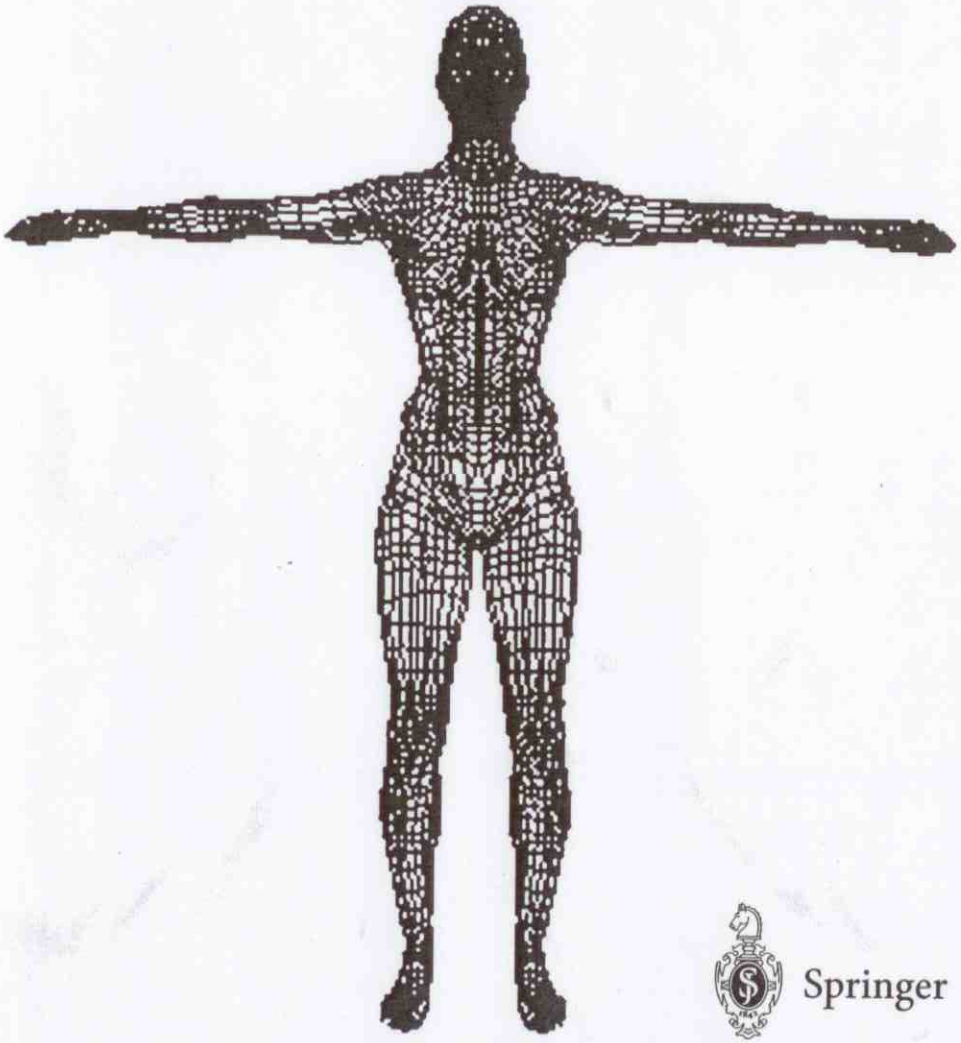
Reinhard Klette • Karsten Schlüns • Andreas Koschan

COMPUTER VISION

Three-Dimensional Data from Images



Springer

9 STRUCTURED LIGHTING

The projection of light patterns into a scene is called *structured lighting*. The light patterns are projected onto the objects which lie in the field of view of the camera. The distance of an object to the camera or the location of an object in space can be determined through analyzing the observed light patterns in the images. The active manipulation of the scene by using light patterns simplifies the 3D reconstruction task enormously as described in this chapter.

Notice that the location (x, y) of a pixel in the image (this is a grid square or grid rectangle) constrains the 3D location of the corresponding object point (X, Y, Z) to a certain sub-space in the scene. This sub-space contains all those scene points which project onto the grid square (x, y) and can be modeled as a four-sided infinite pyramid [1]. If the intrinsic parameters of the camera are known then this sub-space can be described with respect to the camera coordinate system. The apex of the pyramid lies at the projection center of the camera. The four edges of the pyramid pass through the corners of the grid square (x, y).

In the following we regard the grid square as a point which is defined as lying at the center of the grid square. The infinite pyramid shrinks to a ray ℓ in the scene. This ray ℓ can be interpreted as the projection ray corresponding to the object point (X, Y, Z) and the image point (x, y). Therefore, we can constrain the search space for (X, Y, Z) to ℓ which is one of the basic ideas of all structured lighting approaches.

A further general idea of structured lighting consists in intersecting the ray ℓ with an additional ray ℓ' or an additional plane Π which leads to a unique reconstruction of the object point (X, Y, Z). The goal of structured lighting is to introduce the ray ℓ' or the plane Π in such a way that a correspondence analysis is not necessary.

Note, in Chapter 4 the result of the correspondence analysis was a set of matched image points. Since the two matched points correspond to the same object point (X, Y, Z) the intersection of their projection rays leads to the reconstruction of (X, Y, Z). Thus, a second ray ℓ' was found by means of a correspondence analysis. In structured lighting the ray ℓ' or the plane Π are projected

[1] An infinite pyramid is a pyramid having no base.

actively into the scene, and it is assumed that their equations (or equivalent information) are available, e.g. by a geometrical calibration as presented in Chapter 2. For the sake of simplicity the light rays and planes are modeled by straight lines and plane equations, respectively. Hence, the diameter or thickness of the projected light patterns is normally not part of the mathematical models.

As a result of the above discussion structured lighting methods can be regarded as a modification of static binocular stereo. One of the cameras is replaced by a light source which projects the light pattern into the scene. The correspondence problem in the stereo vision pipeline (see the beginning of Chapter 4) does not exist any more since the triangulation is carried out by intersecting the projection ray (camera) and the light ray/plane (light source).

Almost all previously discussed 3D shape recovery methods assumed not very complex requirements for image acquisition. The solution of the reconstruction problem was found by analyzing theoretical, mathematical, and algorithmical issues. Structured lighting simplifies the task by increasing the engineering prerequisites, hence the complexity of the surface reconstruction task is shifted to another level in structured lighting.

Various shapes of light patterns exist, e.g. spot patterns, stripe patterns or color-coded patterns. The position, orientation and shape of the light patterns can be changed or remain static during the image acquisition process.

Structured lighting is especially applied in those fields where the automated three-dimensional measurement of an object has to be carried out with high precision. Of course, this class of techniques is restricted to environments allowing the active projection and detection of light patterns. Even outdoor scenes are recoverable without introducing artificial illumination. For example, sunlight in conjunction with a thin pole produces a stripe of shadow on the objects which yields equivalent information.

Laser light sources, special light stripe projectors, or slide projectors are employed to generate the light patterns. Structured lighting systems which are often called *3D scanners* or *range scanners* exist for a wide range of object sizes and applications. For example, small hand-held 3D scanners are available for reconstructing single teeth, and larger ones are used for "whole human body" reconstructions employing several light sources and cameras.

The 3D coordinates of the scene points in the images are recovered by assuming a known image acquisition geometry and using triangulation which is explained in Section 9.1.1. The methods of 3D object acquisition using structured lighting can be divided into methods which use simple geometric light patterns such as light spots or light stripes and methods which are based on spatial and/or temporal coding of the light patterns. Both classes will be introduced in the following sections.

9.1 PROJECTION OF SIMPLE GEOMETRIC PATTERNS

The motivation for using 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 acquisition geometry is known and the light pattern is located in the image. The simplest and best recognizable light patterns are light spots and light stripes.

In practice the precise detection of these simple geometric patterns is essential because they determine the achievable accuracy of the 3D reconstruction. One way to obtain a pixel accurate (or subpixel accurate) location of the patterns in the image is to calculate the first moments of those image segments which depict the projected light patterns (compare Exercise 3 in Section 2.5).

9.1.1 Light Spot Projection

The *light spot projection technique* is the simplest method to measure distances between the image acquisition system and points on object surfaces in the scene. A single light beam which is modeled as a ray (or line) is projected into the scene. The projection ray and the ray of the light beam are intersected to find the position of the illuminated object point in the 3D space.

(Task 9.1) Let us assume that an object surface is illuminated pointwise by a collimated laser beam and the reflected light is received by the camera. Determine the distance between the receiver and the illuminated object point using *triangulation*.

