutes a competitive advantage of passive techniques compared with active techniques that require extra devices.

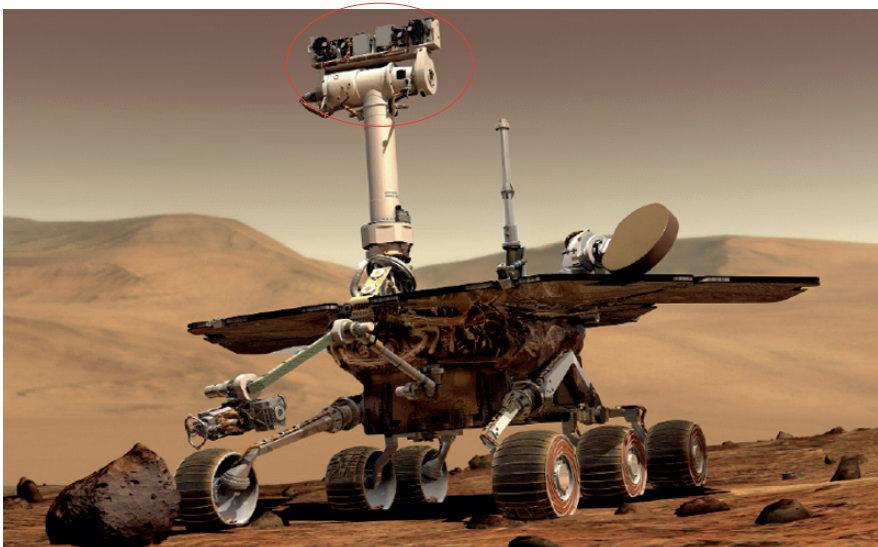
Such passive techniques include stereo vision, trinocular vision (Lehel et al. 1999, Kim 2004, Farag 2004), and many monocular shape-from-X techniques, e.g. 3D shape from texture, motion parallax, focus, defocus, shadows, shading, specularities, occluding contours, and other surface discontinuities. The problem is that recovering 3D information from a single 2D image is an ill-posed problem (Papadopoulos 2001). Stereo vision is still the single passive cue that gives

reasonable accuracy. Human has two eyes, and precisely because of the way the world is projected differently onto the eyes, human is able to obtain the relative distances of objects. The setup of a stereo machine vision system also has two cameras, separated by a baseline distance  $b$ . The 3D world point may be measured by the two projection equations, in a way that is analogous to the way the human eyes work. To interpret disparity between images, the matching problem must be solved, which has been formulated as an ill-posed problem in a general context and which anyway is a task difficult to automate. This correspondence problem results in an inaccurate and slow process and reduces its usefulness in many practical applications (Blais 2004). The other major drawback of this passive approach is that it requires two cameras and it cannot be used on un-textured surfaces which are common for industrially manufactured objects. The requirement of ambient light conditions is also critical in passive visual sensing. The advantage of stereo vision is that it is very convenient to implement and especially suitable for natural environments. A few applications are illustrated in Figs. 2.1 to 2.3.

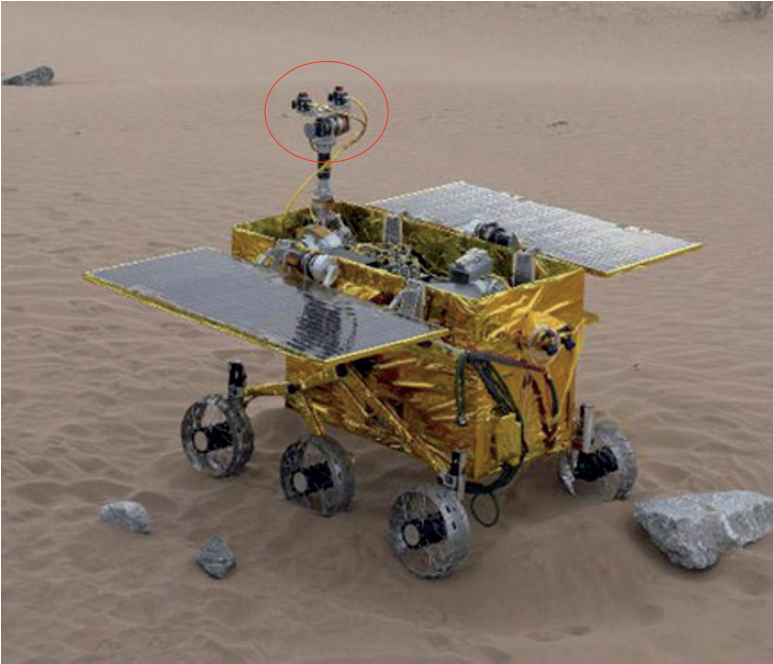
The structure-from-motion algorithms solve the following problem: given a set of tracked 2D image features captured by a moving camera, find the 3D positions and orientations of the corresponding 3D features (structure) as well as the camera motion. Pose estimation, on the other hand, solves the problem of finding the position and orientation of a camera given correspondences between 3D and 2D features. In both problems two-dimensional line features are advantageous because they can be reliably extracted and are prominent in man-made scenes. Taylor and Kriegman (1995) minimized a nonlinear objective function with respect to camera rotation, camera translation and 3D lines parameters. The objective function measures the deviation of the projection of the 3D lines on the image planes from the extracted image lines. This method provides a robust solution to the high-dimensional non-linear estimation problem. Fitzgibbon and Zisserman (1998) also worked towards the automatic construction of graphical models of scenes when the input was a sequence of closely spaced images. The point features were matched in triples of consecutive images and the fundamental matrices were estimated from pairs of images. The projective reconstruction and camera pose estimation was upgraded to a Euclidean one by means of auto-calibration techniques (Pollefeys et al. 1998). Finally, the registration of image coordinate frames was based on the algorithm of iterative closest points (Besl and McKay 1992).



**Fig. 2.1.** Stereo vision for industrial robots



**Fig. 2.2.** Mars Rover in 3D (NASA mission in 2003–2004) (Pedersen 2003, Miller 2003, Madison 2006, Deen and Lore 2005)



**Fig. 2.3.** MR-2 (Prototype of Chinese Moon Explorer in 2007–2008)

### 2.1.2 Active Visual Sensing

In contrast to passive visual sensing, the other class of visual sensing techniques is called active visual sensing. For the above cases of passive techniques (that use ambient light), only visible features with discernable texture gradients like on intensity edges are measured. For the example of the stereo setup, there is a corresponding problem. Matching corresponding points is easy if the difference in position and orientation of the stereo views is small, whereas it is difficult if the difference is large. However, the accuracy of the 3D reconstruction tends to be poor when the difference in position and orientation of the stereo views is small. To overcome the shortcomings of passive sensing, active sensing techniques have been developed in the recent years. These active systems usually do not have the correspondence problem and can measure with a very high precision.

By active sensing, an external projecting device (e.g. laser or LCD/DLP projector) is used to actively emit light patterns that are reflected by the scene and detected by a camera. That is to say they rely on probing the scene in some way rather than relying on natural lighting. Compared with the passive approach, active visual sensing techniques are in general more accurate and reliable.

Generally active 3D vision sensors can resolve most of the ambiguities and directly provide the geometry of an object or an environment. They require minimal operator assistance to generate the 3D coordinates. However, with laser-based approaches, the 3D information becomes relatively insensitive to background illumination and surface texture. Therefore, active visual sensing is ideal for scenes that do not contain sufficient features. Since it requires lighting control, it is usually suitable for indoor environments and both camera and projector need to be pre-calibrated.

Typically, properly formatted light, or another form of energy, is emitted in the direction of an object, reflected on its surface and received by the sensor; the distance to the surface is calculated using triangulation or time-of-flight (Papadopoulos 2001). Typical triangulation-based methods include single/multi-point projection, line projection, fringe and coded pattern projection, and moire effect (Figs. 2.4–2.6). Typical time-of-flight based methods are interferometers and laser range finders.

Moire devices work on the principle that: effectively projecting a set of fringe patterns on a surface using an interference technique, tracking the contours of the fringes allows the range to be deduced. Systems that use point projection, line scanning, and moiré effect are highly accurate, but can be slow. Moire devices are best suited to digitizing surfaces with few discontinuities.

Interferometers work on the principle that: if a light beam is divided into two parts (reference and measuring) that travel different paths, when the beams are combined together interference fringes are produced. With such devices, very small displacements can be detected. Longer distances can also be measured with low measurement uncertainty (by counting wavelengths).

