

# Title goes here

Balázs Bálint



Pázmány Péter Catholic University

Faculty of Information Technology



European Molecular Biology Laboratory

Supervisors: [Lars Hufnagel](#), [Balázs Rózsa](#)

A thesis submitted for the degree of  
*Doctor of Philosophy*  
2017

## Tartalmi kivonat

Absztrakt magyarul.

## Abstract

Abstract in English.



# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Symmetric Mouse SPIM</b>	<b>5</b>
2.1	Motivation . . . . .	5
2.2	Light-sheet microscopy . . . . .	5
2.3	Fluorescence microscopy . . . . .	5
2.4	Confocal microscopy . . . . .	8
2.5	Confocal-theta microscopy . . . . .	9
<b>3</b>	<b>Single-plane illumination microscopy</b>	<b>11</b>
3.1	Optical properties . . . . .	11
3.2	Multi-view fusion techniques . . . . .	16
3.3	Dual Mouse-SPIM . . . . .	16
<b>4</b>	<b>Real-time image processing</b>	<b>19</b>
4.1	Image compression . . . . .	19
4.2	Entropy coding . . . . .	19
4.3	Transform coding . . . . .	24
4.4	Conclusion . . . . .	28
<b>5</b>	<b>Discussion</b>	<b>29</b>
	<b>Acknowledgements</b>	<b>31</b>
	<b>References</b>	<b>33</b>



# Chapter 1

## Introduction

[1]





## Chapter 2

# Symmetric Mouse SPIM

### 2.1 Motivation

### 2.2 Light-sheet microscopy

#### 2.2.1 Fluorescence

#### 2.2.2 Wide-field microscopy

#### 2.2.3 Confocal laser scanning microscopy

#### 2.2.4 Light-sheet microscopy

[1]

#### 2.2.5 MuVi-SPIM

[2]

#### 2.2.6 Mouse-SPIM

[3]

### 2.3 Fluorescence microscopy

Fluorescence microscopy [4] is a microscopy technique that uses fluorophores, a special group of chemicals that are capable of absorbing a photon of well defined energy, and emit a different photon within a stochastically determined time frame. The emitted photon usually has less energy than the absorbed photon, it is shifted towards the red side of the spectrum. This effect is called the Stokes-shift.

Using this phenomenon, it is possible to differentiate between the photons used for excitation and the emitted photons using an appropriate filter. Since the emitted photon

usually has less energy, a low-pass filter can be used that will filter out the higher energy, lower wavelength illumination photons. Even though a very small amount of illumination photons will result in fluorescence ( $< 0.0001\%$ ), the signal to noise ratio of the fluorescence is still very high due to the filtering.

Fluorescent microscopy is especially suitable to image live specimens, since a lot of organisms have naturally fluorescent structures. This phenomenon is called autofluorescence.

Imaging specimens using just autofluorescence although can be useful in some cases, but in most of the time we would want to narrow our interest in a specific part, specific structure of the specimen. This can be achieved in several ways, but the simplest and earliest solution is the use of fluorescent dyes. These dyes, as the name implies, emit fluorescence under a certain excitation wavelength, and have such a structure, that they can specifically bind to certain parts of the cell. Fluorescent dyes although very useful in most cases, have some limitations.

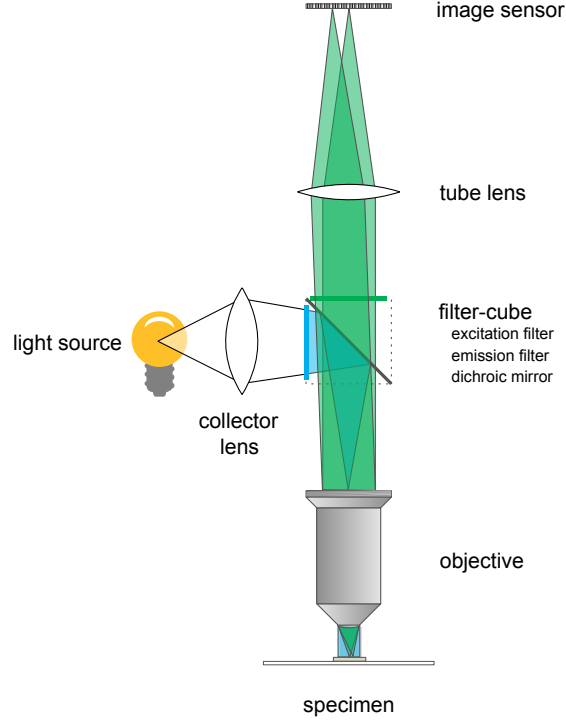
However, fluorescent microscopy took a great leap with the discovery of the green fluorescent protein. This had such an impact on future development of fluorescence microscopy, that three scientists were awarded with a Nobel Prize in chemistry in 2008 for its discovery [5]. Since then, several variants were generated, and so today a whole spectrum of fluorescent proteins are available [6]. The huge benefit of these proteins over the synthetic fluorescent dyes, is that the cell is capable of producing them by itself. Using genetic manipulation techniques, it is possible to introduce the genetic material coding a fluorescent protein into the genome of virtually any organism. These proteins can also be modified and fused with certain proteins of interest. Using this technique, the fluorescent protein will be attached to a functional protein, this way its expression and localisation pattern can be determined.

### **2.3.1 Wide-field fluorescence microscopy**

One of the simplest fluorescent microscopes is a wide-field fluorescent microscope (Figure 2.1). The main difference to a regular microscope, is the introduction of a filter cube. This cube contains an excitation filter, emission filter and a dichroic mirror. The dichroic mirror is used to couple the illumination and detection light paths, and the filters are used to separate the fluorescent light from the excitation light. The image of the sample is focused on the detector of a digital camera.

#### **Resolution of a wide-field microscope**

Resolution of a microscope is given by the minimal distance of two points that can still be distinguished. The smaller this distance is, the better the resolving power of the microscope. However, the resolution is limited by the wave nature of light, and cannot be decreased infinitely.



**Figure 2.1: Wide-field fluorescent microscope.**

A generally accepted method to calculate lateral ( $\sigma_{xy}$ ) and axial ( $\sigma_z$ ) resolution of an optical microscope is described by the Stelzer-Grill-Heisenberg, or SGH theory[7, 8]:

$$\sigma_{xy} = \frac{\lambda}{\sqrt{3 - 2 \cos \alpha - \cos 2\alpha}} \quad (2.1)$$

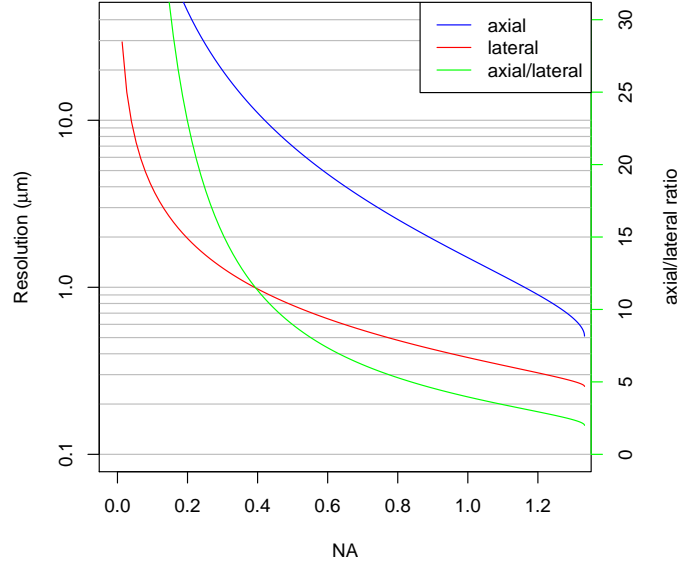
$$\sigma_z = \frac{\lambda}{1 - \cos \alpha} \quad (2.2)$$

Generally, instead of  $\alpha$ , the numerical aperture (NA) is commonly used to characterize a lens' aperture.  $NA = n \cdot \sin \alpha$ , where  $n$  is the refractive index of the medium and  $\alpha$  stands for the angular aperture.

