Unterschrift BetreuerIn

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

# Light-field Imaging in Automatic Inspection

ausgeführt am Institut für Angewandte Physik
der Technischen Universität Wien

unter der Anleitung von
**Univ.-Prof. Dr. Gerhard Schütz**

durch

**Branislav Holländer**

Wien

November 23, 2014

Unterschrift StudentIn

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschliesslich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, am 20.11.2014

# Kurzfassung

Die Lightfeld-Bildgebung ist ein aufkommendes Gebiet im Bereich Computational Imaging. Ihr Anwendungsgebiet reicht dabei von der Fotografie und Mikroskopie bis hin zur industriellen Bildgebung. Die Hauptidee besteht darin, dass gleichzeitig räumliche und winkelabhängige Strahlungsinformation auf einen 2-dimensionalen Sensor aufgenommen wird. Das Ziel dieser Arbeit ist es, ein Bildgebendes System zu entwickeln, das das Linienscanverfahren, das in der automatischen Inspektion genutzt wird, mit Lichtfeldbildgebung kombiniert, und dessen Leistung zu evaluieren.

Zu Beginn der Arbeit wird die Theorie der Lichtfelder diskutiert und eine schlüssige mathematische Formulierung vorgestellt. Danach wird das Prinzip des Systems erklärt und der Winkelbereich, die Winkelauflösung, die Tiefenauflösung und die Tiefenschärfe werden hergeleitet. Ausserdem wird ein alternatives System vorgeschlagen, das eine Schlitzblende verwendet, um den Winkelbereich zu vergrössern. Es werden sowohl die Vor- und Nachteile beider Systeme, als auch Nachbearbeitungssschritte, wie die Flachfeldkorrektur oder das Demosaicing, diskutiert. Schliesslich wird der Tiefenschätzungsalgorithmus kurz beschrieben. Die Systemleisutng und die Tiefenauflösung werden sowohl durch ein Ground-truth-objekt als auch in realistischen Szenarien evaluiert. Die Möglichkeit, verschiedene Ansichten zu generieren, wird durch Optical Variable Devices (OVDs) demonstriert.

# Abstract

Light-field imaging is an emerging area of computational imaging, whose applications range from photography through microscopy to industrial imaging. Its main idea lies in multiplexing spatial and angular radiance information onto a 2-dimensional sensor. The aim of this thesis is to construct an imaging system combining the line-scan acquisition process used in automatic visual inspection with light-field imaging and assess its performance for a variety of inspection scenarios.

First, the thesis examines the theory of light fields, presenting a concise mathematical description. Secondly, the principle of the system is explained and angular range, depth resolution and depth-of-field is derived. Additionally, an alternative system with a slit aperture is proposed to enhance the angular range. Advantages and drawbacks of both setups are discussed, including post-processing steps such as flat-field correction and demosaicing. Finally, a short description of the depth estimation algorithm is given. System performance and depth resolution is evaluated using a ground truth object as well as in real-case scenarios. The multi-view capability is demonstrated on Optical Variable Devices (OVDs).

# Contents

# 1 Introduction

## 1.1 Multiview imaging

Photography can be seen as one of the most important inventions of the 19th century. Since then, it penetrated nearly every aspect of science, technology and our everyday lives. It provides us with an opportunity to capture and reproduce what we see with our eyes as well as the possibility to study what cannot be seen naturally, either because of the insufficient resolution of our eyes (microscopy), their limited spectral range (x-ray imaging) or a completely different imaging mechanism (magnetic resonance imaging). Nevertheless, there is at least one aspect where human vision is superior to most traditional imaging techniques - the fact that we have two eyes enables us (after the necessary post-processing in the brain) to see our environment in three dimensions. Researchers tried to reproduce this capability since the invention of the stereograph by Sir Charles Wheatstone in 1833 [Whe38], however, it was not until the end of 20th century, that computers could at least partially replace the human brain in inferring depth information indirectly from two or more views, mimicking human vision.

As these methods evolved, so did their applicability in *machine vision*. Machine vision deals with automated imaging-based inspection and analysis of manufactured goods, often with high demands on reliability and quality, e.g. for the quality control of various types of materials, medical devices, *printed-circuit boards* (PCBs) or even banknotes. Many of these objects are 3-dimensional, and need to be inspected in their entirety. For these, it is only natural to choose an imaging system providing depth information for their inspection, which in turn often requires obtaining information on multiple viewing and/or illumination angles.

To this day, a broad variety of optical techniques was developed to obtain multiple views. One that ranges among the most prominent is *stereo imaging*, which builds upon the aforementioned stereograph and assumes the sufficiency of two 2-dimensional images to reconstruct depth information in the computer using *stereo matching algorithms*. The goal of these algorithms, namely to find corresponding object points in the stereo image pair and determining their mutual distance (the so-called *disparity*), constitutes a very large topic in the field of computational imaging. A multitude of stereo matching algorithms have been proposed in the past (see e.g. [NSG07], [SS02] or [Tip+13]), which are able to deliver acceptable

performance and results for many machine vision applications. However, most of them are restricted in their accuracy by only using two views of the scene. This may cause problems when dealing with repeatable patterns, as is often the case in automated inspection (take for example printed matter, such as banknotes), as well as with situations where a part of the scene is occluded in one of the views.

*Structure from motion* (SfM) is a technique which explicitly requires movement in order to derive depth information from a sequence of 2-dimensional images (video) [Del+00]. There are two possibilities to do this: either by first deriving trajectories (i.e. correspondences) between certain points of the scene in multiple images and then trying to reconstruct 3-dimensional information, or reconstructing depth directly without a-priori knowledge of correspondences, e.g. by using a maximum likelihood method.

Another popular and widely used, but wholly different technique is holography, originally developed by Gabor in 1948 [Gab48]. Holography differs from many other imaging methods by typically requiring coherent light of a single wavelength to capture both amplitude and phase of the recorded light, which can in turn be used to reconstruct a 3-dimensional image. The reconstruction can be performed either optically by illuminating the hologram with light of the recording wavelength, or computationally [ML02]. The computational method is especially interesting for automated inspection, as it works without any additional illumination steps. A big advantage of holograms is that they inherently contain any possible view from a limited, but continuous viewpoint range. The disadvantage lies in the complex calculations involved in the reconstruction. This makes the implementation of real-time holographic imaging tedious at best.

As it turns out, the usability of the various multiview methods in industrial imaging varies according to the application and the requirements, such as depth resolution or stability of the results. If, however, a system were built which would provide reliable depth information while at the same time eliminating some fundamental disadvantages (e.g. the ambiguity of the correspondence problem in stereo imaging or the dependence on monochromatic light, as in holography), it would constitute a major step in the context of 3-dimensional imaging.

## 1.2 Light fields

Thinking in the context of ray optics, having multiple views of a particular scene corresponds to recording rays entering the imaging device from different directions separately. The more views one wants to obtain, the more separate ray directions have to be captured. In the limiting case of capturing a continuous range of directions on a continuous spatial interval, we speak of recording the *light-field function* on that interval. The light-field function is, as many descriptions of

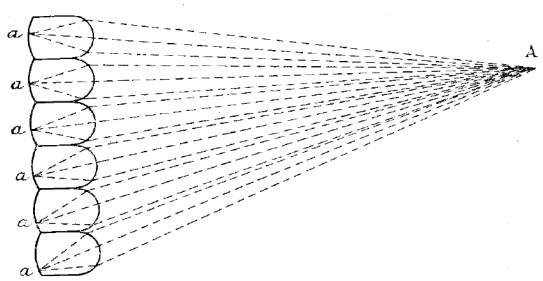


Figure 1.1: Graphical visualization of Lippmann's idea [Lip08]. An arbitrary object point $A$ is imaged on several points a in multiple microscopic images. The positions depend on the position of $A$ w.r.t. the imaging system.

continua in physics, an artificial concept. It provides the basic means to describe a set of light rays by a single function. When capturing a light-field by a light-field imaging device, in fact the light-field function is sampled at discrete points. A variety of such devices have been designed in the past century. The next few sections will provide an overview of those.

## 1.3 Light-field imaging using pinhole- and microlens-arrays

In order to sample the light-field function, one has to capture information in the angular direction as well as in the spatial direction simultaneously on a 2-dimensional sensor (or, as of earlier, on a photoplate). The first invention for such an acquisition came from Gabriel Lippmann, a Nobel Prize laureate for color photography and the advisor of Marie Skłodowska-Curie, in 1908. His main idea was to place a set of microlenses on a specially shaped photo emulsion, creating an optical system similar to an insect's eye. Each of the microlenses produces a microscopic image on the spherically shaped emulsion behind it (albeit each from a slightly different viewing angle), thus effectively creating multiple cameras on one photo emulsion. As one can see in Fig.1.1, an object point A which is in focus w.r.t. the system, is imaged multiple times on the emulsion, each time under a slightly different angle and in a slightly different position in the microscopic images. When looking at the developed plaque from the capturing side, the microscopic images produce a converging bundle of light rays, i.e. an image, at the position of A. Thus, for reconstruction of the 3-dimensional structure of the captured scene, the observer

needs only to look at the emulsion from the same side as it was captured, in an arbitrary distance and lit from behind by a diffuse light source (Lippmann talks about holding the plaque against a piece of diffuse glass), and he is able to see a particular view of the captured scene. This is in contrast to holography, where the observer has to look through the hologram to see the reconstructed object, while the illumination is placed on the same side as during recording. Lippmann called his method integral photography, because the observed image is integrated from the multitude of microscopic images on the emulsion. However, from today's point of view, the terms "integral photography" and "light-field imaging" are synonyms and "light-field imaging" will be used throughout this work to avoid confusion.

While Lippmann himself was unable to manufacture an actual working prototype of his system because of the technology at that time and the complexity of the emulsion surface, Sokolov [Sok11] manufactured an integral imaging system based on a much simpler idea, namely pinhole cameras (see Fig.1.2). By choosing a conical pinhole camera array over the much more complex microlens-based array and placing it directly over the sensor, he constructed the first practical light-field imaging setup. However, because of the pinhole approach, the system suffered greatly from the long exposure times required to do an actual acquisition.

Much later, Ives [Ive30] combined the idea of Sokolov's pinhole-based light-field system with a large main lens placed in front of the lenslet- or pinhole-array. This changed the way one could think about the light-field by changing the structure of the captured data.

In his setup the main lens is focused on the pinholes, while the sensor plane lies slightly behind the pinholes. Thus, the resulting image on the photographic film is completely defocused. The beauty of the system lies in the separation between spatial and angular information on the resulting image. Looking at Fig.1.3, it becomes clear that the image on the image plane is composed of multiple micro-images, each corresponding to a single pinhole. The position in the micro-image determines the angle, at which light passed through the pinhole.

A particular disadvantage of Ives' system is the fact that the spatial resolution of the image is restricted by the number of pinholes. Another technical difficulty which needs to be taken care of is the adjustment of the F-number of the main lens w.r.t. the spacing of the pinholes, in such a way, that the micro-images do not overlap, but are spaced to fill out the image area, to record the maximum amount of information.

Later in the 20th century, more variations of the previous methods surfaced, such as light-field imaging with diffraction lenses [Vil89]. However, in 1992, a true breakthrough in light-field imaging was achieved as Adelson considered processing and enhancing the information contained in the light-fields computationally. In particular, he was the first to render multiple views of a scene out of light-field

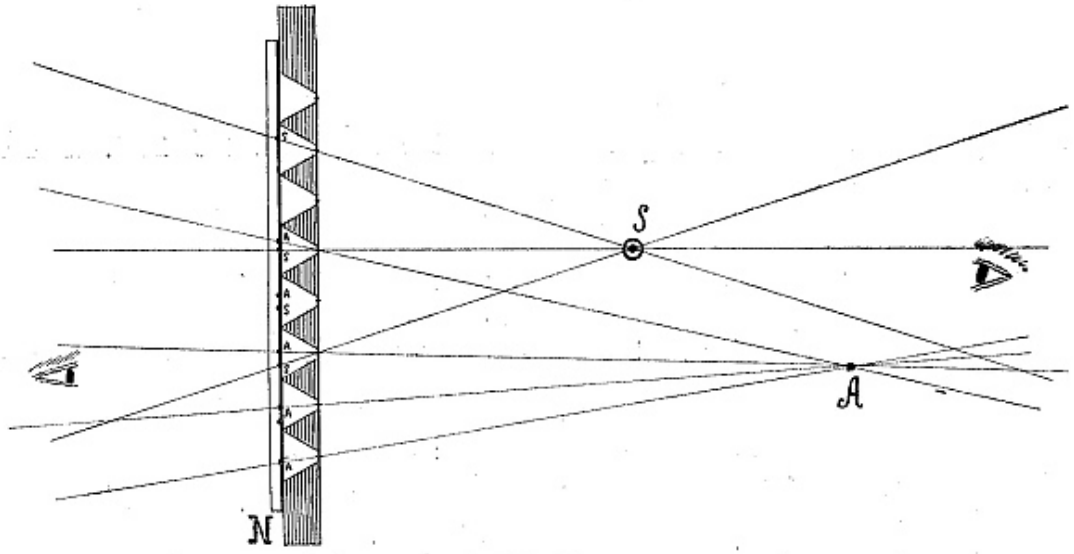


Figure 1.2: Pinhole camera array by Sokolov [Sok11]. $N$ depicts a photo plate with a conical-shaped pinhole array on top of it. Two objects $A$ and $S$ get imaged on different parts of the photographic plate, thus multiple small images of the scene are created. By looking at the exposed plate from the other direction (lit by a diffuse light source), the scene is recreated on the same side as it was imaged.

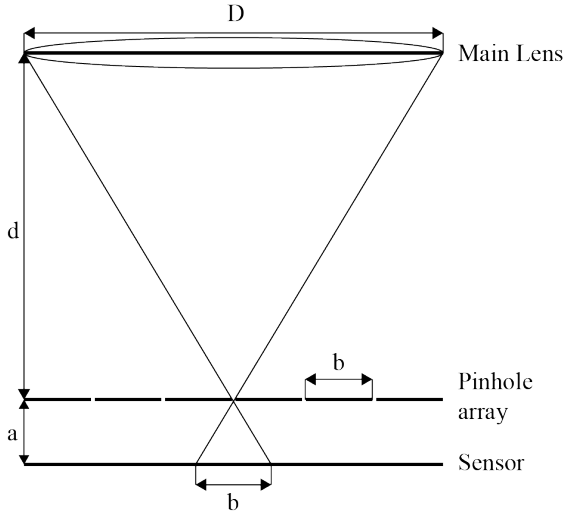


Figure 1.3: Pinhole camera by Ives.

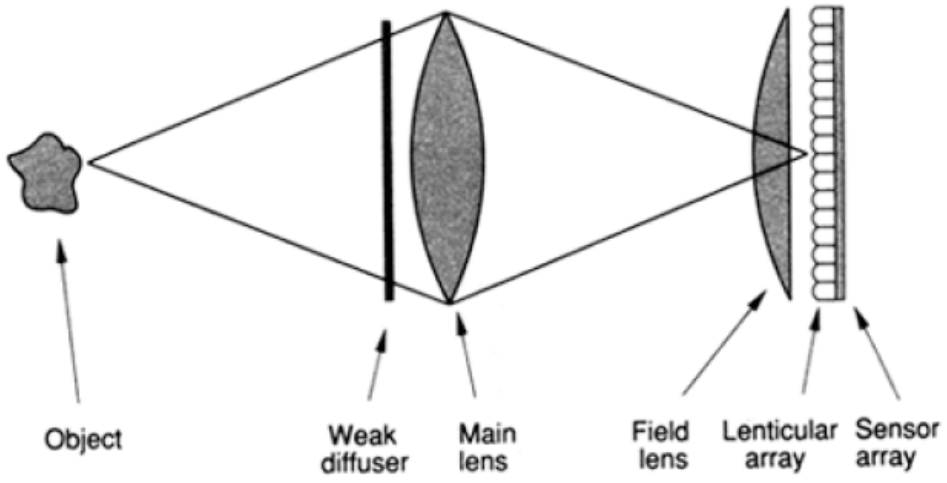


Figure 1.4: Adelson's setup, as presented in [AW92]. From Adelson's paper: "The object at the left is imaged by the main lens onto a lenticular array, behind which lies a sensor array such as a CCD. Each lenticule gathers light over one macropixel and forms a microimage of the main lens aperture on the sensor elements beneath it. The field lens places the main lens aperture at optical infinity from the lenticules so that rays from the center of the main lens are imaged in the center of each microimage. A weak diffuser may also be used (shown here in front of the main lens but really placed in the main lens aperture plane) to prevent aliasing due to the sampling of the image by the lenticules. The weak diffuser is designed to blur the image very slightly with a point spread function roughly as wide as the spacing between lenticules."

data on a computer, and subsequently perform 3-D reconstruction [AW92]. His system used a combination of a large main lens, microlenses and a field lens. On the image in Fig.1.4, one can see that the system is setup such that the image plane of the main lens corresponds with the lenslet plane, with the lenslets focused on infinity. This makes it similar to the pinhole system proposed by Ives, with the additional advantage of more light transmission.

A wide variety of light-field imaging systems using computational post-processing emerged afterwards. For example, Ng [Ng+05] presented a hand-held light-field camera for computational refocusing. In the camera, which is otherwise identical to Adelson's camera, the field lens was ditched in the setup in favor of a lighter, more compact system. Otherwise, the principles of the system remain the same, with the main lens being focused on the microlenses and the microlenses being focused on infinity, resulting in a raw image as seen in Fig.1.5.

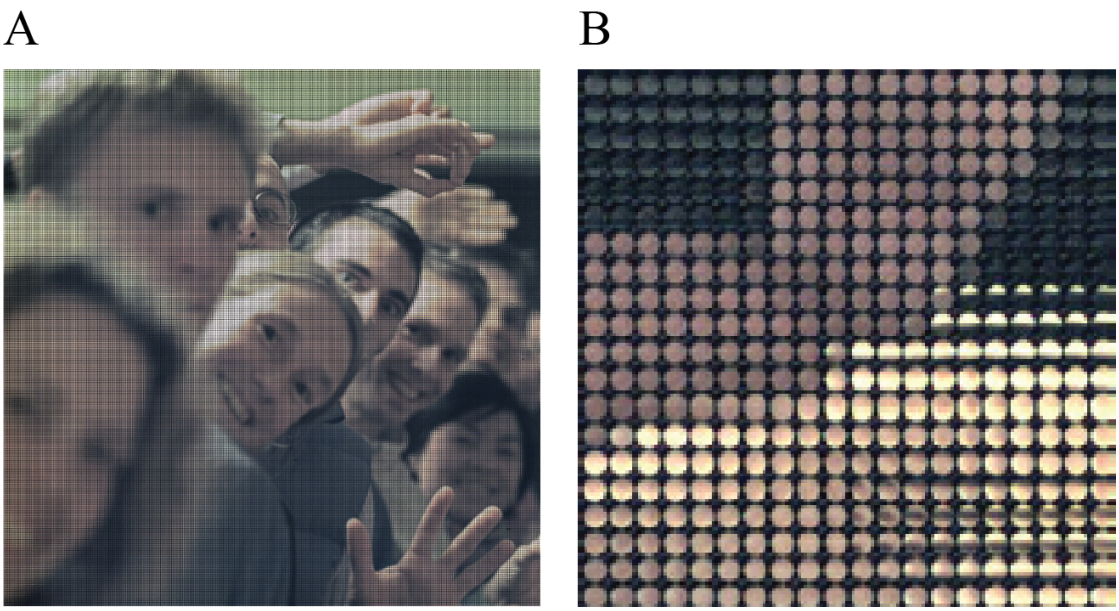


Figure 1.5: Image obtained by Ng's hand-held plenoptic camera: (A) Raw image, (B) Close-up view of the raw image. Notice that the pink pixels (corresponding to the clasped hand in the raw image) are completely defocused. This is an area in the image which is focused by the main lens, whereas the other pixels which contain more or less sharp edges are placed further or nearer to the camera. Courtesy of [Ng+05].

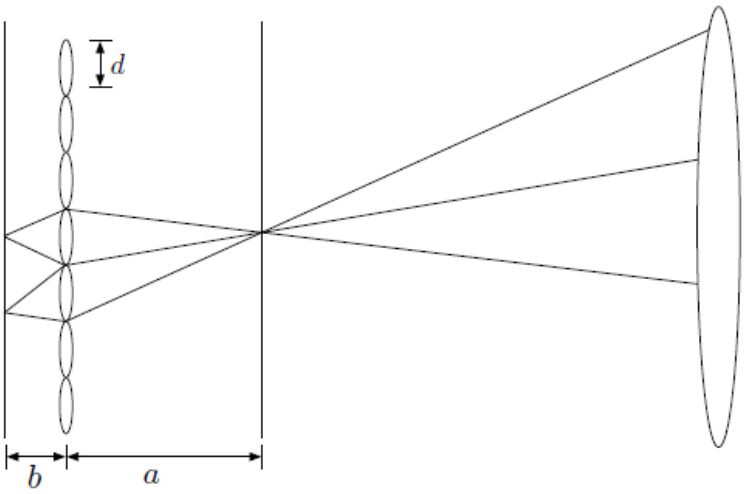


Figure 1.6: The focused plenoptic camera ray diagram, courtesy of [LG09].

A further improvement in light-field imaging was achieved by Georgiev et al. [LG09]. This work addressed the major shortcoming of previous plenoptic cameras, namely their reduced lateral resolution in comparison to standard photographs. Georgiev proposed a novel light-field imaging system, which doesn't constrain lateral resolution to the number of microlenses. It also makes use of a main lens and a microlens array, but in this case, the main lens is not focused on the microlenses, but rather on an intermediate image plane, with the microlenses also being focused on the same image plane (for clarification, see Fig.1.6). The microlenses then act as a relay imaging system with the main lens by imaging individual parts of the image plane on the sensor. If the image plane of the main lens is replaced by the actual imaged scene, the setup is similar to the camera array setup mentioned earlier. The microimages on the sensor are now focused, and the spatio-angular trade-off (i.e. the balance between spatial and angular resolution) is not anymore determined by the number of microlenses and the number of pixels under each microlens, but rather can be freely chosen by varying the geometry of the setup (i.e. the distances $a$ and $b$ between the main image plane, the microlenses and the sensor, respectively). The difference in rendering is that the angular samples of one particular point of the scene are now sampled by different microlenses, rather than by a single microlens, as is the case in Ives' plenoptic camera. Fig.1.7 shows the raw sensor image of the captured light-field as well as a view rendered from this light-field.

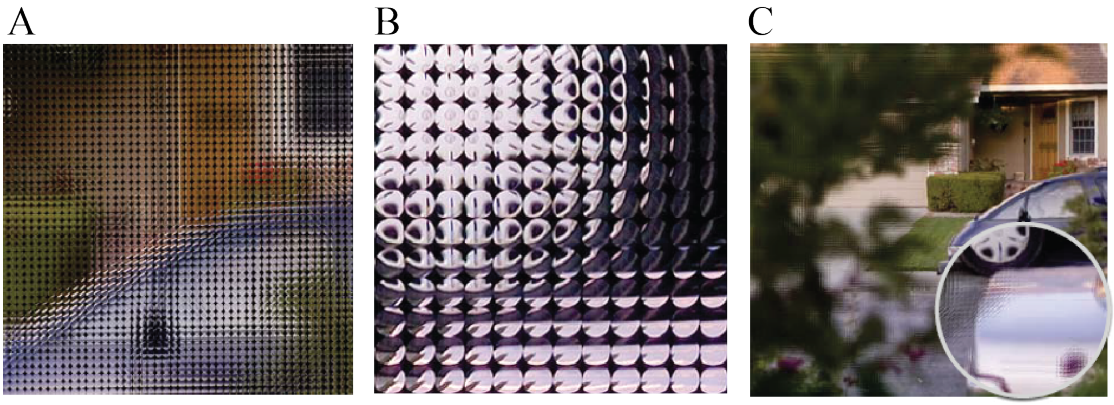


Figure 1.7: Gerogiev's Focused Plenoptic Camera. (A) Raw image recorded on the sensor. (B) Closeup view of the raw image. Obviously, the image is focused, as opposed to Fig.1.5. (C) An arbitrary view rendered from the light-field using the "focused plenoptic algorithm". Courtesy of [LG09].

## 1.4 Camera arrays

When sampling the light-field using only a microlens array, as in Lippmann's setup, actually an array of cameras is being emulated. Thus it is not a far-fetched idea to capture light-fields using a real array of cameras. However, such a setup only samples the light field sparsely, in contrast to the microlens setup, which might introduce aliasing effects when reconstructing the views. Although a camera array is typically large and bulky, it is capable of recording dynamic scenes in a resolution which is much higher than the resolution of microlens- or pinhole-based plenoptic cameras because of the large combined resolutions of all sensors in the system. However, the multitude of cameras used requires precise adjustment, post-processing and registration of the images (e.g. finding the overlaps in all individual images) to produce a light-field which can then further be used for additional computational tasks. Additionally, each camera has to be corrected for aberrations, as opposed to the microlens setups which only require the main lens to have high correction.