For laser range finders, the distance is measured as a direct consequence of the propagation delay of an electromagnetic wave. This method usually provides good distance precision with the possibility of increasing accuracy by means of longer measurement integration times. The integration time is related to the number of samples in each measurement. The final measurement is normally an average of the sample measures, decreasing therefore the noise associated to each single measure. Spatial resolution is guaranteed by the small aperture and low divergence of the laser beam (Sequeira et al. 1995, 1996, 1999). Basically laser range finders work in two different techniques: pulsed wave and continuous wave. Pulsed wave techniques are based on the emission and detection of a pulsed laser beam. A short laser pulse is emitted at a given frequency and the time elapsed between the emission and the received echo is measured. This time is proportional to the distance from the sensor to the nearest object. In a continuous wave laser ranging system, rather than using a short pulse, a continuous laser beam modulated with a reference waveform is emitted and the range is determined as a result of the comparison of the emitted and received laser beams. This type of system can use either amplitude modulation (e.g. sinusoidal signal) or frequency modulation.

Among various 3D range data acquisition techniques in computer vision, the structured light system with coded patterns is based on active triangulation. A very simple technique to achieve depth information with the help of structured light is to scan a scene with a laser plane and to detect the location of the reflected

stripe. The depth information can be computed out of the distortion along the detected profile. More complex techniques of structured light project multiple stripes (Fig. 2.7) or a pattern of grids at once onto the scene. In order to distinguish between stripes or grids they are coded either with different brightness or different colors (Fig. 2.8) (e.g. Coded Light Approach (Inokuchi et al. 1984, Stahs and Wahl 1992) and unique color encoding method). The structured light systems, as well as laser range finders, map directly the acquired data into a 3D volumetric model having thus the ability to avoid the correspondence problem associated with passive sensing techniques. Indeed, scenes with no textural details can be easily modeled. A drawback with the technique of coded stripes is that because each projection direction is associated with a code word, the measurement resolution is low. Fortunately, when this approach is combined with a phase-shift approach, a theoretically infinite height resolution can be obtained. For available products, Fig. 2.9 illustrates some examples of 3D laser scanners and Fig. 2.10 illustrates some examples of 3D Structured Light System.

Therefore, although there are many types of vision sensors available to measure object models by either passive or active methods, structured-light is one of the most important methods due to its many advantages compared with other methods, and thus it is successfully used in many areas for recovering 3D information of an industrial object. This chapter considers typical setups of the structured light system for active visual sensing, using stripe light vision or color-encoded vision. Their system configurations and measurement principles are presented in the following sections.

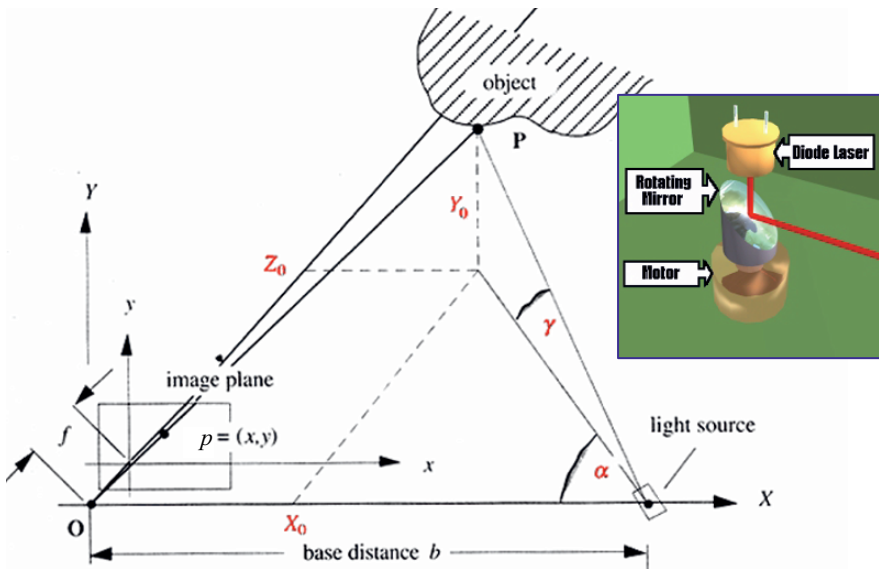
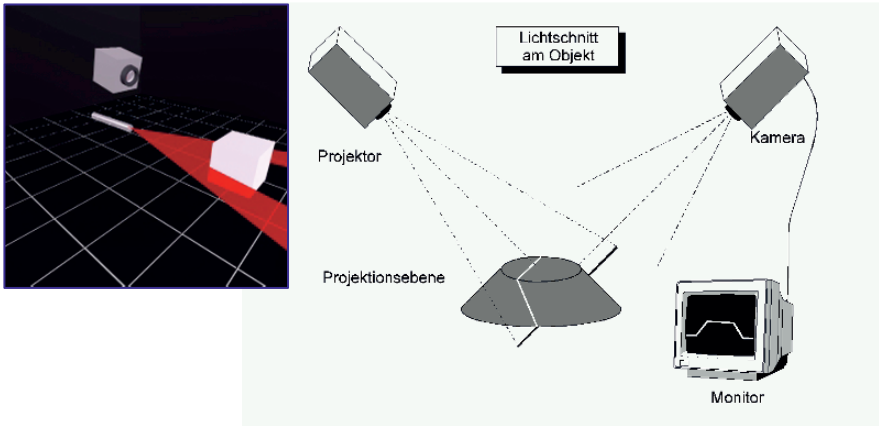
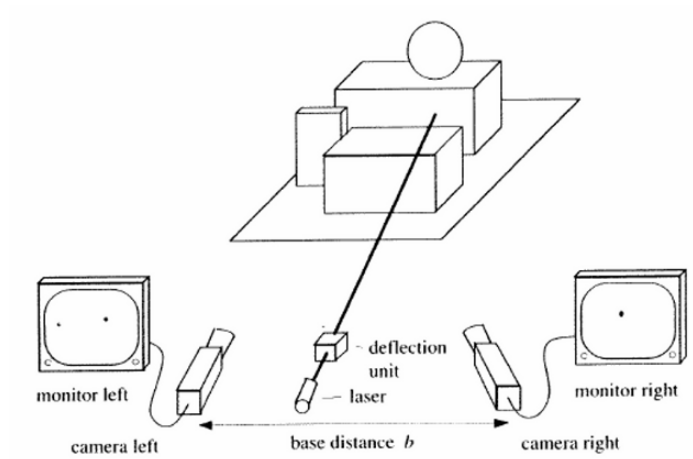


Fig. 2.4. Light spot projection



**Fig. 2.5.** A stripe light scanning system (Intersecting the projection ray with an additional ray or plane will lead to a unique reconstruction of the object point.)



**Fig. 2.6.** Single spot stereo analysis

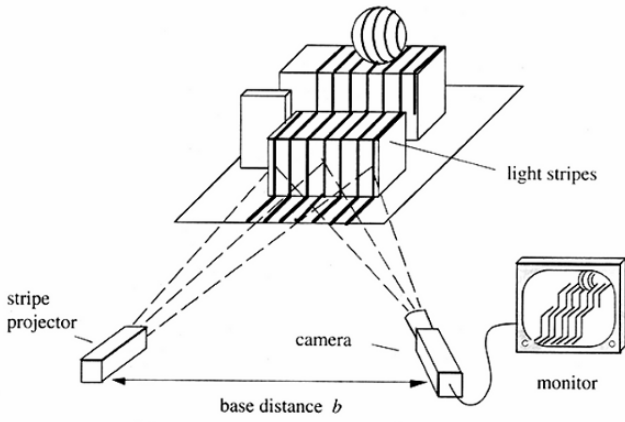


Fig. 2.7. Stripe light vision system

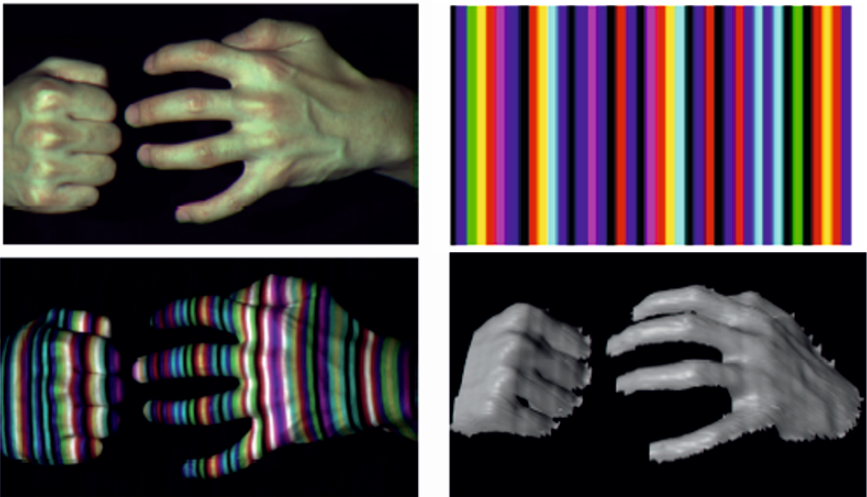
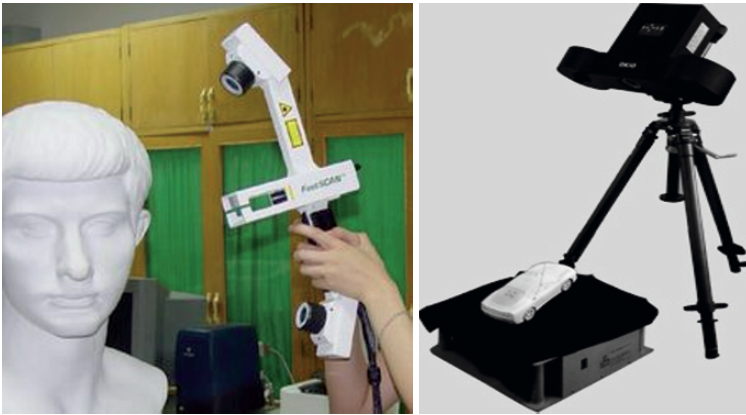