(Solution 9.1, 2D) For the sake of simplicity, at first the discussion is carried out for a plane, i.e. for a "2D object scene" having no vertical dimension (no Y-axis). This allows a simpler illustration of the fundamental principles. Object, light source, and camera lie in one plane (see Fig. 9.1). The goal is to recover the position of the object (i.e. the illuminated object point) in the plane. The angle α is given by the calibration and can be controlled using the deflection system of the laser. The angle β is defined by the projection geometry (compare Exercise (1) in Section 9.4) of the camera. The camera has to be regarded as an one-dimensional device since a 2D image acquisition set-up is assumed.

Let $\mathbf{O}$ be the projection center of the camera and the origin of the camera coordinate system. The base distance b is assumed to be constant and known. The distance d between the camera and the object point $\mathbf{P} = (X_0, Z_0)$ is calculated by using the law of sines:

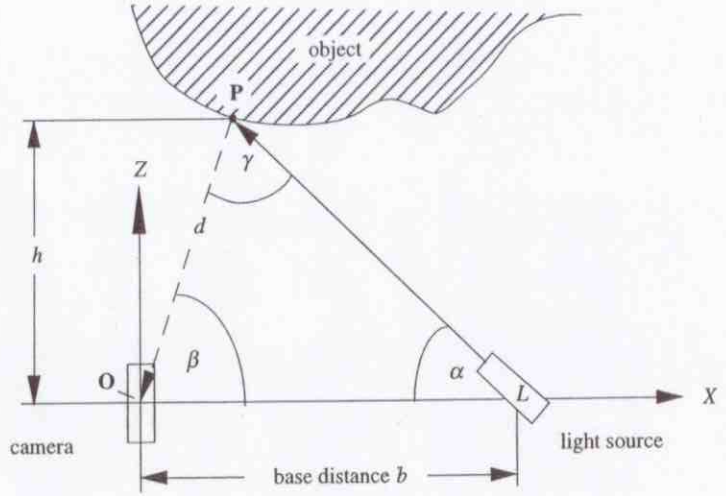


Figure 9.1: Triangulation with a light spot projector.

$$\frac{d}{sin(\alpha)} = \frac{b}{sin(\gamma)} .$$

From $\gamma = \pi - (\alpha + \beta)$ and $sin(\pi - \gamma) = sin(\gamma)$ it follows that

$$\frac{d}{sin(\alpha)} = \frac{b}{sin(\pi - \gamma)} = \frac{b}{sin(\alpha + \beta)} .$$

Thus, the distance d is given as

$$d = \frac{b \cdot sin(\alpha)}{sin(\alpha + \beta)} .$$

The location of the point $\mathbf{P} = (X_0, Z_0)$ can be represented in the camera coordinate system by the two-dimensional polar coordinates (d, β). The transformation into Cartesian coordinates is carried out by

$$X_0 = d \cdot \cos(\beta) \quad \text{and} \quad Z_0 = h = d \cdot \sin(\beta) .$$

The Z-axis coincides with the optical axis of the camera and the image plane lies at $Z = f$. Thus the distance coordinate Z can be determined using the angle β.

(Solution 9.1, 3D) The general case of a 3D triangulation requires to include the vertical dimension, in our case the Y-axis (compare Fig. 9.2). A camera centered

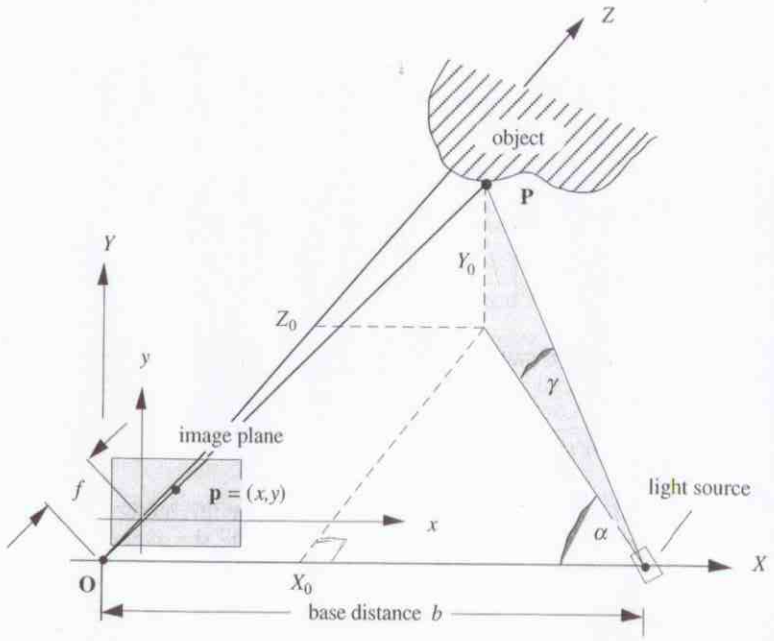


Figure 9.2: Illustration of the light spot projection technique in three dimensions.

XYZ-coordinate system and an image plane lying at $Z = f$ are assumed. The considered object point $\mathbf{P} = (X_0, Y_0, Z_0)$ is projected onto a point $\mathbf{p} = (x, y)$ in the image plane.

In the 3D case the camera and the light source can be arranged in an arbitrary way. Again, the Z-axis is the optical axis of the camera. We assume that the optical center of the light source is located on the X-axis. In contrast to the 2D case the laser beam is no longer restricted to the XZ-plane. The beam can be modeled by a segment of a straight line starting at point $(b, 0, 0)$ and ending at point $\mathbf{P}$.

The angles α and γ define the direction of the laser beam in 3D space (see Fig. 9.2). According to the ray theorem it holds that

$$\frac{X_0}{x} = \frac{Z_0}{f} = \frac{Y_0}{y} ,$$

for the chosen camera centered coordinate system. Using the trigonometry of right triangles it follows that

$$tan(\alpha) = \frac{Z_0}{b - X_0}$$

and

$$Z_0 = \frac{X_0}{x} \cdot f = tan(\alpha) \cdot (b - X_0)$$

resulting in

$$X_0 \left(\frac{f}{x} + tan(\alpha) \right) = tan(\alpha) \cdot b \, .$$

The 3D position of point **P** can be calculated by

$$X_0 = \frac{tan(\alpha) \cdot b \cdot x}{f + x \cdot tan(\alpha)} \, , \quad Y_0 = \frac{tan(\alpha) \cdot b \cdot y}{f + x \cdot tan(\alpha)} \, , \text{ and } \quad Z_0 = \frac{tan(\alpha) \cdot b \cdot f}{f + x \cdot tan(\alpha)} \, .$$

The distance $d_2(\mathbf{P}, \mathbf{O}) = \|(X_0, Y_0, Z_0)\|$ can be assigned to the range image at point (x, y). Notice that the angle γ is not used to calculate the scene point **P**. The reason is that the light beam is regarded as being a plane when the point **P** is calculated. This plane is oriented perpendicular to the XZ-plane (compare Fig. 9.2). The X-axis and this plane subtend the angle α.

(Algorithm 9.1) The procedure in Fig. 9.3 can be taken as a basis for calculating a range image (compare Section 3.2.1) by using the light spot technique.

```
calibrate image acquisition system;
begin
   for angle α:= α_min to α_max do
      for angle γ:= γ_min to γ_max do
         if ( p is visible and detectable ) then
            begin
               determine position p = (x, y);
               calculate coordinates X_0, Y_0, Z_0 for P;
               assign distance d_2(P, O) to point (x, y) of the range image
            end {if}
end
```

Figure 9.3: Generation of a range image employing the light spot projection technique.

(Comment 9.1) For the practical application of this technique the accuracy of the results is influenced by several factors. Problems in practice are

a) the limitation to incomplete range images or depth maps caused by shadows and invisibility, see Fig. 9.4, since a distance measurement is not possible inside the shaded areas when either the camera does not see this area or the laser point does not reach the object surface,

b) the speed of the measurement process which depends on the deflection system and the light spot detection,

c) the accurate calibration of the image acquisition system (see Chapter 2), and

d) the precise detection of the light spot in the image.

These influencing factors apply to a certain degree to the more advanced techniques described later in this chapter as well.

9.1.2 Light Spot Stereo Analysis

The method of *light spot stereo analysis* is based on the combination of the light spot technique with a method of static stereo analysis (compare Chapter 4). A laser spot is projected at a chosen location onto the object surface, and this spot is acquired with two cameras from different directions (compare Fig. 9.5).

