Depth of Field Reduction for Miniature-Camera Imaging

by

Timothée E. Cognard

under the supervision of

Nicholas Devaney (Academic Supervisor)
Chris Dainty (Industrial Supervisor)

A dissertation submitted to the National University of Ireland in fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D.)

Applied Optics Group, School of Physics,
National University of Ireland, Galway

August 2020

0.1 Declaration

The work in this thesis is based on research carried out at the Applied Optics Group, School of Physics, National University of Ireland, Galway. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless otherwise referenced in the text.

Dearbhaím gurb é an taighde go hiomlán atá san ábhar seo atá mé a thíolacadh le haghaidh measúnachta déanta san Grúpa Optaice Feidhmí, Scoil na Fisice, Ollscoil na hÉireann, Gaillimh. Dearbhaím nár baineadh aon chuid de as saothar aon duine eile ach amháin sa mhéid atá luaite agus aitheanta sa téacs.

Timothée Cognard

0.2 Abstract

This thesis investigates the reduction of depth of field for miniature camera systems, such as the ones embedded in smartphones. The work on digital shallow depth of field is already mature and implemented in most modern phones in the form of image processing algorithms. Yet, little research has been performed to provide optical design solutions for depth-of-field management. This thesis proposes a theoretical framework for the understanding of depth of field in the particular case of miniature cameras and two practical solutions to reduce the depth of field. The first one is an apodization phase-mask designed in order to achieve depth super-resolution with about 20% lobe width reduction. The second one is an f/1.0 and an f/1.4 lens. Both solutions are presented and compared. Finally, some preliminary work is presented on how image processing could further improve the hardware contribution of this thesis.

0.3 Dissemination of PhD Research

All contributions carried out by the author in whole or in part over the duration of this Ph.D. study are listed below.

- Cognard, T.E., Devaney, N. and Dainty, C., 2017, June. A Particle Swarm Optimization approach of the Axial Apodization problem. In Applied Industrial Optics: Spectroscopy, Imaging and Metrology (pp. JTu5A-24). Optical Society of America.
- Cognard, T.E., Dainty, C. and Devaney, N., 2018. Estimating axial resolution with diffraction theory. Applied optics, 57(22), pp.E138-E141.
- Ungureanu, A.S., Thavalengal, S., Cognard, T.E., Costache, C. and Corcoran, P., 2017. Unconstrained palmprint as a smartphone biometric. IEEE Transactions on Consumer Electronics, 63(3), pp.334-342.
- Cognard, T.E., Goncharov, A., Devaney, N., Dainty, C. and Corcoran, P., 2018, August. A Review of Resolution Losses for AR/VR Foveated Imaging Applications. In 2018 IEEE Games, Entertainment, Media Conference (GEM) (pp. 1-9). IEEE.

Au vrai, toute création de l'esprit est d'abord *poétique* au sens propre du mot [...]. Et de cette nuit originelle où tâtonnent deux aveugles-nés, l'un équipé de l'outillage scientifique, l'autre assisté des seules fulgurations de l'intuition, qui donc plus tôt remonte, et plus chargé de brève phosphorescence? La réponse n'importe. Le mystère est commun.

Saint-John Perse, Discours au banquet Nobel (1960)

In truth, every creation of the mind is first of all *poetic* in the proper sense of the word [...]. And from that primal night in which two men born blind grope for their ways, the one equipped with the tools of science, the other helped only by the flashes of his imagination, which one returns sooner and more heavily laden with a brief phosphorescence? The answer does not matter. The mystery is common to both.

Saint-John Perse, Speech at the Nobel Banquet (1960)

0.4 Acknowlegements

I would like to thank all the people that contributed to this thesis. All of your contributions to this work have been greatly appreciated.

Colleagues and work

I had the chance to do my research work under the supervision of Nicholas Devaney and Chris Dainty, although I have also received some help from Alexander Goncharov and Peter Corcoran. Of course, I would like to thank all of them for the quality of their work and their help to my research. I would like to thank them for their inspiring conversations on my research projects (technical details, conferences work or intellectual property) and on seemingly unrelated topics (Mamiya lenses, Irish traditions, classical music...). I thank also Claudia Costache and Joe Desbonet for their administrative help.

I would like to acknowledge Science Fundation Ireland (SFI) for the generous funding for my PhD (Project ID 13/SPP/12868). The other SFI and IRC students with whom I did my research also deserve some praise. With no particular order : Asma, Adrian, Shejin, Viktoras, Tudor, Shabab, Hossein, Joe, Aoefa, Faisal... Some of them have already submitted their manuscript, but they have all been great, friendly people to share ideas with.

I also would like to acknowledge XPERI for co-funding my research and providing me with some facilities (computers, darkroom...). I am very grateful to Alex, Cosmin, Petronel, Sharon, Dan, Pat, Brian, Diana, Alin and many others.

As for the University (NUIG), I would like to thank everyone in the Applied Optics group, as well as the staff in charge of the demonstrating labs. I have a total of four Conor to thank in the University as well as Michelle, Fiona, Niamh, Colm, Aron, Laetitia, Ciarrai. Thank you all for the wonderful time !

Family and friends

Special thanks goes to my family. Their support and help, even through the distance of seas and oceans, has been of great importance for me.

I would like to thank a couple of important people for me : Eliza who has been constantly supportive through the writing of this manuscript and always sending me kind thoughts in the difficult times and Adrian-Stefan with whom I had the chance to discuss about anything from philosophy

to geopolitics, including of course, the contribution of the Dacians to history.

Secondly, all of my close friends. With no particular order : Eric, Alexis, Adrian, Sébastien, Charles, Pierre, Elias, Diana, Alin, Bjorn, Alessandro, Conor, Léna, Victoria, Finghin, Anna, Vincent, Patrizio, Claudia, Ciara... This list should be extended to all the people I have met in Ireland that made my stay enjoyable.

Inspirations and models

Last but not least, I have a kind thought for all the people that pushed me in my journey to learn more about physics : my former teachers who gave me some skills and knowledge and the many inspirational physicists and photographers that gave me some interest for the field of Optics such as Augustin Fresnel, René Descartes, Pierre Angénieux, Isaac Newton, Nicolas Tesla, Irving Penn, Brassai and Conrad L. Hall...

0.5 List of Symbols

Unless stated otherwise in the text, these symbols are used with the proposed units.

Light parameters

A: Illumination amplitude

I: Light intensity (cd)

λ: wavelength of the signal (nm)

Diffraction parameters

J_1: Bessel function

ϕ_{Airy}: Airy disk diameter (m)

S_{Airy}: Surface area of the Airy disk (m^2)

Lens parameters

r: radius of the lens pupil or radius of the pinhole (m).

C_{oc}: Circle of confusion (mm)

f: focal length (mm)

N: Numerical aperture

H: Hyperfocal distance (m)

Sensor parameters

H_{sensor}: Height of the medium used to capture the image (m)

W_{sensor}: Width of the medium used to capture the image (m)

d_{sensor}: Distance from which the medium is captured at (m)

D_{sensor}: Diagonal of the sensor (mm)

R_{sensor}: Resolution of the sensor (MPx)

S_{sensor}: Area of the sensor (mm^2)

Screen parameters

H_{screen}: Height of the screen used to observe the image with an emetropic eye (m)

W_{screen}: Width of the screen used to observe the image with an emetropic eye (m)

d_{screen}: Distance from the screen to the eye (m)

R_{screen}: Resolution of the screen (unitless)

Planes of focus

P: plane of focus

ΔP: distance between the first and last plane in focus (m)

P_P: *Punctum proximum*, i.e. first plane of focus (m)

P_R: *Punctum remotum*, i.e. last plane of focus (m)

N_{planes}: Number of planes of focus

$R_{\text{diffraction}}$: Resolution as limited by the diffraction (unitless)

Noise Source parameters

P_N: Probability of photon arrival following a Poisson distribution

N_{photon}: number of photons

σ_N^2: variance of the noise

R_e: Electric Resistance (Ω)

T: Temperature (K)

Δf: Frequency Bandwidth (Hz)

k_B: Boltzmann Constant, $k_B = 1.38 \times 10^{-23} \mathrm{m}^2.\mathrm{kg}.\mathrm{s}^{-2}.\mathrm{K}^{-1}$

J_{surf}: Thermal generation at surface state

J_{gen}: Thermal generation in the bulk silicon

J_{depl}: Thermal generation in the depletion region

q: Charge constant $q = 1.60 \times 10^{19} \mathrm{C}$

S_0: Surface area of a silicon component (m^2)

e^-: Electron symbol

SNR: Signal to Noise Ratio

0.6 Glossary

ALEC: Automatic Lens Correction, lens design software.

APS-C: Sensor Size (25.1×16.7 mm). A smaller APS size is used by Canon and is known as APS-H.

B

BFL: Back Focal Length.

BW: Black and White image or sensor.

C

CCD: Charge-Coupled Device, type of photosensitive device.

CMOS: Complementary Metal-Oxide Semiconductor, type of photosensitive device.

CNN: Convolutional Neural Network.

CRA: Chief Ray Angle.

D

DOF : Depth of Field.

DSLR: Digital Single Lens Reflex.

E

eDoF: Extended Depth of Field.

EPI: Epipolar-Plane Image.

Epipolar geometry: Epipolar geometry is a mathematical and geometrical model providing the geometrical relations between several images of the object scene.

F

FFT: Fast Fourier Transform.

Free-Form Surface: A free-form surface is an optical surface that has no translational or rotational symmetry about axes normal to the mean plane, according to the ISO standard 17450-1:2011.

Full Frame: Sensor size of 24×36 mm.

FWC: Full Well Capacity, maximal number of electrons that a given pixel can hold.

G

GA: Genetic Algorithm, optimization algorithm.

GRIN: Gradient of Refractive Index material.

J

JPEG: Joint Photographic Experts Group

L

LCoS: Liquid Crystal on Silicon.

LEAD: Lens Evaluation And Design, lens design software.

M

Metasurface: Optical metasurfaces are layers of sub-wavelength patterns (typically 10μm) used to scatter light and change the optical wavefront behaviour.

MPx: Mega-pixel, unit for a million pixel image (e.g. 1000px $\times$1000px array).

MTF: Modulation Transfer Function.

MOSFET: Metal Oxide Semiconductor Field-Effect Transistor.

N

NA: Numerical Aperture.

O

Optical Aberration theory : Optical Aberration theory provides mathematical estimates of how optical surfaces can introduce local perturbations to the optical wavefront. Two polynomial sets are used to describe the aberration function, Seidel aberration theory and Zernike aberration theory. The former describes the primary aberrations in terms of ray heights and angles within the paraxial region. The later is a set of orthonormal polynomials.

ORDEAL: Optical Routines for Design and Analysis, lens design software.

OTF: Optical Transfer Function.

P

PDAF: Phase-Detection Autofocus.

Petzval sum: The Petzval sum is a mathematical tool used to measure the field of curvature of a lens system. Each surface of the optical design brings a certain Petzval curvature contribution. The sum of these contributions should ideally be null for a flat sensor.

PSF: Point Spread Function.

PSO: Particle Swarm Optimization.

R

RAM: Random Access Memory.

RAW: Raw information file from the sensor data.

RGB: Red-Green-Blue Bayer Pattern.

RMS: Root Mean Square.

S

SAD: Sum of Absolute Differences, algorithm.

SIFT: Scale-Invariant Feature Transform, algorithm.

SLAMS: Successive Linear Approximation by Maximum Steps, lens design software.

SLR: Single-Lens Reflex.

SNR: Signal to Noise Ratio.

Strehl ratio: The Strehl ratio measures the quality of the resolution of an optical system. The metric varies between 0 and 1 where 1 is the behaviour expected from a diffraction limited lens system. The Maréchal criterion suggests that for an imaging system, the Strehl ratio should be above 0.8.

T

ToF: Time of Flight.

TTL: Total Track Length.

Else...

24×36: 24mm $\times$36mm sensor.

1/3 inch sensor: 4.8 mm $\times$3.6mm sensor.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Foreword

Depth of field is the distance between the closest and farthest object planes that appears acceptably sharp on the image. The notion of acceptably sharp is defined quantitatively in the body of the thesis. The purpose of this thesis is to present the research work done to manage the depth of field. In particular, and this is the main object of this research, we seek here to reduce the depth of field.

The choice of this subject may seem quite surprising. However, depth of field is one of the only visible parameters that distinguishes the image produced by a professional (large sensor) camera from a cheap compact camera (small sensor). Of course, splitting imaging systems into two categories of sensors (small sensors and large sensors) is artificial. Routhly, one could consider all sensors smaller than Super 16mm format (i.e. 7.4×12.5mm) to be small-sensors, and all larger sensors to be large-sensors. Of course the laws of Physics remain similar for both small and large sensors. What does change, however, are the influence of diffraction, aberration theory and defocus blur in the final image.

In particular, despite advances in image processing related to smartphones, the perception of distance through depth of field is a priority to enable mobile imagery to compete with high-end systems.

This thesis sums up the work done during a Ph.D. that started in October 2015. The smartphone industry (and high-tech industry in general) is fast paced and has evolved in these few years. Two main changes influenced this work.

Firstly, multi-cameras became mainstream. In 2012, following some interests for 3D-imaging, some manufacturers implemented dual-camera set-ups. These systems were only found on some high-end systems and had quite limited uses. However, since 2018, dual-camera systems are a given for

most customers. They can provide different fields of view, stereo-vision capabilities or even better night-time performance. Some manufacturers have added more than 2 camera modules and the trend seems to be for 3-camera or more. Massively multi-camera projects have emerged, like the Light L16.

Secondly, image processing has been revolutionized by the use of machine learning techniques. Tasks, like depth estimation, which were performed solely through filtering methods are now optimized with machine-learning techniques. Research on depth estimation from monocular views are advancing rapidly. However, there is still a need for hardware-based systems for depth estimation.

1.2 Structure of the thesis

The organization of this thesis is as follows:

Chapter 2 provides some historical and general background about imaging systems, optical design, miniature cameras and sensor noise.

Chapter 3 provides some theoretical elements about the notion of a perfect camera in theoretical and practical terms (resolution, dynamic range, colour and depth).

Chapter 4 provides some theoretical work about the depth of field. The notion of circle of confusion is revisited, as well as the definition of the hyperfocal distance and depth of field. Some new formulas are proposed to establish the optimal planes of focus for a given configuration. This chapter also shows that the state-of-the-art geometrical-optics approximation used to estimate the error for depth is inaccurate for miniature cameras due to diffraction effect.

Chapter 5 reviews the state of the art methods for estimating depth. These include stereo-imaging, Phase-Difference Auto-Focus (PDAF) pixels, micro-lens arrays, machine learning methods and coded apertures.

Chapter 6 proposes some new work on the design of an apodization filter to reduce the depth of field. The proposed filter is a phase mask generated using Particle Swarm Optimization algorithm. The results are compared with the ones from state-of-the-art Genetic Algorithms. The performance obtained suggests that apodization can have a positive influence on depth-of-field reduction, yet brings many challenges for implementation (e.g. chromatic aberrations).

Chapter 7 offers some new designs for wide aperture lenses. Following the theoretical work done in Chapter 4, the lens design is optimized for depth performance. Two systems are proposed (with an f/1.0 and f/1.4 f-numbers respectively). The tolerancing of the lenses and its properties are discussed.

Chapter 8 provides some comments on image processing and the possible fusion of images from cameras having different apertures. The section lists some of the applications for the lens designed in Chapter 7 and offers a conclusion for the work done through the thesis.

1.3 Novel work in this thesis

The main contributions of the thesis are:

- A new theoretical framework for depth estimation. The film-era framework using print-based circle of confusions is no longer relevant for current use.
- Some new theoretical findings about the optimal planes of focus and depth resolution.
- A novel formula for the axial resolution, based on diffraction theory. This formula is compared with geometric optics formula.
- Some work on axial apodization to establish a phase mask using genetic algorithms.
- The design of a new fast lens for depth estimation with details about the tolerancing.
- Some work on the image processing required for wide aperture aberrated lenses.

All of the contributions are explained in greater detail throughout the thesis and in the conclusion. chapter.

Chapter 2

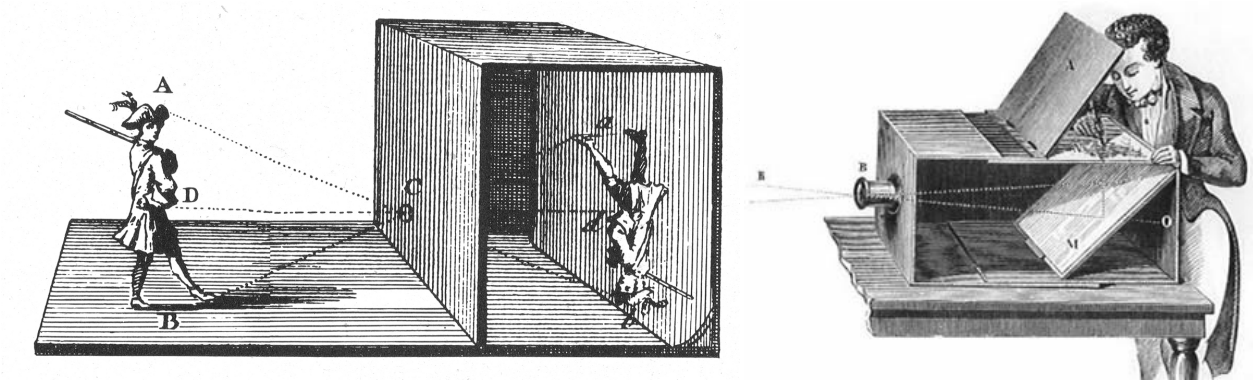
Overview and history

2.1 A brief history of camera systems

The history of photography is almost two centuries long, and numerous systems have been proposed [1–3]. The point of this section is to give a quick overview of the evolution of photography.

2.1.1 Camera Obscura

The origins of cameras are connected to the interest in better paintings. The first camera system, known as the *camera obscura*, was intended for artists as a tool for the faithful reproduction of landscapes. The simplest version of the *camera obscura* was using a simple hole in a dark room to project the inverted image on a wall. This was improved later into a portable wooden box with a mirror. Figures 2.1a and 2.1b present both systems.



(a) Original camera obscura (b) Portable camera obscura

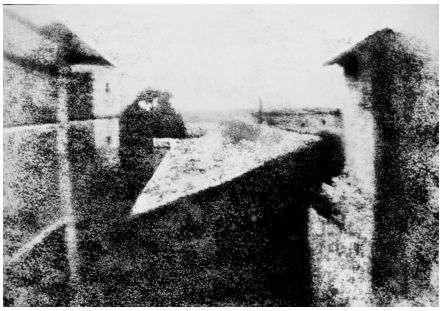
Figure 2.1: Principle of the camera obscura, source - Wikimedia Commons (Public Domain)

2.1.2 The invention of Photography

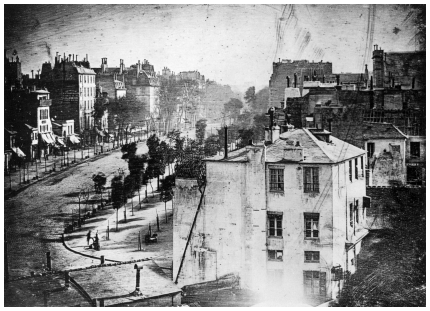
The invention of photography is attributed to the Frenchman Joseph Nicéphore Niépce who, in the early XIXth century, fixed an image onto paper using a sheet of bitumen-coated pewter. The

image shown in figure 2.2a is known as the oldest photograph and is dated from 1827. It is worth mentioning that this wasn't the first attempt of Niépce to obtain a photograph. His correspondence mentions some attempts as early as 1816 using some silver chloride coated paper. However, Niépce did not manage to prevent the obtained image from darkening further as it was exposed to the light for viewing.
Louis Daguerre improved the process and developed the popular daguerreotype device in 1839, together with Nicéphore Niépce. The Daguerre process used a silver plate (commonly a silver coated copper plate) in order to image the scene. The plate was a polished mirror as any surface defect could deteriorate the final image. The plate was then exposed to fumes prior to taking a picture. The original process used only iodine crystal fumes, but diverse halogens, such as bromine and chlorine, have been used to improve light sensitivity. The development of the plate was made using some mercury fumes. A last step consisted in fixing the image using a solution of sodium thiosulphate which removed any remaining silver halides. Hippolyte Fizeau added an extra bath using gold chloride that helped to stabilize the image and brought some warmer tones to the image.

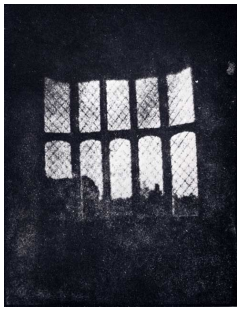
At about the same time, Henry Fox Talbot introduced the first paper-based negative film known as the calotype negative (see picture 2.2c). Patented in 1841, his negative film was made of "salted paper" (paper bathed in a solution of water and sodium chloride) coated with some silver nitrate. The negative film obtained could be used several times to produce positive prints whereas Daguerrotypes had to be rephotographed.



(a) View from the Window at Le Gras by Joesph Nicéphore



(b) Boulevard du Temple by Louis Daguerre

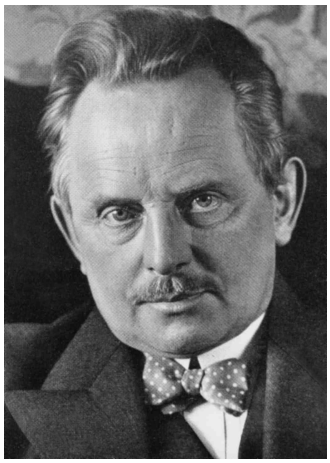


(c) Latticed window at Lacock Abbey by William Fox Talbot

Figure 2.2: First Photographs, source - Wikimedia Commons (Public Domain)

2.1.3 Film Photography

The first negative films were commercialized in 1889 by G. Eastman, the founder of Kodak. The still-used 120 film for medium format was introduced in 1901 with a standard 56mm height. Medium format comprises all film sizes between 24×36mm (full frame) and 100×130mm (large format). Oscar Barnack shown in figure 2.3a, an engineer at Leica, introduced in 1914 the idea of the 24×36mm film format that would be easier to carry for photographers.



(a) Portrait of Oskar Barnack in 1936, source : Oskar Barnack (Public Domain)

(b) Leica Model I (1927), source : Kameraprojekt Graz 2015 (Wikimedia Commons, CC BY-SA 4.0)

Figure 2.3: Leica's engineer Oskar Barnack introduced the 35mm film format, enabling new types of pictures (street photography, sports, news images...).

In the 1970s, the Japanese industry took the lead in the 35mm camera market with Canon, Nikon, Pentax, Minolta (owned by Sony since the early 2000s), Fujifilm and Olympus becoming increasingly popular. Japanese companies were also very popular for medium format cameras with companies such as Mamiya, Bronica, Pentax, Fuji and Kowa. These companies are still leading the DSLR and hybrid camera markets.

2.1.4 Digital Photography

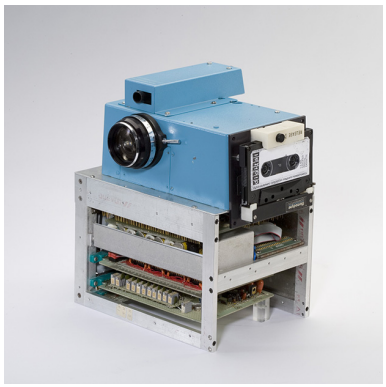
After the invention of the Charge-Coupled Device (CCD) sensor in 1969 in Bell Labs, it was first adopted for scientific (e.g. astronomy) purposes. The initial limitations of digital sensors (cost, weight of the system, low resolution, readout noise and dark noise) progressively disappeared as the technology matured.

2.1.4.1 The prototype era

The first working prototype of a portable digital camera is attributed to Steven Sasson (one of Kodak's engineers) in 1975. The prototype, shown in figure 2.4a, offered a Super-8 sized CCD-sensor with a resolution of 100×100 pixels. The system weighted 3.6kg and required 16 AA batteries. The image was recorded on a cassette tape, which took about 23 seconds. The image could later be displayed on a television set from the camera. Steven Sasson made some public demonstrations of his prototype in 1976. He described the audience as being both curious and sceptical. Looking at photographs on a television system, in particular, was criticized [4].

The first digital camera announced to consumers was the Sony Mavica (Magnetic Video Camera, see figure 2.4b). Despite the claim of M. Morita in 1981 that the camera would be available within two years, Sony did not mass produce the camera until the late 1980s and early 1990s. The 1981 prototype weights 800g, was powered with 3 AA batteries and had the size of a modern DSLR camera. A 9×12mm CCD sensor with a sensitivity of 200 ISO was used. The images (490×570 pixels) were recorded on a small size floppy disk [5].

In the early 1980s, most photography brands started researching digital cameras. Several prototypes are known such as the Sharp ESVC in 1982 or the Pentax Nexa in 1983. The Olympics of Los Angeles in 1984 enabled Canon to test a prototype on the field; the D413 [6]. The colour images with a 0.4MPx could be sent through a phone line in 24 minutes for colour or 6 minutes for black-and-white.



(a) First prototype by Steven Sasson from Kodak - 1975, source : Brett Jordan (CC 2.0)

(b) First prototype by Sony, known as the Mavica - 1981, source : Morio (CC by SA 3.0)

Figure 2.4: First digital cameras.

2.1.4.2 The transition era

The Joint Picture Expert Group (JPEG) format was specified in 1991 and adopted by 1992. Despite being sub-optimal from a compression standpoint [7, 8], it is still the dominant image format.

In 1994, Apple released the Apple QuickTake 100. Despite not being the first commercially available digital camera, it sold better than the previous systems. It was updated to the Quicktake 150 and later the Quicktake 200 in 1997.

Several companies partnered in the 1990s to combine the electronics and optics skills into fully working systems. Kodak licensed Nikon bodies (DCS cameras) to develop Kodak-branded cameras

such as the Nikon F3-based Kodak DCS 100 in 1990. The DCS cameras were designed for high resolution applications. The DCS 460 offered 6.2MPx as early as 1995 despite having a 1.3 crop factor. The E2 and E3 provided 1.3MPx images using a 2/3 inch CCD sensor. Because of the large crop factor, Nikon implemented some relay-lens system to adapt the full frame lenses. The Nikon D1, released in 1999 and using an APS-C sensor is often regarded as the first modern DSLR despite its modest 2.7MPx sensor.

2.1.5 Smartphone Photography

First introduced in a Sharp phone in 2000, mobile phones with an embedded digital camera rapidly became popular with the introduction of the iPhone in 2007. With the increasing interest in social media and image sharing, the embedded cameras account for a large fraction of the photographs taken everyday.

In 2017, over a trillion photographs have been taken using smartphones [9]. Smartphone penetration is over 80% of phones in most developing or developed countries according to Deloitte.

	Developing countries	Developed countries
Smartphone ownership among the population	80%	82%
Checking phone at night	40%	70%
Access phone over 50 times a day	17%	29%
Watches Virtual Reality (VR) content on their phone	22%	50%
Claim to not share personal information online	32%	19%

Table 2.1: Some general statistics about smartphone usage

Dual-camera systems have been increasingly used to improve image quality and versatility since 2014. The first triple cameras were commercially available in 2017. Some systems use four or more camera modules. Multiple camera modules can be used for very different purposes. One can group them into several function categories :

- Depth estimation : Historically, the first use of dual camera systems was related to the stereo-vision principle. Assuming two cameras with the same field of view, one can obtain a depth map from the disparity map (map of distances between two images of a given object point in the scene). Ideally, the resolution for both cameras should be identical; however some smartphones can be found with typically a combination of a 12MPx sensor and a 2MPx sensor.

- Spectrum : Multiple sensors can be used to increase the spectral sensitivity. In most cases the additional module is only used as a monochrome camera to gather about three times

more light compared to an RGB sensor, as the lack of colour filters reduces light absorption. In some cases, the additional module can provide near-infrared sensitivity. In recent times, the additional monochrome camera sensor is being replaced with a larger sized colour sensor.

- Field of view : Since 2018, the most common use for multiple camera modules is to offer several focal lengths. The most common combination for a two lens system is a standard wide angle (equivalent $20-24$mm for a 24×36mm format) and a short telephoto lens (equivalent $40-55$mm for a 24×36mm format). Some phones with 3 focal lengths are commercially available (see Appendix B): an ultra-wide angle, a standard wide angle and a short telephoto lens.

2.2 A brief history of optical systems for cameras

Optics plays an important role in the imaging process, and it is thus of interest to give an overview of the optical design dedicated to camera systems [10].

2.2.1 From pinhole to lens system

In the case of a *camera obscura*, a pinhole is sufficient to produce images with an acceptable contrast and resolution level, although with very little brightness. The pinhole lens uses geometric optics to make an image using a small diameter hole. It thus behaves like a lens with a small aperture. The diameter of the pinhole is a compromise between geometrical optics (the smaller the diameter, the better) and diffraction theory (the larger the diameter, the better). The optimal trade-off between geometrical optics and diffraction was first estimated by Joseph Petzval [11] as radius $r=\sqrt{\frac{\lambda f}{2}}$, with λ the wavelength of the light (in the case of visible light, one can use the central green light to a good approximation) and f the distance from the pinhole to the imaging plane. Although elegant, the solution was partly flawed and underestimates the optimal size of the pinhole. Eugen Lommel [12] proposed a better description of the problem using wave theory. His method was reused by Lord Rayleigh [13] who concluded that the ideal radius is around $r=\sqrt{\lambda f}$ [14].

This type of lens however sacrifices most of the light available; the exposure time was already very long for most portrait applications; Barger [15] mentions a typical 3 to 15 minute exposure time. It is worth noting that this is already an improvement over the eight hours needed to record the photograph "View from the Window at Le Gras" [16].

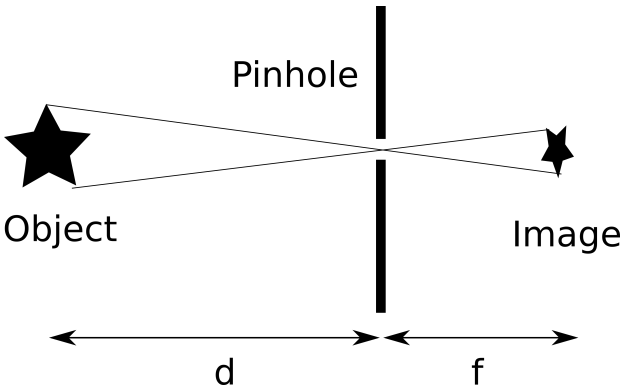
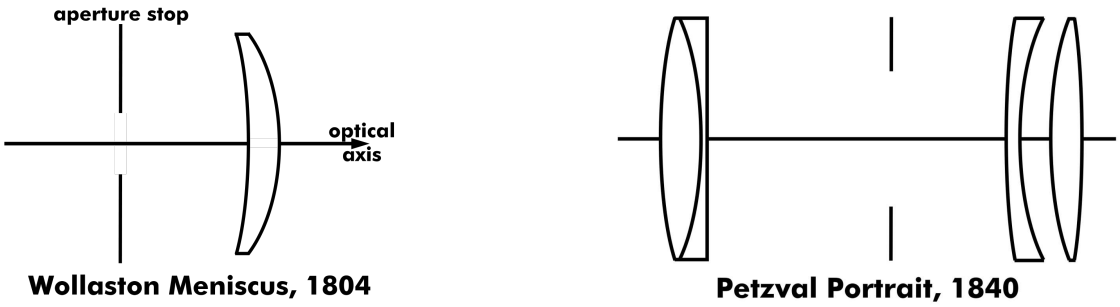


Figure 2.5: Pinhole lens

2.2.2 Increasing the aperture in the late XIXth century

In order to collect more light, optical engineers switched to an optical design made of glass elements. The Wollaston Meniscus was implemented in a *camera obscura* in 1812 (see figure 2.6a). It was later optimized by Charles Chevalier into an achromatic doublet for the Daguerre camera, but its aperture of f/16 as well as the low sensitivity of film made its use for portraits difficult.



(a) Wollaston meniscus lens

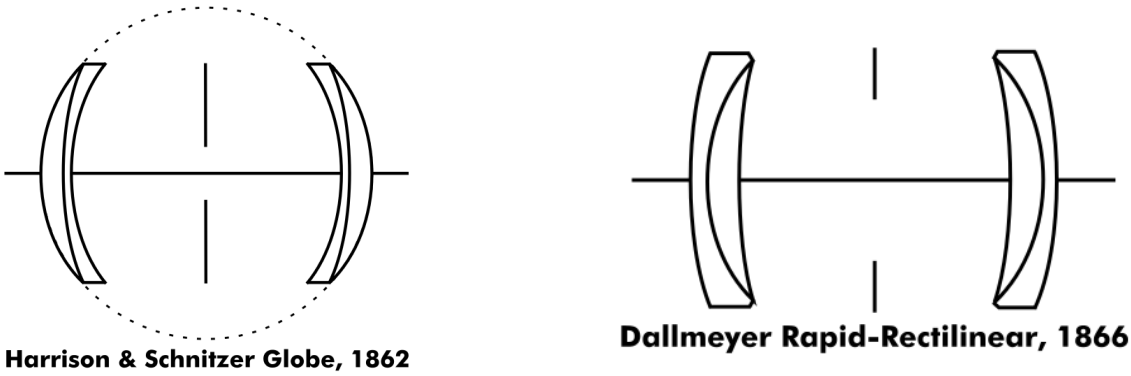
(b) Petzval portrait lens

Figure 2.6: Early lens designs, source - Paul Chin (CC by SA 3.0)

In 1840, Joesph Petzval designed the first fast lens using a double doublet system [17]. His lens is shown in figure 2.6b. With a focal length of 160mm and an aperture of f/3.6, the Petzval portrait lens collected around 20 times more light than the Wollaston-Chevalier lens and enabled a great development of portrait photography.

The Petzval lens has a limited field of view, which makes it undesirable for landscape applications where the Chevalier lens remained the only option until the 1860s [18]. Two main systems then became available, offering a wider field of view : the Harrison and Schnitzer Globe lens created in 1862 offers a typical 80° to 90° field of view with an f/30 to f/16 aperture. This initial system was

updated into the popular Rapid Rectilinear lens and the Steinheil Aplanat [19]. These systems are shown in figure 2.7. The former designed by John Henry Dallmeyer in 1866, the latter designed by Carl August von Steinheil at the same time in Germany. Both lenses provide a narrower field of view with a wider aperture. The Rapid Rectilinear offers typically $\pm 24°$ field angle for an aperture of f/8. Steinheil's wide-angle aplanat offered $\pm 45°$ field of view with an f/18 aperture, the landscape aplanat offered $\pm 30°$ field of view with an f/10 aperture and the portrait aplanat proposed $\pm 12°$ with an f/3.5.



(a) Globe lens

(b) Rapid rectilinear lens

Figure 2.7: Early lens designs, source - Paul Chin (CC by SA 3.0)

In the late XIXth century, the improvements in glass manufacture (Otto Schott), optics understanding (Ernst Abbe) and lens manufacture (Carl Zeiss) resulted in a combined improvement of lens design, glass purity and lens manufacturing.

The Cooke triplet, see figure 2.8, was patented in 1893 offering an f/3.5 aperture. Theoretically, the Cooke triplet has a null Petzval sum and corrects perfectly all third order aberrations. The Cooke triplet can be optimized using several methods [20]. Kingslake [21] proposes some elements for designing a triplet type lens without using a computer.

2.2.3 Interchangeable lenses in the early XXth century

The growth of the camera market resulted in an active development of various single-focal lenses for photographers and cinematographers. The Tessar lens was introduced in 1902 by Paul Rudolph based on his previous Unar and Protar lens designs. This lens was initially proposed with an f/6.3 aperture before being improved to an f/2.8 aperture in 1930. The evolution of Unar and Protar designs into the Tessar lens is shown in figure 2.9.
The Cooke triplet was developed into the 4-element Ernostar lens design in 1924 with an f/2.0 aperture, enlarged to an f/1.8 aperture in 1925. In 1932, Ludwig Jakob Bertele patented for Zeiss

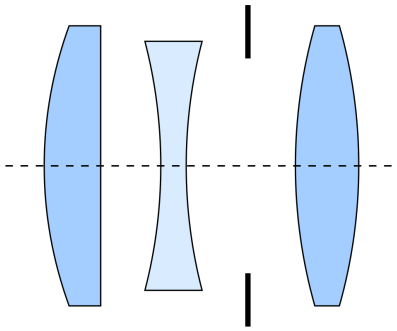


Figure 2.8: Cooke Triplet Lens. Source: Panther (CC by SA 2.5)

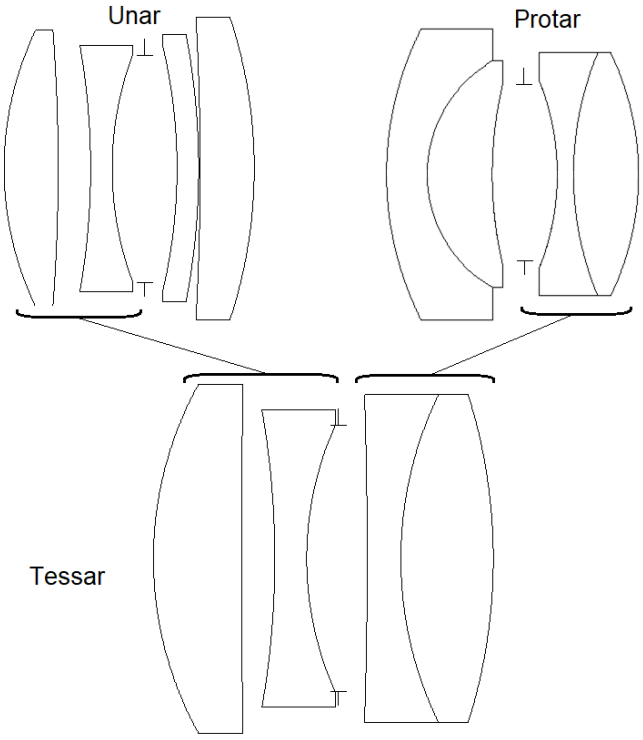


Figure 2.9: Origin of the Tessar lens design

Ikon the Sonnar lens, offering a bright f/1.5 aperture thanks to 7 elements in 3 groups (modified triplet design).

In 1927, Zeiss released the Biotar 50mm f/1.4. Many designs based on the double-Gauss have been proposed, of which Smith [22] offers a survey. A more detailed overview of the double-Gauss lens formulas is proposed in section 7.

Although not directly related to lens design, it is worth noting that the anamorphic system designed by Henri Chrétien was patented in 1926 under the name Hypergonar. It is commonly claimed that

the process was imagined by H. Chrétien after watching the movie *Napoléon* by Abel Gance which used a revolutionary process for the time [23]. The movie was recorded using 3 simultaneous cameras in order to show very wide landscapes or action scenes. A sample image is shown in figure 2.10. The anamorphic process enabled film directors to record wider aspect ratios with standard film and equipment. The process is still used to date.

Figure 2.10: Image from the film Napoléon, by Abel Gance (Public Domain)

The first camera zoom system can be tracked back to Roger Cuvillier. His invention was patented in 1949 under the name PanCinor. The system was composed of only 4 lenses. The lens offers a $\times 3$ magnification. A refinement enabled a five times focal length magnification (with the addition of a 5th lens element) in 1954. Clark and Wright [24] provide further details about early zoom history and evolution.

Pierre Angénieux invented the retrofocus system in 1950 [25] using 6 elements in 5 groups. The lens is presented in figure 2.11. The front group contains a large negative power element, followed by the rear positive power group which is the reverse of a telephoto lens. The first telephoto lens can be attributed to Petzval. For photographic systems, the retrofocus lens has a longer backfocal length compared to its focal length, which is a desirable feature for DSLR cameras.

Starting from the 1950s, the bulk of lens manufacture moved from Europe and the USA to Japan. Japan initially exported some affordable entry-level cameras and lenses and progressively improved its optical design and manufacturing standards to produce better lenses. According to [10], computer aided design reduced the gap in image quality between the lenses sold by experienced designers and more recent companies.

2.2.4 Introduction of computer design

Until the late 1940s, optical design was hand-calculated and ray tracing was computed using logarithmic scales [26]. The earliest attempt to use electronic calculator was made at Harvard by James G. Baker in 1944 with a Mark I calculator. Baker's program enabled to trace rays in typically 2

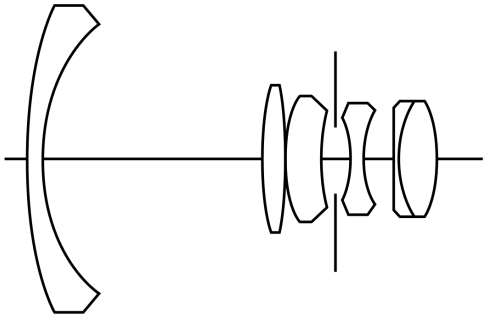


Figure 2.11: Retrofocus lens patented by Pierre Angénieux. Source: Fred the Oyster (CC by SA)

minutes per ray per surface.

At the Institute of Optics of the University of Rochester, Robert E. Hopkins used an IBM 650 in 1953 [27]. The computer was replaced in 1961 with an IBM 7070 and an IBM 7074 in 1963. Gordon Spencer developed the Automatic Lens Correction (ALEC) software in 1963 using FORTRAN [28–30]. In parallel to this work, the Optical Routines for Design and Analysis (ORDEAL) software was used for teaching and research purposes. The ALEC program was extended into the program FLAIR (it is not known what the acronym stands for) when Spencer worked for IBM. In the 1960s, Charles Wynne created the SLAMS (Successive Linear Approximations by Maximum Steps) method [31, 32]. In 1963, Donald P. Feder developed for Eastman Kodak one of the first commercially available lens design software known as LEAD (likely for Lens Evaluation And Design).

Spencer's method [33] was reused in his ACCOS V design program [34], although it is believed that most modern software (Code V, Zemax, Oslo, Hexagon...) uses the same principles. Stavroudis [35] provides an extensive overview of various optical design methods and algorithms from the early 1960s.

In the 1960s, lens makers started investigating the use of lens design softwares to design their lenses. At Zeiss, Erhard Glatzel is credited for creating an alternative lens design method [36]. At Thales Angénieux, Pierre Angénieux and his assistant Jacques Debize started programming an optical design software [37], likely derived from IBM's FLAIR, in 1968.

The use of computers saved considerable amounts of time for ray tracing and lens design and thus enabled better performance through an increase in the complexity of the optical formulas.

Automatic lens design is still an active field with new optimization algorithms [38] or algorithms designed for new tasks (e.g. free-form surfaces [39, 40]).

2.3 Sensor noise and dynamic range : a specific challenge for the smartphone industry

2.3.1 Noise sources

In this section, the most common noise sources for all types of sensors are summarized. This information will be used to estimate the main noise source in small sensor devices.

2.3.1.1 Photon Noise

Assuming the sensor is ideal, there is still a noise source resulting from the nature of light. Photon noise (or shot noise) results from the random arrival time of photons on a sensor and does not depend on the sensor technology. The arrival of photons is statistically governed by a Poisson distribution :

$$P_N = \frac{1}{N} \times (\overline{N_{\text{photon}}})^{N_{\text{photon}}} \times exp(-\overline{N_{\text{photon}}}) \tag{2.1}$$

P_N being the probability of arrival of a number N_{photon} of photons where $\overline{N_{\text{photon}}}$ is the expected average number of photons. The variance σ_N^2 of this noise is given by :

$$\sigma_N^2 = \overline{N_{\text{photon}}} \tag{2.2}$$

Consequently, the RMS noise level σ_N increases with the square root of the number of expected photons, whereas the signal increases linearly with the number of photons. Overall, the signal to noise ratio (SNR) increases as the square root of the number of photons.

$$\text{SNR} = \sqrt{\overline{N_{\text{photon}}}} \tag{2.3}$$

2.3.1.2 Thermal Noise

The thermal noise was measured by John B. Johnson in 1926 and analysed theoretically by Harry Nyquist. For this reason, it is also known as Johnson-Nyquist noise.
Thermal noise characterizes the random fluctuation of the detected voltage because of the Brownian motion of electrons. For a resistance R, a temperature T and a frequency bandwidth Δf it is possible to express the Root Mean Square of the voltage fluctuation with Boltzmann constant k_B:

$$v_T = \sqrt{4k_B T R \Delta f} \tag{2.4}$$

2.3.1.3 Dark Current

The dark current is a current measured even if the sensor is not illuminated. It combines thermal generation of surface states J_{surf}, thermal generation within the bulk silicon J_{gen} and thermal generation in the depletion region J_{depl}.

$$J_{surf} = \frac{qS_o n_i}{2} \propto \frac{qS_o \sqrt{T^3 \exp(-E_g/(k_B T))}}{2} \tag{2.5}$$

$$J_{gen} = \frac{qn_i W}{\tau_G} \tag{2.6}$$

$$J_{depl} = \frac{qD_n n_i^2}{L_n N_A} \tag{2.7}$$

The electrons generated by the dark current accumulate in each pixel of the sensor.

2.3.1.4 Read Noise

Read noise is introduced at the conversion from a charge value to a voltage to a digital value. In the case of a CCD camera, the photons generate some charges in each pixel of the sensor, the charges are transferred to a single amplifier and analogue-to-digital converter.

In the early 2000s, CMOS sensors underperformed their CCD counterparts regarding noise. In 2000, Tian et al. [41] claim that CMOS image sensors have higher temporal noise, higher fixed-pattern noise, higher dark current, smaller full well charge capacitance, and lower spectral response. In 2006, the observation still holds and can be found in Bigas et al. [42]. According to Boukhayma [43] however, modern CMOS sensors featuring four transistors per pixel outperform CCD electronics. In order to better estimate noise properties for specific CMOS and CCD sensors, one can implement the numerical simulations proposed by Konnik and Welsh [44].

2.3.1.5 Fixed-pattern Noise

Fixed-pattern noise, also known as Photo Response Non Uniformity (PRNU), comprises the different responses of the different pixels inside the same sensor. The size of the pixels can slightly vary from one pixel to another. The sensitivity can also be slightly different for each pixel. This noise is usually calibrated by manufacturers and automatically corrected.