An example for such a system can be found in [Wil+05]. In that work, a $12 \times 8$-array of 30 Hz video cameras is used. The cameras are triggered consequently 9 times during each interval to increase the frame rate of the recorded video. Making use of the higher spatio-temporal sampling and optical flow algorithms, video quality was shown to be largely improved. Additionally, in this way a light-field video was recorded for the first time, enabling effects such as the famous "bullet-time" known from the movie Matrix. Additional effects were achieved by

setting different apertures and exposure times for the individual cameras in the array. In this way, "hybrid aperture" images were captured, utilizing different apertures in different parts of the image.

## 1.5 Light-field microscopy

Various approaches to bring light-field technology into the microscopic world have been explored recently, most notably [Lev+06] for the investigation of biological fluorescent samples, as well as [GAS09] for the in-vivo imaging of calcium dynamics in neurons of a "larval zebrafish watching a prey-like stimulus". In this paper, light-field imaging was chosen because of its ability to reconstruct the (fluorescent) sample volume without the need of a scanning technique, thus enabling in-vivo imaging, which was anticipated by the researchers. A custom microscope was built for this task, using a 63x 0.9 NA water immersion objective, as well as a microlens array, which was placed on the intermediate image plane of the microscope, and a *charge-coupled device* (CCD) sensor placed on the image plane of the microlenses. A rate of 4fps was achieved using this setup, which was deemed appropriate for the required task. After the acquisition, the video data were processed with *Independent Component Analysis* (ICA) to localize the stimulus. The method described is in many ways similar to methods used in *Functional Magnetic Resonance Imaging* (fMRI). Another very recent result in this research area is presented in [Pre+14]. In this paper, 3-D calcium neuronal imaging with a resolution capable of recording single neurons was performed in a whole *Caenorhabditis elegans*, as well as on the brain of a larval zebrafish with a reconstructed area of cca. $700 \times 700 \times 200 \mu m$.

These examples of microscopic light-field imaging serve to illustrate the multiplicity of applications of this technology. Concerning industrial microscopy, however, there is only one available known light-field system, produced by the company Raytrix [Gmb14b]. This system uses a standard microscopic objective, together with a microlens array, this time according to the *focused plenoptic camera* principle discussed above, which is capable of delivering higher resolutions than the standard defocused setup. Although this system seems to have very good resolving power, it can only be used in area-scan imaging. Furthermore, its cost is prohibitive for wide-spread usage.

Although light-fields in microscopy continue to be a promising technique because of the ability of in-vivo 3-D imaging with relatively high frequencies, they suffer under the restriction that most microscopic objectives are infinity-corrected, i.e. they possess no intermediate image plane. Thus, without the use of additional optical elements, the recording of a light-field is not possible. Another shortcoming of the technique concerning volume reconstruction is the fact that light-fields

are only strictly defined in a space free of occluders. This leads to artifacts when reconstructing biological samples, which, although usually transparent, tend to scatter light rays in the volume. Another, even more fundamental disadvantage is connected to the wave-like nature of light on small scales. This restricts the resolution not only in the lateral direction, but also along the optical axis, and leads to lower depth resolutions in the reconstructed images. A recent paper [Bro+13] addresses this problem and proposes a deconvolution algorithm, capable of delivering very high-quality volume reconstruction by making use of the knowledge of the *point-spread function* (PSF) of the system.

## 1.6 Techniques related to light-field imaging

There are other techniques which make use of light-fields without mentioning them explicitly, such as the Shack-Hartmann wavefront sensor [PS01]. This is a wavefront sensor measuring perturbations in optical wavefronts by using a microlens array focused on a (typically) CCD sensor. For a perfectly planar wavefront, light is focused exactly behind the microlenses. Lateral deviations of these spots from the central position then indicate the direction of the wavefront at the recorded position. Thus, the system practically records the incoming light-field (its origins lie in the year 1900, even earlier than Lippmann's integral imaging system). Usage of the Shack-Hartmann sensor ranges from determining and improving the optical quality of telescopes and satellite images through testing laser beams for quality all the way to measuring the profile of the cornea in ophthalmology (by projecting a point source onto the retina and measuring the reflected wavefront).

Another approach, similar to our line-scan light-field system (which will be introduced in the next chapter), albeit on a completely different scale, is the *push broom* (PB) principle used in remote sensing [Pet05]. Push Broom imaging uses sensors mounted on an aircraft or satellite recording multiple lines with single optics, where each line obtains a slightly different angular view of the earth's surface. While the individual lines usually record multiple spectral ranges, in the *High Resolution Stereo Camera* (HRSC) on the Mars Express Spacecraft [Jau+07], these are also used to obtain 3-D information about the surface. Although such sensors typically do not provide more than three different views, the system can be considered as a very sparse 3-D light-field acquisition system. Push Broom imaging is, in fact, most related to our previously published mirror-based multi-line-scan camera which provides three views of an inspected object for stereo processing with a large angular difference between the individual views [HŠH13].

## 1.7 Light-fields in machine vision?

Having described these light-field setups, it is important to decide which of these setups would fit into the context of automatic inspection. It turns out that for some inspection scenarios, a good choice would be to use a 4-D light-field acquisition system using microlenses, which can be adapted to different scales and needs. Given a precise setup and sufficient surface structure, it provides digital refocusing and absolute depth estimation with a single camera and a single chip. On the other hand, it suffers from reduced resolution, which is often a prohibiting disadvantage. Furthermore, it requires recalculating the image data in order to generate synthetic views of the objects, adding to the complexity of the algorithmic steps. Additionally, the acquisition speed of microlens systems is relatively slow because of the requirement to illuminate and read out large chip areas. Therefore, the need arises for a system which would provide a compromise between the achievable speed and the amount of data generated, which would be compact and durable to withstand the harsh conditions often encountered in manufacturing facilities and which would be specifically designed for in-line inspection to allow inspection without human intervention.

## 1.8 Thesis overview

The thesis is divided in two main parts:

1. Theoretical background on light fields (Chapter 2).

2. Experimental part focusing on the line-scan light-field imaging system (Chapters 3-5).

In Chapter 2, a rigorous definition of the light-field function is given, various formulations of light fields in literature are discussed, light fields of some simple wavefronts are presented to gain an intuitive understanding and an overview about the behavior of the light-field function under various transformations, such as free-space travel or lens refraction, is provided.

Further, Chapter 3 explains the principle of our line-scan light-field system and derives its various properties, such as how it samples the light field, as well as the achievable depth resolution, angular range and angular resolution. It also introduces an alternative line-scan light-field setup with a slit aperture inserted between the object and the camera.

In the following chapter, the methodology of this work is presented along with a detailed description of the actual setup. Acquisition and illumination is discussed in detail, as well as computational correction methods applied to the light

field. Afterwards, the algorithms used for depth estimation and all-in-focus image composition are discussed.

Chapter 5 presents and discusses the results of visual inspection of a variety of different objects, including a detailed depth resolution analysis using a ground truth object, while Chapter 6 concludes the thesis.

# 2 Description of light-fields

## 2.1 Traditional imaging and light-fields

When speaking of capturing images on a 2-dimensional sensor, as is the case for example in traditional photography, each pixel of the sensor captures light coming from all directions permitted by the optics passing through the respective pixel position. Thus, one can imagine this as photons causing a charge proportional to the *integral* of the energy of all photons hitting the pixel, i.e.

$$I(x,y) = \int_A \int_\theta \int_\phi p_{sensor}(x,y,\theta,\phi) d\phi d\theta dx dy \tag{2.1}$$

where $p_{sensor}(x,y,\theta,\phi)$ is the photon distribution on the sensor plane for photons coming from the direction described by the angles $(\theta,\phi)$ and $A$ corresponds to the area of one pixel. This approach to imaging is thus *integral* in the sense that no information about the radiance in specific directions is retained.

However, there are other possibilities of capturing radiance on a sensor. In fact, it is possible to arbitrarily divide up the area of the sensor to record both *spatial* information, such as in traditional photography, and *angular* information about the direction of photons registered by the system. Such a system mathematically samples a 4-dimensional function, whose two dimensions correspond to spatial and two dimensions to angular resolution, respectively. This function, labeled $\mathcal{L}(x,y,\theta,\phi)$, is called the *4-dimensional light-field function*. There is another type of light-field function, called the *3-dimensional light-field function*, which has one of the angular dimensions omitted, i.e. $\mathcal{L}(x,y,\theta)$. This type of light-field function is used in our line-scan light-field acquisition setup and will be treated in detail in Chapter 3. Of course, sampling the light field instead of a flat image on the same sensor leads to a subsequent reduction in spatial resolution, and is in fact a fundamental limit in light-field imaging (see [Geo+06] for further information). Additionally, the choice of sampling frequency, both in the spatial and the angular direction, is crucial to artifact-free light fields, and has to be chosen to conform with the sampling theorem (more details about this can be found in Sec.4.1.4).

Mathematically, the sampled points of a light-field function constitute the (4-D- or 3-D-)tensor $L$. Even though one cannot directly visualize the light-field tensor because of its high dimensionality, if one chooses a specific angular direction

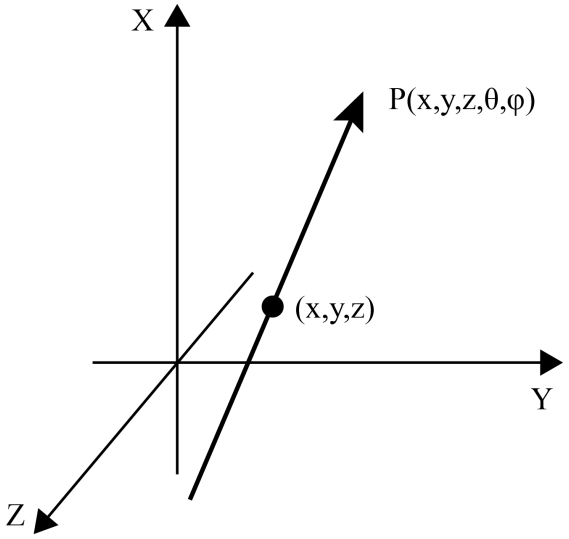


Figure 2.1: The 5-dimensional plenoptic function $P(x, y, z, \theta, \phi)$.

$(\theta_i, \phi_i)$, $L(x, y, \theta_i, \phi_i)$ corresponds to a particular *view* of the scene, as if viewed from the desired direction. In a similar way, one can choose a specific $x_i$ and $\theta_i$, obtaining a 2-dimensional light-field slice $L(x_i, y, \theta_i, \phi)$. Such a slice of $L$ is called *Epipolar-Plane Image-view* or shortly EPI, referencing to the correspondence of this structure and the epipolar geometry known from stereo. Alternatively, one can choose a specific $y_i$ and $\phi_i$ and obtain a similar EPI, this time in the other dimension. In this work, EPIs will be used for obtaining depth information from the light field.

## 2.2 Parametrization

The notion of the light-field function developed from a more general concept, that of the *plenoptic function* [AB91]. The plenoptic function $P(x, y, z, \theta, \phi, t, \lambda)$ is a 7-dimensional function describing every possible light ray passing through a particular space-time. Hereby, $t$ denotes time, $\lambda$ the wavelength, and $(x, y, z)$ is the point through which a light ray with the direction $(\theta, \phi)$ is passing. If wavelength and time are omitted (corresponding to a stationary scene), $P(x, y, z, \theta, \phi)$ becomes a 5-dimensional function (see Fig.2.1). This function contains all radiometric information available in $(x, y, z)$.

It was pointed out by Levoy and Hanrahan [LH96] that one dimension of this function is redundant, if the imaged region of space is free of occluders, as the radiance of a given ray does not change in free space. Thus, if one restricts to such free-space environments, one can express the angular variations of radiance with

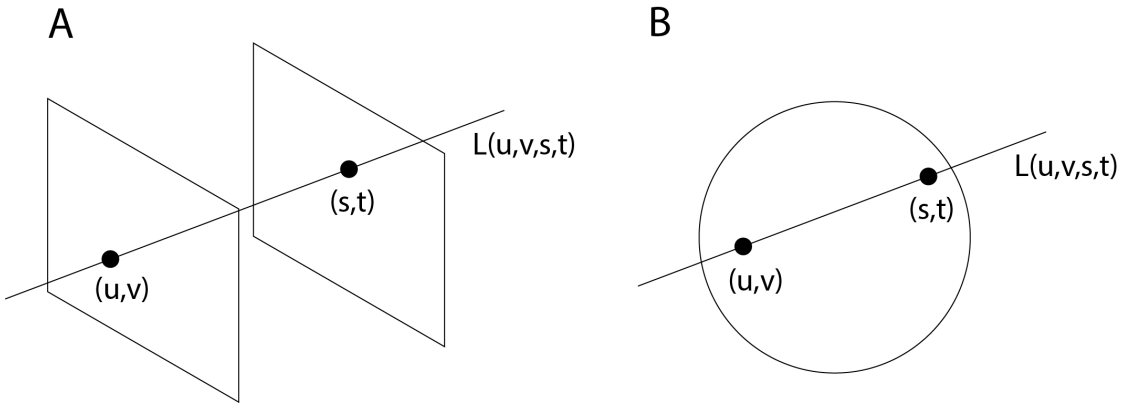


Figure 2.2: Parametrizations of the light-field function: (A) Two-plane parametrization, (B) Sphere light slab parametrization.

a 4-dimension function, the light-field function $\mathcal{L}$ which is far easier to capture.

There are various possibilities of parametrizing this function, and one possibility was already pointed out, namely $\mathcal{L}(x, y, \theta, \phi)$. This parametrization, while being the most common one, restricts to thinking in the means of angles and positions. Another, frequently used parametrization is $\mathcal{L}(u, v, s, t)$, also called the *Two-plane parametrization* (see Fig.2.2A). In this parametrization, the light ray is described by two parallel planes lying at a specified distance from each other. Tuples $(u, v)$ and $(s, t)$ correspond to the two points, where the light ray crosses these planes, respectively. This parametrization is very convenient to work with it from a computational point of view [LH96], since the $(u, v)$-plane can be imagined to correspond to the lens plane of an imaging system and the $(s, t)$-plane to the sensor plane. A shortcoming would be that it clearly cannot parametrize all rays in $\mathbb{R}^3$, e.g. rays lying parallel to the planes or rays travelling in the opposite direction, which, however, is not a major disadvantage, as these light rays aren't imaged anyway.

In theory, any mathematical way of describing lines in $\mathbb{R}^3$ is conceivable as a light-field parametrization. As another example, Figure 2.2B shows the *Sphere light slab* parametrization, mentioned in [WL02]. This parametrization uses intersection points of light rays with a spherical surface.

Of course, one has to keep in mind that all presented parametrizations are dealing with photons traveling in a certain direction along a precisely specified line, ignorant of the quantum-mechanical nature of these particles, their respective Heisenberg uncertainty and in turn, their wave-like properties. Thus, the treatment of light-fields in this way corresponds to ray optics, a very simplified model of optics, which is nevertheless effective when dealing with macroscopic scenes.

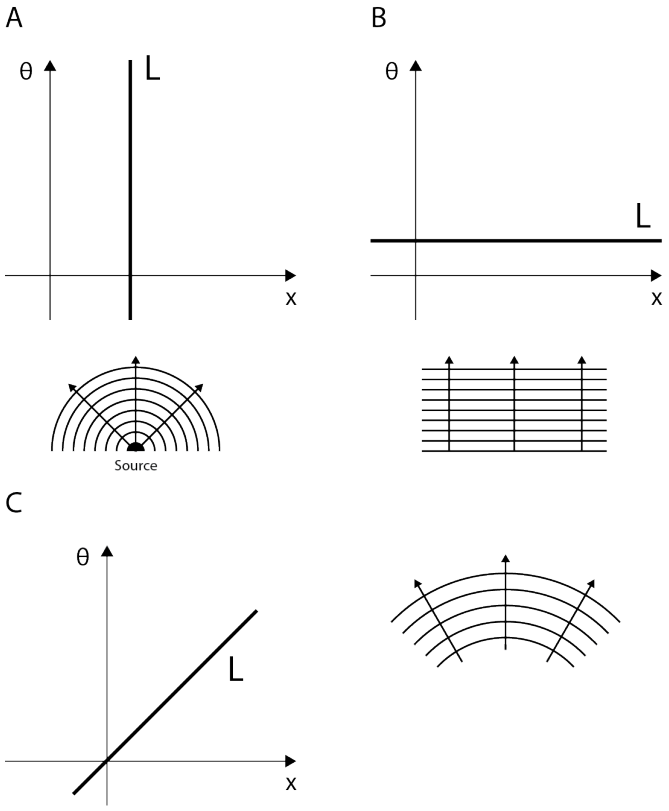


Figure 2.3: 1-dimensional light-field function for: (A) a point source, (B) a plane wave, (C) a spherical wavefront.

## 2.3 Light-fields of simple wavefronts

In this section, some important light-field functions will be presented, which build up a stronger intuition about the light-field function. In the following, we will restrict ourselves to a 1-dimensional world, i.e. to a 2-dimensional light-field, with one spatial and one angular dimension. This simplifies the situation without loss of generality (the same concepts can be applied to 3 dimensions). The discussion closely follows a paper by Oh et al. [Oh+10].

### 2.3.1 Light-field of a point source

The notion of a point source comes from continuous electrodynamics and is defined as an indefinitely small point radiating light in every direction of space. In other

words, the point source emits spherical waves. This leads to the light-field function

$$\mathcal{L}(x, \theta) = R\delta(x - x'), \tag{2.2}$$

where $R$ is the radiance, $x'$ is the location of the point source and $\delta$ is the Dirac delta-distribution. The function is shown graphically in Fig.2.3A. The extension of the equation for $n$ sources is trivial:

$$\mathcal{L}(x, \theta) = \sum_n R_n\delta(x - x'_n). \tag{2.3}$$

In this case, $R_n$ are the radiances of the inidividual sources. However, it must be kept in mind that light-fields provide a ray-optical description of light. Thus, no coherence effects are considered and the light-field can only describe incoherent light. Light emitting from an ideal point source is necessarily (spatially) coherent. Therefore interference effects will play a role and the light-field description does not model the physical reality any more. Although not treated here, there are multiple papers considering a definition of light-fields in wave optics. At the moment, the Wigner distribution, originally defined in statistical mechanics, seems to be the most promising candidate because of its similarities to the light-field function. The reader is referred to [Bas79; Oh+10; ZL09] for further information.

### 2.3.2 Light-field of a plane wave

Plane waves are waves traveling in a single direction whose wavefront is a plane. Equivalently, in ray optics, these are rays in every point of space with the same direction. Therefore, their light-field function reads:

$$\mathcal{L}(x, \theta) = R\delta(\theta - \theta'). \tag{2.4}$$

$R$ stands for the total radiance of the plane wave. Fig.2.3B shows the visualization. It is interesting to note that the light-field function for a plane wave is rotated 90 degrees w.r.t. the light-field function of the point-source. As will be shown in the next section, as the distance from a point-source increases, the light-field function slowly rotates in the direction of the plane-wave function. This is a well-known effect in wave optics: spherical waves emitted from a point source locally resemble plane waves in a large distance from the source.

### 2.3.3 Light-field of a spherical wavefront

The spherical wavefront is the wavefront of a point-source in some distance from it. As it is in some sense lying between the point-source case and the plane-wave

case, its light-field function experiences a shear between 0 and 90 degrees:

$$\mathcal{L}(x, \theta) = R\delta(ax - b\theta). \tag{2.5}$$

Here, $a$ and $b$ are factors that determine the incline of the light-field function and dependent on the curvature of the wavefront. A graphical representation of the function is found in Fig.2.3C. This dependency will be further explained in the next section.

## 2.4 Light-field transforms

In this section, functionals are introduced, which transform the light-field function when being refracted on a lens or traveling through free-space. These functionals, or *light-field transforms*, shall prove useful when studying ray paths in the line-scan light-field system later in Chapter 3. Furthermore, the same theory can also be applied to the wave-optical equivalent of the light-field, the Wigner distribution function, or, alternatively, to the augmented light-field function introduced in [Oh+10]. This makes studying light-field systems using this formalism very versatile. As in the previous section, this discussion is motivated by [Oh+10] as well as [Bas79].

A general light-field transform for a 1-dimensional light-field function is defined as

$$\mathcal{L}_2(x_2, \theta_2) = \int\int T(x_2, \theta_2, x_1, \theta_1)\mathcal{L}_1(x_1, \theta_1)dx_1 d\theta_1, \tag{2.6}$$

where $T$ is the light-field transformation function, which is 4-dimensional for one-dimensional light-fields and 8-dimensional for 4-dimensional light-fields, $\mathcal{L}_1$ is the original light-field function and $\mathcal{L}_2$ is the transformed light-field function. The output of the transform is a new transformed light-field function.

In order to find out the transformation function in concrete cases, a technique known from ray optics as *Ray Transfer Matrix Analysis* (RTMA) is applied (see e.g. [GB94]). This technique makes use of the fact that the most common transformations of light rays are linear, and can be expressed as matrix calculations, with the additional constraint of *paraxial approximation*. In RTMA, a light ray is described as a 2-dimensional vector $l = (x, \theta)$, where $x$ stands for an arbitrary point on the ray, while $\theta$ stands for the angular direction of the ray. Both $x$ and $\theta$ are hereby vectors themselves when dealing with more than one dimension. As in the previous discussions, we will consider $x$ and $\theta$ to be 1-dimensional for simplicity. In this case, the transformation matrices are two-times-two square matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Their exact form can be easily deduced by calculating the result of the well-known equations (e.g. the thin-lens equation) and then solving for $a, b, c, d$.

The transformation of the light ray after passing through the whole system can be calculated by multiplying the matrices of the corresponding optical elements.

Comparing RTMA with light-fields, it becomes clear that these two descriptions share many properties, e.g., identical parametrization of light rays in RTMA and light-field functions. Therefore, it is an obvious choice to derive light-field transformation functions from RTMA matrices. Some examples will be shown next.

### 2.4.1 Free-space transform

The transform which has to be applied when comparing the light-field functions of two different parallel planes, separated by a distance $d$ from each other by a medium with a constant refractive index (e.g. air or water), is called the free-space transform. In RTMA, the free-space travel matrix is defined as

$$\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}. \tag{2.7}$$

This means that the radiance $r(x, \theta)$ is transformed to $r(x + d\theta, \theta)$. From this, the light-field free-space transform follows:

$$\mathcal{L}(x_2, \theta_1) = \int \delta(x_1 - x_2 + d\theta_1)\mathcal{L}(x_1, \theta_1)dx_1. \tag{2.8}$$

The integral over $\theta$ from the general definition was dropped, since that parameter does not change during the transformation. In other words, the direction of the light-field remains constant. However, $x$ shifts in the direction of $\theta$ by the amount given by $d$. Graphically, the free-space transform causes a shear of the light-field function in the direction of $\theta_1$.

### 2.4.2 Refraction on a flat interface

At the interface between two mediums with different refractive indices, optical rays change their direction according to Snell's law. In the paraxial approximation, this change is proportional to the quotient of the refractive indices. The RTMA matrix for refraction at a flat interface reads

$$\begin{pmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{pmatrix}, \tag{2.9}$$

where $n_1, n_2$ are the refractive indices of the two media. The light-field transform is given by

$$\mathcal{L}(x_1, \theta_2) = \int \delta(\frac{n_1}{n_2}\theta_1 - \theta_2)\mathcal{L}(x_1, \theta_1)d\theta_1. \tag{2.10}$$

This time, the integral over the position was dropped, as position doesn't change during refraction. The only changing variable is $\theta$, which gets divided by the quotient of the refractive indices.

### 2.4.3 Thin lens refraction

The last example shows refraction at a thin lens. Thin lenses change the direction of incident light rays depending on their focal length $f$. The thin lens matrix is given by

$$\begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}. \tag{2.11}$$

From this, the light-field transform can be readily derived as

$$\mathcal{L}(x_1, \theta_2) = \int \delta(\theta_1 - \theta_2 + \frac{x}{f}) \mathcal{L}(x_1, \theta_1) d\theta_1. \tag{2.12}$$

Once again, no integration is performed over $x$. This is due to the fact that the thickness of thin lenses is zero. However, the direction of light rays changes w.r.t. the position this time and, similar to the free-form transform, causes a rotation of the light-field function, this time in the direction of $x_1$.

To conclude the discussion, it has to be mentioned that the presented transforms preserve the total radiance of the light-field, since they are derived from RTMA which only describe non-absorbing elements. Thus, the transformation functions mentioned consist of delta-functions. However, the concept of light-field transforms is a more general framework, which can also describe absorbing elements such as apertures. In the case of a light-field passing through an aperture, the transformation function simplifies to $T(x_1, x_2, \theta_1, \theta_2) = T(x, \theta)$, i.e. integration can be neglected, and transforming the light-field simply becomes a multiplication with a (possibly $x-$ or $\theta-$dependent) factor.

# 3 Line-scan light-field acquisition

## 3.1 Line-scan acquisition

Quality control is an important part of the manufacturing process in the industry. In the past, it evolved from manual inspection through supervised machine inspection all the way to the automatic inspection technologies available today. For a wide range of products, including *printed circuit boards* (PCBs), medical equipment or even banknotes, automatic visual inspection is the ideal and fast way of quality tests. A useful feature of visual inspection is that the inspection system can be integrated into the inline assembly pipeline. This is accomplished by scanning an object on a moving gantry perpendicular to its transport direction.

