

## 2 Basics

The following chapter summarizes basics of optics and information theory. For the sake of compactness, only limits and relations are described, which play an important role in the physical and information theoretical optimization of the introduced sensor. Elaborate explanations and mathematical derivations are widely avoided. For a detailed explanation related literature is cited.

The sensor principle presented in this thesis is based on the well known principle of *line triangulation*. In the first section of this chapter, the basic idea of triangulation is explained. The next section introduces the physical limits of optical 3D sensors based on line triangulation. The third section specifies aspects for the information theoretical optimization of optical 3D sensors. The properties that make a 3D sensor *efficient* are discussed in particular.

### 2.1 Triangulation

This thesis intents to develop of a new measurement principle for the fast three-dimensional acquisition of *macroscopic* objects with *scattering* surfaces. In addition to the commercially emerging ‘time-of-flight’-techniques (to be discussed in chapter 3), there is only one basic principle which is commonly used for this task: **Triangulation** (see Fig. 2.1): A *vertical* shift  $\Delta z$  of a surface point, observed with a camera under a defined angle - the ‘*triangulation angle*’  $\theta$  - introduces a *lateral* shift  $\Delta x'$  in the image plane on the camera chip.

Note that in this thesis all *image-sided* quantities are marked by primed characters, whereas *object-sided* quantities are not. Sometimes it is convenient to describe image-sided quantities

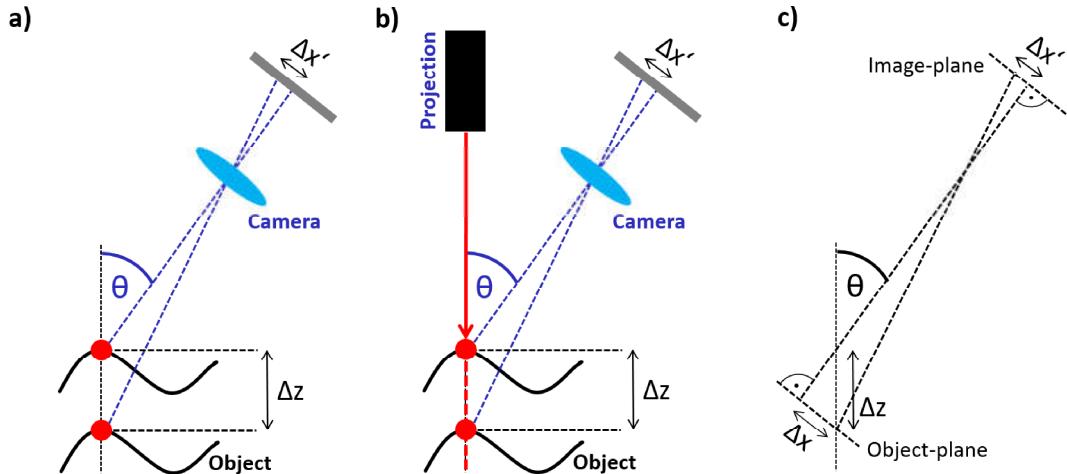


Figure 2.1: Basic principle of triangulation: A vertical shift  $\Delta z$  of a surface point observed with a camera under the triangulation angle  $\theta$  introduces a lateral shift on the camera chip. a) **Passive** triangulation requires a sufficiently structured object surface. b) **Active** triangulation structures the object artificially by proper illumination. c) Sketch to the triangulation formula (Eq. 2.1).  $\Delta x$  is the related lateral shift in the object plane.

in units of the pixel pitch. 1 *pix* is the distance between the centers of two neighbored pixels.

To identify the surface point in the camera picture, the object has to be sufficiently structured, e.g. by texture. If this cannot be ensured for all surfaces of interest, it is possible to structure the surface artificially with the help of proper illumination, e.g. with a laser spot (see Fig. 2.1(b)).

For evaluation  $\Delta x'$  is measured. Together with the known  $\theta$  and scale of the system  $\beta'$ ,  $\Delta z$  can be calculated via

$$\Delta z = \frac{\Delta x}{\sin \theta} = \frac{\Delta x'}{\beta' \sin \theta}, \quad (2.1)$$

where  $\Delta x$  is the related lateral shift in the object-plane of the camera<sup>1</sup>.

As an ultra-precise mounting of components would be required, the triangulation formula Eq. (2.1) is not used in practice without alteration. Moreover, each optical system suffers from aberrations, which are not considered in Eq. (2.1). Instead a calibration taking aberrations, varying magnification, and several other parameters into account is applied. The calibration for the measurement principle which is introduced in this thesis, is described in chapter 8.

### 2.1.1 Correspondence

*Correspondence* describes the relation between two light patterns, e.g. the pattern to be projected and the pattern observed by the camera. This relation has to be found *unambiguously* in order to evaluate correct 3D data. The ‘*correspondence problem*’ is well known in triangulation. Figure 2.2 introduces the correspondence problem. Point-shaped signals are assumed for simplicity but the explanation is also valid for stripes, lines, checkerboards or all other kinds of signal-patterns. A point projected from the coordinate  $(x'_1, y'_1)_P$  in the projector plane, illuminates the object surface at  $(x_1, y_1, z_1)$ , which is imaged by the camera at the position  $(x'_1, y'_1)_C$  on the camera chip. If only one point is projected, it is obvious that the signal on the camera chip ‘*corresponds*’ to this point (see Fig. 2.2(a) and (b)). The shift between  $(x'_1, y'_1)_P$  and  $(x'_1, y'_1)_C$  (often called ‘*disparity*’) can be calculated, which deciphers the depth  $z_1$  in space. Generally this shift cannot be predicted or calculated in advance, since it is strongly dependent on the surface shape to be measured.

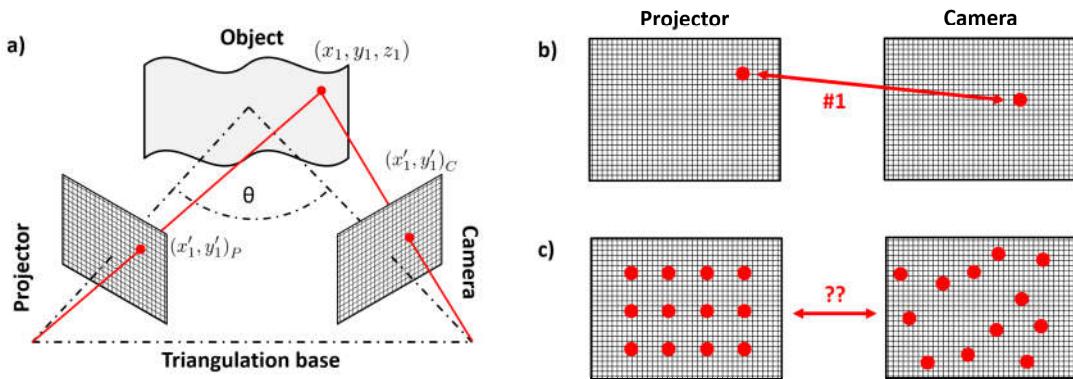


Figure 2.2: Basic description of the correspondence problem

<sup>1</sup> As can be seen in Fig. 2.1(c), a vertical shift  $\Delta z$  introduces a shift of the object-plane as well. For sharp imaging over the whole range  $\Delta z$ , a certain *depth of field* is required. The shift of the object-plane also changes the scale of the system ( $\beta = \beta(z, \theta)$ ). These problems can be avoided by complying the *Scheimpflug condition*, which requires a tilt of the chip (= image plane) inside the camera [4–6].