Fig. 2.8. Coded structured light vision: project a light pattern into a scene and analyze the modulated image from the camera





**Fig. 2.9.** Examples of 3D laser scanners



**Fig. 2.10.** Examples of 3D structured light system (FastScan and OKIO)

## 2.2 3D Sensing by Stereo Vision Sensors

### 2.2.1 Setup with Two Cameras

Binocular stereo vision is an important way of getting depth (3D) information about a scene from two 2-D views of the scene. Inspired by the vision mechanism of humans and animals, computational stereo vision has been extensively studied in the past 30 years, for measuring ranges by triangulation to selected locations imaged by two cameras. However, some difficulties still exist and have to be researched further. The figure illustrated below contains several examples of mobile

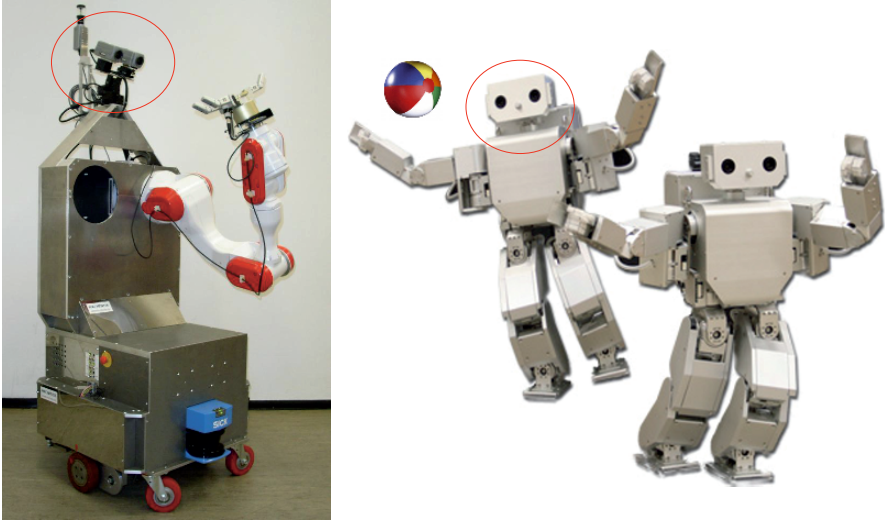


Fig. 2.11. The mobile robots with stereo vision setup at the University of Hamburg

robots that use stereo vision sensors for understanding the 3D environment, which are currently employed in our laboratory at the University of Hamburg (Fig. 2.11).

### 2.2.2 Projection Geometry

In a stereo vision system, the inputs to the computer are 2D-projections of the 3D object. The vision task is to reconstruct 3D world coordinates according to such 2D projected images, so we must know the relationship between the 3D objective world and 2D images (Fig. 2.12), namely the projection matrix. A camera is usually described using the pinhole model and the task of calibration is to confirm the projection matrix. As we know, there exists a collineation which maps the projective space to the camera's retinal plane:  $P^3 \rightarrow P^2$ . Then the coordinates of a 3D point  $\mathbf{X} = [X, Y, Z]^T$  in a Euclidean world coordinate system and the retinal image coordinates  $\mathbf{x} = [u, v]^T$  are related by the following (2.1).

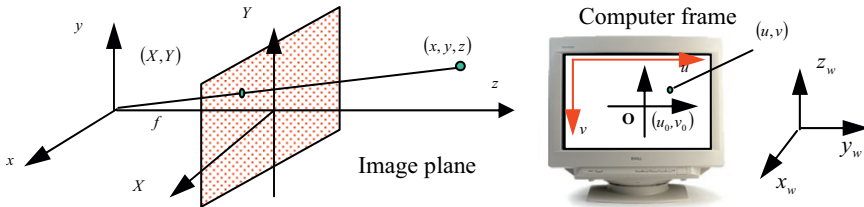


Fig. 2.12. The projection geometry: from 3D world to 2D image

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ & f_y & v_0 \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 \\ & & 1 \\ & & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^\top & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (2.1)$$

where  $\lambda$  is a scale factor,  $c = [u_0, v_0]^\top$  is the principal point,  $f_x$  and  $f_y$  are focal lengths,  $s$  is the skew angle, and  $\mathbf{R}$  and  $\mathbf{t}$  are external or extrinsic parameters.  $\mathbf{R}$  is the  $3 \times 3$  rotation matrix which gives axes of the camera in the reference coordinate system and  $\mathbf{t}$  the translation in the  $X$ ,  $Y$  and  $Z$  directions representing the camera center in the reference coordinate system (Henrichsen 2000). Of course, it is the same reference coordinate system in both views of the stereo couple.

$$\mathbf{R} = \mathbf{R}_x(\theta_x)\mathbf{R}_y(\theta_y)\mathbf{R}_z(\theta_z) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad (2.2)$$

$$\mathbf{t} = [T_x \ T_y \ T_z]^\top. \quad (2.3)$$

Equation (2.1) can be expressed as

$$\lambda \mathbf{x} = \mathbf{M} \mathbf{X}, \quad (2.4)$$

where  $\mathbf{x} = [u, v, 1]^\top$  and  $\mathbf{X} = [X, Y, Z, 1]^\top$  are the homogeneous coordinates of spatial vectors, and  $\mathbf{M}$  is a  $3 \times 4$  matrix, called the perspective projection, representing the collineation:  $P^3 \rightarrow P^2$  (Henrichsen 2000).

The first part of the projection matrix in the collineation (2.1), denoted by  $\mathbf{K}$ , contains the intrinsic parameters of the camera used in the imaging process. This matrix is used to convert between the retinal plane and the actual image plane. In a normal camera, the focal length mentioned above does not usually correspond to 1. It is also possible that the focal length changes during an entire imaging process, so that for each image the camera calibration matrix needs to be reestablished (denoted as *recalibration* in a following section).

### 2.2.3 3D Measurement Principle

If the value of the same point in computer image coordinate shoot by two cameras can be obtained, the world coordinates of the points can be calculated through the projection of two cameras (Fig. 2.13). Then four equations can be obtained from the two matrix formulas and the world coordinates of the point can be calculated (Ma and Zhang, 1998).

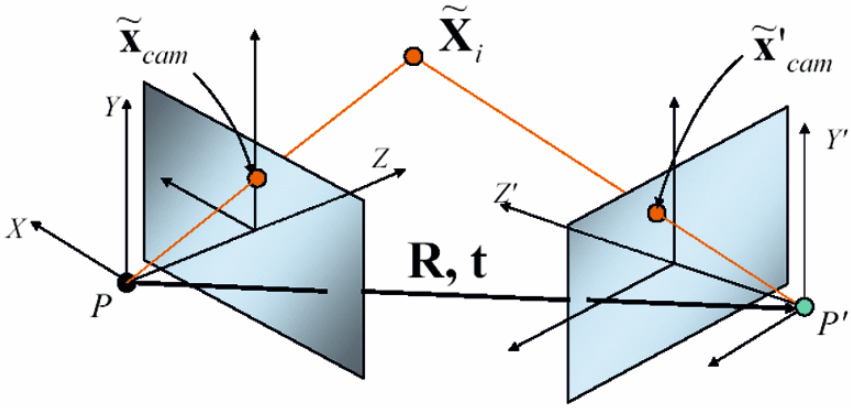


Fig. 2.13. 3D measurement by stereo vision sensors

For the two cameras, we have

$$\lambda_1 \mathbf{x}_1 = \mathbf{M}_1 \mathbf{X} \quad \text{and} \quad \lambda_2 \mathbf{x}_2 = \mathbf{M}_2 \mathbf{X}, \quad (2.5)$$

or

$$Z_{c1} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^1 & m_{12}^1 & m_{13}^1 & m_{14}^1 \\ m_{21}^1 & m_{22}^1 & m_{23}^1 & m_{24}^1 \\ m_{31}^1 & m_{32}^1 & m_{33}^1 & m_{34}^1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.6)$$

$$Z_{c2} \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^2 & m_{12}^2 & m_{13}^2 & m_{14}^2 \\ m_{21}^2 & m_{22}^2 & m_{23}^2 & m_{24}^2 \\ m_{31}^2 & m_{32}^2 & m_{33}^2 & m_{34}^2 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

