



Institut Langevin  
ONDES ET IMAGES



Thesis submitted to obtain the degree of doctor of philosophy from

**Sorbonne Université**

École doctorale EDITE de Paris (ED130)  
Informatique, Télécommunication et Électronique

Laboratoire de Recherche et Développement d'EPITA (LRDE)  
Institut Langevin, ESPCI

---

# **Non-iterative methods for image improvement in digital holography of the retina**

–

*Méthodes non itératives pour l'amélioration des images en holographie  
numérique de la rétine*

---

**Julie RIVET**

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Retinal imaging</b>	<b>3</b>
1.1 The role of the retina . . . . .	3
1.1.1 Some information about the eye . . . . .	4
1.1.2 Retina pathologies . . . . .	5
1.2 Imaging techniques . . . . .	7
1.2.1 Optical coherence tomography . . . . .	7
1.2.2 Blood flow imaging . . . . .	11
1.2.3 Conclusion on imaging techniques . . . . .	13
1.3 Eye motion . . . . .	14
1.3.1 Fixation movements . . . . .	14
1.3.2 State of the art in motion compensation . . . . .	15
1.4 Eye aberrations . . . . .	16
1.4.1 Sources of aberrations in the eye . . . . .	16
1.4.2 Mathematical description of aberrations . . . . .	16
1.4.3 State of the art in aberration compensation . . . . .	20
1.5 Conclusion . . . . .	24
<b>2 Digital holography: OCT and Laser Doppler</b>	<b>25</b>
2.1 Holographic imaging . . . . .	25
2.1.1 Setups . . . . .	25
2.1.2 Acquisition of interferograms . . . . .	26
2.1.3 Digital processing of optically-acquired interferograms . . . . .	31
2.2 Optical coherence tomography . . . . .	33
2.2.1 Calculation of the axial range . . . . .	34
2.2.2 Image resolution . . . . .	35
2.2.3 Improvement of image resolution with rescaling . . . . .	36
2.2.4 Conclusion on holographic OCT setup . . . . .	36
2.3 Laser Doppler imaging . . . . .	38
2.4 Motion compensation . . . . .	39
2.4.1 Lateral movement compensation . . . . .	39
2.4.2 Axial movement compensation . . . . .	41
2.4.3 Conclusion on motion compensation . . . . .	45
<b>3 Estimation of aberrations with a digital wavefront sensor</b>	<b>47</b>
3.1 The Shack-Hartmann wavefront sensor . . . . .	47
3.1.1 General principle . . . . .	47

3.1.2	Theoretical study . . . . .	48
3.1.3	Estimation of simulated aberrations . . . . .	52
3.1.4	Conclusion on the simulations . . . . .	58
3.2	Adaptation to the laser Doppler holography . . . . .	58
3.2.1	Theoretical modelling with the setup . . . . .	58
3.2.2	Implementation . . . . .	60
3.3	Tests and results . . . . .	63
3.3.1	Experimental verification . . . . .	63
3.3.2	Results on patients . . . . .	67
3.3.3	Conclusion and perspectives . . . . .	69
<b>4</b>	<b>Aberration compensation by deep learning</b>	<b>71</b>
4.1	About the use of deep learning for holographic imaging and phase estimation . .	71
4.2	Correction of aberrated eye-fundus images . . . . .	71
4.2.1	Context . . . . .	71
4.2.2	The database . . . . .	72
4.2.3	The network . . . . .	72
4.2.4	The training step . . . . .	77
4.2.5	Results . . . . .	81
4.3	Conclusion and perspectives of eye-fundus correction . . . . .	86
4.4	Digital Gabor hologram rendering . . . . .	87
4.4.1	Synthetic interferograms . . . . .	88
4.4.2	Experimental results . . . . .	92
4.4.3	The inverse problem . . . . .	93
4.5	Conclusions and perspectives on the use of deep learning in correction of aberrations	96
	<b>Conclusions and perspectives</b>	<b>96</b>
	<b>A Statistics for the simulation of a Shack-Hartmann wavefront sensor</b>	<b>99</b>
	<b>B Scientific contributions</b>	<b>103</b>
	<b>References</b>	<b>119</b>



---

## Introduction

---

According to the World Health Organization [1], 217 million people around the world suffer from moderate to severe visual impairment. 38,5 million of the world population are blind. With demographic growth and ageing of the population, this number is estimated to triple by 2050. Monitoring and treatment of vision disorders are therefore major issues in medicine today.

Several methods have been developed all over the years to study the eye and its internal dynamics: the shape of the cornea, the structure of the retina, the blood flow circulation... These characteristics can provide precious information about the vision, but also about some pathologies. They can be obtained by different imaging techniques, more or less invasive.

The technology we are going to focus on in this manuscript is holography. Non invasive, it consists in using the optical wavefield recorded by interference of the reflected light beam sent to the object to be imaged. In the eye, holography is increasingly used for retinal imaging. At the Quinze-Vingts national ophthalmology hospital in Paris, two optical benches have been settled in recent years to develop two real-time digital holography techniques for the retina: holographic optical coherence tomography (OCT) and laser Doppler holography. The first reconstructs depth-resolved, three dimensional structural images of optical reflection and absorption. The second allows the visualization of blood flow through the vessels, in a large depth-of-field image, encompassing the retinal and choroidal networks.

One of the benefits of holography lies in its ability to give access to the entire field without the need for a complex setup. Digital processing and optimization on Graphics Processing Unit (GPU) allow to get a high rendering speed. Yet, the image quality needs to be improved in order to be able to better see the smallest structures like vessels and capillaries. The passage of the light through the cornea causes a distortion of its wavefront. Indeed, this natural lens is deformed differently for each person. This distortion, called aberration, is reflected in the resulting holographic images of the retina, adding more or less blur to them, and decreasing their resolution. In order to overcome this limitation, aberrations of the eye need to be compensated.

This PhD thesis aims at improving the resolution of the images provided by the holographic setups at Quinze-Vingts national ophthalmology hospital, in particular by compensating aberrations. For this purpose, it is necessary to rigorously understand the transformations carried out to obtain images, as well as the effect of aberrations. The study of the state of the art reveals several leads to be explored. We have chosen to investigate thoroughly two of them,

each with their own specificities.

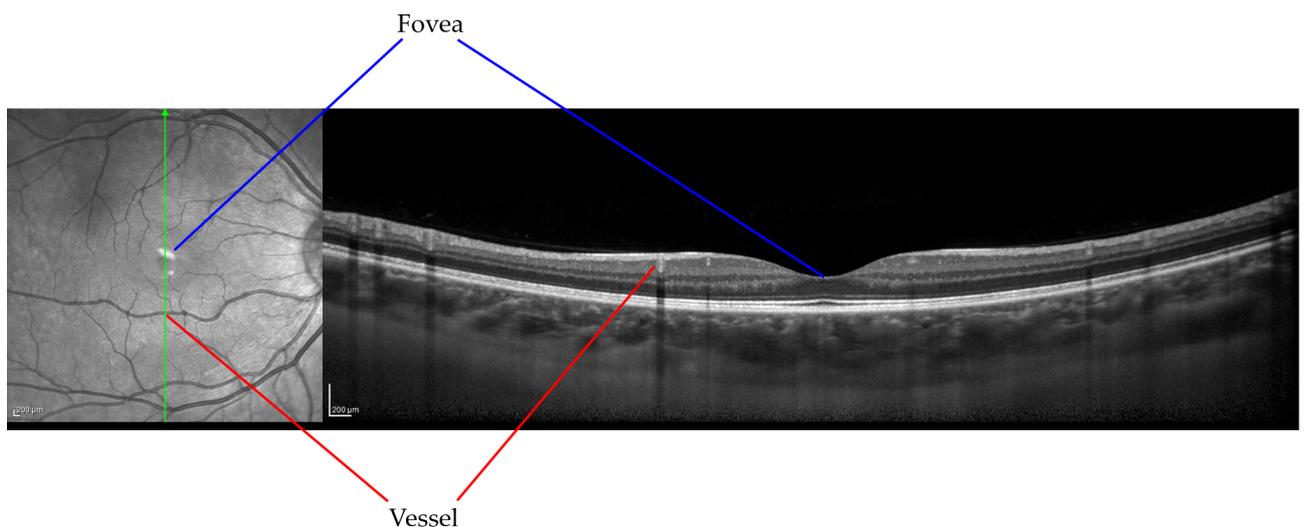
The first method is based on the most used technique to measure aberrations: a wavefront sensor. It consists in analogically measuring the wavefront distortion. In our work, we have transferred this technique from analog to digital, thus to avoid adding any optical element to the setups. Once the aberration measured, the access to the field, provided by digital holography, allows its correction in digital post-processing.

The second method has been studied for the time saving it could bring: the compensation of aberrations would be done in milliseconds. This consists in using a deep neural network, which is a specific sequence of linear and non linear operations, in order to obtain a corrected image directly from an aberrated image.

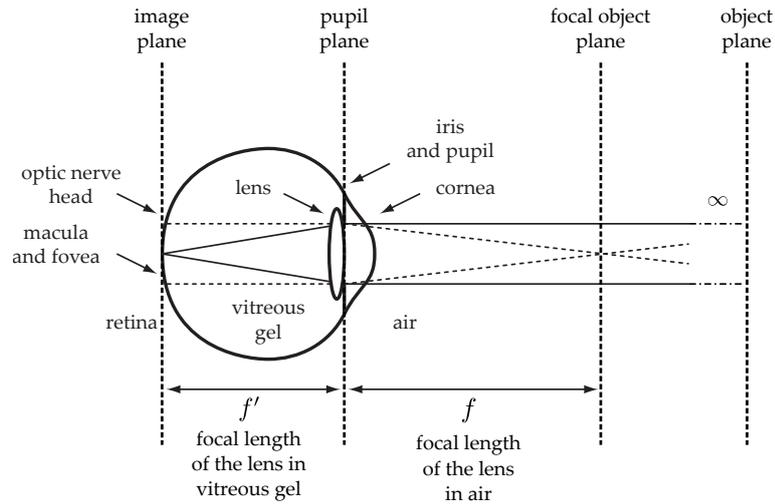
The details and the results of these methods have been documented in this manuscript. The first chapter introduces the main themes of this thesis through the state of the art. The second chapter explains how the holography benches are designed and tuned in order to record the interferograms that will be used for image rendering, as well as the first treatments applied to images, such as motion compensation. Finally, the third and fourth chapters focus respectively on the two methods of estimation and correction of the aberrated wavefront introduced earlier. This work paves the way for new processing methods that will allow a major improvement of the lateral resolution of images in digital holography of the retina.

## 1.1 The role of the retina

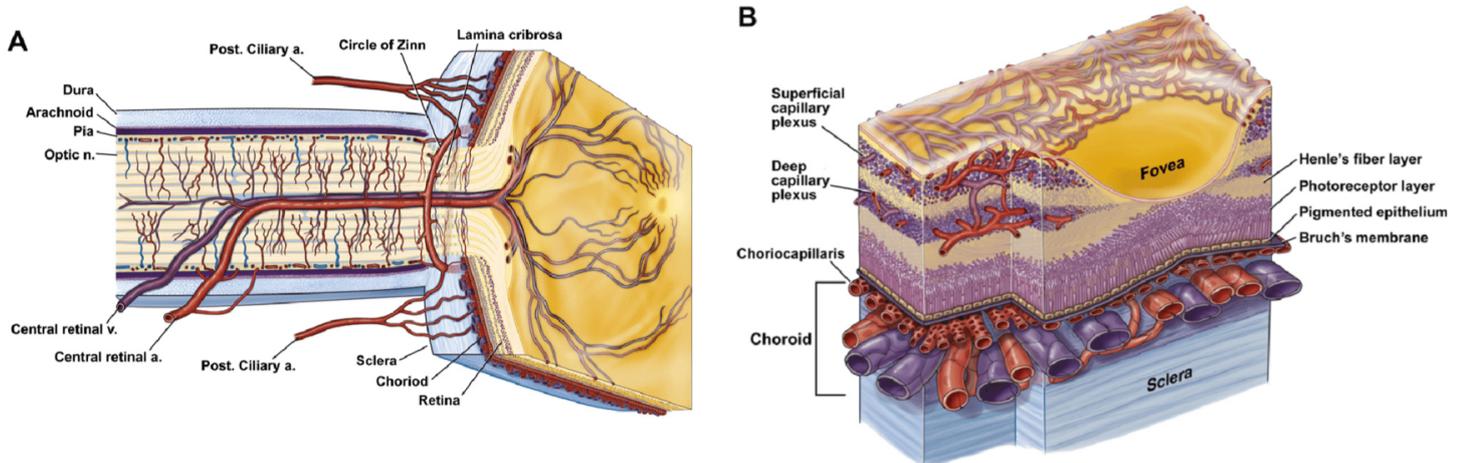
This chapter aims at introducing the different elements that form the basis of the main themes of the manuscript. This section provides general information about the retina and several pathologies that can affect it, to understand its role and the reasons why it is important to image it. An overview of the different imaging techniques is presented in the next part. It focuses on the two families of technologies that we are interested in in this manuscript: structural visualisation with holographic OCT and blood flow imaging with laser Doppler holography. Finally, we discuss about the difficulties brought by these types of imaging, due to motion and aberrations of the eye.



**Figure 1.1** – Eye-fundus and optical coherence tomography (OCT) of a retina taken with Spectralis device. The green arrow on the eye-fundus image corresponds to the scanning position for the second image. The fovea, center of the vision, is well-visible on both images, designated with the blue arrows. The red arrow shows the same vessel.



**Figure 1.2** – Optical configuration of the retina for a perfect eye. Beams coming from an object located at infinity are focused on the retina. Image plane (retina) and focal object plane are conjugated by the optical system of the eye.



**Figure 1.3** – Drawings by Dave Schumick, extracted from [2]. A: Scheme of the cross section of the optical nerve along the superior-inferior axis. Two main arteries from the ophthalmic artery irrigate the fundus: the central retinal artery reaches the eye via the optical nerve, while posterior ciliary arteries supply blood to the choroid. B: Scheme of vasculature in retina and choroid next to the fovea. The choroid is a large and dense vascular network which feeds all the outer retina via the choriocapillaris.

### 1.1.1 Some information about the eye

The eye is a spherical organ of about 2.5 *cm* of diameter which can be compared to an optical system. It is mainly composed of two lenses, a pupil, and a sensor (fig. 1.2).

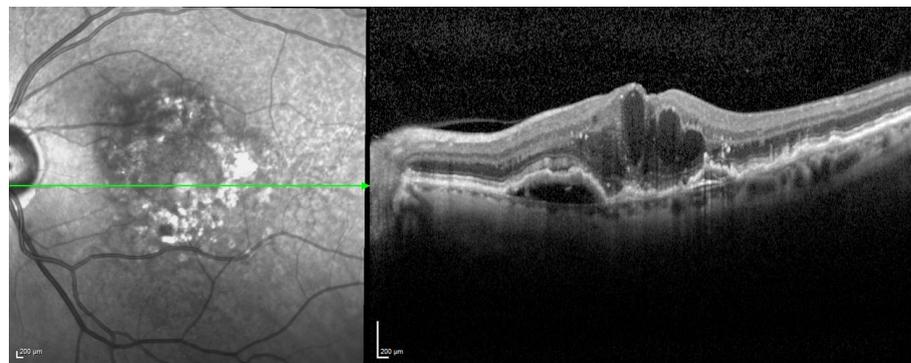
The retina stands for the sensor in this model. The fig. 1.1 represents an image of the surface of the retina (eye-fundus image), and its depth-resolved structure (OCT image). The retina is composed of numerous layers and irrigated by blood vessels from the choroid bringing oxygen and blood (see fig. 1.3). The largest arteries have a typical diameter of 100 – 150  $\mu\text{m}$  while the capillaries are 3.5 – 6  $\mu\text{m}$  large. The retina captures light through photoreceptor cells layers, composed of about 5 million cones and 120 million rods. The light signal is converted into nerve impulse by the photoreceptors, and sent via multiple downstream neurons to the brain via the ganglion cells whose axons form the optic nerve. The ganglion cells transmit the information

to the brain via their axons, which form the optic nerve. In order to have the best signal, light must be focused on the retina. This is the role of the cornea and the crystalline lens.

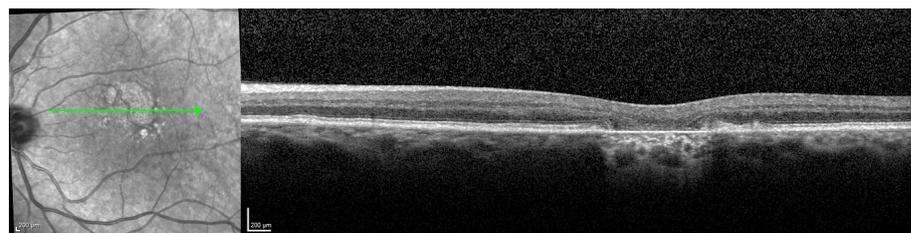
The cornea is the first lens of the system and constitutes the main protection of the eye from the outside. It is a  $0.5\text{ mm}$  thick tissue which is rigid thanks to collagen, its main component. The cornea allows more than 90 % of light transmission in  $550\text{ nm}$  and  $1200\text{ nm}$  wavelength band [3]. The crystalline lens is a deformable lens around  $18\text{ mm}$  away from the retina. Ciliary muscles modify its curvature, which changes its vergence. Then, for normal eyes, this system of two lenses is able to focus the light on the macula, which is the area of the retina containing the majority of photoreceptors. The pupil shrinks and grows to regulate the light intensity entering in the eye.

### 1.1.2 Retina pathologies

Through retina study, physicians can get a lot of information about pathologies involving the vision. The morphology of the retina and the blood flow often characterize diseases. Several pathologies are presented in this part.



(a) Subretinal neovascularization of the macula due to AMD.



(b) Atrophy of the retina due to AMD.

**Figure 1.4** – Eye-fundus and SDOCT images of patients with AMD taken with Spectralis device. Green arrows of the eye-fundus images correspond to OCT images.

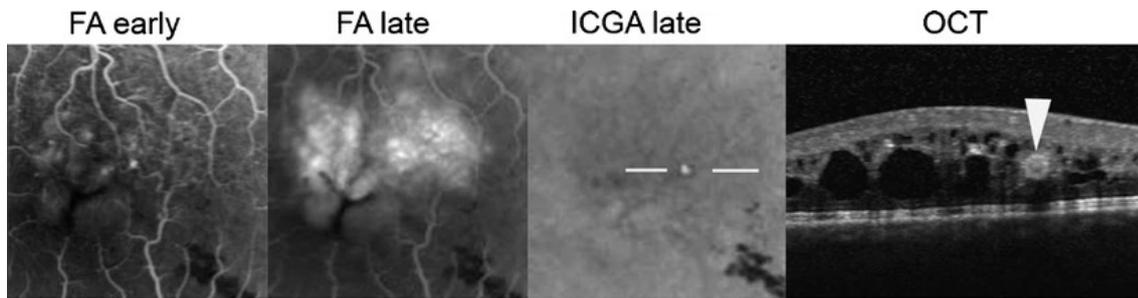
#### Age-related macular degeneration

Age-related macular degeneration (AMD) is the most frequent cause of irreversible decrease of central vision (governed by the macula) for elderly people. This pathology is divided into two types: wet and dry. In any case, the first step of AMD is dry AMD. 85% of people suffering from the disease have dry AMD. It affects retinal pigmentary epithelium, directly responsible for cones and rods, and provokes the appearance of deposits on the pigmentary epithelium (drusen). Most advanced stages provoke chorioretinal atrophy (fig. 1.4(b)). Among the 15%

other people, 80% – 90% are suffering of severe vision loss. Abnormal blood vessels appear on the choroid under the retina (choroidal neovascularization) in the wet AMD form.

### Arterial or vein occlusion

Arterial or vein occlusion often creates new blood vessels in the retina layers (retina neovascularization). Then, vein exudation can occur, provoking fluid accumulation in the macula and swelling (macular edema).



**Figure 1.5** – Early (image taken during the first 2 minutes after contrast agent injection) and late (image taken after 8-10 minutes) fluorescein angiography (FA), late indocyanine green angiography (ICGA) and high-resolution optical coherence tomography (OCT) images of macro-aneurysm (represented by the white arrow) due to retina vein occlusion [4]. The white line in the ICGA image localizes the OCT scan. The macular edema caused by the occlusion is well-visible with FA and OCT.

### Glaucoma

Glaucoma is a variety of degenerative diseases affecting the ganglion cells. It provokes irreversible blindness, starting with the loss of peripheral vision. Glaucoma is the second cause of blindness in the world after cataract. The pathology is often diagnosed by a visual field examination, but this test only reveals advanced-step glaucoma. It has been shown intraocular pressure is highly related to the development of glaucoma [5–7]. Furthermore, the reduction of blood flow has been shown to be linked to glaucoma [8–10].

### Diabetic retinopathy

This pathology caused by diabetes is a major cause of adult blindness. The first stage is non-proliferative and has no symptoms. During the development of the disease, capillary permeability increases, provoking microaneurysms, hemorrhages, exudations, macular edema... Its proliferative stage causes formation of preretinal vessel neovascularization on the optic nerve or retinal surface.

For all these diseases, retina observation and sectioning and blood flow imaging techniques are important to study characteristics and progresses of the pathologies. The improvement of these techniques is crucial to have a better understanding of those pathologies. Several imaging techniques to image the retina are presented in the following part.

## 1.2 Imaging techniques

First observations of the retina have been made possible thanks to funduscopy and fundus photography invented in the middle of the nineteenth century by Hermann von Helmholtz. With this technology, physicians are able to detect both forms of AMD. They can also diagnose diabetic retinopathy and glaucoma.

Later, several optical methods have emerged to provide other types of information about the retina, specially about layers shape with optical coherence tomography and local dynamics inside the vessels with blood flow imaging.

### 1.2.1 Optical coherence tomography

Optical coherence tomography (OCT) [11–13] is a very powerful imaging technique, invented in 1991, for imaging biological tissues. It performs a non-invasive optical depth-sectioning of the samples. OCT has emerged when ultrasounds and microscopy technologies were dominating biological imaging. Microscopy has high axial and transverse resolutions (inferior to  $1 \mu m$ , depending on the aperture), but limited penetration in the tissues (about  $100 \mu m$ ). In contrast, ultrasound imaging can provide images with a penetration depending on the frequency of the sound wave, of several centimeters for tens of megahertz, and several millimeters for a hundred of megahertz [14, 15]. However, it is limited to  $0.1 - 1 mm$  in axial resolution for a wave frequency of tens of megahertz, and around  $20 \mu m$  for a hundred of megahertz.

OCT brings a good compromise between resolution and penetration depth, or microscopy and ultrasounds: it can reach a  $2 mm$  of penetration depth, with an axial resolution of  $1 - 10 \mu m$ .

For a gaussian-shape spectrum beam like the laser, the axial resolution of OCT is defined as:

$$\Delta z = \frac{2 \ln(2)}{\pi} \frac{\lambda^2}{\Delta \lambda}, \quad (1.1)$$

where  $\Delta \lambda$  and  $\lambda$  are the spectral bandwidth and the center wavelength of the light source [16]. The transverse resolution of OCT is:

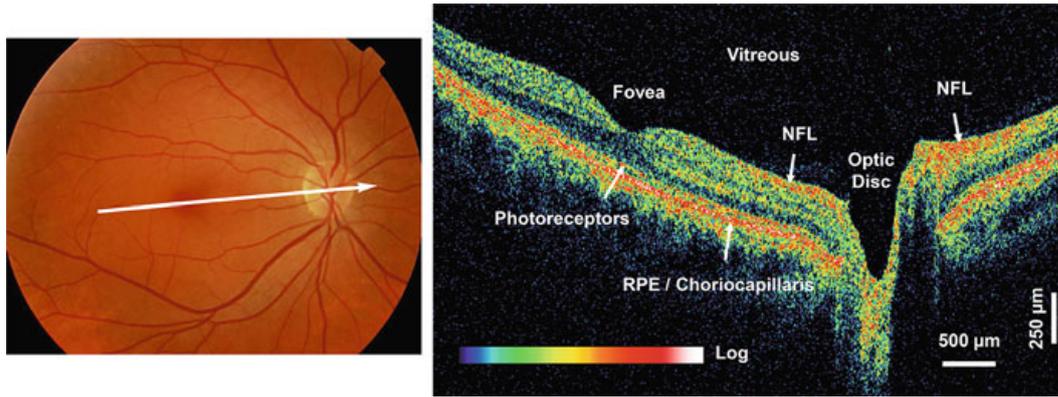
$$\Delta x = \Delta y = \frac{4 \lambda f}{\pi d} = \frac{2 \lambda}{\pi NA}, \quad (1.2)$$

where  $d$  is the diameter of the incident beam on the objective lens and  $f$  is the focal distance [12].  $NA$  is the numerical aperture of the beam. The transverse resolution is only determined by the diffraction limit, which can be assimilated to  $\lambda/NA$ .

These definitions show one aspect of the power of OCT: the axial resolution is independent of the transverse resolution. Then, OCT images can reach the same resolution as in microscopy, with a fine resolution in depth too.

Concerning the penetration depth, we can define the depth of field as two times the Rayleigh range [12]:

$$b = 2z_R = \frac{\pi(\Delta x)^2}{\lambda}. \quad (1.3)$$



**Figure 1.6** – Eye-fundus image and TDOCT image of in vivo human retina, extracted from [24]. The right arrow of the first image corresponds to the depth profile view. Several layers are visible on the OCT image.

This brings a limit to the OCT: improving the transverse resolution implies reducing the penetration depth.

Ophthalmology is the main field where OCT is used, specially due to the properties of high transmittance of ocular media which ease its use. Thanks to OCT, retina layers can be imaged, which is convenient to study pathologies of the eye fundus.

OCT images can help to evaluate macular edema severity, which is the most frequent cause of blindness due to diabetic retinopathy. Macular edema due to vein or arterial occlusion can also be observed with OCT (fig. 1.5). OCT is often performed to identify intraretinal and subretinal fluids. It is also used to control AMD progression, as the pathology's effects are well-visible on OCT images (fig. 1.4).

The main evolutions of OCT bring an increase of resolution and speed. The first OCT, time-domain (TDOCT), was outperformed by the spectral-domain approach (SDOCT), improving signal-to-noise ratio [17,18] and speed. Other types of OCT are also in development, like line-field OCT [19], time-domain full field OCT [20,21], off-axis time-domain full-field OCT [22] and off-axis swept-source full field OCT [23].

### Time-domain OCT (TDOCT)

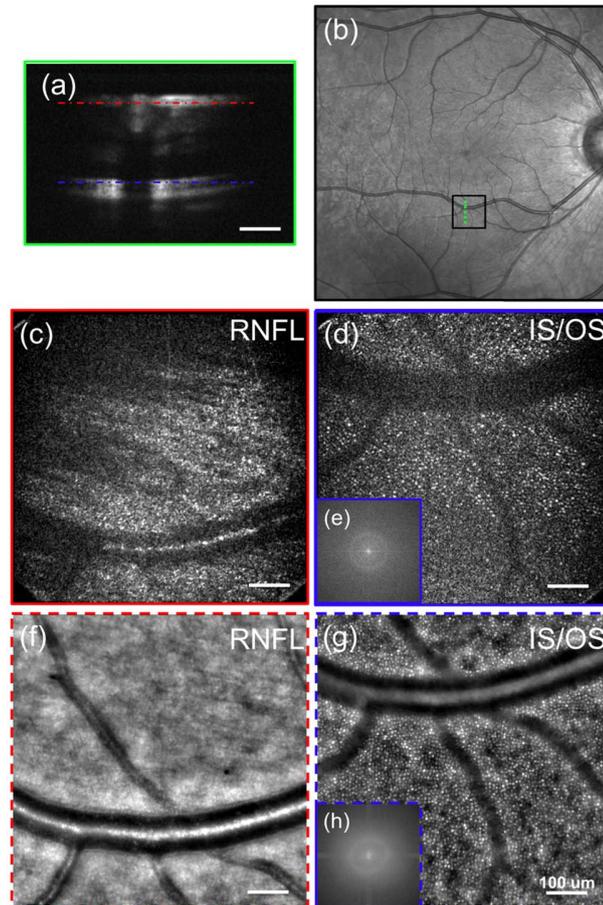
TDOCT is the first generation of OCT which has been commercialized. An example of imaging of the retina is represented on the fig. 1.6. It uses time-domain approach for amplitude and phase retrieval [11,25]. It is generally composed of a Michelson interferometer. A beam emitted by a low-coherence light is divided into two parts. The first one is sent and reflected on the object before reaching the sensor: this is the object arm. The second one is reflected by a mirror and sent to the sensor to interfere with the first wave: this is the reference arm.

Interference between both waves can happen only if the path length difference between both arms is below the coherence length of the source, which is equal to the axial resolution  $\Delta z$  (eq. (1.1)). Thus, by moving the reference mirror, we can reduce the path length difference and image the object. The amplitude of the signal recorded depends on the reflectivity of the sample. The plot of this amplitude as a function of the depth is called A-scan, and corresponds to the reference mirror scanning.

The lateral scan is performed by moving the object or the sensor to get other A-scans.

One row or one column of A-scans give a B-scan which image a depth profile. The C-scan corresponds to all the acquired volume.

TDOCT acquisition is very time-consuming because of these two scans. Therefore, all the researches from there have focused on increasing the speed of OCT (acquisition and computation of images), while trying to improve image quality as well.



**Figure 1.7** – Images of in vivo human retina, extracted from [21]. (a) SDOCT depth image. It only serve to localize retina layers. (b) Eye-fundus photography. The black box indicates the FFOCT field, while the green dashed line represents the OCT scan of (a). (c) Retinal nerve fiber layer (RNFL) and (d) inner/outer segment photoreceptor layer (IS/OS) imaged with FFOCT. RNFL is located by the red dashed line in (a) and IS/OS by the blue dashed line in (a). (e) 2D power spectrum of (d). (f) RNFL and (g) IS/OS imaged with adaptive optics (AO) retinal camera, at the same location as (c) and (d). (h) 2D power spectrum of (g). The axial sectioning of the FFOCT is better than the AO retinal camera: (f) shows images of vessels which do not appear on (c), because FFOCT selects with more precision the layer to observe.

### Time-domain full-field OCT (FFOCT)

One way to increase OCT speed is to parallelize acquisitions using a 2D-camera. Then, the depth-scan is still necessary, but en-face 2D images are recorded. This version of the OCT has been introduced at Institut Langevin in 1998 [20].

FFOCT is based on an interferometric setup too. Microscope objectives have been added to both reference and object arms in order to increase the transverse resolution. To compute a

single en-face image, 4 images are acquired by the camera at 4 different positions of the reference mirror, controlled by a piezoelectric motor, in order to modulate the intensity interference pattern. The phase difference between each images is well-known. The full-field image can be computed from these 4 images, via a 4-phase temporal demodulation [26]. The fig. 1.7 shows the kind of images obtained with FFOCT device.

### Spectral-domain OCT (SDOCT)

Another way to decrease acquisition time is to remove the depth scan. Instead of acquiring one pixel by axial line, the interference spectrum corresponding to the whole line is measured with a spectrometer. The reference mirror stays still, which saves time. This is the main difference between SDOCT and TDOCT.

The signal produced by interferometry and recorded by the camera is :

$$I_0 \propto |E_r|^2 + |E_s|^2 + 2E_r E_s \cos(2k\Delta L), \quad (1.4)$$

where  $I_0$  is the recorded intensity,  $E_r$  is the reference field,  $E_s$  is the signal field,  $k$  is the wave number and  $\Delta L$  is the path length difference between arms. The cosine modulates the signal at a frequency depending on  $\Delta L$ . The spectrometer records the frequencies, corresponding to different  $\Delta L$ . The amplitude of the spectrum is proportional to the sample reflectivity. Then, a map of the different object positions and reflectivities is computed as A-scan. Moving the sample sideways allows the construction of B-scans.

Nowadays, SDOCT is one of the most used OCT technique to image retina of patients. Images acquired with SDOCT are shown in fig. 1.1, fig. 1.5 and fig. 1.4.

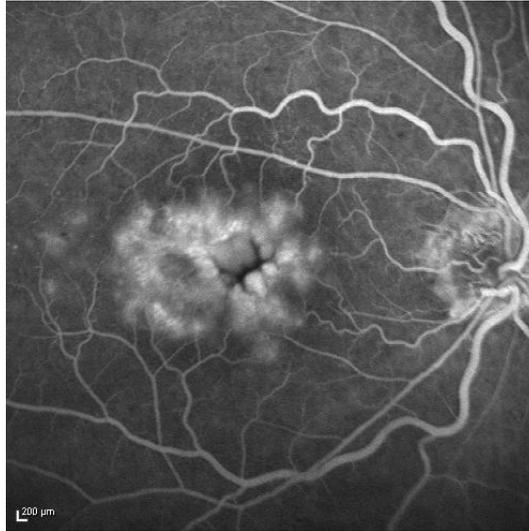
### Swept-source OCT (SSOCT)

Swept-source OCT is an extension of SDOCT. SSOCT uses a narrow bandwidth light source with a frequency sweep to acquire interferometer spectrum, while SDOCT uses a low-coherence light source to record interferometer spectrum by moving the reference mirror. Thus, the modulation is not spatially encoded but depends on the frequency of the light source, varying with time. SSOCT performs a fast scanning limited by the laser sweep speed and the sensor sample frequency.

With the constant need for acceleration of acquisition, FFOCT and SSOCT were combined to form the full-field swept-source OCT (FF-SSOCT) [27, 28]. It allows the acquisition of en-face 2D images with a fast depth mapping with swept-source. Hence the 3D volume is more quickly available.

### Digital holography and FF-SSOCT

The first images of eye fundus in vivo by holographic ophthalmoscopy [29–32] paved the way to the development of time-averaged holography for retinal vessel imaging. Digital holography ensures the recording of entire wave fields. The images are computed from the interference pattern. Then, the reconstruction of images is not limited to a precise depth: numerical refocusing can be achieved with phase and amplitude information. Image quality remains



**Figure 1.8** – Fluorescein angiography (FA) images. FA highlights retina vessels and arteries and the edema in the center of the image, due to hemorrhages of the vessels.

the same regardless the reconstruction depth. However, digital holography does not provide cross-sectional imaging. Digital holography technique will be explained in the section 2.1.

The combination of FF-SSOCT and digital holography allows to get 3D volumes [12, 33, 34].

OCT remains one of the most used imaging technology in ophthalmology. All the variations presented previously are still improving by different teams around the world. The combination of several techniques, like full-field imaging and OCT (FFOCT), FFOCT and SSOCT (FF-SSOCT) or digital holography and FF-SSOCT, are more and more studied: they open up opportunities, each technique bringing its own set of advantages.

### 1.2.2 Blood flow imaging

Blood flow imaging has been developed at the end of the twentieth century. As it has been said in the section 1.1.2, many pathologies involve blood circulation, making the evolution of imaging techniques interesting.

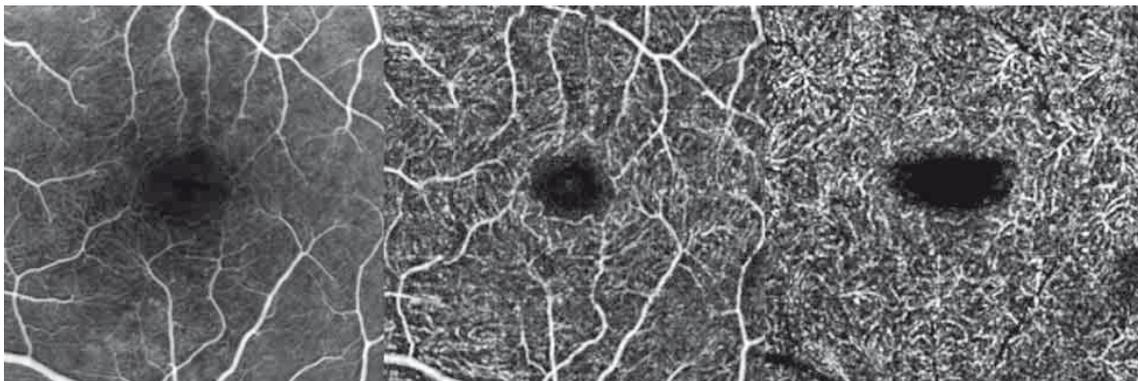
Nowadays, fluorescein angiography (FA) and indocyanin-green angiography (ICG-A) are the most used imaging techniques but also the most invasive ones. They consist in an intravenous injection of fluorescent contrast agent which emits at a precise wavelength in veins and arteries. FA is usually employed in first investigation to observe veins and arteries of the retina (fig. 1.8). The contrast agent in FA is sodium fluorescein which emits at  $520 - 530 \text{ nm}$  after being exposed at a light of  $465 - 490 \text{ nm}$ . To observe the choroidal vasculature, ICG-A (fig. 1.9) is preferred because the indocyanine green emits at  $790 - 805 \text{ nm}$  and then penetrates more easily retina's pigmented tissues. Choroidal veins are clearly visible with ICG-A, while choroidal arteries can be seen only in a short time window just after the contrast agent injection. Besides the fact intravenous injection is invasive for the patients, the dyes can provoke adverse effects or anaphylactic reactions. Hence several other technologies have emerged to study blood flow in a less invasive way.

First images of retina and choroid microvasculature with OCT-angiography (OCT-A) have been shown in 2008 [35]. To obtain an image, several B-scans are acquired with SDOCT or



**Figure 1.9** – Images of ICG-A. Retina and choroid vessels are highlighted by the dye. Choroidal vasculature is not visible.

SSOCT device and amplitude and phase comparisons between frames are computed to detect motion within the vessels. Besides the non-invasive quality of OCT-A, it provides quantitative analysis of blood flow in retina and choroid and using OCT device allows 3D images [36]. OCT technique brings also a depth sectioning that does not exist in FA or ICG-A, to image vessels and arteries from more precise layers of retina or choroid (fig. 1.10). The retinal and choroidal neovasculatures can be localised by OCT-A, which is useful in case of arterial or vein occlusion, wet AMD, or diabetic retinopathy. However, there are some drawbacks inherent to the OCT technology. Deeper vessels can be difficult to image due to too high absorption by tissue and shadowing artifacts, where superficial vessels can hide the deepest ones [37]. OCT-A is also highly affected by eye and body motion. The field of view is limited. Eventually, low amounts of blood flow may not be detected by the device.

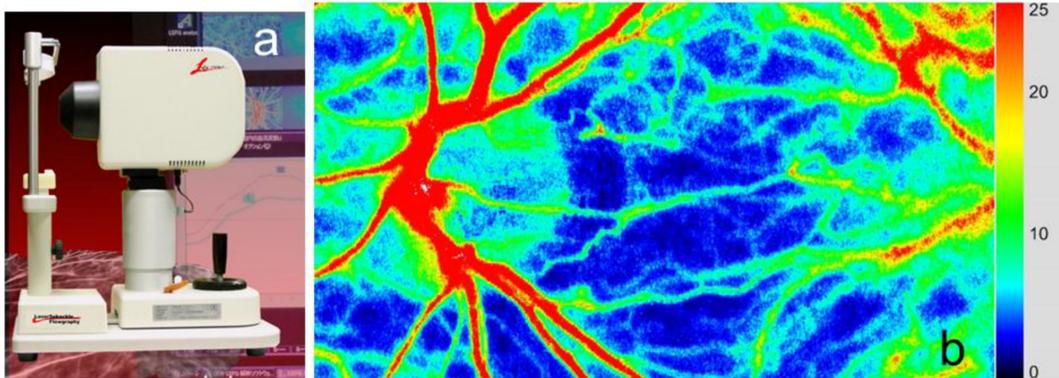


**Figure 1.10** – FA (first image), OCT-A of the inner retinal vascular plexus (second image) and OCT-A of the outer plexus (third image) of the same central macular region, extracted from [38]. OCT-A lets a depth sectioning of the layers, while FA highlights the most superficial vasculatures of the retina without differentiating the depth of imaging.

Laser speckle flowgraphy or laser speckle contrast has been developed by Fercher and Briers [39] in 1981 to observe blood flow (fig. 1.11). Retina images are recorded while illuminated by a laser. The speckle pattern variations between frames are converted into blurs of different intensities, which indicate the local velocity of blood flow. Recently, a study has showed that

one type of glaucoma can be detected by analysing blood flow with laser speckle flowgraphy [40]. Laser speckle contrast requires a simple setup and camera. For now, it has not been used to observe choroid vasculature.

Laser Doppler flowmetry uses self-interferences of a wave backscattered by the retina to provide blood flow measurement [41–43]. Laser Doppler flowmetry researches showed the choroidal blood flow decreases with age [44] and with the evolution of the AMD [45]. The main limitation of this method is the field of view: the measurement is performed on a single point, and obtaining larger field of view takes time.



**Figure 1.11** – Laser speckle flowgraphy NAVI device (a) and image of the retina blood flow. [46]

Ultrasounds can also be used for blood flow imaging. The Doppler effect is measured by sending acoustic waves on the retina and the local fluctuations can be measured. As mentioned in section 2.2, ultrasounds have the advantage of penetrating deep into the tissues. However, the imaging resolution is very low compared to other methods.

Laser Doppler holography [47, 48] is a full-field interferometric instrument which is able to image retina and choroid blood flow. It is very similar to laser Doppler flowmetry, with the addition of an interferometric setup to make the wave interfering with the reference beam. This setup allows full-field acquisition, and post processing of the data by recording the entire field.

### 1.2.3 Conclusion on imaging techniques

The various technologies presented in this part have been developed with a view to increasing speed and improving resolution in eye imaging. With the rise of computation power, optical instruments are increasingly being replaced by calculation on faster computers, which is well illustrated by the use of digital holography to get tomographic or blood flow images of the retina.

Besides problems inherent to the imaging system itself, optical devices are subject to external disturbance, bringing also difficulties in imaging and loss of accuracy. The main obstacles these technologies face are eye motion and eye aberrations.

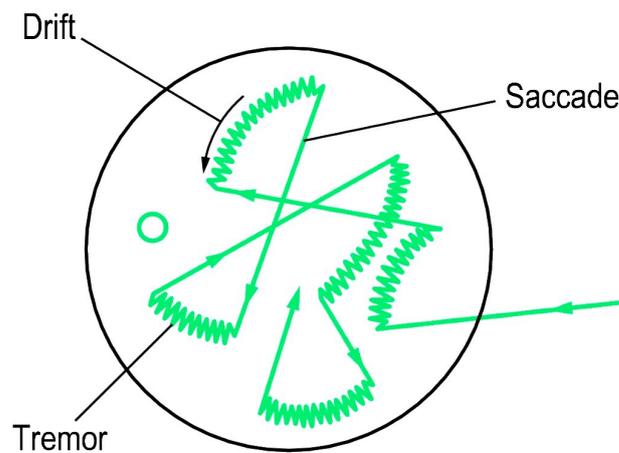
## 1.3 Eye motion

### 1.3.1 Fixation movements

The human eye is constantly subject to movements, even when trying to fixate a precise point. The vision could not be possible without those movements [49]. If photoreceptors receive the same signal without any change, the response of neural cells will be weak and tend to vanish. Constant motion makes the neuronal response strong.

Fixation movements are divided into three main classes depending on their frequency, speed and duration:

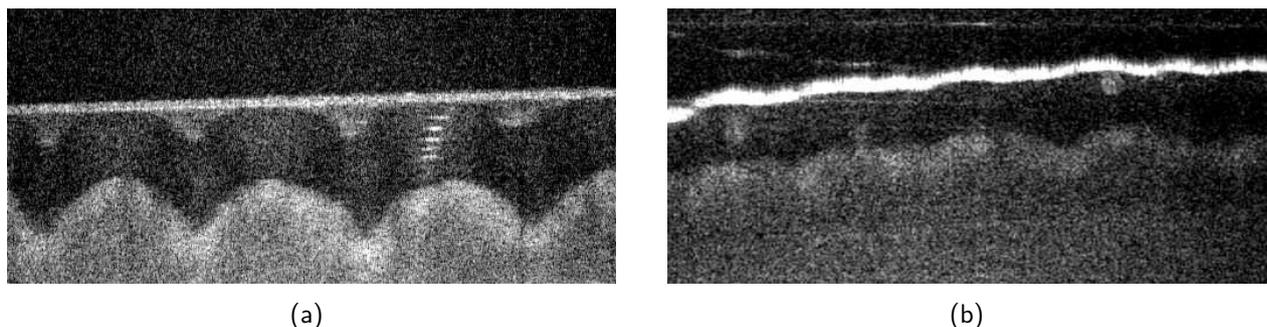
- **Tremor:** this is an aperiodic movement with a characteristic bandwidth between 10 and 200  $Hz$ . It has the smallest amplitude with 20  $arcsec$ .
- **Drift:** this is the dominant movement in the eye. It is slower than the tremor with a characteristic speed between 6 and 30  $arcmin/s$ . It happens simultaneously with tremor. Its mean amplitude varies between 2 and 5  $arcmin$ .
- **Microsaccade:** this is the fastest movement with a displacement speed from 600 to 6000  $arcmin/s$ . The amplitude varies between 2 and 20  $arcmin$ .



**Figure 1.12** – Movements of the eye during fixation of about 10 seconds. The large circle stands for 10 minutes of arc in diameter, while the small one represents the dimension of a photoreceptor. Tremors occur during drift, and saccades (microsaccades) tend to refocus the vision. Credits to [50].

During the fixation, drift occurs, with tremors, moving the focus point of the vision. Microsaccades make up for the distance covered by the drift and focus the image of the object on the fovea. These movements are represented on the fig. 1.12.

Added to heartbeat, respiration and tissue motion, fixation movements can shift or even remove the object of interest from the image during several frames. It can cause blur when averaging the images, or even misencoding of the signal. For example, in OCT, the depth information is crucial for imaging. When an axial drift occurs, the depth information encoded in interferograms is spoiled, decreasing the axial accuracy (fig. 1.13). In the following, we will differentiate lateral motion and axial motion.



**Figure 1.13** – B-scans of a finger tip acquired with SSOCT with stabilization (a) and without stabilization (b). Motion disturbs the rendering. Images extracted from [51].

### 1.3.2 State of the art in motion compensation

Compensating the lateral movement means stabilizing the object of interest in the center of the images. This is done by finding a geometric transformation in order to minimise the distance between the two images to align. This is called registration.

Registration techniques have been developed in the past years, specially in the medical field [52, 53]. The word itself often refers to making two images from the same patient (taken at different times) coincide. Registration methods can also be used to compensate lateral movement due to fixation movements.

Most of the literature concerns radiological modalities, but researches exist in retina imaging. These are specially concerning eye fundus images. Other studies exist to register images from different modalities like OCT and fundus photography [54, 55].

With eye-fundus images, feature-based registration [56–58] is preferred to intensity or area-based registration. Indeed, correspondence between two images through points, lines and contours can be easily found in fundus photography and OCT: optic disc and vessels are salient features. However, too much noise can create false descriptors and mislead the algorithms. If there are not enough prominent details, intensity-based registration methods are used, like cross-correlation [59], phase correlation [60] and mutual information comparison [61] between images involved in registration. These methods give results with a root mean square error of the order of degree. Image stabilization has also been shown by using a tip/tilt mirror [62], reaching a root mean square error of about  $0.34 - 0.53$  *arcmin* or  $1.66 - 2.56$   $\mu m$ .

As far as the axial motion compensation is concerned, it has more impact on OCT device. Indeed, it causes a phase difference between acquisitions and then disturbs the encryption of the depth data within the phase. Axial tracking can be used to stabilize the axial plane at a depth of interest [63], reaching cellular resolution, but it increases system complexity. Another methods consist in computing and retrieving the phase corresponding to the motion after the full volume acquisition [51, 64], which allows an improvement in resolution of the order of  $10$   $\mu m$ .

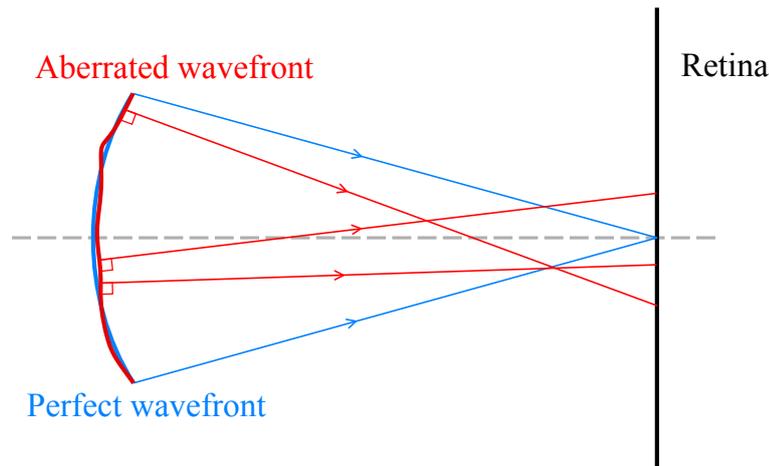
Depending on the imaging conditions, optimizing motion compensation techniques may be more or less necessary. In section 2.4, we will study the motion compensation requirements of our system and select the appropriate algorithms.

The following section presents another difficulty for retina imaging which is the main topic of interest of this manuscript: aberrations caused by the eye.

## 1.4 Eye aberrations

### 1.4.1 Sources of aberrations in the eye

No optical system is perfect and the eye is no exception to the rule. Light wave can be more or less distorted by an optical system, provoking changes in output image. These defects are aberrations, characterised by the transformation of the point spread function (PSF) of the system. There are chromatic aberrations, provoked by the dispersion of the ocular media [65,66], and geometric aberrations, decreasing the image resolution. This last type of aberration will be detailed in the following.



**Figure 1.14** – Comparison between wavefronts from a perfect eye (perfect wavefront) and a normal eye (aberrated wavefront). Light crossing the cornea and lens system sees its wavefront locally distorted. This prevents the light from focusing on a single point on the retina as is the case with a perfect eye.

In the eye, the main source of geometric aberration is the system of dioptrics created by the cornea, the crystalline lens and the tear film [67]. Wavefront of the light going through this combination is distorted, which changes wave phase locally (see fig. 1.14). The distortion occurs even for healthy eyes, but is more penalizing for imaging techniques when the lens is affected by strong aberrations, as for myopia or hypermetropia for example.

### 1.4.2 Mathematical description of aberrations

#### Aberrated phase

Several methods and conventions exist to describe and quantify aberrations. They all evaluate the error to the non-aberrated wavefront. One of them consists in projecting this phase  $\phi_a$  on a polynomial basis:

$$\phi_a(\rho, \theta) = \sum_{i=0}^{\infty} a_i P_i(\rho, \theta), \quad (1.5)$$

with  $\theta$  the azimuth angle,  $\rho$  the radial distance as  $0 \leq \rho \leq 1$ ,  $a_i$  a real coefficient and  $P_i(\rho, \theta)$  the  $i^{\text{th}}$  order of the polynomial  $P$ . The most used polynomial is Zernike polynomial.

### Zernike polynomials

Zernike polynomials are divided between even and odd polynomials. Even polynomials are defined by:

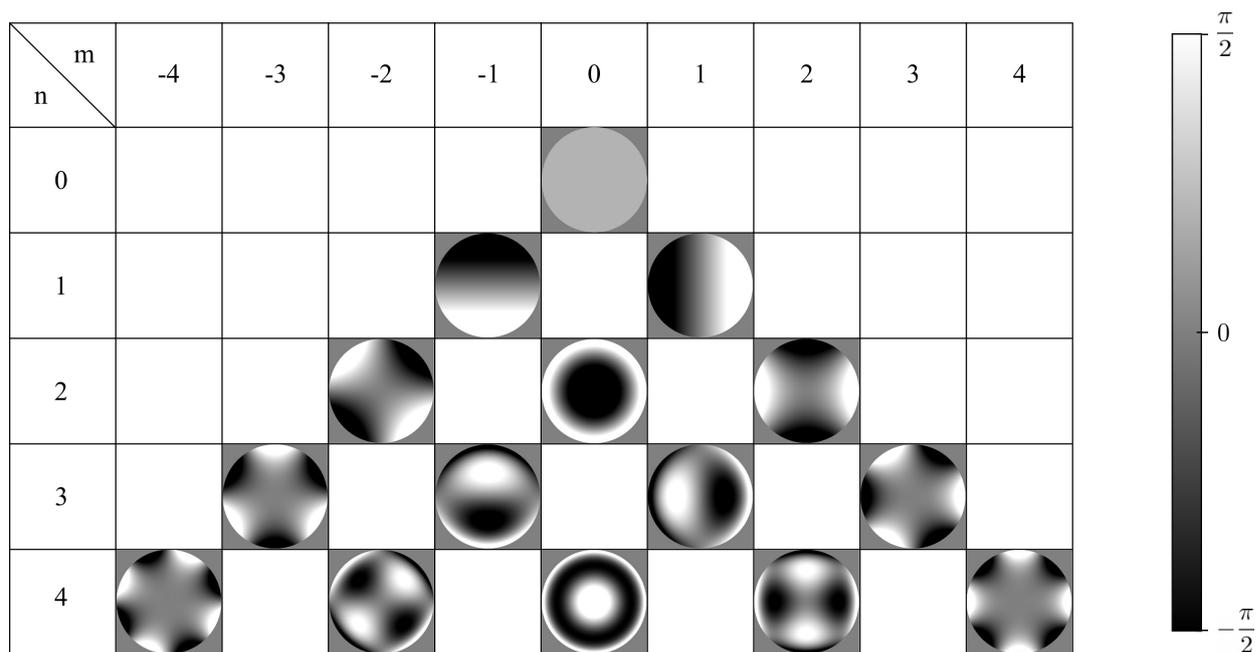
$$Z_n^m(\rho, \theta) = R_n^m(\rho)\cos(m\theta), \quad (1.6)$$

whereas odd polynomials are expressed as:

$$Z_n^{-m}(\rho, \theta) = R_n^m(\rho)\sin(m\theta), \quad (1.7)$$

where  $m$  and  $n$  are integers as  $n \geq m$ , with  $m$  the azimuth frequency and  $n$  the radial order, and  $R_n^m$  a radial polynomial as:

$$R_n^m(\rho) = \sum_{k=0}^{\frac{n-m}{2}} (-1)^k \binom{n-k}{k} \binom{n-2k}{\frac{n-m}{2}-k} \rho^{n-2k}. \quad (1.8)$$



**Figure 1.15** – Wavefronts corresponding to the first Zernike polynomials in OSA/ANSI convention [68], classified by radial and azimuth frequency  $n$  and  $m$ .

Zernike polynomials can be described under Noll [69] or OSA/ANSI [68] conventions. For the rest of this manuscript, we will use the OSA/ANSI standard. The fig. 1.15 illustrates the phase corresponding to the first modes of Zernike polynomials in this convention. The following formula is used to have the  $i^{\text{th}}$  index of OSA/ANSI standard from  $m$  and  $n$ :

$$i = \frac{n(n+2) + m}{2}. \quad (1.9)$$

This formula links  $Z_n^m$  and  $Z_n^{-m}$  to  $Z_i$ .

Zernike polynomials are normalized such that:

$$\int_{\theta=0}^{2\pi} \int_{\rho=0}^1 Z_i^2(\rho, \theta) \rho \, d\rho d\theta = \pi. \quad (1.10)$$

In the human eye,  $Z_4$  ( $Z_2^0$ , corresponding to defocus) and  $Z_3$  and  $Z_5$  ( $Z_2^{-2}$  and  $Z_2^2$ , corresponding to oblique and horizontal astigmatisms respectively) are the biggest contributors to eye aberrations [70]. The vision defaults corrected by glasses correspond to these modes, which are low-order aberrations. High-order aberrations correspond to modes from the third radial order  $n$ .