As soon as more than one signal (point)  $(x'_1, y'_1)_P, \dots, (x'_n, y'_n)_P$  is projected, the correspondence of points remains ambiguous (see Fig. 2.2(c)). This ambiguity cannot be resolved unless an ***additional source of information*** is exploited. Many options with different advantages and drawbacks are available, and indeed, one of the main distinctions between the multitude of triangulation principles is the *approach to solve the correspondence problem*. Representative methods are discussed in chapter 3 and section 10.1.

The measurement principle presented in this thesis is based on *line* triangulation: Instead of isolated points, *one or more lines* are projected onto the surface. Explanations about the principle and related problems are given in chapter 3 and 4. In chapter 5 a new approach for the solution of the correspondence problem of line triangulation is presented.

## 2.2 Optical Basics of Triangulation Sensors, condensed

### 2.2.1 Misleading Terminology: Measurement Uncertainty

*All measurements include errors.*

An *error* is defined by the difference between true value and measured value of a measurand. The exact value of an error is *always unknown*, otherwise it could be subtracted from the measured value. Only an *estimation* of the error can be given that is commonly called '*measurement uncertainty*'.

For a quantitative comparison of errors the term '*measurement uncertainty*' has to be properly defined. A lack of proper definition often leads to misunderstanding, since it remains unclear which kind of error and which statistical property (absolute value, standard deviation, ...) is intended to be described.

The *Guide to the Expression of Uncertainty in Measurement (GUM)* [7] gives explicit rules for the terminology. Simplified, it can be said that two kinds of errors, showing different behavior exist: *systematic errors* and *random errors*. *Systematic errors* cause deviations from the true value, that are always the same for different repeated measurements. Therefore, they affect the so-called '*trueness*' of the measurement. In optical 3D metrology, a poor *trueness* can be caused by a poor calibration of the system, but by the object itself as well, e.g. if the light penetrates a certain distance into the surface before it is scattered (volume scattering). *Random errors* cause deviations from the true value that are different for different measurements. For a large number of measurements, the random errors tend to cancel each other. A large number of measurements can be achieved by repeated measurements of the same measurand ('*time average*' via a time sequence) or simultaneous measurements of different measurands with the same value ('*ensemble average*' via a spatial sequence). The standard deviation of a random error is defined as '*precision*'.<sup>2</sup> A poor precision of an optical 3D measurement is mainly caused by statistical noise on the data, e.g. speckle noise or photon noise (to be discussed in section 2.2.4).

If the *trueness* and the *precision* are jointly considered, they define the '*accuracy*', which describes how close a measurement is to the true value of a measurand. A common example for the graphical explanation of *trueness*, *precision*, and *accuracy* is shown in Fig. 2.3.

For the remainder of this thesis, the term ***measurement uncertainty***  $\delta z$  is defined as the ***precision*** of 3D data.

---

<sup>2</sup> The exact definition is '*repeatability precision*', if repeated measurements were performed under *constant* conditions and '*reproducibility precision*', if repeated measurements were performed under *varying* conditions [8].



### 3 State of the Art: The Basic Principles of Optical 3D Metrology

This chapter introduces a selection of *basic principles* for the acquisition of three-dimensional surface data of macroscopic objects with scattering surfaces. As depicted in chapter 2, most of the applied principles in this regime rely on triangulation. During the following descriptions, special attention is paid to advantages and drawbacks in reference to the desired specs *speed*, *precision*, *high feature resolution and data density* as well as how the *correspondence problem*, introduced in section 2.1.1, is solved. Of course, sensors which apply the basic methods in their ‘pure form’ are rarely found in the literature or in the market. Most available sensors exploit a modification of the ‘pure form’. It is clear, that not all of these modifications can be discussed here. However, some single-shot principles will be discussed in more detail.

A second ‘state of the art section’ (section 10.1) will deal with a few selected methods, that use a similar idea or seem to be strongly related to the methods and approaches introduced in this thesis. To understand similarities and differences, these methods will be discussed after the explanations of chapter 5.

#### 3.1 Preamble: Stereo Vision

The human eye is an extremely paradigmatic invention of nature. It serves as the most important link between the human brain and the outside world and is the archetype for all classical imaging systems. Even more important for this thesis: The whole human vision system is more than the sum of its parts. Two eyes, paired with the well-defined geometrical arrangement in the head, enable the ability of 3D vision by triangulation.

Hence, the human vision system is the *archetype* of all existing triangulation methods. The underlying principle is called ‘**passive stereo-photogrammetry**’, or in short ‘**passive stereo**’. Two imagers (cameras, eyes) observe the object under a certain triangulation angle  $\theta$ . In order to find the correct correspondences in the two images, the object surface has to be

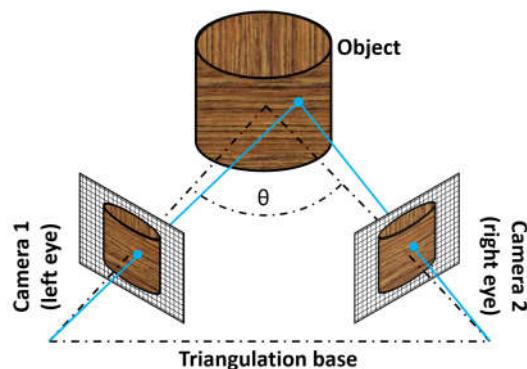


Figure 3.1: Basic principle of passive stereo photogrammetry: An object is observed by two imagers under a triangulation angle. The object surface is sufficiently structured in order to identify correspondences.

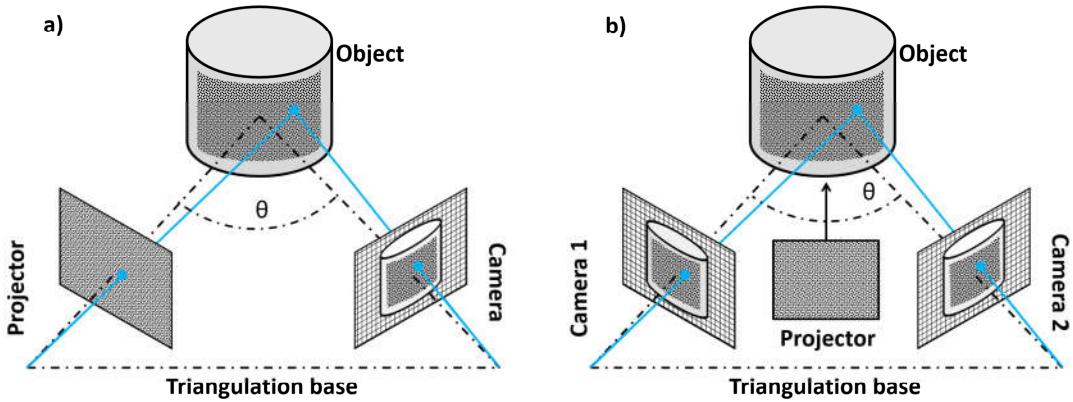


Figure 3.2: Basic principle of active stereo photogrammetry: The object is artificially structured by projection. In order to span a triangulation base, two options are possible: a) One calibrated camera and one calibrated projector. b) Two calibrated cameras and one projector.

sufficiently structured, e.g. by intrinsic texture, see Fig 3.1. If this cannot be ensured, the surface is artificially structured by pattern projection ('*active stereo*'). The pros and cons of different patterns will be discussed later. In order to span a defined triangulation base, it is important that the system has *at least two calibrated components!* This means (see Fig 3.2):

- one calibrated camera and one calibrated projector or
- two calibrated cameras and one uncalibrated projector just for 'synthetic texture' projection (avoids elaborate projector calibration).