The two uncertain numbers can be removed and (2.6) becomes

$$\begin{aligned} (u_1 m_{31}^1 - m_{11}^1)X + (u_1 m_{32}^1 - m_{12}^1)Y + (u_1 m_{33}^1 - m_{13}^1)Z &= m_{14}^1 - u_1 m_{34}^1 \\ (v_1 m_{31}^1 - m_{21}^1)X + (v_1 m_{32}^1 - m_{22}^1)Y + (v_1 m_{33}^1 - m_{23}^1)Z &= m_{24}^1 - v_1 m_{34}^1 \\ (u_2 m_{31}^2 - m_{11}^2)X + (u_2 m_{32}^2 - m_{12}^2)Y + (u_2 m_{33}^2 - m_{13}^2)Z &= m_{14}^2 - u_2 m_{34}^2 \\ (v_2 m_{31}^2 - m_{21}^2)X + (v_2 m_{32}^2 - m_{22}^2)Y + (v_2 m_{33}^2 - m_{23}^2)Z &= m_{24}^2 - v_2 m_{34}^2 \end{aligned} \quad (2.7)$$

or

$$\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \\ q_{41} & q_{42} & q_{43} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}, \quad (2.8)$$

$$\mathbf{Q}[X \ Y \ Z]^T = \mathbf{B}$$

In this linear system (2.8), the only three unknowns ( $X$ ,  $Y$ ,  $Z$ ) can be solved simply by:

$$[X \ Y \ Z]^T = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{B} \quad (2.9)$$

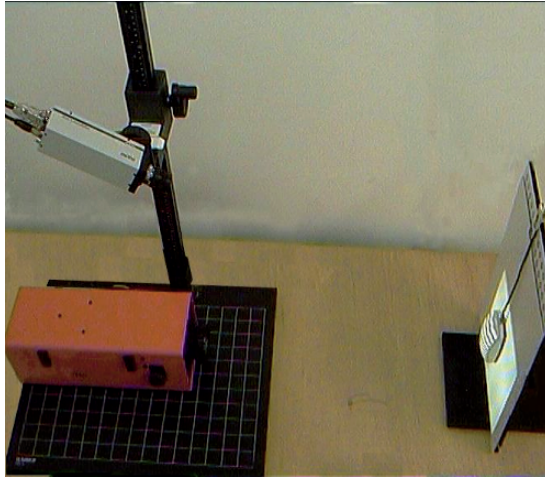
In practice, correction of lens distortion and epipolar geometry for feature matching should be considered for improving the efficiency and accuracy of 3D reconstruction. The epipolar plane contains the three-dimensional point of interest, the two optical centers of the cameras, and the image points of the point of interest in the left and right images. An epipolar line is defined by the intersection of the epipolar plane with image planes of the left and right cameras. The epipole of the image is the point where all the epipolar lines intersect. More intensive technology for stereo measurement is out of the scope of this book, but can be found in many published contributions.

## 2.3 3D Sensing by Stripe Light Vision Sensors

Among the active techniques, the structured-light system features high quality and reliability for 3D measurement. It may be regarded as a modification of static binocular stereo. One of the cameras is replaced by the projector which projects (instead of receives) onto the scene a sheet of light (or multiple sheets of light simultaneously). The simple idea is that once the perspective projection matrix of the camera and the equations of the planes containing the sheets of light relative to a global coordinate frame are computed from calibration, the triangulation for computing the 3D coordinates of object points simply involves finding the intersection of a ray (from the camera) and a plane (from the light pattern of the projector). A controllable LCD (Liquid Crystal Display) or DLP (Digital Light Processing) projector is often used to illuminate the surface with particular patterns. It makes it possible for all the surfaces in the camera's field of view to be digitized in one frame, and so is suitable for measuring objects at a high field rate.

### 2.3.1 Setup with a Switchable Line Projector

The active visual sensor considered in this section consists of a projector, which is a switchable LCD line projector in this study, to cast a pattern of light stripes onto the object and a camera to sense the illuminated area as shown in Fig. 2.14. The 3D measurement is based on the principle of triangulation. If a beam of light is cast, and viewed obliquely, the distortions in the beam line can be translated into height variations. The correspondence problem is avoided since the triangulation is carried out by intersecting the two light rays generated from the projector and seen by the camera.



**Fig. 2.14.** Setup of stripe light vision system

The projector is controllable by a computer to select a specific light pattern. All the patterns are pre-designed with light and dark stripes and switchable during the operation.

### 2.3.2 Coding Method

In structured light systems, the light coding method is used as a technique to solve the correspondence problem (Batlle et al. 1998, Salvi et al. 2004). 3D measurement by structured lighting is based on the expectation of precise detection of the projected light patterns in the acquired images. The 3D coordinates can be triangulated directly as soon as the sensor geometry has been calibrated and the light pattern is located in the image.

For such systems as shown in Fig. 2.14, a Coded Light Approach is most suitable for space-encoding and position detection when using a switchable LCD or mask. It is also an alternative approach for avoiding the scanning of the light and it requires only a small number of images to obtain a full depth-image. This can be achieved with a sequence of projections using a set of switchable lines (light or dark) on the LCD device. All the lines are numbered from left to right.

In a so-called gray-coding (Inokuchi et al. 1984, Stahs and Klahl 1992), adjacent lines differ by exactly one bit leading to good fault tolerance. Using a projector, all lines (e.g. 512 switchable lines) may be encoded with several bits. This can be encoded in 10 projected line images. One bit of all lines is projected at a time. A bright line represents a binary '0', a dark line a '1'. All object points illuminated by the same switchable line see the same sequence of bright and dark illuminations. After a series of exposures, the bit-plane stack contains the encoded number of the corresponding lines in the projector. This is the angle in encoded format. The angle  $\alpha$  is obtained from the column address of each pixel. Thus all the

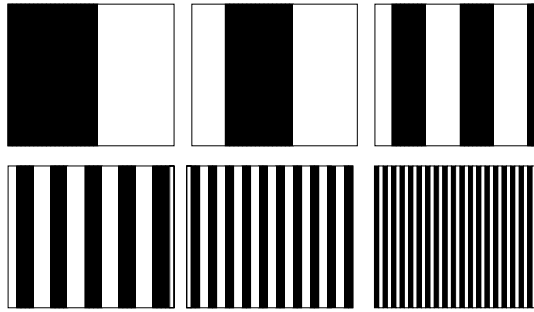


Fig. 2.15. An example of the coding method

information needed to do the triangulation for each pixel is provided by the x-address and the contents of the bit-plane stack. Using look-up-tables can generate a full 3D image within a few seconds. With such a setup, the depth resolution can be further increased using the phase-shift method or the color-coded method.

Figure 2.15 illustrates an example of the coding method. The lines are numbered from left to right. They are called Gray-Code, although they are in binary patterns. Using a controllable projector with  $2^n$  switchable lines, all lines may be encoded with  $n+1$  bits, and projected with  $n+1$  images.

