Towards a computational model of retinal structure and behavior

Matteo Paolo Lanaro

Dottorato di Ricerca in Informatica
XXXII Ciclo

Settore Scientifico Disciplinare INF/01 Informatica
Dipartimento di Informatica
Università degli Studi di Milano, Italy

Relatore: Prof. Alessandro Rizzi

Il Direttore della Scuola di Dottorato in Informatica: Prof. Paolo Boldi

A. A. 2018/2019

Abstract

Human vision is our most important sensory system, allowing us to perceive our surroundings. It is an extremely complex process that starts with light entering the eye and ends inside of the brain, with most of its mechanisms still to be explained. When we observe a scene, the optics of the eye focus an image on the retina, where light signals are processed and sent all the way to the visual cortex of the brain, enabling our visual sensation.

The progress of retinal research, especially on the topography of photoreceptors, is often tied to the progress of retinal imaging systems. The latest adaptive optics techniques have been essential for the study of the photoreceptors and their spatial characteristics, leading to discoveries that challenge the existing theories on color sensation. The organization of the retina is associated with various perceptive phenomena, some of them are straightforward and strictly related to visual performance like visual acuity or contrast sensitivity, but some of them are more difficult to analyze and test and can be related to the submosaics of the three classes of cone photoreceptors, like how the huge interpersonal differences between the ratio of different cone classes result in negligible differences in color sensation, suggesting the presence of compensation mechanisms in some stage of the visual system.

In this dissertation will be discussed and addressed issues regarding the spatial organization of the photoreceptors in the human retina. A computational model has been developed, organized into a modular pipeline of extensible methods each simulating a different stage of visual processing. It does so by creating a model of spatial distribution of cones inside of a retina, then applying descriptive statistics for each photoreceptor to contribute to the creation of a graphical representation, based on a behavioral model that determines the absorption of photoreceptors. These apparent color stim-

uli are reconstructed in a representation of the observed scene. The model allows the testing of different parameters regulating the photoreceptor's topography, in order to formulate hypothesis on the perceptual differences arising from variations in spatial organization.

Contents

Chapter 1

Introduction

Color vision is a process that starts with light entering the eye and ends in the visual cortex of the brain. The most crucial part of this process takes part in a tiny fraction of the retina, a region with a surface of less than 1 cm^2 consisting in the center of our gaze, where we possess the highest discerning power. The retina is the first real stage of visual processing right after the optics of the eye, consisting in 4-6 millions of color photoreceptors, connected to different processing layers leading to the optic nerve and outwards to the brain; its function is to discriminate, identify and estimate light signals, encoding them into signals defining spatial frequency, brightness, spectral composition and every other aspect contributing to our vision. Photon absorption contributes to the photoreceptor's output signal, while the comparison of different signals from cones with different photopigments is required for a correct color discrimination.

The tridimensional nature of color vision is observable in the three classes of cone photoreceptors in the retina, spatially distributed in a layer containing photosensitive cells. The perception of wavelengths in the range of the visible spectrum of light in an observed scene is mediated by cones, performing a discrete sampling of different portions of a retinal image. The evolution and development of our visual system has made this sampling strategy work successfully in natural scenes, based on the fact that spectral reflectance [1] slowly varies across the cones spatial scale, while for prolonged stimuli at elevated spatial frequencies some issues may arise due to the grain of the

---

[1] In optics, reflectance indicates the proportion of incident light that a given surface is capable of reflecting. It is represented by the ratio between the intensity of reflected radiant flux and the intensity of the received radiant flux.

trichromatic distribution of cones. This is evident for example in the recognition of color speckles inside of high frequency white and black patterns [1].

Both the perception of natural scenes and the small fluctuations in color derive from the same topographical layout of those cells, revealing the granularity of the cones mosaic but also the discrete nature of the color and spatial vision processing in the many layers of the retina. Moreover, the structure of cones in the retina is instrumental not only in color sensation, but also in spatial vision and motion processing. In light of these aspects, it seems evident that the topography of the photoreceptors in the retina is an essential factor for the visual processing, and in particular the positioning of the three wavelenght-sensitive classes of cones.

Most of the data available on the retina and the sensation of color comes from psychophysical visual experiments used to verify sensory theories. Usually the test performed involves visual threshold or discrimination of a white or colored signal, visual acuity, wavelength discrimination, brightness discrimination, hue naming, threshold of visual stimuli in the presence of an adapting background and a plethora of various phenomena. The most recent techniques of imaging *in vivo*, using adaptive optics to compensate for the ocular aberrations, have allowed capabilities to examine in detail the mosaic of cones [2, 3] and use previous psychophysical data with the observable structural variations in the photoreceptors mosaic.

This technology has made possible not only to get more information on the organization of the photoreceptor's mosaic in observers with normal color vision and in observers with deficits in the color or spatial vision, but also to observe the variations in density and count of photoreceptors in different age groups and their impact in the perception of color and movement. It is also possible to observe the correlations between structural changes of the photoreceptors layer and the changes in visual perception in patients with congenital or age related retinal anomalies.

1.1   Aims and scope

The aim of this dissertation is to propose a computational model of the human retina capable of contributing to the knowledge of perceptive phenomena related to the spatial distribution of retinal photoreceptor cells, and

providing a visual representation of an observed scene, given the fact that the process of vision happens both at retinal and cortical stages. The topography of photoreceptors is correlated to a wide range of perceptive phenomena, like the vast interpersonal difference in the ratio of M to L cones resulting in a negligible difference in perception and the filling-in taking place in the retinotopic region corresponding to the optic disc, issues that are actually still open. The computational model takes into account the response of photoreceptors to color stimuli and simulates connectivity of retinal neurons, given the fact that said responses are tied to the synaptic interconnections between cells. Simulation of the mechanisms of human color vision and spatial sensitivity is a powerful tool for the imaging industry, helping the design of cameras, sensors and displaying techniques. Understanding the biological correlation between retinal structure and visual sensitivity is a helpful tool for the diagnosis of retinal degenerative diseases.

1.2   Starting point

There are a lot of studies concerning the characteristics of the human retina and how they relate to the sensation of color. The complex structure of the retina is composed by numerous biological elements with different characteristics, and their behavior is regulated by as many parameters. The space variance of the cones sampling structure is one of the main concerns. Log-polar mapping is practically used as an approximation of the retinal mapping, but the way photoreceptors are spatially arranged in their own retinal layer is fundamentally different and regulated by other principles. The basic idea is that differences in the composition of retinal mosaics are resembling of the differences between fingerprints: there are no equal retina or fingerprints between individuals, yet they can be regarded as a sample from the same constrained random process. These characteristics determine the limits of the sampling properties that are observable across all human retinas.

The human eye model of Deering [4] proposes a cone synthesizer that designs cone mosaics from an algorithm that generates cells starting from a seed and then surrounding them with points defining a polygon, constituting the cell boundary. This cone synthesizer is validated by calculation of the neighbor fraction ratio and by matching statistics of density in cones/mm^2

with values in literature.

The most recent works on retinal modeling are focused on the neural behavior [5, 6, 7]; for example, *Virtual Retina* by Wohrer and Kornprobst is a large scale simulation software that transforms a video input into spike trains, designed with a focus on nonlinearities, implementing a contrast gain control mechanism. Few works present a quantitative approach at modeling the visual pipeline. One of the most recent is ISETBIO [8], a framework designed for exploring the properties of the earliest stages of biological visual systems, estimating the effects of human optics, eye movements, cone absorptions and photocurrent, and retinal cell properties.

The color reconstruction simulation has been inspired by demosaicing techniques. Akin to the photodiodes in a camera sensors, human photoreceptors are considered to be able to capture only the informations related to their spectral sensitivity, and for this reason they can not provide exact punctual data of color in that location. A demosaicing process is used to address this limitation, interpolating between different responses spatially distributed, to generate the correct color information. Many algorithms exist to reconstruct the missing color information of the pixels. Some of them interpolate the value according to a fixed scheme (no-adaptive algorithms), while others, more complex, are able to interpolate according to the spatial features of the neighboring elements (adaptive algorithms).

1.3   Methods

In this thesis a quantitative approach has been taken in order to model the functioning of the cone photoreceptors of the retina and the underlying layers, comprising the connections with bipolar and horizontal cells up to the ganglion cells. While the bipolar cells act as a direct pathway from the photoreceptors to a ganglion cell, the cell responsible for sending visual information directly to the brain, the function of the horizontal cells is to horizontally share information to influence the closest bipolar and photoreceptor cells. Every ganglion cell receives synapses directly from a group of photoreceptors, varying from one in the center of the fovea to thousands in the retinal periphery. The color sensation obtained from a single cone is therefore incorporated into a receptive field that has to take into account both the sensation from that single cone and all the cones in a determined

radius from it. The first step of this simulation, which starts from a strictly biologically plausible standpoint, is the creation of a model of spatial distribution of cones inside of a retina, that preserves the spatial characteristics of the cells mosaic present in the literature, and validating it through comparison with real retinas. Subsequently, inside of the model the spatial descriptive statistics are created for each photoreceptor, to contribute to the creation of a graphical representation. This representation is based on a behavioral model that determines how light is sampled and processed in the retina, relating to the absorption of photoreceptors, and how these apparent color stimuli result in a visual representation of a scene that we are observing, also taking into account the connections of the retinal circuitry and the retinal layers underlying the photoreceptors. These latter stages of the model, while starting from a biologically plausible spatial distribution of cone cells, have to simulate the behavior of such complex biological systems that a *black box* simulative approach had to be considered.

The proposed model is organized into a modular pipeline of extensible methods each simulating a different stage of visual processing, from a visual scene through the cone mosaic and to the reconstruction of the color information inside of a visual representation.

1.4   Thesis organization

The organization of the dissertation is as follows:

In chapter 2, after a brief description of the theory of light, an overview of the human visual system is proposed, with details of how visual signals are neurally encoded in the process of vision and examples of visual perceptive phenomena and anomalies.

In chapter 3 is presented the state of the art on the modeling of the human visual system, with attention to biologically-inspired foveated imaging systems, studies concerning the spatial distribution of the cones sampling structure and an overview of models of the retina with their goals and methods, and lastly an introduction to techniques of spatial color reconstruction.

Chapter 4 explores the details of the retinal topography and how the spatial distribution of photoreceptor classes in the retina may affect the sampling of color signals.

In chapter 5 different cone array sampling strategies have been tested

to reproduce the spectral and density characteristics of a real photoreceptor distribution. A method for generating space-variant cone arrays is proposed.

In chapter 6 the proposed cone sampling model is described, going into detail of its computational pipeline. Its objective is to simulate the absorption of light into artificially-generated retinal distribution of the cone's trichromatic mosaic.

Lastly, chapter 7 proposes a tool for reconstruction of the color signals into a space-variant image. It also presents some tests performed with the model to determine how the proposed color reconstruction method behaves with different ratios of L to M cones or visual illusions.

Chapter 2

Color Vision

Human vision is an extremely complex process that starts with light entering the eye and ends inside of the brain. In this chapter, after a brief presentation of the theory of light, the early stages of the human visual system are discussed, from the anatomy and structure of the human eye towards the retina and how light is neurally encoded to allow us to sense colors. Also some aspects of visual perceptive phenomena related to color vision are here discussed with some examples and anomalies.

2.1 Theory of Light

The nature of light is a complex concept that has been debated since ancient times, arguably starting with Euclid in 300BC. In the 17th century, the corpuscolar theory of light hypothesized by Pierre Gassendi and developed by Sir Isaac Newton depicted light as a flux of tiny particles called *corpuscles*, having negligible mass, traveling at high velocities and carrying kinetic energy. The different colors of light were due to the different size of the corpuscles.

Meanwhile in the same years, Christiaan Hughens was developing a mathematical wave theory of light, based on the findings of Robert Hooke. His theory described light as a series of waves departing from a source and that could interfere with each other like sound waves. James Clerk Maxwell, in 1862 found out that light is a particular form of electromagnetic radiation akin to radio waves, propagating with the same speed of 299,792,458 m/s and exhibiting the same properties, with wavelength approximately between

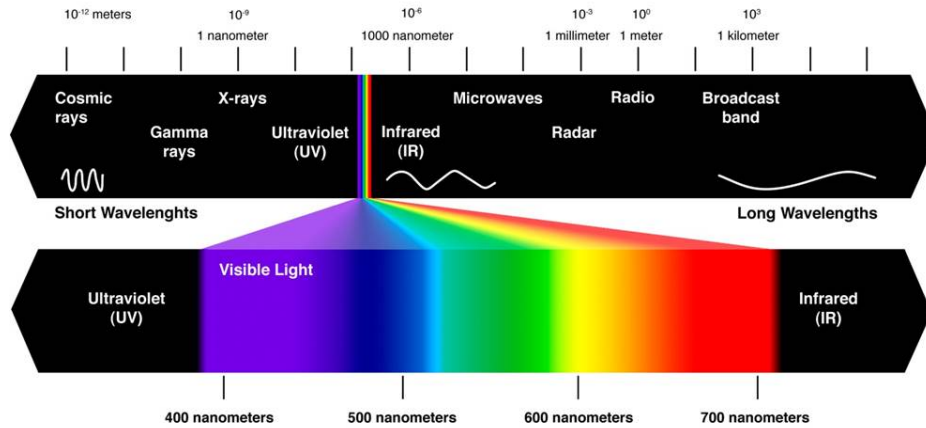380 and 730 nanometers of wavelength, as displayed in Figure2.1.

Figure 2.1: Visible light is an electromagnetic radiation in the range of about 380 to 730 nanometers of wavelength. Image taken from `https://commons.wikimedia.org/wiki/File:Visible_and_invisible_light.jpg`

The current theory of light was advanced by Albert Einstein in 1905 according to the quantum hypothesis of Max Planck, winning him the Nobel Prize in 1921. The *quantum theory* depicted light as composed of discrete quantum particles, later called *photons*. This theory confirms that light can behave both as a particle as well as a wave.

For the measurement of color it is sufficient to only consider light as an electromagnetic radiation propagating as a wave, physically described by its spectral composition in the form of a spectral power distribution (SPD) of the amount of energy in function of its wavelengths. However, in the human eye both wave and particle phenomena are important for the sensation of light and color.

2.2   The Eye

The ocular bulb (Figure 2.2) is an asimmetrical sphere, with front to back diameter of 24-25 mm in adults and an approximate volume of 6.5 cc. It is situated in the orbit of the skull, which contains and protects it. The eye is covered in three concentric coats or layers, from the outermost to the innermost:

- Fibrous tunic, comprising the sclera and the cornea

- Vascular tunic, consisting of the iris, ciliar body and choroid

- The retina

It is also possible to distinguish three cavities: the anterior and posterior chamber, separated by the iris, filled with watery aqueous humor, and a posterior cavity filled with gel-like vitreous humor. The intraocular pressure exerted by the fluids inside of the eye support the shape of the cavities.
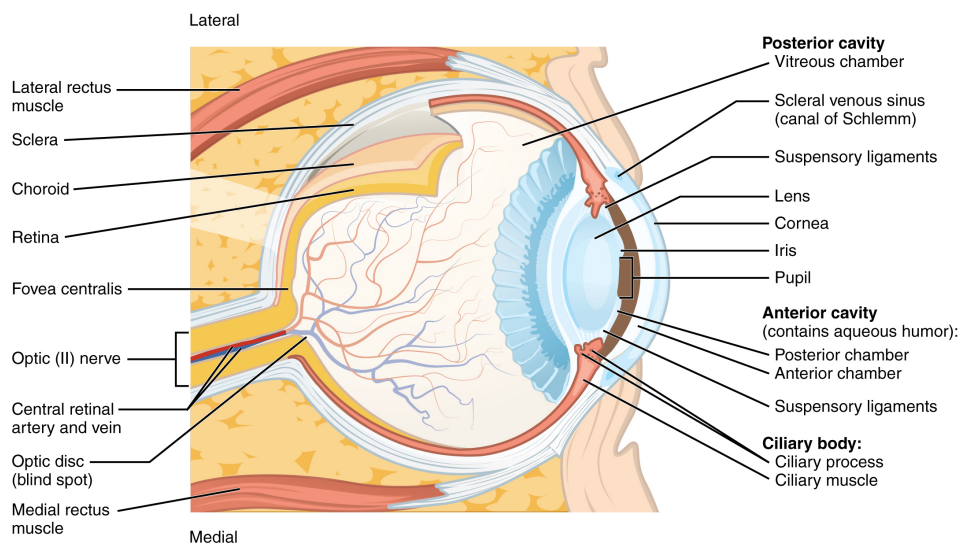

Figure 2.2: Structure of the human eye. Image courtesy of Betts et al. [9]

From an optical point of view, the eye can be considered as a spheroidal *camera obscura* composed by an external convex lens, the cornea, and an internal crystalline lens of biconvex shape, interspersed with fluids. On the anterior surface of the lens lies the iris which, by contracting or relaxing, regulates the quantity of light entering the eye through a circular hole acting as a diaphragm, the pupil. The pupil opening is regulated by an involuntary mechanism that causes its change in size, from a minimum of 2 mm to a maximum of 8 mm, inversely based on the quantity of light. The surfaces of the cornea and lens compose an approximately centered system and their center of curvature define the optical axis of the eye.

Images of objects processed by the optical system of the eye are projected, shrunken and inverted upside down, on the inner surface of the ocular bulb, coated by a thin membrane of neural tissue, the retina. The crystalline lens, by changing its shape according to involuntary contractions of the cil-

iary muscle which supports it, focuses the image so that it is projected in a small central region of the retina, about 6 mm in diameter, called *macula*. At the center of the macula lies a small depression with a diameter of less than 2 mm called *fovea centralis*, the region with the highest the highest density of cones corresponding with the region of highest visual acuity in the retina. When the eye fixates on something, every point of space which is projected on the surface of the retina is visible, constituting the visual field. Properties of vision are not uniform in the visual field, but rapidly change by straying from the visual axis (different from the optical axis, defined as the line passing from the eye's rotation center to the center of the fovea) towards the periphery of the visual field, both for optical properties and for non-uniformity of the retina.

2.3 Structure of the Retina

The retina is a thin membrane of neural tissue that covers the inner part of the ocular bulb, and is regarded as an extension of the brain since they are linked by the optic nerve that directly extends from it. The retina is the only tissue containing the photoreceptors, particular neural cells of elongated shape capable of a process called *phototransduction*, wherein a protein called *opsin* hit by a photon undergoes photoisomerization changing the cell membrane's potential and therefore converting a light stimulus into an electric signal [10]. The photoreceptor cells extend across several anatomic layers of the retina: the *Bacillary* layer contains their outer segments and inner segments, in the second closest layer to the choroid, right after the retinal *pigment epithelium*, a layer of highly pigmented cells directly faced by photoreceptors, away from incidental light. This latter layer has the function of both sustaining the photoreceptors health and integrity and absorbing scattered light. Figure 2.3 illustrates the structure of the retina, with the light coming from the outside of the eye on the left side.

Photoreceptor cells are of two kinds: rods and cones. The cones are overall less numerous in the retina and have a wider inner segment diameter than the rods, which are longer and thinner, with the exception of the fovea where their diameter becomes similar. Rods and cones serve different purposes in the process of vision. The cones are responsible for the *photopic* vision, occurring in conditions of daylight illumination, and are the only cells
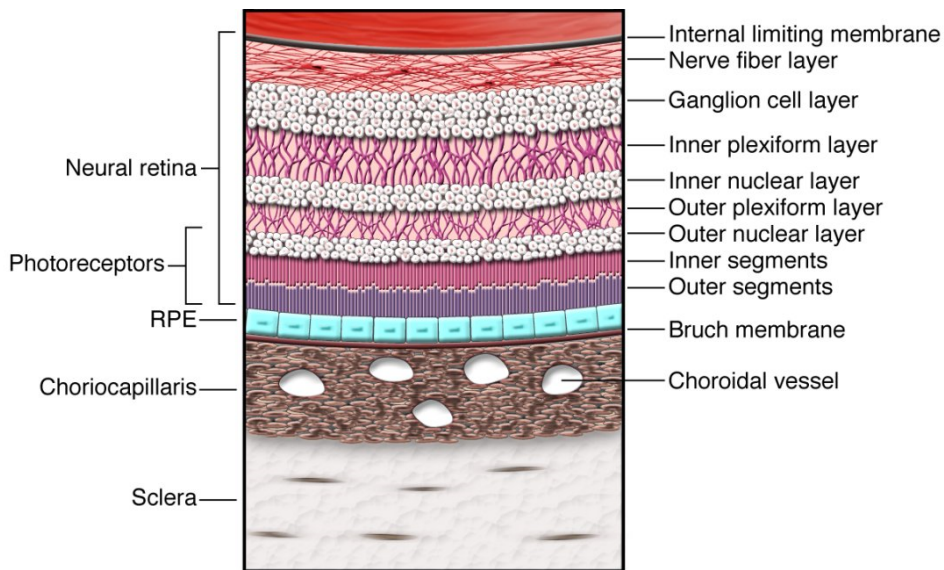
15

Figure 2.3: Photoreceptor and nerve layers in the retina, image courtesy of M. Friedlander [11]

capable of the perception of chromatic signals. Meanwhile, rods are able to detect even the smallest amount of radiant energy, generating responses even to a single photon event, but they can only yield achromatic or grey levels of color sensation. Rods are characteristic of the human *scotopic* vision, which happens in conditions of low illumination or at dusk, characterized by the inability to detect colors and fine details [12]. Photoreceptors will be discussed in more detail in section 2.4.

Going further into the retinal layers after the bacillary layer, the *Outer Nuclear Layer* is comprised of bodies of the photoreceptors, then the *Henle's Fiber Layer*, only present in para- and perifovea, contains bundles of unmyelinated cone and rod photoreceptor axons terminating in the pedicles and spherules [13] that synapse in the retinal *Outer Plexiform Layer*, containing a synaptic network between the terminals of the photoreceptors and the dendrites of horizontal and bipolar cells. The *Inner Nuclear Layer* follows with the bodies of horizontal, bipolar, amacrine and Müller cells, followed by the *Inner Plexiform Layer*, with its name given from the synaptic plexus composed by axon terminals of the aforementioned cells and ganglion cells dendrites. The next layer is the *Ganglion Cell Layer*, containing the bodies of ganglion cells and displaced amacrine cells, ending in the *Nerve Fiber Layer* that, as implied by the name, is constituted by the nerve fibers

of these cells, extending all over the retina until converging into the fibers of the optic nerve.

Retinal regions are also concentrically organized based on their eccentricity from the visual axis in the fovea and were first defined by Polyak in 1941 [14]. Inside of the macula, starting from the center it is possible to distinguish the *foveola* (175 μm radius), the *foveal floor* (200 μm radius), the *parafovea* (μm 1250 μm radius), and the *perifovea* (2750 μm radius). Outside of the macula, Polyak defined regions of the peripheral retina as the *near periphery* (4250 μm radius), the *far periphery* (10000 μm temporal and 16000 μm nasal radius), and the *extreme periphery*, going all the end to the ora serrata.

3 to 5 mm from the center towards the nasal retina lies the optic disc, a retinal region devoid of photoreceptors where the process of vision is absent, hence corresponding to the so called blind spot of the eye. This exact region corresponds to the exit point of the eye for the optic nerve, where all the axons of the ganglion cells converge, and an enter point for the central retinal artery and vein that supply blood to the retinal tissue. The estimate distribution of cones and rods across a horizontal section of the retina can be observed in Figure 2.4, it is possible to see a gap from approximately 16 to 20 degrees of visual angle in which photoreceptors are absent, corresponding to the optic disc.

We are usually unaware of the region corresponding to the physiological blind spot since in our everyday vision it is perceptually filled-in, but there is no universally agreed explanation behind this phenomena.

2.4   Photoreceptors

There are approximately 130 million photoreceptors in the retina carrying on the phototransduction of light signals. Their outer segment acts like an optic fiber channeling incident photons inside of them. The head of an outer segment consists of a membrane folded into floating stacks of disks, and just by looking at their size and shape one can easily identify whether a photoreceptor is a cone or rod. In fact, with the sole exception of the fovea where photoreceptors are tightly packed together, cones are overall less numerous on the retina and appear to be shorter, with larger section and possessing a characteristic conic shape, while the rods have a longer and thinner form
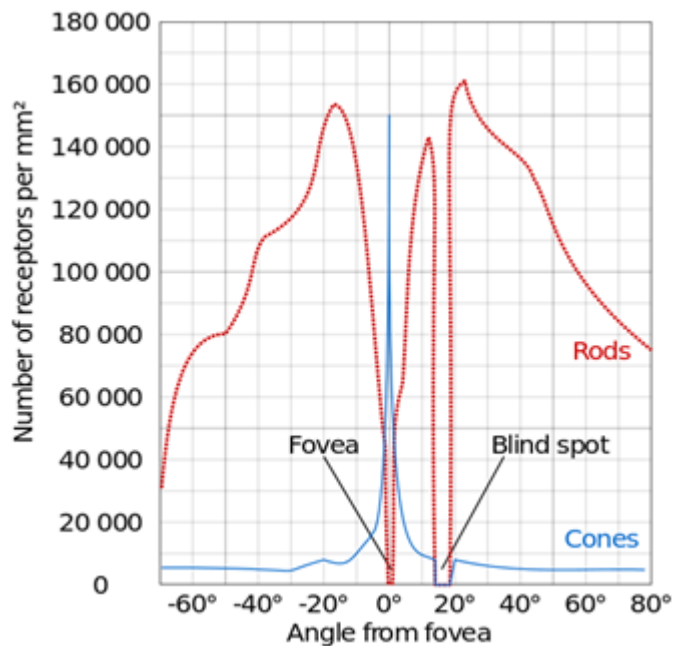
Figure 2.4: Density of photoreceptors across a horizontal section of the retina, image based on B. Wandell [15]

factor. The disks in their outer segment are filled with photopigment, constituted by the union of an organic compound called *retinaldehyde*, or vitamin A, with the heptahelical protein *opsin*. When struck by photons, the photopigment in the disks undergoes a pysiochemical transformation generating a difference in the electric potential of the cell, transmitted to cells in the subsequent layers of the retina via synaptic transmission, perpetrating the process of vision.

The human photoreceptors contain four types of opsins:

- A class extremely sensitive to light (Rh1), only found in rods, called *rhodopsin* and used in night (scotopic) vision

- Three cone opsins, called *photopsins*, used in photopic vision, respectively:

  - A short-wavelength sensitive class (SWS1), found in S cones
  - A medium-wavelength sensitive class (MWS), found in M cones
  - A long-wavelength sensitive class (LWS), found in L cones

18

It is also worth noting that a fifth class of opsins exists, but they do not participate in the process of vision. Discovered in 1997 by Provencio et al. [16], the *melanopsin* is found in ganglion cells labeled as *intrinsically photosensitive retinal ganglion cells* (ipRGCs), and their function is regulating circadian rythms.

Cones and rods are both active in conditions of intermediate illumination, defined as *mesopic* vision. With a progressive decrease in illumination, long-wavelength sensible cones tend to have their response lowered quicker than short-wavelength cones, causing a shift in our color perception towards shorter wavelengths. The photoreceptor luminosity functions are visibile in Figure 2.5 left, and represent the average spectral sensitivity of the perception of brightness at every wavelength, obtained by comparison ot the brightness of a 555 nm monochromatic light, where the eye is most sensitive, with the brightness of another monochromatic source of differing wavelength.
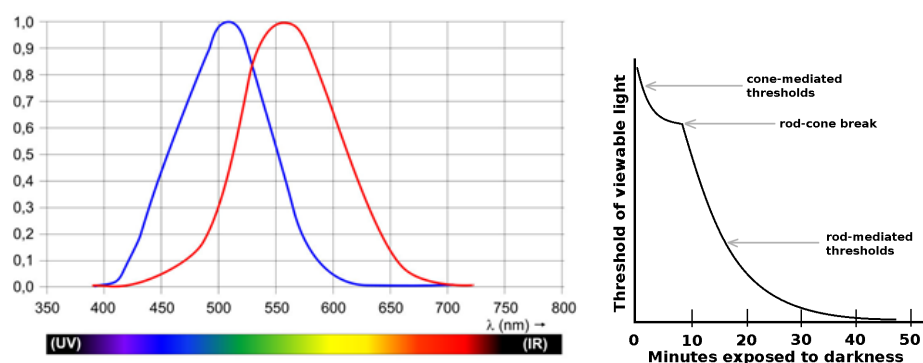

Figure 2.5: **Left**: Scotopic (blue) and photopic (red) luminosity functions. The horizontal axis is wavelength in nm, the vertical axis is relative sensitivity, image taken from `https://commons.wikimedia.org/wiki/File:LuminosityCurve1.svg`. **Right**: Adaptation of the human eye in darkness, public domain image courtesy of `https://wtamu.edu/~cbaird/`

The photopic luminosity function, also called V(λ) was standardized by the CIE (Commission Internationale de l'Éclairage) in 1924, and is used for the conversion of radiant energy into luminous energy, while the scotopic luminosity function V'(λ), adopted later in 1951, shifts towards the blue wavelenghts. This change in spectral sensitivity from mesopic to scotopic manifests itself in different conditions of illumination with a variation of contrast, making red tones appear brighter in daylight and blue-greenish tones

19

brighter at dusk, this effect's name is *Purkinje shift* and is rendered in Figure 2.6.

The duality of mechanisms of scotopic and photopic vision can be visualized in the dark adaptation curve in Figure 2.5 right. After 10 minutes of exposure to darkness, the cones stop working and our vision is mediated by rods only, since their threshold of sensitivity to light is much lower. The cause of this curve is in the different speeds of photopigment regeneration in cones and rods, in dailyght vision the rods are bleached by the intensity of the light and they take approximately 30 minutes to regenerate to their normal sensitivity, while for the cones this time is 7 minutes only.

The photoreceptors organization varies widely from the fovea to the periphery, as will be discussed in detail in chapter 4. In general, outside of the fovea it is possible to find more rods than cones, and a higher ratio between photoreceptors and ganglion cells. The combined effect of this organization manifest itself in the higher sensitivity to light within a few visual degrees from the center of fixation including the majority of the macula surface, since rods are specialized for lower levels of light intensity and each ganglion cell receives information from a larger photoreceptor pool, in a process defined retinal summation. We can personally notice the consequences of this organization during a starry night, when our vision is adapted to darkness. By fixating a bright star, we can find in our peripheral vision another dimmer star: by shifting our gaze towards it can be observed that it will disappear when projected towards the central
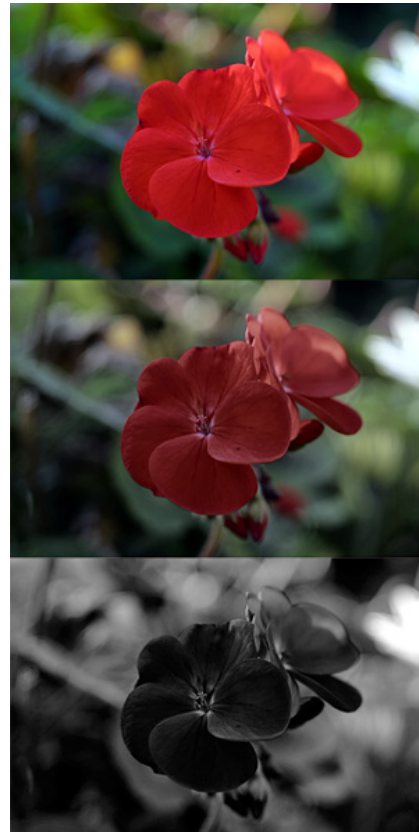

Figure 2.6: Purkinje shift. Red appears brighter in daylight, while the green foliage appears brighter at dusk. Derivative image taken from `https://commons.wikimedia.org/wiki/File:Red_geranium_photoic_mesopic_scotopic.jpg`

region of the retina, but it will appear again when projected a few degrees of visual angle from the center. Moreover, signal from rods is enhanced by retinal convergence, which is the sharing of a single nerve fiber in the retina by several rods.

2.5   Visual Acuity

The ability of our vision to resolve details is known as visual acuity. Visual acuity is limited in the few central degrees of vision by the photoreceptor's spacing [17], while outwards after the macula by the spacing of ganglion cells [18], gathering signals of the photoreceptors in their receptive field. An emmetropic human eye can distinguish patterns of alternating black and white lines with features as small as one minute of arc (1/60 degree or $\pi/(60*180) = 0.000291$ radians), while at higher spatial frequencies the pattern will result in a neutral grey, as well as lower contrast patterns at maximum spatial frequency. Considering a pattern being viewed at distance d from the eye, the length of a distinguishable line is calculated with: *length = angle (in radians) * d = 0.000291 * d.* For instance, when observing an object at a distance of 25cm, the human eye can resolve lines of 0.0727 mm, corresponding to a spatial frequency of 6.88 line pairs per mm.

To specify the resolution and perceived sharpness the Modulation transfer function (MTF) is used, defined as the spatial frequency response of an imaging system or a component, or the contrast at a given spatial frequency relative to low frequencies, and is measured in lines per mm (lp/mm), with higher spatial frequencies corresponding to finer details. The proposition at the basis of this function is that a sinusoidal grating is always imaged in an optical system as a sinusoidal grating, but with lowered contrast [19]. Since in our eyes, before arriving on the retina, light passes through an optical system composed of lenses and fluids in the same way it would in a camera system, the response is lowered at higher spatial frequencies; this means that optical systems can be considered as low-pass filters. The specification of this imaging property can be expressed in a graph showing the demodulation of sinusoidal gratings as a function of their spatial frequency. For example, when observing a pure tone sine wave pattern, at frequencies where the MTF is 100% the pattern is unattenuated, retaining full contrast, at 50% the contrast is half the original value, and so forth.

The psychophisical equivalent of the MTF is the Contrast Sensitivity Function (CSF), first proposed by Mannos and Sakrison [20] and calculated with the formula

$$A(f_r) = 2.6 \cdot (0.0192 + 0.114 \cdot f_r) \cdot e^{-(0.114 \cdot f_r)^{1.1}}$$

where f_r is the given spatial frequency of the sinusoidal pattern in cycles/degree. The function, visible in Figure 2.7, assumes a peak value of 1.0 for $f_r = 8.0$ cycles/degree and declines up until 60 cycles/degree, coherent with the spacing of photoreceptors.
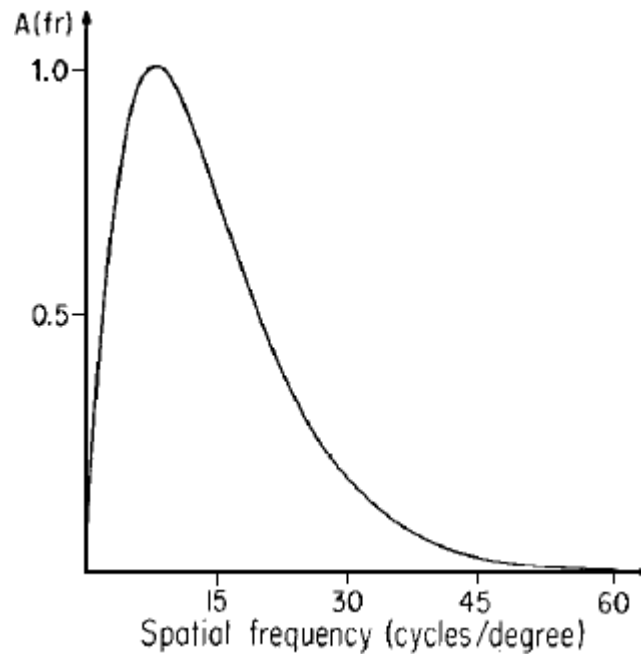

Figure 2.7: Contrast Sensitivity Function as calculated by [20]

2.5.1 Vernier Acuity

Among visual acuity test, there is one in particular called the Vernier acuity test. It consists in showing two vertical misaligned parellel lines to an observer which has to decide whether the top line is aligned to the left or right of the bottom one. For each iteration of the test, the top line is randomly misaligned towards a side and every time the observer picks the correct side, a new arrangement where the lines are closer is presented. This separation

can reach distances inferior to a millimeter, even inferior to the resolving limit of the photoreceptors, but still our visual system is capable of detecting misalignments with accuracies of one second of arc [21]: this phenomena is known as the Vernier alignment acuity.

The Verner acuity is a fine example of hyperacuity, defined as a discrimination ability that surpasses the acuity limit. In the case of the human eye, the visual acuity at the fovea defined by the spacing of individual photoreceptor cells allows to resolve two separated lines with a distance of at least 1 arcminute, nonetheless we are able to detect a misalignment of two lines of a tenth of this value.

Different explanations to this phenomenon have been proposed, involving eye movements and temporal integration of the visual signal. The theories concerning eye movements, and more specifically fixational eye movements (described in detail in the next section), state that these movements are known to cause a blur on the image projected on the retina, but this blur might in turn help to fill the empty spaces between the cones [22]. Findings from Ko et al. [23] seem to indicate an enhancement in fine spatial discrimination in presence of microsaccades. However, cones and rods require a certain interval of time to adapt to any given visual condition, exhibiting an hysteresis behavior: the time required for this excitation-relaxation cycle seems to invalidate the idea of the contribution of small movements to enhancing the acuity of the eye by combination of visual images [24].