This is really fundamental and plays an important role in the rest of this thesis. Nearly all triangulation methods used for optical 3D metrology are *active* methods and follow these rules.

Subsequently, some active methods will be described. Active stereo photogrammetry will be shortly discussed again in section 3.3.1, when '*Microsoft Kinect One*' is introduced.

## 3.2 Sequential Methods

### 3.2.1 Phase Measuring Triangulation

Phase measuring triangulation (PMT) is the *archetype of all sequential triangulation methods*. It is still used with slight modifications in a majority of 3D sensors due to its prominent features: *high precision* and *high density of 3D data* (see below). The basic principle was first described by Srinivasan, Liu and Halioua in 1984 [33]: A pattern, composed of sinusoidal fringes is projected onto the object surface, observed with a camera under the triangulation angle. Due to the depth variations of the surface, the fringes are deformed in the camera image. Hence, the observed phase  $\varphi(x', y')$  of the fringes encodes the depth  $z$  (modulo  $2\pi$ ). The observed intensity in each pixel  $(x', y')$  can be generally expressed as

$$I(x', y') = A(x', y') + B(x', y') \cdot \cos(\varphi(x', y')) . \quad (3.1)$$

Eq. (3.1) contains *three unknowns*: The (wanted) phase  $\varphi(x', y')$  as well as  $A(x', y')$  and  $B(x', y')$ , which contain information about the unknown bias illumination and object reflectivity. *Three unknowns* require at least *three equations* to be solved. These equations are

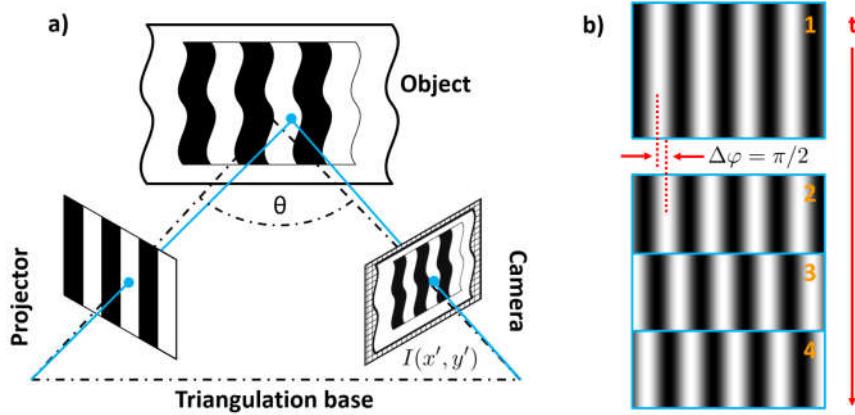


Figure 3.3: Basic principle of phase measuring triangulation (PMT). a) A sinusoidal pattern is projected onto the object. The phase of the fringes in the camera picture deciphers the object height and is evaluated by temporal phase shifting. b) ‘Four-phase shift’ as an example for phase shifting: For subsequent images are acquired. Between each image, the phase of the projected pattern is shifted by  $\Delta\varphi = \pi/2$ .

generated via a *temporal phase shift* procedure [34]:  $F \geq 3$  exposures are taken. Between these exposures the projected phase is shifted by  $\Delta\varphi = \frac{2\pi}{F}$ . The observed intensity  $I_s(x', y')$  in the  $s^{th}$  shift is

$$I_s(x', y') = A(x', y') + B(x', y') \cdot \cos(\varphi(x', y') - \varphi_s), \quad \varphi_s = (s-1)\frac{2\pi}{F}. \quad (3.2)$$

Finally  $\varphi(x', y')$  can be evaluated [34] by

$$\varphi(x', y') = \arctan \frac{\sum_{s=1}^F I_s(x', y') \cdot \sin \varphi_s}{\sum_{s=1}^F I_s(x', y') \cdot \cos \varphi_s} \quad (3.3)$$

For  $F = 4$  (‘four-phase shift’), the algorithm becomes very simple and, in addition, insensitive to second order nonlinearities:

$$F = 4 \quad \Rightarrow \quad \varphi(x', y') = \arctan \frac{I_2(x', y') - I_4(x', y')}{I_1(x', y') - I_3(x', y')} \quad (3.4)$$

However, in practice commonly many more sequential images ( $F \approx 8$  to 30) are required. This is related to the profound *ambiguity problem*. As discussed,  $\varphi(x', y')$  can be only evaluated by modulo  $2\pi$ . For a precise 3D evaluation, the fringe frequency should be high, which produces ambiguities to a large extent. The problem is solved for example by successive phase shifts with different fringe frequencies (‘multi-frequency phase shift’) [35] or by projecting an additional sequence of a binary ‘gray-code’<sup>1</sup> [37], which enables ‘phase unwrapping’. A detailed overview about different strategies is given in [38–40].

Although PMT is not a single-shot principle, it represents the ‘gold standard’ of 3D metrology. This has mainly two reasons: First, it is very precise. A measurement uncertainty down to

<sup>1</sup>The ‘gray-code’ is a cyclic permutation binary code named after Bell Labs researcher Frank Gray [36]; also known as ‘reflected binary code’

several  $\mu\text{m}$  can be achieved, which, for example, enables measurements of dental casts or coins [34, 41] with sufficient precision. Second, PMT is a *purely local* method. Each camera pixel produces a 3D point without relying on information from neighbored pixels. This is why PMT is often described as ‘pixel-dense’ with  $\rho_{3D} = 100\%$ . However, 100% can not fully be reached, since the sampling theorem has to be satisfied. And indeed, even PMT displays artifacts at sharp edges, where the sampling theorem is violated.

Nevertheless, the channel efficiency of ‘very good’ PMT approaches reaches high values around  $\eta_{PMT} \approx 20\%$  to 30% (see section 6.5). Single-shot methods, described in section 3.3, have commonly a much lower channel efficiency. This thesis intends to develop a single-shot principle that reach a channel efficiency similar to the good values of  $\eta_{PMT}$ .

The Institute of Optics in Erlangen investigated, developed and constructed PMT sensors since the late 90’s with focus on medical applications, such as measurements of the human body, face or teeth [34, 41–44]. Figure 3.4 shows a few examples. The sensors are commercialized by the spin-off ‘3D-Shape GmbH<sup>2</sup>’ [45].