Although high NA objectives have better axial resolution (Figure 2.2), this is still around 3–6 times worse than the lateral resolution. In theory, it is possible to achieve isotropic resolution in case of  $\alpha = 180^\circ$ , which means collecting all light emitted from the sample. However, it is currently unachievable to realize such a device that is capable of this. Because of this, 3D imaging with any microscopy technique that use one objective will always produce an anisotropic image.

Another disadvantage of the wide-field microscope, is that it can not be used with thick specimens. Usually this type of microscopy is only used for a single layer of cells, because all the objects in the field of view will appear on the imaging plane, not just the plane in focus. These objects will appear blurred if close to the focus, or just evenly add

to the background noise if they are further from the focus. This is why imaging specimens much thicker than  $10\text{ }\mu\text{m}$  will result in suboptimal image quality.



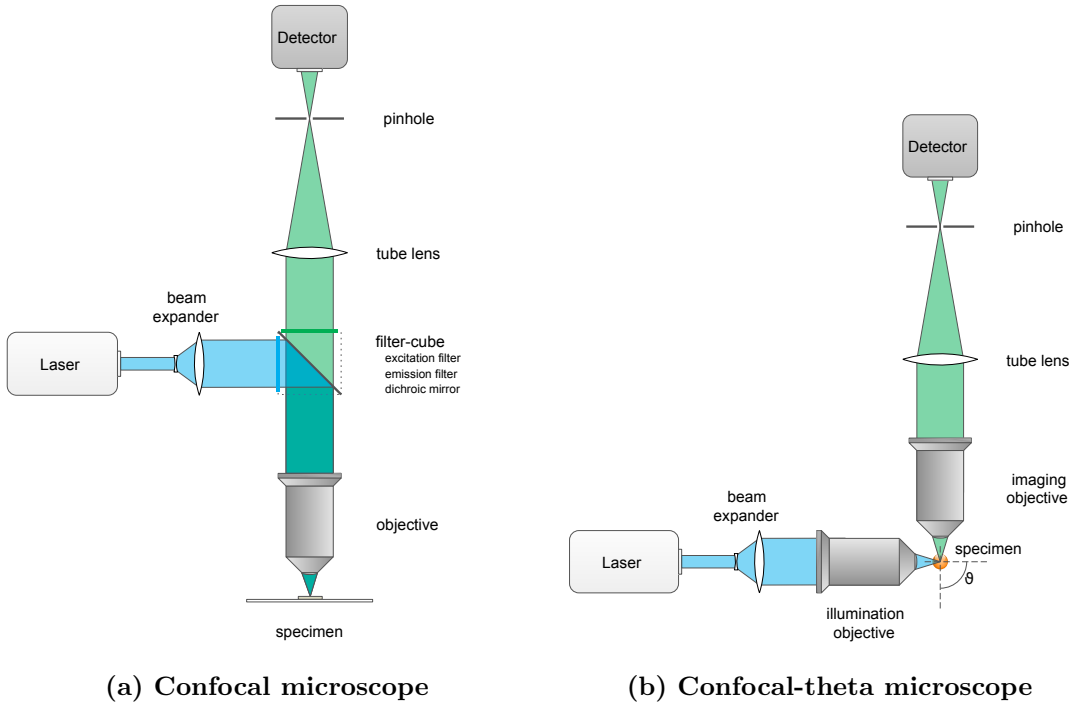
**Figure 2.2: Resolution of a wide-field microscope.** Axial (blue) and lateral (red) resolutions of a wide-field microscope are shown with respect to the numerical aperture (NA). Resolutions are calculated with  $\lambda = 510\text{nm}$ , the emission maximum of GFP and  $n = 1.333$ , the refractive index of water, for water dipping objectives.

## 2.4 Confocal microscopy

Laser scanning confocal microscopy [9] addresses the problem of noise originating from the out of focus planes. The principle for illumination and detection optics is very similar to a wide-field microscope, but for illumination a focused laser light is used.

The biggest change is in the detection method: the confocal microscope uses a pinhole, to exclude light coming from out of focus planes. Since only those rays are taking part in the imaging that originate from the focus, the image quality is highly improved. This microscopy technique is also capable of 3D imaging, with an axial resolution corresponding to a wide-field microscope.

However, the image can not be registered as simply as with the wide-field detection, since at any given time, only a very small portion of the sample will be used in the imaging. Because of this, a more sensitive detection is required, which generally means the use of a photomultiplier. To get the whole image, the focus is scanned through the whole sample (or rather, the sample is scanned through the focus), recording an intensity value for each position. The image is then generated using a computer based on the recorded position and intensity values.



**Figure 2.3: Basic optical components of a laser scanning confocal and confocal-theta microscope.** Both type of microscopes use confocal images detection, which means that a pinhole is used to exclude light coming from out of focus points. Light intensity is measured by a photomultiplier for every voxel in the region of interest. The final image is generated on a computer using the positions and recorded intensity values. A regular confocal microscope (2.3a) uses the same objective for illumination and detection, while a confocal-theta microscope (2.3b) uses a second objective that is rotated by  $\vartheta$  around the focus. In this case,  $\vartheta = 90^\circ$ .

Although this microscopy technique already has 3D capabilities, it's axial resolution is still limited by the SGH theory, since it uses only one objective. Imaging live specimens for an extended period of time with confocal microscopy although possible [10], is not ideal. Since for each voxel imaged, almost the entire specimen has to be illuminated, which results in a very high dose of radiation of the samples. This can be as much as 30–100 times bigger, than the dose used for the actual imaging [11]. High power of laser for an extended time frame can result in bleaching the fluorophores, thus resulting in a lower signal at later times, but the more significant issue is phototoxicity, that is when the cells themselves are damaged by the laser light.

## 2.5 Confocal-theta microscopy

Although confocal microscopy already provides a better resolution in all dimensions, the ratio of the axial and lateral resolution is still very high, due to the single objective illumination and detection. This seriously limits the microscope's 3D imaging capabilities, since in the  $z$  direction (i.e. along the imaging axis) the resolution would be significantly worse than in the other directions.

Confocal-theta microscopy [12] introduces a second objective to a regular confocal

microscope, that is used to illuminate the sample (Figure 2.3b). Since this decouples the illumination and detection, using a filter cube is no longer necessary. The second objective is rotated by  $\vartheta$  around the focus, this is where the name of this setup originates.

Resolution is also improved compared to the regular confocal microscope, because the lateral resolution of the imaging objective now corresponds to the axial resolution of the detection objective. The combined resolution of the two-objective system can be calculated in the following manner [13]:

$$\frac{1}{\sigma_{sys}^2} = \frac{1}{\sigma_{ill}^2} + \frac{1}{\sigma_{det}^2} \quad (2.3)$$

where  $\sigma_{ill} = \sigma_{xy}$  and  $\sigma_{det} = \sigma_z$  for the axial resolution of the system, and reversed for the lateral resolution of the system. This means, that the axial and lateral resolution would be the same (if the same objectives are used), and the resulting point spread function is almost isotropic.

Although this is a big improvement to confocal microscopy, the issue of photobleaching and phototoxicity is still not solved with a confocal theta microscope, which means that longer developmental processes are near impossible to follow with this type of microscopy.