### 2.3.3 Measurement Principle

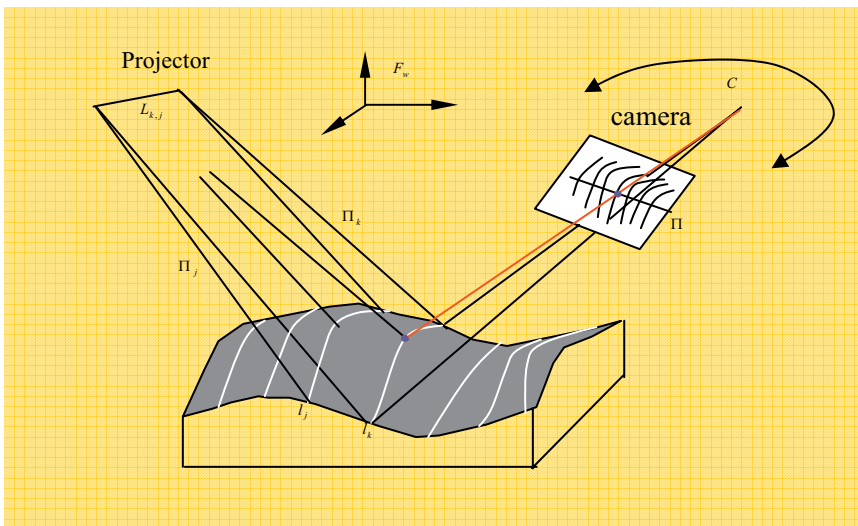


Fig. 2.16. The measurement principle

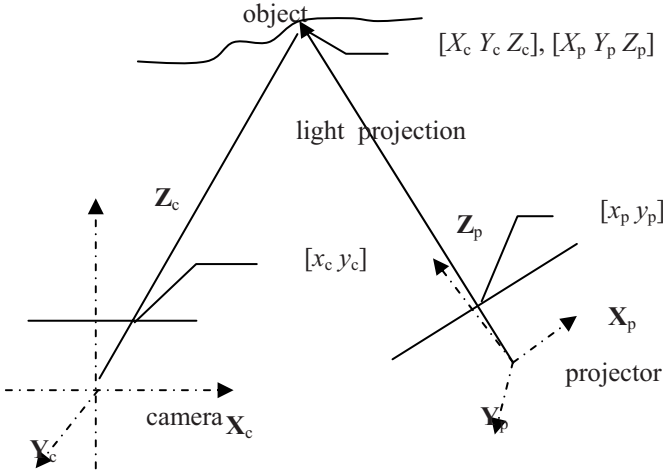


Fig. 2.17. The coordinates in the system

Figure 2.17 illustrates the measurement principle in the stripe light system and Fig. 2.18 illustrates the representation of point coordinates. For the camera, the relationship between the 3D coordinates of an object point from the view of the camera  $\mathbf{X}_c = [X_c \ Y_c \ Z_c \ 1]^T$  and its projection on the image  $\mathbf{x}_c = [\lambda x_c \ \lambda y_c \ \lambda]^T$  is given by

$$\mathbf{x}_c = \mathbf{P}_c \mathbf{X}_c, \tag{2.10}$$

where  $\mathbf{P}_c$  is a  $3 \times 4$  perspective matrix of the camera:

$$\mathbf{P}_c = \begin{bmatrix} v_x & k & x_{c0} & 0 \\ 0 & v_y & y_{c0} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.11}$$

Similarly, the projector is regarded as a pseudo-camera in that it casts an image rather than detects it. The relationship between the 3D coordinates of the object point from the vantage point of the projector  $\mathbf{X}_p = [X_p \ Y_p \ Z_p \ 1]^T$  and its back projection on the pattern sensor (LCD/DMD)  $\mathbf{x}_p = [\kappa x_p \ \kappa]^T$  is

$$\mathbf{x}_p = \mathbf{P}_p \mathbf{X}_p \tag{2.12}$$

where  $\mathbf{P}_p$  is a  $2 \times 4$  inverse perspective matrix:

$$\mathbf{P}_p = \begin{bmatrix} v_p & 0 & x_p^0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.13}$$



The relationship between the camera view and the projector view is given by

$$\mathbf{X}_p = \mathbf{M}\mathbf{X}_c, \quad \mathbf{M} = \mathbf{R}_\theta \mathbf{R}_\alpha \mathbf{R}_\beta \mathbf{T}, \quad (2.14)$$

in which  $\mathbf{R}_0$ ,  $\mathbf{R}_\alpha$ ,  $\mathbf{R}_\beta$ , and  $\mathbf{T}$  are  $4 \times 4$  matrices standing for 3-axis rotation and translation.

Substituting (2.14) into (2.12) yields

$$\mathbf{x}_p = \mathbf{P}_p \mathbf{M} \mathbf{X}_c. \quad (2.15)$$

$$\text{Let } \mathbf{H} = \mathbf{P}_p \mathbf{M} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}, \quad (2.16)$$

where  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are 4-dimensional row vectors. Equation (2.15) becomes

$$(x_p \mathbf{r}_2 - \mathbf{r}_1) \mathbf{X}_c = \mathbf{0}. \quad (2.17)$$

Combining (2.10) and (2.17) gives

$$\begin{bmatrix} \mathbf{P}_c \\ x_p \mathbf{r}_2 - \mathbf{r}_1 \end{bmatrix} \mathbf{X}_c = \begin{bmatrix} \mathbf{x}_c \\ 0 \end{bmatrix} \quad (2.18)$$

or

$$\mathbf{Q} \mathbf{X}_c = \mathbf{x}_{c+} \quad (2.19)$$

where  $\mathbf{Q}$  is a 4 by 4 matrix.

Then the three-dimensional world position of a point on the object surface can be determined by

$$\mathbf{X}_c = \mathbf{Q}^{-1} \mathbf{x}_{c+} \quad (2.20)$$

From the above equations, the 3D object can be uniquely reconstructed if we know the matrix  $\mathbf{Q}$  that contains 13 parameters from the two perspective matrices  $\mathbf{P}_c$  and  $\mathbf{P}_p$  and the coordinate transformation matrix  $\mathbf{M}$ . This means, once the perspective matrices of the camera and projector relative to a global coordinate frame are given from calibration, the triangulation for computing the 3D coordinates of object points simply involves finding the intersection of a ray from the camera and a stripe plane from the projector.

## 2.4 3D Sensor Reconfiguration and Recalibration

Since the objects may be different in sizes and distances and the task requirements may also be different for different applications, a structure-fixed vision sensor does not work well in such cases. A reconfigurable sensor, on the other hand, can change its structural parameters to adapt itself to the scene to obtain maximum 3D information from the environment. If reconfiguration occurs, the sensor should be capable of self-recalibration so that 3D measurement can follow immediately.

### 2.4.1 The Motivation for Sensor Reconfiguration and Recalibration

In an active visual system, since the sensor needs to move from one place to another for performing a multi-view vision task, a traditional vision sensor with fixed structure is often inadequate for the robot to perceive the object features in an uncertain environment as the object distance and size are unknown before the robot sees it. A dynamically reconfigurable sensor can help the robot to control the configuration and gaze at the object surfaces. For example, with a structured light system, the camera needs to see the object surface illuminated by the projector, to perform the 3D measurement and reconstruction task. Active calibration means that the vision sensor is reconfigurable during runtime to fit in the environment and can perform self-recalibration in need before 3D perception.

The concept of self-calibration in stereo vision and camera motion has been studied for more than ten years and there are many useful outputs. It is an attempt to overcome the problem of manual labor. For example, using the invariant properties of calibration matrix to motions, Dias et al. (1991) proposed an optimization procedure for recalibration of a stereo vision sensor mounted on a robot arm. The technique for self-recalibration of varying internal and external parameters of a camera was explored in (Zomet et al. 2001). The issues in dynamic camera calibration were addressed to deal with unknown motions of the cameras and changes in focus (Huang and Mitchell 1995). A method for automatic calibration of cameras was explored by tracking a set of world points (Wei et al. 1998). Such self-calibration techniques normally require a sequence of images to be captured via moving the camera or the target (Kang 2000). With some special setups, two views can also be sufficient for such a calibration (Seo and Hong 2001). All these are passive methods to calibrate the sensor with some varying intrinsic parameters.

For structured light vision systems, most existing methods are still based on static and manual calibration. That is, during the calibration and 3D reconstruction, the vision sensor is fixed in one place. The calibration target usually needs to be placed at several accurately known or measured positions in front of the sensor (DePiero and Trivedi 1996, Huynh 1997, Sansoni et al. 2000). With these traditional methods, the system must be calibrated again if the vision sensor is moved or the relative pose between the camera and the projector is changed. For the active vision system working in an unknown environment, changes of the position and configuration of the vision sensor become necessary. Frequent recalibrations in using such a system are tedious tasks. The recalibration means that the sensor has been calibrated before installation on the robot, but it needs to be calibrated again as its relative configuration is changing.

However, only a few related works can be found on self-calibration of a structured-light system. Furthermore, the self-calibration methods for a passive camera cannot be directly applied to an active vision system which includes an illumination system using structured light in addition to the traditional vision sensor. Among the previous self-calibration works on structured light systems, a self-reference method (Hébert 2001) was proposed by Hébert to avoid using the external calibrating device and manual operations. A set of points was projected on the scene and was detected by the camera to be used as reference in the calibration

of a hand-held range sensor. With a cubic frame, Chu et al. proposed a calibration-free approach for recovering unified world coordinates (Chu et al. 2001). Fofi et al. discussed the problem in self-calibrating a structured light sensor (Fofi et al. 2001). A stratified reconstruction method based on Euclidean constraints by projection of a special light pattern was given. However, the work was based on the assumption that “projecting a square onto a planar surface, the more generic quadrilateral formed onto the surface is a parallelogram”. This assumption is questionable. Consider an inclined plane placed in front of the camera or projector. Projecting a square on it forms an irregular quadrangle instead of a parallelogram as the two line segments will have different lengths on the image plane due to their different distances to the sensor. Jokinen’s method (Jokinen 1999) of self-calibration of light stripe systems is based on multiple views. The object needs to be moved by steps and several maps are acquired for the calibration. The registration and calibration parameters are obtained by matching the 3D maps via least errors. The limitation of this method is that it requires a special device to hold and move the object.