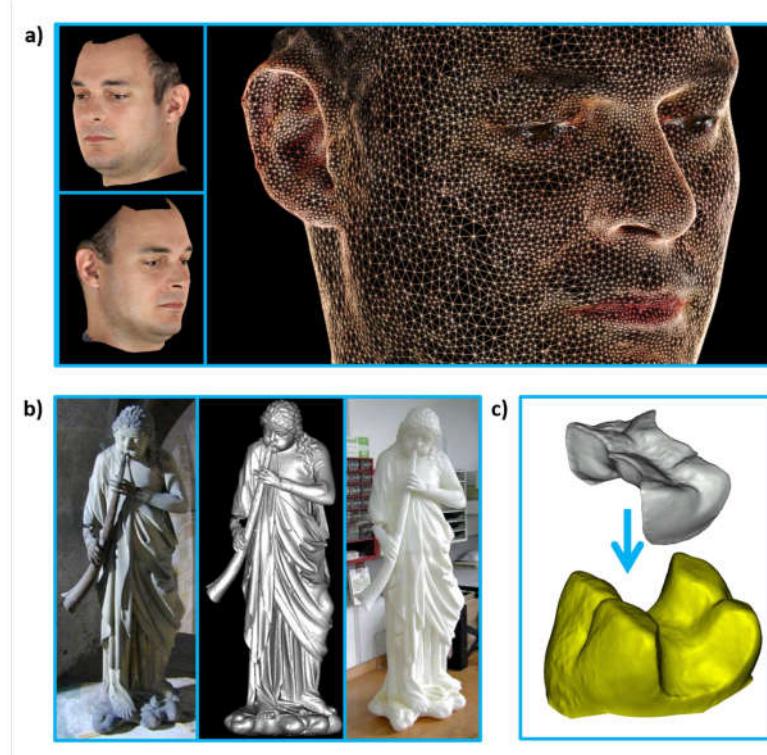


Figure 3.4: Measurement examples from PMT sensors developed at the Institute of Optics in Erlangen. a) Facial scan: 3D model from two different positions and triangular mesh. b) The ‘Trombone Angel’ of the Bamberg cathedral. Left to right: original, 3D model, copy. c) Treated tooth and dental inlay.

### 3.2.2 Other Sequential Methods

In simplified terms one can say that *all sequential methods follow a similar basic approach*: In each coordinate (pixel) on the camera chip, a *temporal sequence* of intensity values deciphers

<sup>2</sup> [www.3d-shape.com](http://www.3d-shape.com)

a coordinate on the projector chip (or the chip of the second camera) to obtain the correspondence. Eventually, a depth value is calculated via triangulation of these two coordinates. This procedure does not necessarily require a sinusoidal pattern. Principally any random pattern (within the limits of the sampling theorem) can be projected.

Sequential methods, that do not use the classical PMT approach, commonly do this all for the same purpose: *speed up the total measurement time for one 3D frame* in order to enable also measurements of moving objects. This is necessarily related to fast hardware. Due to the improvements in chip technology, cameras with high frame rates are affordable nowadays. Mostly, the projector is the bottleneck. For an 8 bit pattern-image with 256 gray values, a DLP projector requires 256 mirror switching cycles until the next pattern can be projected. Depending on the measured scene, this can be too long to project a sequence of patterns without object motion in between. This ‘projector-problem’ is solved in different ways. Some illustrating methods are discussed in the following.

### Binary Phase Shifting

By projecting a binary pattern the projector speed can be significantly increased. The projector pixels (mirrors on DLP) only have to switch *once* per picture in *on-* or *off-state*. Hence, a binary image can be projected 256 times faster than an 8 bit gray image (sufficient luminous intensity assumed).

Multiple options for the design of binary patterns are available. Projecting a binary stripe pattern provokes phase errors in the evaluation. This can be solved by defocusing the projector or by algorithmic methods [46]. The latter may involve low pass filtering. Another possibility is to use dithered patterns [47]. Other methods are illustrated, e.g. in [38–40].

### Temporal Correlation of ‘Random’ Patterns

The discussed ‘projector problem’ can also be solved by using a projector, that does *not* require electronic switching. The group of Kowarschik in Jena has been following this approach for many years. The basic setup is comparable to the setup of Fig. 3.2(b): Two cameras and one projector, which principally does not have to be calibrated into the system. A pseudo random, band-limited pattern is projected, which constantly moves over the object surface. This can be done, for example, by moving a slide with the imprinted pattern on an overhead-projector or by projecting the pattern onto a wobbling mirror [48]. With a high-frame rate camera, a series of images  $I_{t_1}, \dots, I_{t_n}$  is acquired in a short time. Each camera pixel  $(x', y')_{C_1}$  on camera 1 delivers a temporal sequence of gray values  $[I_{t_1}(x', y')_{C_1}, \dots, I_{t_n}(x', y')_{C_1}]$ . The correct correspondence to pixel  $(x', y')_{C_2}$  on camera 2 can be found by (temporal) correlation of the gray value sequence  $[I_{t_1}(x', y')_{C_2}, \dots, I_{t_n}(x', y')_{C_2}]$  with the sequence from camra 1.

A robust result requires  $n \approx 15$  to 30 temporal images [49]. Since all images should be acquired in a short time (compared to the object motion), one of the main challenges in this approach is the need of *bright light sources*. A tricky solution is the use of moving laser speckles from a rotating diffuser as projected pattern [49].

As PMT does, the described method delivers nearly independent 3D points in each camera pixel. If a reduced feature resolution and 3D data density is sufficient (e.g. at very smooth objects), the number of required images can be reduced by observing the temporal evolution of pixel-clusters (e.g.  $5 \times 5$  windows). This ‘spatiotemporal correlation’ [50] is a representative example for the free exchange between space and time when dealing e.g. with correlation information<sup>3</sup>.

---

<sup>3</sup> In other words: The *Space-time-bandwidth-product STBP*, also known as ‘information cube’ is ‘incompressible’, see e.g. [51, 52]

The authors of [53, 54] use a similar criterion as described above to obtain correspondence in the camera images, but project an aperiodic sinusoidal fringe pattern. The fast varying pattern is generated with a rotating slit mask in front of a high power light source (*GOBO-Projector*).

### Spatial Codification

Spatial codification approaches are commonly used in single-shot principles for correlation. In sequential methods as PMT, similar approaches can be used to replace time consuming multi-frequency phaseshifts. In [55], a sinusoidal pattern is superimposed by a two-dimensional codification pattern. Although the principle is fast, the spatial codification consumes space bandwidth. Other multi-shot approaches exploiting spatial codification are given e.g. in [38].

## 3.3 Single-Shot Methods

In this thesis, the term ‘single-shot’ is defined in that way, that ***no temporal sequence*** of exposures is required for the acquisition of one 3D frame. This implies an essential advantage: If the exposure time is short enough, single-shot principles are ***motion-robust***, which enables a free movement of the sensor and the object during the measurement.

Single-shot principles, that do not exploit additional modalities (color, polarization, frequency limitations,...), can ***never*** deliver ‘pixel-dense’ data ( $\rho_{3D} = 100\%$ ) in one 3D frame! This is due to the discussed *three unknowns*, which cannot be calculated from one camera picture. A common way to overcome this dilemma is to project a *binary signal*. With proper thresholding, variations in surface reflectivity or background illumination have nearly no effect on the evaluation of the signal.

However, binary signals cannot be projected ‘pixel-dense’, which is why all single-shot data display a certain sparsity, often concealed by interpolation in the final 3D dataset.

But not all single-shot principles rely on binary projection. Principles, using patterns with different gray values, are able to reach higher data densities, albeit they are more sensitive to texture or reflectance variations.