This enhancement mechanism exists inside of the HVS, but it has to take into account also the information theory: information is either captured or not captured, but it cannot be figured out from data which does not exist. Still, the consensus is that the enhancement of the resolving power in the human eye is dependent exclusively from neural post-processing.

2.6   Eye movements

Eye movements are mechanisms specialized for stabilizing an image on the fovea, compensating for moving objects in our visual world, head movements, or both. They are actuated by three pairs of extraocular muscles which move the eyeball in the eye's orbit and can be classified into:

- Pursuit: is an involuntary movement that compensates for drifts of a visual target, by moving the eyes in a direction that reduces the drift.

- Vergence: the eyes move in opposite directions to direct the point of sight, to track in three dimensions and compensate for changes of depth.

- Optokinetic reflex: in conjunction with the vestibular reflex, compensates for movements of the head and body that affect the entire visual field, e.g. locomotion, head rotation.

- Saccades: are binocular movements that suppress reflexes to rapidly move to a new fixation point. These movements are so quick that no visual information can be correctly processed, so our visual system performs an active suppression of vision during saccades.

Our gaze focuses on a visual target for periods of time called fixations, usually lasting 200 to 300 ms, separated by saccadic movements. Eye movements are the most frequent kind of movement that humans make, during our life we generate more saccades than heartbeats. In the real world we are seldom presented with just a single potential target: we must choose between many, and some will have more significance than others.

Figure 2.8: An example of a scan path over an image. Light blue circles indicate fixations with diameter proportional to duration of the fixation, yellow arrow indicate the saccades and the blue path is the actual movement of the gaze point. Image obtained with SR Research Data Viewer `https://www.sr-research.com/data-viewer/`

When we see a visual scene, our gaze usually shifts to the region with the highest information density, then quickly to other locations determined subjectively, without a proper awareness of the decision process. This series of fixations generates a scanpath, defined as a pattern of exploratory eye movements, an example of which is visible in figure 2.8.

During fixations our eyes are never perfectly still, but perturbed by small undetectable ocular motions called "fixational eye movements". Their role is to contrast the effects of neural adaptation during unvarying stimuli, preventing the effect of perceptual fading, like in the Troxler's effect optical illusion in Figure 2.9. Three types of eye motion can be distinguished dur-

Figure 2.9: Troxler's effect. When fixating the center point for enough time, the surrounding circle starts to fade and eventually disappear until our gaze shifts from that point. Image courtesy of `https://www.illusionsindex.org/i/troxler-effect`

ing a fixation: tremor, drift and microsaccades. Tremor is defined as an aperiodic, wave-like motion of the eyes with frequency of 30 to 100 Hz and angular extent of 10-20 seconds of arc, about the diameter of one cone in the fovea [25]. Visual tremor is also reported to be generally independent in both eyes, generating a physical limit to the ability of matching corresponding visual points in the retina during steroscopical vision. Drifts are a movement that happens in conjuction with tremor, resulting in a slow motion of the eye causing the image to move across an extent of 5 minutes of arc. Microsaccades are tiny and unnoticeable jerking movements of the eye occurring during voluntary fixation, usually moving the retinal image

across 2 to 25 minutes of arc and happening in about 25 ms [26]. These movements have been theorized to be functional during fixations the same way that saccadic movements are during the exploration of a scene.

2.7   Trichromatic Vision

The theory of trichromatic vision was first explored by Young in 1802 and later developed by Helmholtz 50 years later, also referred to as the Young-Helmholtz theory. It states that there are three independent variables in color vision, like in the three colors of additive synthesis red, green and blue, or the three colors of subtractive synthesis, cyan, yellow and magenta. The tristimulus theory was later confirmed in the experiments of Wright and Guild in the early 20's, with the goal of generating the data necessary for constructing a formal, quantitative, rigorous system for characterizing colors. Subjects with normal color vision were shown a circular hole divided in two: on one side it was projected a sample of monochromatic light of known wavelength, on the other side were projected three linearly independent monochromatic lights ($r_\lambda = 650nm$, $g_\lambda = 530nm$ and $b_\lambda = 460nm$) with the possibility of regulating their intensity. By manually adjusting the intensities of the three monochromatic lights, the observers had to reproduce the color shown in the other side of the circle, containing the reference sample wavelength. Their results have shown that by mixing three primary lights, it is possible to obtain almost all of the fundamental colors of the chromatic spectrum. The Wright and Guild experiments were conducted separately, and since their results agreed well with each other, they were later combined by the CIE to produce the RGB color matching functions, namely $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$, presented in Figure 2.10. Each of these functions defines the amount of three primary lights ($r_\lambda = 645nm$, $g_\lambda = 526nm$ and $b_\lambda = 444nm$) required to match monochromatic targets of equal energy.

For some particular wavelengths, however, the equivalent color could not be obtained, this was solved by adding small quantities of primary light to the reference light in order to obtain perceptive equivalence, explaining the values below zero of the color matching functions. The RGB functions are, however, only estimates derived from psychophysical experiments. Trichromacy in humans is attributed to the physiological characteristic of the photoreceptors in the eye, namely the three types of cones in the retina
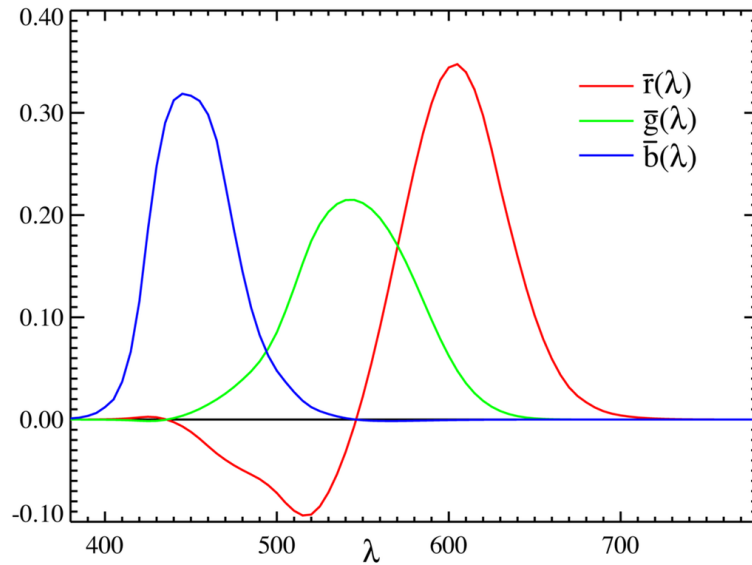
Figure 2.10: The CIE RGB color matching functions $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$. Horizontal axis is wavelength in nm, the vertical axis is tristimulus value. Public domain image taken from `https://commons.wikimedia.org/wiki/File:CIE1931_RGBCMF.png`

(L, M and S) and their fundamental spectral sensitivities, $\bar{l}(\lambda)$, $\bar{m}(\lambda)$ and $\bar{s}(\lambda)$. The three cone submosaics with their different sensitivities perform a sampling of the retinal image and perform the first step in the neural coding of color starting from spectral information. Stockman and Sharpe in 2000 [27] derived spectral sensitivities of L and M cones from photoreceptor excitation trying to establish the three primary lights that would uniquely excite each class of cones to produce physiological photoreceptor spectral sensitivities. In Figure 2.11 are reported the normalized LMS spectral sensitivities functions, with peaks of maximum absorption in the 440, 545 and 565 nm frequency of the visible spectrum for respectively the S, M and L cones.

Using these sensitivity functions, every spectral power distribution (SPD) can be reconducted to a set of three tristimulus values, and every perceived color can be identified by said values. Since the operation to obtain these values is an integration on the frequencies of the spectrum, it might happen that different SPDs match into the same set of tristimulus values. Physiologically, this is explained by the cone opsins responding only to the rate of
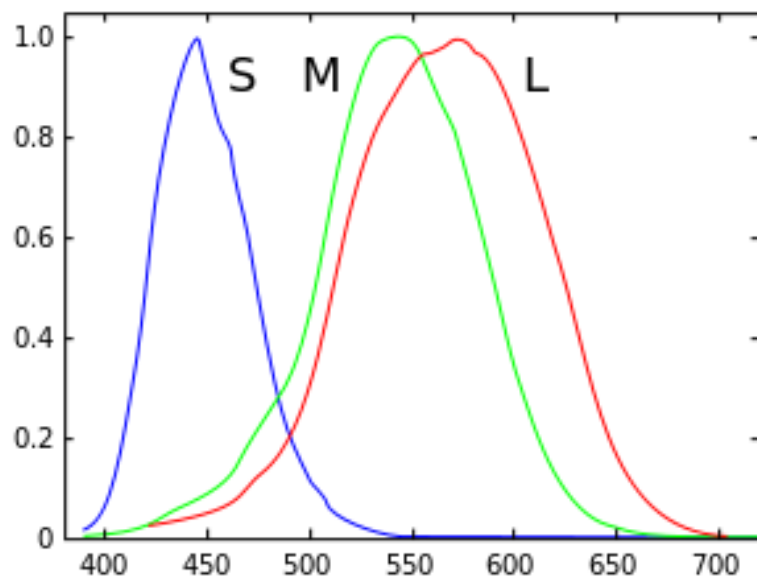
Figure 2.11: The relative spectral sensitivity functions for the (red) L, (green) M, and (blue) S retinal cone cells. The horizontal axis is wavelength in nm, the vertical axis is relative sensitivity. Image courtesy of Stockman et al. [28].

captured photons rather than the actual wavelength, so that different lights with different SPD appear identical if they produce the same absorption in photoreceptors. This occurence is called metamerism and evidences that there is no biunivocal correspondence between SPD and tristimulus values, meaning that while it is possible to get tristimulus values from and SPD, it's impossible to reconstruct the spectral distribution from the values.

In Figure 2.11 a strong overlap of the M and L functions is immediately noticeable. This translates to a very low maximum M to L signal ratio, and it is possible to observe in Figure 2.12 how sampling an image directly with the three functions results in a set of colors with poor saturation. Image courtesy of [29].

Trichromatic color vision is based on the comparison of the photopigment absorptions in cone photoreceptors, with each class of cones contributing to a particular hue of a visual stimulus through two opponency mechanisms, as explained in the quantitative model of Hurvich and Jameson [30], based on Hering's theory of color vision [31]. The color-opponent process theory states that there are certain combinations of colors that our visual system
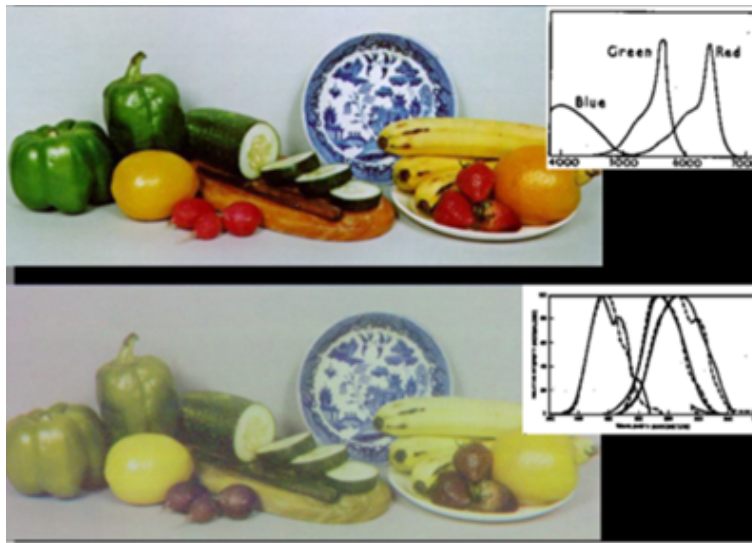
Figure 2.12: The same image acquired with camera film (top) and sampled with the three sensitivity functions of Figure 2.11. Image courtesy of [29].

is unable to perceive together, modeled into three opponent channels: red and green, blue and yellow, black and white. This process happens when signals from cones are combined in the subsequent layers of the retina in the bipolar and ganglion cells. One example of this theory is the phenomena of complementary-color afterimage, happening when photoreceptors adapt to an overstimulation and start to lose sensitivity due to opsin bleaching: when staring at a red patch for a period of time and then immediately shifting our gaze to a white surface, an equally shaped green patch can be observed on that surface.

2.7.1 Color matching functions

In 1931, the CIE and his members decided to render unique the color matching functions previously introduced in Figure 2.11, in order to create a lighting standard that could have been used to univocate every description of color. Those color matching functions (CMF), in fact, define a coordinate system for tristimuls value that can describe color starting from its spectral power distribution.

This coordinate system is called XYZ and is based on the three CMFs $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$, visible in Figure 2.13 and inspired by color matching
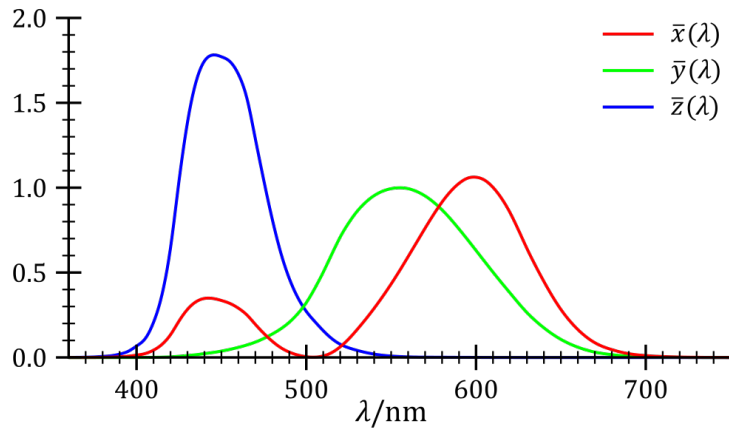
Figure 2.13: The XYZ standard color-matching functions. Horizontal axis is wavelength in nm, vertical axis is tristimulis value. Image taken from https://it.m.wikipedia.org/wiki/File:CIE_1931_XYZ_Color_Matching_Functions.svg

experiments but chosen with some particular constraints in mind:

- $\bar{y}(\lambda)$ is an approximation of the scotopic luminosity function $V'(\lambda)$, corresponding to the perceived brightness of monochromatic stimuli of the same size and duration.

- All the three CMFs have the same area bounded by their graph, and it equals to 1.

- The CMFs do not have any negative value.

- $\bar{z}(\lambda)$ has value equal to zero for most of the visible spectrum of light.

This standardized CIE CMFs are not without issues. The lack of negative values is in order to simplify the design of instruments for measuring measuring tristimulus coordinates, but this also constitutes a disadvantage because there is actually no combination of primary lights that yield the CMFs with direct measurement, as they would have to possess negative energy value in some wavelengths. In 1951, Judd [32] improved upon the $V'(\lambda)$ function by increasing the sensitivity at shorter wavelengths, creating the Judd modified CIE $V'(\lambda)$. This adjustment still retains a flaw of assuming a macular pigment density too high for a 2° visual angle. The Judd modified function was subsequently improved by Vos in 1978 [33] to produce

the Judd-Vos modified CIE $V'(\lambda)$, but even if said function is supposed to be more accurate than the standard CIE function, they have never been officially adopted and are only used in scientific fields.

The three CMFs are used to compute the XYZ tristimulus values for any given spectral power distribution $P(\lambda)$, in the case of an emissive source, by calculating their integrals over the visible spectrum:

$$X = \int_\lambda P(\lambda)\,\bar{x}(\lambda)\,d\lambda \tag{2.1}$$

$$Y = \int_\lambda P(\lambda)\,\bar{y}(\lambda)\,d\lambda \tag{2.2}$$

$$Z = \int_\lambda P(\lambda)\,\bar{z}(\lambda)\,d\lambda \tag{2.3}$$

In the case of a reflective or transmissive sample, where the sample does not emit light, instead of an SPD there will be a spectral reflectance $S(\lambda)$ that must be multiplied by the SPD of a reference illuminant $I(\lambda)$, since the appearance of the color sample is influenced by how it is illuminated. There are infinitely possible illuminants and it is impossible to separate the illuminant from the illuminated surface, hence a reference illuminant is used.

$$X = \frac{1}{N} \int_\lambda S(\lambda)\,I(\lambda)\,\bar{x}(\lambda)\,d\lambda \tag{2.4}$$

$$Y = \frac{1}{N} \int_\lambda S(\lambda)\,I(\lambda)\,\bar{y}(\lambda)\,d\lambda \tag{2.5}$$

$$Z = \frac{1}{N} \int_\lambda S(\lambda)\,I(\lambda)\,\bar{z}(\lambda)\,d\lambda \tag{2.6}$$

$$N = \int_\lambda I(\lambda)\,\bar{y}(\lambda)\,d\lambda \tag{2.7}$$

Since the functions that are supposed to be in these integrals come from measurements or empirical tests, they are never represented as mathematical equations, but by discrete samples with constant spacing, ranging from 1 to 20 nm usually. In order to calculate the tristimulus with discrete samples, instead of integrals are used summations:

$$X = \frac{1}{N} \sum S_i I_i \bar{x}_i \Delta\lambda \qquad (2.8)$$

$$Y = \frac{1}{N} \sum S_i I_i \bar{y}_i \Delta\lambda \qquad (2.9)$$

$$Z = \frac{1}{N} \sum S_i I_i \bar{z}_i \Delta\lambda \qquad (2.10)$$

$$N = \sum I_i \bar{y}_i \Delta\lambda \qquad (2.11)$$

Both spacings of the discrete samples and the standard observer functions must match in order to perform the summations. The three tristimulus values are then used to obtain a chromaticity value defined as the intersection between the three XYZ coordinates and the plane $X + Y + Z = 1$. The chromaticity is in fact described by the parameters x and y, while Y is the luminance of a color:

$$x = \frac{X}{X + Y + Z} \qquad (2.12)$$

$$y = \frac{Y}{X + Y + Z} \qquad (2.13)$$

$$Y = Y \qquad (2.14)$$

The x,y coordinates are then used to find the chromaticity on the CIE 1931 color space chromaticity diagram, as shown in Figure 2.14, representing how the human eye perceives light of a given spectrum. On the curved boundary of the diagram lie the colors obtained from a single wavelength, also defined as spectral locus.

2.7.2   RGB color space

The RGB color space is an additive color model based on the three base components *Red, Green* and *Blue*. In an additive color space, each component constitutes a source of light energy and the final perception of color is determined by the sum of the three components. This is the principle at the basis of the emission of a display monitor or a projector, where the final value of each pixel is determined by the amount of the three red, green and blue emissions in the pixel region. The opposite principle is instead applied to subtractive color spaces, like the CYMK (*Cyan, Magenta, Yel-*
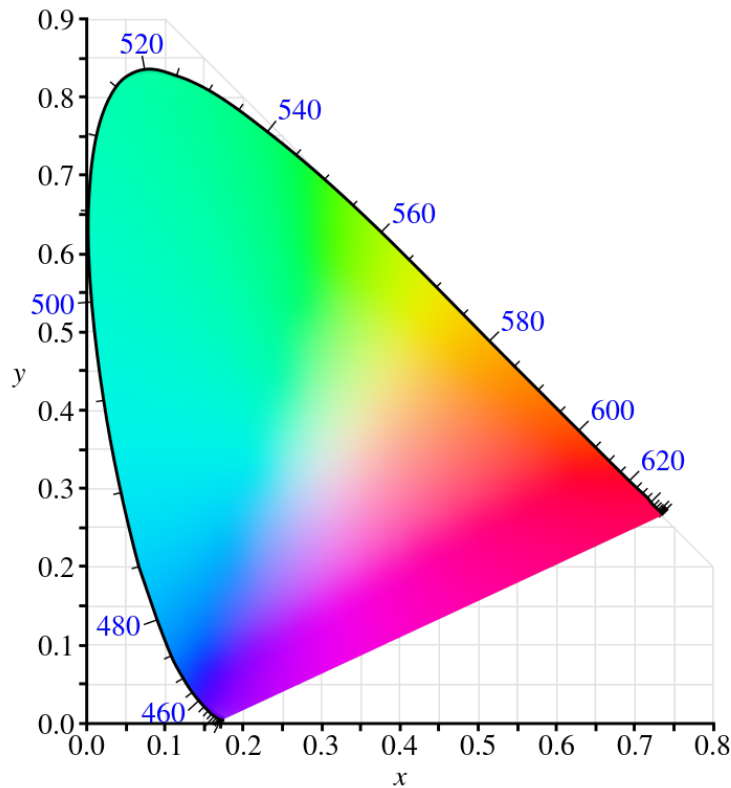
Figure 2.14: The CIE 1931 color space chromaticity diagram, on the x,y subspace. The diagram is the representation of the plane obtained at tristimulus values $X + Y + Z = 1$. Public domain image taken from `https://commons.wikimedia.org/wiki/File:CIE1931xy_blank.svg`.

low, Black) which is used in printing and where each component subtracts reflective energy from the surface where it is applied.

Colorimetrically, an RGB color space can be defined as all colors obtainable from the combination of three primary colors with different intensities. On the CIE diagram, this is equal to defining three chromaticity coordinates and, thanks to the second Grassman's law stating that a point corresponding to an obtainable color from the mixture of two other colors of adequate intensities in a diagram lies on the line between these two colors [34], every combination of obtainable colors can be found inside a triangle with the three primary colors as vertexes.

For example, Figure 2.15 shows the triangular sRGB color space with the three vertexes corresponding to the three primaries Red, Green and Blue. It
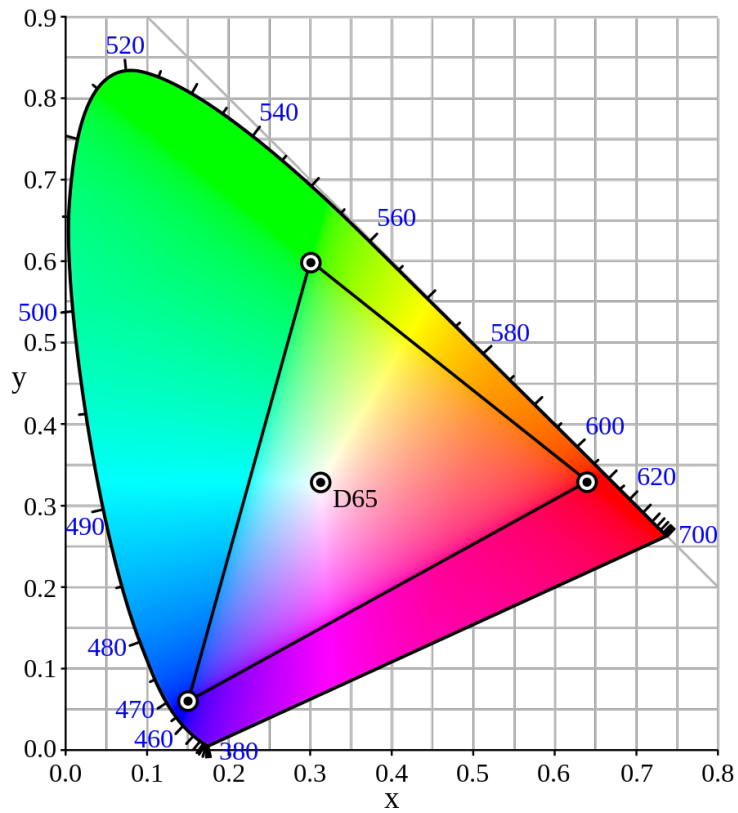
Figure 2.15: The triangular sRGB color space displayed as a subset of the CIE chromaticity diagram. Image taken from `https://commons.wikimedia.org/wiki/File:CIExy1931_sRGB.svg`.

is apparent that every RGB color space is a subset of the full chromaticity diagram. In order to convert an XYZ color into the appropriate RGB values, it must be multiplied by the reverse of the transformation matrix $[M]$:

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = [M]^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.15}$$

where $[M]$ is standardized for each desired RGB standard and is dependent by the reference white of the color space. For example, the reference matrix for sRGB is

$$M = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \qquad (2.16)$$

The obtained $v \in [r, g, b]$ values are not the final result since they are linear. To obtain the final nonlinear $V \in [R, G, B]$ values they must in fact be subjected to gamma correction with the following formula

$$V = v^{\frac{1}{\gamma}} = \begin{cases} 12.92v & if\ v \leq 0.0031308 \\ 1.055v^{\frac{1}{2.4}} - 0.055 & otherwise \end{cases} \qquad (2.17)$$

2.8   Color Vision Deficiency

Color vision deficiency (CVD), also referred to as color blindness or daltonism, is a condition of altered color perception, or inability to discriminate color [35]. While CVD can be acquired after brain or retinal damage or due to diseases that affect the eye, the leading cause of CVD is genetic inheritance, due to an alteration in the X chromosome that regulates the encoding of information on the opsin molecules of the photoreceptors, leading to anomalous trichromacy, dichromacy or monochromacy.

Anomalous trichromacy is associated to a deviation or abnormality in the peak sensitivity in one of the three classes of cones photopigments. Whether (L) long, (M) medium or (S) short-wavelength sensitive cones are affected, the three types of anomalous trichromacy are called respectively *protanomaly*, *deuteranomaly* and *tritanomaly*, and are characterized by a shifted maximum absorption of wavelenght, so that M and L absorption curves of an anomalous dichromat are more similar than in normal vision. Dichromacy is characterized by the total absence of one of the three classes of cones, namely *protanopia* for missing L cones, *deuteranopia* for M and *tritanopia* for S. Dichromats are unable to discriminate hues that appear different to a normal-vision individual, like a green from a red from a yellow, given that they only perceive the entirety of color from only two cone classes. Tritan phenotypes are also associated with a progressive S-cone dystrophy and accompanied by a disruption in the regularity of the cone mosaic, caused by the loss of S cones [36]. Monochromacy is a more severe

form of CVD resulting in a loss of two or all three cone pigments and is by
far the rarest form of CVD. In the case of a loss of function of all the three
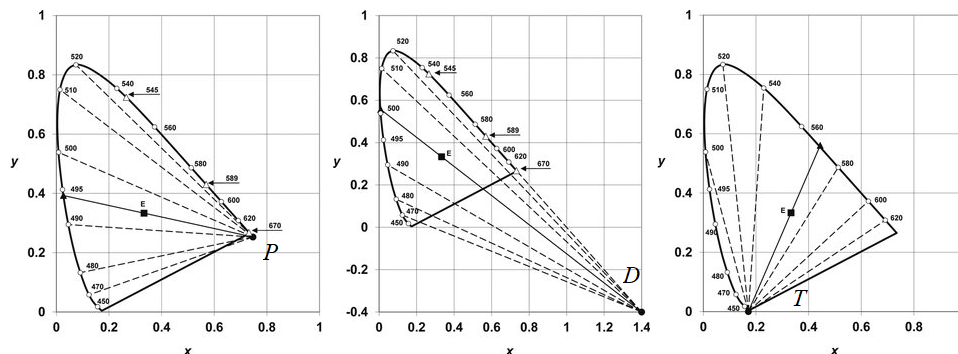classes of cones we refer to it as *achromatopsia*.

Figure 2.16: Confusion lines in the CIE diagram for protanope (P), deutera-
nope (D) and tritanope (T). All points from the same confusion line appear
the same to a dichromat. P, D and T points are the confusion points for the
three dichromacies, according to *x, y* coordinates. Image courtesy of [37].

Chromatic equivalent stimuli for a dichromat can be identified on the
CIE chromaticity diagram as per Figure 2.16. Given a stimulus, the non
discriminable colors are all the points of a line passing through the point
corresponding to the stimulus. These lines are called *confusion lines* and
they meet at a point termed *confusion point*, different for the three kinds of
dichromacy. The approximate coordinates of the three confusion points on
the CIE diagram are

$$(x_p = 0.747, y_p = 0.253), (x_d = 1.080, y_d = 0.080), (x_t = 0.171, y_t = 0)$$

but these positions are not fixed, since they are subject to a high variability
for each color deficient person.

2.8.1 Clinical tests

In order to have a diagnosis on the type and severity of color vision deficien-
cies, several clinical tests have been developed. They are usually classified
under four sub-types [38]:

1. Confusion chart tests, commonly referred to as pseudoisochromatic
   (PIC) plates. The first PIC test in commercial production was Dr.

Shinobu Ishihara's Test [39], published in 1917. It consists of 38 plates containing a circle of dots of variable size and colors, which form a number or a shape that is clearly visible for normal vision subjects, and invisible or difficult to see for subjects with forms of CVD. The PIC test actually used for screening and diagnosis in aviation and military is the American-Optical Hardy-Rand-Ritter (AO-HRR), first published in 1954 [40], and currently published by Richmond Products [41]. A PIC test especially designed for diagnosis and evaluation of blue-yellow dyschromatopsia is the Lanthony's Tritan Album (1987), consisting of 6 plates with squares of arranged grey dots, with the dots of one of the four corners are colored to form a smaller square. Of the many other PIC tests, some examples worth of mentioning are the Dvorine [42] (1944) test, the Boström-Kugelberg [43] (1944) and the Böstrom [44] (1950).

2. Arrangement tests, where the subject has to sort colors based on hue or into groups. In the Farnsworth Munsell 100 Hue test [45] (FM100), the subject has to arrange 85 colored caps with increasing hue in the correct chromatic order. The test results are obtained by inserting into a graph the numbers on the bottom side of the caps, revealing the type of CVD and its severity. Since the FM100 takes a long time to be performed, a shorter version of the test has been developed, called FD15, and it is composed by only 15 colored caps. Other noteworthy arrangement tests are the Lanthony desaturated d15 and Adams desaturated d15, which use different chromas than FD15.

3. Matching tests, where the subject has to adjust two colors until they match. The most famous example is the Nagel Anomaloscope [46], a device invented by ophthalmologist Willibald Nagel used to identify red-green color blindness. The subject sees a disc divided in half with one side showing a yellow light (589nm) which can vary in brightness while the other side shows a mix of red (671nm) and green (546nm) light, with a variable ratio, the subject must then change the brightness of the yellow side to match the color showing on the other side of the circle. A normal vision subject can only get a match only in a determined red/green ratio, but when this ratio changes, only color blind subjects can match the yellow by changing its brightness.

4. Naming tests, where the subject has to identify one or more colors, are mostly used in occupational fields where color naming is needed, like textile or paint industries, aviation, railways, maritime. The Farnsworth Lantern was developed for use in the US Navy, it is a lantern with two vertically aligned lights showing desaturated red and green and a yellowish white, close to the protan and deutan confusion lines. The subject is asked to identify 9 pairs of these colors shown for 2 seconds each. Since this test is no longer commercially available, it has been replaced by the Stereo Optical OPTEC 900.

2.8.2 Computer tests

With the widespread diffusion of Personal Computers, computerized color vision testing was investigated [47] and it has now become commonly used in clinical practice. The main problem in computerized color vision testing is that computer displays use additive color system, each pixel composed of the sum of Red, Green and Blue lights, instead of the subtractive color system used in inks, paints and filters. Moreover, each display has its own different color gamut, brightness and contrast characteristics, so that they have to be calibrated to be used in color vision testing.

There are numerous available programs that can test for color vision defects and assess their severity, some noteworthy examples are the Rabin Cone Test [48] used by the US Air Force, the ColorDx Color Vision Test, also used in pediatry, the Colour Assessment & Diagnosis (CAD) test, used in the UK civil aviation.

2.8.3 Daltonization methods

In 1996 Atkinson [49] filed a US Patent in which a computer system can test a user for CVD and applies a compensatory color palette to the computer display. Image processing techniques used to improve images for people with CVD are called Daltonization methods, they consist in a series of color correction algorithms that modify content in order to make it accessible for CVD observers. Since then, a multitude of approaches have been taken to help color blind people, either via algorithms, applications or physical commercial products.

Daltonize [50] is a user-assisted recoloring technique for protanopes and

deuteranopes. It is based on the algorithm proposed by Brettel et al. [51], users have to provide three parameters that specify the stretching of the red-green channel and its projection into the luminance channel and the yellow-blue channel. In [52], a daltonization algorithm for protanopes is proposed, featuring an iteration technique for the selection of adaptation parameters and a color clustering method to avoid color matching between the original and daltonized image.

In the same year, Huang et al. [53] developed a recoloring method for red-green color blindness that, contrary to other previous works, aims to preserve the original colors of the base image as much as possible, producing images that appear to have a more natural look to the color blind users. This approach was also taken by Kuhn et al., [54] which presented a deterministic image-recoloring technique for dichromats based on mass-spring optimization.

In 2010 Machado and Oliveira [55] developed a real-time contrast enhancement technique for dichromats that uses Gaussian pairing and predominant component analysis. Their approach is based on the observation that whenever dichromats experience some significant loss of color contrast, most of this contrast can be recovered by working on a perceptually uniform color space, and orthographically projecting the original colors onto a plane aligned with the direction that maximizes contrast loss.

Their transformation is performed in the CIE L*a*b* color space in two steps: the direction that maximizes the loss of contrast is estimated using Gaussian pairing and the recoloring is realized by determining and correcting any sudden variation of the direction of the vector estimated to maximize the loss of contrast.

In 2012 Kotera [56] proposed a spectral-based image Daltonization algorithm for the dichromats. It extracts the visible and invisible spectra to dichromatic vision, and shifts invisible spectra to a visible spectral region, reintegrating it into the fundamental spectra of source image.

In a recent paper, Joschua Simon-Liedtke et al. [57] compiled a checklist of guidelines that should be considered when designing and evaluating a recoloring tool, hereby reported:

"Recolouring should (i) preserve the naturalness of the image in general, (ii) keep individual colour characteristics as much as possible, and (iii) sustain colour communicability consistently across the workflow. (iv) It should

state the colour tasks it intends to improve, and (v) the target images it uses. (vi) It should be customizable for individual CVD types and severities, and (vii) account for colour management across different media devices. (viii) Finally, it must be tested on different types of images (ix) by real observers with CVDs."

A novel approach is also taken by [58], proposing a strategy of daltonization that enhances chromatic edges and contrast to highlight areas and objects colored in confusion colors, rather than shifting the color of objects and areas altogether. This is achieved by computing an edge map obtained from the gradient of the error image between the original and its simulation.

Due to their nature, mobile phone apps can be used to help people with CVDs in various ways. There are countless apps, designed with color blind people in mind, which offer various tools to help with daily tasks, some of them also implementing Daltonization algorithms. The tools usually implemented by mobile phone apps have different purposes, they can help identify a color from an image or a live picture from the camera, adjust or shift colors to make them easily recognizable by people with color blindness, find colors on a picture that match a chosen color and highlight them, or help in finding harmonizing colors.

In the last few years, wearable devices have started to be developed in order to assist people with abnormal color vision. The EnChroma glasses, for example, use a series of optical filters in their lenses to modify chromatic and luminous aspects of the color appearance of light to human vision [59]. By acting like a multi-band filter that cuts out specific wavelengths of light, they increase red-green color discrimination for protanomalous and deuteranomalous observers. In 2014 Tanuwidjaja et al. [60] developed Chroma, an augmented-reality wearable support for color blindness implemented on the Google Glass device. It can operate on four distinct modes: highlighting a range of colors, comparing and contrasting two different colors, applying a Daltonization algorithm and outlining areas strongly affected by the person's color blindness.

Chapter 3

State of the art

3.1 Foveated Imaging Systems with Space-variant Resolution

Space-variant images, defined as images whose resolution changes across the image, have been studied for about three decades and have surpassed conventional approaches in robotics applications to lower the amount of processing required, especially when real-time constraints make it necessary to utilize resource-economic image representations and processing methodologies. They are especially used in studying visual attention, target tracking, motion estimation of camera systems, and 3D perception.

The reasoning beyond the use of space-variant images is that in most cases, execution time and computer memory space must be spent in the first place to eliminate the redundant information in acquired images, in order to extract the features required by the execution of an assigned task. If this reduction in information was to be performed at the lowest level of a visual process, that is, at the sensor level, the efficiency of the system in terms of execution time and later computation, would be greatly improved. On the other hand, the information filtered out at the sensor level is no longer available for further computation.

This approach has both biologic and computational complexity evidence. Biologically, we see evidence of the processing in the human visual system. Each region in the retina can be directly mapped to a specific portion of visual cortex where it is processed. The fovea is the region of highest visual acuity, and even though it only covers only a 1% region of the retina, it

maps to over 50% of the visual cortex [10]. Therefore, our brain would have to be significantly larger if our retina was covered with photoreceptors at the same density as it is in the fovea.

Computational complexity analyzes for foveal vision systems has been performed by Bandera [61], presenting a first order analysis where he uses a complexity measure of the amount of data captured multiplied by the computation time. Using this metric, a Cartesian array of NxN pixels has an $O(n^4)$ complexity while an exponential foveal array has a complexity of $O((\log_2 n)^3)$.