2.3.2 Signal to Noise Ratio

The Signal-to-Noise-Ratio (or SNR) is the metric that enables comparing the level of information to the level of noise for a given system. It is used in multiple fields including imaging.

2.3.2.1 What is the minimum SNR accepted ?

The required SNR depends on the application. For detection, a low SNR may be sufficient, but for imaging applications, one would expect the SNR to be as high as possible. A DSLR with good lighting conditions gives a typical $SNR \approx 100$. In order to better appreciate the visual differences between several SNR values, a visual comparison is provided in figure 2.12.

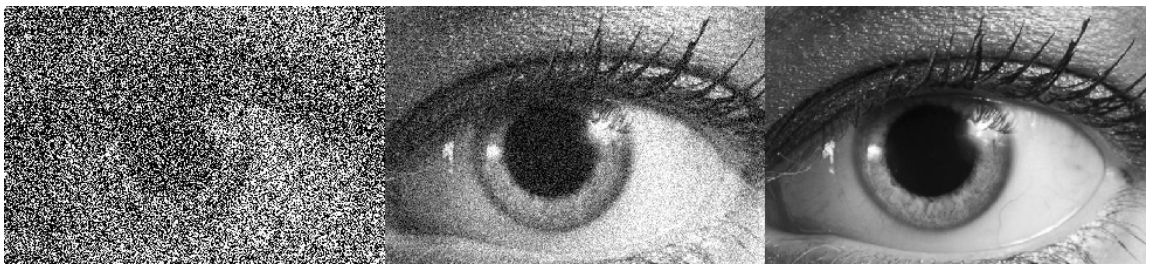


Figure 2.12: Visual comparison for various SNR. From left to right : $SNR = 1$, $SNR = 10$, $SNR = 100$. The image on the right was captured using the Nikon D800E with a Full-Well-Capacity of 55×10^3 electrons, which translates to a SNR of 99.5 in this case where the average pixel collects 18% of the full-well capacity. - source : T. Cognard

2.3.2.2 Peak SNR for sensors

Let us compute the peak values for the SNR depending on the sensor size. At best a sensor should suffer only from photon noise, which is the physical limitation for all sensors and technologies.
The key parameter limiting the SNR is the capacity of a pixel to store electrons. As a matter of fact, the size of a pixel is directly connected to its ability to store electrons. The typical value is named Full-Well Capacity (FWC) and corresponds to the number of electrons before the pixel is saturated. Some rough FWC values are provided in table 2.2, which have been found in several datasheets from commercial sensors [**?**].

Sensor size	Typical Full Well Capacity	Peak SNR	Mean SNR
Full Frame	60 000e^-	245	104
APS-C	35 000e^-	187	79
4/3”	20 000e^-	141	60
1/3”	3 000e^-	55	23

Table 2.2: Full-Well Capacity and SNR. The Peak SNR is obtained using the photon noise formula for the Full Well Capacity [45]. The Mean SNR is obtained using the photon noise formula for 18% of the Full Well Capacity, which corresponds to the typical exposure target for digital cameras.

Using the full-well capacity to compute an SNR is not meaningful since it corresponds to a white image. A more realistic approach is to use the standard used in the camera industry for exposure

computation: the target signal is to get $\text{Gray}_{\text{small Target}} = 18\%$ of the full well capacity FWC on average for the image. With this definition, based on the Shot Noise alone, the mean SNR can be computed for various sensor sizes.

$$\text{SNR} = \sqrt{\text{Gray}_{\text{small Target}}\text{FWC}} \tag{2.8}$$

The results calculated from Equation 2.8 are shown as mean SNR in table 2.2 and the best theoretical SNR (corresponding to the Shot noise of the FWC) is presented in the best SNR column.

References

[1] A. Davenport. *The History of Photography: An Overview*. University of New Mexico Press, Boston, 2 edition, 1991.

[2] H. Gernsheim. *A Concise History of Photography*. Dover photography collections. Dover Publications, New-York, 3 edition, 1986.

[3] A. Cox. *Manual of Photo-Technique*. Photographic Optics. Focal Press Limited, 1966.

[4] Waffles. The worlds first digital image was in 1957. Available at `http://wafflesatnoon.com/first-digital-image/`. Published: 16-08-2015.

[5] Lionel Gérard Colbere. Rétro photo - sony mavica : des balbutiements à la réalisation. Available at `https://www.lesnumeriques.com/photo/retro-photo-sony-mavica-des-balbutiements-a-la-realisation-pu114347.html`. Published: 31-03-2015.

[6] Lionel Gérard Colbere. Rétro-photo : de 1981 à 1984, les premiers pas. Available at `https://www.lesnumeriques.com/photo/retro-photo-de-1981-a-1984-les-premiers-pas-pu115281.html`. Published: 17-08-2015.

[7] Sophie Triantaphillidou, Elizabeth Allen, and R Jacobson. Image quality comparison between jpeg and jpeg2000. ii. scene dependency, scene analysis, and classification. *Journal of Imaging Science and Technology*, 51(3):259–270, 2007.

[8] Elizabeth Allen, Sophie Triantaphillidou, and R Jacobson. Image quality comparison between jpeg and jpeg2000. i. psychophysical investigation. *Journal of Imaging Science and Technology*, 51(3):248–258, 2007.

[9] Caroline Cakebread. People will take 1.2 trillion digital photos this year thanks to smartphones. Available at `https://www.businessinsider.com/12-trillion-photos-to-be-taken-in-2017-thanks-to-smartphones-chart-2017-8?r=US&IR=T`. Published: 01-09-2017.

[10] Rudolf Kingslake. *A history of the photographic lens*. Academic Press Inc., The Institute of Optics, University of Rochester, New York, 1989.

[11] Joseph Petzval. Bericht über dioptrische untersuchun. *Wien, Sitz. Ber.*, 26:33–90, 1857.

[12] Eugen von Lommel. *Die Beugungserscheinungen einer kreisrunden Oeffnung und eines kreisrunden Schirmchens*. Akad. Franz, 1885.

[13] Lord Rayleigh. X. on pin-hole photography. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 31(189):87–99, 1891.

[14] John William Strutt and Lord Rayleigh. Some applications of photography. *Nature*, 44(1133):249–254, 1891.

[15] M Susan Barger and William B White. *The daguerreotype: Nineteenth-century technology and modern science*. Johns Hopkins University Press, Baltimore, 2000.

[16] Harald Johnson. The first photo. Available at `https://petapixel.com/2013/10/02/first-photo/`. Published: 02-10-2013.

[17] Schade Willy. Petzval-type photographic objective, March 7 1950. US Patent 2,500,046.

[18] Rudolf Kingslake. A classification of photographic lens types. *Journal of the Optical Society of America*, 36(5):251–255, 1946.

[19] R. Kingslake. The development of the photographic objective1. *Journal of the Optical Society of America*, 24(3):73–84, Mar 1934.

[20] Darko M Vasiljevic. Optimization of the cooke triplet with various evolution strategies and damped least squares. In *Optical Design and Analysis Software*, volume 3780, pages 207–215. International Society for Optics and Photonics, 1999.

[21] Rudolf Kingslake. Lens design without a computer. In *Computer-Aided Optical Design*, volume 147, pages 58–62. International Society for Optics and Photonics, 1978.

[22] Warren J Smith, Ellis Betensky, David Williamson, Juan C Miñano, and R John Koshel. The past, present, and future of optical design. In *International Optical Design Conference 2006*, volume 6342, page 63422Y. International Society for Optics and Photonics, 2006.

[23] Jean-Jacques Meusy and Alan Williams. Henri chretien, bernard natan, and the hypergonar. *Film History*, 15:11–31, 2003.

[24] Alan Dennis Clark and WD Wright. *Zoom lenses*. Adam Hilger, London, 1973.

[25] Angenieux Pierre. Wide-angle photographic objective lens assembly, August 18 1953. US Patent 2,649,022.

[26] Robert R Shannon. *The art and science of optical design*. Cambridge University Press, Cambridge, 1997.

[27] Robert E Hopkins and Gordon Spencer. Creative thinking and computing machines in optical design. *Journal of the Optical Society of America*, 52(2):172–176, 1962.

[28] Donald P Feder. Automatic lens design methods. *Journal of the Optical Society of America*, 47(10):902–912, 1957.

[29] Donald P Feder. Automatic lens design with a high-speed computer. *Journal of the Optical Society of America*, 52(2):177–183, 1962.

[30] Donald P Feder. Automatic optical design. *Applied Optics*, 2(12):1209–1226, 1963.

[31] CG Wynne. Lens designing by electronic digital computer: I. *Proceedings of the Physical Society*, 73(5):777, 1959.

[32] Jonathan Maxwell and Prudence MJH Wormell. Charles gorrie wynne. 18 may 1911–1 october 1999: Elected frs 1970. Available at `https://doi.org/10.1098/rsbm.2001.0030`, 2001.

[33] Gordon H Spencer. A flexible automatic lens correction procedure. *Applied Optics*, 2(12):1257–1264, 1963.

[34] Daniel G Albert. The accos v lens design program: A summary. In *Recent Trends in Optical Systems Design and Computer Lens Design Workshop*, volume 766, pages 224–227. International Society for Optics and Photonics, 1987.

[35] Orestes N Stavroudis. Automatic optical design. In *Advances in Computers*, volume 5, pages 227–255. Elsevier, 1964.

[36] Juan L Rayces. Ten years of lens design with glatzel's adaptive method. In *1980 International Lens Design Conference*, volume 237, pages 75–84. International Society for Optics and Photonics, 1980.

[37] Jacques Debize. Thales angenieux: 42 years of cine 35 mm zoom leadership. In *Optical Design and Engineering*, volume 5249, pages 261–272. International Society for Optics and Photonics, 2004.

[38] Dabo Guo, Liang Yin, and Guang Yuan. New automatic optical design method based on combination of particle swarm optimization and least squares. *Optics Express*, 27(12):17027–17040, 2019.

[39] Gerwin Damberg and Wolfgang Heidrich. Efficient freeform lens optimization for computational caustic displays. *Optics Express*, 23(8):10224–10232, 2015.

[40] Xianglong Mao, Jinpeng Li, Fengbiao Wang, Rong Gao, Xing Li, and Yongjun Xie. Fast design method of smooth freeform lens with an arbitrary aperture for collimated beam shaping. *Applied Optics*, 58(10):2512–2521, 2019.

[41] Hui Tian. *Noise analysis in CMOS image sensors*. PhD thesis, Applied Physics, Standford University, 8 2000.

[42] M Bigas, Enric Cabruja, Josep Forest, and Joaquim Salvi. Review of cmos image sensors. *Microelectronics journal*, 37(5):433–451, 2006.

[43] Assim Boukhayma. Ultra low noise cmos image sensors. Technical report, EPFL, 2016.

[44] Mikhail Konnik and James Welsh. High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial. *arXiv preprint arXiv:1412.4031*, 2014.

[45] William J. Claff. Dxomark derived sensor characteristics. Available at `http://www.photonstophotos.net/Charts/Sensor_Characteristics.htm`. Accessed: 01-2020.

Chapter 3

The notion of a Perfect Camera, Theoretical and Practical Limits

3.1 The notion of a Perfect Camera

3.1.1 Technical properties

Theoretically, a perfect sensor should record the position, ray direction, time of arrival, wavelength and polarization of each photon collected during the exposure time with infinite accuracy. In practice however, acquiring all of these parameters at the same time for all photons is not possible. Let's provide a brief overview of the challenges faced in achieving such properties:

- The accuracy of photon location on the sensor is limited by the pixel size.
- The ray direction can be estimated using micro-lens arrays (cf. Section 5). One has to compromise lateral resolution to achieve better directional estimation.
- The time of arrival of a photon can be determined using a high frequency system. Unfortunately, such systems (e.g. photomultiplier tubes) cannot handle a high rate of photon arrival.
- The wavelength in photographic systems is usually measured using a Bayer colour filter pattern which only has a limited spectral resolution.
- Polarization is rarely measured for photographic applications, although photographers use polarization filters to manage distracting reflections on shiny surfaces.
- Quantum efficiency of the sensor must be as close as possible to 100%. The notion of Detective Quantum Efficiency (DQE) explained in detail by Dainty and Shaw [1], provides a way to measure the efficiency of sensors.

3.1.1.1 Spatial Resolution

The spatial resolution achieved by the sensor and lens should be as high as possible and compatible with one another since both elements can limit the resolution of the image. The sensor, usually because of the number of pixels available. The lens because of aberrations and diffraction effects. Aberrations affect the Modulation Transfer Function (MTF) of the optical system, which may result in contrast loss and/or lowered maximum resolution. Diffraction affects the resolving power of the lens.

One should also mention the use of optical low-pass filters on top of sensors in order to prevent aliasing effect (also known as Moiré). As explained with the sampling theory, any signal of a given frequency is incorrectly estimated if the sampling is less than twice the signal frequency. This property results in aliasing effect which can be corrected if all spatial frequencies above twice the sensor spatial frequency are removed (i.e. an optical low-pass filter).

Obtaining colour information through a Bayer filter array partly reduces the spatial resolution achieved, because of the interpolation required to estimate the colour for all pixels. Other colour technologies such as Foveon enable a higher resolution at the expense of greater noise [2]. Riutort-Mayol and al. [3] provide an in-depth study of the noise sources in Foveon sensors.

3.1.1.2 Dynamic Range

Dynamic range characterizes the ability of a camera to capture at the same time the shadows and highlights of a given image.

Dynamic range is traditionally expressed in stops among the photographic community, where each stop corresponds to multiplying or dividing the amount of collected light by two. Having 10 stops of dynamic range means detecting light levels over a range of 2^{10}. Some dynamic range values and the associated contrast ratios can be found in table 3.1.

Photoreceptor	Dynamic range (stops)	Ratio
JPEG	8	256 : 1
RAW file	14	16384 : 1
Cinema camera	15	32768 : 1
Human eye (including variation of the pupil)	24	16777216 : 1

Table 3.1: Dynamic range of different photo-receptors

A typical problem of digital sensors as opposed to film is the problem of pixel saturation at high light levels. To put it simply, film can be overexposed without losing too much detail in the highlights, while sensors cannot. The behaviour of film emulsions is known as a D log E curve in the

highlights which corresponds to a logarithmic behaviour to the exposure of photons as shown in figure 3.1. The variable D stands for density and E stands for exposure. The overall behaviour can be explained with statistics: if most silver receptors have already been activated by photons, the likelihood of finding an available receptor dramatically decreases.

The exposure scale is the part of the curve where the density and the exposure are linearly related. The slope of the curve measures the contrast of the image:

$$\gamma = \frac{dD}{d(log_{10}E)} \Leftrightarrow D = \gamma(log_{10}E - log_{10}E_0) \tag{3.1}$$

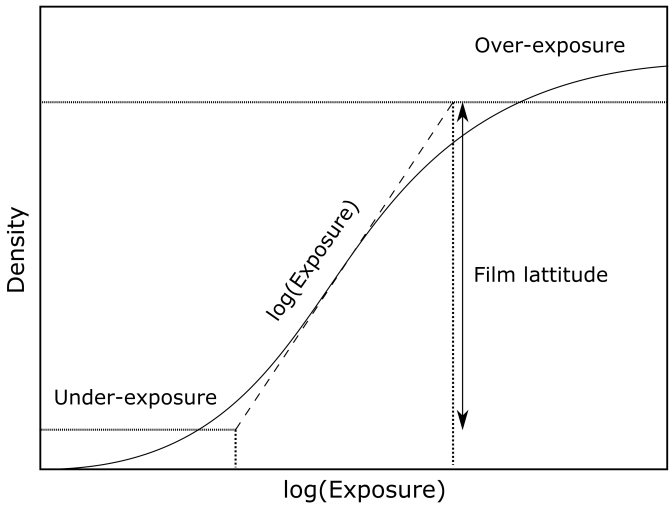


Figure 3.1: Exposure and reaction density

On the other hand, Charge Coupled Devices (CCD) or Complementary Metal-Oxide-Semiconductor (CMOS) technologies can store only a limited number of electrons before being saturated. Saturated pixels are the pure white areas in the image.

Photographers consider that it is preferable to make an error on the side of overexposing film and underexposing digital sensor since saturated pixels cannot be adjusted contrary to film.

3.1.1.3 Sensitivity

Sensitivity is the ability of a camera to detect low levels of light. Ultimately a camera is limited to the detection of single photons. In practice however, the sensitivity is limited by the existence of noise. The read noise alone can be as high as a few electrons (c.f. section 2.3). In the final image, image processing can be used to attenuate some of the visible noise at the expense of some resolution and accuracy (excessive blur and artefacts). In low-light situations, the perceived performance of miniature cameras has partly improved thanks to denoising techniques [4].

For a given system (fixed quantum efficiency and sensor size) the easiest approach to improve the sensitivity is to increase the light flux with the lens aperture. This can benefit the system sensitivity at the expense of optical aberrations.

3.1.1.4 Colour reproduction

Colour reproduction characterizes the capability of a device to record and reproduce as faithfully as possible the genuine colours of the scene. Since each photon has an associated wavelength, physical matching of colors corresponds to recording each photon with the associated wavelength. In this regard, the use of a Bayer pattern [5] provides an easy solution, corresponding to human vision based on trichromatic sensitivity. A Bayer pattern is a colour filter matrix as shown in figure 3.2.

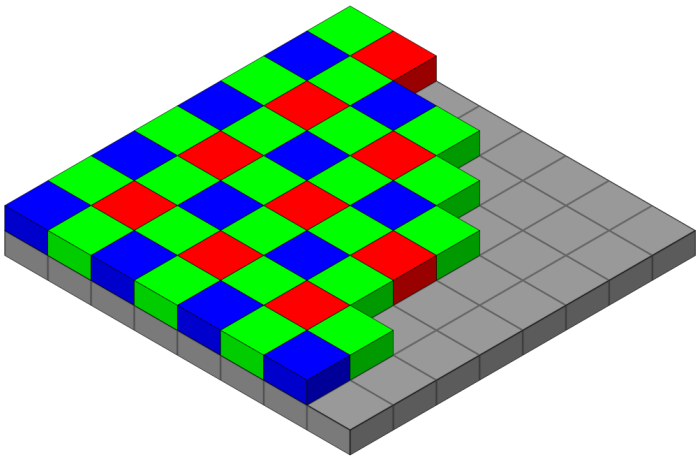


Figure 3.2: Bayer pattern on a sensor, source : Cburnett (Public Domain).

The Bayer pattern attenuates the optical signal as shown in figure 3.3 and as a consequence alternative patterns have been proposed to provide a better light efficiency. Luo [6] proposes a sparse pattern which takes advantage of the lower colour resolution in the human eye as opposed to the contrast resolution [7]. Wang and Menon [8] present a diffraction-based alternative to colour filters, at the expense of a greater image processing operation.
Colour accuracy also depends on physiological perception. In practice, two different mixes of red green and blue can look similar to the human eye [10, 11]. Colour spaces are based on the trichromatic properties to offer an as-complete-as-possible set of colours from red, green and blue. These values are usually refered to as X,Y,Z or sometimes normalized as $x = \frac{X}{X+Y+Z}$, $y = \frac{Y}{X+Y+Z}$ and $z = \frac{Z}{X+Y+Z}$ as in figure 3.4. The spectrum locus, which is used to assess the range of colours available is shown in the same figure together with some colour spaces (sRGB, ProPhoto RGB, Adobe RGB 1998).

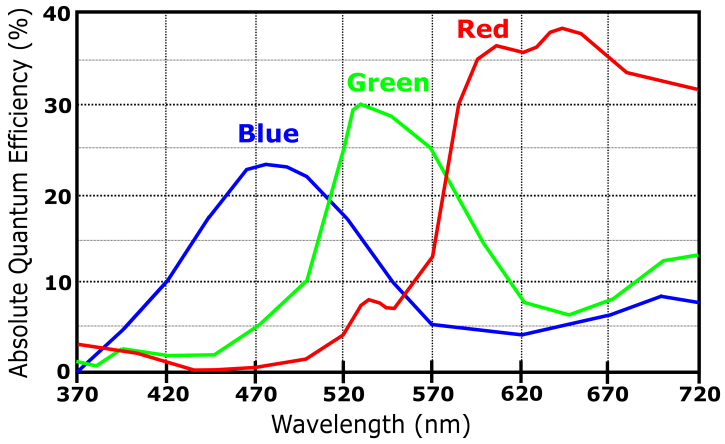


Figure 3.3: Absorption spectrum of Bayer filter, values obtained from Olympus [9].

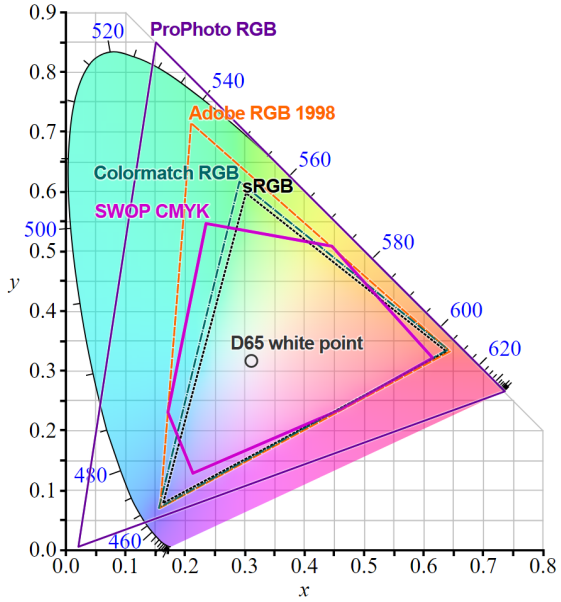


Figure 3.4: Colour spaces (polygons) and spectrum locus (2D-spectrum of all the visible colours). The x-y axis corresponds to the normalized tristimulus values. Source: BenRG and cmglee (CC by SA 3.0)

Even more complex is the psychological perception of colours [12] since the context can influence the human perception of colours. The simplest example is the research on white-balance algorithms [13].

Besides human vision, there is some interest in detecting near-infrared wavelengths. To do so, Bayer patterns can be adapted to include near-infrared filters [14, 15]. Biometrics is a possible application for such systems, as presented in [16].

3.1.1.5 Field of View and Depth of Field

The perfect camera should provide unlimited possibilities when it comes to the field of view and the depth of field. For this purpose, most professional photographers use several lenses to adjust for the needed focal length and aperture. While in principle some zoom lenses cover most of a photographer's needs, they are usually slow (f/5.6 or even f/6.3) and have poor aberration correction. On the over hand, wide-aperture lenses are usually single focus. Finally, the choice of the focal length and focus must happen prior to taking the picture in most systems, although one might want to adjust these parameters later on. The solution currently used by some photographers is to frame wider and crop inside the image to digitally zoom in. As for depth management, there are still very limited options available.

Light-field cameras have been proposed and marketed (Lytro) as a solution to select the depth of field and refocus after taking the image, at the expense of the spatial resolution of the image. At the moment interchangeable-lens camera systems are the only ones enabling all of the creativity through an extensive range of focal lengths (from 8mm to 1200mm) and lens types (fish-eye, portrait, macro, tilt-shift...). The framing, the focus and the blur control are important parts of the creative process.

3.1.2 Commercial constraints for smartphone cameras

In addition to these technical aspects, consumer devices need to take into account the commercial constraints. The main ones for smartphone devices are power consumption, form factor and cost.

3.1.2.1 Power Consumption

Power consumption should be kept as low as possible for embedded camera systems. The battery is an increasingly important concern for customers. Unfortunately, some research suggests that the camera applications use a significant part of the battery and processing power (mostly to run image-processing algorithms). In this regard, the use of dedicated GPU-style chips for photography processing or video encoding have enabled a better autonomy without reducing the performance. Some dedicated systems are also embedded to run new machine learning based algorithms.

3.1.2.2 Form factor

The form factor of the camera module is critical, especially the thickness. The weight is not a real issue and does not influence the lens design. As for the thickness, one should keep in mind that thinner phones sell better (as long as they keep structural integrity).

From a system design point of view, the limited thickness of the phone constrains the lens design (number of lens elements). Beyond the lens, the sensor size is also limited: even if a manufacturer

wanted to sell a premium smartphone with a full frame sensor and lens, it would not fit in an 8mm thin phone.

3.1.2.3 Cost

Cost should be kept as low as possible from a manufacturer point of view. This is done both with larger volumes (economy of scale) and with cost-effective components. The size of the sensor plays a large role in this equation although it is a critical element in image quality. For high-end phones, the sensor size has increased dramatically in the last few years (from 1/3.2" on the iPhone 5 to 1/2.55" on the iPhone 11). The lens cost is relatively low since most lens elements are made of plastic and moulded.

Overall, the typical cost for a smartphone camera is believed to be in the 5$ to 20$ range, with the main camera lens being usually the most expensive module of the system.

3.2 Background on Lateral Resolution

In this section, we focus on the number of cycles per pixel in the final image which can ultimately be limited by several factors: diffraction, optical aberrations and sensor pixel pitch.

3.2.1 Diffraction theory and influence of diffraction on the lens resolving power

Assume a perfect lens of finite diameter D. Because of the wave nature of light, the image of a point is not a point but a Fraunhofer diffraction pattern with a size inversely proportional to the diameter of the lens [17].
For a pupil of size D, with a lens of focal length f with an illumination amplitude A, the radial intensity $I(r)$ is given by:

$$I(r) = \left(\frac{A}{\lambda f}\right)^2 \left[\frac{2J_1(kDr/f)}{kDr/f}\right]^2 \tag{3.2}$$

where J_1 is a Bessel function of the first kind and $k = 2\pi/\lambda$.

Using this intensity profile, known as an Airy pattern, it is possible to find the first minima of the function for $r = 1.22$. As a result, the lens resolving power is limited by:

$$\phi_{\text{Airy}} = 1.22\frac{\lambda f}{D} \tag{3.3}$$

with ϕ_{Airy} the radius of the Airy diffraction pattern, λ the wavelength, f the focal length and D the diameter of the pupil.

Resulting from equation 3.3, we can use the Rayleigh criterion, shown in figure 3.5, to obtain the theoretical lens resolving power for various apertures and sensor sizes.

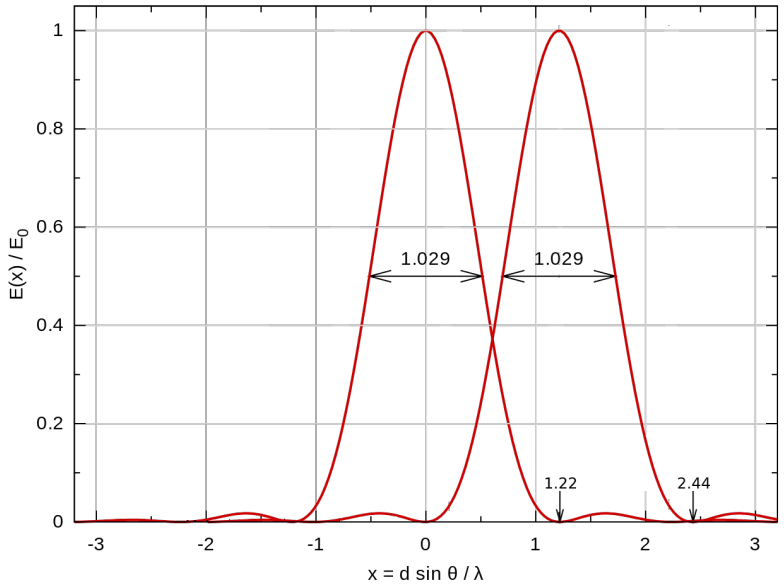


Figure 3.5: Airy functions and Rayleigh criterion, source: Edgar Bonet (CC by SA 4.0)

The Rayleigh criterion states that the first diffraction minimum of the image of one source point coincides with the maximum of another. It is possible to use this definition to find the diffraction-limited resolving power of a lens. To do so, one can compute the area of the central peak of the Airy pattern corresponding to this criterion.

$$S_{\text{Airy}} = \pi \times \left(\frac{\phi_{\text{Airy}}}{2} \right)^2 \tag{3.4}$$

The diffraction resolving power $R_{\text{diffraction}}$ of the lens can be approximated with the fraction of the sensor surface by the Airy surface.

$$R_{\text{diffraction}} = \frac{S_{\text{Sensor}}}{S_{\text{Airy}}} \tag{3.5}$$

3.2.2 Theoretical resolving power

The results from equation 3.5 for several apertures are gathered in table 3.2. The amount of dots that can be resolved through a diffraction limited lens varies according to the lens aperture, i.e. f-number and sensor size.

Sensor size	Resolving Power at f/2	Resolving Power at f/4	Resolving Power at f/8
Full Frame	610 million dots	153 million dots	38 million dots
APS-C	260 million dots	65 million dots	16 million dots
Four-Thirds	159 million dots	40 million dots	10 million dots
1/3”	12 million dots	3 million dots	0.8 million dots

Table 3.2: Diffraction resolving power for various sensor sizes. As assumed in equation 3.5, the lens is considered diffraction limited and the number of dots that can be imaged is computed.

It is even more interesting to compare these resolutions with the current pixel pitch from commercially available sensors that are provided in table 3.3.

Sensor size	Pixel size (μm)	Typical Pixel resolution (MPx)	Largest aperture for a diffraction-limited s
Full Frame	3.5-6	24-60	f/8
APS-C	3.5-5	18-24	f/7.1
Four Thirds	3-4	16-24	f/5.6
1/3”	0.8-1.2	12-27	f/1.7

Table 3.3: Diffraction resolving power and typical sensor resolution (2020)

While large-sensor camera systems are often limited by the pixel pitch or the optical aberrations, the small-sensor systems are mostly diffraction limited. The analysis in this chapter is related to the image plane, however the diffraction effect along the optical axis will be investigated in chapter 4.

3.3 Nyquist frequency and optical design

3.3.1 Shannon-Nyquist sampling theorem

The Shannon-Nyquist sampling theorem provides an estimate of the distance between two samples needed to reconstruct accurately a signal with a maximum spatial frequency F. For such an F frequency, samples should be at a maximum distance of $\frac{1}{2F}$.

3.3.2 Aliasing and optical design

In the case of a limited sampling rate compared to the frequency F, the measure is imperfect and leads to an aliasing effect. Aliasing, also known as Moiré patterns in imaging, is the wrong estimation of the spatial frequency of the signal. The signal has a lower estimated spatial frequency than it should.

Aliasing is an issue for digital cameras with low resolution. Until the early 2010s, most digital cameras provided an optical anti-aliasing filter. The aliasing filter is an optical low-pass filter.

Some cameras with in-body image stabilization systems offer a switchable anti-aliasing effect using small sensor movements to overcome aliasing.

For smartphone photography, most of the phones released do not have an anti-aliasing filter.

3.3.3 Matching the optical design with the sensor resolution

In order to match the pixel density and the optical design, one must take into consideration the spatial cutoff frequency of the lens and match it with the Nyquist frequency of the sensor. In table 3.4 the cutoff frequency is computed for a diffraction limited lens. In other cases, the Modulation Transfer Function (MTF) curves from the lens design provides some guidance about the suited sensor specifications. Some combinations are provided in table 3.4.

Sensor Resolution (MPx)	Pixel Size (μm)	Cutting frequency (cycles/mm)	Largest aperture for a diffraction-limited system (f-number)
0.3	7.59	66	f/11.3
2.0	2.94	170	f/4.4
5.0	1.86	269	f/2.8
8.0	1.47	340	f/2.2
12	1.20	417	f/1.8
16	1.04	481	f/1.6
20	0.93	538	f/1.4
27	0.80	625	f/1.2
32	0.73	680	f/1.1
38	0.67	742	f/1.0

Table 3.4: Cut-off frequency and equivalent f-number for a diffraction limited system for several commercially available sensors.

References

[1] J.C. Dainty and R. Shaw. *Image science: principles, analysis and evaluation of photographic-type imaging processes*. Academic Press, 1974.

[2] Jiashu Zhang. Film-like images with super-fine details using foveon x3 quattro. In *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, page IM3E.4. Optical Society of America, 2017.

[3] Gabriel Riutort-Mayol, Ángel Marqués-Mateu, Ana Elena Seguí, and José Luis Lerma. Grey level and noise evaluation of a foveon x3 image sensor: A statistical and experimental approach. *Sensors*, 12(8):10339–10368, 2012.

[4] Veli-Tapani Peltoketo. Benchmarking of mobile phone cameras. *Computer Science 16 Telecommunication Engineering*, 2016. University of Vaasa.

[5] Bryce E Bayer. Color imaging array, July 20 1976. US Patent 3,971,065.

[6] Gang Luo. Color filter array with sparse color sampling crosses for mobile phone image sensors. In *International Image Sensor Workshop*, pages 162–165, 2007.

[7] CF Stromeyer III. Edge-contingent color after effects: Spatial frequency specificity. *Vision Research*, 12(4):717–733, 1972.

[8] Peng Wang and Rajesh Menon. Ultra-high-sensitivity color imaging via a transparent diffractive-filter array and computational optics. *Optica*, 2(11):933–939, 2015.

[9] Renato Turchetta. Introduction to CMOS Image Sensors. `https://www.olympus-lifescience.com/en/microscope-resource/primer/digitalimaging/cmosimagesensors/`, 2020. [Online; accessed 19-June-2020].

[10] David H Krantz. Color measurement and color theory: I. representation theorem for grassmann structures. *Journal of Mathematical Psychology*, 12(3):283 – 303, 1975.

[11] David H Brainard and Andrew Stockman. *Color Vision Mechanisms*. McGraw-Hill New York, NY, 2010.

[12] Nigel W Daw. The psychology and physiology of colour vision. *Trends in Neurosciences*, 7(9):330–335, 1984.

[13] Francesca Gasparini and Raimondo Schettini. Color correction for digital photographs. In *12th International Conference on Image Analysis and Processing, 2003. Proceedings.*, pages 646–651. IEEE, 2003.

[14] Tudor Nedelcu, Shejin Thavalengal, and Peter Corcoran. Hybrid sensor visible+ near infrared demosaicing. In *2015 IEEE 5th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pages 252–255. IEEE, 2015.

[15] Zahra Sadeghipoor, Yue M Lu, and Sabine Süsstrunk. A novel compressive sensing approach to simultaneously acquire color and near-infrared images on a single sensor. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1646–1650. IEEE, 2013.

[16] Shejin Thavalengal. Contributions to practical iris biometrics on smartphones. *Biometrics*, 4(4):192–199, 2015.

[17] Joseph W Goodman. *Introduction to Fourier optics*, volume 1. Roberts and Company Publishers, 3 edition, 2005.

Chapter 4

Theoretical considerations about depth-of-field

4.1 Definitions

4.1.1 What is depth of field ?

At its core the definition of the depth of field is quite simple: it is the depth of the object space for which a point imaged by the optical system will be considered "in focus" from an imaging system perspective; however, the precise definition of "in focus" needs to be quantified.

4.1.2 Planes of focus

For a given optical system and focusing distance, three planes of focus are of interest, which are marked in figure 4.1:

- The plane of focus PF, which is the object space plane for which the lens has been focused.
- The near distance plane of focus P1, which is the first plane "in-focus" from the lens.
- The far distance plane of focus P2, which is the last plane "in-focus" from the lens.

Obviously, some image space planes (P'1, P'2, P'F) are associated with each of the above-mentioned planes.
The depth of field is the distance between P1 and P2. The depth of focus is the distance between P'1 and P'2.

4.1.3 Hyperfocal distance

Hyperfocal distance is the distance such that any object between half of hyperfocal distance and infinity appears in-focus. Figure 4.2 shows the hyperfocal distance and the minimum focusing distance (half the hyperfocal distance). The hyperfocal distance relies on the definition of this "acceptable level of focus" which is the definition of in-focus from an optical point of view. The

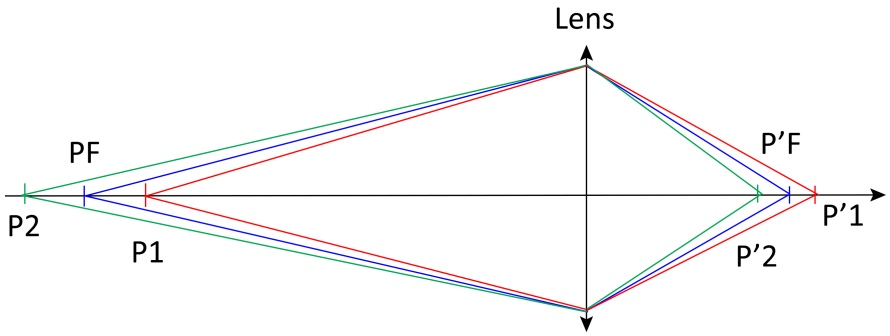


Figure 4.1: Principal planes of focus in the object and image space

in-focus part of the image is the one that is conjugated with the sensor plane. The image of a defocused point is a blur that lowers the Modulation Transfer Function (MTF) [1].

The circle of confusion is the pattern for which all the imaging points are not distinguished as they are seen by an observer on the final image. The geometrical optics model has some limits and does not explain the Arago spot [2], but can be used to understand the circle of confusion and hyperfocal distance. For specific applications using MTF curves can be an alternative to the notion of circle of confusion.

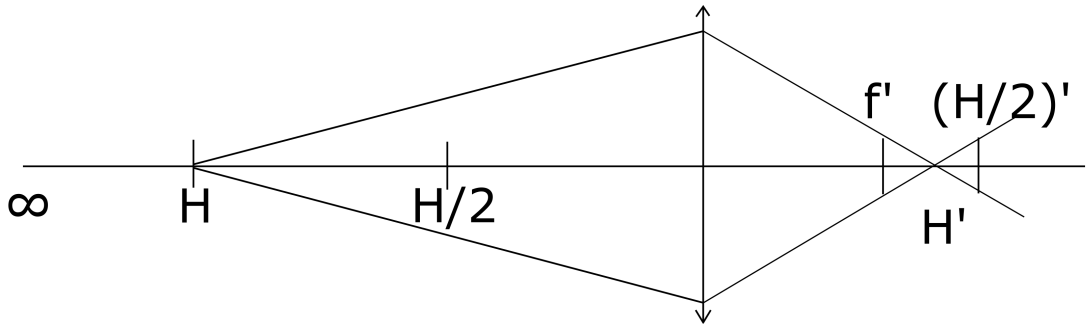


Figure 4.2: Principal planes of focus in the object and image space

The depth of field is the part of the image that achieves a certain level of resolution at least. This level of resolution was assessed during the film era with the notion of circle of confusion. Based on empirical estimates, it is considered that on the display, the circle of confusion should be of order $0.2mm$ for an emetropic eye [3].
This image display value can be translated for sensors and films. In the case of 24×36 film format cameras to produce 10×15cm sized prints, one can compute the minimal resolution expected from an image to look sharp on the print (corresponding to 1.4 million dots in the image) and then

deduce the value of this circle of confusion, with what is known by the photographic community as the "Zeiss formula", despite not having an established link with the Zeiss company.

$$C_{\text{oc}} = \frac{D_{\text{sensor}}}{1730} \tag{4.1}$$

In equation 4.1, the circle of confusion C_{oc} and the sensor diagonal D_{sensor} have mm units. The number 1730 is derived from the 20×15cm print resolution and is unitless.

4.1.4 An updated definition of the circle of confusion

In the digital age, and with the wide variety of uses of pictures (internet, prints, cinema or video games...) this notion of circle of confusion needs to be re-examined. The notion of circle of confusion relies on the observation conditions for the image.
Over a couple of decades, the dominant use for images changed from printing to screen-displaying. Compared with prints, displayed images can be resized, altered by compression or even digitally sharpened/upsampled. In the generic case of a viewer watching a screen of size $H_{\text{screen}} \times W_{\text{screen}}$, of resolution R at a given distance d_{screen}, one must find a digital equivalent of the circle of confusion. A more common metric is the resolution of the desired image, and this is a way to update it. Assuming that the observation is limited by the image resolution (in pixels), one can simply consider the circle of confusion to be the dimension of a sensor pixel mapped onto the display resolution.

The hyperfocal distance H (in m) is usually given with the formula 4.2:

$$H = \frac{f^2}{N \times C_{\text{oc}}} \tag{4.2}$$

where N is the numerical aperture of the lens, C_{oc} is the circle of confusion of the system in m and the focal length f is also in m.

Let's assume that we want to achieve a certain resolution R_{sensor} from a given sensor of dimensions $H_{\text{sensor}} \times W_{\text{sensor}}$. The circle of confusion can be obtained from the equivalent size of the minimum disk needed to achieve the resolution R_{sensor}.

In the case of a 3/2 sensor ratio, we can express the surface S_{sensor} of the sensor as:

$$S_{\text{sensor}} = H_{\text{sensor}} \times W_{\text{sensor}} = 1.5 \times H^2_{\text{sensor}} \tag{4.3}$$

The resolution of the sensor R_{sensor} then corresponds to the fraction:

$$R_{\text{sensor}} = \frac{4S_{\text{sensor}}}{\pi C^2_{\text{oc}}} \tag{4.4}$$

Which leads to the exact formula:

$$H = \frac{f^2 \times \sqrt{R_{\text{sensor}}}}{N \times H_{\text{sensor}}} \times \sqrt{\pi/1.5} \tag{4.5}$$

For photographic applications, it is of great interest to approximate $\sqrt{\pi/1.5} \approx 1$ as well as notice that some subunits can be used instead of m and pixel. In practice, the use of millimetres and megapixels (both are adapted to the size of sensor and its typical resolution) provides a very simple formula.

As an example, a 50mm lens at f/8 on a 24×36 full frame sensor with a desired resolution of 36MPx leads to $H \approx \frac{50^2 \times \sqrt{36}}{8 \times 24} \approx 78$m.

Of course this formula is consistent with the film era tabulated values for the particular case of 1.4MPx sized prints.

4.2 Computation of the depth of field

Since the definition of in-focus needs a resolution value, we can work out a formula using the resolution and the sensor parameters.
If we assume that the focus is at a distance P and the hyperfocal distance of the optical system is given by H, the first plane P_1 of the image to be sharp enough to be considered in focus is at a distance:

$$P_1 = \frac{HP}{H + P} \tag{4.6}$$

the last plane P_2 to be sharp enough to be considered in focus is at a distance:

$$P_2 = \frac{HP}{H - P} \tag{4.7}$$

The depth of field is the distance between the first and last planes of sharpness, given by $\Delta P = P_2 - P_1$.
With a direct replacement of the above expressions of P_1 and P_2 one finds:

$$\Delta P = \frac{2HP^2}{H^2 - P^2} \tag{4.8}$$

In the particular case of portrait photography (far from the hyperfocal distance, so $P \ll H$), this expression can be simplified to:

$$\Delta P = 2\frac{P^2}{H} \tag{4.9}$$

4.3 Interpretation for a digital sensor

Assuming a required minimum sensor resolution R_{sensor}, and N_{planes} the numbe of planes of focus, it is possible to use equations 4.5 and 4.8 to obtain:

$$\Delta P = \frac{2\sqrt{\pi/1.5}f^2\sqrt{R}P^2}{\frac{\pi f^4 R}{1.5N_{\text{planes}}H_{\text{sensor}}} - P^2 N_{\text{planes}} H_{\text{sensor}}} \tag{4.10}$$

The approximation in equation 4.9 gives:

$$\Delta P = \frac{2P^2 N_{\text{planes}} H_{\text{sensor}}}{\sqrt{\pi/1.5}f^2\sqrt{R}} \tag{4.11}$$

It is worth mentioning that these computations do not include the existence of the Bayer pattern (which limits the resolution that can be achieved). The signal theory would also push towards the use of a higher sampling in order to satisfy Nyquist sampling.

4.4 Optimal focusing planes

One may ask the question: how many focusing planes are necessary for a given lens and how to choose them optimally ?

4.4.1 From finite distance to infinity

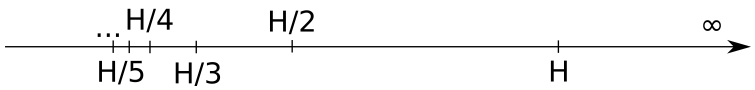


Figure 4.3: Optimal focusing planes

Considering the hyperfocal distance H, a possible way to optimize the focusing planes is to recursively compute the optimal focusing planes from infinity to the nearest plane of focusing for the lens (*Punctum Proximum*, or "near plane", P_P) which are shown in figure 4.3. This series of planes of focus are P^0, P^1, P^2,..., $P_{N_{\text{planes}}}$, with N_{planes} being a positive integer. By definition, focusing at distance $P^0 = H$ means that the area between $H/2$ and ∞ is in focus.