Such an image acquisition system, also called *line-scan imaging system*, consists of a transport mean for the object, such as a conveyor belt, which moves the object of inspection at a constant speed, as well as a line-scan camera. Line-scan cameras only possess one or a few sensor lines which are read-out at a constant frequency, adjusted to the speed of the transport. They can be placed either orthogonally or in any particular angle w.r.t. the transport plane, as long as the sensor lines remain oriented orthogonally to the transport direction.

During the acquisition process, the read-out lines are stacked together to produce the final image (also see Fig.3.1). One can imagine this as transferring one dimension of the image into the time domain. Such a setup has multiple advantages: (i) it allows for faster scan rates than area-scan setups, (ii) it allows for stronger, focused illumination, because only a narrow line has to be illuminated instead of a larger area, (iii) it is telecentric in the transport direction, meaning that the dimensions of the inspected object in the image are independent from its distance to the camera, enabling absolute measurements. A slight downside of line-scan imaging is its restriction to stationary objects. This can cause artifacts when facing vibrations or any other movement of the object during the acquisition.

As mentioned previously, the read-out frequency of the sensor has to be adjusted to the transport to obtain comparable resolution in both spatial directions. The required frequency is computed by

$$f = \frac{v}{x}, \tag{3.1}$$

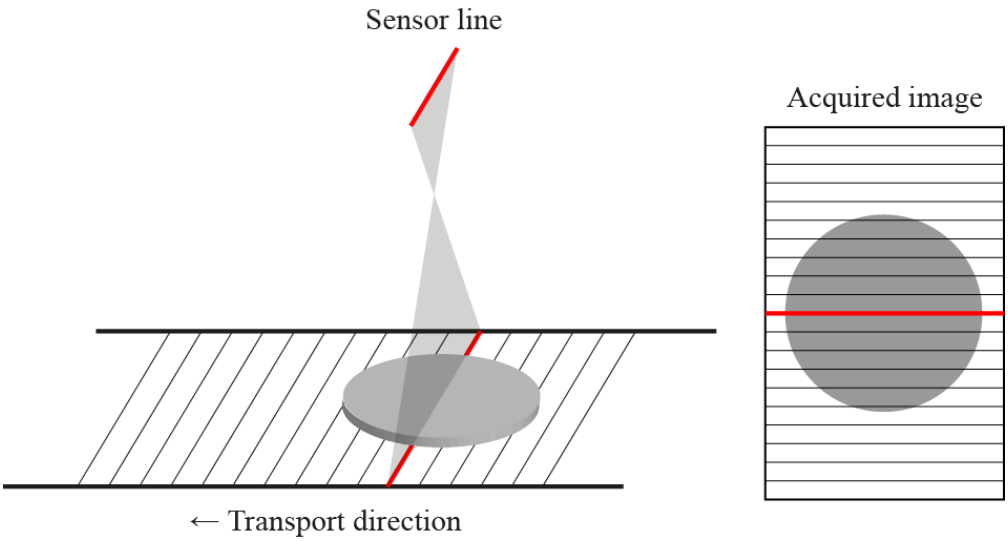


Figure 3.1: The line-scan acquisition process. An object moving with a constant speed on the transport is scanned by a single-line sensor. Lines can be stacked after the acquisition to form an image of the inspected object, if necessary.

where $f$ is the frequency in [px/s], $v$ is the transport speed in [mm/s] and $x$ is the resolution of the system in [mm/px]. Note that it is only possible to achieve equal resolutions on a single z-plane $z_0$, called the zero-plane, as the resolution perpendicular to the transport varies with depth according to the laws of perspective. This causes deformations of the object in the final image if the change in resolution w.r.t. the depth is large and/or the object shows large variations in its height profile.

In the past, line-scan systems have been implemented for traditional 2-D imaging, including x-ray imaging, as well as for stereo imaging, and industrial solutions are readily available [Gmb14a].

## 3.2 Line-scan light-field acquisition

The idea of combining line-scan acquisition with light-field imaging, presented for the first time in [Što+14] was developed in order to make use of the largely increased information volume that light-fields provide as opposed to traditional imaging in automatic inspection. During the course of this work, a line-scan light-field system delivering full-resolution images in the direction orthogonal to the transport while at the same time also providing different angular views was developed. Possible usage of this system ranges from hologram inspection to the

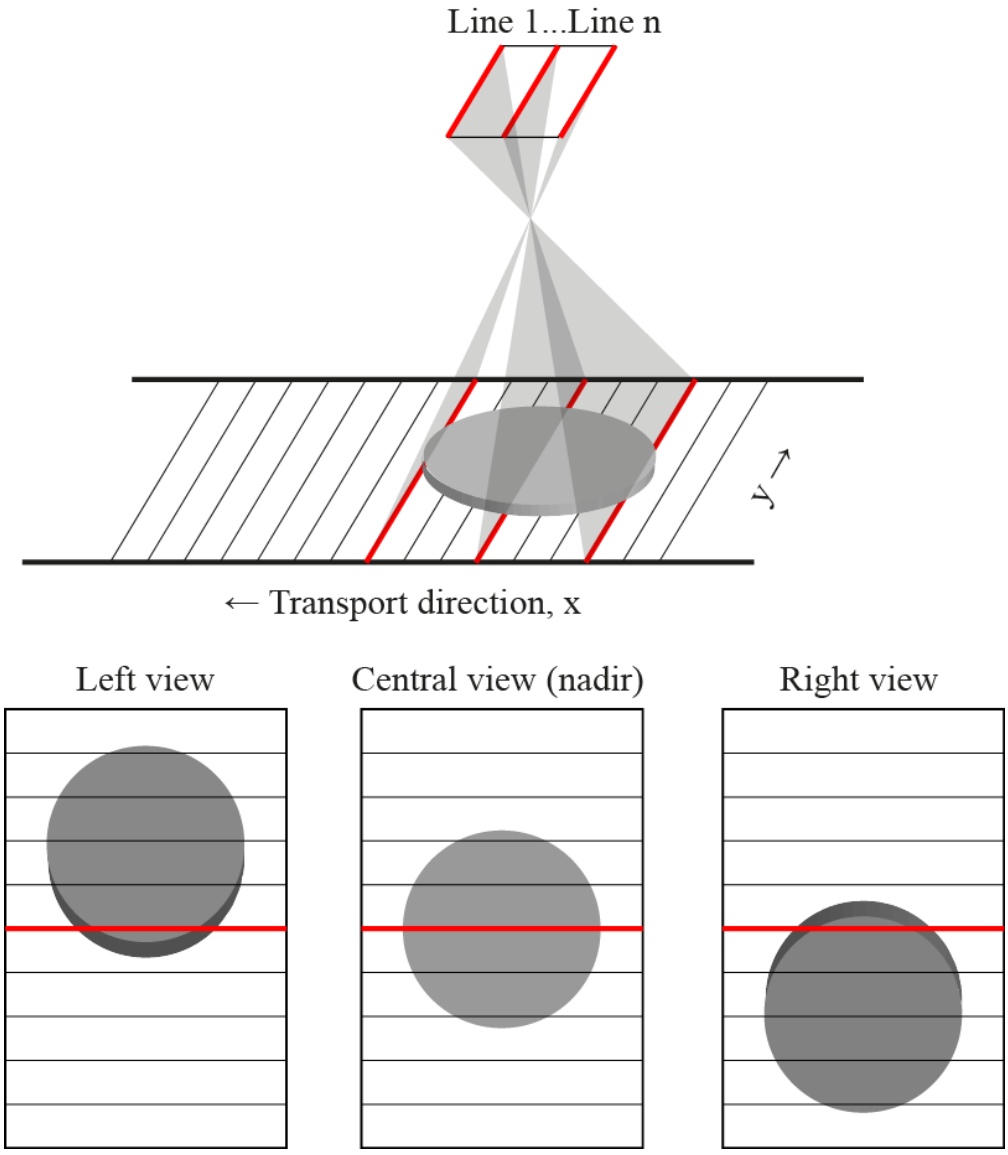


Figure 3.2: Schematic drawing of the line-scan light-field acquisition. As the object moves along the transport direction, it is scanned by multiple sensor lines simultaneously. Each sensor line captures the object from a slightly different angle, resulting in multiple views of the object, shifted by the distance between the scan lines on the object plane.

inspection of 3-dimensional objects using depth estimation and 3-D reconstruction. At the same time, light-fields acquired by the system can be used to enhance signal-to-noise ratio and depth-of-field.

The system works using a transport for the inspected object, as in every line-scan setup, and by placing an *area-scan* camera above it such that its sensor lines are directed orthogonally to the transport direction $x$. Instead of reading out the whole sensor repeatedly, however, only $n$ lines are read out from the camera, with $n < 16$. These lines are then treated as separate line-scan acquisitions, resulting in $n$ different views of the same object, albeit from a slightly different angle, as dictated by the optics (see Fig.3.2). Note that not all lenses are suitable for the system. Telecentric lenses, for example, do not maintain perspective, and thus there is no change in viewing angle across the sensor.

An important difference as compared to other light-field imaging approaches is the fact that the system only acquires a 3-D slice of the light-field, i.e. $L(x, y, \theta)$. No angular information is provided orthogonally to the transport direction, in the direction of $y$. Although this seems to be an issue, we didn't find it to be limiting for our goal of visual inspection and especially 3-D reconstruction. On the other hand, full spatial resolution is provided in both directions, which isn't the case for microlens-based setups, while preserving their compactness (unlike with camera arrays). Additionally, there is little need for calibration, since the sensor lines are guaranteed to be parallel to each other by the manufacturer, and only a single lens is used.

It can be argued that when considering only the most extreme (left-most and right-most) views of the system, the line-scan light-field setup is comparable to a stereo setup. However, there is a number of significant differences:

1. In standard line-scan stereo systems, different views are usually looking at the same line at one time.

2. Resulting from the above, stereo systems require precise mechanical adjustment.

3. In order to obtain a certain depth resolution, the baseline (angular range) of stereo systems is usually much larger (see the next section).

4. Stereo systems don't provide the same amount of information as light-field systems. This information is helpful not only in 3-D reconstruction and depth measurements, but also enables sophisticated effects such as computational refocusing.

In this sense, the line-scan light-field system is most closely related to a narrow-baseline (small angular range) stereo setup.

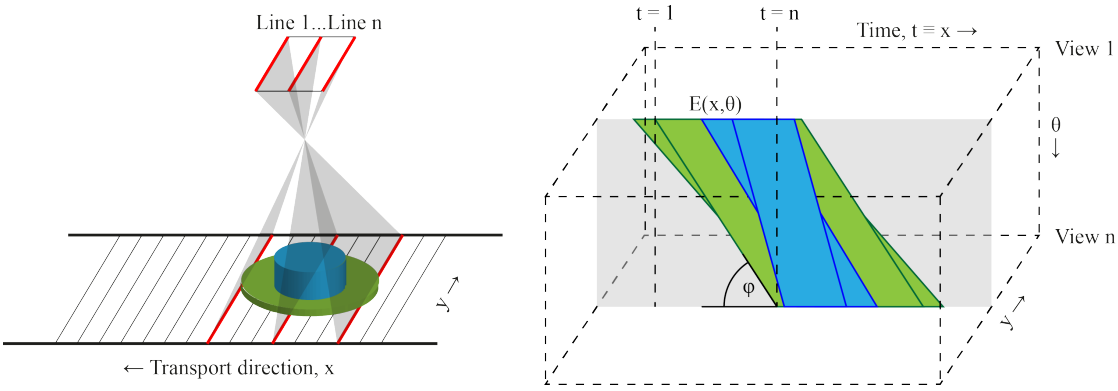


Figure 3.3: The process of light-field construction. At each time step, one line from each view is acquired. Slicing the light-field orthogonally to the views forms the EPI $E(x, \theta)$ which reveals the course of movement of the objects through the field-of-view. This in turn enables depth estimation.

Furthermore, the approach can be compared to the aforementioned Structure from Motion (SfM) approach, as the speed of movement of feature points in the view-port can be directly converted to depth information. This will be explicitly discussed in detail in Sec.4.3.

### 3.2.1 Structure of line-scan light-field data

In Fig.3.3, it is shown explicitly how the acquisition process constructs the light-field structure. One can think of this as building up a specific slice of the light-field $\mathcal{L}(x, y, \theta)$, the EPI $E(x, \theta)$, over time for each $y$. Object points imaged by the camera form sloped structures in EPI, which are then used to estimate their depth. In the figure, a structure $K$, which is seen in View 1 at time $t = 1$ and in View $n$ at time $t = n$, forms a slope in EPI with the incline $\phi$, the so-called disparity slope. We will denote by $l$ the *line stride*, which describes how far away the read-out lines of the sensor are from each other. If $l = 1$, subsequent lines in a certain area are read-out, and for objects on the zero-plane $z_0$ (which usually coincides with the focal plane), $\phi = 45°$ holds. Objects under $z_0$ move slower through the viewport, therefore $\phi > 45°$, while for objects behind $z_0$, $\phi < 45°$ holds. For later processing, it is important to ensure that structures in $E(x, \theta)$ form perfect lines and are clearly recognizable by the depth estimation algorithms. Methods how to ensure this as well as problems which may arise will be discussed in Chapter 4.

A practical example of $L(x, y, \theta)$ and the corresponding slices is shown in Fig.3.4. In this figure, a particular view of the scene, $L(x, y, \theta_i)$ is shown, together with the

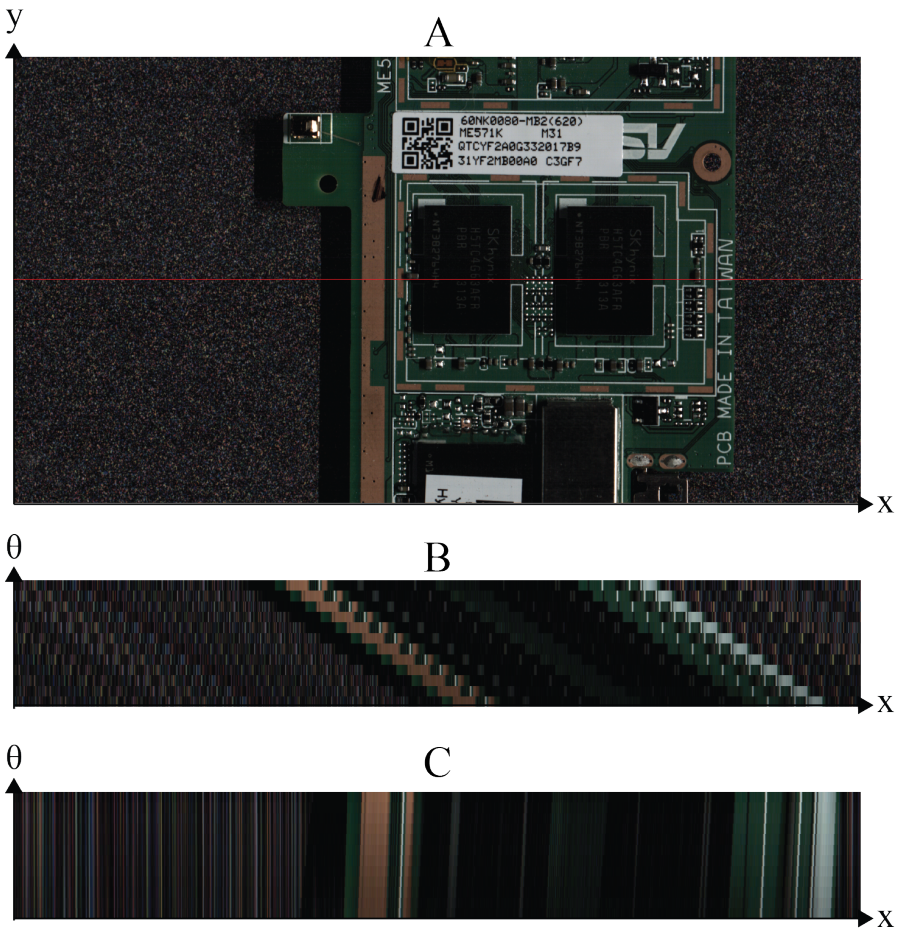


Figure 3.4: (A) Central view of a PCB board, (B) unaligned EPI of the highlighted line, with a line stride of $l = 80$ and (C) the aligned EPI. In (C), the individual horizontal pixel lines from (B) were shifted according to the line stride $l$, so that structures on $z_0$ (e.g. the background) are vertical, while nearer structures (e.g. the PCB board) are inclined to the right. This is beneficial for visualizing the differences in the incline of neighbouring slopes as well as for computational purposes, as pixels belonging to the same object point are placed approximately under each other.

EPI for the highlighted line, $E(x, \theta)$.

## 3.3 Light-field sampling in the line-scan light-field system

Up to now, we only described the system and postulated that it captures 3-D light-fields. In this section, we introduce a mathematical description of the way the system samples light-fields using the light-field transforms introduced earlier. The result will be derived using a 2-dimensional light-field, ignoring the $y$-dimension orthogonal to the transport, because no angular information is recorded in that direction.

We recall here two light-field transforms defined in Sec.2.4. The free-space transform of length $d$ of an arbitrary two-dimensional light-field is defined as

$$\mathcal{L}_2(x_2, \theta_2) = \int\int \delta(x_1 - x_2 + d\theta_2)\delta(\theta_1 - \theta_2)\mathcal{L}_1(x_1, \theta_1)d\theta_1 dx_1 \tag{3.2}$$

and the thin-lens transform is defined as

$$\mathcal{L}_2(x_2, \theta_2) = \int\int \delta(x_1 - x_2)\delta(\theta_1 - \theta_2 - \frac{1}{f}x_2)\mathcal{L}_1(x_1, \theta_1)d\theta_1 dx_1. \tag{3.3}$$

Once illuminated, the inspected object reflects light in a specific way, resulting in a light-field that needs to be captured by the system. Before that, the light rays travel a distance $a$ from the object to the lens. Thus it follows from the above formulas that

$$\mathcal{L}_2(x_2, \theta_2) = \mathcal{L}_1(x_2 - a\theta_2, \theta_2), \tag{3.4}$$

where $\mathcal{L}_1$ is the original light-field and $\mathcal{L}_2$ is the transformed light-field. Afterwards, the light-field is refracted at the lens and afterwards travels the distance $b$ from the lens to the sensor. This gives for the final light-field at the sensor

$$\begin{aligned}\mathcal{L}_4(x_4, \theta_4) = &\int\int\int\int \delta(x_3 - x_4 + b\theta_4)\delta(\theta_3 - \theta_4)\delta(x_2 - x_3)\delta(\theta_2 - \theta_3 - \frac{1}{f}x_3) \\ &\mathcal{L}_1(x_2 - a\theta_2, \theta_2)d\theta_3 dx_3 d\theta_2 dx_2 \\ &= \mathcal{L}_1(x_4 - b\theta_4 - a\theta_4 - \frac{a}{f}x_4 + \frac{ab}{f}\theta_4, \theta_4 + \frac{1}{f}x_4 - \frac{b}{f}\theta_4).\end{aligned} \tag{3.5}$$

At this point, the assumption is made that the inspected object is approximately

in focus and thus for the $a$ and $b$ the relation

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f} \tag{3.6}$$

holds. Inserting this above, we obtain for $\mathcal{L}$ (omitting the now useless indices):

$$\mathcal{L}_4(x, \theta) = \mathcal{L}_1(-\frac{a}{b}x, -\frac{b}{a}\theta + \frac{1}{f}x). \tag{3.7}$$

This is the final light-field function on the sensor. It shows that for an arbitrary "input" light-field, the $x$-coordinate is magnified by the magnification factor $M = \frac{a}{b}$, while it also moves on the opposite side of the optical axis. This a trivial result in optics. On the other hand, angular information shifts according to the refracting power of the lens $\frac{1}{f}$ and $\frac{1}{M}$ because of the lens. The sensor only samples the light-field at discrete points $x_i$. It follows from the above that with an increased pixel density, both the lateral and angular resolution improve.

For the intensity measured at the sensor, the light-field function has to be integrated over the angular coordinate:

$$I(x) = \int \mathcal{L}(x, \theta) d\theta. \tag{3.8}$$

The integration range is $\frac{D}{b}$, with $D$ being the diameter of the lens. If a Lambertian object with diffuse reflection is assumed, $\mathcal{L}$ is constant over $\theta$ and for the intensity we obtain

$$I_{lambert}(x) = \frac{D}{b}\mathcal{L}(-\frac{a}{b}x, 0). \tag{3.9}$$

The angle $\theta_i$ which is sampled on a particular pixel $x_i$ of the sensor is $\theta_i = \arctan(\frac{x}{b})$.

For specular reflections, a term is added to the light-field function which corresponds to the specularity in a certain angular range. For most objects, a specular lobe with a (more or less) gaussian profile over the angular dimension can be assumed.

### 3.3.1 Angular range, angular resolution and depth resolution

The depth resolution of a 3-D imaging system is the most important parameter for depth estimation, since it gives an indication of which objects might be suitable for inspection as well as providing the basic means of recalculating disparity maps, generated by the depth estimation algorithm, to depth maps. Ultimately, the depth resolution of a light-field system can only be assessed experimentally, since it depends on the depth algorithm as well as the inspected object, but an approximate

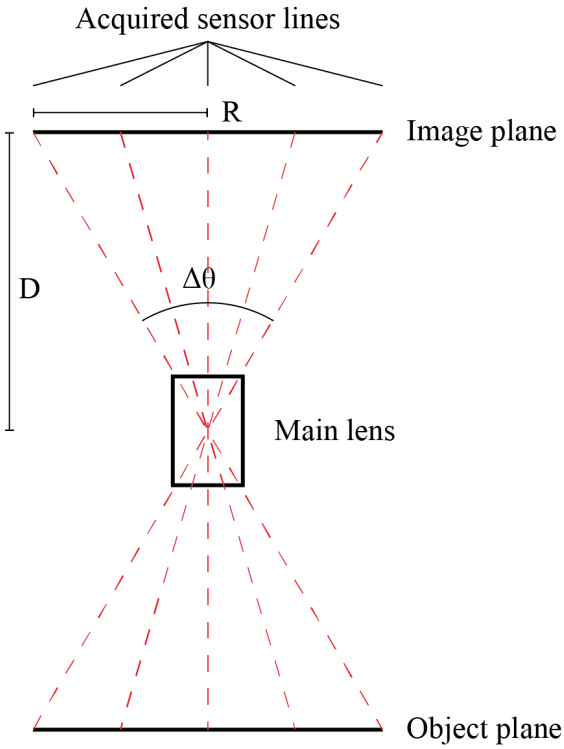


Figure 3.5: Angular range and angular resolution.

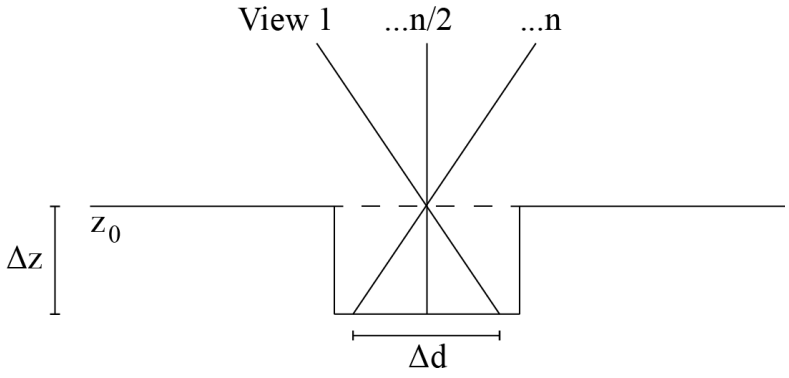


Figure 3.6: The relationship between the disparity $\Delta d$ and the depth $\Delta z$, relative to the zero-plane $z_0$.

figure can be given by comparing our system to a stereo system, again taking into account the left-most and right-most views, and deriving the depth resolution in the standard way used in stereo. In the results section, it will be shown that the real depth resolution exceeds this estimate, proving that depth estimation can largely benefit from the additional information provided by light-fields.

Prior to deriving the depth resolution, the angular range of the system has to be known, i.e. the angular difference between the left-most and right-most views. This parameter is also important in the inspection of *optically variable devices* (OVDs) such as holograms — for a large angular range, these are more likely to exhibit visible changes in appearance.

The angular range $\Delta\theta$ is, in principle, given by the angular difference between the chief rays from the left-most and right-most captured sensor lines (see Fig. 3.5). It is calculated by

$$\Delta\theta = 2\arctan(D/R), \tag{3.10}$$

with $D$ being the distance from the image plane to the main lens, and $R$ being the distance of the left- (or right-)most line from the central line of the sensor (we assume symmetry).

The angular resolution, i.e. the sampling frequency of the light-field in the angular dimension, is another important parameter for depth estimation. Very sparse sampling hinders correct depth estimation on surfaces with repetitive patterns. This is related to the sampling theorem, and will be discussed in Sec.4.3.