The most common space-variant image structures is Log-Polar mapping: Log-polar mapping is a geometrical image transformation that attempts to emulate the topological reorganization of visual information from the retina to the visual cortex of primates. It defines a pixel to be an arc slice between two radii while the spacing of radii distances follows a logarithmic relationship. The log-polar mapping creates high resolution pixels at the center of the frame and low-resolution pixels on the periphery.
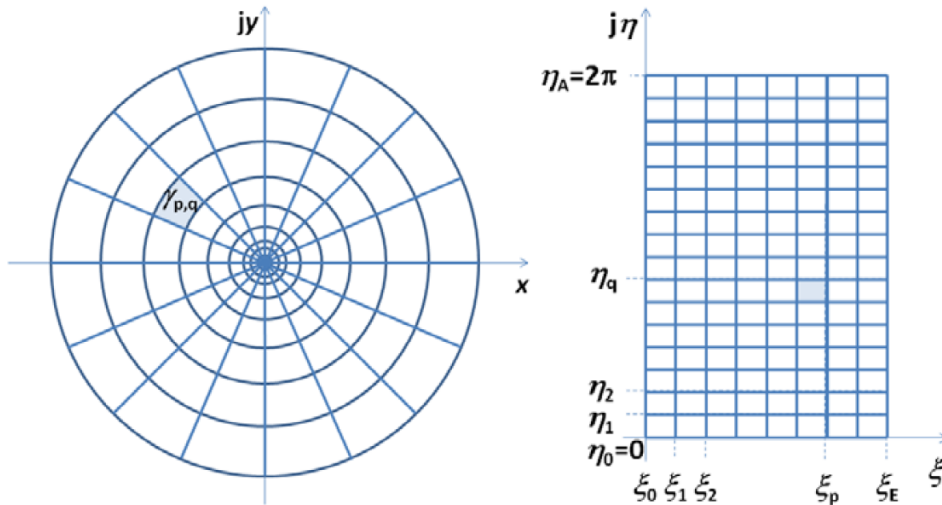
Figure 3.1: Log-polar transformation. The retinal plane (left) is mapped onto the cortical plane (right) by $\omega = log\,(z)$, where concentric circumferences and radial lines become straight. In the rectangular grid, each position corresponds to a section of a concentric annulus in the retinal plane. Image courtesy of [62].

It can also be found in the literature under different names, such as

log-polar transformation or the $log(z)$ model. The reason for this last denomination comes from the fact that the mapping can be mathematically modeled by the complex logarithmic function log(z), where z is the complex variable representing points on the image plane.

Considering the complex retinal and cortical (log-polar) planes, represented by the variables z = x + jy and ω = ξ + jη, respectively (j is the complex imaginary unit). The complex log-polar mapping is:

$$\omega = log\left(z\right)$$

and the log-polar coordinates ξ (eccentricity) and η (angle) are given by:

$$\xi = log\left(z\right) = log\left(\sqrt{x^2+y^2}\right)$$

$$\eta = arg\left(z\right) = atan2\left(x, y\right)$$

where atan2(y, x) denotes the two-argument arctangent function that considers the sign of x and y in order to determine the quadrant of the resulting angle. This mapping transforms concentric circumferences and radial lines in the retinal plane into straight lines along the ξ and η directions in the cortical plane, respectively (Figure3.1).

Properties and practical implication of this mapping are:

- Conformal mapping: the space-variant image preserves oriented angles between curves and neighborhood relationships. While this theory predicts that Cartesian image processing operations can be applied to log-polar images, in practical applications specific algorithms are required in most cases.

- An elegant trade-off solution, between the criteria of wide field of view, high visual resolution and amount of data. Especially in robotics, the lesser size of log-polar images facilitates visual data processing, and the highest resolution at the center (foveal predominance) implies that targets can be tracked without segmentation from the background, while incidentally providing an easier segmentation process.

- Biological plausibility: since it approximates the receptive field distribution and retinal-cortical mapping in the visual system of mammals, it can be seen as a support tool to neurophysiology, or an inspiration

from the natural world for engineering strategies.

- Rotation and scaling invariance: when rotated or scaled, patterns in the log-polar image only undergo translations, preserving their shape. This is helpful for rotation- and scale-invariant pattern recognition and motion estimation in active tracking scenarios. Still, in contrast to traditional square arrays, this representation is not location-invariant.

There are different ways to obtain log-polar images either from conventional images or directly from a scene, using software and/or hardware-based solutions. One of the visual processes where log-polar imaging is most suitable is probably active target tracking, some advantages have also been found in estimating the observer's motion using log-polar images, due to its polar geometric nature which fits particularly well with time-to-collision computation and other navigation tasks in mobile robots. Binocular depth estimation has also been considered with a joint usage of log-polar imaging and active vergence movements.

A different approach adopted by Geisler and Perry [63] is the use of a space variant pyramid, by which an image is decomposed into a pyramid of 2D arrays of coefficients representing different spatial frequency bands. The first level of the pyramid contains the greatest number of coefficients and the highest spatial frequency band. Each successive level of the pyramid contains one fourth the number of coefficients of the previous level, and encodes the band of spatial frequencies centered at one half of the center spatial frequency of the previous level. Geisler's approach was used by Alonso et al. [64] for a real-time foveation technique by eliminating the perceived image intelligibility for image obfuscation purposes, while simultaneously preserving privacy in face deidentification.

3.1.1 Case Studies

The first mathematical descriptions and computational implementations of space variant sampling are due to Sandini and Tagliasco [65], which created a model based on a discrete distribution of elements whose radii increase linearly with eccentricity in the visual field, that was supposed to be implemented in chip fabrication. In their paper, they described a special type of mapping, called "retino-cortical", characterizing the human representation

of visual stimuli in the visual cortex and they proved that a logarithmic-polar law applied to retinal coordinates can be used to mimic this biological transformation. They modeled two distinct regions: the fovea and the periphery. The fovea uses a high-resolution, uniformly sampled rectangle in the center of the field of view, while the periphery uses the log-polar mapping to cover the visual field with circular elements with low overlapping, simulating the density of retinal cells. To minimize the overlapping, they made it so that the intersections of the circular elements' contours counted as vertices of hexagons. The linear relationship between radius R and eccentricity E was expressed, in terms of number N of equal size elements at a given eccentricity, by the following relation:

$$R = \frac{2\pi}{3N} E$$

This structure obtained (example in Figure 3.2) is algorithm-independent, so that any local computation can be applied, like spatial filtering or edge detection, without compromising the spatial compression of visual information. From a computational point of view, this process is equivalent to removing an increasing amount of high-frequency components from the center to the periphery, but introducing noise artifacts represented by the sharp discontinuities between the circles.

The simulation was carried out by computing the average of the light intensity distribution over the single areas. One of their main findings is that cross correlation procedures applied to space variant descriptors present clear and deep peaks, simplifying the research of corresponding points in different images.

In their work, Robert-Inacio and Yushchenko [66] use an hexagonal cell model to achieve eye saccade mimicking because of the similarities in shape between hexagonal cells and retinal cells. Such a sampling leads to setting up foveated image processing with good results in terms of data preservation, utilizing 15 times less data and performing in a really fast way. The foveated images are used to compute a sequence of points of interest: by following this sequence, a system of vision can mimick eye saccades when focusing successively at each point of interest. This process is used for modeling visual attention, as it takes into account detection of points of interest.

Tistarelli and Grosso [67] investigated the use of space-variant sensing
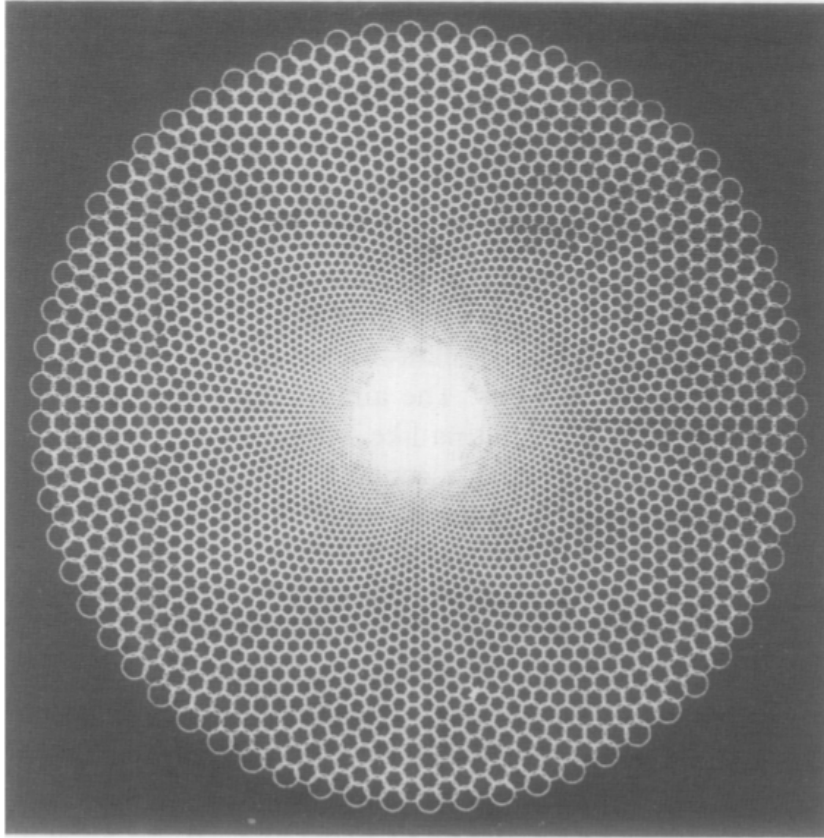
Figure 3.2: Structure of the model proposed by Sandini and Tagliasco. Example with 7000 elements obtained with parameter $N = 64$. Image courtesy of [65].

in face recognition. In particular, they developed a hybrid approach, coding each face image by few space-variant descriptors and then applying principal component analysis for data reduction. A framework for practical use of space variant descriptors on large-scale applications is proposed. The reported experimental results show that the prior extraction and match of salient points allows to boost correlation procedures commonly used to compare space variant descriptors.

Rojer and Schwartz [68] use the continuous log-polar transformation for a space-variant sensor design based on the conformal mapping of the half disk, $\omega = \log(z+a)$, with a > 0 to displace the singularity at the center of the field of view, which characterizes the anatomical structure of the non-human primate and human visual systems. This model allows the designer

to select the desired field of view and the number of pixels in the outermost periphery. Using this model, they show how to duplicate the field of view and resolution found in the human visual system.

Yamamoto [69] presented an implementation for a foveated robot vision system called Fovia (FOVeated Image Application). They used log-polar mapping to simulate both the photoreceptor sampling and their receptive fields. The sampling function defines two distinct regions: a foveal, uniformly sampled region and a peripheral region where receptive field size increases as a linear function of eccentricity. They demonstrate the modularity of the system and its potential as a testbed for active vision by incorporating two different attentional mechanisms and quantitatively evaluating their performance on artificial and natural images, as shown in Figure 3.3.

Figure 3.3: Simulation of a gaze movement pattern with an active foveated vision system, exploring a grayscale image of Einstein. Image courtesy of [69].

Jurie [70] improved log-polar mapping by specifying the range of the logarithmic function used for the mapping and by introducing algorithms to perform basic operations on the space-variant images, allowing real-time applications like face tracking with low-cost hardware. The grey level values of log-polar pixels are computed using the log-map and rectangular pixels gray level values; the originality of the encoding is that log-polar pixels are not obtained by aggregating rectangular pixels, but are a fractional part of rectangular pixels, producing more regular images.

Concluding, over the last few decades, fundamental properties of space-variant image representation have been studied and exploited in appropriate algorithms. Log-polar sampling provides interesting benefits in several scenarios. One criticism that can be posed is that, while it is clear that

space variant descriptors can characterize important cognitive processes like recognition, the practical adoption of space variant approaches suffers from computational problems and turns out to be much less appealing than methods based on regular image sampling. A good reason why foveated vision deserves study and investigation lies in its biological inspiration: it can be used to broaden our still limited understanding of human vision and the related brain processes. It is worth noting that despite the biological approach, the works here discussed that adopt log-polar mapping do not take into account the displacement of post-receptoral cells, mainly the ganglion cells, when mapping the visual world onto the brain.

3.2   Modeling space variance in the retina

Log-polar mapping is practically used as an approximation of the retinal mapping, but the way photoreceptors are spatially arranged in their own retinal layer is fundamentally different and regulated by other principles. The basic idea is that differences in the composition of retinal mosaics are resembling of the differences between fingerprints: there are no retina or fingerprint equal between individuals, yet they can be regarded as a sample from the same constrained random process. These characteristics determine the limits of the sampling properties that are observable across all human retinas. This section reports known studies concerning the space variance of the cones sampling structure.

Watson [71], in modeling human spatial vision, made a first approximation by observing that spatial processing is identical across the visual field but changes according to a bi-dimensional scaling factor s given by the function

$$s = 1 + ke$$

where k is a constant estimated at 0.4 and e is eccentricity in degrees. This function determines that a sensor at the center of the visual field, the fovea, has frequency f, width w and density d; while as the eccentricity e increases, density and frequency rapidly drop becoming d/s and f/s and width increases by ws, as illustrated in Figure 3.4.

In order to produce a cone sampling array, Ahumada [72] started from non-human primate data to create a model that generates a list of coor-
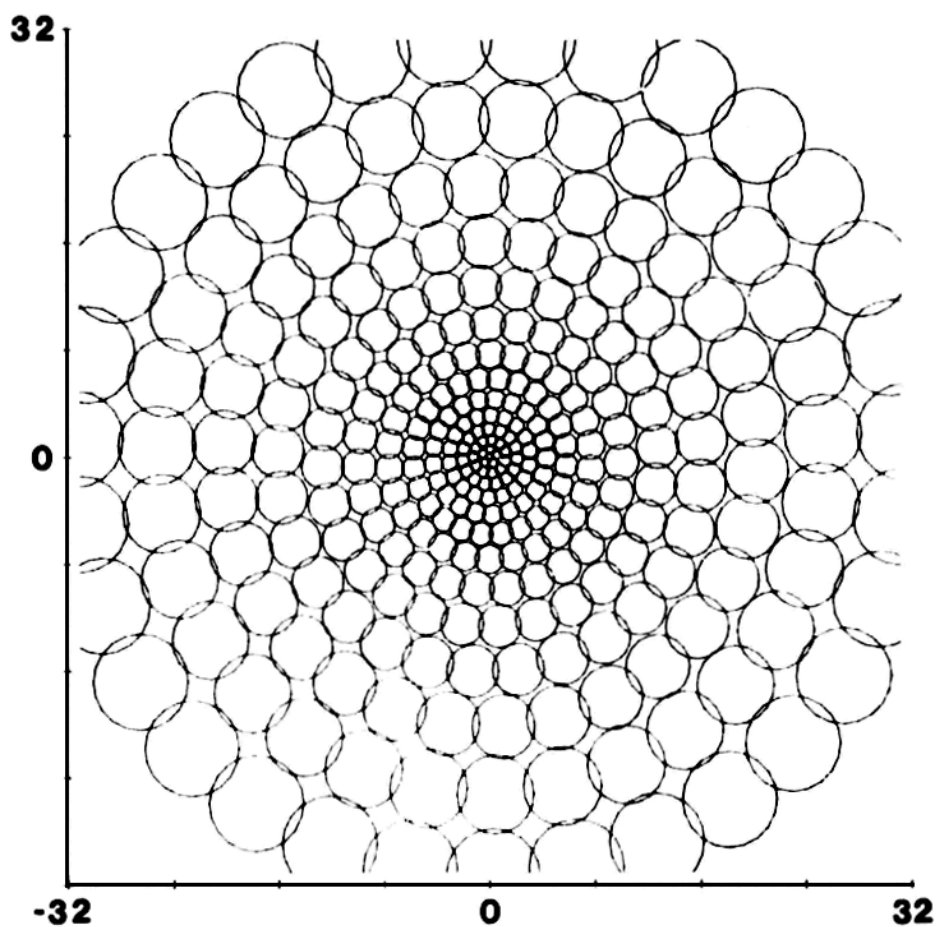
Figure 3.4: Sampling structure for a base frequency of 1 cycle/degree. Axes are in visual degrees from the fovea. Image courtesy of [71]

dinates corresponding to the center of cones. Each point is considered as the center of a disk region of variable radius selected from a uniform random distribution of radii, the circular regions are then distributed starting from a seed of two adjacent discs, with the restriction that discs can't overlap. Then, iteratively, discs are added to the array until the packing is complete, and ultimately a small normally distributed jitter is computed on the coordinates of the central points, to allow the distribution to become a jittered-lattice model. The authors attempted at generating a space-varying parametric model to extend the modeling capabilities beyond the foveola, by variation of the mean and standard deviation of the radius of the cone disc and standard deviation of the postpacking jitter in function of the distance

49

from the center of the retina, but they found out that past the central fovea the model did not account for additional disorder processes.

Curcio and Sloan continued to explore the possibilities of modeling the human photoreceptor topography by proposing a model of cones distribution based on a regular array of points subjected to spatial compression and a jitter, validating them by fitting with actual cone mosaics [73]. Their performed analysis was based on the distribution of distance and angles of neighboring cones, with a comparison of real cone mosaics with artificially generated ones, demonstrating the existence of local anisotropies in cone cells spacing.

Wang's attempt to model the sampling properties of the cone mosaic [74] consisted in a polar arranged array of cones subjected to a jitter according to the standard deviation of a Gaussian normal distribution, with the additional constraint of a minimal spacing rule, visible in Figure 3.5. By comparing the power spectrum of distributions of human foveal cones and the generated sampling arrays, similarities are presented, as well as exhibiting basic features of real foveal cone mosaics.
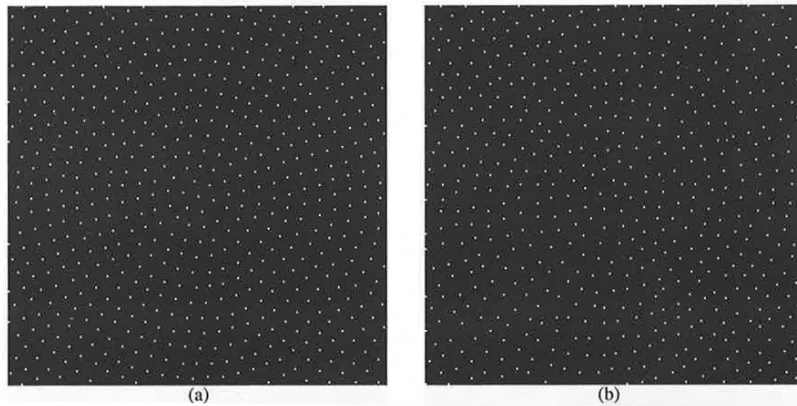
Figure 3.5: Sampling arrays generated by the polar-arranged, jittered-lattice model of Wang et al. Image courtesy of [74]

In the human eye model of Deering [4], the cone mosaic is modeled from an algorithm that generates cells starting from a seed and then surrounding them with points defining a polygon, constituting the cell boundary, as observable in Figure 3.6. With each iteration, new cells are added adjacent to the seed or previously generated cells and are subjected to both attractive

and repulsing forces to adjust their position, modifying their boundaries. This cone synthesizer is validated by calculation of the neighbor fraction ratio and by matching statistics of density in cones/mm^2 with values in literature.
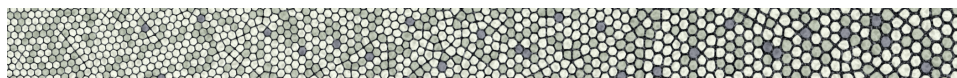
Figure 3.6: Structure of first 1.3 visual degrees of synthesized retinal cones by Deering. Image courtesy of [4].

Another method to generate space variant cone mosaics has been proposed by Bradley et al. in 2014 [75], using heuristics to position cones along isodensity contours around the fovea, making sure that cones are no closer than the spacing implied by a density model.

ISETBIO [8] is an image-computable method combined with an inference engine designed for exploring the properties of the earliest stages of biological visual systems, estimating the effects of human optics, eye movements, cone absorptions and photocurrent, and retinal cell properties. The proposed method for generating cone mosaics aims at retaining the quasi-hexagonal cone packing of the fovea while decresing the mosaic density with eccentricity, subjecting each cone and its neighbors to mutually repulsive forces.

3.3   Models of the Retina

The complex organization of the retina is constituted by numerous biological components with different characteristics, and their behavior is regulated by as many parameters. There are several studies concerning the characteristics of the human retina and how they relate to the sensation of color, so modeling the retina and its properties is a task undertook by a large number of vision scientists. This section does not intend to be an exhaustive survey but a starting point to consider some of the models developed in this domain and their different fields and purposes.

Models of the retina have the goal to pre-process an input signal into an accurate representation of a natural scene. Therefore, the majority of them are based on the concept of ecological vision [76], since the pre-processing of

the retina acts like an optimization of the information representation derived from a visual scene. Alleysson and Guyader [77] propose a distinction of retinal models based on their modeled elements and application. Biological models are based on Hodgkin and Huxley model of neural transmission [78], in which the transmission of synapses between neurons is modeled. A more general system approach of neural modeling is based on geometric spaces of neural activities, where the behavior of the neural model is determined using geometric rules. Such models are also known as Neurogeometry models, after the work of Petitot [79]. Lastly, there are information theory models, that consider the retina as an information transmission system. In table 3.1 a summary of the considered models is presented.

Paper	Goal	Modeled element	Results
1993, Beaudot et al. [80]	To provide an efficient implementation on a conventional computer architecture of some early visual processing that occurs in the eye.	Non-homogeneous filtering by the crystalline lens, photoreceptor coupling, chromatic sampling and digital filtering with an irregular spatial sampling.	Provides the fundamental equations of the retina functions to model the retina circuitry through a spatio-temporal filter.
2005, Hans van Hateren [81]	To develop a model of the visual system that could function well in natural, outdoor-lighting conditions, using horizontal cell measurements.	A model for the sensitivity regulation in outer retina, consists of a nonlinearity cascaded with three feedback control loops	Cone responses are the major factor regulating sensitivity in the retina.

2006, Hans van Hateren [82]	A model that provides range compression, as well as luminance-dependent noise suppression and a global tone mapping algorithm for rendering HDR images.	A dynamical model of response characteristics of the human cones.	A cone model that describes the fast adaptation performed by the enzymatic amplifier inside the cone that transduces light into electrical signals, well-suited for rendering HDR videos.
2007, Hateren and Snippe [83]	To establish a tractable computational scheme that can be used as a preprocessing module for studying and modeling visual information processing in downstream parts of the human retina and beyond.	Include the pigment bleaching at the cone outer segment to better simulate cone response.	A model for human L- and M-cones that produces credible responses at intensities ranging from mid-mesopic levels (1 td) up to high-photopic levels.

2007, Hérault and Durette [84]	To create a model of the retina with respect to sampling, spatiotemporal filtering, color-coding and non-linearity.	Spatiotemporal filtering, Color multiplexing, Color decoding, Irregular sampling, Space-variant sampling, Photoreceptors compression, Ganglion cells compression.	A model of retinal characteristics including properties and their consequences on the processing of visual information.
2009, Wohrer and Kornprobst [5]	A large scale simulation software that transforms a video input into spike trains, allowing large scale simulations in reasonable processing times and keeping a strong biological plausibility.	A linear model of filtering in the Outer Plexiform Layer, a shunting feedback at the level of bipolar cells accounting for rapid contrast gain control, and a spike generation process modeling ganglion cells.	A software tool that transforms realistic large-scale video input into spike trains.
2010, Benoit et al. [85]	Showing the advantages of using a model in order to develop efficient and fast bio-inspired modules for low-level image processing.	Spatio-temporal filtering, frequency and orientation based analysis.	A software tool with image processing modules for contour enhancement, moving contour extraction, motion analysis and motion event detection.

| 2019, Cottaris et al. [8] | To provide a computational-observer model of the human spatial contrast-sensitivity function. | Human optics, eye movements, cone absorptions and photocurrent, retinal cell properties. | Derived contrast-sensitivity functions agree well with ones derived using traditional ideal-observer approaches, when the mosaic, optics, and inference engine are matched. |

Table 3.1: Summary of considered models of the retina.

3.4   Spatial color reconstruction

In digital imaging, demosaicing is the procedure used to reconstruct a color image starting from the monochromatic RAW image. Similarly to the human eye, when light entering in the lens of a camera reaches the sensor, it is converted into electrical pulses and then into digital values, to be stored in a RAW file. This file holds the original monochrome image data captured by the camera, referred only to the luminance image. In order to obtain a full-color image, a conversion is performed from the raw data using a Color Filter Array (CFA). The most widespread scheme among the Color Filter Arrays is the *Bayer pattern* (Figure 3.7), a 2x2 matrix, where 50% of the photodiodes capture the green (G), 25% are used for red (R) and the remaining 25% for blue (B). The additive mixture of these three components allows the reproduction of images. Each photodiode is able to capture the information related to only one color and can not provide the data for the formation of whole (intended as RGB component) pixel. To address this limitation, a demosaicing process is used, which interpolates between the pixels of the image, to generate the missing color information. In this way an RGB image can be generated to be displayed on the monitor.

The most used CFA mounted on current digital cameras is the Bayer Pattern. However, many others have been developed with the purpose of overcoming the Bayer Pattern limits. In fact, according to the characteristics
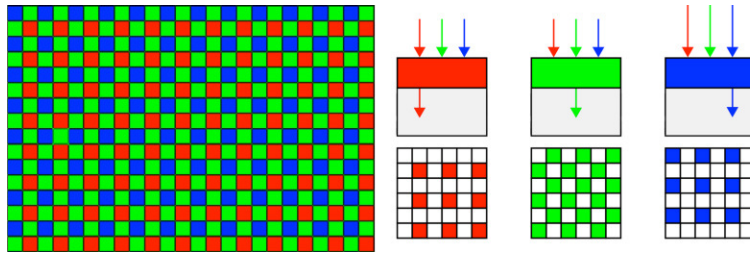
Figure 3.7: Example of CFA using the Bayer pattern.

of the input image, different pattern could reduce some artifacts. In [86] a set of different pattern are presented. We show them in figure 3.8.
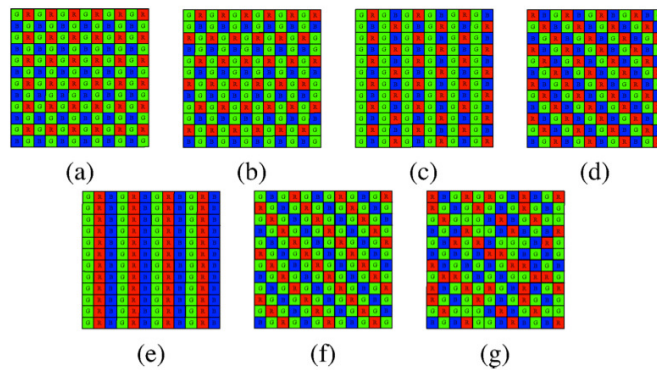
Figure 3.8: a) Bayer CFA. b) Lukac and Plataniotis CFA [87], c) Yamanaka CFA [88]. d) Diagonal stripe CFA [87] e) Vertical stripe CFA [87], f) Modified Bayer CFA [87], g) HVS-based CFA [89]

The patterns showed in Figure 3.8 are all RGB, with the same pixel size. Other solutions have been proposed in the past, or are actually subject of research. According to [90], Fujifilm designed a new pattern where the sub-components of the sensor pixels have varying size: the green is larger than the red and blue, and a white filter is added.

There are many algorithms to reconstruct the missing color information of the pixels. Some of them reconstruct the missing value interpolating according to a fixed scheme (no-adaptive algorithms), while others, more complex, interpolate according to the spatial features of the neighbor pixels (adaptive algorithms). Bilinear interpolation [91] is one of the simplest interpolation methods. It obtains the value of the missing pixel simply calculating the average of the adjacent pixels. It works separately on the three channels, without taking in account the possible correlation among

the RGB values. This method works well on regions of the same color, but it generates artifacts near the edges.

Brainard [92] proposed an approach based on Bayesian methods, starting from the responses from an interleaved array of trichromatic sensors and a model of statistical distribution in natural images, which minimizes the expected reconstruction error.

Alleysson [93] developed a model for spatial multiplexing of color, based the assumption that the human visual system has some solution to the problem of color reconstruction, since its subsampling does not affect the spatial resolution and color accuracy of the reconstructed image. He did this by separating luminance and chrominance in the Fourier domain, coding luminance entirely and subsampling chromatic information.

Chapter 4

Spatial distribution of photoreceptors

The regularity and development of the retinal cone cells mosaic have been the subject of different theories. Wassle and Riemann [94] proposed two models of mosaic generation based on mechanisms of self-regulation of an original random pattern. The first contemplated repulsive forces acting between cells to determine their final arrangement, the second was based on competition for territory for each cell with its neighbors, the same as for the treetops competing for a surface area to get sunlight from. In analyzing the retinal mosaic from a non-human primate, Yellott [95] postulated that the distribution of photoreceptors in conforms to a Poisson disc distribution. He treated cones as an array of sampling points in a two-dimensional space and performed spectral analysis on it, and observed that the spectral properties of the cones mosaic is representative of a Poisson disc point process, with the only difference of an additional minimum distance ruling between the center of a cell and its nearest neighbors, due to the cell possessing a body size, unlike a point distribution in space. Galli-Resta et al. performed further investigation in this direction, analyzing the spatial features of ground squirrel retinal mosaics [96]. Their findings suggest that the array of rods and S cones can be adequately described by a d_{min} minimal-spacing rule in conjunction with a desired density of photoreceptors. Poisson disc distributions are now regarded as one of the best sampling methods, by virtue of their blue-noise spectral characteristics [97].

4.1 Topography of the retina

The number of photoreceptors in the retina is highly variable, for the purpose of this work data from Curcio et al. [98] will be used, reporting to have an average of 92 millions of rods and 4.6 millions of cones. The *fovea centralis* is a 2 mm wide pit located in the center of the retina and is the region with the peak density of cones, amounting at 199,000 cones/mm^2. In the center of the fovea the rods are missing, in an area corresponding to roughly 1.5 to 2 visual degrees, visible in Figure 4.1. Another peculiarity of this region is that bipolar and ganglion cells are not directly in front of photoreceptors but are displaced towards the outskirts of the fovea, so that light radiation focalized by the eye's optical system is delivered directly without passing through the other retinal layers. Moreover, in the macular region of the retina, midget ganglion cells are known to form a *private-line* circuit in which ON and OFF midget bipolar cells contact a single L or M cone and in turn provide synaptic output to only a single ON or OFF midget ganglion cell partner, establishing a clear anatomical basis for an L or M cone pure receptive field center [99], so that each limb of each cone signal has a single reserved bipolar-ganglion pathway to the brain in a way that they are individually processed, effectively explaining why the fovea is the region with the highest visual acuity in the retina.

By moving only of 120 to 150 μm from the center, the density of cones halves and keeps on declining, unevenly across the eye's meridians. This decline is faster in the vertical meridian than in the horizontal, but eventually it slows down in the periphery. Since the temporal region of each eye overlaps in human stereoscopic vision, it is possible to notice that at the same eccentricities, cone density in nasal retina is 40-45% higher than in the corresponding temporal region. Toward the far periphery of the retina, cone densities levels off or slightly increases by 13 to 17% more than the corresponding lowest density along the same meridian. The graphs in Figure 4.2 report the mean cone density in function of eccentricity along meridians.

As for the rods, their density rapidly increases right after the foveal rod-free zone, with densities up to 100,000 rods/mm^2. This region, called rod ring, possess an horizontally oriented elliptic shape and is situated at 4.5 mm or 18 degrees from the foveal pit, roughly the same eccentricity of the optic disc. Along the rod ring in the superior retina lies the region with highest
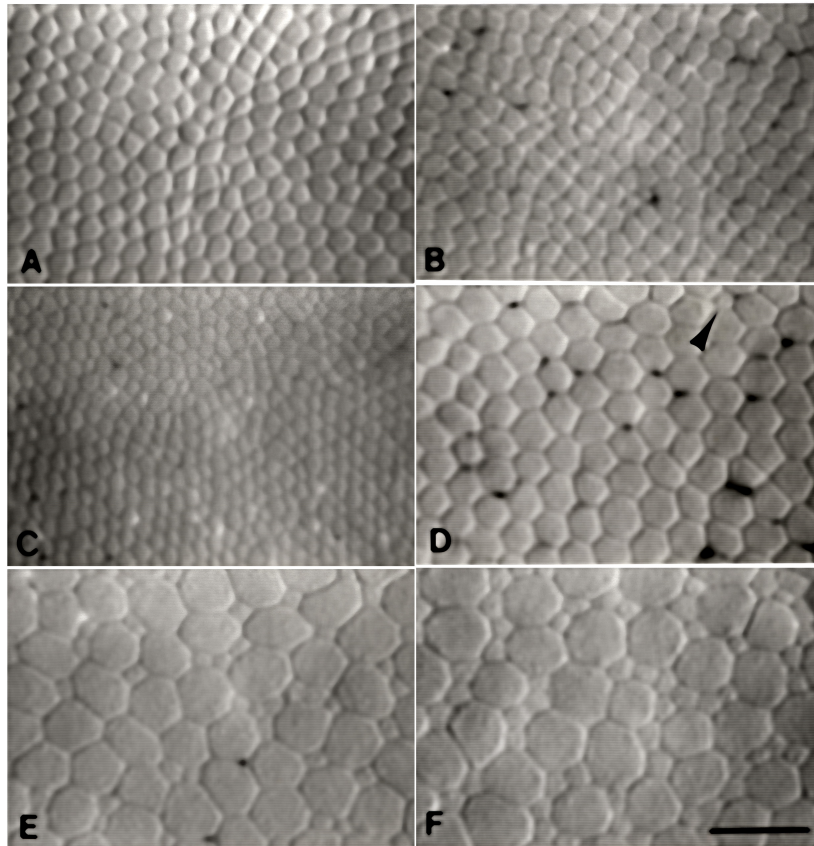
Figure 4.1: Sections of the cone mosaic inner segments. A-C are foveal centers containing only cones, to be analyzed in Chapter 5. D is the edge of the rod-free zone, with arrow pointing to a single rod, E is the point of equal rod and cone density, F is the foveal slope, where rods outnumber cones by 4:1. Image courtesy of [98].

rod density, called the hot spot, with an average of 176,000 rods/mm^2. At higher eccentricities, rod density steadily declines towards the far periphery, reaching minimum values of 30,000 rods/mm^2.

All of the photoreceptor signals ultimately convey into the optic nerve, considered part of the central nervous system and composed of ganglion cells axons and glial cells providing sustenance to the tissue. In comparison to the hundred of millions photoreceptor cells, the average human optic nerve is composed of a million fibers, so that their ratio is about 100:1. This indicates a compression in the visual signal, with ganglion cells mediating signals from a higher number of photoreceptors the higher the eccentricity,
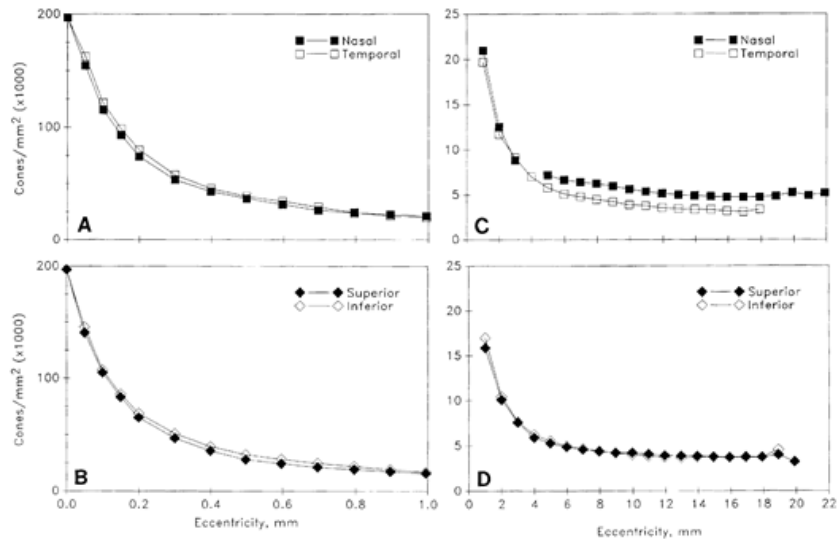
Figure 4.2: Mean of cone density at different eccentricities in the retina in the horizontal (A,C) and vertical (B,D) meridian. Image courtesy of [98].

compatible with our visual acuity.

4.2 Color sampling

The three submosaics of retinal cone cells allows to perceive different hues in a mechanism of opponency inside of the receptive field of the ganglion cells. However, spatial organization of the three classes of cones is not constant between individuals and mechanisms that determine their development and regularity are still object of research.

The S cones mosaic is independent from the L and M mosaics and it has observed to be distributed in a non-random fashion, taking the semblance of an almost hexagonal array [100]. The amount of S cones has been measured to be approximately 7% of the total cone population, and their highest concentration is found in a ring at 0.1 - 0.3 mm eccentricity. Moreover, S cones are inherently absent from the center of the fovea, the region of peak cone density, with a diameter measured to be from 100 µm [101] up to 750 µm according to Hendrickson and Youdelis [102].