For a given plane of focus $P_{N_{\text{planes}}}$, the next plane of focus $P^{N_{\text{planes}}+1}$ can be computed as follows using the intervening plane P_i:

$$P_i = \frac{P^{N_{\text{planes}}+1}H}{H + P^{N_{\text{planes}}+1}} \tag{4.12}$$

$$P^{N\ \text{planes}+1} = \frac{HP_i}{H+P_i} P^{N\ \text{planes}+1} = \frac{H(\frac{HP^{N\ \text{planes}+1}}{H+P^{N\ \text{planes}+1}})}{H+(\frac{HP^{N\ \text{planes}+1}}{H+P^{N\ \text{planes}+1}})} \tag{4.13}$$

Recursively, one can then compute all the optimal focusing planes using the hyperfocal distance.

$$P^0 = H \tag{4.14}$$

$$P^{N_{\text{planes}}+1} = \frac{\frac{H^2 \times P^N_{\text{planes}}}{H+P^N_{\text{planes}}}}{\frac{H(H+P^N_{\text{planes}})+P^N_{\text{planes}} \times H}{H+P^N_{\text{planes}}}} = \frac{H^2 P^N_{\text{planes}}}{H^2+2HP^N_{\text{planes}}} \tag{4.15}$$

So in particular,

$$P^1 = \frac{H}{3} \tag{4.16}$$

$$P^2 = \frac{H}{5} \tag{4.17}$$

A quick observation of the series of the planes of focus brings:

$$\forall N_{\text{planes}}, P^N_{\text{planes}} = \frac{H}{2N_{\text{planes}}+1} \tag{4.18}$$

Computation of the first plane of focus (the *punctum proximum*) P_P

$$P_P \geq \frac{H}{2N_{\text{planes}}+2} \tag{4.19}$$

Which finally yields:

$$N_{\text{planes}} \geq \frac{H}{2P_P} - 1 \tag{4.20}$$

4.4.2 From finite distance to finite distance

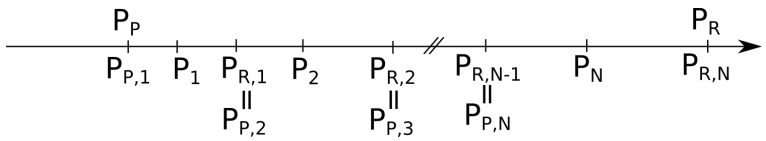


Figure 4.4: Optimal focusing planes between a distance P_P and P_R

Assuming that we have a *punctum proximum* (near point) P_P and a *punctum remotum* (far point) P_R. Let's assume we want to have N_{planes} planes of focus. The planes are named P_1, P_2, ..., $P_{N\ \text{planes}}$ as shown in figure 4.4. The first plane in-focus corresponds to the closest plane, so:

$$P_1 = \frac{H \times P_P}{H - P_P} \tag{4.21}$$

$$\forall n \in [\![1, N_{\text{ planes}} - 1]\!], P_{n+1} = \frac{H^2 P_n}{H^2 - 2HP_n} \tag{4.22}$$

Combining equation 4.21 and 4.22 brings:

$$\forall n \in [\![1, N_{\text{ planes}} - 1]\!], P_n = \frac{H^{(2n-1)} \times P_P}{H^{(2n-1)} - (2n-1) \times H^{(2n-2)} \times P_P} \tag{4.23}$$

For the last plane $P_{N_{\text{ planes}}}$ in particular, we know that the last plane in-focus corresponds to the plane P_R, so:

$$P_{N_{\text{ planes}}} = \frac{H \times P_R}{H + P_R} \tag{4.24}$$

Combining equation 4.23 and 4.24 yields:

$$\frac{H^{(2N_{\text{ planes}}-1)} \times P_P}{H^{(2N_{\text{ planes}}-1)} - (2N_{\text{ planes}} - 1) \times H^{(2N_{\text{ planes}}-2)} \times P_P} = \frac{H \times P_R}{H + P_R} \tag{4.25}$$

$$\Leftrightarrow H^{(2N_{\text{ planes}}-1)} \times [H \times (P_P - P_R) + 2N_{\text{ planes}} \times P_P \times P_R] = 0 \tag{4.26}$$

$$\Rightarrow H = 0 \text{ , or } H = \frac{2N_{\text{ planes}} \times P_P \times P_R}{P_R - P_P} \tag{4.27}$$

Obviously, the first solution does not make sense. The second solution is an elegant way to compute the optimal hyperfocal distance required to complete the task.

Since we have proposed the equation 4.5 for a 3/2 sensor ratio, one can compute H from the optimal focal length f, resolution $R_{\text{ sensor}}$, f-number $F_{N_{\text{ planes}}}$, and sensor size $H_{\text{ sensor}}$ (sensor height).

$$H = \frac{f^2\sqrt{R}}{F_{N_{\text{ planes}}} \times H_{\text{ sensor}}} \times \sqrt{\frac{\pi}{1.5}} \tag{4.28}$$

4.4.3 Typical photographic application

For most photographic applications the range of interest is comprised between 50cm and ∞. It happens to also be a convenient way to estimate the number of focusing planes from the hyperfocal distance since, using equation 4.28, $N_{\text{ planes}} = \lfloor H \rfloor$.

For a given $f = 4$mm, 1/3inch sensor-size camera, which are typical smartphone size conditions, table 4.1 provides information about the number of focusing planes and the thickness of the in-focus plane assuming that the lens is focused at 1m.

Number of planes of focus	Thickness at 1m	Parameters
5	551mm	6MPx, f/2.2
7	352mm	10MPx, f/2.0
8	223mm	10.5MPx, f/1.8
9	251mm	12MPx, f/1.7
10	179mm	13MPx, f/1.6
12	151mm	14.5MPx, f/1.4
14	130mm	17MPx, f/1.3
16	115mm	19MPx, f/1.2
18	102mm	20MPx, f/1.1
20	93mm	20.5MPx, f/1.0
25	76mm	20.5MPx, f/0.8
30	64mm	22.5MPx, f/0.7

Table 4.1: Number of planes of focus as computed from equation 4.28.

From the previous table and with some diffraction optics considerations, one can produce the Figure 4.5.

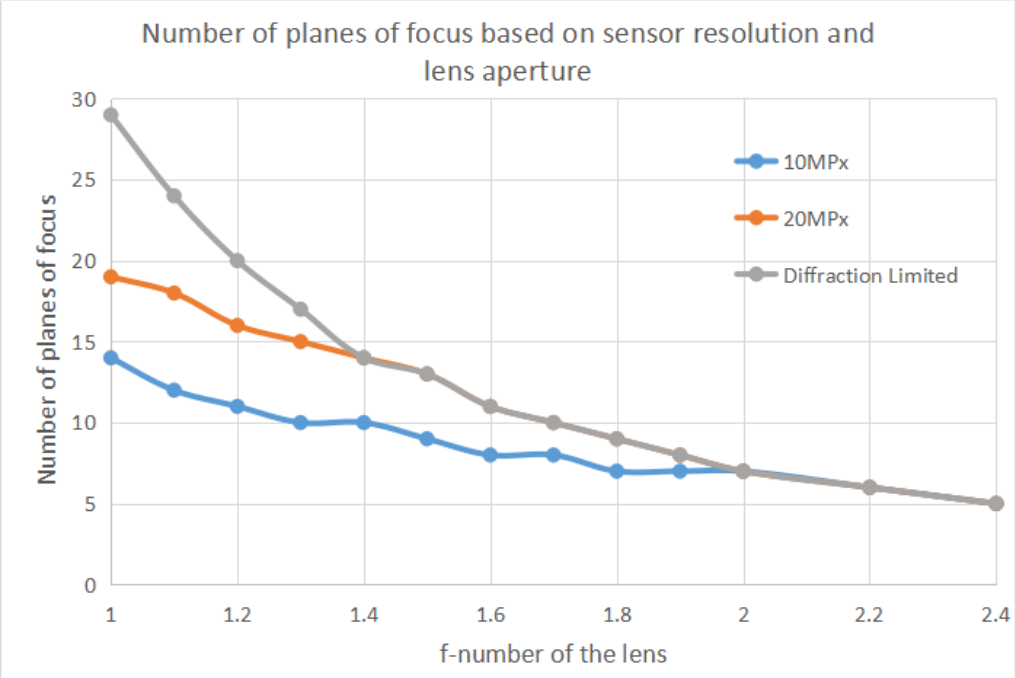


Figure 4.5: Number of depth planes for camera systems based on the resolution and f-number. In this graph, the sensor size and focal length of the lens do not vary and are similar to those chosen for table 4.1.

The main conclusions from Figure 4.5 are as follows:

- For a given sensor resolution, assuming that the resolution is pixel limited, a smaller f-number improves the depth estimation of the system.

- For a given lens f-number, assuming that the resolution is pixel limited, a higher resolution improves the depth estimation of the system.
- From a designer point of view, it is better to target a diffraction limited aperture of f/1.2 rather than an aberration limited aperture of f/1.0.

The conclusions from this theoretical work will be put to use in the chapter 7.

4.5 Estimating Axial Resolution with Geometrical Optics Theory

Contrary to the case of the lateral resolution, where the diffraction limit is widely accepted as a resolution criterion, axial resolution is often computed with a geometrical approach. In this section we provide both the geometrical approach and a novel diffraction theory-based criterion. Both resolutions are then compared.

4.5.1 Geometrical Optics Method

4.5.1.1 Principle

The geometrical approach relies on the assumptions that at the scale of the lens system, both diffraction and the optical aberrations can be neglected. In this configuration, the focus shift causes an increase in the diameter of the Point Spread Function (PSF) related to the distance from the subject as shown in Fig. 4.6. The estimation of the size of the PSF, noted Φ_{PSF}, can be made using geometric theory.

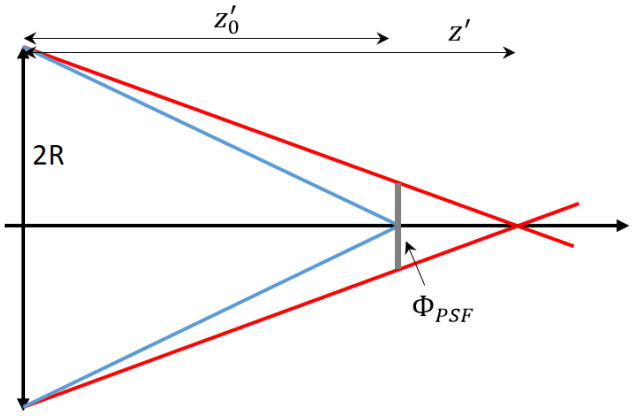


Figure 4.6: PSF produced by defocus from the image plane

4.5.1.2 General computation

Assume two points along the optical axis, z and z_0, having images z' and z'_0 respectively through a lens of focal length f and radius R. $\delta = |z - z_0|$ and $\delta' = |z' - z'_0|$ are the distance between the points in object and image space, as shown in Fig. 4.7. The sensor has a pixel size Φ_{px}.

From the previous assumptions, it is possible to obtain the formula using the Lens Formula:

$$z'_o = |\frac{z_o f}{z_o + f}| \tag{4.29}$$

It is of interest to differentiate locally this equation in order to obtain the value of z', which is the conjugate of a point z close to the point z_o.

$$\frac{\partial z'_o}{\partial z_o} = \frac{\partial |\frac{f z_o}{f + z_o}|}{\partial z_o} = \frac{f^2}{(f + z_o)^2} \approx \frac{f^2}{z_o^2} \tag{4.30}$$

As a consequence of equation 4.30:

$$z' = z'_o + \frac{\partial z'_o}{\partial z_o}(z - z_o) \approx z'_o + \frac{f^2}{z_o^2}\delta \tag{4.31}$$

The depth variation δ produces a depth variation δ' in the image space. For $\delta = z - z_0$, one can express

$$\delta' = z' - z'_0 = \frac{f^2 \delta}{z_0^2} \tag{4.32}$$

$$z' = z'_0 + \delta' = \frac{z_0 f}{z_0 + f} + \frac{f^2 \delta}{z_0^2} \tag{4.33}$$

Using the Intercept theorem in a triangle (cf. Fig. 4.6) , one gets:

$$\frac{\Phi_{PSF}}{2R} = \frac{z'_o - z'}{z'} \Rightarrow \frac{\Phi_{PSF}}{2R} = \frac{\delta'}{z'} = \frac{\frac{f^2 \delta}{z_0^2}}{\frac{z_0 f}{z_0 + f} + \frac{f^2 \delta}{z_0^2}} \tag{4.34}$$

$$\Rightarrow 2R\frac{f^2 \delta}{z_0^2} = \Phi_{PSF}(\frac{z_0 f}{z_0 + f} + \frac{f^2 \delta}{z_0^2}) \tag{4.35}$$

$$\delta = \frac{\Phi_{PSF} z_0^3}{f(z_0 + f)(2R - \Phi_{PSF})} \tag{4.36}$$

Further approximation and simplification can be made as proposed by Trouvé [4] under some conditions:

$$\begin{cases} z_0 \gg f \\ 2R \gg \Phi_{PSF} \end{cases} \Rightarrow \delta \approx \frac{\Phi_{PSF} z_0^2}{2fR} \tag{4.37}$$

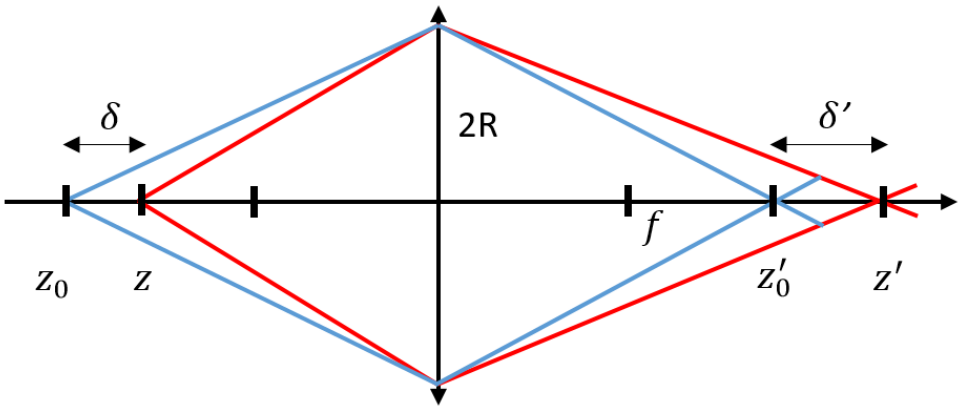


Figure 4.7: Drawing of the lens and notations

4.5.1.3 Validity of approximation

One can examine the conditions used to obtain equation 4.37.

In the case of a smartphone, one can find typically $f \sim 10^{-3}$m, and $z_0 > 10^{-1}$m (minimum focusing distance) as well as $R \sim 10^{-3}$m and $\Phi_{PSF} \sim 10^{-6}$m which validates both conditions.
In the case of a DSLR, one can find typically f between 10mm and 300mm with $z_0 > 10 \times f$ (this admits the exception of "macro" lenses, which aim at imaging subject with a great magnification) as well as $R \sim 10^{-2}$m and $\Phi_{PSF} \sim 10^{-6}$m which validates both conditions.
The approximation tends to the exact formula towards infinity. However, at close distances, the error created can be around 10% in some cases (c.f. Fig. 4.8).

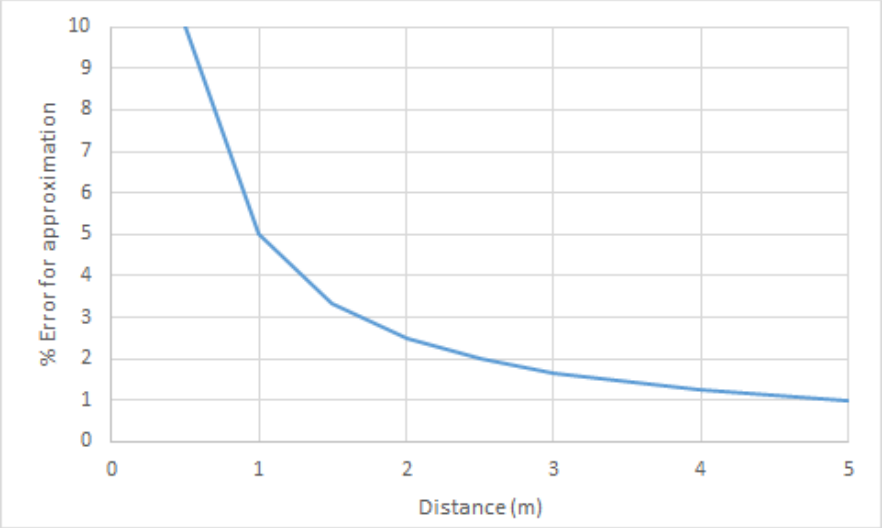


Figure 4.8: Error due to approximating formula 4.36 with formula 4.37, in the case of a DSLR system, with $f = 50$mm and an $f/4$ aperture, pixel size $\Phi_{px} = 4\mu$m.

4.5.1.4 Pixel size and pixel criterion

The geometrical approach to the axial resolution is pixel dependent. In this approach, the limit comes from the discrimination of a point corresponding to a perfectly focused image, from a PSF corresponding to a defocused image. The criterion for the minimum size of the defocused PSF varies from one author to another: Trouvé proposes a criterion of Φ_{px} (the size of a single pixel) and Blayvas-Kimmel-Rivlin (BKR) [5, 6] proposes $1.22\sqrt{\Phi_{px}^2 + (\frac{\lambda f}{2R})^2}$. This criterion might also vary depending on the nature of the image, as knowledge of the light source can enable sub-pixel sampling.

4.5.2 Diffraction Theory

Diffraction is the fundamental limit for a perfect lens. The lateral diffraction effect is well known as being an Airy disk pattern [7] and the Rayleigh criterion is usually adopted as the criterion for the resolution of an optical system.

Axial resolution suffers from a similar diffraction effect and the light intensity along the axis consequently behaves as $\text{sinc}^2(z)$. Using an analogy of the Rayleigh criterion, one can establish the axial resolution criterion as being the distance between the peak intensity and the first zero.

4.5.2.1 Computation of the Axial Intensity

The axial intensity can be computed in detail around the focal plane, as shown in Born and Wolf [8] with the valid assumption of Fraunhofer diffraction. With the Cartesian coordinate system (x, y, z) centred on the image plane, the intensity can be written along the optical axis as:

$$I(z) = \text{sinc}^2(\frac{z\pi}{2\lambda}(\frac{R}{f})^2)(\frac{\pi R^2|A|}{\lambda f^2})^2 \tag{4.38}$$

where A the amplitude of the light and λ the wavelength.

4.5.2.2 Resolution in the Image Space

The equivalent of the Rayleigh criterion, which states that the distance between the two nearest separable points should be at least equal to the radius of first zero of the intensity, is verified for $\text{sinc}^2(X) = 0 \Leftrightarrow X = \pi$, and thus:

$$\frac{\Delta_z \pi}{2\lambda}(\frac{R}{f})^2 = \pi \Leftrightarrow \Delta_z = \frac{2\lambda f^2}{R^2} \tag{4.39}$$

4.5.2.3 Resolution in the Object Space

From the resolution in the image space, it is possible to compute the resolution in the object space using the the Lens Formula. The resolution in the object space is of great interest as it provides the depth map estimation error.

Let's assume that the object is located at a distance z from the lens of focal length f and radius R, of which the image is formed on the plane z'.
Assuming that the axial resolution around the plane z' is Δ_z in the image space, then the object points z and z_0 in the object space can be obtained from the Lens Formula:

$$\begin{cases} z = \frac{fz'}{f-z'} \\ z_0 = \frac{f(z'+\Delta_z)}{f-(z'+\Delta_z)} \end{cases} \tag{4.40}$$

The resolution is consequently[1]:

$$z - z_0 = \frac{f^2\Delta_z}{(f - z' - \Delta_z)(f - z')} \tag{4.41}$$

Replacing z' with its expression, the resolution formula leads to A.15:

$$\delta = \frac{\Delta_z}{1 - \frac{2z}{z-f} + \frac{z^2}{(z-f)^2} - \frac{\Delta_z}{f} + \frac{\Delta_z z}{f(z-f)}} \tag{4.42}$$

Inserting the image-space resolution from equation A.12 into equation A.15, one can obtain the following axial intensity resolution equation in the case of diffraction limit:

$$\delta = \frac{\lambda f^2}{(\frac{R^2}{2} - \lambda f) + \frac{\lambda f z - z R^2}{z-f} + \frac{z^2 R^2}{(z-f)^2}} \tag{4.43}$$

This formula confirms that sensors with larger f and R produce shallower depth-of-field.

4.5.3 Comparison of both methods

4.5.3.1 Summary of formulas for Geometrical Optics and Diffraction theories

Table 4.2 presents a summary of the axial resolution formulas obtained with geometrical optics and diffraction theory, along with the assumptions made for each formula.

4.5.3.2 Comparison of the errors

The geometrical optics formula behaves similarly for both Trouvé and BKR formula, with the slope being the only variable. The diffraction behaviour differs mainly at close distances, yet remaining compatible with the geometrical optics results. In both Fig. 4.9 and Fig. 4.10, the diffraction theory predicts a greater error than the geometrical optics formula.

[1]N.B.: The approximation $\Delta_z \approx 0 \Leftrightarrow z - z_0 \approx \frac{f^2\Delta_z}{(f-z')^2}$ is valid only for a close distance to the lens.

Table 4.2: **Summary of formulas**

FORMULA	VALUE
Geometrical Optics *Trouvé* • No lateral diffraction • No Aberrations • Pixel size dependent *BKR* • Lateral diffraction • Uniform Aberrations • Pixel size dependent	**Exact**: $\delta = \frac{\Phi_{PSF} z_0^3}{f(z_0+f)(2R-\Phi_{PSF} f)}$ **Approximation**: $\delta \approx \frac{\Phi_{PSF} z_0^2}{2Rf}$ $\Phi_{PSF} = \beta\Phi_{px}$, $\beta = 1$ is an empirical parameter $\Phi_{PSF} = 1.22\sqrt{\Phi_{px}^2 + (\frac{\lambda f}{2R})^2}$
Diffraction Formula • Axial diffraction optics • No lateral diffraction • No Aberrations • Pixel size independent	$\delta = \frac{\lambda f^2}{\frac{R^2}{2} - \lambda f + \frac{\lambda f z - z R^2}{z-f} + \frac{z^2 R^2}{(z-f)^2}}$

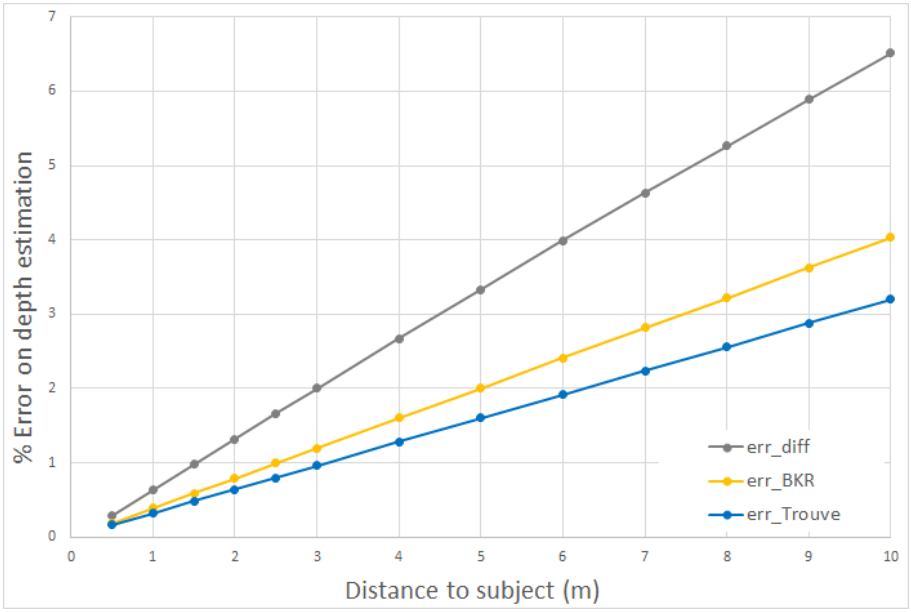


Figure 4.9: Percentage of error on the estimation of the object distance from the camera, for a DSLR system, with $f = 50$mm and an $f/4$ aperture, $\Phi_{px} = 4\mu$m.

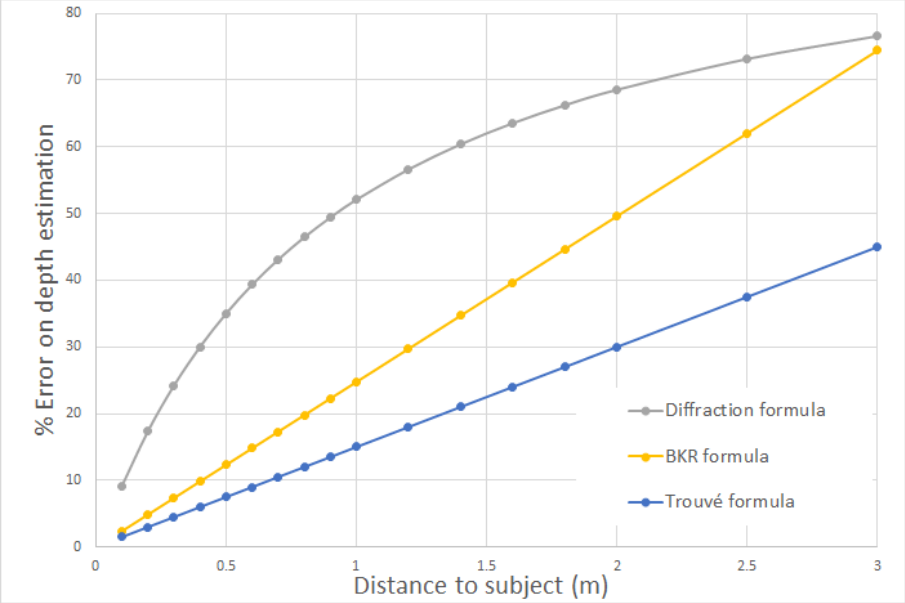


Figure 4.10: Percentage of error on the estimation of the object distance from the camera, for a smartphone system, with $f = 4$mm and an $f/2$ aperture, $\Phi_{px} = 1.2\mu$m.

4.5.4 Criterion to choose between Diffraction and Geometrical formula

The diffraction formula provides a pixel independent result for the given optical parameters of the lens system. In order to choose the appropriate formula, an equivalent pixel size C can be used:

$$\begin{cases} \Delta_z = \frac{2\lambda f^2}{R^2} \\ \frac{C}{R} = \frac{\Delta_z}{f} \end{cases} \Rightarrow C = \frac{2\lambda f}{R} \tag{4.44}$$

The comparison between C and the pixel size Φ_{px} provides a criterion for the limiting factor of the resolution. If $C > \Phi_{px}$ the diffraction effect is predominant over the pixel size and the diffraction formula should be used. In practice, the diffraction formula is relevant for small apertures, for instance for an f-number greater than f/2.8 for a full frame sensor, assuming a $\Phi_{px} = 4\mu$m pixel size. In the particular case of Fig. 4.9, axial diffraction effect is the limit and not pixel size.

For mobile phone imaging, the proposed diffraction formula should be used instead of the geometrical optics approach which over-estimates the axial resolution achieved, resulting in an erroneous depth map precision.

4.5.4.1 Comparison between lateral resolution limit and axial resolution limit

The lateral resolution limit provides the optimal pixel size for a given aperture such that the Rayleigh criterion is respected. The criterion provided in equation 4.44 establishes an equivalent for the axial diffraction. The optimum pixel sizes for the lateral and axial resolution differs for a given aperture, and lateral resolution requires smaller pixels than axial resolution.

4.5.4.2 Aberrated Geometrical Resolution

In a more general case, one would expect the system to have aberrations. The approach proposed by BKR, using an optical-aberration point-spread function is a first step to take aberrations into consideration. However, this method is based on the assumption that the optical aberrations are uniform both along the z-axis and across the image, which is wrong in general. Optical aberrations (e.g. coma, astigmatism,...) are field variable and most systems exhibit variable optical properties at different focusing distances. It appears that an encircled or ensquared energy approach are the only ways to estimate precisely the influence of aberrations. In the case of spherical aberrations, Seidel polynomials or wavefront theory can be used to estimate the impact of the aberration on the depth estimation, as shown by Miks and Pokorny [9].

4.5.5 Practical application for the case of smartphone photography

From the axial resolution limit, it is interesting to compute the resolution in the object space to get a better understanding of the challenges of small sensors. In table 4.3 the object space axial resolutions for a typical smartphone lens (focal length $4mm$) for various apertures and a $50mm$ lens are compared. Depth mapping with the information from a single camera with a small sensor is limited by this axial resolution.

Focal length and Aperture	Axial resolution (image space)	Axial resolution at 1m (object space)
4 mm f/1	0.55μm	3.3cm
4 mm f/2	2.2μm	12cm
4 mm f/4	8.8μm	35cm
4 mm f/8	35.2μm	68cm
50 mm f/2	2.2μm	0.1cm

Table 4.3: Axial resolution for several focal lengths and apertures. The diffraction formula is used for the smartphone lens of focal length $4mm$ and the geometrical optics formula is used for the lens of focal length $50mm$.

References

[1] JR Jenness. The modulation transfer function of an optical system with a rotationally symmetric circle of confusion. *Applied Optics*, 5(3):421–423, 1966.

[2] Michel Gondran and Alexandre Gondran. Energy flow lines and the spot of poisson–arago. *American Journal of Physics*, 78(6):598–602, 2010.

[3] Jeff Conrad. An introduction to depth of field. *Large Format Photography*, pages 1–9, 2006.

[4] Pauline Trouvé-Peloux, Frédéric Champagnat, Guy Le Besnerais, and Jérôme Idier. Theoretical performance model for single image depth from defocus. *Journal of the Optical Society of America A*, 31(12):2650–2662, Dec 2014.

[5] Ilya Blayvas, Ron Kimmel, and Ehud Rivlin. Role of optics in the accuracy of depth-from-defocus systems. *Journal of the Optical Society of America*, 24(4):967–972, Apr 2007.

[6] Ralf Blendowske. Role of optics in the accuracy of depth-from-defocus systems: comment. *Journal of the Optical Society of America*, 24(10):3242–3244, Oct 2007.

[7] J. W. Goodman. *Introduction to Fourier optics*, volume 1. Roberts and Company Publishers, 3 edition, 2005.

[8] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, 1970. Fourth Edition, chapter VIII : Elements of the theory of diffraction, section 8.8 : The three dimensional light distribution near focus.

[9] Antonín Mikš and Petr Pokorný. Spherical aberration of an optical system and its influence on depth of focus. *Applied Optics*, 56(17):5099–5105, Jun 2017.

Chapter 5

Review of Passive Depth-Estimation Methods

> Can one ask photography to render all the variety experienced in directly looking at objects? Is it possible to capture in a photographic print things such that one can observe as if it was an open window on the real world? The answer seems to be yes; one can expect from photography much more than from a man's hand.
>
> M.G. Lippmann, 1908

The quest for an accurate representation of the world is centuries old, as discussed in Chapter 2. This quest is to not only capture 2D images, but also to provide a sense of depth. In visual arts, some perspective can be found in Roman paintings (see figure 5.1) or mosaics [1]. The paintings from the medieval period are often religious and mostly ignore perspective. The Italian architect Filippo Brunelleschi is credited with introducing perspective in Renaissance paintings [2]. His theoretical work on perspective led to wide use by painters and artists, such as Michelangelo (see figure 5.2). Auguste Berthier experimented in 1896 with kinegrams to bring a sense of depth to images [3]. From his work, some drawing books are made, like the one shown in figure 5.3. As for photography, the research by G. Lippmann on integral imaging in 1908 [4] was the first attempt to bring depth to the photographic process. The set-up described by Lippmann uses a form of micro-lens array. The discovery, however, is disputed with E. Estanave who proposed a prototype of the same idea about two weeks after Lippmann's article was submitted [5]. The process patented by Estanave has been used commercially as early as the 1930s, mainly by the photographer M. Bonnet.

A depth map is a grayscale image that contains the distances between the viewpoint and the surfaces from the objects of the scene. We propose here an overview of the existing methods used to

Figure 5.1: Roman still life from the home of Julia Felix, Pompeii. The painting shows some examples of shades and volume perspective. Source: Wikimedia Commons (Public Domain)

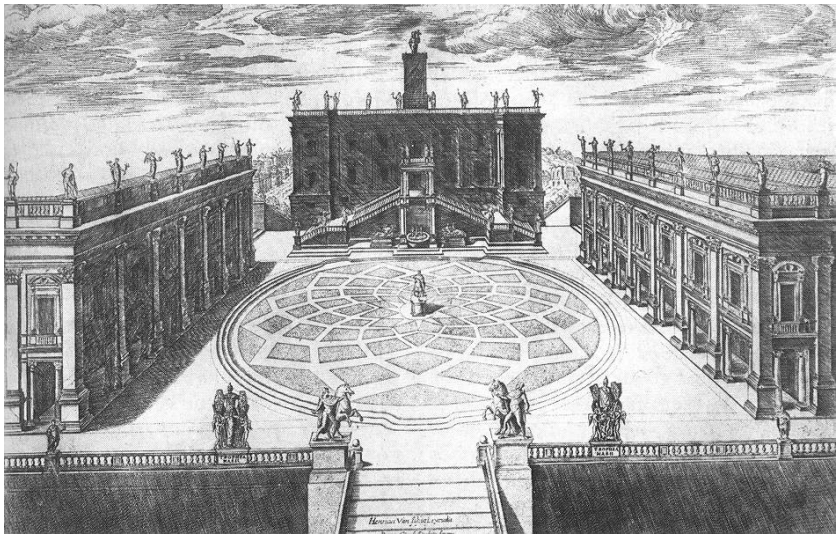


Figure 5.2: Drawing of the Capitoline Hill by Michelangelo, circa 1540. Source: Wikimedia Commons (Public Domain)

achieve depth map estimation. Overall, the typical approach to depth of field management relies on a two-step process.

Firstly, a system is used to estimate the depth map of the scene. This system can take many forms (the following sections will provide an overview of the main methods). Ideally, the depth map has infinite axial resolution and a lateral resolution at least similar to the one of the images (in some cases, such as Time-of-Flight (ToF), super-resolution techniques are employed [6]).

Secondly, a set of image processing algorithms is used to either refocus, extend the depth of field

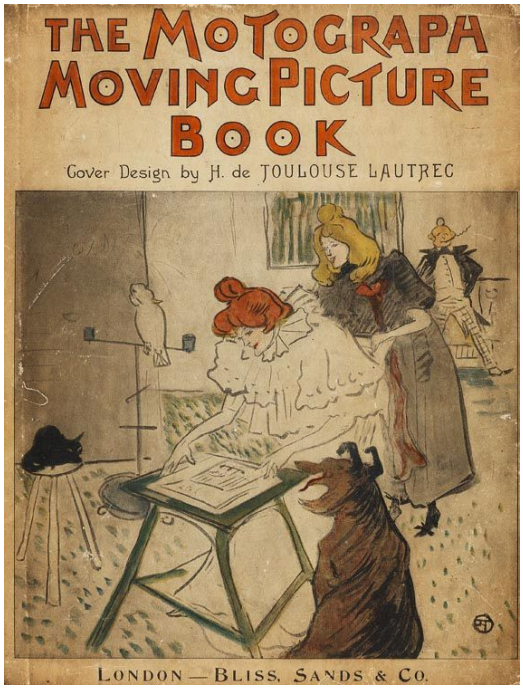


Figure 5.3: Cover of a kinegram book from 1896. The drawings from the cover are from Henri de Toulouse-Lautrec. Source: F.J. Vernay (Public Domain)

or reduce the depth of field. The results are usually limited by the quality of the depth map or the amount of noise in the image (for deconvolution algorithms).

5.1 Estimating the depth map

This overview mostly focuses on passive depth estimation methods. This choice is related to the slightly different use case for active illumination methods. However, the principle of active illumination is briefly mentioned and some references are provided.

5.1.1 Active illumination systems

At their core, illumination methods rely on some light projection (often in the near infrared) and analysis of the reflected light. Active illumination systems can be classified into two main categories: structured light methods and non-structured light methods.

5.1.1.1 Structured illumination methods

The structure of the projected light, i.e. the spatial variation of the projected light intensity, can be used to estimate how much a given object distorts the light pattern [7]. For instance, if a rectilinear grid pattern is projected onto a round-shaped object, then the shape of the object will appear to be illuminated with curved light strips. Flash illumination can be considered [8, 9] as

the simplest form of projected light, although the pattern is plain. Geng [10] describes the state of the art of the main patterns used and gives some guidance on their implementation. A structured illumination method has been engineered for iPhone Face ID technology [11] where a $30,000$ dot projector is used to estimate the depth features of a user's face.

5.1.1.2 Other methods

Some alternatives measure the time for light pulses to reach objects in the scene and to be reflected back. This technique can take several forms such as LIDAR [12, 13] or Time-of-Flight (ToF) [14–17]. Miniature Time-of-Flight sensors are used in a few smartphones; for sample the Samsung Galaxy S10 use a time of flight camera for depth estimation [18].

5.1.2 Stereo-vision

Camera arrays are a common technique used to achieve depth estimation, and their principle is very similar to that of human vision.

5.1.2.1 Set-up and calibration

Dual-camera systems are inspired by the human vision system [19, 20]. Two cameras with a similar field of view are placed side by side, separated by a distance b from lens centre-to-centre, called the baseline. The baseline introduces a shift for an object point on the two images produced. The estimation of the shift and a triangulation method can reconstruct the depth of a given pixel. The map of the estimated shift for all pixels is refereed to as the disparity map.

The use of stereo cameras long preceded the introduction of digital sensors and image processing. Figure 5.4 shows an example of a stereo-film camera. During the film era, stereo-cameras were used for visual stereo viewers but not to compute a depth map. The most challenging aspect of stereo-system manufacturing is to align both cameras (x-axis, y-axis, pitch, roll, yaw) and focus the lenses together [21]. Most systems use an image registration step before estimating the disparity map to compensate for possible misalignment.

In order to align the images from both cameras, one needs to estimate the transformation matrix E between both points of view. The intrinsic matrix K for each camera and the fundamental matrix F (translation between both cameras) are linked to the matrix E through the equation:

$$\mathbf{E} = \mathbf{K}^T \times \mathbf{F} \times \mathbf{K} \tag{5.1}$$

The matrix $\mathbf{K}$ is defined using the physical parameters of the camera, i.e. lens focal length f, pixel densities k_u and k_v (usually, both cameras have the same sensor and thus $k_u = k_v$) as well as the

Figure 5.4: A Soviet stereo-camera from the 1960s. The set-up uses 120 film and records 2 side by side images. The third lens is used for framing and focusing. Source: Bilby (CC by SA 3.0)

position of the image x_0 and y_0.

$$\mathbf{K} = \begin{pmatrix} k_u f & 0 & k_u x_0 \\ 0 & k_v f & k_v y_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.2}$$

The transformation matrix (a combination of some rotation matrices and a translation matrix) can be calibrated with a set of targets.

For specific tasks, the alignment is done directly using hardware. Such is the case for the cinematic industry: lenses are produced as a pair to ensure similar performance/aberrations [22] and the stereo-vision rigs are aligned precisely ahead of the film shooting [23, 24]. As a side-note, the cinematic industry commonly uses a beamsplitter to accommodate for the camera size while keeping a reduced baseline (about the same as the human inter-pupilary distance). In the smartphone industry, the cameras are usually placed side-by-side with no yaw rotation.

5.1.2.2 Depth map estimation from stereo-cameras

The depth map estimation process is a four-step process [25]:

First, some points of interest are found in both images. These points can be obtained with algorithms such as SAD [26] or SIFT [27]. Figure 5.5 shows an example of SIFT features on an image. These points are usually detected with some gradient-based filters.

Second, a selection and matching step is carried out. The points that are detected in only one of the two images are removed in order to avoid mismatches. All the detected points are matched with points as close as possible in the other image. The points that cannot be matched are removed from the pool.

Third, the disparity map is estimated for all the points: the distance between the points in both the left and right image is used to compute the depth for the set of points, knowing the parameters of the system.

Fourth, since the set of points detected is not dense and regular enough to cover all parts of the image, a regularisation step is needed to interpolate depth information for the areas of the image where extracted features are sparse. Segmentation of the RGB image can be used for this purpose.

Figure 5.5: Example of SIFT features for an office image. Source: Engjw (CC by SA 3.0)

5.1.2.3 Limits for stereo-vision systems

In the smartphone field, dual cameras provide a depth-of-field reduction based on the two cameras [28, 29]. In fact, for such applications, the size of the baseline is critical. A large baseline is preferable for increased precision, at the expense of the depth-map coverage: the sides of the two frames don't match since both lenses have a parallel optical axis. This requires the image to be cropped during the registration step [30]. The trade-off between depth precision and the need to crop the image usually favours a small baseline with consequently reduced depth precision.

In the particular case of smartphones, the two cameras might be slightly different in order to offer additional features. A practical example is the color + monochrome dual-camera configurations implemented to collect more light with the monochrome camera [31]. The different spectral response of the second camera might degrade the feature matching resolution.

5.1.2.4 Larger camera arrays

The use of larger than stereo-vision, multi-camera systems was introduced in 2005 by Levoy [32]. The Huawei P20 Pro, released in 2018 by Huawei [33], is the first phone equipped with a tri-camera rear system. Several systems have since been launched featuring 3 to 5 cameras. The largest array commercially available is the Light L16 [34].

The PiCam array of cameras proposed by Pelican Imaging in 2013 introduced a 16-sensor array for smartphones [35], but was never commercialised. From the information available on the prototype cameras, all 16 cameras shared the same CMOS sensor and featured a single element lens. The recent Light L16 also uses 16 cameras [36], however each of them is using a dedicated CMOS sensor and a 6 element lens design. The cameras are organized into 3 groups of focal lengths : 28mm, 70mm and 150mm. The images from the different modules overlap and are combined to produce high resolution images (up to 80 MPx) with image-processing-based zoom and depth of field management.

5.1.3 Depth from Defocus

5.1.3.1 General case

Defocus entails a blur in the out-of-focus part of a given image. The amount of blur highly depends on the lens and sensor parameters as explained in Chapter 4. More importantly, for a given system, the amount of blur in an image is related to the distance between the object and the plane of focus.

The shape of the out-of-focus blur can be quite complex because of the influence of optical aberrations and possible mechanical vignetting. For a circular aperture, an Airy pattern is expected, which is often approximated with a Gaussian blur. As a consequence, most algorithms [37–48] use a Gaussian blur model.

Estimating depth from monocular images is a form of deblurring problem since the aim is to estimate the amount of blur for each pixel. This is an ill-posed problem [49]. Since some noise can affect the image obtained, one should also apply regularization. In the literature, there are several techniques to estimate the blur such as edge estimation [50] and Fourier spectrum analysis [51].

A major caveat also needs to be addressed ; in most cases, it is impossible to determine if an object is in front or behind the focal plane. Both front-focused and back-focused parts of the image are blurred. In the general case, for a given blur size, there are two possible distances (respectively in front of and behind the plane of focus). The blur size however is not symmetrical around the focal plane. The blur is significantly larger closer to the sensor as opposed to the blur toward infinity [52] as shown in figure 5.6. In the particular case of a blur size greater than the one expected for a point at infinity, one can determine the object position uniquely. Also, for objects close to the camera lens, the blur size can prove too large for depth estimation.

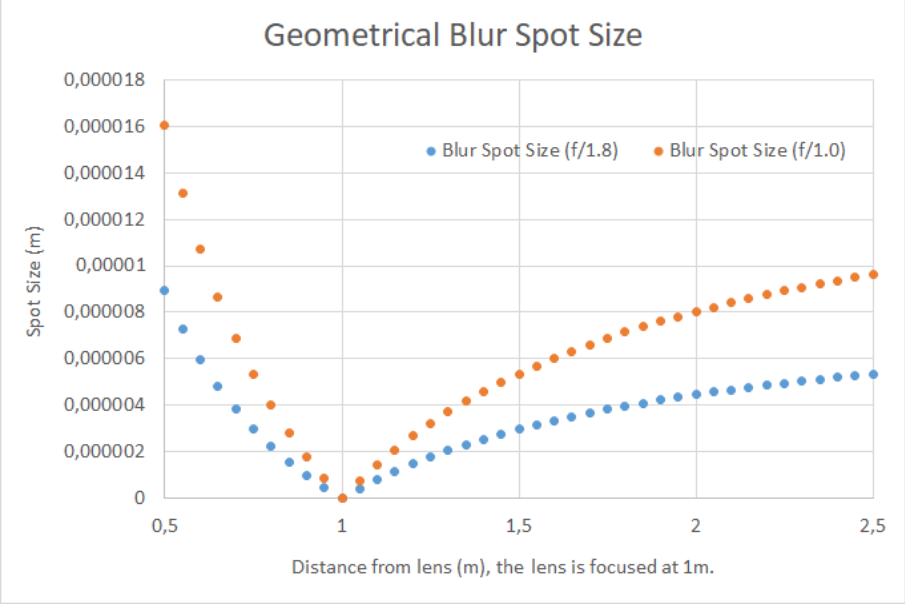


Figure 5.6: Geometrical Blur Size computed for two lenses with focal length $f = 4$mm and apertures f/1.8 and f/1.0 respectively, focused at 1m. The formula used for the computation was proposed in [53].

Since Depth from Defocus is currently the dominant monocular depth-estimation method for smartphone devices, it is of great importance to point out some of its limitations. Depth from defocus only works if the blur size can be estimated. This is only possible if the imaged scene contains high-frequency spatial information. A typically challenging depth estimation is that of a white wall. Since the area is uniform and does not contain texture information, it is impossible to estimate how defocused it is.

5.1.3.2 Depth from image stack

In order to provide a more robust estimation of the depth map, one might consider using a series of images instead of a single photograph. Typically, the camera is programmed to take a burst of images with several focusing distances [54]. This can be used together with some pixel coding

mechanism [55].

The image stack technique is commonly used to produce some all-in-focus images [56] but can also provide a depth map.

Alignment of the images is a key step in this process since the refocusing affects the lens magnification (the effect is known as focus breathing) and sometimes the user moves the camera slightly during the focusing time. Since the points of interest from one image might be blurred in another image, the use of features for alignment is sometimes impossible. In such cases, a Euclidean-Transform estimation provides good results [57].

Out of the image stack, one must then estimate the relative sharpness of each pixel of the image among the available images. To do so, it is possible to blur the images with a Gaussian filter prior to detecting the edges using a Laplacian edge detection kernel. The use of the Gaussian blur removes some of the artefacts created by high frequencies. Equation 5.3 shows a typical 3×3 Laplacian kernel that can be convolved with the image to carry out this filtering.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \tag{5.3}$$

Once the sharpest areas of each image have been identified, for a given area of an image I, one can estimate their depth D using the formula:

$$D = I\frac{f_{max} - f_{min}}{N} \tag{5.4}$$

With f_{max} the maximal focusing distance, f_{min} the minimal focusing distance and N the number of images in the focal stack.

Repeating the process for all the areas of the image theoretically leads to a full depth map. In practice however, areas with less texture might not be identified properly and noise can reduce the accuracy. As for the texture-less areas, it is common to extrapolate the depth information from the neighbouring regions for which the depth is known using image segmentation. For the noise issue, it is common to denoise the images before estimating the depth. Wlodek et al. [57] report observing some artefacts unless the images are denoised and smoothed, and suggest the use of bilinear filters.