This chapter studies the problems of “self-calibration” and “self-recalibration” of active sensing systems. Here self-recalibration deals with situations where the system has been initially calibrated but needs to be automatically calibrated again due to a changed relative pose. Self-calibration refers to cases where the system has never been calibrated and none of the sensor’s parameters including the focal length and relative pose are known. Both of them do not require manual placements of a calibration target. Although the methods described later in this book are mainly concerned with the former situation, they can also be applied to the latter case if the focal lengths of the camera and the projector can be digitally controlled by a computer.

The remainder part of this section investigates the self-recalibration of a dynamically reconfigurable structured light system. The intrinsic parameters of the projector and camera are considered as constants, but the extrinsic transformation between light projector and camera can be changed. A distinct advantage of the method is that neither an accurately designed calibration device nor the prior knowledge of the motion of the camera or the scene is required during the recalibration. It only needs to capture a single view of the scene.

### 2.4.2 Setup of a Reconfigurable System

For the stripe light vision system (Fig. 2.16), to make it adaptable to different objects/scenes to be sensed, we incorporated two degrees of freedom of relative motion in the system design, i.e. the orientation of the projector (or the camera) and its horizontal displacement. This 2DOF reconfiguration is usually adequate for many practical applications. Where more DOFs are necessary, the recalibration issues will be addressed in the next section. In Fig. 2.18, the camera is fixed whereas the projector can be moved on a horizontal track and rotated around the  $y$ -axis. The  $x$ - $z$  plane is orthogonal to the plane of the projected laser sheet. The 3D

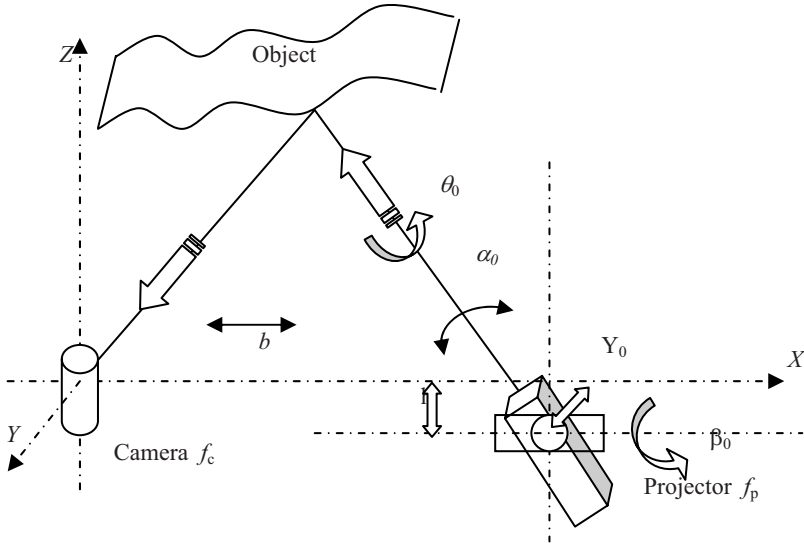


Fig. 2.18. A reconfigurable system

coordinate system is chosen based on the camera (or the projector) center and its optic axis.

For such a reconfigurable system, the two perspective matrices, which contain the intrinsic parameters such as the focal lengths ( $v_c, v_p$ ) and the optical centers ( $x_{c0}, y_{c0}, x_{p0}$ ), can be determined in the initial calibration stage. Based on the formulation in the previous section, the dynamic recalibration task is to determine the relative matrix  $\mathbf{M}$  in (2.14) between the camera and the projector. There are six parameters, i.e.

$$\mathbf{u} = [\theta \quad \alpha \quad \beta \quad X_0 \quad Y_0 \quad Z_0] \tag{2.21}$$

Since the system considered here has two DOFs, only two of the six parameters are variable while the other four are constants which can be known from the initial calibration. If the X-Z plane is not perpendicular to the plane of the projected laser sheet, its angle  $\Phi$  can also be identified at this stage. As the angle  $\theta_0 = (90^\circ - \Phi)$  is small and the image can be rectified by rotating the corresponding angle accordingly during recalibration, it can be assumed that  $\theta_0 = 0$ . The displacement in the y-direction between the camera center and the projector center,  $Y_0$ , and the rotation angle  $\beta_0$  are also small in practice. They do not affect the 3D reconstruction as the projected illumination consists of vertical line stripes. Therefore, we may assume that  $Y_0 = 0$  and  $\beta_0 = 0$ . Thus, the unknown parameters are reduced to only two ( $\alpha_0$  and  $b$ ) for the dynamic recalibration. Here  $h$  is a constant and  $\alpha_0$  and  $b$  have variable values depending on the system configuration.

For such a 2DOF system, the triangulation (2.18) for determining the 3D position of a point on the object surface is then simplified as (see Fig. 2.18)

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \frac{b - h \cot(\alpha)}{v_c \cot(\alpha) + x_c} \begin{bmatrix} x_c \\ y_c \\ v_c \end{bmatrix}, \quad (2.22)$$

where  $v_c$  is the distance between the camera sensor and the optical center of the lens,  $\alpha = \alpha(i) = \alpha_0 + \alpha_p(i)$  is the projection angle, and

$$\alpha_p(i) = \tan^{-1}\left(\frac{x_p(i)}{v_p}\right), \quad (2.23)$$

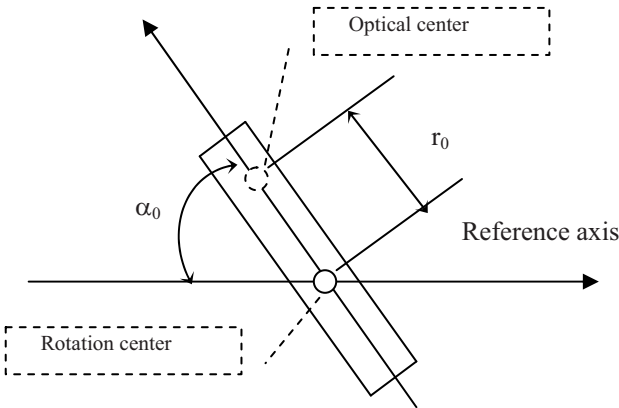
where  $i$  is the stripe index and  $x_p(i)$  is the stripe coordinate on the projection plane  $x_p(i) = i \times \text{stripe width} + x_p(0)$ .

If the projector's rotation center is not at its optical center,  $h$  and  $b$  shall be replaced by:

$$h' = h - r_0 \sin(\alpha_0) \quad \text{and} \quad b' = b - r_0 \cos(\alpha_0) \quad (2.24)$$

where  $r_0$  is the distance between the rotational center and the optical center (Fig. 2.19).  $h$  and  $r_0$  can be determined during the initial static calibration. Figure 2.20 illustrates the experimental device which is used to determine the camera's rotation axis.

For the reconfigurable system, the following sections show a self-recalibration method which is performed if and when the relative pose between the camera and the projector is changed. The unknown parameters are determined automatically by using an intrinsic cue, the geometrical cue. It describes the intrinsic relationship between the stripe locations on the camera and the projector. It forms a geometrical constraint and can be used to recognize the unknown parameters of the vision system.