Angular sampling of the line-scan light-field system is not equidistant. As can be seen in Fig.3.5, the spacing between the sampled lines is constant. This results in a constant $\tan\theta$, but not $\theta$. We illustrate this for the two shown sampled lines, located at distances $x$ and $2x$ from the central line. For the line nearer to the

center, the relation

$$\tan\theta_1 = \frac{x}{D} \tag{3.11}$$

$$\theta_1 = \arctan\frac{x}{D} \tag{3.12}$$

holds, with $D$ again being the image plane distance from the main lens. For the line further from the central line, the relation

$$\theta_2 = \arctan\frac{2x}{D} - \arctan\frac{x}{D} \tag{3.13}$$

holds, which proves that $\theta_1 \neq \theta_2$. This could lead to the wrong assumption that the structures in the EPI will actually form curves, which change their slope depending on the angular resolution. However, this is not the case, since the parameter determining the slope in the EPI is not the angle $\theta$, but rather the *disparity* $\Delta d$. The disparity in stereo imaging is the measured displacement of the same object point in the two images in pixels. Looking at Fig.3.6, an object which would be imaged on the exact same spot in both views, if its height was on the zero-plane $z_0$, will be separated by the disparity $\Delta d$, if its depth relative to the zero-plane is $\Delta z$. $\Delta d$ is thus related to the depth $\Delta z$ by

$$\Delta d = \tan\theta \Delta z, \tag{3.14}$$

where $\theta$ is the angle between the two views. As $\tan\theta$ is sampled equidistantly, the disparity changes linearly across the views, and for a constant $\Delta z$ the structure in the EPI forms a line.

The basic depth resolution can then be quantified as the minimum $\Delta z$ required to cause a disparity $\Delta d = 1$ px between the left- and right-most views. It is given by the aforementioned proportionality factor $\tan\theta$ between depth and disparity, which again depends on two factors:

1. The distance between the left-most and right-most lines of the sensor, which are read-out, and

2. The distance from the sensor to $z_0$ (which is usually the plane on which the system is focused).

The larger this factor, the larger the depth resolution. Typically, stereo systems achieve a depth resolution of $0.1 - 1\times$ the lateral resolution. A line-scan light-field system, however, provides much more information than a stereo system. Therefore, its depth resolution is higher and the depth estimates tend to be more robust.

### 3.3.2 Lateral optical resolution

As the line-scan light-field setup is using a standard industrial area-scan camera for imaging, its optical resolution doesn't differ from a standard imaging system with the same camera, although the short working distance and therefore a comparatively large numerical aperture are responsible for a (theoretically) much higher resolution than is usual in photography, approaching microscopy. Assuming the Rayleigh criterion, the resolution can be calculated by the well-known formula [Luk08]:

$$R = 2.44\lambda F = 1.22F, \tag{3.15}$$

where $R$ is the spatial resolution in µm, $\lambda$ is the wavelength of light, assumed to be 500 nm, and $F$ is the F-number of the main lens aperture.

Because of lens aberrations, the optical resolution is in practice always lower than the calculated maximum. Microscopic lenses, which tend to be very well corrected for a certain (small) working distance, typically achieve the highest resolutions. In our case, the inspected objects tend to have a depth which by far exceeds the depth-of-field of most microscopic lenses. We therefore decided to use photographic macro lenses in the setup, which have lower resolutions, but are corrected for a larger number of working distances. Even with these corrections, however, the resolution tends to decrease the further away one moves from the center of the sensor.

In our case, the system isn't limited by optical resolution, and our chosen resolution (20 µm/px and 40 µm/px) rather proves to be the limiting factor at the aperture numbers we use. We also limit the field-of-view, so that the system remains resolution-limited (rather than diffraction-limited) for all acquired views.

### 3.3.3 Depth-of-field

The *depth-of-field* (DoF) of an imaging system determines an interval along the optical axis in which the system remains focused, i.e. its circle of confusion (CoC) $c$ is smaller than the pixel diameter. A similar concept to DoF is the *depth-of-focus* (dof), which is the allowable movement of the image plane along the optical axis which retains the system in focus. The DoF is related to the dof via the magnification $M$ of the lens. In order to derive the DoF of a standard imaging system, let us define by $D_1$ and $D_2$ the front and back plane of the DoF and by $d_1$ and $d_2$ the front and back plane of the dof, as in Fig.3.7. Looking at the sketch, it is obvious that $D_1$ and $D_2$ do not have the same distance from the focal spot because of the different sizes of the projected circles of confusion $C_1$ and $C_2$.

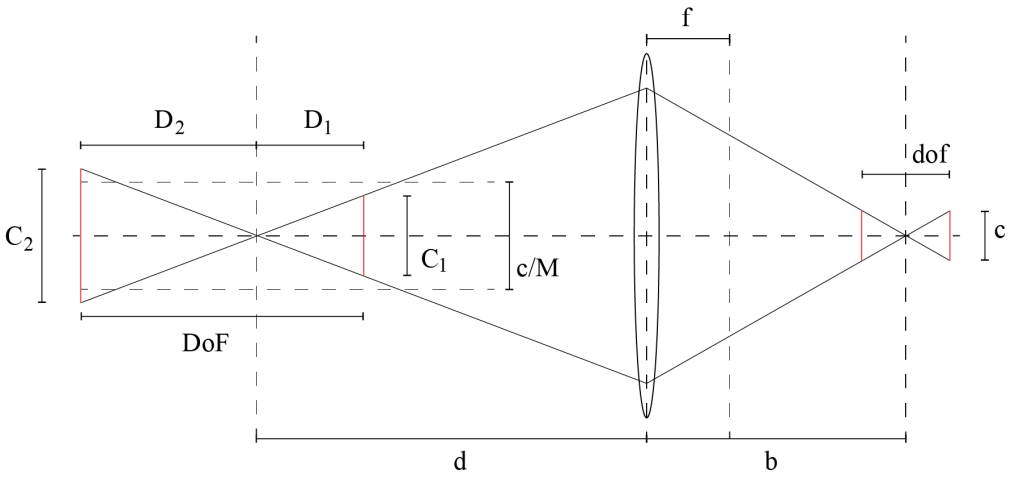


Figure 3.7: Parameters determining the DoF.

The front and back planes of the DoF can be derived using the sketch as

$$D_1 = \frac{Fcd^2}{f^2 + Fcd} \tag{3.16}$$

$$D_2 = \frac{Fcd^2}{f^2 - Fcd} \tag{3.17}$$

which results in

$$DoF = D_1 + D_2 = \frac{2Fcd^2f^2}{f^4 - F^2c^2d^2}, \tag{3.18}$$

where $F$ is the F-number of the lens, $c$ is the CoC (with dimensions of one pixel), $f$ is the focal length and $d$ is the focusing distance.

Making the assumption that $c$ is very small, which is generally the case, the DoF is roughly proportional to the F-number, i.e. $DoF \propto F = f/A$. A wide-open aperture reduces the DoF while a smaller aperture, with the extreme case of the camera obscura, prolongs the DoF (theoretically) to infinity. Naturally, this holds only until the lens aperture is stopped down to a degree where light diffraction begins to play a role and the resolution in the focal plane is reduced. Therefore, the DoF is also inversely proportional to optical resolution. As a typical example, consider light microscopes, which exhibit very detailed optical resolution, while limiting the DoF to micrometers. Also, the formula for DoF derived here strictly holds only for paraxial approximation. This is not a restriction in our case, as we use this approximation in the whole work.

While small DoFs are often desirable in photography, in automatic inspection, there is need for a large DoF in order to inspect objects with largely varying

height profiles. Therefore it is necessary to close the aperture enough to have acceptable resolution for the whole inspected object. This in turn reduces the maximal resolution, so that a compromise has to be found. The best compromise is generally to find the greatest aperture-number, for which the system is not yet diffraction limited (i.e. the resolution is limited by the number of pixels, rather than by diffraction). Another shortcoming of large aperture numbers (small aperture diameters) is the low amount of light entering the system, leading to a small *signal-to-noise ratio* (SNR), and subsequently to low contrast and strong noise in the images.

Note that it is not possible to improve depth defocus by deconvolution, since depth blurring is not a convolution process (the blurring kernel is *not* translation-invariant). Nevertheless, the larger amount of data in light-fields as compared to traditional machine vision systems enable us to achieve a larger depth-of-field without sacrificing the SNR. The corresponding technique will be explained in Sec.4.4.

## 3.4 Enhancing the angular range using a slit

One of the shortcomings of the previously proposed system is the large field-of-view required for an acceptable depth resolution. This in turn requires homogeneous illumination over a large area as well as a precise synchronization between the camera and the transport, which is sometimes hard to ensure in practice.

A simple way to narrow down the field-of-view while at the same time preserving the desired depth resolution consists of placing a narrow slit aperture in the optical path of the system, aiming to allow only light rays from a very narrow range of angles to hit a certain sensor line.

Two slit aperture configurations are conceivable: (i) placing the slit between the object plane and the lens, and (ii) placing the slit between the lens and the image plane. As the two placements are conjugate (they are related by the magnification of the lens), the two setups can be expected to give approximately the same results, although the inner slit aperture must not be placed on the telecentric plane, i.e. at focal distance from the lens. Furthermore, in practice, it is more convenient to place the slit outside of the camera, allowing for easier adjustment and more flexibility. Additionally, placing the aperture inside, while being more compact, requires placing the slit extremely close to the sensor to achieve the desired depth resolution, resulting in the need of removing the cover glass of the sensor. Vignetting can also be an issue. Therefore, only the outer slit configuration will be discussed here.

Figure 3.8 shows the light ray paths in the slit aperture setup. Compared to the standard line-scan light-field system, the slit aperture increases the angular

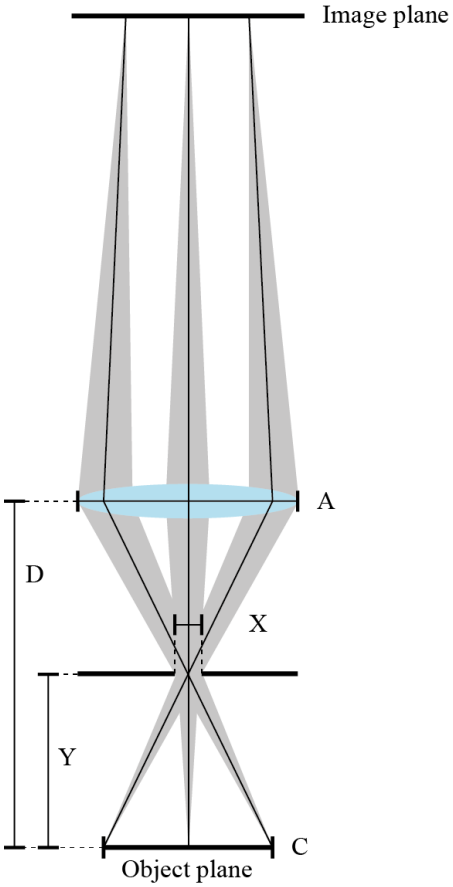


Figure 3.8: A schematic drawing of the slit aperture system, with $A$ being the main lens aperture width, $X$ the slit diameter, $C$ the maximal field-of-view, $Y$ the distance from the object plane to the slit, and $D$ the distance from the object plane to the main lens.

differences between neighboring lines.

### 3.4.1 Angular range

According to Fig.3.8, the angular range of the slit aperture setup can be defined as:

$$\alpha_R = 2\arctan(\frac{C}{2Y}), \tag{3.19}$$

where $C$ is the field-of-view on the object plane and $Y$ is the vertical distance from the object plane to the slit. In fact, the angular range is even higher, if one considers object lines lying even further away from the center. However, in that case, some of the light from these lines is blocked by the main aperture. Additionally, these lines have a different angular resolution, which causes a curve in the EPI structures. This makes the extreme views undesired when considering depth estimation, while for OVD inspection, they can be considered in some cases.

A big advantage of the setup is, that, according to the formula, the angular range can be arbitrarily chosen by varying the vertical placement of the slit $Y$. This theoretically allows for very large angular ranges, although there is also the limitation by the main lens aperture. Furthermore, practical placement of the aperture can be difficult due to lack of space (with the illumination also having to fit in).

### 3.4.2 Lateral optical resolution of the slit aperture setup

With the addition of the slit, lateral optical resolution also changes, although only in the transport direction $x$. Since the resolution is associated with the numerical aperture and therefore the acceptance angle of light, it depends on the placement of the slit and gets worse as the slit is placed further away and as it narrows. It can be calculated as

$$R = \frac{0.61\lambda}{\sin(2\arctan(\frac{X}{2Y}))}, \tag{3.20}$$

where once again the Rayleigh criterion has been used. Similarly, $\lambda = 0.5$ nm.

The resolution is inversely related to the angular range, forcing to make a compromise between those two parameters.

### 3.4.3 Drawbacks of the slit aperture setup

As already pointed out, the slit aperture setup is an alternative to the standard line-scan light-field setup. When deciding which setup to use in real-world cases, it is therefore necessary to carefully evaluate advantages as well as disadvantages

of both w.r.t. to the inspection scenario. Drawbacks of the slit aperture setup can be largely related to drawbacks of the pinhole camera, since in transport direction, the acceptance angle of light is very narrow.

Firstly, by using the very narrow ($<$1 mm) slit, the amount of light hitting the sensor dramatically decreases. This can be somewhat compensated by the fact that the area which needs to be illuminated decreases, but in many cases, this is insufficient and requires amplification of the camera signal. Secondly, as mentioned previously, the system's resolving power becomes anisotropic, resulting in a non-circular point-spread function (PSF). This effect becomes visible when approaching the resolution limit of the system, or in the case of a defocused object. A possible compensation can be achieved by placing a second slit orthogonally to the first one inside of the camera, although this then also reduces the resolution in the $y$ direction while also further reducing the light intensity. Note that the anisotropic effect also increases the depth-of-field in $x$-direction, since the DoF is approximately inversely proportional to the resolution.

# 4 Methodology

## 4.1 Setup

As already mentioned, the basic experimental setup consists of a precise conveyor belt as a transport means, a fast area-scan camera, which is capable of reading out selected lines from the sensor fast, together with an appropriate lens and a varying type of illumination depending on the surface characteristics of inspected object. Tab.4.1 contains used components, while Tab.4.2 provides an overview of the two optical setups used to obtain the results presented in this thesis. The camera is oriented such that sensor lines are orthogonal to the transport. Fig.4.1 shows the actual setup.

### 4.1.1 Slit aperture setup

The major difference between the original setup and the slit aperture setup is the addition of the slit and the fact that a dense light-field is acquired, since the slit enhances the angular difference between successive lines. Further, a macro lens with a large aperture is used in order to make the most out of the advantages of the slit system. A brief summary of the slit aperture setup is found in Tab.4.3.

### 4.1.2 Illumination

In every imaging process (except for situations in which the light source itself is being imaged, e.g. fluorescent imaging), careful consideration should be applied to the choice of illumination. In automatic inspection, where the type of objects to be inspected is usually known beforehand, it is important to adjust the illumination to suit that particular object. Whereas Lambertian objects (objects which follow the Lambertian cosine law of reflection) are generally less sensitive to the choice of illumination, imaging quality of highly specular objects such as metallic or plastic surfaces largely depends on correct illumination.

Two types of illumination were used in this project (see also Tab.4.1):

1. Directional illumination using one high-intensity LED and two symmetric line converters oriented orthogonally to the transport, and

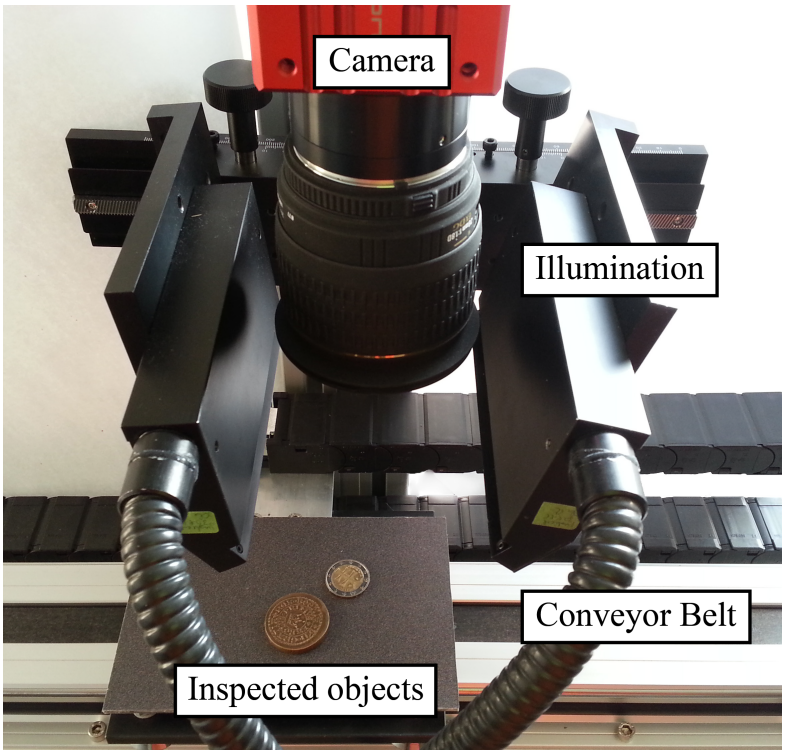


Figure 4.1: A picture of the setup. In the upper section the camera and the optics is placed, while in the lower section, one can see the conveyor belt with a pair of coins as sample objects. The (directional) illumination is placed on both sides of the camera.

| Camera | |
|---|---|
| Name | Allied Vision Bonito CL-400C |
| Sensor | Alexima AM41 CMOS area-scan |
| Resolution | $2320 \times 1728$ px |
| Pixel pitch | 7 µm |
| CFA | Bayer pattern |
| Max. line rate | cca. 300 kHz |
| Shutter | global |
| Color resolution | 8 bit |
| Lens mount | Nikon F-mount |
| **Lenses** | |
| Name | SIGMA EX DG Macro 50 mm F2.8 |
| Focal length | 50 mm |
| Min. focal distance | 18.8 cm |
| Name | SIGMA EX DG Macro 28 mm F1.8 |
| Focal length | 28 mm |
| Min. focal distance | 20 cm |
| **Illuminations** | |
| Name | Volpi intraLED 5 with dual branch fiber optic line light |
| Type | Directed |
| Name | Metaphase MB-DALL306 line light |
| Type | Diffuse axial |
| **Transport** | |
| Name | Isel Zahnriemenvorschub LEZ 2 |
| Controller | Isel MC1-40 |
| Trigger | Isel iMS Magnetic length measuring system |
| Trigger resolution | 20 µm |

Table 4.1: Components used in the experimental setup.

| High-resolution setup | |
|---|---|
| Lens | SIGMA EX DG Macro 50 mm F2.8 |
| Working distance | 240 mm (object-to-camera) |
| Lateral resolution | 20 µm/px |
| Field-of-view | 19.2 mm x 46.4 mm (960 lines a 2320 pixels) |
| Angular range | 6.2° |
| Acquired lines | 12 line pairs with 80 lines distance in-between |
| Depth resolution (equivalent stereo) | 184.6 µm / disparity |
| **Low-resolution setup** | |
| Lens | SIGMA EX DG Macro 28 mm F1.8 |
| Working distance | 346 mm (object-to-camera) |
| Lateral resolution | 40 µm/px |
| Field-of-view | 38.4 mm x 92.8 mm (960 lines a 2320 pixels) |
| Angular range | 7.17° |
| Acquired lines | 12 line pairs with 80 lines distance in-between |
| Depth resolution (equivalent stereo) | 319.1 µm / disparity |

Table 4.2: Two setups of the line-scan light-field system.

| Slit aperture setup | |
|---|---|
| Lens | Sigma 50mm F1.4 EX DG HSM |
| Illumination | Metaphase compact diffused |
| Slit width | $250\mu m$ |
| Distance slit - object | 11mm |
| Number of views | 32 (dense) |
| Working distance | 215mm (object-to-camera) |
| Lateral resolution | $40\mu m/px$ |
| Field-of-view | 1.28mm x 92.8mm (32 lines a 2320 pixels) |
| Angular range | 3.3° |
| Depth resolution (equivalent stereo) | $694\mu m$ / disparity |

Table 4.3: Components and dimensions of the slit aperture setup.

2. Diffuse axial illumination with a line of LEDs using a beam splitter and a diffuser.

Lambertian surfaces can be imaged using both types of illumination and produce comparable results. Highly specular and reflective surfaces are viewed in the *dark-field* when illuminated with directional illumination, meaning that only a small range of suitable surface normals reflects light rays to the camera. To obtain *bright-field* (the opposite case, with most of the surface being visible in the image) on such surfaces, diffuse illumination has to be used.

For depth estimation, one has to strive for a *bright-field* illumination with suppressed specular reflections. Since specular reflections are only visible in some views, but not in others, the search for correlated pixels across the views is more difficult if these reflections exceed the dynamic range of the camera. Lambertian surfaces usually do not introduce such problems. For specular surfaces, a possiblity used to mitigate narrow-angle specular reflections is using polarized light and a polarizer on the camera, rotated 90 degrees w.r.t. the polarization of the incoming light. This way, direct reflections are suppressed, while the ambient contribution remains.

Another illumination effect causing lowered image quality is *speckle*. Speckle is defined as spatially varying light intensity on a surface due to coherence effects, and manifests itself as high frequency noise.

Speckle can also appear when using white partially coherent or otherwise incoherent light sources, such as LEDs, given specific conditions. These conditions include illuminating from a very narrow angle, combined with imaging with a large aperture number and a specular micro-rough surface, the explanation being that light waves forming in similar positions on the LED have similar wavelengths, but different phases. If the inspected surface is highly reflective, it reflects light in a very narrow angle. Thus, closing the aperture only allows this specific angular range to enter the camera. As the light is focused on the sensor by the camera lens, it interferes, causing speckle.

While some optical measurement techniques make use of interference and speckle for surface classification, depth estimation and surface roughness measurement ([Wya02; Rib69; Tch+10]), they are undesirable in light-field imaging and therefore are considered as artifacts.

In our experiments, speckle formation when imaging highly reflective surfaces was overcome by using *diffuse* axial illumination, ensuring object illumination from a wide angle, as well as imaging with aperture numbers of f/2.8-f/8, which didn't cause any problems with coherence artifacts.

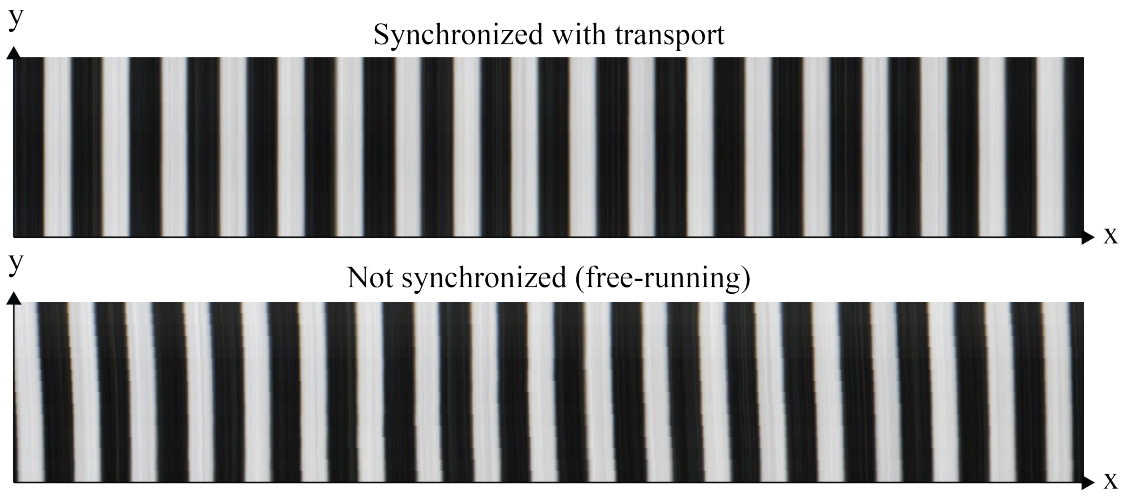


Figure 4.2: Example of a non-triggered acquisition. The upper image shows the EPI of a particular line of the FFC calibration object (with black and white stripes) with triggered acquisition, while the lower image shows the same EPI without any synchronization with the transport. Obviously, the structures in the lower EPI exhibit a varying slope, which is inappropriate for depth estimation.

### 4.1.3 Transport

As one dimension of the light-field (namely the direction parallel to the transport $y$) is acquired in the time domain, synchronization of the transport is crucial to recording correct data. The speed of the transport determines the slope of the structures in the EPI image and the resolution in the $y$-dimension, and consequently, in depth estimation algorithms, determines the search range of slopes found in the EPI. It is also important for a measurement of the exact dimensions of objects in the $y$-dimension. On the other hand, *variations* of speed $\frac{dv}{dt}$ in the transport also prohibit correct depth estimation, since they alter the structures in the EPI image in a way that they no more form straight lines, but rather form curves with varying slopes. This violates the assumption of the depth estimation algorithm, that structures in the EPI image are straight, and makes it hard to estimate the correct depth. Fig.4.2 compares the EPI of a perfectly synchronized acquisition to that of a non-synchronized acquisition.