5.1.3.3 Shape of the aperture and coded apertures

In the general case, the aperture of the lens is either a disk or a regular convex n-gon. In the case of an n-gon, the usual range is $n \in [5, 10]$ in commercial photographic lenses. Although, there is a

smartphone system with $n = 4$ and some film-era lenses with up to $n = 24$ aperture blades. Table 5.1 provides examples of lenses with various numbers of blades.

Polygon type	Number of blades	Lens (release date)
Circle	0	Most smartphones
Square	2	Samsung Galaxy S9 (2018)
Pentagon	5	Canon EF 50mm f/1.8 II (1990)
Hexagon	6	Nikon 58mm f/1.4 (1959)
Heptagon	7	Canon EF 50mm f/1.8 STM (2015)
Octogon	8	Canon EF 50mm f/1.2 L (2006)
Nonagon	9	Nikon 58mm f/1.2 (1977)
Decagon	10	Leica 50mm f/0.95 Noctilux-M (2006)
Dodecagon	12	Carl Zeiss Jena Tessar 50mm f/2.8 (1955)
Pentadecagon	15	Jupiter-9 85mm f/2.0 (1951)
Heptadecagon	17	Carl Zeiss Jena Biotar 58mm f/2.0 (1946)
Icositetragon	24	Novoflex Noflexar 400mm f/5.6 (1955)

Table 5.1: Geometry of the aperture for several lenses.

The purpose of the aperture mechanism is to reproduce, as closely as possible, a circular aperture which is desired by most photographers for the smoothness of the background blur. This explains why most of the blades are rounded, especially for portrait lenses for which the blur quality is critical. To the best of our knowledge, only a few experimental lenses offer the possibility to use non-circular apertures, such as Lomography's Petzval lens.

In the case of a coded system however, the use of a circular aperture is not always the optimal choice depending on the application. A variety of masks has been proposed for visible and other wavelengths such as gamma rays shown on figure 5.7. Levin suggests a pseudo-random grid shaped pupil for all-in-focus application [58]. Zhou et al. suggest that instead of taking an image with a single coded aperture for depth estimation and out-of-focus deblurring, one should use two apertures. The 2009 paper [59] explains how to achieve this using a gradient search and the 2011 paper [60] provides with additional information about the performance of the pupil. Haim et al. [61] propose a phase mask optimized using deep learning techniques and claim to out-perform depth from monocular images. Wang et al. [62] implement an adaptive coded aperture with a two-step process : first a scan of the scene using a pinhole camera and second, a Liquid Crystal on Silicon (LCOS) is used to create the coded mask. Levin [64] investigates the analytical noise-related errors from several coded apertures. A general review of coded apertures is provided by Cieslak et al. [63].

The measured signal from a sensor follows the equation:

$$o = h \star s + n \tag{5.5}$$

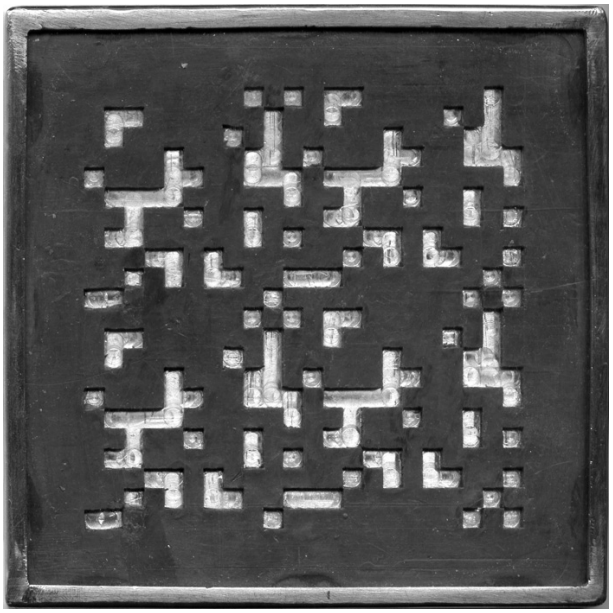


Figure 5.7: Sample of a coded aperture mask designed for a gamma-ray camera. Source SPECT (CC by SA)

where o is the output signal, h the known impulse response (in our case, the PSF), s the input signal and n the noise that affects the signal.

The deconvolution process consists of finding s from o. This can be done typically using either Wiener filtering or the Lucy-Richardson algorithm. Wiener filtering is optimized for Gaussian noise whereas Lucy-Richardson is designed for Poisson noise. In photographic applications, the Poisson noise mode is more suitable. For instance, Levin implements a Lucy Richardson deconvolution [58].

Since deconvolution, even knowing the PSF h, is an ill-posed problem, the deconvolved image is often looking unnatural with "ringing" artefacts. Some modified Lucy-Richardson algorithms have been proposed to address specifically this issue [65–67]. However, using some prior information, one can regularize the deconvolution and improve the output image. A regularization takes the form of an added term corresponding to a model.

$$s = \arg\min |h \star s - o|^2 + \lambda \sum_i \rho(\nabla s_i) \tag{5.6}$$

In this case, the choice of ρ depends on the prior. For a Gaussian prior, $\rho(x) = |x|^2$. For a sparse gradient prior $\rho(x) = |x|^{0.8}$. The sparse prior seems to be the most suitable for natural images [68].

5.1.4 Lensless techniques

The field of lensless imaging devices is limited to basic depth estimation or low resolution imaging tasks. Lensless imaging relies on a dictionary approach (i.e. a list of calibrated source points is

used in the algorithm) to recover an image from the signal measured on the sensor [69, 70]. Such systems can be used to estimate a coarse depth map using homothetic properties such as the scale and location of the the projected coded aperture on the sensor as shown in figure 5.8. Although lensless imaging provides partial information about the depth and point source of the signal [71] it cannot compete with lenses for photographic applications.

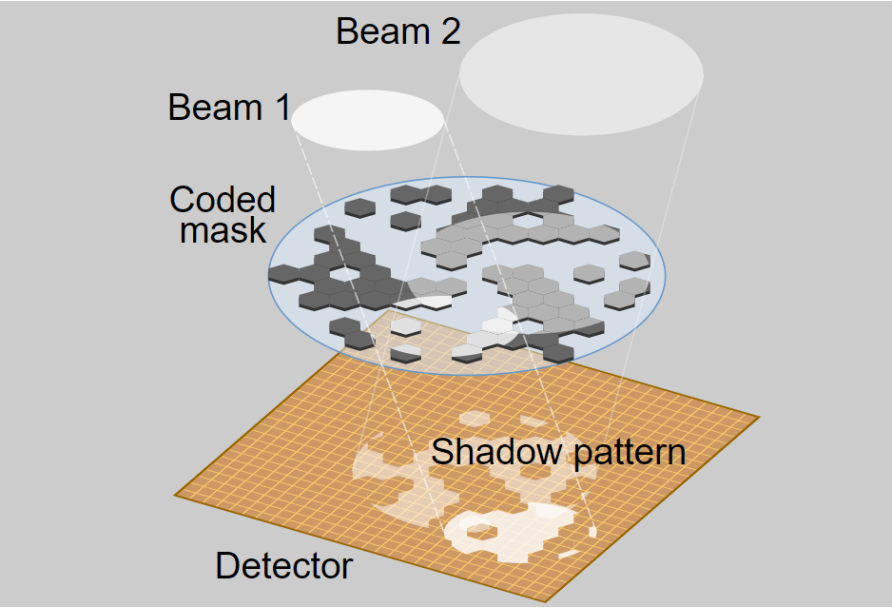


Figure 5.8: Principle of operation of a coded aperture system. Image inspired from HURA hexagonal coded aperture mask used in the SPI instrument of the INTEGRAL space telescope. Source: Cmglee (CC by SA 3.0)

5.1.5 Deep Learning and depth map from monocular images

Depth-map estimation from monocular images using deep learning has received a lot of attention in recent years [72–76]. This method is data driven and as such, only few remarks can be made as for the performance and accuracy of the results.

Geiger et al. lists the following methods that are available to perceive depth [77]:

- Position in the image
- Occlusion
- Texture information
- Perspective
- Apparent size of objects

- Shading and illumination
- Defocus blur
- Aerial perspective (i.e. colour shift and contrast loss observed in very far away objects due to refractive index variations among the atmosphere).

5.1.5.1 Image datasets

There are three main datasets used for neural network training for depth estimation from natural images: KITTI [78], Make 3D [79] and NYUD-V2. Table 5.9 provides some information about each dataset.

Dataset	Size of dataset (number of images)	Annotation	Scene
NYUD-V2	407K RAW	Depth + Segmentation	Indoor
Make 3D	500K	Depth	Outdoor
KITTI	94K	Depth aligned with RAW data	Street

Figure 5.9: Information about the main datasets used for monocular depth estimation. RAW stands for an uncompressed file with 10 to 16 bits of information per pixel (whereas the JPEG format provides 8 bits).

For the standard KITTI dataset [78], the aerial perspective does not appear, the resolution is too low for the use of texture and blur information and the range of distances provides only limited perspective, occlusions and shading. Position in the image and apparent size of the object are the main cues available for machine-learning purposes. For the MonoDepth network proposed in [80], the vertical position seems to be the primary depth cue for the neural network since the typical image displays a car on a road and so the height of the car inside the image is related to its distance from the camera.

5.1.5.2 Limits

Despite offering very appealing results, deep learning methods suffer from severe drawbacks:

- Limited resolution: The KITTI dataset is made of images with a 0.7MPx resolution (1382×512px). This is sufficient for automotive applications but far from the requirements of the photographic community. If the machine learning methods are patch-based, this is not of critical importance. However, the patch-based approach can lower the performance of the machine-learning algorithm since the algorithm cannot estimate the depth information from the full image.
- Database: the quality of the database is limited by the diversity of real life situations it features. Building a high resolution database containing images for many different contexts is both crucially important and costly.

- Lack of transparency: The maximum depth map error is unknown. Contrary to traditional techniques relying on the laws of physics, there is no way to ensure that the maximum error for a given network falls in a specific range.

These limitations make deep learning depth estimation methods mostly suited for specific applications (e.g. autonomous driving) and should be combined with other methods to prevent errors.

5.1.6 PDAF sensor technology

The Phase-Difference detection Auto-Focus (PDAF) sensor technology was introduced in the early 2010s [81] and consists of subdividing each pixel into two sub-pixels; instead of having one square photodiode, two rectangular photodiodes are used. Because of the presence of the lenslet array, the two sub-pixels capture some coarse angular information about the light rays falling onto the sensor, and this additional information has only limited impact on the sensor noise [82]. The principle is illustrated in figure 5.10.

Theoretically, for an in-focus plane, the image formed by the left pixels is aligned with the image formed by the right pixels. However, for out-of-focus planes, a shift can be measured. If the two images are misaligned, the sign of the shift provides some information about the order of the focal plane with respect to the object. If the object is behind the focal plane, the shift of the left-pixels-image compared to the right-pixels-image is positive (the lights rays go away from the optical axis). A negative shift indicates that the object is closer to the camera compared to the focal plane. The magnitude of the shift can provide some coarse information about the distance from the focal plane to the object, although the magnitude is not linearly related to the distance.

Some improvements have been proposed to adjust the focus in a more reliable way despite the noise of the phase information. Chan et al. [83] demonstrate that the use of a statistical model to characterize the relation between the phase shift and the lens movement for several object depths can improve the reliability.

For smartphone sensors, PDAF pixels are used in numerous devices such as the Google Pixel phones [84]. Garg et al. provide an example of depth resolution for the PDAF technology as well as a deep-learning based method for estimating depth [85].

5.1.7 Micro-lens Array

5.1.7.1 Principle

As mentioned in the introduction of this chapter, the first light-field imaging idea is commonly attributed to G.M. Lippmann in his work of 1908 [4]. The core idea of micro-lens array systems

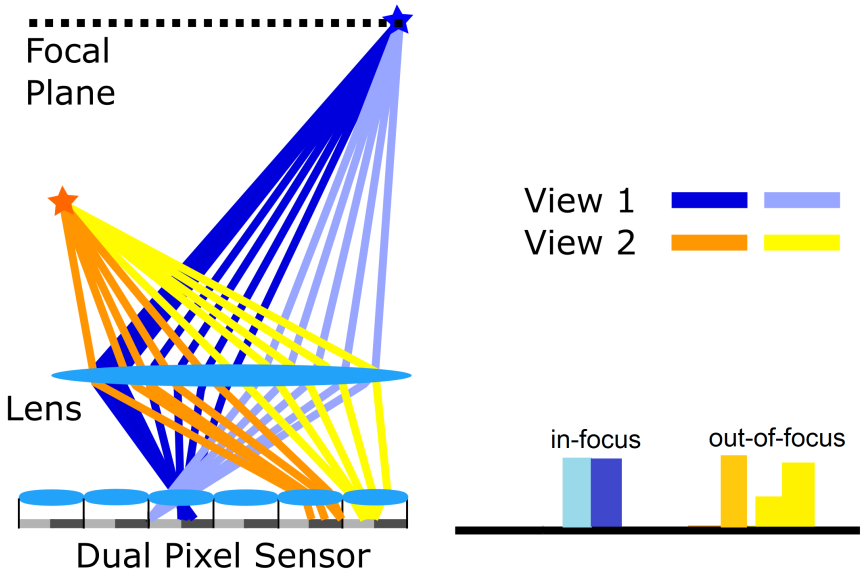


Figure 5.10: Detection of an out-of-focus point using PDAF sensor technology. The light rays in blue converge to a single pixel whereas the light rays in orange show some shift, which corresponds to a defocus.

is that by placing an array of small lenses on the light path before it reaches the sensor, one can observe the light ray angles. This information can later be used to compute the light-field and recover some information about the depth of the scene.

This can be done if several conditions are met:

- A micro-lens array is placed between the sensor and the lens.
- The size of an individual micro-lens covers several pixels i.e. the number of pixels is far greater than the number of lenses.
- The f-number of the micro-lenses and of the main lens are matched (there is no need for additional optical elements since the lens is close to the sensor). If the main lens aperture is too fast, the sub-images are too large and overlap (creating artefacts in the reconstruction process). If the main lens aperture is too slow, the sub-images are too small and some resolution is lost.

There are two possible positions for the micro-lens array as shown in Figure 5.11. [86–88]

5.1.7.2 Methods

The main advantage of lightfield based methods being that no camera calibration is required to proceed to depth estimation. As for estimating depth from a light-field image, three main techniques are available [89].

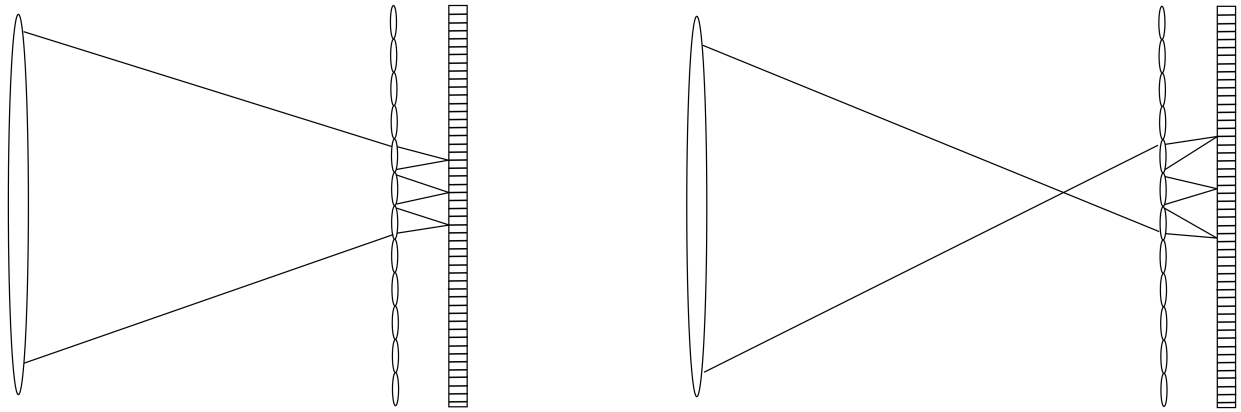


Figure 5.11: Two possible positions for the micro-lens array. The first one is very similar to the Shack-Hartman wavefront sensing configuration.

The **stereo-vision approach** provides only a coarse depth map based on sub-aperture images observed with a short baseline. The small size of the baseline is determined by the micro-lens array, which at around 10μm leads to very poor depth estimation. Such sub-aperture images have been used in Ng's work [88].

The **epipolar-plane image (EPI)** and some theoretical properties of the light field can be used to achieve a better depth estimation.

In recent years, **machine learning methods (such as Convolutional Neural Networks) have been used to estimate depth from the 4D lightfield**. This method is data driven and uses end-to-end learning from the 4D lightfield to the 2D depth map.

In 2013, Wanner [90] shows the need for a dataset for algorithm benchmarking purposes. In 2016, Rerabek [91] presented 118 light field images captured with a Lytro Illum device. Each image is coded with 10 bits and has a 5368×7728 pixels resolution. Unfortunately, such as database is too small for most deep learning methods. Wang [92] claims that patch-based Convolutional Neural Network could benefit from light field information for classification.

For all depth estimation techniques, a regularization step is needed to handle the noise introduced by the sensor. Regularization methods can involve Markov-Random-Field (MRV), Total-Variation (TV), Active Wavefront Sampling (AVS) and Total-Generalized-Variation (TGV).

5.1.7.3 Practical implementations

Although promising on paper, micro-lens arrays sacrifice a large amount of pixels for angular resolution. Light-field cameras refer to the number of pixels from the sensor as a number of rays, and use the number of pixels as a metric for the output image resolution. The limits of this trade-off will be highlighted later. Limiting the spatial resolution for depth estimation greatly limits the interest of micro-lens arrays for the photographic community. For existing devices the resolution offered is typically an order of magnitude (i.e. 10 times) smaller compared to state of the art sensors.

For this reason, micro-lens arrays are used more in industrial applications. Some commercial options have been made available for the photography market, such as the Lytro devices. The commercial success was limited. The first device, released in 2012, only offered 1MPx images (from a 10 Megaray sensor) and the higher-end Lytro Illum delivered 4MPx images (from a 40 Megaray sensor) in 2014. Since the lens aperture needs to match the micro-lens array aperture, the Illum only offered a fixed lens (9.5mm to 77.8mm zoom with an f/2.0 aperture).

5.2 Using the depth map information for digitally adjusting the depth and focus

Once the depth map is estimated, and assuming the resolution is sufficient both on the lateral and axial dimensions, the image can be processed with the intent of reducing the depth of field, extending the depth of field or refocusing.

5.2.1 Extending the depth of field

Extending the depth of field from a stack of images is extensively discussed in the scientific literature. Anish et al. [**?**] survey the image fusion algorithms for image stacks. Liu et al. [**?**] show that a CNN-method can be used to fuse the images. However, the colour filter array can produce, thanks to longitudinal chromatic aberration a shift between the images recorded in the red, green and blue channel. The use of this colour information for depth of field extension is discussed by Tisse et al. [**?**].

If depth of field can be extended, some applications no longer require a focusing mechanism for the lens. This topic is discussed by Bakin et al. [**?**] with a close focusing application in mind (for bar code readers of business cards scanning).

5.2.2 Refocusing

Refocusing can be considered as the combination of a depth of field extension and a depth of field reduction for a different plane of focus. The method proposed by Bando et al. [**?**] relies on the generation of an all-in-focus image and then the selective blur of the image. Zhang et al. [**?**] use the same principle, but deblur the image with a blind deconvolution method.

Refocusing is also used with light field cameras: the light field can be used to extract the image corresponding to another depth plane thanks to epipolar geometry [93, 94]. The resolution remains unchanged regardless of the refocusing. In order to refocus the image, most of the techniques go through an all-in-focus image that is selectively blurred to obtaine an image simulating a focus at a different depth.

5.2.3 Reducing the depth of field

To a good approximation, one can simulate the defocus blur with simple Gaussian blur. In the case of a system with a circular aperture and free from aberrations, the theoretical defocus blur is a Poisson spot.

Blurring the image is not as simple as one might think as the blur applied at a given depth can spread to sharp areas of the image. Edges between defocused areas of the image and the plane of focus are especially critical for digital blur. Colours can also be affected in the blurring process.

Several methods have been proposed that can be classified into three main categories: scattering-based methods, gathering-based techniques and layered methods [95]. The scattering method diffuses the intensity of each given pixel to its neighbours using a circle of confusion metric. The gathering based method uses filtering on groups of pixels. The layered technique computes the blur using several scales, such as a pyramid approach [96]. Alternatively, Zhang et al. [97] propose a method to take into account the occluded pixels from the background in order to avoid mixing the background blur with the frontground objects.

It is not obvious that a simulated blur can match the aesthetics of a wide aperture lens. As such, the work proposed by Hauser et al. [98] introduces a systematic assessment method for the quality of computational bokeh and offers a comparison with an f/2.8 DLSR lens. For a given blur size, at least four elements have to be measured:

- Quality of the background/foreground segmentation in the image: all the out-of-focus elements must be out-of-focus, all the in-focus elements must be in-focus.

- Blur gradient: the ability to progressively blur the image along a transition from an in-focus to an out-of-focus area.
- Noise consistency: using some form of blur on a noisy image changes the noise properties. The use of a selective blur can result in inhomogeneous noise texture across the image, which can be disturbing for a viewer.
- Repeatability: since the process is image-processing-based, the depth map can vary from one image to another, depending for instance on lighting conditions.

5.2.4 Conclusion

The methods reviewed in this chapter provide a better understanding of the state-of-the-art work available for depth of field managment. Some methods are of no practical use for photographic application (e.g. lensless imaging) whereas others are already implemented (PDAF pixels). The passive depth estimation methods require some form of image processing, which can be challenging for miniature cameras because of the SNR.

References

[1] Kevin R Brooks. Depth perception and the history of three-dimensional art: Who produced the first stereoscopic images? *i-Perception*, 8(1):2041669516680114, 2017.

[2] Denis Favennec and Emmanuel Riboulet-Deyris. Douce perspective. *Une histoire de science et d'art, Ellipses, Paris*, 2007.

[3] Auguste Berthier. Images stéréoscopiques de grand format. *Cosmos*, 34(590):205–210, 1896.

[4] M. G. Lippmann. Epreuves réversibles- photographies intégrales. *Bulletin de la Société Belge d'Astronomie*, 13:245–254, 1908. Comptes-rendus, vol 146.

[5] Jean-Marc Fournier. La photographie intégrale.

[6] Sebastian Schuon, Christian Theobalt, James Davis, and Sebastian Thrun. High-quality scanning using time-of-flight depth superresolution. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–7. IEEE, 2008.

[7] Markus Rossi. 3 types of 3d sensing for smartphones and self-driving cars. Available at `https://spectrum.ieee.org/transportation/sensors/3-types-of-3d-sensing-for-smartphones-and-selfdriving-cars`, 2019. MWC Barcelona.

[8] Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. In *ACM transactions on graphics (TOG)*, volume 23, pages 679–688. ACM, 2004.

[9] Changyin Zhou, Alejandro Troccoli, and Kari Pulli. Robust stereo with flash and no-flash image pairs. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 342–349. IEEE, 2012.

[10] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011.

[11] Sam Costello. How to use face id on your iphone ? Available at `https://www.lifewire.com/face-id-4151714`, 2020. Lifewire, 29-01-2020.

[12] Michael Himmelsbach, Andre Mueller, Thorsten Lüttel, and Hans-Joachim Wünsche. Lidar-based 3d object perception. In *Proceedings of 1st international workshop on cognition for technical systems*, volume 1, 2008.

[13] Clément Mallet and Frédéric Bretar. Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of photogrammetry and remote sensing*, 64(1):1–16, 2009.

[14] Felix Heide, Wolfgang Heidrich, Matthias Hullin, and Gordon Wetzstein. Doppler time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(4):36, 2015.

[15] S Burak Gokturk, Hakan Yalcin, and Cyrus Bamji. A time-of-flight depth sensor-system description, issues and solutions. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 35–35. IEEE, 2004.

[16] Jiejie Zhu, Liang Wang, Ruigang Yang, James E Davis, et al. Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps. *IEEE transactions on pattern analysis and machine intelligence*, 33(7):1400–1414, 2010.

[17] Yan Cui, Sebastian Schuon, Derek Chan, Sebastian Thrun, and Christian Theobalt. 3d shape scanning with a time-of-flight camera. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1173–1180. IEEE, 2010.

[18] What is tof camera technology on galaxy and how does it work ? Available at `https://www.samsung.com/global/galaxy/what-is/tof-camera/`. Samsung website, 2019.

[19] David Marr and Tomaso Poggio. A theory of human stereo vision. Technical report, Massachusetts Institute of Technology Cambridge Artificial Intelligence Lab, USA, 1977.

[20] David Marr and Tomaso Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 204(1156):301–328, 1979.

[21] Joss Knight and Ian Reid. Active visual alignment of a mobile stereo camera platform. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 4, pages 3203–3208. IEEE, 2000.

[22] Marc Spiwak. Angenieux optimo 3d-ready lens pairs. Available at `https://www.bhphotovideo.com/explora/content/angenieux-optimo-3d-ready-lens-pairs`. B and H photography website, Accessed: 15-08-2019.

[23] Frederik Zilly, Josef Kluger, and Peter Kauff. Production rules for stereo acquisition. *Proceedings of the IEEE*, 99(4):590–606, 2011.

[24] Frédéric Devernay and Paul Beardsley. Stereoscopic cinema. In *Image and Geometry processing for 3-D cinematography*, pages 11–51. Springer, 2010.

[25] Hajer Fradi and Jean-Luc Dugelay. Improved depth map estimation in stereo vision. In *Stereoscopic Displays and Applications XXII*, volume 7863, page 78631U. International Society for Optics and Photonics, 2011.

[26] Raad H Thaher and Zaid K Hussein. Stereo vision distance estimation employing sad with canny edge detector. *International Journal of Computer Applications*, 7(3):38–43, 2014.

[27] Yuanxiu Zhou, Yan Song, and Jintao Lu. Stereo image dense matching by integrating sift and sgm algorithm. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:3, 2018.

[28] J. Jang J. Jo and J. Paik. Image fusion using asymmetric dual camera for digital zooming. *Imaging and Applied Optics*, 2016.

[29] J. Heo S-W. Jung S. K. Kim, B. Kang and O. Choi. Photometric stereo-based single time-of-flight camera. *Optics Letters*, 01 2014.

[30] Nicolas S Holliman. Mapping perceived depth to regions of interest in stereoscopic images. In *Stereoscopic displays and virtual reality systems XI*, volume 5291, pages 117–128. International Society for Optics and Photonics, 2004.

[31] Koichi Yoshikawa. Image processing apparatus, image processing method, and program to create a composite image from color image data and monochrome image data, April 29 2014. US Patent 8,711,256.

[32] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. *ACM Transactions on Graphics (TOG)*, 24(3):765–776, 2005.

[33] Holly Brockwell. Huawei P20 review. Available at `https://www.techradar.com/reviews/huawei-p20-review`, 2018. Online; December 04, 2018.

[34] Furkan E Sahin and Rajiv Laroia. Light l16 computational camera. In *Applied Industrial Optics: Spectroscopy, Imaging and Metrology*, pages JTu5A–20. Optical Society of America, 2017.

[35] Kartik Venkataraman, Dan Lelescu, Jacques Duparré, Andrew McMahon, Gabriel Molina, Priyam Chatterjee, Robert Mullis, and Shree Nayar. Picam: An ultra-thin high performance monolithic camera array. *ACM Transactions on Graphics (TOG)*, 32(6):166, 2013.

[36] Rajiv Laroia and Sapna A Shroff. Camera device including multiple optical chains and related methods, January 14 2016. US Patent 20,160,014,314.

[37] M. H. Hayes S. Lee and J. Paik. Distance estimation using a single computational camera with dual off-axis color filtered apertures. *Optics Express*, 11 2013.

[38] J. Li H. Zhou X. Yan, H. Qin and T. Yang. Multi-focus image fusion using a guided-filter-based difference image. *Applied Optics*, 03 2016.

[39] J.H. Kim N. Singhal and Ltd. S.D. Cho Samsung Electronics Co. Apparatus and method for generating bokeh effect in out-focusing photography. Available at `https://www.google.com/patents/US20110280475`, 11 2011. US 13/106323.

[40] N. Cohen E. Ben-Eliezer G. Michrowski, G. Shabtay and E. Goldenberg DigitalOptics Corporation Europe Ltd. Scene background blurring including determining a depth map. Available at `https://www.google.com/patents/US8363085`, 01 2013. US 8363085.

[41] O. Lobachev N. Friedrich and M. Guthe. Faking it: Simulating background blur in portrait photography using a coarse depth map estimation from a single image. volume CSRN 2602. Vaclav Skala-UNION Agency, 2016.

[42] C. Molgaard T.E. Bishop, A. Lindskog and F. Doepke Apple Inc. Photo-realistic shallow depth-of-field rendering from focal stacks. Available at `https://www.google.com/patents/US20170070720`, 03 2017. US-20170070720.

[43] K. Li and Ltd. Y. Liu Huawei Device Co. Image stitching method and apparatus. Available at `https://www.google.com/patents/US20110158509`, 06 2011. US20110158509.

[44] K.G. Larkin D.P. Morgan-Mar and M.R. Arnison Canon Kabushiki Kaisha. Bokeh amplification. Available at `https://www.google.com/patents/US20140152886`, 06 2014. US20140152886.

[45] T.K. Kim J.K. Lee and Ltd T.S. Jeong Samsung Electronics Co. Method for performing out-focus depth information and camera using the same. Available at `https://www.google.com/patents/US9154684`, 10 2015. US9154684.

[46] S.H. Hong Blackberry Ltd. J.N. Fransson. Method and system for generating shallow depth of field effect. Available at `https://www.google.com/patents/US9025874`, 05 2015. US9025874.

[47] I.A. Gulliver Google Inc. Accurate simulation of shallow depth of field using contrast detection. Available at `https://www.google.com/patents/US9237277`, 01 2016. US9237277.

[48] O. Gigushinski G. Shabtay, E. Goldenberg and N. Cohen Corephotonics Ltd. Dual aperture zoom digital camera. Available at `https://www.google.com/patents/US9185291`, 11 2015. US9185291.

[49] James G Nagy, Katrina Palmer, and Lisa Perrone. Iterative methods for image deblurring: a matlab object-oriented approach. *Numerical Algorithms*, 36(1):73–93, 2004.

[50] Ali Karaali and Claudio Rosito Jung. Edge-based defocus blur estimation with adaptive scale selection. *IEEE Transactions on Image Processing*, 27(3):1126–1137, 2017.

[51] Scott McCloskey, Yuanyuan Ding, and Kwong Wing Au. Fourier domain blur estimation method and system, December 22 2015. US Patent 9,218,648.

[52] Hans Strasburger, Michael Bach, and Sven P Heinrich. Blur unblurreda mini tutorial. *i-Perception*, 9(2):2041669518765850, 2018.

[53] Pauline Trouvé-Peloux, Frédéric Champagnat, Guy Le Besnerais, and Jérôme Idier. Theoretical performance model for single image depth from defocus. *Journal of the Optical Society of America A*, 31(12):2650–2662, 2014.

[54] Sun-Jung Kim, Beom Su Kim, Hong Il Kim, Tae-Hwa Hong, and Joo-Young Son. The method for defocusing selfie taken by mobile frontal camera using burst shot. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1816–1820. IEEE, 2016.

[55] Xing Lin, Jinli Suo, Gordon Wetzstein, Qionghai Dai, and Ramesh Raskar. Coded focal stack photography. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2013.

[56] Chao Zhang, John Bastian, Chunhua Shen, Anton Van Den Hengel, and Tingzhi Shen. Extended depth-of-field via focus stacking and graph cuts. In *2013 IEEE International Conference on Image Processing*, pages 1272–1276. IEEE, 2013.

[57] J Wlodek, KJ Gofron, and YQ Cai. Achieving 3d imaging through focus stacking. In *AIP Conference Proceedings*, volume 2054, page 050001. AIP Publishing, 2019.

[58] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70, 2007.

[59] Changyin Zhou, Stephen Lin, and Shree Nayar. Coded aperture pairs for depth from defocus. In *2009 IEEE 12th international conference on computer vision*, pages 325–332. IEEE, 2009.

[60] Changyin Zhou, Stephen Lin, and Shree K Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. *International journal of computer vision*, 93(1):53–72, 2011.

[61] Harel Haim, Shay Elmalem, Raja Giryes, Alex M Bronstein, and Emanuel Marom. Depth estimation from a single image using deep learned phase coded mask. *IEEE Transactions on Computational Imaging*, 4(3):298–310, 2018.

[62] Xuehui Wang, Feng Dai, Yike Ma, Ke Gao, and Yong Dong Zhang. Scene-adaptive coded aperture imaging. *Multimedia Tools and Applications*, 78(1):697–711, 2019.

[63] Michał J Cieślak, Kelum AA Gamage, and Robert Glover. Coded-aperture imaging systems: Past, present and future development–a review. *Radiation Measurements*, 92:59–71, 2016.

[64] Anat Levin. Analyzing depth from coded aperture sets. In *European Conference on Computer Vision*, pages 214–227. Springer, 2010.

[65] Yu-Wing Tai, Ping Tan, and Michael S Brown. Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1603–1618, 2010.

[66] Wang Yongpan, Feng Huajun, Xu Zhihai, Li Qi, and Dai Chaoyue. An improved richardson–lucy algorithm based on local prior. *Optics & Laser Technology*, 42(5):845–849, 2010.

[67] Sergey Chalkov, Natalie Meshalkina, and Chang-Su Kim. Post-processing algorithm for reducing ringing artefacts in deblurred images. In *23rd International Technical Conference on Circuits/Systems, Computers and Communications*, 2008.

[68] Anat Levin, Rob Fergus, Fredo Durand, and William T Freeman. Deconvolution using natural image priors. *Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory*, 3, 2007.

[69] Ganghun Kim, Kyle Isaacson, Rachael Palmer, and Rajesh Menon. Lensless photography with only an image sensor. *Applied Optics*, 56(23):6450–6456, 2017.

[70] Vivek Boominathan, Jesse K Adams, M Salman Asif, Benjamin W Avants, Jacob T Robinson, Richard G Baraniuk, Aswin C Sankaranarayanan, and Ashok Veeraraghavan. Lensless imaging: A computational renaissance. *IEEE Signal Processing Magazine*, 33(5):23–35, 2016.

[71] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, 2018.

[72] Ruomei Yan and Ling Shao. Blind image blur estimation via deep learning. *IEEE Transactions on Image Processing*, 25(4):1910–1921, 2016.

[73] Saeed Anwar, Zeeshan Hayder, and Fatih Porikli. Depth estimation and blur removal from a single out-of-focus image. In *BMVC*, volume 1, page 2, 2017.

[74] Marcela Carvalho, Bertrand Le Saux, Pauline Trouvé-Peloux, Andrés Almansa, and Frédéric Champagnat. Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks? In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[75] Shanghang Zhang, Xiaohui Shen, Zhe Lin, Radomír Měch, Joao P Costeira, and José MF Moura. Learning to understand image blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6586–6595, 2018.

[76] Parikshit Sakurikar, Ishit Mehta, and PJ Narayanan. Defocus magnification using conditional adversarial networks. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1337–1346. IEEE, 2019.

[77] Tom van Dijk and Guido CHE de Croon. How do neural networks see depth in single images? *arXiv preprint arXiv:1905.07005*, 2019.

[78] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012.

[79] Ashutosh Saxena, Min Sun, and Andrew Y Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):824–840, 2008.

[80] Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 270–279, 2017.

[81] S Uchiyama. Superiority of image plane phase detection af. Technical report, ITE Technical Report, 36.38, 2012.

[82] Masahiro Kobayashi, Michiko Johnson, Yoichi Wada, Hiromasa Tsuboi, Hideaki Takada, Kenji Togo, Takafumi Kishi, Hidekazu Takahashi, Takeshi Ichikawa, and Shunsuke Inoue. A low noise and high sensitivity image sensor with imaging and phase-difference detection af in all pixels. *ITE Transactions on Media Technology and Applications*, 4(2):123–128, 2016.

[83] Chin-Cheng Chan and Homer H Chen. Improving the reliability of phase detection autofocus. *Electronic Imaging*, 2018(5):1–5, 2018.

[84] Xiangyu Xu, Deqing Sun, Sifei Liu, Wenqi Ren, Yu-Jin Zhang, Ming-Hsuan Yang, and Jian Sun. Rendering portraitures from monocular camera and beyond. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 35–50, 2018.

[85] Rahul Garg, Neal Wadhwa, Sameer Ansari, and Jonathan T Barron. Learning single camera depth estimation using dual-pixels. *arXiv preprint arXiv:1904.05822*, 2019.

[86] Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1547–1555, 2015.

[87] Wei Yuan, Li-Hua Li, Wing-Bun Lee, and Chang-Yuen Chan. Fabrication of microlens array and its application: a review. *Chinese Journal of Mechanical Engineering*, 31(1):16, 2018.

[88] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, Pat Hanrahan, et al. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005.

[89] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, 2017.

[90] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *VMV*, volume 13, pages 225–226. Citeseer, 2013.

[91] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. In *8th International Conference on Quality of Multimedia Experience (QoMEX)*, number CONF, 2016.

[92] Ting-Chun Wang, Jun-Yan Zhu, Ebi Hiroaki, Manmohan Chandraker, Alexei A Efros, and Ravi Ramamoorthi. A 4d light-field dataset and cnn architectures for material recognition. In *European Conference on Computer Vision*, pages 121–138. Springer, 2016.

[93] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen. The lumigraph. In *Siggraph*, volume 96, pages 43–54, 1996.

[94] Chi Zhang, Guangqi Hou, Zhaoxiang Zhang, Zhenan Sun, and Tieniu Tan. Efficient auto-refocusing for light field camera. *Pattern Recognition*, 81:176–189, 2018.

[95] Brian A Barsky and Todd J Kosloff. Algorithms for rendering depth of field effects in computer graphics. In *Proceedings of the 12th WSEAS international conference on Computers*, volume 2008. World Scientific and Engineering Academy and Society (WSEAS), 2008.

[96] Martin Kraus and Magnus Strengert. Depth-of-field rendering by pyramidal image processing. In *Computer graphics forum*, volume 26, pages 645–654. Wiley Online Library, 2007.

[97] Benxuan Zhang, Bin Sheng, Ping Li, and Tong-Yee Lee. Depth of field rendering using multilayer-neighborhood optimization. *IEEE transactions on visualization and computer graphics*, 2019.

[98] Wolf Hauser, Balthazar Neveu, Jean-Benoit Jourdain, Clément Viard, and Frédéric Guichard. Image quality benchmark of computational bokeh. *Electronic Imaging*, 2018(12):340, 2018.

Chapter 6

The apodization approach to depth of field reduction

6.1 Overview

Apodization is a technique used to shape the 3D Point Spread Function (PSF) through the design of a 2D mask usually placed in the pupil plane. Apodization can have imaging or non-imaging purposes. In this chapter, we consider a depth-of-field application, which is part of imaging techniques. However, several fields use apodizing filters mostly to control the profile of a light beam, such as LASERs. One could consider the Lyot coronograph to be a primitive form of apodization (blocking the light flux from a bright star to better observe the surrounding planets).

Although popular for some applications, amplitude masks such as the one used in the Lyot coronograph result in a signal attenuation. This is never desirable and is only acceptable for applications with a controlled illumination (e.g. microscopy techniques with a chosen light source).

In 2D, apodization can be used to reduce some aberrations or achieve super-resolution [1–4]. In this study however, the main concern is axial rather than lateral resolution. Thus we are exploring the capabilities of phase-mask design to reduce the axial PSF size. If the image-space PSF is narrowed along the optical axis, the conjugated object space area of focus is also narrowed, which translates into a shallower depth of field.

When it comes to depth of field reduction, the literature does not provide a systematic way to design a phase mask to tailor the 3D PSF. Most methods are only used for lateral PSF shaping [3, 5–9]. For amplitude apodization, diffraction theory points to rings as the best solution [10]. For phase apodization, although non annular masks can be used, such as Walsh filters [11], most of the literature explores annular masks.

From a 2D phase or amplitude mask, one can reconstruct the 3D PSF through propagation theory [12, 13]. Thus, we used the algorithm proposed by Lin et al. [14] to simulate the 3D PSF from a set of conditions and physical parameters. The Matlab code was provided by Lin and adapted to test sets of pupils and to provide a merit function.

From the simulated 3D PSF, one can then design a merit function and use an optimization algorithm to design an optimal mask for a set of conditions. For this purpose, we decided to use genetic algorithms and particle swarm optimization since they outperform the more common gradient descent optimization algorithm.

As shown below, when it comes to depth-of-field reduction, the results given in this chapter are not of practical use for photographic applications. The z-axis profile can be compressed either at the expense of the amount of signal (amplitude mask absorbing most of the light) or with a fairly limited range. We typically obtained 20% compression along the z-axis. As a comparison, a full stop of aperture compresses the z-axis profile by a factor 2. For a typical f/2.0 lens, a phase mask could produce the axial PSF of an f/1.8 lens.

Despite this limited usefulness of apodized lenses in photographic applications, we believe that our contribution can be valuable for ophthalmic contact lenses [15, 16].

6.2 Literature review of Apodization techniques

6.2.1 Apodization for super resolution

We have demonstrated in section 4 that the axial resolution is limited. However, there is interest in several research fields in beating this axial limit. A major caveat needs to be addressed regarding "super resolution" techniques. Although numerous techniques can be used to overcome the diffraction limit, these techniques always trade some aspects of the performance. Microscopy for instance, aims at imaging sections as thin as possible which can be achieved at the expense of some of the light efficiency.
An early wave of the research around pupil apodization focused on using apodization techniques to achieve super-resolution. Both lateral [17] and axial super-resolution [18] have been considered using phase or amplitude masks in the pupil plane.

Pupil filtering cannot improve the diffraction limited Optical Transfer Function (OTF) of an optical system as shown by Davis [19] but can improve the Signal to Noise Ratio (SNR) of a microscopy device. The reduction of the axial extent of the PSF enables the capture of more high spatial frequency signal in the focal plane which has a positive impact on the SNR for high frequencies.

6.2.2 Apodization for depth of field estimation and extension

A very active subject of research is depth-of-field estimation and extension. In practice depth estimation from defocus is the first step for any processing such as depth-of-field extension or reduction. The depth of field extension requires to have some information even in the out-of-focus areas in order to recover the scene.

Since the pupil is known, one can recover part of the information by deconvolving the out-of-focus areas of an acquired image with the known PSF. The choice of the pupil pattern depends on the system and in particular on the SNR of the signal detected by the sensor, as others have demonstrated [20, 21]. The circular pattern is not always the optimal blur pattern for this application, and numerous masks have been proposed to improve the image quality obtained by deconvolution depending on the noise parameters of the system.

6.2.3 Apodization for depth-of-field reduction

Research into super resolution in other fields can find valuable applications in the consumer imaging field. One of the main problems of consumer oriented imaging devices is the small size of the sensor and the quite small aperture of the lenses used, which is unable to achieve a shallow depth of field image with a satisfying bokeh.

Numerous image processing attempts have been made for single sensor and multi-sensor cameras to decrease depth of field [22–25] but until now the problem of the transition between in focus and out of focus areas of an image has not been solved and edges bring artefacts to the image result.

If image processing alone cannot manage to produce a robust depth of field reduction then a solution can hopefully be found in the combination of optics and image processing. It is claimed that the super-resolution pupils proposed by Toraldo di Francia [26] can reduce the depth of field without limit, at the cost of more intense side-lobes. The in-focus plane is imaged, whereas other planes surrounding the in-focus plane are imaged with a lower flux. With the apodization mask, one attempts to lower the light flux around the imaging plane in order to get closer to the behaviour of a wide-aperture lens, as shown in figure 6.1.

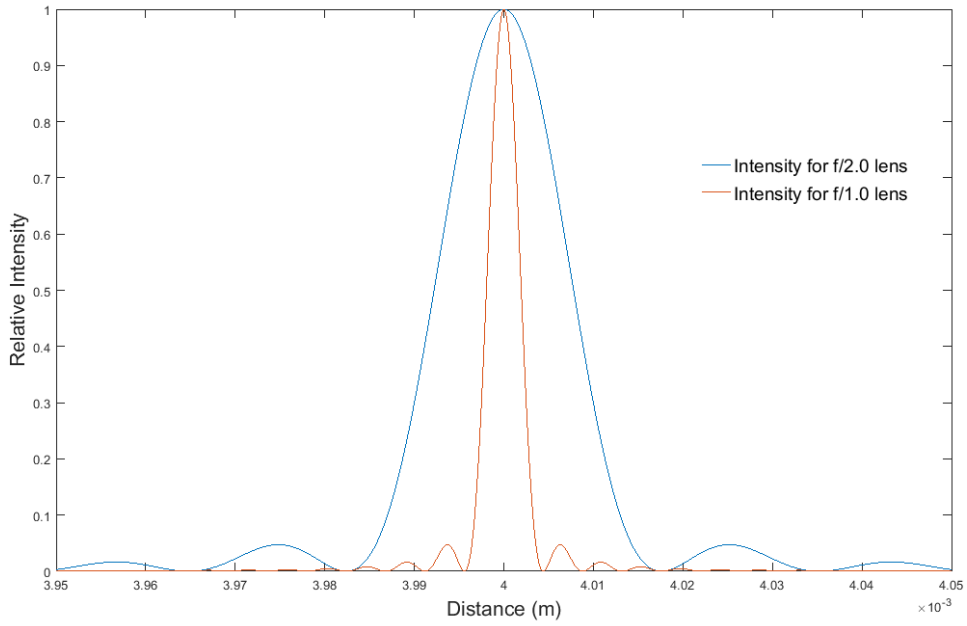


Figure 6.1: Axial intensity for a 4mm focal length lens imaging an object located at infinity, with respective f/2.0 and f/1.0 apertures.