### Aberrated phase expressed with Zernike polynomials

The phase corresponding to aberrations can be written as follows:

$$\phi_a(\rho, \theta) = \frac{2\pi}{\lambda} \sum_{i=0}^{\infty} a_i Z_i(\rho, \theta), \quad (1.11)$$

with  $\lambda$  the wavelength and  $Z_i$  the  $i^{\text{th}}$  index Zernike polynomial in the OSA/ANSI standard. Zernike polynomials form an orthonormal basis. Then, the  $a_i$  coefficients can be expressed as:

$$a_i = \frac{1}{\pi} \int_{\theta=0}^{2\pi} \int_{\rho=0}^1 \phi_a(\rho, \theta) Z_i(\rho, \theta) \rho \, d\rho d\theta. \quad (1.12)$$

The mean of  $\phi_a$  is the first coefficient  $a_0$ . Its variance, called squared wavefront error (SWFE), is defined by:

$$SWFE = \sum_{i=1}^{\infty} a_i^2. \quad (1.13)$$

It is noticeable that the SWFE does not take into account the mean  $a_0$ , and starts from index 1. The wavefront error (WFE) is defined as:

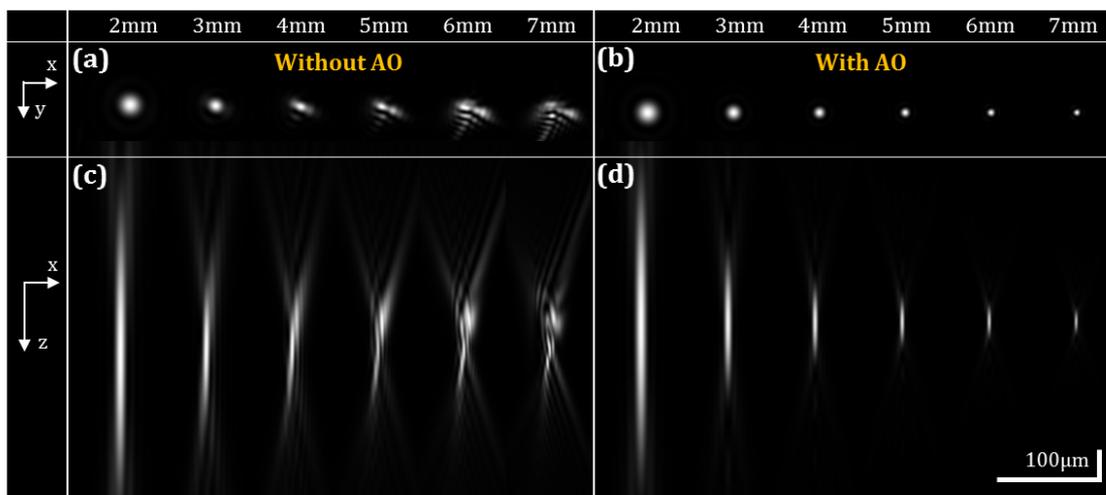
$$WFE = \sqrt{SWFE}. \quad (1.14)$$

### Impulse response

We consider  $P(x, y)$  the pupil function in the pupil plane:

$$P(x, y) = p(x, y) e^{i\phi_a(x, y)} = p(\rho, \theta) e^{i\phi_a(\rho, \theta)}, \quad (1.15)$$

where  $x$  and  $y$  are the Cartesian coordinates corresponding to the polar coordinates  $\rho$  and  $\theta$ , and  $p(x, y) \in \mathbb{R}$ . Then, the response to a punctual source located in the focal image plane is



**Figure 1.16** – PSF of a normal eye for different pupil diameters in en-face plane  $(x, y)$  (a) and in-depth plane  $(x, z)$  (c). PSF of a normal eye corrected by adaptive optics, equivalent to a perfect eye, for different pupil diameters in en-face plane  $(x, y)$  (b) and in-depth plane  $(x, z)$  (d). Without adaptive optics (AO), equivalent to a normal eye, the larger the NA is, the more distorted the PSF of the system is. With AO, equivalent to the perfect eye, the PSF precision increases with the pupil diameter. Credits to [71].

called point spread function (PSF) and defined as:

$$PSF(x, y) = |FT2D\{P(x', y')\}|^2, \quad (1.16)$$

where FT2D is the 2-dimension Fourier transform:

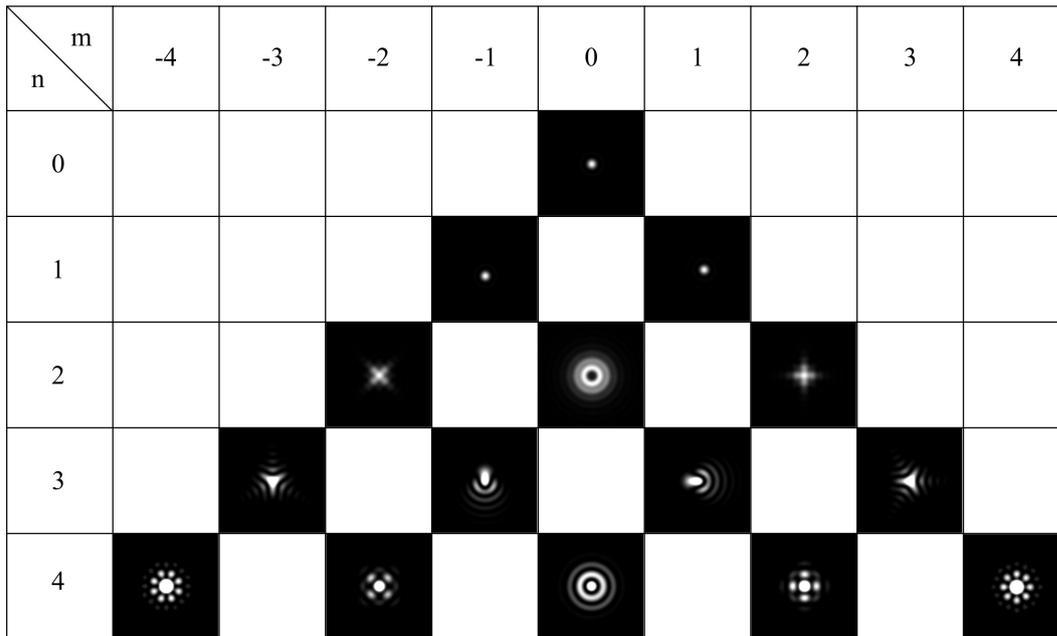
$$FT2D\{P(x', y')\}(x, y) = \iint_{-\infty}^{+\infty} P(x', y') e^{2i\pi(x'x+y'y)} dx' dy'. \quad (1.17)$$

PSFs are used to define the power of resolution of a system. If there are aberrations, it becomes:

$$PSF(x, y) = |FT2D\{P(x', y') e^{j\phi_a(x', y')}\}|^2. \quad (1.18)$$

For a perfect eye, the PSF size is only limited by diffraction. However, for a normal eye, the larger the pupil is, the more distorted the PSF is [71]. This is illustrated by the fig. 1.16, where the perfect eye is represented by a normal eye whose PSF has been corrected by adaptive optics (see below section 1.4.3). The fig. 1.17 illustrates the effect of each Zernike mode on the PSF.

The effects of the aberrations on the PSF show the deformations that can be caused on an image: as the image of a point can be distorted like in fig. 1.16 and fig. 1.17, the image of the retina can be blurred. Therefore, aberrations need to be removed to increase the resolution of retina images. In the next part, a state of the art of the methods developed to compensate aberrations is presented.



**Figure 1.17** – PSF corresponding to the first Zernike polynomials in OSA/ANSI convention [68], classified by radial and azimuth frequency  $n$  and  $m$ .

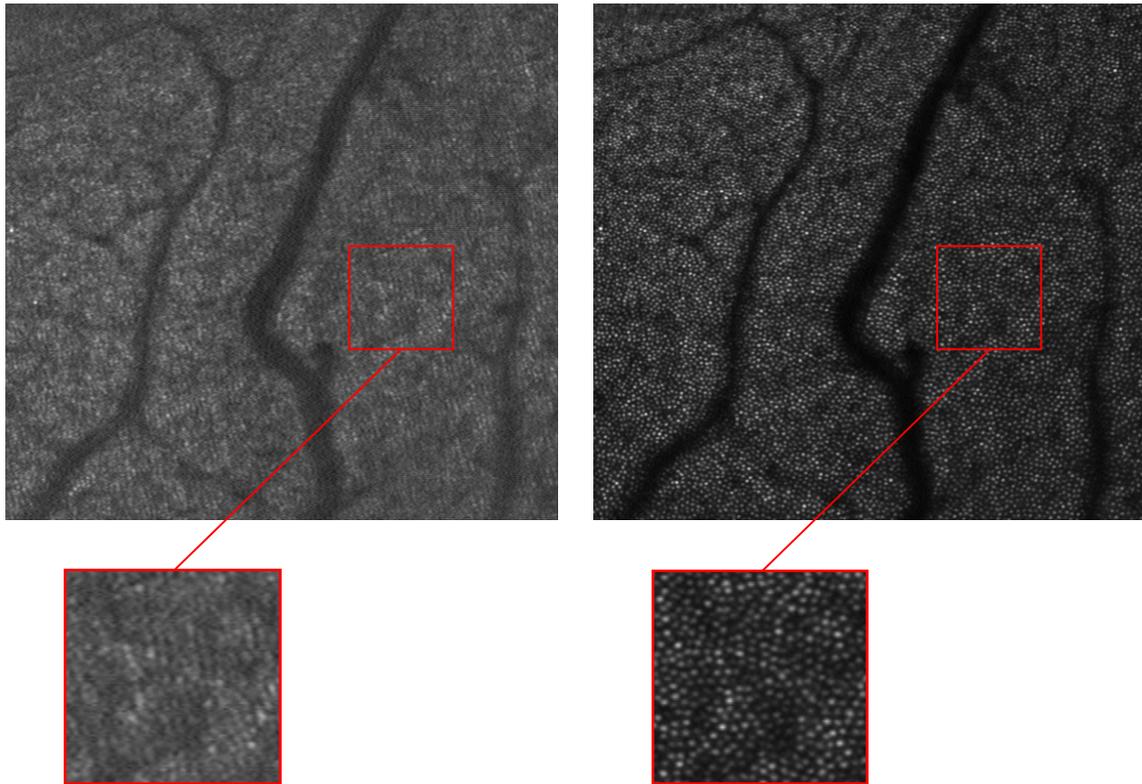
### 1.4.3 State of the art in aberration compensation

#### Sensor based hardware AO (HAO)

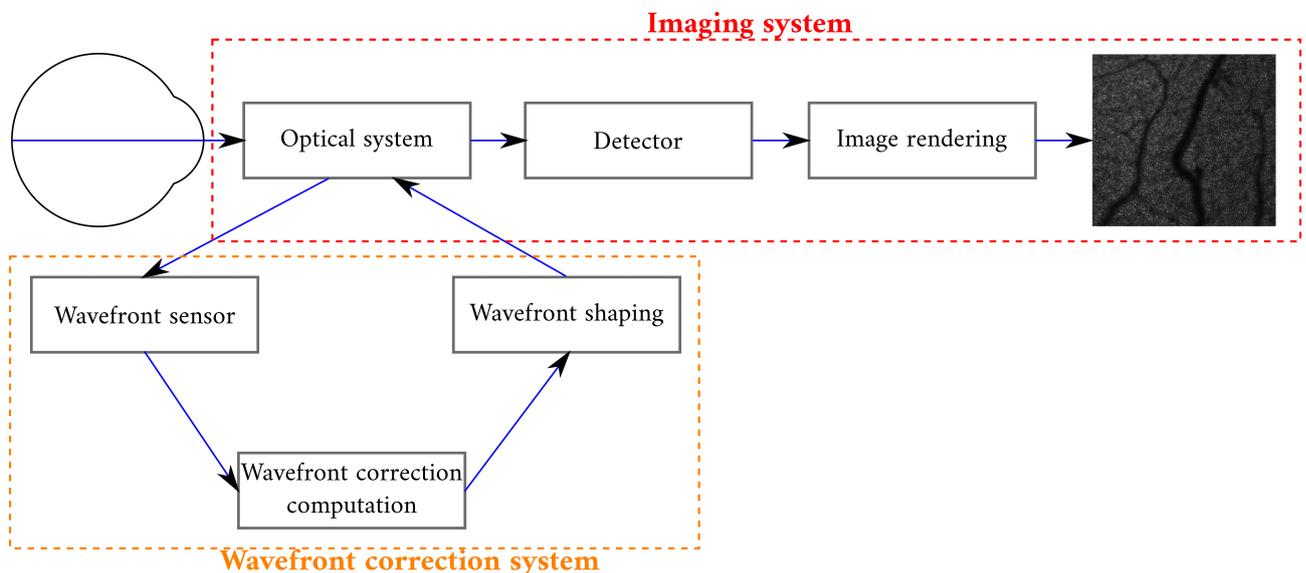
Adaptive optics (AO) is a method created to correct aberrations. Its effects on the PSF of an optical system is well-illustrated on the fig. 1.16. AO has been designed first for astronomy to correct star images from telescope observation, which are spoiled by atmosphere turbulence [74]. Then, it has been adapted to ophthalmology to have more resolved images of retina and photoreceptors. AO has been implemented in ophthalmologic devices like AO flood illumination ophthalmoscope (AO-FIO) [75] or AO scanning laser ophthalmoscope (AO-SLO) [76]. The fig. 1.18 illustrates the effect of adaptive optics on imaging: the photoreceptors are much more visible after the correction by the AO-SLO

The most commonly used AO methods rely on hardware system of aberration detection and correction (fig. 1.19). Some of the light used for imaging is taken to evaluate aberrations. Wavefront distortion is measured by a sensor, and the corresponding correction is sent to a wavefront corrector to change the shape of the wave in the optical system, before detection and image formation.

The Shack-Hartmann (SH) wavefront sensor [77, 78] is the most used wavefront sensor in ophthalmology. It measures wavefront distortions using a network of lenses (fig. 1.20). The principle of SH wavefront sensor will be detailed in section 3.1. Other wavefront sensors have been studied, in order to improve the axial resolution comparing to SH sensor and to handle high spatial frequency aberrations [79–82]. However, they add complexity to an already complex system. Besides, they need more light to be efficient, which means reducing the amount of light in the main path to make images. As far as hardware correctors are concerned, several possibilities exist: liquid crystal phase plates, programmable lenses or deformable mirrors (most used nowadays). Their main aim is to settle local delays to the wave to compensate distortions brought by aberrations.

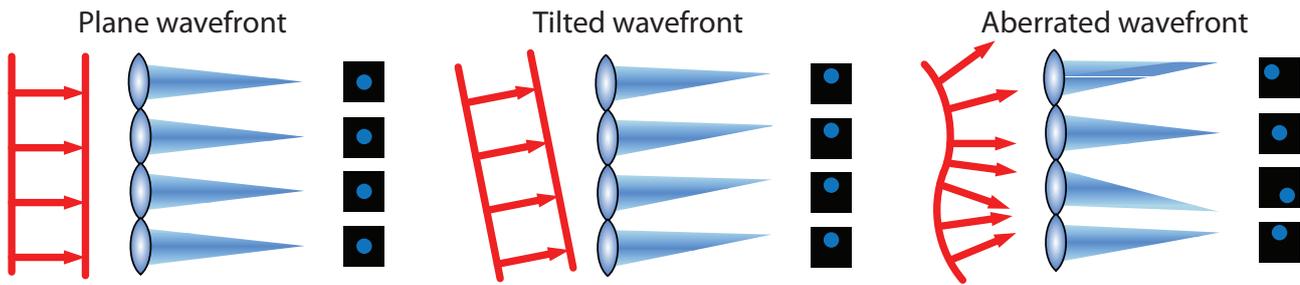


**Figure 1.18** – Images recorded with the PSI AO-SLO device of retina vessels and photoreceptors [72]. Adaptive optics has only been used on the second image. It has significantly improved the resolution of photoreceptors.



**Figure 1.19** – Sensor based hardware AO (HAO) for correction of aberrations [73]. This is divided between an imaging part and a wavefront correction part. The imaging system aims at producing an image while the wavefront correction system changes the shape of the wavefront hardware-wise. The arrows correspond to the path of the signal containing the information.

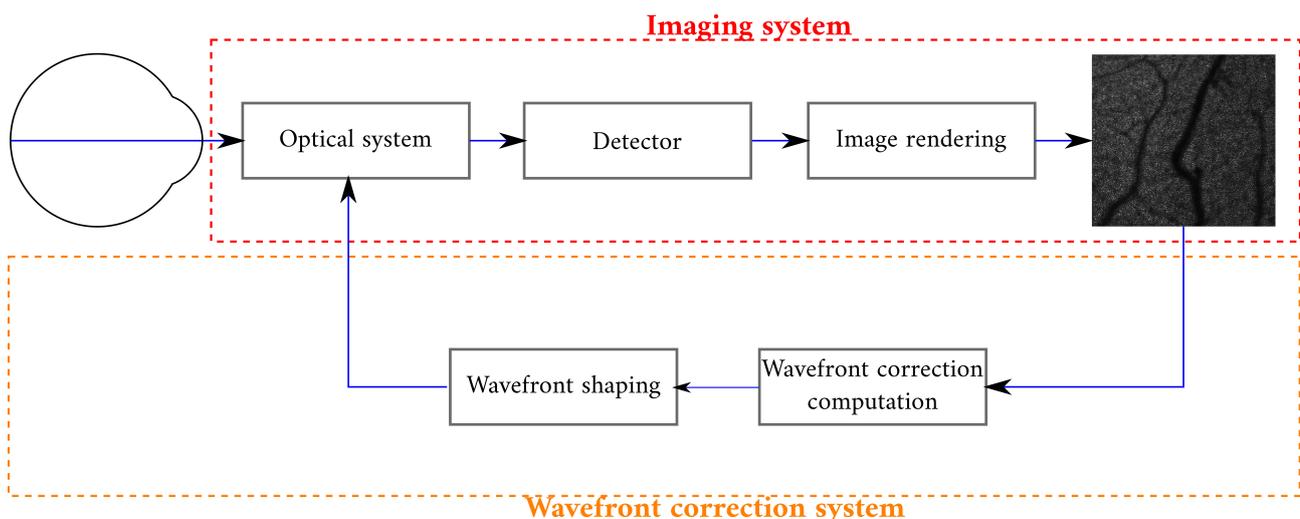
One of the main drawbacks of HAO is the complexity of the setups: they have often a lot of parts to calibrate carefully to make the device work. Besides some of the light is lost for the imaging because of the wavefront measurement. In order to overcome these negative aspects, several teams turned to the development of other AO methods.



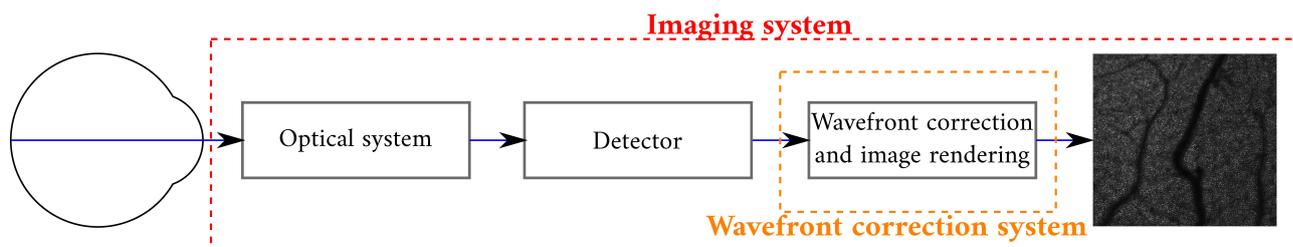
**Figure 1.20** – Principle of the Shack-Hartmann wavefront sensor. It is based on the use of a lens network. When the wavefront is plane, the beams crossing the lenses focus in the middle of each subaperture, on the optical axis. For a tilted wavefront, the focus points are shifted in the same direction. For an aberrated wavefront, the position of each focus point depends on the local distortion of the wavefront. Hence from shifts of focus points within the subapertures, it is possible to go back to the phase.

### Sensorless AO (SAO)

The desire to simplify the setup gave birth to a way of correction without any wavefront sensor: the sensorless AO (see fig. 1.21). This technique estimates the wavefront from the reconstructed image. It uses image-based algorithms, aiming at optimizing image quality calculated from a metric like image intensity [83, 84], sharpness [85] and spatial frequency content [85–87]. The choice of the metric is very important, depending on the imaging technique and the structures of interest. However, the main problem of SAO remains the speed of convergence. It requires a lot of iterations over the process of wavefront estimation. Some help can be given to the algorithm by weighting more some aberration modes, like defocus and astigmatism, because of their importance in human eyes. Nevertheless, it is often not sufficient because of the quick variations of aberrations (dynamic aberrations [67]). This technique remains interesting mainly for reducing the system complexity and increasing the amount of light dedicated to the image rendering.



**Figure 1.21** – Sensorless AO (SAO) for correction of aberrations [73]. It is very similar to the HAO (fig. 1.19), except the wavefront sensor has been removed. Instead, the correction is based on the image. The compensation to add to the wave is calculated and applied thanks to a hardware corrector.



**Figure 1.22** – Computational AO (CAO) for correction of aberrations [73]. The light reflected by the retina is sent through the optical system and captured by a sensor. The correction is calculated from this signal and directly applied to produce a corrected image.

### Computational AO (CAO)

With incoherent imaging, spatial frequencies below the noise level are impossible to restore. Therefore, wavefront correction must be hardware-based. However, in coherent imaging, access to the field allows correction after acquisition.

The increase of computer power over the last few decades has led to new computational AO (CAO) methods (fig. 1.22). It removes the last remain of hardware from SAO. The principle of the correction is to find the best aberration correction filter, directly calculated with the output signal from the detector. Instead of optimizing parameters of the hardware wavefront corrector like in SAO, CAO techniques optimize parameters of the digital filter.

Aberration compensation was demonstrated by measurement of a guide star hologram [88,89]. This method specially applies to objects with point-like scatterers such as photoreceptors in the case of the retina.

Another approach is very similar to the one used in SAO: it consists in an image-based correction. A metric characterizing the image is defined such as sharpness [90–92], image intensity [93,94], spatial frequency content of the image [93,94] or entropy [95], and an iterative process optimizes it. This method is very time-consuming, preventing real-time image rendering.

Reconstruction of digital holograms in subapertures from wavefront measurement with a digital wavefront sensor for line-field OCT has been shown possible [19,96]. It is very similar to a Shack-Hartmann sensor. It estimates the local slopes of the wavefront by computing correlations between images of subdivisions in the Fourier plane. However, this method can only estimate low-order aberrations: to estimate higher orders, more subdivisions are necessary, which decreases the accuracy of the correlations.

### Computational methods using machine learning for phase estimation

Machine learning has skyrocketed for the last years, due to CPU and GPU improvements. With the desire of simplifying setups and acquisition and processing times, optics field has become increasingly interesting by these new methods to solve several problems, including image reconstruction and phase recovery.

As seen before in this manuscript, most of imaging techniques go through a reconstruction stage, which is becoming increasingly fast. A deep neural network (DNN) has been used to reconstruct images from holograms in holography [97]. By training the network with holograms of MNIST database, it makes possible to reconstruct the original images of handwritten numbers or URSAF targets. Optical tomography reconstruction of refractive beads and cells with DNN

has also been investigated [98]. Another team works on the improvement of image quality in holographic microscopy (without image reconstruction) [99, 100].

More and more studies focus on phase estimation and recovery. Several teams have created a digital wavefront sensor (like in SAO or CAO) using neural networks. First emerging works have mimicked wavefront sensors taking on the exact same role, *i.e.* estimating Zernike coefficients from displacements within subapertures [101]. Then, the studies have moved on something more complex. Paine et al. have used convolutional neural networks (CNN) to estimate a first approximation of Zernike coefficients from PSF as a starting point for a more accurate estimation [102]. Nishizaki et al. have gone even further by calculating Zernike coefficients from single intensity images with aberrations for point sources and extended sources [103]. Networks used in those cases are derivatives of Inception v3 [104] which is a classification CNN.

In parallel, a lot of teams have developed machine learning methods to directly estimate the shape of the phase. For incoherent light, the reconstruction has been shown in [105]. The wavefront of the beam is modified with a spatial light modulator (SLM) and then sent through a diffuser. The recorded light is sent as input of a CNN or a machine learning algorithm. The output of the algorithm is the estimation of the SLM phase. For coherent source, phase recovery has been performed from speckle pattern with training and testing on databases like MNIST, ImageNet, Faces-LFW and the Caltech computer vision database [106–110]. The networks used are often derivatives of UNet network [111], first designed for medical image segmentation, and there used for these reconstructions [105, 109, 110].

The main problem of deep learning for image and phase reconstruction is the need of a large amount of images, which can be a major limitation specially in the medical field. Moreover, techniques presented in this part are used for now on simple images (often from MNIST dataset) or with a low resolution. The challenge is to process and improve higher resolution medical images.

## 1.5 Conclusion

In this first chapter, state of the art of imaging methods for depth sectioning and blood flow imaging has been presented. Most of these methods are highly affected by motion or aberrations, spoiling or even preventing image reconstruction. Several correction technologies have been tested. Digital methods are more and more used, taking advantage of major advances in computer science in order to decrease time of acquisition and processing. This thesis goes in the same direction. Digital holography will be presented as our way to image the retina, and motion and aberrations compensation algorithms will be developed, aiming at the best image quality in the shortest time possible.

## 2.1 Holographic imaging

For several years, the research on eye diseases has been strengthened by the opening of a clinical investigation center in the Quinze-Vingts national ophthalmology hospital in Paris. The development of new imaging devices to ease the diagnostic and the study of eye pathologies has been encouraged. In this context, two digital holographic setups have been settled: a digital holographic FF-SSOCT and a laser Doppler holography device.

As far as the OCT is concerned, FF-SSOCT seems to be the most promising technique to decrease acquisition and computation times [27]. Indeed, full fields are acquired by the sensor array of a camera, and the sweep of the laser source ensures a fast depth scanning. The choice of digital holography has been made for several reasons. Holography setups are quite simple: they consist in an interferometric setup. Then, they do not require lots of adjustments and space. The camera records interferences between two waves and a computer processes data. Acquiring these interferograms gives access to the whole field and allows the image to be reconstructed at different reconstruction distances from the camera. Besides, it gives much more possibilities for post-processing. However, interferograms represent a lot of data (in some cases, giga voxels per second to process). Optimizing calculations on GPU is the solution we have chosen to handle these huge amounts. Hologram reconstruction is completely carried by Holovibes<sup>1</sup>.

In the following, both setups are presented, as well as image reconstruction techniques and first processes to enhance image quality.

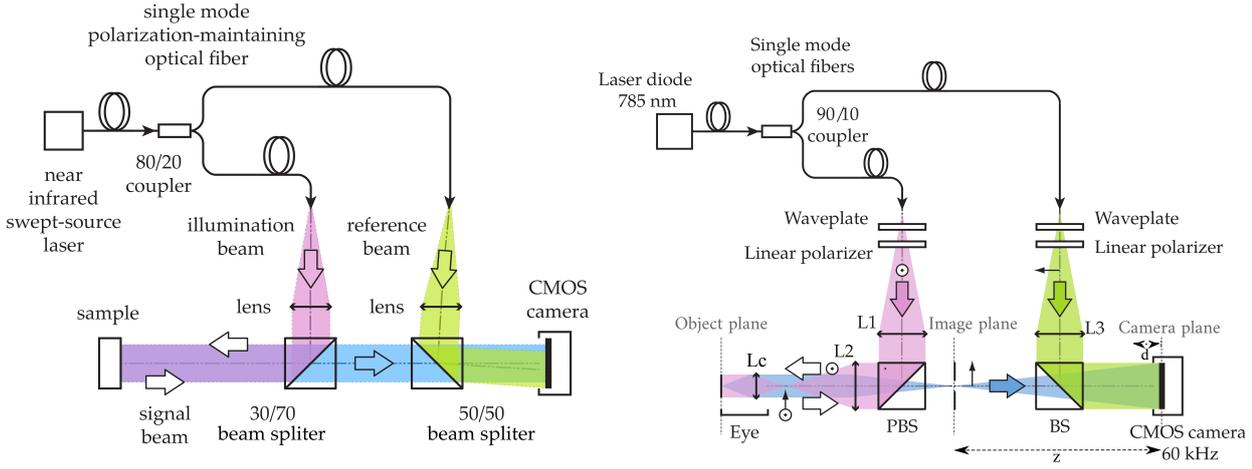
In this chapter, some parts are directly extracted from the article *Motion compensation in digital holography for retinal imaging* attached at the end of the manuscript in appendix B.

### 2.1.1 Setups

The instruments that we used during this project have been settled by Michael Atlan and Léo Puyo at the Quinze-Vingts national ophthalmology hospital. They are very similar Mach-Zehnder interferometers. Both setups are illustrated in fig. 2.1. For the holographic OCT setup

---

<sup>1</sup><http://holovibes.com>



**Figure 2.1** – Schemes of the holographic off-axis OCT (first image) and in-line laser Doppler (second image) setups. In both cases, the reference wave from the source interferes with the object wave which has been reflected on the sample.

(first setup in fig. 2.1), the beam is emitted by a tunable laser whose wavelength varies from  $\lambda_1 = 870 \text{ nm}$  to  $\lambda_2 = 820 \text{ nm}$  with a sweep time  $T = 0.5 \text{ s}$ . For the laser Doppler setup (second setup in fig. 2.1), it is a laser of wavelength  $\lambda = 785 \text{ nm}$ . In the first setup, the camera records  $1024 \times 1024$ -pixel images at a frame rate of  $\omega_S/(2\pi) = 512 \text{ Hz}$  with 16 bit/pixel quantization. In the second setup, the camera records images corresponding to the  $(x, y)$  plane, at a frame rate depending on the speed of the blood flow to observe (this point will be explained in the section 2.3).

### 2.1.2 Acquisition of interferograms

Mach-Zehnder interferometer consists in making interfere two beams from the same laser source. The source is split between reference arm and object arm. The light wave from the object arm is backscattered once the sample is reached and interfere in the camera plane with the beam in the reference arm.

The rest of this section explains the image formation process in more detail. The explanation is based on the second scheme of fig. 2.1. This reasoning can be applied in the same way on the first scheme of fig. 2.1.

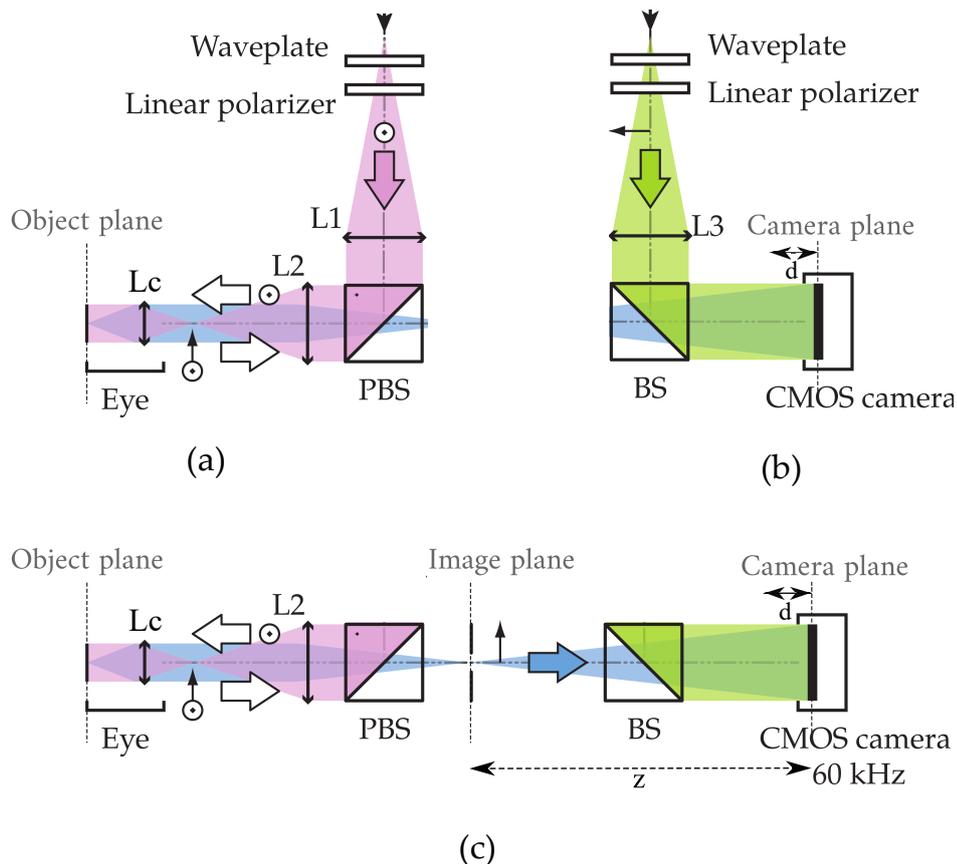
#### Formalism

$E_0(x, y)$  is the reference field from the laser source, such as  $E_0(x, y) = A\delta(x, y)$ , with  $A$  a constant and  $\delta(x, y)$  the dirac distribution, defined as:

$$\begin{cases} \delta(x, y) = (0, 0), & \text{if } (x, y) \neq (0, 0), \\ \iint_{-\infty}^{+\infty} \delta(x', y') dx' dy' = 1 & \text{else.} \end{cases} \quad (2.1)$$

To simplify the explanations, we divide the setup (fig. 2.1) in three parts: the pink optical path, from reference to retina, the blue optical path, from retina to camera, and the green optical path, from reference to camera. The fig. 2.2 zooms in these different parts.

In the pink wave direction of propagation, image focal plane related to the lens L1 is the



**Figure 2.2** – Zooms on the pink optical path from reference to retina (a), green optical path from reference to camera (b), and blue optical path from retina to camera (c). The pink optical path includes L1, L2 and Lc lenses. The green optical path includes L3 lens. The blue optical path includes Lc and L2 lenses.

same as object focal plane of the lens L2. Image focal plane related to the lens L2 is the same as object focal plane of the lens Lc (cornea + crystalline lens). The retina is in the object plane, which is the same as image focal plane of the lens Lc.

In the blue wave direction of propagation,  $z$  is the distance between camera plane and image plane, defined as the image focal plane of the lens L2.

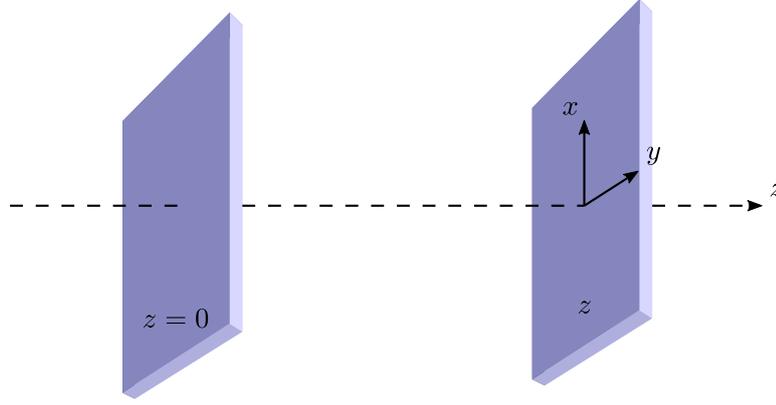
In the green wave direction of propagation,  $d$  is the distance between camera plane and the image focal plane of the lens L3.

In the following, fields are expressed in 2 dimensions ( $x$  and  $y$ ). The 2-dimension Fourier transform is noted  $FT2D$ . The symbol  $\propto$  means "proportional to" and  $*$  stands for the convolution. We consider the retina as a plane locally. We will use formulas of field propagation between planes recalled in the next paragraph, to describe the light path in our system.

### Wave propagation formulas

This part presents the formulas used to compute the propagation of the waves, deriving from the theory of Fourier optics [112].

Let us denote  $M$  and  $N$  the number of pixels of lateral dimensions ( $x$  and  $y$  respectively),  $d$  the pixel pitch in meters and  $\lambda$  the wavelength. The first plane is located at  $z = 0$  and the second one at  $z$  (fig. 2.3).  $E_0(x, y)$  is the field in the plane  $z = 0$  and  $E_z(x, y)$  is the field in the plane  $z$ .



**Figure 2.3** – Schemes of two planes spaced by  $z$ .

For  $z \geq \frac{Nd^2}{\lambda}$ , the Fourier-Fresnel formula (from Huygens-Fresnel theorem) is used to propagate waves from plane  $z = 0$  to plane  $z$ :

$$E_z(x, y) = \frac{e^{2i\pi z/\lambda}}{i\lambda z} e^{i\pi(x^2+y^2)z/(\lambda z)} FT2D\{E_0(x', y') e^{i\pi(x'^2+y'^2)z/(\lambda z)}\} \left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right). \quad (2.2)$$

For  $z \leq \frac{Nd^2}{\lambda}$ , the angular spectrum formula (from Huygens-Fresnel theorem) is used to propagate waves from plane  $z = 0$  to plane  $z$ :

$$E_z(x, y) \propto FT2D^{-1}\{FT2D\{E_0(x, y)\} \times K_z(x', y')\}, \quad (2.3)$$

where  $K_z(x, y)$  is the kernel defined by:

$$K_z(x, y) = \exp\left[\frac{2i\pi z}{\lambda} \sqrt{1 - \left(\frac{x\lambda}{Md}\right)^2 - \left(\frac{y\lambda}{Nd}\right)^2}\right]. \quad (2.4)$$

The Fraunhofer diffraction to infinity is:

$$E_z(x, y) = \frac{e^{2i\pi z/\lambda}}{i\lambda z} FT2D\{E_0(x', y')\} \left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right). \quad (2.5)$$

In order to explain the components of the image recorded by the camera, we explicit field propagation through object arm (pink and blue paths) and the reference arm (green path) in the next parts.

### **Pink optical path, from reference to retina**

To get the field  $E_1(x, y)$  in the image focal plane of L1 lens, Fraunhofer diffraction (proportional to 2-dimension Fourier transform) is computed from  $E_0(x, y)$ , as for  $E_2(x, y)$  in the image focal plane of L2 lens, from  $E_1(x, y)$ :

$$\begin{cases} E_1(x, y) \propto FT2D\{E_0(x', y')\}(x, y), \\ E_2(x, y) \propto FT2D\{E_1(x', y')\}(x, y). \end{cases} \quad (2.6)$$

Then,  $E_2(x, y)$  is the mirror image of  $E_0(x, y)$ :  $E_2(x, y) \propto E_0(-x, -y)$ . In the same way, from  $E_2(x, y)$ ,  $E_c(x, y)$  in the image focal plane of the cornea can be expressed as:

$$E_c(x, y) \propto FT2D\{E_0(-x', -y')\}(x, y) = FT2D\{A\delta(-x', -y')\}(x, y) = A. \quad (2.7)$$

Once reflected on the retina,  $E_r(x, y)$  in retina plane becomes:

$$E_r(x, y) = E_c(x, y) \times O(x, y)e^{i\phi(x, y)} \propto O(x, y)e^{i\phi(x, y)}. \quad (2.8)$$

$O(x, y)e^{i\phi(x, y)}$  is the information to extract to form holographic images of the retina. From here, the retina is considered as a new punctual source. Aberrations created by the cornea in the pink path are neglected. Therefore, we only have to express aberrations in the blue path.

### Aberrations

We consider that aberrations  $\phi_a(x, y)$  are created in the infinity plane of the retina whose field is  $E_{r\infty}(x, y)$ . They are added to the phase of the wave in this plane. Thus, we use the Fraunhofer diffraction to calculate the field:

$$E_{r\infty}(x, y) \propto FT2D\{E_r(x', y')\}(x, y)e^{i\phi_a(x, y)}. \quad (2.9)$$

Then, the reverse propagation is carried out from  $E_{r\infty}(x, y)$  to get the aberrated field in the retina plane:

$$E_{ra}(x, y) \propto O(x, y)e^{i\phi(x, y)} * FT2D^{-1}\{e^{i\phi_a(x', y')}\}(x, y). \quad (2.10)$$

### Blue optical path, from retina to camera

In the image plane (corresponding to the image focal plane of the lens L2), by applying the same reasoning as before:

$$E_{c2}(x, y) \propto E_{ra}(-x, -y). \quad (2.11)$$

To obtain the field  $E_z(x, y)$  in the camera plane, angular spectrum or Fourier-Fresnel propagation must be used, depending on the distance between camera and image planes. In the laser Doppler holography setup, angular spectrum method is used to propagate  $E_{c2}(x, y)$ :

$$E_z(x, y) \propto FT2D\left\{FT2D^{-1}\{E_{ra}(-x, -y)\} \times K_z(x', y')\right\}, \quad (2.12)$$

where  $K_z(x, y)$  is a kernel of propagation, defined previously in **Wave propagation formulas** subsection. In the case of OCT setup, Fourier-Fresnel propagation is used, but that does not change the overall purpose of these explanations.

### Green optical path, from reference to camera

Using Fraunhofer diffraction and Fourier-Fresnel propagation, we get in the camera plane:

$$E_{0d}(x, y) \propto e^{-i\pi\frac{(x^2+y^2)}{\lambda d}} E_0(-x, -y) * FT2D^{-1}\left\{e^{-i\pi\frac{(x'^2+y'^2)}{\lambda d}}\right\}\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right) = 1. \quad (2.13)$$

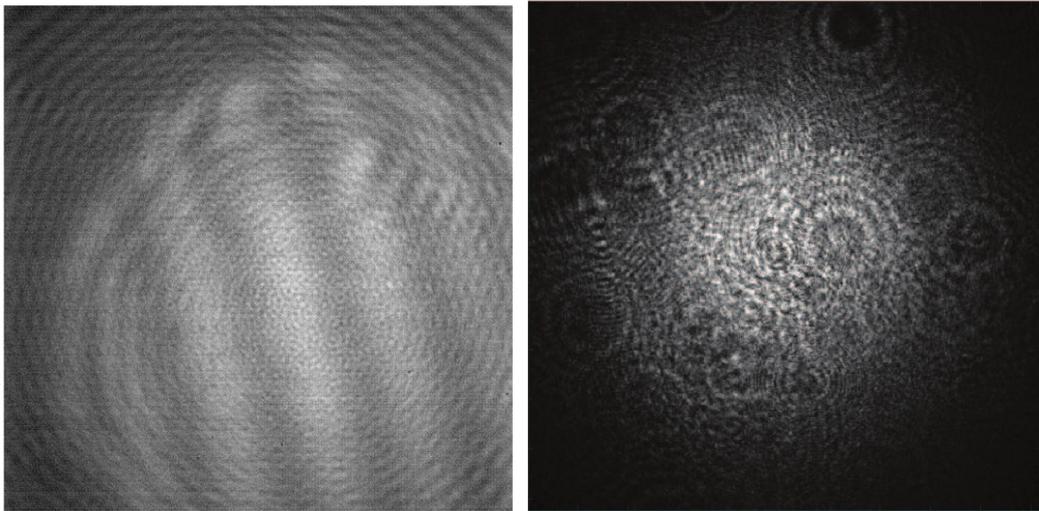
The field from the reference projected in the camera plane does not depend on  $x$  and  $y$ .

### Recorded intensity

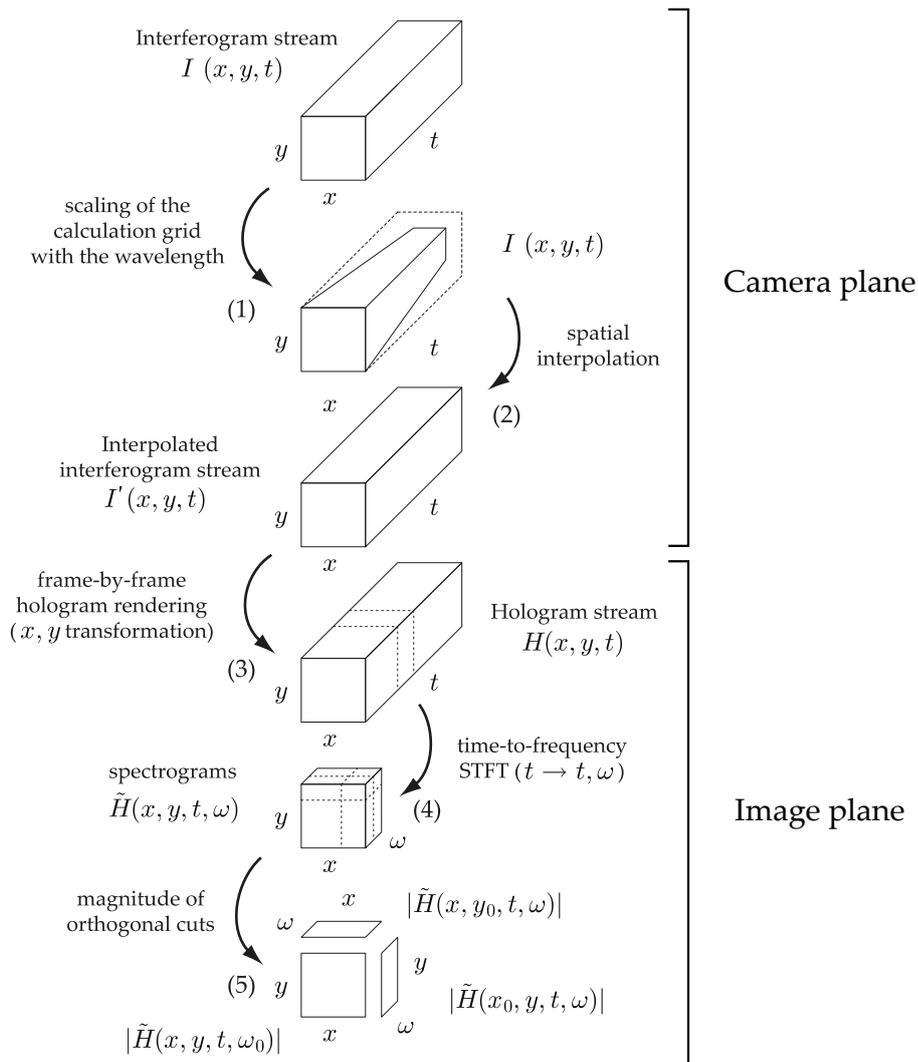
The camera records the intensity  $I(x, y)$  of the sum of the fields, called interferogram:

$$\begin{aligned} I(x, y) &= |E_z(x, y) + E_{0d}|^2 \\ &= |E_z(x, y)|^2 + |E_{0d}|^2 + E_z(x, y)E_{0d}^* + E_z^*(x, y)E_{0d}. \end{aligned} \quad (2.14)$$

Laser Doppler and OCT images can be created from the phase difference between both optical fields, noted  $\phi$  (see section 2.1.3, section 2.2 and section 2.3). The fig. 2.4 shows recorded interferograms from OCT and laser Doppler setups. The OCT has an off-axis configuration: the beam from the reference arm is slightly deviated from the object beam. It is well-visible on the first image of fig. 2.4 where the brighter beam (reference) has not the same center as the other ray.



**Figure 2.4** – Interferograms of off-axis FF-SSOCT (left) and laser Doppler imaging (right). The off-axis configuration is well-visible with the delimitation of the smallest diameter beam on the first image.



**Figure 2.5** – Sketch of the process of image propagation. In the case of the swept-source, pixels of the interferograms  $I(x, y, t)$  have not the same size for each  $t$  (1). Thus, rescaling is performed (2). The data  $I(x, y, t)$  in the camera plane are propagated in the image plane and become holograms  $H(x, y, t)$  (3). Frequency dependency is expressed with a time Fourier transform on  $H(x, y, t)$  (4). Eventually, amplitude of the block is taken to obtain images (5).

### 2.1.3 Digital processing of optically-acquired interferograms

From the recorded interferogram  $I(x, y)$ , we perform several processes to obtain final images in OCT or laser Doppler holography. The different processing steps are detailed in the following, and illustrated by the fig. 2.5.

#### Interferogram rescaling with wavelength

In order to avoid the distortion of the signal, the impact of the sweep in OCT needs to be considered. In fact, the size of the pixels in the image plane depends on the wavelength of the beam and of the distance between camera and image planes [113, 114]:

$$d' = \frac{\lambda z}{Nd}, \quad (2.15)$$

where  $d$  and  $d'$  are the lateral size of a pixel from camera plane and from image plane, respectively,  $z$  is the hologram distance of reconstruction (distance between both planes) and  $N$  is the number of pixels in one lateral dimension. Because of the sweep of the source, the pixels of the image plane shrink with wavelength (see arrow (1) of fig. 2.5). To circumvent lateral field variation with wavelength, each interferogram is resampled by linear interpolation of the calculation grid with a different pitch (see arrow (2) of fig. 2.5). The rescaled interferogram is:

$$I'(x, y, t) = I(x\lambda/\lambda_1, y\lambda/\lambda_1, t), \quad (2.16)$$

where  $\lambda$  is the current wavelength. In the case of laser Doppler imaging, the optical wavelength  $\lambda$  is kept constant, hence the interferogram does not need to be rescaled:  $I' \equiv I$ .

### Hologram computation

The propagation of the interferogram from camera to image plane, which is equivalent to the retina plane, is carried out by an angular spectrum propagation integral for laser Doppler holography setup (see arrow (3) of fig. 2.5), which gives the hologram  $H(x, y)$ :

$$\begin{aligned} H(x, y) &\propto FT2D^{-1} \{FT2D\{I'(-x, -y)\} \times K_z(x', y')\} \\ &= H_{const}(x, y) + H_{obj}(x, y) + H_{conj}(x, y), \end{aligned} \quad (2.17)$$

where  $H_{const}(x, y)$  is the constant part based on  $|E_z(x, y)|^2 + |E_{0d}|^2$ ,  $H_{obj}(x, y)$  is the interest part composed from  $E_z(x, y)E_{0d}^*$  and  $H_{conj}(x, y)$  is made from the conjugate  $E_z^*(x, y)E_{0d}$ . As it has been noticed in section 2.1.2, for OCT setup, the Fourier-Fresnel transform is applied to propagate fields between camera and image plane, instead of the angular spectrum method. Though that does not change the decomposition of  $H(x, y)$  in three parts.

In the case of swept-source OCT, its off-axis configuration creates a spatial separation between the constant term and the cross-beating interferometric contributions [115] (fig. 2.6). Then, parts of interest can be spatially filtered.

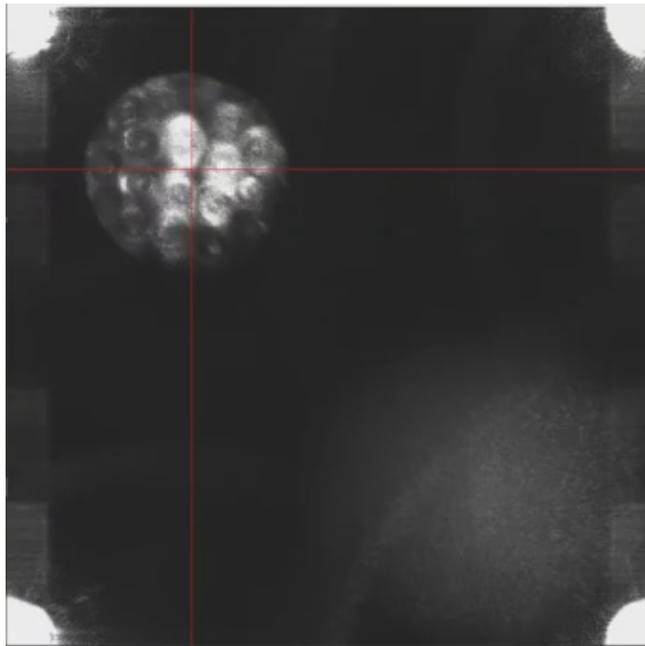
For the laser Doppler setup, constant part is removed by filtering low frequencies (section 2.3). Conjugate part  $H_{conj}(x, y)$  is out of focus in the reconstruction plane. Thus, it does not affect significantly image quality. The object part  $H_{obj}(x, y)$  is:

$$H_{obj}(x, y) \propto E_{ra}(x, y) \propto O(x, y)e^{i\phi(x, y)} * FT2D^{-1}\{e^{i\phi_a(x', y')}\}(x, y). \quad (2.18)$$

Therefore, we get the part of interest  $O(x, y)e^{i\phi(x, y)}$  which constitutes the object, and the aberration part  $FT2D^{-1}\{e^{i\phi_a(x', y')}\}(x, y)$ .

### Image formation

The phase  $\phi$ , which is also the phase difference between object and reference fields, contains different information about the sample. In swept-source OCT, the instantaneous beating frequency  $\partial\phi/\partial t = \omega$  scales up linearly with axial depth  $z$  [13], whereas in laser Doppler, it describes local velocities of the scatterers [48]. In both cases, frequency must be expressed to get to the desired information (depth for OCT and velocities for laser Doppler).



**Figure 2.6** – En-face image of 1.5 mm glass beads by off-axis FF-SSOCT. The off-axis configuration allows a spatial separation between the object part, its conjugate and the constant part of the interference signal. The image of interest is on the top left side of the image. It is the image of the beads. Its conjugate is on the bottom right side of the image. The reconstruction is done in the image plane. The conjugate part is not well-focused on this plane, which causes its blurring. The constant part is located at the four corners of the image.

By acquiring stream of interferograms in time, we can express  $I(x, y)$  as a function of time  $I(x, y, t)$ . Then, interferograms are transformed into holograms  $H(x, y, t)$ . Short-time Fourier transform (STFT) (see arrow (4) of fig. 2.5) is performed to get spectrograms  $\tilde{H}(x, y, t, \omega)$  depending on frequency:

$$\tilde{H}(x, y, t, \omega) = \int H(x, y, \tau) g_T(t - \tau) e^{-i\omega\tau} d\tau, \quad (2.19)$$

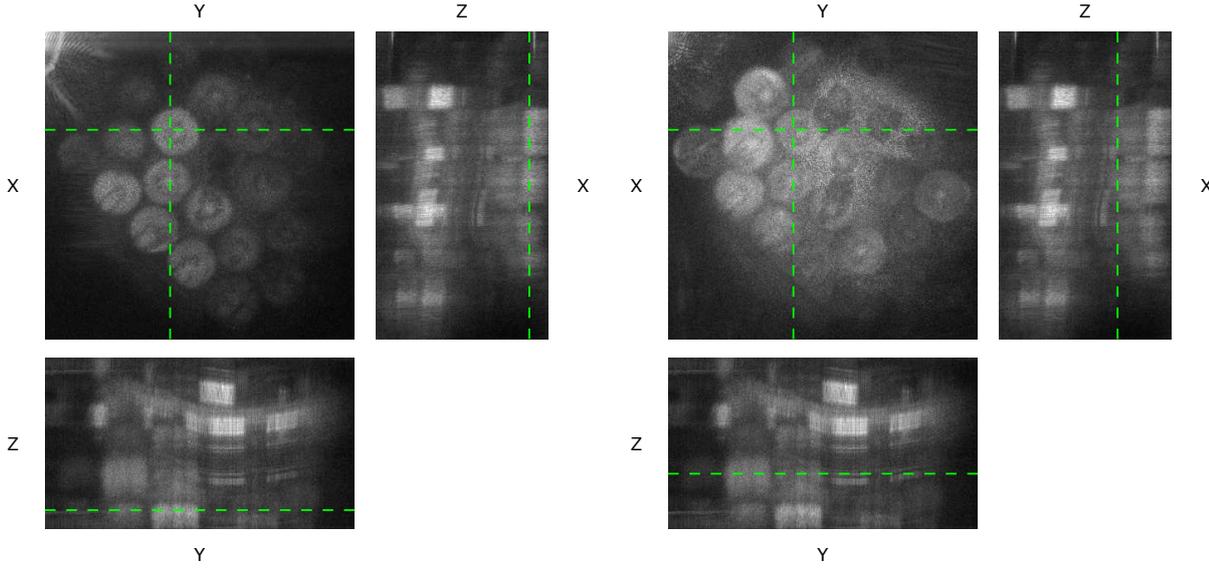
where  $g_T(t)$  is a time gate of width  $T$  at time  $t$ . Then, the intensity image  $|\tilde{H}|^2$  is formed (see arrow (5) of fig. 2.5).