Unfortunately, every real transport has mechanical restrictions which cause inaccuracies in its speed (this includes e.g. inertial effects and stepping motors). Therefore, especially for low speeds, it is not possible to run the camera and the transport independently from each other. The camera has to adjust its capture frequency to short-time changes in transport speed. This can be accomplished by

triggering the acquisitions from the transport. The conveyor belt used in our experiments has a trigger resolution of 20 µm (compare Tab.4.1). The trigger signal is produced by a magnetic length meter. It is fed into the PC, where the trigger frequency can be further divided if needed. Afterwards, it is transmitted through the CameraLink interface to the camera. This reduces most of the transport artifacts in the data, although other minor effects causing curved structures in the EPI images may remain, which will be discussed in Sec.4.2.3.

### 4.1.4 Acquisition

Before the actual acquisition, the camera is configured to read-out the selected lines. For every view, two consecutive lines are acquired to obtain full color information, as the camera sensor uses a Bayer-Color Filter Array (CFA, more on this can be found in Sec.4.2.2). It is possible to capture a *dense* or *sparse* light-field in the angular direction depending on the captured lines. If a continuous block of lines is read-out, a dense light-field is obtained and vice-versa. The advantage of a dense light-field is its high information content, which can improve depth estimation. It does however limit the depth resolution by constraining the angular range, which was deemed impractical for most inspection purposes. The acquisitions in this work were all done using sparse acquisitions, which is made possible by the fact that the sensor can deliver individual lines separated by an arbitrary distance from each other.

The speed of the acquisition is, in principle, limited to the lighting conditions as well as to the maximum frame frequency of the camera, and is dependent on the number of views (i.e. lines) which are read-out. In our experiments, 12 views were chosen as a good compromise, enabling precise depth estimation as well as acceptable speed. Because of the ability to read out the same number of lines from the sensor equally fast, independently of their position, the angular range can be chosen freely without sacrificing speed.

After the read-out, raw data is transferred to the PC, where further processing ensures correct calibration, and subsequently, the depth estimation algorithm is performed.

## 4.2 Post-processing

A diagram of the processing pipeline for the light-field acquisition can be found in Fig.4.3. The processing software was implemented in MATLAB, as this language together with its infrastructure was considered most useful for rapid prototyping and the signal processing algorithms required. It was run on a PC with Windows 7 operating system and an Intel Xeon CPU with 3.6 GHz frequency.

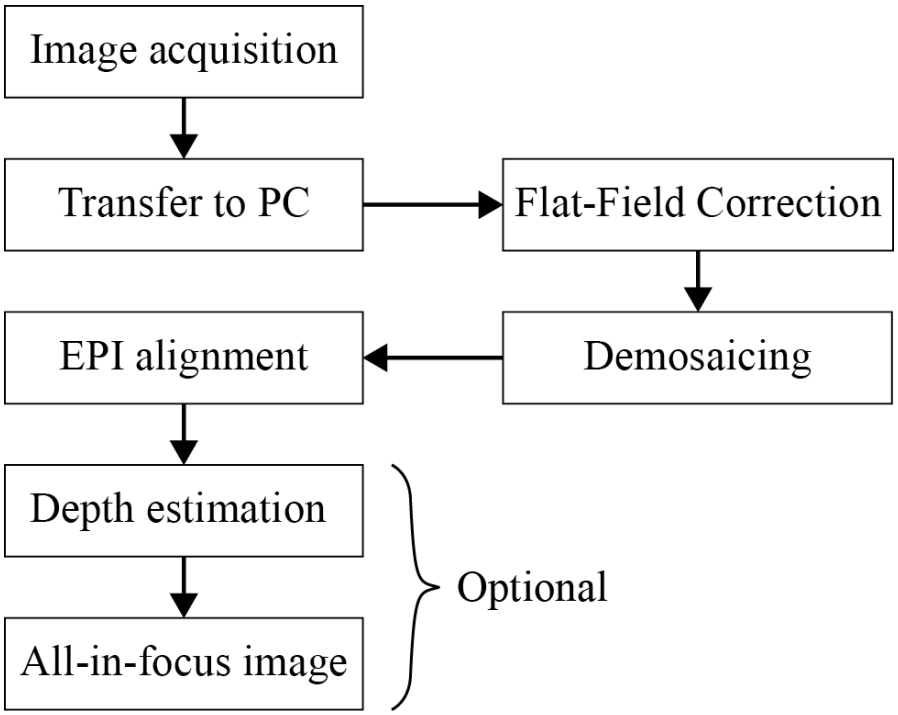


Figure 4.3: The processing pipeline. After the acquisition, the images are transferred to the PC, where FFC and demosaicing are applied. The EPI is then aligned s.t. pixels corresponding to the same object point in the $z_0$-plane are placed over each other. Finally, depth estimation and all-in-focus image generation can be optionally performed.

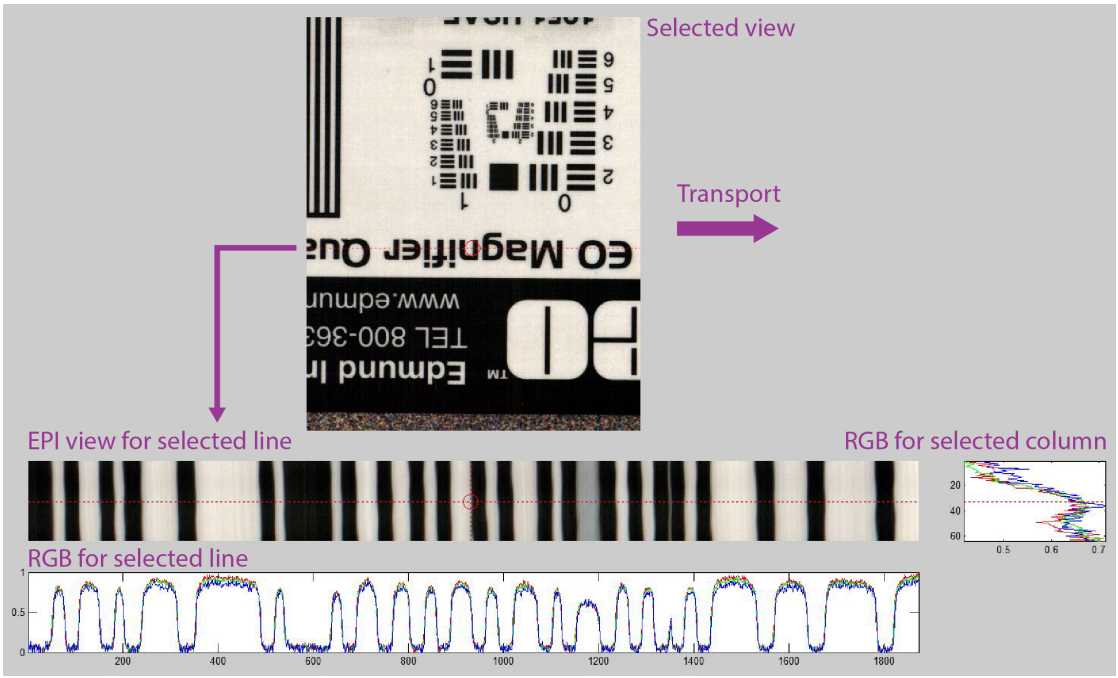


Figure 4.4: A screenshot of the acquisition software.

A screenshot of the evaluation part of the software is shown in Fig.4.4.

### 4.2.1 Flat-field correction and white-balancing

In every color camera, individual color channels have to be calibrated for the current lighting conditions before capturing in order to reproduce colors accurately. If linear sensitivity is assumed for every color channel, it is sufficient to adjust the amplification factors for each channel. This procedure is called *white-balancing*. Similarly, dark analog noise produced by the sensor — the so-called *fixed-pattern noise* (FPN) has to be suppressed and light sensitivity for individual pixels has to be balanced out. This is accomplished with the so-called *flat-field correction*. Note that this definition of flat-field correction, which is usual in the field of machine vision, collides with the definition of flat-field correction in microscopy. In microscopy, flat-field correction denotes the objective correction of Petzval field curvature, which causes inhomogenous sharpness over the field-of-view. In the described system, white-balancing and flat-field correction are combined into a single step. In Fig.4.5, a comparison between an acquisition without white-balancing and flat-field correction and one with flat-field correction applied is shown. While the left image delivers false colors and suffers from FPN, the right (corrected) image is corrected for all these errors.

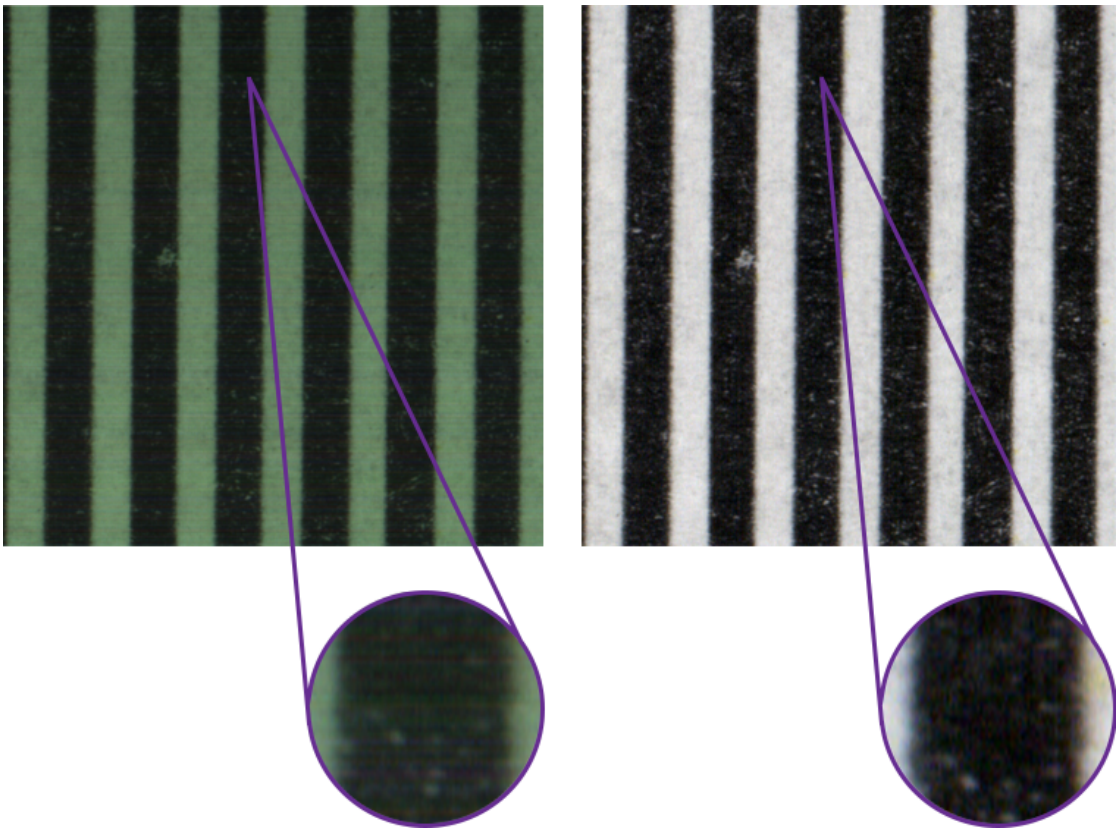

Figure 4.5: Two acquisitions of the striped calibration object. Left: no correction. Right: flat-field correction applied. Both acquisitions are of the same area, the central view is shown on both sides.

The flat-field correction is applied to the raw images right after their transfer to the PC, before the demosaicing step. It consists of two steps:

1. A calibration step, for which a striped object containing white and black stripes lying in the focal plane of the system is recorded, with $N$ raw frames (in most cases, 100 frames are sufficient) being acquired. Afterward, the correction factors for each pixel are calculated using the following formulas:

$$FFC_{dark}(x,y) = mean_i(I_i(x,y)) - std_i(I_i(x,y)) \tag{4.1}$$

$$= \frac{\sum_{i\in N} I_i(x,y)}{N} - \sqrt{\frac{\sum_{i\in N}(I_i - \bar{I})^2}{N-1}} \tag{4.2}$$

$$FFC_{bright}(x,y) = mean_i(I_i(x,y)) + std_i(I_i(x,y)) \tag{4.3}$$

$$= \frac{\sum_{i\in N} I_i(x,y)}{N} + \sqrt{\frac{\sum_{i\in N}(I_i - \bar{I})^2}{N-1}}, \tag{4.4}$$

   where $FFC_{dark}$ and $FFC_{bright}$ are the flat-field correction factors for each pixel $(x,y)$, $I_i$ is the i-th acquired image and $\bar{I}$ is the mean of $I$ over the $N$ acquisitions. The idea in this formula is that as the object moves, every pixel on the sensor records both white and black areas of the calibration object. Thus, $mean_i(I_i(x,y)) - std_i(I_i(x,y))$ should correspond to the black areas, while $mean_i(I_i(x,y)) + std_i(I_i(x,y))$ corresponds to the white areas. An alternative approach would be to record both correction factors separately, using black and white surfaces respectively.

   It has to be stressed that these formulas only apply to calibrating the sensor using a striped pattern such as the one in Fig.4.5. An alternative solution would be to divide up the calibration in two steps, acquiring a completely dark surface and a completely white surface in succession and calculating the correction factors using these two acquisitions.

   After the calibration, the correction factors are stored for later use in the correction step.

2. A correction step, which is performed after every acquisition and consists of applying the correction factors to the images as:

$$I_{corr}(x,y) = A * \left[\frac{I(x,y) - FFC_{dark}(x,y)}{FFC_{bright}(x,y) - FFC_{dark}(x,y)}\right]^{\gamma}. \tag{4.5}$$

   Hereby $I_{corr}$ is the flat-field corrected image. Simultaneously to flat-field correction, gamma correction is applied, with $\gamma$ being the gamma correction

factor and $A$ being a compression factor. In general, $\gamma < 1$, so that the high intensity values are compressed to allow for a higher dynamic range.

### 4.2.2 Demosaicing

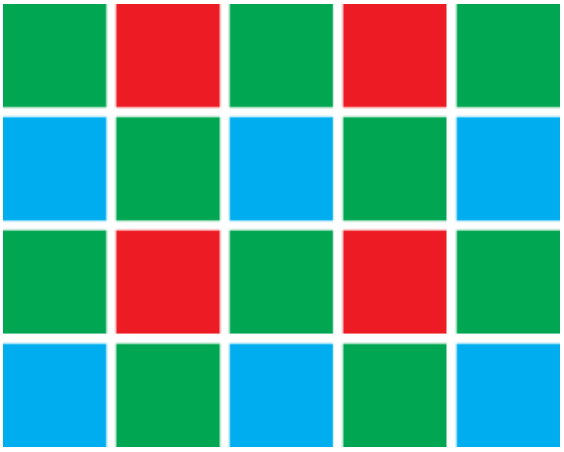

Figure 4.6: Bayer pattern structure.

The camera used in the system (Bonito CL-400, see Tab.4.1) uses a Color Filter Array (CFA) placed over the sensor. This means that neighboring pixels have different colour filters and capture light of different wavelengths (red, green and blue). While several patterns exist for the CFA, this camera uses the most usual one — Bayer pattern. As can be seen from Fig.4.6, this pattern consists of pixels clustered in 2-times-2 blocks, where 2 pixels have a green filter, 1 is red and 1 is blue. The increased number of green pixels is due to the sensitivity of the human eye on the green light frequency. The Bayer CFA is the standard way of color capture in area-scan cameras (see e.g. [LP05]). For obtaining full RGB information, a so-called demosaicing algorithm has to be performed over the raw (in our case, flat-field corrected) data. This algorithm interpolates missing RGB values by using available color information from the neighboring pixels. During this process, no resolution is recovered, therefore the resolution of a Bayer CFA sensor is often given as

$$R_{bayer} = R_{pixel}/2, \tag{4.6}$$

where $R_{bayer}$ is the resolution in [px/mm] of the Bayer CFA camera and $R_{pixel}$ is the resolution of a gray-scale sensor with the same dimensions. However, this is a very rough approximation. Firstly, as one can see from Fig.4.6, the resolution of the green channel is 2-times higher than the resolution of the other two channels. Secondly, resolution can be gained by making the (often valid) assumption that

the color channels are correlated, and borrowing information from the neighboring pixels (an example of such an algorithm can be found in [MHC04]). Naturally, there are scenes which do not fulfill this assumption, such as holographic films, or even fluorescent microscopic imaging.

In the case of line-scan light-field imaging, standard demosaicing algorithms are not feasible. This is because individual lines in our system correspond to different angular views and should not be interpolated, as this causes loss of angular resolution (as different views are composed of concatenated lines coming from different sensor locations) as well as loss of color information. While this issue also persists in traditional photography, it is much more severe in the case of line-scan light-fields, as such a system uses short working distances in order to increase the angular resolution. Additionally, as already mentioned above, we use sparse light-field acquisitions, during which only some lines are read-out from the sensor. These lines can potentially be spaced far away from each other, having a large angular difference. Moreover, in the slit aperture system, which will be described later and is especially designed to increase angular difference between neighboring lines, artifacts are generated in the image, which are not acceptable. Thus, a custom demosaicing algorithm had to be designed to cope with these issues.

The employed algorithm makes use of the fact that the camera is able to read-out pairs of consecutive lines, while the gaps between those pairs can be arbitrarily large. Keeping in mind that one line of the sensor always contains just green and red, or blue and green information, we treat those pairs of consecutive lines as one view, which contains full RGB information. In this way, the angular difference between individual views can be chosen deliberately, as there is no need for interpolating between them. Fig.4.7 shows how this is accomplished. As red and blue color information is contained in only half of the pixels, a convolution with an averaging kernel $(0.5, 0, 0.5)$ is applied to both lines. In this way, two new lines are formed, which contain the corresponding color information in the other half of the pixels. Next, the color channels are combined together to form full RGB pixels. Because there are two green channel values for each pixel, we take their mean as the G-value for the pixel.

### 4.2.3 Imperfections in the setup

Unfortunately, in the real world, no imaging system is perfect, and apart from a fundamental resolution limit caused by the Heisenberg Uncertainty principle, there are a variety of manufacturing as well as mechanical adjustment factors involved in the quality of the captured images. Optical aberrations such as geometrical distortion, Petzval field curvature or lateral and longitudinal chromatic aberrations as well as improper illumination all reduce the effective resolution and cause artifacts in the images. While some of these problems can be addressed by a more suitable

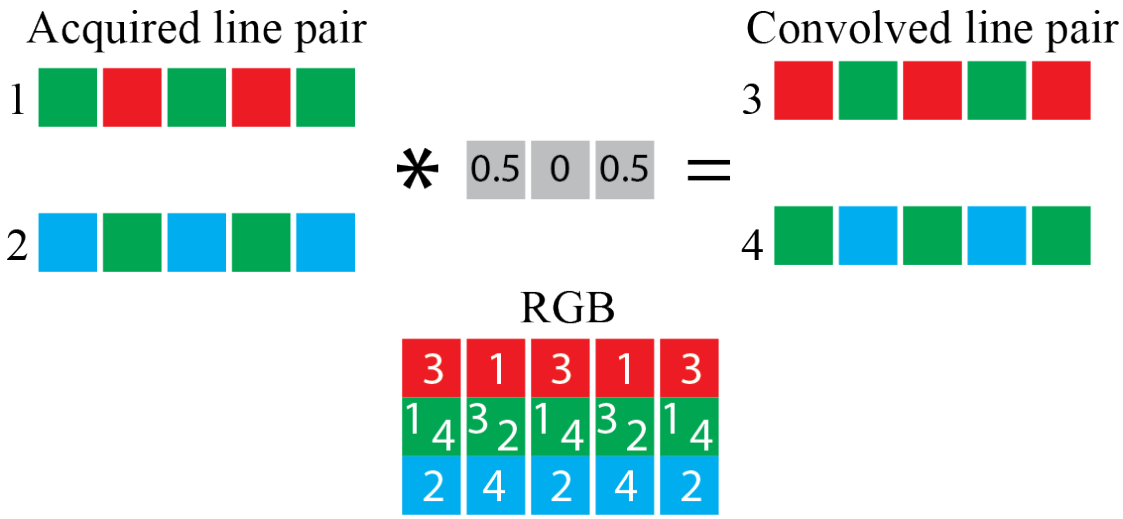


Figure 4.7: The demosaicing process shown demonstratively on one image line: for each line of the resulting image, two neighboring lines of the sensor are read-out. These contain the complete RGB color information. The process of interpolating between the pixels and selecting the appropriate ones is shown here. First, both sensor lines are convolved with the averaging kernel to compute the colors on the missing pixels. Afterwards, the corresponding pixels (marked here by the sensor line number) are collected to form the RGB signal. Note that each pixel contains two green channel values; these are averaged during the process.

or more precise setup or higher precision components, others are inherent. In this section, the relationships of some of these problems to the EPI-images and thus to the quality of most light-field algorithms will be explained. We also discuss possible post-processing steps to mitigate these.

### Geometrical distortion

When considering real (thick) lenses, it can be observed that the transformation from the object space to the image space is not rectilinear, as is desired for performing optical measurements, but is rather modeled using (often) high-order polynomials. This effect is often radially symmetric because of the shape and manufacturing process of the lens. As a result, the image can appear to be bent 'inward' or 'outward'. The former case is called *Pincushion distortion* while the latter is called *Barrel distortion*. Fig.4.8 illustrates these two distortion types. Pincushion distortion can be associated with tele lenses while barrel distortion mostly appears on wide angle lenses (including the fish-eye lens, which knowingly utilizes

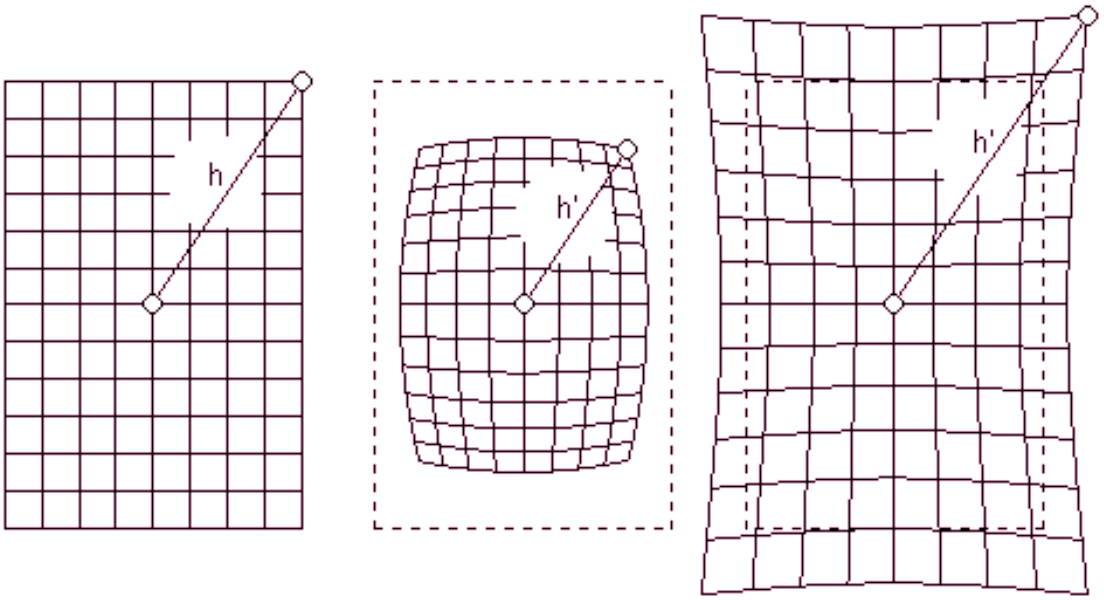


Figure 4.8: Barrel and pincushion distortion compared with the original image. Courtesy of [Wal14].

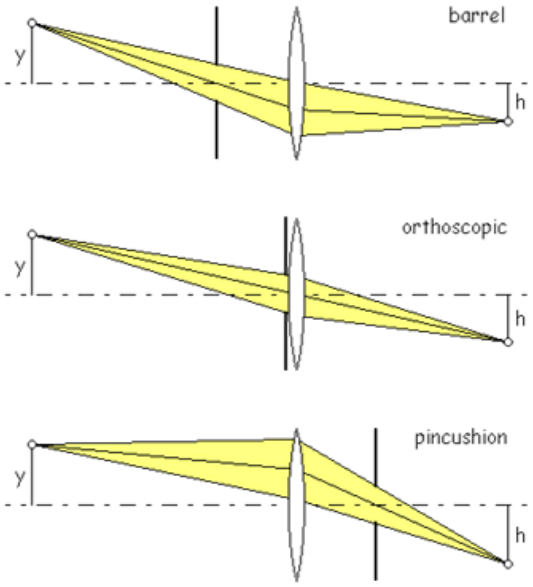


Figure 4.9: Origins of barrel and pincushion distortion. Hereby, y is located on the object plane and h on the image plane. Both effects are associated with the objective aperture. If the aperture is placed in front of the lens, barrel distortion occurs. The opposite is true for pincushion distortion. Courtesy of [Wal14].

the distortion to image a very wide viewing angle onto a spatially limited image plane). The origins of both distortion types are explained visually in Fig.4.9. The distortion effect can be quantitatively described by

$$\Delta M = \frac{M(I) - M}{M}, \tag{4.7}$$

where $M$ is the axial lateral magnification and $M(I)$ is the off-axis magnification at the image plane (see also [SD14]). Visible effects of geometrical distortion mostly occur for a lens whose aperture is wide open, as the level of the distortion increases to the edges of the lens. Therefore, the effects on standard photographs are mostly minor and can be corrected using off-the shelf post-processing software. This is also true for the line-scan light-field setup. However, when using a slit to increase the angular range, as shown in Sec.3.4, it is important to fully open the lens aperture, resulting in visible distortions.