## Chapter 3

# Single-plane illumination microscopy

The main principle behind single plane illumination microscopy, that is illuminating the sample from the side by a very thin light-sheet, dates back to the early 20<sup>th</sup> century, when Siedentopf and Zsigmondy first described the ultramicroscope [14]. This microscope used sunlight as an illumination source, that was guided through a precision slit to generate a thin light-sheet. This allowed Zsigmondy to visualize gold nanoparticles floating in and out of the light-sheet. Since these particles are much smaller than the wavelength of the light, the device was called an ultramicroscope. His studies with colloids together with the development of the ultramicroscope led Zsigmondy to win the Nobel Prize in 1925.

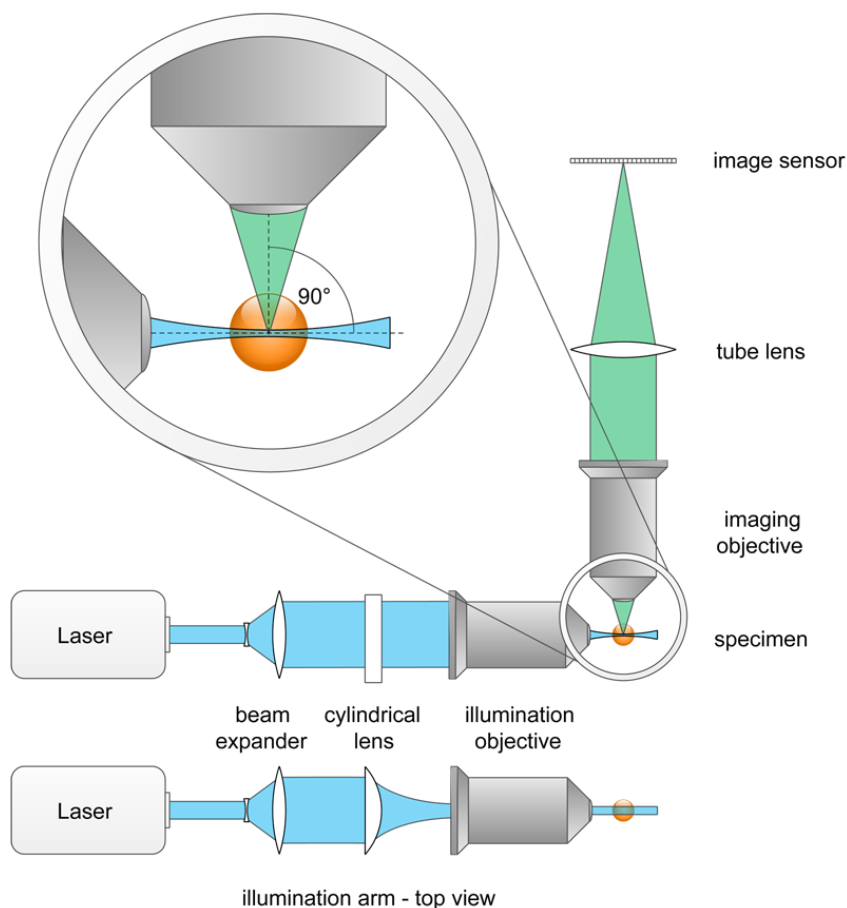
Since then however, light-sheet microscopy was seldom used, but in the last decade it was reinvented and combined with fluorescent microscopy. The first notable light-sheet fluorescent microscope (LSFM) was developed at EMBL in 2004 [1], that demonstrated the benefits of using a light-sheet in imaging developmental processes in three dimension.

Since then, light-sheet based imaging has gained more and more popularity, as it can be adapted and applied to a wide variety of problems. It was numerously proven to be a better choice than confocal microscopy [11, 15] especially in developmental biological applications [16]. It can also be used with a wide variety of specimens of different sizes, such as zebrafish embryo [17], mouse brain [18] or drosophila embryo [2]. It is also possible to use light-sheet microscopy in super-resolution, allowing for individual molecule localization [?].

### 3.1 Optical properties

A selective-plane illumination microscope (SPIM) uses a light-sheet to illuminate only a thin section of the sample (Figure 3.1). This illumination plane is perpendicular to the imaging axis of the detection objective and coincides with the focal plane. This way the image is taken of only that specific plane that is illuminated, thus providing much better

signal to noise ratio. In case of conventional wide-field fluorescent microscopy, where the whole specimen is illuminated, light scattering from different regions contribute to a significant background noise. With selective-plane illumination, this problem is intrinsically solved, and it also provides a true sectioning capability. This makes SPIM especially suitable for 3D imaging.



**Figure 3.1: Basic optical components of a SPIM.** A selective plane illumination microscope uses two objectives orthogonally aligned. One objective is used to generate a thin light-sheet that illuminates the sample from the side, while the other is used for detection. To generate an image of the specimen, a suitable tube lens is used to focus the light on the sensor of a detection unit (e.g. sCMOS camera). The light-sheet is generated by the illumination objective, using a beam that is previously shaped by a cylindrical lens.

### 3.1.1 Detection

The detection unit of a SPIM is basically equivalent to a detection unit of a wide-field microscope, without a dichroic mirror (Figure 3.1). Most important components are the objective together with the tube lens, filter wheel, and a sensor, typically a CCD or sCMOS camera.

One of the most important aspects that determine the resolution of the microscope is the detection objective. Since in developmental biology specimens require a water-based



solution, these objectives are usually water dipping objectives directly submerged in the medium. Since the refraction index of water ( $n = 1.333$ ) is greater than the refraction index of air, these objectives tend to have a higher  $NA$ , which results in higher resolution. This, however, also depends on the sensor used, mainly on the pixel size ( $d_{sensor}$ ).

The magnification is typically  $10\times$ ,  $20\times$ ,  $40\times$  or  $100\times$  but these values are sound only when the objective is used together with the prescribed tube lens. These lenses are specially made to be used with the specific objectives, and are corrected for any aberrations. They typically have a focal length of 160–200mm.

### 3.1.2 Illumination

#### Using cylindrical lens

The light-sheet can be generated using a cylindrical lens, which focuses the laser beam in only one direction, and creating a thin sheet in the proximity of the focal point. However, to achieve light-sheets that are thin enough, one would need to use cylindrical lens with low focal lengths, but these are hardly accessible in well corrected formats. For this reason, its more common to use a longer focal length cylindrical lens in conjunction with a microscope objective, which is well corrected for chromatic and spherical aberrations [19]. This way, the light-sheet length, thickness and width can be adjusted for the specific imaging tasks.

For paraxial waves, i.e. waves with nearly parallel wave front normals, a general wave equation can be approximated with the paraxial Helmholtz equation [13, 20]

$$\nabla_T^2 + i2k \frac{\partial U}{\partial z} = 0 \quad (3.1)$$

where  $\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ ,  $U(\vec{r})$  is the wave-function,  $k = \frac{2\pi}{\lambda}$  is the wavenumber and we assume, that the light spreads in  $z$  direction.

A simple solution to this differential equation is the Gaussian beam:

$$U(r, z) = A_0 \cdot \frac{W_0}{W(z)} \cdot e^{-\frac{r^2}{W^2(z)}} \cdot e^{-i \cdot \phi(r, z)} \quad (3.2)$$

where  $A_0$  is the amplitude of the wave,  $W_0$  is the radius of the beam waist (the thinnest location on the beam),  $r = \sqrt{x^2 + y^2}$  is the distance from the center of the beam,  $W(z)$  is the radius of the beam  $z$  distance from the waist, and  $\phi(r, z)$  is the combined phase part of the wave-function. Furthermore:

$$W(z) = W_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2} \quad (3.3)$$

where the parameter  $z_0$  is called the Rayleigh-range. This has the following connection with the beam waist:

$$z_0 = \frac{\pi W_0}{\lambda} \quad (3.4)$$

Which means, the thinner the beam waist, the shorter the Rayleigh-range, that is the beam divergence is faster for more focused beams.