6.2.4 Possible applications

Axial apodization, and more generally 3-D PSF apodization, is of great interest for numerous fields such as microscopy [27, 28], electromagnetism [29, 30], ophthalmic systems (for depth of field extension only) [15, 16] and optical design with the possible reduction of simple aberrations such as third order spherical aberration [31].

6.3 Types of Apodization masks

Because of the above-presented applications, it is of great interest to be able to tailor the PSF along the z-axis. Several pupil shapes can be used to achieve a PSF tailoring.
In 1952, Toraldo di Francia introduced the idea of annular filtering to achieve super-resolution along the z-axis [26]. This work shows that there is virtually no limit to the axial resolution that can be achieved using concentric annular pupils such as the example in figure 6.2, at the cost of the amount of signal transmitted by the system and lateral resolution performance.

6.3.1 Amplitude masks

Numerous variations [32, 33] around the original idea of Toraldo di Francia have been made, mainly for pupil manufacturing reasons. Several papers have shown that the apodization could be made with an absorbing apodizer, i.e. an amplitude mask. Mino and Okano pioneered research into amplitude masks for depth of field extension in 1971 [34], where they studied the dependence of

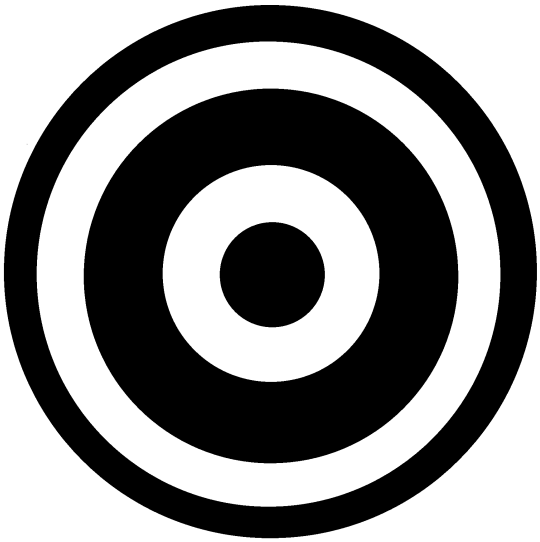


Figure 6.2: Example of a concentric annular pupil. Black areas block the light (no transmission), whereas white areas transmit the light. Most of the concentric annular pupils feature rings with an equal area, and thus the concentric radii follows inverse squareroot series.

the Optical Transfer Function (OTF) on the focusing distance. They introduced two kinds of amplitude pupils, T_A and T_B, with the respective properties of extending the depth of field and reducing the depth of field. The use of amplitude masks is not suitable for small sensors as it reduces the amount of light collected and thus the SNR.

Coded apertures presented in Chapter 5.1.3.3 are a generalized version of the typical amplitude mask. The former being a 2D varying pattern whereas the later only features 1D variation along the radius.

6.3.2 Walsh phase masks

The concept of depth-of-field extension through binary amplitude masks has been used widely in combination with image processing to extend the depth of field as shown previously. To keep the ease of manufacturing but improve the light efficiency of the pupil, binary amplitude pupils have been replaced to phase masks with 0 and π binary masks [35–37]. The family of functions used for such masks was introduced by Walsh in 1923 [38]. Figure 6.3 shows an example of the pupil W23. The optimization of such masks has first been proposed in 2004 by Cagigal [39] and more recently by Hazra and Mukherjee [40, 41]. The most interesting property of Walsh pupils is that they can be easily manufactured (only two different phases), although this binary constraint limits the optimization.

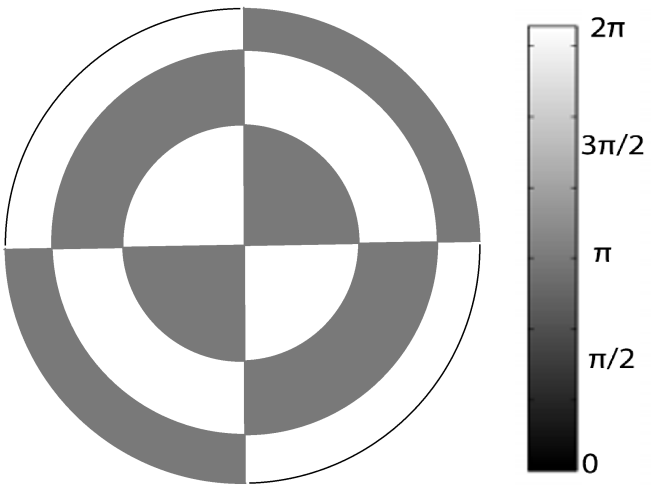


Figure 6.3: Example of a Walsh phase mask. This particular pupil is known as W23.

6.3.3 Generalized phase masks

Without the constraint of producing masks with a very small number of phase shifts, it is possible to generalize the idea of phase masks with only the constraint of a discrete mask pattern and number of phase shifts. In practice, with the use of LCoS devices[1] it is possible to create a high resolution phase image with 256 levels [42]. In microscopy, Sales introduced an optimization process for phase-only pupils, which results in a better Strehl ratio[2] than amplitude masks [43]. The idea of a phase mask has been explored for various applications [44, 45], including microscopy [46, 47] and astronomy [48]. To our knowledge it has never been used for small sized sensor applications and thus we will discuss the use of a phase mask for this particular case.

6.3.4 Wavefront Coding functions

A different idea connected to pupil apodization through phase masks is wavefront coding to achieve defocus invariance. The image obtained with wavefront coding is not usable before a deconvolution step has been made.
The field was initiated by the concept of cubic phase masks introduced in 1995 by Dowski and Cathey [49]. The cubic phase mask is shown in figure 6.4. The main advantages of wavefront coding as opposed to previous methods are that it does not suffer from the loss of optical power entailed by amplitude masks and that it has a better optical resolution at all distances. Following this initial solution, various phase mask functions have been proposed such as logarithmic [50], exponential [51], high order [52], polynomial [53], free form rational [54], cosine [55] and sine [56].

[1]LCoS or Liquid Crystal on Silicon is a reflective spatial light modulator technology.

[2]The Strehl ratio is a criterion quantifying the aberrations of an optical system as opposed to an ideal, non-aberrant, equivalent system.

The wavefront coding approach is however limited to the depth of field *extension* and not applicable to depth of field *reduction*.

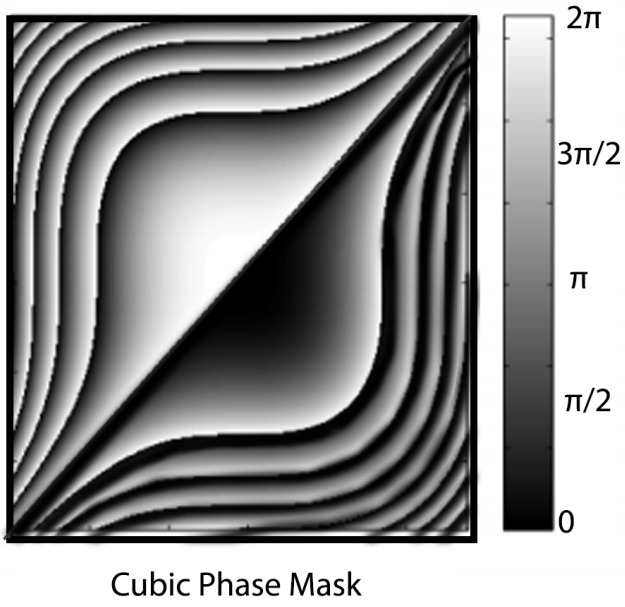


Figure 6.4: Cubic phase mask. The phase varies between 0 and 2π.

6.3.5 Analytic solution for phase masks

Luneberg introduced in 1944 [57] a set of apodization problems. These problems are oriented towards super resolution on the lateral plane and can be adapted for z-axis apodization. In 1962, R. Barakat suggested an analytical approach [58] that can be transposed to the axial apodization problem.

Let us introduce the following notations: U the amplitude of the field around the focal plane, such that $U(0)$ is the amplitude at the focal plane z_0. We will compute the intensity along the z axis. $P(\beta, \rho)$ is the desired pupil with an intensity null at a distance β from the focal plane, while ρ defines the radius of the rotationally symmetric pupil.
The desired system must respect several conditions in order to narrow the main axial intensity lobe while demanding that the central illuminance be maximum:

Maximum intensity in the central lobe, $U(0, \beta)$ maximum:

$$U(0,\beta) = \int_0^1 P(\beta,\rho)\exp(-\frac{i\beta\rho^2}{2})\rho d\rho \tag{6.1}$$

Intensity null at $z = z_0 + \beta$:

$$U(\beta, \beta) = \int_0^1 P(\beta, \rho)\exp(-\frac{i\beta\rho^2}{2})\rho d\rho = 0 \tag{6.2}$$

No absorption in the pupil plane:

$$\int_0^1 |P(\beta, \rho)|^2 \rho d\rho = 1 \tag{6.3}$$

The optimization can be carried out using a Lagrange multiplier and an analytic solution could be achieved from the resolution of a Fredholm Equation. However this last step only provides trivial solutions or shifted focus solutions. Adding extra constraints to the problem leads to a non solvable equation, thus making the Barakat analytic approach unusable for phase masks.

6.4 Numerical/Algorithmic Approach

Since we want to reduce the depth of field through the use of a phase only mask, we will need to find the optimal phase only pupil. Some random or exhaustive searches could be used in the case of a $0 - \pi$ phase mask as for 10 rings, there are $2^10 = 1024$ possible combinations. However in the underestimated case of 10 rings with 10 levels of phase, there are 10^{10} combinations, which would take an approximate 43 years of computations for a realistic 0.1sec of computation per pupil (30 rings precision requires 0.15 ± 0.02sec).

6.4.1 Description of the problem

From an algorithmic approach, the apodization problem is an optimization problem. There is no trivial function connecting the merit function defined the pupil. However, it is possible to compute the 3D Point Spread Function resulting from a given configuration and obtain a merit value using a merit function.

The following assumptions have been made in order to reduce the number of variables:

- The pupils that we want to explore have symmetry of revolution. In practice, we are studying the impact of a vector on the depth of focus by generating a pupil through revolution symmetry.
- For our design, we decided to use a phase modulation. This assumption has been made mainly for SNR concerns. This is in particular suitable for small sized sensors under a not-controlled lighting environment, such as low light photography applications for smartphone camera devices.

6.4.2 Computation of a three dimensional diffracted field

Using the diffraction formula introduced by McCutchen in 1964 [59], Lin and *al.* proposed a direct computation of the three dimensional diffracted field near the focal plane [60], shown in figure 6.5.

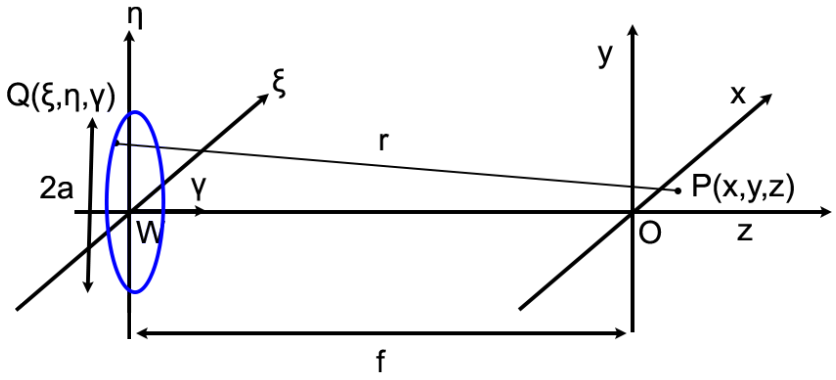


Figure 6.5: Geometry of the 3D diffraction calculation, and notations used by Lin et al. [60].

Assumptions:

1. We assume that the computation will be made near the focal plane.
2. We assume that the pupil point is close to the axis as $r \approx f + z$.
3. Given the geometrical configuration we have $\xi^2 + \eta^2 + \gamma^2 = f^2$ where ξ corresponds to the x-axis, η corresponds to the y-axis and γ corresponds to the z-axis as shown in figure [**?**].

Approximation of r:

By definition, the distance between the source point and the observed point, noted r is given by $r^2 = (x-\xi)^2 + (y-\eta)^2 + (z-\gamma)^2 = x^2 + y^2 + z^2 - 2(x\xi + y\eta + z\gamma) + \xi^2 + \eta^2 + \gamma^2$.
Using $\xi^2 + \eta^2 + \gamma^2 = f^2$ (assumption 3), we have $r^2 = x^2 + y^2 + z^2 - 2(x\xi + y\eta + z\gamma) + f^2$.
Replacing $f^2 + z^2$ with the equivalent expression $(f+z)^2 - 2fz$, we can use a limited development around the greatest term $(f+z)^2$ ($\sqrt{1+x} \approx 1 + \frac{x}{2}$ for $x \ll 1$) for the square root:

$$r \approx f + z - \frac{fz}{f+z} + \frac{x^2+y^2}{2(f+z)} - \frac{x\xi + \eta y + \gamma z}{f+z} \tag{6.4}$$

Knowing that $z + f - \frac{fz}{f+z} = z + \frac{f(f+z)-fz}{f+z} = z + \frac{f^2}{f+z}$, we finally reach the equation 6.5 provided:

$$r \approx z + \frac{f^2}{f+z} + \frac{x^2+y^2}{2(f+z)} - \frac{x\xi + \eta y + \gamma z}{f+z} \tag{6.5}$$

This approximation is known as the Fresnel approximation, which is valid for [61]:

$$\begin{cases} z^3 \gg \frac{\pi}{4\lambda}[(x-\xi)^2 + (y-\eta)^2]^2_{\max} \\ z < \frac{2D^2}{\lambda} \\ \text{Fraunhofer Approximation domain} \end{cases} \tag{6.6}$$

For a circular aperture, the direct fast-Fourier-transform (FFT) computation provides the result in figure 6.7. On the plot, a pixel along the horizontal axis corresponds to 2μm and a pixel along the vertical axis corresponds to 0.01μm.

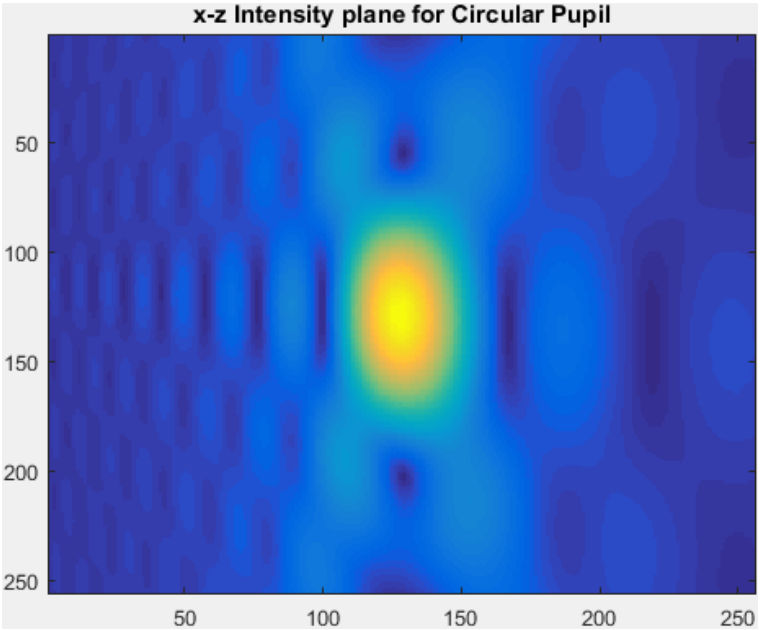


Figure 6.6: x-z intensity profile for a standard circular pupil, as computed with a Fast 3D FFT from the pupil.

As one would expect, a narrower central lobe along the z-axis is traded against a wider central lobe along the x-axis. Figure

As one would expect, a narrower central lobe along the z-axis is traded against a wider central lobe along the x-axis. Figure 6.8 compares the apodized pupil with a standard lens.

As a side note, the 3D-FFT shows that spherical aberration widens the central lobe along the z-axis, whereas astigmatism widens the central lobe along the x-axis, both behaviours are reported in figure 6.9. The first result has been used in the patent of N. George and W. Chi [62] related to extended depth of field.

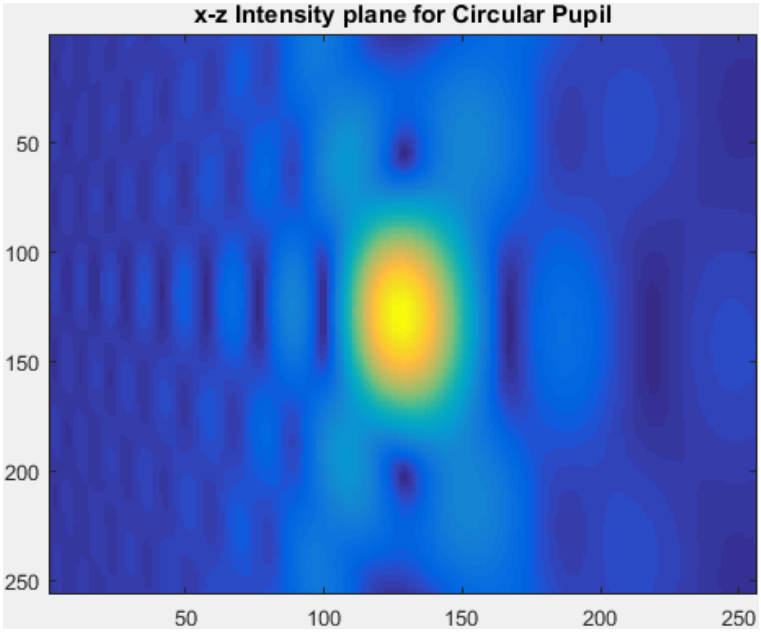


Figure 6.7: x-z intensity profile for a standard circular pupil, as computed with a Fast 3D FFT from the pupil.

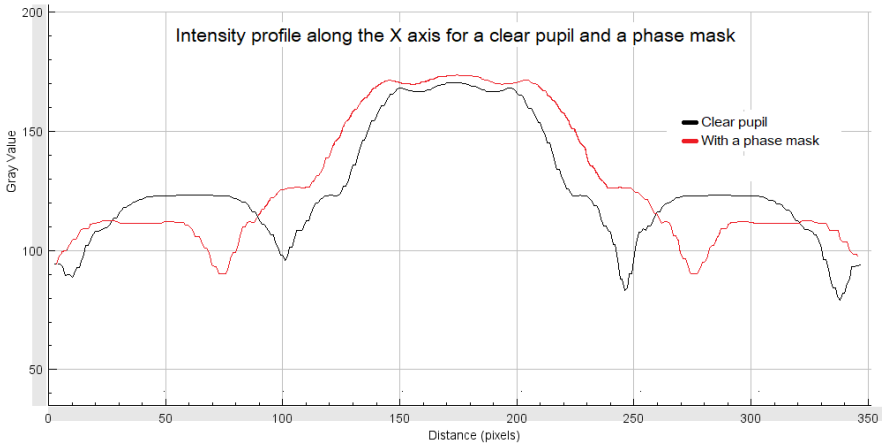
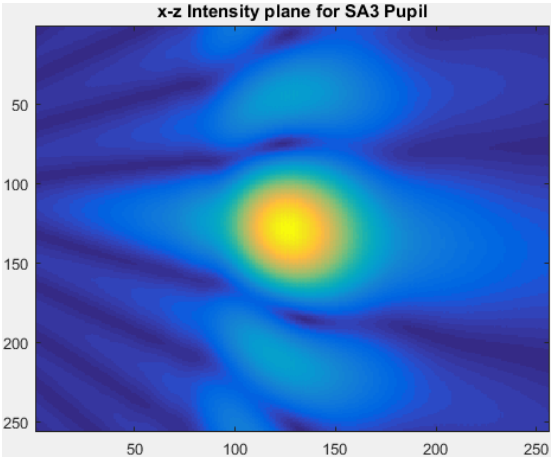


Figure 6.8: Comparison of the intensity profile along the x-axis for a clear aperture (in black) and for an optimized phase-mask (in red). The apodized pupil has a 30% wider central lobe.

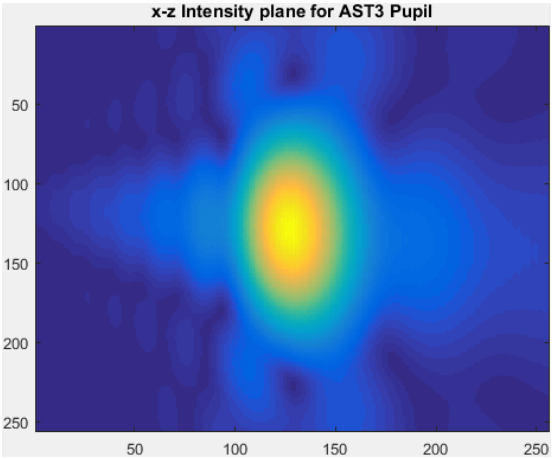
6.4.3 Definition of a merit function

The method of Hazra [63] used a dynamic merit function, varying with the iteration, to speed up convergence of the optimization. However this merit function only relies on the integral of the intensity, which cannot be adapted to specific apodization requirements. For this reason we introduce a new, static, merit function using:

1. The distance from the maximum of the central lobe to the desired position along the z-axis. During the optimization process a focus shift can occur which might be unacceptable for



(a) x-z intensity profile for a circular pupil with spherical aberration.



(b) x-z intensity profile for a circular pupil with astigmatism aberration.

Figure 6.9: Influence of optical aberrations on depth of field.

certain applications.

2. The maximum slope of the main lobe. We would like to maximize this slope to have a very clear transition between in-focus and out-of-focus areas.

3. The proportion of the intensity curve under a certain sidelobe-threshold, which corresponds to the amount of noise that can be accepted (see Noise Threshold on **Fig 4**). This quantity should be maximized.

4. The proportion of the intensity curve above a certain peak-threshold corresponding to the useful signal (see Signal Threshold in **Fig 4**). This value must be minimized in the case of "super-resolution" applications.

A visual depiction of these metrics is proposed in figure 6.10.

In order to get a more robust estimate of the merit function even with a few points we have introduced a polynomial fit before computing the metric. Since the intensity along the axis over the considered interval is symmetric with respect to the plane of the focus, the polynomial degree needed to fit the curve is even. For the interval considered a polynomial of degree 6 should be sufficient but since the lobes can be shifted, we have used a polynomial of degree 8. With such a fit, it seems that the computation of 30 points for the intensity provides good scores for the optimization, versus the 100 points recommended by [63]. Since the intensity computation is the bottleneck of the algorithm, this speeds up the algorithm execution.

A dynamic merit function has also been tested in order to improve the convergence. However, according to our observations, coding the merit function dynamically does not bring better results

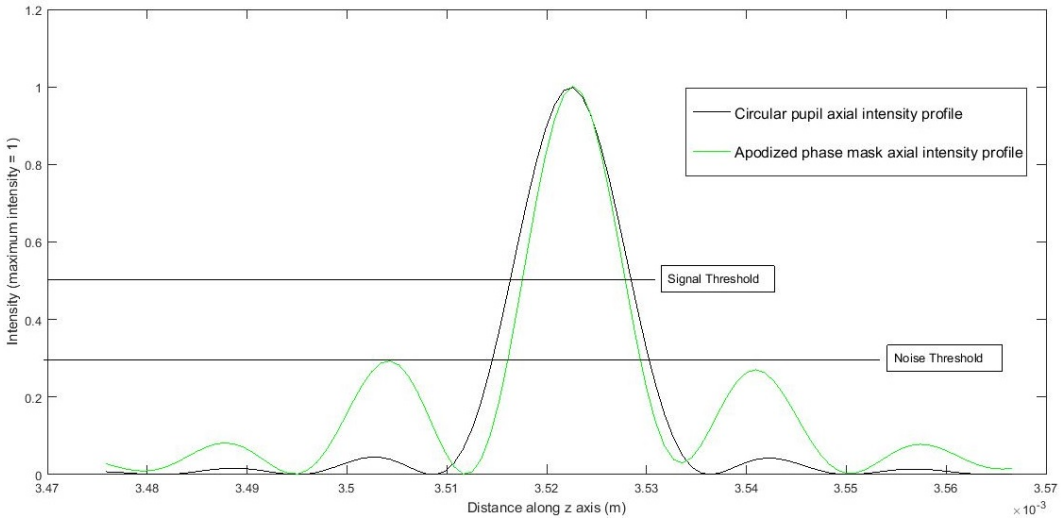


Figure 6.10: Definition of a merit function based on an example of observed intensity profile

in the case of a PSO algorithm. This can be explained by the easier convergence of PSO as opposed to Genetic Algorithms.

6.4.4 Extensive search and Optimized Extensive search

One may wonder if it is possible to test extensively all the ring configurations possible for a given number of phase levels. The number of phase masks to test is reported in table 6.1. The PSF corresponding to each of these masks must be computed, and then analysed to obtain a merit function. In most cases a $32 \times 32 \times 32$ sample of the PSF provides an accurate representation of the behaviour, and this PSF size will be used for the rest of this section. To put the numbers from table 6.1 into perspective, our desktop computer (i7-2600, 16GB of RAM) processes around 800 phase masks a minute. With a more powerful system, one might be able to process up to a few thousands of masks a minute.

Number of pupil rings	Number of phase levels	Number of pupils to test
4 rings	2 levels	16
4 rings	5 levels	625
4 rings	10 levels	10^4
6 rings	2 levels	64
6 rings	5 levels	1.6×10^4
6 rings	10 levels	10^6
8 rings	2 levels	256
8 rings	5 levels	3.91×10^5
8 rings	10 levels	10^8
10 rings	2 levels	1024
10 rings	5 levels	9.77×10^6
10 rings	10 levels	10^{10}
12 rings	2 levels	4096
12 rings	5 levels	2.44×10^8
12 rings	10 levels	10^{12}
14 rings	2 levels	1.63×10^4
14 rings	5 levels	6.10×10^9
14 rings	10 levels	10^{14}
16 rings	2 levels	6.55×10^4
16 rings	5 levels	1.53×10^{11}
16 rings	10 levels	10^{16}
20 rings	2 levels	1.05×10^6
20 rings	5 levels	9.54×10^{13}
20 rings	10 levels	10^{20}
32 rings	2 levels	4.30×10^9
32 rings	5 levels	2.33×10^{22}
32 rings	10 levels	10^{32}

Table 6.1: Number of masks to be tested in the case of an extensive search.

The number of masks is $N_{levels}^{N_{rings}}$.

The number of phase levels N_{levels} is critical for the computation time. Mathematically however, one could try to approximate the exploration of all the possible masks with a coarse exploration (e.g. pupils with only 2 phase levels) that is progressively refined. This technique is a form of optimized search. Assuming N_{opt} is the number of phase levels of the coarse mask, the number of tested masks drops from $N_{levels}^{N_{rings}}$ to $\lceil log_{N_{opt}}(N_{levels}) \rceil \times N_{opt}^{N_{rings}}$. The practical results of this approximation can be seen in table 6.2.

Extensive search	Number of pupils to test	Optimized search	Number of pupils to test
4 rings, 5 levels	625	4 rings, 3×2 levels	48
4 rings, 10 levels	10^4	4 rings, 4×2 levels	64
8 rings, 5 levels	3.91×10^5	8 rings, 3×2 levels	768
8 rings, 10 levels	10^8	8 rings, 4×2 levels	1024
10 rings, 5 levels	9.77×10^5	10 rings, 3×2 levels	3072
10 rings, 10 levels	10^{10}	10 rings, 4×2 levels	4096
16 rings, 5 levels	1.53×10^{11}	16 rings, 3×2 levels	1.97×10^5
16 rings, 10 levels	10^{16}	16 rings, 4×2 levels	2.62×10^5
20 rings, 5 levels	9.54×10^{13}	20 rings, 3×2 levels	3.15×10^6
20 rings, 10 levels	10^{20}	20 rings, 4×2 levels	4.19×10^6

Table 6.2: Comparison of the number of masks to test in the case of an extensive search and an optimized search.

6.4.5 Optimization algorithms

Since there is no derivable function connecting the pupil with its merit function, it is not possible to use a gradient descent type of optimization algorithm. Tabu Search [64], Simulated Annealing [65], Genetic Algorithm and Particle Swarm Optimization seem to be the best candidates to find the global minimum.

However Tabu Search and Simulated Annealing are not suitable for our problem for two reasons:

- Speed: These algorithms are known to be among the slowest. Some further improvements have been made but are still slower than Genetic Algorithms for high dimensional problems [66]
- Result: Neither algorithm is designed for high multidimensional spaces. Thus the convergence to a global minimum is not guaranteed.

Various refinements have been proposed to the previous methods [67–69], some of them including some Genetic Algorithm principles [70, 71], but still not solving these two concerns.

6.5 Phase Mask design using Genetic Algorithms

The Genetic Algorithm (GA) was introduced using Darwinian evolution as a model. Starting with a random population (in our case a population of phase masks), the algorithm iteratively hybridizes the best elements using random crossover and mutations. Each new population should thus be better on average than the previous generation.

The Genetic Algorithm requires some tuning mainly for two parameters: the ideal size of the population should be determined and the frequency of crossover should be tuned. For our tests

we have fine tuned these parameters starting from the parameters used by [63].

6.5.1 Limitations

Genetic Algorithms have an inherent limitation in the case of phase pupils. Since each ring behaves independently, the selection of a ring among the best pupils does not necessarily improve the newly created pupil. This fundamental limitation of the Genetic Algorithm is overcome only in the case of a great number of steps, which means a very long processing time.

In order to overcome this convergence limitation, a dynamic merit function has been proposed [63]. The idea of a dynamic merit function in this case is to progressively reduce the scope of possible values to force the algorithm to converge with a global minimum instead of remaining trapped in a local one. It is important to point out that by using a dynamic merit function, the search for a global minimum may be too restricted to only a part of the possible solutions. The way of defining the dynamic variation of the merit function has an impact on the performance of the algorithm.

6.6 Phase Mask design using Particle Swarm Optimization

The Particle Swarm Optimization (PSO) algorithm is a heuristic global optimization algorithm invented in 1995 as an alternative to Genetic Algorithms [72].

Implementation is based on a population initially randomly distributed for each of the dimensions of the problem (each of the phase values in our case) and with an initially null variation (variation used to compute the next particle position). The score of all pupils is computed and used to compute a variation vector moving the particle toward an expected better position.

$$\begin{aligned}\text{Variation}_i = v \times [w \times R_{\text{phase}} \\ + c_1 \times (\text{Best Position}_i - \text{Position}_i) \\ + c_2 \times (\text{Best Position}_{\text{all particles, all time}} - \text{Position}_i) \\ + c_3 \times (\text{Best Position}_{\text{neighbors of i}} - \text{Position}_i)]\end{aligned} \quad (6.7)$$

The variation amplitude is controlled with a speed parameter v that can be varied dynamically. For each particle the variation (and thus new position) is based on the history of the particle (best score of the particle), the best scores of its neighbours and the overall best score of the population. A random contribution is also added to help explore neighbouring positions. The PSO algorithm aims at making this population evolve until it converges to a global minimum.

A particular case to mention is that of borders. In the case of a particle new position being outside the selected boundaries (in our case, the phase is assumed to be in the $[0-2\pi]$ range), two solutions can be used: the simplest one is to suppress the particle and create a new one instead, losing the information associated with the particle; another one is to bounce the particle back inside the boundaries either through bouncing borders or through a toroidal space representation.
The coefficients used to manage the optimization behaviour are: the maximum variation allowed (can be a function of the step of the process), and the different coefficients used to balance the relative contributions of the personal best, neighbours and all time best attractions.

Numerous variants of the original algorithm have been made since [73–82] in order to reduce the risk of reaching a local minimum instead of the desired global minimum [83].

6.6.1 Limitations

Particle Swarm Optimization is designed to converge rapidly. The algorithm itself needs to be fine tuned in order to have a smooth convergence toward the global minimum. The behaviour of the algorithm depends on the merit function and thus may need a fine tuning in case of a modification of the physical constraints or targets.

In the case of a number of rings $n > 15$ we have noticed that the score achieved worsened with the number of rings suggesting the reach of a local minimum.

6.7 Results from different algorithmic approaches and discussion

Numerous comparisons have been made in various fields to evaluate the relative performance of GA and PSO algorithms [84, 85] The PSO based algorithm converges very quickly to a solution. From our observations, a satisfying solution can be obtained after about 10 to 20 steps and the algorithm converges in less than 50 steps.

6.7.1 Comparison of performance for a given time

	PSO score	GA score
5 rings, 80 sec	85.11	98.04
10 rings, 80 sec	82.55	98.03
15 rings, 120 sec	87.53	97.3
20 rings, 200 sec	90.59	95.86

Table 6.3: Normalized results from GA and PSO for different settings and time constraints, evaluated with the defined merit function. Percentage of the width of the diffraction limited central lobe. The diffraction-limited system scores 100. The lower the better.

The main lobe can be compressed to typically $85 - 90\%$ of its original width, in less than 200 seconds of optimization (i7-2600, 16GB RAM). Without any speed constraint, the results are slightly better.

6.7.2 Phase Pupil Results

In order to illustrate the results from different algorithms, one can take the case of 8 rings with 5 levels of phase. In this case, 3.91×10^5 pupils have to be tested for an extensive search. On the computer used, this corresponds to around 8 hours of computation time. The purpose of this experiment is to find how good the pupil found with each algorithm is compared to the "best" possible result (from extensive search).

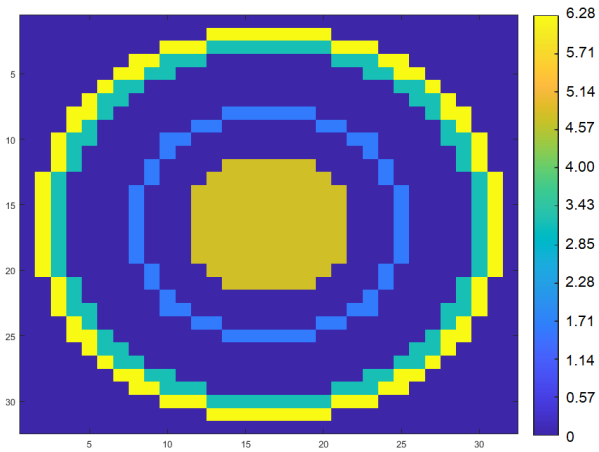
Optimization Algorithm	Compression Coefficient (%)
Extensive Search	70.5%
Genetic Algorithm	84.2%
PSO	78.8%

Table 6.4: Comparison of the results from Genetic, PSO algorithms and best pupil.

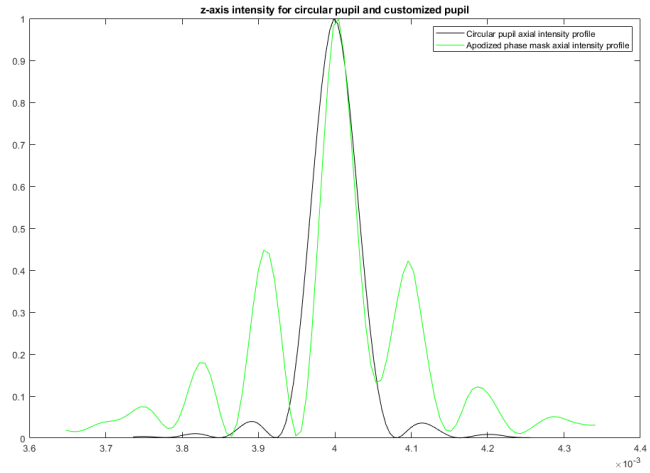
6.8 Conclusions from testing

Comparing the results from both the GA and PSO algorithms, the results from PSO are better than the ones from GA, especially for quick optimizations. This result is coherent with the observations made by [86, 87], stating that the PSO approach is faster than the Genetic Algorithm while remaining easy to implement.

The merit function created to solve the apodization problem can be customized at will and adapted to various scenarios, including extension of the depth of field (eDoF). The literature of extended

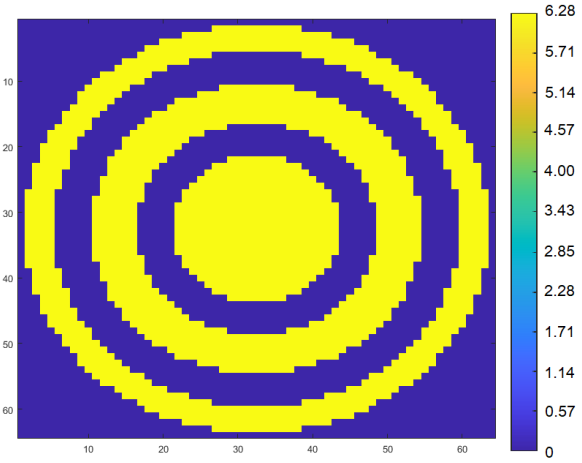


(a) Best pupil according to extensive search.

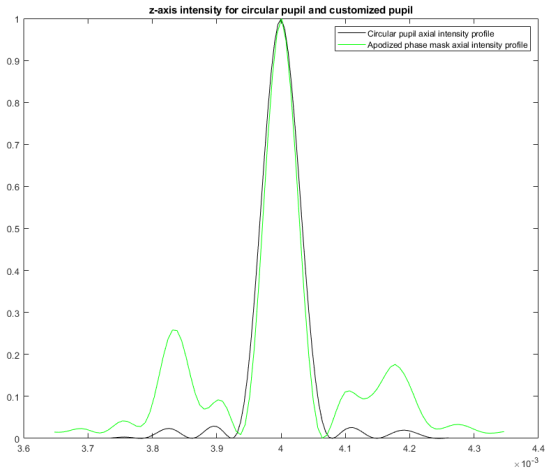


(b) Function corresponding to the pupil.

Figure 6.11: Extensive Search.



(a) Best pupil according to Genetic Algorithm search.

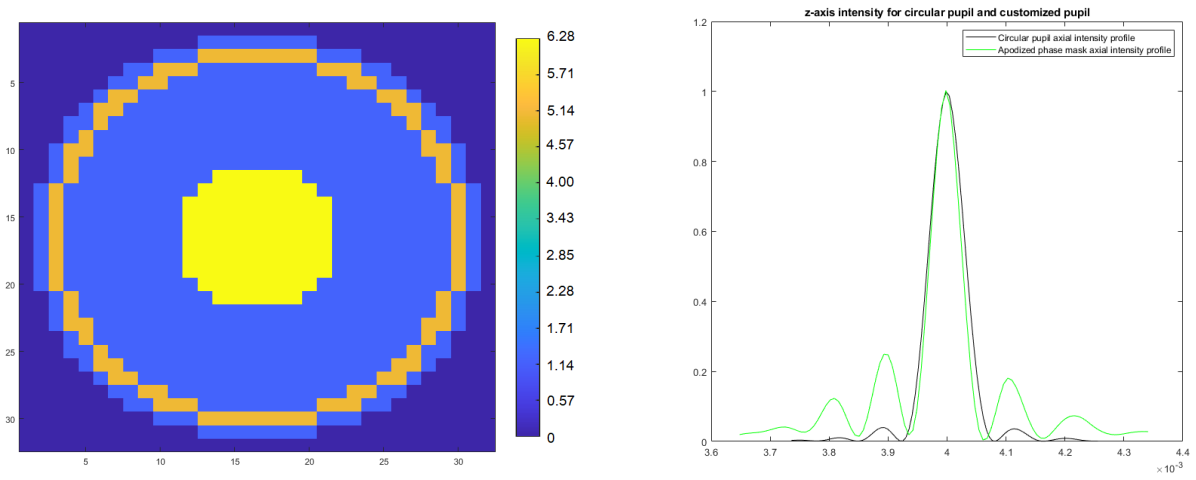


(b) Function corresponding to the pupil.

Figure 6.12: Genetic Algorithm Search.

Depth of Field is rich, containing other approaches using phase masks such as cubic phase masks [49], exponential phase pupil mask [88] and the asymmetric phase masks [89].
The final aim of the depth of field reduction work is to be usable for small sized cameras. For a smartphone camera in particular, the typical focal length used is about 4mm with a typical f/1.7-f/2.0 aperture.

In order to carry out some measurement of the possible improvement of the depth of field reduction, the formula $\Delta_z \propto \frac{\lambda}{N^2}$ (with N the numerical aperture) can be used for estimating an equivalent aperture.



(a) Best pupil according to PSO search.

(b) Function corresponding to the pupil.

Figure 6.13: Particle Swarm Optimization Search.

Lens Aperture	Compression Coefficient (%)	Equivalent Aperture
f/2.00	80%	f/1.79
f/2.00	75%	f/1.73
f/2.00	70%	f/1.67
f/1.80	80%	f/1.61
f/1.80	75%	f/1.56
f/1.80	70%	f/1.51
f/1.70	80%	f/1.52
f/1.70	75%	f/1.47
f/1.70	70%	f/1.42

Table 6.5: Equivalent aperture for several lens apertures and compression coefficients.

The difference in equivalent aperture is mitigated by possible aberrations such as chromatic aberration [90].

This result can be explained, from a physics point of view, by the formula connecting the depth of field as the inverse of the square dimension of the pupil $\Delta_z \propto \frac{\lambda}{N^2}$ (with N the numerical aperture). Because of this square relation, a variation of the depth of focus results from a square root variation of the aperture. From the observation, it is not possible to reduce the depth of field using a phase mask for more than half a stop of f-number.

6.9 Conclusion

In this chapter, the use of a phase mask for depth of field reduction is studied. A new PSO-based algorithm is tested in order to design the phase mask. The results obtained in the given configuration shows that it outperforms the state of the art algorithm.

References

[1] Haifeng Wang, Colin JR Sheppard, Koustuban Ravi, Seng Tiong Ho, and Guillaume Vienne. Fighting against diffraction: apodization and near field diffraction structures. *Laser & Photonics Reviews*, 6(3):354–392, 2012.

[2] Arnold Jan Den Dekker and A Van den Bos. Resolution: a survey. *Journal of the Optical Society of America A*, 14(3):547–557, 1997.

[3] Andra Naresh Kumar Reddy, Mahdieh Hashemi, and Svetlana Nikolaevna Khonina. Apodization of two-dimensional pupils with aberrations. *Pramana*, 90(6):77, 2018.

[4] Carlos J Zapata-Rodríguez, Mahdieh Hashemi, Mahin Naserpour, Amin Moazami, and Isaac Suarez. Metalenses with high-na, enhanced resolution and apodization. In *2017 19th International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2017.

[5] Kazumasa Tanaka, Kouichi Sakamoto, Toshiyuki Tanaka, and Mitsuo Taguchi. Method of designing an apodizer. *Applied Optics*, 34(28):6332–6336, 1995.

[6] Emiel H Por. Optimal design of apodizing phase plate coronagraphs. In *Techniques and Instrumentation for Detection of Exoplanets VIII*, volume 10400, page 104000V. International Society for Optics and Photonics, 2017.

[7] Manuel P Cagigal, José E Oti, Vidal F Canales, and Pedro J Valle. Analytical design of superresolving phase filters. *Optics communications*, 241(4-6):249–253, 2004.

[8] A Naresh Kumar Reddy and D Karuna Sagar. Two-point resolution of asymmetrically apodized optical systems. *Optica Pura y Aplicada*, 46(3):215–222, 2013.

[9] Rajib Kumar Panigrahi and Amit Kumar Mishra. A novel algorithm for apodization and super-resolution in fourier imaging. In *2011 International Conference on Devices and Communications (ICDeCom)*, pages 1–5. IEEE, 2011.

[10] Hubert F. A. Tschunko. Imaging performance of annular apertures. 3:apodization and modulation transfer functions. *Applied Optics*, 18(22):3770–3774, Nov 1979.

[11] LN Hazra. Walsh filters for tailoring of resolution in microscopic imaging. *Micron*, 38(2):129–135, 2007.

[12] CW McCutchen. Generalized aperture and the three-dimensional diffraction image. *Journal of the Optical Society of America*, 54(2):240–244, 1964.

[13] Kyoji Matsushima and Tomoyoshi Shimobaba. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Optics Express*, 17(22):19662–19673, 2009.

[14] J Lin, X-C Yuan, SS Kou, CJR Sheppard, OG Rodríguez-Herrera, and JC Dainty. Direct calculation of a three-dimensional diffracted field. *Optics Letters*, 36(8):1341–1343, 2011.

[15] David Miller and Leroy Meshel. Annular mask contact lenses, May 26 1998. US Patent 5,757,458.

[16] John M Evans, Chun-Shen Lee, Praful C Doshi, and Jerome A Legerton. Methods for fabricating annular mask lens having diffraction-reducing edges, October 12 1999. US Patent 5,965,330.

[17] Phanindra Narayan Gundu, Erwin Hack, and Pramod Rastogi. Apodized superresolution - concept and simulations. *Optics Communications*, 249(1-3):101–107, may 2005.

[18] CJR Sheppard and ZS Hegedus. Axial behavior of pupil-plane filters. *Journal of the Optical Society of America A*, 5(5):643–647, 1988.

[19] Brynmor Davis, William Karl, Anna Swan, M Unlu, and Bennett Goldberg. Capabilities and limitations of pupil-plane filters for superresolution and image enhancement. *Optics Express*, 12(17):4150–4156, 2004.

[20] Frederic Diaz, François Goudail, Brigitte Loiseaux, and Jean-Pierre Huignard. Comparison of depth of focus enhancing pupil masks based on a snr criterion after deconvolution. volume 27, 2010.

[21] Changyin Zhou and Shree K. Nayar. Aperture evaluation for defocus deblurring and extended depth of field. In *Aperture Evaluation for Defocus Deblurring and Extended Depth of Field*, 2012.

[22] Jiaze Wu, Changwen Zheng, Xiaohui Hu, and Fanjiang Xu. Rendering realistic spectral bokeh due to lens stops and aberrations. *The Visual Computer*, 29(1):41–52, 2013.

[23] Kazuya Kodama, Hiroshi Mo, and Akira Kubota. Virtual bokeh generation from a single system of lenses. In *ACM SIGGRAPH 2006 Research Posters*, SIGGRAPH 06, New York, NY, USA, 2006. ACM.