A simple way to correct distortions is to find a more suitable lens. During the experiments, we found that most lenses in the upper price range are well-corrected for geometrical distortion. Also, when dealing with the slit, usually only few central lines of the chip are used, which suffer far less from distortions.

Another possibility consists of warping all views w.r.t. a central (nadir) view as a post-processing step. However, this wasn't considered for two main reasons:

- When dealing with wide angular ranges, occlusions often prevent successful warping by hiding the respective object points in some views.
- This kind of post-processing hinders real-time performance.

### Defocusing of lateral views

Another effect caused by imperfections of the optics is the reduction of resolution in lateral views, caused by multiple optical aberrations including spherical aberration and astigmatism. These effects are caused by the violation of the paraxial assumption and increase further from the optical axis, just as geometrical distortion. If they are significant, lateral views become blurred and the depth resolution is reduced because it becomes harder to estimate the correct slope in the EPI, even if the central view is perfectly focused. Because of this effect, aspherical lenses with acceptable correction levels are preferred for line-scan light-field imaging.

### Line-scan effects

There are also effects in line-scan light-field imaging, which do not stem from optics, but simply from imprecision of the mechanical setup. This includes not placing the sensor lines orthogonally to the transport, or, in the case of the setup

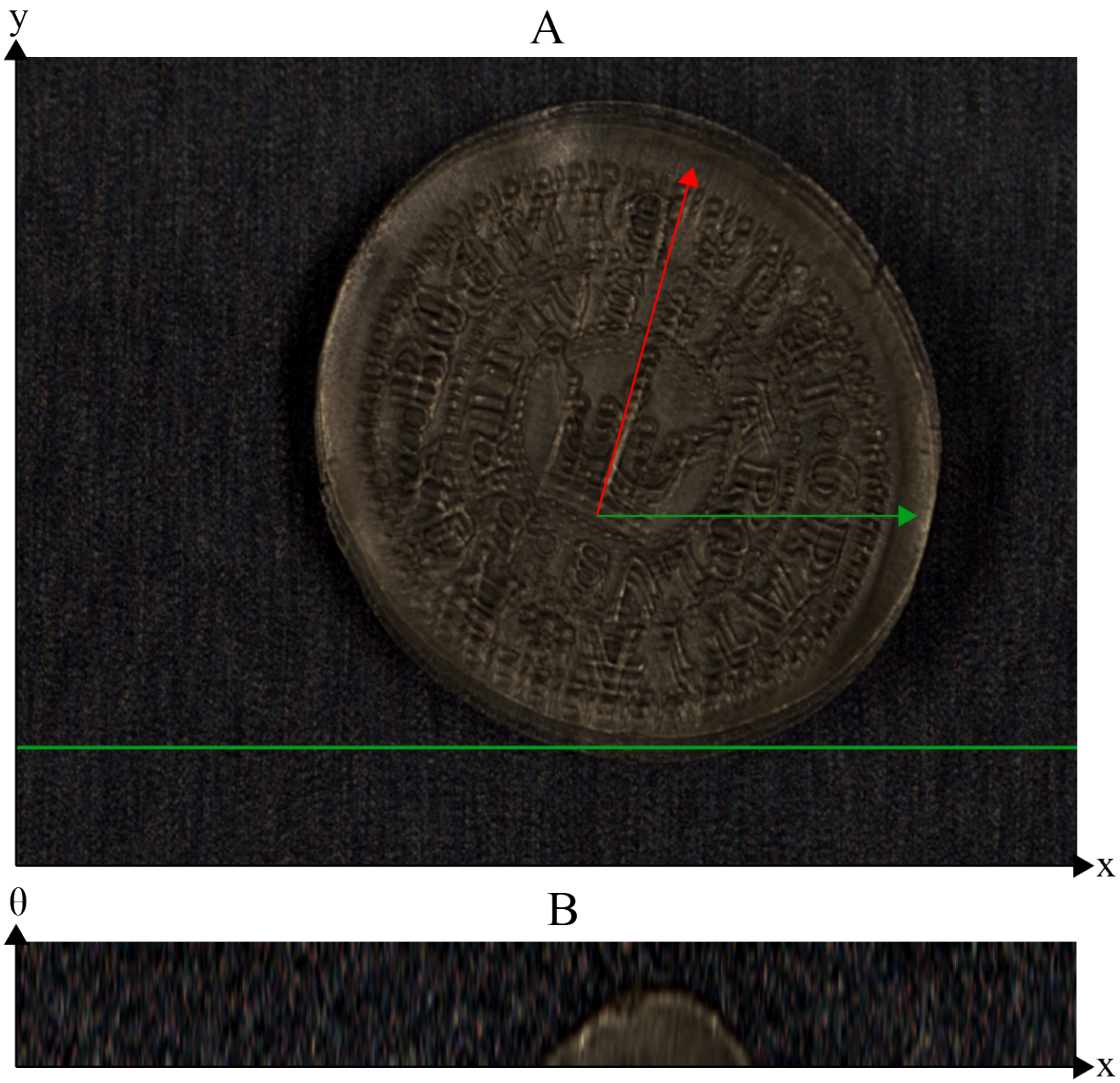


Figure 4.10: Three overlapped views of a coin (A) and the EPI of the green highlighted line (B) displaying an issue with the camera that wasn't oriented orthogonally to the transport. This causes structures in EPI to disappear in certain views, and appear in the neighboring EPI. As the surface of the coin is higher than the $z_0$-plane, a horizontal shift of the object points across the views would be expected (green arrow). However, because of the incorrect camera orientation, the object additionally moves in the transverse direction (y), as indicated by the red arrow.

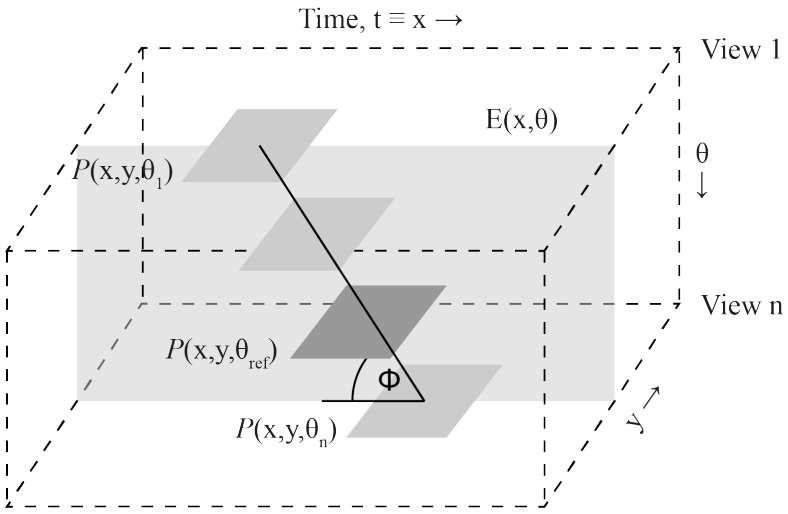


Figure 4.11: Graphical representation of the depth estimation algorithm. For a pre-defined range of slopes $\phi_i$, patches $\mathcal{P}$ are compared with the reference patch $\mathcal{P}_{ref}$, with each patch lying on a single line with incline $\phi_i$. This procedure is repeated for each slope $\phi_i$, each pixel $(x, y)$, as well as for each color channel.

with the slit aperture, having an aperture whose width is not constant along its length. In both cases the light-field is not recorded correctly.

In the case of incorrect camera orientation, objects become skewed, as each time a different object line is being captured. Furthermore, the EPI structures do not span the entire angular dimension, as the object feature moves from one EPI line to another (see Fig.4.10). This has to be avoided, either by exact mechanical adjustment (as it is the case for our system), or by calibration, because it reduces the applicability of depth estimation algorithms (usually, these algorithms expect the structure to be visible in the same EPI in all views).

In the case of an imperfect slit aperture, this causes varying angular range and resolution in different lines, violating the linearity of structures in EPI. As it is hard to manufacture a slit with high precision, it is generally recommended to use the tradition setup without the slit aperture whenever possible.

## 4.3 Depth estimation

The depth estimation algorithm used in this work was developed by AIT and is largely based upon the algorithm described in [Što+14]. This is essentially a multi-view stereo matching algorithm comparing image patches in different views. More precisely, multiple hypotheses are generated for each pixel in an arbitrarily chosen

reference view, which correspond to multiple depth levels. These hypotheses are then tested, with the best hypothesis yielding the final estimate.

Formally, the process can be defined by constructing cost maps $C_i(x, y, \phi_i)$ for each pixel $(x, y)$ and a specified range of disparity slopes $\phi_i$ in EPI. Finally, the disparity map $D$ is computed by

$$D(x, y) = \min_i C_i(x, y, \phi_i). \tag{4.8}$$

$D(x, y)$ corresponds to the disparity determined in each pixel. The final step is to convert the determined disparity to absolute depth, which was described in Sec.3.3.1 and includes a multiplication with a factor determined by calibration. For relative measurements, this step can be omitted.

In order to compute $C_i(x, y, \phi_i)$, a cost function is required. The cost for a given slope $\phi_i$ in the EPI is minimal when features on the slope are most similar based on some similarity measure. In our case, *Census Transforms* (CT) of different views are checked for similarity by means of the Hamming distance. CT is a binary filter transform defined by comparing each pixel in an image with its $m \times m$-neighbourhood in a binary fashion. The result of the comparison is saved as a $(m \times m - 1)$-bit descriptor. For example, if a $3 \times 3$-neighbourhood is used, 1 byte (8 bits) is stored for each pixel. After performing this transform for each view in the light-field, the census transforms of the corresponding patches are compared with an arbitrary reference view using the Hamming distance (see Fig.4.11).

CT was first introduced in [ZW94] and was found to be comparable to other widely used stereo measures, such as *Normalized Cross-Correlation* (NCC) or the *Sum of Absolute Differences* (SAD) for depth estimation from EPIs in [Što+14]. At the same time, on many hardware platforms, it is also computationally very efficient and suitable for real-time processing. Other advantages include robustness against brightness changes and its locality, making it suitable for parallel implementation.

In order to formally describe the cost map computation, we define by $\mathcal{P}(x, y, \theta)$ an $m \times m$ image patch that is centered around $(x, y)$ in a particular view $\theta$. A set of image patches along a certain slope $\phi$ in the EPI is defined by

$$\Omega(x, y, \phi) = \mathcal{P}_i(x + \phi(\theta_i - \theta_{ref}), y, \theta_i), i = 1..n, \tag{4.9}$$

where $\theta_{ref}$ is the index of the reference view and $n$ denotes the total number of views. As the $x$-coordinate is not guaranteed to be an integer, interpolation has to be used in general. In this work, linear interpolation was applied to the original images for each tested hypothesis, before performing the Census Transform.

The cost function for the Census Transform in multi-view stereo matching is

defined by

$$C(x, y, \phi) = \sum_i \mathcal{H}(\mathcal{P}(x, y, \theta_i), \mathcal{P}(x, y, \theta_{ref})); \mathcal{P} \in \Omega(x, y, \phi), i = 1..n, \qquad (4.10)$$

where $\mathcal{H}$ is the Hamming distance, $\mathcal{P}(x, y, \theta_{ref})$ is the reference patch and $\mathcal{P}(x, y, \theta_i)$ is the corresponding image patch in the view $\theta_i$. This procedure corresponds to summing the Hamming distances between the corresponding patches on the slope $\phi$ and the reference patch, $\mathcal{P}(x, y, \theta_{ref})$, respectively.

Before choosing the minimal cost, the cost stack $C(x, y, \phi)$ is low-pass filtered in a post-processing step. This regularizes the cost function and thus reduces artifacts caused by noise or other interfering effects. A simple box filter was used in the view domain $(x, y)$ in order to attain real-time performance and was found to produce satisfactory results.

Notice however, that the depth estimation approach taken here is one optimized for speed, while at the same time striving for acceptable result quality. As it is a local approach, it doesn't really deal with outliers caused by noise or occlusions, or absence of image structure hindering estimation of slopes in the EPI domain, which can significantly alter the outcome of the algorithm (although averaged out in the results). In fact, the assumption is made that no occlusions occur in the recorded light-fields. This assumption was found to be valid for most inspected 2.5D relief objects, as long as the maximal disparity remains relatively small. Furthermore, no interpolation of missing depth values was used. Nevertheless, the algorithm produced a dense depth map.

The algorithm was performed for all three channels with successive weighting of the results. Weights of $R = 1/4, G = 2/4, B = 1/4$ were used in order to account for denser information content in the green channel.

## 4.4 SNR-improvement, Refocusing and All-in-focus images

Good *signal-to-noise ratio* (SNR) is a prerequisite for further processing using *digital signal processing* (DSP) algorithms. In imaging, high SNR is often achieved either using smaller F-numbers to improve light conditions and reduce analog and digital noise, or by increasing the exposure time. However, long exposure times are inappropriate for line-scan imaging, resulting in motion blur and limiting system performance, while small F-numbers reduce the depth-of-field and can also lead to more severe optical aberrations.

To tackle these shortcomings, a solution often used in line-scan acquisitions systems is to acquire more than one image of an object by using sensors with

multiple lines and treating each of these lines as equivalent line-scan acquisitions, similar to the line-scan light-field approach. After the acquisition, the individual images are then aligned and integrated, allowing for theoretical increase of SNR at best proportional to the number of integrated acquisitions $n$. This approach is called *time delay and integration* (TDI) and is already implemented in some CCD sensors on analogue level, before analog-to-digital conversion is performed to circumvent digitizing errors [WYS92]. An implementation of TDI was also proposed for RGB CMOS sensors, however offering less efficient SNR improvement of about $\sqrt{n}$ [Bod+12].

Knowing that in an imaging system, each line of a sensor looks on an object in a different angle (except for telemetric lenses), it becomes clear that the applicability of TDI is largely reduced to objects lying close to the focal plane of the camera. Objects lying on different depths are imaged at different resolutions normally to the transport and acquisitions of these cannot be added naively integrated. Applying TDI on such objects results in motion blur.

As the line-scan light-field system performs multiple acquisitions of the same object while simultaneously obtaining depth information, it can be used to intelligently integrate correct pixels from these acquisitions to improve SNR. Note that refocusing the light-field is equivalent to summing pixel values along slopes with varying angle in the EPI. To introduce this formally, we define

$$E_\phi(x, y, \theta) = E(x + (\theta - \bar{\theta})\phi, y, \theta), \tag{4.11}$$

where $\bar{\theta}$ is a (arbitrarily chosen) reference view. This corresponds to a "shift" in the viewing domain by a amount proportional to $\phi$. Then,

$$I_\phi(x, y) = \sum_{\theta=1}^{n} E_\phi(x, y, \theta) \tag{4.12}$$

denotes a refocused view, where $n$ is the number of views. Refocusing is possible along the entire DoF. As a result of summing up pixels from different views, $I_\phi$ enhances SNR while providing shallower DoF. More precisely, SNR is improved by $\sqrt{n}$. This can be derived from the fact that the added signals are correlated, thus growing proportionally to $n$, while noise is usually uncorrelated, only growing by $\sqrt{n}$. Thus,

$$\mathrm{SNR}_{improved} = \frac{S}{N}\mathrm{SNR} = \frac{n}{\sqrt{n}}\mathrm{SNR} = \sqrt{n}\mathrm{SNR}. \tag{4.13}$$

Such refocusing thus improves SNR while at the same time reducing the DoF. However, it is possible to keep the DoF constant together with the improvement in SNR. Since the depth and hence the optimal integration slopes in the EPI are already known after performing depth estimation, one can combine the different

refocused images by refocusing each pixel "optimally" along the correct slope. This produces an *all-in-focus* image of the scene which enhances SNR by $\sqrt{n}$ without the shortcomings of TDI. It is worth pointing out that while TDI works on analog level, all-in-focus image construction is applied only after digitalization. Thus, all-in-focus images are equivalent to a so-called digital TDI, with the important difference of locally adaptive integration along optimal EPI slopes.

One can also think of all-in-focus images as a way to improve the DoF. As a consequence of the above, the use of an F-number $F$ which is $\sqrt{n}$-times larger yields approx. the same SNR in the all-in-focus image, while expanding the DoF (remember that DoF is roughly proportional to $F$).

# 5 Results and discussion

In this chapter, experimental results obtained by the line-scan light-field system are presented and discussed. For this, a variety of objects suitable for inspection scenarios was acquired. These were divided into three categories:

1. ground-truth objects,
2. real 2.5-D/3-D objects with a non-flat surface, and
3. *optical variable devices* (OVDs), e.g. holograms.

A summary of parameters used for the algorithm is found in Tab.5.1.

## 5.1 Ground-truth objects

In order to find out the practical depth resolution, we used a 3-D printed ground-truth object containing 5 slopes with different gradients (see Tab.5.2 for a summary of the different slope gradients). A random pattern was applied on the object's surface in order to provide our matching algorithm with enough structure for correct depth estimation.

The object was subsequently acquired using both the high- and low-resolution versions of the system (Fig.5.1 and Fig.5.4, respectively). Afterwards, the depth estimation was performed. In both cases, directional illumination was used, and the object was imaged using an aperture number of $F = 5.6$, resulting in different DoFs for the two setups. While the high-resolution setup has a DoF of approx. 1.4 mm, the low-resolution setup has a DoF of approx. 7 mm. In both cases, the system was focused on 0 mm, which results in considerable defocus for the high-resolution setup in some parts of the ground-truth object (see Fig.5.1D). On the other hand, the low-resolution setup remains mostly in focus on the whole object (see Fig.5.4B).

From the qualitative depth map comparison (Fig.5.1F and Fig.5.4D), it is evident that the overall shape of the object was reconstructed correctly.

In order to make a quantitative assessment, the height values of the steepest slope and the middle slope were further plotted against the predicted slopes (Figs. 5.2, 5.3, 5.5 and 5.6). To take into account deformations of the ground-truth

| All objects | |
|---|---|
| Cost function | Census transform |
| Patch size | $3 \times 3$ |
| Cost computation interpolation | bilinear |
| **Grond-truth high resolution** | |
| Cost stack filtering | mean |
| Filter domain | $5 \times 5$ |
| Filter iterations | 3 |
| Disparity step per hypothesis | 1 |
| **Grond-truth low resolution** | |
| Cost stack filtering | mean |
| Filter domain | $3 \times 3$ |
| Filter iterations | 3 |
| Disparity step per hypothesis | 1 |
| **Coin** | |
| Cost stack filtering | mean |
| Filter domain | $5 \times 5$ |
| Filter iterations | 3 |
| Disparity step per hypothesis | 0.5 |
| **Coin group with slit aperture system** | |
| Cost stack filtering | mean |
| Filter domain | $5 \times 5$ |
| Filter iterations | 3 |
| Disparity step per hypothesis | 0.2 |
| **Plaster** | |
| Cost stack filtering | mean |
| Filter domain | $3 \times 3$ |
| Filter iterations | 3 |
| Disparity step per hypothesis | 0.5 |
| **EUR10 banknote** | |
| Cost stack filtering | mean |
| Filter domain | $5 \times 5$ |
| Filter iterations | 3 |
| Disparity step per hypothesis | 1 |

Table 5.1: Summary of depth estimation algorithm parameters.

object after the printing, the height of the object was also measured at various points along the slopes using a dial indicator with a resolution of 0.05 mm. The

| Length across slope [mm] | 100 | Slope 1 | Slope 2 | Slope 3 | Slope 4 | Slope 5 |
|---|---|---|---|---|---|---|
| Width of each slope [mm] | 10 | | | | | |
| Height left [mm] | | 0 | 0 | 0 | 0 | 0 |
| Height right [mm] | | 1 | 2 | 3 | 4 | 5 |
| Gradient [%] | | 1 | 2 | 3 | 4 | 5 |

Table 5.2: Summary of slopes on the ground-truth object.

measurements are shown in the corresponding figures.

Looking at the height measurement plots, the estimated slope deviates slightly from the theoretical slope, but exact measurement with the dial indicator shows that these deviations can largely be attributed to deformations of the object itself. There is a slight inaccuracy of the algorithm on the lower part of the highest slope, but as this inaccuracy is consistent in both the high- and low-resolution plots, it can also be suspected to be a deformation in the object between the measurement and the acquisition.

Due to the dense sampling in the angular domain, it is a non-trivial task to derive theoretically the maximal depth resolution of our system. On one hand, one may use as a lower limit the resolution obtained with one disparity step in a standard stereo, but as already discussed before, the resolution of the light-field system is in fact higher. Moreover, the depth estimation algorithms available nowadays can mostly work with sub-disparity steps, assuming smoothness in the cost domain. As the measurement performed using the ground-truth object was deemed to produce roughly correct depth values, it was decided to measure the resolution using the standard deviations of depth values obtained for each height along the slopes, i.e. the standard deviations of depth values across each slope. The standard deviation interval was then taken as the depth resolution at that height. Tab. 5.3 summarizes the measurement.

Comparing the two resolutions used, it is rather interesting that the low-resolution setup has seemingly higher depth resolution (approx. 70 μm on the steepest slope and 45 μm on the middle slope on average, compare Tab.5.3) than the high-resolution setup (93 μm on average on the middle slope and 155 μm on average on the steepest slope). This can be explained twofold: Firstly, the low resolution proves to be beneficial in depth estimation of the ground-truth object, as the structures printed on it mostly match the size of the pixels in this case. Therefore, depth can be exactly estimated in the majority of pixels. This decreases the deviations from the exact value. Secondly, the angular range of the low-resolution setup is actually higher than that of the high-resolution setup, further contributing to depth estimation when using sub-disparity measurement (when only considering depth for one disparity, the depth resolution is lower, since the lateral resolution

| | Resolution [μm] | | | |
|---|---|---|---|---|
| | highest | lowest | average | std |
| High-resolution steepest slope | 60.4 | 305.5 | 154.6 | 47.5 |
| High-resolution middle slope | 47.9 | 155.1 | 93.2 | 22.1 |
| Low-resolution steepest slope | 39.5 | 109.6 | 69.9 | 18.5 |
| Low-resolution middle slope | 25.7 | 78.1 | 44.6 | 11.2 |

Table 5.3: Quantitative results of the depth resolution measurement. For each measured slope (Figs. 5.2, 5.3, 5.5 and 5.6), the minimal, maximal, and average resolution is shown, as well as its standard deviation. The resolution hereby corresponds to the measured standard deviations and their variations along the respective slopes. The standard deviations are calculated across the $y$-dimension of each slope, assuming that the height of the slope on one $x$-position is constant.

is also lower).

Another important result deduced from the measurement is that the depth resolution seems to be almost constant in the low-resolution setup, while the depth resolution of the high-resolution setup varies greatly in different heights. This can be attributed to the larger depth-of-field of the low-resolution setup, which spans almost the whole depth range, while the other setup has a very limited depth-of-field compared to the total height of the object. This causes structures on higher positions to become blurred, limiting the applicability of the algorithm, which relies on sharp spatial gradients in the images. Thus, the depth resolution becomes as low as 300 μm on the steepest slope on its upper part, which is very blurry.

Although the previous discussion seems to undermine the importance of high spatial resolution in depth estimation, this is not the case. Rather, it seems that the structure on the selected ground-truth object is more suited to the low-resolution setup. This means that the high-resolution setup requires a larger matching window to match the same structures as the low-resolution setup. An additional argument in favor of the low-resolution setup is its higher baseline and therefore a higher stereo depth resolution (compare Tab.4.2). On the other hand, depth estimation is highly dependent on the type of inspected object. Naturally, to obtain more detailed depth maps, it is necessary to work with high spatial resolutions. Objects that contain many detailed structures can thus largely benefit from increased spatial resolution. Furthermore, when inspecting metallic objects, it is often beneficial to increase spatial resolution so that the structure of the material itself can be observed, thus enabling more detailed pattern matching. Of course, this tends to limit the depth-of-field, so that a compromise has to be found in most real-life cases. That being said, we will focus on testing the system in more

realistic scenarios in the next section.

## 5.2 Real 3-dimensional objects

Real testing objects in this study were chosen to be representative of objects which are suitable for inspection using the system. Non-flat objects like coins and plasters were considered. Additionally, a folded banknote was taken as an example of a printed object containing periodic texture. In print inspection, high speeds (up to 10 m/s) are often required. These present a major problem mechanically, as paper tends to fold at these speeds. Measuring the depth of the paper would allow to compute a warping pattern which would correct for the folding, making print inspection a lot easier. Additionally, features such as intaglio printing, very popular in security print, could become measurable.