Different information is extracted from eq. (2.19), depending on whether OCT or laser Doppler holography setup is used. The next sections develop the final steps to obtain images with these two technologies.

## 2.2 Optical coherence tomography

We form OCT images by calculating eq. (2.19) with a time gate  $g_T$  of 256 points ( $T = 0.5$  s). Taking  $|\tilde{H}(x, y, t, \omega_0)|^2$  at a precise frequency  $\omega_0$  provides an en-face image at the corresponding depth. For given  $x$  or  $y$ , we obtain depth profiles. The fig. 2.7 shows typical images provided by our setup.

After being able to acquire those images, we have studied properties of our OCT system. Experimental tests were fundamental to determine the performance of the OCT, and to verify lateral and axial resolution, as well as the axial range. To carry out these experiments, we used numerous volume samples (glass beads of different sizes wrapped in tape, resolution targets,



**Figure 2.7** – OCT images of glass beads of 1.5 mm of diameter at two different depths.  $(x, y)$  represents the en-face plane, and  $z$  is the depth. The green dashed lines indicate the cuts of the planes. Bright objects on images show high reflection, which means they belong to the corresponding  $z$  plane.

tangerine quarters...) which allowed us to evaluate the improvements to be made.

### 2.2.1 Calculation of the axial range

In swept-source OCT, the phase  $\phi$  between object and reference waves can be defined as:

$$\phi(t) = \frac{\beta z}{c}t, \quad (2.20)$$

where  $z$  is the detuning pathlength,  $c$  the light speed and  $\beta$  the constant sweep speed defined as  $\beta = (\omega_2 - \omega_1)/T$ . It follows that the instantaneous beating frequency between object and reference waves scales linearly with the depth [13]:

$$\omega = \frac{\partial \phi}{\partial t} = \frac{\beta z}{c}. \quad (2.21)$$

Hence the short-time Fourier transform is necessary to get this frequency dependency from  $H(x, y, t)$ , which means also the depth of imaging. With the Shannon theorem, we know the bandwidth is bounded by the sampling frequency  $\omega_s$ , which is equal to  $\omega_s = \frac{2\beta\Delta z}{c}$  thanks to eq. (2.21), where  $\Delta z$  is the axial range. Then, it follows:

$$\Delta z = \frac{\omega_s c}{2\beta} = \frac{\omega_s c T}{2(\omega_2 - \omega_1)} = \frac{\omega_s c T}{2 \times 2\pi c \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right)} = \frac{\omega_s}{2\omega_{sweep}} \frac{\lambda_1 \lambda_2}{\Delta \lambda}, \quad (2.22)$$

with  $\omega_{sweep} = 2\pi/T$ . We can simplify this equation by writing:

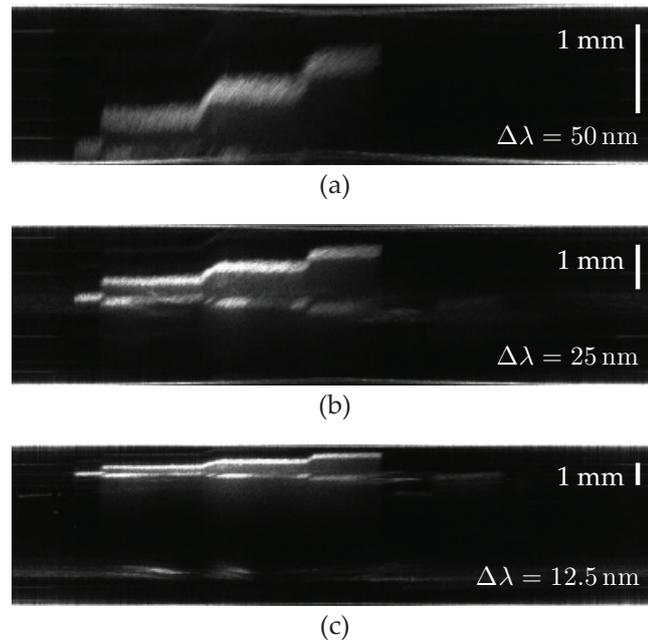
$$\Delta z = \frac{N}{2} \Lambda, \quad (2.23)$$

where  $\Lambda = \lambda_1 \lambda_2 / \Delta\lambda$ , and  $N = \omega_s / \omega_{sweep}$  is the number of interferograms measured during one frequency sweep. For tunable lasers with non-Gaussian output spectra [116], the theoretically limiting axial resolution is the round-trip coherence length  $\bar{\lambda}^2 / \Delta\lambda = 14.3 \mu\text{m}$ , where  $\bar{\lambda} = (\lambda_1 + \lambda_2) / 2$  is the central frequency of the sweep. This round-trip coherence length is about the same length as the axial pitch:  $\bar{\lambda}^2 / \Delta\lambda \simeq \lambda_1 \lambda_2 / \Delta\lambda$ .

Thus, eq. (2.23) can be written such as:

$$\Delta z = \frac{N}{2} \Lambda \approx \frac{N}{2} \frac{\bar{\lambda}^2}{\Delta\lambda}. \quad (2.24)$$

In our case,  $\bar{\lambda} = 845 \text{ nm}$  and  $N = 256$ . Therefore, we have a direct relation between the axial range and the sweep range. We can verify this relation by measuring the axial field of view for different wavelength tuning ranges. The fig. 2.8 represents the axial field of view of a sample for  $\Delta\lambda = 50 \text{ nm}$ ,  $\Delta\lambda = 25 \text{ nm}$  and  $\Delta\lambda = 12.5 \text{ nm}$ . We can observe the axial fields of view are close to the theoretical values calculated from eq. (2.24), namely  $1.8 \text{ mm}$ ,  $3.7 \text{ mm}$  and  $7.3 \text{ mm}$  respectively.



**Figure 2.8** – Axial range  $\Delta z$  for a wavelength tuning range  $\Delta\lambda$  of  $50 \text{ nm}$  (a),  $25 \text{ nm}$  (b), and  $12.5 \text{ nm}$  (c). Each scale is  $\sim 313 \mu\text{m}$  height, which was measured with a numerical calliper and confirmed by full-field OCT. The axial field of view  $\Delta z$  of  $1.8 \text{ mm}$  (a),  $3.7 \text{ mm}$  (b), and  $7.3 \text{ mm}$  (c) is in agreement with the theoretical value from Eq. 2.24 for  $N = 256$ . Axial scalebar :  $1 \text{ mm}$ .

### 2.2.2 Image resolution

By calculating the axial resolution of our device, we obtained :

$$\delta z = \frac{\Delta z}{N} = \frac{\lambda_1 \lambda_2}{2\Delta\lambda} = 7.5 \mu\text{m}, \quad (2.25)$$

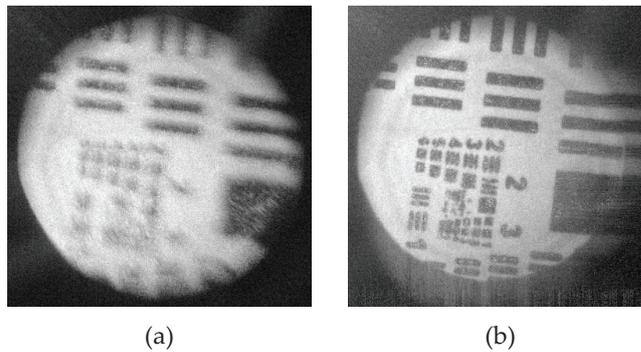
where  $N = 256$  is the number of acquisitions in time for one sweep,  $\Delta z = N \frac{\lambda_1 \lambda_2}{2\Delta\lambda}$  is the axial range, and  $\Delta\lambda = \lambda_1 - \lambda_2 = 50 \text{ nm}$  is the sweep range. We were able to confirm this value experimentally by constructing a calibrated sample. The sample used for this calibration consists in three scales of around  $313 \mu\text{m}$  each of tape layers. They have been visualized with the OCT (fig. 2.8). Each step of the fig. 2.8 (a) consists of about 42 pixels, which corresponds to the resolution calculated in eq. (2.25). Likewise, the axial resolution corresponds to the number of pixels for fig. 2.8 (b) and fig. 2.8 (c), with other values of the sweep range.

In digital holography, lateral resolution is limited by optical aberrations created by the eye. By compensating these aberrations, we could bring that resolution close to the diffraction limit.

As a comparison, one can currently find commercial devices with an axial resolution of  $4 \mu\text{m}$  and a lateral resolution of  $3 \mu\text{m}$ .

### 2.2.3 Improvement of image resolution with rescaling

One of the first improvement we have realized has been explained in the section 2.1.3, namely the rescaling of interferograms. It allows us a significant gain, specially in lateral resolution. The fig. 2.9 represents en-face view of a target we have imaged with the OCT. The rescaling improves lateral resolution of the image. This correction has also an impact on the axial resolution. The fig. 2.10 shows that the rescaling allows even a better separation between depth planes: in image (a), beads on the right seem almost at the same elevation than beads on the left, because the depth accuracy is low, while in image (b), right beads clearly belong to a different layer. The fig. 2.11 also shows an improvement in depth sectioning. The rescaling is necessary to obtain a much better image quality. This functionality has been implemented in real-time on Holovibes.

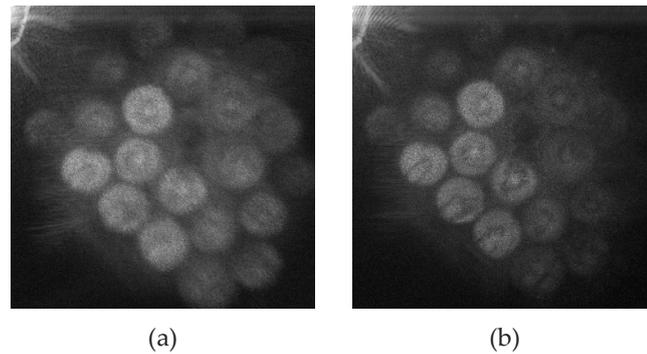


**Figure 2.9** – En-face image of a USAF resolution target by SSOC, before (a) and after (b) rescaling of the acquisitions.

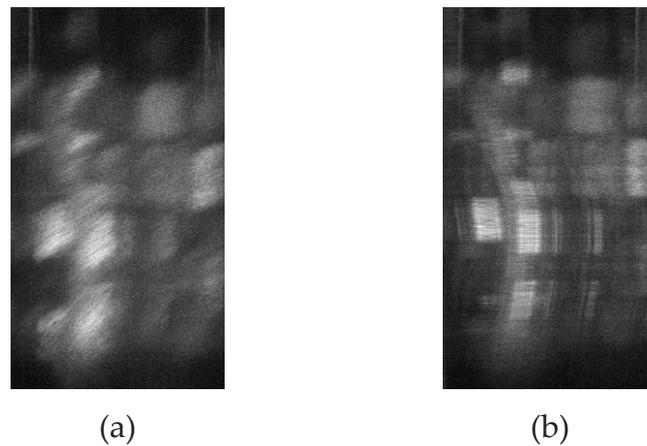
### 2.2.4 Conclusion on holographic OCT setup

We have built an holographic OCT setup, verifying its main parameters and improving its resolution. Several tests on phantoms (glass beads of different sizes, mandarin orange peel, tap layers, plastic objects...) have been performed.

OCT rendering time has improved a lot during the past decades. With devices like the Spectralis OCT, from Heidelberg Engineering (see fig. 1.1 and fig. 1.4), 2D-images (B-scans)



**Figure 2.10** – En-face image of glass beads of 1.5 mm of diameter, before (a) and after (b) rescaling of the acquisitions.



**Figure 2.11** – Profile image of glass beads of 1.5 mm of diameter, before (a) and after (b) rescaling of the acquisitions.

are rendered in real-time. Conversely, holographic swept-source OCT has the potential to perform 3D imaging in real-time.

With a rendering of 10 Giga voxels per second, Hologibes makes possible the visualization of the evolution of samples in real-time. Improvements like the rescaling or the motion compensation techniques we will see further are implemented gradually, making the software evolving. Real-time movies of several samples observed with our OCT and Hologibes are available on our Youtube channel Digital Holography<sup>2</sup>.

However, for now, we have not been able to observe a living eye yet for several reasons. The first reason is the sensitivity of our device to motion, that we have improved with motion compensation techniques as we will explain in the section 2.4. The second is aberrations, also impacting too much our system. The third and most important issue is the signal acquisition time. In order to perform accurate holographic swept-source OCT, the total wavelength sweep time set to record one volume should be lower than about 5 ms. This time is 100 times longer in our case, which makes rendering impossible. We have concentrated our work on motion and aberrations compensation, that we have applied on the holographic device which was already able to image human eye (the laser Doppler holography), which is the reason why the OCT has not got other improvements.

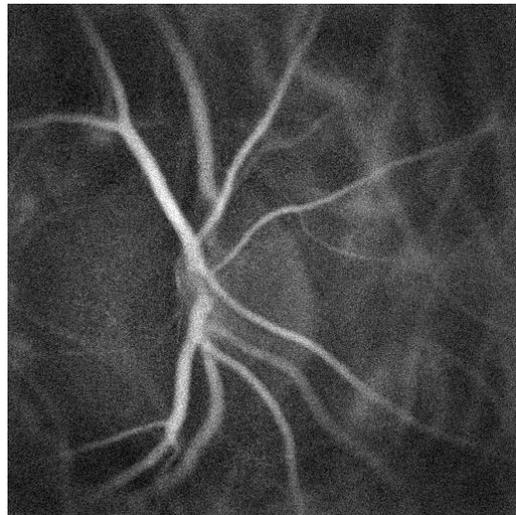
<sup>2</sup>[https://www.youtube.com/channel/UC\\_DamX84B2Y375nANrD\\_Tjw](https://www.youtube.com/channel/UC_DamX84B2Y375nANrD_Tjw)

## 2.3 Laser Doppler imaging

In laser Doppler case,  $|\tilde{H}(x, y, t, \omega_0)|^2$  obtained with eq. (2.19) represents the image of the fluctuations into blood vessels whose speed corresponds to the  $\omega_0$  frequency. However, the amount of information contained in this spectrogram alone is not sufficient to reconstruct an image with a satisfying signal-to-noise ratio. Thus, spectrograms need to be integrated over frequency. This zero-order moment is the Doppler image (see example in fig. 2.12) corresponding to the acquisition  $I(x, y, t)$ :

$$M_0(x, y, t) = \int_{\omega_1}^{\omega_2} |\tilde{H}(x, y, t, \omega)|^2 d\omega, \quad (2.26)$$

where  $[\omega_1, \omega_2]$  is the frequency range for the integration of the spectrum, with  $\omega_1 > 0$ . The choice of  $\omega_1$  and  $\omega_2$  is very important. First, as we have seen in section 2.1.3, this frequency range selection makes possible the filtering of the constant part of the interferograms. Then, because of the high speed of blood cells circulation, the integration is made on very high frequencies of the spectrum. Besides, taking high frequencies allows also a motion filtering: the motion provoked by respiration, heart beats and other often constitutes the lowest frequency part of the spectrum. However, in some cases, the frequency bandwidth of the motion overlaps some lower speed vessels, which spoils their visualization. Finally, the blood in the vessels visible on Doppler holography images has a flow velocity that depends directly on this frequency interval. If the integration is done on lower frequencies, the visible vessels are the ones with the lowest blood flow velocity. If the integration is done on the highest frequencies, the apparent vessels are the ones with the highest blood flow velocity.

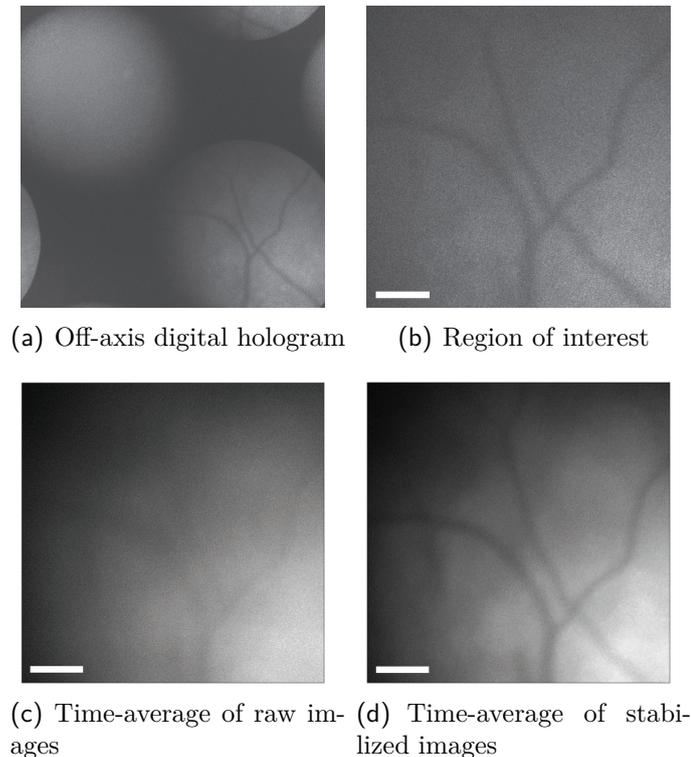


**Figure 2.12** – Image of retina vessels by laser Doppler holography. This corresponds to the zero-order moment of the spectrum.

The maximum frequency  $\omega_m$  of this interval of integration is related to the sample frequency  $\omega_s$  by the Shannon-Nyquist theorem:  $\omega_s$  must be at least twice  $\omega_m$ . For example, if the frequency of acquisition of the camera is equal to  $60 \text{ kHz}$ ,  $\omega_m$  will be inferior or equal to  $30 \text{ kHz}$ , which means only vessels whose blood flow is lower than  $30 \text{ kHz}$  will be revealed by laser Doppler holography. Hence, the faster the camera, the faster the observable vessels are.

## 2.4 Motion compensation

As we said previously, movement is a significant barrier to good imaging. It is responsible for blurring images and spoils their visualization. It has also heavier consequences on holograms. For laser Doppler holography, motion frequency can be confused with blood cells circulation. For OCT, signals can be misencoded by axial shifts. For these reasons, work on motion compensation is required.



**Figure 2.13** – Retinal image stabilization by lateral  $(x, y)$  motion compensation of intensity holograms. The white scale bars indicate  $200 \mu\text{m}$ .

### 2.4.1 Lateral movement compensation

To compensate lateral motion (in  $(x, y)$  plane), we use a cross-correlation-based stabilization algorithm. Even though feature-based methods have proved their efficiency in classical eye-fundus imaging [56–58], it is difficult to use them on our images because of the noise which can be significant. Besides, the movement we are trying to compensate is a translation, so a rigid registration method like correlation-based algorithms is sufficient. In the following, the algorithm of lateral motion compensation is detailed, and the results obtained on laser Doppler holographic images are presented.

#### Algorithm

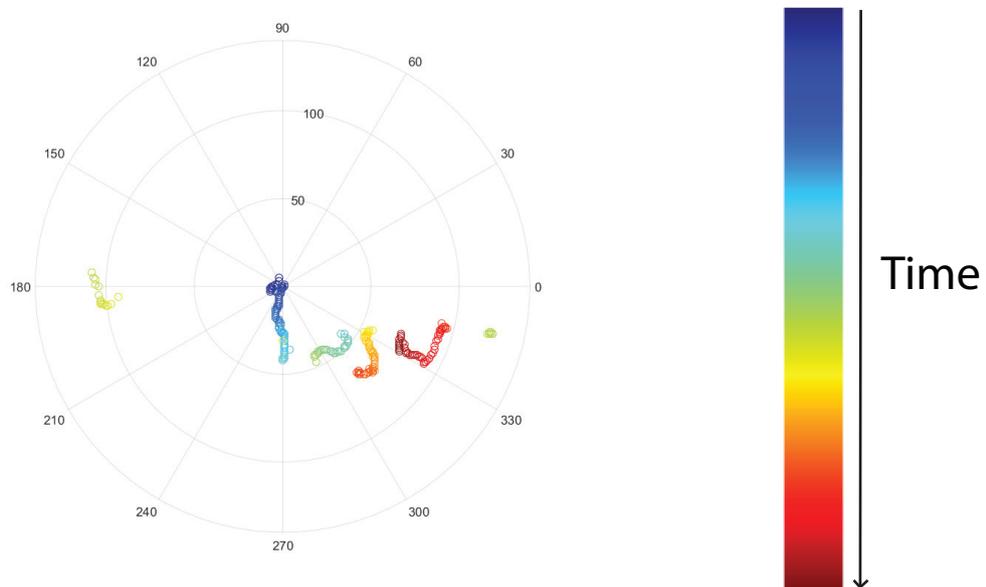
A reference image  $M_{01}(x, y, t)$  is made by integrating the spectrogram on 10 frequencies, and compared to a moving average  $M_{02}(x, y, t + \tau)$  of 10 frequencies in the neighborhood of the current image. The quantitative comparison is performed with a normalized and centered cross-correlation [117]. The correlation matrix  $\gamma$  takes the form:

$$\gamma(u, v) = \frac{\sum_{(x,y) \in D_{u,v}} M'_{01}(x, y, u, v) M'_{02}(x, y, u, v)}{\sqrt{\sum_{(x,y) \in D_{u,v}} M'_{01}(x, y, u, v)^2 M'_{02}(x, y, u, v)^2}}, \quad (2.27)$$

where:

$$\begin{aligned} M'_{01}(x, y, u, v) &= M_{01}(x, y) - \overline{M_{01,u,v}}, \\ M'_{02}(x, y, u, v) &= M_{02}(x - u, y - v) - \overline{M_{02,u,v}}, \end{aligned}$$

and where  $D_{u,v}$  is the region of overlap of images and  $\overline{M_{01,u,v}}$  and  $\overline{M_{02,u,v}}$  are the average of  $M_{01}(x, y)$  and  $M_{02}(x - u, y - v)$  respectively, in the region  $D_{u,v}$ . The difference between the position of the maximum value and the center of the correlation matrix  $\gamma$  gives the displacement between the two images. The resulting shifts describe the lateral motion of the retina.



**Figure 2.14** – Displacements of the eye with time. Drifts are observable each time the points follow each other. They are separated by saccades. The two first saccades take away the focus from the center. The third and the fourth one try to refocus on the center. The total time for these movements is 4 seconds.

## Results

The fig. 2.13 shows retinal vessels acquired by laser Doppler imaging. The fig. 2.13(a) is the whole  $2048 \times 2048$  image. The cross-beating interferometric contribution, which is the object of interest in the right bottom side of the image, is spatially separable from the other parts because of the off-axis configuration. The fig. 2.13(b) shows the focus on retinal vessels. Vessels are moving in the image and sometimes disappear during several frames. Then, vessels do not appear clearly on the time-averaged image of 324 consecutive frames (fig. 2.13(d)). After lateral stabilization, vessels are more visible on the time-averaged image: motion compensation is efficient. Although motion compensation is less efficient when the image changes too often, it improves overall image quality.

We did not go further by comparing the results obtained with other methods because this correlation-based algorithm is sufficient for our imaging system. The aim is just to get the image stabilized so we can work on it with resolution enhancement algorithms.

With the shifts calculated from correlations, a map of displacements can be computed. The map corresponding to the displacements of the fig. 2.13 is represented on the fig. 2.14, where the fixation movements (specially drifts and saccades) are well-visible.

Lateral motion compensation has been implemented in real-time by optimizing the cross-correlations on GPU. It allows the stabilization of the object of interest in the center of the image.

We have mainly shown lateral motion compensation method on laser Doppler images. As laser Doppler and OCT setups are very similar, it can be assumed that this algorithm will be equally effective for holographic OCT.

## 2.4.2 Axial movement compensation

As explained previously, OCT is particularly sensitive to axial shifts, which makes axial compensation critical. In the following, we explain the algorithm of axial movement compensation and show the results obtained on samples.

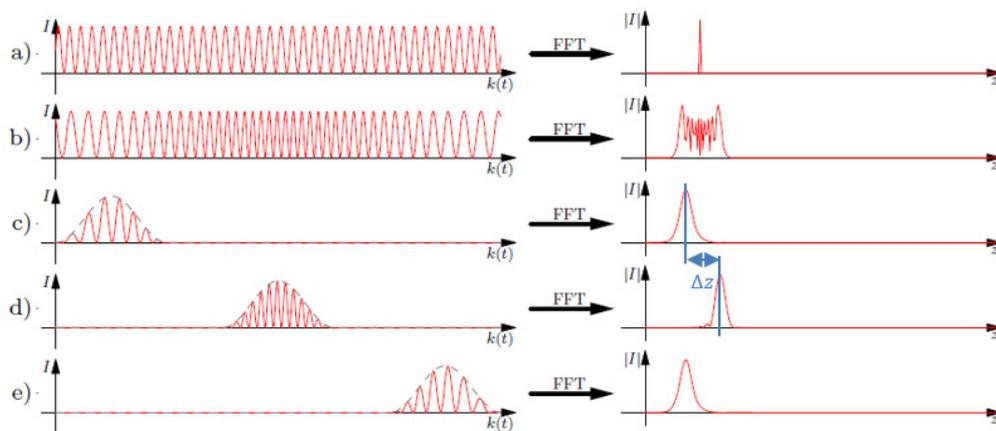


Fig. 2. Schematic representation of the influence of dispersion on an FD-OCT signal and the short-time Fourier transforms. a) Spectrum of the interference from one reflecting surface in the sample. The width of the reconstructed OCT-Signal (point spread function) is determined by the spectral width. b) By dispersion, here visible by a chirp in the sine, the peak in the Fourier transform is broadened and modulated. c)-e) By filtering a small region of the spectrum in which the chirp is neglectable, shifted peaks are restored. However, these peaks are broadened compared to the unchirped case due to the reduced spectral bandwidth introduced by the windowing.

**Figure 2.15** – Extracted from [51]. Underlying principle of the axial motion compensation algorithm. The  $\Delta z$  distance, calculated by correlation between the two signals, is proportional to the derivative of the phase shift with respect to time. Then, integrating the successive  $\Delta z$  allows to recover the phase shift provoked by axial motion.

### Algorithm

Axial motion during the recording process has a negative impact on the reconstruction of OCT images: when an axial drift occurs, the depth information encoded in the beating frequency of the interferogram is inaccurate, adding an offset phase to the signal and decreasing in-depth accuracy. The principle of the compensation method, illustrated in the fig. 2.15, is to identify where the phase shift due to the axial motion occurs by using STFT [51] on 20-point windows. Indeed, a global axial displacement of the sample during acquisition can be spotted in the time Fourier domain, where it corresponds to a frequency shift. In the following, frequencies  $\omega$  and distances  $z$  are considered equivalent as it has been shown in the eq. (2.21).

For  $y = y_0$ , the reference image  $|\tilde{H}_1(x, y_0, t'_0, \omega)|$  and the sliding image  $|\tilde{H}_2(x, y_0, t_0, \omega)|$  are defined as the modulus of the STFT of  $H(x, y_0, t)$  whose sub-windows are fixed or moving, respectively:

$$\begin{aligned} |\tilde{H}_1(x, y_0, t'_0, \omega)| &= |\int H(x, y_0, \tau) g_T(t'_0 - \tau) e^{-i\omega\tau} d\tau|, \\ |\tilde{H}_2(x, y_0, t_0, \omega)| &= |\int H(x, y_0, \tau) g_T(t_0 - \tau) e^{-i\omega\tau} d\tau|, \end{aligned} \quad (2.28)$$

where  $g_T$  is a 20-point Hanning window, and  $t_0$  and  $t'_0$  are the centers of the moving and the reference windows, respectively. The quantitative comparison between  $|\tilde{H}_1(x, y_0, t'_0, \omega)|$  and  $|\tilde{H}_2(x, y_0, t_0, \omega)|$  is performed by calculating the cross-correlation shown in eq. (2.27). Its maximum position gives a frequency shift  $\Delta\omega(t_0)$  corresponding to  $\frac{\partial\phi}{\partial t}(t_0)$ , the derivative of the phase caused by movement at time  $t_0$ . Axial motion compensation is then performed by subtracting the cumulated phase shift  $\int_{t'_0}^{t_0} \Delta\omega(t) dt$  from the phase  $\phi(t_0)$ . This procedure is repeated for all times  $t_0$  in the time gate  $g_T$  used in eq. (2.19) of width  $\omega_s T / (2\pi) = 256$  points [51].

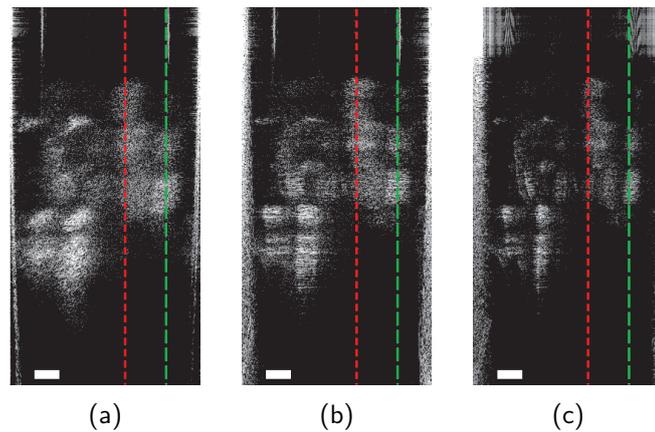
Although no test can be done on a human retina, tests on phantoms are able to prove the efficiency of the method.

### Results

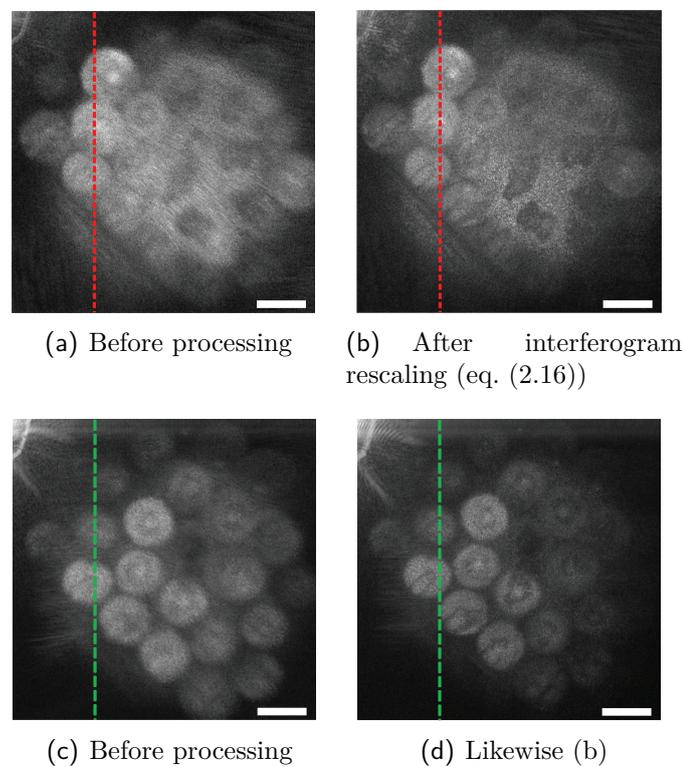
The fig. 2.16 and the fig. 2.17 are swept-source OCT images of samples composed of 1.5 mm diameter glass beads rolled in a single tape layer. The fig. 2.16 shows the in-depth profile  $(x, \omega)$  of this sample corresponding to the y-axis represented by dotted lines in the fig. 2.17. The fig. 2.17 shows the en-face  $(x, y)$  images at two different depths corresponding to the dotted lines in the fig. 2.17. The fig. 2.17(a) and the 2.17(b) correspond to the red dotted line (depth of 1.1 mm) in the fig. 2.16, and the fig. 2.17(c) and the 2.17(d) correspond to the green dotted line (depth of 1.5 mm) in the fig. 2.16. In the fig. 2.17(a) and the 2.17(b), the target layer is located between tape and beads: the beads on the left are at the same elevation as the tape layer on the right which starts to be sectioned. In the fig. 2.17(c) and the 2.17(d), six beads are observed in the layer.

A difference of lateral resolution is visible on images of fig. 2.17. The interferogram rescaling allows the contours of the beads to be cleaner, as we have seen in the section 2.2.3. Comparing fig. 2.16(a) and fig. 2.16(b) shows an improvement in axial resolution. Axial motion correction can be seen in fig. 2.16(c): the accuracy is improved and the different layers are better separated.

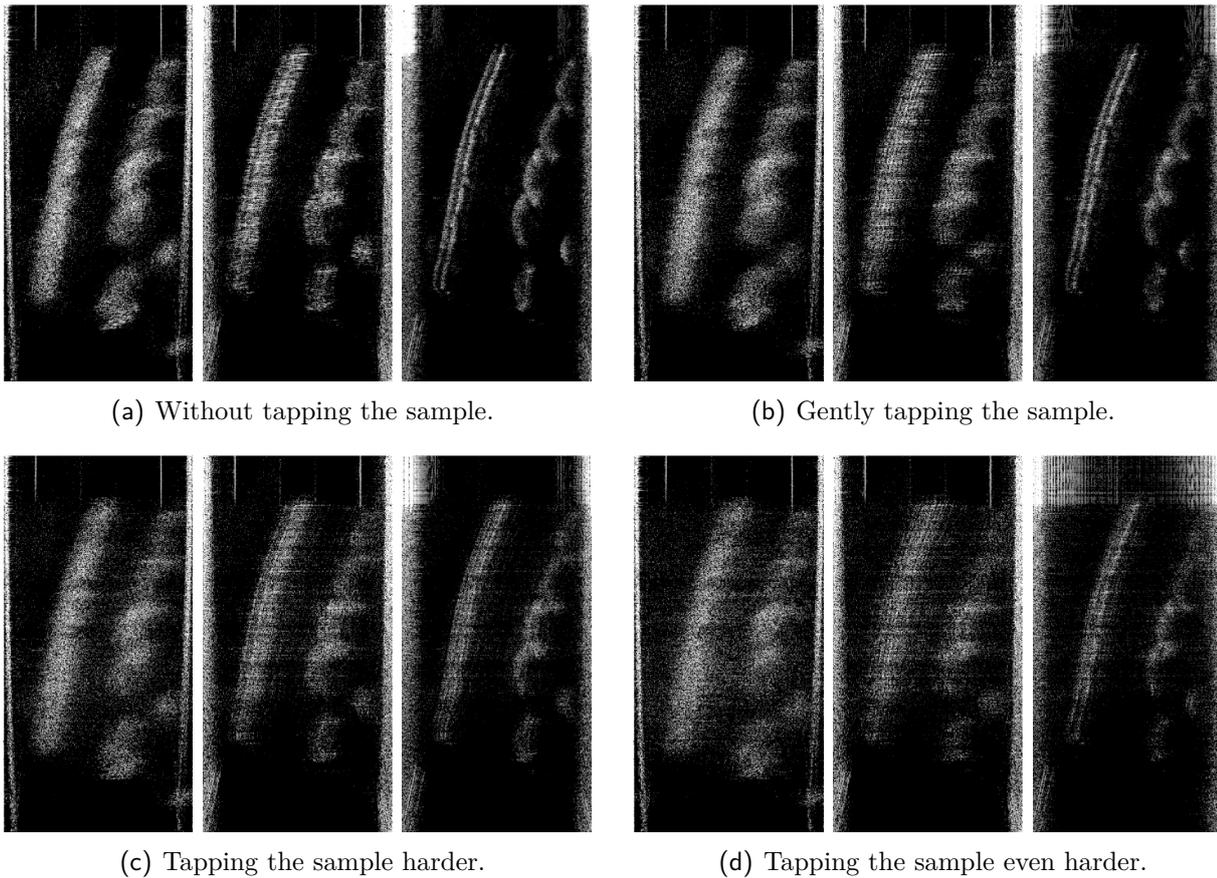
The cause of the axial motion visible on these images is the setup. We have also tried



**Figure 2.16** – Axial motion compensation in swept-source holographic OCT of glass beads rolled in a tape layer. The registered holograms are slices in the  $(x, \omega)$  plane, for the  $y$ -cut in fig. 2.17. Figures represent the sample: (a) without any correction, (b) with interferogram rescaling (eq. (2.16)), and (c) with interferogram rescaling and axial motion compensation. The white scale bars indicate 0.2 mm.



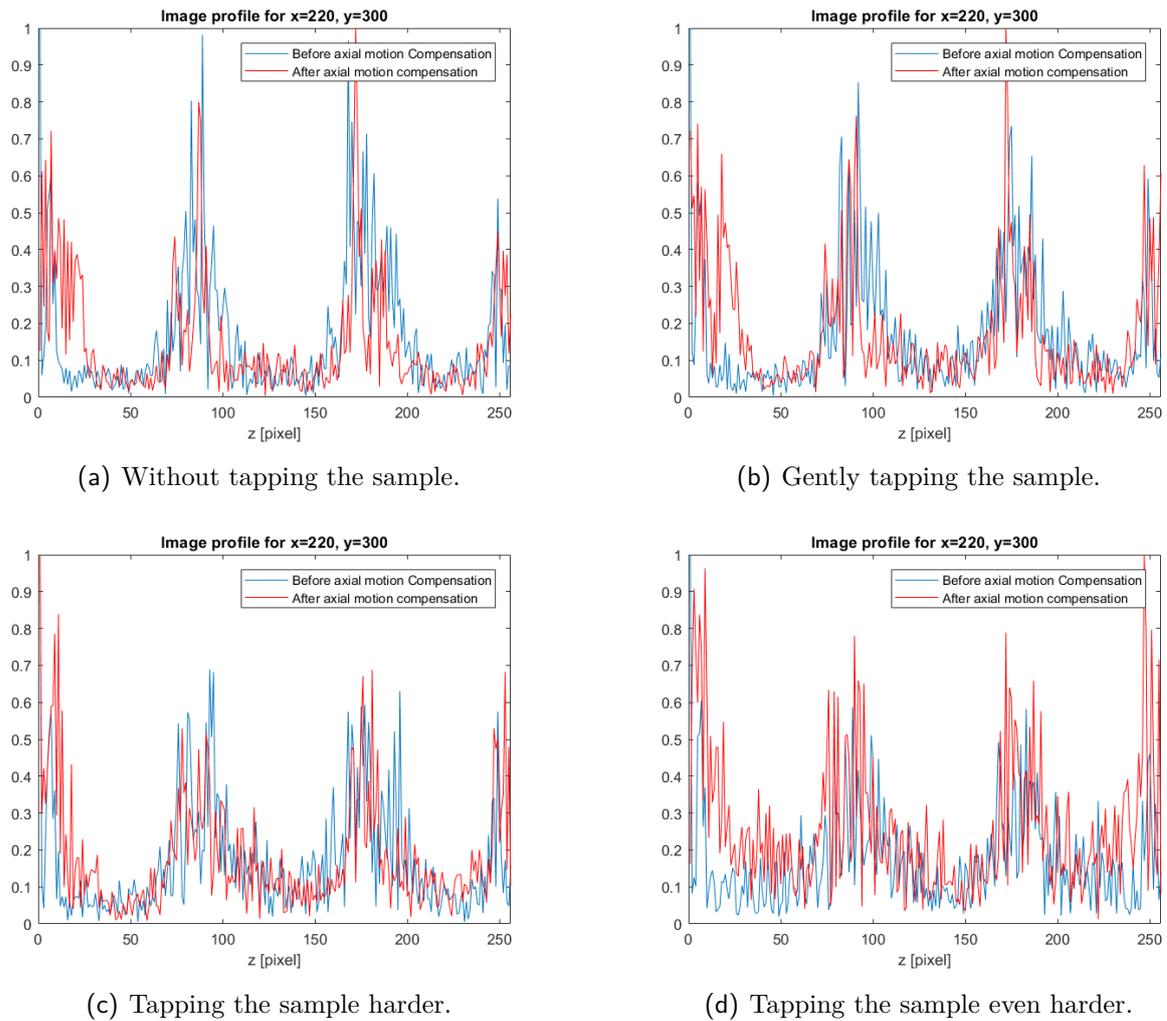
**Figure 2.17** – (a,b) En-face  $(x, y)$  images of glass beads for the first red  $z$ -cut in fig. 2.16, and (c,d) for the second green  $z$ -cut in fig. 2.16. The white scale bars indicate 1.5 mm.



**Figure 2.18** – OCT profile images of glass beads wrapped in tape without any correction (first image of the three), with rescaling (second image) and with rescaling and axial motion compensation (third image), for fixed  $y$ . The fig. 2.18(a) represents the tape layer and the surface of the beads with the only movement provoked by the noise of the setup. For the fig. 2.18(b), fig. 2.18(c) and fig. 2.18(d), the sample was tapped stronger and stronger to provoke axial motion.

to provoke more movement by tapping on the sample. The sample constituted by beads of 1.5 mm of diameter wrapped into tape has been tapped periodically at more or less strong intensity. The results are shown in fig. 2.18 and fig. 2.19. The fig. 2.18 represents 3 images of the same sample without any correction, with rescaling and with rescaling and axial motion compensation, at four levels of axial movement: with the setup noise only (fig. 2.18(a)), with a light tapping (fig. 2.18(b)), with a stronger one (fig. 2.18(c)) and with an even stronger one (fig. 2.18(d)). The fig. 2.19 represents the cut on normalized profiles of the fig. 2.18 at row  $x = 220$ , which corresponds approximately to the central row of the images. It compares the cut before any correction with the image after rescaling and axial motion compensation.

We can see a clear improvement after the axial motion compensation: the contours of the tape layer and the beads are more precise than without correction. Even in the worst case with motion which provokes important blurring of the image (fig. 2.18(d)), the axial movement compensation improves a lot the image quality. Axial motion compensation reduces the background noise and enhances the edges, improving the resolution by a factor of 2.



**Figure 2.19** – Normalized depth profiles of the intensity images of the fig. 2.18 for fixed  $x$  and  $y$ . They correspond to  $x = 220$ , which is the cut located approximately on the center of the images of the fig. 2.18. The blue curve represents the profile without any correction, and the red curve stands for after rescaling and axial motion compensation. In all cases, the corrections improve the resolution.

### 2.4.3 Conclusion on motion compensation

We have presented lateral and axial motion compensation algorithms in laser Doppler holography and holographic OCT respectively, with cross-correlation stabilization methods. These methods are suited to our images and cancel efficiently the effects of motion. Besides, the reported results are non-iterative and compatible with real-time processing at high throughput on GPU. A real-time version of lateral motion compensation has been implemented, not yet for axial motion compensation. The reported results in motion compensation pave the way towards the design of high resolution computational imaging for the retina in real-time by digital holography.



---

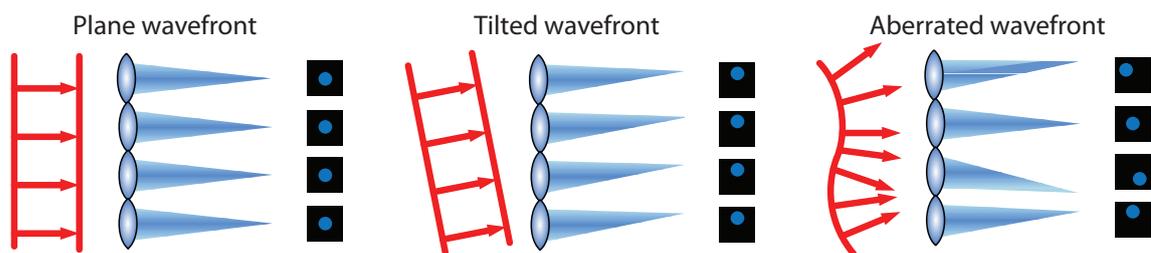
 Estimation of aberrations with a digital wavefront sensor
 

---

### 3.1 The Shack-Hartmann wavefront sensor

After making the first corrections mentioned in the previous chapter, we started working on improving the resolution of the images. The main point to be improved is the handling of aberrations. The objective of this work is to estimate the aberrated phase of the field which has been caused by a deformation of the cornea, the lens and the tear film [67]. This optical system in the eye is considered to be located in the pupil plane, in the Fourier plane of the retina. The method detailed in this chapter is based on the Shack-Hartmann wavefront sensor.

#### 3.1.1 General principle

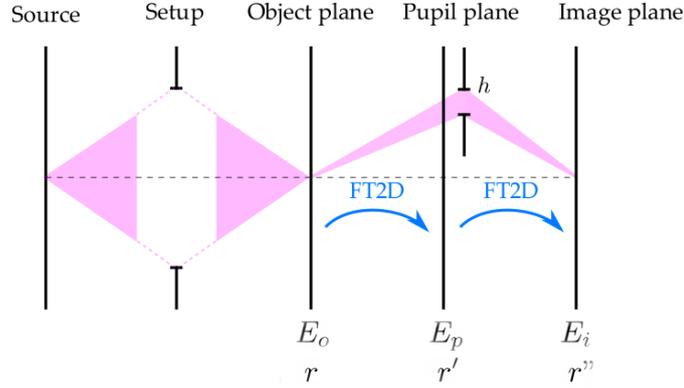


**Figure 3.1** – Illustration of the Shack-Hartmann wavefront sensor principle. When a plane wave goes through the micro-lenses network, each lens focuses the light on its optical axis. If the wavefront is tilted, the focus points are moved in the same direction. If the wavefront is aberrated, the focus points are shifted in different ways. From these shifts, the wavefront slope can be estimated.

The Shack-Hartmann wavefront sensor is the most used wavefront sensor in adaptive optics [77,78]. It uses a micro-lens network to focus different parts of the wave on the sensor. The shift observed between the focus point and the optical axis is recorded and allows the evaluation of the wavefront distortions. The principle of Shack-Hartmann wavefront sensor is illustrated on fig. 3.1.

We will first study the optical configuration of this wavefront sensor, to adapt it in a second step to our setup and implement it digitally.

### 3.1.2 Theoretical study



**Figure 3.2** – Scheme of the optical configuration for using a Shack-Hartmann wavefront sensor. In this configuration, the pupil is in the Fourier plane of the object plane and the image plane is in the Fourier plane of the pupil. Thus, the fields in the image plane and in the object plane are considered identical (the only transform between them is a flip). We consider the setup as a “black box”, and the image of a point source is a point in the object plane.  $h$  is a pupillary transmittance, here represented by a window.

In this part, we study theoretically how the wavefront sensor works. The calculation presented in the following has been classical and well known for several years, but it remains interesting to develop it further in order to understand how it can be applied to our imaging system.

We will rely on the configuration of the fig. 3.2. The Shack-Hartmann sensor is based on the partition of the pupil plane by a network of lenses. We will study the theoretical effect of the partitioning of the pupil plane on the field intensity in the image plane. For this purpose, we first establish a relationship between the intensities in the image plane ( $I_i$ ) and the object plane ( $I_o$ ).

In the following,  $\mathbf{r}$ ,  $\mathbf{r}'$  and  $\mathbf{r}''$  stand for the vectors of coordinates in the object plane, in the pupil plane and in the image plane, respectively. For any transformation  $f$ , we note  $f^-(\mathbf{r}) = f(-\mathbf{r})$ . To lighten the notations, only in this part,  $FT2D\{E\}$  is replaced by  $\tilde{E}$  and the time dependency of the fields in object and image planes does not appear ( $E_o(\mathbf{r}, t) = E_o(\mathbf{r})$  and  $E_i(\mathbf{r}'', t) = E_i(\mathbf{r}'')$ ).

The field intensity in the image plane is expressed as follows:

$$\begin{aligned} I_i(\mathbf{r}'') &= \langle |E_i(\mathbf{r}'')|^2 \rangle_t \\ &= \langle E_i(\mathbf{r}'') E_i^*(\mathbf{r}'') \rangle_t, \end{aligned}$$

where  $\langle \rangle_t$  is the time average of the signal.

We can express  $E_i$  from the field  $E_o$  of the object plane:

$$\begin{aligned}
E_i(\mathbf{r}'') &= FT2D\{\tilde{E}_o \times h\}(\mathbf{r}'') \\
&= (E_o^- * \tilde{h})(\mathbf{r}'') \\
&= \int_{\mathbf{r}_1} E_o(-\mathbf{r}_1) \tilde{h}(\mathbf{r}'' - \mathbf{r}_1) d\mathbf{r}_1,
\end{aligned} \tag{3.1}$$

where  $h$  is the pupillary transmittance. Likewise,

$$E_i^*(\mathbf{r}'') = \int_{\mathbf{r}_2} E_o^*(-\mathbf{r}_2) \tilde{h}^*(\mathbf{r}'' - \mathbf{r}_2) d\mathbf{r}_2. \tag{3.2}$$

Hence,

$$\begin{aligned}
I_i(\mathbf{r}'') &= \langle \iint E_o(-\mathbf{r}_1) E_o^*(-\mathbf{r}_2) \tilde{h}(\mathbf{r}'' - \mathbf{r}_1) \tilde{h}^*(\mathbf{r}'' - \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \rangle_t \\
&= \iint \langle E_o(-\mathbf{r}_1) E_o^*(-\mathbf{r}_2) \rangle_t \tilde{h}(\mathbf{r}'' - \mathbf{r}_1) \tilde{h}^*(\mathbf{r}'' - \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2.
\end{aligned} \tag{3.3}$$

For a conventional use of a Shack-Hartmann wavefront sensor, the source is spatially incoherent, which means that it is composed of independent source points that do not interfere with each other. With this assumption, we can simplify the expression of the mutual intensity in the object plane:

$$J_o(\mathbf{r}_1, \mathbf{r}_2) = \langle E_o(\mathbf{r}_1) E_o^*(\mathbf{r}_2) \rangle_t = \delta(\mathbf{r}_1 - \mathbf{r}_2) I_o(\mathbf{r}_1), \tag{3.4}$$

where  $\delta(\mathbf{r})$  is the Kronecker function.

With this hypothesis, the eq. (3.3) can be simplified:

$$\begin{aligned}
I_i(\mathbf{r}'') &= \iint \delta(\mathbf{r}_2 - \mathbf{r}_1) I_o(-\mathbf{r}_1) \tilde{h}(\mathbf{r}'' - \mathbf{r}_1) \tilde{h}^*(\mathbf{r}'' - \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \\
&= \int I_o(-\mathbf{r}_1) \tilde{h}(\mathbf{r}'' - \mathbf{r}_1) \tilde{h}^*(\mathbf{r}'' - \mathbf{r}_1) d\mathbf{r}_1 \\
&= (I_o^- * |\tilde{h}|^2)(\mathbf{r}'').
\end{aligned} \tag{3.5}$$

With the assumption of spatially incoherent source, the intensity of the field in the image plane is expressed as the convolution of the intensity of the field in the object plane and the square of the pupillary transmittance.

For the Shack-Hartmann wavefront analysis, the pupil plane is parsed into several sub-apertures. First of all, let us consider the pupillary transmittance  $h_1$  which delimits one of these sub-apertures, such as  $h_1(\mathbf{r}') = \Pi_R(\mathbf{r}')$ , where  $\Pi_R(\mathbf{r}')$  is a window of size  $R$ . Let us take another transmittance  $h_2$  shifted by  $\Delta\mathbf{r}'$ , such as  $h_2(\mathbf{r}') = h_1(\mathbf{r}') * \delta(\mathbf{r}' - \Delta\mathbf{r}')$ . With a Fourier transform, we get  $\tilde{h}_2(\mathbf{r}'') = \tilde{h}_1(\mathbf{r}'') \times e^{-2i\pi\mathbf{r}''\Delta\mathbf{r}''}$ . Then,  $|\tilde{h}_2(\mathbf{r}'')|^2 = |\tilde{h}_1(\mathbf{r}'')|^2$ .

Hence with the eq. (3.5), we notice the intensity of the field in the image plane is not modified by a translation of the transmittance  $h$ . The image intensity is always reconstructed in the same way regardless of the position of the sub-aperture. This means the parsing in sub-apertures

does not have any impact on the reconstructed image for each sub-aperture.

This property is a consequence of the spatially incoherent source assumption. If this hypothesis is not validated, then, for  $h_1$  and  $h_2$  previously defined:

$$I_{ih1}(\mathbf{r}'') = \iint \langle E_o(-\mathbf{r}_1)E_o^*(-\mathbf{r}_2) \rangle_t \tilde{h}_1(\mathbf{r}'' - \mathbf{r}_1)\tilde{h}_1^*(\mathbf{r}'' - \mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2, \quad (3.6)$$

and

$$I_{ih2}(\mathbf{r}'') = \iint \langle E_o(-\mathbf{r}_1)E_o^*(-\mathbf{r}_2) \rangle_t \tilde{h}_1(\mathbf{r}'' - \mathbf{r}_1)\tilde{h}_1^*(\mathbf{r}'' - \mathbf{r}_2)e^{2i\pi(\mathbf{r}_1-\mathbf{r}_2)}d\mathbf{r}_1d\mathbf{r}_2. \quad (3.7)$$

In this case,  $I_{ih1} \neq I_{ih2}$ , and each sub-aperture cannot be treated the same way. Thus, the assumption made on the source is essential for the proper functioning of the wavefront analyzer. In the rest of this part, this hypothesis will be kept.

Now let us suppose the transmittance  $h$  is no longer just a window, but is composed of an amplitude  $P$  and a phase  $\phi_a$  as follows:

$$h(\mathbf{r}') = P(\mathbf{r}') \times e^{i\phi_a(\mathbf{r}')}. \quad (3.8)$$

$|\tilde{h}(\mathbf{r}'')|^2$  is the image of  $h$  in the image plane, and it is one of the object of interest of the eq. (3.5).  $|\tilde{h}(\mathbf{r}'')|^2$  can also be seen as the image of a point source through the system. Let us consider the centroid of  $|\tilde{h}(\mathbf{r}'')|^2$ :

$$\begin{aligned} \mathbf{b} &= \frac{\int_{-\infty}^{+\infty} \mathbf{r}'' |\tilde{h}(\mathbf{r}'')|^2 d\mathbf{r}''}{\int_{-\infty}^{+\infty} |\tilde{h}(\mathbf{r}'')|^2 d\mathbf{r}''} \\ &= \frac{\int_{-\infty}^{+\infty} \mathbf{r}'' \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}''}{\int_{-\infty}^{+\infty} \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}''}. \end{aligned} \quad (3.9)$$

We will develop this expression of the centroid.