Intensity of the emitted fluorescence is based on the intensity of the excitation light. In case of a Gaussian beam:

$$I(r, z) = U(r, z) \cdot U^*(r, z) = |A_0|^2 \cdot \left( \frac{W_0}{W(z)} \right)^2 \cdot e^{-\frac{2r^2}{W^2(z)}} \quad (3.5)$$

Apart from the circular Gaussian beam, the elliptical Gaussian beam is also an eigenfunction of Helmholtz equation (3.1):

$$U(x, y, z) = A_0 \cdot \sqrt{\frac{W_{x,0}}{W_x(z)}} \sqrt{\frac{W_{y,0}}{W_y(z)}} \cdot e^{-\frac{x^2}{W_x^2(z)}} \cdot e^{-\frac{y^2}{W_y^2(z)}} \cdot e^{-i \cdot \phi(x, y, z)} \quad (3.6)$$

This beam still has a Gaussian profile along the  $x$  and  $y$  axes, but the radii are uncoupled, which results in an elliptical beam. Since the beam waist is different along the two axes, the Rayleigh range is also different:

$$z_{x,0} = \frac{\pi W_{x,0}^2}{\lambda} \quad (3.7)$$

$$z_{y,0} = \frac{\pi W_{y,0}^2}{\lambda} \quad (3.8)$$

Intensity of the beam is the following:

$$I(x, y, z) = U(x, y, z) \cdot U^*(x, y, z) = |A_0|^2 \cdot \frac{W_{x,0}}{W_x(z)} \cdot \frac{W_{y,0}}{W_y(z)} \cdot e^{-\frac{2x^2}{W_x^2(z)}} \cdot e^{-\frac{2y^2}{W_y^2(z)}} \quad (3.9)$$

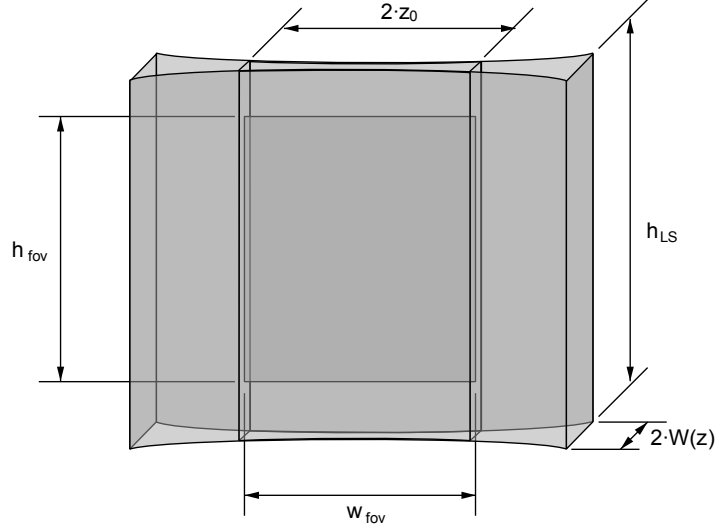
where

$$W_x(z) = W_{x,0} \sqrt{1 + \left( \frac{z}{z_{x,0}} \right)^2} \quad \text{and} \quad W_y(z) = W_{y,0} \sqrt{1 + \left( \frac{z}{z_{y,0}} \right)^2} \quad (3.10)$$

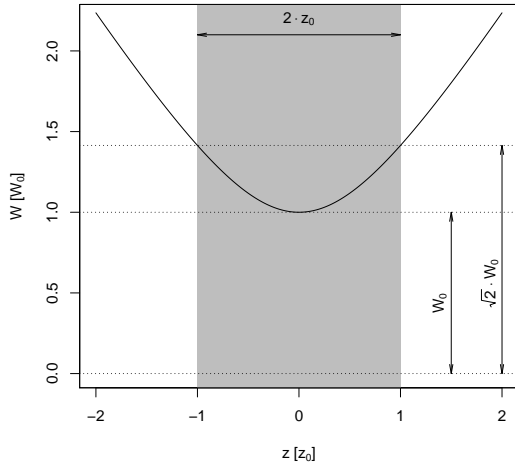
Since the illumination is uneven, the usable field of view is smaller than the actual illuminated region (Figure 3.2a). The width of the field of view  $w_{fov}$  is determined by the Rayleigh length, since this is in a direct relation with the beam divergence. To stay in the optimal region, the light-sheet should only be used in the range of 1 Rayleigh length on both sides of the beam waist (Figure 3.2b). In this range, the ratio between the thickest (at  $z = z_0$ ) and the thinnest (at  $z = 0$ ) part of the beam  $W(z)$  will be  $\sqrt{2} \approx 1.4142$  which is still acceptable.

Light-sheet height is determined by the profile of the beam along the vertical axis (Figure 3.2c). Since this is a Gaussian function (see Equation 3.2), only a small part in

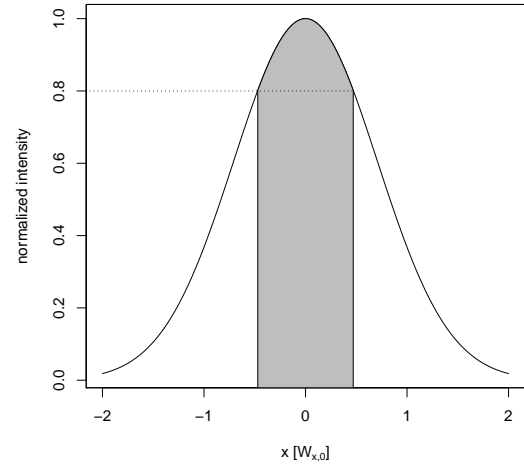
the middle can be used for imaging, because towards the sides the intensity dramatically drops. To allow a maximum 80% drop of intensity at the edges, the light-sheet height is  $h_{fov} = 2 \cdot 0.472 \cdot W_{x,0}$



(a)



(b)



(c)

**Figure 3.2: Light-sheet dimensions.** 3.2a shows a light sheet, with the field of view indicated. Since the light-sheet intensity is uneven, the field of view has to be confined to a smaller region. 3.2b The width and thickness of the field of view depends on the Rayleigh length of the beam ( $z_{y,0}$ ). 3.2c Height of the field of view is determined by the Gaussian profile of the elliptical beam.

## **Using focused beam scanning**

Since using cylindrical lenses it's not possible to generate a homogeneous light-sheet, moreover at higher magnification the Rayleigh range would be too small, we also consider using focused beam scanning to generate the light-sheet (digital scanned light-sheet microscopy, DSLM). To generate a scanning beam, a galvanometer controlled mirror is used to alter the beam path. This can quickly turn around its axis which will result in an angular sweep with the laser beam. To change the angular movement to translation, a scan lens is used to generate an intermediate scanning plane. This plane is then imaged to the specimen by the tube lens and the illumination objective, resulting in a scanned focused beam.

This method to generate the light-sheet has several advantages compared to a static light-sheet. The height of this sheet is not determined by the cylindrical lens, but it can be dynamically modified. Also, the intensity is uniform through the whole height of the light-sheet.