In the following, several single-shot principles will be described. As this is relevant for the remaining part of the thesis, the discussions will be more detailed and also consider sensors, which are commercially available.

### 3.3.1 Microsoft Kinect One (Active Stereo)

The ‘*Microsoft Kinect One*’ 3D sensor (released in 2010) was originally developed for an interactive controlling of computer games. The user (player) is positioned in front of the sensor and controls his avatar by body movements (e.g. playing virtual tennis by moving his arm). Since also fast movements have to be detected, a single-shot approach is applied.

The basic setup of the sensor is comparable to Fig. 3.2(a): One camera and one calibrated projector. An additional color camera allows the acquisition of color texture. The measurement principle is based on active stereo photogrammetry. A pseudo random, speckle-like dot-pattern is projected onto the object, see Fig. 3.5(a) and b). The wavelength used is in infra-red, which allows for a simultaneous color texture acquisition.

Since the single dots are indistinguishable, the correspondences between projected pattern and camera image have to be found by comparing the *spatial arrangement of dots* in a defined correlation window (e.g. the red window in Fig. 3.5(a) and b)). This consumes space-bandwidth, which is lost for high feature resolution of the 3D image, as displayed in Fig. 3.5(c).

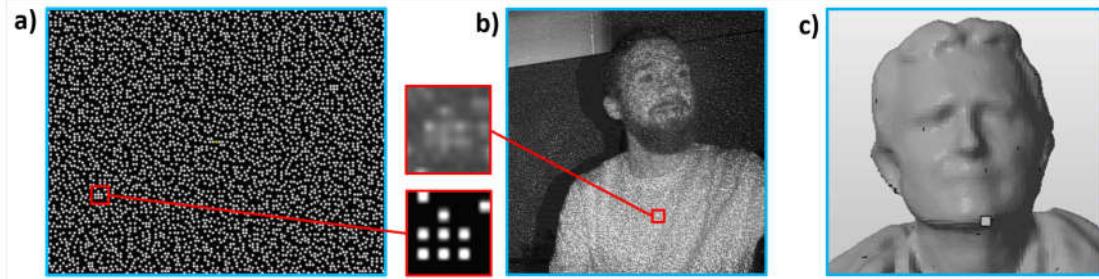


Figure 3.5: Basic principle of ‘Microsoft Kinect One’. A pseudo random, speckle-like pattern is projected onto the object. The projected wavelength is not visible for the human eye (IR). a) Part of the projected pattern [59]. b) Pattern projected on a person [62]. Red frames: Correspondences between (a) and (b). c) Measurement result of a human face.

For smooth objects and the originally intended gaming application, this might be tolerable. For metrology applications, like the measurement of mechanical parts, this fact is unfavorable, as all high frequencies, including edges, are smoothed out due to low pass filtering.

Nevertheless, ‘Kinect One’ became one of the ‘standard sensors’ in 3D vision. Since the public release of the Microsoft Kinect SDK (Software Development Kit), the sensor found its way into scientific applications; mainly in medicine, art preservation or some fields of computer vision. Last but not least this is because of the large distribution, the low price ( $\sim 100$  USD) and the (obviously uncritical) user community.

As ‘Microsoft Kinect One’ is a commercially sold product, the basic principle is only described in very few scientific publications. For more detailed information, it is recommended to review the related patent applications [56–58] as well as some technical websites [59–62].

In 2014, Microsoft released the second generation of its ‘Kinect’ sensor, which is based on a ‘Time-of-Flight’ approach (see section 3.5.1).

### 3.3.2 Fourier Transform Profilometry

Fourier transform profilometry (FTP) is one of the few principles that is able to acquire a ‘pixel-dense’ point cloud in one single-shot. As discussed, this is not possible until something else has to be condoned. In the case of FTP, this is a bandwidth limitation of the object surface. The basic principle was first described by Takeda et. al. in 1983 [63]. A narrow sinusoidal fringe pattern is projected onto the object surface, as shown in Fig. 3.6(a). The acquired camera image is Fourier-transformed (Fig. 3.6(b)). If the object surface is ‘smooth’ enough, the spectrum  $Q_1$  around the carrier frequency  $f_0$  can be directly separated from the other spectra in the Fourier domain. After filtering and Fourier-back-transformation, the phase variation  $\Delta\varphi(x, y)$  can be separated from the unwanted amplitude variation and the object height  $z(x, y)$  is calculated.

Where is the limit?  $Q_1$  has to be separable in the Fourier domain. According to [63], this is the case if

$$\left| \frac{\partial \varphi(x, y)}{\partial x} \right|_{max} < \frac{1}{3} \cdot 2\pi f_0 . \quad (3.5)$$

This restricts the maximal allowed object slope to

$$\left| \frac{\partial z(x, y)}{\partial x} \right|_{max} < \frac{1}{3} \cdot \frac{l_0}{d} , \quad (3.6)$$

where  $l_0$  is the stand-off distance and  $d$  is the triangulation base. The triangulation angle  $\theta$

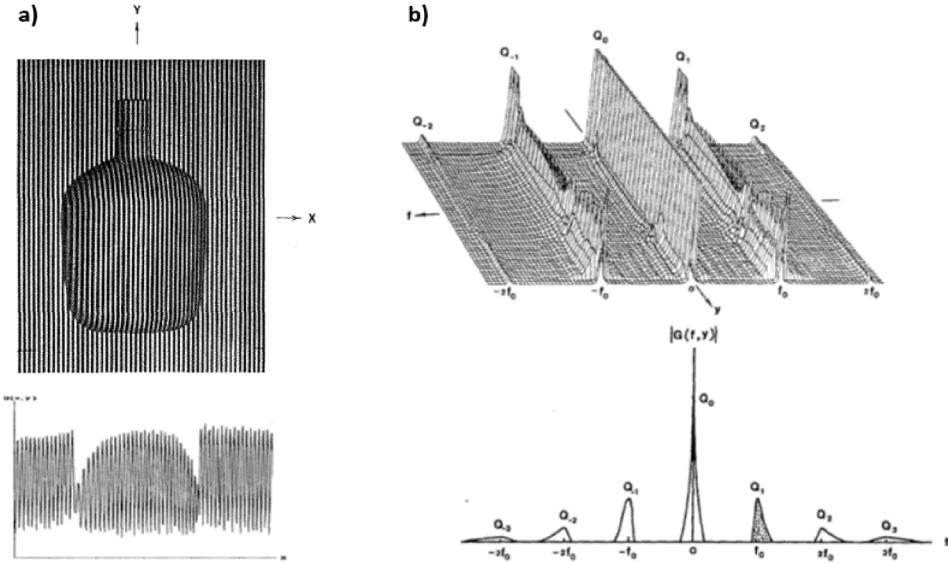


Figure 3.6: Original figures, taken from [63]: a) Object with deformed projected pattern. Cross section below. b) Related Fourier spectrum. Spectrum  $Q_1$  around the carrier frequency  $f_0$  has to be separable from other spectra.

can be defined as  $\tan \theta = d/l_0$ . For a common triangulation angle of  $\theta = 10^\circ$ , the object slope has to be smaller than 1.9. This is only satisfied by relatively smooth objects. In the meantime several methods are developed to overcome this limitation. Some of them e.g. rely on the subsequent acquisition of a second image, which makes the principle not single-shot anymore. Another interesting modification that enables the measurement of surface discontinuities is the ‘frequency-multiplex FTP’ [64]. An overview of different methods is given e.g. in [65].