With Parseval theorem, for two functions  $f$  and  $g$  belonging to  $L^2(\mathbb{R})$  (vector space of square-integrable functions), we have:

$$\int_{-\infty}^{+\infty} f(\mathbf{r}')g^*(\mathbf{r}')d\mathbf{r}' = \int_{-\infty}^{+\infty} \tilde{f}(\mathbf{r}'')\tilde{g}^*(\mathbf{r}'')d\mathbf{r}'' \quad (3.10)$$

We replace  $\tilde{g}$  by  $\tilde{h}$  in this expression, so we get also  $g = h$ . Noting  $\tilde{f}(\mathbf{r}'') = \mathbf{r}''\tilde{h}(\mathbf{r}'')$ , and using the property on the derivative of the Fourier transform, we have:

$$f(\mathbf{r}') = FT2D^{-1}\{\mathbf{r}''\tilde{h}(\mathbf{r}'')\} = -\frac{i}{2\pi} \frac{\partial}{\partial \mathbf{r}'} FT2D^{-1}\{\tilde{h}(\mathbf{r}'')\} = -\frac{i}{2\pi} \frac{\partial h(\mathbf{r}')}{\partial \mathbf{r}'}. \quad (3.11)$$

We can simplify this expression:

$$\frac{\partial h(\mathbf{r}')}{\partial \mathbf{r}'} = \frac{\partial(P(\mathbf{r}')e^{i\phi_a(\mathbf{r}')})}{\partial \mathbf{r}'} = \frac{\partial P(\mathbf{r}')}{\partial \mathbf{r}'} e^{i\phi_a(\mathbf{r}')} + i \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} h(\mathbf{r}'). \quad (3.12)$$

Therefore, we have  $f$  and  $g$ :

$$\begin{cases} f(\mathbf{r}') = -\frac{i}{2\pi} \frac{\partial P(\mathbf{r}')}{\partial \mathbf{r}'} e^{i\phi_a(\mathbf{r}')} + \frac{1}{2\pi} \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} h(\mathbf{r}'), \\ g(\mathbf{r}') = h(\mathbf{r}'). \end{cases} \quad (3.13)$$

Using the eq. (3.10), we deduce that the numerator  $\int_{-\infty}^{+\infty} \mathbf{r}'' \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}''$  of eq. (3.9) is equal to  $\int_{-\infty}^{+\infty} f(\mathbf{r}') g^*(\mathbf{r}') d\mathbf{r}'$ . By replacing  $f$  and  $g$ , this numerator can be written as:

$$\int_{-\infty}^{+\infty} \mathbf{r}'' \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}'' = -\frac{i}{2\pi} \int_{-\infty}^{+\infty} \frac{\partial P(\mathbf{r}')}{\partial \mathbf{r}'} e^{i\phi_a(\mathbf{r}')} h^*(\mathbf{r}') d\mathbf{r}' + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} |h(\mathbf{r}')|^2 d\mathbf{r}'. \quad (3.14)$$

Finally, by replacing  $h$  by its expression in eq. (3.8), we get:

$$\begin{aligned} \int_{-\infty}^{+\infty} \mathbf{r}'' \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}'' &= -\frac{i}{2\pi} \int_{-\infty}^{+\infty} P(\mathbf{r}') \frac{\partial P(\mathbf{r}')}{\partial \mathbf{r}'} d\mathbf{r}' + \frac{1}{2\pi} \int_{-\infty}^{+\infty} |P(\mathbf{r}')|^2 \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} d\mathbf{r}', \\ &= -\frac{i}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{2} \frac{\partial P^2(\mathbf{r}')}{\partial \mathbf{r}'} d\mathbf{r}' + \frac{1}{2\pi} \int_{-\infty}^{+\infty} |P(\mathbf{r}')|^2 \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} d\mathbf{r}'. \end{aligned} \quad (3.15)$$

In this study, the support of the pupil  $P$  is finite. Thus, the value of  $P^2$  at  $+\infty$  and  $-\infty$  is equal to 0. Then, the numerator in eq. (3.15) can be simplified:

$$\int_{-\infty}^{+\infty} \mathbf{r}'' \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}'' = \frac{1}{2\pi} \int_0^{\mathbf{N}_{subap}} |P(\mathbf{r}')|^2 \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} d\mathbf{r}', \quad (3.16)$$

where  $\mathbf{N}_{subap}$  is the sub-aperture size. Using Parseval theorem, the denominator of eq. (3.9) can be written as:

$$\int_{-\infty}^{+\infty} \tilde{h}(\mathbf{r}'') \tilde{h}^*(\mathbf{r}'') d\mathbf{r}'' = \int_0^{\mathbf{N}_{subap}} |P(\mathbf{r}')|^2 d\mathbf{r}'. \quad (3.17)$$

Eventually, the centroid is:

$$\mathbf{b} = \frac{\frac{1}{2\pi} \int_0^{\mathbf{N}_{subap}} |P(\mathbf{r}')|^2 \frac{\partial \phi_a(\mathbf{r}')}{\partial \mathbf{r}'} d\mathbf{r}'}{\int_0^{\mathbf{N}_{subap}} |P(\mathbf{r}')|^2 d\mathbf{r}'}. \quad (3.18)$$

Therefore, the centroid is the average of the derivative of the aberrated phase weighted by the sub-aperture intensity. In other words, it is the average of the weighted phase slope over the sub-aperture.

Thus, the position of the centroid differs from one aberration to another, since the slope of the sub-apertures is different. Then, the focus point in sub-aperture images (or sub-images)

moves depending on these slopes. From eq. (3.5) and the previous analysis on the centroid, we can deduce that the intensity in the image plane can be assimilated to the flipped intensity in the object plane, shifted according to the centroid displacement, and to which is added the diffraction by the sub-aperture of transmittance  $h$ .

With this reasoning, a fundamental property of the Shack-Hartmann wavefront sensor is obtained: under the assumption of spatial incoherence of the source and with the configuration of the optical planes in fig. 3.2, the image of each sub-aperture is the field intensity in the object plane, which has been flipped, translated and diffracted. From the translations, we can estimate the approximate shape of the aberrated phase.

### 3.1.3 Estimation of simulated aberrations

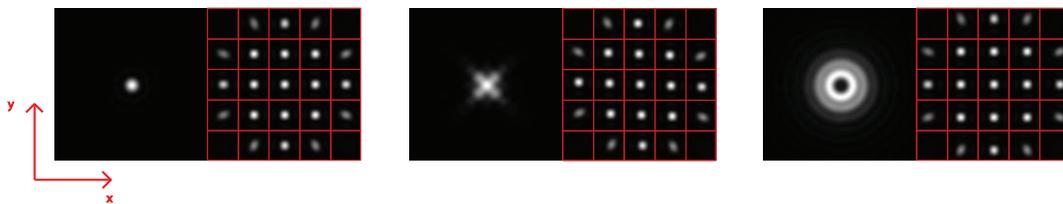
We have explained the theoretical operation of the Shack-Hartmann wavefront sensor. In this part, we test its behavior on simulated data. We simulate aberrations on a point source, then, on an extended object represented by an image. The estimation of the aberrated wavefront is calculated from the translations of the sub-aperture images.

#### Simulation with a point source

For this simulation, the configuration of the fig. 3.2 is taken. Besides, we suppose that a point as a source provides exactly a point in the object plane, so we consider only the object plane, the pupil plane and the image plane. We also suppose the pupil is circular.

Let us consider a point in the object plane. It is projected in the pupil plane by a 2D Fourier transform and an aberrated phase  $\phi_a = \frac{2\pi}{\lambda} \sum_{i=1}^{n_{modes}} a_i Z_i$  is added. Then, an image is produced in the image plane by propagating the field from the pupil plane to the image plane. The fig. 1.15 of Chapter 1 illustrates the shape of the amplitude of the field in the image plane for each Zernike mode.

To realize the digital Shack-Hartmann wavefront sensor, the field in the pupil plane is parsed in  $n_{subap} \times n_{subap}$  sub-apertures, where  $n_{subap}$  is the number of sub-apertures in lateral size. As explained in the section 3.1.2, the intensity of each sub-aperture image (or sub-image) is a point, like the object field.



**Figure 3.3** – Illustration of parsing into  $5 \times 5$  sub-apertures for a point source without any aberration, with astigmatism and with defocus (from left to right). For each part of the illustration, the first image represents the point source in the image plane. The second image represents the sub-images in the image plane. Every focal spot is in the center of each sub-image when there is no aberration. Each aberration is associated to a set of sub-image shifts.

Let  $d_x$  and  $d_y$  be vectors of size  $n_{subap}^2 \times 1$ , containing the horizontal and vertical pixel displacements of the points relative to the reference sub-image (usually one of the central sub-image). Each Zernike mode provides one pair  $(d_x, d_y)$ . For example, the fig. 3.3 shows the shifts corresponding to oblique astigmatism and defocus.

The vector  $d$  of size  $2n_{subap}^2 \times 1$  is defined as the concatenation of  $d_x$  and  $d_y$ . Each mode of Zernike results in a unique  $d$ . So, if we set the value of the  $i^{th}$  Zernike coefficient  $a_i$  to a constant  $a_{ref}$  while the other Zernike coefficients are null, we get a displacement vector  $d_i$  corresponding to one aberration. For  $i \in [1, n_{modes}]$ , we get a set of  $d_i$  vectors corresponding to  $a_i = a_{ref}$  and  $\forall j \neq i, a_j = 0$ . We can gather these vectors in a reference matrix  $M_{ref}$  of size  $2n_{subap}^2 \times n_{modes}$ :

$$M_{ref} = (d_1, d_2, \dots, d_{n_{modes}}). \quad (3.19)$$

Each  $d$  provoked by an aberrated phase can be decomposed as:

$$d = M_{ref} \times a, \quad (3.20)$$

where  $a$  is the vector of Zernike coefficients  $a_i$  of size  $n_{modes} \times 1$ , with  $a_i \in \mathbb{R}$ . As a reminder, the aberrated phase  $\phi_a$  is equal to  $\frac{2\pi}{\lambda} \sum_{i=1}^{n_{modes}} a_i Z_i$ . Therefore, the main goal is to compute the estimate of  $a$ , with  $d$  and the Moore-Penrose inverse of  $M_{ref}$ :

$$\hat{a} = M_{ref}^\dagger d. \quad (3.21)$$

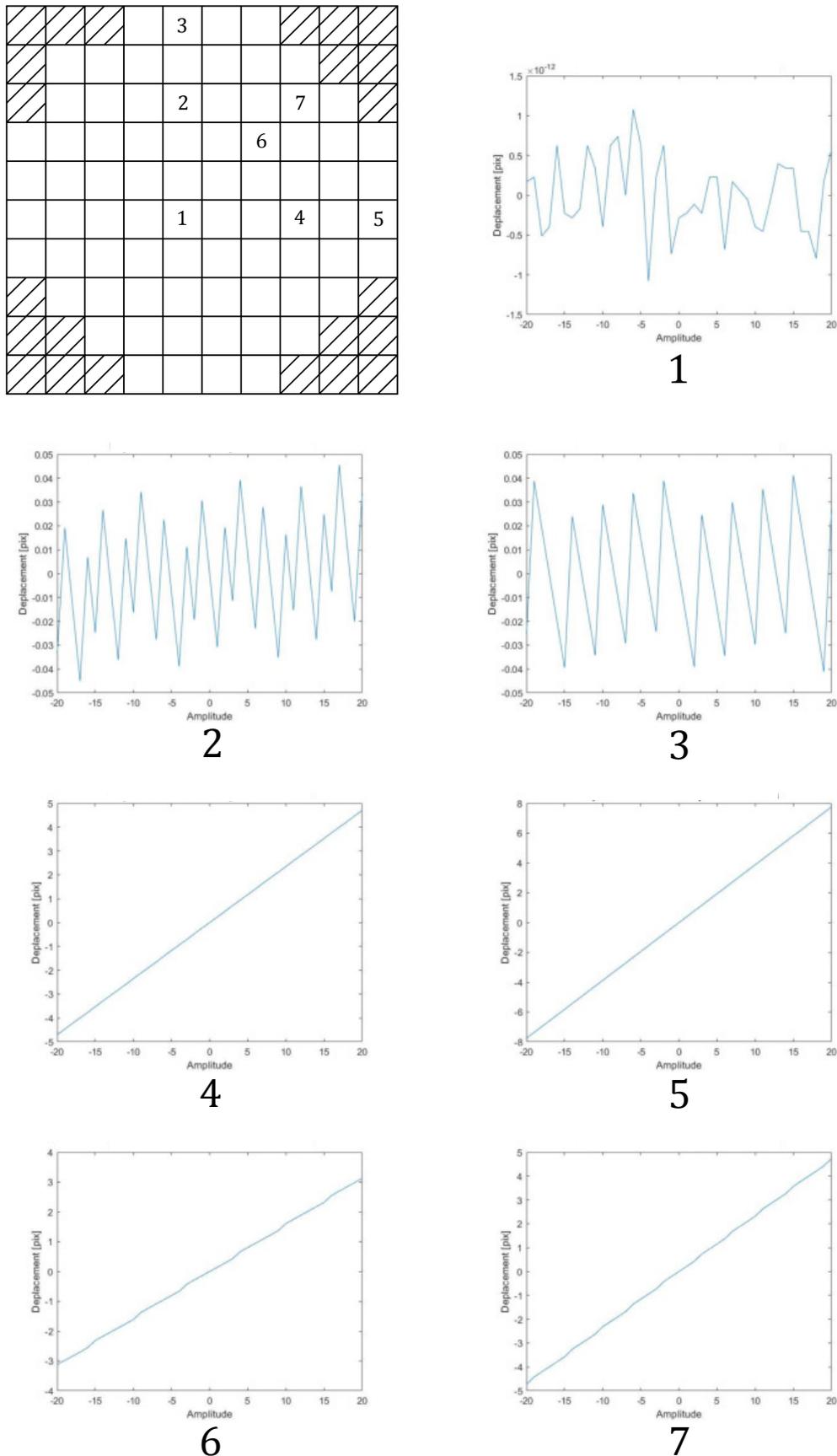
### **Important remarks:**

- Because of the circular pupil, when parsing the pupil plane, some sub-apertures in the corners contain little or no information. We have decided to suppress sub-apertures with less than 50% information. It is done on the fig. 3.3, where sub-apertures in the four corners have been removed. Then, when we consider  $n_{subap}^2$  sub-apertures, in reality some of them have been removed. But to make it easier, we will keep  $n_{subap}^2$  sub-apertures notation.
- While only  $1 \times 1$  parsing is necessary to see tip and tilt, and  $2 \times 2$  for astigmatism and defocus, more sub-apertures are required to estimate high-order aberrations. More generally, for detecting  $n^{th}$  radial order aberrations, it is necessary to have at least  $n \times n$  sub-apertures:  $n_{subap} \geq n$ .
- For tip and tilt, the sub-images are moving exactly in the same way. Then, the shift between sub-images and the central sub-image is null: tip and tilt cannot be detected. Therefore, from now the first estimated aberration order is the oblique astigmatism.

### **Simulation with an extended object**

Let us consider an image in the object plane. As for a point source, aberrations can be added, and parsing into sub-apertures provides a grid of shifted images. With normalized cross-correlations [117] between each sub-image and the reference, the displacements  $d$  can be computed, which gives an estimation of  $a$  with the eq. (3.21).

We have implemented sub-pixel correlation by fitting the maximum of the correlation matrix and its surroundings with a parabola. The coordinates  $x_{m\_subpix}$  and  $y_{m\_subpix}$  are computed



**Figure 3.4** – The grid in the top left represents the parsing into  $10 \times 10$  sub-apertures. Cross-out boxes stand for the sub-apertures which contain less than 50% information. Due to the circular pupil, the corner sub-apertures are not taken into account. Each curve is associated to a number corresponding to a sub-aperture in the grid. The curves represent the evolution of the horizontal position of the focus spot in regard to the reference sub-aperture (number 1) depending on the amplitude of a defocus. For this aberration, the farther from the pupil center the sub-aperture is, the more the focal spot shifts away from the center.

from the coordinates  $(x_m, y_m)$  of the peak of the correlation matrix  $C$  with this parabolic interpolation:

$$\begin{aligned} x_{m\_subpix} &= x_m + \frac{0.5 \times (C(x_{m-1}, y_m) - C(x_{m+1}, y_m))}{C(x_{m-1}, y_m) + C(x_{m+1}, y_m) - 2C(x_m, y_m)}, \\ y_{m\_subpix} &= y_m + \frac{0.5 \times (C(x_m, y_{m-1}) - C(x_m, y_{m+1}))}{C(x_m, y_{m-1}) + C(x_m, y_{m+1}) - 2C(x_m, y_m)}, \end{aligned} \quad (3.22)$$

where  $x_{m-1}$  and  $x_{m+1}$  are the coordinates in  $x$  before and after the position of the maximum  $x_m$  (likewise for  $y_{m-1}$  and  $y_{m+1}$ ). The displacement vector  $d$  is the difference with the center of  $C$  and  $(x_{m\_subpix}, y_{m\_subpix})$ . Then,  $\hat{a}$  can be calculated from  $M_{ref}$  and this measured  $d$  with eq. (3.21).

### Preliminary checks

Several tests must be performed in order to check some basic properties of the system.

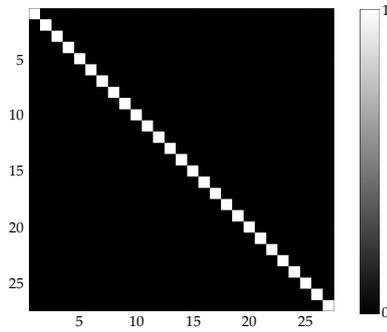
First, we have verified that the shifts are the same for a point source and a extended object, when the same aberration is applied. For this purpose, we have measured the absolute error between estimated shift with a point source and with an extended object, for several random aberrations. For 1000 measurements, we have obtained a mean error of 0.7 pixels with a standard deviation of 1.5 pixels. Therefore, we can conclude that the shifts are the same with an accuracy of the order of a pixel.

Secondly, the linearity of the system must be tested: for the same Zernike mode, the displacement within a sub-image must scale up linearly with the amplitude of the mode. This is illustrated by the fig. 3.4: successive defocus with increasing amplitude have been applied on a point source. The curves represent the horizontal shift between the central reference sub-image and the others. The reference sub-image is the sub-image 1.

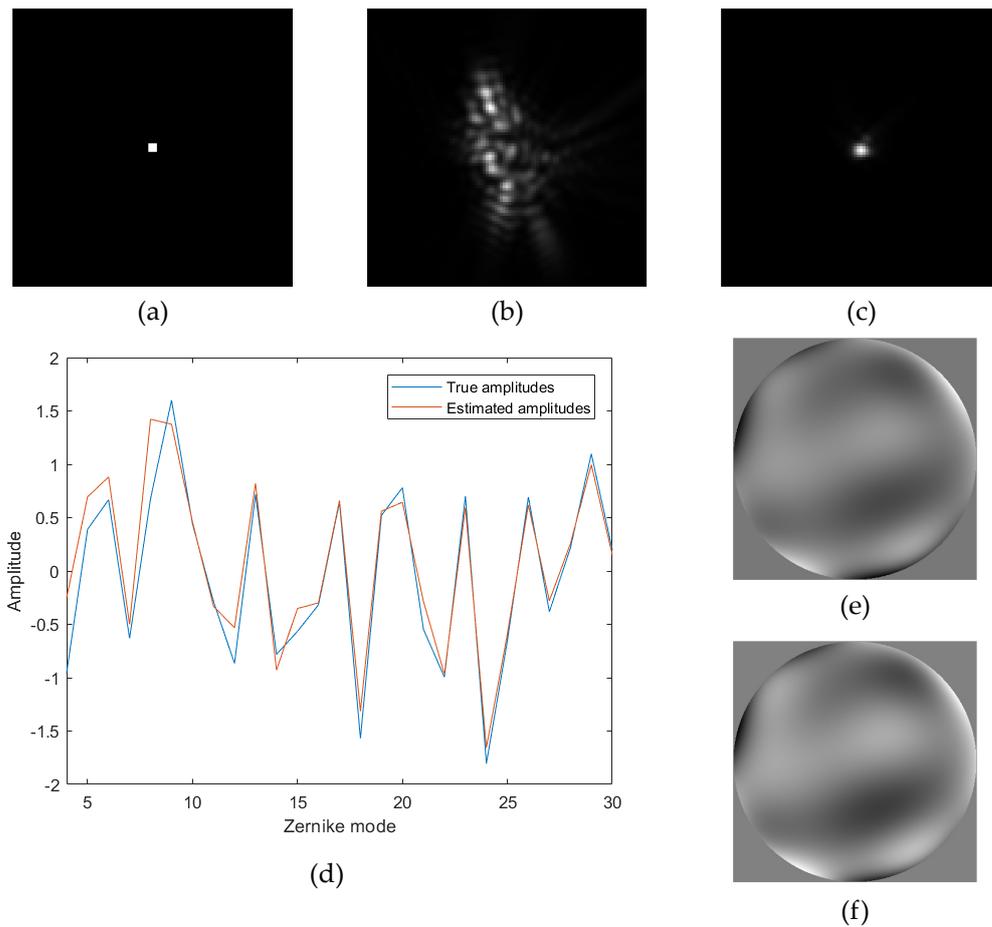
For a defocus, the central sub-images do not move. The closer to the edges the sub-image is, the steeper the slope of the curve must be. The movement of the sub-image 1 is negligible ( $10^{-12}$  pixels) as expected since it is the reference. The horizontal displacement for sub-images 2 and 3 is also small ( $10^{-2}$  pixels). In fact for these sub-images, only the vertical displacement (not represented here) is higher for a defocus. Similarly, the sub-images 4 and 5 have only a horizontal displacement. The slope of the curve of the sub-image 5 is almost twice steeper than the sub-image 4, as expected. Similarly, the slope of the curve of the sub-image 7 is almost twice steeper than the sub-image 6. Because the sub-images 4 and 7 are located at the same horizontal position, they should have the same horizontal displacement curve, which is well verified on the curves. The same verification has been be also successful for vertical shifts. Therefore, the linearity of the system is well respected.

Eventually, since the Moore-Penrose inverse of the matrix  $M_{ref}$  is calculated, care must be taken to ensure that the matrix conditioning is not too high. Conditioning is defined as the ratio of the maximum singular value to the minimum singular value of the matrix. If it is too high, inversion of the matrix may create instability in the resolution of the system. We can check the consistency of  $M_{ref}$  by studying  $M_{ref}^\dagger M_{ref}$ . If this product is not equal to the identity matrix, then, it will bring uncertainty in the calculation of  $\hat{a}$ . Respecting  $n_{subap} \geq n$ , with  $n$  the radial order of the highest considered order aberration, we have verified  $M_{ref}^\dagger M_{ref} = I_{n_{modes}}$ ,

with  $I_{n_{modes}}$  the identity matrix of size  $n_{modes}$ . The fig. 3.5 represents an example of the result of this product for  $10 \times 10$  sub-apertures and 27 Zernike modes.



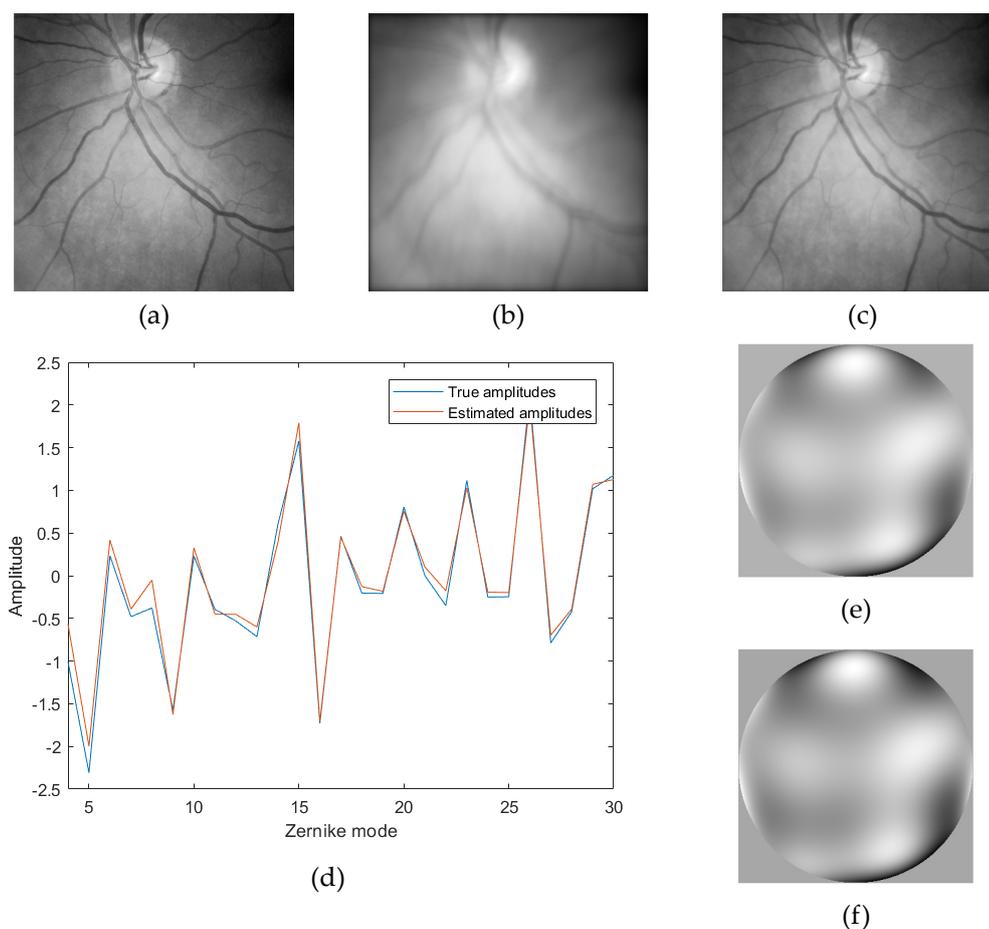
**Figure 3.5** – Product between  $M_{ref}^\dagger$  and  $M_{ref}$  for 27 Zernike modes, from the 4<sup>th</sup> (first astigmatism) to the 30<sup>th</sup>.



**Figure 3.6** – Estimation of the phase for a point source, with a Shack-Hartmann using  $10 \times 10$  sub-apertures. (a) Source. (b) Aberrated PSF. (c) Corrected PSF. (d) Amplitudes of the Zernike coefficients of the added phase in radians (blue) vs estimated amplitudes (orange). (e) Added phase shape. (f) Estimated phase shape.

### Estimation of the aberrated phase

Tests have been realized by applying an aberrated phase to point source and extended object in the pupil plane. Random sequences of Zernike coefficients following a normal distribution  $N(0, 1)$  (average of 0 and standard deviation of 1) have been generated and have constituted aberrated phases. We have used our digital wavefront sensor in order to estimate these random coefficients. The errors (absolute value of the difference between ground truth and estimation) have been computed for each coefficient. Then, we have repeated this estimation 100 times, for both types of source, and using  $6 \times 6$ ,  $7 \times 7$ ,  $8 \times 8$ ,  $9 \times 9$ , and  $10 \times 10$  sub-apertures. We have gathered the corresponding mean error and standard deviation over these 100 iterations on tables, in the appendix A.



**Figure 3.7** – Estimation of the phase for an extended object, with a Shack-Hartmann using  $10 \times 10$  sub-apertures. (a) Image without aberration. (b) Aberrated image. (c) Corrected image. (d) Amplitudes of the Zernike coefficients of the added phase in radians (blue) vs estimated amplitudes (orange). (e) Added phase shape. (f) Estimated phase shape.

On these tables, we can see that in general, mean error and standard deviation decrease while the number of sub-apertures increases. Indeed, by increasing the number of sub-apertures, the accuracy of the estimation is improved as there are more observations to solve the problem and perform the estimation. We can also notice that the error seems slightly more important for the first Zernike modes. These modes stand for lower frequency patterns in the phase, so they need more sub-apertures to be estimated which may increase the possible sources of error and can explain poorer results in average. In any case, the mean error and the standard deviation never

exceed 1 and 0.7 rad, respectively. They are more between 0.2 and 0.3 rad, which is satisfying, particularly in view of the qualitative results.

The fig. 3.6 shows an example of estimation for a point source. The source and the aberrated image are represented on (a) and (b) respectively. The estimation of the Zernike coefficients allows the estimation of the aberrated phase. The figure (c) has been obtained by subtracting the estimated aberrated phase. The point has almost been completely recovered with the successful estimation.

The fig. 3.7 illustrates the estimation for an extended object. The image (a) is in the object plane, and the image (b) represents the intensity of its projection in the image plane. The estimation of Zernike coefficients is close to the real aberrated phase, which makes possible the correction in the image (c).

### 3.1.4 Conclusion on the simulations

In this section, we have studied the theoretical properties of a typical Shack-Hartmann wavefront sensor, and adapted it into a digital version. We have been able to successfully estimate phase on simulated data with this method. The results are promising for point source and extended object.

Now, the challenge is to verify theoretically if the digital Shack-Hartmann wavefront sensor can be used with our setup of laser Doppler holography. Then, if that is the case, the implementation will be discussed and further tests will be performed.

## 3.2 Adaptation to the laser Doppler holography

### 3.2.1 Theoretical modelling with the setup

For this part, we base our work on the holographic data provided by the laser Doppler holographic setup to illustrate theoretical modelling. As seen in the previous section, some hypotheses are necessary to apply the Shack-Hartmann wavefront sensor. The most critical one is the spatial incoherence of the field in the object plane.

Due to the poor signal to noise ratio, we have seen that the moment needs to be computed to obtain laser Doppler images. Thanks to this operation, we are able to neglect the speckle in the signal of the object plane, and then, to fulfill the conditions for the use of the Shack-Hartmann wavefront sensor. It remains to check if a digital wavefront sensor can work with this moment instead of the only intensity, to estimate the slope of the wavefront.

#### Hypotheses

The first hypothesis formulated to apply the Shack-Hartmann wavefront sensor concerns the coherence of the source. In the previous case, it was assumed that the source was spatially incoherent, allowing us to admit the spatial incoherence of the field in the object plane. In reality, the field in the retina is not spatially incoherent. Because of the propagation of coherent light in a scattering media, the intensity  $I$  of this field can be decomposed as:

$$I = E[I] + \sigma_I. \quad (3.23)$$

$E[I]$  is the statistical mean value of the intensity over time and  $\sigma_I$  the standard deviation of  $I$  which has the same amplitude as  $E[I]$ . The speckle provoked by  $\sigma_I$  can be observed in the final images. However, by calculating the moment to obtain the Doppler images, we average several images of the data blocks. Then, the speckle effect is less visible than the average  $E[I]$  which forms the image. In that case, we can consider  $E[I]$  as an incoherent part which predominates over coherent part. Therefore, the hypothesis of incoherence of the field in the retina plane can be applied on the incoherent part of the field, knowing the coherent part can be neglected with the final processes.

The second hypothesis concerns the arrangement of the planes. In our case, as we have seen in the section 2.1.2, we consider that the plane of aberrations is the Fourier plane of the retina plane. Then, it is the same situation as in section 3.1.2.

### The wavefront sensor for laser Doppler holography

As we have seen in the section 2.1.3, the interferograms  $I(x, y, t)$  are propagated in the image plane to get holograms  $H(x, y, t)$ . The configurations in section 2.1.3 and section 3.1.2 are slightly different. In the case of the section 2.1.3, the holograms are propagated from the image plane rather than the object plane. For this study, we will note  $E_{io}(\mathbf{r}'', t) = H(x, y, t)$  the original field in the image plane,  $I_{io}(\mathbf{r}'') = \langle |E_{io}(\mathbf{r}'', t)|^2 \rangle_t$  the corresponding intensity,  $E_p(\mathbf{r}', t) = FT2D\{E_{io}(\mathbf{r}'', t)(\mathbf{r}')\}$  the field in the pupil plane, and  $E_i(\mathbf{r}'', t) = FT2D^{-1}\{E_p \times h\}(\mathbf{r}'', t)$  the field in the image plane after propagation from the pupil plane. Repeating the calculations of the section 3.1.2 with these slight changes, we reach the equation:

$$I_i(\mathbf{r}'') = I_{io}(\mathbf{r}'') * |FT2D^{-1}\{h\}(\mathbf{r}'')|^2. \quad (3.24)$$

By taking into account the parsing of the pupil and the effect of aberrations, we come to the same conclusion we had in section 3.1.2 for the centroid displacement. Therefore, with this configuration, each sub-image corresponds to the shifted  $I_{io}(\mathbf{r}'')$ .

However, in this setup of laser Doppler imaging, the computation of  $I_i(\mathbf{r}'')$  or  $I_{io}(\mathbf{r}'')$  is not sufficient to get proper image. Several steps which are explained in the section 2.3 must be performed to obtain the final Doppler images:  $M_{io}(x, y, t) = \int_{\omega_1}^{\omega_2} |\tilde{H}(x, y, t, \omega)|^2 d\omega$ . In the rest of this section, we will study the effect of these successive operations on the wavefront analysis.

First, we consider the moment  $M_i$  in the image plane of the fig. 3.2 such as

$$\begin{aligned} M_i(\mathbf{r}'', t) &= \langle |FT_t\{E_i(\mathbf{r}'', t)\}|^2 \rangle_\omega \\ &= \langle FT_t\{E_i(\mathbf{r}'', t)\} FT_t\{E_i(\mathbf{r}'', t)\}^* \rangle_\omega \\ &= \int_\omega FT_t\{E_i(\mathbf{r}'', t)\} FT_t\{E_i(\mathbf{r}'', t)\}^* d\omega. \end{aligned} \quad (3.25)$$

With a window  $h$  in the pupil plane, we get:

$$E_i(\mathbf{r}'', t) = E_{io}(\mathbf{r}'', t) * FT2D^{-1}\{h(\mathbf{r}'')\} = \int_{r_1} E_{io}(\mathbf{r}_1, t) FT2D^{-1}\{h\}(\mathbf{r}'' - \mathbf{r}_1) d\mathbf{r}_1, \quad (3.26)$$

and

$$FT_t\{E_i(\mathbf{r}'', t)\} = \int_{\mathbf{r}_1} FT_t\{E_{io}(\mathbf{r}_1, t)\} FT2D^{-1}\{h\}(\mathbf{r}'' - \mathbf{r}_1) d\mathbf{r}_1. \quad (3.27)$$

Then,

$$M_i(\mathbf{r}'', t) = \langle \iint_{\mathbf{r}_1, \mathbf{r}_2} FT_t\{E_{io}(\mathbf{r}_1, t)\} FT_t\{E_{io}(\mathbf{r}_2, t)\}^* FT2D^{-1}\{h\}(\mathbf{r}'' - \mathbf{r}_1) FT2D^{-1}\{h\}^*(\mathbf{r}'' - \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \rangle_{\omega}. \quad (3.28)$$

Because of the hypothesis of spatial incoherence of the source, we can show that we have:

$$\langle FT_t\{E_{io}(\mathbf{r}_1, t)\} FT_t\{E_{io}^*(\mathbf{r}_2, t)\} \rangle_{\omega} = \delta(\mathbf{r}_1 - \mathbf{r}_2) \int |FT_t\{E_{io}(\mathbf{r}_1, t)\}|^2 d\omega. \quad (3.29)$$

Then, the eq. (3.28) can be simplified:

$$M_i(\mathbf{r}'', t) = \langle |FT_t\{E_{io}(\mathbf{r}'', t)\}|^2 \rangle_{\omega} * |FT2D^{-1}\{h\}(\mathbf{r}'')|^2 = M_{io}(\mathbf{r}'', t) * |FT2D^{-1}\{h\}(\mathbf{r}'')|^2. \quad (3.30)$$

As we said previously, by taking into account the parsing of the pupil and the effect of aberrations, the centroid of  $|FT2D^{-1}\{h\}(\mathbf{r}'')|^2$  moves with the sub-aperture phase slope.

Therefore, in each sub-image, we will be able to observe the displacements of the moment  $M_{io}(\mathbf{r}'', t)$  which is equal to  $\langle |FT_t\{E_{io}(\mathbf{r}'', t)\}|^2 \rangle_{\omega} = \int_{\omega_1}^{\omega_2} |\tilde{H}(x, y, t, \omega)|^2 d\omega$ . The vector  $d$  representing the displacements can be measured from cross-correlation between the moments, and the phase can be estimated with eq. (3.21).

### 3.2.2 Implementation

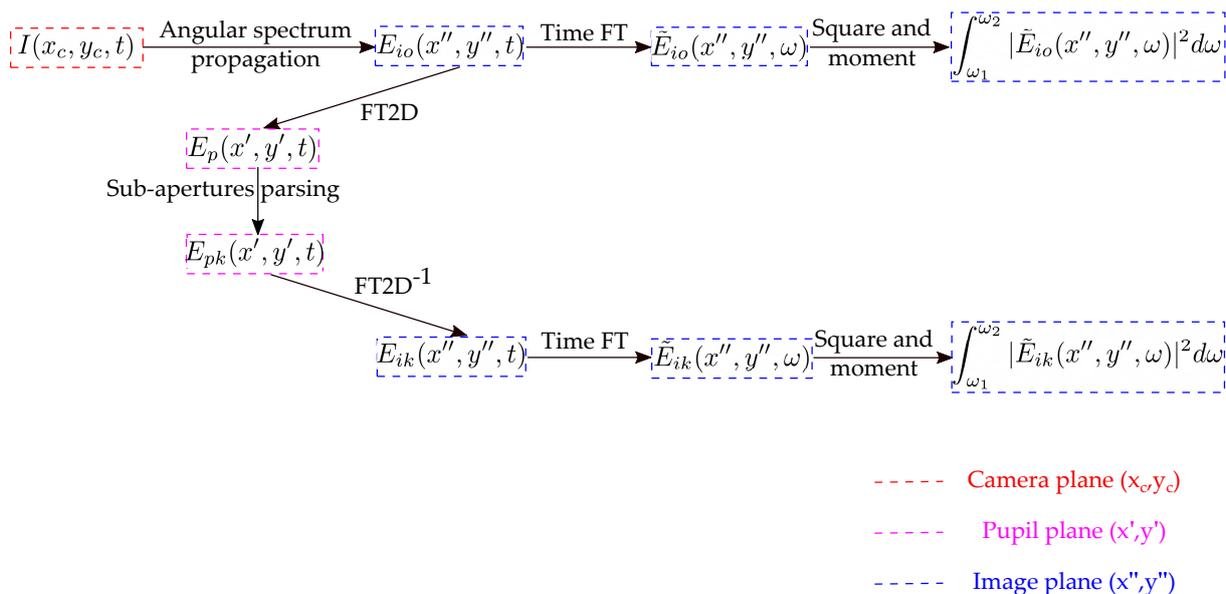
The main processing steps to get sub-images are summed up in the fig. 3.8.

All these processes are implemented with Matlab. In this part, I give some details about implementation options we have added. The critical point concerns the correlations between sub-images: it has to be done properly to compute the right phase.

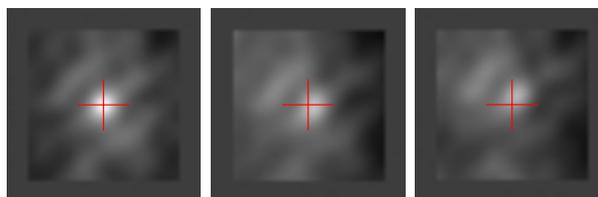
First we apply a gaussian filter to the sub-images before the correlation, like in the section 2.4.1. We can also apply a filtering with a singular value decomposition (SVD) to remove specular reflections and artifacts related to the movement for example.

In order to avoid as many errors as possible, we set several pixels of the correlation matrix to zero so as to keep only the pixels in the center: even with high amplitudes of aberration, the displacement of the maximum remains small, as illustrated by the fig. 3.9. We generally remove at least 80% of the pixels on each lateral dimension. This precaution prevents us from having too high displacement values, caused by local maxima due to artifacts or side effects.

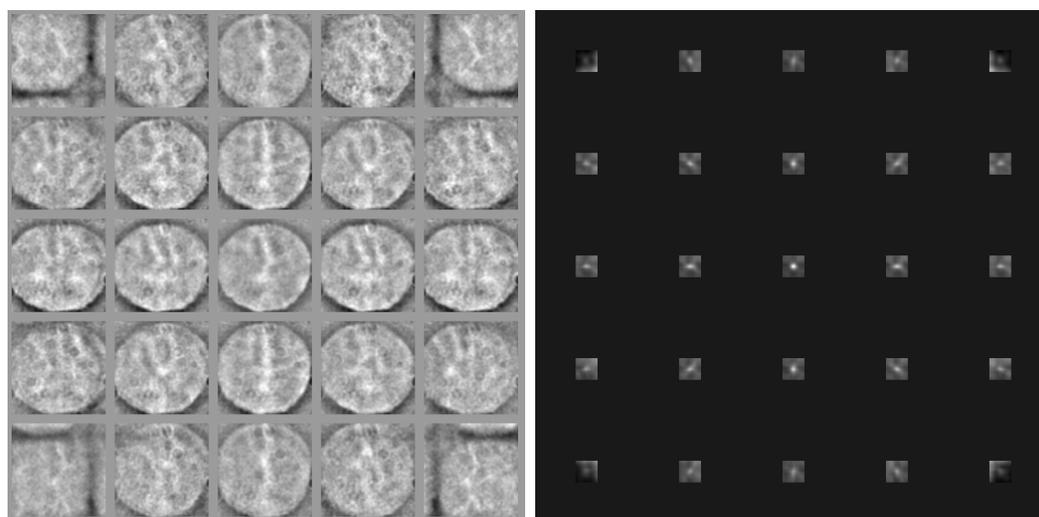
The fig. 3.10 shows another problem related to the sub-images. On the left figure, not to mention edge effects on the four corner sub-images, a circular limiting aperture is well visible on each sub-image. We suspect this is due to the size of the patient's pupil since it is not visible



**Figure 3.8** – Summary of the main steps to get sub-images. The interferograms  $I(x_c, y_c, t)$  are propagated from camera plane to image plane to get holograms  $E_{io}(x'', y'', t) = H(x'', y'', t)$ . The steps on the top after  $E_{io}(x'', y'', t)$  sum up the usual processing to get images. To compute sub-images, the holograms are propagated in the pupil plane with a 2D Fourier transform. The holograms in the pupil plane  $E_p(x', y', t)$  are parsed into sub-apertures  $E_{pk}(x', y', t)$ , with  $k \in [1, n_{subap}^2]$ . Each sub-aperture  $E_{pk}(x', y', t)$  is transferred in the image plane. Then, every sub-image  $E_{ik}(x'', y'', t)$  is processed to get images.



**Figure 3.9** – Examples of correlation matrices. The red cross is located at the center of the matrix. The maximum of the correlation maps is represented by the white value. The displacement remains small even to describe high amplitude aberrations.



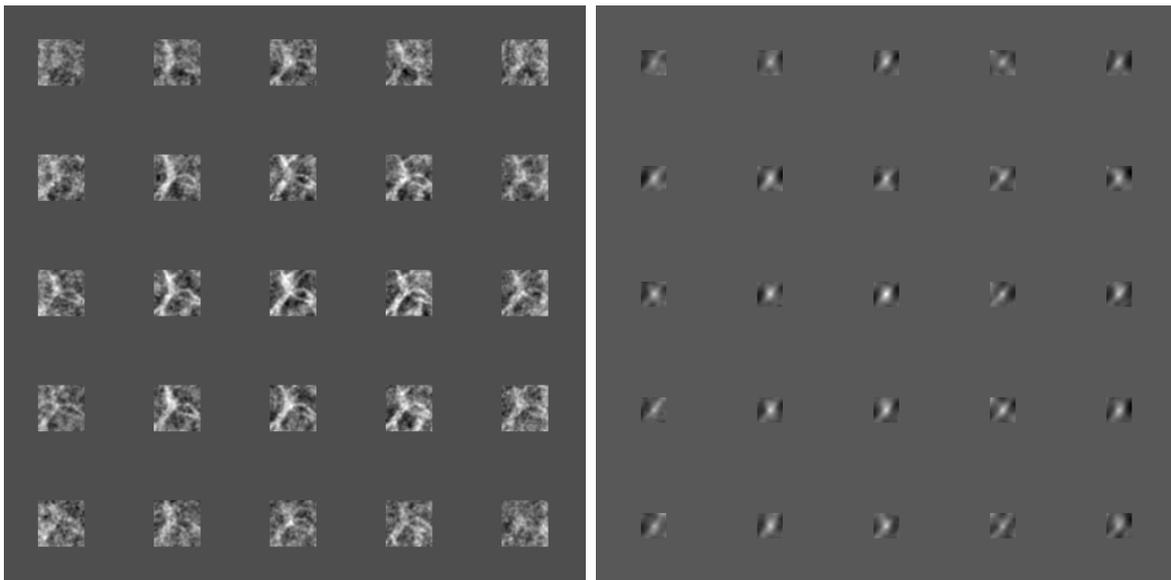
**Figure 3.10** – Example of sub-images and corresponding cropped correlation matrices, for  $5 \times 5$  sub-apertures.

in the same way for every person. However, this disturbs the result of the correlations: the correspondence between sub-images due to the limiting aperture produces a high-energy peak in the correlation map which covers the peak related to the intended displacement. In order to hide this phenomenon, we have decided to do the same we did with the correlation matrix: we keep only the center of the image, removing about 60% of the pixels on each lateral dimension of the image. Even if it seems to be a substantial part of the image, only a little with several vessel structures is required to spot movement.

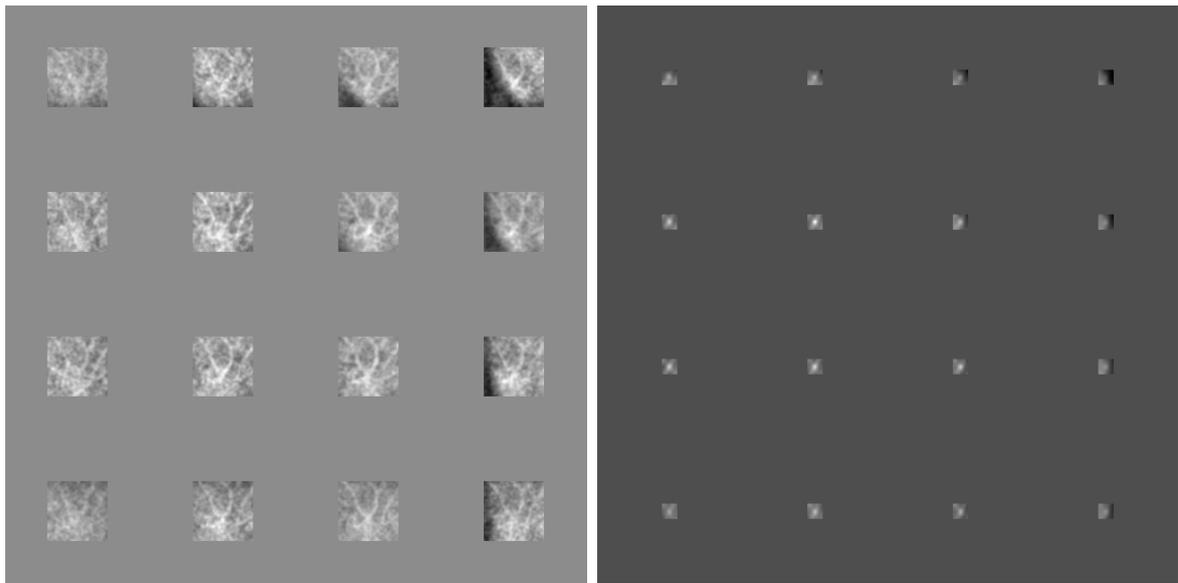
The fig. 3.11 illustrates a parsing with these zero margins in sub-images and correlation matrices. From these views, we can remove some extra sub-apertures if it is needed. Indeed, on the figure, it is noticeable vessels of interest within sub-images are fading as they are farther from the central sub-apertures. It attenuates the corresponding correlation peaks. In some cases, it is necessary to remove from calculations some other sub-apertures in addition to the four corners. Besides, it also highlights that the more sub-apertures there are, the more off-center sub-apertures we will have to suppress because of the fading. For this reason, we have decided to work with only 4 or 5 lateral sub-apertures, removing the four corner sub-apertures by default.

Nevertheless, sometimes, removing only four sub-apertures is not sufficient because of fading as explained before, but also due to other artefacts. In some acquisitions, the shape of the pupil is still visible even with the zero-margin mask, as it is illustrated in the fig. 3.12. In those cases, more sub-apertures have to be removed, even though it can reduce drastically the precision of the estimation or the maximum possible number of modes.

Default parameters have worked pretty well for most of the tested cases as it is shown in the section 3.3: the masks for sub-images and correlations, and 4 or 5 lateral sub-apertures. However, for some patients, it is necessary to take a closer look to settings.



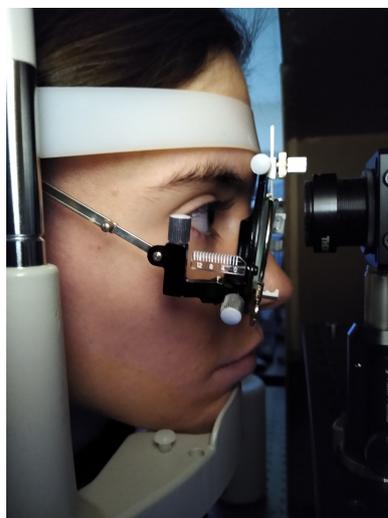
**Figure 3.11** – Example of cropped sub-images and corresponding cropped correlation matrices, for  $5 \times 5$  sub-apertures.



**Figure 3.12** – Example of cropped sub-images and corresponding cropped correlation matrices, for  $4 \times 4$  sub-apertures. Here all the sub-images on the right column show a limiting aperture, which spoils the correlation: the corresponding peaks in the correlation matrices are barely or even not visible.

### 3.3 Tests and results

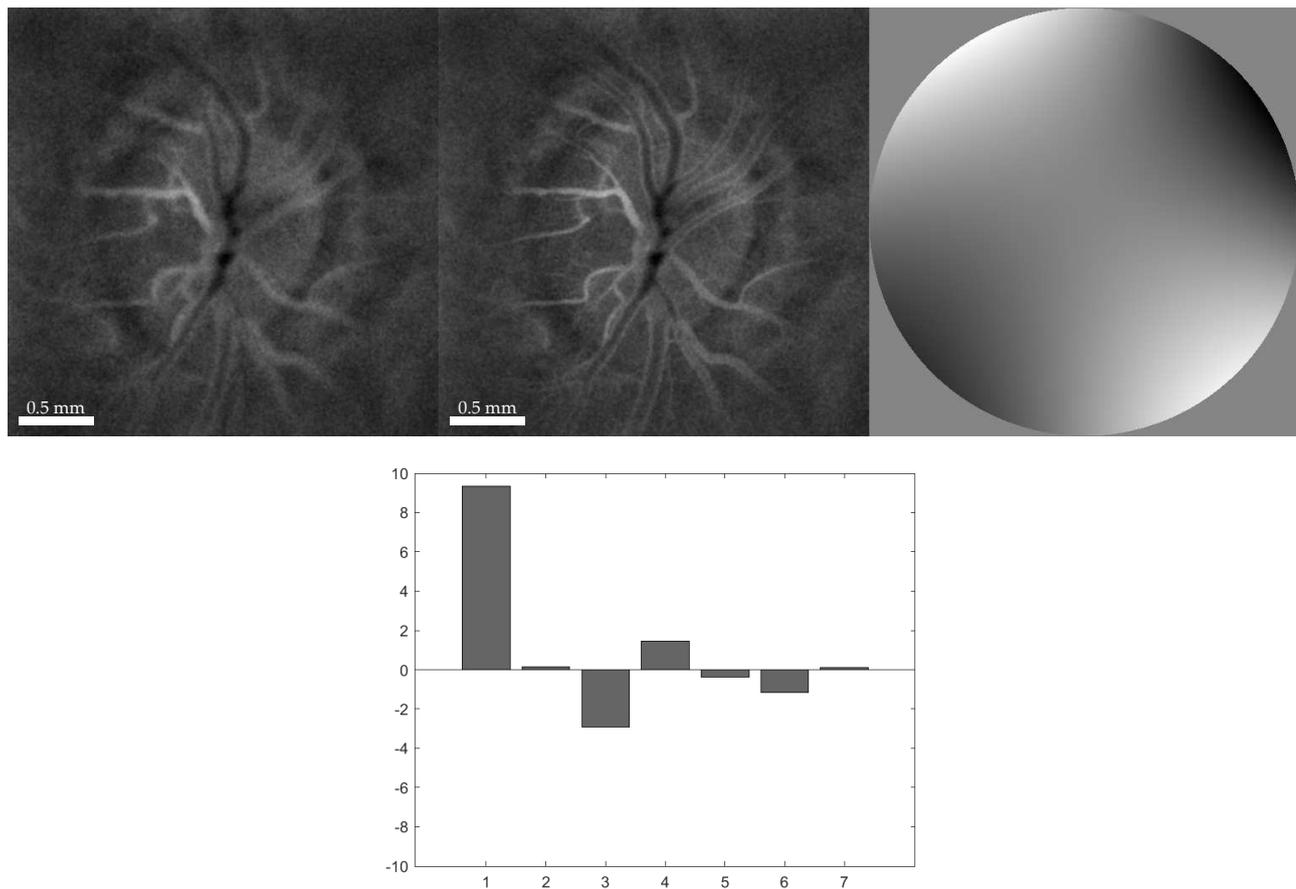
All the laser Doppler images from this section have been obtained by averaging consecutive frames of registered computed moments.



**Figure 3.13** – Photo of the verification experiment. Astigmatic glasses are placed and adjusted in front of the emmetropic eyes using holders.

#### 3.3.1 Experimental verification

The experimental checks have been performed on a emmetropic eye. Astigmatic glasses are placed just in front of the eyes. We can change the intensity of the astigmatism (expressed in diopters) and its orientation. With the algorithm, we can visualize images before and after

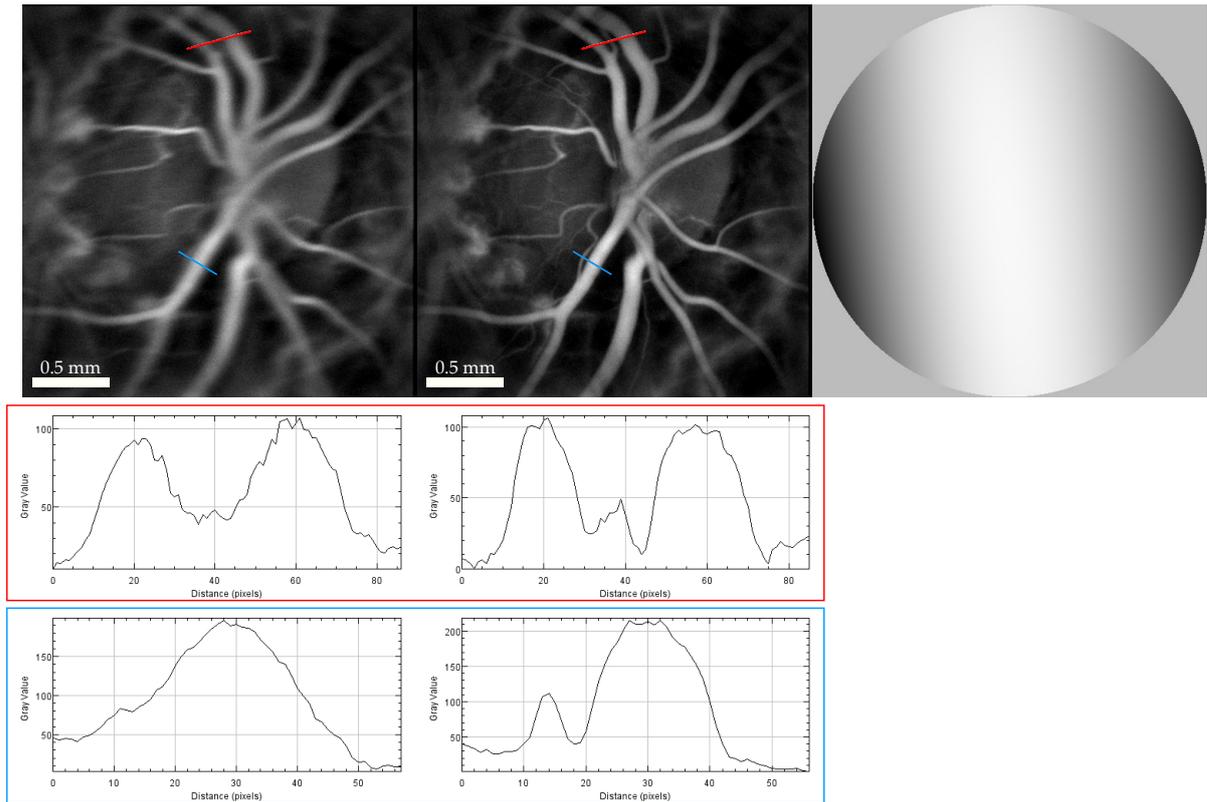


**Figure 3.14** – Correction of a laser Doppler image of the model retina, where 6-diopter oblique astigmatism has been added. The first figure represents the image before aberration compensation, while the second one represents the corrected image. The third figure illustrates the shape of the wavefront. The graph stands for the amplitude of the coefficient for Zernike modes, corresponding to this phase estimation. Amplitudes are expressed in radian and Zernike mode numbering starts with oblique astigmatism.