## **3.2 Multi-view fusion techniques**

### **3.2.1 Why multi view**

### **3.2.2 registration**

**Image based registration**

**Bead based registration**

**Affine transformation**

### **3.2.3 fusion**

**Average**

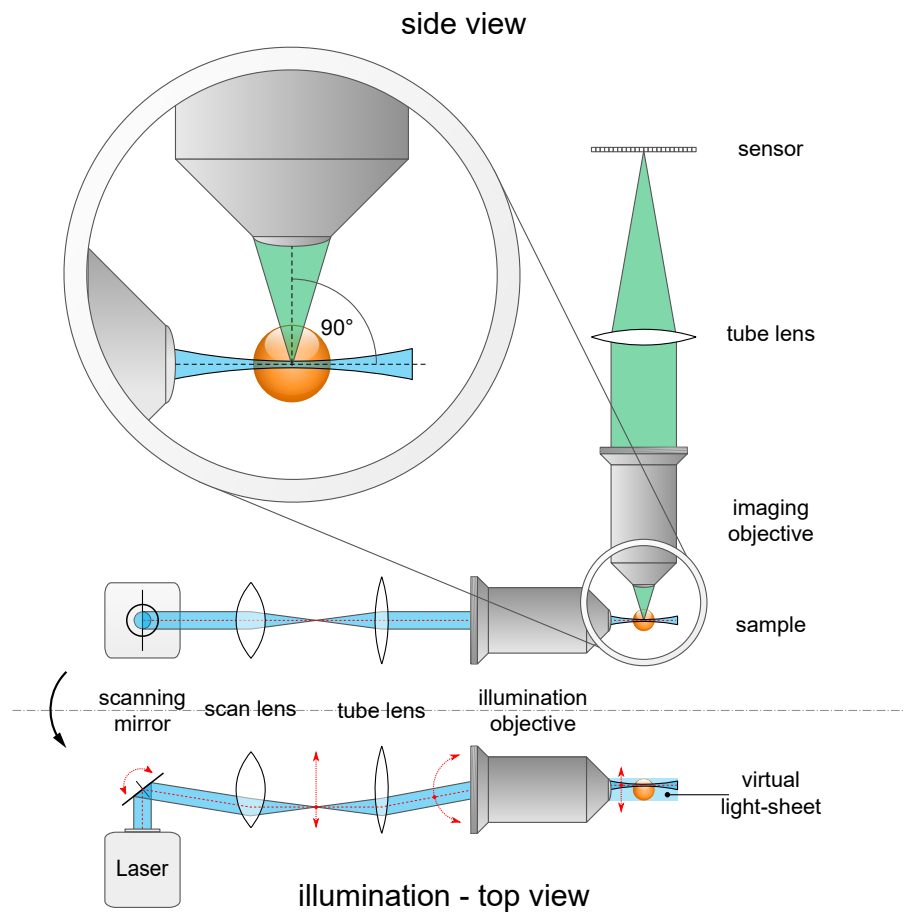
**Sigmoidal weighted average**

**Fourier mixing**

**Multi-view deconvolution**

## **3.3 Dual Mouse-SPIM**

### **3.3.1 Previous Mouse-SPIM**



**Figure 3.3: DSLM illumination.** DSLM illuminates a specimen by a circularly-symmetric beam that is scanned over the field of view. This creates a virtual light-sheet, which illuminates a section of a specimen just like the SPIM. Light-sheet in DSLM is uniform over the whole field of view and its height can be dynamically altered by changing the beam scan range.



## Chapter 4

# Real-time image processing

### 4.1 Image compression

Image compression is an important tool for everyday life, however it's rarely used in the context of scientific imaging because of fear of information loss. This preconception is mainly due to the famous blocking artifacts found in many highly compressed JPEG images, however not all image compression algorithms introduce artifacts, and in fact many lossless algorithms exist that would be suitable for such images. Nowadays, when data production is in an exponential growth compression is again in highlight, without it it would be extremely difficult to maintain many scientific projects that produce images at a high data rate.

In this paper I will review the basics of image compression based on Sayood's textbook, Introduction to Data Compression [21]. I will introduce some basics of information theory and entropy, followed by discussing two widely used entropy coding algorithms, Huffman coding and arithmetic coding. Section 3 will be about transform coding, specifically Discrete Cosine Transform (DCT) and wavelet transform while also touching upon techniques based on differential pulse code modulation. Finally I will show how some of the most widely used image compression standards use these methods to achieve effective image compression.

### 4.2 Entropy coding

#### 4.2.1 Information and Entropy

For the purpose of data compression it is useful to quantify the amount of *information* contained within a piece of data. The first rigorous definition of information was presented in an extremely influential paper by Shannon, published in two parts in 1948 [22, 23].

First, let's define the amount of self-information contained in the outcome of a random

experiment:

$$I(A) = \log_b \frac{1}{P(A)} = -\log_b P(A) \quad (4.1)$$

2 independent events:

$$P(A, B) = P(A) \cdot P(B) \quad (4.2)$$

self-information is additive:

$$I(A, B) = \log_b \frac{1}{P(A, B)} \quad (4.3)$$

$$= \log_b \frac{1}{P(A) \cdot P(B)} \quad (4.4)$$

$$= \log_b \frac{1}{P(A)} + \log_b \frac{1}{P(B)} \quad (4.5)$$

entropy for random variable  $X$     average or expected self-information for the random variable

$$H(X) = \sum_i P(A_i) I(A_i) = - \sum_i P(A_i) \log_b P(A_i) \quad (4.6)$$

entropy rate for data source  $S$     average information output by the data source

#### 4.2.2 Huffman coding

Huffman coding is a prefix-free, optimal code that is widely used in data compression. It was developed by David A. Huffman as a course assignment on the first ever course on information theory at MIT, and was published shortly afterwards [24]. It is a variable length binary code which assigns different length codewords to letters of different probabilities. It is able to achieve optimal compression, which means the total length of the coded sequence will be minimal.

Although it produces a variable length code which can introduce some issues with decoding, it is still uniquely decodeable. It achieves this property by using prefix-free codewords, meaning that none of the codewords are prefixes of any other codewords. This property can be exploited when decoding the codeword, since during this procedure the number of bits for the next codeword can not be determined in advance. However if no codeword is a prefix of another codeword, by simply reading the successive bits one by one until we reach a valid codeword, it's possible to uniquely decode the message.

Let's take the example in Table 4.2. Five letters are coded in binary code by Code #1 and by Code #2. Code 1 is not a prefix code, and because of this when reading the encoded sequence we can not be sure when we reach the end of a codeword. Decoding the sequence 0000 for example could be interpreted as 4 letters of  $a_1$  or 2 letters of  $a_3$ .

The Huffman coding procedure is based on two observations regarding optimal and prefix-free codes:

1. For a letter with higher frequency the code should produce shorter codewords, and



**Table 4.1:** Examples of a random binary code (#1) and a prefix-free binary code (#2). Code #2 is uniquely decodable, while for code #1 it's necessary to introduce boundaries between codewords to be able to distinguish them.

Letter	Code #1	Code #2
$a_1$	0	10
$a_2$	11	11
$a_3$	00	00
$a_4$	10	010
$a_5$	111	011

**Table 4.2:** Huffman code table

Letter	Probability	Codeword
$a_2$	0.4	$c(a_2)$
$a_1$	0.2	$c(a_2)$
$a_3$	0.2	$c(a_2)$
$a_4$	0.1	$c(a_2)$
$a_5$	0.1	$c(a_2)$

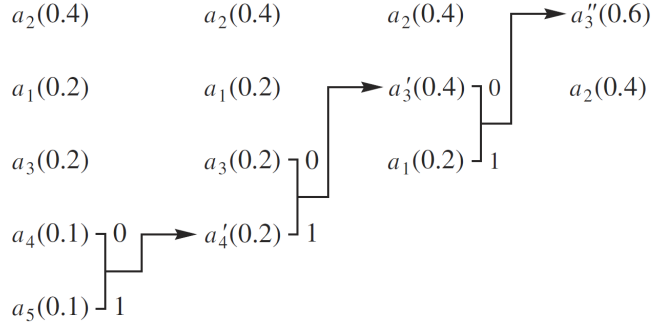
for letters with lower frequency it should produce longer codewords.

2. In an optimum code, the two least frequent codewords should have the same lengths.

From these statements the first is trivial to see that is correct. If the more frequent letters would have longer codewords then the less frequent letters, the average codeword length (weighted by the probabilities) would be larger than in the opposite case. Thus, more frequent letters must not have longer codewords than less frequent letter.