Whether the remaining cones end up being M or L during the retinal development is entirely regulated by the X chromosome gene, which determines the corresponding cone opsin. While S cone cells are of relatively easy

61

characterization, there are not many differences between L and M cones, given that their pigment is 96% identical. The first to actually examine their mosaic were Hofer et al. [103] using a combination of adaptive-optics imaging, manipulating the optical wavefront to avoid aberrations, and retinal densitometry to selectively bleach the photopigments in the cones using different wavelengths. L and M cones are generally randomly distributed in the retina, in fact an evident departure from the average ratio of 2:1 L to M has been observed, both interpersonally and in the same subject in different retinal locations, as in Figure 4.3, in addition to a tendency towards clumping. Also, a protan color-vision defective retinal mosaic was estimated to have an average L:M ratio of 0.5:1.
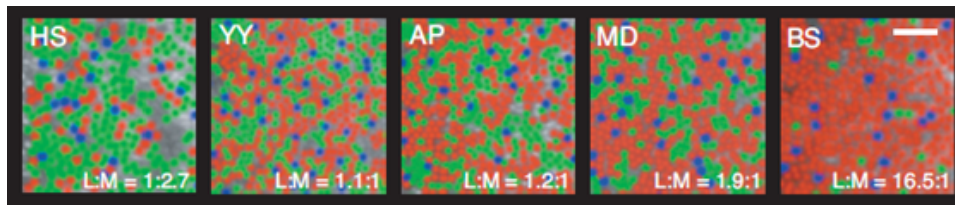
Figure 4.3: False color images of retinal mosaics of five different subjects, showing the location of L (red), M (green), and S (blue) cones in patches of retina at approximately 1-deg retinal eccentricity and the correspondent L:M cone ratio. Image courtesy of [104].

In spite of this high variance in spatial distribution of cones in different subjects, there has been no notable difference in subjective perception. A yellow wavelength has been used to evaluate interpersonal differences in perception, since it is neutral in the red-green opponent mechanism and its perception is thought to be mainly regulated by differences in the excitation of L and M cones [105]. Even in subjects with widely different L:M ratios, the wavelength perceived as yellow remains almost constant, indicating that this variability is offset by some perception mechanism probably involving cortical elaboration of L versus M cone opponency in the midget pathway, but it is however still controversial [99].

Hagstrom et al. [106] analyzed the cone pigment mRNA at different locations in the eyes of 23 male donors 5 to 96 years old, finding variations in L to M cone ratios ranging from 0.82 and 3.00 and significant differences in said ratio in function of the eccentricity, with M cone population dropping by 15% from the center to the far periphery. In their 1999 article, Roorda

and Williams [2] describe how to take in vivo images of the cones in two subjects they used a CCD sensor coupled with a Hartmann-Shack wavefront sensor to measure the aberrations of the eye's optics and compensated for it with a deformable mirror. By taking different pictures with either fully bleached, selectively bleached and dark adapted photopigments, they were able to characterize the individual cones, to find average L to M ratios of 1.15 and 3.79.

In order to investigate the perception of small spatial scale stimuli after elaboration of the cortical circuitry, Hofer et al. [104] characterized the locations of S, M and L cones in patches of retina at the same eccentricity in 5 different subjects with normal color vision, as per Figure 4.3. In order to stimulate an area corresponding to half the diameter of an individual cone they used adaptive optics to shine brief, tiny flashes of monochromatic light (500, 550 and 650nm) in selected locations of the characterized retinal regions, with the subjects reporting the hue appearance of the flash.

In Figure 4.4 are presented the experiment results for a 550 nm flash. It is possible to observe that contrarily to plausible expectations, a large number of hue categories were necessary to describe the perception of the same stimulus, including white, reportedly caused by the excitation of multiple classes of cones since it contains all colors, but also blue and purple, even for 550 nm and 600 nm flashes of light set to an intensity threshold specific for M and L cones, since S cones require a higher intensity of light to be stimulated. This suggests that contrary to the standard model of color opponency, M and L cones may have a role in the perception of the blue sensation. Moreover, two subjects reported an "indescribable" category, meaning that they perceived the flash but had no definite perceptual response.

Subjects with more L cones in their retina reported a higher number of red sensations, while subjects whose retina is rich in M cones reported more green than red sensations. This large number of required hue categories and the tendency to correlate green and red sensations with the varying L to M ratio is constant for all of the tested wavelength of light. The interpersonal difference in the identification of hues is different from previous experiments performed without the aid of adaptive optics, where statistics of reported color sensations was constant between individuals. In fact, this dependency of expected green and red responses from L to M ratio is not present in color
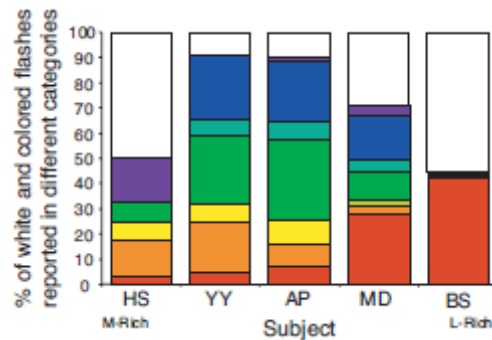
Figure 4.4: The color sensation reported by subjects when presented a small spot of 550 nm light. Image courtesy of [104].

appearance in large scale stimuli.

What can be evidenced from this experiment is that stimulating cones with the same photopigment can bring to different color sensations, even without stimulation of other retinal regions or different wavelength-sensitive cones. This challenges the elementary theory of color opponency and trichromatic vision, suggesting that spatial distribution of stimuli, photoreceptor mosaic and L to M ratio have some kind of role in this change of paradigm.

In [107], it is argued that there are interactions between pattern vision and color vision since they are both processed in the early stages of visual encoding, leading to a trade-off between the acuity of color and pattern vision. Moreover, when considering our vision as trichromatic, one assumption is that the spatial structure of the image has no role in the sensing of a scene. A retinal image is considered to be composed by three different cone mosaics, each with its own elaboration of the light spectra in the form of spot-like cone excitations. However, this model of vision does not consider the spatial variation of the three cone classes submosaics: at each retinal location there is only one cone, and the sensation of color comes from comparison between cones of different classes at different spatial locations, meaning that space and color are mutually dependent in the neural processing of color [105].

Alleysson and Meary, in proposing a Neurogeometry of color vision [108], find an obstacle in the randomness of the cone mosaic and the implication that post-receptoral elaboration would need to adapt to the local arrangement of cones in the mosaic. Since the chromatic information is subsampled, the visual system must perform some sort of demosaicing, so a knowledge

64

of the spatial position of each cone is required. The interaction between the physiology of neurons in the visual system and the phenomenology of color perception is still not clearly explained, so they suggest that the interlaced sampling performed by the cone mosaic could be key to formulate a neurogeometry of color vision.

Summarizing, many aspects are still not clearly explained by research, like how spatial distribution and L to M cone ratio affects the sampling of a retinal image [105, 104]. Further investigative methods to explore these processes can be quantitative and modelling approaches, to statistically analyse how distribution of cones affects visual perception.

Chapter 5

The model: Blue-noise sampling for cone distribution

The process of sampling consists in the reduction of a continuous signal into a discrete one, or the selection of a subset from a discrete set of signals. For sampling to be effective two competing goals must be satisfied: samples should be uniformly distributed in a way that there are no discontinuities; but at the same time, regular or repeating patterns must be avoided to prevent aliasing, like moiré patterns in a digital image (Figure 5.1).

In the human retina, the mosaic formed by the cone photoreceptor cells performs sampling on the retinal optical projection of the scene, achieving the first neural coding of the spectral information from the light that enters the eye. In order to solve the sampling problem, the human retina has adopted an elegant solution in the form of an arrangement of photoreceptors that is neither perfectly regular nor perfectly random. Local analysis of mosaics in the foveal region show that cones are arranged in hexagonal or triangular clusters, but extending this analysis to larger areas or eccentricities shows characteristics such as parallel curving and circular rows of cones associated with rotated local clusters.

It is still unclear how the spatial distribution and mean density of cones can affect the sampling of a retinal image [109]. An interesting evidence of this open issue is the experiment from Hofer [104] which tested the perception of stimuli of small spatial scale, described in Section 4.2.
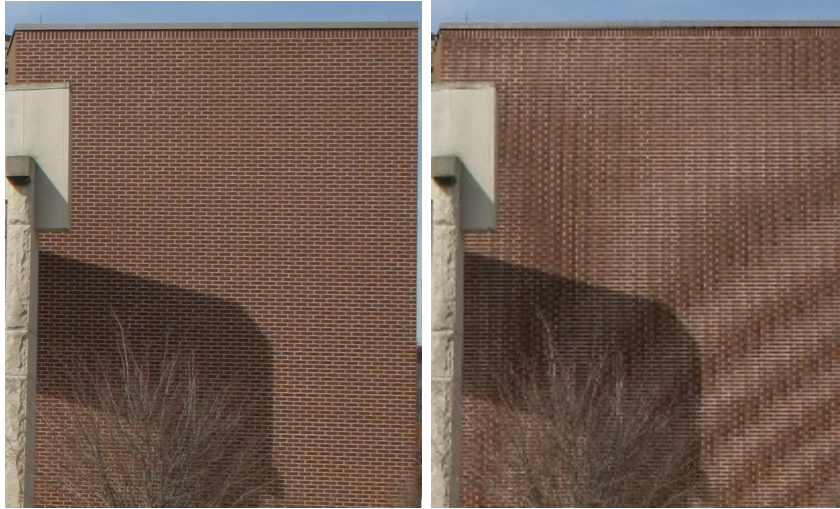
Figure 5.1: Left: a properly sampled image. Right: an undersampled image presenting moiré patterns. Images taken from `https://commons.wikimedia.org/wiki/File:Moire_pattern_of_bricks.jpg`

In this chapter it will be demonstrated that the sampling properties of the cone photoreceptor mosaic can be modeled by a blue-noise algorithm, and that they can be used to generate sampling arrays with the same features of the retinal cone mosaics. More specifically, an algorithm is identified that is capable of generating sampling arrays with the same range of densities in the retina, and specific metrics are used to compare the spatial and spectral properties of the real and generated cones distribution.

5.1 Sampling algorithms

The term *blue noise*, originating from Ulichney [110], refers to an even, isotropic, yet unstructured distribution of points. Blue noise was first recognized as an integral component in image dithering since it manages to capture the intensity of an image by means of its local point density, without the introduction of artificial structures of its own. It promptly became widespread in numerous scientific fields, particularly in computer graphics, where its isotropic properties enable a high-quality sampling of multidimensional signals, while its absence of structure prevents aliasing. Its visual efficacy, used to some extent in stippling and pointillism, has been argued

to be linked to the presence of a blue-noise arrangement of photoreceptors in the retina, according to Yellott [95]. Over the years, both the characteristics and the generation of blue noise distributions have been the subject of numerous research efforts in the computer graphics field.

It is argued that the very first approach at algorithmically generating point distribution that mantain a good balance between spatial irregularity and density control was obtained through error diffusion [111, 110], which was particularly well suited for low-level hardware implementation, for example in printers.

Meanwhile, uniform and regularity-free distributions started being of interest in the context of anti-aliasing, belonging in the computer rendering field [112]. In 1986, Cook [113] first proposed a dart-throwing algorithm to create Poisson disc distributions with a minimal-spacing rule, where points are never closer than a certain distance threshold. Subsequently, there were many efforts to provide improvements and modifications from this very algorithm [114, 115, 116, 117, 118]. Current advances in the development of Poisson disc algorithms led to very efficient and versatile implementations [119, 120], also running on modern GPUs to improve computational times [121, 122, 123]

Other than Cook, it is thanks to the founding works of Dippé and Wold [124], Mitchell [125] and Shirley [126] that in the computer graphics community there was awareness of the fact that noise and aliasing are phenomena tightly related to sampling. Since then, a large number of approaches based on optimization have been proposed, and the majority of them can be distinguished into two main branches: (1) on-line optimization-based approaches [115, 119, 127, 128, 122, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138], and (2) off-line optimization [139, 140, 141, 142, 143, 144], where the near-optimal solution is already prepared in the form of lookup tables, used directly in runtime.

The idea of using precalculated patterns computed offline in order to allow a fast and high-quality blue noise generation was first proposed by Dippé and Wold [124]. One of the major concerns of this technique is the possibility of aliasing artifacts due to repeated patterns, so Cohen et al. [145] advised the use of non-periodic Wang tiles, subsequently leading to an improved hierarchical sampling [140] and many other tile-based alternatives [139, 146, 141, 147]. In 2014, Wachtel et al. [142] proposed

a tile-based method that incorporates spectral control over distribution of samples. More recently, Ahmed et al. [144] introduced a 2-D square tile-based sampling method with the characteristic of having only one sample per tile and controllable Fourier spectra.

The present work uses as reference the approach called *Blue Noise Through Optimal Transport* (BNOT), developed by de Goes et al. [134], since it allows to achieve the best Blue Noise distribution known today.

5.2 Methods

The mosaics of cone photoreceptors used for the measurements have been obtained from images of patches of real human retinas previously published in the literature, and are visible in Figures 5.2. They were acquired from the electronic versions of the papers where available, and saved as png images. The pictures of patches are from different subjects of various ages, and were obtained with different techniques through their respective works: from histological tissue prepared for electronic microscopic imaging in [98, 148, 101, 149], to the most recent *in vivo* imaging techniques, characterized by the use of adaptive optics like deformable mirrors coupled with a wavefront sensor to compensate for the ocular aberrations of the eye [2, 150, 3, 151].

The center of the of the cells inner segments were manually plotted using WebPlotDigitizer [152] and saved as a text file containing the x and y coordinates of the centers. This work is in fact based on a relatively small dataset due to the short availability of ample collections of retinal images with a visible cone mosaic, an issue probably caused by the different imaging techniques and, where applicable, tissue preparation and hopefully there will be larger and more accessible datasets in the future. A manual plotting of the points was chosen because most of the articles had not available cone coordinates, and the ones who did had insufficient spatial precision. In the analysis of points distribution, the distance between the cone centers in space was converted in actual μm on the retina by multiplying them with the appropriate scale factor of the image, determined by the size of the sample window's side. Conversion from visual degrees was performed according to the model from Drasdo and Fowler [153], with one degree of visual angle equal to 288 μm on the retina. The obtained cone spacing values are compatible with Wyszecki and Styles [154], with the exception of data
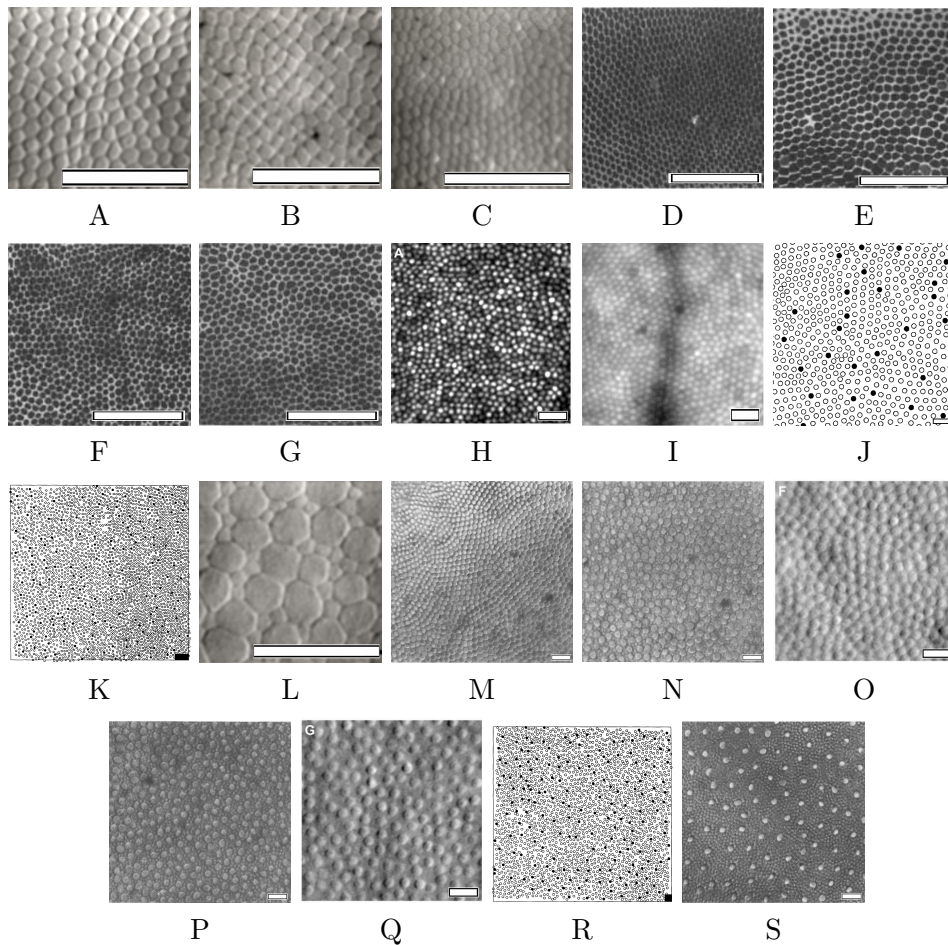
Figure 5.2: The pictures of the patches of retina used for the analysis, sorted by growing eccentricity from the foveal center, scale bars are 20 μm. **Retinal locations**: A,B,C,D,E,F,G: Foveal center. H: 1° temporal to fixation. I: 1° nasal from the foveal center. J: edge of the rod-free zone, about 0.35 mm from the foveal center. K: 0.36 mm eccentricity. L: foveal slope, 0.66 mm temporal to foveal center of A. M: 0.7 mm from the foveal center. N: 1 mm from the foveal center. O: 5° temporal to fixation. P: about 2.25 mm from the foveal center. Q: 10° temporal to fixation. R: 3 mm eccentricity from the foveal center. S: 9.5 mm from the foveal center. **Image sources**: H,O,Q: Images from Scoles et al. [3]. I: Image from Roorda & Williams [2]. A,B,C,L: Images from Curcio et al. [98]. M,N,P,S: Images from Jonas et al. [148]. J,K,R: Images from Curcio et al. [101]. D,E,F,G: Images from Gao & Hollyfield [149].

from [149] exhibiting lower cone spacing values, probably attributable to post mortem shrinkage of the tissue. Additionally, retinas J, K and R have

been cropped during analysis because they didn't fully cover the sampling
window, and would have included uncharacterized areas.

5.2.1 Analysis of point process

In this section, basic notions from Stochastic Point Processes are briefly
introduced. A point process $\mathcal{S}$ consists of a stochastic generation of points
in a given domain Ω (in this case, $[0,1)^s$). A realization of a point process
with n samples is denoted by $P_n := \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}\} \subset \Omega$. A point
process $\mathcal{S}$ can be defined *stationary* if it is invariant by translation, and cen
be defined isotropic if it is invariant by rotation. More formally, if we assume
that $\mathbf{P}$ is a *probability measure*, $\mathcal{S}$ is stationary if $\forall \mathbf{x} \in \mathbb{R}^s$

$$\mathbf{P}(\mathcal{S}(\Omega)) = \mathbf{P}(\mathcal{S}(\Omega - \mathbf{x})), \tag{5.1}$$

and *isotropic* if any rotation or translation of $\mathcal{S}$ possess the same statistical
properties. It is also possible to define the *density* of a point set as the
average number of point samples inside a region B of volume V_B around a
sample $\mathbf{x}$.

$$\lambda(x) := \frac{B(x)}{V_B} . \tag{5.2}$$

This density has the characteristic of remaining constant for isotropic and
stationary point processes. A sampler which generates sets with a non con-
stant density is usually called a *non-uniform* sampler. In order to char-
acterize isotropic stationary point processes, the *Pair Correlation Function*
(PCF) is a widely used tool. Such function is defined as a characterization
of the distribution of pair distances of a point process. Oztireli [136] devised
a simplified estimator for this measure in the particular case of isotropic and
stationary point processes. The PCF of a pointset P_n in the unit domain
$[0,1)^s$ is obtained by

$$\varrho(r) = \frac{1}{n^2 r^{s-1}} \sum_{i \neq j} k_\sigma(r - d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})), \tag{5.3}$$

where $d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is a distance measure between $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$. Additionally,
the factor k_σ is used to smooth out the function. In his work, Oztireli relies
on this smoothing function to assume ergodicity for all sets. He uses the
Gaussian function as a smoothing kernel, instead of a box or triangle kernel.

This estimator is used to compute a PCF with 3 parameters, the minimal r, $r_{\min}$, the maximal r, $r_{\max}$ and the smoothing value σ, usually chosen empirically. Moreover, as the number of samples increases, the distances between samples will be very different for similar distributions. To mitigate this issue, the distances are normalized during estimations by using the maximal possible radius for n samples ([118], Eq (5)).

Figure 5.3 shows how the PCF of several point processes characterizes the spectral content of the point distribution: a pure uniform sampling, Green-Noise and Pink-Noise samplers obtained using [147], a jittered sampler (for N samples, subdivision of the domain into regular $\sqrt{N} \times \sqrt{N}$ square tile and a uniform random sample is drawn in each tile), a Poisson-Disk sampler [117] and a Blue-noise sampler (BNOT) [134].

5.3   Results and discussion

Regularity index, or *conformity ratio* is a quantitative method used for assessing spatial regularity of photoreceptor distributions [94, 155, 156]. A *k-d* tree structure has been used to find the nearest neighbor for each point, the euclidean distance was calculated for each pair found this way and all the results are classified in histograms. Each distribution of distances to nearest neighbor can be described by a normal Gaussian distribution described by the equation

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2} . \tag{5.4}$$

where μ is the mean of the distribution and σ the standard deviation of the measurements. The regularity index is expressed by the ratio of the mean μ by the standard deviation σ. This index is reported to be 1.9 for a full random sampling and the more regular the arrangement, the higher the value, usually 3-8 for retinal mosaics.

Regularity indexes for retinal data are shown in Table 5.1. In contrast with previous claims, our calculated indexes range from 8 to 12. In the lower bound there is data obtained from [101], which instead of a retinal image shows the marked locations of the inner segments of photoreceptors; meanwhile in the upper bound, close to 12, most of the data is from foveal centers in [149], with the sole exception of retina D, where the different sizes of the photoreceptor profiles reflect different levels of sectioning through the
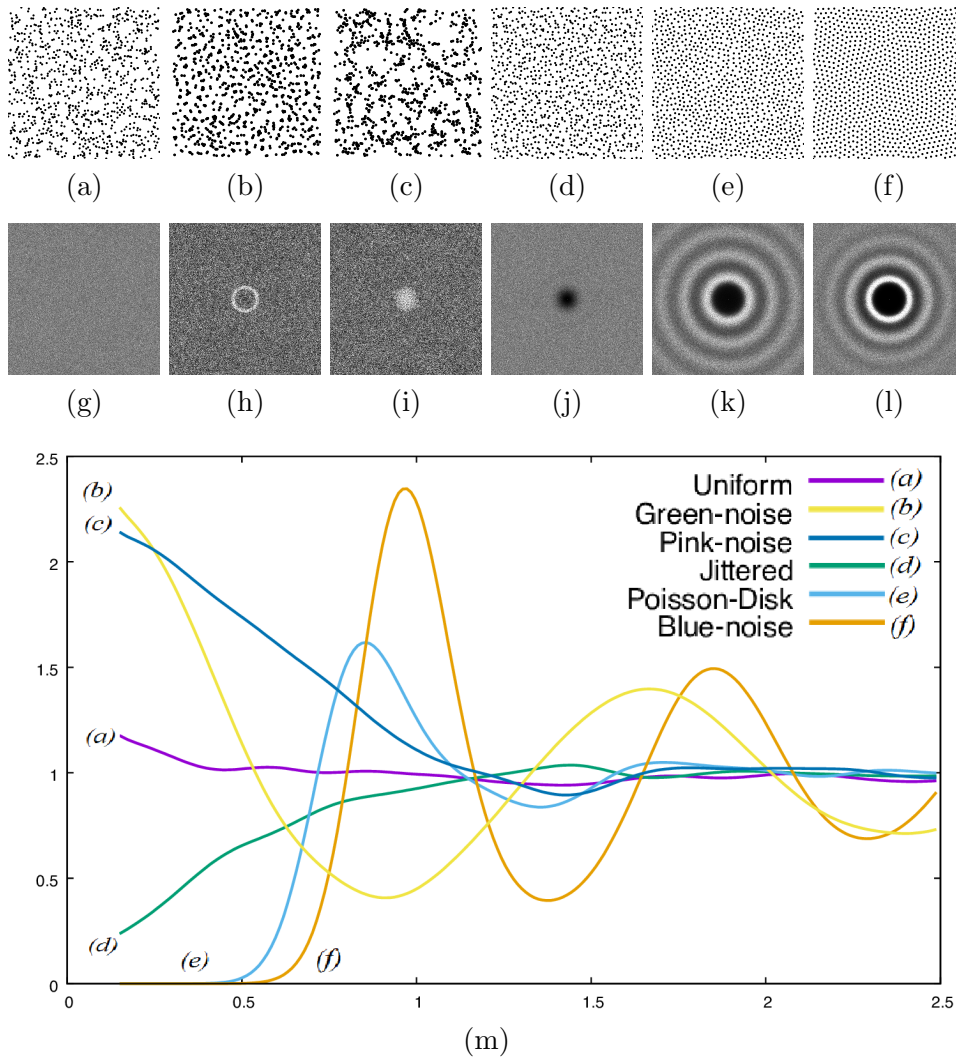
Figure 5.3: Pair Correlation Function of various 2-D samplers. *Fist row from left to right:* Realizations of 1024 samples from a uniform (*a*), a Green-Noise sampler (*b*), a Pink-Noise sampler (*c*), a jittered (*d*), a Poisson-disk (*e*) and a Blue-Noise sampler (*f*). The second row shows the Fourier spectrum (power spectrum) of each sampler ((*g*) − (*l*), spectrum computed on 4096 samples). The PCFs capture the spectral content of each sampler as shown in (*m*), horizontal axis is r, vertical axis is $\varrho(r)$.

inner segments.

The indexes for data generated with Green noise, Pink noise and BNOT samplers are presented in the same table. As expected, the indexes for Green and Pink noise are assimilable to those of a full random sampling, in fact

Data	μ	σ	RI
A	2.15	0.20	10.50
B	1.97	0.20	9.87
C	1.46	0.13	11.36
D	1.50	0.16	9.51
E	1.63	0.13	12.03
F	1.63	0.14	11.95
G	1.85	0.15	12.26
H	3.93	0.35	11.04
I	5.06	0.43	11.65
J	4.60	0.42	10.87
K	4.03	0.51	7.97
L	5.06	0.51	9.88
M	3.83	0.42	9.20
N	6.01	0.61	9.76
O	6.83	0.62	10.98
P	7.22	0.79	9.09
Q	8.58	0.89	9.66
R	9.04	1.07	8.47
S	12.73	1.45	8.79
GreenNoise_512	0.02	0.01	1.31
GreenNoise_1024	0.01	0.01	1.38
PinkNoise_512	0.02	0.01	1.42
PinkNoise_1024	0.01	0.01	1.44
BNOT_1050	0.03	0.01	21.45
BNOT_2050	0.02	0.01	22.92
BNOT_4050	0.01	0.01	23.31

Table 5.1: μ, σ and regularity indexes of retinal mosaics

they are even lower, averaging 1.3 and 1.4 respectively; meanwhile, for the BNOT data, the indexes values are much higher, more than the double of the highest values for retinal RIs. It is not very surprising that, thanks to the the uniformity optimization of BNOT, the indexes are this high; but still very far from the infinite RI of regular lattices. Given the fact that fully regular hexagonal or square patterns are proven to possess poor sampling properties and therefore not suitable for simulating cones distribution, in the scope of this work a higher RI indicates that BNOT performs better at generating point processes than the other analyzed methods.

A more recent and reliable method for assessing the goodness of these point processes is the previously mentioned Pair Correlation Function. In

Table 5.2 are reported the l_∞ distances between the generated point sets and the measured PCF, also plotted in Figure 5.4 . From two PCFs ϱ and ϱ_2, their l_∞ distance is denoted as the maximal distance between the two functions

$$l_\infty(\varrho, \varrho_2) = max_r |\varrho(r) - \varrho_2(r)|, \qquad (5.5)$$

where r is a given radius. The l_∞ distance has been adopted since Oztireli used this measure to compare PCFs [136], two distributions can be considered the same if this distance is under 0.1.

Data	Jitter 1024	DT 1024	BNOT 1024
A	1.14	0.85	0.41
B	0.89	0.52	0.55
C	0.95	0.60	0.48
D	0.64	0.33	0.83
E	1.13	0.77	0.30
F	1.15	0.80	0.30
G	1.11	0.74	0.33
H	0.94	0.52	0.51
I	1.13	0.76	0.30
J	1.17	0.81	0.26
K	0.76	0.25	0.78
L	0.63	0.48	0.86
M	0.96	0.60	0.47
N	0.60	0.26	0.92
O	0.91	0.57	0.52
P	0.81	0.52	0.64
Q	0.66	0.31	0.78
R	0.69	0.24	0.80
S	0.58	0.38	0.92

Table 5.2: l_∞ distances between pairs of PCFs. If this difference is under 0.1, the two distribution can be considered to be the same. It is possible to observe that the Dart Throwing and BNOT samplers are the closest from the measured distribution.

The closest results are from comparison with BNOT and Dart Throwing samplers, moreover, the higher the measured RI for the retinal distribution of photoreceptors, the lower the distance from BNOT PCF. The opposite happens when comparing with Dart throwing algorithm, the closer to the reported RI of 8, the lower the l_∞ distance. This evidences that not only the indexes are actually higher than the ones previously measured, but also

that the most effective method to simulate these distributions comes from Blue-noise samplers. Figures 5.5 through 5.8report the results of the Nearest neighbor analysis and the PCFs of analyzed retinas.
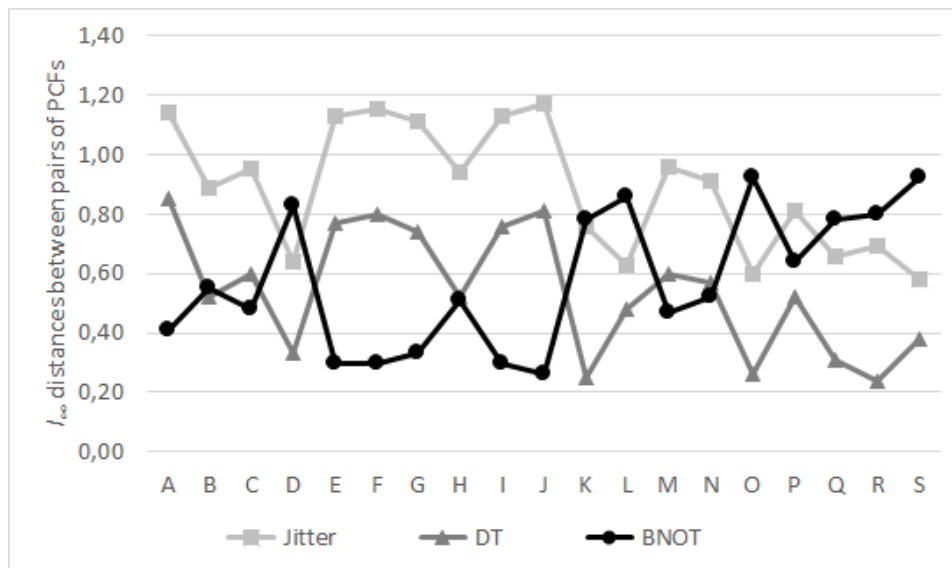
Figure 5.4: l_∞ distances between pairs of PCFs. Horizontal axis are the retinal locations, top axis is l_∞ distance between the location and the sampling algorithm. If this difference is under 0.1, the two distributions can be considered to be the same. It is possible to observe that the Dart Throwing and BNOT samplers are the closest from the measured distribution.

This study, however, has some limitations in the source material for the retinal images that should be noted. Jonas et al. [148] used a tissue preparation that left the fovea unusable and reduced cell counts by half. Gao and Hollyfield [149] sampled only the foveal and equatorial retina and thus missed the import of macular rod loss. Additionally, an important work that was not included is the one from Curcio et al. [157], which established rod loss in central macula as a defining feature of human retinal aging. This study used retinal flat mounts with preserved foveas and unbiased sampling, and computer-assisted counts of inner segments for accuracy. Data for cone counts in this study were also replicated *in vivo*, but rods are generally not visible in current adaptive optics systems.
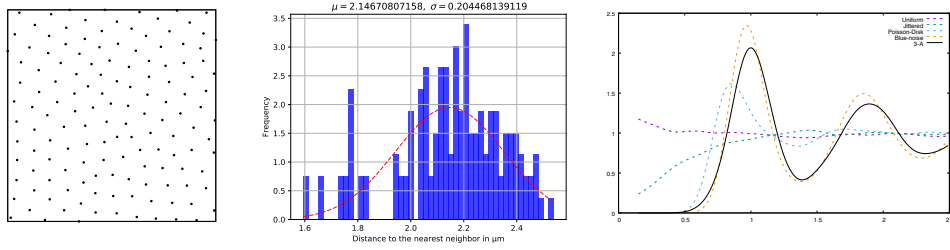
A: Foveal center. Original image from Curcio et al. [98].
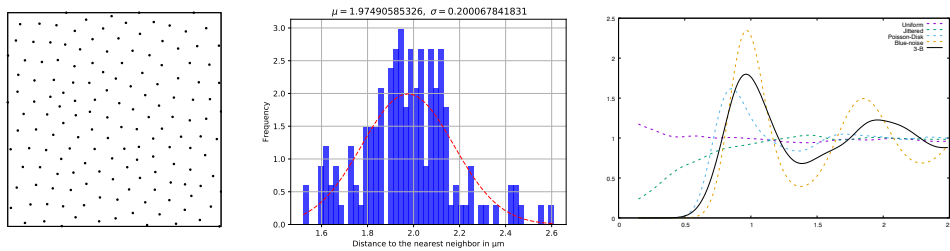
B: Foveal center. Original image from Curcio et al. [98].
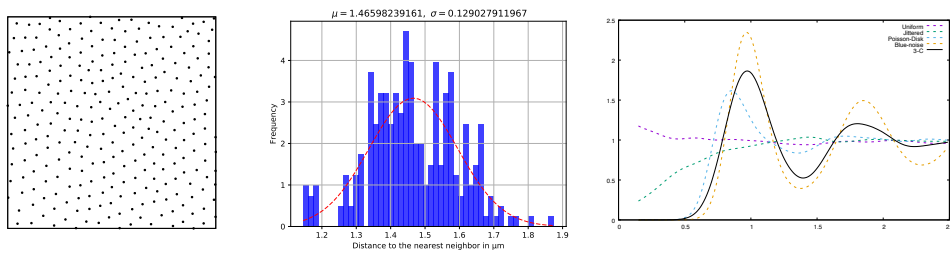
C: Foveal center. Original image from Curcio et al. [98].
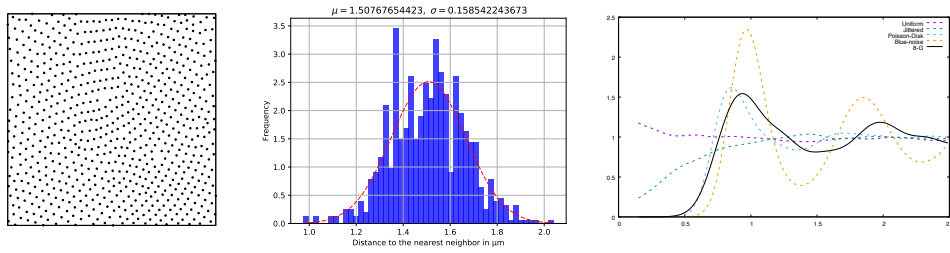
D: Foveal center. Original image from Gao & Hollyfield [149].
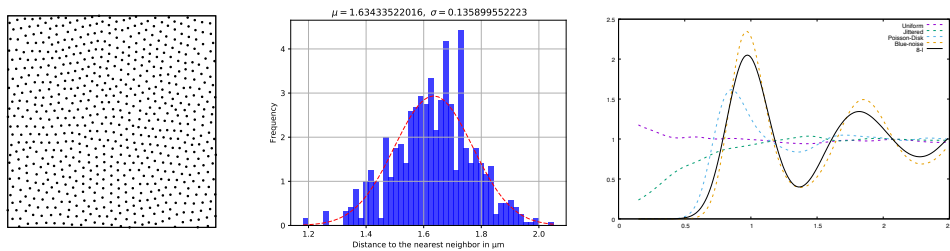
E: Foveal center. Original image from Gao & Hollyfield [149].