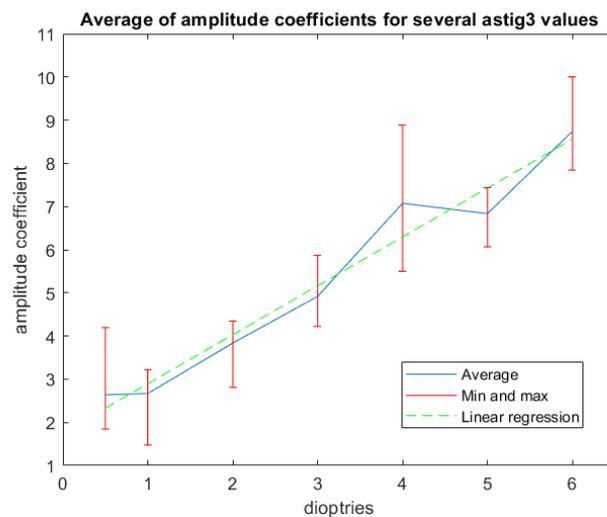
correction, with the shape of the estimated phase and the amplitude of Zernike coefficients. An example of the rendering is shown on fig. 3.14. On this example, it can be seen that the estimated phase is indeed an astigmatism: the first coefficient which represents the oblique astigmatism is equal to 9 radians. When the absolute value of coefficients is lower than 1 radian, we consider them as null as their effects are not significant.

A clear improvement of the image resolution is well-visible on the first tests. For fig. 3.15, 6-diopter horizontal astigmatism has been added. The figure show profiles before and after aberration compensation. The resolution of the vessels has been improved and it even reveals other vessels which were barely visible before the correction.

One of the first test performed concerns the increase of the amplitude of Zernike coefficient with the increase of the diopter number. Then, we put successively, in the same conditions of test, astigmatic lenses of 0.25, 0.5, 1, 2, 3, 4, 5 and 6 diopters, oriented to obtain a vertical astigmatism. For 0.25 diopters, the results were not meaningful (not visible with our device) so we removed them from the study. We gather results in the fig. 3.16 which represents the variation of the mean amplitude of the vertical astigmatism with the number of diopters. The average has been computed on 25 images, extracted from a few seconds recording. The red



**Figure 3.15** – Correction of a laser Doppler image of the model retina, where 6-diopter horizontal astigmatism has been added. The first figure represents the image before aberration compensation. The gray-level profiles corresponding to red and blue lines are shown just below. The second one represents the corrected image. The gray-level profiles corresponding to red and blue lines are shown just below. The third figure illustrates the shape of the wavefront.

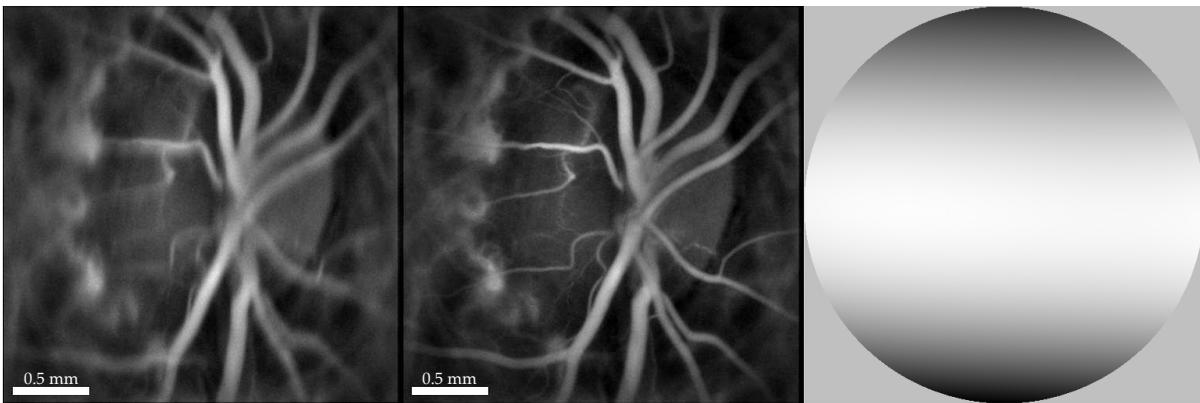


**Figure 3.16** – Evolution the amplitude of the second astigmatism in radian for several values of vertical astigmatism. The blue curve stands for the average value of 25 samples representative of a few seconds of acquisition. The minimum and maximum values of this 25 samples are represented in red. They are representative of the dynamic aberrations which have occurred during this time. The green dotted curve is the linear regression of the blue curve.

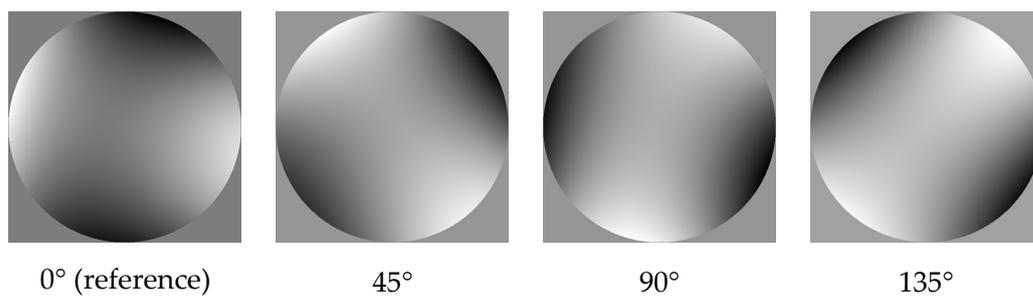
points stand for the minimum and maximum amplitude values among these 25 frames.

The curve is linear as we expected. The slope which is about  $1 \text{ rad.m}^{-1}$  is the same for all astigmatism we have tested. Adding other aberrations affects a bit this slope, but we have not quantified this modification yet. The minimum and maximum values reflect dynamic aberrations: the value of the astigmatism can vary during the acquisition due to dynamic aberrations. However, here the variation can reach almost 4 radians. It seems to be due to an error of measurement that we have not quantified yet. It could be caused by the variation of other modes of Zernike, as we said previously.

As for the link between amplitude coefficients and diopters for astigmatism, we have been able to measure a relationship between reconstruction distance  $z$  and amplitude coefficient for defocus, validated not only on the model eye, but also on each case we have tested so far. Then, we have been able to digitally compensate the defocus by reconstructing images at precise  $z$  distance, so we did not take the risk to influence the other aberrations during the estimation.



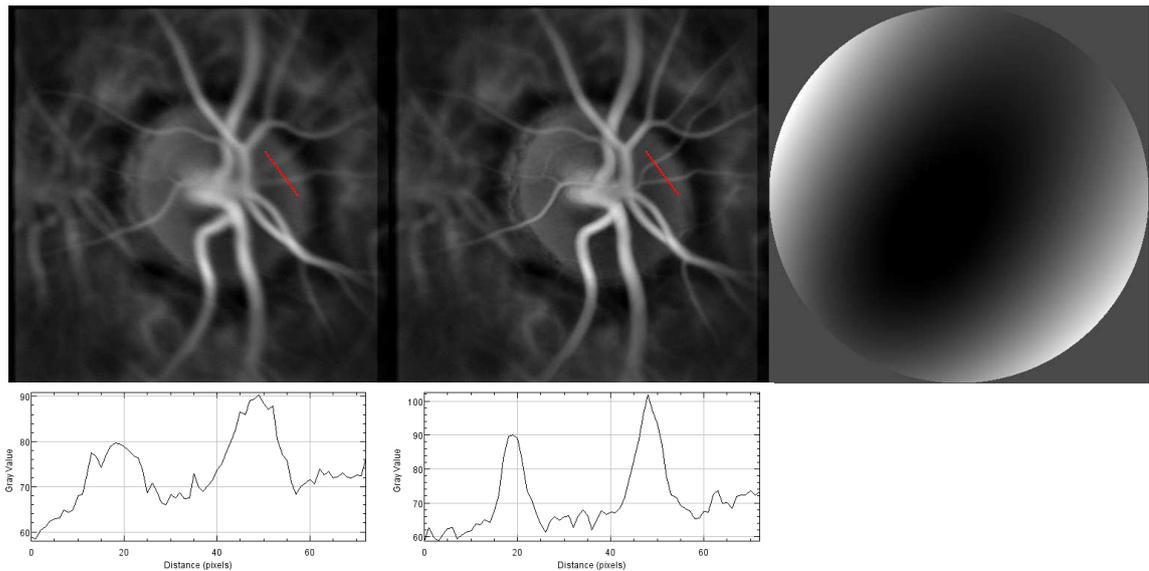
**Figure 3.17** – Correction of a laser Doppler image of the model retina, where 6-diopter vertical astigmatism has been added. The first figure represents the image before aberration compensation, while the second one represents the corrected image. The third figure illustrates the shape of the wavefront.



**Figure 3.18** – Estimated shape of the wavefront for different orientations of the astigmatism.

For another test, we have modified the orientation of the astigmatism. On fig. 3.15, the astigmatism provokes a blurring effect horizontally oriented, as we can see on the aberrated image. On the other side, the vertical astigmatism is also visible with the orientation of the blurring in the fig. 3.17.

To check the orientation of the aberration, we have set and measured a 4-diopter astigmatism as a reference (first illustration of the fig. 3.18). Then, we have rotated it  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ . The fig. 3.18 shows the shape of the wavefront in each case, which corresponds to the rotation we applied.



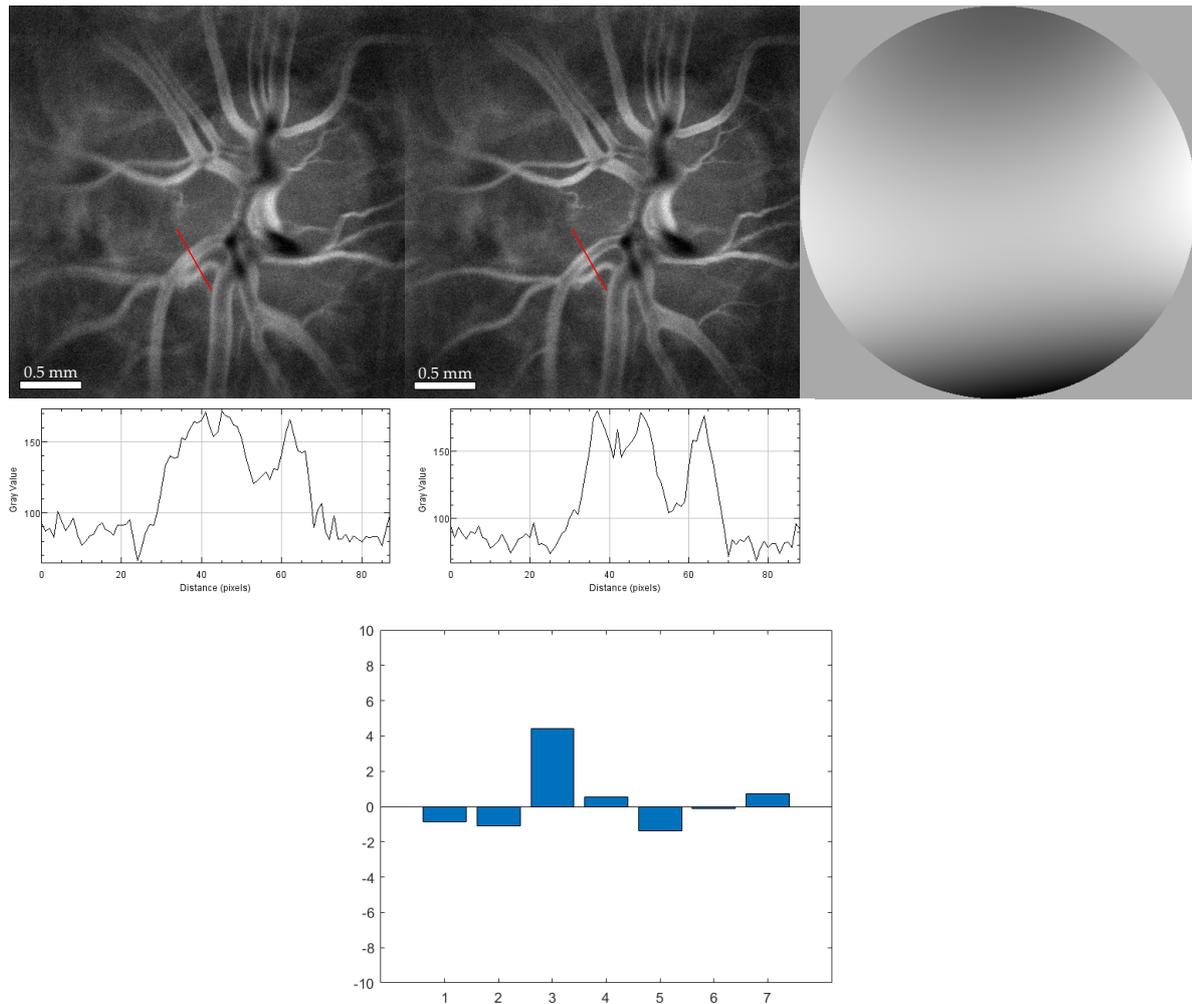
**Figure 3.19** – Correction of a laser Doppler image of a patient's retina. The first figure represents the image before aberration compensation. The gray-level profile corresponding to the red line is shown just below. The second one represents the corrected image. The gray-level profile corresponding to the red line is shown just below. The third figure illustrates the shape of the wavefront.

### 3.3.2 Results on patients

With this method, we have been able to improve resolution on several images of patients. The finest vessels are more accurate and sometimes even appear where there were none. The edges of the large vessels are more resolved. In the following of this part, several examples are presented to illustrate the results.

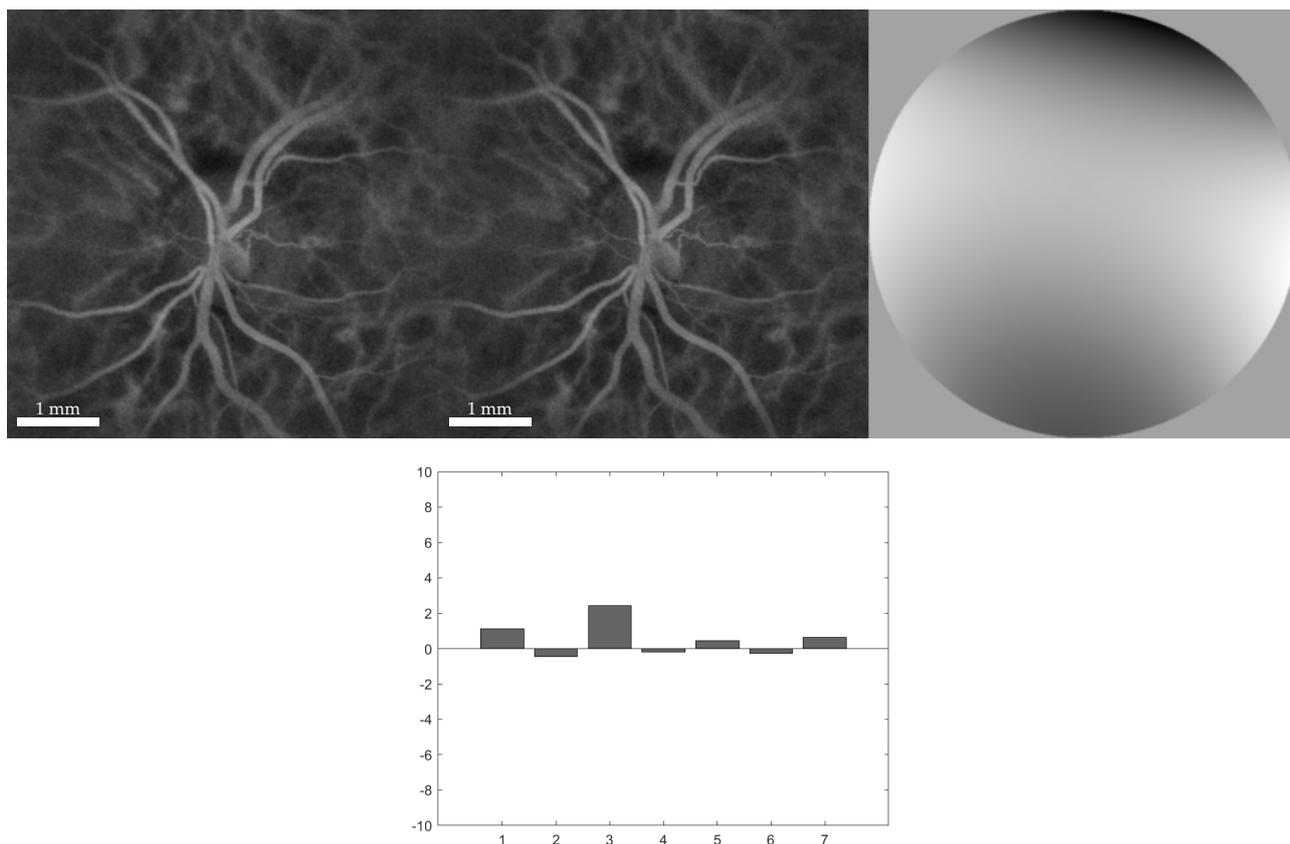
On the fig. 3.19, the algorithm corrects a defocus and an oblique astigmatism. It makes some vessels' edges be more accurate, as we can observe with the profiles: the correction improves the resolution of the two vessels by a factor of two. The algorithm estimates mainly a defocus and an oblique astigmatism.

On the fig. 3.20, the separation between the blood vessels is more pronounced. Besides, the larger vessel's walls are more defined: we can see two peaks at around 37 and 48 pixels of the distance scale, while it is more difficult on the first profile. The graph shows a vertical astigmatism of about 4 radians, which corresponds approximately to 2 diopters. Otherwise, the other modes are under 1 in absolute value (so not significant), except the number 5 corresponding to the vertical coma which is slightly above  $-2$  radians.



**Figure 3.20** – Correction of a laser Doppler image of a patient's retina. The first figure represents the image before aberration compensation. The gray-level profile corresponding to the red line is shown just below. The second one represents the corrected image. The gray-level profile corresponding to the red line is shown just below. The third figure illustrates the shape of the wavefront. The graph stands for the amplitude of the coefficient for Zernike modes, corresponding to this phase estimation. Amplitudes are expressed in radian and Zernike mode numbering starts with oblique astigmatism.

In the fig. 3.21, the field of view is more extended than the previous examples. The algorithm estimates mostly astigmatism (approximately 1 radian of oblique and 2 radians of vertical). This time, it has a slight impact on the image. This can be explained by the width of the field of view. For now, we apply the same correction for the whole reconstruction plane. However, some parts of the image do not belong to the reconstruction plane and may not have the same aberrations. It is well understandable by considering the defocus: the retina is in relief, then, the defocus of each part of the image can be different. Then, if we consider a large field of view, the estimation is less impacting. In the future, we will develop this algorithm with anisoplanatism.



**Figure 3.21** – Correction of a laser Doppler image of a patient's retina. The first figure represents the image before aberration compensation, while the second one represents the corrected image. The third figure illustrates the shape of the wavefront. The graph stands for the amplitude of the coefficient for Zernike modes, corresponding to this phase estimation. Amplitudes are expressed in radian and Zernike mode numbering starts with oblique astigmatism.

### 3.3.3 Conclusion and perspectives

We have been able to implement a digital wavefront sensor adapted to laser Doppler holography. First tests on patients are encouraging. It paves the way to the increase of image resolution, and may improve the understanding of some pathologies with a better observation of the vessels.

Several perspectives have been already evoked. We must better quantify the errors of measurement. For now, because we just get first results, we have not worked on it yet. We must also get more data on even more patients, and the knowledge of their aberrations could help us to compare to our algorithm's response. We could decrease the computation time by optimizing the code. For now, to get corrected images from a batch of 16 interferograms, it takes around 6 *min*. But this time is compressible by working on the structure of the algorithm. Finally, we must work on anisoplanatism in order to apply different compensations for each part of the image. It could improve the quality of our estimation.

This method is a step in the processing chain to produce high-resolution holographic images. An evolution of this system could be to include everything in a single process, to obtain high-resolution holographic images directly from interferograms. The aim would be to simplify and speed up the processing chain. With the rapid emergence of neural networks, this possibility does not seem so far away.



## 4.1 About the use of deep learning for holographic imaging and phase estimation

Beyond correcting the wavefront, we would consider using a neural network to replace the whole processing chain of holographic image rendering in the future. For now, however, we need to start studying the aberration compensation step, which we will focus on in this chapter.

Wavefront estimation by deep learning is getting more and more studied, as seen in the Chapter 1. Its interest comes, among other things, from the fact that the networks are able to find characteristics of the system that have not been noticed yet, by inducing empirically the resolution of the problem. Besides, it has a very low computation time once the network has been trained. This seems to be an interesting approach to consider, with a view to possibly boosting performance and find new features to our system.

The principle of deep neural networks [118–120] is conceptually simple, inspired by the human brain. The network learns a transformation to apply to input data in order to get the expecting result, by training itself on a lot of examples and learning from its errors. This highlights one of the major limitation of networks: they need to be trained on enough data representative of the system. Collecting huge amounts of clean data is not always easy, specially in biomedical field. Besides, a with large database, training can be long, which makes it difficult to set the large number of parameters required for the model.

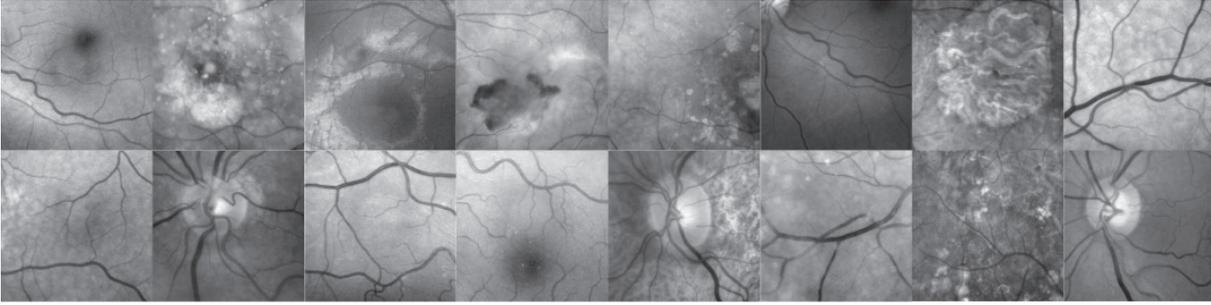
In this chapter, we first test the use of a network to correct aberrations. In a second part, we are getting slightly off topic by working on another aspect of the training, which could nevertheless be useful for aberration correction.

## 4.2 Correction of aberrated eye-fundus images

### 4.2.1 Context

The goal is to be able to correct aberrations on our holographic images. However, for now, we do not have enough diversified holographic data to train a network. Thus, we have decided to

work first on simulated data based on eye-fundus images.



**Figure 4.1** – Examples of images extracted from the database.

## 4.2.2 The database

To respect the constraints of deep neural networks, we must have a substantial database. We could have used an existing database like ImageNet. However, we wanted to be as close as possible to our objective which is the correction of aberrations on holographic retina images. Therefore, we have worked on a database of eye-fundus images, acquired with Spectralis device of the company Sanotek. For the training, we have gathered 28685 images that we have selected, cropped at the same size ( $256 \times 256$  for the first tests, then,  $512 \times 512$ ), centered and normalized between 0 and 1. Some images of the database are represented on fig. 4.1.

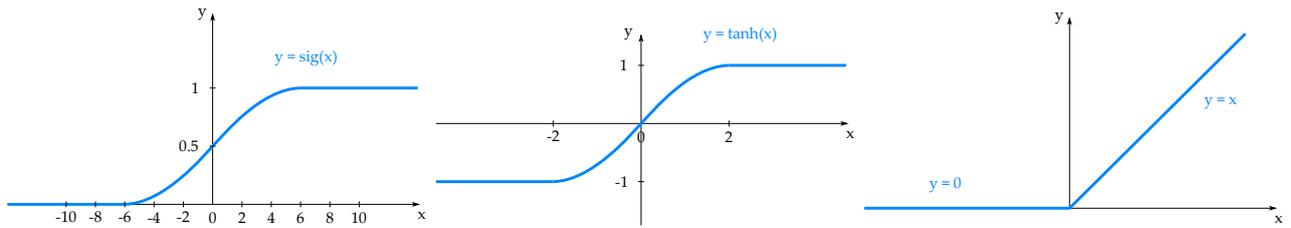
The images of the database are supposed to be in the retina plane, and the aberrations are considered in the Fourier plane of this object plane. Then, the input of the network is the aberrated image in the retina plane, and the ground truth is the image without any aberration.

## 4.2.3 The network

### Organization of a deep neural network

Neural networks that we note  $f$ , are composed of several layers of linear and non-linear functions. What is called a neuron is at the very least made up of one linear function (combination function) and one non-linear function (activation function). The combination function used to be a simple linear combination. For several years, convolutions have become widespread as combination functions thanks to the improvement of computation time and performances they provide: rather than treating each element of the input data separately, convolution processes together several inputs within its kernel. For example, if the input is an image, rather than processing pixel by pixel, the convolution uses clusters of pixels. It is considered to be a filter. The size of its kernel needs to be adapted to the size of the information: kernels that are too large may over-average the information contained in the clusters while smaller ones reduce the interest of convolution. As far as activation functions are concerned, they bring non-linearity to the network, and are often thresholding functions. Classical examples are sigmoid, hyperbolic tangent, Heaviside step function and rectified linear unit (ReLU). Some of these functions are represented on fig. 4.2.

Each layer of a network aims at expressing the input data into another space or under different dimensions. This is performed with undersampling and oversampling layers added to the neurons.



**Figure 4.2** – Some of the most used activation functions: sigmoid (sig), hyperbolic tangent (tanh) and rectified linear unit (ReLU). The derivative of sig and tanh is considered as non-zero only around 0, while the derivative of ReLU is non-zero on half of the real domain.

The network’s inputs  $\mathbf{X}$  are vectors, matrices or tensors, on which the layers are applied. If we note  $(f_1, f_2, \dots, f_N)$  the  $N$  layers of the network, the output of the network is:

$$\hat{\mathbf{Y}} = f(\mathbf{X}) = f_N(\dots f_2(f_1(\mathbf{X}))), \quad (4.1)$$

where each layer is composed of neurons, which in turn are composed of combination, activation, and possibly sampling functions. The output of the network  $\hat{\mathbf{Y}}$  has not necessarily the same size as the input.

### Choosing the most suitable network

To choose a network, it is necessary to keep in mind its aim, and the shape of the data in input and in output. The state of the art in deep learning is often divided into several categories depending on the networks function: classification, image segmentation, reconstruction, generation... But sometimes, some networks used for one task can be also used for another one. That is the case of the network that we have decided to use.

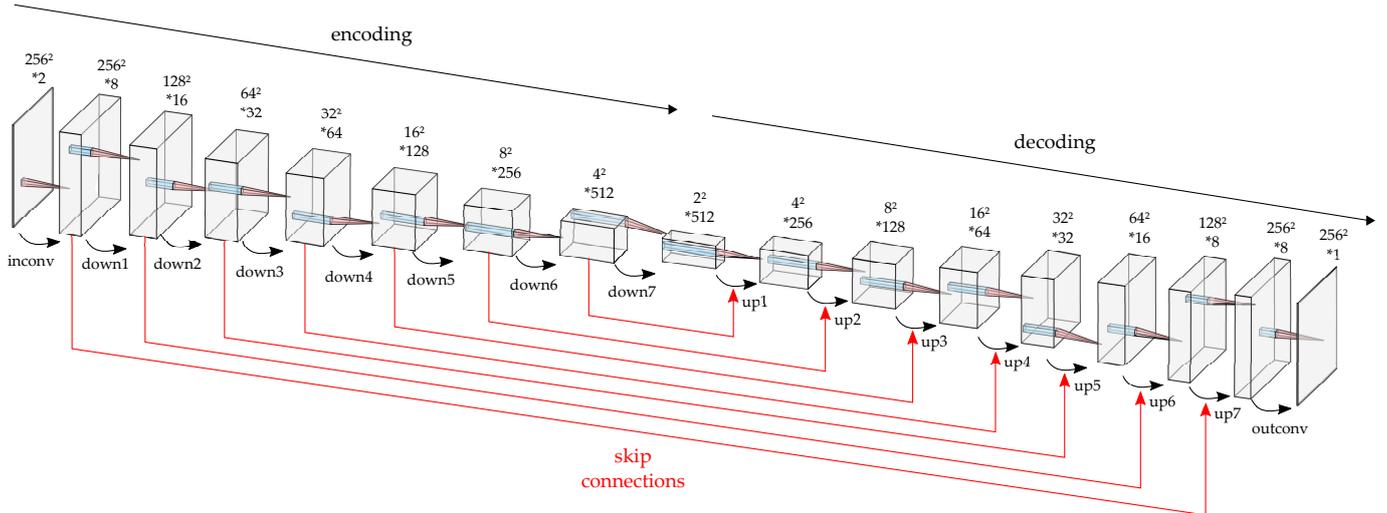
The task that we would like to perform is akin to image reconstruction, with images as input and output. As we have seen in the Chapter 1, several teams are working on phase recovery from speckle pattern [106–110]. Barbastathis team [108–110], whose work is very close to our case, is using a network derived from the UNet [111]. Therefore, we have decided to use a UNet to perform our tests.

### General structure of the UNet

The UNet is a convolutional neural network whose primary use is the segmentation of biomedical images. Since then, its use has been extended to fields other than segmentation, such as image reconstruction.

The UNet is composed of two parts: an encoder and a decoder, with what is called a “bottleneck” between them. In the encoding part, the resolution of the features (results of the blocks) decreases while their number of channels increases. In the decoding part, the resolution of the features increases while their number of channels decreases. The name of the network comes from the symmetry between these two parts facing each other.

Increasing the number of layers can be beneficial for solving more complex problems. However, when the network has too many layers, a phenomenon known as the vanishing gradient problem is more often seen. Each neural network’s weight is updated proportionally to the gradient of the error function, following the principle of the gradient-based optimization methods



**Figure 4.3** – Scheme of the UNet structure used for aberration compensation. The feature maps are represented by parallelepipeds, with their size above (*vertical*  $\times$  *horizontal*  $\times$  *channels*). The input is  $256 \times 256 \times 2$  image for real and imaginary part of the initial gray-scale image. The output is a  $256 \times 256$  gray-scale image. The blocks, which are composed of layers, are represented by the black arrows below the features. The network is composed of 1 convolution block and 4 downsampling blocks in the encoding part, and 1 convolution block and 4 upsampling blocks in the decoding part.

used in the backpropagation step of the learning (the algorithm is detailed in section 4.2.4). In some cases, the gradient becomes smaller and smaller and even vanishes, which prevents the weights from updating, and stops the learning. The UNet has two elements to combat gradient vanishing.

First, the UNet uses ReLU as activation function of each layer. ReLU is defined as  $y = \max(0, x)$  (fig. 4.2). Comparing to other activation functions like sigmoid or hyperbolic tangent, the derivative of ReLU is non-zero over a much larger area: for the sigmoid and the hyperbolic tangent, their derivative is near 0 everywhere in the real domain, except around 0, while the derivative of ReLU is non-zero for half of the real domain. Then, it propagates the gradient during backpropagation step on a much larger area, which limits the vanishing gradient effect.

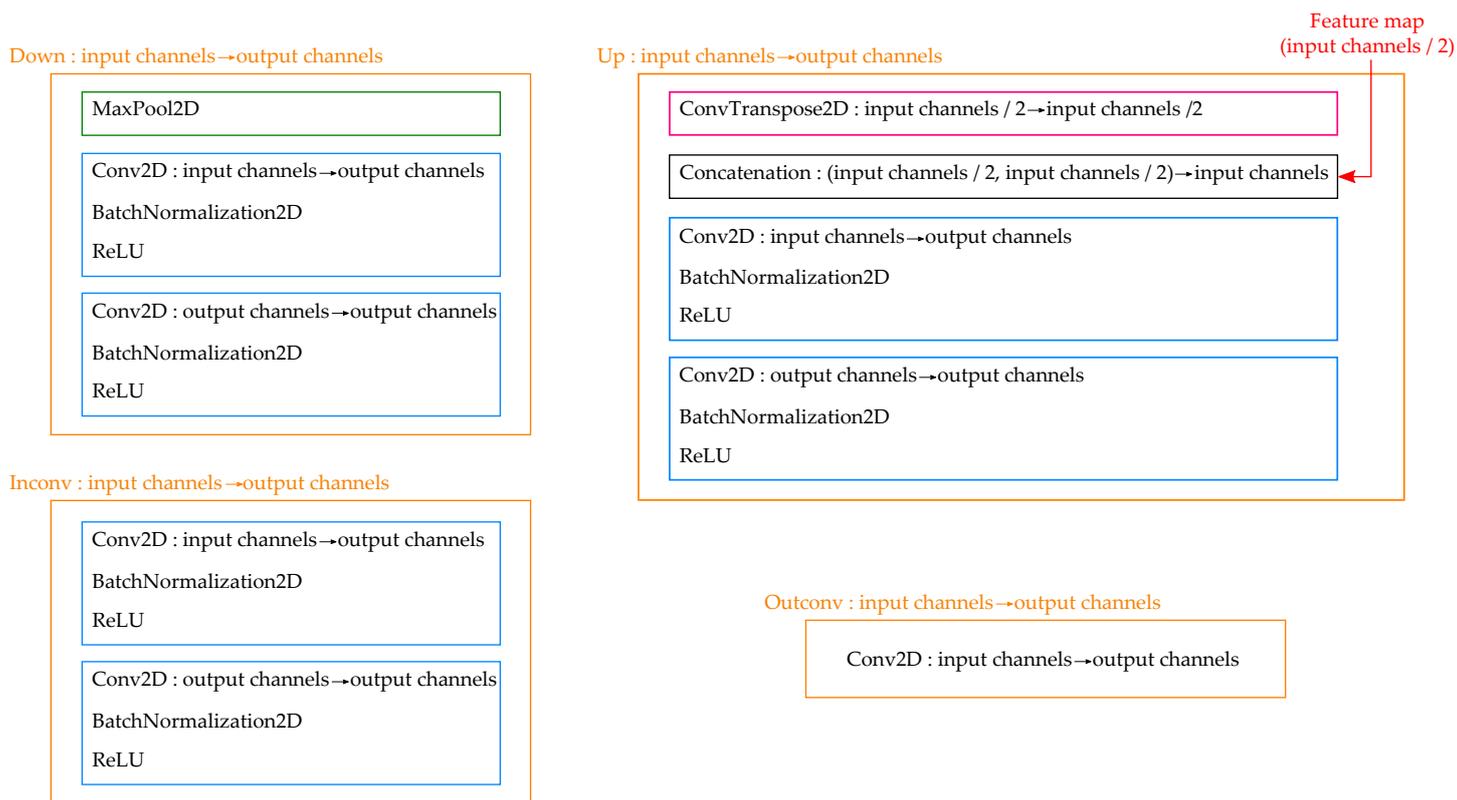
Secondly, the features of the encoding part are propagated in the decoding part through skip connections: the features of the downsampling part are concatenated to the features of the upsampling part. It allows the missing context to be added to the second part, and improve the propagation of the gradient during the backpropagation step.

### Organization of the blocks

The architectures of the UNet can differ from one problem to another: several layers can be added or removed, their size can be changed, as well as their parameters. The configuration that we use is illustrated in the fig. 4.3, and the composition of each block is detailed in the fig. 4.4. Let us note  $x_{in} \times y_{in} \times ch_{in}$  and  $x_{out} \times y_{out} \times ch_{out}$  the dimensions of the input and output features of a block respectively.

Each convolutional layer is a sequence of 2D convolution, 2D batch normalization and ReLU.

The 2D convolution is the basic module of convolutional neural networks. In practice, convolutions are filters applied to the input feature maps. In our UNet,  $ch_{out}$  filters of size  $3 \times 3 \times ch_{in}$



**Figure 4.4** – Composition of each block of the UNet (fig. 4.3): down, up, inconv and outconv. Input channels and output channels are arguments of each block, representing the number of channel of the input and the output feature maps. Down block is composed of one MaxPooling and two convolution layers. Up block is composed of one ConvTranspose and two convolution layers. The input feature map of the convolution layers is the concatenation of the output of the ConvTranspose and the feature map of the corresponding skip connection. Inconv block is composed of two convolution layers. Outconv is only one 2D convolution.

are applied (it can also be considered as a kernel of size  $3 \times 3 \times ch_{in} \times ch_{out}$ ), with no stride (the kernel is moved from pixel to pixel). The parameters  $ch_{in}$  and  $ch_{out}$  are arguments of the convolution we choose. A zero-padding is used to keep  $x_{in} = x_{out}$  and  $y_{in} = y_{out}$ : the input is surrounded by one layer of zeros for the first two dimensions  $x$  and  $y$ . Then, applying the 2D convolution layer on  $x_{in} \times y_{in} \times ch_{in}$  features, we get  $x_{in} \times y_{in} \times ch_{out}$  by using  $ch_{out}$  filters of size  $3 \times 3 \times ch_{in}$ .

As it has been explained previously, the activation function used in the UNet is ReLU. The 2D batch normalization, as the name suggests, performs the normalization of each batch between blocks, allowing the use of higher learning rates, without taking too much care of the initialization of the data. It also adds some regularization to the network to struggle against overfitting [121].

The first block of transformations is noted *inconv*. It is composed of two convolutional layers. The arguments of the first convolution are  $ch_{in} = 2$  and  $ch_{out} = 8$ . The arguments of the second convolution are  $ch_{in} = 8$  and  $ch_{out} = 8$ . Then, from  $256 \times 256 \times 2$ , we get a  $256 \times 256 \times 8$  feature by using a  $3 \times 3 \times 2 \times 8$  kernel for the first convolution, and  $3 \times 3 \times 8 \times 8$  for the second convolution.

After *inconv*, the first part of the network is a sequence of 7 *down* blocks. A *down* block is composed of one MaxPool2D and two convolutional layers. The MaxPool2D is a downsampling

function. It divides the dimensions of the input feature map by taking the maximum value over a moving patch of pixels in  $(x, y)$  dimensions. In our UNet, the size of this patch is  $2 \times 2$  (with no stride, no padding). We choose the output channel size of the *down* block to be the double of the input channel size ( $ch_{out} = 2 * ch_{in}$ ) except for the last *down* block, where  $ch_{out} = ch_{in}$ . Then, if the size of the input of a *down* block is  $x_{in} \times y_{in} \times ch_{in}$ , the output size is  $x_{in}/2 \times y_{in}/2 \times 2 * ch_{in}$  ( $x_{in}/2 \times y_{in}/2 \times ch_{in}$  after MaxPool2D,  $x_{in}/2 \times y_{in}/2 \times 2 * ch_{in}$  after the first convolution and still  $x_{in}/2 \times y_{in}/2 \times 2 * ch_{in}$  after the second convolution).

After the encoding part, the decoding part begins, including 7 *up* blocks. Each is composed of one ConvTranspose2D and two convolutional layers. The ConvTranspose2D is an upsampling function. It doubles the dimensions of the input feature map with a bilinear interpolation over a moving patch of pixels in  $(x, y)$  dimensions. This is a weighted upsampling: the weights are learned by the network. Thus, it has a kernel of weights, like a 2D convolution. In our network, the size of this kernel is  $2 \times 2 \times ch_{in} \times ch_{out}$ , applied with a stride of 2. We choose the output channel size of the *up* block to be half of the input channel size ( $ch_{out} = ch_{in}/2$ ) except for the last *down* block, where  $ch_{out} = ch_{in}$ . Then, if the size of the input of a *up* block is  $x_{in} \times y_{in} \times ch_{in}$ , the output size is  $2 * x_{in} \times 2 * y_{in} \times ch_{in}/2$ .

The particularity of this part is the concatenation of previous features with skip connections. Therefore, the first argument of *up* blocks (input channel size) is twice the channel size of the input feature map: its dimensions are  $x_{in} \times y_{in} \times ch_{in}/2$  before ConvTranspose2D and  $2 * x_{in} \times 2 * y_{in} \times ch_{in}/2$  after. With the concatenation, the feature map is concatenated (over the channel dimension) with a previous one with the same dimensions: its dimensions are  $x_{in} \times y_{in} \times ch_{in}$  after concatenation. Then, the two convolutions are applied which provide an output feature map of dimension  $2 * x_{in} \times 2 * y_{in} \times ch_{in}/2$ .

As the *up* block is more complex, here is an example of the dimensions of a feature, at each step from before to after the block *up1*. The dimensions of the feature before this block are  $2 \times 2 \times 512$ . The first step of ConvTranspose2D transforms the feature into another one of dimensions  $4 \times 4 \times 512$ . Then, it is concatenated to the output of the block *down6*, of shape  $4 \times 4 \times 512$ , with respect to the last dimension. The result is a  $4 \times 4 \times 1024$  feature. Then a first 2D convolution is applied, and we obtained a  $4 \times 4 \times 256$  feature. The final result after the second convolution is a  $4 \times 4 \times 256$  feature.

Finally, the last block is *outconv*, which is simply a 2D convolution, reducing the number of channels from 8 to 1, to get the final gray-scale image.

This network accumulates a total of 14 767 768 weights to learn: almost 15 billions of degrees of freedom, which illustrates pretty well the complexity of the system, and then, the difficulty of the optimisation problem. This network used to compute the results of the section 4.2.5 has been implemented with the Pytorch library of Python. Despite its impressive size, the network delivers on its promise of speed, with a computation time of around 30 *ms* to produce an output image from an input image.

The network of Barbastathis team is based on a UNet. The only difference is the use of dense connections, linking the features together within each block [122]. This is a way to propagate easier the gradient during the backpropagation step, but also increases the complexity of the network. Thus, we have decided to start with a “simple” UNet.

In the following, we explain how we have trained this network and give some details about several parameters.

### 4.2.4 The training step

Once the network has been implemented, the training can begin. For the supervised networks that we will deal with in this chapter, the output of the network  $\hat{\mathbf{Y}} = f(\mathbf{X})$  is compared to the ground truth  $\mathbf{Y}$  with a loss function  $L(\hat{\mathbf{Y}}, \mathbf{Y})$ . The problem is to find the weights that minimize the error function  $L(\hat{\mathbf{Y}}, \mathbf{Y})$ , so that the output of the network  $\hat{\mathbf{Y}}$  is as close as possible to the ground truth  $\mathbf{Y}$ . The gradient of this error is backpropagated through the network by an optimization algorithm (gradient descent type) to modify the weights that govern the combination transformations within the network. The networks are thus multi-scale optimization algorithms, trained with multiple iterations on the database to modify the weights to fit the transformation to the system.

The weights of each layer of the network are initialized according to a uniform distribution described in [123]. It is directly implemented on the modules in Pytorch (Conv2D, ConvTranspose2D...) and we have not changed these default parameters. For our training, the weights are updated following the stochastic gradient descent algorithm, detailed in the next part.

The training is launched on a batch of images  $B$  randomly extracted from the database. The length of the batch depends on the performance of the CPU/GPU. Of course, with larger batches, the training is faster because all the processes are performed in parallel on the images of the batches. However, the CPU/GPU RAM (Random Access Memory) is not infinite and batch length need to be adapted, also depending on the input images size, to avoid computer memory problem. Besides, parallelization on GPU instead of CPU increases drastically the speed of data processing.

Then, the batch of data is sent through the network (direct way), and the error over this batch is calculated. We have decided to work with the mean-square error (MSE) as loss function on the batch of data  $B$ :

$$L_B(\hat{\mathbf{Y}}, \mathbf{Y}) = E \left[ (\hat{\mathbf{Y}} - \mathbf{Y})^2 \right], \quad (4.2)$$

where  $E$  designates the expectation. This error is backpropagated through the network, updating the weights by using the stochastic gradient descent algorithm. Once all random batches have been sent to train the network, the mean error corresponding to one iteration over the database (called epoch) is computed. The network is trained during several epochs, and corresponding mean errors over the training database are computed to check that training is running properly.

In the following, we explain the mechanisms of the stochastic gradient descent during the backpropagation, which is very important to understand the update of the weights. Next, we mention several problems that can occur during training.

#### Stochastic gradient algorithm

First, here are some notations. Let us consider a network with  $N$  layers. We will use only linear combination as combination function rather than convolution. It does not change the aim of the explanation, it is only to simplify calculations. Let us note  $\mathbf{X}^{(n)}$  the network's feature map at the  $n^{th}$  layer. For example,  $\mathbf{X}^{(0)}$  is the input, and  $\mathbf{X}^{(N)}$  is the output ( $\hat{\mathbf{Y}} = \mathbf{X}^{(N)}$ ). To simplify the computation, we will consider only  $\mathbf{X}$  as a vector, but it is also generalizable to

matrices or tensors. For the following and to make notations clearer, we will note the matrices in bold and capital letters, the vectors in bold and lower case letters, and the real values in lower case letters.

We note  $\mathbf{z}^{(n)} = \mathbf{W}^{(n)}\mathbf{x}^{(n)}$  the result of the linear combination of the  $n^{\text{th}}$  layer, with  $\mathbf{W}^{(n)}$  the weights at the same layer. We suppose that the first component  $x_0$  of  $\mathbf{x}$  is equal to 1, in order to have a bias. Weights are the parameters the network has to learn. We can also write without vectors:  $z_j^{(n)} = \sum_i w_{ij}^{(n)} x_i^{(n)}$ . We note  $g$  the activation function such as  $g^{(n)}(\mathbf{z}^{(n)}) = \mathbf{x}^{(n+1)}$ . Without vector notation:  $x_j^{(n+1)} = g^{(n)}(z_j^{(n)})$ .

Similarly, we can write the loss function:

$$L_B(\hat{\mathbf{y}}, \mathbf{y}) = L_B(\mathbf{x}^{(N)}, \mathbf{y}) = E \left[ (\mathbf{x}^{(N)} - \mathbf{y})^2 \right]. \quad (4.3)$$

The aim of the stochastic gradient descent is to update the weights of the network by propagating this loss  $L_B$ . The parameters will be updated with the partial derivative of the loss with respect to the parameters of the concerned layer:

$$w_{ij}^{(n)} \leftarrow w_{ij}^{(n)} - \lambda \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial w_{ij}^{(n)}}, \quad (4.4)$$

where  $\lambda \in \mathbb{R}$  is the learning rate.

Let us compute the error we will use for the update of the last layer corresponding to  $w_{ij}^{(N-1)}$ . The partial derivative of interest can be written as:

$$\frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial w_{ij}^{(N-1)}} = \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial z_j^{(N-1)}} \frac{\partial z_j^{(N-1)}}{\partial w_{ij}^{(N-1)}} = \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial z_j^{(N-1)}} x_i^{(N-1)} = \delta_j^{(N-1)} x_i^{(N-1)}, \quad (4.5)$$

with the previous definitions, and with  $\delta_j^{(N-1)} = \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial z_j^{(N-1)}}$ .

We can explicit  $\delta_j^{(N-1)}$ :

$$\begin{aligned} \delta_j^{(N-1)} &= \frac{\partial}{\partial z_j^{(N-1)}} \left[ \frac{1}{2M} \sum_{i=0}^{M-1} (x_j^{(N)} - y_i)^2 \right] \\ &= \frac{x_j^{(N)} - y_j}{M} \frac{\partial g^{(N-1)}(z_j^{(N-1)})}{\partial z_j^{(N-1)}} \\ &= \frac{x_j^{(N)} - y_j}{M} g'^{(N-1)}(z_j^{(N-1)}), \end{aligned} \quad (4.6)$$

where  $M$  is the size of  $\mathbf{x}^{(N)}$  and  $\mathbf{y}$ , and  $g'$  is the derivative of  $g$ .

Therefore, the weights of this first layer can be updated:

$$w_{ij}^{(N-1)} \leftarrow w_{ij}^{(N-1)} - \lambda \delta_j^{(N-1)} x_i^{(N-1)}. \quad (4.7)$$

Now let us see for the next layer ( $N - 2$ ).

$$\frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial w_{ij}^{(N-2)}} = \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial z_j^{(N-2)}} \frac{\partial z_j^{(N-2)}}{\partial w_{ij}^{(N-2)}} = \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial z_j^{(N-2)}} x_i^{(N-2)} = \delta_j^{(N-2)} x_i^{(N-2)}. \quad (4.8)$$

We can decompose  $\delta_j^{(N-2)}$ :

$$\delta_j^{(N-2)} = \sum_k \frac{\partial L_B(\hat{\mathbf{y}}, \mathbf{y})}{\partial z_k^{(N-1)}} \frac{\partial z_k^{(N-1)}}{\partial z_j^{(N-2)}}. \quad (4.9)$$

The second factor in the sum of eq. (4.9) can be calculated with the definition of  $z$ :

$$\frac{\partial z_k^{(N-1)}}{\partial z_j^{(N-2)}} = w_{jk}^{(N-1)} g'^{(N-2)}(z_j^{(N-2)}). \quad (4.10)$$

We can notice the first factor is simply  $\delta_k^{(N-1)}$ . Then, eq. (4.9) is:

$$\delta_j^{(N-2)} = g'^{(N-2)}(z_j^{(N-2)}) \sum_k \delta_k^{(N-1)} w_{jk}^{(N-1)}. \quad (4.11)$$

We can go on iteratively. Thus, the update of the weights for the layers from  $n = N - 1$  to  $n = 0$  is:

$$w_{ij}^{(n)} \leftarrow w_{ij}^{(n)} - \lambda \delta_j^{(n)} x_i^{(n)}, \quad (4.12)$$

where, for  $n = N - 1$ :

$$\delta_j^{(N-1)} = \frac{x_j^{(N)} - y_j}{M} g'^{(N-1)}(z_j^{(N-1)}), \quad (4.13)$$

and for  $n \in [0, N - 2]$ :

$$\delta_j^{(n)} = g'^{(n)}(z_j^{(n)}) \sum_k \delta_k^{(n+1)} w_{jk}^{(n+1)}. \quad (4.14)$$

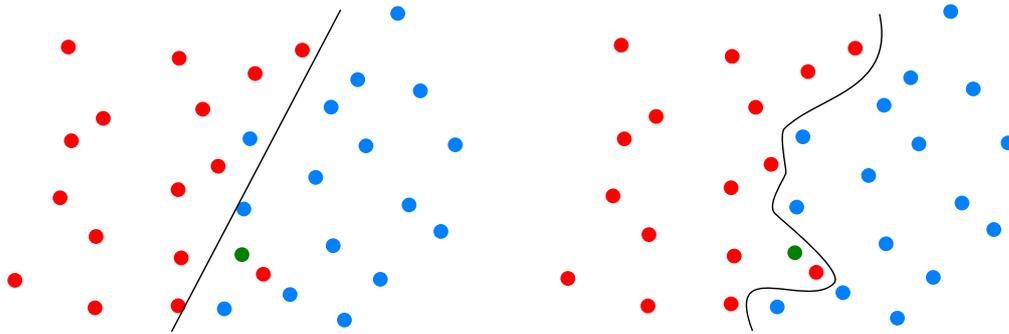
In our case, all the activation functions are the same (ReLU), so  $g^{(n)} = g$ . The derivative of  $g$  appears in the expressions of  $\delta_j^{(N)}$  and  $\delta_j^{(n)}$ . Here, the phenomenon of disappearance of the gradient discussed in section 4.2.3 is well visible: if  $g'(z_j^{(n)})$  is close to zero,  $\delta_j^{(n)}$  is close to zero too. Then, the weights of the corresponding layer  $n$  are not updated. Moreover, if  $\delta_j^{(n)}$  is close to zero, the  $(n - 1)^{th}$  layer is also impacted, because one of the term in its sum disappears (see eq. (4.14)), which causes a progressive fading of the gradient from layer to layer. Therefore, the selection of an appropriate activation function is essential to reduce this effect.

### Difficulties in training

Apart from gradient vanishing, several other problems can happen during the training.

Monitoring the curve of the loss on the training database over the epochs helps identifying problems. Indeed, the network learns when the loss decreases on average (see example on fig. 4.6). First, a compromise has to be found to set the value of the learning rate. If the value is too high, the loss will skyrocket. If it is too low, the curve of the loss will decrease very slowly, almost always stagnating, which means the network does not learn a lot. Of course, when the curve stagnates after falling, it means that the learning process is over. For our case, we have used a learning rate of 0.1 to obtain the results of the section 4.2.5.

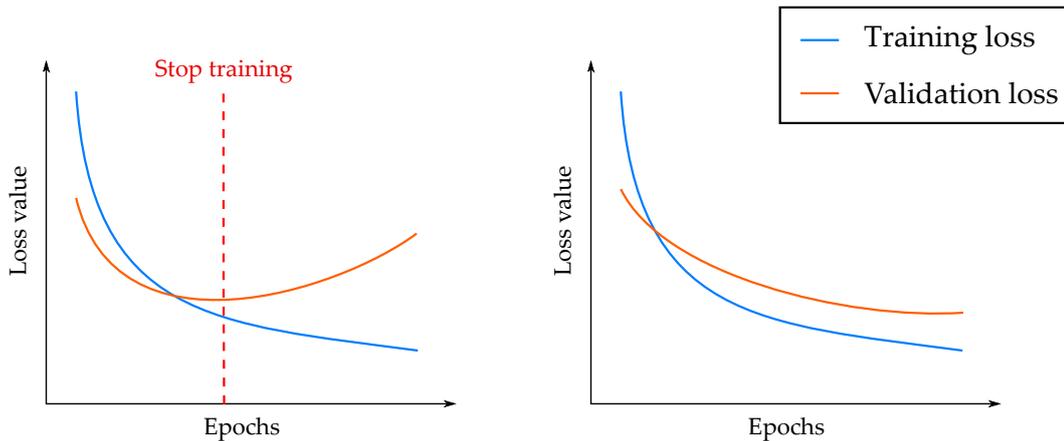
One of the main problem which can occur in the training is overfitting. Overfitting appears when the network learns “too well” from the training: it fits exactly to the training database, and cannot generalize at all when a new data is presented to it. It is illustrated by the simple example of the fig. 4.5. In this example, in order to decrease again and again the training error, the algorithm fits the solution too well to the training data. Therefore, the training error in the second image is lower than the first one, but it cannot handle the new points as well. The principle is exactly the same for more complex systems as the network.



**Figure 4.5** – Illustration of overfitting. The sets of points are divided into two classes: red and blue. Learning from these sets, an algorithm compute the limit between the two sets. After two or three iterations, the algorithm computes the limit of the image on the left. After more and more iterations, it gives the limit of the second image. The green point is an additional data, not used for the algorithm’s learning, and which belongs to the blue points. The first state of the algorithm can well classify it, when the second one makes a mistake.

A simple way to verify that overfitting does not occur is to use a validation dataset. Usually, the initial database is divided into two parts in order to do this test: 90 – 85 – 80% of the database for training, and 10 – 15 – 20% for validation. We have divided our database of 28685 images into 24383 images for training (85% of the initial database) and 4302 images for validation (15% of the initial database). The training database is used to modify the weights (direct propagation and backpropagation in the network), while the validation data is just used to test the network (only direct propagation) and does not affect the weights. The losses for training and validation datasets are monitored epoch by epoch. If the validation loss begins to increase in average, it means the network begins to fit too much to the training data: the training must be stopped (see fig. 4.6).