### 3.3.3 Flying Triangulation (Multi-Line Triangulation with Registration)

In line triangulation or ‘light sectioning’ a narrow straight light line is projected onto the object surface. Since the line direction is perpendicular to the triangulation-plane (see Fig. 3.7), the line is deformed in the camera picture due to the surface height variations. From this deformation, the object shape along the line is calculated.

Similar to PMT for multi-shot principles, line triangulation is very paradigmatic for single-shot principles because it creates a nearly perfect 3D profile: As discussed, the object surface is scanned in line direction with best possible feature resolution, since no neighborhood information is exploited. Perpendicular to the lines, only a small amount of spatial bandwidth is consumed by sub-pixel interpolation of the line intensity maximum (see section 6.1). Therefore, light sectioning is able to measure high object frequencies, e.g. of mechanical parts. And indeed, a multitude of single-line triangulation (SLT) scanners (often called ‘laser scanners’ because of the light source used) is available for industrial applications – mainly to be mounted on robot-arms or above band-conveyors.

The serious drawback of line triangulation is its *small data density*. 3D data can be only generated along the projected line. For a 1 Megapixel camera ( $1000 \times 1000$  pix) this relates to  $\rho_{SLT} = 0.1\%$ .

Projecting multiple lines, like shown in Fig. 3.7, helps to increase the data density. However, common multi-line triangulation (MLT) sensors are *only* able to project  $\sim 10$  lines relating to  $\rho_{MLT} \approx 1\%$ . The reason for this small amount of projected lines is the *correspondence*

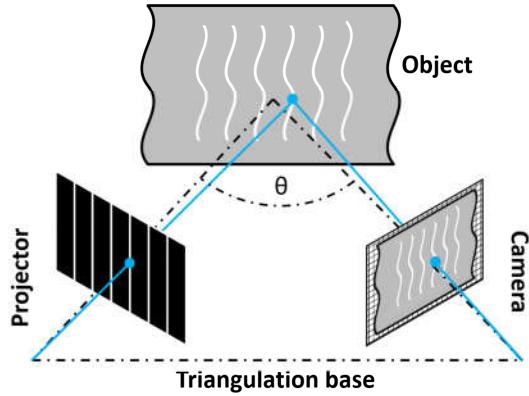


Figure 3.7: Basic principle of (multi-) line triangulation. One or more narrow light lines are projected onto the object surface, which is observed under the triangulation angle. Due to the surface height variations, the line is deformed in the camera picture. The object shape along the line is calculated from this deformation.

*problem:* Straight, narrow lines are indistinguishable! Codification approaches, which can help to distinguish the lines, are undesired in order to preserve high feature resolution.

A workaround for this problem was presented by the Erlangen Institute of Optics in 2009: '**Flying Triangulation**' [66–70]. The method is based on multi-line triangulation with  $\sim 10$  narrow lines *plus* a sophisticated registration algorithm. The basic principle is demonstrated in Fig. 3.8: In order to fill the gaps in the final 3D image, a temporal series of sparse 3D profiles is

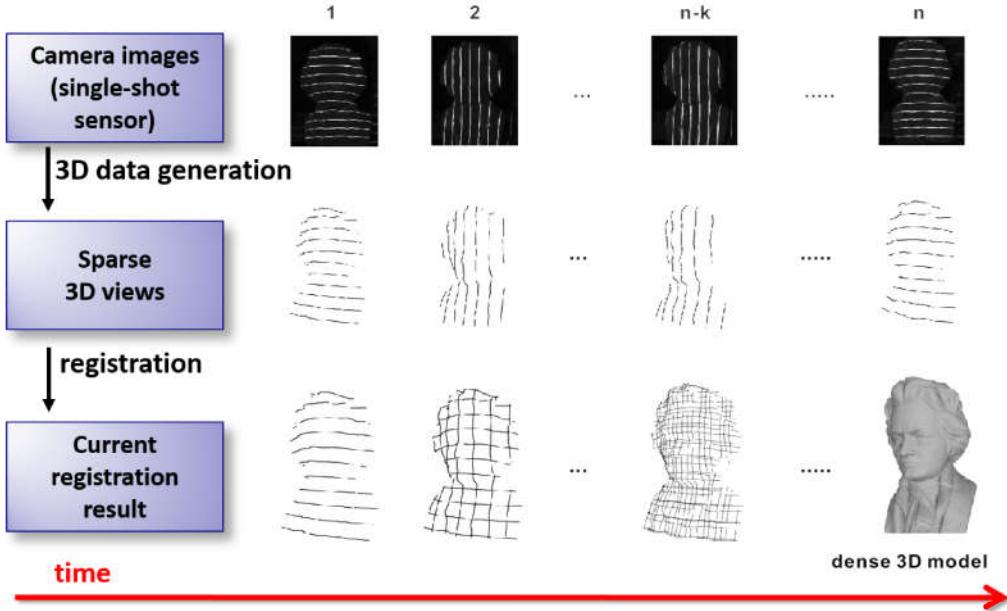


Figure 3.8: Basic principle of ‘Flying Triangulation’.

acquired, while the sensor is moved by hand around the object surface. The sparse 3D profiles are subsequently registered to each other in order to compensate for the unknown movement. The registration algorithm is based on an ‘iterative closest point’ (ICP) approach. It relies on the existence of common measured surface points in two subsequent 3D profiles, so-called ‘temporal corresponding points<sup>4</sup>’. Flying Triangulation ensures the occurrence of ‘temporal corresponding points’ by an alternating projection of horizontal and vertical line patterns. The registration works ‘on line’ and is displayed as interactive user feedback on the computer screen during the measurement. Color texture information is acquired with an additional color camera by flash exposure. 3D models measured with Flying Triangulation are shown in section 4.3. Videos of full measurement sequences can be seen on [2].

Although listed at single-shot principles, Flying Triangulation is not a single-shot principle in the common sense: It relies on a *temporal series of exposures* for a full 3D scan. However, each 3D profile for itself is acquired in one single-shot, which makes the principle robust against motion. During the measurement a free (rigid) movement of object *and* sensor is allowed.

**Remark:** As described, Flying Triangulation presents a *workaround* for the profound correspondence problem in multi-line triangulation. This thesis presents methods and approaches to *solve* the correspondence problem. However, at the beginning of this thesis, the starting point of all considerations was the improvement of Flying Triangulation in terms of reliability and data density per single-shot.

In the next chapter, a detailed description of the correspondence problem in multi-line triangulation is given, using the example of Flying Triangulation. Moreover, several sensors and innovations for Flying Triangulation, developed within the framework of this thesis, are discussed.

### 3.3.4 Artec (Spatial Coded Multi-Line Triangulation)

In the last years, the ‘Artec<sup>5</sup>’ company emerged to one of the most well known producers for hand held 3D scanners. ‘Artec’ scanners are omnipresent in newspapers or TV documentations about archeology, forensic applications or 3D printing applications.

As the scanners are commercially sold products, the underlying principle is not published in detail. However, information can be found in relating patents, like [71] and -for the experienced audience- in product videos, like [72, 73].