Models of the geometry of a stereo image acquisition system as discussed in Chapter 4 can be used for the two cameras. The three-dimensional position of the spot in the scene can be determined analogously to the static stereo analysis namely based on the positions of the projected laser spot in the two image planes. Notice that for both methods, static stereo analysis and light spot projection, this

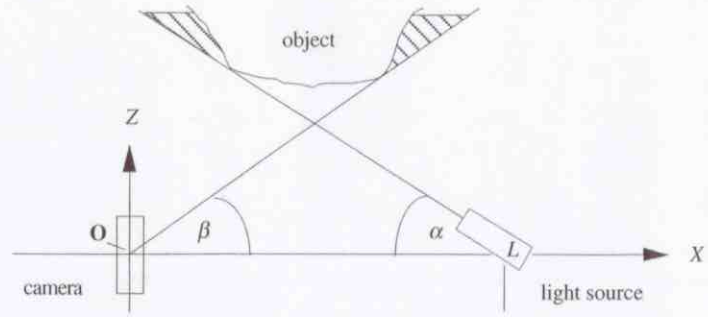


Figure 9.4: Visibility problems of the light spot projection technique.

determination of the 3D scene point is called triangulation since it is the same type of calculation from a geometrical point of view.

The advantage of the light spot stereo analysis technique is that the correspondence analysis is considerably simplified in both images compared to the static stereo analysis. The image of the laser spot on the object surface is assumed to be the brightest point in both images. Therefore, these two image points can be matched as corresponding points without any ambiguity problem.

The influence of the scene illumination on the laser point detection can be reduced by using special laser light filters for the image acquisition and by using image subtraction. For image subtraction the scene is taken at the beginning of the image acquisition process. This image is subtracted from the current image containing the laser spot. Under utilizing the epipolar geometry described in Section 4.1 the search space for the corresponding point in the second image can be reduced to an one-dimensional search interval along the epipolar line. Contrary to an "ordinary" method of the static stereo analysis a stereo image has to be generated and analyzed for each object point. Alternatively the detection of the laser spot can be achieved by using special hardware which provides directly the x- and y-position without a software based search process.

The advantage of the light spot stereo analysis technique compared to the light spot technique is that no calibration of the laser deflection system is necessary. The precision of the positioning of the laser beam does not influence

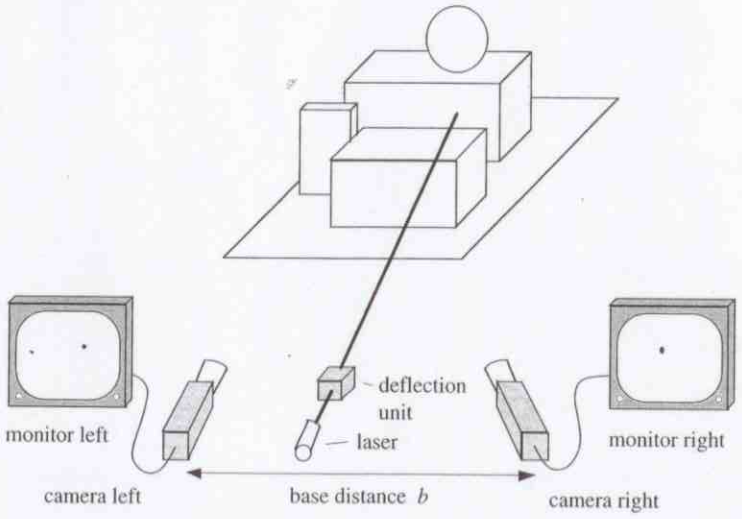


Figure 9.5: General arrangement for a method based on light spot stereo analysis.

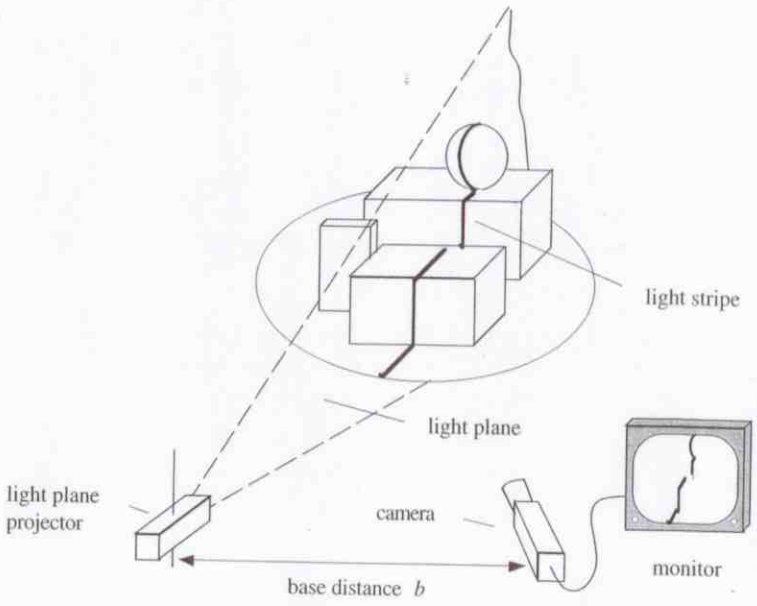


Figure 9.6: Image acquisition set-up for the light stripe projection technique.

the accuracy of the reconstruction because the orientation of the laser beam does not affect the calculation. Therefore, less requirements with respect to the light projection system are necessary which leads to a considerable reduction of the expenses for the realization of the image acquisition set-up.

For the software implementation of the technique a rough calibration of the deflection system is suggested because the search space in both images can be minimized with that rough knowledge of the geometry of the projected light spot in the image planes.

9.1.3 Light Stripe Projection

The *light stripe projection technique* or *light striping technique* represents an extension of the light spot projection technique (compare Section 9.1.1). This technique projects a light plane into the object scene (see Fig. 9.6). The idea is to intersect the projection ray of the examined image point with the light plane. The intersection of the light plane with the object surface is visible as a light stripe in the image. Therefore, a larger set of depth values can be recovered from a single

image which results in a faster reconstruction process compared to the single spot techniques. A light plane can be generated, for example, by using laser light in conjunction with a cylindrical lens, a laser plane projector, or a slide projector with a slit mask.

The use of laser light leads to a very bright and narrow light stripe which is important for an appropriate depth resolution. When employing a slide projector the slit has to be sufficiently wide so that the light stripe is sufficiently bright in the image. The slit mask is cheaper than the laser and in contrast to the laser no eye safety requirements are necessary.

(Task 9.2) Determine the distances between the illuminated object points in the scene and a reference plane using a light stripe projection technique. The reference plane is defined by the vertical axes (Y-axes) of the coordinate system of the light plane projector and the camera which are assumed to be parallel.

(Solution 9.2) The light plane is perpendicular to the reference plane and perpendicular to the plane spanned by the optical axes of the light plane projector and the camera. The optical axes are assumed to be coplanar but must not be parallel. The angle α_{opt} subtended by the optical axes will be estimated later. We assume that the camera image is digitized as a matrix of size $M \times N$ and that the

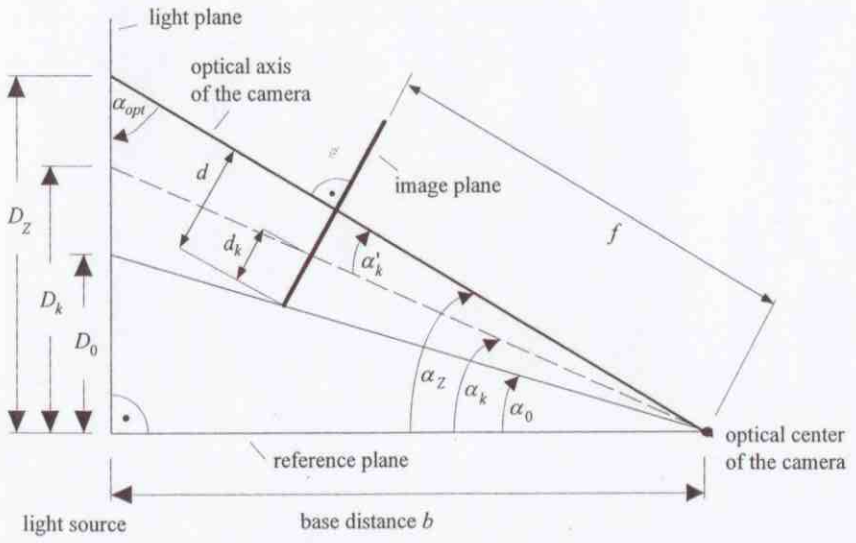


Figure 9.7: Schematic representation of a particular geometry for the light stripe technique which simplifies the calibration significantly.

column indices of this matrix are $x = 0, 1, \ldots, M-1$. The described technique follows an approach given in K.S. Fu et al. (1987).