[24] Benjamin Huhle, Timo Schairer, Philipp Jenke, and Wolfgang Strasser. *Realistic Depth Blur for Images with Range Data*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[25] M. Kraus and M. Strengert. Depth-of-field rendering by pyramidal image processing. *Computer Graphics Forum*, 26(3):645–654, 2007.

[26] G. Toraldo di Francia. Nuovo pupille superresolventi. 7:366–372, 1952.

[27] Manuel Martinez-Corral, Amparo Pons, and Maria-Teresa Caballero. Axial apodization in 4pi-confocal microscopy by annular binary filters. *Journal of the Optical Society of America*, 19(8):1532–1536, Aug 2002.

[28] Noé Alcalá Ochoa and J. E. A. Landgrave. Non-iterative method for designing super-resolving pupil filters. *Optics Express*, 19(23):23613–23620, Nov 2011.

[29] D Mugnai, A Ranfagni, and R Ruggeri. Pupils with super-resolution. *Physics Letters A*, 311(2-3):77–81, 2003.

[30] D. Mugnai and A. Ranfagni. Further remarks on super-resolving pupils. *Journal of Applied Physics*, 102(3), 2007.

[31] Sudhakar Prasad, V. Paul Pauca, Robert J. Plemmons, Todd C. Torgersen, and Joseph van der Gracht. Pupil-phase optimization for extended-focus, aberration-corrected imaging systems. volume 5559, pages 335–345, 2004.

[32] Jeffrey A. Davis, Juan Carlos Escalera, Juan Campos, Andres Marquez, Maria J. Yzuel, and Claudio Iemmi. Programmable axial apodizing and hyperresolving amplitude filters with a liquid-crystal spatial light modulator. *Optics Letters*, 24(9):628–630, May 1999.

[33] Manuel Martinez-Corral, Pedro Andres, Jorge Ojeda-Castaneda, and Genaro Saavedra. Tunable axial superresolution by annular binary filters. application to confocal microscopy. *Optics Communications*, 119(5):491 – 498, 1995.

[34] M. Mino and Y. Okano. Improvement in the otf of a defocused optical system through the use of shaded apertures. *Applied Optics*, 10(10):2219–2225, Oct 1971.

[35] Vidal F. Canales, Jose E. Oti, and Manuel P. Cagigal. Three-dimensional control of the focal light intensity distribution by analytically designed phase masks. *Optics Communications*, 247(1-3):11–18, 2005.

[36] Zeev Zalevsky, Amir Shemer, Alexander Zlotnik, Eyal Ben Eliezer, and Emanuel Marom. All-optical axial super resolving imaging using a low-frequency binary-phase mask. *Optics Express*, 14(7):2631–2643, Apr 2006.

[37] Manuel Martinez-Corral. Point spread function engineering in confocal scanning microscopy. *Proc. SPIE*, 5182:112–122, 2004.

[38] J. L. Walsh. A closed set of normal orthogonal functions. *American Journal of Mathematics*, 45(1):5–24, 1923.

[39] Manuel P. Cagigal, Jose E. Oti, Vidal F. Canales, and Pedro J. Valle. Analytical design of superresolving phase filters. *Optics Communications*, 241(4-6):249–253, 2004.

[40] Pubali Mukherjee and Lakshminarayan Hazra. Far-field diffraction properties of annular walsh filters. *Advances in Optical Technologies*, 2013, 2013.

[41] Lakshminarayan Hazra and Pubali Mukherjee. Self-similarity in walsh filters and wave-optical engineering. In Kallol Bhattacharya, editor, *International Conference on Optics and Photonics 2015*. SPIE-Intl Soc Optical Eng, jun 2015.

[42] Hajime Nagahara, Changyin Zhou, Takuya Watanabe, Hiroshi Ishiguro, and Shree K Nayar. Programmable aperture camera using LCoS. In *Computer Vision–ECCV 2010*, pages 337–350. Springer, 2010.

[43] Tasso RM Sales and G Michael Morris. Axial superresolution with phase-only pupil filters. *Optics communications*, 156(4):227–230, 1998.

[44] Manuel Martinez-Corral, Amparo Pons, and Maria-Teresa Caballero. Axial apodization in 4pi-confocal microscopy by annular binary filters. *Journal of the Optical Society of America*, 19(8):1532–1536, Aug 2002.

[45] Jeffrey A. Davis, Juan Carlos Escalera, Juan Campos, Andres Marquez, Maria J. Yzuel, and Claudio Iemmi. Programmable axial apodizing and hyperresolving amplitude filters with a liquid-crystal spatial light modulator. *Optics Letters*, 24(9):628–630, May 1999.

[46] Gilbert Boyer. New class of axially apodizing filters for confocal scanning microscopy. *Journal of the Optical Society of America*, 19(3):584–589, Mar 2002.

[47] M. Martinez-Corral, C. Ibanez-Lopez, G. Saavedra, and M. T. Caballero. Axial gain resolution in optical sectioning fluorescence microscopy by shaded-ring filters. *Optics Express*, 11(15):1740–1745, Jul 2003.

[48] D. Rouan, P. Riaud, A. Boccaletti, Y. Clenet, and A. Labeyrie. The four-quadrant phase-mask coronagraph. i. principle. *Publications of the Astronomical Society of the Pacific*, 112(777):1479, 2000.

[49] Edward R. Dowski and W. Thomas Cathey. Extended depth of field through wave-front coding. *Applied Optics*, 34(11):1859–1866, Apr 1995.

[50] Sherif S. Sherif, W. Thomas Cathey, and Ed R. Dowski. Phase plate to extend the depth of field of incoherent hybrid imaging systems. *Applied Optics*, 43(13):2709–2721, May 2004.

[51] Qingguo Yang, Liren Liu, and Jianfeng Sun. Optimized phase pupil masks for extended depth of field. *Optics Communications*, 272(1):56 – 66, 2007.

[52] Albertina Castro and Jorge Ojeda-Casta neda. Asymmetric phase masks for extended depth of field. *Applied Optics*, 43(17):3474–3479, Jun 2004.

[53] Nicolas Caron and Yunlong Sheng. Polynomial phase mask for extending depth-of-field optimized by simulated annealing. volume 6832, pages 68321G–68321G–10, 2007.

[54] Feng Zhou, Ran Ye, Guangwei Li, Haitao Zhang, and Dongsheng Wang. Optimized circularly symmetric phase mask to extend the depth of focus. *Journal of the Optical Society of America*, 26(8):1889–1895, Aug 2009.

[55] Richard Barakat and Agnes Houston. The effect of a sinusoidal wavefront on the transfer function of a circular aperture. *Applied Optics*, 5(11):1850–1852, Nov 1966.

[56] Hui Zhao and Yingcai Li. Optimized sinusoidal phase mask to extend the depth of field of an incoherent imaging system. *Optics Letters*, 35(2):267–269, Jan 2010.

[57] Rudolf Karl Luneburg and Max Herzberger. *Mathematical theory of optics*. Univ of California Press, 1964.

[58] Richard Barakat. Solution of the luneberg apodization problems∗. *Journal of the Optical Society of America*, 52(3):264–275, Mar 1962.

[59] C. W. McCutchen. Generalized aperture and the three-dimensional diffraction image. *Journal of the Optical Society of America*, 54(2):240–244, Feb 1964.

[60] J. Lin, 5 X.-C. Yuan, * S. S. Kou, 3 C. J. R. Sheppard, 3 O. G. Rodriguez-Herrera, , and J. C. Dainty. Direct calculation of a three-dimensional diffracted field. volume 36. Optics Letters, 2011.

[61] Joseph W Goodman. *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

[62] Nicholas George and Wanli Chi. Extended depth of field using a multi-focal length lens with a controlled range of spherical aberration and a centrally obscured aperture, February 26 2008. US Patent 7,336,430.

[63] L. N. Hazra and N. Reza. Optimal design of toraldo super-resolving filters. *SPIE, Novel Optical Systems Design and Optimization XIII*, (77870D), 2010.

[64] Fred Glover. Tabu search - part i. *ORSA Journal on computing*, 1(3):190–206, 1989.

[65] Scott Kirkpatrick, C Daniel Gelatt, Mario P Vecchi, et al. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[66] Lester Ingber and Bruce Rosen. Genetic algorithms and very fast simulated reannealing: A comparison. *Mathematical and Computer Modelling*, 16(11):87 – 100, 1992.

[67] Fast simulated annealing. *Physics Letters A*, 122(3):157 – 162, 1987.

[68] Eugeniusz Nowicki and Czeslaw Smutnicki. A fast tabu search algorithm for the permutation flow-shop problem. *European Journal of Operational Research*, 91(1):160 – 175, 1996.

[69] L. Ingber. Very fast simulated re-annealing. *Mathematical and Computer Modelling*, 12(8):967 – 973, 1989.

[70] Fred Glover, James P. Kelly, and Manuel Laguna. Genetic algorithms and tabu search: Hybrids for optimization. *Computers and Operations Research*, 22(1):111 – 134, 1995.

[71] D. Adler. Genetic algorithms and simulated annealing: a marriage proposal. In *IEEE International Conference on Neural Networks*, pages 1104–1109 vol.2, 1993.

[72] James Kennedy and Russell Eberhart. Particle swarm optimization. volume 4, pages 1942–1948. IEEE, 1995.

[73] P. J. Angeline. Using selection to improve particle swarm optimization. In *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, pages 84–89, May 1998.

[74] Xiao-Feng Xie, Wen-Jun Zhang, and Zhi-Lian Yang. Dissipative particle swarm optimization. In *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, volume 2, pages 1456–1461, 2002.

[75] Liping Zhang, Huanjun Yu, and Shangxu Hu. *A New Approach to Improve Particle Swarm Optimization*, pages 134–139. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[76] F. van den Bergh and A. P. Engelbrecht. A cooperative approach to particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8(3):225–239, June 2004.

[77] Bo Liu, Ling Wang, Yi-Hui Jin, Fang Tang, and De-Xian Huang. Improved particle swarm optimization combined with chaos. *Chaos, Solitons and Fractals*, 25(5):1261 – 1271, 2005.

[78] Yan Jiang, Tiesong Hu, ChongChao Huang, and Xianing Wu. An improved particle swarm optimization algorithm. *Applied Mathematics and Computation*, 193(1):231 – 239, 2007.

[79] Marco Dorigo, Mauro Birattari, Christian Blum, Maurice Clerc, Thomas Stützle, and Alan Winfield. *Ant Colony Optimization and Swarm Intelligence: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008, Proceedings*, volume 5217. Springer, 2008.

[80] Z. H. Zhan, J. Zhang, Y. Li, and H. S. H. Chung. Adaptive particle swarm optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(6):1362–1381, Dec 2009.

[81] Gai-Ge Wang, Amir Hossein Gandomi, Xin-She Yang, and Amir Hossein Alavi. A novel improved accelerated particle swarm optimization algorithm for global numerical optimization. *Engineering Computations*, 31(7):1198–1220, Sep 2014.

[82] Ke-Lin Du and M. N. S. Swamy. *Particle Swarm Optimization*, pages 153–173. Springer International Publishing, Cham, 2016.

[83] S. S. Aote, M. M. Raghuwanshi, and L. Malik. A brief review on particle swarm optimization: Limitations and future directions. *International Journal of Computer Science Engineering*, 2(5):196–200, 2013.

[84] Daniel W. Boeringer and Douglas H. Werner. Particle swarm optimization versus genetic algorithms for phased array synthesis. *IEEE Transactions on Antennas and Propagation*, 52(3):771–779, 2004.

[85] Russell C. Eberhart and Yuhui Shi. *Comparison between Genetic Algorithms and Particle Swarm Optimization*, pages 611–616. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[86] Emad Elbeltagi, Tarek Hegazy, and Donald Grierson. Comparison among five evolutionary-based optimization algorithms. *Advanced Engineering Informatics*, 19(1):43 – 53, 2005.

[87] M. Clerc and J. Kennedy. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73, Feb 2002.

[88] Qingguo Yang, Liren Liu, and Jianfeng Sun. Optimized phase pupil masks for extended depth of field. *Optics Communications*, 272(1):56 – 66, 2007.

[89] Albertina Castro and Jorge Ojeda-Casta neda. Asymmetric phase masks for extended depth of field. *Applied Optics*, 43(17):3474–3479, Jun 2004.

[90] Maria J. Yzuel, Juan C. Escalera, and Juan Campos. Polychromatic axial behavior of axial apodizing and hyperresolving filters. *Applied Optics*, 29(11):1631–1641, Apr 1990.

Chapter 7

Design of an ultra-fast lens for depth estimation

In order to improve the depth capabilities of an optical system, we have shown in Chapter 4 that the axial resolution is related to the aperture of the lens and the lateral resolution. Following the work in Chapter 6, it is important to be able to compare the results from phase mask apodization with the possibilities brought by lens design. Namely, this chapter will focus on achieving depth of field reduction through lower f-number lens design.

7.1 Introduction

The number of smartphone users worldwide is estimated at around 2.5 billion with 1.45 billion smartphones produced in 2018. The camera module of smartphones is an important element of the device. According to Deloitte TMT 2016 [1], 90% of all pictures are taken with a smartphone camera and 2.5 trillion photos were shared or stored on-line globally in 2016, of which about 1.875 trillion photos were shared through social media and messaging services. Camera improvement is also one of the main reasons for people to change their phones.

However, the camera module remains an optical design challenge because of its small form factor, the optical materials available and the cost constraints [2]. In 2020, the typical cost for a smartphone camera module is in the $\$ 10-30$ range, including the sensor and associated electronics. However some manufacturers propose multiple camera systems which increase the cost of the camera. Multiple camera modules are used to enable new functionalities thanks to computational imaging [3] (i.e. several focal lengths, depth mapping, extended spectrum sensitivity) or improved performance.

Despite a wide number of attempts to mimic the aesthetic qualities of high end portraits such as the background blur produced by wide aperture lenses, smartphone camera systems can only achieve limited depth estimation as shown in [4] and consequently provide poor depth of field

management and blurring capabilities.

In this chapter, we explore the possibilities for ultra-fast miniature camera lenses, with a potential smartphone application in mind. Over the last decade, optical design for smartphones has been incremental and required the aperture stop to be placed in front of the optics. This chapter explores the possibility of a double-Gauss type system with a central stop position and a considerably wider pupil. The aim of this investigation is not to deliver unprecedented lateral resolution in the image plane but rather to propose acceptable lateral resolution with a wide aperture. As such, our target is not the diffraction limit. The original contribution of this chapter resides both in the ultra-wide aperture of the system and in the use of a central stop position for a miniature lens.

Section 7.2 provides an overview of the state of the art for smartphone lens design. Section 7.3 justifies the choice of the starting optical design and gives some experimental comparisons between three designs depending on the number of optical elements. Section 7.6 presents the proposed design for an f/1.0 aperture smartphone lens, with a performance analysis and tolerancing reports. Section 8.1.1 discusses the possibility of lenses with an aperture larger than f/1.0 and the possible technological improvements for miniature lenses.

7.2 State of the art of optical design for miniature lenses

7.2.1 Sensor size in smartphone industry

For a given resolution, the pixel size is an important parameter of the camera image quality. For a given sensor resolution, bigger sensors (i.e. bigger pixels) improve the SNR in low light and increase the depth resolution. As of 2019, most smartphones are equipped with sensors having a diagonal between 1/4 inch (4.5mm diagonal) and 1/2.5 inch (7.2mm diagonal), with 1/3 inch sensors (6.0mm diagonal) being typical for their main camera. High-end phones tend to have larger sensor sizes and smaller sized sensors are often used to create long focal length auxiliary cameras.

7.2.2 Flat sensor designs

As of 2019, the typical lens f-number for miniature cameras is in the f/1.8-f/2.0 range with a few lenses having f/1.7, f/1.6 or even a variable f/1.5 aperture [5]. The trend for the aperture of lenses in the mobile phone industry is an increase in aperture size over the years, from a typical value of f/2.8 in the late 2000s. The choice of a wider aperture is critical for the increase in angular resolution which is ultimately limited by diffraction to $\theta_R = \frac{\lambda}{D}$. The resolution achieved by miniature camera lenses is essentially diffraction limited and as such, the aperture increase has enabled higher resolution images. The image quality is theoretically limited by shot noise which is inversely proportional to the light flux collected on the detector. For photometric reasons, larger

apertures improve the SNR in the images produced, at the expense of increased optical aberrations.

State of the art miniature lens designs feature a frontal aperture stop or second element aperture stop [6] and typically 5 to 6 elements, the last one functioning as a field flattener [7]. The frontal position of the stop reduces aberration correction compared to a central stop position. Central aperture optical systems have a typically better aberration control since some of the aberrations cancel out due to symmetry. Each surface used in the current optical designs is aspheric. Since the lenses are moulded and not ground, highly aspherical surfaces are both easier and cheaper to produce compared to conventionally ground glass surfaces while exhibiting some interesting optical properties making them suitable for aberration correction. However, such lenses are currently made of polymer materials which offer less diversity in refractive index and Abbe number compared to glasses.

Recently, folded designs have started to appear for some smartphones to enable an additional telephoto lens complementing the main wide angle camera. Such systems are either single-folded using an angled mirror or prism [8] or double-folded [9, 10]. The latter uses, for instance, a combination of a prism and mirror to achieve a longer focal length with fast aperture within the thickness constraints of a smartphone camera.

7.2.3 Curved sensor designs

Curved sensors remove or reduce the need for field flattening of the image plane. As a result, researchers have proposed wide-aperture lens designs adapted for such sensors [11, 12]. Such systems can reach an f/1.2 aperture while remaining compact [13]. As pointed out by [11], the use of a curved sensor is well adapted for non-symmetrical lens designs. Contrary to interchangeable lens systems (e.g. DSLR) and zoom lenses, miniature cameras have a fixed lens and a fixed focal length associated with a single imaging plane curvature. This makes the use of curved sensors suitable for smartphones.

7.2.4 Non-smartphone applications

Miniature lenses are mostly produced for smartphone and tablet applications, but can find other uses: security cameras, car systems, industrial inspection, video or photographic devices. In the context of security cameras, the wide aperture is often desired as it provides a greater light flux in dim light. To our knowledge, for such a focal length and sensor size, an f/1.2 aperture is the commercially available state of the art. For video purposes, subject isolation and good photometric capabilities are often desired. Since the resolution expected is typically lower than that expected for still images, making wide aperture lenses is highly desirable for action cameras or video blog cameras.

7.3 Considerations for the design of an f/1.0 smartphone lens

7.3.1 Design considerations among the double-Gauss systems

The chosen starting point for the design is a double-Gauss type system. The system has been extensively used for photographic lenses in the second half of the XXth century and is still used in modern entry-level lenses. The double Gauss lens design traditionally comprises 6 to 8 elements in a quasi-symmetrical design. The lens cannot be designed perfectly symmetrical since it would result into a 1:1 magnification ratio. The 4 central elements (2 before the stop and 2 after the stop) are cemented doublets. The external lenses are usually singlets.

A very rough approximation of the double-Gauss lens design is the triplet model proposed in [14]. In this model the 2 central cemented doublets are playing the role of the diverging lens from a triplet.

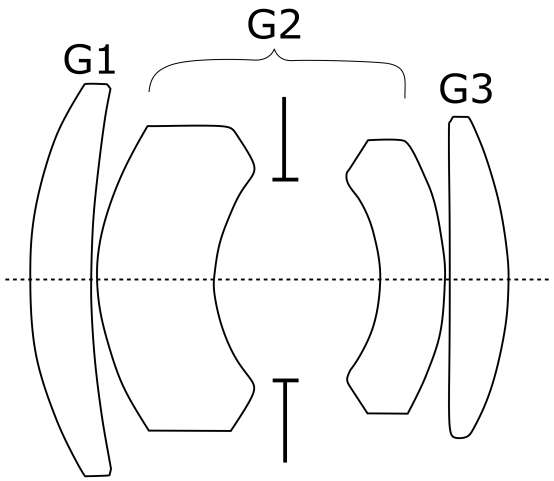


Figure 7.1: Double-Gauss lens including 6 elements, displayed as a triplet model.

The number of elements used for the double-Gauss system depends on the targeted f-number of the system. Most of the f/2.0 lenses feature 6 elements while, f/1.4 lenses usually have 7 elements. Faster lenses tend to have extra elements for better aberration correction.

The double-Gauss system provides good correction of optical aberrations (and is consequently adapted to fast lenses) but has a rather limited field of view making it suitable for only a limited range of focal lengths. For a 24×36mm sensor size, this is typically in the 35mm to 135mm focal range.

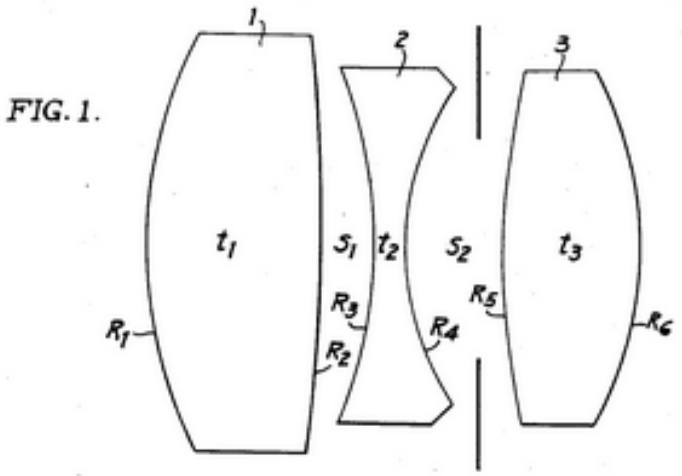


Figure 7.2: Original Triplet lens proposed by Cooke in the patent [15] (Public Domain).

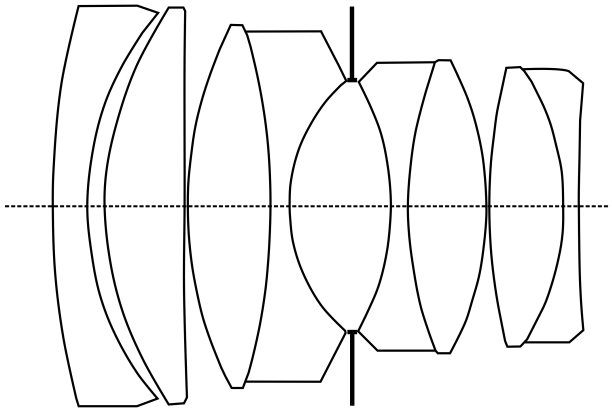


Figure 7.3: Example of fast double-Gauss lens system: the Leica Noctilux 50mm f/0.95.

The double-Gauss lens design has several limitations for miniature camera applications. The first is the range of focal lengths available. Most smartphones offer wide angle or ultra-wide angle lenses. There are several reasons for this choice: the limited thickness of the phone, the limited distance from the user to the phone in the case of a self-portrait, the possibility to digitally crop inside a wide angle image but not to extrapolate the information for a wider angle field of view. As of 2020, the typical 35mm equivalent focal length for phone camera standard lenses are in the 20mm to 24mm range.

7.3.2 Number of optical elements

State of the art double-Gauss lenses typically feature a 6 element design. However, a greater number of elements is usually required to correct the greater optical aberrations resulting from a wider aperture.

A quick analysis, with few ray traces, show that 6 elements are not sufficient to correct all optical aberrations, especially on the external part of the field of view. However, an 8 element design could provide an even more homogeneous performance across the frame.

7.3.3 The difficult question of tolerancing

Tolerancing is always a real limit for the design of lenses. Manufacturing of the lenses can only be done up to a certain level of precision. For this reason, the lens design should look into the limitations of the manufacturing. Some lens designs that, in principle, should provide great image quality cannot deliver a useful image once produced.

The topic of lens tolerancing is extremely challenging to grasp: firstly because it brings together several fields of knowledge (mechanical design, lens manufacturing, assembly statistics...) and secondly because a large part of this information is unavailable (confidential statistics or manufacturing process).

A first attempt at tolerancing lenses is to set some error ranges in an optical design software. Although helpful, one should remember that this is not the full story. Manufacturers may require additional factors to be taken into consideration such as the ratio between the thickness of the lens at the center and edges, or the overall shape of the lens.

7.4 Use of Zemax software

7.4.1 Zemax Layout and features

Zemax is an optical design software. Four main tabs have been used through the design process: Setup, Analyse, Optimize and Tolerance.

The Setup tab gathers all the controls related to the object position and size, the points used in the image space to evaluate the performance, the wavelengths as well the lens aperture.

The Analyse tab features a range of metrics, graphs and tools to estimate the lens performance. Of these, the following have been extensively used in the design process:

- Rays and Spots menu
- Aberration menu
- MTF menu

- Encircled energy menu
- Extended scene analysis

The optimize tab contains an optimization wizard to design a merit function. This merit function can be customized at will to include a broad range of aberration metrics for each surface of the lens system. The Zemax user manual gives additional information regarding to the available functions and metrics.

The optimization process is limited to three choices : optimize, global search and hammer current. From a user perspective, it seems that the former is designed for quicker but less accurate optimization, whereas the latter is designed for a longer and more global search.

The tolerance tab provides the lens designer with some Monte Carlo simulations in order to estimate how much the lens performance can be affected by the production process.

7.4.2 How Zemax was used during this research

The first step for designing the lens is to setup the working environment of the system (wavelengths, object plane, image plane, sensor size). Most of the time a starting point can be found from the litterature. In the case where no starting points are available, one should start with a smaller aperture than the final one. In our case, typically f/4.0. Then, one can vary only a few surfaces to find a coarse system with spherical surfaces.

Assuming such a coarse system exists, the next step aims at increasing the aperture to the desired value. Of course, this will result in more optical aberrations. As a consequence, new degrees of freedom must be provided for optimization purposes. Zemax offers a function to find which surface benefits the most from being converted to aspheric. As a rule of thumb, it seems to be a good practice to start from the surfaces far from the lens pupil which have more optical power.

Occasionally the optimization result runs into a configuration that does not enable further aperture variations or impossible to manufacture shapes. As a rule of thumb, this is usually a consequence of large apsheric coefficients. One could stop down the system, reduce the value of the aspheric coefficients and run some optimization again.

7.5 Early design

7.5.1 Selection of a starting point

In the photographic lens literature [16], only double-Gauss systems are used for f/1.0 or larger apertures. Such lenses can be optimized using computer software as has been shown as early as 1967 [17]. The standard double-Gauss system is designed with a very large back-focal length in order to accommodate the mirror used in DSLR systems. For mobile phone application, large back-focal distances are not possible and the lens system must be re-designed to reduce this distance.

Double-Gauss lenses usually come with a minimum of 6 elements, usually 2 cemented doublets are used around the lens stop as can be seen in figure 7.4. The wider aperture systems usually include extra elements at the back of the lens to better correct aberrations. A 7 element double-Gauss system would typically include 3 frontal elements and 4 rear elements. For a smartphone application, the last element will also be used as a field flattener. The proposed lens is shown in Figure 7.15, with a total track length of 8.2mm which is acceptable for a smartphone application.

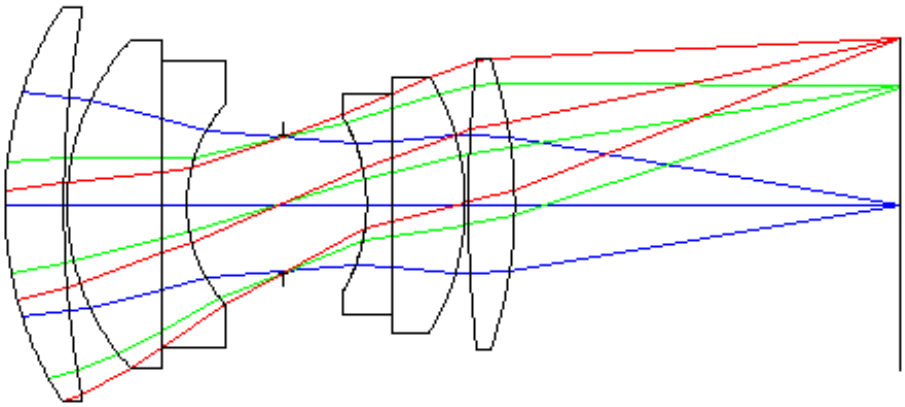


Figure 7.4: Example of DSLR optical design for a 6 elements double-Gauss lens system as proposed by [18]. Source: Doc Phil (Public Domain).

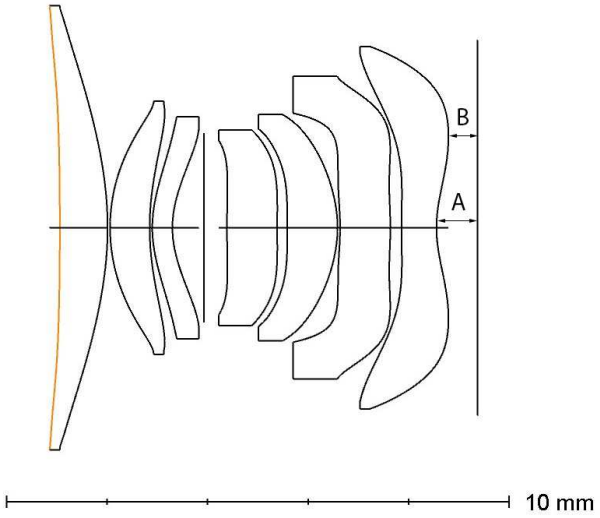


Figure 7.5: The proposed lens design for an f/1.0 smartphone lens with a 5mm focal length and 8.2mm total track length. A: Distance from last lens element centre to image plane. B: Distance from last element to image plane.

The double-Gauss optical design uses the central position of the stop in the system to balance the aberrations and compensate most of the coma, distortion and lateral chromatic aberration. In a perfectly symmetrical system, the odd Zernike polynomials are compensated almost perfectly (it is possible to compensate perfectly for a given aberration with one to one magnification, but not for all at the same time). However, to a good approximation, this property remains valid for quasi-symmetrical systems such as the double-Gauss type of lenses.

In practice, the smartphone industry does not use cemented doublets since they require additional steps in the manufacturing process. Instead, all elements are singlets which enables extra degrees of freedom from an optical design point of view.

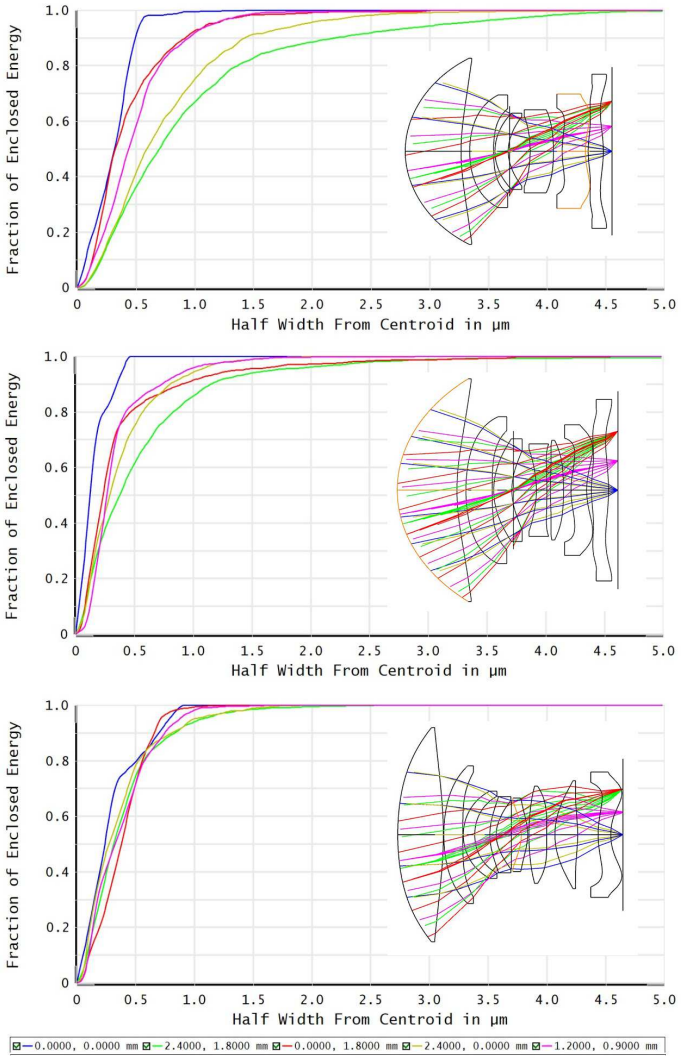


Figure 7.6: Comparison of the en-squared energy performance for 6-elements, 7-elements and 8-elements lens designs optimized for 5 image points for a 1/3 inch sensor. No limits have been imposed on the shape of the elements and manufacturability.

7.5.2 Number of elements for an f/1.0 lens

There is no theoretical way to find the number of optical elements needed to correct the aberrations of a system for a given aperture and field of view. Our proposed design features 7 elements, although one may wonder if 6-element or 8-element systems could be used to achieve similar or better performance.

The optimal number of lens elements must be kept low in order to improve manufacturability. Current standards for production are 5-6 lens elements. A number of lens elements greater than 8 is currently out of reach for cost, manufacturing and tolerancing reasons. Although a first lens system has been explored using an 8 element design, the final system was re-engineered to a 7 element design since it was more robust to the production process and ultimately presented more

relaxed tolerances.

Three optical systems with central aperture have been designed with performance analysis at 5 field points with the results shown in Figure 7.6. The target focal length for this sensor size is 5mm, which corresponds to 36mm full frame equivalent. The designs show a good example of the limits brought about by a low number of elements. In the case of 6 elements, the central performance is acceptable, but the edges and corners suffer from a significant drop in resolution. In the case of 7 elements, the periphery of the image gets some improvement but the corners cannot match the central performance. The 8-elements design shows a very even resolution across the frame but a tolerance study shows that 8-elements design are more difficult to manufacture with state of the art lens production precision.

7.5.3 Opto-mechanical constraints imposed by manufacturing

When designing the lens, several constraints must be taken into consideration. One of them is the back-focal length (the distance between the last optical element and the imaging plane). Although smartphone systems do not require extra space for flip mirrors like DSLRs, some distance is still required for the micro-lens array and colour filters. This distance is typically 0.6mm. Since the last lens is often used as a field flattener, the back-focal distance is critical. The effectiveness of the field flattener is reduced if placed further away from the image plane.

7.5.4 On-board image flaws correction and optical design choices

Smartphone cameras are used for both still images and videos. For still images, image processing applied to the raw data can correct many optical flaws in the system such as distortion, vignetting and lateral chromatic aberration. For videos however, the correction must be real-time at typically 30 frames per second. The real-time correction limits the use of on-board image processing. It is for instance possible to correct distortion in real-time but not vignetting since some advanced denoising is needed.

7.5.5 Conclusions from preliminary work

The preliminary work concluded that a minimum of seven lens elements are required, that the minimum back focal length must accommodate for a micro-lens array and colour filters and that distortion can be accepted and corrected in real-time for video applications but vignetting is a more critical factor for the lens design.

7.6 Double-Gauss inspired miniature lens

7.6.1 Overview of the lens

The lens system proposed, presented in Figure 7.15, is designed for a 1/3 inch sensor with size 4.8mm × 3.6mm. The system has been optimized for a 1m distance from the subject to the camera, which corresponds to a typical portrait distance. Nowadays, the typical thickness for a smartphone is in the $7-10$mm range. Most camera lenses thus have a Total Track Length (TTL) around $5-6$mm even if some smartphones are thicker (e.g. Energizer Power Max P18K Pop [19]). The proposed system has a TTL of 8.2mm from frontal element to the sensor. The first lens element has been designed to be rather flat for optimal compactness.

In order to limit vignetting, the actual imaging circle has been extended to cover a 1/2.7 inch sensor (5.37 × 4.04mm), however the resolution drops quickly towards the edges and only a 1/3 inch sensor is recommended for this lens design. Figure 7.7 illustrates the vignetting effect on a 4.8mm × 3.6mm sensor format.

The proposed lens has been designed with a target focal length of 5mm, corresponding to a 36mm equivalent focal length for a 24 × 36mm full frame sensor. The field of view is 54 × 38 degrees.

Figure 7.7: Vignetting for the proposed 1/3 inch sensor size format. On the left, the raw image, on the right, the corrected image both calculated using Zemax Image Simulation Toolbox.

7.6.2 Choice of materials

The proposed system comprises 5 elements made of K26R and 2 elements made of EP5000. The simulations are made for a 25 degree Celsius temperature. Under these standard conditions, K26R has a refractive index $n = 1.53$ and an Abbe number $V = 56$ whereas EP5000 has a refractive index $n = 1.63$ and an Abbe number $V = 24$. As such, K26R corresponds to a crown glass and EP5000 corresponds to a flint glass. Both plastic materials are industry-standards and can be moulded in order to achieve aspherical lens elements.

7.6.3 Role of each element

The TTL is 8.3mm, with a back-focal length designed to be at least 0.82mm from the centre of the lens (distance A in Figure 7.15), or 0.6mm from the imaging plane (distance B in Figure 7.15).

The asphericity of the lens elements is described using polynomials, of which the first 7 terms are used in addition to the radius of curvature [20]. From left to right in Figure 7.15 and Figure 7.9, one can observe:

- The first and second lens elements have a positive optical power.
- The third lens element has little optical power, it mostly helps to reduce the spherical aberration and coma introduced by the first two lenses.
- The lens stop is in the front part of the lens system.
- The fourth lens element has little power and helps correct some of the distortion.
- The fifth lens element has negative optical power and also corrects for distortion.
- The last two elements act as field flattener as well as reducing astigmatism and coma aberrations.

As stated in the section 7.6.1, this layout is thicker than what is currently accepted in most phones, although not critical since it could fit in a $10-12$mm thick phone. The minimum element thickness is 0.4mm (third element) and the maximum thickness is 1.0mm (sixth element). The air distance is reduced to a minimum, being almost zero except around the stop of the system (0.63mm before and 0.46mm after) and around the last element (0.223mm before and 0.82mm after).

7.6.4 Performance

The lens exhibits some pincushion distortion. The distortion is around -2% for the 1/3 inch sensor size and up to -9% for the larger 1/2.7 inch size. Several aberrations are still visible such as spherical aberration, astigmatism and to a lesser extent some coma and field curvature. The Seidel coefficients are shown in Figure 7.9.

The resolution achieved by the lens across the frame is described in figure 7.10. Comparing the proposed system to state of the art lenses is only relevant to a certain extent. The designed lens does not aim at greater resolution, rather improved low-light performance and better portrait capabilities. The lens design has been done with a realistic target given the aperture of the lens. To be clear, the design cannot achieve diffraction limited resolution. According to preliminary work, it seems that such a performance is only possible with an extra optical element.

The proposed lens performance is comparable to the work proposed by Steinich and Blahnik [7] which was designed for a 5MPx sensor. Although 5MPx is below state of the art smartphone device capabilities, it is enough for video purposes, depth mapping application or social media activities.
The Modulation Transfer Function (MTF) corresponds to a diffraction limited f/2.4 lens for the centre of the frame and worsens slightly at the edges of the image. MTF curves are presented in Figure 7.11 where the MTFs are computed using the diffraction method from the Zemax MTF Analysis Toolbox. The f/1.0 diffraction limit is visible and the cut off frequency is chosen to be 900 cycles/mm, roughly equivalent to the cut off frequency of an f/2.0 lens. The performance across the image is not perfectly uniform. The MTF undergoes a drop inside the image as shown in Figure 7.10. Maintaining the homogeneity of the lens performance across the frame is challenging with only 7 lens elements and an 8-element design would be preferable here. Depending on the application, one could either preferably optimize for the centre of the image (portrait application) or lower the resolution across the frame for better uniformity.

The Chief Ray Angle (CRA) for the proposed lens system peaks at 28 degrees as reported in figure 7.8. The 16.5MPx sensor OV16850 is designed for a chief ray angle of 33.4 degrees. The result obtained is close to that proposed in state of the art [21]. The sensor CRA criterion depends on the micro-lens array used in front of the sensor, which can be customized if needed.

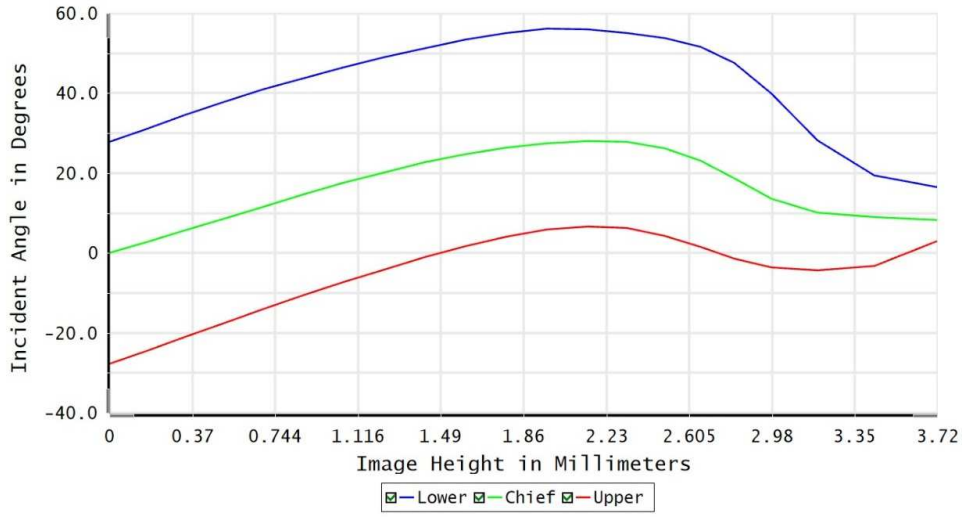


Figure 7.8: Chief Ray Angle across the frame.

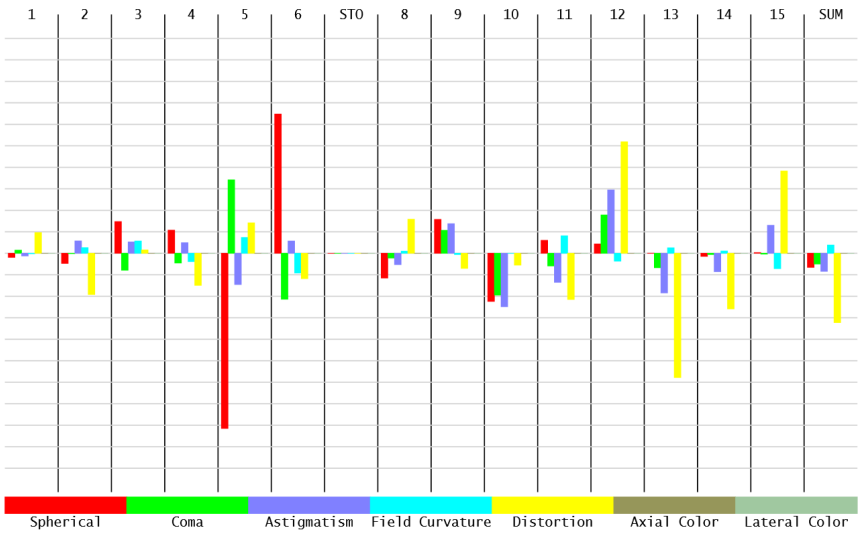


Figure 7.9: Seidel aberration diagram showing the aberration contribution of each lens surface. Wavelength: 550nm, maximum scale 5mm.

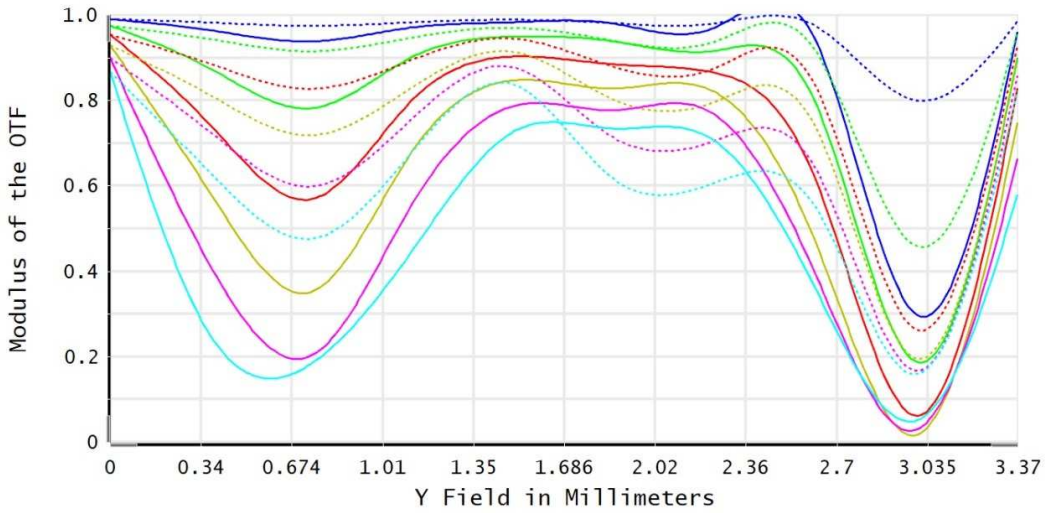


Figure 7.10: Geometrical Modulation Transfer Function (MTF) inside the image frame. Blue: 10 cyc/mm, Green: 20 cyc/mm, Red: 30 cyc/mm, Yellow: 40 cyc/mm, Pink: 50 cyc/mm, Turquoise: 60 cyc/mm. Plain: Tangential rays, Dashed: Sagittal rays.