From this references it becomes obvious, that ‘Artec’ uses a special kind of spatial coded multi-line triangulation: The correspondence problem is solved by thickness-variations of the projected lines. From line to line the thickness is changed in different periods (see. Fig. 3.9), which makes the lines distinguishable. Exploiting the epipolar-geometry of the sensor additionally helps to resolve ambiguities.

Figure 3.9 deals with the projected pattern. Figure 3.9(a) is taken from the patent [71], and shows an ‘*illustration of a representative light structure*’ (pattern). Figure 3.9(b) consist of two screenshots from a product video [72] on the official ‘Artec’ YouTube-channel. The pattern, projected by the scanner, is clearly visible and resembles the pattern of Fig. 3.9(a).

To distinguish between the lines, the binary thickness-code has to be read by the computer. This requires neighborhood information, since a line segment of a certain length has to be visible in the camera picture. The consumed space bandwidth is lost for the resolution of high

---

<sup>4</sup> In this case the word ‘corresponding’ refers to *temporal correspondence*: The same surface point, measured with a *moving* camera in subsequent exposures. In this thesis, ‘correspondence’ is mostly used for *spatial correspondence*: The same surface point, measured simultaneously from different directions.

<sup>5</sup> [www.artec3d.com](http://www.artec3d.com)

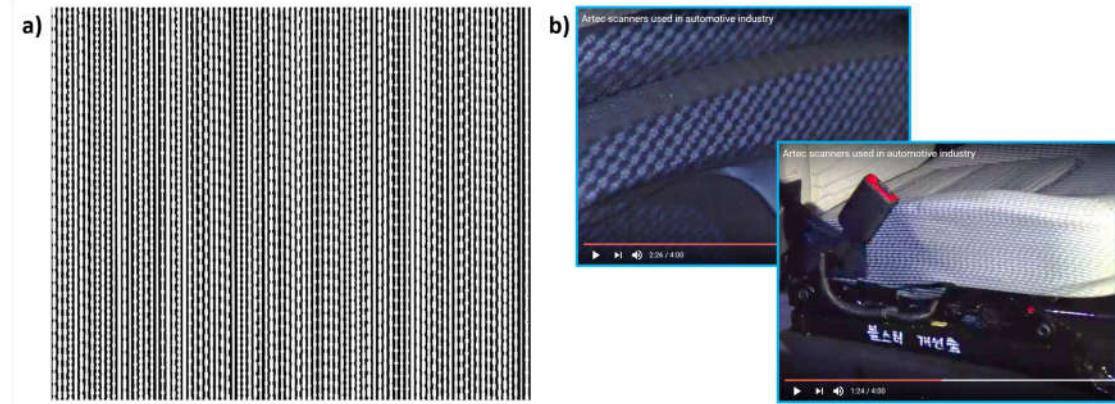


Figure 3.9: Pattern projected by Artec sensors. a) Original figure taken from patent [71], showing a representative pattern to be projected. b) Screenshots from a product video [72]. The projected pattern is clearly visible and resembles the pattern shown in (a).

object frequencies. Features, smaller than the required segment, cannot be measured in line direction.

The codification also consumes space bandwidth in the perpendicular direction due to the thickness variations. This limits the number of projected lines and therefore the data density of the sensor principle. According to Fig. 3.9(a), the projected pattern is assumed to consist of  $\sim 80$  coded lines. With this an upper bound for the data density can be calculated to  $\rho_{\text{Artec}} < 8\%$  for a 1-Megapixel camera ( $1000 \times 1000 \text{ pix}$ ). The real data density is estimated smaller, since only *independent* points are considered. The value above has to be divided by the minimal line segment length (unknown) that ensures the identification of the ‘code’. For a reasonable value of  $10 \text{ pix}$  to  $15 \text{ pix}$ , the data density shrinks to  $\rho_{\text{Artec}} < 0.8\%$ . As already known from other sensors, the low data density is concealed to a large extent by interpolation and high resolution color texture, which is acquired during an additional flash exposure with a color camera.

Although not indicated by the references, it is not impossible that ‘Artec’ also uses a short temporal sequence of exposures or 3D frames to increase the data density or to use the information for correspondence. The drawbacks of this approach have already been discussed.

The approach of coded multi-line triangulation is not exclusively followed by Artec. Similar approaches with similar advantages and drawbacks are e.g. [74, 75]. The Erlangen Institute of Optics investigated a line codification approach as well [6, 76–78] to increase the number of projected lines in Flying Triangulation. Here, the thickness of the projected lines remains constant, but the intensity profile is coded with a sinusoidal pattern of different frequencies to distinguish the lines. Such intensity codings can help to save ‘space’ on the camera chip, but reduce the *SNR* in the camera image.

### 3.3.5 Color Coded Triangulation

Instead of their shape, lines can also be distinguished by their *color*. The color coding methods follow exactly this approach. Color cameras are required for the measurement. Most cameras acquire color texture information with a ‘Bayer pattern’ - a pattern of small color filters, mounted in front of the camera chip. Each small filter is positioned directly in front of one pixel. One  $2 \times 2$  ‘elementary cell’ of the filter pattern consists of two green, one red and one blue filter

(rarely: two yellow, one magenta and one cyan)<sup>6</sup>. This has two major drawbacks: First, the effective chip resolution of the camera is reduced. Each image point contains information from at least three pixels. Most cameras are able to export a red, green and blue channel picture in the original chip resolution. However, these pictures are interpolated. Second, pixels behind the green, red and blue filters have different *SNRs*. Depending on the light source, the transmitted intensity strongly varies between the different filters. If e.g., a line intensity maximum has to be calculated, this is a source of noise. However, expensive ‘three-chip-cameras’ can solve the discussed problems to a certain extend.

Another problem of the color coded triangulation is the acquisition of the intrinsic object color texture. Since object texture is not ‘spectrally pure’, the reflected color is strongly dependent on the used light source. This can be seen as an artifact in the 3D model. In many sensor principles an additional flash exposure is used for the object texture acquisition. The lost single-shot ability is the price.

Several approaches for color coded triangulation can be found in literature.

The most intuitive approach is to project different color coded lines. The authors of e.g. [79, 83–85] use this method. But this is not all. In many cases, the pattern *period* of a line color is varied, too, like shown in Fig. 3.10(a). This helps to reduce the number of projected colors, because additional correlation information is created. However, distinguishing lines by pattern periods requires neighborhood information. Spatial bandwidth is consumed. The 3D models display larger artifacts at edges, were no neighborhood information is available.