As an initial step a calibration has to be carried out to calculate the depth values. A goal is to make the calibration process as simple as possible. Firstly, the length b is measured which is the distance between the optical centers of the light plane projector and the camera. Without using a special calibration procedure this can only be done approximately.

Let us define two angles α_Z and α_0. The angle α_Z which is subtended by the reference plane and the optical axis of the camera (compare Fig. 9.7) can be estimated as follows. It is assumed that a planar calibration object which is oriented parallel to the reference plane can be shifted along the optical axis of the light plane projector. Hence the calibration object stays parallel to the reference plane during the movement. The calibration object is shifted in such a way that the light stripe is projected onto the center column $x = M/2$ in the image.

Next, the distance D_Z between the calibration object and the reference plane is measured and the angle α_Z is determined by

$$\alpha_Z = arctan\left(\frac{D_Z}{b}\right). \tag{9.1}$$

The two optical axes subtend an angle of $\alpha_{opt} = 90° - \alpha_Z$.

The angle α_0 is subtended by the reference plane and the plane which is defined by the first (left) image column and the optical center of the camera. Notice that every image column can be modeled as a straight line segment in 3D space (compare Fig. 9.7). For the determination of the angle α_0 the distance between the calibration object and the reference plane is reduced until the light stripe is projected into column $x = 0$ in the image plane. The distance D_0 is measured, and it follows that α_0 is determined by

$$\alpha_0 = arctan\left(\frac{D_0}{b}\right). \tag{9.2}$$

As soon as the angles α_Z and α_0 are known it follows that the horizontal length $2 \cdot d$ of the image plane can be determined by

$$2 \cdot d = 2 \cdot f \cdot tan(\alpha_Z - \alpha_0), \tag{9.3}$$

where f is the effective focal length of the camera.

The length $2 \cdot d$ describes the width of the visible portion of the CCD-sensor. Equation (9.3) can be simplified by applying the addition formula

$$tan(x \pm y) = \frac{tan(x) \pm tan(y)}{1 \mp tan(x) \cdot tan(y)} \tag{9.4}$$

giving

$$tan(arc\,tan(x) \pm arctan(y)) = \frac{x \pm y}{1 \mp x \cdot y}$$

and resulting in

$$d = f \cdot \frac{(D_Z - D_0) \cdot b}{b^2 + D_Z \cdot D_0} \,. \tag{9.5}$$

It was assumed that the digital image has M columns hence the distance d_k between column k and column 0 can be calculated by

$$d_k = k \frac{d}{M/2} = \frac{2kd}{M}. \tag{9.6}$$

The angle α_k which is subtended by an arbitrary image column (the projection of an arbitrary stripe) and the reference plane (compare Fig. 9.7) can be easily determined since it holds that

$$\alpha_k = \alpha_Z - \alpha'_k \tag{9.7}$$

with

$$tan(\alpha'_k) = \frac{d - d_k}{f} \,.$$

Using equation (9.6) it follows that

$$\begin{aligned} tan(\alpha'_k) &= \frac{d \cdot (M - 2k)}{M \cdot f} \\ &= \frac{(D_Z - D_0) \cdot b \cdot (M - 2k)}{M \cdot (b^2 + D_Z \cdot D_0)} \end{aligned} \tag{9.8}$$

with $0 \le k \le M - 1$. Hence, if the point on the light stripe belonging to scene point **P** is detected in a column k then the angle α'_k and α_k can be recovered. Finally, the distance D_k between the reference plane and the scene point **P** is given by

$$\begin{aligned} D_k &= b \cdot tan(\alpha_k) \\ &= b \cdot tan\left(\alpha_Z - arctan\left(\frac{d \cdot (M - 2k)}{M \cdot f}\right)\right) \\ &= b \cdot tan\left(\alpha_Z - arctan\left(\frac{(D_Z - D_0) \cdot b \cdot (M - 2k)}{M \cdot (b^2 + D_Z \cdot D_0)}\right)\right) \end{aligned} \tag{9.9}$$

or

$$\begin{aligned} D_k &= \frac{D_Z - \dfrac{(D_Z - D_0) \cdot b^2 \cdot (M - 2k)}{M \cdot (b^2 + D_Z \cdot D_0)}}{1 + \dfrac{D_Z \cdot (D_Z - D_0) \cdot (M - 2k)}{M \cdot (b^2 + D_Z \cdot D_0)}} \\ &= \frac{(D_Z^2 + b^2)MD_0 + 2k(D_Z - D_0)b^2}{(D_Z^2 + b^2)M - 2k(D_Z - D_0)D_Z} \end{aligned} \tag{9.10}$$

for $0 \le k \le M - 1$. Equations (9.9) and (9.10) state that a distance can be calculated by detecting the image column of the projected scene point **P**. Notice that equation (9.7) does not contain the effective focal length f of the camera.

(Algorithm 9.2) The realization of the algorithm is trivial since it just consists in implementing equation (9.10). After the calibration is carried out the distance value for every column can be stored in a look-up table.

(Comment 9.2) An advantage of the above described arrangement is the simple way to calibrate the system and to determine the distance values. As a drawback we have the assumption that the vertical axes of the light plane projector and the camera are coplanar. Furthermore, the optical axes have to be coplanar and the calibration plane has to be moved exactly parallel to the reference plane. Deviations from these assumptions limit the accuracy.

Example 9.1: Figure 9.8 shows on the left the look-up table entries for three different image acquisition set-ups of the above method where the angle α_{opt} was varied. Therefore, the curves refer to different camera orientations. The parameters of the first image acquisition set-up are presented in Table 9.1.

The angles α_{opt} of the two other curves are 36.56° and 46.56°. The parameters M, b, f, and d were left unchanged. Of course, the values of D_Z and D_0 change. They can be read from the curves. The curves demonstrate that the range of recoverable distances decreases when the angle α_{opt} is increased. On the other hand, the depth resolution is improved in this case.

The right hand plot of Fig. 9.8 shows the differences $D_{k+1} - D_k$ between adjacent distance entries in the look-up tables. If the light stripe in the image is not detected in subpixel accuracy then the shown values represent the best achievable depth resolution. The closer the object point to the reference plane the higher is the depth resolution. Todays 3D scanners can achieve a much higher resolution by incorporating subpixel accuracy and interpolation (compare Section 9.2.1). { end of Example 9.1 }

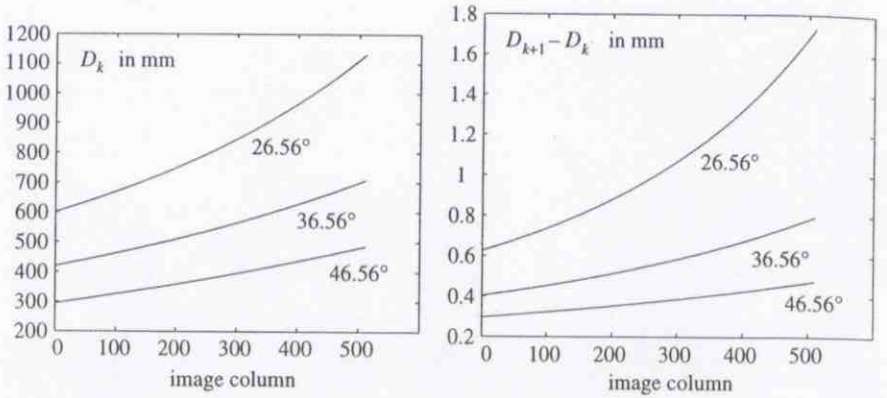


Figure 9.8: The left hand plot shows a graphical representation of the three look-up tables used in Example 9.1 for the light stripe projection technique. The graphics on the right depicts the differences of adjacent distance entries in the look-up tables.

Alternatively to the calibration method discussed above the curves shown in Fig. 9.8 could be found by curve fitting. The underlying model of the curve is given in equation (9.10). The control points for the curve fitting are measured by positioning the calibration plane at a suitable number of distances. The x-component of each control point is the detected image column. The y-component is the current position (distance) of the calibration plane to the reference plane. The advantage of this approach is that the base distance b does not have to be known in advance. Moreover, the position of the reference plane can be chosen arbitrarily, however notice that its orientation must not be changed from the initial definition. Altogether, this leads to a simpler calibration set-up.