The second statement at first glance might be so intuitive, so let's consider the following situation. The two least frequent codewords do not have the same lengths, that is the least frequent is longer. However, because this is a prefix code, the second longest codeword is not a prefix of the longest codeword. This means, if we truncate the longest codeword to the same length as the second longest, they will still be distinct codes and uniquely decodable. This way we have a new coding scheme which requires less space on average to code the same sequence as the original code, from which we can conclude the original code was not optimal. Therefore, for an optimal code, statement 2 must be true.

To construct such a code, the following iterative procedure can be used. Let's consider an alphabet with five letters  $A = [a_1, a_2, a_3, a_4, a_5]$  with  $P(a_1) = P(a_3) = 0.2$ ,  $P(a_2) = 0.4$  and  $P(a_4) = P(a_5) = 0.1$  (Table 4.2). The entropy for this source is 2.122 bits/symbol. Let's order the letters by probability, and consider the two least frequent. Since the codewords assigned to these should have the same lengths, we can assign their codewords



**Figure 4.1:** Building the binary Huffman tree. The letters are ordered by probability, these will be the final leave of the tree. To join the to branches at every iteration we join the to nodes with the smallest probability, and create a new common node with the sum of the probabilities. This process is continued until all nodes are joined in a root node with probability of 1. Now, if we traverse down the tree to each leaf, the codeword will be defined by their position.

**Table 4.3:** Huffman code table

Letter	Probability	Codeword
$a_2$	0.4	1
$a_1$	0.2	01
$a_3$	0.2	000
$a_4$	0.1	0010
$a_5$	0.1	0011

as

$$c(a_4) = \alpha_1 * 0$$

$$c(a_5) = \alpha_1 * 1$$

where  $c(a_i)$  is the assigned codeword for letter  $a_i$  and  $*$  denotes concatenation. Now we define a new alphabet  $A'$  with only four letters  $a_1, a_2, a_3, a'_4$ , where  $a'_4$  is a merged letter for  $a_4$  and  $a_5$  with the probability  $P(a'_4) = P(a_4) + P(a_5) = 0.2$ . We can continue this process of merging the letters until all of them are merged and we have only one letter left. Since this contains all of the original letter, its probability is 1. We can represent the end result in a binary tree (see Figure 4.1), where the leaves are the letter of the alphabet, nodes are the merged letters, and the codewords are represented by the path from the root node to each leaf (compare with Table 4.3). The average length of this code is

$$l = 0.4 \times 1 + 0.2 \times 2 + 0.2 \times 3 + 0.1 \times 4 + 0.1 \times 4 = 2.2 \text{ bits/symbol} \quad (4.7)$$

A measure of the efficiency of this code is its redundancy—the difference between the entropy and the average length. In this case, the redundancy is 0.078 bits/symbol. The redundancy is zero when the probabilities are negative powers of two.

**Table 4.4:** Example alphabet for arithmetic coding

Letter	Probability
$a_1$	0.7
$a_2$	0.1
$a_3$	0.2

### 4.2.3 Arithmetic coding

Although in this case the redundancy of the Huffman code is minimal, however in cases where a few symbols have very high probability compared to the rest, the redundancy increases. This is simply because even for the most frequent letter the shortest codeword the Huffman code can produce is of length 1.

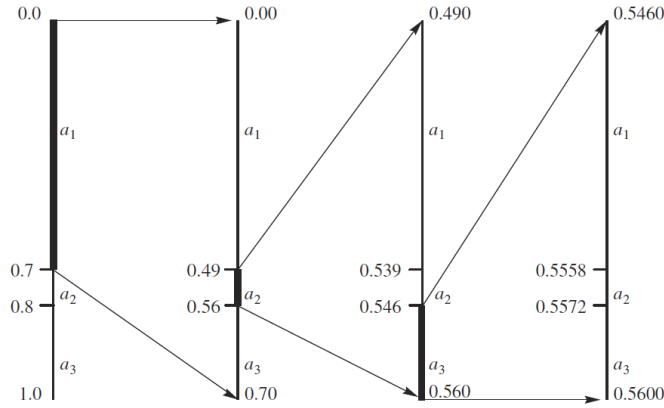
Let's consider the following example: take alphabet  $A = [a_1, a_2]$  with  $P(a_1) = 0.95$  and  $P(a_2) = 0.05$ . The first order entropy for this source is  $-0.95 \log 0.95 - 0.05 \log 0.05 = 0.2864$  bits/symbol, however if we assign the Huffman code we will have to use  $c(a_1) = 0$  and  $c(a_2) = 1$ , which means the average length will be 1 bit/symbol. This means in order to code this sequence using a Huffman code, we will need more than 3 times the number of bits promised by the entropy.

To get around this fundamental limitation, a coding scheme must be used which does not use discrete codewords at all. Arithmetic coding is the most well-known example of such a scheme. The idea is due to Peter Elias. He developed it during the same course on information theory in which Huffman developed his coding method, but he never published it.

In order to distinguish a sequence of symbols from another sequence of symbols we need to tag it with a unique identifier. One possible set of tags for representing sequences of symbols are the numbers in the unit interval  $[0,1)$ . Because the number of numbers in the unit interval is infinite, it should be possible to assign a unique tag to each distinct sequence of symbols. In order to do this, we need a function that will map sequences of symbols into the unit interval.

A straightforward algorithm to arithmetically encode a given input string is the following: Partition the unit interval  $[0, 1)$  into sub-intervals and assign one subinterval to each symbol in the input alphabet. The sizes of the sub-intervals are chosen to be proportional to the symbol frequencies. A string of symbols can be encoded by applying this process recursively: The sub-interval from the previous step is subdivided again using the same proportions. Figure 4.2. shows an example of arithmetic coding using the symbol frequencies given in Table 4.4.

This kind of coding although easy to understand, it's actually quite cumbersome to directly implement on a computer, where only a certain floating point precision can be achieved. For real life use this precision is often unsatisfactory, so some extra steps are involved in the coding.



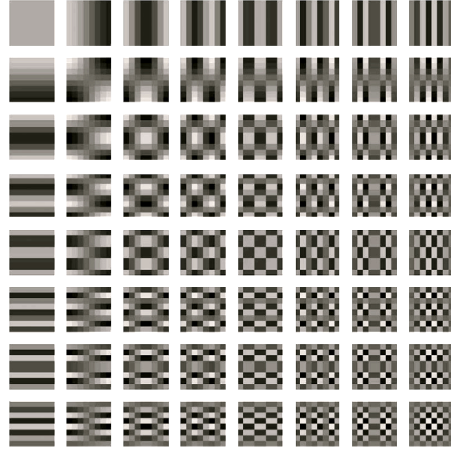
**Figure 4.2:** Arithmetic coding scheme in practice with the alphabet from Table 4.4 on the sequence  $a_1, a_2, a_3$

Since after each step the next subinterval is a subset of the previous interval, if the coding interval after a certain number of steps is contained in either the upper or lower half of the unit interval it will remain there for the rest of the coding. We can exploit this fact by rescaling that interval to the unit interval and writing either a 0 or 1 bit to the output depending on the position of the subinterval. By continuing this scheme it's possible to reach arbitrary precision even using a computer.

The decoding process is analogous to encoding. The decoder keeps track of the current lower and upper bounds. It mimics the rescaling operations of the encoder based on the bits of the encoded binary number.

### 4.3 Transform coding

The methods outlined in the previous section are effective at compressing the data in an optimal way and reaching the first order entropy, however they assume nothing about the structure of the data. For image compression it's important to also consider this, since most images have a relatively high level of autocorrelation, meaning that neighboring values have a high chance of being similar, although not necessarily equal. Transform coding is a technique that by itself does not compress the data, however by exploiting some knowledge of the structure it transforms the data effectively reducing its first order entropy. When regular entropy coding is then performed on the transformed data, because of the reduced first-order entropy, these techniques can achieve a higher rate of compression. At decompression, after decoding the entropy coder, the reverse transformation is applied to reveal the original data. In this section I will introduce two of the most important algorithms for transform coding, Discrete Cosine Transform and Wavelet Transform.