### 5.2.1 Historical coin

The first experiment was done inspecting a historical coin using the high-resolution setup. This should serve as an example for inspecting and reconstructing metallic objects. These objects are traditionally hard to image, since their multiple reflections hinder proper correlation. Fig.5.7 shows the achieved results. The system was focused on the top of the coin surface, using $F = 5.6$ and diffuse axial illumination, which is suitable for metallic objects, producing bright-field for a wide range of surface normals.

Having a closer look at the displayed depth map, one can see recognize detailed features of the coin, such as its elevated border or the protruding crown in the middle. These features are also visible in the 3-D reconstructed model. Another interesting comparison can be made between a hypothetic TDI image (Fig.5.7C) and the all-in-focus image produced by our method (Fig.5.7D). Considerable motion blur is visible in the TDI image, whereas the all-in-focus image reduces noise but preserves detail in all focal planes. Admittedly, real TDI systems usually show smaller baselines, which reduces this problem somewhat, but it still occurs for objects which show high height variations.

### 5.2.2 Plaster

A similar experiment using the same setup, albeit with directional illumination, was performed on a medical plaster (Fig.5.8). In this case, the goal was to detect small height variations in the holes of the plaster. The system was focused on the top of the plaster surface using $F = 4$ and directional illumination.

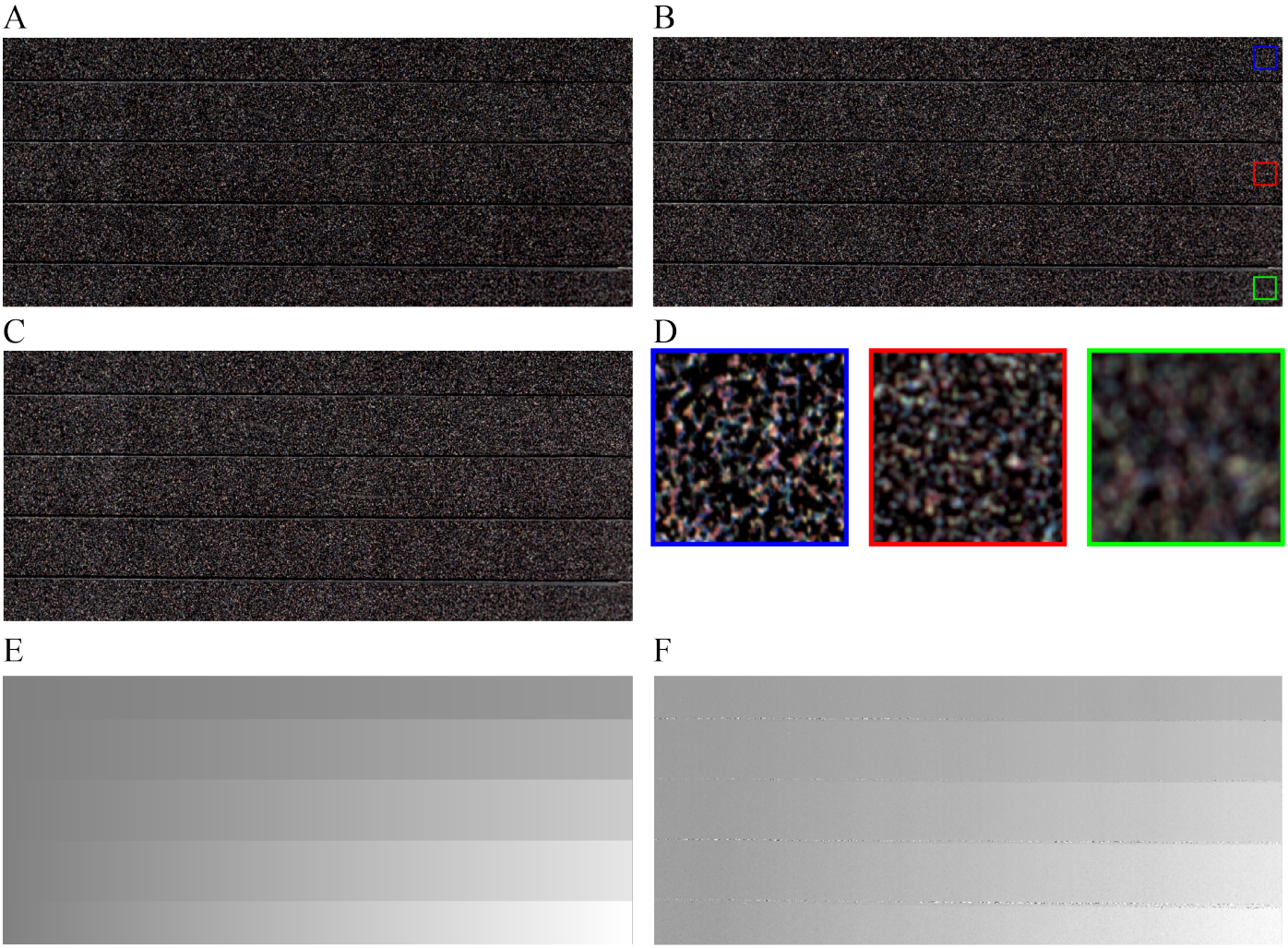


Figure 5.1: Ground-truth object acquired using the high-resolution setup (20 µm/px). Slopes rise from left to right, with the shallowest slope on the top. (A), (B) and (C) show the left-most, central, and right-most views of the object, respectively. In (D), the highlighted parts from (B) are magnified. On the bottom of the figure, a qualitative comparison between the theoretical ground-truth (E) and the real depth map (F) is shown. A summary of the depth estimation algorithm parameters for each object in the results section can be found in Tab.5.1.

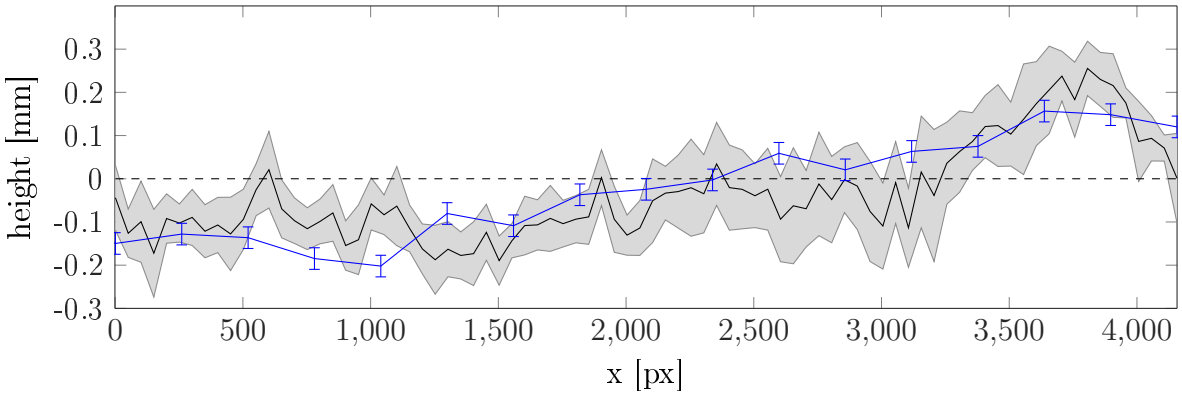


Figure 5.2: Height measurement along the steepest slope with the high-resolution setup. The black dashed line corresponds to the predicted height. The slope was plotted horizontally rather than with the real gradient in order to make the comparison with the measurement easier. The black line shows the estimated mean height (with the mean being computed from height values measured across the slope) while the grey area corresponds to the standard deviation. The blue line corresponds to the height values measured with the dial indicator.

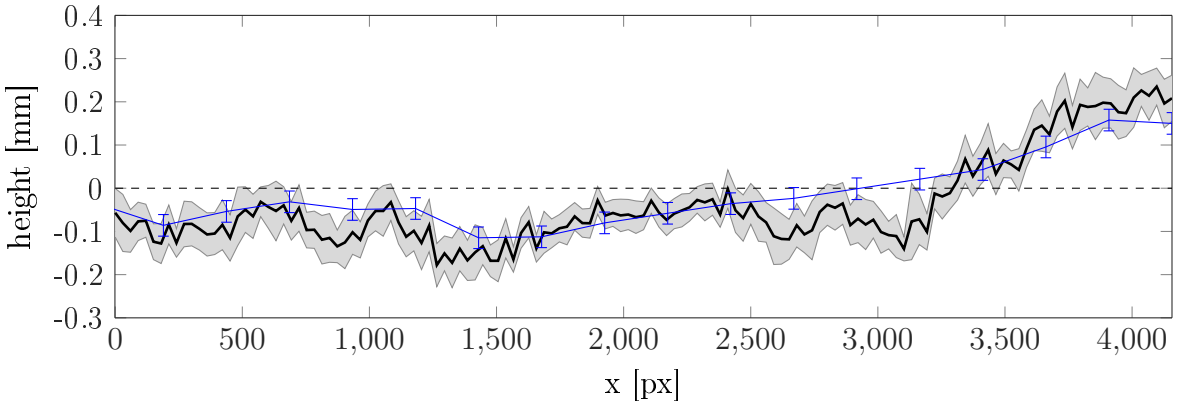


Figure 5.3: Height measurement along the middle slope with the high-resolution setup. The black dashed line corresponds to the predicted height. The slope was plotted horizontally rather than with the real gradient in order to make the comparison with the measurement easier. The black line shows the estimated mean height (with the mean being computed from height values measured across the slope) while the grey area corresponds to the standard deviation. The blue line corresponds to the height values measured with the dial indicator.

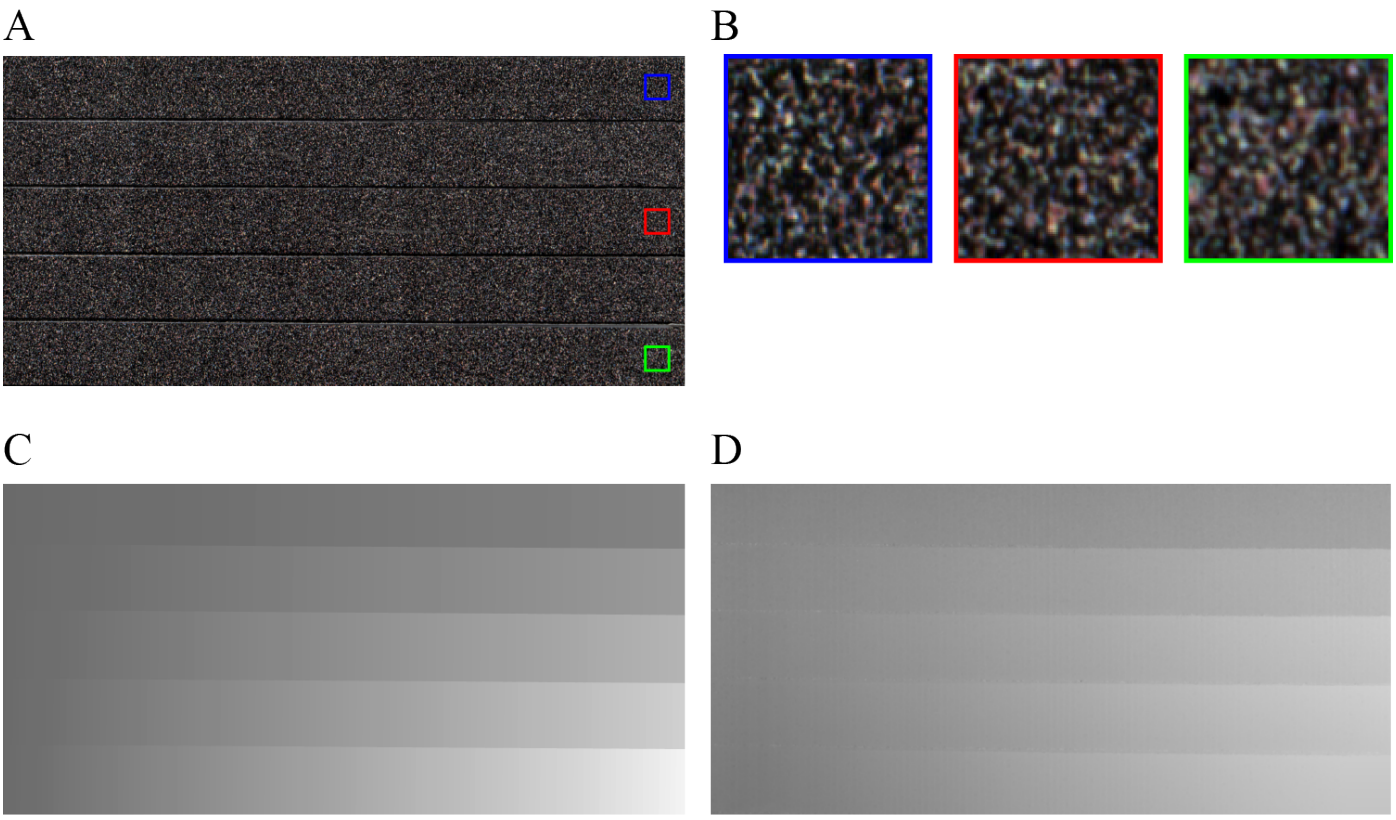


Figure 5.4: Acquisition of the ground-truth object using the low-resolution setup (40 µm/px). Slopes rise from left to right, with the shallowest slope on the top. (A) shows the central views of the object. In (B), the highlighted parts from (A) are magnified. On the bottom of the figure, a qualitative comparison between the theoretical ground-truth (C) and the real depth map (D) is shown. A summary of the depth estimation algorithm parameters for each object in the results section can be found in Tab.5.1.

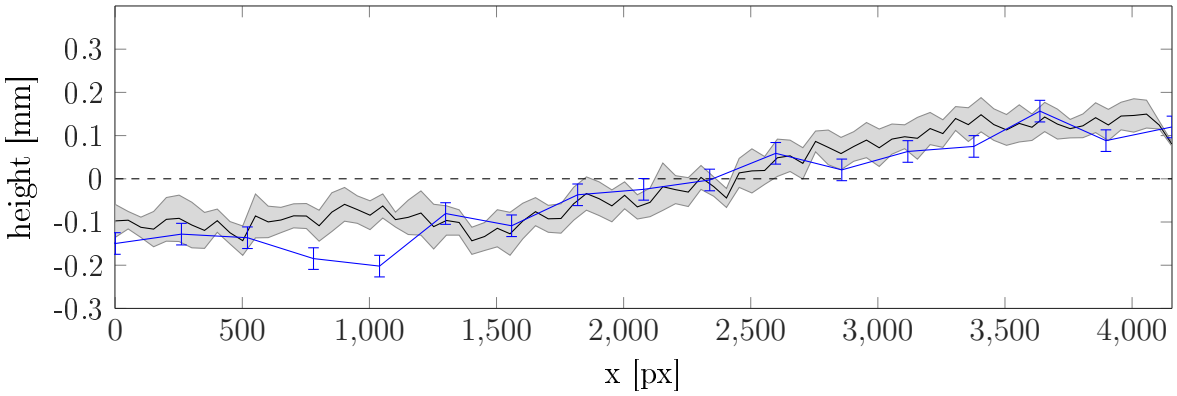


Figure 5.5: Height measurement along the steepest slope with the low-resolution setup. The black dashed line corresponds to the predicted height. The slope was plotted horizontally rather than with the real gradient in order to make the comparison with the measurement easier. The black line shows the estimated mean height (with the mean being computed from height values measured across the slope) while the grey area corresponds to the standard deviation. The blue line corresponds to the height values measured with the dial indicator.

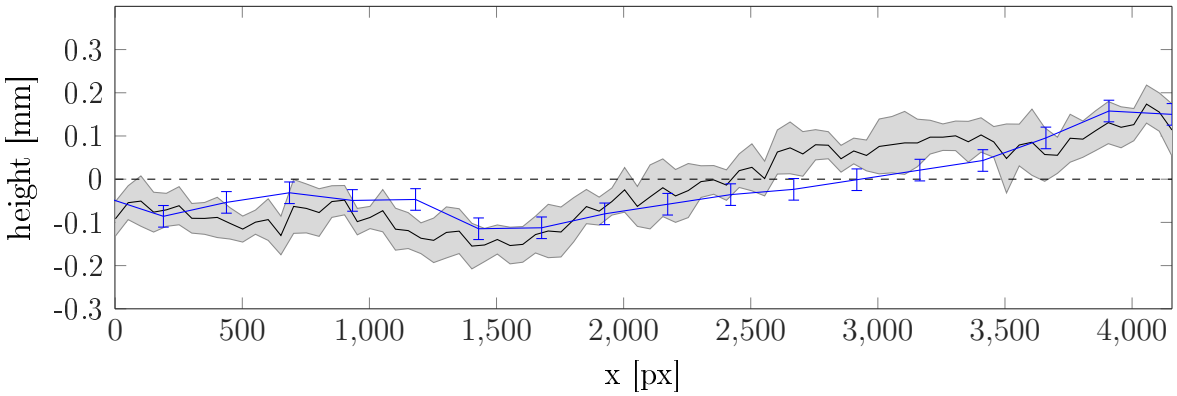


Figure 5.6: Height measurement along the middle slope with the low-resolution setup. The black dashed line corresponds to the predicted height. The slope was plotted horizontally rather than with the real gradient in order to make the comparison with the measurement easier. The black line shows the estimated mean height (with the mean being computed from height values measured across the slope) while the grey area corresponds to the standard deviation. The blue line corresponds to the height values measured with the dial indicator.

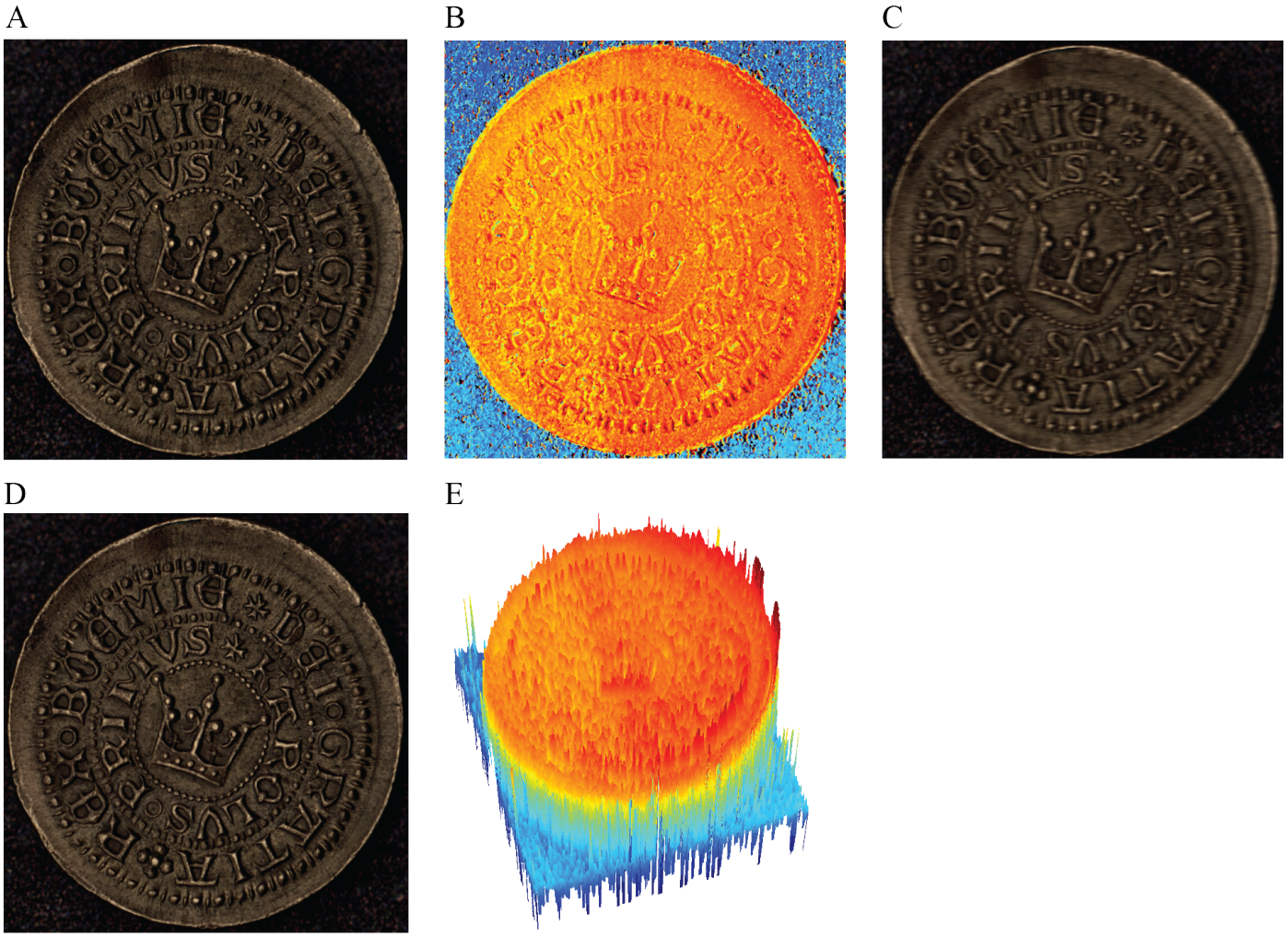


Figure 5.7: Surface reconstruction of a historical coin using the high-resolution setup. (A) shows the central view from the system, (B) shows the color-coded disparity map (blue - bottom, red - top), (C) shows a TDI acquisition using a hypothetical digital TDI system with the same parameters as ours. (D) shows the all-in-focus image generated by integrating the correct slopes. In (E), a 3-D model generated from (B) is displayed. A $7 \times 7$ medial filter was applied to (B) before 3-D visualization.

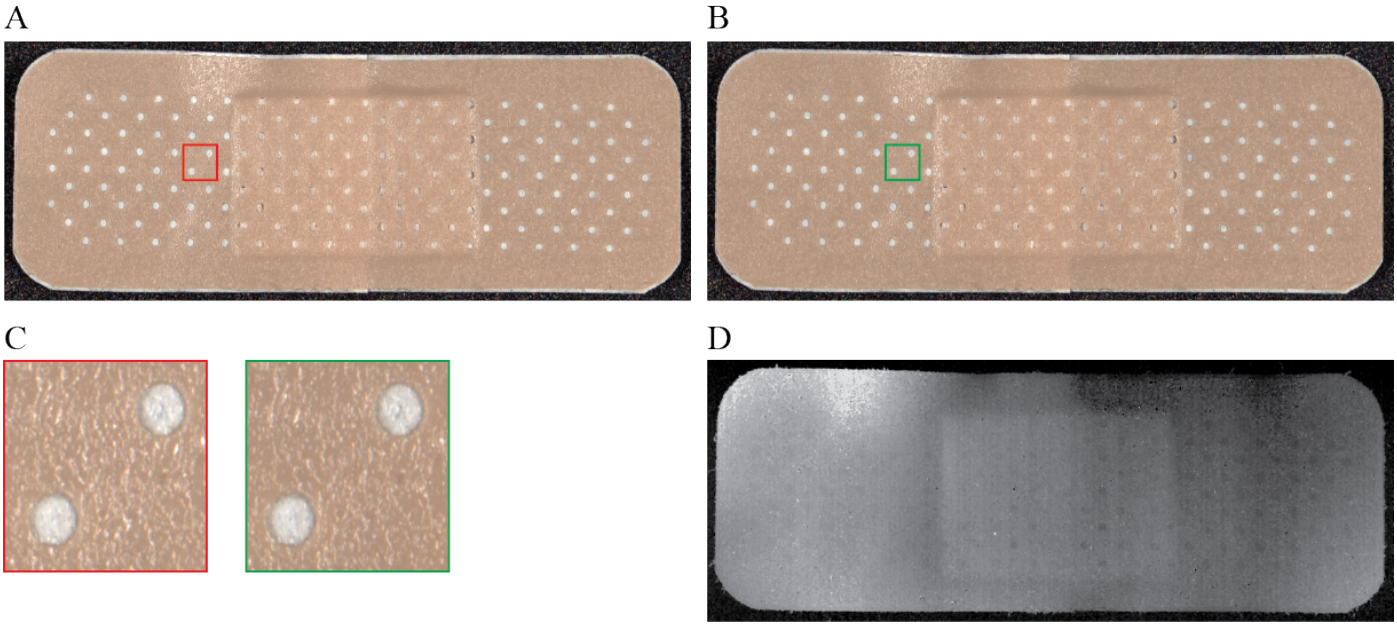


Figure 5.8: Depth measurement of a plaster using the high-resolution setup. (A) shows the left-most view, (B) the right-most view. (C) displays magnified images of the highlighted zones, while (D) shows the final disparity map.

As the material inside the holes is a textile, the experiment was successful. On textiles in general, the algorithm is expected to work very well, as these materials mostly have a very visible macroscopic structure. Another remarkable result was the correct and detailed matching of the surface material. This material (a kind of plastic) has significant specular component and as can be seen in the figure, produces micro-reflections on the surface. Interestingly, these reflections prove beneficial in this case, since their pattern does not change dramatically when changing the viewing angle, despite employing directional illumination. Note that the holes are visible in the depth map despite being only approx. 0.1 mm lower than the rest of the surface. This proves that the depth resolution of the system is highly material-dependent (note on the ground-truth object, 0.1 mm seemed to be the limit for depth resolution).