Up to now the light stripe projection technique can recover distance values D_k for object points which are illuminated by the light plane. To obtain a 3D reconstruction the object can be rotated on a turntable. Each rotation leads to a *3D*

Calibration parameters				Calculated parameters		
M	b	D_Z	D_0	α_{opt}	f	d
512	40 cm	80 cm	60 cm	26.56°	30 mm	3.75 mm

Table 9.1: Calibration parameters and calculated parameters of a typical image acquisition set-up for the light stripe projection technique.

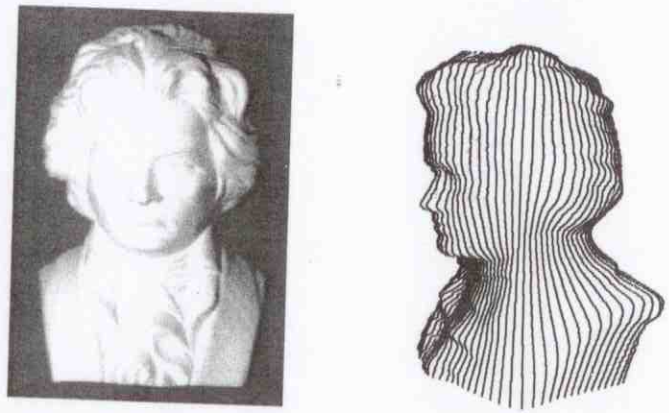


Figure 9.9: Left: Beethoven plaster statue. Right: Recovered 3D profiles on the statue using the light stripe projection technique.

profile of distance values. If the axis of rotation is oriented in such a way that it is parallel to the reference plane and lies "inside" the light plane then the 3D object can be easily represented in cylindrical coordinates.

Figure 9.9 and Fig. 9.10 illustrate reconstruction results of a Beethoven plaster statue which was rotated on a turntable. The 360°-reconstruction was obtained by using 81 rotation steps. Hence a number of 81 3D profiles were reconstructed.

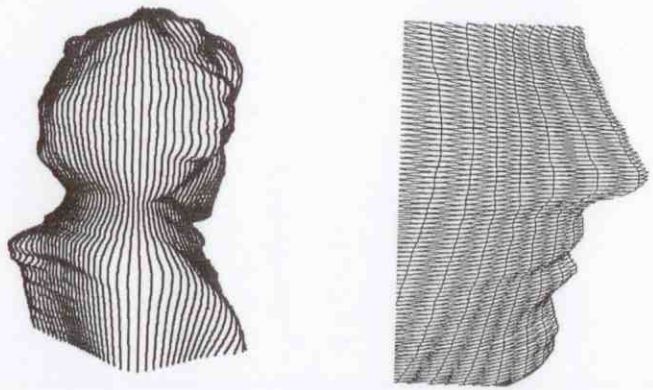


Figure 9.10: Left: Rear view of the recovered 3D profiles of the Beethoven statue shown on the left of Fig. 9.9. Right: Recovered triangular mesh around nose and mouth of the statue.

The left image of Fig. 9.9 shows the input object. The right image of Fig. 9.9 and the left image of Fig. 9.10 illustrate the recovered 3D profiles along the detected light stripes. After the 3D profiles are calculated a surface has to be fitted to the data. One way to derive a surface model is to generate triangles between adjacent 3D profiles and adjacent points on the 3D profiles. The plot shown on the right hand side of Fig. 9.10 presents the recovered triangular mesh around nose and mouth of the statue.

9.1.4 Static Light Pattern Projection

Instead of using single light rays or planes we can project several rays or planes at the same time on the examined object surfaces (*static light pattern projection technique*, see Fig. 9.11) to reduce the number of images. These light rays or planes are visible as a set of light spots or stripes in the image. The fundamental principle is based on the techniques described in the previous sections. Popular patterns are dotted lines, parallel lines, dot matrices, as well as concentric circles.

The main question is how to uniquely identify and index the light stripes in the camera image when several light rays or planes are projected simultaneously into the scene. It has to be determined which pattern element corresponds to which illuminated point in the image in order to compute the triangulation. Using

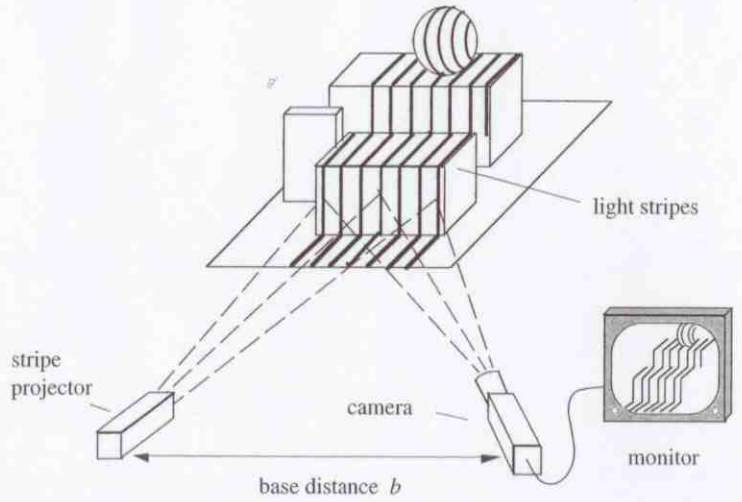


Figure 9.11: The principle of simultaneously projecting several light stripes.

smoothness assumptions the elements of a regular pattern can be addressed by taking neighborhood relations into account.

The projection of static light patterns can also be used to carry out planarity checks on smooth surfaces for industrial quality control, for example. Indexing of the pattern elements and a subsequent depth value calculation are not required for this application. The planarity check could be realized by comparing the measured image positions with the image of a highly planar surface.

9.2 PROJECTION OF ENCODED PATTERNS

If a simple geometric pattern, for example a light spot or a light stripe, is projected onto an object, then a large number of images has to be generated for determining the surface geometry of the object. When dealing with complex surfaces several light patterns cannot be simultaneously projected into a scene because a unique matching between the projected pattern and the pattern visible in the camera image cannot be ensured, for example due to possible occlusions. However, a unique matching can be obtained by a spatiotemporal or spatial coding of the patterns. In the following the calculation of depth values by means of the analysis of binary encoded and color encoded patterns will be described. The main principle of triangulation underlying the depth recovery will not be changed.

9.2.1 Binary Encoded Light Stripes and Phase Shifting

For the *binary encoded light stripe projection technique* a set of light planes is projected onto the examined objects. The individual light planes are indexed by an encoding scheme for the light patterns. These light patterns lead to a unique code for every plane. In principle every (unique) binary code could be employed. Then, 2^n light planes are uniquely encodable by using n patterns, i.e. taking n images.

The so-called *Gray code* (named after Frank Gray) is often used as a simple and robust indexing scheme. To avoid confusion notice that we are not dealing with a gray value code. A Gray code $G(i)$ is a binary code where i is an integer, and the binary representations of $G(i)$ and $G(i+1)$ differ exactly in one bit. Therefore, a code error in neighboring light planes can be easily detected. Figure 9.12 illustrates the principle of coding.

By applying a threshold operation to the images the scene areas which are illuminated by the current pattern are detectable. The obtained binary images are bit-planes which form a so-called *bit-plane stack*. After completing the image

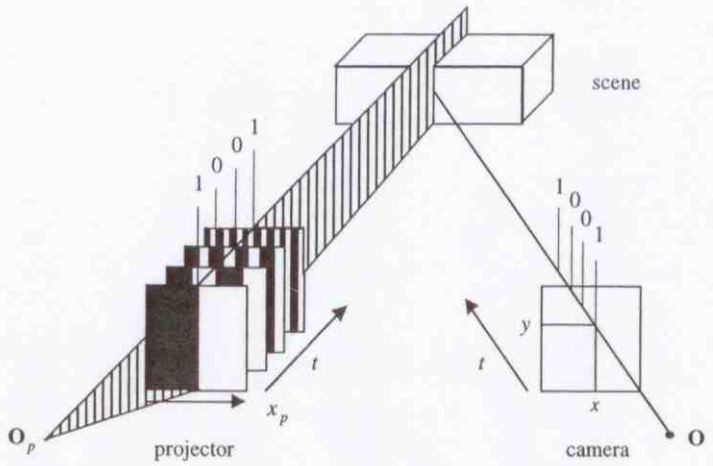


Figure 9.12: Principle of binary encoded light stripe technique.

acquisition process the bit-plane stack contains a sequences of n bits for all projected scene points and the corresponding image points (x, y).

Figure 9.13 shows an image sequence of a cube that were taken to create a Gray code bit-plane stack of 256 light planes. The light planes were horizontally projected onto the object unlike in the previous examples. The first image of the figure (upper left) shows the object scene where all light planes are switched on, hence an usual intensity image was acquired. The remaining $n = 8$ images of Fig. 9.13 are necessary to encode the $2^n = 256$ planes. The images are arranged row-wise in the order they were taken during image acquisition.