It exists several ways to fight against overfitting. First and quite intuitively, the smaller the database, the faster the network begins to fit to the data. Then, gathering huge databases prevent from overfitting. In our case, with the 24383 images used for training, the validation loss has not increased during the training. Other ways to counter overfitting, that we will



**Figure 4.6** – Losses during training and validation. The first figure illustrates an example when there is overfitting: the validation loss increases from the red dotted line while the training loss goes on decreasing. The training must be stopped at the dotted line level. The second figure illustrates an example where the training is going well: both losses are decreasing.

not detail here, are data augmentation (applying geometric transformations to the database to increase its size), regularization (applying a regularization on the weights or the loss function to constraint the possible responses of the network) and dropout (randomly ignoring some paths of the network during the training). We did not need to use them for the results presented in section 4.2.5.

The last important parameter we will discuss here is the momentum. Sometimes, the loss falls within local minima. In order to get out of it, a momentum  $\beta$  is added, which is a moving average of the gradient:

$$(\delta_j^{(n)} x_i^{(n)})^t \leftarrow \beta (\delta_j^{(n)} x_i^{(n)})^{t-1} + (1 - \beta) (\delta_j^{(n)} x_i^{(n)})^t, \quad (4.15)$$

where  $(\delta_j^{(n)} x_i^{(n)})^t$  means  $\delta_j^{(n)} x_i^{(n)}$  computed at the time  $t$  ( $t^{\text{th}}$  iteration). We have settled the momentum to 0.9.

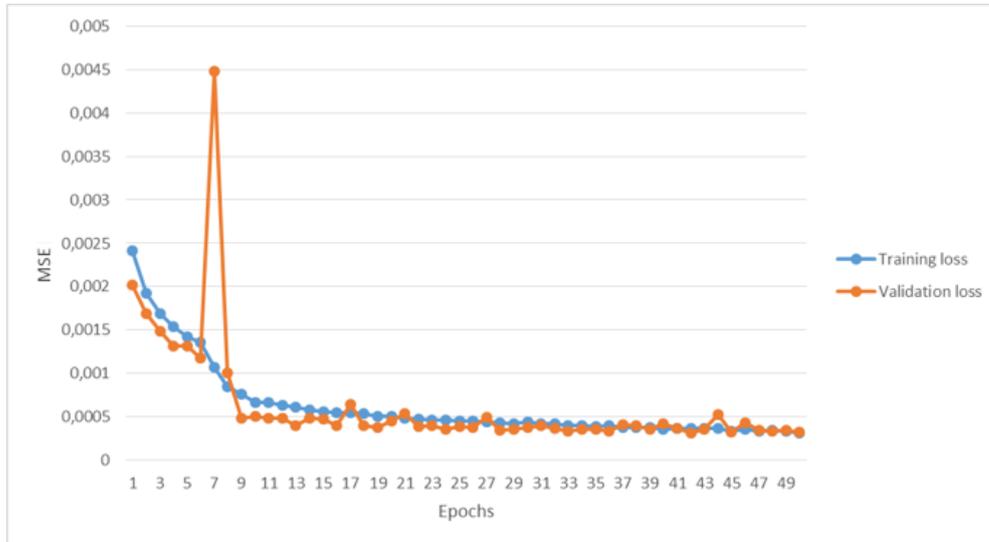
What makes difficult the training is the amount of parameters to set, for the training and the network. Besides, with large dataset like ours, the training can be long (on the order of a day) even though it is optimized on GPU (we set the batch size to 16 for  $256 \times 256$  images). The results of the next section have been obtained with the network and the parameters discussed in the previous parts. These are preliminary results that pave the way for further study.

## 4.2.5 Results

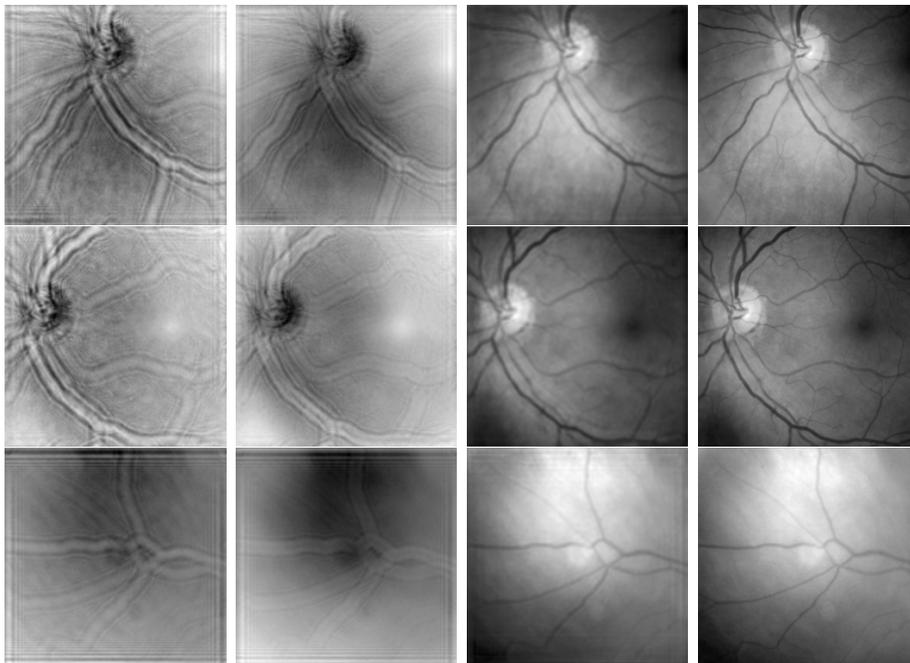
As seen in the previous sections, the input is two  $256 \times 256$  gray-scale images. They are real and imaginary parts of the aberrated image. As a reminder, the basic images composing the databases are intensity images of the retina to which aberrations have been added. Therefore, they have real and imaginary parts. This was our starting point.

We began with a simple case at first, applying a single aberration (defocus, the 5<sup>th</sup> Zernike mode, with a fixed coefficient) to all the training images. The loss functions corresponding to this training are represented on fig. 4.7. One value of the validation loss is abnormally high, which may be due to high errors for some images of the database. However, both losses decrease in average.

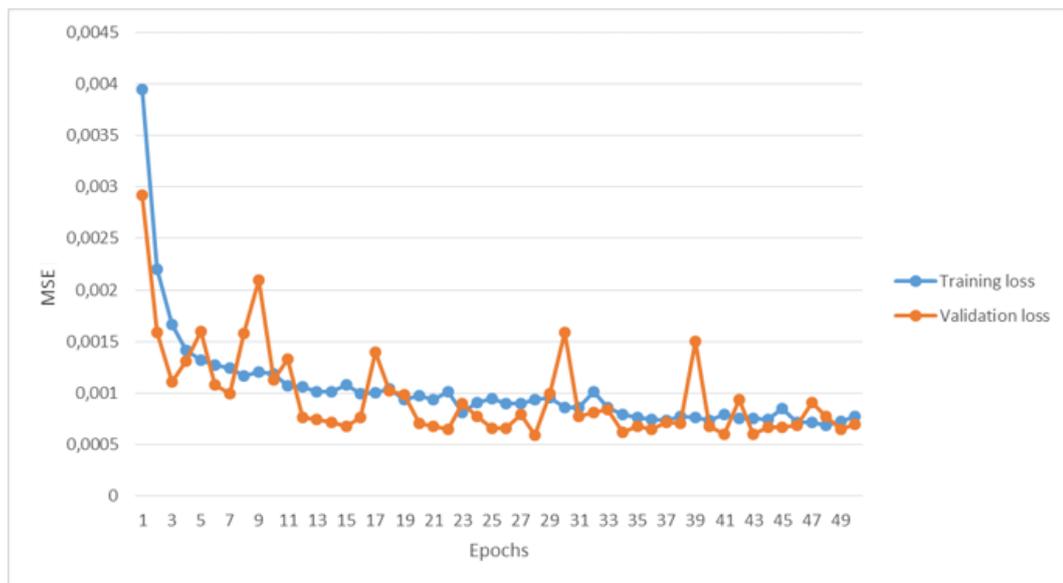
This training gave good results on the test images, to which the same aberration has been applied. An example is shown on three of the test images on fig. 4.8. The resolution of the U-Net reconstruction is slightly lower than the reference image, but the result remains very close to the reference and most of the blood vessels are well represented. The MSE between the output of the network and the reference (computed on normalized images, whose values are between 0 and 1) is equal to  $5.0 \times 10^{-4}$  for the first case,  $5.1 \times 10^{-4}$  for the second case and  $5.6 \times 10^{-4}$  for the last case.



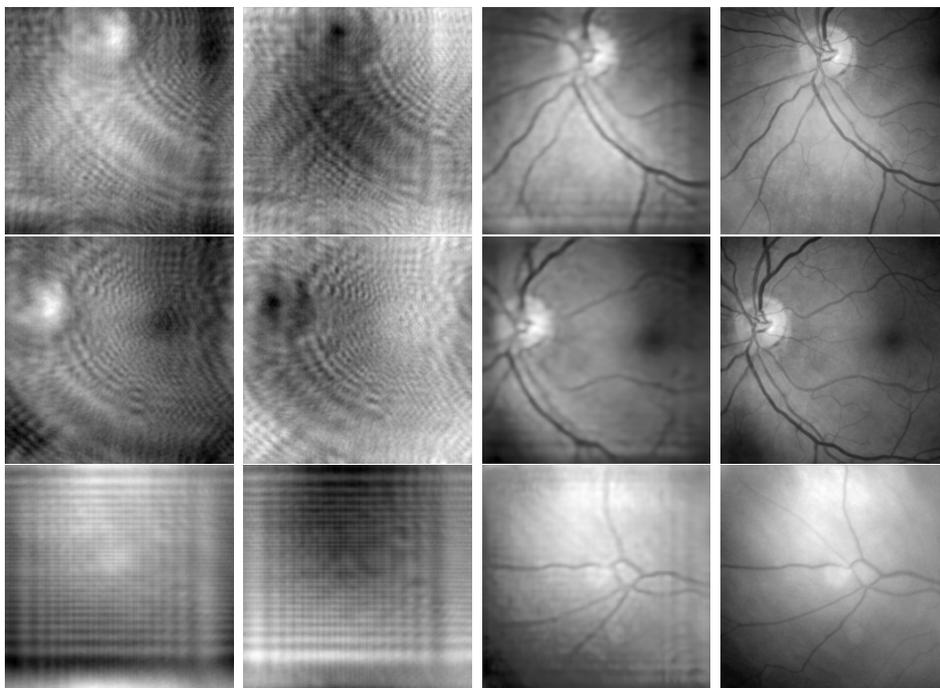
**Figure 4.7** – Loss functions over the number of epochs, for training and for validation with one defocus.



**Figure 4.8** – Result of the network on three  $256 \times 256$  test images for a defocus. First input of the network, real part of the aberrated image (first column), second input of the network, imaginary part of the aberrated image (second column), output of the network, corrected image (third column), and reference image (fourth column). The MSE between the images of the third and fourth columns is  $5.0 \times 10^{-4}$  for the first row,  $5.1 \times 10^{-4}$  for the second row and  $5.6 \times 10^{-4}$  for the last row.



**Figure 4.9** – Loss functions over the number of epochs, for training and for validation. A stronger defocus has been applied to all the database images.

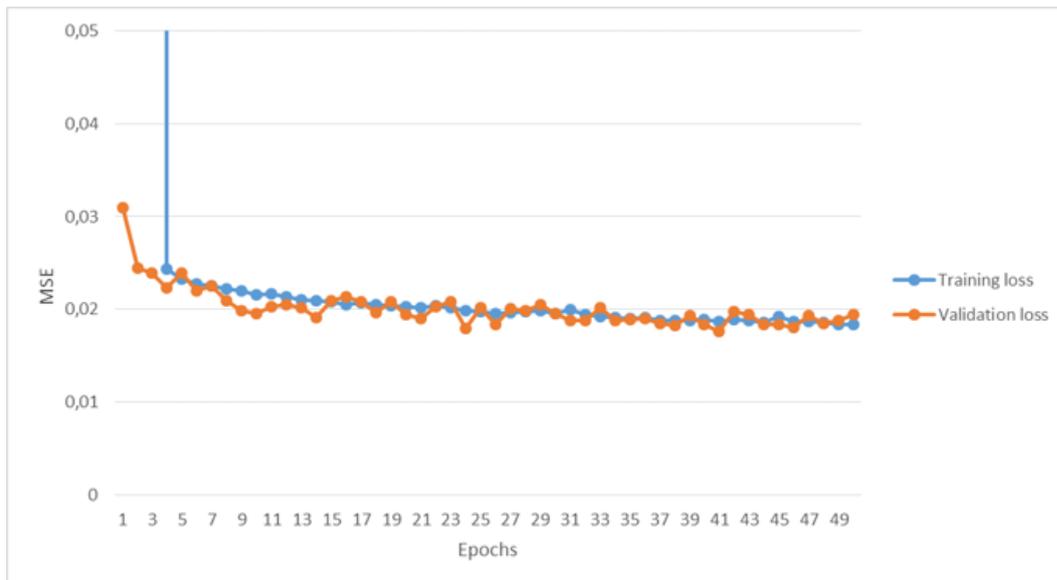


**Figure 4.10** – Result of the network on three  $256 \times 256$  test images for a very high defocus. First input of the network, real part of the aberrated image (first column), second input of the network, imaginary part of the aberrated image (second column), output of the network, corrected image (third column) and reference image (fourth column). The MSE between the images of the third and fourth columns is  $2.6 \times 10^{-3}$  for the first row,  $2.3 \times 10^{-3}$  for the second row and  $2.3 \times 10^{-3}$  for the last row.

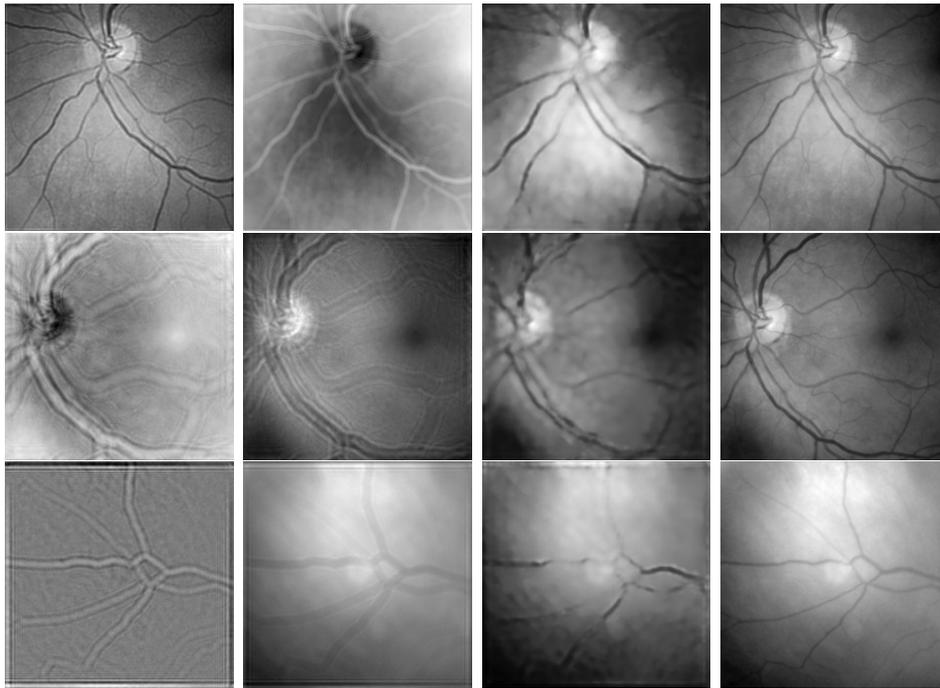
For a higher defocus applied to the training and test bases, the reconstruction is done, but the resolution is lowered. The losses are represented on fig. 4.9 and the fig. 4.10 gives an example of the correction on the same images than the previous case. Both losses decrease in average. However, for last epochs, they are higher than the losses of the previous case (fig. 4.7), which shows that the results are worse than before. The MSE calculated on the images of the fig. 4.10 are also higher than previously, by a factor 10:  $2.6 \times 10^{-3}$  for the first case,  $2.3 \times 10^{-3}$  for the second case and  $2.3 \times 10^{-3}$  for the last case. The results seem encouraging for the future though: this was a test and the amplitude of aberration applied here is about 1000 times stronger than the aberrations we can find in human eye. And even with this strong defocus, the network is able to give some excellent results.

We performed the same experiments by training the network with astigmatisms (one by one) of one amplitude and the same kind of results have been observed. We then decided to study the results of the correction on images on which we had applied the same mode of aberration but at different amplitudes than those learned. To do this, we randomly picked amplitudes over a bounded interval following a uniform law, which we applied to the same aberration (defocus). Thus, the defocus of the images varied from one image to another, for training and testing databases.

The losses are represented on fig. 4.11. Both losses rapidly decrease at the beginning, and then, much more slowly, reaching a value just below  $2 \times 10^{-2}$ , which is way higher than the previous cases of training. Therefore, lower resolution results are expected. Indeed, the results on the same model images (fig. 4.12) show an estimation with a low accuracy. The computed MSE is equal to  $7.9 \times 10^{-3}$  for the first case,  $2.7 \times 10^{-3}$  for the second case and  $1.3 \times 10^{-2}$  for the last case. The network did not learn enough cases for each of the randomly drawn amplitudes, and therefore, provides an approximate answer, based on what it thinks is most likely.



**Figure 4.11** – Loss functions over the number of epochs, for training and for validation. A random defocus has been applied to all the database images.



**Figure 4.12** – Result of the network on three  $256 \times 256$  test images for a defocus. The amplitudes of the defocus were randomly drawn over a bounded interval following a uniform law. First input of the network, real part of the aberrated image (first column), second input of the network, imaginary part of the aberrated image (second column), output of the network, corrected image (third column) and reference image (fourth column). The MSE between the images of the third and fourth columns is  $7.9 \times 10^{-3}$  for the first row,  $2.7 \times 10^{-3}$  for the second row and  $1.3 \times 10^{-2}$  for the last row.

After several tests, we have decided to work only with the intensity image in input (one channel instead of two for the input data, and appropriate settings for *inconv*). Indeed, after our first results, we noticed that since the original image had no phase, writing with a real part and an imaginary part was not very useful. However, it proves it can work and we will be able to use real and imaginary parts on our digital holograms in the future.

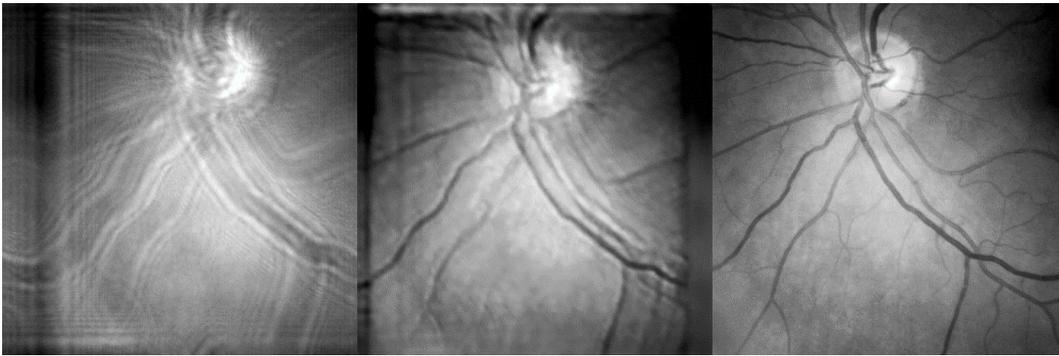
We also tried to extend the size of the input images. We tried successfully on  $512 \times 512$  images, by doing slight changes of blocks size. The network used is represented on the fig. 4.15 of the section 4.4. The results on a test image are shown on fig. 4.13. Here, like before, the network has been trained only on one amplitude defocus, with  $512 \times 512 \times 1$  gray-scale intensity images. The results are similar to those obtained with  $256 \times 256 \times 2$  images: for a not high defocus, the network reconstruction is satisfying.

The last test we have performed in this thematic has been to train the network with only one aberration, and to test on images with an aberration which “fluctuates” around the learned aberration. We used the data of [67], regrouping the aberrated phase  $\phi_a(t)$  (which contains several modes of aberration) of 50 people over time. We have chosen one eye among the 50, and applied the average over time  $\langle \phi_a \rangle$  of its phase to all the images of our training database. Then, we have applied successively all the  $\phi_a(t)$  to the images of the testing database. The fig. 4.14 shows one image from the testing database, on which one  $\phi_a(t)$  has been applied. The correction of the network improves a lot the image quality. However, it makes some important mistakes, like adding vessels in the right center of the image.

The overall computation time is about 30 *ms* on average to get a batch of output images



**Figure 4.13** – Result of the network on a  $512 \times 512$  test image for a defocus. The first image is the input of the network, which is the intensity of the aberrated image, the second one is the output, the corrected intensity image, and the third one is the reference image.



**Figure 4.14** – Result of the network trained with one aberration  $\langle \phi_a \rangle$ , on a  $512 \times 512$  test image on which one aberration  $\phi_a(t)$  has been applied. The first image is the input of the network, which is the intensity of the aberrated image, the second one is the output, the corrected intensity image, and the third one is the reference image.

from a batch of input images. The promise of a short computing time is therefore respected.

### 4.3 Conclusion and perspectives of eye-fundus correction

This work on the correction of aberrations with deep learning is a preliminary work. We have tested a lot of sets of parameters, more or less conclusive, but still many other tests must be achieved to strengthen the study. However, we can already draw several conclusions.

Even though the studied cases were simple, the first results show it is possible for a network to deduce a phase from an intensity image, which is quite promising. Nevertheless, a UNet may not be sufficient to solve our problem. We got a glimpse of it by training the network with a multitude of amplitudes of defocus, which has not been a success. One solution would be to expand the database with as many examples of aberrations and amplitudes as necessary. However, there is an infinite number of combinations between amplitude and modes of aberration to be taken into account. Variations of the UNet or even other architectures could be investigated. We could also work with a classification network, and rather than directly correct images, estimate the aberrated phase as a wavefront sensor does.

In any case for now, because of the instability and the “black box” effect that is created by

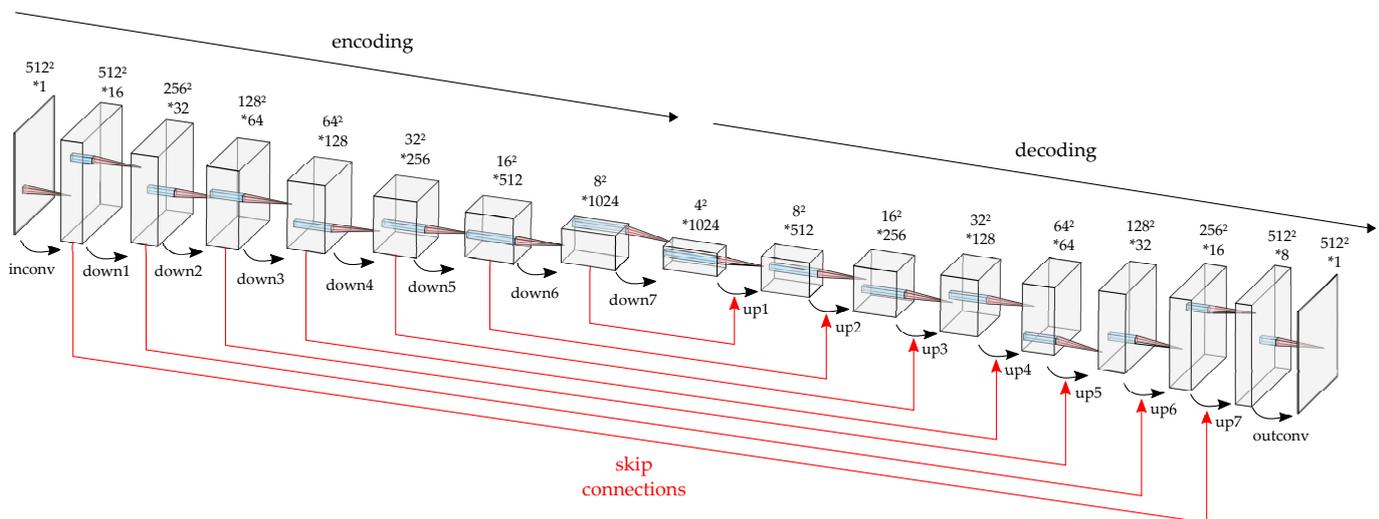
deep learning and is still work in progress in the mathematical community, it might not be easy to use it to reconstruct biomedical images. There is still an important risk of error, like creating vessels that do not exist like in fig. 4.14, that physicians cannot afford to take.

Besides, this work was just effective on eye-fundus images, while we want to use it on digital data. It is more complicated for now with our setup because we would need a huge amount of a large representative panel of data from our digital holography setup. However, we have started to do some tests on holographic data for hologram reconstruction, which are presented in the next section.

## 4.4 Digital Gabor hologram rendering

In this part, we show the reconstruction of so-called Gabor holograms using a UNet trained on an image database containing random amplitude points at random positions. The imaging model is based on propagation from the object to the camera, and vice versa, using the angular spectrum method (see section 2.1.2). We teach the network to infer this transformation from images of interference between several points. The learning on these simulated data allows to get rid of the constraint of the possibly insufficient number of experimental data. Once the network is trained, it is tested on synthetic interferograms simulated from images, and on interferograms of experimental data.

This work differs slightly from correcting aberrations. However, it may be possible to use the results for that aim, as it is explained in the section 4.5.

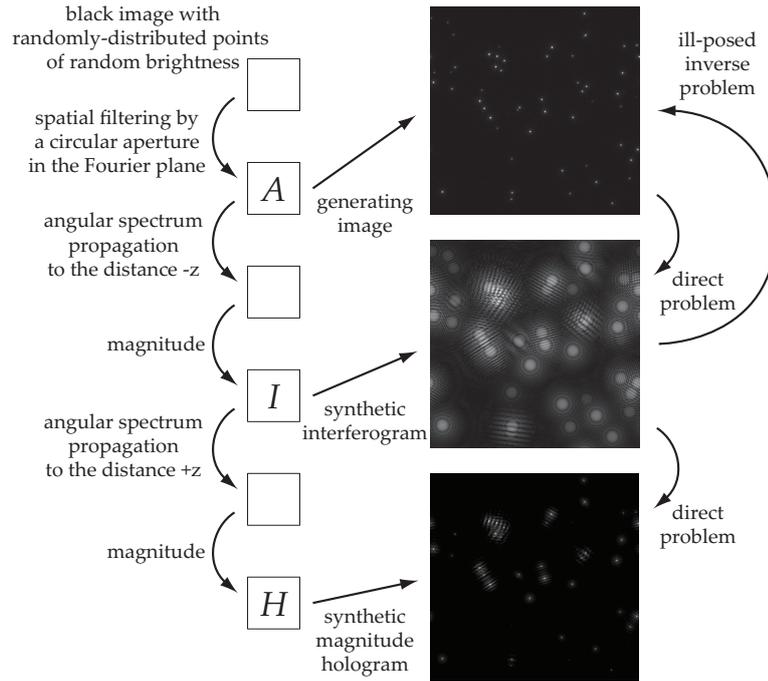


**Figure 4.15** – Scheme of the UNet structure which has been used for hologram reconstruction. The feature maps are represented by parallelepipeds, with their size above (*vertical*  $\times$  *horizontal*  $\times$  *channels*). The input is  $512 \times 512 \times 1$  image for the initial gray-scale image. The output is a  $512 \times 512$  gray-scale image. The blocks, which are composed of layers, are represented by the black arrows below the features. The network is composed of 1 convolution block and 7 downsampling blocks in the encoding part, and 1 convolution block and 7 upsampling blocks in the decoding part.

### 4.4.1 Synthetic interferograms

#### UNet

We use the same type of network than previously (UNet). The network used for this part is illustrated in fig. 4.15. The only differences are in the number of layers and the dimensions of some features. It is composed of 7 *down* blocks and 7 *up* blocks. The size of the input image is  $512 \times 512 \times 1$ , same as the output. The first block *inconv* generates 16 channels from the input. The last block *outconv* provides the output from a  $512 \times 512 \times 8$  feature. For the rest of the network, the parameters are similar.



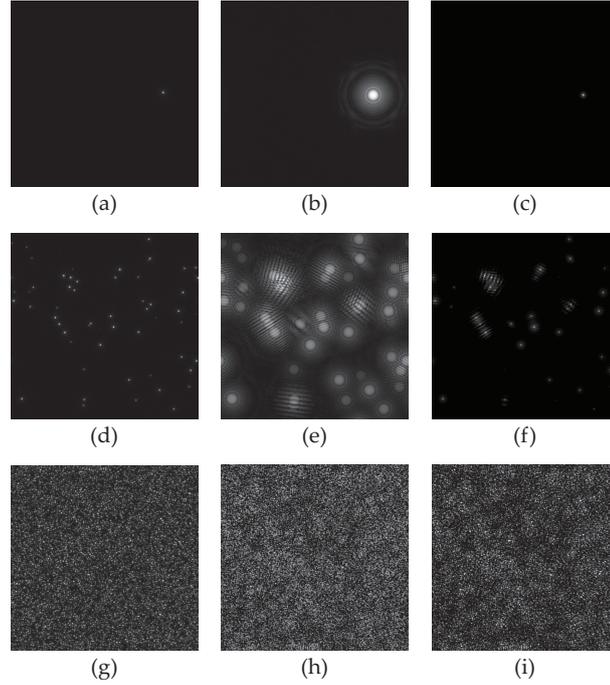
**Figure 4.16** – Steps to create a synthetic image database.  $I$  can be observed from  $A$  and  $H$  is calculated from  $I$ : this is the direct problem. The inverse problem consists in estimating  $A$  ( $\hat{A}$ ) from  $I$ . This is an ill-posed problem, because several  $\hat{A}$  can correspond to one  $I$ , due to the magnitude calculation to get  $I$  from the originate  $A$ .

#### Training database

The fig. 4.16 illustrates the construction of the database. A  $512 \times 512$  real positive image is generated to represent an object consisting of  $N$  source points of random brightness, randomly localised on a black background. The light emitted by these sources in the object plane is spatially filtered through a circular aperture in the Fourier plane of the object, constituting an image  $A$  in the image plane.  $A$  then describes the amplitude of the transmission function of the object (see first image in fig. 4.16).

The field described by  $A$  is then propagated using angular spectrum propagation (see ??) at a distance  $-z$  from the image plane. The amplitude of this field constitutes the image which is called a synthetic Gabor interferogram  $I$  :

$$I(x, y) = \left| \iint \tilde{A}(k_x, k_y) e^{-ik_z z} e^{ik_x x} e^{ik_y y} dk_x dk_y \right|, \quad (4.16)$$



**Figure 4.17** – Example of triplets  $(A, I, H)$  from the database: generates  $A$  images (1<sup>st</sup> column), synthetic interferograms  $I$  (2<sup>nd</sup> column) and amplitude of synthetic holograms  $H$  (3<sup>rd</sup> column). The number of random point sources is  $N = 1$  (1<sup>st</sup> row),  $N = 58$  (2<sup>nd</sup> row) and  $N = 26214$  (3<sup>rd</sup> row).

where  $(x, y)$  are the coordinates of the image pixels, and  $\tilde{A}(k_x, k_y)$  is the 2D Fourier transform of  $A(x, y)$ .  $k_x, k_y$ , and  $k_z$  are the projections of the wave vector  $k = 2\pi/\lambda$  (with  $\lambda$  the wavefront) on  $x$  and  $y$  (lateral) and  $z$  (axial) axes respectively, such as  $k_z^2 = k^2 - k_x^2 - k_y^2$ .

$I$  is measured experimentally by the camera, and the corresponding hologram  $H$  is calculated by propagating  $I$  to a distance  $+z$  with the angular spectrum method, and taking the amplitude of the field thus obtained :

$$H(x, y) = \left| \iint \tilde{I}(k_x, k_y) e^{+ik_z z} e^{ik_x x} e^{ik_y y} dk_x dk_y \right|, \quad (4.17)$$

where  $\tilde{I}(k_x, k_y)$  is the 2D Fourier transform of  $I(x, y)$ .

At the end of this process, we obtain a triplet of positive images  $(A, I, H)$  of size  $512 \times 512$  as illustrated in the fig. 4.17. The aim of the study is to teach the UNet the transformation to go from the interferogram  $I$  to the hologram  $H$ , with only this database of points.

### Results on synthetic data

For the training, 50000 pairs  $(I, H)$  have been generated. The number of point sources  $N$  over this database varies on a logarithmic scale from 1 to  $10\% \times 512^2 \approx 26214$ . The distance of propagation  $z = 0.065 \text{ m}$  is the same during all the training. 15% of the database has been used only for validation. The cost function used for training is the mean-square error, computed between the estimated hologram  $\hat{H}$ , which is the network output, and the expected hologram  $H$ .

The network has been trained on 70 epochs with a learning rate of 0.1. The training and the

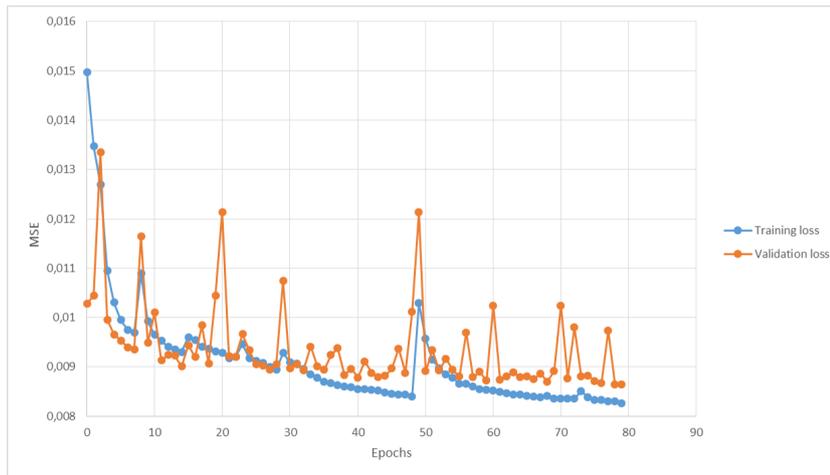


Figure 4.18 – Loss functions over the number of epochs, for training and for validation.

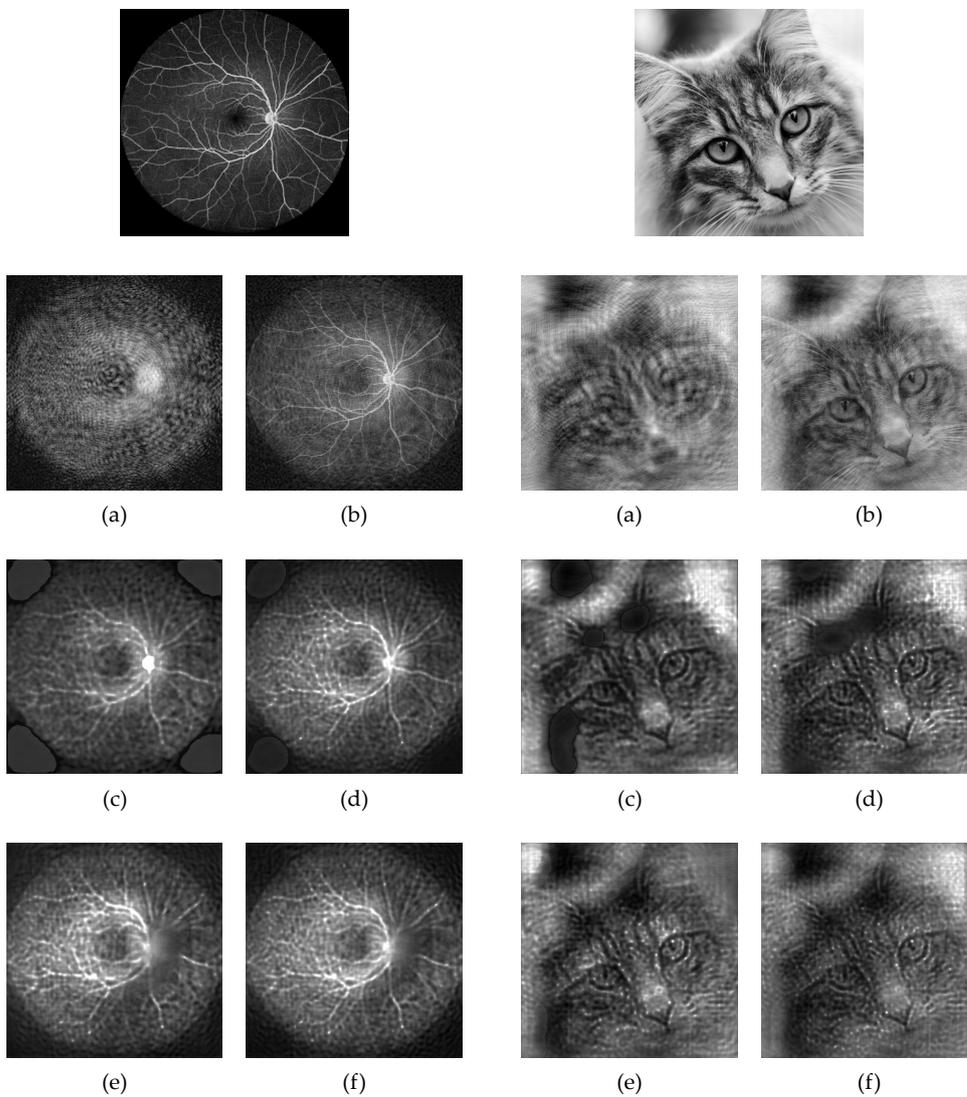
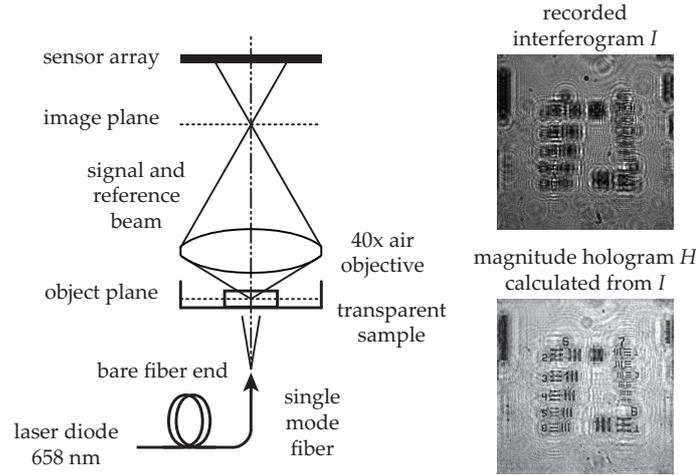


Figure 4.19 – Tests on simulated data: image  $A$  (top image), amplitude of the interferograms  $I$  calculated with eq. (4.16) (a), amplitude of the holograms  $H$  calculated with eq. (4.17) (b), and output images  $\hat{H}$  from the network after training over 10 (c), 40 (d), 50 (e) and 70 (f) epochs.



**Figure 4.20** – Holographic in-line microscope setup used to acquire images. Interferograms  $I$  are recorded by the sensor.

validation losses are presented in the fig. 4.18. Both curves tend to decrease in general, with only a few bumps, that always end up going back down. The decrease is slower and slower. We decided to stop the training when the validation loss seemed to stagnate.

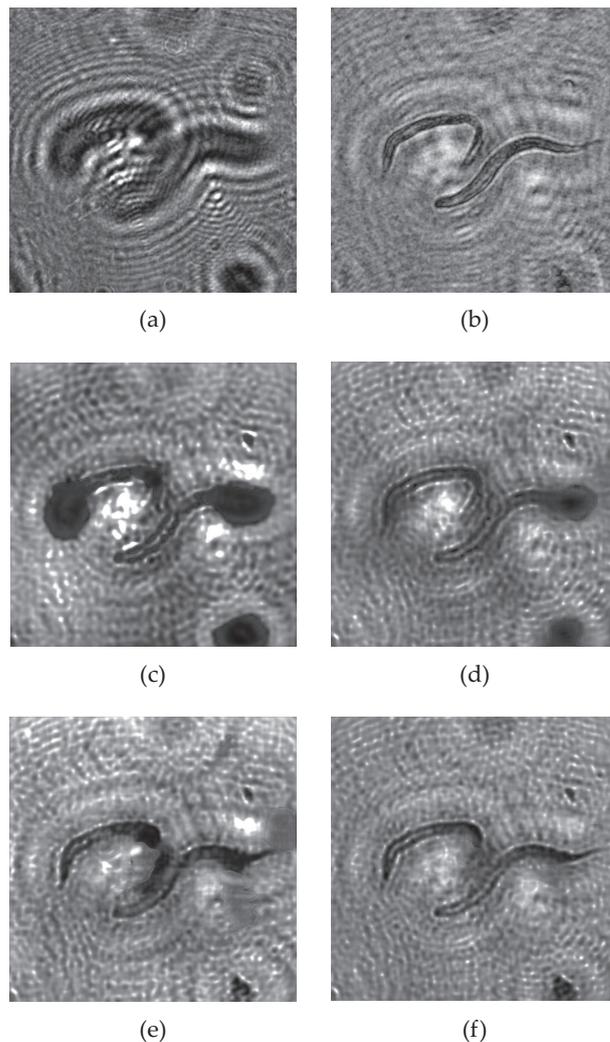
Tests have been performed on simulated interferograms: the process described on the fig. 4.16 has been applied on images  $A$  (this time, not points, but representing several objects like blood vessels or even cats), with  $z = 0.065\text{ m}$ , to get  $(I, H)$  pairs to test. Some examples of the results of the related tests are shown on fig. 4.19.

The network is reconstructing the holograms well, as we can see in the table below: the MSE values are quite close to those in the section 4.2.5, which reflects a good accuracy. The network seems to have learned well the propagation process based only on the example of source point propagation, even at the earliest epochs. As we go through the epochs, some artefacts disappear (in general the saturated parts, the four corners in the first image...).

However, we can also notice that some elements of the images seem less well represented in further epochs: it is the case of some vessels on the right in the first set of images for example. By aiming at minimizing the cost function, the network considers some elements less important, depending on the image, which can be a problem. This is reflected in the MSE values, which are not all decreasing with the number of training epochs. Therefore, it is necessary to test and choose carefully the versions of the network, with respect to the number of training epochs, before using it.

In any case, this example shows that by learning from PSFs, the network is able to assimilate physical properties of the system and apply them to more complex images.

MSE between $\hat{H}$ and $H$				
Number of training epochs	10	40	50	70
Blood vessels (fig. 4.19)	$4.9 \times 10^{-3}$	$6.6 \times 10^{-3}$	$1.6 \times 10^{-2}$	$1.2 \times 10^{-2}$
Cat (fig. 4.19)	$3.0 \times 10^{-2}$	$2.7 \times 10^{-2}$	$3.8 \times 10^{-2}$	$2.7 \times 10^{-2}$
Roundworms (fig. 4.21)	$2.6 \times 10^{-2}$	$1.5 \times 10^{-2}$	$1.7 \times 10^{-2}$	$7.7 \times 10^{-3}$



**Figure 4.21** – *C. elegans* roundworms images: recorded interferograms  $I$  (a), amplitude of holograms  $H$  calculated with eq. (4.17) (b), output images  $\hat{H}$  after 10 (c), 40 (d), 50 (e) and 70 (f) training epochs.

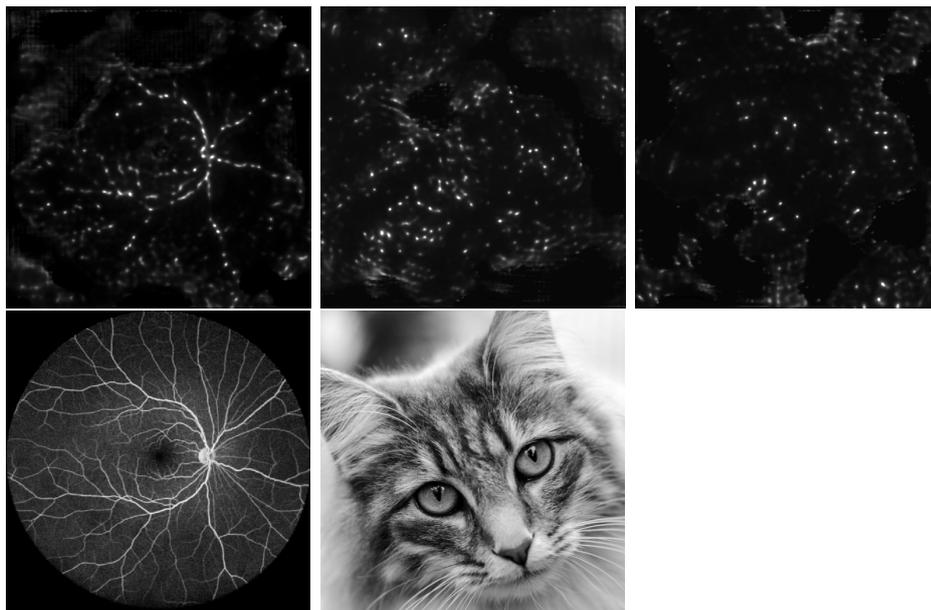
#### 4.4.2 Experimental results

After obtaining results on synthetic images, we have tested the network on experimental interferograms. The fig. 4.20 illustrates the setup we have used to acquire interferograms. The object is placed on a sample holder. A laser of wavelength  $\lambda = 658 \text{ nm}$  illuminates the sample from below. The light transmitted through the sample is focused by a lens, and a Gabor interferogram of size  $2048 \times 2048$  is measured by a camera located at a reconstruction distance  $z = 0.065 \text{ m}$  from the focal plane (image plane). A region of interest of size  $512 \times 512$  is isolated to form the  $I$  interferogram that will be processed by the network.

We have measured Gabor interferograms of roundworms *C. elegans*. The fig. 4.21 shows the results obtained by performing the classical hologram calculation (eq. (4.17)) and with the network. As for the synthetic images, the network produces a very good hologram reconstruction.

The granular aspect of the images reconstructed by neural networks is due to the training: by having learned the transformation on points, the network treats the image as independent points that interfere with each other, and not as correlated sets to propagate. By varying the size of the numerical aperture, which is responsible for the filtering of the field in the Fourier

plane to form  $A$ , the size of the elementary “grains” of the image can be changed.



**Figure 4.22** – Estimation  $\hat{A}$  of the network (first row) trained with  $(I, A)$  pairs formed with random points, for test images of blood vessels from fig. 4.19 (first image), cat from fig. 4.19 (second image) and roundworms from fig. 4.21. The second row are the corresponding  $A$  images (the third image has no  $A$  because it has not been simulated). The MSE between  $A$  and  $\hat{A}$  is equal to  $3.9 \times 10^{-2}$  for the first image, and  $3.2 \times 10^{-1}$  for the second one.

### 4.4.3 The inverse problem

Artifacts around the roundworms can be seen in the image  $H$  in the fig. 4.21b. They are due to the superposition of a twin image [124–126]. The beating wave between the field of the diffracted object and the reference wave is summed to its conjugate which creates artifacts in the holographic reconstruction. Since the network has learned with  $(I, H)$ , the reconstruction  $\hat{H}$  also includes this superposed twin image. So, the next step is to train the network with  $(I, A)$  couples, in order to remove this twin image.

The removal of the twin image using neural networks has already been demonstrated [127]. In this study, the network receives as input a hologram of a reconstructed real image with the twin image, and produces as output a corrected hologram. This output is compared to a hologram from which artifacts have been removed using an iterative algorithm [128].

Our training approach would allow us to train the network with  $A$  images, devoid of twin image, and independently of the database. With the training on  $(I, H)$  couples, the eq. (4.17) transformation has been taught to the network. With the training on the  $(I, A)$  couples, we could thus teach it directly the problem of image formation, without intermediary, possible source of perturbations.

For now, training only with  $(I, A)$  pairs does not give any conclusive result. By learning on  $(I, A)$  pairs, the network tries to retrieve source points, but in small quantities, which is reflected in the results of the fig. 4.22: the output images are only dark background with only few points. The MSE between  $A$  and  $\hat{A}$  is equal to  $3.9 \times 10^{-2}$  for the first image of fig. 4.22, and  $3.2 \times 10^{-1}$  for the second one. These are higher than the values observed in the table of section 4.4.1.

In reality, the problem is ill-posed. The difficulty lies in the surjectivity of the transformation from  $I$  to  $A$ : for one  $I$ , there exists several  $A$  that can correspond. Therefore, the system must be constrained to help the network to find the right solution, for example by adding a penalty to the cost function. However, it is difficult to find the right penalty to apply: we tried some of them like  $L_0$ ,  $L_1$  and  $L_2$  norms, without success. Then, we had the idea of using the Generative Adversarial Networks (GAN) system to constrain our images.

Generative Adversarial Networks (GAN) [129] are a system of two networks working together: the generator and the discriminator. The role of the generator is to generate images that could fool the discriminator in its task. The discriminator classifies images in two categories (like true or false). As both networks are trained, they become better and better: the generator provides images closer and closer to the real images, and the discriminator is more and more difficult to fool. We would like to use this system to add a constraint to our network.

The principle of the GAN that we would want to use is summarized in the fig. 4.23. The generator is the previously used UNet, which estimates an  $A$  for a given  $I$ . The cost function  $E1$  is a mean-square error (MSE) between  $A$  and  $\hat{A}$ .

$$E1 = MSE(A, \hat{A}) = \mathbb{E} \left[ (A - \hat{A})^2 \right]. \quad (4.18)$$

Then, the image  $\hat{A}$  is sent to the discriminator. The discriminator is a classifier which aims at differentiating true  $A$  images from false ones. It will provide a constraint to our system. Its response  $R1$  is a vector of 2 components:  $R1_0$  which corresponds to the false image class and the component  $R1_1$  corresponding to the true image class. The purpose of the generator is to produce an image that would deceive the discriminator. Thus, during the training of the generator, we try to minimize the  $E2$  error (eq. (4.19)).

$$E2 = -\log \left( \frac{\exp(R1_1)}{\exp(R1_0) + \exp(R1_1)} \right) = -\log(\mathcal{A}), \quad (4.19)$$

Indeed, we want  $R1_1 \gg R1_0$ , which would mean that the output of the discriminator is classified as a true image. In this case,  $\mathcal{A}$  tends towards 1, and thus  $E2$  tends towards 0. Conversely, if  $R1_0 \gg R1_1$ , our goal is not reached because the output is considered false, and therefore,  $\mathcal{A}$  tends towards 0, and  $E2$  tends towards infinity.

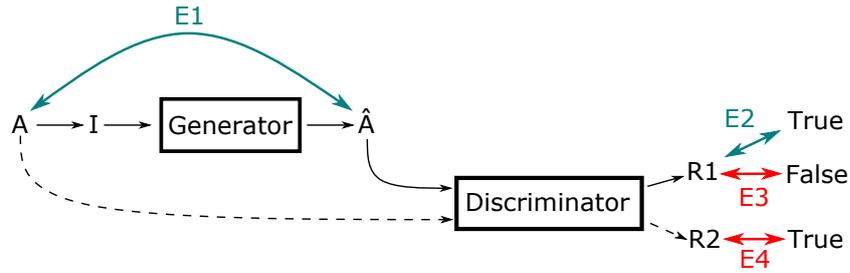
The total cost function  $E_{generator}$  used to train the UNet is composed of  $E1$ , the attachment to the data, and  $E2$ , the constraint brought by the discriminator.

$$E_{generator} = E1 + \alpha E2, \quad (4.20)$$

where  $\alpha$  is a coefficient used to weight the impact of the constraint.

For generator training, cost functions  $E3$  and  $E4$  are computed:

$$E3 = -\log \left( \frac{\exp(R1_0)}{\exp(R1_0) + \exp(R1_1)} \right), \quad (4.21)$$



**Figure 4.23** – Scheme of the GAN we use. The interferogram  $I$ , calculated from  $A$ , is at the input of the generator. It outputs the estimated image  $\hat{A}$ , which is sent to the input of the discriminator. The output of the latter is a response  $R1$ . Likewise  $A$  is sent as an input to the discriminator, which returns a  $R2$  response. The comparison between  $A$  and  $\hat{A}$  is  $E1$ . The comparison between  $R1$  and “True” class is  $E2$ . These two comparisons are used to train the generator. The comparison between  $R1$  and “False” class is  $E3$ . The comparison between  $R2$  and “False” class is  $E4$ . These two comparisons are used for the training of the discriminator.

and

$$E4 = -\log \left( \frac{\exp(R2_1)}{\exp(R2_0) + \exp(R2_1)} \right), \quad (4.22)$$

where  $R_2$  is the response from the discriminator to the image  $A$ . This time, the aim is to train the discriminator to differentiate between true and false images. Thus, the minimization of  $E3$  allows to improve its recognition of false images, while the minimization of  $E4$  allows to learn to recognize the true  $A$  images. The total cost function  $E_{discriminator}$  used to train the discriminator is

$$E_{discriminator} = (E3 + E4)/2. \quad (4.23)$$

Thus, the generator and the discriminator are trained in turn, minimizing the cost functions corresponding to each by gradient descent.

In this model, the required constraint would be provided by the discriminator, which would force the UNet to produce the right images. The discriminator that we have used until now is a simple combination of 3 linear layers, each separated by one Leaky ReLU layer, which is a ReLU layer authorizing a small positive gradient for negative values. The output of the last linear layer goes through a sigmoid layer which provides the final classification.

Unfortunately for now, we have not yet obtained satisfying results. The error functions  $E1$  and  $E2$ , and  $E3$  and  $E4$  are difficult to control together: when one is convergent and decreases for each iteration, the other is often divergent. This may be due to the classifier which could be too simple compared to the generator (UNet). We are testing the system with more complex classifiers such as VGG [130] and ResNet [131], which could more easily compete, at least in terms of complexity, with the UNet.

## 4.5 Conclusions and perspectives on the use of deep learning in correction of aberrations

We have shown in the previous section that it is possible to teach the network a simple transformation through learning based on the system impulse response. Tests on experimental images have been conclusive, opening up many possibilities to be explored to improve the quality of holographic reconstructions and to increase the variety and complexity of problems that the network can solve with this learning method.

The transformation that the network is learning is a propagation from one plane to another plane further away. We can consider this transformation as a defocus. In this way, the network could also learn other types of aberrations. That would avoid the database problem of our holographic setup: we would not need a lot of various holographic data to learn the correction of an aberration. The learning on simulated PSF would be sufficient.

Still it remains the problem of the learning of only one aberration. We have seen in the section 4.2.5 that it is difficult for a network which has learned several amplitudes of the same aberration, to correct images with these amplitudes of aberrations, due to their variety. So, it is hard to imagine that the network, as it is, could compensate several modes of aberration. However, the simplification of the database could suggest that the number of degrees of freedom of aberrations might be increased. Indeed, more weights of the network would be devoted to the transformation itself instead of the effect of the transformation on features specific to each image. Then, the network would have less difficulty to learn more aberrations. But it remains an hypothesis as it has not been tested.

In any case, the accuracy of the results obtained in this chapter suggests that the compensation of aberrations will be difficult with current means: the direct correction of images may lack of the necessary precision to be used in the field of biomedical imaging. However, other approaches could be investigated, through other UNet-type architectures for aberration compensation, or classification architecture for the estimation.

---

## Conclusions and perspectives

---

The aim of this PhD thesis was to find a way to improve holographic image quality, in particular by compensating aberrations. After having highlighted of the state of the art to compare with other existing retinal imaging techniques, this manuscript provides an overview of our setups located at the Quinze-Vingts national ophthalmology hospital in Paris: holographic swept-source optical coherence tomography and laser Doppler holography. As a first step, we have made some improvements in the image rendering, with the rescaling of interferograms, and axial and lateral motion compensation. Then, for the same purpose of improving the image resolution, we have mainly studied two methods for the correction of aberrations in images of the Doppler holographic setup.