### 5.2.3 EUR10 Banknote

Another experiment was performed to assess the stability of the algorithm against periodic structures (Fig.5.9). For this purpose, a banknote seems to be a perfect inspection object given the fact that the so-called Guilloché patterns are very

popular in banknote design. As discussed elsewhere, periodic patterns are problematic in stereo depth reconstruction, particularly with large matching windows, although this is also the case for our $3 \times 3$ matching window, as some structures on the banknote are smaller than that. The acquisition was performed using the low-resolution setup. Directional illumination was used together with an aperture of $F = 5.6$. The banknote was folded to simulate a realistic inspection scenario.

Looking at Fig.5.9D and E, it is apparent that the algorithm is very stable in high spatial frequency patterns, but less stable in lower frequencies. While the disparity map in Fig.5.9D doesn't show any significant correlation with the pattern on the banknote, Fig.5.9E does. This is likely caused by the specific spatial frequencies of the pattern. While (D) contains a high-frequency texture, which allows for reliable depth measurement in each pixel, the frequencies in (E) are lower, enabling depth measurement only in areas with notable gradients. In general, sharp gradients increase the probability that the correct slope will be detected. Another improvement could be expected with a larger matching window. By applying the cost aggregation described in Sec.4.3, the depth values in each pixel even out, nevertheless, small differences remain in the depth map. Notice also the increased height of the number "10". This is not an artifact, but rather correct detection of the intaglio print.

### 5.2.4 Group of coins imaged using the slit aperture setup

The last depth measurement was performed using the slit aperture system and a set of coins, placed on top of each other (Fig.5.10). Directional illumination was used for the acquisition and the aperture number was $F = 4$. This acquisition has particularly homogeneous illumination over the whole field-of-view due to the small area that has to be lit. Additionally, the depth-of-field is relatively high, and depth can be estimated on the whole image area, with a height difference of about 10 mm between the lowest and highest point of the scene.

The setup handled the large high differences and the strong reflections of the coins very well, showing noticeable depth variations between the figures on the upper coin and its ground surface. Although there is higher noise on the lower coin, most likely attributed to the blur caused by a lower depth-of-field, the border of the coin as well as individual letters can still be recognized in the depth map. Also, a low amount of noise is found on the stone ground plate.

The comparison between the central view, TDI simulation and all-in-focus image in Fig.5.10C shows that while the all-in-focus image clearly improves the signal-to-noise ratio compared to the central view, it also doesn't introduce the motion blur seen in the TDI image.

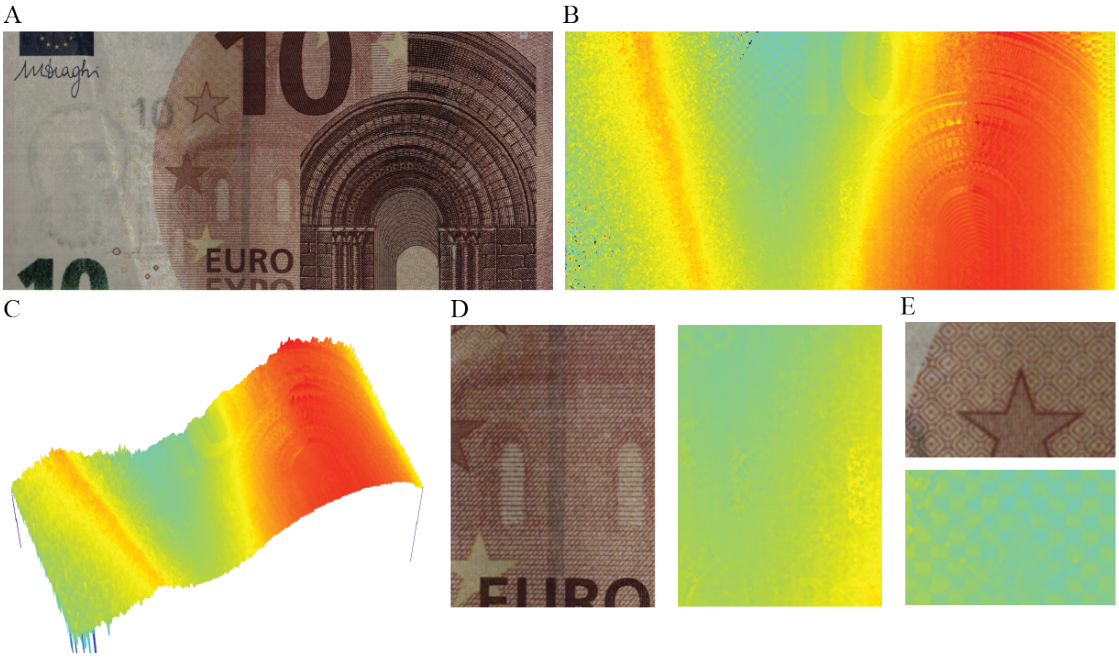


Figure 5.9: Low-resolution setup acquisition of a EUR10-banknote. (A) shows the central view, while (B) shows the color-coded disparity map. In (C), a 3-D model of the disparity map is displayed, which shows the general trend of the folded banknote. (D) and (E) show two different areas of the banknote, magnified.

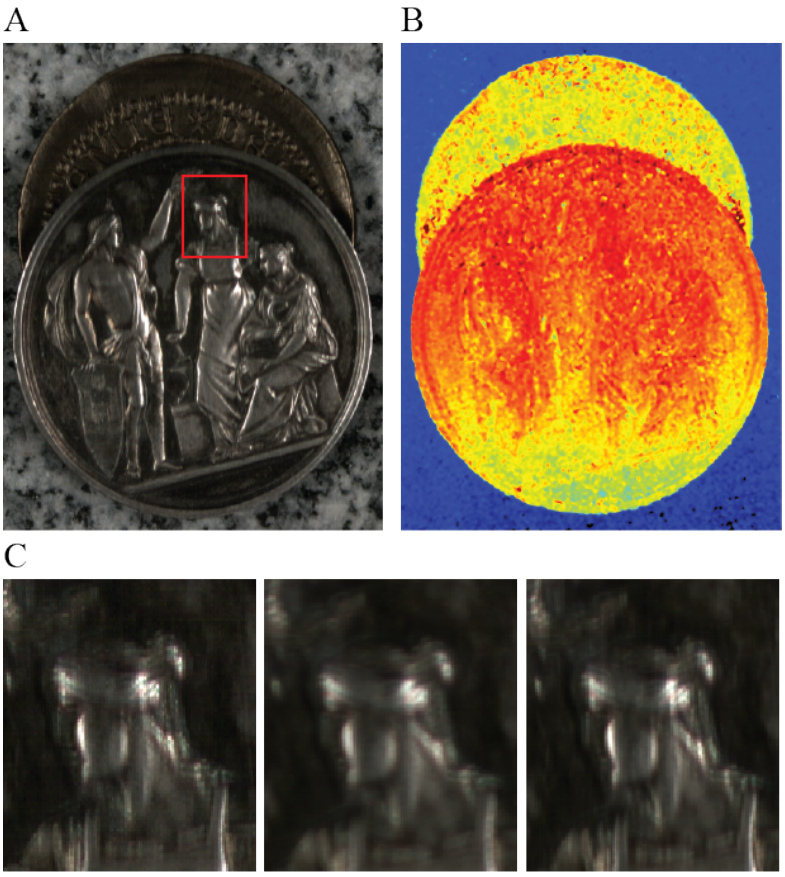


Figure 5.10: Slit-aperture setup acquisition of two coins placed over each other. (A) shows the central view, while (B) shows the obtained disparity map. In (C), a zoomed-in part of the central view is shown (left), along with a simulation of TDI (middle) and the all-in-focus image (right).

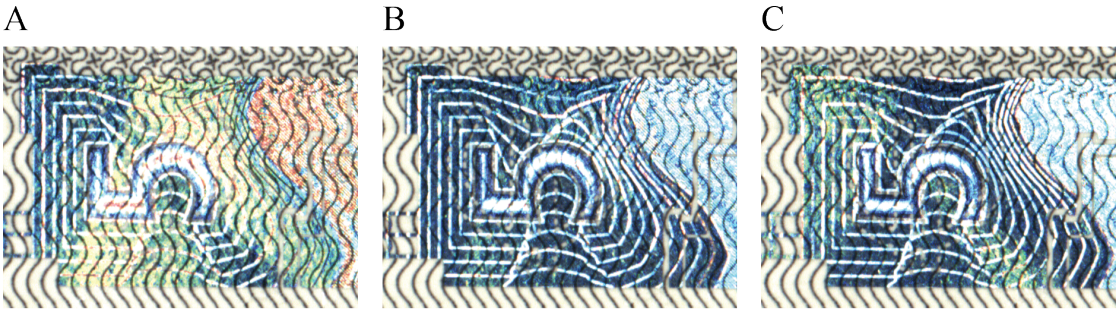


Figure 5.11: Cutout from the diffractive stripe of the new EUR5 banknote, displaying the number 5 with changing coloring. From left to right: left view, central view and right view.

## 5.3 Optical variable devices

Even though 3-dimensional objects are the main focus of line-scan light-field inspection, it is interesting to apply the principle to the inspection of other viewing angle-dependent flat objects, such as *optical variable devices* (OVDs). These become more and more prevalent in the security industry. In this section, two view-dependent objects are presented: an diffractive pattern on the EUR5 banknote (Fig.5.11) and a hologram (Fig.5.12). In each figure, three different views (left, central and right) are shown, highlighting the radical changes in appearance those objects exhibit w.r.t. the different views.

It has to be noted that these changes cannot only be attributed to the change in viewing angle. Instead, changes in viewing angle *together* with the changes in illumination angle have to be taken into account. As the illumination used for the experiments was directional, different viewing angles capture different maxima of the diffraction pattern produced by the hologram. When using diffuse illumination, appearance changes are far less apparent.

Using directional illumination combined with light-fields enables effective OVD acquisition, since the number of views and the respective viewing angles can be adjusted for each OVD to precisely match its diffraction pattern. For the experiments in this section, 12 views were extracted. In Fig.5.11, 80 sensor lines spacing between each view was sufficient to display significant color shift. This corresponds to an angular range of 6.2° (compare Tab.4.2). For Fig.5.12, a line spacing of 120 lines was necessary to reach the first diffraction maximum (angular range of 6.8°).

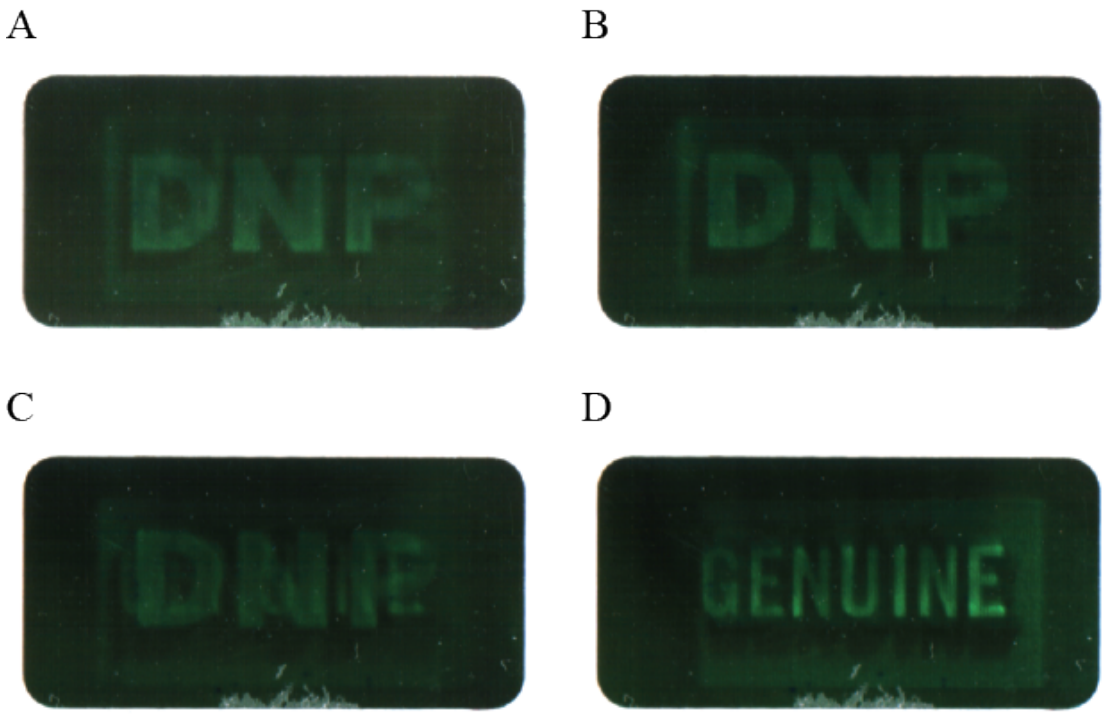


Figure 5.12: Security hologram by Dai Nippon Printing (DNP). Views 1, 4, 8, and 12 are shown. While the change between the first two views is only marginal, a gradual shift between the two 3-dimensional texts appears more apparent in the remaining views.

# 6 Conclusions

In this study, it was shown that light-fields can be successfully applied to industrial inspection for a variety of objects and materials. For this purpose, a line-scan light-field system was presented, which captures the 3-D light-field of an object moving with constant velocity under the camera. As the inspection system only consists of a off-the-shelf industrial-grade area-scan camera together with a suitable optics and a transport means, it is easy to install and operate. Obtaining a light-field of the inspected object instead of a simple image enables surface reconstruction and depth estimation, digital refocusing and depth-of-field extension, together with multi-view capability. For depth estimation, it was shown that suitable depth estimation algorithms exist which can reliably reconstruct reliefs. Census transform (CT) was used as a similarity measure in this work because of its speed and robustness against variations in illumination, although other (and more high-quality) algorithms such as normalized cross-correlation (NCC) could also be applied, if computational power is not a limiter. Additionally, for optical variable devices such as security holograms, it is sufficient to compare multiple views, e.g. to check for correct printing.

A theoretical background on light-fields was presented and used in conjunction with the line-scan light-field system to derive how the light-field is sampled by the system. This rigorous approach has the additional advantage of being usable in wave optics, when the light-field function is switched with its wave-optical equivalent, the Wigner function, or the previously mentioned augmented light-field. In that case, only the description of the light-field itself changes, but not the description of the system itself.

The system was tested using two optical configurations and two illumination configurations. For both optical configurations, the result of depth estimation was tested against a 3-D printed ground truth. This yielded a maximal depth resolution of approx. 100 µm, although a strong dependence on image sharpness and feature size was found. Additionally, multiple real-life test objects were acquired which show the versatility of the system. It was shown that using diffuse axial illumination, accurate depth estimates can be obtained on specular surfaces, such as metal coins. Directional illumination is mostly suitable for OVDs, whose reflection characteristics depend not only on the viewing angle, but also on the illumination angle. For Lambertian surfaces, the choice of illumination is not as crucial, as long as the imaged area is lit homogeneously and with sufficient power.

In addition to the standard configurations, an alternative system was proposed which enhances the angular range, while at the same time reducing the field-of-view (and thus the illuminated area) by inserting a slit aperture between the lens and the object. This configuration was demonstrated on a stack of metal coins and found to be equally capable of producing depth maps, although the lack of light might be a problem if fast acquisition speeds have to be achieved.

An advantage of the line-scan light-field setup compared to other systems used for depth estimation, such as stereo systems, is that light-fields provide more angular information than a stereo pair of images and can therefore reach higher depth resolutions, while reducing the problem of insufficient angular sampling. The angular range is variable and can be chosen just by an electronic re-configuration of the camera.

It was shown here that very good results can be achieved with the system with a simple mechanical adjustment and without elaborated calibration. This can prove to be of benefit in comparison to line-scan stereo systems or camera arrays, which often require complex calibrating procedures to produce precise results.

A disadvantage can be seen in the lower speed of this system compared to stereo due to a larger amount of data to be transferred and processed. This is currently an area of research and can be significantly enhanced by implementing the algorithms on a specific hardware (e.g. *field-programmable gate arrays* (FPGA) or *digital signal processors* (DSP)). Also, even sparser sampling of the light-field is thinkable, with only 4 or 8 lines instead of 12, as was used here. One aspect we did not deal with here is how the quality of the reconstruction relates to the sampling density in angular direction.

Another disadvantage is related to the fact that lenses always feature some kind of optical aberration, especially further away from the optical axis, so that this can worsen the quality of the obtained views. This however can be diminished by applying an appropriate software correction and further by the choice of a high-quality lens which produces acceptable results in the whole angular range. This however requires additional computational power, which might be better spent on the depth estimation, if the issue isn't significant.

Light-fields offer great advantages to a range of different imaging areas, including industrial imaging, although their full potential hasn't been explored as of yet. They provide high amounts of data ready to be processed and used. Therefore it is not unusual to see great interest for such systems in the industry. Personally, I hope that light-field imaging systems become more and more prevalent in every area of imaging and that possibilities related to them will be progressively revealed.

# Bibliography


Adelson, E.H. and J.R. Bergen. "The plenoptic function and the elements of early vision". In: *Computational Models of Visual Processing*. MIT Press, 1991, pp. 3–20.

Adelson, E.H. and J.Y.A. Wang. "Single lens stereo with a plenoptic camera". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 14.2 (Feb. 1992), pp. 99–106.

Bastiaans, M.J. "Wigner distribution function and its application to first-order optics". In: *J. Opt. Soc. Am.* 69.12 (Dec. 1979), pp. 1710–1716.

Bodenstorfer, E. et al. "High-speed line-scan camera with multi-line CMOS color sensor". In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. June 2012, pp. 9–14.

Broxton, M. et al. "Wave optics theory and 3-D deconvolution for the light field microscope". In: *Opt. Express* 21.21 (Oct. 2013), pp. 25418–25439.

Dellaert, F. et al. "Structure from motion without correspondence". In: *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. Vol. 2. 2000, 557–564 vol.2.

Gabor, D. "A new microscopic principle". In: *Nature* 161.4098 (1948), pp. 777–778.

Georgiev, T. et al. "Spatio-angular resolution tradeoff in integral photography". In: *Eurographics Symposium on Rendering*. 2006, pp. 263–272.

Gerrard, A. and J.M. Burch. *Introduction to matrix methods in optics*. Dover Books on Physics. Dover, 1994. ISBN: 9780486680446.

GmbH, Chromasens. *Chromasens 3D-PIXA: High-resolution colour camera for 3D inspection*. `http://www.chromasens.de/en/news/chromasens-3d-pixa-high-resolution-colour-camera-3d-inspection`. [Online; accessed 28-August-2014]. 2014.

GmbH, Raytrix. *Raytrix. 3d light field cameras*. `http://raytrix.de/`. [Online; accessed 25-August-2014]. 2014.

Grosenick, L., T. Anderson, and S.J. Smith. "Elastic source selection for in vivo imaging of neuronal ensembles". In: *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*. June 2009, pp. 1263–1266.

Holländer, B., S. Štolc, and R. Huber-Mörk. "Multi-view line-scan inspection system using planar mirrors". In: vol. 8791. 2013,

Ives, H.E. "Parallax panoramagrams made with a large diameter lens". In: *J. Opt. Soc. Am.* 20.6 (June 1930), pp. 332–340.

Jaumann, R. et al. “The high-resolution stereo camera (HRSC) experiment on Mars Express: Instrument aspects and experiment conduct from interplanetary cruise through the nominal mission”. In: *Planetary and Space Science* 55 (May 2007), pp. 928–952.

Levoy, M. and P. Hanrahan. “Light field rendering”. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH ’96. New York, NY, USA: ACM, 1996, pp. 31–42. ISBN: 0-89791-746-4.

Levoy, M. et al. “Light field microscopy”. In: *ACM Trans. Graph.* 25.3 (July 2006), pp. 924–934. ISSN: 0730-0301.

Lippmann, G. “Epreuves reversibles. Photographies integrals”. In: *Comptes-Rendus Academie des Sciences* 146 (1908), pp. 446–451.

Lukáč, R. *Single-Sensor Imaging: Methods and Applications for Digital Cameras*. Image Processing Series. Taylor & Francis, 2008. ISBN: 9781420054538.

Lukáč, R. and K.N. Plataniotis. “Color filter arrays: design and performance analysis”. In: *Consumer Electronics, IEEE Transactions on* 51.4 (Nov. 2005), pp. 1260–1267. ISSN: 0098-3063.

Lumsdaine, A. and T. Georgiev. “The focused plenoptic camera”. In: *Proc. IEEE ICCP*. 2009, pp. 1–8.

Malvar, H.S., L.-W. He, and R. Cutler. “High-quality linear interpolation for demosaicing of bayer-patterned color images”. In: *International Conference of Acoustic, Speech and Signal Processing*. Institute of Electrical and Electronics Engineers, Inc., May 2004.

Milgram, J.H. and W. Li. “Computational reconstruction of images from holograms”. In: *Appl. Opt.* 41.5 (Feb. 2002), pp. 853–864.

Nalpantidis, L., G. Ch. Sirakoulis, and A. Gasteratos. “Review of stereo matching algorithms for 3D vision”. In: *16th International Symposium on Measurement and Control in Robotics*. 2007, pp. 116–124.

Ng, R. et al. *Light field photography with a hand-held plenoptic camera*. Tech. rep. Apr. 2005.

Oh, S.B. et al. “Rendering wave effects with augmented light field”. In: *Computer Graphics Forum* 29.2 (2010), pp. 507–516. ISSN: 1467-8659.

Petrie, G. “Airborne push broom line scan: An alternative to digital frame cameras”. In: *GEOInformatics* 8 (2005), pp. 50–57.

Platt, B. C. and R. Shack. “History and principles of Shack-Hartmann wavefront sensing.” In: *Journal of Refractive Surgery* 17.5 (Oct. 2001). ISSN: 1081-597X.

Prevedel, R. et al. “Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy”. In: *Nat. Meth.* 11.7 (July 18, 2014), pp. 727–730. ISSN: 1548-7091.

Ribbens, W.B. “Interferometric surface roughness measurement”. In: *Appl. Opt.* 8.11 (Nov. 1969), pp. 2173–2176.

Scharstein, D. and R. Szeliski. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms". In: *Int. J. Comput. Vision* 47.1-3 (Apr. 2002), pp. 7–42. ISSN: 0920-5691.

Sokolov, A.P. *Autostereoscopy and integral photography by Professor Lippmann's method*. Moscow State Univ. Press, 1911.

Spring, K.R. and M.W. Davidson. *Geometrical Distortion*. `http://www.microscopyu.com/tutorials/java/aberrations/distortion`. [Online; accessed 30-August-2014]. 2014.

Štolc, S. et al. "Depth and all-in-focus images obtained by multi-line-scan light-field approach". In: *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics. 2014, pp. 902407–902407.

Tchvialeva, L. et al. "Surface roughness measurement by speckle contrast under the illumination of light with arbitrary spectral profile". In: *Optics and Lasers in Engineering* 48.7–8 (2010), pp. 774–778. ISSN: 0143-8166.

Tippetts, B. et al. "Review of stereo vision algorithms and their suitability for resource-limited systems". In: *Journal of Real-Time Image Processing* (2013), pp. 1–21. ISSN: 1861-8200.

Vilums, I.J. *Optical imaging system using lenticular tone-plate elements*. US Patent 4,878,735. 1989.

Walree, P. van. *Distortion*. `http://toothwalker.org/optics/distortion.html`. [Online; accessed 30-August-2014]. 2014.

Wang, Z. and Z. Liang. "Sphere light field rendering". In: vol. 4681. 2002, pp. 357–365.

Wheatstone, C. "Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision". In: *Philosophical Transactions of the Royal Society of London* 128 (1838), pp. 371–394.

Wilburn, B. et al. "High performance imaging using large camera arrays". In: *ACM Trans. Graph.* 24.3 (July 2005), pp. 765–776. ISSN: 0730-0301.

Wong, H.-S., Y.L. Yao, and E.S. Schlig. "TDI charge-coupled devices: Design and applications". In: *IBM J. Res. Dev.* 36.1 (Jan. 1992), pp. 83–105. ISSN: 0018-8646.

Wyant, J.C. "White light interferometry". In: vol. 4737. 2002, pp. 98–107.

Zabih, Ramin and John Woodfill. "Non-parametric local transforms for computing visual correspondence". In: *Proceedings of the Third European Conference on Computer Vision (Vol. II)*. ECCV '94. Stockholm, Sweden: Springer-Verlag New York, Inc., 1994, pp. 151–158. ISBN: 0-387-57957-5.

Zhang, Z. and M. Levoy. "Wigner distributions and how they relate to the light field". In: *IEEE International Conference on Computational Photography (ICCP)*. 2009.

# Acknowledgements

I feel obligated to express my gratitude to all the people I have been working with on the ideas of this thesis, and I would like to thank them for their support. First and foremost, I would like to offer my special thanks to Dr. Svorad Štolc and Prof. Gerhard Schütz, who both guided me throughout the whole thesis with their insightful comments and great ideas, who encouraged me and who made this thesis possible in the first place. I would also like to thank the entire staff of the High-Performance Image Processing business unit of the Austrian Institute of Technology for enabling me to do my research at their facility and who were always helpful and supportive.

Finally, I would like to thank my parents for their constant support and encouragement throughout my study. I dedicate this thesis to them.