For example, the first image in this sequence (middle image in the top row) was acquired when the upper half of the light planes were switched on and the lower half of the light planes were switched off. For a better illustration of the projected patterns Fig. 9.14 visualizes the status (on or off) of the light planes. A depicted signal level "high" refers to "light plane on" and "low" refers to "light plane off". The left hand part of each binary curve in Fig. 9.14 corresponds to the upper region of the images, and the right hand part corresponds to the lower region.

During the surface reconstruction process the bit-plane-stack is used to uniquely address the light plane corresponding to every image point. A calibration procedure has to be performed for obtaining the equations of all light planes. Notice that the simple calibration scheme discussed in Section 9.1.3 is not suitable for the binary encoded light stripe technique. One reason is that the light planes cannot be regarded as lying parallel to each other.

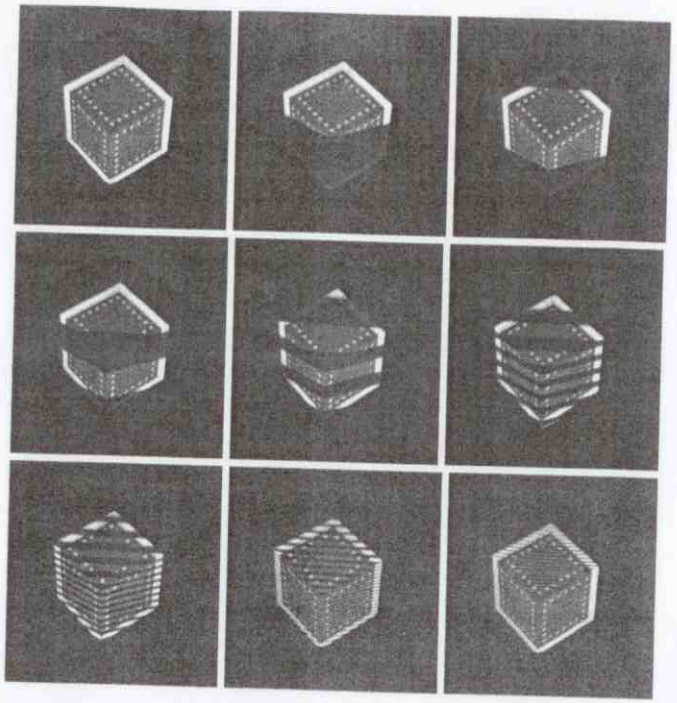


Figure 9.13: Binary encoded illumination. The entirely illuminated scene CUBE is shown at the top left. The other images show the generated images for different projected light patterns (with kind permission by A. McIvor, Industrial Research Ltd., Auckland).

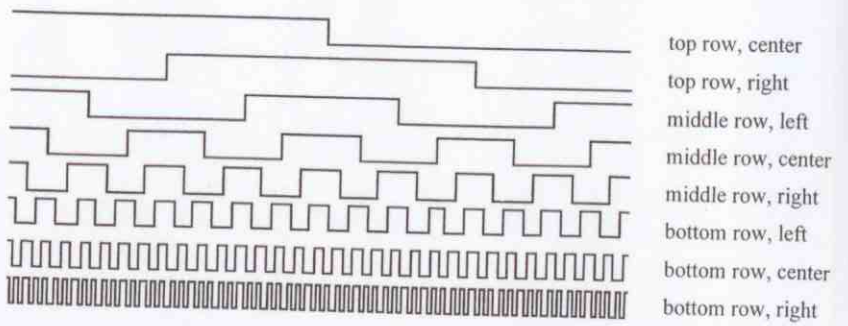


Figure 9.14: "On/off" status of the light planes during the image acquisition process. The eight different signals refer to the image sequence shown in Fig. 9.13. The signal level "high" refers to "light plane on" and "low" refers to "light plane off".

An alternative way for calibration is to place at least two planar objects of known orientation and position in the scene. The light plane equations are calculated from the light stripes on the known planes. After calibration the 3D coordinates of the examined object points are obtained using conventional triangulation (compare Section 9.1.1).

The resolution of the depth measurement can be increased by employing the so-called *phase shift method* which is based on the idea of interpolating between adjacent light planes. One possible way is to fuse two adjacent light planes to one macro light plane, as a first step. Thus the number of projected light planes is reduced by a factor of 2. The task of this initial step is to compute the indices of all macro light planes by using the Gray code technique as described above. Then the macro light planes are shifted four times by the width of one single light plane and four additional images are acquired (see Fig. 9.15).

During the preparation of the interpolation process the shifted pattern is regarded as a continuous sine shaped signal. The parameters of this signal are calculated by fitting the measured image irradiances to a sine model. The obtained phase parameter describes the relative depth information inside the macro light planes. The determination of the phase parameter is very sensitive to noise. Therefore a large number of (actually) identical images are taken in every step

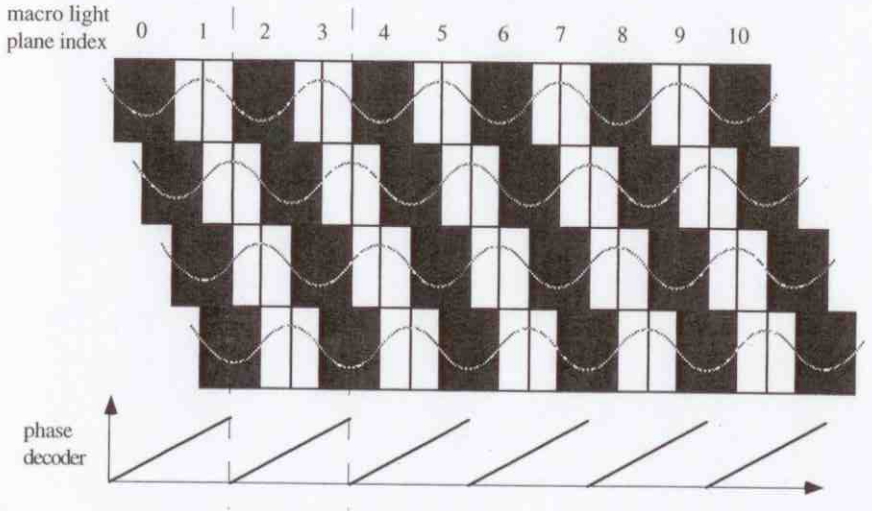


Figure 9.15: Sketch of the spatial arrangement of light patterns used to implement a simple phase shift method.

Figure 9.16: Beethoven plaster statue and a recovered range image. The surface recovery result was obtained by using the binary encoded light stripe technique and an LCD-projector which produces 320 light planes.

and an average is calculated. Notice that phase shifts can also be achieved by a mechanical movement of the projection unit.

The advantage of the binary encoded light stripe projection compared to the "ordinary" light stripe projection is that much less images are required for the same number of measured object points in the scene. Furthermore, we obtain a dense depth map or range image, and a rotation of the object is not required. Of course, a rotation is mandatory to recover a complete *3D model* of the object. However a more complete 3D model can be calculated from a set of dense depth maps or range images.

Usually computer controllable *LCD-projectors* (liquid crystal display projectors) are used to generate the light planes. There exist special projectors for structured lighting that are manufactured with high precision and can generate more than 1000 very bright light planes. Fig. 9.16 illustrates on the right a reconstructed range image of a Beethoven plaster statue (shown on the left) by using an LCD-projector which generates 320 light planes.

9.2.2 Color Encoded Light Stripe Projection

With exception of the static light pattern technique all structured lighting approaches discussed so far were based on a sequential acquisition of a set of

images. The reconstruction of moving or non-rigid objects is not possible when several images have to be taken consecutively. In principle the static light pattern technique can cope with this class of objects but there exists no way for a direct measurement of dense depth information. The idea of the binary encoded light stripe technique combined with the additional utilization of color leads to the projection of a single color pattern. This approach is used to recover depth data from a single color image. It is obvious that many more distinguishable states can be coded by colors than by a binary code. The *color encoded light stripe projection technique* described here uses patterns consisting of red, green, blue, and white light planes.

Because only three primary colors and white are utilized the segmentation of the light stripes is simple. The primary colors should produce maximum image irradiance values in the corresponding color channels, and the white plane should cause a maximum in all three color channels. Without making further assumptions this condition is only fulfilled if the examined object surfaces have no color, i.e. the objects are white or gray.