**Figure 4.3:** Basis functions of the  $8 \times 8$  2D DCT, computed as the outer product of the 1D basis vectors.

### 4.3.1 Discrete Cosine Transform

Discrete Cosine Transform [25] is closely related to the well known Fourier transform. The main idea behind this is to represent a function by a weighted sum of different sine and cosine functions. This provides a different view, instead of looking at the data in the time domain, we gain information about the frequencies that compose the signal.

Discrete Cosine Transform is a variant of the discrete Fourier transformation, however instead of making the signal periodic which can introduce large jumps at the edges, the signal is extended in a symmetric way. Since this will give a smooth transition even at the boundaries, the transform does not have to include so many high frequency components. Also, because of the symmetric extension, it's possible to represent the functions only by using the cosine bases, resulting in fewer coefficients. Overall, the DCT is much better suited for compression, than DFT.

The DCT base functions are defined in the following way:

$$c_{i,j} = s_i \cdot \cos \frac{(2j+1)i\pi}{2n} \quad \text{with } s_i = \begin{cases} \sqrt{\frac{1}{n}} & \text{if } i = 0 \\ \sqrt{\frac{2}{n}} & \text{otherwise} \end{cases} \quad (4.8)$$

The scaling factors  $s_i$  are chosen so that the  $L_2$  norm of each basis vector is 1 and so the transform is orthonormal. The inverse transform can therefore be found by simply transposing the transform matrix.

Applying the DCT to two dimensional functions, i.e. images is very similar to the two dimensional Fourier transform. The base functions are the outer products of the 1D base functions (Figure 4.3.), and the transform can actually be performed separately for each dimension.

Computation effort in a naive implementation is  $O(n^2)$  but since DCT is based on the Fourier transform, an  $O(n \log n)$  algorithm is also possible, analogous to the fast Fourier transform. This is still larger than a linear scaling with the data size, which can be very

inconvenient. Therefore, in practice the DCT is usually applied to smaller blocks of the image, such as  $4 \times 4$ ,  $8$  or  $16 \times 16$ .

Although DCT is an effective way of reducing the first-order entropy, it has a potential shortcoming by its use of floating point arithmetic. Because of the finite machine precision inherent rounding errors will occur, which means the reverse transformation can not generate the exact original data. For many application, such as photography, this still can be acceptable, but for scientific image data lossy compression is generally not accepted.

### 4.3.2 Discrete Wavelet Transform

Methods based on Fourier analysis, such as the DCT introduced in the previous section, give excellent localization in frequency space: They tell us exactly which frequencies occur in the data, which is very useful for data compression. However, they give no spatial localization: They do not tell us where in the signal these frequencies occur. Every DCT base function affect the thole image domain, which means distinct local structures can have a global effect on the final outcome. In case of an edge for example, it's necessary to include a high frequency component with a large coefficient, but since every DCT base function has an impact on the whole domain, this will have to be compensated on smoother regions by also increasing the coefficients of other factors. This can negatively impact compression performance.

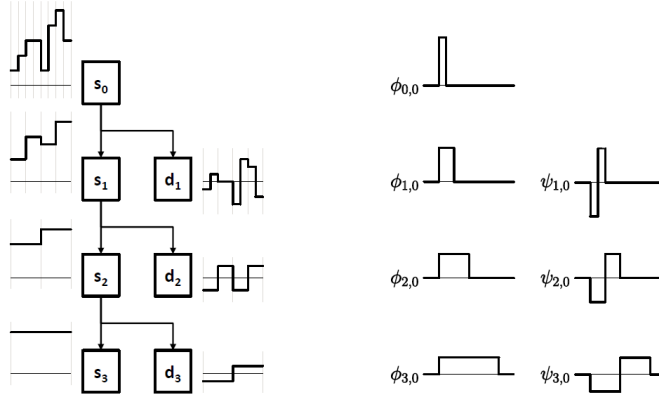
A solution for this is to use different base functions, namely ones with finite support. This way we will not only be able to get information about the frequency, but also about the localization of that frequency in some extent.

One option for such a set of local basis functions, and certainly the most popular one, is the multi-resolution analysis based on wavelets. The term “multi-resolution analysis” in the context of wavelets was introduced in the late 1980s by Stéphane Mallat [26], though research on wavelets had been ongoing for several years before that.

The idea behind the wavelet multi-resolution analysis is to build a basis out of translated and scaled versions of one underlying function called the *mother wavelet*  $\psi$ . The mother wavelet is non-zero only in a small region, leading to the locality properties. It is translated to cover the whole domain. It also covers only a small frequency band, and is scaled to cover higher or lower frequencies. The family of translated and scaled functions  $\psi_{l,i}$  is generated from according to

$$\psi_{l,i}(t) = \sqrt{2^l} \psi(2^l t - i), \quad l, i \in \mathbb{Z} \quad (4.9)$$

Incrementing  $l$  halves the width of the resulting function, which thus corresponds to a higher frequency band. Changing  $i$  moves the function along the x axis. The size of each step scales with the width of the function, defined by  $l$ . The normalization factor  $\sqrt{2^l}$  is chosen so that the  $L_2$  norm stays constant. The mother wavelet can be chosen so that the  $\psi_{l,i}$  are pairwise orthogonal and thus form a basis of some function space.



**Figure 4.4:** Multi-resolution wavelet decomposition, using the Haar wavelets as an example. Left: decomposition to multiple levels of low pass and high pass coefficients corresponding to the scaling and the wavelet functions respectively. Right: some of the base wavelet functions used for the decomposition.

However, representing a function in this basis will generally require an infinite number of basis functions  $\psi_{l,i}$ : To represent a constant component, i.e. content of frequency zero, the wavelet must be infinitely scaled. To address this, it is necessary to introduce an additional scaling function  $\phi$  which complements the wavelet. It is scaled and translated the same way as the mother wavelet.

The oldest wavelets are the Haar wavelets, and because of their simplicity, they provide a good example on how wavelet transformation works. The Haar wavelet scaling  $\phi$  and wavelet  $\psi$  functions are the following:

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

$$\psi(t) = \begin{cases} -1 & \text{if } 0 \leq t < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.11)$$

Clearly, all wavelet functions  $\psi_{l,i}$  are orthogonal. Additionally, the scaling functions  $\phi_{l,i}$  at a fixed level  $i$  are orthogonal. The scaling functions  $\phi_{l,i}$  are also orthogonal to the wavelet functions  $\psi_{k,i}$ ,  $k \geq l$  at the same and all finer levels.

Figure 4.4. (left) schematically shows the decomposition of a signal into low-pass components corresponding to the scaling function, and high-pass components corresponding to the wavelet. The signal  $s_0$  can be represented by the translated scaling function at a finer scale  $l_0$ , which can be decomposed to the sum of a coarser approximation  $s_1$  corresponding to the scaling function and the detail part  $d_1$  corresponding to the wavelet functions. This decomposition to a sum of coarser approximation and detail can be continued multiple times, thus resulting in the previously mentioned multi-resolution analysis.

Of course, for image compression applications, it is desirable to extend this transfor-

mation to a 2D version, that can be applied to the images before performing the arithmetic coding. Similarly to the DCT, the wavelet transform is also separable, and can be performed as a sequence of 1D transforms along different directions. Separable in a sense, that it does not matter in which order the individual 1D DWTs are applied to get the same 2D transformation.