The estimation of aberrations with a digital Shack-Hartmann wavefront sensor is well adapted to our optical system as its mathematical model has been developed on its basis. It provides good results on tests and even on the first retina images of patients. The computation time is reasonable and bodes well for the future: the computation of the wavefront sensing algorithm increases only by a factor 2 the image rendering of laser Doppler holograms. This time may be compressed by further parallelization and optimization of the algorithm. Therefore, this technique is very well suited to our system and improves the resolution of the images.

The deep learning method is promising in terms of computation time. Indeed, the network provides corrected images from aberrated images in 30 ms. However, the results remain far from being exploitable for the moment. And this will certainly remain the case for a “simple” UNet which does not sufficiently remember the variety of training cases, and which then provides a general and average image of the possible solutions. In this case, perhaps we should rather look at another use of deep learning, to estimate aberrations rather than correcting the images directly.

This work has nevertheless identified several interesting points that deserve to be raised. The network is able to estimate and compensate phase from an intensity image. Even in the hardest cases, when the network has learned multiple phases, it provides an improvement of the image quality. Finally, the last study on the database suggests opportunities for learning that could be adapted to our case. Thus, this work on deep learning remains an introduction, which nevertheless brings interesting perspectives that can be explored later, such as the compensation of scattering effects provoking a twin-image in holography.

Both methods have been developed for use on Doppler holography images. The great simi-

larity between laser Doppler and OCT setups allows us to think that it will be possible to apply these methods to holographic OCT, once the latter is able to image the retina.

In conclusion, we have introduced several methods for image quality improvement in digital holography, and validated them experimentally. The resolution of holographic images has been improved by robust non-iterative methods: lateral and axial tracking and compensation of translations, and measurement and compensation of optical aberrations. The digital wavefront sensor can be considered as a very well adapted method. The use of a neural network has provided interesting results, even though its performance is not sufficient for now. This allows us to be optimistic that structures on holographic images of the retina will be more visible and sharper, which could ultimately provide very valuable information to clinicians.



## APPENDIX A

## Statistics for the simulation of a Shack-Hartmann wavefront sensor

For a point source, on 100 iterations										
sub-apertures	Mean error					Standard deviation				
	6 × 6	7 × 7	8 × 8	9 × 9	10 × 10	6 × 6	7 × 7	8 × 8	9 × 9	10 × 10
4 <sup>th</sup> mode	0,61	0,62	0,74	0,58	0,70	0,48	0,47	0,55	0,45	0,49
5 <sup>th</sup> mode	0,63	0,56	0,55	0,43	0,42	0,50	0,46	0,36	0,31	0,31
6 <sup>th</sup> mode	0,72	0,70	0,62	0,64	0,70	0,52	0,52	0,44	0,46	0,50
7 <sup>th</sup> mode	0,49	0,48	0,38	0,37	0,45	0,37	0,35	0,28	0,27	0,31
8 <sup>th</sup> mode	0,87	0,68	0,54	0,44	0,59	0,64	0,49	0,44	0,31	0,44
9 <sup>th</sup> mode	0,96	0,64	0,53	0,43	0,54	0,66	0,44	0,44	0,36	0,39
10 <sup>th</sup> mode	0,47	0,44	0,33	0,34	0,38	0,37	0,35	0,29	0,25	0,29
11 <sup>th</sup> mode	0,55	0,42	0,36	0,28	0,29	0,39	0,35	0,28	0,23	0,22
12 <sup>th</sup> mode	0,44	0,35	0,27	0,16	0,20	0,37	0,28	0,26	0,15	0,16
13 <sup>th</sup> mode	0,42	0,40	0,27	0,20	0,20	0,34	0,33	0,21	0,16	0,16
14 <sup>th</sup> mode	0,57	0,40	0,29	0,28	0,15	0,40	0,31	0,26	0,24	0,13
15 <sup>th</sup> mode	0,53	0,45	0,31	0,35	0,30	0,41	0,35	0,23	0,28	0,22
16 <sup>th</sup> mode	0,62	0,52	0,38	0,32	0,30	0,54	0,45	0,25	0,29	0,21
17 <sup>th</sup> mode	0,32	0,19	0,19	0,14	0,10	0,21	0,13	0,14	0,12	0,09
18 <sup>th</sup> mode	0,24	0,19	0,11	0,13	0,09	0,20	0,16	0,09	0,10	0,09
19 <sup>th</sup> mode	0,22	0,16	0,11	0,12	0,08	0,19	0,11	0,10	0,10	0,07
20 <sup>th</sup> mode	0,28	0,17	0,12	0,11	0,10	0,24	0,13	0,12	0,09	0,08
21 <sup>th</sup> mode	0,28	0,23	0,19	0,21	0,20	0,20	0,18	0,14	0,15	0,15
22 <sup>th</sup> mode	0,23	0,23	0,21	0,24	0,12	0,19	0,16	0,17	0,18	0,09
23 <sup>th</sup> mode	0,21	0,15	0,08	0,13	0,06	0,21	0,12	0,07	0,09	0,05
24 <sup>th</sup> mode	0,21	0,16	0,09	0,10	0,09	0,19	0,13	0,08	0,08	0,07
25 <sup>th</sup> mode	0,20	0,13	0,07	0,10	0,06	0,16	0,10	0,05	0,08	0,05
26 <sup>th</sup> mode	0,20	0,14	0,09	0,12	0,06	0,17	0,11	0,08	0,10	0,06
27 <sup>th</sup> mode	0,30	0,20	0,18	0,16	0,12	0,26	0,15	0,15	0,13	0,10
28 <sup>th</sup> mode	0,39	0,23	0,23	0,15	0,15	0,30	0,17	0,18	0,12	0,11
29 <sup>th</sup> mode	0,27	0,27	0,16	0,18	0,13	0,24	0,22	0,12	0,12	0,13
30 <sup>th</sup> mode	0,33	0,21	0,14	0,13	0,08	0,27	0,17	0,12	0,11	0,07
Average	0,43	0,35	0,28	0,25	0,25	0,33	0,27	0,22	0,20	0,19

For an extended point source, on 100 iterations										
sub-apertures	Mean error					Standard deviation				
	$6 \times 6$	$7 \times 7$	$8 \times 8$	$9 \times 9$	$10 \times 10$	$6 \times 6$	$7 \times 7$	$8 \times 8$	$9 \times 9$	$10 \times 10$
4 <sup>th</sup> mode	0,44	0,57	0,69	0,60	0,76	0,33	0,40	0,46	0,42	0,52
5 <sup>th</sup> mode	0,43	0,42	0,43	0,36	0,41	0,33	0,33	0,30	0,26	0,28
6 <sup>th</sup> mode	0,55	0,63	0,69	0,62	0,79	0,43	0,50	0,47	0,46	0,55
7 <sup>th</sup> mode	0,38	0,41	0,40	0,33	0,42	0,31	0,30	0,27	0,24	0,30
8 <sup>th</sup> mode	0,72	0,52	0,45	0,48	0,47	0,47	0,43	0,38	0,33	0,36
9 <sup>th</sup> mode	0,72	0,54	0,42	0,39	0,48	0,56	0,39	0,31	0,28	0,35
10 <sup>th</sup> mode	0,35	0,35	0,31	0,33	0,39	0,30	0,27	0,22	0,24	0,27
11 <sup>th</sup> mode	0,39	0,32	0,27	0,23	0,27	0,28	0,24	0,18	0,16	0,20
12 <sup>th</sup> mode	0,31	0,21	0,09	0,07	0,07	0,23	0,16	0,08	0,06	0,07
13 <sup>th</sup> mode	0,27	0,20	0,15	0,10	0,08	0,20	0,17	0,14	0,10	0,06
14 <sup>th</sup> mode	0,37	0,26	0,18	0,11	0,07	0,27	0,24	0,16	0,10	0,07
15 <sup>th</sup> mode	0,43	0,32	0,25	0,25	0,25	0,32	0,27	0,18	0,18	0,17
16 <sup>th</sup> mode	0,33	0,35	0,28	0,21	0,20	0,26	0,30	0,23	0,16	0,14
17 <sup>th</sup> mode	0,30	0,15	0,10	0,08	0,06	0,22	0,11	0,09	0,07	0,06
18 <sup>th</sup> mode	0,21	0,12	0,06	0,09	0,06	0,15	0,11	0,06	0,07	0,04
19 <sup>th</sup> mode	0,23	0,10	0,07	0,06	0,03	0,18	0,08	0,06	0,05	0,03
20 <sup>th</sup> mode	0,24	0,12	0,08	0,06	0,05	0,20	0,10	0,07	0,06	0,04
21 <sup>th</sup> mode	0,24	0,22	0,15	0,17	0,19	0,17	0,17	0,11	0,11	0,13
22 <sup>th</sup> mode	0,22	0,14	0,14	0,15	0,11	0,17	0,10	0,10	0,11	0,08
23 <sup>th</sup> mode	0,14	0,10	0,04	0,08	0,02	0,12	0,08	0,04	0,06	0,02
24 <sup>th</sup> mode	0,20	0,13	0,04	0,06	0,05	0,15	0,09	0,03	0,04	0,03
25 <sup>th</sup> mode	0,16	0,09	0,04	0,07	0,03	0,12	0,08	0,03	0,06	0,03
26 <sup>th</sup> mode	0,13	0,11	0,07	0,05	0,03	0,11	0,09	0,07	0,06	0,03
27 <sup>th</sup> mode	0,23	0,16	0,12	0,08	0,06	0,20	0,14	0,11	0,08	0,06
28 <sup>th</sup> mode	0,28	0,14	0,19	0,13	0,12	0,22	0,11	0,14	0,09	0,08
29 <sup>th</sup> mode	0,18	0,17	0,12	0,16	0,07	0,13	0,13	0,09	0,12	0,06
30 <sup>th</sup> mode	0,20	0,13	0,09	0,06	0,04	0,13	0,11	0,09	0,05	0,04
Average	0,32	0,26	0,22	0,20	0,21	0,24	0,20	0,17	0,15	0,15

These two tables represent the mean error and the standard deviation of the distance between estimated and true Zernike coefficients (in radians), for simulated data. The first table is for a point source and the second one stands for an extended source. For each iteration of the simulation, the coefficients were drawn randomly according to a normal distribution  $N(0, 1)$ , and estimated by our implementation of the wavefront sensor. Mean error and standard deviation have been calculated on 100 iterations, for  $6 \times 6$ ,  $7 \times 7$ ,  $8 \times 8$ ,  $9 \times 9$ , and  $10 \times 10$  sub-apertures. The last line is the computation of the average value for each column.



# APPENDIX B

---

## Scientific contributions

---

In this appendix, the scientific contributions brought during this PhD thesis are presented.

The oral communications performed during this PhD are:

- *Motion compensation in digital holography for retinal imaging*, at “Journées d’imagerie optique non conventionnelle“ (JIONC) in Paris in **March 2018**.
- *Motion compensation in digital holography for retinal imaging*, at Optdiag in Paris in **May 2018**.
- *Deep neural networks for aberration compensation in digital holographic imaging of the retina*, at SPIE Photonics West in San Francisco in **February 2019**.
- *Motion compensation in digital holography for retinal imaging*, at IEEE International Symposium on Biomedical Imaging (ISBI) in Venice in **April 2019**.
- *Aberration compensation in digital holography of the retina*, “ma thèse en 180 secondes“, at “Journée de l’EDITE“ in Paris in **September 2019**.
- *Aberration compensation in digital holography of the retina*, during a seminary at ONERA in Châtillon in **September 2019**.
- *Compensation numérique d’aberrations en imagerie holographique de la rétine*, at “Journées d’imagerie optique non conventionnelle“ (JIONC) in Paris in **March 2020**.

The poster presented during this PhD are:

- *Motion compensation in digital holography for retinal imaging*, at Optdiag in Paris in **May 2018**.
- *Deep neural networks for aberrations compensation in retinal imaging acquired by digital holography*, Data Science Summer School at Ecole Polytechnique in Palaiseau in **June 2018**.

- *Deep neural networks for aberrations compensation in retinal imaging acquired by digital holography*, at “Journée de l’EDITE” in Paris in **July 2018**.

The publications are:

- Julie Rivet, Guillaume Tochon, Serge Meimon, Michel Paques, Thierry Géraud and Michael Atlan. Motion compensation in digital holography for retinal imaging. In Proceedings of the 16th IEEE International Symposium on Biomedical Imaging (ISBI), pages 1428-1431, Venice, Italy, April 2019a. <http://dx.doi.org/10.1109/ISBI.2019.8759564>
- Julie Rivet, Guillaume Tochon, Serge Meimon, Michel Paques, Thierry Géraud and Michael Atlan. Deep neural networks for aberration compensation in digital holographic imaging of the retina. In Proceedings of the SPIE Conference on Adaptive Optics and Wavefront Control for Biological Systems V, number 10886-34, San Francisco, CA, USA, March 2019b. SPIE. <http://dx.doi.org/10.1117/12.2509711>

The articles published, under review or put on arxiv related to this PhD thesis are attached on the next page. These are:

- *Swept-source, off-axis Fresnel transform holographic optical coherence tomography with an output throughput of 10 Giga voxels per second in real-time*
- *Motion compensation in digital holography for retinal imaging*, published in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019).
- *Experimental digital Gabor hologram rendering by a model-trained convolutional neural network*

# Swept-source, off-axis Fresnel transform holographic optical coherence tomography with an output throughput of 10 Giga voxels per second in real-time

E. Charpentier,<sup>1</sup> F. Lapeyre,<sup>1</sup> J. Gautier,<sup>1</sup> L. Waszczuk,<sup>1</sup> J. Rivet,<sup>1</sup>  
L. Puyo,<sup>1</sup> J.P. Huignard,<sup>1</sup> M. Fink,<sup>1</sup> S. Meimon,<sup>2</sup> and M. Atlan<sup>1</sup>

<sup>1</sup> *Centre National de la Recherche Scientifique (CNRS) UMR 7587,*

*Institut Langevin. Paris Sciences et Lettres (PSL) University, Université Pierre et Marie Curie (UPMC),*

*Université Paris 7. École Supérieure de Physique et de Chimie Industrielles ESPCI Paris - 1 rue Jussieu. 75005 Paris. France*

<sup>2</sup> *ONERA the French Aerospace Lab, Châtillon, France*

(Dated: May 20, 2019)

We demonstrate swept-source optical coherence tomography in real-time by high throughput digital Fresnel hologram rendering from optically-acquired interferograms with a high-speed camera. The interferogram stream is spatially rescaled with respect to wavelength to compensate for field-of-view dilation inherent to discrete Fresnel transformation. Holograms are calculated from an input stream of 16-bit, 1024-by-1024-pixel interferograms recorded at up to 512 frames per second with a digital camera. All calculations are performed by a NVIDIA TITAN Xp graphics card on single-precision floating-point complex-valued arrays (32-bit per quadrature). It allows sustained computation of 1024-by-1024-by-256-voxel volumes at 10 billion voxel/s, from which three perpendicular cuts are displayed in real-time at user-selected locations, up to 38 frames per second.

Optical coherence tomography (OCT) refers to a wide range of non-invasive high resolution optical interferometric imaging technologies which have found major applications in ophthalmology. Point-scanning (flying spot) OCT allows for confocal gating of the detected photons, filtering of the out-of-focus light. However, nowadays, shot noise-limited point scanning confocal OCT modalities [1] have reached a throughput limit due to the maximum permissible exposure in ophthalmic applications. The high complexity of OCT systems that implement adaptive optics [2] hinders their wide spread use in clinical research. Full-field swept-source OCT (FF-SS-OCT) is an OCT implementation that can acquire data several orders of magnitude faster with a sensor array. FF-SS-OCT allows for acquisition of volumes in a single laser sweep, eliminating artifacts from imperfect phase reproducibility during laser sweeps [3]. Finally, FF-SS-OCT allows for off-axis schemes, that improve image quality by spatial filtering of interferometric terms [4]. Line-field OCT is a hybrid method, benefiting from the advantages of confocal and full field detection schemes, while suffering from their respective limitations. Images are recorded with a single line camera; a lateral scanning and the laser sweep allow for three-dimensional imaging. The partial confocal filter of line-field OCT permits high quality and high-speed imaging at increased illumination levels with respect to bidimensional scanning schemes [5]. With a setup complexity slightly lower than its confocal counterparts, line-scanning OCT offers good promises of high image quality through digital refocusing of in-depth acquisitions [1, 6].

Imaging speed is crucial in optical coherence tomography (OCT) systems for imaging of dynamic samples [7]. The presence of motion artefacts deteriorates the resolution of the OCT images. Consequently, acquisition speed and spatial resolution are ultimately linked [8]. Holographic OCT with a swept-source laser on a sensor array can acquire data several orders of magnitude faster

than scanning OCT modalities [3]. As the technology of high-speed cameras, wavelength-swept lasers, and parallel computing improve, holographic OCT should become a cost-effective alternative to complex OCT systems that implement adaptive optics [2]. The main advantages of holographic OCT lie in: 1- a reduced system complexity, in which no moving parts are involved, 2- the available detection throughput of high-speed CMOS cameras, 3- the possibility to perform aberration correction in post-processing [3], 4- the computation of 3D motion fields [9]. A fast, robust and versatile digital image acquisition and rendering software is a key requirement for the development of digital holographic imaging. Video-rate holographic image rendering was made possible by streamline processing of optically-acquired interferograms on graphics processing units (GPUs) [10, 11]. When real-time performance is required, hardware acceleration of computations is an efficient way of increasing the throughput [12, 13]. Special-purpose field-programmable gate array (FPGA) chips were reportedly used for high throughput holographic image rendering [14], at the price of less versatility than with general-purpose graphics processing units. Here, we demonstrate digital hologram rendering for swept-source optical coherence tomography with an output throughput of 10 billion voxels per second.

Digitized interferograms were measured with the setup sketched in Fig. 1. It is an off-axis (angle  $\sim 1^\circ$ ), interferometer used for optical detection of an object field  $E$  beating against a separate local oscillator field  $E_{LO}$ , in reflective geometry. A tunable laser (Broadsweeper BS-840-2-HP from Superlum) emits a 40 mW radiation whose angular optical frequency  $\omega$  is swept linearly with time from  $\omega_1$  to  $\omega_2$ , during a sweep time  $T = 2\pi/\omega_{\text{sweep}} = 1$  s. These bounds are linked to the start wavelength  $\lambda_1 = 870 \pm 2$  nm and the end wavelength  $\lambda_2 = 820 \pm 2$  nm respectively via the relation  $\omega = 2\pi c/\lambda$ , where  $c$  is the speed of light, and define a tuning range  $\Delta\lambda = |\lambda_2 - \lambda_1| = 50$  nm. As a consequence, the angular wavenumber

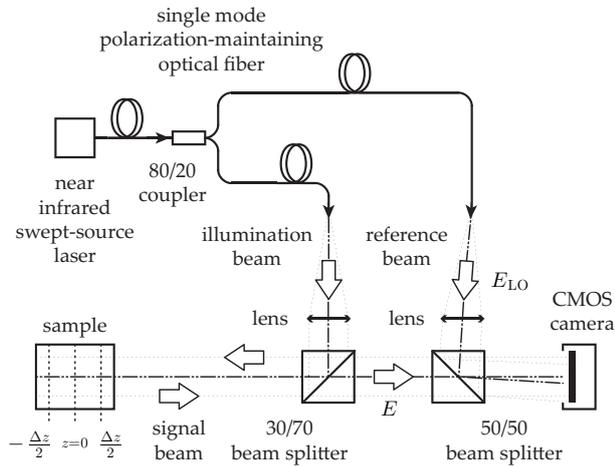


FIG. 1: Sketch of the wavelength-tuning laser holographic interferometer prototype used for demonstration.

$k = 2\pi/\lambda$  of the laser radiation is swept linearly with time. Interference patterns were recorded with a Adimec Quartz camera (pixel size  $d = 12 \mu\text{m}$ ), via a CoaXPress framegrabber Bitflow Cyton CXP4, at a frame rate of  $\nu_S = \omega_S/(2\pi) = 256 \text{ Hz}$ , with 16-bit/pixel quantization. Each raw digitized interferogram of  $1024 \times 1024$  pixels at position  $(x, y)$  and time  $t$  is noted  $I(x, y, t) = |E + E_{LO}|^2$ , where the object field is  $E = \mathcal{E} \exp(i\omega t)$ , the reference field is  $E_{LO} = \mathcal{E}_{LO} \exp(i\omega t - i\phi)$ , and  $\phi$  is the phase detuning between both optical fields. Off-axis configuration allows for spatial separation of the self-beating  $|E|^2$ , and  $|E_{LO}|^2$  and cross-beating interferometric contribution,  $EE_{LO}^* = \mathcal{E}\mathcal{E}_{LO}^* \exp(i\phi)$ , and its complex conjugate [15].

In swept-source OCT, a linear variation of the instantaneous optical angular frequency  $\omega_L = \beta t$  with time  $t$  during a sweep results in a phase variation  $\phi = 2kz$  of the interferometric beat between the wave backscattered by a diffuser at axial position  $z$  and the reference wave

$$\phi = \frac{2\beta z}{c} t \quad (1)$$

of the cross-beating part of the interferogram scaling up linearly with the sweep speed  $\beta = (\omega_2 - \omega_1)/T$  of the laser angular frequency and with the detuning pathlength  $z$  between the object and reference waves [16]. Hence, the instantaneous beat frequency

$$\omega = \frac{\partial \phi}{\partial t} = \frac{2\beta z}{c} \quad (2)$$

of the interferogram encodes the optical reflectivity signal at a detuning pathlength  $z$  of the interferometer. Hence the Fourier transform of the temporal trace of the beat signal scales as the optical reflectivity against  $z$  [16]. The total restituted axial range  $\Delta z$  can be derived from Eq. 2 and the Shannon theorem: beat frequencies measured for detuning pathlengths are bounded by the sampling bandwidth:  $\omega_S = 2\beta\Delta z/c$ . For a wavelength sweep from

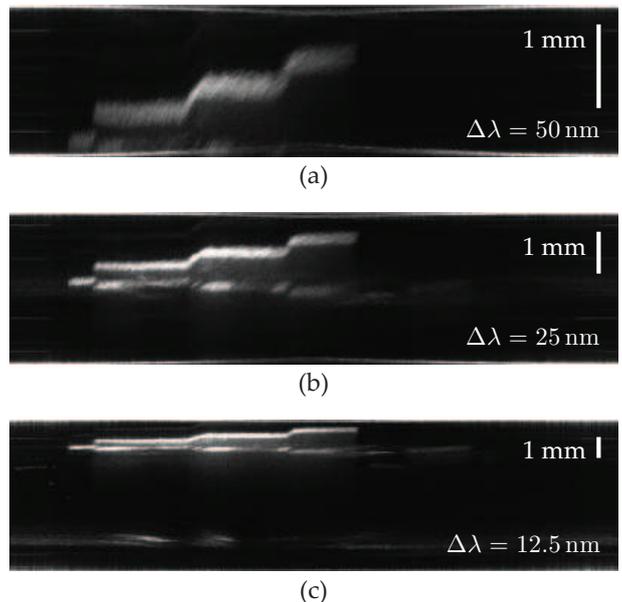


FIG. 2: Axial range  $\Delta z$  vs. wavelength tuning range  $\Delta\lambda$  of 50 nm (a), 25 nm (b), and 12.5 nm (c). The height difference between adjacent scales is  $\sim 313 \mu\text{m}$ , which was measured by full-field OCT. The axial field of view  $\Delta z$  of 1.8 mm (a), 3.7 mm (b), and 7.3 mm (c) is in agreement with the theoretical value from Eq. 3 for  $N_z = 256$ . Axial scalebar : 1 mm.

$\lambda_1$  to  $\lambda_2$  during the time  $T$ , this translates to

$$\Delta z = \frac{N_z}{2} \Lambda \quad (3)$$

where  $\Lambda = \lambda_1\lambda_2/\Delta\lambda = 14.3 \mu\text{m}$  is twice the axial pitch and  $N_z = \omega_S/\omega_{\text{sweep}} = 256$  is the number of interferograms measured during one frequency sweep. The variation of the axial range  $\Delta z$  versus wavelength tuning range  $\Delta\lambda$  is illustrated experimentally in Fig. 2. For tunable lasers with non-Gaussian output spectra [7], the theoretically limiting axial resolution is the round-trip coherence length  $\delta z \approx \bar{\lambda}^2/\Delta\lambda = 14.3 \mu\text{m}$ , where  $\bar{\lambda} = (\lambda_1 + \lambda_2)/2$  is the central frequency of the sweep. This round-trip coherence length is about the same length as the axial pitch:  $\delta z \simeq \Lambda$ .

The sequence of steps reported in Fig. 3 from the input stream of recorded interferograms to the stream of tomographies illustrate the image rendering pipeline. A prerequisite to avoid distortion and loss of resolution of tomograms calculated from Fresnel-transform holograms is to enforce constant values of pixel dimensions in the image plane for all sampling wavelengths of the sweep [17, 18]. In contrast with angular spectrum propagation of the wave field [19, 20], the lateral size of a reconstructed pixel of a hologram calculated by Fresnel transformation (Eq. 6)

$$d' = \frac{\lambda z}{N_x d} \quad (4)$$

scales up linearly with the wavelength. This chromatic variation can be counterbalanced by resampling each in-

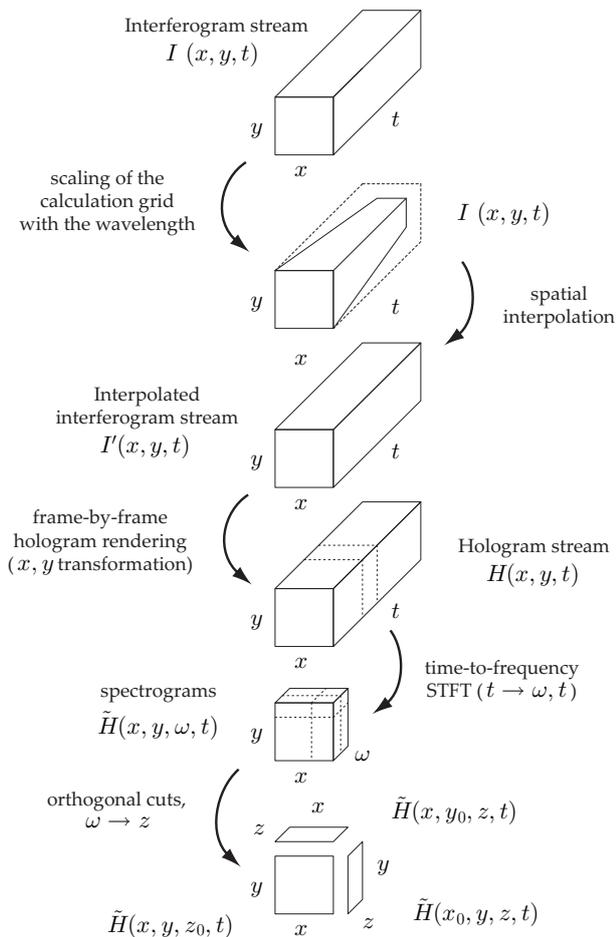


FIG. 3: Flowchart of the main data processing steps. Each interferogram  $I$  from the input data stream is rescaled by linear interpolation to  $I'$  to compensate for chromatic distortion. Holograms  $H$  are calculated by Fresnel transformation from rescaled interferograms  $I'$ . Temporal short-time Fourier transforms at each pixel location  $(x, y)$  turn holograms  $H$  into spectrograms  $\tilde{H}$ . Orthogonal cuts of the magnitude of each spectrogram  $|\tilde{H}|$  are displayed at video rate.

interferogram with a wavelength-dependent pitch  $d_\lambda$ . Using a variable sampling pitch  $d_\lambda = \lambda d / \lambda_1$  makes the reconstruction pixel  $d'$ , and hence the lateral field of the reconstructed hologram, wavelength-independent. In practice, each interferogram  $I$  was resampled by linear interpolation to  $I'$  onto a calculation grid with the same number of points ( $N_x \times N_y$ ) with a pitch  $d_\lambda$ .

$$I'(x, y, t) = I(x\lambda/\lambda_1, y\lambda/\lambda_1, t) \quad (5)$$

Fig. 3 illustrates experimental chromatic stretching compensation by linear interferogram resampling, and Fig. 4 shows its effect on the lateral resolution of a tomographic hologram of a resolution target for a wavelength sweep from  $\lambda_1 = 870$  nm to  $\lambda_2 = 820$  nm. Image rendering of complex-valued holograms  $H(x, y, t)$  from the stream of rescaled interferograms  $I'(x, y, t)$  was performed by dis-

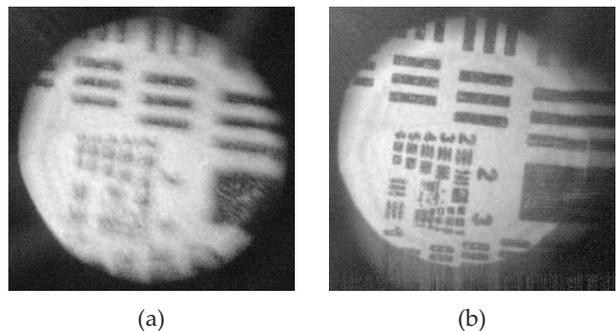


FIG. 4: illustration of chromatic compensation by interferogram resampling. Hologram rendering without interpolation in Visualization 1.

crete Fresnel transformation [20]

$$H(x, y, t) = \frac{i}{\lambda_1 z} \exp(-ik_1 z) \iint I'(x', y', t) \times \exp\left[\frac{-i\pi}{\lambda_1 z} ((x - x')^2 + (y - y')^2)\right] dx' dy' \quad (6)$$

where the parameter  $z$  corresponds to the sensor-to-object distance for a flat reference wavefront and in the absence of lens in the object path. Eq. 6 is used for reconstruction parameters  $z \geq z_b$ , where  $z_b = N_x d^2 / \lambda_1$ . Eq. 6 is used with parameters  $k_1$  and  $\lambda_1$  for the reconstruction of all the digital holograms throughout the sweep, regardless of the wavelength to which they correspond, in consequence of wavelength rescaling (Eq. 5). Demodulation of the axial depth  $z$  of the tomograms consists in forming the beat frequency spectrum of the holograms. For that purpose, temporal short-time Fourier transforms  $\tilde{H}(x, y, \omega, t)$  are calculated from the stream of holograms  $H(x, y, t)$ .

$$\tilde{H}(x, y, \omega, t) = \int H(x, y, \tau) g(t - \tau) e^{-i\omega\tau} d\tau \quad (7)$$

where  $g(t)$  is a rectangular time gate of  $N_z = 256$  consecutive images. Then, the envelope of  $\tilde{H}$  is formed, and axial pathlength detunings  $z$  are calculated from beat frequencies  $\omega$  via Eq. 2. Three orthogonal cuts  $|\tilde{H}(x_0, y, z, t)|^2$ ,  $|\tilde{H}(x, y_0, z, t)|^2$ , and  $|\tilde{H}(x, y, z_0, t)|^2$  of the magnitude of the rendered volume of a static semi-transparent phantom at arbitrary locations  $(x_0, y_0, z_0)$  are displayed in Fig. 5. Swept-source digital holographic OCT was implemented in the software Holovibes (www.holovibes.com). The software has five independent threads dedicated to 1- image acquisition, 2- holographic rendering (spatial demodulation), 3- time-to-frequency analysis (temporal demodulation), 4- image display, and 5- image saving. The acquisition of the raw video stream from the camera is buffered to avoid any image drop, in order to ensure the consistency of the temporal demodulation (Eq. 7). The propagation integral in Eq. 6 is computed by the function `cuffft2d()`; the discrete Fourier transform in Eq. 7 used to create local spectrograms is computed by the function `cuffft()`. Holovibes is written

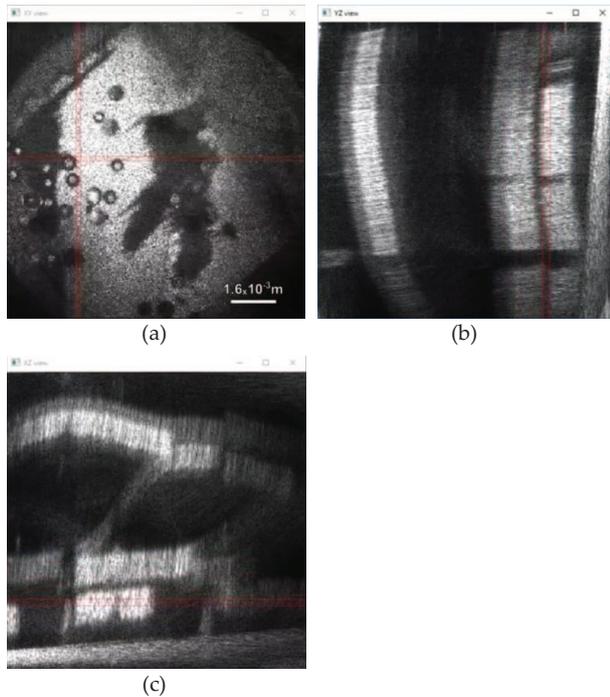


FIG. 5: Real-time image rendering and visualization in swept-source holographic optical coherence tomography. Visualization 2. Upper-left window : en-face section. Upper-right and Bottom-left windows : orthogonal B-scans. The measured sample is a static phantom made of 400-800 microns glass beads embedded in several layers of 65 micron-thick tape.

rendering	input throughput	output throughput
Fresnel	1 GB/s	10 Gvoxel/s
Fresnel + CC	512 MB/s	10 Gvoxel/s

TABLE I: Typical input/output throughput benchmarks with a Titan Xp card. Input data: 16-bit  $1024 \times 1024$ -pixel digitized interferograms. Fresnel : Fresnel transformation (Eq. 6). Time-frequency short-time Fourier transform on 256 frames (Eq. 7). CC : Chromatic compensation by interferogram resampling (Eq. 5).

in C++ and compiled with Microsoft Visual Studio 2017 and NVIDIA CUDA toolkit 9.1, all calculations are performed on  $2 \times 32$ -bit single precision floats per complex value.

In conclusion, we have demonstrated real-time computation and visualization of off-axis Fresnel transform digital holograms from an input stream of 16-bit, 1024-by-1024-pixel interferograms at a rate of 512 frames per second. The resulting stream of holograms was further processed by short-time Fourier transformation to form local spectrograms from 256 frame stacks at the maximum rate of 38 volumes per second. All calculations were performed on single-precision floating-point complex-valued arrays with one NVIDIA Titan Xp graphics card. The resulting software, Holovibes, was used in a swept-source holographic optical coherence tomography experiment for image rendering by Fresnel transformation and chromatic

compensation of the pitch variation with wavelength. The reported results demonstrate the scalability of digital holography for high-throughput computational volumic imaging in real-time.

This work was supported by LABEX WIFI (Laboratory of Excellence ANR-10-LABX-24) within the French Program Investments for the Future under Reference ANR-10-IDEX-0001-02 PSL, and European Research Council (ERC Synergy HELMHOLTZ, grant agreement #610110). The Titan Xp used for this research was donated by the NVIDIA Corporation.

- 
- [1] Daniel J Fechtig, Abhishek Kumar, Wolfgang Drexler, and Rainer A Leitgeb. Full range line-field parallel swept source imaging utilizing digital refocusing. *Journal of Modern Optics*, 62(21):1801–1807, 2015.
  - [2] Omer P Kocaoglu, Timothy L Turner, Zhuolin Liu, and Donald T Miller. Adaptive optics optical coherence tomography at 1 mhz. *Biomedical optics express*, 5(12):4186–4200, 2014.
  - [3] Dierck Hillmann, Hendrik Spahr, Carola Hain, Helge Sudkamp, Gesa Franke, Clara Pfäffle, Christian Winter, and Gereon Hüttmann. Aberration-free volumetric high-speed imaging of in vivo retina. *Scientific Reports*, 6:35209–, October 2016.
  - [4] Dierck Hillmann, Gesa Franke, Christian Lühns, Peter Koch, and Gereon Hüttmann. Efficient holoscopy image reconstruction. *Optics Express*, 20(19):21247–21263, 2012.
  - [5] Daniel J Fechtig, Branislav Grajciar, Tilman Schmoll, Cedric Blatter, Rene M Werkmeister, Wolfgang Drexler, and Rainer A Leitgeb. Line-field parallel swept source mhz oct for structural and functional retinal imaging. *Biomedical optics express*, 6(3):716–735, 2015.
  - [6] Laurin Ginner, Tilman Schmoll, Abhishek Kumar, Matthias Salas, Nastassia Pricoupenko, Lara M Wurster, and Rainer A Leitgeb. Holographic line field en-face oct with digital adaptive optics in the retina in vivo. *Biomedical Optics Express*, 9(2):472–485, 2018.
  - [7] Thomas Klein and Robert Huber. High-speed oct light sources and systems. *Biomedical optics express*, 8(2):828–859, 2017.
  - [8] Boris Považay, Bernd Hofer, Cristiano Torti, Boris Hermann, Alexandre R Tumlinson, Marieh Esmaeelpour, Catherine A Egan, Alan C Bird, and Wolfgang Drexler. Impact of enhanced resolution, speed and penetration on three-dimensional retinal optical coherence tomography. *Optics express*, 17(5):4134–4150, 2009.
  - [9] Hendrik Spahr, Clara Pfäffle, Peter Koch, Helge Sudkamp, Gereon Hüttmann, and Dierck Hillmann. Interferometric detection of 3d motion using computational subapertures in optical coherence tomography. *Optics express*, 26(15):18803–18816, 2018.
  - [10] Tomoyoshi Shimobaba, Yoshikuni Sato, Junya Miura, Mai Takenouchi, and Tomoyoshi Ito. Real-time digital holographic microscopy using the graphic processing unit. *Opt. Express*, 16(16):11776–11781, 2008.
  - [11] Lukas Ahrenberg, Andrew J. Page, Bryan M. Hennelly, John B. McDonald, and Thomas J. Naughton. Using commodity graphics hardware for real-time digital hologram view-reconstruction. *J. Display Technol.*, 5(4):111–

- 119, 2009.
- [12] Wenjing Gao and Qian Kemao. Parallel computing in experimental mechanics and optical measurement: A review. *Optics and Lasers in Engineering*, 50(4):608 – 617, 2012.
- [13] ID Reid, JJ Nebrensky, and PR Hobson. Challenges in using gpus for the reconstruction of digital hologram images. In *Journal of Physics: Conference Series*, volume 368, page 012025. IOP Publishing, 2012.
- [14] Takashi Kakue, Tomoyoshi Shimobaba, and Tomoyoshi Ito. High-speed parallel phase-shifting digital holography system using special-purpose computer for image reconstruction. 9495:94950N, 2015.
- [15] E.N. Leith and J. Upatnieks. Reconstructed wavefronts and communication theory. *JOSA*, 52(10):1123–1128, 1962.
- [16] AF Fercher, W Drexler, CK Hitzenberger, and T Lasser. Optical coherence tomography - principles and applications. *Reports of Progress in Physics*, 66:239–303, February 2003.
- [17] Pietro Ferraro, Sergio De Nicola, Andrea Finizio, Giovanni Pierattini, and Giuseppe Coppola. Recovering image resolution in reconstructing digital off-axis holograms by fresnel-transform method. *Applied Physics Letters*, 85(14):2709–2711, 2004.
- [18] Vít Lédl, Pavel Psota, František Kaván, Ondřej Matoušek, and Pavel Mokrý. Surface topography measurement by frequency sweeping digital holography. *Applied Optics*, 56(28):7808–7814, 2017.
- [19] Lingfeng Yu and Myung K Kim. Wavelength-scanning digital interference holography for tomographic three-dimensional imaging by use of the angular spectrum method. *Optics letters*, 30(16):2092–2094, 2005.
- [20] Tatiana Latychevskaia and Hans-Werner Fink. Practical algorithms for simulation and reconstruction of digital inline holograms. *Applied optics*, 54(9):2424–2434, 2015.

# MOTION COMPENSATION IN DIGITAL HOLOGRAPHY FOR RETINAL IMAGING

Julie Rivet<sup>1,2</sup> Guillaume Tochon<sup>1</sup> Serge Meimon<sup>3</sup> Michel Paques<sup>4</sup> Thierry Géraud<sup>1</sup> Michael Atlan<sup>2</sup>

<sup>1</sup> EPITA Research and Development Laboratory (LRDE)

<sup>2</sup> ESPCI Paris, PSL Research University, Sorbonne Université, CNRS, Institut Langevin

<sup>3</sup> ONERA – the French Aerospace Lab

<sup>4</sup> Institut de la Vision, CIC 1423, INSERM UMR-S 968, CNRS, Sorbonne Université

## ABSTRACT

The measurement of medical images can be hindered by blur and distortions caused by the physiological motion. Specially for retinal imaging, images are greatly affected by sharp movements of the eye. Stabilization methods have been developed and applied to state-of-the-art retinal imaging modalities; here we intend to adapt them for coherent light detection schemes. In this paper, we demonstrate experimentally cross-correlation-based lateral and axial motion compensation in laser Doppler imaging and optical coherence tomography by digital holography. Our methods improve lateral and axial image resolution in those innovative instruments and allow a better visualization during motion.

**Index Terms**— Motion Compensation, Cross-correlation, Digital Holography, Optical Coherence Tomography, Laser Doppler Holography.

## 1. INTRODUCTION

Ophthalmology is an ancient science which has benefited from technological progresses in optics. The first images of eye fundus in vivo by holographic ophthalmoscopy [1, 2] paved the way to the development of time-averaged holography for retinal vessels imaging. This non-invasive and low-exposure optical interferometric imaging technology has made accessible the retinal imaging, leading to better diagnosis of several diseases like glaucoma, macular degeneration and diabetes. Later, retinal blood flow contrasts were also successfully rendered by narrowband [3] and wideband [4] digital holography.

Digital holography relies on the acquisition and the processing of interference images created by the superposition of two light waves from the same source. The first wave is a reference, whereas the second is back-scattered by the retina. Several facets of digital holography have been developed during the past few years to take maximum advantage of phase variations between both waves which contain numerous information about the retina. Optical Coherence Tomography (OCT) [5, 6] permits optical sectioning of retinal layers. On the other hand, laser Doppler imaging [7] focuses on the detection of local dynamics.

Digital holography faces several issues altering images quality. Eye motion during image acquisition distorts the

signal. In order to stabilize images of retinal features, image registration can be used. It consists in finding the geometric transformation to align two images. Medical image registration techniques have been developed during the past years [8]. Although most of the literature in this field concerns radiological modalities, several researches have been led in retinal imaging. The most widely used technique is feature-based registration of eye fundus images [9, 10]. However in our case, feature-based registration is not suitable because objects of interest can disappear or even be replaced by another structure if the movement of the eye is too large, which could bias the search of feature points. This article is focused on cross-correlation stabilization to compensate lateral and axial motion of the whole image structure in laser Doppler holography and swept-source holographic OCT, respectively.

## 2. HOLOGRAPHIC IMAGING

### 2.1. Setup

Optical interferograms are recorded with a laser Doppler instrument [3] in human eyes, and with a holographic OCT instrument in synthetic phantoms. For the laser Doppler setup, the source is a laser diode emitting at the wavelength  $\lambda = 785$  nm. The camera records  $2048 \times 2048$ -pixel images corresponding to the  $(x, y)$  plane, at a frame rate of  $\omega_S/(2\pi) = 80$  Hz with 8 bit/pixel quantization. For OCT, the beam is emitted by a tunable laser and varies from  $\omega_1$  to  $\omega_2$ , which are linked to the wavelengths  $\lambda_1 = 870$  nm and  $\lambda_2 = 820$  nm by  $\omega = 2\pi c/\lambda$ , where  $c$  is the speed of light, with a sweep time  $T = 0.5$  s. The camera records  $1024 \times 1024$ -pixel images at a frame rate of  $\omega_S/(2\pi) = 512$  Hz with 16 bit/pixel quantization. Digitized interferograms from the camera are processed in real-time with the Holovibes<sup>1</sup> software to compute and visualize holograms, and saved for offline processing.

### 2.2. Acquisition of interferograms

A Mach-Zehnder interferometer has been used to record interferograms. It consists in making interfere two beams from the same laser source. The source is split between reference and object arms. The light wave from the object arm is backscattered once the sample is reached and interferes in the camera

<sup>1</sup><http://holovibes.com>

plane with the beam in the reference arm. The interferogram  $I$  acquired in the camera plane  $(x, y)$  at time  $t$  is expressed as:

$$I(x, y, t) = |E(x, y, t) + E_{\text{LO}}(x, y, t)|^2, \quad (1)$$

where the object field is  $E$  and the reference field is  $E_{\text{LO}}$ . The phase difference between both optical fields is noted  $\phi$ . The off-axis configuration of the setups allows a spatial separation between the constant term, and the cross-beating interferometric contribution, which will produce the image of interest, and its conjugate [11] (see Fig. 2a).

### 2.3. Processing of optically-acquired interferograms

*Interferogram rescaling with wavelength.* In order to avoid the distortion of the signal, the impact of the sweep in OCT needs to be considered to propagate the interferograms  $I$  from the camera plane to the retina or image plane. In fact the size of the pixels in the image plane depends on the wavelength of the beam and of the distance between camera and image planes [12, 13]:  $d' = \lambda z / (Nd)$ , where  $d$  and  $d'$  are the lateral size of a pixel from camera plane and from image plane, respectively,  $z$  is the distance between both planes and  $N$  is the number of pixels in one lateral dimension. Because of the sweep of the source, the pixels of the image plane shrink with wavelength (arrow (1) of Fig. 1). To circumvent lateral field variation with wavelength, each interferogram is resampled by linear interpolation of the calculation grid with a different pitch (arrow (2) of Fig. 1). The rescaled interferogram is:

$$I'(x, y, t) = I(x\lambda/\lambda_1, y\lambda/\lambda_1, t), \quad (2)$$

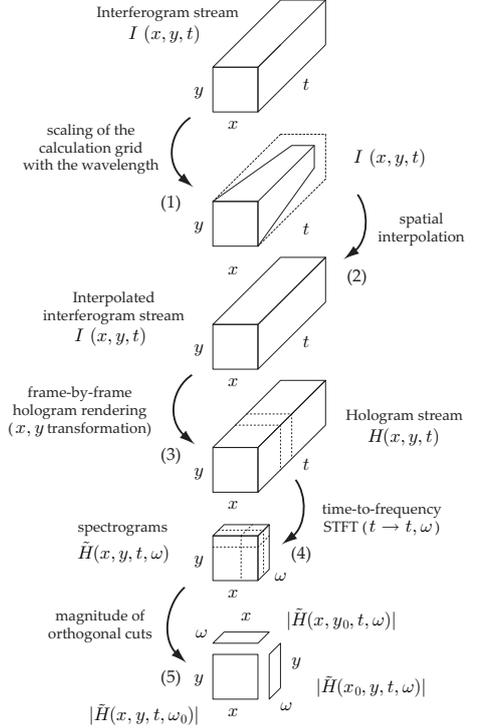
where  $\lambda$  is the current wavelength. In the case of laser Doppler imaging, the optical wavelength  $\lambda$  is kept constant, hence the interferogram does not need to be rescaled:  $I' \equiv I$ .

*Spatial demodulation by Fresnel transform.* The propagation of the fields from camera to image plane is carried out by a Fresnel transform [11] (arrow (3) of Fig. 1), which gives the hologram  $H(x, y, t)$ :

$$H(x, y, t) = \frac{i}{\lambda z} \exp(-ikz) \iint I'(x', y', t) \times \exp\left[\frac{-i\pi}{\lambda z} ((x - x')^2 + (y - y')^2)\right] dx' dy'. \quad (3)$$

The argument of the complex-valued hologram in the cross-contribution region is the phase difference  $\phi$  between the optical fields  $E$  and  $E_{\text{LO}}$  in the image plane.

*Temporal demodulation by Fourier transform.* In swept-source OCT, the instantaneous beating frequency  $\partial\phi/\partial t = \omega$  scales up linearly with axial depth  $z$  [5], whereas in laser Doppler, it describes local velocities of the scatterers [7]. Hence for both methods, temporal signal demodulation is performed by short-time Fourier transform (STFT) (arrow (4) of Fig. 1). Time-and-space-dependent spectrograms



**Fig. 1:** Sketch of the process of image propagation. The raw data in the camera plane are transferred in the image plane and are expressed with depth.

$\tilde{H}(x, y, t, \omega)$  are calculated from the stream of intermediate holograms  $H(x, y, t)$ :

$$\tilde{H}(x, y, t, \omega) = \int H(x, y, \tau) g_T(t - \tau) e^{-i\omega\tau} d\tau, \quad (4)$$

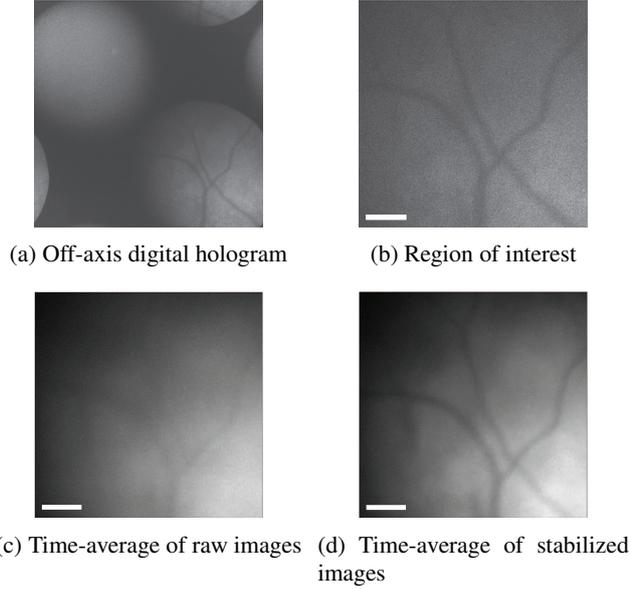
where  $g_T(t)$  is a time gate of width  $T$  at time  $t$ . Then, the envelope  $|\tilde{H}|$  of the STFT of  $H$  is formed (arrow (5) of Fig. 1). The quantities  $|\tilde{H}(x, y, t, \omega)|$  are the images in laser Doppler and OCT on which motion compensation will be performed.

## 3. MOTION COMPENSATION METHODS

### 3.1. Lateral Motion Compensation

We form Doppler-contrast images at the Nyquist frequency  $\omega_S/2$  by calculating Eq. (4) with a time gate  $g_T$  of two points ( $T = 25$  ms). To compensate lateral motion (in  $(x, y)$  plane), we use a cross-correlation-based stabilization algorithm. The human eye is constantly subject to deviations, saccades and tremors. Added to heartbeat and respiration motion, those movements can shift or even remove the object of interest from the image during several frames. Thus, a processing chain has been built in order to keep important structures in the center of the en-face image.

A reference image  $f_1$  is made by averaging  $n = 10$  consecutive images, and compared to a moving average  $f_2$  of



**Fig. 2:** Retinal image stabilization by lateral  $(x, y)$  motion compensation of intensity holograms (Section 3.1). The white scale bars indicate  $200 \mu\text{m}$ .

$n = 10$  images in the neighborhood of the current image:

$$\begin{aligned} f_1(x, y) &= \sum_{t=1}^n |\tilde{H}(x, y, t, \omega_S/2)| / n, \\ f_2(x, y) &= \sum_{t=1}^n |\tilde{H}(x, y, t + \tau, \omega_S/2)| / n. \end{aligned} \quad (5)$$

The quantitative comparison is performed with a normalized and centered cross-correlation [14]. The difference between the position of the maximum value and the center of the correlation matrix  $\gamma$  gives the displacement between the two images. Resulting shifts describe retina lateral motion.

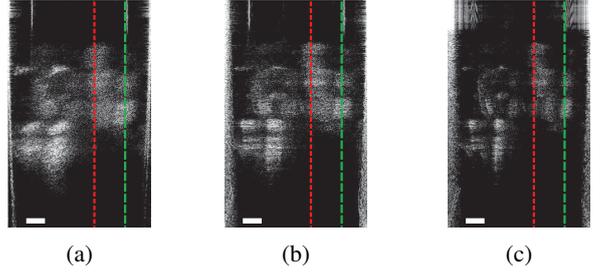
### 3.2. Axial Motion Compensation

We form OCT images by calculating Eq. (4) with a time gate  $g_T$  of 256 points ( $T = 0.5$  s). Axial motion during the recording process has a negative impact on the reconstruction of OCT images: when an axial drift occurs, the depth information encoded in the beating frequency of the interferogram is inaccurate, adding an offset phase to the signal and decreasing in-depth accuracy. The principle of the compensation method is to identify where the phase shift due to the axial motion occurs by using STFT [15] on 20-point windows. Indeed, a global axial displacement of the sample during acquisition can be spotted in the time Fourier domain, where it corresponds to a frequency shift.

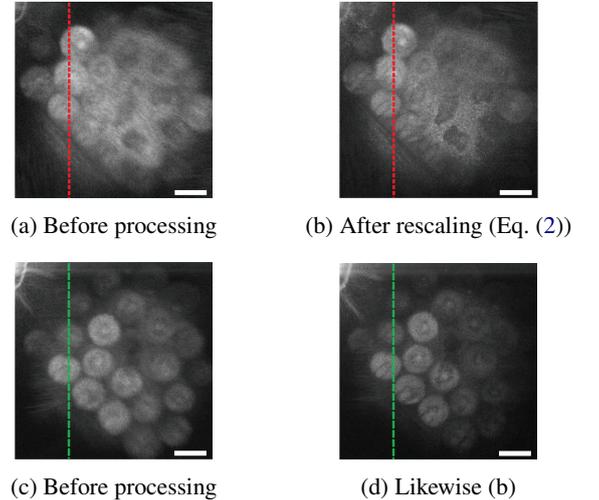
For  $y = y_0$ , the reference image  $f_1$  and the sliding image  $f_2$  are defined as the modulus of the STFT of  $H(x, y_0, t)$  whose sub-windows are fixed or moving, respectively:

$$\begin{aligned} f_1(x, \omega) &= |\int H(x, y_0, \tau) g_T(t'_0 - \tau) e^{-i\omega\tau} d\tau|, \\ f_2(x, \omega) &= |\int H(x, y_0, \tau) g_T(t'_0 - \tau) e^{-i\omega\tau} d\tau|, \end{aligned} \quad (6)$$

where  $g_T$  is a 20-point Hanning window, and  $t'_0$  and  $t''_0$  are the centers of the moving and the reference windows, respec-



**Fig. 3:** Axial motion compensation in swept-source holographic OCT of glass beads rolled in a tape layer. The registered holograms are slices in the  $(x, \omega)$  plane, for the  $y$ -cut in Fig. 4. Figures represent the sample: (a) without any correction, (b) with interferogram rescaling (Eq. (2)), and (c) with interferogram rescaling and axial motion compensation (Section 3.2). The white scale bars indicate  $0.2$  mm.



**Fig. 4:** (a,b) En-face  $(x, y)$  images of glass beads for the first red  $z$ -cut in Fig. 3, and (c,d) for the second green  $z$ -cut in Fig. 3. The white scale bars indicate  $1.5$  mm.

tively. The quantitative comparison between  $f_1$  and  $f_2$  is performed by calculating cross-correlation whose maximum position gives a frequency shift  $\Delta\omega(t'_0)$  corresponding to the derivative of the phase caused by movement at time  $t'_0$ . Axial motion compensation is then performed by subtracting the cumulated phase shift  $\int_{t'_0}^{t''_0} \Delta\omega(t) dt$  from the phase  $\phi(t'_0)$ . This procedure is repeated for all times  $t'_0$  in the time gate  $g_T$  used in Eq. (4) of width  $\omega_S T / (2\pi) = 256$  points [15].