Keep in mind that the spectral power distribution $E(\lambda)$ of the illumination and the spectral reflectance factor $R(\lambda)$ of the surface material are only "seen" as an integral of the product $E(\lambda) \cdot R(\lambda)$ by the three sensors of a color camera (compare Section 2.2.2). If the object color is white or gray then its spectral reflectance factor $R(\lambda)$ can be modeled as being constant. If the object is colored then it depends on $E(\lambda)$, $R(\lambda)$ and the spectral sensitivity of the camera whether the four colors can be disentangled or not. In this respect surface colors having a low saturation are more suitable than surfaces showing pure colors.

The actual problem of the color encoded light stripe projection is the unique indexing of the light stripes in the image (compare Section 9.1.4). If all light planes are visible in the image and if each light stripe is completely visible then indexing would be trivial. However, in general, some light stripes are invisible or only partially visible due to occlusions. In this case it has to be ensured that each light stripe can be uniquely assigned to the corresponding light plane by incorporating the local neighborhood of the stripe.

For the color encoded light stripe technique the knowledge about the arrangement (coding) of the color pattern is used to solve this problem. The color pattern is divided into several distinguishable subpatterns (color codes). For the generation of the color pattern,

the entire number of projected stripes (i.e. the width of the light planes),

the number of light planes in every subpattern,

the number of subpatterns, and

the number of colors

have to be taken into account. A color code is selected that neighboring stripes have different colors. This restriction reduces the possible arrangements of the color code. Let L be the entire number of used colors and K the number of stripes of every subpattern, then the number M_0 of possible codes is

$$M_0(K,L) = L \cdot (L-1)^{K-1}.$$

The formula is derived as follows. For the first stripe in a subpattern we can choose among L colors. For the remaining $K-1$ positions we can choose among $L-1$ colors. If the colors red (R), green (G), blue (B), and white (W) are used for the color stripes ($L = 4$) and if every subpattern shall consist of $K = 6$ stripes, then $M_0 = 972$ different subpatterns can be generated according to the above formula. If the scene shall be illuminated with 512 stripes, then only 86 of these subpatterns are needed because $6 \cdot 86 \geq 512$. Bear in mind that the colors at the borders of two adjacent subpatterns have to be different. The designed color stripe code could be physically realized as a slide by using a film recorder attached to a computer. Then the slide is projected onto the object with a common slide projector.

The algorithm for finding the correct index of the color stripes in the image searches along a path orthogonal to the stripes. If an entire subpattern is visible then the index of each light stripe belonging to the subpattern can be found using its position in the complete color pattern. Starting with this information the color stripes in the remaining incompletely projected subpatterns are recognized. For instance, if the subpattern (GBWRWG) was completely found, then it is known from the chosen order of the subpatterns that the subpattern (RWBGRB) should lie next to it on the left. If instead only the pattern (RWRB) is visible in the image, then it follows that the blue and the green color stripes are occluded. The three-dimensional coordinates of the object points are calculated by a triangulation (compare Section 9.1.1).

The advantage of the color encoded light stripe projection is that the geometry of an object can be reconstructed from only one color image. The disadvantage is that the colors of recoverable objects are constrained to a certain extent.

9.2.3 Active Color Stereo Analysis

The main problem of the technique described in the previous section is the robust identification of the color pattern. One way to avoid this problem is to employ a second camera. As for the color encoded light stripe technique a color pattern is projected onto the objects. The scene is taken by two cameras placed at two different positions (stereo image acquisition). Therefore ambiguous color patterns

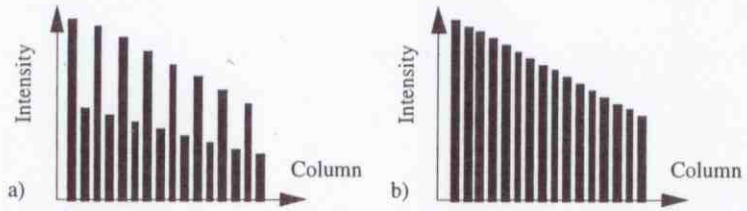


Figure 9.17: Contrasting illumination (a) and continuous illumination (b).

are no longer a problem since the patterns do not change between the two views. The color stripes can be matched by a correspondence analysis (compare Section 4.3). The advantage of this *active color stereo analysis* compared to a normal static stereo analysis consists in the possibility of matching corresponding points even if the surfaces show no texture.

By using two cameras actually any arbitrary color code can be used. However, the choice of the suitable color pattern is of great importance to assure a robust matching of the points in the two images. There exist two different ways how to design the color pattern. The pattern can be generated such that a high contrast arises between neighboring stripes (*contrasting illumination*) or such that continuous stripes arise between the stripes (*continuous illumination*).

Figure 9.17 illustrates the spatial arrangement of one color component inside a subpattern with contrasting or continuous illumination where a vertical projection of the color pattern is assumed. Neighboring color stripes can be distinguished well with contrasting illumination and smooth changes of the distances in the scene. However, for depth discontinuities, i.e. at object edges, the contrasting

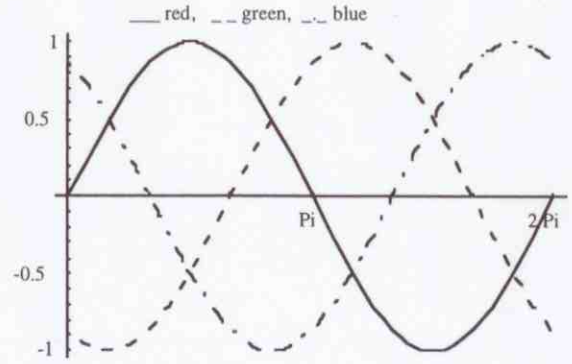


Figure 9.18: Sketch of the intensities of the three color components in a subpattern.

illumination can lead to a lower contrast between neighboring stripes in the image. On the other hand, for continuous illumination the contrast is generally low but high at object edges.

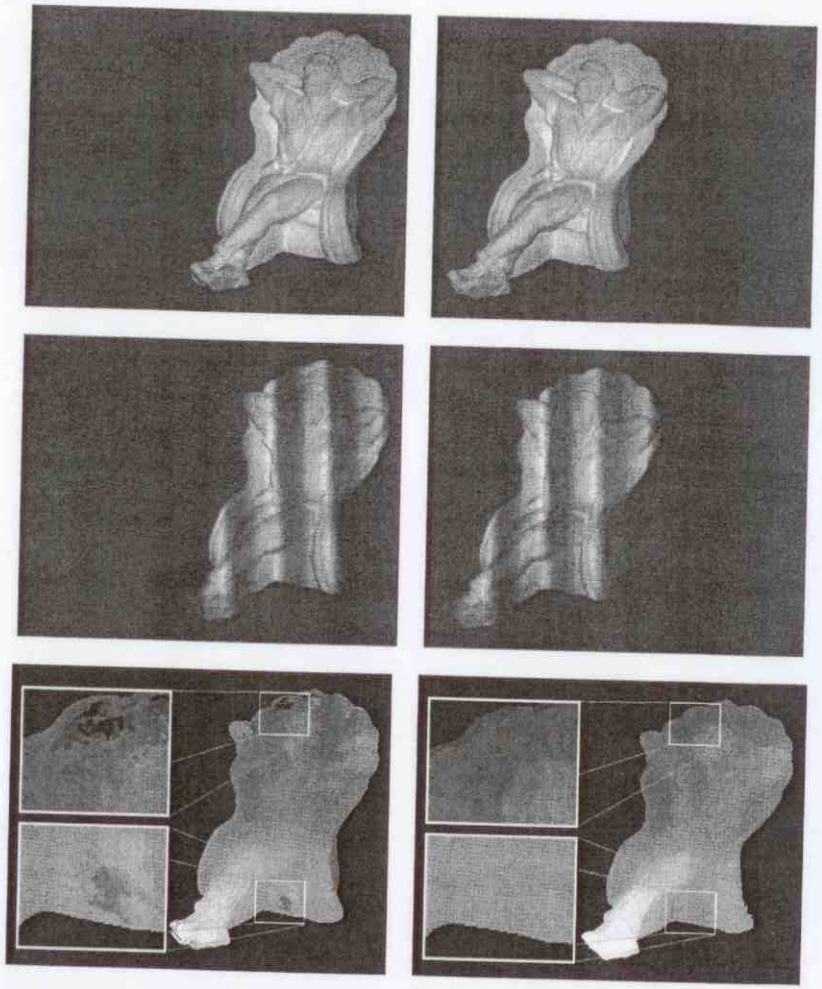


Figure 9.19: Top row: Gray value representation of the stereo color image TOM. Middle row: Gray value representation of the stereo image with projection of a color code (original images see Color Images 7 and 8). Lower left: Disparity map calculated without the projected color code. Lower right: Disparity map calculated by using the projected color pattern.