Figure 5.5: From left to right: The point samples extracted from the retinal locations of Figure 5.2, Nearest neighbor analysis with mean and standard deviation values, Pair Correlation Function.
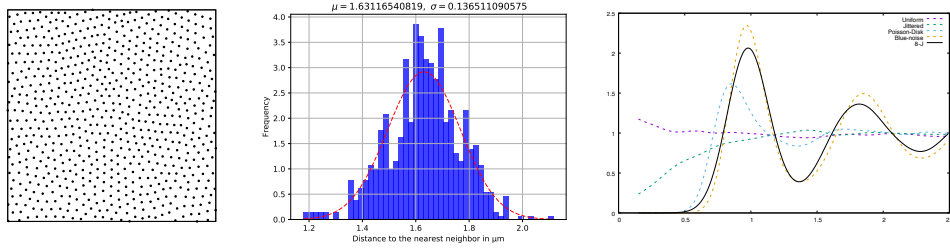
F: Foveal center. Original image from Gao & Hollyfield [149].
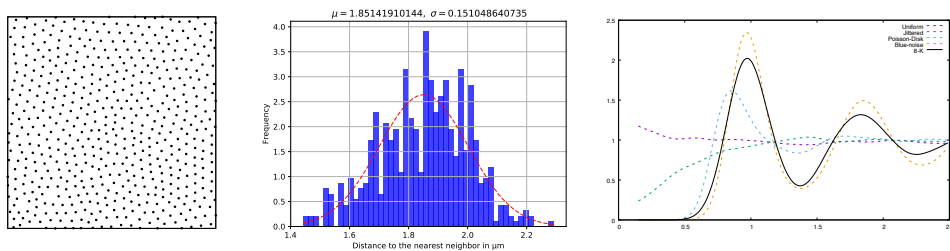
G: Foveal center. Original image from Gao & Hollyfield [149].
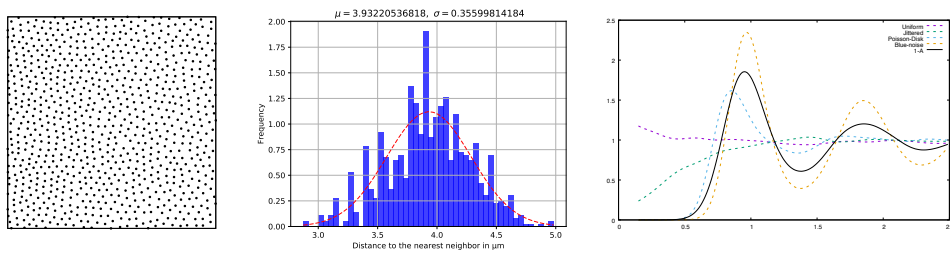
H: 1° temporal to fixation. Original image from Scoles et al. [3].
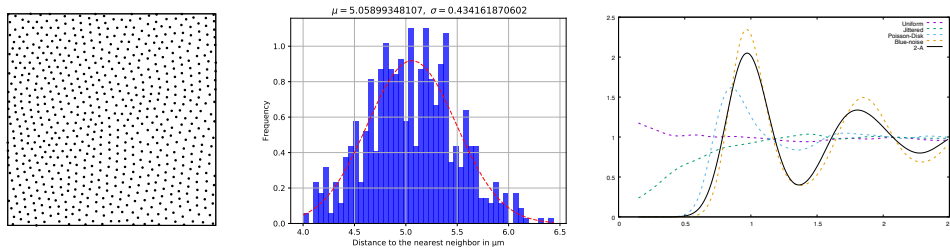
I: 1° from foveal center. Original image from Roorda & Williams [2].
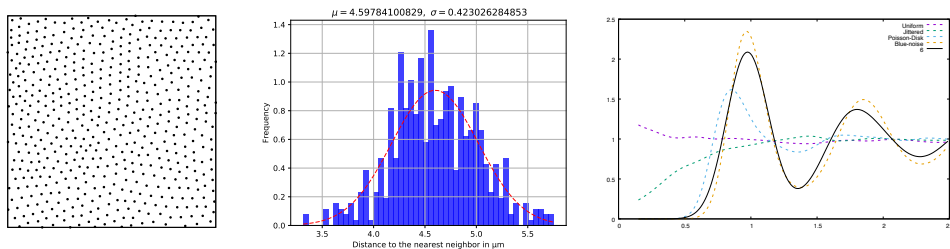
J: 0.35 mm from the foveal center. Original image from Curcio et al. [101].

Figure 5.6: From left to right: The point samples extracted from the retinal locations of Figure 5.2, Nearest neighbor analysis with mean and standard deviation values, Pair Correlation Function.
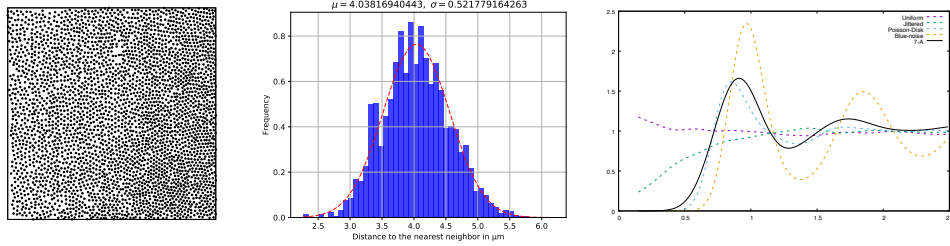
K: 0.36 mm eccentricity. Original image from Curcio et al. [101].
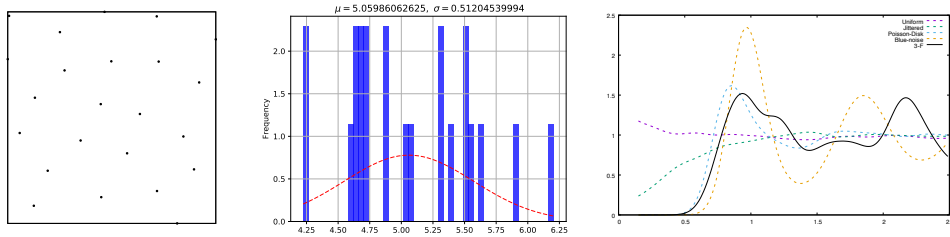
L: 0.66 mm from foveal center of A. Original image from Curcio et al. [98].
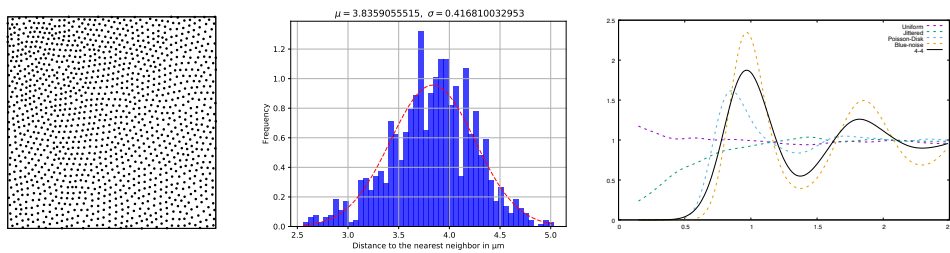
M: 0.7 mm from the foveal center. Original image from Jonas et al. [148].
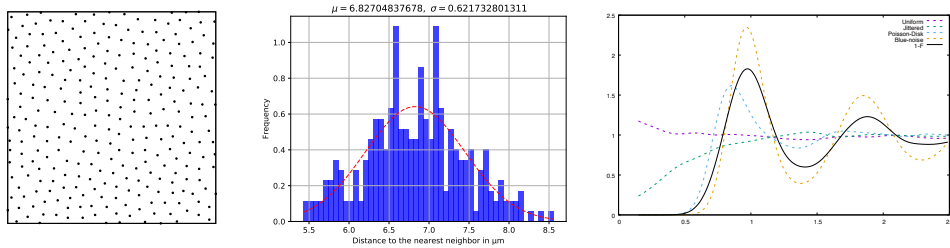
N: 1 mm from the foveal center. Original image from Jonas et al. [148].
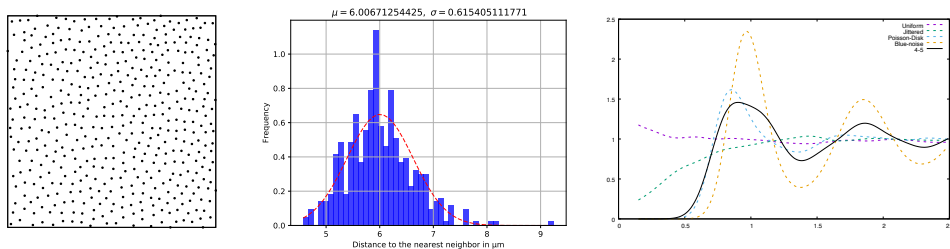
O: 5° temporal to fixation. Original image from Scoles et al. [3].

Figure 5.7: From left to right: The point samples extracted from the retinal locations of Figure 5.2, Nearest neighbor analysis with mean and standard deviation values, Pair Correlation Function.
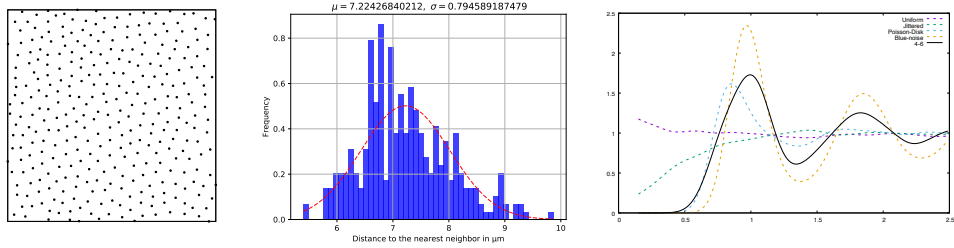
P: About 2.25 mm from the foveal center. Original image from Jonas et al. [148].
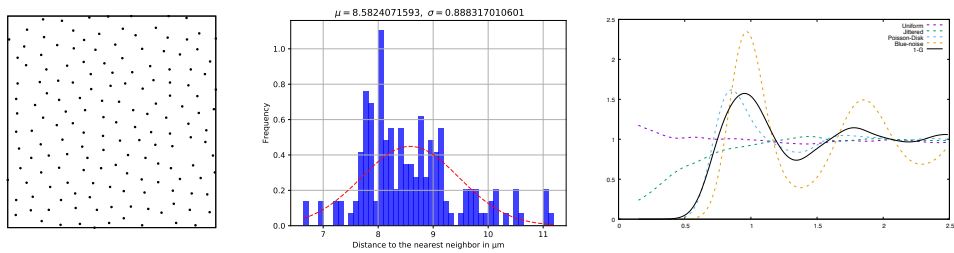
Q: 10° temporal to fixation. Original image from Scoles et al. [3].
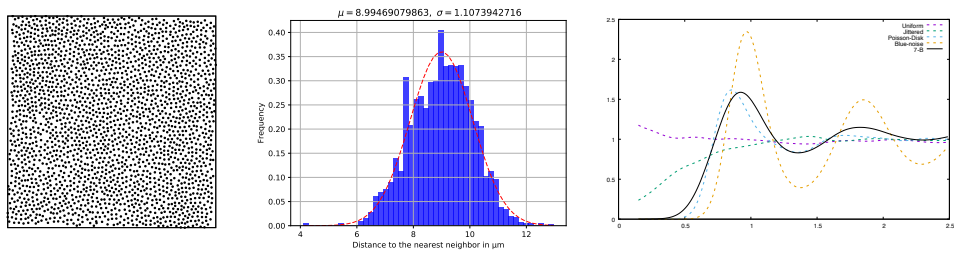
R: 3 mm from the foveal center. Original image from Curcio et al. [101].
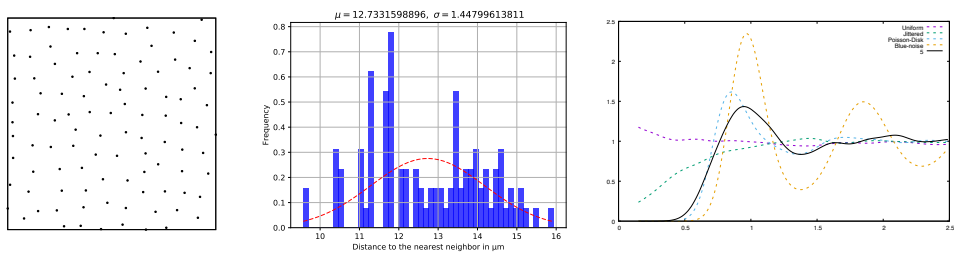
S: 9.5 mm from the foveal center. Original image from Jonas et al. [148].

Figure 5.8: From left to right: The point samples extracted from the retinal locations of Figure 5.2, Nearest neighbor analysis with mean and standard deviation values, Pair Correlation Function.

5.4 Generation of cone mosaic

In order to generate cone sampling mosaics in the retina, Curcio's data [98] on photoreceptor density was used to generate retinal patches with the BNOT [134] blue noise sampling algorithm. The first step in this process is to create a gradient to be used as reference ramp for the algorithm. One characteristic of blue-noise sampling algorithms is that they can be applied adaptively over a grayscale image to produce a sampling with the same density properties of the input image.
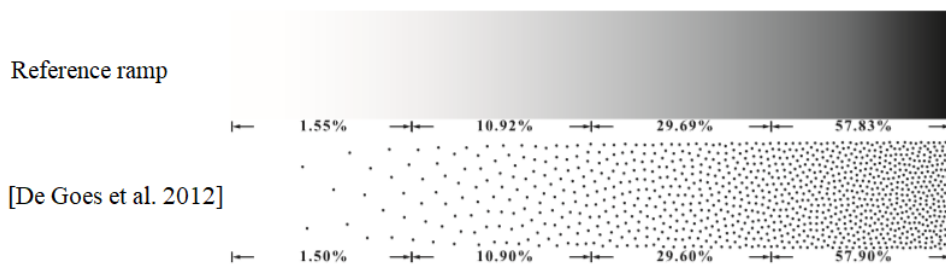
Figure 5.9: Blue noise sampling of a quadratic density function with 1000 points. The percentages in each quarter indicate RGB density in the image, and point density in the example. Image courtesy of [134].

The values used to create the grayscale gradient were acquired starting from the number of cones per square millimeter along the temporal meridian, and are reported in Table 5.3. In order to convert the cones per square mm values into the grayscale RGB values, the point of highest density at the center of the fovea with 196890 cones per square mm, reported in the table as 0.00 visual degree, was considered to be the region of maximum density of the gradient and so it was used as the starting RGB value of (0,0,0). From there, RGB values were defined at every reported eccentricity with the formula

$$255 - (d \cdot 255)/d_{max}$$

where d is the density at any given eccentricity and d_{max} is the maximum density. The obtained RGB values were used to create a gradient with the software GIMP (GNU Image Manipulation Program) for the first 4 degrees of visual angle of the retina, while keeping the proportions of the spacing of the degrees data, that is not uniformly sampled but at intervals that become

Eccentricity	Density	RGB
0	196890	0
0.18	162414	44
0.36	121619	97
0.54	98424	127
0.71	80004	151
1.07	57712	180
1.43	45420	196
1.79	38380	205
2.14	34223	210
2.50	29027	217
2.86	24144	223
3.21	21165	227
3.57	19699	229
4.00	18736	230

Table 5.3: RGB grayscale values of the gradient of cone densities. Eccentricity is expressed in degrees of visual angle, Density in cones/mm^2.

larger with increasing eccentricities.

Once the gradient has been generated, it has been applied radially across an image to produce the reference ramp. A 2000x1000 pixels resolution image has been created containing the equivalent of 4 degrees of visual angle, going from -1 to 3 degrees horizontally, as visibile in Figure 5.10.

In order to obtain the synthesized retinal patches, the image gradient is used as reference ramp for the application of the BNOT algorithm. The image is divided in square regions of 1000x1000 pixels due to the requirements of the algorithm. The region with the fovea is used as reference in setting the parameters for the sampling, to obtain a photoreceptor count according to retinal data from [158], in this case it was observed that a maximum rank parameter K=15 yields the closest numbers. The generated patches are described as a list of x,y coordinates in the interval [0,1], corresponding to the center of the inner segments of cone photoreceptors, and are plotted in Figure 5.11.

A visual comparison between random sampling, blue-noise sampling and real cone retinal sampling, obtained by foveal cone data from Wang et al. [158], is presented in Figure 5.12. A 1000x1000 pixel test image has been processed by the proposed retinal model, to be discussed in detail in the following chapters. Comparison between outputs can be obtained in terms of
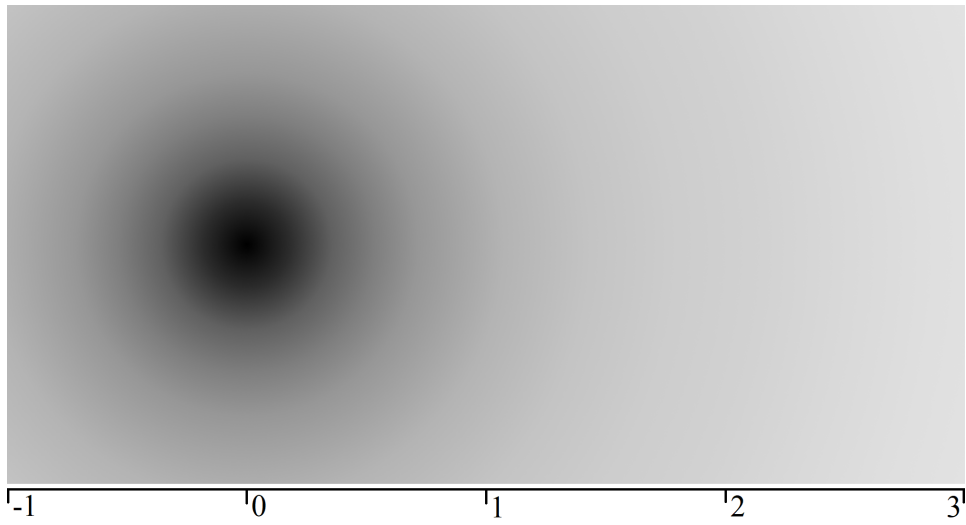
Figure 5.10: The reference ramp for the densities of the cone photoreceptors in 4 degrees of visual angle. The dark spot on the left part of the image corresponds to the fovea, with the highest density.
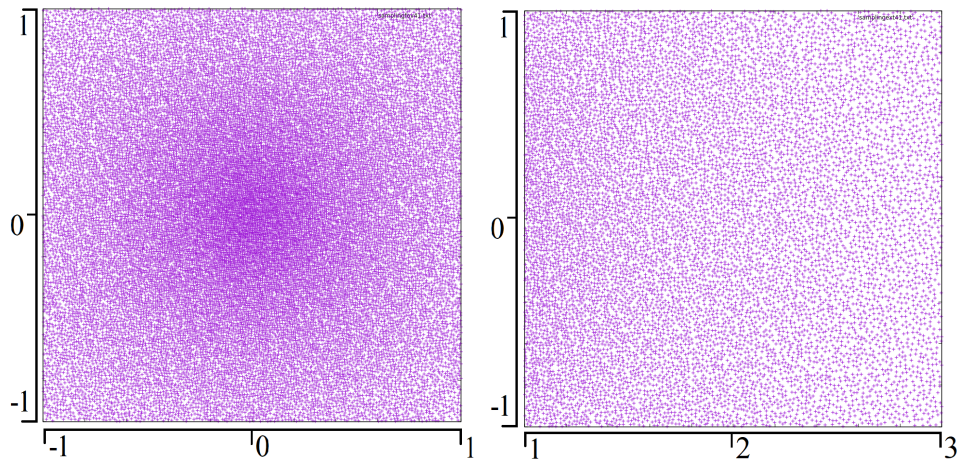
Figure 5.11: The plots of the center of the photoreceptor cells in two synthesized patches of retina, each spanning 2 degrees of visual angle, with one degree of visual angle equal to 288 μm on the retina according to the model from [153]. The plots correspond to the foveal region (left) and the parafoveal region towards the temporal meridian (right).

peak signal-to-noise ratio (PSNR), since Guerrero-Colón et al. [159] adopted this measure to compare performance of different Space-variant sampling

(a) Original Image
MSE = 0.0
PSNR = 361.20

(b) Random Sampling
MSE = 158.43
PSNR = 26.13
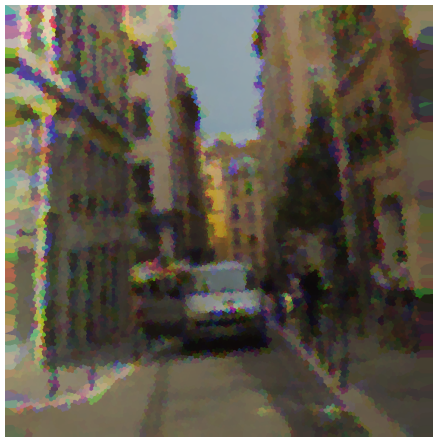
(c) Real retinal cone locations
MSE = 176.59
PSNR = 25.66

(d) Blue-noise Sampling
MSE = 151.30
PSNR = 26.33

Figure 5.12: Comparison between different sampling strategies with MSE and PSNR computed from the original image. (*a*) Original test image. (*b*) Test image sampled with 21085 random samples. (*c*) Test image sampled with 21085 retinal cone locations obtained from data from Wang et al. [158]. (*d*) Test image sampled with blue-noise distribution (21085 samples) obtained with the proposed method.

methods.

PSNR computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is used as a quality measurement between the original

and a compressed image: the higher the PSNR, the better the quality of the compressed, or reconstructed image. PSNR is calculated as $10log_{10}(255^2\sigma_e^2)$, where $\sigma_e^2)$ is the Mean Square Error (MSE). The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error, meaning that two identical images will have MSE = 0. It is worth noting that PSNR and MSE metrics are not used to establish a perceptive difference, but it is only used to compare different sampling methods.

Summarizing, in this chapter it was demonstrated how point processes generated by blue noise sampling algorithms can be used to simulate features of a human retinal cone distribution with a certain degree of similarity with data found in existing literature, and how they can be used to accurately model local patches of retina. The analogies between this modeling technique and the unique distribution of neuron cells in the photoreceptor layer can be a starting point to determine the mechanisms underlying the development of retinal tissue, or to figure out how this spatial distribution affects the sampling of a retinal image, and its implications on human vision. Given the possibility of blue noise algorithms to generate adaptive sampling, it is theoretically possible to develop a space-variant model describing the placement of photoreceptors in the whole retina. However, such sampling would be difficult to validate, since imaging of a whole retina is difficult to obtain and analyze; the only way way would be through analysis of local patches and their relative density, or with isodensity maps. Recent advancements in sampling algorithms offer the possibility of applying a smooth sampling across local patches in the retina to obtain an adaptive sampling. Given the PCF and spectra of local patches, a point process with the same properties can be reproduced [160] and correlated by means of a heat map or gradient representing their interpolation in space [161]. The retinal imaging data available is limited, so for a more accurate modeling of the cone sampling, large datasets would be beneficial. Moreover the available data is often not consistent, due to different imaging techniques and tissue preparation.

Chapter 6

The model: Cone classes and absorption

Human vision is a complex biological system. In this chapter the focus is on the spatial and chromatic structure of the initial stage of vision: the first layers of the human retina. The human retina is not simply a sensor, but it consists of a complex structure that beside sensing the external world, has the task of preparing the sensed information to be sent over the optic nerve to subsequent elaborations in the visual cortex.

The model here presented has the goal of being an operative tool for further investigations on retinal functional mechanisms. The focus of the model is mainly on the variable structure of the first layers of the retina: variability is found not only on the different spatial concentrations of cones in function of their eccentricity, but also on the high inter-subjective variance in the L:M ratio of the actual cones population.

The first step of this simulation, which starts from a strictly biologically plausible standpoint, was the creation of a model of spatial distribution of cones inside of a retina, that preserves the spatial characteristics of the cells mosaic present in the literature, and validating it through comparison with real retinas (see previous chapter). Subsequently, inside of the model the spatial descriptive statistics will be created for each photoreceptor, to contribute to the creation of a graphical representation using a reconstruction tool, to be discussed in Chapter 7. These latter stages of the model have to simulate the behavior of biological systems of such complexity that a *black box* approach had to be considered.

The image reconstruction tool allows to simulate the effect of varying L:M ratios, visualizing its consequences on the sensed image. The output of the model is in fact an image realized by interpolation of the spatial sampling of the cones: it is not to be intended as the final *perceived* image, since the model does not account for any visual cortical stage, but it has the purpose of visualizing the immediate effect of L:M ratio variation in relation to the uneven foveal/peripheral cone distribution. The output generated by the model's simulation can be used as input for more complex explanatory models, i.e. accounting for contrast sensitivity function, spatial chromatic acuity, or many other biological characteristics.

6.1   The model pipeline

The pipeline of the proposed model is presented in Figure 6.1. An RGB image in PNG format is used as input for a representation of a visual stimulus, and the desired output is a space-variant representation of an observed scene. The model is constituted by separate methods each corresponding to a different stage of color vision. The first step is simulation of the spatial sampling performed by the cones on the retina. This is performed with a blue-noise sampling algorithm, due to shared sampling properties with the cone mosaic [162], allowing to generate the locations of the center of the cones inner segment over a square region representing a patch of retina. The synthetic retina and the input image are then used as input to our proposed algorithms. In the model, Cone classes are determined for every sampling point in the synthetic retina patch, and a Voronoi diagram is used to simulate the polygonal aperture of the cone inner segment. Absorption of light is simulated inside of their area of influence as the average of all the [0,255] values in the corresponding RGB channel, assuming L = red, M = green and S = blue. This generates an intermediate image with an "interleaved" sampling, since in our vision each cone class does not sample every location of the retinal image [107].

A tool was also developed to propose an interpolation approach in reconstructing the correct sensed color information from the neighborhood of each cone, mimicking the horizontal connections in the outer and inner plexiform layers of the retina. At the end, the final space-variant image displays the "demosaiced" result, produced by the simulated color-pattern interaction

bound to happen in color vision. The tool is described in detail in Chapter 7.
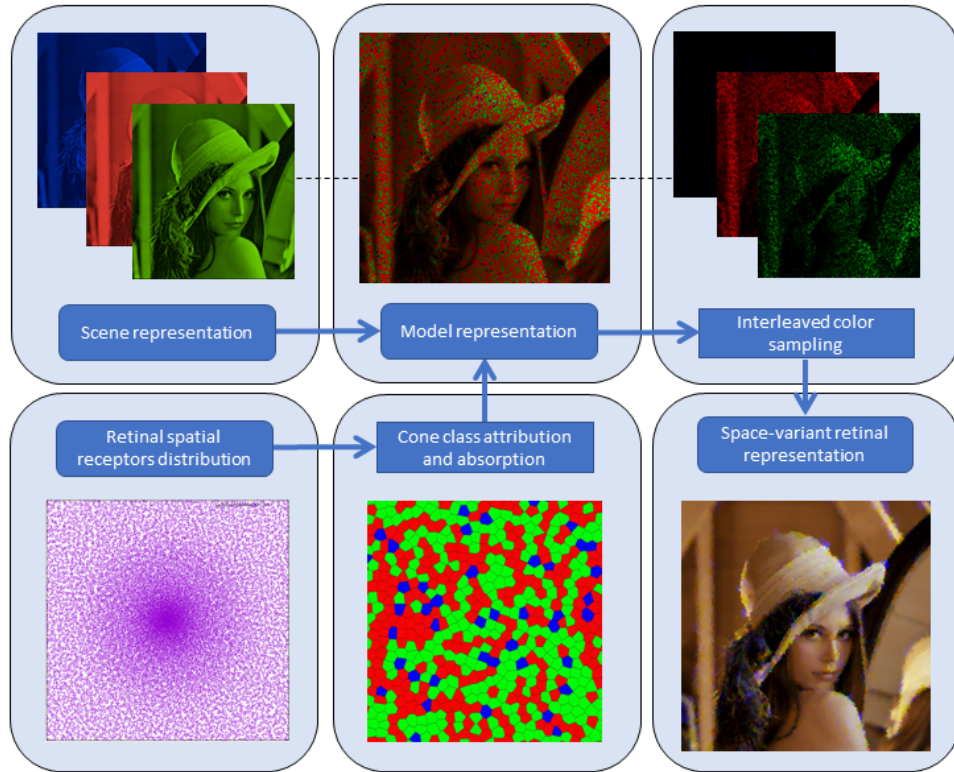
Figure 6.1: The proposed model pipeline.

6.1.1 Projection Geometry

Since the image used as input in the model is considered as directly projected onto the retinal surface, a conversion is needed that takes into account the viewing geometry given a specific viewing distance, given that the model works directly on pixels in a digital image. Suppose that there is an image with a black and white pattern to be observed, and the stripes of this pattern possess a measurable size, for example 1 cm. If the image is observed at a close distance, there is no doubt that the stripes will be clearly distinguishable and defined in an emmetropic eye. However, as the distance increases, there will be a point where the stripes become blurred and instead of a grating, only a medium gray will be perceived. When this happens is because the visual angle of the projected grating becomes smaller than the

visual angle of a single foveal photoreceptor, and their resolving power is not enough.

The visual angle can be easily predicted and calculated given the distance between the object and the observer and the object's size. The visual angle is inversely proportional to the viewing distance, and directly proportional to the object's size: a distant object will appear smaller than a closer object and a bigger object will appear larger than a smaller object. The formula to calculate the visual angle knowing the object size and distance is

$$V = 2 \cdot atan \left(\frac{\frac{S}{2}}{D} \right)$$

where V is the visual angle, S and D are the object's size and distance in the same units. The synthesized fovea in section 5.4 corresponds to 2 degrees of visual angle, hence supposing that the input image to the model is 2x2 cm the correct viewing distance is 57.29 cm.

6.2 Distribution of the three cone classes

Once generated (see Chapter 4), the spatial distribution of cones is applied over an input image of $X \times Y$ resolution. Starting from the input image, a given number of points called *seeds* are created. Each point P is defined by the pair (x, y) such that $x \in [0, X)$ and $y \in [0, Y)$ and they represent the coordinates of the center of the inner segment of cones, the coordinates are represented as a set of pairs to easily remove any possible cone collision. Two options to generate such pair are given, the first option has been implemented to take an input file with the normalized coordinates of points such that those coordinates are included between zero and one, like the one created with the method described in section 5.4, while the second option is used to create an uniform distributed set of pairs using the Mersenne Twister engine [163]. When coordinates are given as an input file, they are stretched over the image resolution. Once the seeds are created, they can be used to define a Voronoi diagram. The Voronoi diagram is a spatial structure in which the space is partitioned into regions [164]. Voronoi regions are defined by a seed s and all the points belonging to the region of s that are closest to s. More formally, given a space $\mathbb{X}$ which admits a metric, like an Euclidean vector space with its Euclidean distance, a subset of $\mathbb{X}$ of points

$P = \{p_1, ..., p_n\}$ with $n > 2$, the region R of a given point p_i is defined by:

$$V(p_i) = \{x \in \mathbb{X} \mid \parallel x - x_i \parallel \leq \parallel x - x_j \parallel \quad \forall \; j \neq i, j \in I_n, i \in I_n\} \quad (6.1)$$

where I_n is a set of indices of length n. The Voronoi seeds are used to represent the cone mosaic. An example containing five computation can be seen in Figure 6.2.
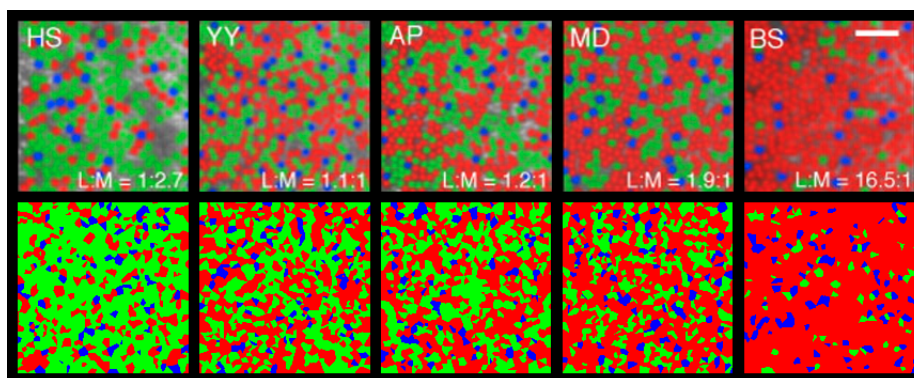
Figure 6.2: On the top row, the retinal mosaics from the same retinal location in five different subjects, courtesy of Hofer et al. [104]. On the bottom row, the computed Voronoi mosaic simulating the same L:M ratios.

Inside of the computational model, a Voronoi diagram is represented by the class *Voronoi*, which includes as attributes a vector of points called *seeds* to represent the seeds location and their associated cone type, the image resolution, the $L : M$ cone ratio, a counter for each cone type and a flag that tells the model if the inspected image is projected into the center of the fovea. This last flag is used only to let the model know that there will be a S-cones free zone in the center of the image, to simulate the presence of foveal tritanopia [165].

6.3 Photoreceptor absorption

The three submosaics of retinal cone cells allows to perceive different hues in a mechanism of opponency inside of the receptive field of the ganglion cells. However, spatial organization of the three classes of cones is not constant between individuals and mechanisms that determine their development and regularity are still object of research. The S cones mosaic is independent

from the L and M mosaics and it has observed to be distributed in a non-random fashion, taking the semblance of an almost hexagonal array [100]. Whether the remaining cones end up being M or L during the retinal development is entirely regulated by the X chromosome gene, which determines the corresponding cone opsin. L and M cones are generally randomly distributed in the retina, in fact an evident departure from the average ratio of 2:1 L to M has been observed, both interpersonally and in the same subject in different retinal locations, as illustrated in the top row of Figure 6.2, in addition to a tendency towards clumping. In spite of this high variance in spatial distribution of cones in different subjects, there has been no notable difference in subjective perception, but this compensation mechanism seems to not be consistent when small-scale spatial stimuli were involved, as previously evidenced in [104].

In the experiment of Brainard et al. [105], a yellow wavelength has been used to evaluate interpersonal differences in perception, since it is neutral in the red-green opponent mechanism and its perception is thought to be mainly regulated by differences in the excitation of L and M cones. Even in subjects with widely different L:M ratios, the wavelength perceived as yellow remains almost constant, indicating that this variability is offset by some perception mechanism probably involving cortical elaboration of L versus M cone opponency in the midget pathway, but it is however still controversial [99].

In the model, a variable L:M cone ratio distribution has been implemented to study how different ratios of cones family can change our sensation of color. The cone type assignment is executed after the Voronoi seeds distribution is generated. After a spatial distribution is created, a procedure to assign the cone type to each cone is launched. Such procedure takes in input the set of cones and generate the seeds in the Voronoi diagram used to visually represent the spatial structure of cone mosaic. Then, the proportion of cones type assignment is created over an auxiliary queue, which contains the indexes of shuffled seeds, again with the Mersenne Twister engine. The ratio of short waveband has been chosen to be 8% of the population, according to experimental data of Curcio et. al[101]. There is an exclusion zone in the center of the fovea in which S-cones are not present. The auxiliary queue provides indices for the Voronoi seeds vector. When inspecting the queue starting from any index, if the inspected cone is in the exclusion zone

it is discarded and appended to the back of the queue, otherwise it's chosen to be an S-cone, and the seed with that index is promoted to a short waveband cone and it is not appended to the back of the queue. This function is repeated until 8% of the population of cones has been chosen to be a S-cone. The ratio between L-cones and M-cones is defined by the user when invoking the reconstruction method or is standardized to $2 : 1$. In that case, first two third of the remaining population of seeds in the queue is selected and assigned to become long waveband cones, while the remaining cones are chosen to be medium waveband cones. Note that *seeds* is an attribute

---
Algorithm 1 Cone Assignment

---
    procedure ASSIGN CONE TYPE
        Queue $Q \leftarrow$ shuffled indices of *seeds*
        for $0.8 \cdot seeds.size()$ **do**
            $i \leftarrow Q.pop()$
            if $seeds[i]$ is not in the exclusion zone **then**
                $seeds[random\_index].type \leftarrow SHORT$
            else
                $Q.push(i)$
        for Number of M-cones **do**
            $i \leftarrow Q.pop()$
            $seeds[i].type \leftarrow MEDIUM$
        for Number of L-cones **do**
            $i \leftarrow Q.pop()$
            $seeds[i].type \leftarrow LONG$

---

of the Voronoi diagram: a vector of objects of class *Point* which is used to represent the cones among with their location and type.