## 4. RESULTS

Fig. 2 shows retinal vessels acquired by laser Doppler imaging. Fig. 2a is the whole  $2048 \times 2048$  image. The cross-beating interferometric contribution, which is the object of interest in the right bottom side of the image, is spatially separable from the other parts because of the off-axis configuration. Fig. 2b shows the focus on retinal vessels. Ves-

sels are moving in the image and sometimes disappear during several frames. Then, vessels do not appear clearly on the time-averaged image of 324 consecutive frames (Fig. 2c). After lateral stabilization, vessels are more visible on the time-averaged image: motion compensation is efficient. Although motion compensation is less efficient when the image changes too often, it improves overall image quality.

Fig. 3 and Fig. 4 are swept-source OCT images of samples composed of 1.5 mm diameter glass beads rolled in a single tape layer. Fig. 3 shows the in-depth profile  $(x, \omega)$  of this sample corresponding to the y-axis represented by dotted lines in Fig. 4. Fig. 4 shows the en-face  $(x, y)$  images at two different depths corresponding to the dotted lines in Fig. 4. Fig. 4a and 4b correspond to the red dotted line (depth of 1.1 mm) in Fig. 3, and Fig. 4c and 4d correspond to the green dotted line (depth of 1.5 mm) in Fig. 3. In Fig. 4a and 4b, the target layer is located between tape and beads: the beads on the left are at the same elevation as the tape layer on the right which starts to be sectioned.

A difference of lateral resolution is visible on images of Fig. 4. The interferogram rescaling allows the contours of the beads to be cleaner. Comparing Fig. 3a and Fig. 3b shows an improvement in axial resolution, which is also noticeable on en-face images: in Fig. 4c, beads on the right seem almost at the same elevation than beads on the left, because the depth accuracy is low, while in Fig. 4d, right beads clearly belong to a different layer. Axial motion correction can be seen in Fig. 3c: the accuracy is improving and the different layers are better separated. Besides, the optical bench will be strengthened to reduce mechanical noise, in order to further improve the axial accuracy of depth images.

## 5. CONCLUSION AND PROSPECTS

We have demonstrated lateral and axial motion compensation in laser Doppler holography and holographic OCT by cross-correlation stabilization, respectively. This method is suited to our images and cancels efficiently the effects of motion. Besides, the reported results are non-iterative and compatible with real-time processing at high throughput on graphics processing units. A real-time implementation of lateral motion compensation has been implemented<sup>2</sup>. The reported results in motion compensation pave the way towards the design of high resolution computational imaging for the retina in real-time by digital holography.

Aside from motion, retinal imaging suffers from phase distortions caused by the eye: the light backscattered by the retina crosses lenses in the eye, which can cause aberrations. As a future work, we will try to correct optical aberrations.

**Funding:** This work has been supported the European Research Council (ERC Synergy HELMHOLTZ, grant agreement #610110).

<sup>2</sup><https://youtu.be/RhVPXBnPhXc>

## References

- [1] Joseph L Calkins and Carl D Leonard, "Holographic recording of a retina using a continuous wave laser," *Investigative Ophthalmology & Visual Science*, vol. 9, no. 6, pp. 458–462, 1970.
- [2] R Lawrence Wiggins, K Diane Vaughan, and Gerhart B Friedmann, "Fundus camera holography of retinal microvasculature," *Arch Ophthalmol*, vol. 88, pp. 75–79, 1972.
- [3] Manuel Simonutti et al., "Holographic laser doppler ophthalmoscopy," *Optics letters*, vol. 35, no. 12, pp. 1941–1943, 2010.
- [4] Mathilde Pellizzari et al., "High speed optical holography of retinal blood flow," *Optics letters*, vol. 41, no. 15, pp. 3503–3506, 2016.
- [5] Adolf F Fercher, Wolfgang Drexler, Christoph K Hitzenberger, and Theo Lasser, "Optical coherence tomography-principles and applications," *Reports on progress in physics*, vol. 66, no. 2, pp. 239, 2003.
- [6] David Huang et al., "Optical coherence tomography," *Science (New York, NY)*, vol. 254, no. 5035, pp. 1178, 1991.
- [7] Caroline Magnain et al., "Holographic laser doppler imaging of microvascular blood flow," *JOSA A*, vol. 31, no. 12, pp. 2723–2735, 2014.
- [8] Max A. Viergever et al., "A survey of medical image registration – under review," *Medical Image Analysis*, vol. 33, pp. 140–144, 2016.
- [9] Chengyin Liu, Jiayi Ma, Yong Ma, and Jun Huang, "Retinal image registration via feature-guided gaussian mixture model," *JOSA A*, vol. 33, no. 7, pp. 1267–1276, 2016.
- [10] Roziana Ramli et al., "Feature-based retinal image registration using d-saddle feature," *Journal of healthcare engineering*, vol. 2017, 2017.
- [11] Nicolas Verrier and Michael Atlan, "Off-axis digital hologram reconstruction: some practical considerations," *Applied optics*, vol. 50, no. 34, pp. H136–H146, 2011.
- [12] Pietro Ferraro et al., "Controlling image size as a function of distance and wavelength in fresnel-transform reconstruction of digital holograms," *Optics Letters*, vol. 29, no. 8, pp. 854–856, 2004.
- [13] Vít Lédl, Pavel Psota, František Kaván, Ondřej Matoušek, and Pavel Mokrý, "Surface topography measurement by frequency sweeping digital holography," *Applied optics*, vol. 56, no. 28, pp. 7808–7814, 2017.
- [14] Dirk Padfield, "Masked object registration in the Fourier domain," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2706–2718, 2012.
- [15] Dierck Hillmann et al., "Common approach for compensation of axial motion artifacts in swept-source OCT and dispersion in Fourier-domain OCT," *Optics express*, vol. 20, no. 6, pp. 6761–6776, 2012.

# Experimental digital Gabor hologram rendering by a model-trained convolutional neural network

J. Rivet,<sup>1</sup> A. Taliercio,<sup>1</sup> C. Fang,<sup>1</sup> G. Tochon,<sup>2</sup> T. Géraud,<sup>2</sup> JP. Huignard,<sup>1</sup> and M. Atlan<sup>1</sup>

<sup>1</sup> *Centre National de la Recherche Scientifique (CNRS) UMR 7587,*

*Institut Langevin. Paris Sciences et Lettres (PSL) Research University. Fondation Pierre-Gilles de Gennes, Institut National de la Santé et de la Recherche Médicale (INSERM) U 979, Université Pierre et Marie Curie (UPMC), Université Paris 7. École Supérieure de Physique et de Chimie Industrielles ESPCI Paris - 1 rue Jussieu. 75005 Paris. France*

<sup>2</sup> *EPITA Research and Development Laboratory (LRDE),*

*14-16 rue Voltaire, F-94270 Le Kremlin-Bicetre, France*

(Dated: May 20, 2020)

Digital hologram rendering can be performed by a convolutional neural network, trained with image pairs calculated by numerical wave propagation from sparse generating images. 512-by-512 pixel digital Gabor magnitude holograms are successfully estimated from experimental interferograms by a standard UNet trained with 50,000 synthetic image pairs over 70 epochs.

Convolutional neural networks already have demonstrated their potential for digital hologram rendering from optically-acquired interferograms in free-space propagation conditions [1–4] and through scattering media [5–7]. Our aim here is to determine whether an auto-encoder convolutional neural network, a UNet [8], can be trained over a synthetic database for digital hologram rendering from experimental interferograms. A model of wave propagation is used to create synthetic Gabor interferograms and synthetic Gabor magnitude holograms from random images. This image formation model is based on angular spectrum propagation and magnitude calculation of the wave field from the object to the sensor array, and from the sensor to the object.

In contrast with previously reported computational image rendering schemes with convolutional neural networks, where image formation is statistically inferred through experimental data [1–4], in our approach it is inferred from synthetic data created by physical modeling of wave interference and propagation. Since the UNet training strategy relies on the strong use of a large and diverse database [8], training on synthetic data alleviates the need for numerous experimental data and data augmentation.

The convolutional neural network used in this study is (sketched in Fig. 1) is a standard UNet [8] with an input image of  $512^2$  pixels, a depth of 7 down sampling blocks and 7 up sampling blocks. Convolution kernels are 3-by-3-by- $n$  pixels, where  $n$  is the number of channels of the input feature map. The first set of 16 kernels generates a feature map of  $n = 16$  channels from the input image which has only  $n = 1$  channel. In the down sampling part, the lateral size of the features is divided by two and the number of channels  $n$  is multiplied by two between blocks. In the up sampling part, the lateral size of the features is multiplied by two and the number of channels  $n$  is divided by two between

blocks. Mirror features from the down sampling part are concatenated to their up sampling counterparts. The UNet is trained with 50,000 image pairs (among which 15% are used for validation purposes). The chosen loss function is the mean-square error between predicted image  $H'$  and actual training output  $H$  during the validation process. It is used to measure their inconsistency; the optimization (or deep learning) of the network consists in finding the set of network weights for which this loss function is minimum. The learning rate controls how much the weights of the network are adjusted with respect to the gradient of the loss function.

We construct a database of training input and output image pairs by the procedure illustrated in the flowchart from Fig. 2. A square generating image  $A$  of  $512^2$  pixels that describes the amplitude transmission function of a synthetic object is constructed by setting a given number  $N$  of source points at random locations with random brightness on a black background, and spatial filtering by a circular aperture in the Fourier plane. The diameter of the aperture is one half of the diagonal of the reciprocal plane. The values of the array  $A$  are positive real numbers. A synthetic Gabor interferogram  $I$  is calculated from this generating image  $A$  by angular spectrum propagation [10] of the wave field described by the transmission function  $A$  with a distance parameter  $-z$ , followed by a rectification consisting of taking the magnitude of the complex-valued array points

$$I(x, y) = \left| \iint \mathcal{F}A(k_x, k_y) e^{-ik_z z} e^{ik_x x} e^{ik_y y} dk_x dk_y \right| \quad (1)$$

where  $(x, y)$  are the pixels of arrays, and  $\mathcal{F}A(k_x, k_y)$  is the two-dimensional Fourier transform of  $A(x, y)$ . The wave vector projections ( $k_x$ ,  $k_y$ , and  $k_z$ ) along lateral and axial directions ( $x$ ,  $y$ , and  $z$ ) satisfy  $k_z^2 = k^2 - k_x^2 - k_y^2$ , with  $k = 2\pi/\lambda$ , and  $\lambda$  is the optical wavelength. A synthetic magnitude hologram  $H$  is calculated from each synthetic interferogram  $I$  by angular spectrum propagation of the

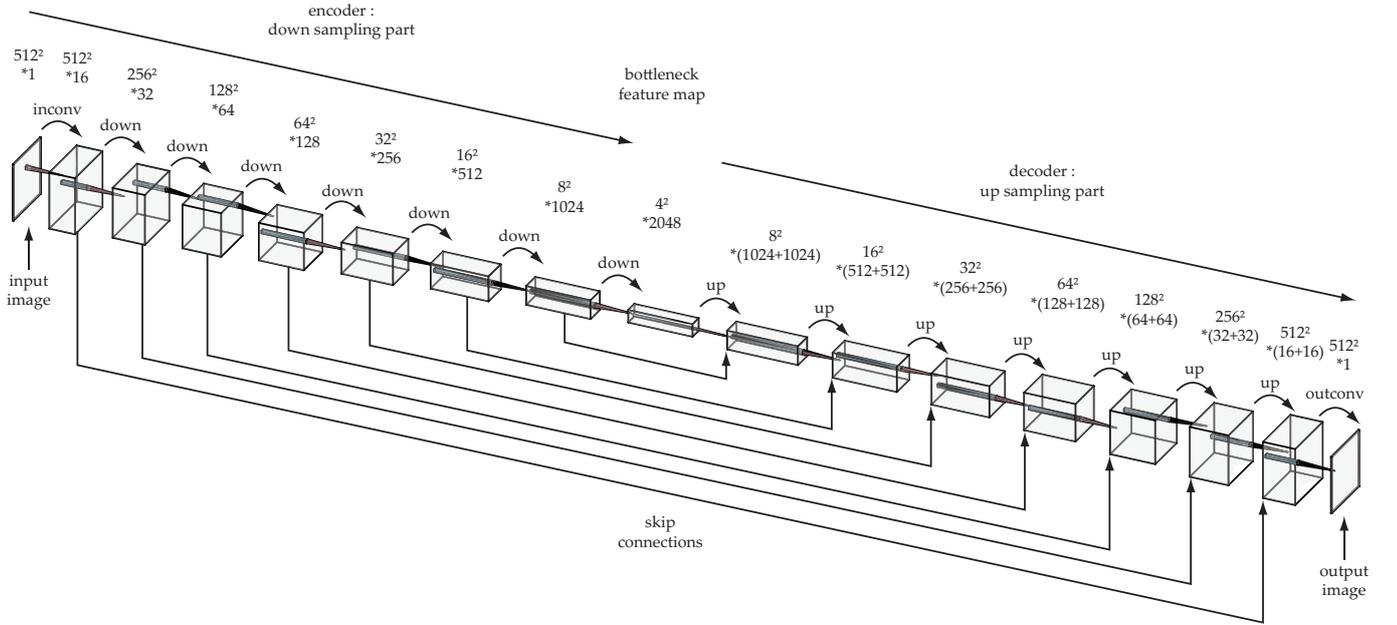


FIG. 1: Topology of the convolutional neural networks trained with synthetic, input/output image pairs, and used for experimental hologram rendering. Standard UNet [8] with a positive real input (and output) image of  $512^2$  pixels (width = height = 512 pixels), a depth of 7 down sampling blocks in the encoding part and 7 up sampling blocks in the decoding part. Convolution kernels are 3-by-3-by- $n$  pixels, where  $n$  is the number of channels of the input feature map. The first set of kernels generates  $n = 16$  channels from the input image. In the down sampling part, the lateral size of the features is divided by two and the number of channels is multiplied by two between feature maps. Down sampling transitions (noted "down") include max pooling, and two iterations of convolution and rectification (ReLU) [9]. In the up sampling part, the lateral size of the features is multiplied by two and the number of channels is divided by two between blocks. Up sampling transitions (noted "up") include a convolution transpose, and two iterations of convolution and rectification. Mirror features from the down sampling part of the network are concatenated to their up sampling counterparts through skip connections that allow feature maps to pass through the bottleneck. The boxes represent feature maps, the numbers on top of each box are their respective width, height, and depth  $n$ . Flowchart courtesy of <http://alexlenail.me/NN-SVG/LeNet.html>

wave field described by  $I$  with a distance parameter  $+z$ , followed by rectification.

$$H(x, y) = \left| \iint \mathcal{F}I(k_x, k_y) e^{+ik_z z} e^{ik_x x} e^{ik_y y} dk_x dk_y \right| \quad (2)$$

where  $\mathcal{F}I(k_x, k_y)$  is the two-dimensional Fourier transform of  $I(x, y)$ . These operations generate a positive, real-valued image triplet  $(A, I, H)$ , displayed in Fig. 3. We ought to teach wave field propagation to a UNet, by deep learning over a large training database of  $M$  randomly generated input/output image pairs  $(I, H)$ . The number of source points  $N$  in each generating image  $A$  is logarithmically-spaced from 1 to one-tenth of  $512^2$ .

By following the same construction procedure as for the generation of the training database, image couples  $(I, H)$  are generated from a set of arbitrary images  $A$  for validation purposes. The training procedure is stopped after 70 iterations of the optimization process over the whole training database (epochs), with a learning rate of 0.1, when the network output  $H'$  for an input image  $I$  becomes similar to the model-rendered magnitude hologram  $H$ .

Gabor interferograms  $I$  measured from a preparation of *C. elegans* roundworms with a digital holographic Gabor microscope, sketched in Fig. 4, are then used to compare the network output  $H'$  to magnitude holograms  $H$  reconstructed by angular spectrum propagation (Eq. 2). In the experiments, the radiation wavelength  $\lambda$  is 658 nm, the pixel pitch is 5.5 microns, and the reconstruction distance is  $z = 0.065$  m. 512-by-512-pixel interferograms  $I$  are cropped from 2048-by-2048-pixel frames in a region of interest of the sensor array. A database of image couples  $(I, H)$  is then constructed from a set of recorded Gabor interferograms  $I$  and their magnitude hologram counterparts  $H$ , reconstructed by angular spectrum propagation from  $I$ , followed by rectification (Eq. 2). Examples of network estimates  $H'$  at several training iterations (epochs) for an input interferogram  $I$ , alongside the calculated magnitude hologram  $H$  (Eq. 2) are displayed in Fig. 5. All the training dataset  $(I, H)$  is calculated for  $z = 0.065$  m. It is worth remarking that training the network over several reconstruction distances degrades the prediction accuracy.

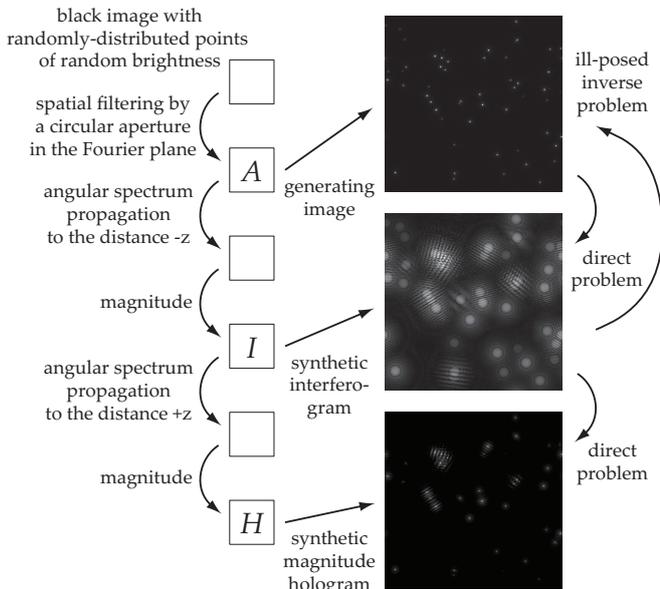


FIG. 2: Flowchart of the synthetic database creation. An image pair  $(I, H)$  is calculated numerically from a random generating image  $A$ . This process is iterated for each random generating image to create the whole training database.

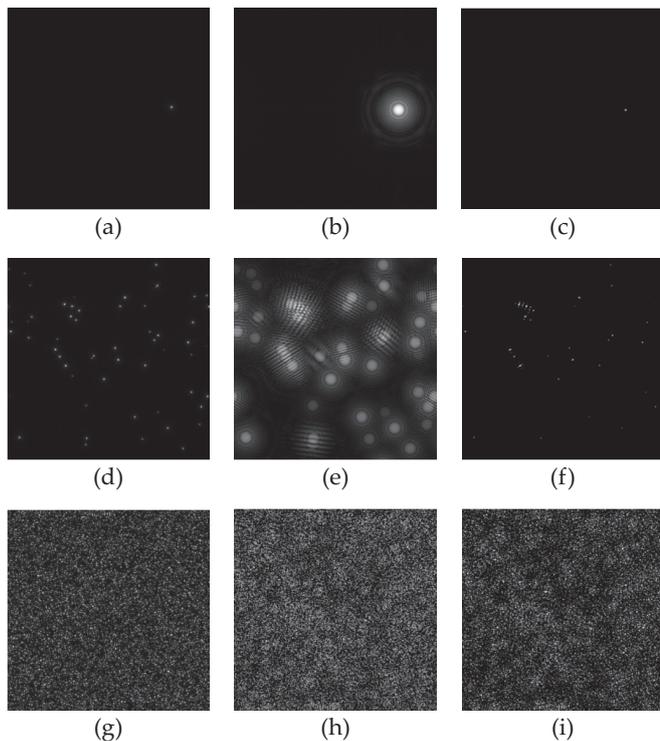


FIG. 3: Examples of database of image triplets  $(A, I, H)$ . Generating images  $A$  (left column, a,d,g), synthetic interferograms  $I$  (center column, b, e, h), synthetic magnitude holograms  $H$  (right column, c, f, i). The number of random points in the generating images is  $N = 1$  (top row, a, b, c),  $N = 58$  (center row, d, e, f),  $N = 512^2/10$  (bottom row, g, h, i). A movie of generated image triplets  $(A, I, H)$  illustrating the distribution of the whole range of number of source points is displayed in [Visualization 1](#).

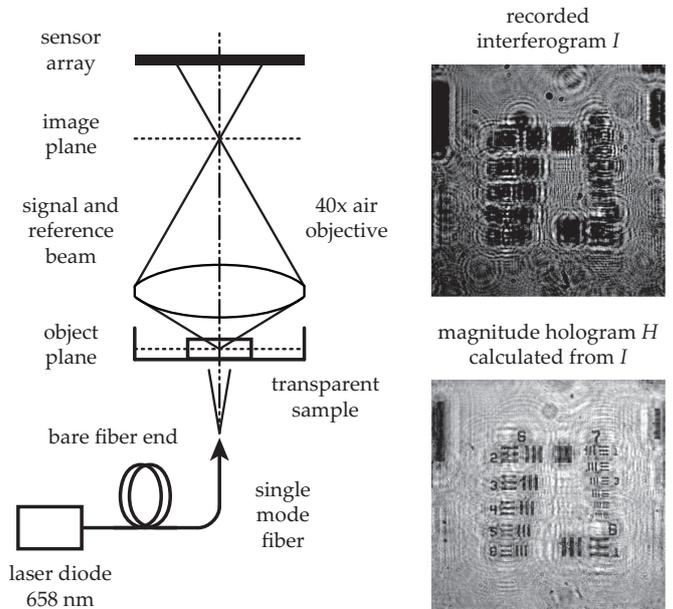


FIG. 4: Optical arrangement of the Gabor inline holographic microscope used to image transparent worms. A petri dish with growth medium placed in the object plane is illuminated by red laser light. It transmits light collected by a microscope objective, which creates an image conjugate of the sample in a plane between the lens and the sensor array of a camera. The recorded interferogram  $I$  (top, right) of a resolution target placed in the object plane, yields the magnitude hologram  $H$  (bottom, right) via Eq. 2.

Training the network with synthetic interferograms and reconstructed holograms  $(I, H)$  image pairs teaches the network to estimate the solution of the transformation of Eq. 2, for a given depth  $z$ , which already has an analytical solution. Yet this solution is cluttered with a spurious contribution. The quality of single-shot magnitude holograms reconstructed from Gabor in-line interferograms is degraded by the superposition of a twin image [11–13]: the ripples observed in the neighborhood of the worms in Fig. 5(b) are the twin image of the roundworms in focus. The sum of the diffracted object wave beating against the reference wave, and their conjugate are present in the recorded interferogram  $I$ , hence the object wave reconstructed  $+z$  is stained with an additive diffraction pattern, which creates a twin image at the reconstruction distance  $-z$ . Those ripples are also present in the image  $H'$  in Fig. 5(f), estimated by the neural network.

The convolutional neural network proves capable of mimicking standard hologram rendering with a high level of accuracy (Fig. 5(f) vs. Fig. 5(b)). We also wanted to assess whether it would also provide high quality estimates of solutions to the twin-image problem. Twin-image elimination by neural network rendering

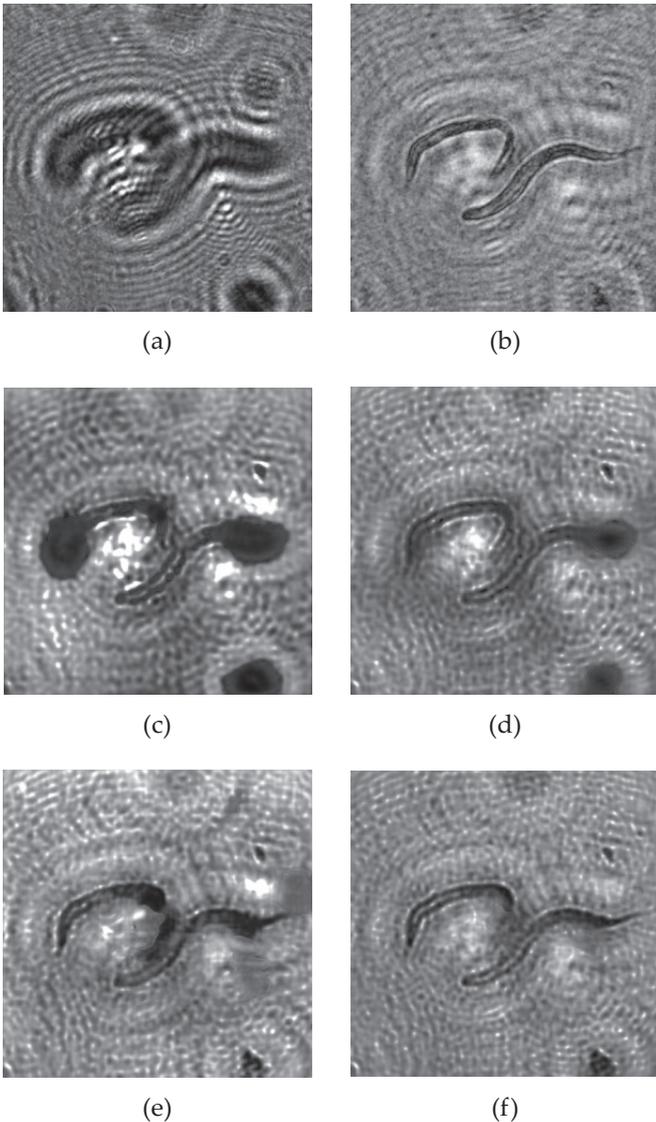


FIG. 5: (a) Optically-acquired digital interferogram  $I$  of transparent worms. (b) Magnitude hologram  $H$  calculated by angular spectrum propagation from  $I$ . Output image estimate  $H'$  after 10 (c), 40 (d), 50 (e), and 70 (e) training epochs, over 50,000 synthetic image pairs  $(I, H)$ . A movie of image triplets  $(I, H, H')$  is displayed in [Visualization 2](#).

was previously reported for hologram estimation by a convolutional neural network [3]. In this approach, the network was trained with interferograms measured experimentally and with calculated holograms from which the twin-image was removed by an experimental and numerical iterative multi-height phase recovery scheme [14]. This suggests that UNets may be able to estimate solutions to ill-posed inverse problems beyond the ones for which the normal operator is a convolution [15]. The inverse problem that needs to be solved is to determine the possible positive real-valued images (object amplitude transmission functions) to reproduce

a given measured Gabor amplitude interferogram. Our network was also trained with  $(I, A)$  pairs instead of  $(I, H)$ , ie. onto the inverse problem of image formation (Fig. 2), switching the calculated magnitude holograms  $H$  for generating images  $A$ , naturally devoid of twin image. Yet it did not enable the neural network to estimate twin-image-free magnitude holograms  $H'$  from inline interferograms inputs  $I$ . This approach failed to reconstruct twin-image-free Gabor holograms. This problem is most often ill-posed, which means that many object transmission estimates may produce the same Gabor amplitude interferogram. Yet the direct problem, which is the formation of an interferogram by a given transmission function, has an analytical formulation. Adding regularization constraints [16] has emerged as the standard procedure for iterative image reconstruction algorithms [11–13, 15]. It may be also prove useful for hologram rendering by convolutional networks.

In conclusion, digital image rendering in Gabor holography can be performed by a convolutional neural network trained with a fully synthetic database formed by image pairs generated randomly, and linked by a numerical model of in-line angular spectrum propagation of a scalar wave field from the object to the sensor array, and magnitude calculation. Gabor holograms of microscopic worms are successfully predicted from experimental interferograms by a UNet trained with 50,000 random image pairs. Two main caveats apply to the use of a standard Unet for image rendering : the results were obtained for a fixed reconstruction distance, and twin-image elimination could not be achieved by training the network with image pairs from the inverse problem.

This work was supported by LABEX WIFI (Laboratory of Excellence ANR-10-LABX-24) within the French Program Investments for the Future under Reference ANR-10-IDEX-0001-02 PSL, and European Research Council (ERC Synergy HELMHOLTZ, grant agreement #610110). The Titan Xp used for this research was donated by the NVIDIA Corporation.

We are thankful to Vincent Galy for providing the round worms *C. elegans*, and we acknowledge valuable assistance from Armelle Rancillac, Nicolas Letort and Stéphanie Rind.

- 
- [1] Ayan Sinha, Justin Lee, Shuai Li, and George Barbasthis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, 2017.
  - [2] Ryoichi Horisaki, Kazuki Fujii, and Jun Tanida. Single-shot and lensless complex-amplitude imaging with inco-

- herent light based on machine learning. *Optical Review*, 25(5):593–597, 2018.
- [3] Yair Rivenson, Yibo Zhang, Harun Günaydin, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141, 2018.
- [4] Hao Wang, Meng Lyu, and Guohai Situ. eholonet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Optics express*, 26(18):22603–22614, 2018.
- [5] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tamida. Learning-based imaging through scattering media. *Optics express*, 24(13):13738–13743, 2016.
- [6] Shuai Li, Mo Deng, Justin Lee, Ayan Sinha, and George Barbastathis. Imaging through glass diffusers using densely connected convolutional networks. *optica*, 5(7):803–813, 2018.
- [7] Yunzhe Li, Yujia Xue, and Lei Tian. Deep speckle correlation: a deep learning approach towards scalable imaging through scattering media. *arXiv preprint arXiv:1806.04139*, 2018.
- [8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [9] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [10] Joseph Goodman. *Introduction to Fourier optics*. McGraw-hill, 2008.
- [11] Loïc Denis, Dirk Lorenz, Eric Thiébaud, Corinne Fournier, and Dennis Trede. Inline hologram reconstruction with sparsity constraints. *Optics letters*, 34(22):3475–3477, 2009.
- [12] Tatiana Latychevskaia and Hans-Werner Fink. Solution to the twin image problem in holography. *Physical review letters*, 98(23):233901, 2007.
- [13] Wenhui Zhang, Liangcai Cao, David J Brady, Hua Zhang, Ji Cang, Hao Zhang, and Guofan Jin. Twin-image-free holography: A compressive sensing approach. *Physical review letters*, 121(9):093902, 2018.
- [14] Alon Greenbaum and Aydogan Ozcan. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Optics express*, 20(3):3129–3143, 2012.
- [15] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.
- [16] James R Fienup. Reconstruction of an object from the modulus of its fourier transform. *Optics letters*, 3(1):27–29, 1978.

---

## References

---

- [1] Rupert RA Bourne, Seth R Flaxman, Tasanee Braithwaite, Maria V Cicinelli, Aditi Das, Jost B Jonas, Jill Keeffe, John H Kempen, Janet Leasher, Hans Limburg, et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, 5(9):e888–e897, 2017.
- [2] Anand-Apte B and Hollyfield J G. Developmental anatomy of the retinal and choroidal vasculature. *The Retin. Its Disord*, 179, 2011.
- [3] Edward A Boettner and J Reimer Wolter. Transmission of the ocular media. *Investigative ophthalmology & visual science*, 1(6):776–783, 1962.
- [4] Alexandre Bourhis, Jean-François Girmens, Severin Boni, Florence Pecha, Catherine Favard, José-Alain Sahel, and Michel Paques. Imaging of macroaneurysms occurring during retinal vein occlusion and diabetic retinopathy by indocyanine green angiography and high resolution optical coherence tomography. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 248(2):161–166, 2010.
- [5] Douglas R Anderson. What happens to the optic disc and retina in glaucoma? *Ophthalmology*, 90(7):766–770, 1983.
- [6] Harry A Quigley. Neuronal death in glaucoma. *Progress in retinal and eye research*, 18(1):39–57, 1999.
- [7] William H Morgan, Martin L Hazelton, and Dao-Yi Yu. Retinal venous pulsation: Expanding our understanding and use of this enigmatic phenomenon. *Progress in retinal and eye research*, 55:82–107, 2016.
- [8] Josef Flammer, Mona Pache, and Thérèse Resink. Vasospasm, its role in the pathogenesis of diseases with particular reference to the eye. *Progress in retinal and eye research*, 20(3):319–349, 2001.
- [9] Ali S Hafez, Regina Bizzarro, Denise Descovich, and Mark R Lesk. Correlation between finger blood flow and changes in optic nerve head blood flow following therapeutic intraocular pressure reduction. *Journal of glaucoma*, 14(6):448–454, 2005.
- [10] Yukihiko Shiga, Hiroshi Kunikata, Naoko Aizawa, Naoki Kiyota, Yukiko Maiya, Yu Yokoyama, Kazuko Omodaka, Hidetoshi Takahashi, Tomoki Yasui, Keiichi Kato,

- et al. Optic nerve head blood flow, as measured by laser speckle flowgraphy, is significantly reduced in preperimetric glaucoma. *Current eye research*, 41(11):1447–1453, 2016.
- [11] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *science*, 254(5035):1178–1181, 1991.
- [12] James G Fujimoto and Wolfgang Drexler. *Optical coherence tomography: Technology and applications*, 2015.
- [13] Adolf F Fercher, Wolfgang Drexler, Christoph K Hitzenberger, and Theo Lasser. Optical coherence tomography-principles and applications. *Reports on progress in physics*, 66(2):239, 2003.
- [14] Thomas L Szabo. *Diagnostic ultrasound imaging: inside out*. Academic Press, 2004.
- [15] David L Hykes, Wayne R Hedrick, and Dale E Starchman. *Ultrasound physics and instrumentation*. Churchill Livingstone New York, 1985.
- [16] Eric A Swanson, David Huang, Michael R Hee, James G Fujimoto, CP Lin, and CA Puliafito. High-speed optical coherence domain reflectometry. *Optics letters*, 17(2):151–153, 1992.
- [17] Johannes F De Boer, Barry Cense, B Hyle Park, Mark C Pierce, Guillermo J Tearney, and Brett E Bouma. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Optics letters*, 28(21):2067–2069, 2003.
- [18] Michael A Choma, Marinko V Sarunic, Changhuei Yang, and Joseph A Izatt. Sensitivity advantage of swept source and fourier domain optical coherence tomography. *Optics express*, 11(18):2183–2189, 2003.
- [19] Laurin Ginner, Tilman Schmoll, Abhishek Kumar, Matthias Salas, Nastassia Pricoupenko, Lara M Wurster, and Rainer A Leitgeb. Holographic line field en-face oct with digital adaptive optics in the retina in vivo. *Biomedical optics express*, 9(2):472–485, 2018.
- [20] Emmanuel Beaurepaire, A Claude Boccara, Martial Lebec, Loïc Blanchot, and Hervé Saint-Jalmes. Full-field optical coherence microscopy. *Optics letters*, 23(4):244–246, 1998.
- [21] Peng Xiao, Viacheslav Mazlin, Kate Grieve, José-Alain Sahel, Mathias Fink, and A Claude Boccara. In vivo high-resolution human retinal imaging with wavefront-correctionless full-field oct. *Optica*, 5(4):409–412, 2018.
- [22] Helge Sudkamp, Peter Koch, Hendrik Spahr, Dierck Hillmann, Gesa Franke, Michael Münt, Fred Reinholz, Reginald Birngruber, and Gereon Hüttmann. In-vivo retinal imaging with off-axis full-field time-domain optical coherence tomography. *Optics letters*, 41(21):4987–4990, 2016.
- [23] Dierck Hillmann, Hendrik Spahr, Helge Sudkamp, Carola Hain, Laura Hinkel, Gesa Franke, and Gereon Hüttmann. Off-axis reference beam for full-field swept-source oct and holoscopy. *Optics express*, 25(22):27770–27784, 2017.

- [24] Michael R Hee, Joseph A Izatt, Eric A Swanson, David Huang, Joel S Schuman, Charles P Lin, Carmen A Puliafito, and James G Fujimoto. Optical coherence tomography of the human retina. *Archives of ophthalmology*, 113(3):325–332, 1995.
- [25] Robert C Youngquist, Sally Carr, and David EN Davies. Optical coherence-domain reflectometry: a new optical evaluation technique. *Optics letters*, 12(3):158–160, 1987.
- [26] Arnaud Dubois. *Handbook of full-field optical coherence microscopy: Technology and applications*. CRC Press, 2016.
- [27] Tim Bonin, Gesa Franke, Martin Hagen-Eggert, Peter Koch, and Gereon Hüttmann. In vivo fourier-domain full-field oct of the human retina with 1.5 million a-lines/s. *Optics letters*, 35(20):3432–3434, 2010.
- [28] Boris Považay, Angelika Unterhuber, Boris Hermann, Harald Sattmann, Holger Arthaber, and Wolfgang Drexler. Full-field time-encoded frequency-domain optical coherence tomography. *Optics Express*, 14(17):7661–7669, 2006.
- [29] H Ohzu and T Kawara. Application of holography in ophthalmology. In *Holography in Medicine and Biology*, pages 133–146. Springer, 1979.
- [30] AN Rosen. Holographic funduscopy with fibre optic illumination. *Optics & Laser Technology*, 7(3):127–129, 1975.
- [31] R Lawrence Wiggins, K Diane Vaughan, and Gerhart B Friedmann. Fundus camera holography of retinal microvasculature. *Archives of Ophthalmology*, 88(1):75–79, 1972.
- [32] Joseph L Calkins and Carl D Leonard. Holographic recording of a retina using a continuous wave laser. *Investigative Ophthalmology & Visual Science*, 9(6):458–462, 1970.
- [33] Dierck Hillmann, Christian Lührs, Tim Bonin, Peter Koch, and Gereon Hüttmann. Holoscopy—holographic optical coherence tomography. *Optics letters*, 36(13):2390–2392, 2011.
- [34] Dierck Hillmann, Gesa Franke, Christian Lührs, Peter Koch, and Gereon Hüttmann. Efficient holoscopy image reconstruction. *Optics express*, 20(19):21247–21263, 2012.
- [35] Jeff Fingler, Carol Readhead, Daniel M Schwartz, and Scott E Fraser. Phase-contrast oct imaging of transverse flows in the mouse retina and choroid. *Investigative ophthalmology & visual science*, 49(11):5055–5059, 2008.
- [36] Andrew Koustenis, Alon Harris, Josh Gross, Ingrida Januleviciene, Aaditya Shah, and Brent Siesky. Optical coherence tomography angiography: an overview of the technology and an assessment of applications for clinical research. *British Journal of Ophthalmology*, 101(1):16–20, 2017.
- [37] Richard F Spaide, James G Fujimoto, and Nadia K Waheed. Image artifacts in optical coherence angiography. *Retina (Philadelphia, Pa.)*, 35(11):2163, 2015.
- [38] Richard F Spaide, James M Klancnik, and Michael J Cooney. Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography. *JAMA ophthalmology*, 133(1):45–50, 2015.
- [39] AF Fercher and J David Briers. Flow visualization by means of single-exposure speckle photography. *Optics communications*, 37(5):326–330, 1981.

- [40] Anna Sophie Mursch-Edlmayr, Nikolaus Luft, Dominika Podkowinski, Michael Ring, Leopold Schmetterer, and Matthias Bolz. Laser speckle flowgraphy derived characteristics of optic nerve head perfusion in normal tension glaucoma and healthy individuals: a pilot study. *Scientific reports*, 8(1):1–7, 2018.
- [41] CE Riva, S Harino, BL Petrig, and RD Shonaf. Laser doppler flowmetry in the optic nerve. *Experimental eye research*, 55(3):499–506, 1992.
- [42] Charles E Riva, Stephen D Cranstoun, Juan E Grunwald, and Benno L Petrig. Choroidal blood flow in the foveal region of the human ocular fundus. *Investigative ophthalmology & visual science*, 35(13):4273–4281, 1994.
- [43] Charles E Riva, Martial Geiser, Benno L Petrig, and Ocular Blood Flow Research Association. Ocular blood flow assessment using continuous laser doppler flowmetry. *Acta ophthalmologica*, 88(6):622–629, 2010.
- [44] Juan E Grunwald, Seenu M Hariprasad, and Joan DuPont. Effect of aging on foveolar choroidal circulation. *Archives of ophthalmology*, 116(2):150–154, 1998.
- [45] Juan E Grunwald, Seenu M Hariprasad, Joan DuPont, Maureen G Maguire, Stuart L Fine, Alexander J Brucker, Albert M Maguire, and Allen C Ho. Foveolar choroidal blood flow in age-related macular degeneration. *Investigative ophthalmology & visual science*, 39(2):385–390, 1998.
- [46] Tetsuya Sugiyama. Basic technology and clinical applications of the updated model of laser speckle flowgraphy to ocular diseases. In *Photonics*, volume 1, pages 220–234. Multidisciplinary Digital Publishing Institute, 2014.
- [47] Manuel Simonutti, Michel Paques, José-Alain Sahel, Michel Gross, Benjamin Samson, Caroline Magnain, and Michael Atlan. Holographic laser doppler ophthalmoscopy. *Optics letters*, 35(12):1941–1943, 2010.
- [48] Caroline Magnain, Amandine Castel, Tanguy Boucneau, Manuel Simonutti, Isabelle Ferrezou, Armelle Rancillac, Tania Vitalis, José-Alain Sahel, Michel Paques, and Michael Atlan. Holographic laser doppler imaging of microvascular blood flow. *JOSA A*, 31(12):2723–2735, 2014.
- [49] Susana Martinez-Conde, Stephen L Macknik, and David H Hubel. The role of fixational eye movements in visual perception. *Nature reviews neuroscience*, 5(3):229–240, 2004.
- [50] Michel Millodot. *Dictionary of Optometry and Visual Science E-Book*. Elsevier Health Sciences, 2014.
- [51] Dierck Hillmann, Tim Bonin, Christian Lührs, Gesa Franke, Martin Hagen-Eggert, Peter Koch, and Gereon Hüttmann. Common approach for compensation of axial motion artifacts in swept-source oct and dispersion in fourier-domain oct. *Optics express*, 20(6):6761–6776, 2012.
- [52] JB Antoine Maintz and Max A Viergever. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998.
- [53] Max A Viergever, JB Antoine Maintz, Stefan Klein, Keelin Murphy, Marius Staring, and Josien PW Pluim. A survey of medical image registration—under review, 2016.

- [54] Radim Kolar and Pavel Tasevsky. Registration of 3d retinal optical coherence tomography data and 2d fundus images. In *International Workshop on Biomedical Image Registration*, pages 72–82. Springer, 2010.
- [55] Marzieh Golabbakhsh and Hossein Rabbani. Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model. *IET Image Processing*, 7(8):768–776, 2013.
- [56] Roziana Ramli, Mohd Yamani Idna Idris, Khairunnisa Hasikin, A Karim, Noor Khairiah, Ainuddin Wahid Abdul Wahab, Ismail Ahmedy, Fatimah Ahmedy, Nahrizul Adib Kadri, and Hamzah Arof. Feature-based retinal image registration using d-saddle feature. *Journal of healthcare engineering*, 2017, 2017.
- [57] Jiayi Ma, Junjun Jiang, Chengyin Liu, and Yansheng Li. Feature guided gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration. *Information Sciences*, 417:128–142, 2017.
- [58] Lauri Laaksonen et al. Spectral retinal image processing and analysis for ophthalmology. 2016.
- [59] Artur V Cideciyan. Registration of ocular fundus images: an algorithm using cross-correlation of triple invariant image descriptors. *IEEE Engineering in Medicine and Biology Magazine*, 14(1):52–58, 1995.
- [60] R Kolar, V Harabis, and J Odstrečilik. Hybrid retinal image registration using phase correlation. *The Imaging Science Journal*, 61(4):369–384, 2013.
- [61] Philip A Legg, Paul L Rosin, David Marshall, and James E Morgan. Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation. *Computerized Medical Imaging and Graphics*, 37(7-8):597–606, 2013.
- [62] Qiang Yang, Jie Zhang, Koji Nozato, Kenichi Saito, David R Williams, Austin Roorda, and Ethan A Rossi. Closed-loop optical stabilization and digital image registration in adaptive optics scanning light ophthalmoscopy. *Biomedical optics express*, 5(9):3174–3191, 2014.
- [63] Michael Pircher, Erich Götzinger, Harald Sattmann, Rainer A Leitgeb, and Christoph K Hitzenberger. In vivo investigation of human cone photoreceptors with slo/oct in combination with 3d motion correction on a cellular level. *Optics express*, 18(13):13935–13944, 2010.
- [64] Laurin Ginner, Abhishek Kumar, Daniel Fechtig, Lara M Wurster, Matthias Salas, Michael Pircher, and Rainer A Leitgeb. Noniterative digital aberration correction for cellular resolution retinal optical coherence tomography in vivo. *Optica*, 4(8):924–931, 2017.
- [65] David A Atchison and Dion H Scott. Monochromatic aberrations of human eyes in the horizontal visual field. *JOSA A*, 19(11):2180–2184, 2002.
- [66] David A Atchison and George Smith. Chromatic dispersions of the ocular media of human eyes. *JOSA A*, 22(1):29–37, 2005.

- [67] Jessica Jarosz, Pedro Mécê, Jean-Marc Conan, Cyril Petit, Michel Paques, and Serge Meimon. High temporal resolution aberrometry in a 50-eye population and implications for adaptive optics error budget. *Biomedical optics express*, 8(4):2088–2105, 2017.
- [68] American National Standards Institute. *American National Standard for Ophthalmics: Methods for Reporting Optical Aberrations of Eyes; ANSI Z80. 28-2004*. Optical Laboratories Association, 2004.
- [69] Robert J Noll. Zernike polynomials and atmospheric turbulence. *JOsA*, 66(3):207–211, 1976.
- [70] Jessica Jarosz, Pedro Mécê, Jean-Marc Conan, Cyril Petit, Michel Paques, and Serge Meimon. High temporal resolution aberrometry in a 50-eye population and implications for adaptive optics error budget. *Biomedical optics express*, 8(4):2088–2105, 2017.
- [71] Pedro Mécê, Elena Gofas-Salas, Kate Grieve, Cyril Petit, Frédéric Cassaing, José Sahel, Michel Paques, and Serge Meimon. Visualizing and enhancing axial resolution in non-confocal adaptive optics ophthalmoscopy. In *Ophthalmic Technologies XXIX*, volume 10858, page 108580P. International Society for Optics and Photonics, 2019.
- [72] Pedro Mécê, Kassandra Groux, Jules Scholler, Olivier Thouvenin, Mathias Fink, Kate Grieve, and Claude Boccara. Curved-full-field oct for high-resolution imaging of living human retina over a large field-of-view. *arXiv preprint arXiv:2001.06893*, 2020.
- [73] Stephen A Burns, Ann E Elsner, Kaitlyn A Sapoznik, Raymond L Warner, and Thomas J Gast. Adaptive optics imaging of the human retina. *Progress in retinal and eye research*, 68:1–30, 2019.
- [74] Horace W Babcock. The possibility of compensating astronomical seeing. *Publications of the Astronomical Society of the Pacific*, 65(386):229–236, 1953.
- [75] Junzhong Liang, David R Williams, and Donald T Miller. Supernormal vision and high-resolution retinal imaging through adaptive optics. *JOSA A*, 14(11):2884–2892, 1997.
- [76] Austin Roorda, Fernando Romero-Borja, William J Donnelly III, Hope Queener, Thomas J Hebert, and Melanie CW Campbell. Adaptive optics scanning laser ophthalmoscopy. *Optics express*, 10(9):405–412, 2002.
- [77] Ben C Platt and Roland Shack. History and principles of shack-hartmann wavefront sensing. *Journal of refractive surgery*, 17(5):S573–S577, 2001.
- [78] Larry N Thibos and Xin Hong. Clinical applications of the shack-hartmann aberrometer. *Optometry and vision science: official publication of the American Academy of Optometry*, 76(12):817–825, 1999.
- [79] Roberto Ragazzoni. Pupil plane wavefront sensing with an oscillating prism. *Journal of modern optics*, 43(2):289–293, 1996.
- [80] Ignacio Iglesias, Roberto Ragazzoni, Yves Julien, and Pablo Artal. Extended source pyramid wave-front sensor for the human eye. *Optics express*, 10(9):419–428, 2002.
- [81] Stéphane R Chamot, Chris Dainty, and Simone Esposito. Adaptive optics for ophthalmic applications using a pyramid wavefront sensor. *Optics Express*, 14(2):518–526, 2006.

- [82] Marco Lombardo and Giuseppe Lombardo. New methods and techniques for sensing the wave aberrations of human eyes. *Clinical and Experimental Optometry*, 92(3):176–186, 2009.
- [83] Heidi Hofer, Nripun Sredar, Hope Queener, Chaohong Li, and Jason Porter. Wavefront sensorless adaptive optics ophthalmoscopy in the human eye. *Optics express*, 19(15):14160–14171, 2011.
- [84] Yifan Jian, Sujin Lee, Myeong Jin Ju, Morgan Heisler, Weiguang Ding, Robert J Zawadzki, Stefano Bonora, and Marinko V Sarunic. Lens-based wavefront sensorless adaptive optics swept source oct. *Scientific reports*, 6:27620, 2016.
- [85] Xiaolin Zhou, Phillip Bedggood, Bang Bui, Christine TO Nguyen, Zheng He, and Andrew Metha. Contrast-based sensorless adaptive optics for retinal imaging. *Biomedical optics express*, 6(9):3577–3595, 2015.
- [86] Delphine Débarre, Martin J Booth, and Tony Wilson. Image based adaptive optics through optimisation of low spatial frequencies. *Optics Express*, 15(13):8176–8190, 2007.
- [87] James Polans, Brenton Keller, Oscar M Carrasco-Zevallos, Francesco LaRocca, Elijah Cole, Heather E Whitson, Eleonora M Lad, Sina Farsiu, and Joseph A Izatt. Wide-field retinal optical coherence tomography with wavefront sensorless adaptive optics for enhanced imaging of targeted regions. *Biomedical optics express*, 8(1):16–37, 2017.
- [88] Steven G Adie, Nathan D Shemonski, Benedikt W Graf, Adeel Ahmad, P Scott Carney, and Stephen A Boppart. Guide-star-based computational adaptive optics for broadband interferometric tomography. *Applied physics letters*, 101(22):221117, 2012.
- [89] Changgeng Liu, Xiao Yu, and Myung K Kim. Phase aberration correction by correlation in digital holographic adaptive optics. *Applied optics*, 52(12):2940–2949, 2013.
- [90] Alexander Stadelmaier and Jürgen H Massig. Compensation of lens aberrations in digital holography. *Optics letters*, 25(22):1630–1632, 2000.
- [91] JR Fienup and JJ Miller. Aberration correction by maximizing generalized sharpness metrics. *JOSA A*, 20(4):609–620, 2003.
- [92] Paritosh Pande, Yuan-Zhi Liu, Fredrick A South, and Stephen A Boppart. Automated computational aberration correction method for broadband interferometric imaging techniques. *Optics letters*, 41(14):3324–3327, 2016.
- [93] Steven G Adie, Benedikt W Graf, Adeel Ahmad, P Scott Carney, and Stephen A Boppart. Computational adaptive optics for broadband optical interferometric tomography of biological tissue. *Proceedings of the National Academy of Sciences*, 109(19):7175–7180, 2012.
- [94] Yang Xu, Yuan-Zhi Liu, Stephen A Boppart, and P Scott Carney. Automated interferometric synthetic aperture microscopy and computational adaptive optics for improved optical coherence tomography. *Applied optics*, 55(8):2034–2041, 2016.
- [95] Dierck Hillmann, Hendrik Spahr, Carola Hain, Helge Sudkamp, Gesa Franke, Clara Pfäffle, Christian Winter, and Gereon Hüttmann. Aberration-free volumetric high-speed imaging of in vivo retina. *Scientific reports*, 6:35209, 2016.

- [96] Abhishek Kumar, Wolfgang Drexler, and Rainer A Leitgeb. Subaperture correlation based digital adaptive optics for full field optical coherence tomography. *Optics express*, 21(9):10850–10866, 2013.
- [97] Hao Wang, Meng Lyu, and Guohai Situ. eholonet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Optics express*, 26(18):22603–22614, 2018.
- [98] Ulugbek S Kamilov, Ioannis N Papadopoulos, Morteza H Shoreh, Alexandre Goy, Cedric Vonesch, Michael Unser, and Demetri Psaltis. Learning approach to optical tomography. *Optica*, 2(6):517–522, 2015.
- [99] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017.
- [100] Yair Rivenson, Yibo Zhang, Harun Günaydin, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141–17141, 2018.
- [101] Hong Guo, Nina Korablinova, Qiushi Ren, and Josef Bille. Wavefront reconstruction with artificial neural networks. *Optics express*, 14(14):6456–6462, 2006.
- [102] Scott W Paine and James R Fienup. Machine learning for improved image-based wavefront sensing. *Optics letters*, 43(6):1235–1238, 2018.
- [103] Yohei Nishizaki, Matias Valdivia, Ryoichi Horisaki, Katsuhisa Kitaguchi, Mamoru Saito, Jun Tanida, and Esteban Vera. Deep learning wavefront sensing. *Optics express*, 27(1):240–251, 2019.
- [104] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [105] Ryoichi Horisaki, Kazuki Fujii, and Jun Tanida. Single-shot and lensless complex-amplitude imaging with incoherent light based on machine learning. *Optical Review*, 25(5):593–597, 2018.
- [106] Yunzhe Li, Yujia Xue, and Lei Tian. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica*, 5(10):1181–1190, 2018.
- [107] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. Learning-based imaging through scattering media. *Optics express*, 24(13):13738–13743, 2016.
- [108] Shuai Li, Mo Deng, Justin Lee, Ayan Sinha, and George Barbastathis. Imaging through glass diffusers using densely connected convolutional networks. *Optica*, 5(7):803–813, 2018.
- [109] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, 2017.
- [110] George Barbastathis, Aydogan Ozcan, and Guohai Situ. On the use of deep learning for computational imaging. *Optica*, 6(8):921–943, 2019.
- [111] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

- [112] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.
- [113] Pietro Ferraro, Sergio De Nicola, Giuseppe Coppola, Andrea Finizio, Domenico Alfieri, and Giovanni Pierattini. Controlling image size as a function of distance and wavelength in fresnel-transform reconstruction of digital holograms. *Optics Letters*, 29(8):854–856, 2004.
- [114] Vít Lédl, Pavel Psota, František Kaván, Ondřej Matoušek, and Pavel Mokrý. Surface topography measurement by frequency sweeping digital holography. *Applied optics*, 56(28):7808–7814, 2017.
- [115] Nicolas Verrier and Michael Atlan. Off-axis digital hologram reconstruction: some practical considerations. *Applied optics*, 50(34):H136–H146, 2011.
- [116] Thomas Klein and Robert Huber. High-speed oct light sources and systems. *Biomedical optics express*, 8(2):828–859, 2017.
- [117] Dirk Padfield. Masked object registration in the fourier domain. *IEEE Transactions on image processing*, 21(5):2706–2718, 2011.
- [118] Yoshua Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.
- [119] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [120] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [121] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [122] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [123] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [124] Loïc Denis, Dirk Lorenz, Eric Thiébaud, Corinne Fournier, and Dennis Trede. Inline hologram reconstruction with sparsity constraints. *Optics letters*, 34(22):3475–3477, 2009.
- [125] Tatiana Latychevskaia and Hans-Werner Fink. Solution to the twin image problem in holography. *Physical review letters*, 98(23):233901, 2007.
- [126] Wenhui Zhang, Liangcai Cao, David J Brady, Hua Zhang, Ji Cang, Hao Zhang, and Guofan Jin. Twin-image-free holography: a compressive sensing approach. *Physical review letters*, 121(9):093902, 2018.
- [127] Yair Rivenson, Yibo Zhang, Harun Günaydın, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141–17141, 2018.

- 
- [128] Alon Greenbaum and Aydogan Ozcan. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Optics express*, 20(3):3129–3143, 2012.
- [129] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [130] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [131] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.