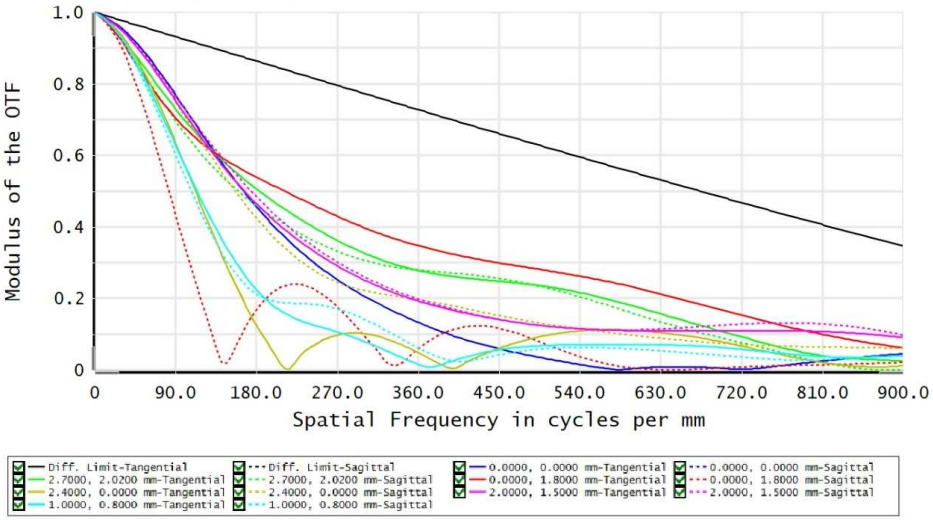


Figure 7.11: Modulation Transfer Function (MTF) on a scale from 0 cycles/mm to 900 cycles/mm which corresponds to the cut off frequency of an f/2.0 smartphone lens.

7.6.5 Tolerancing

Tolerancing aims at simulating the real life performance obtained from a system built in an industrial process. It takes into account the manufacturing uncertainty for each element (radius of curvature, asphericity, thickness, local variation of the Abbe number and refractive index) and the assembly of the elements (centring, tilt of the elements, spacing between elements) to assess the manufacturability of the lens system [22].

For this lens design, the tolerancing has been estimated using the Zemax tolerancing tool. The tolerances used can be found in Table 7.1 and have been obtained from discussion with manufacturing experts and values found in the literature [23]. The tolerances used by Steinich and Blahnik [7] are more strict than the ones used for the proposed system.

Surface Tolerances	
Radius	0.1 fringe
Thickness	2μm
Decenter X	2μm
Decenter Y	2μm
Tilt X	0.4μm
Tilt Y	0.4μm
S+A Irregularity	0.1 fringe
Element Tolerances	
Decenter X	2μm
Decenter Y	2μm
Tilt X	0.1 degree
Tilt Y	0.1 degree
Glass Tolerances	
Refractive Index	0.001
Abbe Number	1%

Table 7.1: State of the art industrial tolerances for the smartphone industry.

Imaging systems	100.0% of systems produce an image	
Nominal spot size	1.99μm	1
Mean RMS	2.70μm	1.36
Standard Deviation	0.48μm	0.24
Distribution of the results		
98% systems	3.88μm	1.95 of nominal spot size
90% systems	3.39μm	1.70 of nominal spot size
80% systems	3.31μm	1.55 of nominal spot size
50% systems	2.60μm	1.30 of nominal spot size
20% systems	2.26μm	1.13 of nominal spot size
10% systems	2.16μm	1.09 of nominal spot size
2% systems	2.06μm	1.04 of nominal spot size

Table 7.2: Number of imaging systems and Monte-Carlo performance, simulated on 10 000 systems with the state-of-the-art tolerances.

The tolerancing performance reported in Table 7.2 shows that the system is manufacturable, in the simulations, most of the systems (over 98%) deliver an image spot size of less than twice the nominal system. This ensures that the image quality of the system is homogeneous through manufacturing and that later image processing can be applied to the images. When it comes to tolerancing, the system under-performs the lens proposed by C. Sheil and A. Goncharov [10] which has a worst spot size of 1.77 times the nominal spot size (versus 2 times for the proposed system).

7.7 Variations of the design for specific needs

The purpose of this section is to present several scenarios based on the previously proposed starting point. The first subsection proposes an f/1.0 system with a low resolution and good tolerancing

properties. The second subsection proposes an f/1.1 near-diffraction limited system.

7.7.1 Performance for an f/1.0 system

The proposed lens design in Figure 7.12 has the same overall dimensions and properties as the starting point. The third element of the lens exhibits a curvature that might require special attention for the manufacturing process (taking the element out of the mold might damage the lens). The system can find some applications in niche markets such as video cameras and low resolution still images (security or industrial imaging) where low light performance is more critical than the resolution.

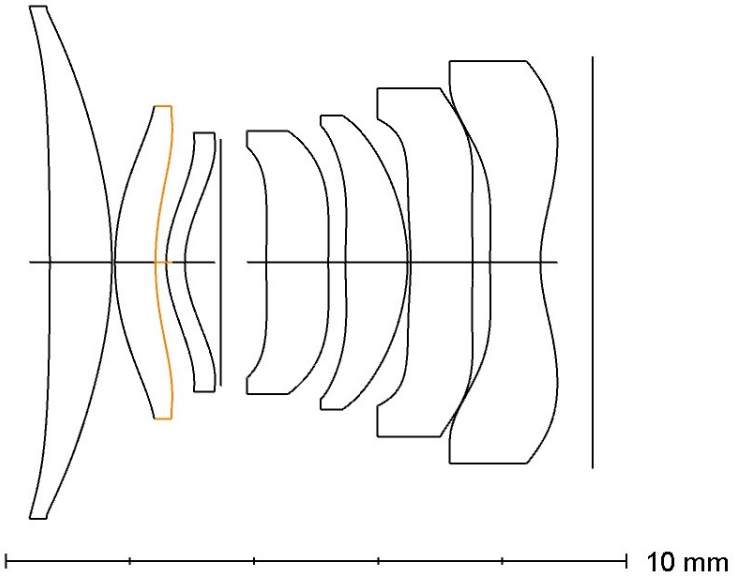


Figure 7.12: Proposed lens design.

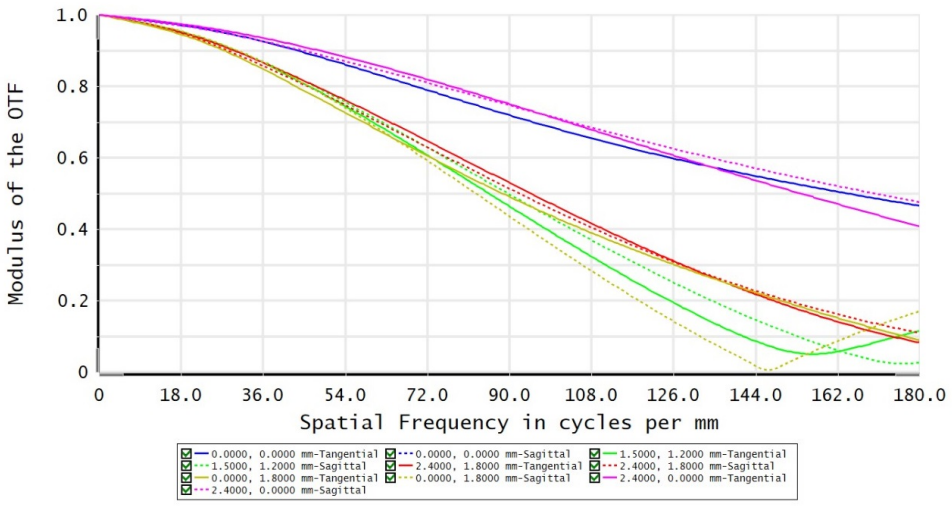


Figure 7.13: MTF curves for a maximum 180 cycles/mm Nyquist frequency.

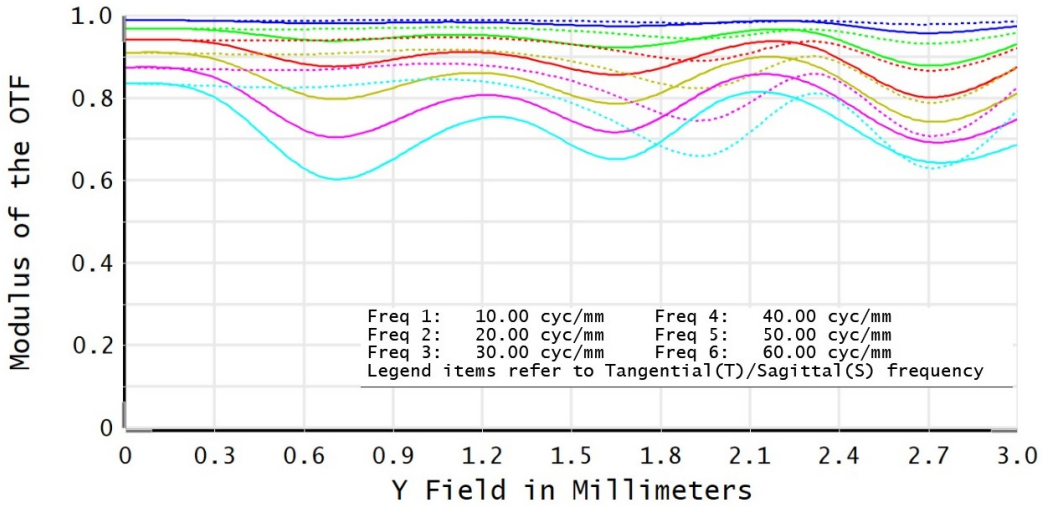


Figure 7.14: MTF curves versus Field.

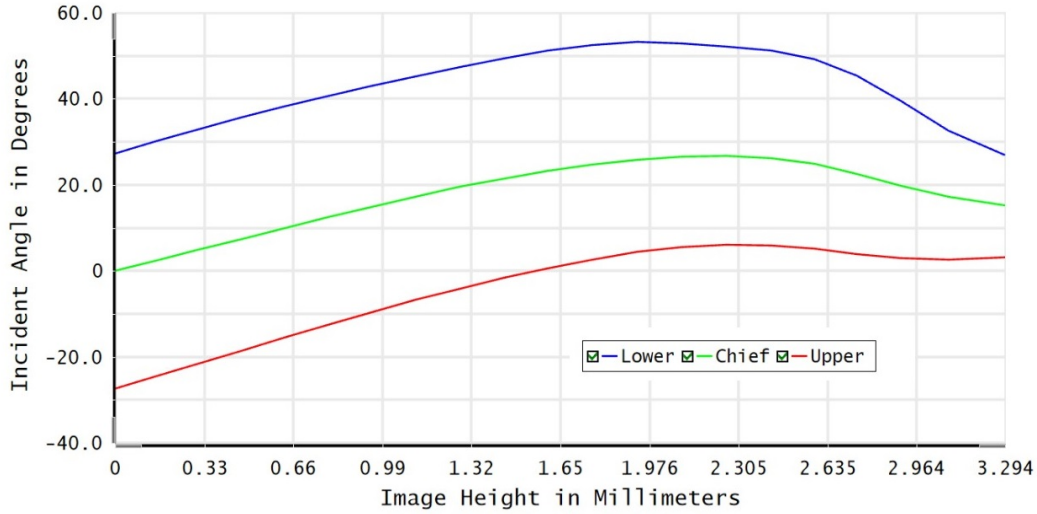


Figure 7.15: Chief Ray Angle.

The lens tolerances can be found in Table 7.3.

Imaging systems	100.0% of systems produce an image	
Nominal spot size	2.21μm	1
Mean RMS	2.59μm	1.17
Standard Deviation	0.12μm	0.24
Distribution of the results		
98% systems	3.24μm	1.47 of nominal spot size
90% systems	2.96μm	1.34 of nominal spot size
80% systems	2.81μm	1.27 of nominal spot size
50% systems	2.54μm	1.15 of nominal spot size
20% systems	2.36μm	1.07 of nominal spot size
10% systems	2.28μm	1.03 of nominal spot size
2% systems	2.20μm	1.00 of nominal spot size

Table 7.3: Number of imaging systems and Monte-Carlo performance, simulated on 10 000 systems with the state-of-the-art tolerances.

7.7.2 Near Diffraction limited system

One may wonder for which aperture the proposed starting point can achieve near-diffraction limited performance. From early research, it seemed that an f/1.1 lens could deliver a near-diffraction limited resolution as shown in figure 7.16. However, such a system fails to achieve high performance once the tolerances are taken into account.

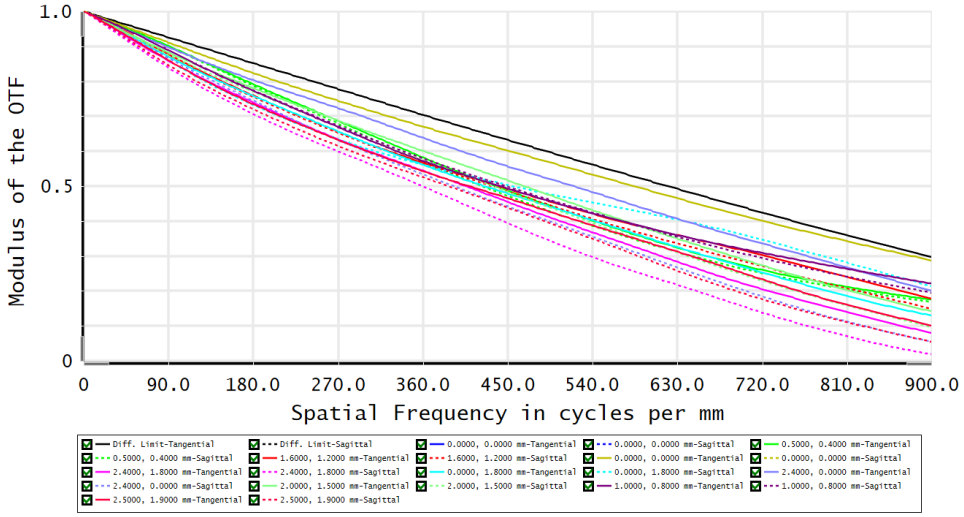


Figure 7.16: Theoretical MTF curves for a theoretical f/1.1 design.

Depending on the tolerances, we believe that our 7 element system can perform in a near diffraction limited regime for apertures between f/1.2 and f/1.4. We present a system designed for an f/1.4 aperture and the same tolerances as described in 7.1. The layout of the system is presented in figure 7.17. The resolution is diffraction limited on axis, and some coma is exhibited at the edges of the frame. Overall, the performance is close to the diffraction limit as shown in figure 7.18. The maximal distortion is -6% with a moustache style (varying from -6% to $+1\%$) as presented in figure 7.19.

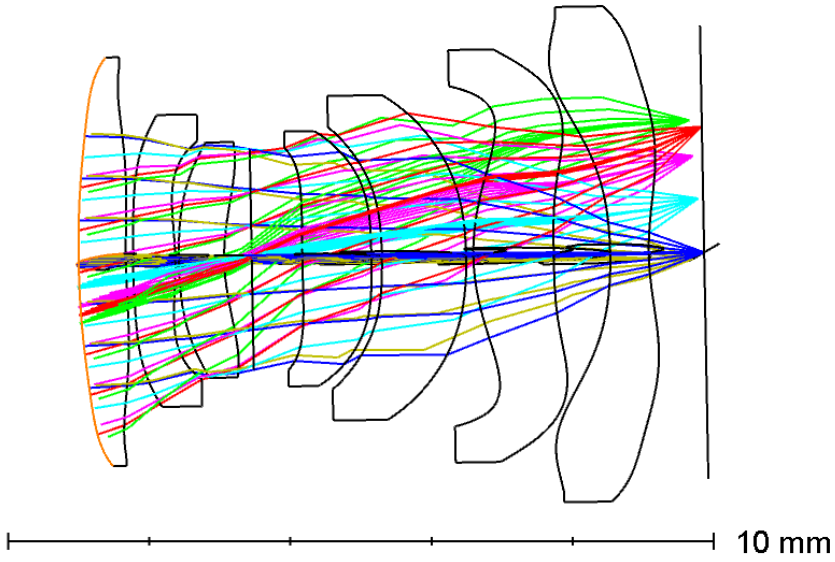


Figure 7.17: Proposed f/1.4 design.

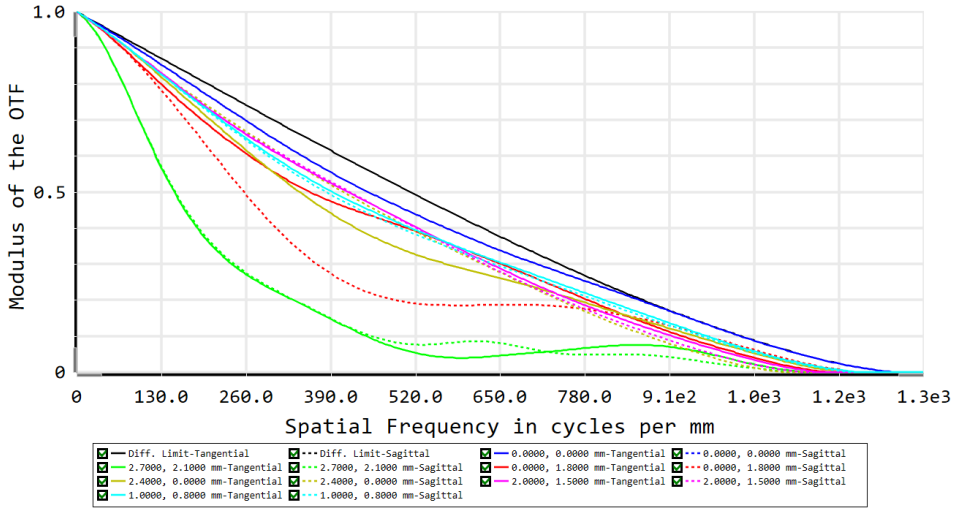


Figure 7.18: Theoretical MTF curves for the f/1.4 design.

Figure 7.19: Image simulation including vignetting simulation for a 3.7mm height sensor. There is some distortion with a moustache effect. Vignetting is well controlled.

As for the tolerancing, a Monte-Carlo simulation returns the results displayed in table 7.4.

Imaging systems	MTF (average on 5 points)
Nominal MTF	0.95541817
Best	0.91564243
Worst	0.62981036
Mean	0.81080678
Standard Deviation	0.06535317
Percentage of systems	MTF (average on 5 points)
Best 90%	0.72451317
Best 80%	0.75001631
Best 50%	0.82005900
Best 20%	0.87143251
Best 10%	0.88767781

Table 7.4: Number of imaging systems and Monte-Carlo performance, simulated on 100 systems with the state-of-the-art tolerances.

7.8 Conclusion

In this chapter two different lens designs have been proposed and one might wonder how much they improve the performance over the state-of-the-art. We expect both systems to improve the depth performance beyond the lenses typically used in current smartphones. The first system with a wide f/1.0 f-number favours a shallow depth of field at the expense of the high spatial frequencies. The second system, with a different trade-off, favours the high spatial frequencies at the expense of the f-number (f/1.4).

From the graph 4.5 proposed in chapter 4 we can suggest that both lenses offer 12 to 14 planes of focus instead of 8 for a f/1.8 lens.

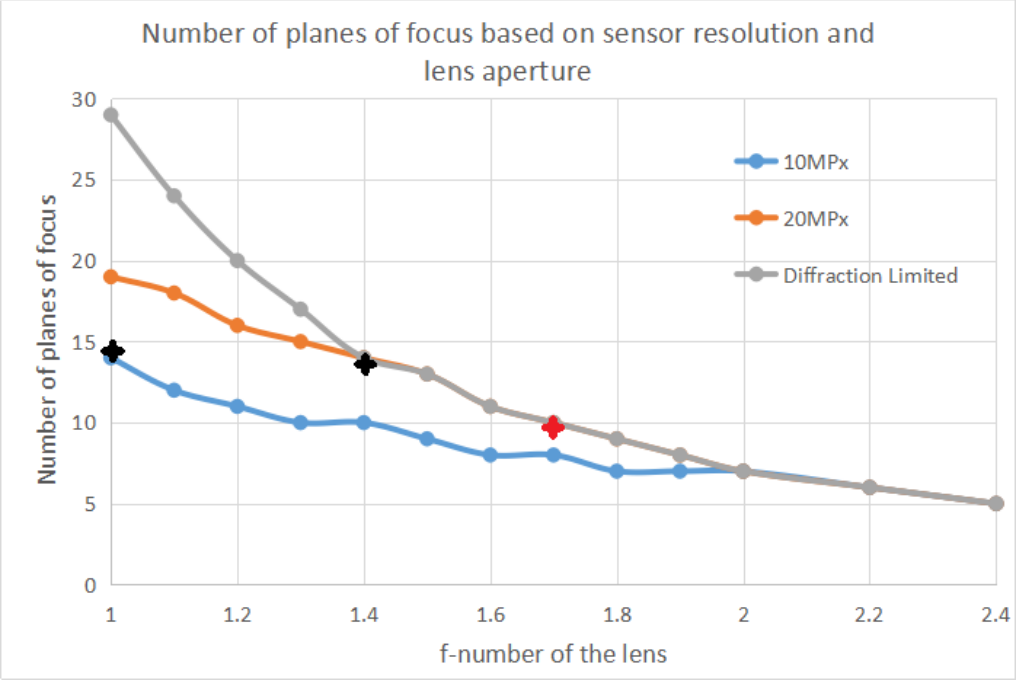


Figure 7.20: Number of depth planes for camera systems based on the resolution and f-number. The state of the art is marked with a red cross, the proposed lenses are marked with some black crosses.

References

[1] Deloitte. 3.5 million photos shared every minute in 2016. Available at https://www2.deloitte.com/uk/en/pages/press-releases/articles/3-point-5-million-photos-shared-every-minute.html. published: 2016.

[2] Jane Bareau and Peter Clark. The optics of miniature digital camera modules. In *International Optical Design*, page WB3. Optical Society of America, 2006.

[3] Joseph N. Mait, Gary W. Euliss, and Ravindra A. Athale. Computational imaging. *Advances in Optics and Photonics*, 10(2):409–483, Jun 2018.

[4] Timothée E Cognard, Chris Dainty, and Nicholas Devaney. Estimating axial resolution with diffraction theory. *Applied Optics*, 57(22):E138–E141, 2018.

[5] Vlad Savov. The galaxy s9s dual-aperture camera is great marketing. Available at https://www.theverge.com/2018/2/27/17057322/samsung-galaxy-s9-dual-aperture-camera-mwc-2018. The Verge, published: 27-02-2018.

[6] Pil Sun Jung, Dong Young Kim, and Chi Ho An. Lens optical system and imaging device, September 28 2017. US Patent App. 15/463,495.

[7] Thomas Steinich and Vladan Blahnik. Optical design of camera optics for mobile phones. *Advanced Optical Technologies*, 1(1-2):51–58, 2012.

[8] Yu-Shan Chang, Long Hsu, and Kuang-Lung Huang. Hybrid lens design for ultrawide-angle imaging with a high strehl ratio in a folded system. *Applied Optics*, 57(19):5514–5522, 2018.

[9] Conor J Sheil and Alexander V Goncharov. Double-folded catadioptric lens for smartphone portraiture photography. In *Optical Design and Engineering VII*, volume 10690, page 106900O. International Society for Optics and Photonics, 2018.

[10] Conor J Sheil and Alexander V Goncharov. Large aperture camera lens with minimalistic refocus for smartphone portraiture photography. *Optics Communications*, 440:207–213, 2019.

[11] Christophe Gaschet, Bertrand Chambion, Stéphane Gétin, Gaid Moulin, Aurélie Vandeneynde, Stéphane Caplet, David Henry, Emmanuel Hugot, Wilfried Jahn, Thibault Behaghel, et al. Curved sensors for compact high-resolution wide field designs. In *Novel Optical Systems Design and Optimization XX*, volume 10376, page 1037603. International Society for Optics and Photonics, 2017.

[12] Brian Guenter, Neel Joshi, Richard Stoakley, Andrew Keefe, Kevin Geary, Ryan Freeman, Jake Hundley, Pamela Patterson, David Hammon, Guillermo Herrera, et al. Highly curved image sensors: a practical approach for improved optical performance. *Optics Express*, 25(12):13010–13023, 2017.

[13] Dmitry Reshidko and Jose Sasian. Optical analysis of miniature lenses with curved imaging surfaces. *Applied Optics*, 54(28):E216–E223, Oct 2015.

[14] Walter Mandler. Design of basic double gauss lenses. In *1980 International Lens Design Conference*, volume 237, pages 222–233. International Society for Optics and Photonics, 1980.

[15] Litten Walter, George H Aklin, and Fred E Altman. Photographic objective of the cooke triplet type, April 11 1950. US Patent 2,503,751.

[16] Herbert Gross, Wolfgang Singer, Michael Totzeck, F Blechinger, and B Achtner. *Handbook of optical systems*, volume 4. Wiley-Vch Weinheim, 2005.

[17] M. J. Kidger and C. G. Wynne. The design of double gauss systems using digital computers. *Applied Optics*, 6(3):553–563, Mar 1967.

[18] Wikipedia. Double-gauss lens. Available at `https://en.wikipedia.org/wiki/File:Double_gauss.png`. published: 13-06-2007.

[19] NewBeat. Energizer 'thick boi' phone gives 50 days of battery life. Available at `https://www.bbc.com/news/newsbeat-47387390`. BBC News, published: 27-02-2019.

[20] James P. McGuire. Manufacturable mobile phone optics: higher order aspheres are not always better. In *International Optical Design Conference and Optical Fabrication and Testing*, page ITuF6. Optical Society of America, 2010.

[21] Yuke Ma and V Borovytsky. Design of a 16.5 megapixel camera lens for a mobile phone. *Open Access library journal*, 2(3), 2015.

[22] Sangjin Jung, Dong-Hoon Choi, Byung-Lyul Choi, and Ju Ho Kim. Tolerance optimization of a mobile phone camera lens system. *Applied Optics*, 50(23):4688–4700, Aug 2011.

[23] Pengbo Chen and Xingyu Gao. Optical design of the 13 mega-pixels mobile phone camera. *3rd International Conference on Materials Engineering, Manufacturing Technology and Control*, 2016.

[24] Experimental Optics. 50 mm f/0.75 - the fastest photographic lens in the world! Available at `https://www.kickstarter.com/projects/1745517054/50-mm-f-075-the-fastest-photographic-lens-in-the-w/description`. published: 20-01-2016.

[25] Wikipedia. Carl zeiss planar 50mm f/0.7. Available at `https://en.wikipedia.org/wiki/Carl_Zeiss_Planar_50mm_f/0.7`. published: 05-05-2012.

[26] Duncan T. Moore. Gradient-index optics: a review. *Applied Optics*, 19(7):1035–1038, Apr 1980.

[27] Yufeng Yan and Jose Sasian. Miniature camera lens design with a freeform surface. In *Optical Design and Fabrication 2017 (Freeform, IODC, OFT)*, page JTu3A.3. Optical Society of America, 2017.

[28] Shuming Wang, Pin Chieh Wu, Vin-Cent Su, Yi-Chieh Lai, Mu-Ku Chen, Hsin Yu Kuo, Bo Han Chen, Yu Han Chen, Tzu-Ting Huang, Jung-Hsi Wang, et al. A broadband achromatic metalens in the visible. *Nature nanotechnology*, 13(3):227, 2018.

[29] Sourangsu Banerji, Monjurul Meem, Berardi Sensale-Rodriguez, and Rajesh Menon. Imaging with flat optics: metalenses or diffractive lenses? *arXiv preprint arXiv:1901.05042*, 2019.

Chapter 8

Conclusion and Future Work

It is important to summarize briefly the findings from the previous chapters.

Chapter 4 explains in detail how to measure the depth of field and how to estimate the number of distinguishable planes of focus of a lens system, which is a good metric for the quality of depth estimation. For a given camera and minimum focusing distance, the more planes of focus between the minimum focusing distance and infinity, the better for the depth map estimation.

In this regard, the work in chapter 6 on apodization points to an improvement from typically f/1.8 to f/1.6. The 2D resolution is not improved by the use of a phase mask, and is likely to be traded against the depth resolution. Comparing an apodized mask solution to the state of the art should increase the number of depth planes from 0.5m to infinity from 9 to 11.

The lens design work from chapter 7 brings two new options for the lenses. The first option is a medium resolution f/1.0 lens. The second option is a near-diffraction limited f/1.4 lens. Both lenses theoretically bring about 14 depth planes, which is superior to the phase mask results.

Of course, it is possible to combine both techniques into a single lens. Assuming that the 2D resolution is left unchanged, one could expect to get about 17 depth planes for an apodized f/1.0 lens. However, it is difficult to estimate precisely the performance without taking into account the tolerancing for the phase mask and the alignment of the mask inside the lens.

In order to provide a better understanding of the perceived depth from the number of depth planes, it is possible to compute an "equivalent aperture" for a full-frame DSLR with an equivalent 28mm focal length. To put things into perspective, the state-of-the-art systems correspond to an f/14 aperture. The proposed system brings the perception of depth closer to that of a DSLR with an f/10 aperture as shown in figure 8.1.

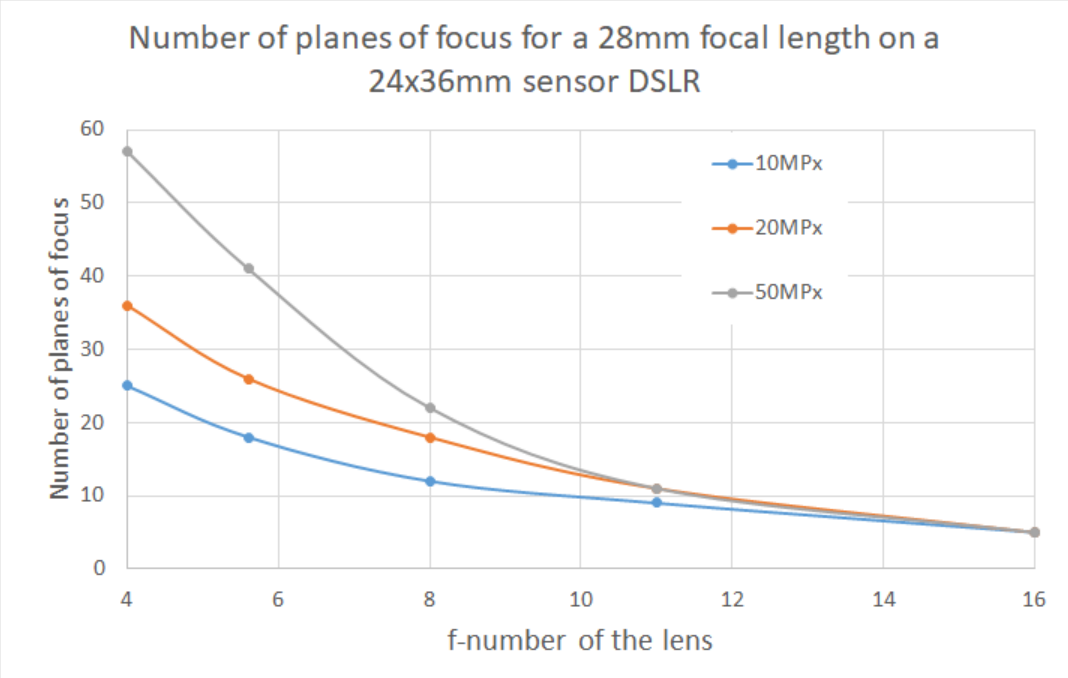


Figure 8.1: Number of depth planes for full frame $24 \times 36mm$ cameras based on the resolution and f-number.

8.1 Image Processing and possible future work

This section offers a few remarks on how optical design and image processing can further improve the contribution from the previous chapters of this thesis.

8.1.1 Going beyond the f/1.0 limit and possible future work

8.1.1.1 On the trend of apertures

One can wonder what is the maximal aperture that can be achieved for a miniature camera. To the best of our knowledge, the widest aperture photographic lenses commercialized were f/0.75 lenses with fixed aperture [**?**], suffering from a large amount of spherical aberration. Outside of the photography market, the lenses made available to Stanley Kubrick by Zeiss offered a good image quality for an f/0.7 aperture [**?**]. This performance was only achieved by adding a lens group in front of the double-Gauss system.

However, the number of optical elements is limited by the TTL of the phone. The phone thickness has been halved in ten years and is rather stable since the mid 2010s at around $7 - 8$mm. The trend might be slightly reversed with the introduction of folded screens and could increase the acceptable TTL length for the optical system.

8.1.1.2 Technical solutions for wider aperture

Several technical solutions can help improve the lens performance in the longer run, with applications to wide-aperture lenses.

- Special materials: a first improvement comes from the materials. Currently, plastic materials used for mobile phone lens design offer a limited range of refractive index and Abbe number. These two parameters are critical for the image quality and chromatic aberrations of the system. The manufacturer LG proposed a lens design including a glass element in its LG V30 phone. The development of higher refractive index plastics could bring some improvement to mobile imaging lenses. Another approach is the one offered by GRIN materials [**?**] where the index varies across the lens radius or thickness. Such variations can be used to better control aberrations and bring extra degrees of freedom to the optical design.
- Better manufacturing precision for both moulding and manufacturing of the lens system.
- Free-form lens elements (i.e. aspheric, non-rotationally symmetric lens elements) can bring some extra degrees of freedom for the optical design as shown by Yan and Sasian [**?**].
- Meta-surfaces are considered [**?**] as a possibility for replacing some of the lens elements. At this stage, the mass production of large sized meta-surfaces hasn't been explored and the efficiency and colour sensitivity performances of these surfaces does not make them suitable for most traditional imaging applications. Banerji et al. [**?**] suggests that diffractive lenses are easier to manufacture while providing superior performance.

8.1.2 How to merge the information from multi-camera systems ?

A more extensive work needs to be done on the particular topic of merging better the depth information from several sensors.

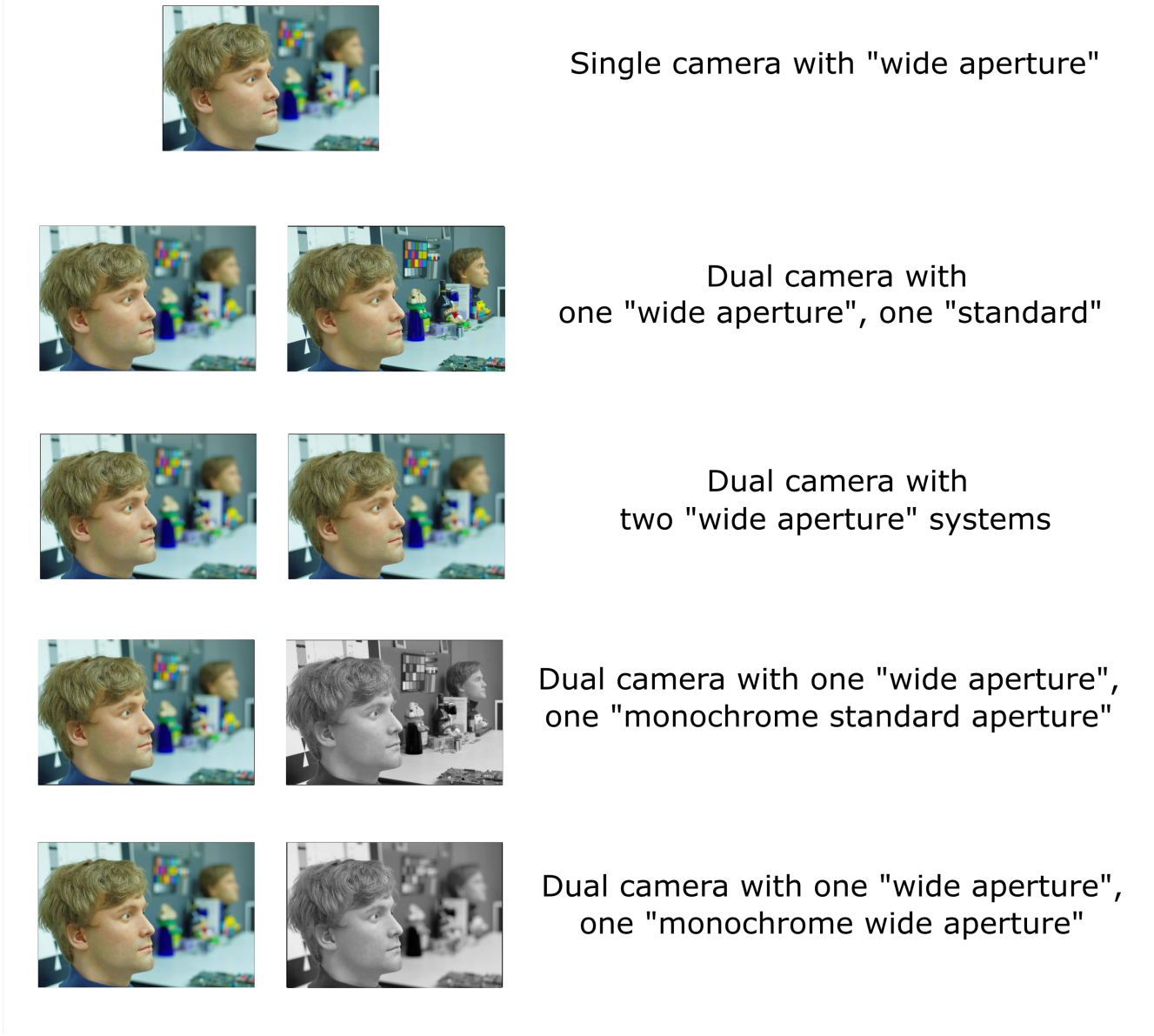


Figure 8.2: Several systems proposed to reduce the depth of field. The images are shot on a monochrome or a color sensor with either an f/1.0 f-number or an f/1.8 f-number. The combinations are proposed based on the statistics performed in Appendix B.

It will be assumed in this section that one of the lens systems proposed in chapter 7 can be manufactured and implemented in a smartphone. The target is to further improve the images and the depth map to achieve realistic and aesthetically pleasing shallow depth of field images. Despite being partially subjective, the notion of "aesthetically pleasing" can be defined in measurable metrics as proposed by Hauser [1].

Several systems can be proposed featuring a combination of a wide aperture camera lens (either f/1.0 or f/1.4), a standard aperture camera lens (typically f/1.8) and possibly different sensors

(colour and monochrome). The classification of the available options for single or dual camera systems with no focal length variations is shown in figure 8.2. This classification is only valid for single camera or dual camera systems, but since 2019, 3 or 4 cameras can be found on smartphones. Triple camera systems would bring even more possibilities for testing.

8.1.3 Building a database of images

In order to develop and test the performance of any algorithm, a database of images adapted to the physical parameters of the system needs to be created.

For a single lens system, the images from Figure 8.2 and Figure 8.3 have been produced with a 28mm focal length on a full frame DSLR sensor with apertures corresponding to the equivalent of a 1/3 inch sensor at f/2.0 (f/15 aperture used for a 24×36 sensor) and f/0.7 (f/5 aperture used for a 24×36 sensor).

Simulating a dual camera system is a more complicated task and can be performed in several ways:

- An existing dual-camera smartphone can be used to save both images in a RAW format, assuming that the focal length and spectrum are similar.
- A system with a beam-splitting mirror and two cameras is an alternative, although it is more costly and requires fine tuning (the narrow baseline of $1 - 2$cm is a challenge).

The images from the wide-aperture lens design can be simulated either using the Zemax toolbox (image simulation) or using Matlab together with Zemax.

8.2 Single image improvement and machine learning techniques

Deep learning methods have been getting researchers' attention for several decades. However, until the mid-2010s, most of the research using machine learning techniques performed worse than traditional methods. Several key factors have dramatically improved the performance of deep-learning based algorithms.

The amount of labelled data available for neural network training has exploded in the last few years. People nowadays exchange unprecedented amounts of data on-line through cloud solutions and social media. Most of this data is thus available for researchers from large IT companies such as Google, Apple, Facebook, Twitter, Microsoft, Nvidia, Intel and many more. Despite the possible unethical uses of this data (e.g. Cambridge Analytica's profiling), this information can

lead to new algorithms and applications.

As shown in the state of the art review in Section 5.1.5, deep learning methods can be used to estimate the depth. They can also be used to improve images [2] or to upsample images [3, 4]. We recommend the thesis of Pitkänen [5] for a state-of-the-art review of the field of machine-learning-based image enhancement.

Theoretically, one can expect similar machine learning algorithms to offer some form of image enhancement for optically aberrated lenses. Some attempts have been made in the field of lensless imaging devices, where a machine learning system is used to classify some of MNIST digits [6]. A more interesting example of such a technique is the one proposed by Nikonorov et al. [7], which compensates some of the lens aberrations produced by a Fresnel lens through machine learning algorithms.

It would be interesting to explore how much can machine learning algorithms be used to compensate certain optical aberrations (e.g. spherical aberration). With such algorithms, one could relax the optical design constraints and thus simplify the lens system.

8.3 Dual lens system

8.3.1 Theoretical considerations

In the case of a dual lens system, depth can be estimated through two different methods, which can ultimately be combined: depth from defocus and stereovision. In the state of the art camera systems, the depth from defocus performs worse than the stereovision technique. Thus, only stereovision is used. Assuming that one of the lenses has a wider aperture, one can expect depth from defocus to match the stereovision technique. For stereovision, the depth can be estimated using equation 8.1 which can be found in [8].

$$\Delta_z = \frac{z^2}{b \times f} \times \Delta_D \tag{8.1}$$

with Δ_z the depth estimation error (expressed in pixel unit), z the plane of focus (m), b the baseline size (m), f the focal length of the lens (m) and Δ_D the disparity error (in pixel unit). It is common to take $\Delta_D = 3$px which translates to a depth estimation error of around 150mm at a 1m distance for a baseline $b = 5$mm and a focal length $f = 4$mm. This result is remarkably close to the case of the proposed wide-aperture lens as has been shown in table 4.3.

8.3.2 Using a depth map estimation step

As stated in section 5, shallow depth of field images are usually produced from a large depth of field image that is partially blurred thanks to a depth map. With a two camera system using one wide aperture and one standard aperture lens, it is possible to extract a depth map and blur selectively the image with the largest depth of field.

Both depth from defocus and depth from stereovision have been explored by others, as shown in section 5. The combination of the depth from defocus depth map and the depth from stereovision depth map can be done through averaging, although a more detailed study of the topic could bring better solutions.

8.3.3 Using a sharpness transfer

A more natural solution might be to transfer the high frequency information from the standard aperture lens image to the wide aperture lens image. Basically, this translates into sharpening parts of a blurred image instead of blurring parts of a sharp image.

To do so, a two step process has been explored: the first step is an estimation of the parts of the blurred image that should be in-focus, the second step is a fine detail transfer.

8.3.3.1 In-Focus-Map estimation

We call in-focus-map the map of all the objects belonging to the focusing plane in the wide aperture lens image.
Various estimations of the in-focus-map can be used. The bottom line idea is to use a metric to measure the higher spatial frequencies inside the image for segmentation. The difference between the image and a re-blurred version of the image can be used to estimate the depth as shown by Zhuo and Sim [9]. Since the estimation is made on low resolution images, certain fine details are lost. As a consequence, the estimation may mistakenly identify parts of the image as being out of focus. From early testings, it seems that artefacts are visible on skin texture as well as *bokeh* edges. An example of bokeh edge artefact is shown in figure 8.3.

8.3.3.2 Detail Transfer

Once the in-focus-map has been estimated, one can proceed to transferring the high frequency information to the in-focus parts of the blurred image. The higher frequencies of the image can be obtained for instance from a wavelet transform [10].

(a) (b) (c) (d)

Figure 8.3: Comparison of the background blur between the wide aperture and standard lens simulated as detailed in section 7. (a) Full size image at f/1.0 (b) f/2.0 crop (c) f/1.0 crop. The in-focus map (d) should be fully dark in this case since all of the crop is out-of-focus. However, the edges around the *bokeh* balls are detected as if they are partly in-focus.

8.4 Additional benefits from a dual-camera system

Since numerous dual-camera combinations are possible, it is of interest to understand how the dual-camera set-up can also improve the photometric properties of the camera.

8.4.1 Noise level and high dynamic range

In the case of a Shot noise limited sensor, the Signal to Noise Ratio (SNR) is proportional to the square root of the amount of light collected. The lack of a Bayer pattern in a monochrome sensor triples the light flux reaching the sensor. The aperture also has a significant impact on the amount of light collected. Compared to an f/2.0 lens, an f/1.0 lens brings 4 times more light which corresponds to improving the SNR of the final image by a factor two.

The dual aperture set-up could be used to obtain higher dynamic range. To do so, the wide aperture lens can be used in order to retrieve some low light information with low noise while the standard lens can capture the brighter parts of the image. Depending on the scene, it could make sense to use the higher resolution camera for the dynamic range of the image subject.

8.4.2 The case for a dual camera featuring a standard monochrome camera and a shallow depth of field colour camera

The combination of the image from a monochrome camera paired with a standard lens and the depth from a colour camera with a wide aperture is especially interesting for 3 reasons:

- Resolution: the lack of a Bayer pattern ensures that there is no de-bayer interpolation in the image processing. This can provide a superior sharpness compared to the state of the art camera systems.

- Photometry: the lack of a Bayer pattern maximizes the amount of flux transmitted to the sensor. It has been estimated by Wang and Menon [11] that this corresponds to a 3.12 factor in light flux.
- Capturing settings: it happens that the previously mentioned 3.12 factor in light flux is approximately similar to the difference between an f/1.0 lens and an f/1.7 or f/1.8 lens (respectively 2.89 and 3.24). Thus, the same camera settings (gain and shutter speed) can be applied to both sensors.
- Colour resolution: due to the distribution of cones and rodes in the the human eye, the eye is less sensitive to colour variations, as opposed to luminance variations. Thus, the use of a high resolution monochrome image together with a lower resolution colour image brings the best perceived resolution [12] from a human vision point of view.

8.5 Conclusion

To some, depth estimation for miniature cameras might be a marginal feature or even a counter-intuitive quest: infinite depth of field is an engineer's dream for decades, whereas most photography enthousiasts praise shallow depth of field. As the distinction between smartphone cameras and DSLR cameras progressively blurs out, shallow depth of field remains as one of the features that smartphones still cannot match. This thesis has made a step in the direction of providing shallow depth of field for smartphone users.

In this thesis, we have proposed some contribution to the fields of theoretical optics and applied optics.

The work done in theoretical optics brings a new look on a century year old question: what does "in-focus" means? A new theoretical framework and an updated definition of the hyperfocal distance have been proposed. Several formulas have been established for the practical use of the hyperfocal distance in estimating the optimal focusing planes. This work has been used in order to design a depth estimation solution improving the state of the art from a typical f/1.7 aperture to a typical f/1.4 aperture regarding depth.