**Fig. 2.19.** The case when the rotational center is not at the optical center

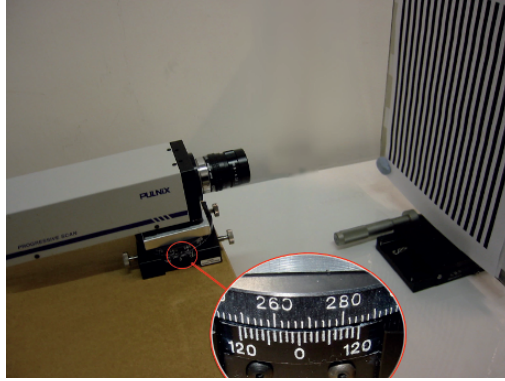


Fig. 2.20. The device to calibrate the rotational center

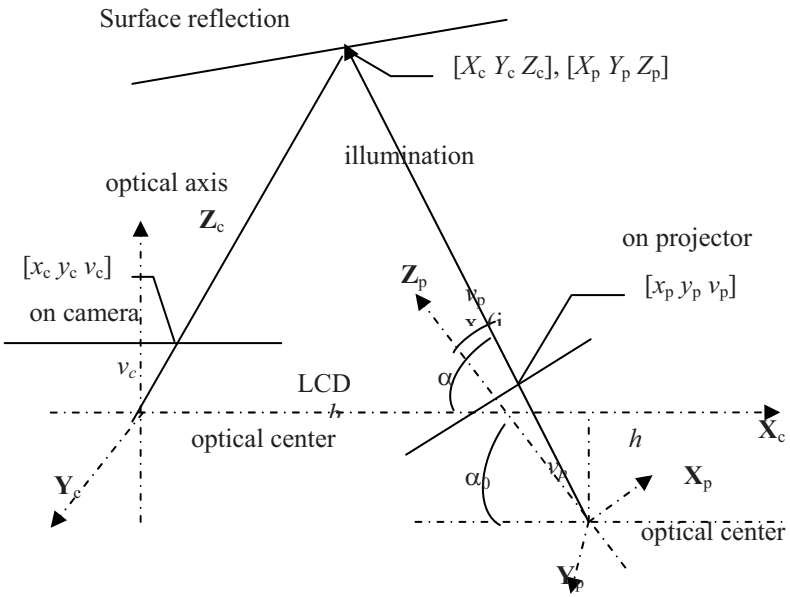


Fig. 2.21. The spatial relationship in the system

### 2.4.3 Geometrical Constraint

Assume a straight line in the scene which is expressed in the camera coordinate system and projected on the X-Z plane:

$$Z_c = C_1 X_c + C_2. \quad (2.25)$$

The geometrical constraint between projecting and imaging of the scene line is obtained by substituting (2.22) into (2.25):

$$[b - h \cot(\alpha)] (v_c - C_1 x_c) - C_2 [v_c \cot(\alpha) + x_c] = 0. \quad (2.26)$$

If  $h = 0$ , the above can be simplified as

$$b(v_c - C_1 x_c) - C_2 [v_c \cot(\alpha_0 + \alpha_{pi}) + x_c] = 0. \quad (2.27)$$

For two point pairs  $(x_{ci}, \alpha_{pi})$  and  $(x_{cj}, \alpha_{pj})$ ,

$$C_1 \left[ x_{cj} \frac{v_c \cot(\alpha_0 + \alpha_{pi}) + x_{ci}}{v_c \cot(\alpha_0 + \alpha_{pj}) + x_{cj}} - x_{ci} \right] = v_c \left[ \frac{v_c \cot(\alpha_0 + \alpha_{pi}) + x_{ci}}{v_c \cot(\alpha_0 + \alpha_{pj}) + x_{cj}} - 1 \right]. \quad (2.28)$$

Denote  $F_{ij} = \frac{v_c \cot(\alpha_0 + \alpha_{pi}) + x_{ci}}{v_c \cot(\alpha_0 + \alpha_{pj}) + x_{cj}}$ . The coordinates of four points yield

$$\frac{x_{cj} F_{ij} - x_{ci}}{x_{cl} F_{kl} - x_{ck}} = \frac{F_{ij} - 1}{F_{kl} - 1}. \quad (2.29)$$

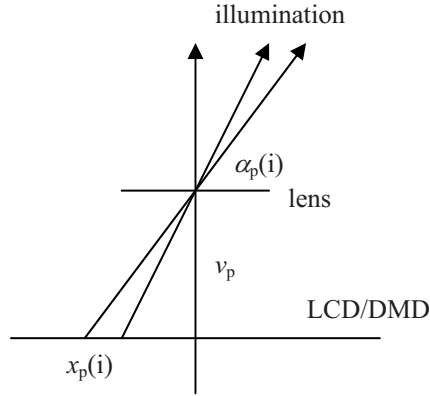
With (2.29), if the locations of four stripes are known, the projector's orientation  $\alpha_0$  can be determined when assuming  $h = 0$ .

If  $h \neq 0$ , from (2.27), the parameters  $v_c$ ,  $h$ , and  $v_p$  are constants that have been determined at the initial calibration stage.  $x_c = x_{ci} = x_c(i)$  and  $\alpha_{pi} = \alpha_p(i)$  are known coordinates on the sensors. Therefore,  $\alpha_0$ ,  $b$ ,  $C_1$ , and  $C_2$  are the only four unknown constants and their relationship can be defined by three points.

Denote  $A_0 = \tan(\alpha_0)$  and  $A_i = \tan(\alpha_{pi})$ . The projection angle of an illumination stripe is (illustrated in Fig. 2.22)

$$\cot(\alpha_0 + \alpha_{pi}) = \frac{1 - \tan(\alpha_0) \tan(\alpha_{pi})}{\tan(\alpha_0) + \tan(\alpha_{pi})} = \frac{1 - A_0 A_i}{A_0 + A_i} = \frac{v_p - A_0 x_p}{v_p A_0 + x_p}, \quad (2.30)$$

where  $x_p = x_p(i)$  is the stripe location on the projector's LCD. The x-coordinate value of the  $i$ th stripe,  $x_p(i)$ , can be determined by the light coding method. The stripe coordinate  $x_p$  and the projection angle  $\alpha_{pi}$  are illustrated in Figs. 2.21 and 2.22.



**Fig. 2.22.** The projection angle

Equation (2.27) may be written as

$$(bA_0 - C_2 - h)v_c v_p + (hC_1 - bC_1 A_0 - C_2 A_0)v_p x_c + (b + A_0 C_2 + hA_0)v_c x_p - (C_2 + bC_1 + hC_1 A_0)x_c x_p = 0 \quad (2.31)$$

or

$$x_c = W_1 + W_2 x_p + W_3 (x_c x_p), \quad (2.32)$$

where

$$W_1 = \frac{V_2 V_3 - V_1 V_4}{V_2 V_4} - \frac{V_3}{V_4}, \quad W_2 = -\frac{V_3}{V_2}, \quad \text{and} \quad W_3 = \frac{V_2}{V_4} \quad (2.33)$$

$$V_1 = bA_0 - C_2 - h, \quad (2.34)$$

$$V_2 = b + A_0 C_2 + hA_0, \quad (2.35)$$

$$V_3 = hC_1 - bC_1 A_0 - C_2 A_0, \quad (2.36)$$

$$V_4 = C_2 + bC_1 + hC_1 A_0, \quad (2.37)$$

Equation (2.32) is the relationship between the stripe locations on the camera and the projector and is termed the geometrical constraint.

#### 2.4.4 Rectification of Stripe Locations

Within a view of the camera, there can be tens or hundreds of stripes from the scene. The stripes' coordinates  $(x_c, x_p)$  on the image and the projector should satisfy (2.32)



in theory. In practice, however, the coordinates ( $x_c$ ) obtained from the image processing may not satisfy this constraint, due to the existence of noise. To reduce the effect of noise and improve the calibration accuracy, the stripe locations on the image can be rectified by using a curve fitting method.

Let the projection error be

$$Q_{err}(W_1, W_2, W_3) = \sum_{i=1}^m [x'_c(i) - x_c(i)]^2 = \sum_{i=1}^m [W_1 + W_2 x_p + W_3 (x_c x_p) - x_c]^2. \quad (2.38)$$

Then  $W_1$ ,  $W_2$ , and  $W_3$  may be obtained by minimizing the projection error  $Q_{err}$  with respect to  $W_k$

$$\frac{\partial Q_{err}}{\partial W_k} = 0, \quad k = 1, 2, 3. \quad (2.39)$$

Using (2.38) in (2.39) gives

$$\begin{bmatrix} m & \sum_{i=1}^m x_p(i) & \sum_{i=1}^m x_c(i)x_p(i) \\ \sum_{i=1}^m x_p(i) & \sum_{i=1}^m x_p^2(i) & \sum_{i=1}^m x_c(i)x_p^2(i) \\ \sum_{i=1}^m x_c(i)x_p(i) & \sum_{i=1}^m x_c(i)x_p^2(i) & \sum_{i=1}^m x_c^2(i)x_p^2(i) \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m x_c(i) & \sum_{i=1}^m x_p(i)x_c(i) & \sum_{i=1}^m x_p(i)x_c^2(i) \end{bmatrix}^T \quad (2.40)$$

$$\text{Or } \mathbf{GW} = \mathbf{X}, \quad \mathbf{W} = \mathbf{G}^{-1}\mathbf{X}. \quad (2.41)$$

The stripe location on the camera coordinate is, thus, rectified as

$$x'_c = \frac{W_1 + W_2 x_p}{1 - W_3 x_p}. \quad (2.42)$$

### 2.4.5 Solution Using the Geometrical Cue

Equation (2.31) can be written as

$$v_c v_p V_1 + v_p x_c V_2 + v_c x_p V_3 - x_p x_c V_4 = 0. \quad (2.43)$$

For an illumination pattern with  $n$  ( $n \geq 3$ ) stripes received on the image plane, (2.39) can be expressed as

$$\begin{bmatrix} v_p v_c & v_c A_1 & v_p X_1 & -A_1 X_1 \\ v_p v_c & v_c A_2 & v_p X_2 & -A_2 X_2 \\ & \dots & \dots & \\ v_p v_c & v_c A_n & v_p X_n & -A_n X_n \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = 0, \quad (2.44)$$

$$\text{or } \mathbf{A} \cdot \mathbf{V} = \mathbf{0}, \quad (2.45)$$

where  $\mathbf{A}$  is an  $n \times 4$  matrix,  $X_i = x_c(i)$ ,  $A_i = x_p(i)$ , and  $\mathbf{V}$  is a  $4 \times 1$  vector formed from (2.34) to (2.37). The following theorem is used for solving (2.44).