Since the model works on digital images it operates in RGB color space, and there is an association between each seed and the RGB channel associated with the cone class, respectively red for L cones, green for M cones and blue for S cones. When calculating the absorption, for each Voronoi cell the average color of the surface is calculated and the value of the channel associated to the cone class is stored in the data structure. An image example of the tessellation of an image into photoreceptor cells with associated cone classes and their relative RGB channel average is presented in Figure 6.3

Original Image
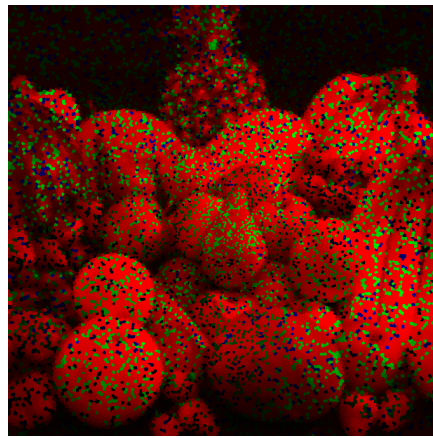
L:M = 4:1
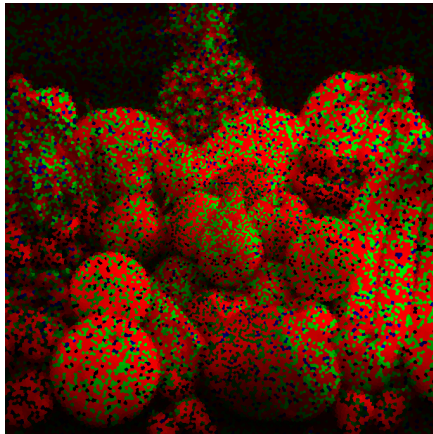
L:M = 2:1
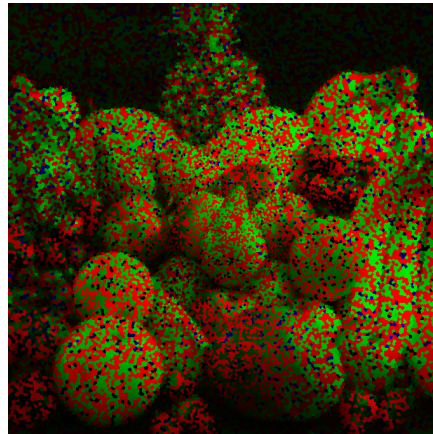
L:M = 1:1
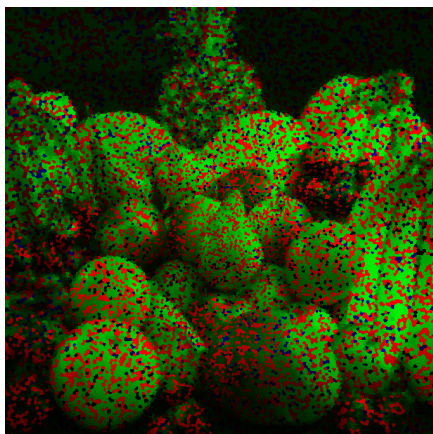
L:M = 1:2
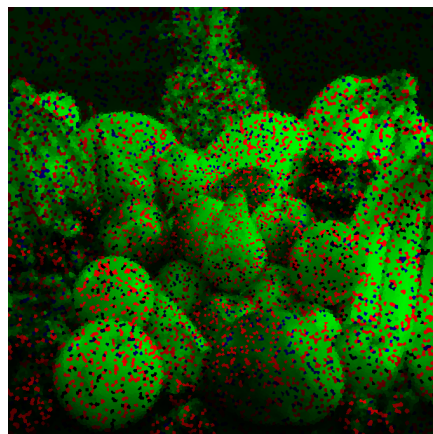
L:M = 1:4

Figure 6.3: Tessellation of an image into photoreceptors cells with associated cone classes and their relative RGB channel average for five different L to M cone ratio. Original image courtesy of Bill Ebbesen `https://commons.wikimedia.org/wiki/File:Culinary_fruits_front_view.jpg`

6.4 Conclusion

In this chapter was presented a modeling approach organized into a modular pipeline of extensible methods, each simulating a different stage of visual processing, mainly focused on the spatial sampling of a visual scene performed by the cone photoreceptor mosaic, and the reconstruction of color information inside of a space-variant image, obtained from the undersampled encoding performed by the three submosaics corresponding to the three different cone classes.

The method proposed is based on cone distributions created with a blue-noise sampling algorithm, since a direct comparison of mosaics generated with this method shows that they possess the closest sampling properties to real retinal cone mosaics [162]. Moreover, the model can be used with sets of coordinates obtained from real retinal cone locations, making it possible to use automatic identification of photoreceptors locations from retinal images [166]. The model allows the testing of different parameters that regulate the photoreceptor's topography including L to M cone ratio, to formulate hypothesis on perceptual differences arising from variations in spatial organization. Color sampling performed by the cone classes has been evaluated inside the cone apertures in the corresponding region of the Voronoi diagram, due to similarities in shape with foveal and parafoveal cones. Moreover, the Voronoi structure is used as the basis for the image reconstruction tool, since perceptually we are unable to distinguish our own cone spacing. The image reconstruction tool, while only providing only a straightforward representation of the mosaicked image with basic interpolation, will allow to visualize the effect of the parameter changes in the model with an output image representation.

Chapter 7

The model: Image reconstruction tool

The image reconstruction tool has been designed to provide a representable output of the sampling performed by the model. It does not mean to provide a cortical representation of a sampled scene but only to provide a space-variant image of the sampled regions in the visual field. Color reconstruction from photoreceptor signal is a biological process of such complexity that a *black box* approach has been considered. The image reconstruction tool allows to simulate the effect of varying L:M ratios, visualizing its consequences on the sensed image. The output of the model is in fact an image realized by interpolation of the spatial sampling of the cones: it is not to be intended as the final *perceived* image, since the model does not account for any visual cortical stage, but it has the purpose of visualizing the immediate effect of L:M ratio variation in relation to the uneven cone distribution. More complex visualization methods can take its place in order to simulate higher level mechanisms that in the actual vision system contribute to the final color sensation.

7.1   k-NN Reconstruction

There are three main functions for the color reconstruction tool: a function to display the image by showing in each Voronoi region the color based on what type of cone is associated to a certain region, a function that reconstruct color by simply summing the color information of the two closest seeds

of different type and a more advanced function that works like the second but searches for the closest $2k$ L-cones or M-cones and the closest k S-cones, as represented in Figure 7.1, and then reconstruct the final displayed color simply by taking an average of the color information of all the seeds of the same class.
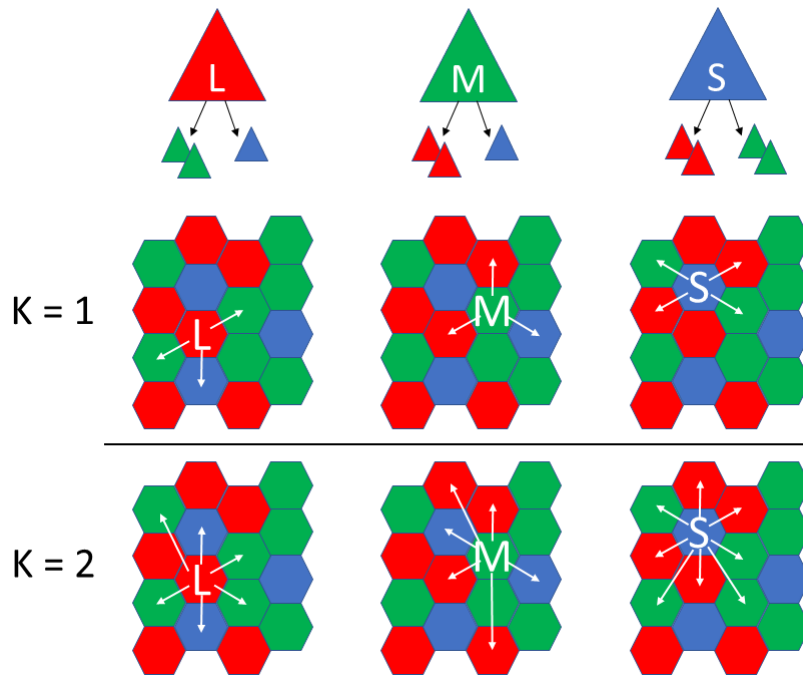
Figure 7.1: The schema for *k-search* color reconstruction: the red triangles represent the L-cones, the green triangles represent the M-cones and the blue triangles represents the S-cones (top row). The search is performed spatially, with each cone looking for the k closest S cones and the 2k closest M or L cones, according to the class of the corresponding cone (middle and bottom row).

The model stores for each type of cone only the corresponding RGB channel value. The red channel is associated to the L-cones, the green channel is associated with M-cones and the blue channel is associated to the S-cones. This way, the function that reconstructs the input image without sharing color information, the *rgb cone color* function, will represent the average color per region based on the channel associated to the cone type. The other two functions work almost in the same way, the difference is that one will perform an Euclidean Nearest neighbor search, and the other will perform an Euclidean *k-Nearest neighbor* search: the algorithm has in input

an image `img` and an integer `k` and it starts by initializing three lists of indices of cones, one for each cone type. The indices are gathered from the *seeds* vector which contains all the seeds used to generate the Voronoi regions and so the cone mosaic. Then, a list of pixels for storing color information is created which contains the color of all the cones sampled from the image. At this point the algorithm isolates the RGB channels not correlated to the cone class: it will save only the red information for long cones, the green information for medium cones and the blue information for short cones. Now the color reconstruction will start and it will save all the information in the data structure, including the color list. For each cone, it will search for the other different k or $2k$ cones, according to Figure 7.1, averaging the associated channel and adding them to the query cone color in the color list. At the end, all the positions in the color list will be filled and the resulting image can be returned. The search algorithm is analyzed in detail in the Appendix 8.

In this section we present some image results obtained from the color reconstruction tool. We remind that this does not intend to be a cortical representation, since the reconstruction phase is only to have a representable output from the retinal sampling performed by the model at the variation of parameters. It can be considered a simulation and not an explanation of biological processes, since the image reconstruction can only test the predictive capabilities of the visualized output in relation to the photoreceptors' absorption of a visual signal and their spatial distribution. In the following subsections we will present results at the variation of the L to M cone ratio and k parameter in subection 7.2, present a test to investigate issues related to space variance and gaze movements in subection 7.4, perform a test to check the effect of L to M ratio on edges and neutral yellow in subsection 7.3 and perform more tests on visual illusions in subsection 7.5. The input image of every test performed in this section is to be considered as colorimetrically calibrated, since the image result from the tool only is qualitative rather than quantitative.

7.2 Testing L to M cone ratio

The proposed model pipeline was applied to two different test images, a natural one and an image with four big square regions colored in red, green, blue

and yellow. For the spatial sampling we used the same foveal distribution of cones for every image, and every image has its own spatial configuration of cone classes. The test image used with resolution 960 x 631 is presented in Figure 7.2.

Figure 7.2: The test image used, resolution 960 x 631. Image taken from `https://pixabay.com/photos/zinnia-flower-meadow-flowers-4060396/`

Figures 7.3 and 7.4 show the original images and the output representations with varying L to M cones ratio. The output images appear mostly similar, so a qualitative analysis of image data has been performed in the test in subsection 7.3 to determine the impact of the L to M cone ratio on result images.

The images in Figure 7.5 demonstrate the results at the variation of the cone cross-talking parameter k. The natural image has been reconstructed with the same cone seeds with different k values, showing a progressive gradual enhancement of the high frequency components of the image, at the cost of an increased spatial scattering of luminance informations of the image, making it more blurred.
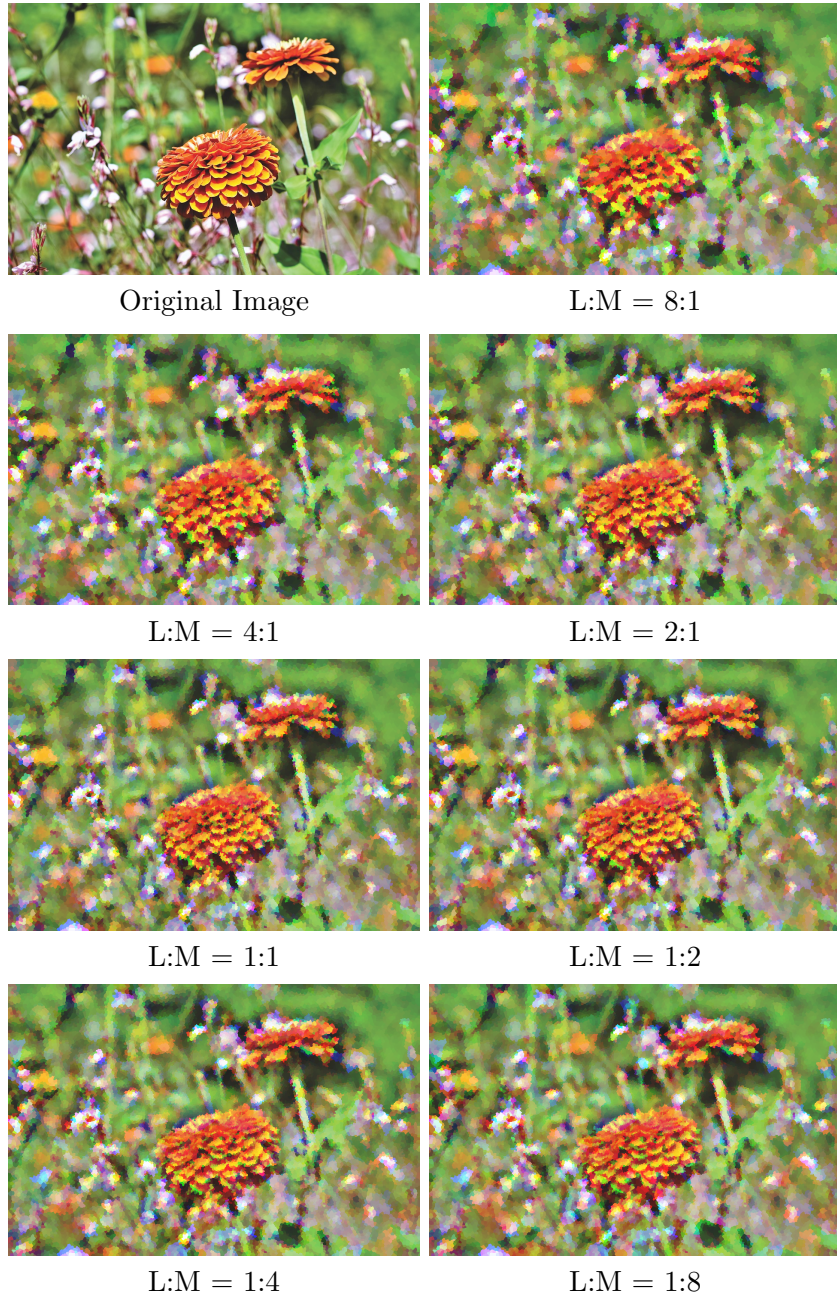
Figure 7.3: Output examples of the original image (top left) with same parameters and varying L:M cone ratio.

| Original Image | L:M = 16:1 | L:M = 8:1 | L:M = 4:1 |
| L:M = 1:1 | L:M = 1:4 | L:M = 1:8 | L:M = 1:16 |
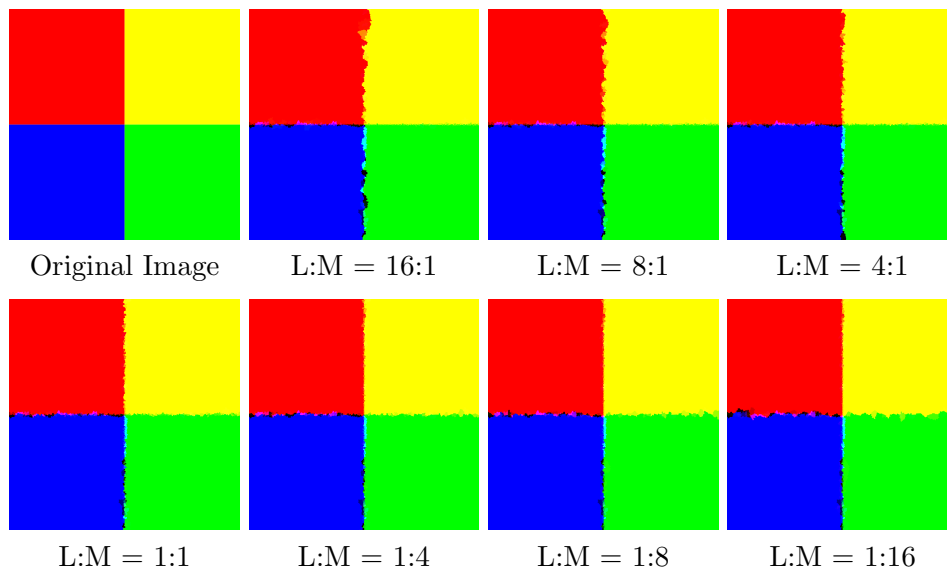
Figure 7.4: Output examples of the original image (top left) with same parameters and varying L:M cone ratio.

| Original Image | $k = 1$ | $k = 2$ |
| $k = 4$ | $k = 8$ | $k = 16$ |

Figure 7.5: Output examples of the original image (top left) with same parameters and varying k.

7.3 Effect of L to M cone ratio on edges

This test is inspired by the work of Brainard et al. [105]. In their experiment, a 0.52 deg circular spot of light was presented to observers with previously characterized L to M cone ratio at the eccentricity of 1 degree of visual angle. The observer had to adjust the wavelength of the test light so that it would be neither green nor red, and subsequently five wavelengths equally spaced around the previously determined test wavelength were presented in 100 trials in random order.

In this test an image of a yellow (RGB 255, 255, 0) spot over a black background is given as the input image to the model. The spot starts at the same size of the experiment in [105], then it decreases by 75%, 50% and 25% (Top row of Figure 7.6. For each of those four images, there will be thirty trials and for each trial the generated cone mosaic will be different, while retaining the same L to M cone ratio. The test is repeated for seven different L to M ratios, from 4:1 to 1:4. For each output image, a dot product normalized with the norm of each image's channel to the original image is performed to understand how similar is the computed image to the original one. With this technique, the result is between 0 and 1. The more it tends to 0, the more the output image and the original image will be dissimilar. On the other side, the more the dot product tends to be 1, the more the two images will be similar.
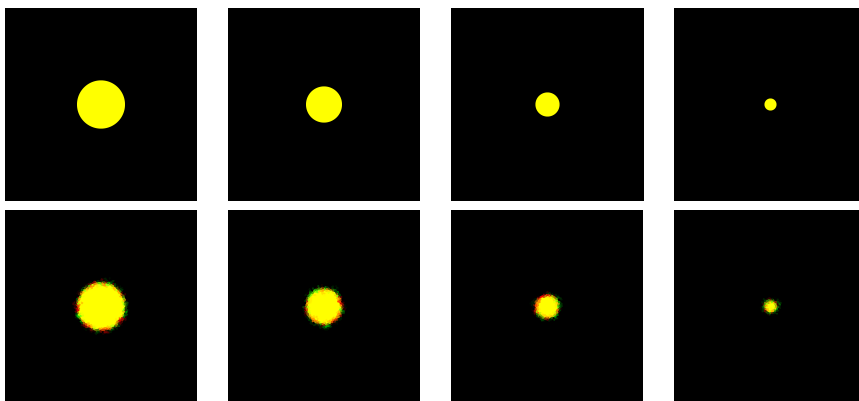
Figure 7.6: The four input images (top row) and four output examples with a distribution L:M of $2:1$ and the same random seed (bottom row).

The results in the bottom row of Figure 7.6 show splotches of red and green localized in the regions sampled around the edges of the circle, caused

by the spatial distribution of the cone mosaic and classes.

L:M	4:1	3:1	2:1	1:1	1:2	1:3	1:4	S.D.
128px	0,9005	0,9024	0,9088	0,9112	0,9070	0,9013	0,8874	0,0078
256px	0,9429	0,9482	0,9522	0,9553	0,9530	0,9485	0,9462	0,0043
384px	0,9635	0,9663	0,9685	0,9703	0,9692	0,9658	0,9631	0,0028
512px	0,9735	0,9758	0,9781	0,9794	0,9783	0,9758	0,9738	0,0023

Table 7.1: Averages of the dot for each L to M ratio and circle size. Last column is the standard deviation for the averages of each circle size.

Table 7.1 shows the averages of the dot product for each L to M ratio and circle size, visualized in the graph in Figure 7.7. It is possible to observe that as the circle becomes bigger, the dot product tends to be higher, meaning that as the circle is sampled by a higher number of cones, the image information comprising the average amount of yellow tends to be more similar to the original image. This is consistent with our visual acuity, the more an object appears bigger in our retinal image, the more details and color information we are able to sense. For each circle size, the best result comes from the more balanced 1:1 ratio, then as the ratios become more unbalanced the resulting dot product tends to gradually decrease. This decrease appears to be faster for the smallest circles, as evidenced by the higher standard deviation in the last column of Table 7.1.
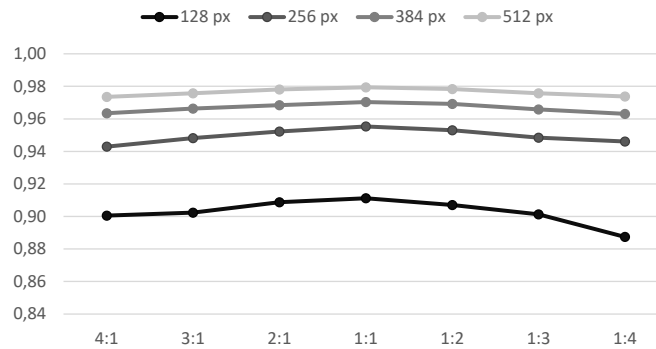
Figure 7.7: Averages of the dot product. Horizontal axis is L:M ratio, vertical axis is the average of dot products.

7.4 Ocular movements

The color reconstruction tool enables to reproduce an image as sensed by the retina. This image has space-variant properties, meaning that the resolution of the image changes across the image itself, biologically represented by the fact that while the distance from the center of the fovea increases, the image is more and more blurred, fitting well with the topography of cones in the retina. Figure 7.8 presents an example of a space-variant sampling performed by the model compared with a uniform sampling with the same number of elements. Color sampling with Voronoi cells is used to obtain this representation as this model is most suitable in terms of spatial description, being very close in shape to cones. Furthermore, each sampling region does not have the full color information over that area.
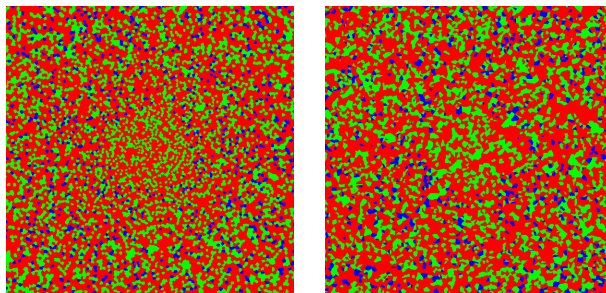
Figure 7.8: Space-variant color sampling performed by the model (left) and the same number of sampling elements with uniform density (right).

While the amount of data required to produce this space-variant image is lower than the original image, a customized algorithm has been necessary to reduce computational time, given the nature of the data structure. This approach is highly valued in foveated imaging systems, where real time computation is a desirable outcome, but in most of the cases the undelying data structure is stored in a regular array, visualized in a log-polar projection or equivalent. In our everyday vision, we are usually not aware of this phenomena since the cortical representation of a scene we are observing is stable even when our eyes or head moves, instead of perceiving the world moving around us. There are two interesting issues related to space variance and gaze:

1. When we shift our gaze in a different direction, we are examining a different region of space with the foveal visual acuity. In the cortical

representation of the scene, the information from the gaze starting point "fades", while the information from the gaze end point starts building up. This is more of a large-scale issue and involves a saccadic movement lasting 20-200 ms depending on the amplitude.

2. During fixations our eyes are never perfectly still, but perturbed by small undetectable ocular motions called *fixational eye movements*. Their role is to contrast the effects of neural adaptation during unvarying stimuli, preventing the effect of perceptual fading, like in the Troxler's effect optical illusion. Tremor is defined as an aperiodic, wave-like motion of the eyes with frequency of 30 to 100 Hz and angular extent of 10-20 seconds of arc, about the diameter of one cone in the fovea [25]. Visual tremor is also reported to be generally independent in both eyes, generating a physical limit to the ability of matching corresponding visual points in the retina during steroscopical vision. Drifts are a movement that happens in conjuction with tremor, resulting in a slow motion of the eye causing the image to move across an extent of 5 minutes of arc. Microsaccades are tiny and unnoticeable jerking movements of the eye occurring during voluntary fixation, usually moving the retinal image across 2 to 25 minutes of arc and happening in about 25 ms [26].

Figure 7.9: (*a*) The original test image. (*b*) Test image sampled with a fovea with 23417 cones. (*c*) average image of the five images obtained with the shifted center of fixation. (*d*) Oversampled image with the center fixation and four different fixations around the center, 110313 cones.

These movements have been theorized to be functional during fixations the same way that saccadic movements are during the exploration of a scene.

To test this theory, it is possible to use the same image in the model with different foveal center coordinates. A shift in our gaze produces a resampling of the image in different points, leading to a more accurate visual information. Two approaches are proposed:

1. Producing different output images with different shifted centers of fixation and averaging the results.

2. Multiplying the number of sampling points within the same sampling window: this leads to an oversampled image, since the amount of information is multiplied accordingly and is spatially more pervasive.

Figure 7.9 shows the simulated images obtained with the two different discussed approaches. Figure 7.9(*a*) presents the original test image used for the simulation. Figure 7.9(*b*) shows the simulated image using a foveal region with 23417 cones. Figure 7.9(*c*) is the average image produced using five

different images obtained with the same foveal sampling of (B) translated orthogonally around the original sampling. The last Figure 7.9(d) is the oversampled image (110313 unique sampling points) using the five fixations at the same time, producing an image with the most natural appearance compared to Figure 7.9(c). This simulation seems to point in the direction that a mechanism involving fixational eye movements and spatial summation could be held accountable for a better perception of a scene [167].

7.5 Testing color reconstruction with visual illusions

Young- Helmholtz's theory of trichromatic vision, introduced in section 2.7, while able to explain phenomena of additive mixture of light, is unable to correctly give an explanation of some visual effects, like the complementary-color afterimage. In 1872, Hering [31] gave an explanation to these effects by proposing his opponent color theory, based on the fact that certain hues are never perceived to appear together, like red and green, or yellow and blue. In the opponent color theory, in fact, information coming from the eyes is divided into three different opponent channels, namely red-green, blue-yellow and black-white, with the last one relative to brightness. Even if this theory can seem to be in contrast with the trichromacy theory, and Helmholtz and Hering were openly in disagreement with each other, they were later combined in order to have a better understanding of how our visual system works, since they are both correct yet they refer to different stages of vision.

The majority of theories on visual illusions usually deal only with assimilation or contrast effects. Left side of Figure 7.10 shows an example of the simultaneous contrast effect, where two identical grey patches presented on different backgrounds appear to be different, in particular the patch on the white background appears slightly darker than the patch on the black background, which appears slightly brighter. Following the opponent theory of color vision, this effect also takes place in a chromatic version (Figure 7.10, right side), where a green background causes the grey patch do have a reddish component, a blue background causes the patch to have a yellowish component, and vice versa.

Opposite to simultaneous contrast is the assimilation effect, first de-
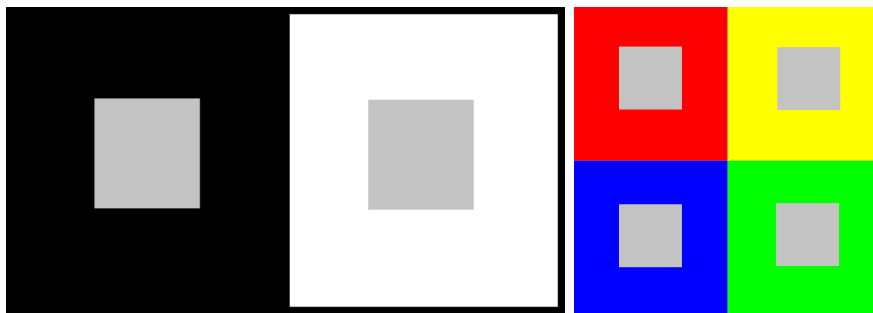
Figure 7.10: Examples of simultaneous contrast. Left: the grey patch on the left side is identical to the corresponding patch on the right, but it appears brighter. Right: simultaneous contrast on chromatic background
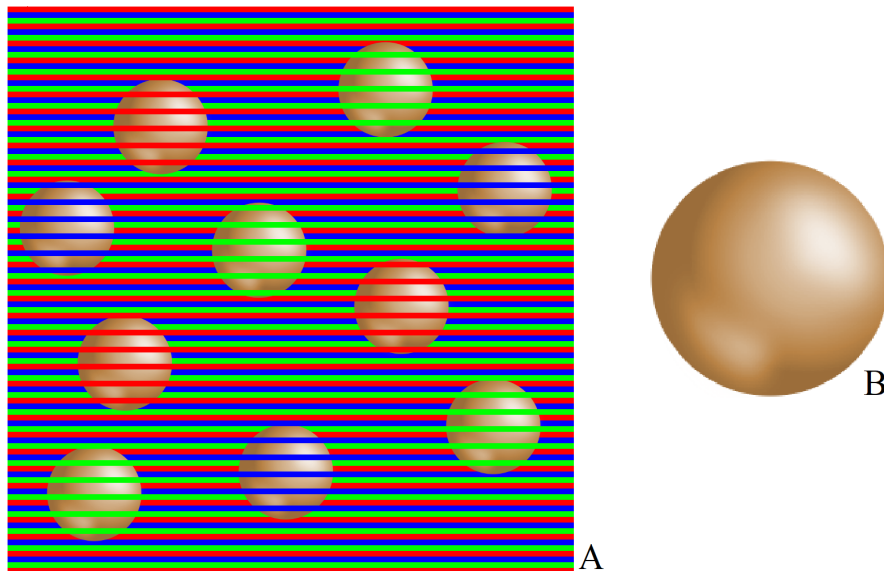
Figure 7.11: Example of color assimilation effect. (**A**) The spheres appear to be yellowish, reddish, and purpleish (**B**) but in in reality they have exactly the same light-brown color.

scribed by Van Bezold in 1876 and illustrated by the example in Figure 7.11. Basically the color of the spheres becomes similar to the grid of lines covering it. While color contrast appears when a color is surrounded by a different color and the differences between colors appear exaggerated, color assimilation happens when two colors are perceptually grouped, and their apparent color difference is reduced [168]. In this section we test the color reconstruction tool with visual illusions concerning color assimilation, color

contrast and fine spatial gratings.

7.5.1 Sphere assimilation illusion

The image in Figure 7.11 shows an example of color assimilation effect. In this subsection the 1000x1000 pixel resolution image will be tested on the model with the variation of L to M cone ratio to determine the impact on result images, the test is performed on seven different L to M ratios, from 8:1 to 1:8.
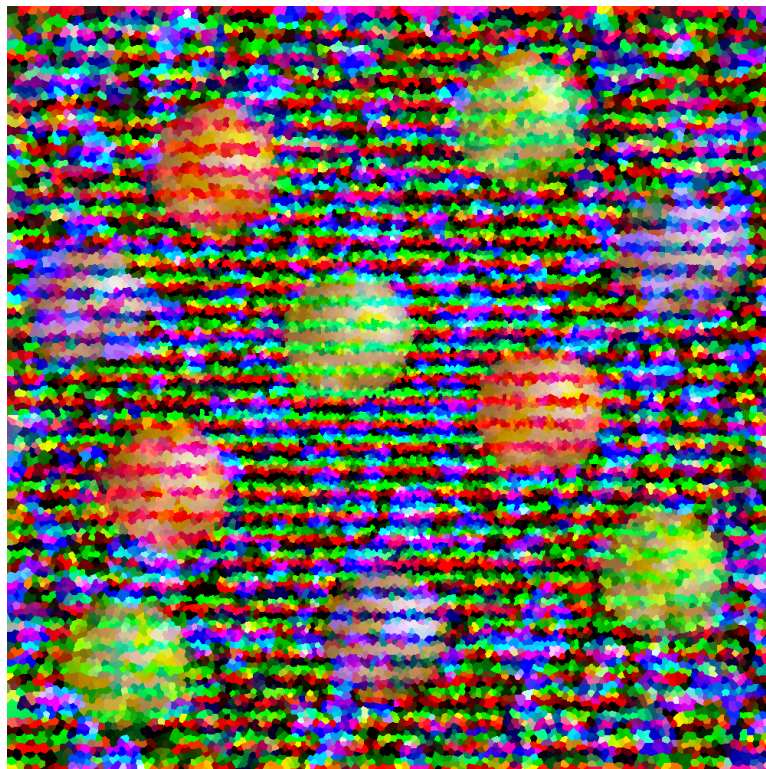
Figure 7.12: Output image obtained from Figure 7.11, with L:M cone ratio of 1:1.

Figure 7.12 shows an example of an output image, obtained with L:M cone ratio 1:1. For each output image, a layer mask has been applied that would cover all the non-sphere region of the image and replace it with transparency, in order to exclude white pixels that are not part of the sphere from the output analysis. The analysis consists in calculating the average RGB value of the sphere surface, comparing it to the original value, to observe

Table 7.2: Average RGB value of the surface of the sphere overlapped with a red line in the assimilation illusion, result images obtained with a variation of L:M ratio. Δ_e is calculated between the base sphere and the corresponding result image.
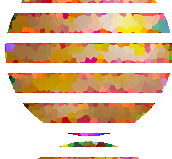
Image	L:M ratio	R	G	B	Δ_e
	Base image	91	72	51	**0**
	8:1	92	52	37	**12.88**
	4:1	93	57	37	**10.33**
	2:1	94	58	37	**10.22**
	1:1	93	61	37	**8.46**
	1:2	94	62	37	**8.53**
	1:4	92	64	37	**7.22**
	1:8	94	64	37	**7.98**

Table 7.3: Average RGB value of the surface of the sphere overlapped with a green line in the assimilation illusion, result images obtained with a variation of L:M ratio. Δ_e is calculated between the base sphere and the corresponding result image.
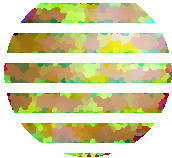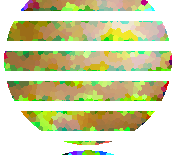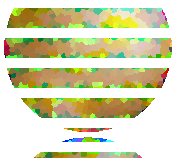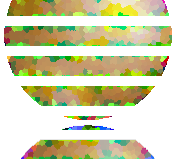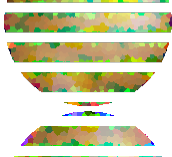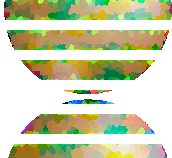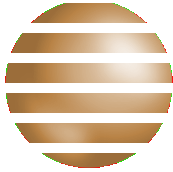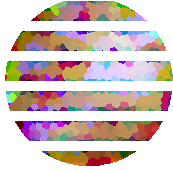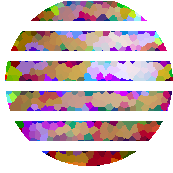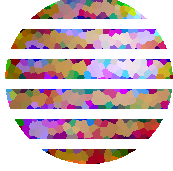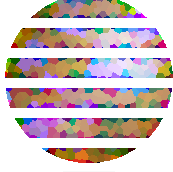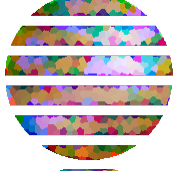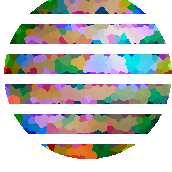
Image	L:M ratio	R	G	B	Δ_e
	Base image	91	72	51	**0**
	8:1	82	81	43	**12.84**
	4:1	81	80	43	**12.41**
	2:1	81	78	43	**10.93**
	1:1	79	78	43	**11.62**
	1:2	78	76	43	**10.55**
	1:4	74	76	43	**12.14**
	1:8	67	76	43	**15.11**

Table 7.4: Average RGB value of the surface of the sphere overlapped with a blue line in the assimilation illusion, result images obtained with a variation of L:M ratio. Δ_e is calculated between the base sphere and the corresponding result image.