### 4.3.3 Pixel prediction

## 4.4 Conclusion

In this paper I have introduced some of the most important techniques for image compression. Many of these are used in different image compression algorithms, such as JPEG [27], JPEG-LS [28] and JPEG2000. The common point for each of these formats is that they first transform the images to effectively reduce first order entropy, either by predicting the pixel values and coding only the differences (JPEG-LS), or by using either Discrete Cosine Transform (JPEG) or Discrete Wavelet Transform (JPEG2000). In the case of lossy standards after this transformation step a quantization is also performed depending on the desired quality setting for the coder. Finally as the last step the transformed and quantized coefficients are compressed by an arithmetic coder, such as Huffman coding in JPEG or arithmetic coding in JPEG2000. The end result in each case is a highly optimized compression that greatly reduces the file size.



## Chapter 5

# Discussion



## Acknowledgements

I would like to thank Lars Hufnagel for this opportunity to work in his group, and involving me in developing this microscope system. He gave me a lot of support, and helped in understanding and solving the possible difficulties that we came across during the development process.

I greatly appreciate the help of Uroš Kržič, who gave me an insight on the principles of selective plane illumination microscopy, and also prepared some illustrations for this work.

I thank Stefan Günther for providing the *Drosophila m.* embryos, and Ulla-Maj Fiuza for providing the *Phallusia m.* embryos.

Finally, I would like to thank Gustavo Quintas Glasner de Medeiros and Nils Norlin, as they both participated in the microscope design and program development, and both of them inspired me greatly. Without their help, I couldn't have written this paper.



# References

- [1] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, “Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy,” *Science*, vol. 305, no. 5686, pp. 1007–1009, 2004. [Online]. Available: <http://www.sciencemag.org/content/305/5686/1007.abstract> 3, 5, 11
- [2] U. Krzic, S. Gunther, T. E. Saunders, S. J. Streichan, and L. Hufnagel, “Multiview light-sheet microscope for rapid in toto imaging,” *Nature Methods*, vol. 9, no. 7, pp. 730–733, Jul. 2012. [Online]. Available: <http://www.nature.com/nmeth/journal/v9/n7/full/nmeth.2064.html> 5, 11
- [3] P. Strnad, S. Gunther, J. Reichmann, U. Krzic, B. Balazs, G. de Medeiros, N. Norlin, T. Hiiragi, L. Hufnagel, and J. Ellenberg, “Inverted light-sheet microscope for imaging mouse pre-implantation development,” *Nature Methods*, vol. 13, no. 2, pp. 139–142, Feb. 2016. 5
- [4] J. W. Lichtman and J.-A. Conchello, “Fluorescence microscopy,” *Nature Methods*, vol. 2, no. 12, pp. 910–919, Dec. 2005. [Online]. Available: <http://www.nature.com/nmeth/journal/v2/n12/full/nmeth817.html> 5
- [5] R. F. Service, “Three Scientists Bask in Prize’s Fluorescent Glow,” *Science*, vol. 322, no. 5900, pp. 361–361, Oct. 2008. [Online]. Available: <http://www.sciencemag.org/content/322/5900/361> 6
- [6] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, “A guide to choosing fluorescent proteins,” *Nature Methods*, vol. 2, no. 12, pp. 905–909, Dec. 2005. [Online]. Available: <http://www.nature.com/nmeth/journal/v2/n12/full/nmeth819.html> 6
- [7] S. Grill and E. H. K. Stelzer, “Method to calculate lateral and axial gain factors of optical setups with a large solid angle,” *Journal of the Optical Society of America A*, vol. 16, no. 11, pp. 2658–2665, Nov. 1999. [Online]. Available: <http://josaa.osa.org/abstract.cfm?URI=josaa-16-11-2658> 7
- [8] E. Stelzer and S. Grill, “The uncertainty principle applied to estimate focal spot dimensions,” *Optics Communications*, vol. 173, no. 1–6, pp. 51–56, Jan. 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030401899006446> 7
- [9] P. Davidovits and M. D. Egger, “Photomicrography of Corneal Endothelial Cells in vivo,” *Nature*, vol. 244, no. 5415, pp. 366–367, Aug. 1973. [Online]. Available: <http://www.nature.com/nature/journal/v244/n5415/abs/244366a0.html> 8
- [10] S. Aldaz, L. M. Escudero, and M. Freeman, “Live Imaging of Drosophila Imaginal Disc Development,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 32, pp. 14 217–14 222, Aug. 2010. [Online]. Available: <http://www.pnas.org/content/107/32/14217> 9
- [11] E. G. Reynaud, U. Kržič, K. Greger, and E. H. Stelzer, “Light sheet-based fluorescence microscopy: more dimensions, more photons, and less photodamage,” *HFSP Journal*, vol. 2, no. 5, pp. 266–275, Oct. 2008. 9, 11
- [12] E. H. Stelzer and S. Lindek, “Fundamental reduction of the observation volume in far-field light microscopy by detection orthogonal to the illumination axis: confocal theta microscopy,”

- Optics Communications*, vol. 111, no. 5–6, pp. 536–547, Oct. 1994. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0030401894905339> 9
- [13] U. Kržič, “Multiple-view microscopy with light-sheet based fluorescence microscope,” Ph.D. dissertation, Ruperto-Carola University of Heidelberg, Germany, 2009. 10, 13
- [14] H. Siedentopf and R. Zsigmondy, “Über Sichtbarmachung und Größenbestimmung ultramikroskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser,” *Annalen der Physik*, vol. 315, no. 1, pp. 1–39, 1902. [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/andp.19023150102/abstract> 11
- [15] J. Huiskens and D. Y. R. Stainier, “Selective plane illumination microscopy techniques in developmental biology,” *Development*, vol. 136, no. 12, pp. 1963–1975, Jun. 2009. [Online]. Available: <http://dev.biologists.org/content/136/12/1963> 11
- [16] M. Weber and J. Huiskens, “Light sheet microscopy for real-time developmental biology,” *Current Opinion in Genetics & Development*, vol. 21, no. 5, pp. 566–572, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959437X11001444> 11
- [17] P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. K. Stelzer, “Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy,” *Science*, vol. 322, no. 5904, pp. 1065–1069, Nov. 2008. [Online]. Available: <http://www.sciencemag.org/content/322/5904/1065> 11
- [18] H.-U. Dodt, U. Leischner, A. Schierloh, N. Jährling, C. P. Mauch, K. Deininger, J. M. Deussing, M. Eder, W. Zieglgänsberger, and K. Becker, “Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain,” *Nature Methods*, vol. 4, no. 4, pp. 331–336, Mar. 2007. [Online]. Available: <http://www.nature.com/nmeth/journal/v4/n4/full/nmeth1036.html> 11
- [19] K. Greger, J. Swoger, and E. H. K. Stelzer, “Basic building units and properties of a fluorescence single plane illumination microscope,” *Review of Scientific Instruments*, vol. 78, no. 2, pp. 023 705–023 705–7, Feb. 2007. [Online]. Available: [http://rsi.aip.org/resource/1/rsinak/v78/i2/p023705\\_s1](http://rsi.aip.org/resource/1/rsinak/v78/i2/p023705_s1) 13
- [20] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. John Wiley & Sons, Mar. 2007. 13
- [21] K. Sayood, *Introduction to Data Compression, Fourth Edition*, 4th ed., ser. The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufmann, 2012. 19
- [22] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948. 19
- [23] —, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, Oct. 1948. 19
- [24] D. A. Huffman, “A Method for the Construction of Minimum-Redundancy Codes,” *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952. 20
- [25] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete Cosine Transform,” *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, Jan. 1974. 25
- [26] G. Mallat, “A theory for multiresolution signal decomposition : the wavelet representation,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1989. 26
- [27] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*. Springer Science & Business Media, Dec. 1992. 28
- [28] M. Weinberger, G. Seroussi, and G. Sapiro, “The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS,” *IEEE Transactions on Image Processing*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000. 28