In practice, two different solutions have been proposed: the first one features an apodization phase mask, the second one is a new 7-element lens design with a wider aperture. The former has several limitations (optimized for a single wavelength, no tolerances on the manufacturing capabilities) whereas the later is toleranced and should be compatible with existing sensors, at the only expense of the form factor (which is about 20$ larger compared to state of the art).

Furthermore, some suggestions are made on how to implement such solutions into a commercial system. As proposed one could balance lateral resolution, axial resolution and colour resolution in a multi-camera system in order to optimize the photometry (and thus the SNR of the image), the 2D image and the depth map.
Finally and to broaden the spectrum of application of this work, miniature cameras are no longer limited to webcams and smartphones. Miniature cameras are becoming dominant in numerous fields such as security systems, production lines, autonomous cars, drones or biometrics. Understanding the physics of miniature cameras, such as the notion of depth of field and hyperfocal distance, can benefit research and applications beyond the particular case of smartphones.

References

[1] Wolf Hauser, Balthazar Neveu, Jean-Benoit Jourdain, Clément Viard, and Frédéric Guichard. Image quality benchmark of computational bokeh. *Electronic Imaging*, 2018(12):340–1, 2018.

[2] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: weakly supervised photo enhancer for digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 691–700, 2018.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[4] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5439–5448, 2017.

[5] Perttu Pitkänen. *Automatic Image Quality Enhancement Using Deep Neural Networks*. PhD thesis, University of Oulu, Finland, 2019. Faculty of Information Technology and Electrical Engineering.

[6] Ganghun Kim, Kyle Isaacson, Rachael Palmer, and Rajesh Menon. Lensless photography with only an image sensor. *Applied optics*, 56(23):6450–6456, 2017.

[7] Artem Vladimirovich Nikonorov, MV Petrov, Sergey Alexeevich Bibikov, Viktoriya Vitalievna Kutikova, AA Morozov, and Nikolai L'vovich Kazanskii. Image restoration in diffractive optical systems using deep learning and deconvolution. *Computer Optics*, 41(6):875–887, 2017.

[8] Silvio Giancola, Matteo Valenti, and Remo Sala. *A survey on 3d cameras: Metrological comparison of time-of-flight, structured-light and active stereoscopy technologies*. Springer, 2018.

[9] Shaojie Zhuo and Terence Sim. Defocus map estimation from a single image. *Pattern Recognition*, 44(9):1852–1858, 2011.

[10] Hasan Demirel and Gholamreza Anbarjafari. Image resolution enhancement by using discrete and stationary wavelet decomposition. *IEEE transactions on image processing*, 20(5):1458–1460, 2010.

[11] Peng Wang and Rajesh Menon. Ultra-high-sensitivity color imaging via a transparent diffractive-filter array and computational optics. *Optica*, 2(11):933–939, Nov 2015.

[12] Don Williams and Peter D Burns. Rethinking image color correction, validation and testing. In *Archiving Conference*, volume 2016, pages 175–180. Society for Imaging Science and Technology, 2016.

Appendix A

Axial Intensity Computation and Theoretical Resolution

A.1 Computation of the Axial Intensity

The axial intensity can be computed in detail as shown in Born and Wolf [1]:

It is possible to compute the 3D amplitude and phase of the light around the focal plane. One can replace the Cartesian coordinate system (x, y, z) centred on the image plane and optical axis with a set of new polar variables defined as:

$$\begin{cases} u = \frac{2\pi}{\lambda}(\frac{a}{f})^2 z \\ v = \frac{2\pi}{\lambda}(\frac{a}{f})^2 \sqrt{x^2+y^2} \end{cases} \tag{A.1}$$

The amplitude $U(u, v)$ can be computed as follow:

$$U(u,v) = \frac{2\mathbf{i}\pi a^2 A}{\lambda f^2} e^{\mathbf{i}(\frac{f}{a})^2 u} \int_0^1 J_0(v\rho) e^{-\mathbf{i}u\rho^2/2} \tag{A.2}$$

The intensity can be written using the Lommel functions:

$$\begin{cases} U_n(u,v) = \sum_{m=0}^{\infty}(-1)^m(\frac{u}{v})^{n+2m} J_{n+2m}(v) \\ V_n(u,v) = \sum_{m=0}^{\infty}(-1)^m(\frac{u}{v})^{-n-2m} J_{-n-2m}(v) \end{cases} \tag{A.3}$$

$$I(u,v) = (\frac{2}{u})^2[1 + V_0^2(u,v) + V_1^2(u,v) - 2V_0(u,v) \times cos(\frac{u+\frac{v^2}{u}}{2}) - 2V_1(u,v) \times sin(\frac{u+\frac{v^2}{u}}{2})]I_0 \tag{A.4}$$

Where the intensity is given by:

$$I_0 = (\frac{\pi a^2 |A|}{\lambda f^2})^2 \tag{A.5}$$

Along the axis, the parameter v is null ($x = 0, y = 0$):

$$\begin{cases} V_0(u,0) = 1 \\ V_1(u,0) = 0 \end{cases} \tag{A.6}$$

Replacing in the equation A.4 with the terms computed in A.6.

$$I(u,0) = 2(\frac{2}{u})^2 \times [1 - cos(\frac{u}{2})]I_0 \tag{A.7}$$

The equation can be re-written using a double cosine formula:

$$cos(\frac{u}{2}) = cos^2(\frac{u}{4}) - sin^2(\frac{u}{4}) \tag{A.8}$$

$$1 - cos(\frac{u}{2}) = 1 - cos^2(\frac{u}{4}) + sin^2(\frac{u}{4}) = 2sin^2(\frac{u}{4}) \tag{A.9}$$

$$I(u,0) = 4(\frac{2}{u})^2 \times sin^2(\frac{u}{4})I_0 = sinc^2(\frac{u}{4})I_0 \tag{A.10}$$

Replacing the variable u with z:

$$I(z) = sinc^2(\frac{z\pi}{2\lambda}(\frac{a}{f})^2)(\frac{\pi a^2|A|}{\lambda f^2})^2 \tag{A.11}$$

This intensity profile can be plotted as follows:

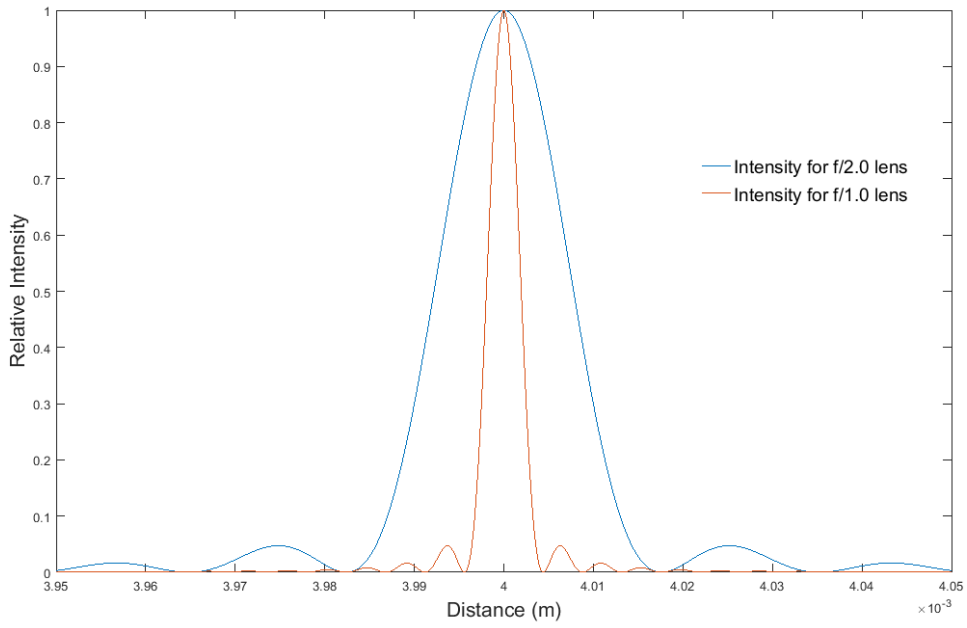


Figure A.1: Axial intensity for a $4mm$ focal length lens imaging an object located at infinity, with respective f/2.0 and f/1.0 apertures.

A.2 Resolution in the Image Space

The axial resolution can be defined using an equivalent of the Rayleigh criterion, which states that the distance between the two nearest separable points should be at least equal to the radius of first zero of the intensity.

This condition is verified for $sinc^2(X) = 0 \Leftrightarrow X = \pi$, and thus:

$$\frac{\Delta_z \pi}{2\lambda}(\frac{a}{f})^2 = \pi \Leftrightarrow \Delta_z = \frac{2\lambda f^2}{a^2} \tag{A.12}$$

A.3 Resolution in the Object Space

From the resolution in the image space, it is possible using the Lens Formula to compute the resolution in the object space. The resolution in the object space is of great interest as it provides the depth map estimation error.

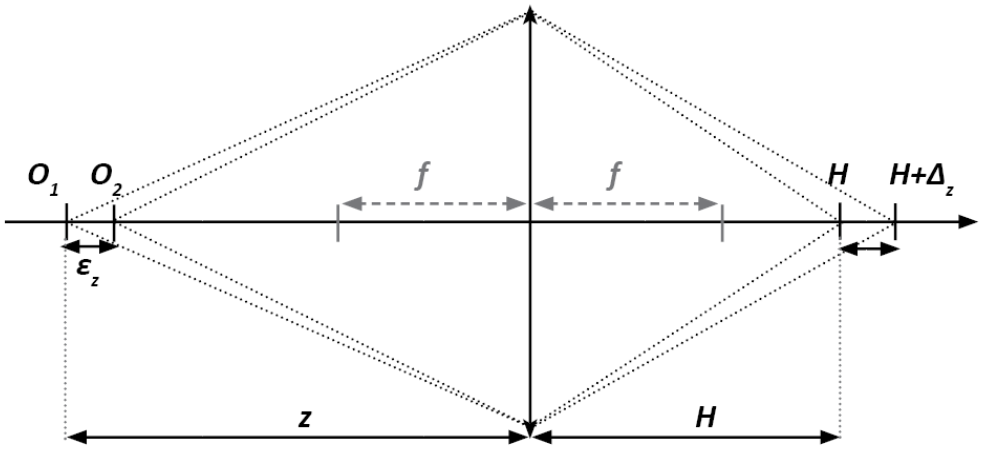


Figure A.2: Diagram of the system with notations.

Let's assume that the object O_1 is located at a distance z from the lens of focal length f and radius a. Using the Lens Formula, the image is formed on the plane $H = \frac{z \times f}{z - f}$.
Assuming that the axial resolution around the plane H is Δ_z in the image space, then the object points O_1 and O_2 in the object space can be obtained from the Lens Formula:

$$\begin{cases} O_1 = \frac{f \times H}{f - H} = z \\ O_2 = \frac{f \times (H + \Delta_z)}{f - (H + \Delta_z)} \end{cases} \tag{A.13}$$

The resolution is consequently:

$$O_2 - O_1 = \frac{f \times (H + \Delta_z)}{f - (H + \Delta_z)} - \frac{f \times H}{f - H} = \frac{f^2 \Delta_z}{(f - H - \Delta_z)(f - H)} \tag{A.14}$$

The approximation $f - H - \Delta_z \approx f - H \Leftrightarrow O_2 - O_1 \approx \frac{f^2 \Delta_z}{(f-H)^2}$ is valid only for a close distance to the lens but cannot be made in the case of a smartphone device.
Replacing H with it's expression, the resolution formula leads to:

$$\epsilon_z = \frac{\Delta_z}{1 - \frac{2z}{z-f} + \frac{z^2}{(z-f)^2} - \frac{\Delta_z}{f} + \frac{\Delta_z z}{f(z-f)}} \tag{A.15}$$

Injecting the image space resolution from equation A.12 in equation A.15, one can obtain the following axial intensity resolution equation in the case of diffraction limit:

$$\epsilon_z = \frac{2\lambda f^2}{a^2 - \frac{2za^2}{z-f} + \frac{z^2 a^2}{(z-f)^2} - 2\lambda f + \frac{2\lambda f z}{(z-f)}} \tag{A.16}$$

References

[1] M. Born and E. Wolf, "Principles of Optics - Electromagnetic Theory of Propagation, Interference and Diffraction of Light", Fourth Edition - Pergamon Press, Chapter 8: *Elements of the theory of diffraction*, Section 8.8: *The Three-dimensional Light Distribution near Focus* (1970)

Appendix B

State of the art of multi-camera smartphone technology

Multi-camera smartphones were almost non-existent prior to 2015. Since then, dual cameras have been very popular from 2015 to 2017. The first triple cameras appeared in 2018, and 4+ camera systems appeared in 2019. As for 2019, the trend is still an increase in the number of camera modules.

B.1 How to classify multi-camera smartphones ?

In order to classify multi-camera smartphones in a systematic way, one can group them by the function provided by additional cameras. To date, four main functions exist: depth estimation, spectrum broadening, focal length choice and focusing distance choice.

B.1.1 Depth multi-cameras (DMC)

One of the earliest forms of multi-camera system is designed for depth estimation purposes through stereovision techniques. An example of early system is the Pelican Imaging Pi Cam, an array of 4×4 lenses on a single sensor providing depth estimation. In most modern systems two cameras only are used. Often, the second camera has a lower resolution compared to the main camera, although in some higher-end systems both sensors are similar.

B.1.2 Spectrum multi-cameras (SMC)

Under the broader name of spectrum multi-cameras we group all devices having some differences in the perceived spectrum between the main camera and the side camera. Most SMCs consist of a color + monochrome camera. This combination has proven to be efficient at improving low light performance thanks to the greater amount of light collected in the monochrome sensor. In some rare projects, near infra-red is proposed as an extension of the visible spectrum.

B.1.3 Focal multi-cameras (FLMC)

FLMCs are all the systems comprising a standard lens and either wider or telephoto side lens. Most of the triple cameras offer both a wide angle lens and a telephoto system. Due to the form factor limitations, camera modules tend to offer a short telephoto (typically twice the focal length of the main lens). The apertures of both wide angle and telephoto modules are usually half a stop to a full stop slower compared to the main lens.

B.1.4 Focus multi-cameras (FSMC)

A few cameras have been proposed with dedicated macro lenses. It is not unlikely that in some cases, manufacturers will use optical designs optimized for different focal lengths depending on the application (portrait, landscape...)

B.2 Statistics on a database of phones from 2015 to 2019

In order to provide a better insight into the trends and state of the smartphone camera industry, we gathered a total of 127 phones offering a multi-camera system and examined for each the possible function. To the best of our knowledge, around 600 phones have been released with 2 or more cameras: 8 in 2015, 17 in 2016, 81 in 2017, 228 in 2018 and 267 in 2019. Among these, many present similar specifications. As a matter of fact, manufacturers tend to release several devices for different target markets with a similar camera system. As an example, Samsung released in 2019 a total of 30 phones, but the combination of a standard 12MPx, a telephoto 12MPx and an ultra-wide 16MPx is found in 9 phones (Note 10, Note 10+, S10, S10+, Fold, and their 5G versions). In most cases, we have avoided accounting multiple times for the same camera system.

Figure B.1 displays the proportion of dual-camera, triple-camera and 4+-camera systems (i.e. 4 or more cameras) in the database. Figure B.2 shows the proportions of systems in the database for each year. Figure B.3 reports the type of each camera from the database (note that since each system has multiple cameras, the sample contains a total number of around 300 cameras). Figure B.4 classifies the role of each camera system. From these last two figures, we can safely conclude that depth and focal length are the two main functions offered as of 2019.
Figure B.5 provides a visual representation of the different types of multi-camera systems together with the typical specifications found in our database.

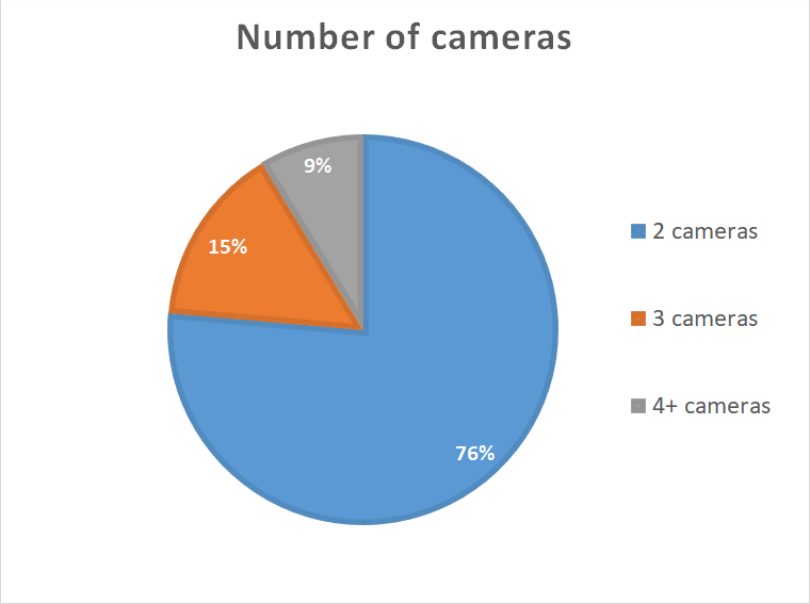


Figure B.1: Types of multi-camera system found in the database

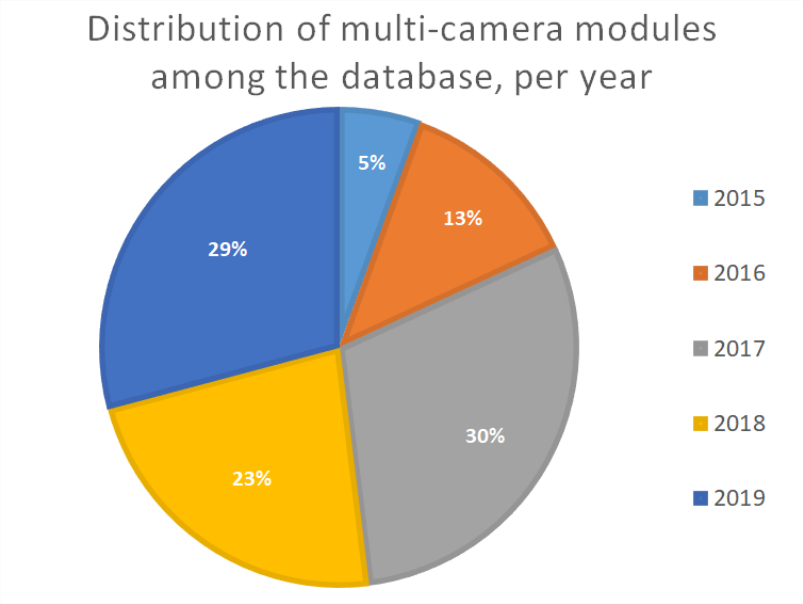


Figure B.2: Proportion of systems found in the database for each year

B.3 A historical evolution of main camera specifications

It is of interest to look at the evolution of the main camera specifications for smartphones in the past 4 years. If we needed to summarize it with one keyword, it would be diversification.

The main sensor size, as presented in figure B.6, has mostly increased since larger sensors provide better low light performance. Since multi-camera phones are getting accessible even at lower price points, some manufacturers prefer multi-camera systems with smaller sensor over a large single

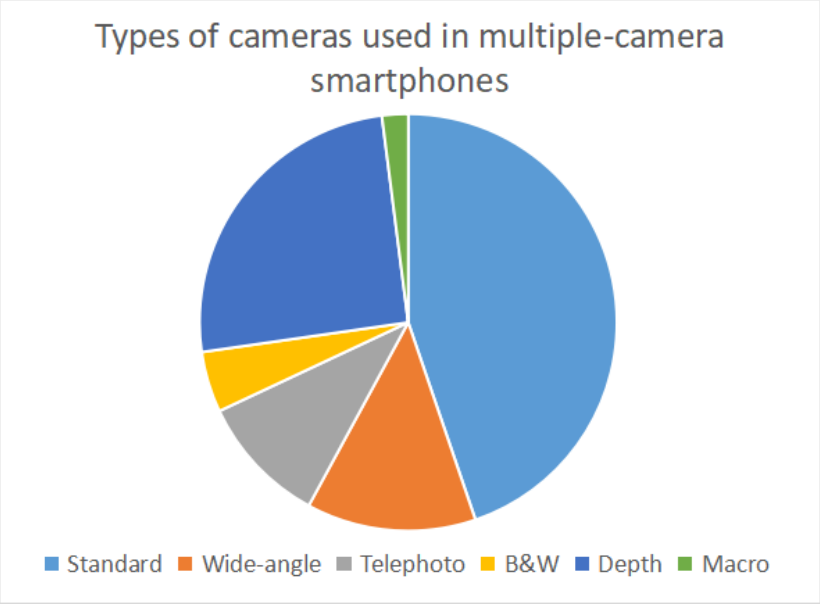


Figure B.3: Percentage of cameras used for each function in 127 multiple-camera devices from 2015 to 2019

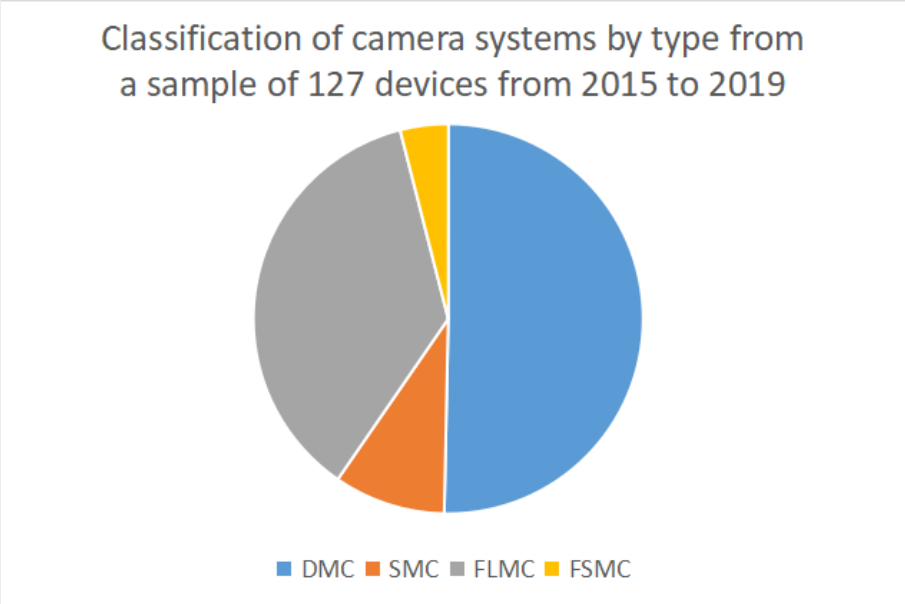


Figure B.4: Classification of systems in a set of 127 multiple-camera devices from 2015 to 2019

camera.

The evolution of pixel resolution shown in figure B.7 indicates an overall increase of the number of pixels, with a wider gap between high-end smartphones (typically over 40MPx) and the entry-level options (still around 12MPx).

The pixel size presented in figure B.8 faces two contrary effects. Since the sensor size is increasing on average, the pixel size tends to increase for high-end smartphones. However, since the pixel resolution is increasing overall, and for the smartphones keeping a typical 1/3 inch sensor size,

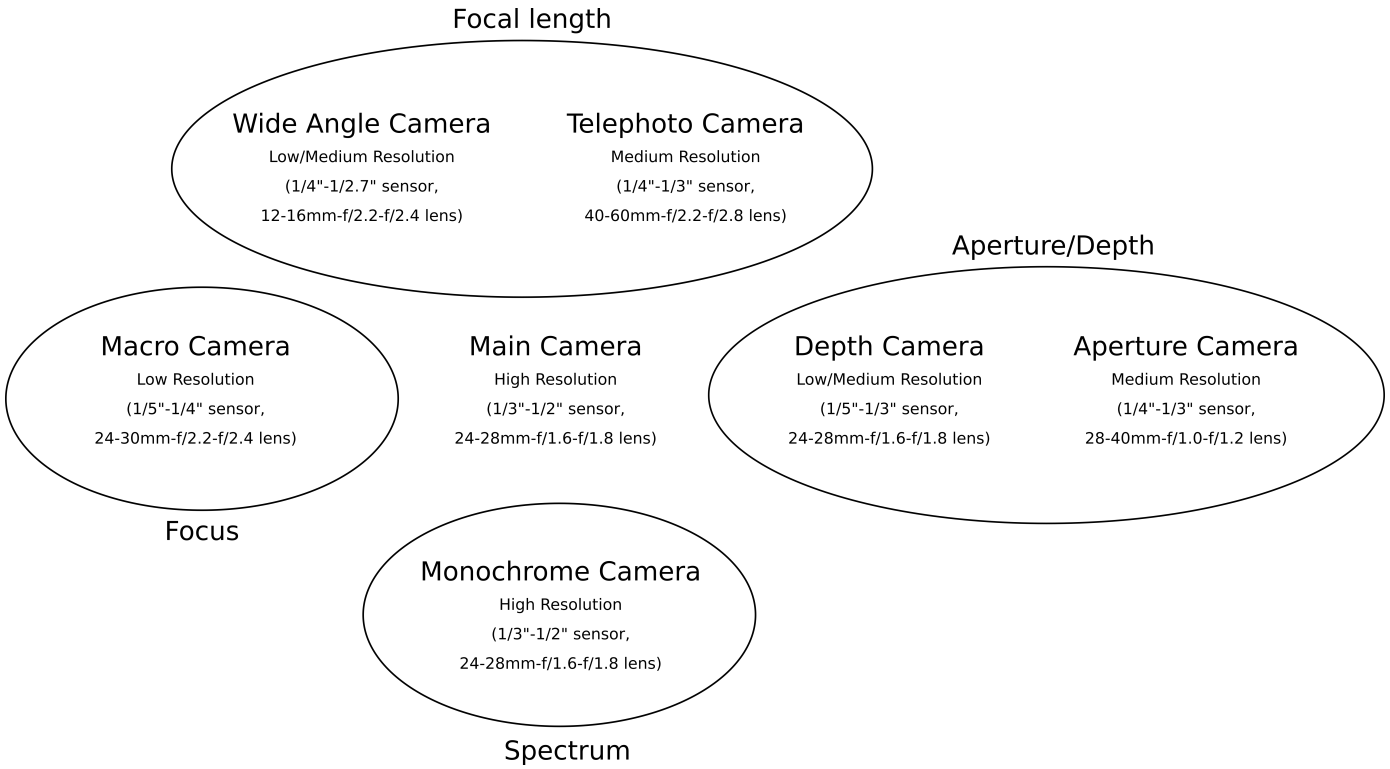


Figure B.5: Classification of systems and specifications found in the database. The focal length provided is the equivalent 24×36mm format focal length. Since the sensor size varies widely from one camera module to another, it is the only way to offer a fair comparison.

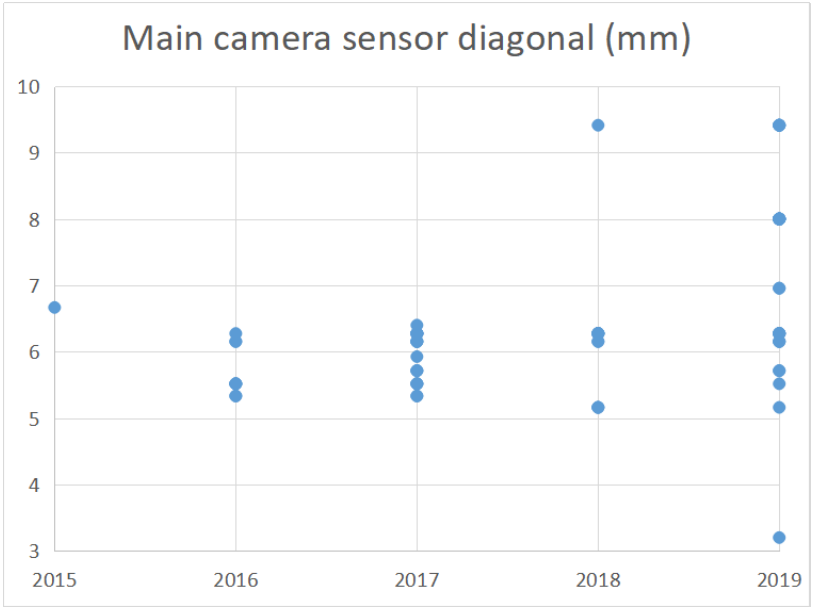


Figure B.6: Sensor diagonal (in mm) from devices in the database, ploted against the release year. For some devices, the sensor diagonal has been estimated from the pixel count and pixel size.

the pixel size is reduced. The trend is also observed for telephoto modules. These systems use a smaller sensor size (1/4 inch sensor for most cameras) with a higher pixel density.

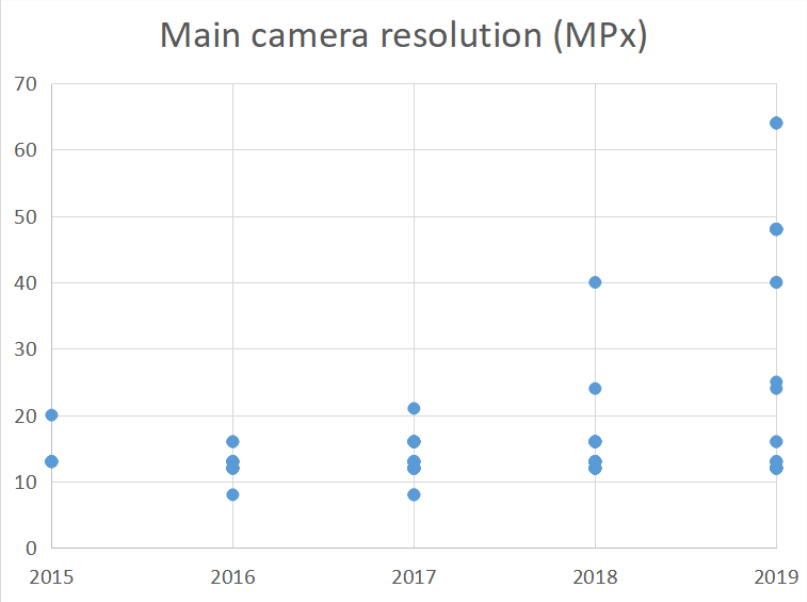


Figure B.7: Number of millions of pixels for the devices in the database, ploted against the release year.

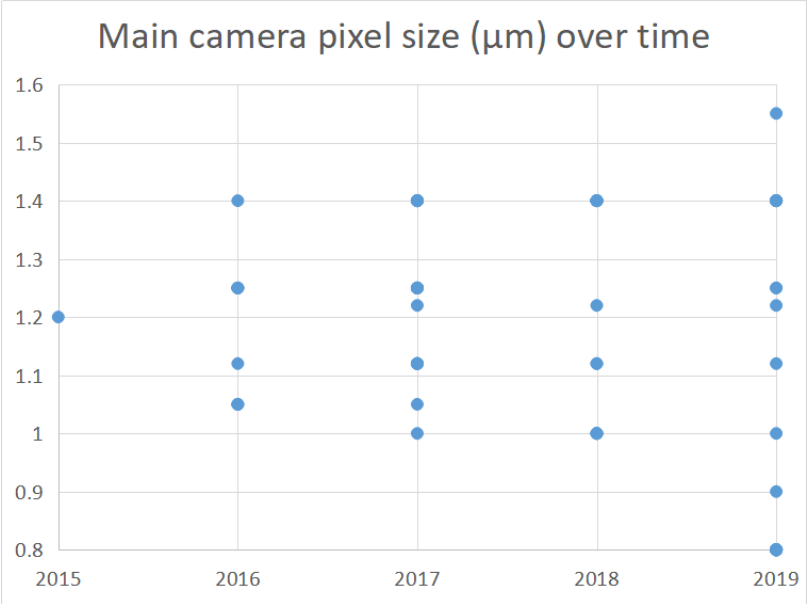


Figure B.8: Pixel size (in μm) from devices in the database, ploted against the release year. All the systems in the database have square pixels, except the few dual pixels used for focusing purposes. For some devices, the pixel size has been estimated from the pixel count and sensor size.

Figure B.9 shows a trend toward wider aperture lens systems. The f/1.5 systems offered by Samsung are the fastest options to date. They do come however with a variable aperture mechanism (f/1.5-2.4) to improve the image quality for landscape applications.
In order to make an apples-to-apples comparison, one should keep in mind that these apertures are expressed regardless of the sensor size whereas the pupil size (and consequent depth properties)

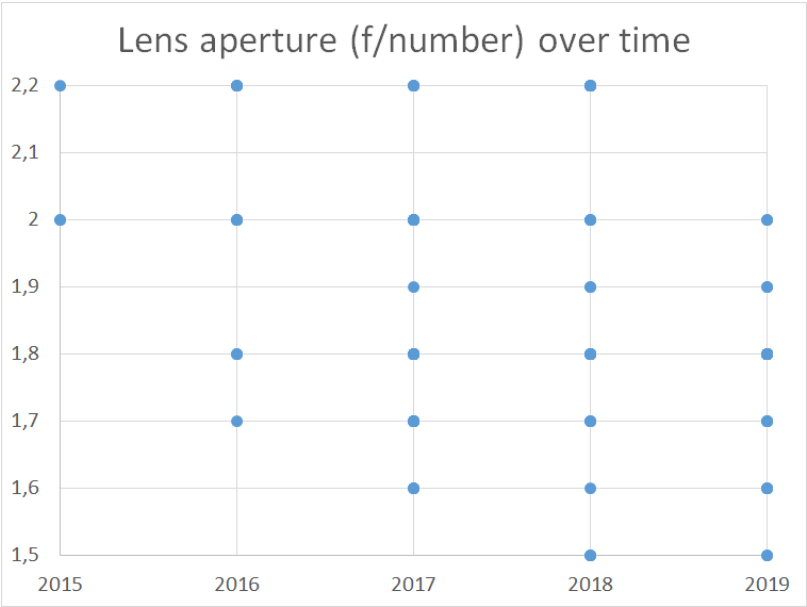


Figure B.9: Aperture of the systems in the database, regardless of the sensor size.

are sensor-size-related. We decided to take the sensor size into account and convert all apertures to their equivalent 35mm counterparts (i.e. to a 24×36mm sensor size). The results are shown in figure B.10. Most lenses have a $24 - 28$mm equivalent 35mm focal length.

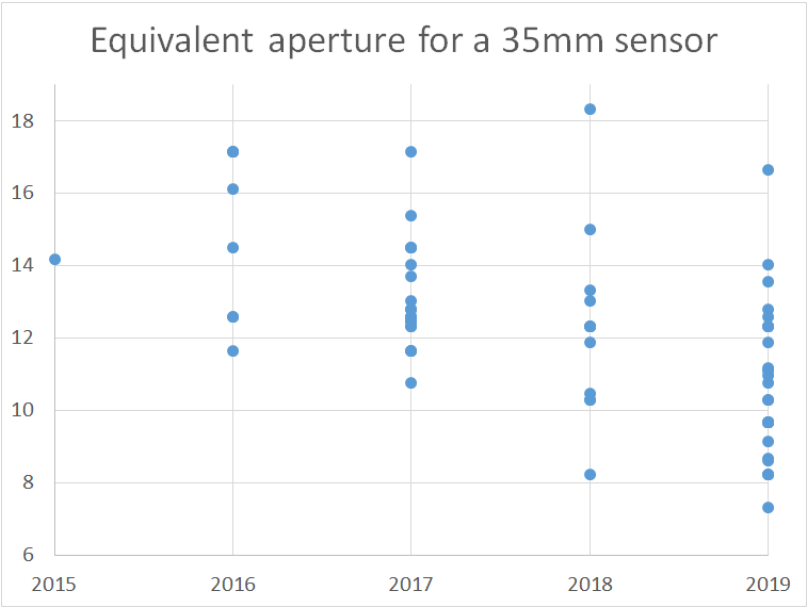


Figure B.10: Equivalent aperture of the systems in the database, converted to a 24×36mm sensor size. Note that the lenses with wide apertures usually have a shorter focal length (typically 24mm equivalent versus typically 30mm).

B.4 List of smartphone database used for statistics

Smartphone	Cameras	Cameras	Function	Year
Huawei Mate 20 Pro	3	Standard, Wide, Telephoto	Focal Length	2019
Huawei P20	2	Standard, Monochrome	Spectrum	2019
Samsung Galaxy A9	4+	Standard, Wide, Telephoto, Depth	Depth, Focal Length	2019
Nokia 9 PureView	4+	Standard, Depth	Depth	2019
Oppo Reno 2	4+	Standard, Wide, Telephoto, Monochrome	Spectrum, Focal Length	2019
Motorola One Macro	3	Standard, Depth, Macro	DepthFocus	2019
OPPO K5	4+	Standard, Wide, Depth Macro	Depth, Focal Length, Focus	2019
LG V50S ThinQ 5G	2	Standard, Wide	Focal Length	2019
Google Pixel XL4	2	Standard, Telephoto	Focal Length	2019
Xiaomi Redmi 8	2	Standard, Depth	Depth	2019
Oppo A5	4+	Standard, Wide, Depth	Depth, Focal Length	2019
Samsung Galaxy A20s	3	Standard, Wide, Depth	Depth, Focal Length	2019
Sony Xperia 5	3	Standard, Wide, Telephoto	Focal Length	2019
Apple iPhone 11 Pro Max	3	Standard, Wide, Telephoto	Focal Length	2019
Realme 5 Pro	4+	Standard, Wide, Depth Macro	Depth, Focal Length, Focus	2019
Samsung Galaxy A50	3	Standard, Wide, Depth	Depth, Focal Length	2019
Samsung Galaxy M30s	3	Standard, Wide, Depth	Depth, Focal Length	2019
Xiaomi Redmi Note 8 Pro	4+	Standard, Wide, Depth, Macro	Depth, Focal Length, Focus	2019
Samsung Galaxy A70s	3	Standard, Wide, Depth	Depth, Focal Length	2019
Apple iPhone 11	2	Standard, Wide	Focal Length	2019
Samsung Galaxy Note10+	4+	Standard, Wide, Telephoto, Depth	Depth, Focal Length	2019
Oppo A9 (2020)	4+	Standard, Wide, Depth	Depth, Focal Length	2019
OnePlus 7T	3	Standard, Wide, Telephoto	Focal Length	2019
Xiaomi Mi 9 Pro	3	Standard, Wide, Telephoto	Focal Length	2019
Nokia 7.2	3	Standard, Wide, Depth	Depth, Focal Length	2019
Huawei P30 Pro	4+	Standard, Wide, Telephoto, Depth	Depth, Focal Length	2019
Samsung Galaxy S10	3	Standard, Wide, Telephoto	Focal Length	2019
Asus ROG Phone II ZS660KL	2	Standard, Wide	Focal Length	2019
Huawei nova 5T	4+	Standard, Wide, Depth, Macro	Depth, Focal Length, Focus	2019
Huawei Y9 Prime	3	Standard, Wide, Depth, Macro	Depth, Focal Length, Focus	2019
Xiaomi Redmi K20 Pro	3	Standard, Wide, Telephoto	Focal Length	2019
Oppo F11 Pro	2	Standard, Depth	Depth	2019
Realme 3 Pro	2	Standard, Depth	Depth	2019
Samsung Galaxy A10s	2	Standard, Depth	Depth	2019
Xiaomi Mi 9	3	Standard, Wide, Telephoto	Focal Length	2019
Oppo F11	2	Standard, Depth	Depth	2019

Smartphone	Cameras	Cameras	Function	Year
Samsung Galaxy M30	3	Standard, Wide, Depth	Depth, Focal Length	2019
Xiaomi Pocophone F1	2	Standard, Depth	Depth	2018
Samsung Galaxy Note9	2	Standard, Telephoto	Focal Length	2018
Apple iPhone XS Max	2	Standard, Telephoto	Focal Length	2018
LG G7 ThinQ	2	Standard, Wide	Focal Length	2018
Samsung Galaxy J6+	2	Standard, Depth	Depth	2018
Nokia 5.1 Plus (Nokia X5)	2	Standard, Depth	Depth	2018
LG V40 ThinQ	3	Standard, Wide, Telephoto	Focal Length	2018
Samsung Galaxy J8	2	Standard, Depth	Depth	2018
Honor Play	2	Standard, Depth	Depth	2018
vivo Y93	2	Standard, Depth	Depth	2018
Apple iPhone XS	2	Standard, Telephoto	Focal Length	2018
Samsung Galaxy S9+	2	Standard, Telephoto	Focal Length	2018
Huawei Y9	2	Standard, Depth	Depth	2018
Nokia 6.1 Plus	2	Standard, Depth	Depth	2018
Nokia 8.1	2	Standard, Depth	Depth	2018
OnePlus 6T	2	Standard	Depth	2018
Oppo A7	2	Standard, Depth	Depth	2018
Oppo A3s	2	Standard, Depth	Depth	2018
Oppo F9 (F9 Pro)	2	Standard, Depth	Depth	2018
Huawei P20 lite	2	Standard, Depth	Depth	2018
vivo Y91	2	Standard, Depth	Depth	2018
BlackBerry KEY2 LE	2	Standard, Depth	Depth	2018
Honor 8C	2	Standard, Depth	Depth	2018
TECNO Camon 11 Pro	2	Standard, Depth	Depth	2018
Huawei Mate 20 X	3	Standard, Wide, Telephoto	Focal Length	2018
Realme U1	2	Standard, Depth	Depth	2018
Motorola Moto G6	2	Standard, Depth	Depth	2018
Motorola One (P30 Play)	2	Standard, Depth	Depth	2018
Samsung Galaxy A8s	3	Standard, Telephoto, Depth	Depth, Focal Length	2018
Apple iPhone 8 Plus	2	Standard, Telephoto	Focal Length	2017
Apple iPhone X	2	Standard, Telephoto	Focal Length	2017
Samsung Galaxy Note8	2	Standard, Telephoto	Focal Length	2017
LG G6	2	Standard, Wide	Focal Length	2017
LG V30	2	Standard, Wide	Focal Length	2017
Xiaomi Mi A1 (Mi 5X)	2	Standard, Telephoto	Focal Length	2017
OnePlus 5T	2	Standard	Depth	2017
Huawei Mate 10 Pro	2	Standard, Monochrome	Spectrum	2017
Nokia 8	2	Standard, Monochrome	Spectrum	2017
Huawei P10	2	Standard, Monochrome	Spectrum	2017
Huawei Mate 10	2	Standard, Monochrome	Spectrum	2017
Xiaomi Mi 6	2	Standard, Telephoto	Focal Length	2017
Huawei P smart	2	Standard, Depth	Depth	2017
Honor 7X	2	Standard, Depth	Depth	2017
Motorola Moto X4	2	Standard, Wide	Focal Length	2017
Lenovo K8 Note	2	Standard, Depth	Depth	2017
Asus Zenfone 4 Max ZC554KL	2	Standard, Wide	Focal Length	2017

Smartphone	Cameras	Cameras	Function	Year
Huawei P10 Plus	2	Standard, Monochrome	Spectrum	2017
Honor 9	2	Standard, Monochrome	Spectrum	2017
Honor View 10	2	Standard, Monochrome	Spectrum	2017
Razer Phone	2	Standard, Telephoto	Focal Length	2017
ZTE Blade Z Max	2	Standard, Depth	Depth	2017
Huawei nova 2	2	Standard	Depth	2017
Blackview P6000	2	Standard, Depth	Depth	2017
vivo X20	2	Standard, Depth	Depth	2017
Oppo R11 Plus	2	Standard,	Depth	2017
Lava Z90	2	Standard, Depth	Depth	2017
Blackview A10	2	Standard, Depth	Depth	2017
Archos Diamond Omega	2	Standard, Telephoto	Focal Length	2017
Asus Zenfone 4 ZE554KL	2	Standard, Wide	Focal Length	2017
Samsung Galaxy C7	2	Standard, Depth	Depth	2017
Asus Zenfone 3 Zoom ZE553KL	2	Standard, Telephoto	Focal Length	2017
Oppo R11	2	Standard	Depth	2017
Meizu M6 Note	2	Standard, Depth	Depth	2017
Lenovo K8 Plus	2	Standard, Depth	Depth	2017
Asus Zenfone 4 Pro ZS551KL	2	Standard, Telephoto	Focal Length	2017
Blackview S8	2	Standard, Depth	Depth	2017
LG Q8 (2017)	2	Standard, Wide	Focal Length	2017
Apple iPhone 7 Plus	2	Standard, Telephoto	Focal Length	2016
LG V20	2	Standard, Wide	Focal Length	2016
LG G5	2	Standard, Wide	Focal Length	2016
Huawei Mate 9	2	Standard, Monochrome	Spectrum	2016
Honor 6X	2	Standard, Depth	Depth	2016
Xiaomi Redmi Pro	2	Standard, Depth	Depth	2016
Huawei Mate 9 Pro	2	Standard, Monochrome	Spectrum	2016
Xiaomi Mi 5s Plus	2	Standard, Monochrome	Spectrum	2016
LG X cam	2	Standard, Depth	Depth	2016
TECNO Phantom 6	2	Standard, Depth	Depth	2016
Huawei Mate 9 Porsche Design	2	Standard, Monochrome	Spectrum	2016
Honor Magic	2	Standard, Monochrome	Spectrum	2016
vivo Xplay6	2	Standard, Depth	Depth	2016
ZTE Grand X Max 2	2	Standard, Depth	Depth	2016
Gionee S9	2	Standard, Depth	Depth	2016
Blackview A10	2	Standard, Depth	Depth	2016
HTC One M9+	2	Standard, Depth	Depth	2015
HTC One M8s	2	Standard, Depth	Depth	2015
ZTE Axon Pro	2	Standard, Depth	Depth	2015
ZTE Axon	2	Standard, Depth	Depth	2015
ZTE Axon Elite	2	Standard, Depth	Depth	2015
ZTE Axon Lux	2	Standard, Depth	Depth	2015
XOLO Black 3GB	2	Standard, Depth	Depth	2015