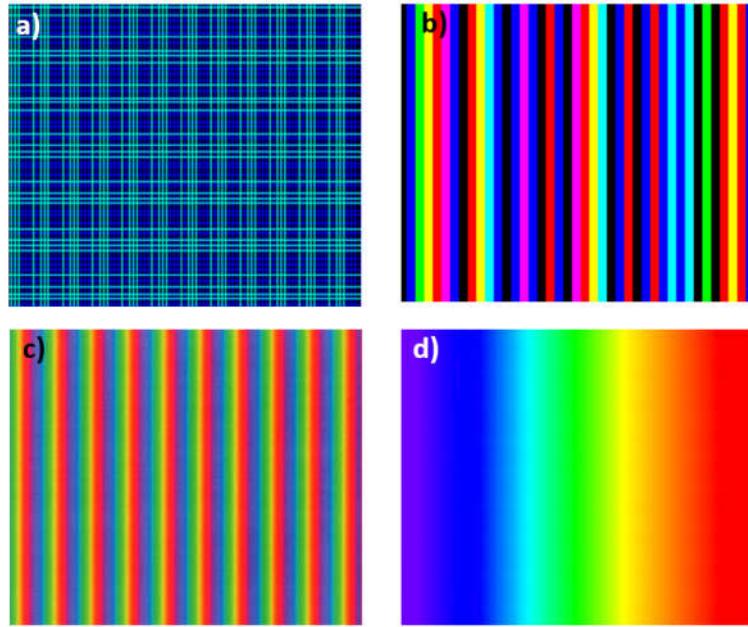


Figure 3.10: Examples for projected patterns in color coded triangulation approaches. a) Color coded lines with varied pattern period, taken from [79]. b) Color coded stripes, taken from [80]. c) Color coded and phase shifted sinusoidal fringes, taken from [81]. d) Continuous color spectrum, as it is used in [82].

<sup>6</sup> The reason, that green appears twice as often as red and blue is related to the spectral emission curve of the sun and the spectral sensitivity of the human eye (photopic  $V_\lambda$ -Curve). Both have their maximum in green. For a transformation of color images to gray images, it is in many cases sufficient just to look at the green channel.

Other approaches use the projection of color coded stripes, proposed e.g. by [80, 86, 87]. In order to correctly distinguish between the projected colors, the projected stripes have to have a certain width (see Fig. 3.10(b)). To calculate 3D data, the edges between two stripe colors are triangulated. Edge detection always involves differentiation operations, which are very noise sensitive. Note as well, that some approaches like [80] again rely on information of the extended neighborhood.

Another option to be discussed is inspired by PMT-phase shifting (see section 3.2.1): Instead of projecting three phase shifted sinusoidal patterns subsequently, it is also possible to use a different color for each phase shifted pattern (red, green, blue) and project everything in one picture, as shown in Fig. 3.10(c). Methods that follow this idea are described e.g. in [81, 88]. Note that the projection of several phase shifts in one colored image does not include the solution of the ambiguity problem, which still requires measurements with multiple fringe-frequencies (multi-shot, as in PMT).

The last option to be discussed is ‘Rainbow Triangulation’ [82, 89], invented by Häusler et.al. in 1993. The projected ‘pattern’ consists of a continuous spectrum of wavelengths (‘rainbow’), that can be generated by the dispersion of a white light beam through a prism. Such a projection, like shown in Fig. 3.10(d) can be seen as a pattern of *densely* projected lines, each coded with a *unique* spectral color. This means, that the principle delivers ‘*pixel-dense*’ 3D data in each single-shot *without correspondence ambiguities*. However, projecting a bright spectrum is technically elaborate and requires high power broadband light sources. Nevertheless, the principle might become more important in the near future, by applying newly developed supercontinuum lasers.

### 3.4 Discussion of Triangulation Principles

In the last two sections, it became obvious that there is no ‘perfect’ triangulation sensor. Each approach has its benefits and drawbacks. It also became clear, that the decision, if a sensor property is *really* a drawback is strongly dependent on the desired application. For the discussion of the preceding examples, the *beneficial properties of triangulation sensors* are divided in *three groups*:

- **No temporal bandwidth restrictions.** This excludes all methods relying on a temporal sequence of exposures for the generation of one 3D frame. Such methods are not single-shot.
- **No spatial bandwidth restrictions** (within the limits of the sampling theorem). This excludes all methods relying on spatial codification, color codification or requiring relatively ‘smooth’ object surfaces. Such methods are not able to measure high frequency components of the object surface.
- **No ambiguity problem.** Such methods are principally able to reach a high density of 3D data. However, if extensive spatial codification is exploited for the solution of the ambiguity problem, the 3D data density might be reduced.

Having in mind this classification and looking at the discussed sensor principles again, it is remarkable, that each sensor principle displays *not more than two* of the described properties at the same time. Figure 3.11 shows a graphical representation.

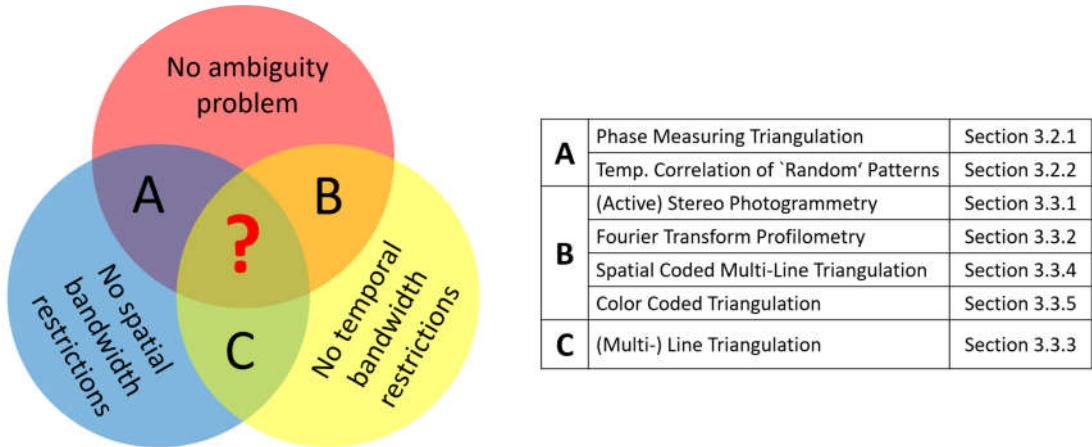


Figure 3.11: Categorization of triangulation principles.

The main question is:

'Why is there no sensor, that displays all three properties at once?'

'Is this fundamental?'.

The answer is: 'yes'.

This answer was the motivation and starting point of the presented thesis.

In chapter 5 an approach is introduced, which indeed displays all three properties at once. As mentioned, this would not be possible without purchasing something.

### 3.5 Other Methods, not based on Classical Triangulation

As discussed, most optical measurement principles for the acquisition of macroscopic objects with scattering surfaces are based on triangulation. However, a few principles, enabling this task, are not. For the sake of completeness, the most important approaches are discussed in this section.

#### 3.5.1 Time-of-Flight

Besides triangulation a distance can be also measured by *runtime*. Runtime-based approaches for distance measurement are already well known from Radar, Sonography (Ultrasound) or Interferometry. *Time-of-Flight* (ToF) 3D cameras, having similar measurement tasks like the system presented in this thesis, are mostly based on the so-called PMD chip technology (*Photonic Mixing Device*), invented by Schwarte [90]: A *temporally modulated* electromagnetic wave is radiated onto the object and reflected from its surface. In each pixel of the PMD chip, the phase shift  $\Delta\varphi$  between outgoing and incoming wave is measured, which deciphers the object distance  $z$  by

$$z = \frac{1}{2} \cdot \frac{c}{f} \cdot \frac{\Delta\varphi}{2\pi}, \quad (3.7)$$

where  $f$  is the modulation frequency. Commonly, a sinusoidal modulation is used, but also other modulations, like separated pulses, are possible. The basic principle of a PMD chip can be understood as follows: Each pixel consists of at least two light-sensitive modulation