**Theorem 2.1** (The rank of the matrix  $\mathbf{A}$ )  $\text{Rank}(\mathbf{A}) = 3$ .

*Proof.* Consider the  $3 \times 3$  matrix  $\mathbf{A}_{lt}$  in the left-top corner of the  $n \times 4$  matrix  $\mathbf{A}$ . If  $\det(\mathbf{A}_{lt}) \neq 0$ , then  $\text{rank}(\mathbf{A}) \geq 3$  is true.

$$\mathbf{A}_{lt} = \begin{bmatrix} v_p v_c & v_c A_1 & v_p X_1 \\ v_p v_c & v_c A_2 & v_p X_2 \\ v_p v_c & v_c A_3 & v_p X_3 \end{bmatrix} \quad (2.46)$$

With row operations, it may be transformed to

$$\mathbf{A}_{lt}' = \begin{bmatrix} v_p v_c & v_c A_1 & v_p X_1 \\ 0 & v_c (A_2 - A_1) & v_p X_2 - v_p X_1 \\ 0 & 0 & v_p (X_3 - X_1) - v_p (X_2 - X_1) \frac{A_3 - A_1}{A_2 - A_1} \end{bmatrix}. \quad (2.47)$$

With (2.32) and (2.33), we have

$$U_2 = \frac{(V_2 V_3 - V_1 V_4)}{V_4^2} = \frac{v_c v_p (1 + A_0^2) C_2 (C_1 b + C_2 + h)}{(C_2 + b C_1 + h C_1 A_0)^2}. \quad (2.48)$$

Suppose that the observed line does not pass through the optical center of either the camera or the projector (otherwise it is not possible for triangulation), i.e.

$$C_2 \neq 0, \text{ and } C_1 b + C_2 + h = h - Z(0, b) \neq 0. \quad (2.49)$$

Hence  $U_2 \neq 0$ .

For any pair of different light stripes illuminated by the projector, i.e.  $A_i \neq A_j$ , from (2.32),

$$X_i = U_1 + \frac{U_2}{U_3 + f_p A_i} \quad (2.50)$$

we have  $X_i \neq X_j$ , and

$$\mathbf{A}_{lt}'(1,1) = v_c v_p \neq 0, \quad (2.51)$$

$$\mathbf{A}_{lt}'(2,2) = v_c (A_2 - A_1) \neq 0, \quad (2.52)$$

$$\begin{aligned} \mathbf{A}_{lt}'(3,3) &= v_p(X_3 - X_1) - v_p(X_2 - X_1) \frac{A_3 - A_1}{A_2 - A_1} = \\ &= \frac{v_p U_2 f_p^2 (A_1 - A_3)(A_2 - A_3)}{(U_3 + f_p A_1)(U_3 + f_p A_2)(U_3 + f_p A_3)} \neq 0 \end{aligned} \quad (2.53)$$

Hence,  $\text{rank}(\mathbf{A}) \geq \text{rank}(\mathbf{A}_{lt}) = \text{rank}(\mathbf{A}_{lt}') = 3$ .

On the other hand, rewrite the matrix  $\mathbf{A}$  with four column vectors, i.e.

$$\mathbf{A} = [c_{m1} \quad c_{m2} \quad c_{m3} \quad c_{m4}], \quad (2.54)$$

where

$$c_{m1} = [v_p v_c \quad v_p v_c \quad \dots \quad v_p v_c]^T, \quad (2.55)$$

$$c_{m2} = [v_c A_1 \quad v_c A_2 \quad \dots \quad v_c A_n]^T, \quad (2.56)$$

$$c_{m3} = [v_p X_1 \quad v_p X_2 \quad \dots \quad v_p X_n]^T, \quad (2.57)$$

$$c_{m4} = [-X_1 A_1 \quad -X_2 A_2 \quad \dots \quad -X_n A_n]^T. \quad (2.58)$$

With the fourth column,

$$\begin{aligned} c_{m4} = \{-X_i A_i\} &= \left\{ -\frac{U_2 + U_1 U_3}{f_p} - U_1 A_i + \frac{U_3}{f_p} X_i \right\} \\ &= \{\tau_1 v_c v_p + \tau_2 v_c A_i + \tau_3 v_p X_i\} = \tau_1 c_{m1} + \tau_2 c_{m2} + \tau_3 c_{m3}. \end{aligned} \quad (2.59)$$

This means that the matrix's 4<sup>th</sup> column,  $c_{m4}$ , has a linear relationship with the first three columns,  $c_{m1} - c_{m3}$ . So the maximum rank of matrix  $\mathbf{A}$  is 3, i.e.  $\text{rank}(\mathbf{A}) \leq 3$ .

Therefore, we can conclude that  $\text{rank}(\mathbf{A}) = 3$ .  $\square$

Now, considering three pairs of stripe locations on both the camera and the projector,  $\{(X_i, A_i) \mid i = 1, 2, 3\}$ , (2.44) has a solution in the form of

$$\mathbf{V} = k [v_1 \ v_2 \ v_3 \ v_4]^T, \quad k \in \mathbf{R}, \quad (2.60)$$

There exists an uncertain parameter  $k$  as the rank of matrix  $\mathbf{A}$  is lower than its order by 1. Using singular value decomposition to solve the matrix equation (2.44) to find the least eigenvalue, the optimal solution can be obtained. In a practical system setup, the  $z$ -axis displacement  $h$  is adjusted to 0 during an initial calibration, and (2.60) gives a solution for the relative orientation:

$$\begin{cases} A_0 = \frac{v_3}{v_4} \\ b_c = \frac{A_0 v_2 - v_1}{A_0 v_1 - v_2} \\ C_1 = (A_0 + b_c) \frac{v_4}{v_1} - b_c \end{cases}, \quad (2.61)$$

where  $b_c = C_2/b$ .

The orientation of the projector is

$$\alpha_0 = \tan^{-1}(A_0). \quad (2.62)$$

By setting  $b = 1$  and solving (2.33) and (2.60), the 3D reconstruction can be performed to obtain an object shape (with relative size). If we need to obtain the absolute 3D geometry of the object, (2.60) is insufficient for determining the five unknowns,  $b$ ,  $C_1$ ,  $C_2$ ,  $A_0$ , and  $k$ . To determine all these parameters, at least one more constraint equation is needed. In our previous work (Chen 2003), the focus cue or the best-focused distance is used below for this purpose.

## 2.5 Summary

Compared with passive vision methods which feature low-cost and are easy to set up, the structured light system using the active lighting features high accuracy and reliability. The stripe light vision system can achieve a precision at the order of 0.1 mm when using a high resolution camera and employing a good sub-pixel method. A specialized projector can be designed for achieving low cost and high accuracy. However, the limitation of the stripe light vision system is that it requires the scene to be of uniform color and static within the acquisition period. To reconstruct one 3D surface, it needs about one second to capture 8–12 images and several seconds to compute the 3D coordinates. The two methods of stereo vision and stripe light vision have both been summarized in this chapter.

To make the vision system flexible for perceiving objects at varying distances and sizes, the relative position between the projector and the camera needs to be changeable, leading to reconfiguration of the system. A self-recalibration method for such a reconfigurable system needs to be developed to determine the relative matrix between the projector and the camera, which is mainly concerned in this chapter. We thus presented a work in automatic calibration of active vision systems via a single view without using any special calibration device or target. This is also in the field of “self-calibration” and “self-recalibration” of active systems. Here self-recalibration deals with situations where the system has been initially calibrated but needs to be calibrated again due to a changed relative pose (the orientation and position) between the camera and projector. Self-calibration refers to cases where the system has never been calibrated and none of the sensor’s parameters including the focal length and relative pose are known. Although the method described in this chapter is mainly concerned with the former situation, it can also be applied to the latter case. Some important cues are explored for recalibration using a single view. The method will be applicable in many advanced robotic applications where automated operations entail dynamically reconfigurable sensing and automatic recalibration to be performed on-line without operators’ interference.