Image	L:M ratio	R	G	B	Δ_e
	Base image	94	74	51	**0**
	8:1	81	65	62	**13.16**
	4:1	80	60	62	**16.28**
	2:1	79	58	62	**17.67**
	1:1	77	62	62	**15.47**
	1:2	73	63	62	**15.70**
	1:4	64	63	62	**17.99**
	1:8	64	64	62	**17.66**
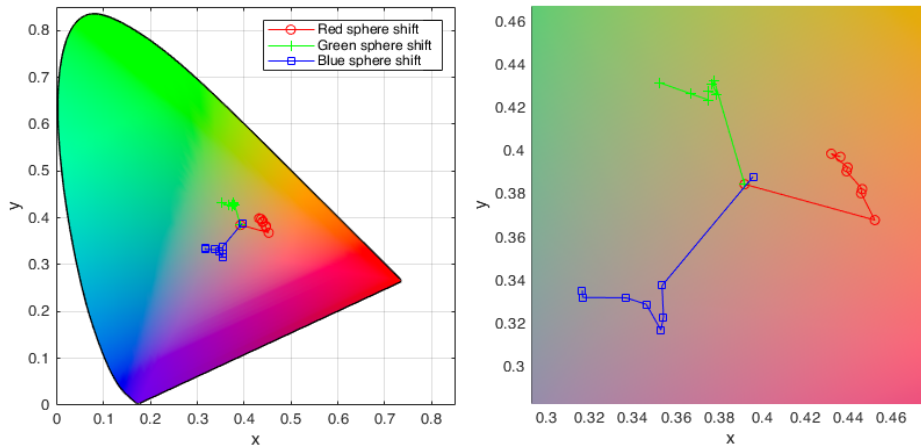
Figure 7.13: CIE 1931 x,y coordinates of the average color for each sphere and each L:M ratio. Each series start from the central original color and shifts according to L:M ratio going from 8:1 to 1:8. Right figure shows a detail of the region of the diagram with the coordinates.

if the assimilation illusion manifests itself in the output image. The difference in color between original and output image has been quantified by the color difference formula CIE76 [169], with Δ_e value corresponding to the Euclidean distance between two colors in CIELAB color space. Results are presented for illusory red, green and blue spheres in tables 7.2 through 7.4. It is possible to observe a consistent behavior in the three different sphere color computations: for each sphere, the RGB channel corresponding to the illusory color shows an increase in the average value, while the other two remaining channels exhibited a decrease in value, thus showing in the resulting image the appearance of color assimilation. In Figure 7.13 each average RGB value has been converted and plotted into a CIE diagram, in order to visualize the shift in color. From the central starting point, it is possible to observe how for each illusory shift in color the coordinates change in the direction of the color that overlaps the sphere.

7.5.2 Chromatic simultaneous constrast

The image in Figure 7.10 (right) presents an example of simultaneous color contrast. In this subsection a the image will be tested on the model with a variation of L to M cone ratio to determine the impact on result images,

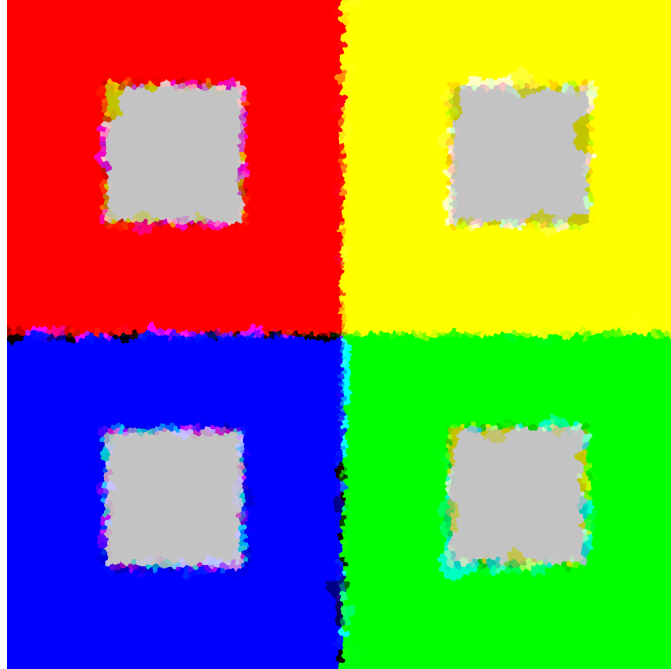the test is performed on seven different L to M ratios, from 8:1 to 1:8.

Figure 7.14: Output image obtained from Figure 7.10 (right), with L:M cone ratio of 1:1.

Figure 7.14 shows an example of an output image, obtained with L:M cone ratio 1:1. For each output image, the four squares have been cropped and analyzed separately by the color surrounding the square. The analysis consists in calculating the average RGB value of the sphere surface, comparing it to the original value, to observe if the simultaneous color contrast illusion manifests itself in the output image. Results are presented for the squares surrounded by red, green, blue and yellow in tables 7.5 and 7.6.

Table 7.5: Grey squares in the result images of the color contrast illusion produced with the color reconstruction tool at the variation of L:M cone ratio. Each column shows squares with a different color background.
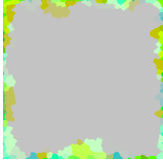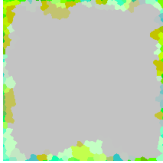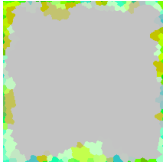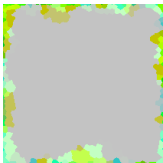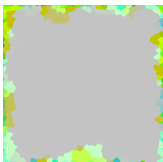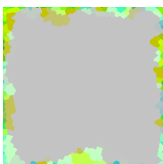
L:M ratio	Red	Green	Blue	Yellow
8:1				
4:1				
2:1				
1:1				
1:2				
1:4				
1:8				

Table 7.6: Average RGB values of the result grey squares in Table 7.5 at the variation of L:M cone ratio.

L:M ratio	Red background	Green background	Blue background	Yellow background
Original image	R = 195 G = 195 B = 195	R = 195 G = 195 B = 195	R = 195 G = 195 B = 195	R = 195 G = 195 B = 195
8:1	R = 196 G = 177 B = 183	R = 190 G = 201 B = 180	R = 190 G = 180 B = 199	R = 196 G = 199 B = 177
4:1	R = 196 G = 182 B = 183	R = 190 G = 198 B = 180	R = 189 G = 185 B = 199	R = 196 G = 197 B = 177
2:1	R = 196 G = 186 B = 183	R = 188 G = 196 B = 180	R = 189 G = 187 B = 199	R = 196 G = 197 B = 177
1:1	R = 197 G = 188 B = 183	R = 187 G = 196 B = 180	R = 188 G = 188 B = 199	R = 196 G = 197 B = 177
1:2	R = 197 G = 189 B = 183	R = 185 G = 196 B = 180	R = 188 G = 189 B = 199	R = 196 G = 196 B = 177
1:4	R = 197 G = 189 B = 183	R = 184 G = 196 B = 180	R = 185 G = 190 B = 199	R = 197 G = 196 B = 177
1:8	R = 199 G = 190 B = 183	R = 181 G = 196 B = 180	R = 180 G = 190 B = 199	R = 198 G = 196 B = 177

Unfortunately, as also evidenced in the graphs in Figure 7.15, in neither square region there has been an increase in the RGB channel value opposite to the background color (with opponents green-red and yellow-blue), and the results are concordant to the ones of the assimilation illusion. Since the model does not account for color opponency, the produced output images do not show any hint of simultaneous contrast.

Figure 7.15: Average RGB value for each channel for the four squares with different colored backgrounds of Table 7.5, with varying L:M cone ratio.

7.5.3 Hamburger image

The image of an hamburger in Figure 7.16 has the characteristic of being composed only of interleaved red and grayscale pixels, so we expect to only see red, white and pink colors inside of it.
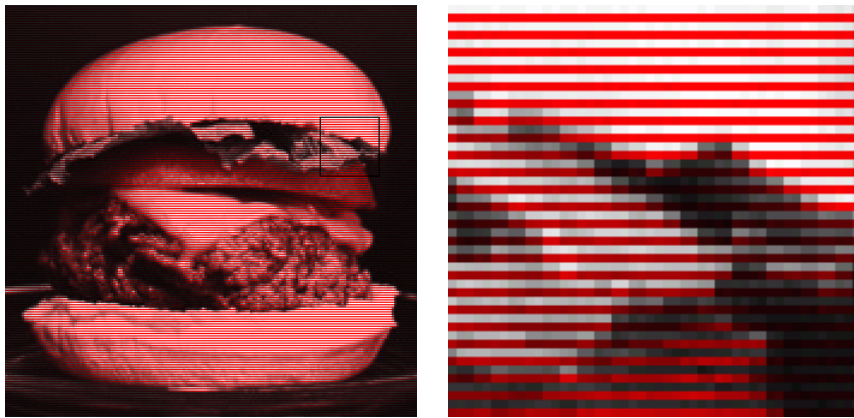
Figure 7.16: The hamburger image, with detail on the right. The image is only composed of red and grayscale pixels. Image courtesy of Chris Taylor https://t-a-y-l-o-r.com/land/

$k = 2$

$k = 4$

$k = 8$

$k = 16$
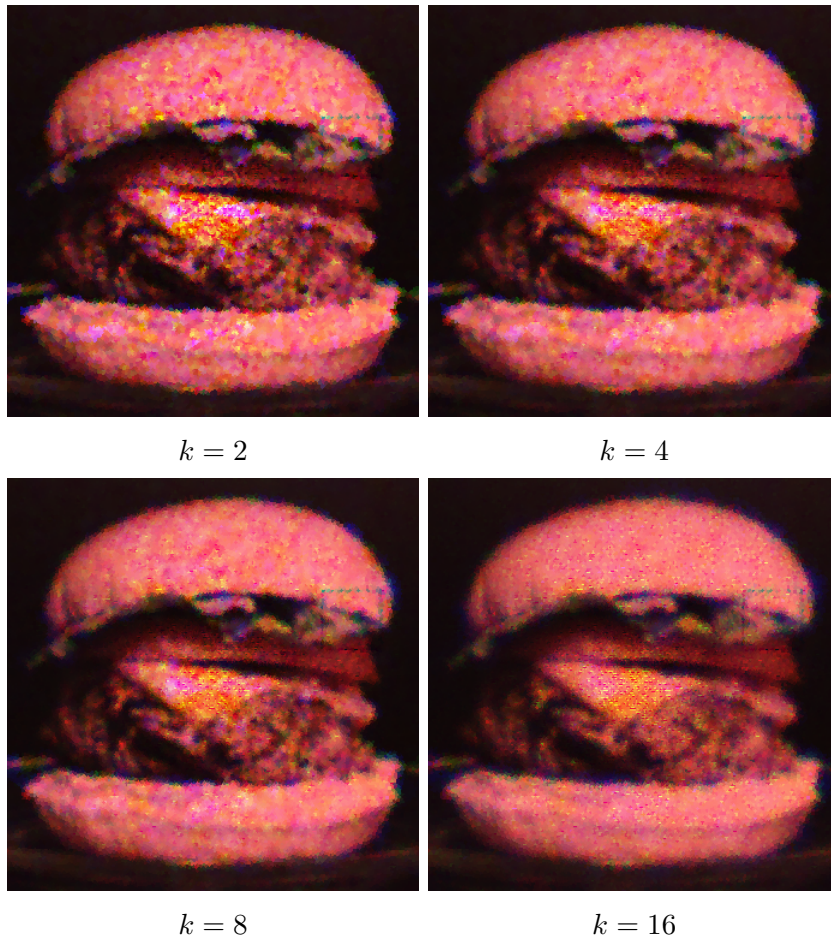
Figure 7.17: Image results of computation of the hamburger image with different k value.

However, to an human observer, this image seems to contain other colors beside red, namely yellow from the cheese, brown from the meat and green from the lettuce. This effect has been first demonstrated by Edwin Land in his experiments on color vision [170] by using two projectors, one with a black and white image and the other with the same image but applying a red filter in front of it.

$k = 2$ $k = 4$

$k = 8$ $k = 16$
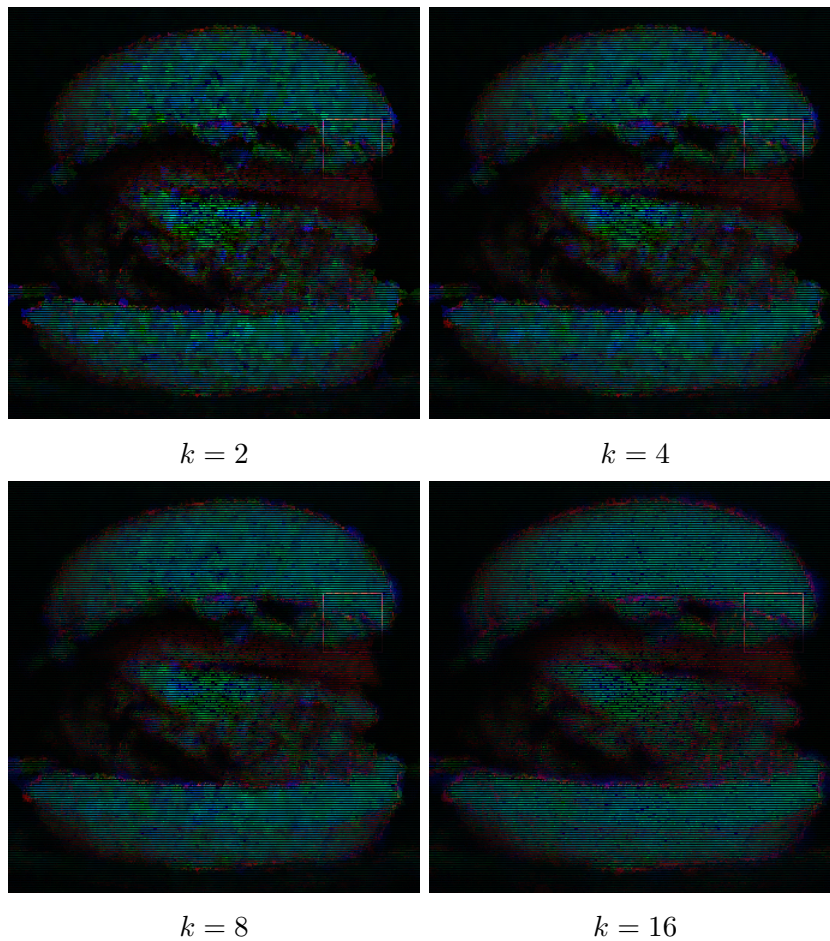
Figure 7.18: Image difference between the computed and original hamburger image with different k value.

Figure 7.17 shows the image results for the model processing on the image. For this computation we used a model fovea of 2 visual degrees with 23417 cone seeds and different k values. It is possible to see that while the image still tends to have a heavy red color cast, splotches of the correct colors appear on their respective regions, like yellow on the cheese and green

Original $k = 2$ $k = 4$ $k = 8$ $k = 16$
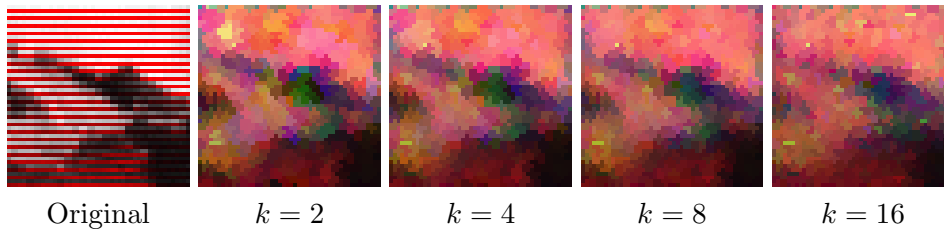
Figure 7.19: Detail of the salad region of the hamburger image computed with different k values.

on the salad. Further details can be found in image 7.18, showing image subtraction between the computed and the original image. The difference has been computed pixel-wise, given a computed and original pixel, each representing an RGB vector $C = [r_1, g_1, b_1]$ and $O = [r_2, g_2, b_2]$, the pixel obtained is $R = C - O = [r_2 - r_1, g_2 - g_1, b_2 - b_1]$.

Thanks to the image difference, it is possible to see how colors beside red are being added to the original image, generating new colors that were previously illusory, as visible in the details of Figure 7.19 evidencing the presence of green pixels in the region correspoding to the salad and yellow pixels in the region corresponding to the bread.

7.5.4 Illusory blue and yellow

The following Figure 7.20 show image results for computations performed on Akioshi Kitaoka's visual illusions "Illusory Yellow Distortion" and "Illusory Blue Distortion". In the yellow image the words appear to be yellowish, though the image only consists of blue, white and gray pixels, meanwhile on the blue image the words appear to be bluish, but the pixels are only yellow, white and gray. Computation by the model show that although the spatial grating and text become intelligible enough (notice that the parameters used imply a vision roughly of a 2 cm inch image at 57.29cm, hence the loss in quality), it is still possible to perceive the illusory text that is composed by splotches of the illusory color, now correctly revealed in the image.

Tables 7.7 and 7.8 evidence the detail of the same region in the two different illusions, before and after computation by the model. In Table 7.7 the original and computed image are separated into their RGB channels: it is possible to observe the different spatial organization and values of pixels in the blue channel. The same is true in Table 7.8, where the original
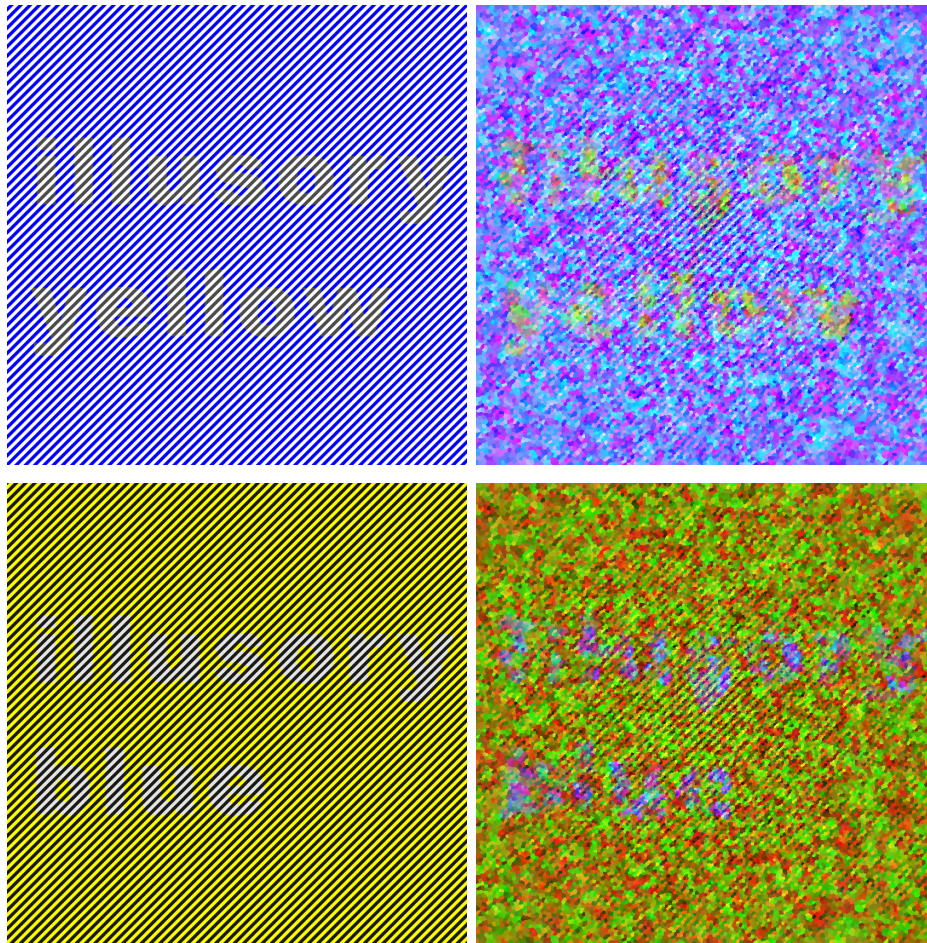
Figure 7.20: Yellow and Blue illusory distortion illusions and their image results after computation by the model. Image courtesy of Akioshi Kitaoka http://www.psy.ritsumei.ac.jp/~akitaoka/.

and computed image are separated into their CMYK channels, in order to evaluate the variation of pixels in the yellow channel. Since the same model fovea of 2 degrees was used ($k = 2$), the illusory color splotches are localized in the same regions in both images. Figure 7.20 presents a fine example of the aliasing caused by the S cone submosaic. The splotchiness of the text in both resulting images is mostly caused by the sampling performed by the highly spatial irregularity of the S cones. Since this sampling process is discrete, every information between cones is not directly sensed: when the spatial frequency of a stimulus is higher than the sampling capabilities of the cone mosaic, it is possible to observe a low-frequency, irregular and

splotchy pattern [95].

Table 7.7: Detail of the illusory blue distortion. Top row is input image, bottom row is output image. The images are separated into their RGB channels.

Detail	R	G	B
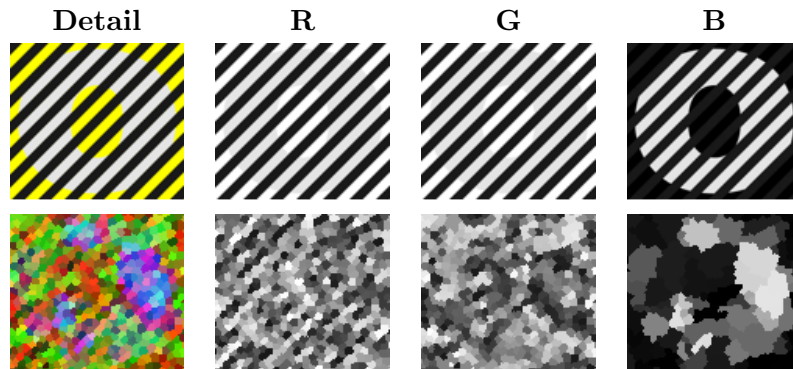
Table 7.8: Detail of the illusory yellow distortion. Top row is input image, bottom row is output image. The images are separated into their CMYK channels.

Detail	C	M	Y	K
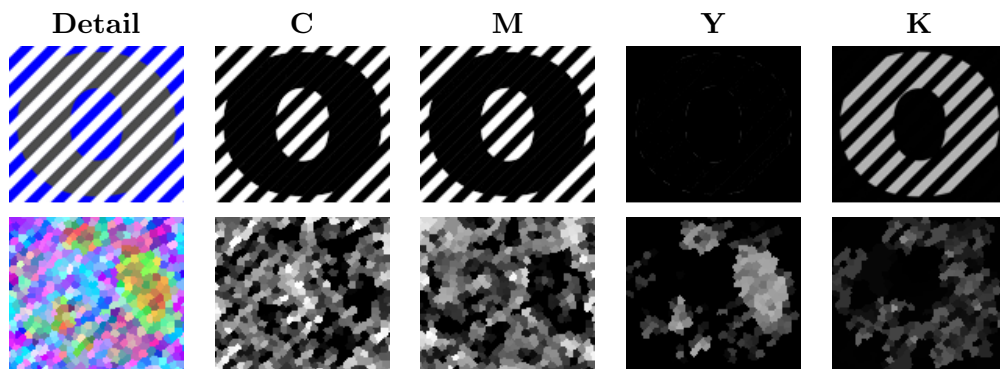
Chapter 8

Conclusions

Color is a fundamental aspect of our visual perception. Luminous visual signals are first neurally encoded by the sampling performed by photoreceptors on the retina, inside our eyes. In chapter 2, an overview is given of the human visual system and the behavior of the processes that lead to the sensation of color. Chapter 3 presented the current state of the art on biologically-inspired foveated imaging systems and how they can be used to model space-variance in the retina. Chapter 4 describes in detail the topography of photoreceptors in the retina, and how their spatial distribution affects the sampling of color signals, with particular attention to the highly inter-subjective variance in the L to M cone ratio. In chapter 5 an investigation on different sampling strategies to reproduce spectral and density characteristics of the cone photoreceptors mosaic has been performed, demonstrating that spatial distribution of points created with blue-noise sampling algorithms possess a high degree of similarity to actual photoreceptor data, and can therefore be used to model local patches of retina. Chapter 6 presents the proposed computational model to investigate how the variable structure of the retina affects the sampling of a scene. The modeled spatial variability accounts not only for the space-variant distribution of photoreceptors in function of their distance from the center of the fovea, but also for the population of the three cone classes according to a user-defined L to M cone ratio parameter. In order to visualize the effects of the parameter variation on a sensed scene, an tool has been developed to perform image reconstruction by means of interpolation of the spatial sampling of cones. This output does not represent the final perceived image,

122

since the model does not take into account any later cortical elaboration performed by the brain, but allows to visualize and evaluate the effects of variations in photoreceptor distribution. The proposed k-Nearest Neighbor algorithm for image reconstruction from samples has been optimized to perform from a computational time of $\Theta(qn)$ dependant on the number of q samples, to a constant time of $\Theta(k)$. The model intends to be a starting point of a modular framework of retinal modeling on which to simulate different behaviors concerning human spatial vision and the detection of color information. In chapter 7 several tests have been performed to evaluate variation of the different model parameters on images and to analyze the performance on visual illusions, demonstrating the model performance on the sensation of color assimilation effect.

Acknowledgments

Thanks to Prof. Giovanni Staurenghi, Prof. Reiner Eschbach and Prof. Christine Curcio for their kindness in accepting to be referees for this dissertation.

Thanks to my advisor Prof. Alessandro Rizzi for his guidance, ideas, remarks and encouragement through the last four years.

Thanks to the colleagues of MIPS Lab Alice Plutino, Fatme Hachem, Elefelious Getachew Belay, Corrado Mio and Simone Liberini for the helpful conversations and support.

Thanks to Prof. Marco Tarini, Dr. Cristian Bonanomi, Luca Giuliani, Dario Ostuni and Prof. Elena Casiraghi for their valuable comments and help.

Thanks to Prof. Victor Ostromoukhov, Dr. David Coeurjolly and Dr. Hélène Perrier of LIRIS lab at the University of Lyon for their scientific collaboration.

Appendix A: k-NN algorithm analysis

In this appendix, describing joint work with Dario Ostuni, the *k-Nearest neighbor* algorithm will be analyzed and discussed.

Concerning the model, the most computational resources intensive task in the whole process is the k-Nearest Neighbor search for image reconstruction. The goal of the k-NN algorithm can be describes as: given a point $p \in \mathbb{R}^2$ and a finite set of points $X \subset \mathbb{R}^2$, find a subset $Y \subseteq X$ such that $|Y| = k$ and $\forall x \in X \setminus Y, \forall y \in Y : |p - y| \leq |p - x|$.

Solving the problem with a *naïve* approach, that is sorting the $n = |X|$ points by their distance to p and then taking the first k as our set Y, it would take $O(n \log(n))$ time. This method can be improved, since only the k nearest elements are needed and the order is unimportant, by using a selection algorithm instead of a sorting algorithm. Since selection can be executed in $O(n)$ time[171], only the distances between p and the elements of X should be generated and then by selecting the first k, it will give a running time of $\Theta(n)$.

Supposing that the *k-NN* problem will be solved for q input points, by using the improved algorithm it will take $\Theta(qn)$ time. Given that for the image reconstruction tool both q and n are in the order of *millions*, a single computation would take time in the order of hours, if not days.

8.1 Algorithm description

In the general case, $\Theta(n)$ is the best possible time, but the data structure of the retinal sampling has a characteristic that can be exploited to make the algorithm faster: locally, sampling regions possess similar enough density d.

125

The updated algorithm works in the following way: first, a regular grid is created with cells with sides of length l and, for each cell, the information of which points of X are contained in that cell is stored: this can be done in $\Theta(n)$ time. Since a uniform density of points has been assumed, each cell $C_{x,y}$ will have an expected number of points contained in it of $\mathbb{E}[|C_{x,y}|] = l^2 d$. Next, when finding the k nearest points to the query point p, the function *search* of algorithm 2 is invoked with arguments p, the query point, and C_{x_0,y_0}, the cell containing p. The function returns Y, which is the answer to the query.

Algorithm 2 *k-NN* search

---

    function SEARCH(p, C_{x_0,y_0})
        $r \leftarrow 0$
        $e \leftarrow \infty$
        Queue $Q \leftarrow \varnothing$
        $Y \leftarrow \varnothing$
        while $lr < e$ **do**
            insert into Q all elements of cells $C_{x,y}$ with $|x - x_0| + |y - y_0| = r$
            $r \leftarrow r + 1$
            while $Q \neq \varnothing$ **do**
                extract the front of Q in q
                insert q in Y
                if $|Y| = k$ **then**
                    $e = l(r+1)\sqrt{2}$
        $Y \leftarrow SELECT(Y, k)$
        return Y

8.2 Proof of correctness

In this subsection it is proven that the algorithm is correct. It is possible to notice that r, during the computation, represents the current lower bound to the distance on the x and y axes of all points that will go in Q. The algorithm cannot terminate until k points have been added to Y, so let's fast forward to the time where $|Y| = k$. At this time the algorithm will have found exactly k points, so f denotes the maximum element in Y. Knowing that $f_x < l(r+1)$ and $f_y < l(r+1)$ and that $\forall a \in Y \setminus \{f\}$ $|p - f| \geq |p - a|$, at this time $k-1$ points which are not farther than f will be found. Since points are being added to Q not by following an order on the distance but rather by

following the order of the x and y coordinates, there could be some point t which is the true k-th nearest point that has not been added yet to Q. How distant can this point possibly be? Knowing that it must be $|p-t| \leq |p-f|$, but $|p-f| < \sqrt{2(l(r+1))^2}$, meaning that $|p-t| < l(r+1)\sqrt{2}$, but since the ending condition has been set exactly to that value, if such a t exists, it will be surely found. Since the k-th nearest point will be found, and since all the other points in Y are not farther than this point, the algorithm terminates with Y containing the k closest points to p, since at the end only the first k points will be selected.

8.3 Expected computational complexity

Let's recall that $\forall x, y \; \mathbb{E}[|C_{x,y}|] = l^2 d$. A function $g(r)$ can be defined that will tell how many points have been visited up to the r-th iteration. It's possible to see that

$$\mathbb{E}\left[g(r)\right] = (2r+1)^2 l^2 d \tag{8.1}$$

Subsequently, how many iterations are needed before at least k points are encountered? Or, in other words, which is the smallest r such that

$$\mathbb{E}\left[g(r)\right] \geq k \tag{8.2}$$

$$(2r+1)^2 l^2 d \geq k \tag{8.3}$$

$$4r^2 + 4r + l^2 d + 1 - k \geq 0 \tag{8.4}$$

$$r = \frac{-2 + \sqrt{4 - (l^2 d + 1 - k)}}{4} \tag{8.5}$$

Since r must be an integer, it becomes

$$r = \left\lceil \frac{-2 + \sqrt{4 - (l^2 d + 1 - k)}}{4} \right\rceil \tag{8.6}$$

When r reaches this value, the algorithm will set

$$e = \left\lceil r\sqrt{2} \right\rceil \tag{8.7}$$

$$e = \left\lceil \left\lceil \frac{-2 + \sqrt{4 - (l^2 d + 1 - k)}}{4} \right\rceil \sqrt{2} \right\rceil \tag{8.8}$$

By setting

$$l = \frac{1}{\sqrt{d}} \tag{8.9}$$

This is obtained

$$e = \left\lceil \left\lceil \frac{-2 + \sqrt{k-2}}{4} \right\rceil \sqrt{2} \right\rceil \tag{8.10}$$

So, it is possible to say that

$$e = \Theta(\sqrt{k}) \tag{8.11}$$

Now, how many operations are being done in the first r iterations? Adding only $f(r)$ points to an array (which takes $\Theta(f(r))$ time) and selecting from k points (which, as previously stated, can be done in $\Theta(k)$ time). This means that the $h(r)$ function, which gives the running time of the algorithm, is

$$h(r) = \Theta(g(r)) + \Theta(k) \tag{8.12}$$

Knowing that at most e iterations will be performed, the expected value obtained is

$$\mathbb{E}[h(r)] = \Theta(\mathbb{E}[g(e)]) + \Theta(k) \tag{8.13}$$

$$\mathbb{E}[h(r)] = \Theta(\mathbb{E}\left[g(\Theta(\sqrt{k}))\right]) + \Theta(k) \tag{8.14}$$

$$\mathbb{E}[h(r)] = \Theta(\left(2\Theta(\sqrt{k}) + 1\right)^2) + \Theta(k) \tag{8.15}$$

$$\mathbb{E}[h(r)] = \Theta(k) \tag{8.16}$$

So it is possible to say that the algorithm has an expected time computational complexity of $\Theta(k)$.

Appendix B: Visual interface for the model

In this appendix, a visual interface for the model is proposed, developed using the Node.js and Express frameworks. The pipeline of the model is divided into 4 steps, each with its own web page. In the first step (Figure 8.1), the sampling array can be generated: at the top of the page it is possible to choose the generation method, each one of them has its own parameters, placed in the square directly on the right side. Note that in every step of the web application, the methods are modular and can be added or modified separately. Once the parameters have been set, the full array is visible on the bottom left square canvas where, using a square selection tool that can be dragged and resized, it is possible to select a sampling window from the generated sampling, highlighted in the square on the right that displays a zoom of the selection. Moreover, on the bottom left it is possible to decide the size of the sampling window and the location of the foveal center, if present. Lastly, in the bottom right, under the zoomed sampling window, a recap of the sampling window attributes is listed.

Figure 8.2 shows the following step where the sampling array becomes a cone distribution that will be used in the next step to sample an image. It is possible to use the basic RGB sampling, where cone absorption is calculated directly from RGB values, or to use a spectral sampling, in which case the spectral absorption curves will be requested. Each selectable method allows to choose the L to M cone ratio, then it is possible to generate the cone distribution that will be visualized on the bottom left, together with a recap of the parameters.

Next, in the third step presented in Figure 8.3, it is possible to upload the image to be sampled by the previously generated cone distribution. If
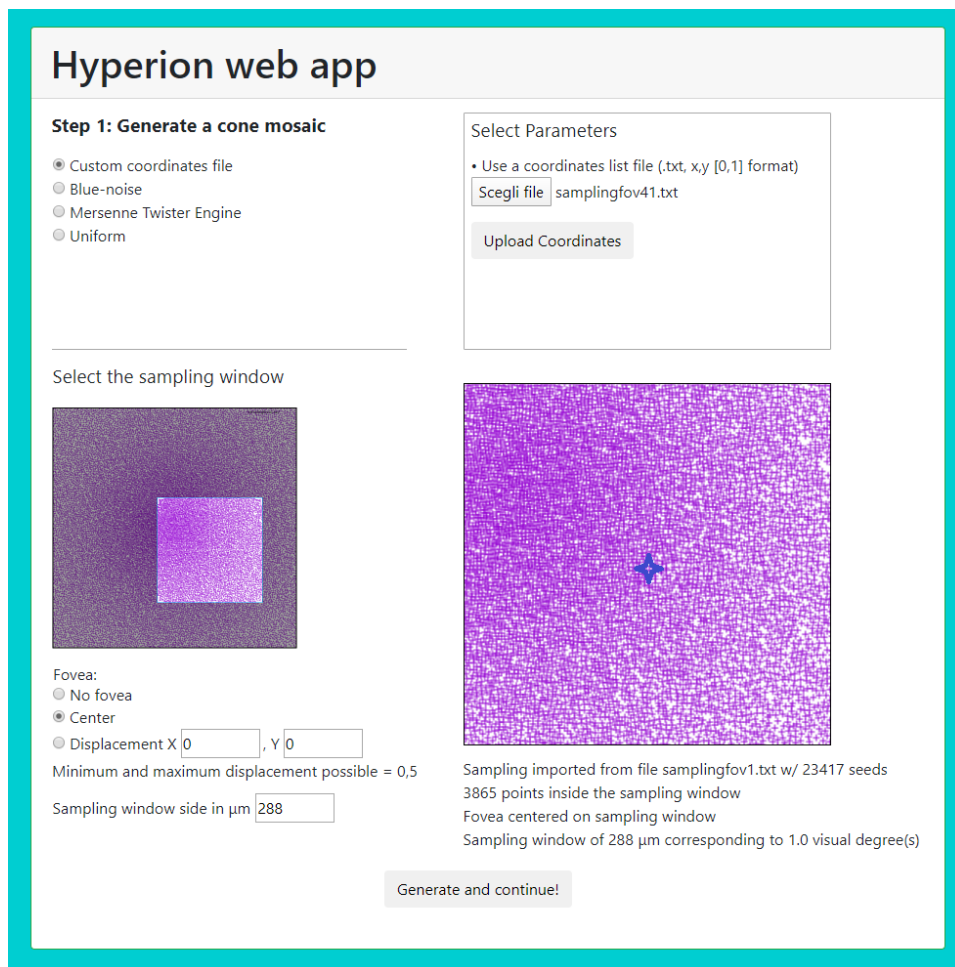
Figure 8.1: First step of the visual interface, the cone array generation.

a spectral sampling method has been selected, it will only be possible to upload an hyperspectral image, otherwise any .png image can be uploaded, better if square in size. Once uploaded, both the original image and the one sampled by the cone distribution will be displayed, together with their histograms of the distribution of pixel values in the RGB channels. At the bottom of the page, the usual recap is displayed, and with the "Continue" button it is possible to proceed to the fourth and final step.