For the generation of a continuous illumination the intensity in the color components can be modeled by "sawtooth functions" or sine functions. Figure 9.18 shows a subpattern in which the intensity of the three color components are determined by sine functions that are shifted by $2/3\ \pi$. The subpattern is generated by a combination of the three components (see Color Image 6). The total color code is a periodical arrangement of several identical subpatterns.

The color stripe pattern is projected into the scene and a stereo image is generated from two different positions. To improve the color quality in both images a white balance should be carried out for the two cameras before the image acquisition starts (compare Section 2.3.2). With the standard stereo geometry (compare Section 4.1) the correspondence analysis simplifies considerably because corresponding pixels always lie on the same rows in the images.
One approach to match corresponding points is, for instance, the block matching method for color stereo analysis which was described in detail in Section 4.3.2. The extension of the block matching approach is that a color stripe pattern is projected onto the examined objects. Then the method from Section 4.3.2 can be applied to the stereo images (without any modifications) .

In the following the achievable improvement of the surface reconstruction will be illustrated. In Fig. 9.19 a gray value representation of a stereo image pair TOM is shown in the top row and the same scene with an overlaid color pattern is shown in the middle row (compare Color Images 7 and 8). The dense disparity maps calculated by using the block matching method are visualized in the bottom row of Fig. 9.19. The left and right image illustrate the recovered disparity map without and with employing active illumination, respectively.

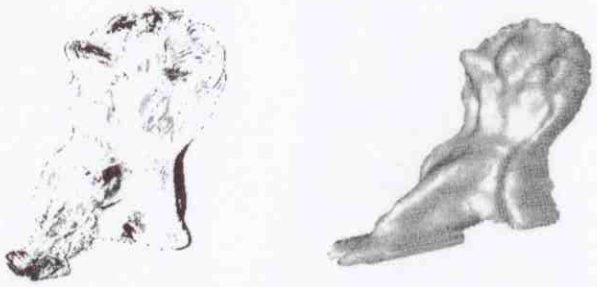


Figure 9.20: Left: Differences (scaled) between the two disparity maps in Fig. 9.19. Right: Shaded plot of the scene TOM which is reconstructed using the projected color pattern.

Errors in the disparity map on the left are visible for instance, around the head ar the foot of TOM. Two enlargements are superimposed to underline the differenc between the disparity maps. Another representation of the differences is shown c the left of Fig. 9.20 where darker intensities represent higher errors. The rig hand side of Fig. 9.20 visualizes a 3D plot of the recovered geometry reco structed from the stereo image pair by overlaying the color pattern.

The advantage of color active stereo analysis compared to the color enc ded light stripe projection is that on the one hand no knowledge about the projec ted color code is required and on the other hand in principle no assumptions ar made on the object colors. Usual stereo analysis methods are not able to dete mine the correct dense depth maps in homogeneous image regions. This proble is solved by the (active) projection of the color pattern into the scene.

9.3 REFERENCES

General overviews about scene analysis techniques using structured lighting can for example, be found in

Jarvis, R.A.: *A perspective on range finding techniques for computer vision* IEEE Transact. on Pattern Analysis and Machine Intelligence **5** (1983), pp 122-139, and
Shirai, Y.: *Three-Dimensional Computer Vision*. Springer-Verlag, Berlin, 1987.

Descriptions about the geometrical background of the techniques are given in

Fu, K.S., Gonzalez, R.C., Lee, C.S.G.: *Robotics: Control, Sensing, Vision, and Intelligence*. McGraw-Hill, Singapore, 1987,
Suk, M, Bhandarkar, S.M.: *Three-Dimensional Object Recognition from Range Images*. Springer-Verlag, Tokyo, Japan, 1992, and
Wechsler, H.: *Computational Vision*. Academic Press, Boston, USA, 1990.

The technique of the light spot stereo analysis is described in

Gerhardt, L.A., Kwak, W.I.: *An improved adaptive stereo ranging method for three-dimensional measurements*. Proc. International Conference on Computer Vision and Pattern Recognition, Miami Beach, 1986, pp. 21-26.

The binary encoded light approach is introduced in

Wahl, F.: *A coded light approach for depth map acquisition*. Proc. 8. DAGM Symp. Musterkennung, G. Hartmann (Ed.), Paderborn, Germany, 1986, pp. 12-17, and

Stahs, T., Wahl F.: *Fast and robust range data acquisition in a low-cost environment*. Proc. of SPIE **1395**, Close-Range Photogrammetry Meets Machine Vision, Zurich, Switzerland, 1990, pp. 496-503.

an LCD-projector is used for structured lighting, then besides the lens distortion f the camera (compare Section 2.1.4) also the lens distortion of the projector has be modeled. A calibration technique which takes both lens distortions into ccount is given in

Valkenburg, R.J., McIvor, A.M.: *Accurate 3D measurement using a structured light system*. Image and Vision Computing **16** (1998), pp. 99-110.

nvestigations about the theoretically and practically achievable accuracy using ght stripe techniques can be found in

Lin, J.C., Chi, Z.-C.: *Accuracy analysis of a laser/camera based 3-d measurement system*. Proc. of SPIE **449**, 3rd International Conference on Robot Vision and Sensor Controls, Cambridge, Massachusetts, 1983, pp. 158-170,

McIvor, A.M.: *The accuracy of range data from a structured light system*. Report 190, Industrial Research Limited, Auckland, New Zealand, 1994.

Yang, Z.M., Wang, Y.F.: *Error Analysis of 3D Shape Construction from Structured Lighting*. Pattern Recognition **29** (1996), pp. 189-206.

echniques for active color stereo analysis (Section 9.2.3) can be found in

Chen, C.-S., Hung, Y.-P., Chiang, C.-C., Wu, J.-L.: *Range data acquisition using color structured lighting and stereo vision*. Image and Vision Computing **15** (1997), pp. 445-456, and in

Koschan, A., Rodehorst, V., Spiller, K.: *Color stereo vision using hierachical block matching and active color illumination*. Proc. Int. Conference on Pattern Recognition ICPR´96, Vienna, Austria, 1996, pp. 835-839.

he first article employs dynamic programming for the correspondence analysis ee Section 4.3.2) and the second one uses a block matching approach (compare ection 4.3.3).

.4 EXERCISES

1) Section 9.1.1 described the 2D triangulation using two known angles α and β. low, consider the calculation of the angle β. Let f be the effective focal length of he camera. The point **P** is projected onto a point **p** in the image plane $Z = -f$ of he camera which has the X-coordinate x. The optical center is **O** (compare Fig.

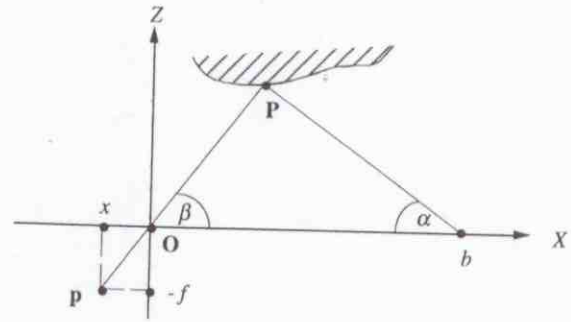


Figure 9.21: Determination of the angle β.

9.21). Derive the formula to determine the angle β from the position of the projected scene point **p** in the image plane.

(2) As described in Section 9.1.1 the angle γ is not taken into account when the range data is calculated (3D case). How can γ be used to verify the reconstruction result?

(3) Discuss ways to improve the achievable depth resolution of the method introduced in Section 9.1.3.

(4) A light plane is projected into a scene consisting of two polyhedral objects

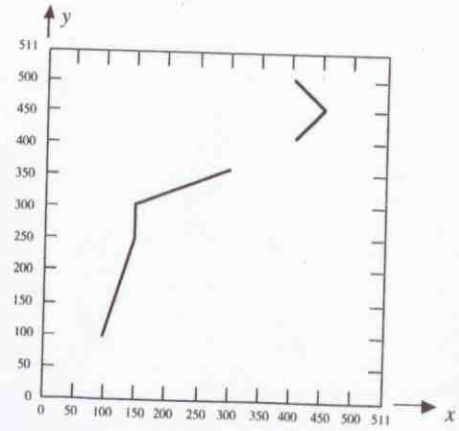


Figure 9.22: Light stripes visible in the image.