Figure 8.4 displays the last step of the model, namely the color reconstruction. Like the previous steps, it is possible to select a reconstruction method from a list of possible implemented methods. In the example of the figure we are selecting the k-NN reconstruction method: once the value
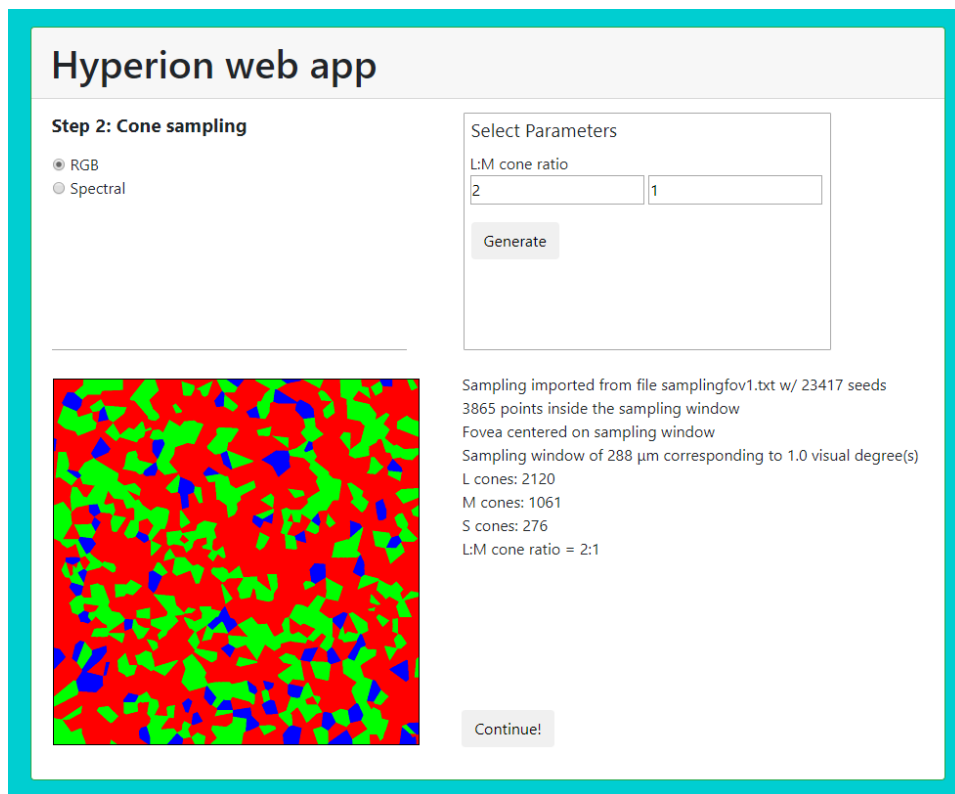
Figure 8.2: Second step of the visual interface, the generation of a cone sampling

of k has been confirmed, all of the image data will be displayed, including the input, the sampled and the reconstructed image, together with their respective histograms to allow an easy comparison. At the bottom, a recap of all the parameters used to produce the final reconstructed image will be displayed, and a button allows to start the web application again.
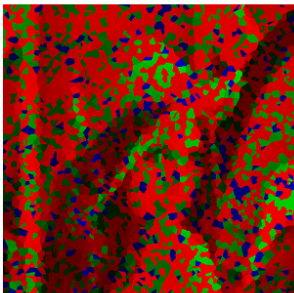
Figure 8.3: Third step of the visual interface, where an image is uploaded and sampled by the cone mosaic.
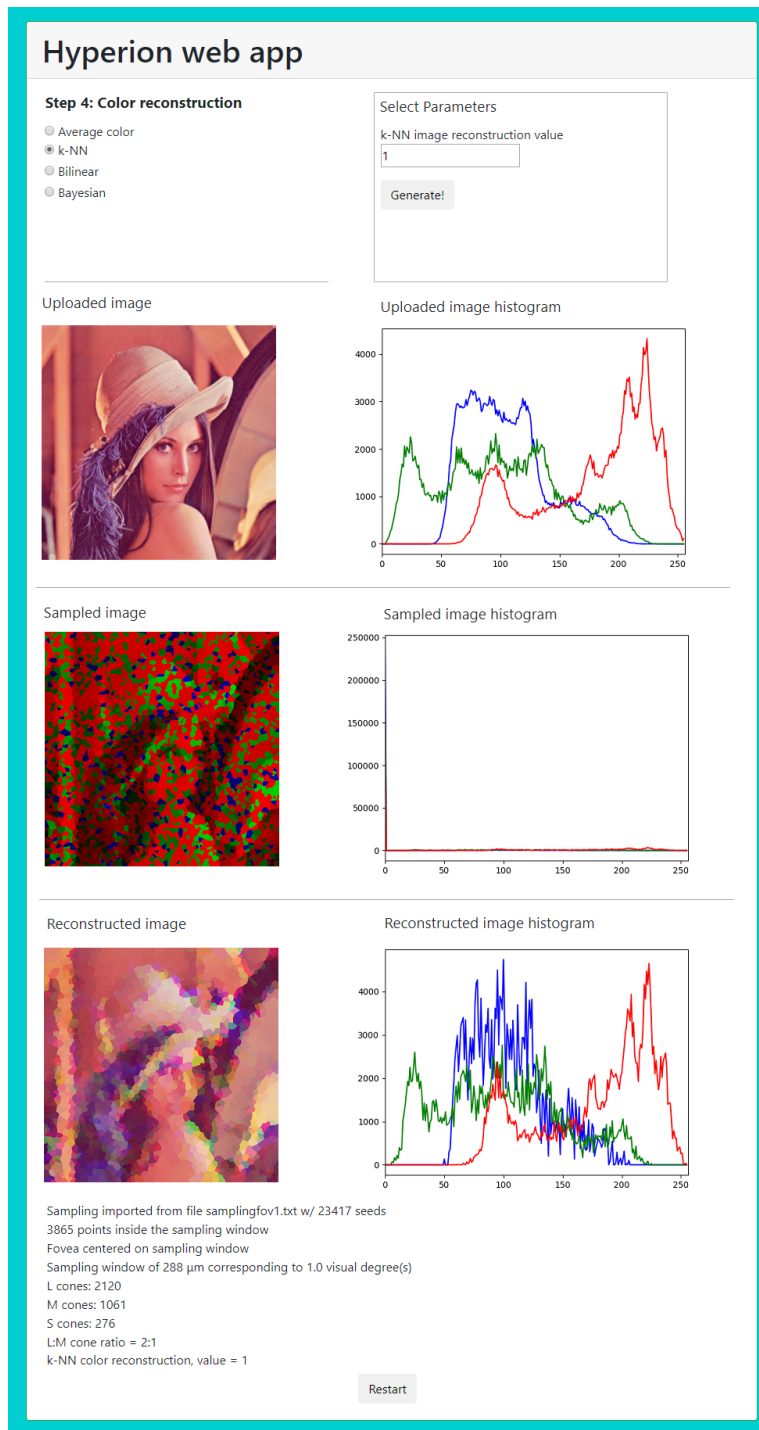
Figure 8.4: Final step of the visual interface, where it is possible to select a color reconstruction method and where final results are reported.

Bibliography

[1] David Brewster. Xxxvii. on the undulations excited in the retina by the action of luminous points and lines. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(3):169–174, 1832.

[2] Austin Roorda and David R Williams. The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520, 1999.

[3] Drew Scoles, Yusufu N Sulai, Christopher S Langlo, Gerald A Fishman, Christine A Curcio, Joseph Carroll, and Alfredo Dubra. In vivo imaging of human cone photoreceptor inner segmentsin vivo imaging of photoreceptor inner segments. *Investigative ophthalmology & visual science*, 55(7):4244–4251, 2014.

[4] Michael F Deering. A human eye retinal cone synthesizer. In *ACM SIGGRAPH 2005 Sketches*, page 128. ACM, 2005.

[5] Adrien Wohrer and Pierre Kornprobst. Virtual retina: a biological retina model and simulator, with contrast gain control. *Journal of computational neuroscience*, 26(2):219–249, 2009.

[6] Christian Morillas, Begoña Pino, Francisco Pelayo, et al. Towards a generic simulation tool of retina models. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 47–57. Springer, 2015.

[7] Christian Morillas, Francisco Pelayo, et al. A conductance-based neuronal network model for color coding in the primate foveal retina. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 63–74. Springer, 2017.

[8] Nicolas P Cottaris, Haomiao Jiang, Xiaomao Ding, Brian A Wandell, and David H Brainard. A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine. *Journal of vision*, 19(4):8–8, 2019.

[9] J Gordon Betts, Peter DeSaix, Eddie Johnson, Jody E Johnson, Oksana Korol, Dean H Kruse, Brandon Poe, James A Wise, Kelly A Young, et al. Anatomy and physiology. 2014.

[10] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, and A James Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[11] Martin Friedlander. Fibrosis and diseases of the eye. *The Journal of clinical investigation*, 117(3):576–586, 2007.

[12] Deane Brewster Judd, Gunter Wyszecki, et al. Color in business, science, and industry. 1975.

[13] Brandon J Lujan, Austin Roorda, Robert W Knighton, and Joseph Carroll. Revealing henle's fiber layer using spectral domain optical coherence tomography. *Investigative ophthalmology & visual science*, 52(3):1486–1492, 2011.

[14] Stephen Lucian Polyak. The retina. 1941.

[15] Brian Wandell and Stephen Thomas. Foundations of vision. *Psyccritiques*, 42(7), 1997.

[16] Ignacio Provencio, Ignacio R Rodriguez, Guisen Jiang, William Pär Hayes, Ernesto F Moreira, and Mark D Rollag. A novel human opsin in the inner retina. *Journal of Neuroscience*, 20(2):600–605, 2000.

[17] Ethan A Rossi and Austin Roorda. The relationship between visual resolution and cone spacing in the human fovea. *Nature neuroscience*, 13(2):156, 2010.

[18] Nicola Quinn, Lajos Csincsik, Erin Flynn, Christine A Curcio, Szilard Kiss, SriniVas R Sadda, Ruth Hogg, Tunde Peto, and Imre Lengyel. The clinical relevance of visualising the peripheral retina. *Progress in retinal and eye research*, 68:83–109, 2019.

[19] Gerald Westheimer. Visual acuity and spatial modulation thresholds. In *Visual psychophysics*, pages 170–187. Springer, 1972.

[20] James Mannos and David Sakrison. The effects of a visual fidelity criterion of the encoding of images. *IEEE transactions on Information Theory*, 20(4):525–536, 1974.

[21] Stanley A Klein and Dennis M Levi. Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation. *JOSA A*, 2(7):1170–1190, 1985.

[22] Matthias H Hennig and Florentin Wörgötter. Eye micro-movements improve stimulus detection beyond the nyquist limit in the peripheral retina. In *Advances in neural information processing systems*, pages 1475–1482, 2004.

[23] Hee-kyoung Ko, Martina Poletti, and Michele Rucci. Microsaccades precisely relocate gaze in a high visual acuity task. *Nature neuroscience*, 13(12):1549, 2010.

[24] HB Barlow. Reconstructing the visual image in space and time. *Nature*, 279(5710):189, 1979.

[25] Lorrin A Riggs, Floyd Ratliff, Janet C Cornsweet, and Tom N Cornsweet. The disappearance of steadily fixated visual test objects. *JOSA*, 43(6):495–501, 1953.

[26] Floyd Ratliff and Lorrin A Riggs. Involuntary motions of the eye during monocular fixation. *Journal of experimental psychology*, 40(6):687, 1950.

[27] Andrew Stockman and Lindsay T Sharpe. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737, 2000.

[28] Andrew Stockman, Donald IA MacLeod, and Nancy E Johnson. Spectral sensitivities of the human cones. *JOSA A*, 10(12):2491–2521, 1993.

[29] Alessandro Rizzi and John J McCann. On the behavior of spatial models of color. In *Color Imaging XII: Processing, Hardcopy, and Ap-*

plications, volume 6493, page 649302. International Society for Optics and Photonics, 2007.

[30] Leo M Hurvich and Dorothea Jameson. An opponent-process theory of color vision. *Psychological review*, 64(6p1):384, 1957.

[31] Ewald Hering. *Zur lehre vom lichtsinne*, volume 68. K. Akademie der Wissenschaften, 1878.

[32] Deane B Judd. Report of us secretariat committee on colorimetry and artificial daylight. *Proceedings of the 12th Session of the CIE, 1951*, 1:11, 1951.

[33] Johannes J Vos. Colorimetric and photometric properties of a 2 fundamental observer. *Color Research & Application*, 3(3):125–128, 1978.

[34] Hermann Grassmann. Zur theorie der farbenmischung. *Annalen der Physik*, 165(5):69–84, 1853.

[35] Lindsay T Sharpe, Andrew Stockman, Herbert Jägle, and Jeremy Nathans. Opsin genes, cone photopigments, color vision, and color blindness. *Color vision: From genes to perception*, pages 3–51, 1999.

[36] Rigmor C Baraas, Joseph Carroll, Karen L Gunther, Mina Chung, David R Williams, David H Foster, and Maureen Neitz. Adaptive optics retinal imaging reveals s-cone dystrophy in tritan color-vision deficiency. *JOSA A*, 24(5):1438–1447, 2007.

[37] Humberto Moreira, Leticia Álvaro, Anna Melnikova, and Julio Lillo. Colorimetry and dichromatic vision. In *Colorimetry and Image Processing*. IntechOpen, 2017.

[38] Stephen J Dain. Clinical colour vision tests. *Clinical and Experimental Optometry*, 87(4-5):276–293, 2004.

[39] Shinobu Ishihara et al. Tests for color blindness. *American Journal of Ophthalmology*, 1(5):376, 1918.

[40] LeGrand H Hardy, Gertrude Rand, and M Catherine Rittler. H–r–r polychromatic plates. *JOSA*, 44(7):509–523, 1954.

[41] Barry L Cole, Ka-Yee Lian, and Carol Lakkis. The new richmond hrr pseudoisochromatic test for colour vision is better than the ishihara test. *Clinical and Experimental Optometry*, 89(2):73–80, 2006.

[42] Israel Dvorine. *Dvorine color perception testing charts*, volume 1. Waverly Press, Incorporated, 1944.

[43] CG Bostrom and Kugelberg I. *Plates for Testing Colour Vision*. Kifa, Stockholm, 1944.

[44] CG Bostrom. *Plates for Testing Colour Vision*. Kifa, Stockholm, 1950.

[45] Dorothy Nickerson, Walter C Granville, and J Opt. The farnsworth-munsell 100-hue and dichotomous tests for color vision. *J. opt*, 1943.

[46] Wilibald A Nagel. I. zwei apparate für die augenärztliche funktionsprüfung. *Ophthalmologica*, 17(3):201–222, 1907.

[47] JA Parker, LA Emerton, et al. Computerized colour vision testing. *Canadian journal of ophthalmology. Journal canadien d'ophtalmologie*, 29(3):125–128, 1994.

[48] Jeff Rabin. Quantification of color vision with cone contrast sensitivity. *Visual neuroscience*, 21(03):483–485, 2004.

[49] Holly G Atkinson. Method and system for color vision deficiency correction, 1996. US Patent 5,589,898.

[50] R Dougherty and A Wade. Daltonize. *Accessed on Oct/06, (http://www.vischeck.com/daltonize)*, 2009.

[51] Hans Brettel, Françoise Viénot, and John D Mollon. Computerized simulation of color appearance for dichromats. *JOSA A*, 14(10):2647–2655, 1997.

[52] Christos-Nikolaos Anagnostopoulos, George Tsekouras, Ioannis Anagnostopoulos, and Christos Kalloniatis. Intelligent modification for the daltonization process of digitized paintings. In *International Conference on Computer Vision Systems*, 2007.

[53] Jia-Bin Huang, Yu-Cheng Tseng, Se-In Wu, and Sheng-Jyh Wang. Information preserving color transformation for protanopia and deuteranopia. *IEEE Signal Processing Letters*, 14(10):711–714, 2007.

138

[54] Giovane R Kuhn, Manuel M Oliveira, and Leandro AF Fernandes. An efficient naturalness-preserving image-recoloring method for dichromats. *IEEE transactions on visualization and computer graphics*, 14(6):1747–1754, 2008.

[55] Gustavo M Machado and Manuel M Oliveira. Real-time temporal-coherent color contrast enhancement for dichromats. In *Computer Graphics Forum*, volume 29, pages 933–942. Wiley Online Library, 2010.

[56] Hiroaki Kotera. Optimal daltonization by spectral shift for dichromatic vision. In *Color and Imaging Conference*, volume 2012, pages 302–308. Society for Imaging Science and Technology, 2012.

[57] Joschua Simon-Liedtke, David R Flatla, and Eskild Narum Bakken. Checklist for daltonization methods: Requirements and characteristics of a good recolouring method. *Electronic Imaging*, 2017(18):21–27, 2017.

[58] Joschua Simon-Liedtke, Ivar Farup, and Reiner Eschbach. On the edge: A scalable daltonization method focusing chromatic edges and contrast. *Electronic Imaging*, 2017(18):28–35, 2017.

[59] Andrew W Schmeder and Donald M McPherson. Multi-band color vision filters and method by lp-optimization, 2013. US Patent App. 14/014,991.

[60] Enrico Tanuwidjaja, Derek Huynh, Kirsten Koa, Calvin Nguyen, Churen Shao, Patrick Torbett, Colleen Emmenegger, and Nadir Weibel. Chroma: a wearable augmented-reality solution for color blindness. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 799–810. ACM, 2014.

[61] Cesar Bandera and Peter D Scott. Foveal machine vision systems. In *Conference Proceedings., IEEE International Conference on Systems, Man and Cybernetics*, pages 596–599. IEEE, 1989.

[62] V Javier Traver and Alexandre Bernardino. A review of log-polar imaging for visual perception in robotics. *Robotics and Autonomous Systems*, 58(4):378–398, 2010.

139

[63] Wilson S Geisler and Jeffrey S Perry. Real-time foveated multiresolution system for low-bandwidth video communication. In *Human vision and electronic imaging III*, volume 3299, pages 294–306. International Society for Optics and Photonics, 1998.

[64] Víctor E Alonso, Rogerio Enríquez-Caldera, and Luis Enrique Sucar. Foveation: an alternative method to simultaneously preserve privacy and information in face images. *Journal of Electronic Imaging*, 26(2):023015, 2017.

[65] Giulio Sandini and Vincenzo Tagliasco. An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14(4):365–372, 1980.

[66] F Robert-Inacio and L Yushchenko. Visual attention model for computer vision. *Biologically Inspired Cognitive Architectures*, 7:26–38, 2014.

[67] Enrico Grosso, Andrea Lagorio, Luca Pulina, and Massimo Tistarelli. Towards practical space-variant based face recognition and authentication. In *2nd International Workshop on Biometrics and Forensics*, pages 1–6. IEEE, 2014.

[68] Alan S Rojer and Eric L Schwartz. Design considerations for a space-variant visual sensor with complex-logarithmic geometry. In *[1990] Proceedings. 10th International Conference on Pattern Recognition*, volume 2, pages 278–285. IEEE, 1990.

[69] Hiroyuki Yamamoto, Yehezkel Yeshurun, and Martin D Levine. An active foveated vision system: Attentional mechanisms and scan path covergence measures. *Computer Vision and Image Understanding*, 63(1):50–65, 1996.

[70] Frédéric Jurie. A new log-polar mapping for space variant imaging.: Application to face detection and tracking. *Pattern Recognition*, 32(5):865–875, 1999.

[71] Andrew B Watson. Detection and recognition of simple spatial forms. In *Physical and biological processing of images*, pages 100–114. Springer, 1983.

[72] Albert J Ahumada Jr and Allen Poirson. Cone sampling array models. *JOSA A*, 4(8):1493–1502, 1987.

[73] Christine A Curcio and Kenneth R Sloan. Packing geometry of human cone photoreceptors: variation with eccentricity and evidence for local anisotropy. *Visual neuroscience*, 9(2):169–180, 1992.

[74] Yi-Zhong Wang, Larry N Thibos, and Arthur Bradley. Modeling the sampling properties of human cone photoreceptor mosaic. In *Vision Science and its Applications*, page FB1. Optical Society of America, 2001.

[75] Chris Bradley, Jared Abrams, and Wilson S Geisler. Retina-v1 model of detectability across the visual field. *Journal of vision*, 14(12):22–22, 2014.

[76] James J Gibson. The ecological approach to the visual perception of pictures. *Leonardo*, 11(3):227–235, 1978.

[77] Jeanny Herault. *Biologically inspired computer vision: fundamentals and applications.* John Wiley & Sons, 2015.

[78] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

[79] Jean Petitot. The neurogeometry of pinwheels as a sub-riemannian contact structure. *Journal of Physiology-Paris*, 97(2-3):265–309, 2003.

[80] William Beaudot, Patricia Palagi, and Jeanny Hérault. Realistic simulation tool for early visual processing including space, time and colour data. In *International Workshop on Artificial Neural Networks*, pages 370–375. Springer, 1993.

[81] Hans van Hateren. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of vision*, 5(4):5–5, 2005.

[82] J Hans Van Hateren. Encoding of high dynamic range video with a model of human cones. *ACM Transactions on Graphics (TOG)*, 25(4):1380–1399, 2006.

141

[83] JH Van Hateren and HP Snippe. Simulating human cones from mid-mesopic up to high-photopic luminances. *Journal of vision*, 7(4):1–1, 2007.

[84] Jeanny Hérault and Barthélémy Durette. Modeling visual perception for image processing. In *International Work-Conference on Artificial Neural Networks*, pages 662–675. Springer, 2007.

[85] Alexandre Benoit, Alice Caplier, Barthélémy Durette, and Jeanny Hérault. Using human visual system modeling for bio-inspired low level image processing. *Computer vision and Image understanding*, 114(7):758–773, 2010.

[86] Wei-Jen Yang, Kuo-Liang Chung, Wei-Ning Yang, and Le-Chung Lin. Universal chroma subsampling strategy for compressing mosaic video sequences with arbitrary rgb color filter arrays in h. 264/avc. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4):591–606, 2013.

[87] Rastislav Lukac and Konstantinos N Plataniotis. Color filter arrays: Design and performance analysis. *IEEE Transactions on Consumer Electronics*, 51(4):1260–1267, 2005.

[88] S Yamanaka. Solid state camera, 1977. US Patent 4 054 906.

[89] Manu Parmar and Stanley J Reeves. A perceptually based design methodology for color filter arrays [image reconstruction]. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, pages iii–473. IEEE, 2004.

[90] Petapixel. http://petapixel.com/2014/01/20/fujifilm-varies-pixel-sizes-proposed-image-sensor-design, April 2016.

[91] James E Adams Jr. Interactions between color plane interpolation and other image processing functions in electronic photography. In *IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*, pages 144–151. International Society for Optics and Photonics, 1995.

[92] David H Brainard et al. Bayesian method for reconstructing color images from trichromatic samples. In *Proceedings of the IS&T 47th Annual Meeting*, pages 375–380, 1994.

[93] David Alleysson, Sabine Susstrunk, and Jeanny Hérault. Linear demosaicing inspired by the human visual system. *IEEE Transactions on Image Processing*, 14(4):439–449, 2005.

[94] H Wassle and HJ Riemann. The mosaic of nerve cells in the mammalian retina. *Proceedings of the Royal Society of London B: Biological Sciences*, 200(1141):441–461, 1978.

[95] John I Yellott. Spectral consequences of photoreceptor sampling in the rhesus retina. *Science*, 221(4608), 1983.

[96] L Galli-Resta, E Novelli, Z Kryger, GH Jacobs, and BE Reese. Modelling the mosaic organization of rod and cone photoreceptors with a minimal-spacing rule. *European Journal of Neuroscience*, 11(4):1461–1469, 1999.

[97] Ares Lagae. Wang tiles in computer graphics. *Synthesis Lectures on Computer Graphics and Animation*, 4(1):1–91, 2009.

[98] Christine A Curcio, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson. Human photoreceptor topography. *Journal of comparative neurology*, 292(4):497–523, 1990.

[99] Wallace B Thoreson and Dennis M Dacey. Diverse cell types, circuits, and mechanisms for color vision in the vertebrate retina. *Physiological reviews*, 99(3):1527–1573, 2019.

[100] Barry B Lee, Paul R Martin, and Ulrike Grünert. Retinal connectivity and primate vision. *Progress in retinal and eye research*, 29(6):622–639, 2010.

[101] Christine A Curcio, Kimberly A Allen, Kenneth R Sloan, Connie L Lerea, James B Hurley, Ingrid B Klock, and Ann H Milam. Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *Journal of Comparative Neurology*, 312(4):610–624, 1991.

143

[102] Anita E Hendrickson and Cristine Yuodelis. The morphological development of the human fovea. *Ophthalmology*, 91(6):603–612, 1984.

[103] Heidi Hofer, Joseph Carroll, Jay Neitz, Maureen Neitz, and David R Williams. Organization of the human trichromatic cone mosaic. *Journal of Neuroscience*, 25(42):9669–9679, 2005.

[104] Heidi Hofer, Ben Singer, and David R Williams. Different sensations from cones with the same photopigment. *Journal of Vision*, 5(5):5–5, 2005.

[105] David H Brainard, Austin Roorda, Yasuki Yamauchi, Jack B Calderone, Andrew Metha, Maureen Neitz, Jay Neitz, David R Williams, and Gerald H Jacobs. Functional consequences of the relative numbers of l and m cones. *JOSA A*, 17(3):607–614, 2000.

[106] Stephanie A Hagstrom, Jay Neitz, and Maureen Neitz. Variations in cone populations for red–green color vision examined by analysis of mrna. *NeuroReport*, 9(9):1963–1967, 1998.

[107] David H Brainard. Color, pattern, and the retinal cone mosaic. *Current Opinion in Behavioral Sciences*, 30:41–47, 2019.

[108] David Alleysson and David Meary. Neurogeometry of color vision. *Journal of Physiology-Paris*, 106(5-6):284–296, 2012.

[109] Elise W Dees, Alfredo Dubra, and Rigmor C Baraas. Variability in parafoveal cone mosaic in normal trichromatic individuals. *Biomedical optics express*, 2(5):1351–1358, 2011.

[110] R. Ulichney. *Digital Halftoning*. MIT Press, 1987.

[111] R. W. Floyd and L. Steinberg. An adaptive algorithm for spatial grey scale. *Proc. Soc. Inf. Display*, 17:75–77, 1976.

[112] Franklin C. Crow. The aliasing problem in computer-generated shaded images. *Commun. ACM*, 20(11):799–805, 1977.

[113] Robert L. Cook. Stochastic sampling in computer graphics. *ACM Trans. Graph.*, 5(1):51–72, 1986.

[114] Don P. Mitchell. Generating antialiased images at low sampling densities. In *ACM SIGGRAPH*, pages 65–72, 1987.

[115] M. McCool and E. Fiume. Hierarchical Poisson disk sampling distributions. In *Proc. Graphics Interface '92*, pages 94–105, May 1992.

[116] Thouis R. Jones. Efficient generation of Poisson-disk sampling patterns. *Journal of Graphics, GPU, & Game Tools*, 11(2):27–36, 2006.

[117] Robert Bridson. Fast Poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH sketches*, 2007.

[118] Manuel N. Gamito and Steve C. Maddock. Accurate multidimensional Poisson-disk sampling. *ACM Trans. Graph.*, 29:8:1–8:19, 2009.

[119] Daniel Dunbar and Greg Humphreys. A spatial data structure for fast Poisson-disk sample generation. *ACM Trans. Graph.*, 25(3):503–508, July 2006.

[120] Mohamed S. Ebeida, Andrew A. Davidson, Anjul Patney, Patrick M. Knupp, Scott A. Mitchell, and John D. Owens. Efficient maximal Poisson-disk sampling. *ACM Trans. Graph.*, 30:49:1–49:12, August 2011.

[121] Li-Yi Wei. Parallel Poisson disk sampling. *ACM Trans. Graph. (SIGGRAPH)*, 27:20:1–20:9, August 2008.

[122] John Bowers, Rui Wang, Li-Yi Wei, and David Maletz. Parallel Poisson disk sampling with spectrum analysis on surfaces. *ACM Trans. Graph.*, 29:166:1–166:10, 2010.

[123] Ying Xiang, Shi-Qing Xin, Qian Sun, and Ying He. Parallel and accurate Poisson disk sampling on arbitrary surfaces. In *SIGGRAPH Asia Sketches*, pages 18:1–18:2, 2011.

[124] Mark A. Z. Dippé and Erling Henry Wold. Antialiasing through stochastic sampling. In *ACM SIGGRAPH*, pages 69–78, July 1985.

[125] D.P. Mitchell. Spectrally optimal sampling for distributed ray tracing. In *Proc. SIGGRAPH '91*, volume 25, pages 157–164, 1991.

[126] P.S. Shirley. Discrepancy as a quality measure for sample distributions. In *Proc. Eurographics '91*, pages 183–194, September 1991.

[127] Ares Lagae and Philip Dutré. A comparison of methods for generating poisson disk distributions. *Computer Graphics Forum*, 27(1):114–129, March 2008.

[128] Michael Balzer, Thomas Schlömer, and Oliver Deussen. Capacity-constrained point distributions: A variant of Lloyd's method. *ACM Trans. Graph.*, 28(3):86:1–8, 2009.

[129] Mohamed S. Ebeida, Scott A. Mitchell, Anjul Patney, Andrew A. Davidson, and John D. Owens. A simple algorithm for maximal poisson-disk sampling in high dimensions. *Comp. Graph. Forum*, 31(2pt4):785–794, May 2012.

[130] Zhonggui Chen, Zhan Yuan, Yi-King Choi, Ligang Liu, and Wenping Wang. Variational blue noise sampling. *IEEE Transactions on Visualization and Computer Graphics*, 18(10):1784–1796, October 2012.

[131] Christian Schmaltz, Pascal Gwosdek, Andreas Bruhn, and Joachim Weickert. Electrostatic halftoning. *Comput. Graph. Forum*, 29(8):2313–2327, 2010.

[132] Thomas Schlömer, Daniel Heck, and Oliver Deussen. Farthest-point optimized point sets with maximized minimum distance. In *Symp. on High Performance Graphics*, pages 135–142, 2011.

[133] Raanan Fattal. Blue-noise point sampling using kernel density model. *ACM Trans. Graph.*, 30(3):48:1–48:12, 2011.

[134] Fernando de Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun. Blue noise through optimal transport. *ACM Trans. Graph.*, 31(6):171:1–171:11, 2012.

[135] Y. Zhou, H. Huang, L.-Y. Wei, and R. Wang. Point sampling with general noise spectrum. *ACM Trans. Graph.*, 31(4):76:1–76:11, 2012.

[136] A. Cengiz Öztireli and Markus Gross. Analysis and synthesis of point distributions based on pair correlation. *ACM Trans. Graph. (Proc. of ACM SIGGRAPH ASIA)*, 31(6):to appear, 2012.

[137] Daniel Heck, Thomas Schlömer, and Oliver Deussen. Blue noise sampling with controlled aliasing. *ACM Trans. Graph.*, 32(3):25:1–25:12, 2013.

[138] Bernhard Reinert, Tobias Ritschel, Hans-Peter Seidel, and Iliyan Georgiev. Projective blue-noise sampling. *Computer Graphics Forum*, 35(1):285–295, 2016.

[139] Victor Ostromoukhov, Charles Donohue, and Pierre-Marc Jodoin. Fast hierarchical importance sampling with blue noise properties. *ACM Trans. Graph.*, 23(3):488–495, 2004.

[140] Johannes Kopf, Daniel Cohen-Or, Oliver Deussen, and Dani Lischinski. Recursive Wang tiles for real-time blue noise. *ACM Trans. Graph.*, 25(3):509–518, 2006.

[141] Victor Ostromoukhov. Sampling with polyominoes. *ACM Trans. Graph.*, 26(3):78:1–78:6, 2007.

[142] Florent Wachtel, Adrien Pilleboue, David Coeurjolly, Katherine Breeden, Gurprit Singh, Gaël Cathelin, Fernando de Goes, Mathieu Desbrun, and Victor Ostromoukhov. Fast tile-based adaptive sampling with user-specified Fourier spectra. *ACM Trans. Graph.*, 33(4), 2014.

[143] Abdalla Ahmed, Hélène Perrier, David Coeurjolly, Victor Ostromoukhov, Jianwei Guo, Hui Huang Dongming Yan, and Oliver Deussen. Low-discrepancy blue noise sampling. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2016)*, 35(6):247:1–247:13, 2016.

[144] A. Ahmed, T. Niese, H. Huang, and O. Deussen. An adaptive point sampler on a regular lattice. *ACM Trans. Graph.*, 36(4):138:1–138:13, jul 2017.

[145] Michael F. Cohen, Jonathan Shade, Stefan Hiller, and Oliver Deussen. Wang tiles for image and texture generation. In *ACM SIGGRAPH*, pages 287–294, 2003.

[146] Ares Lagae and Philip Dutré. An Alternative for Wang Tiles: Colored Edges versus Colored Corners. *ACM Trans. Graph.,*, 25(4):1442–1459, 2006.

[147] Abdalla GM Ahmed, Hui Huang, and Oliver Deussen. Aa patterns for point sets with controlled spectral properties. *ACM Transactions on Graphics (TOG)*, 34(6):212, 2015.

[148] Jost B Jonas, Ulrike Schneider, and Gottfried OH Naumann. Count and density of human retinal photoreceptors. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 230(6):505–510, 1992.

[149] H Gao and JG Hollyfield. Aging of the human retina. differential loss of neurons and retinal pigment epithelial cells. *Investigative ophthalmology & visual science*, 33(1):1–17, 1992.

[150] Tianjiao Zhang, Pooja Godara, Ernesto R Blanco, Russell L Griffin, Xiaolin Wang, Christine A Curcio, and Yuhua Zhang. Variability in human cone topography assessed by adaptive optics scanning laser ophthalmoscopy. *American journal of ophthalmology*, 160(2):290–300, 2015.

[151] Kevin SK Wong, Yifan Jian, Michelle Cua, Stefano Bonora, Robert J Zawadzki, and Marinko V Sarunic. In vivo imaging of human photoreceptor mosaic with wavefront sensorless adaptive optics optical coherence tomography. *Biomedical optics express*, 6(2):580–590, 2015.

[152] Ankit Rohatgi. Webplotdigitizer. *URL http://arohatgi.info/WebPlotDigitizer/app*, 2011.

[153] N Drasdo and CW Fowler. Non-linear projection of the retinal image in a wide-angle schematic eye. *The British journal of ophthalmology*, 58(8):709, 1974.

[154] Gunter Wyszecki and Walter Stanley Stiles. *Color science. Concepts and Methods, Quantitative Data and Formulae, 2nd Edition*, volume 8. Wiley New York, 1982.

[155] Stephen J Eglen. Cellular spacing: Analysis and modelling of retinal mosaics. In *Computational Systems Neurobiology*, pages 365–385. Springer, 2012.

[156] JE Cook. Spatial properties of retinal mosaics: an empirical evaluation of some existing measures. *Visual neuroscience*, 13(1):15–30, 1996.

[157] CA Curcio, C Leigh Millican, KA Allen, and RE Kalina. Aging of the human photoreceptor mosaic: evidence for selective vulnerability of rods in central retina. *Investigative ophthalmology & visual science*, 34(12):3278–3296, 1993.

[158] Yiyi Wang, Nicolas Bensaid, Pavan Tiruveedhula, Jianqiang Ma, Sowmya Ravikumar, and Austin Roorda. Human foveal cone photoreceptor topography and its dependence on eye length. *BioRxiv*, page 589135, 2019.

[159] Jose A Guerrero-Colón, Luis Mancera, and Javier Portilla. Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids. *IEEE Transactions on Image Processing*, 17(1):27–41, 2007.

[160] Yahan Zhou, Haibin Huang, Li-Yi Wei, and Rui Wang. Point sampling with general noise spectrum. *ACM Transactions on Graphics (TOG)*, 31(4):76, 2012.

[161] Riccardo Roveri, A Cengiz Öztireli, and Markus Gross. General point sampling with adaptive density and correlations. In *Computer Graphics Forum*, volume 36, pages 107–117. Wiley Online Library, 2017.

[162] Matteo Paolo Lanaro, Hélène Perrier, David Coeurjolly, Victor Ostromoukhov, and Alessandro Rizzi. Towards human retinal cones spatial distribution modeling. *docs.lib.purdue.edu/modvis/2019/session02/5/*, 2019.

[163] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

[164] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. John Wiley & Sons, 2009.

[165] Edward Nevill Willmer and William David Wright. Colour sensitivity of the fovea centralis. *Nature*, 156(3952):119, 1945.

[166] Kaccie Y Li and Austin Roorda. Automated identification of cone photoreceptors in adaptive optics retinal images. *JOSA A*, 24(5):1358–1363, 2007.

[167] 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, 2004.

[168] Baoxia Liu and Clifton M Schor. Effects of partial occlusion on perceived slant difference. *Journal of vision*, 5(11):4–4, 2005.

[169] Comission Internationale de l'Eclairage. Recommendations on uniform color spaces, color-difference equations, psychometric color terms. *Paris: CIE*, 1978.

[170] Edwin H Land. Color vision and the natural image. part i. *Proceedings of the National Academy of Sciences of the United States of America*, 45(1):115, 1959.

[171] Manuel Blum, Robert W. Floyd, Vaughan Pratt, Ronald L. Rivest, and Robert E. Tarjan. Time bounds for selection. *Journal of Computer and System Sciences*, 7(4):448 – 461, 1973.