Chapter II. The Travels of a Photon: Natural Image Statistics and the Retina

And there was light. Vision starts when photons reflected from objects in the world impinge on the retina. Light is transduced into electrical signals by photoreceptor cells, one of the astounding feats of evolution, rapidly allowing the organism to make inferences about distant objects and events. The structure of the environment plays a critical role in dictating the pattern of connections and responses throughout the visual system and marks the beginning of our journey.

2.1. Natural images are special

Let us consider a digital grayscale image of 100 x 100 pixels, and let us further restrict ourselves to a gray world where each pixel can take 256 shades of gray. Such small colorless image patches constitute a far cry from the complexity of real visual input. Yet, even under these constraints, there is an extremely large number of possible images. There are 256 possible one-pixel images, 256 x 256 possible two-pixel images, etc. All in all, there are $256^{10,000}$ possible 100 x 100 images. This is a pretty big number; there are more of these image patches than the current estimate for the total number of stars in the universe.

Now take a digital camera, a rather old one with a sensor comprising only 100 x 100 pixels, turn the settings to gray images with 8 bits ($2^8=256$), and go around shooting random pictures (Figure II-1). If you are very fast and shoot one picture per second, and if you spend an entire week without sleeping or pausing to eat, just collecting pictures in the city, at the beach, in the forest, or at home, you will have accrued less than a million pictures, a very tiny fraction of a percent of all possible image patches. Yet, you will start to notice rather curious patterns. It turns out that the distribution of

![Figure II-1. Natural images are special. Sixty-four example grayscale patches of 100x100 pixels extracted from photographs. These patches constitute a tiny subset of all possible random 100x100 image patches.](image-url)
natural image patches that you collected in the world tends to have peculiar properties that span an interesting subset of all possible image patches.

In principle, any of the $256^{10,000}$ grayscale patches could show up in the natural world. However, there are strong correlations and constraints in the way natural images look. A particularly striking pattern is that there tends to be a strong correlation between the grayscale intensities of any two adjacent pixels (Figure II-2). In other words, grayscale intensities in natural images typically change in a smooth manner and contain surfaces of approximately uniform intensity. Those surfaces are separated by edges that represent discontinuities, where such correlations between adjacent pixels break, and which tend to be the exception rather than the rule. Edges play a critical role in vision, yet they constitute a small fraction of the image.

One way of quantifying these spatial patterns is to compute the autocorrelation function. To simplify, consider an image in only one dimension. If $f(x)$ denotes the grayscale intensity at position $x$, then the autocorrelation function $A$ measures the average correlation as a function of the separation $\Delta$ between two points:

$$A(\Delta) = \int f(x)f(x-\Delta)dx$$

where the integral goes over the entire image. This definition can be readily extended to more dimensions and colored images. The autocorrelation function of a natural image typically shows a strong peak at small pixel separations followed by a gradual drop (for a review of the properties of natural images, see (Simoncelli and Olshausen, 2001)).

Another way of evaluating the spatial correlations in an image is to compute its power spectrum. Intuitively, one can convert those correlations from the pixel domain into the frequency domain. If there is a lot of power at high frequencies, that implies large changes across small pixel distances as one
might observe when there is an edge. Conversely, a lot of power at low frequencies implies more gradual changes and smoothness in the pixel domain. If $P$ denotes power and $f$ denotes the spatial frequency, natural images typically show that power decreases with $f$ approximately as:

$$P \sim \frac{1}{f^2}$$

There is significantly more power at low frequencies than at high frequencies in natural images. Such a function is called a power law. Power laws are pervasive throughout multiple natural phenomena and have interesting properties. One important property of power laws is scale invariance. If we change the scale of the image, its power spectrum will still have the same shape defined by the equation above.

2.2. Efficient coding by allocating more resources where they are needed

One of the reasons why we are interested in characterizing the properties of natural images is the conjecture that the brain is especially well adapted to represent the real world. This idea, known in the field as the efficient coding principle, posits that the visual system is particularly good at representing the type of variations that occur in Nature. If only a fraction of the $256^{10,000}$ possible image patches are present in any typical image, it may be smart to use most of the neurons to represent the fraction of this space that is occupied. Brain sizes are constrained by evolution and it is tempting to assume that they are not filled with neurons that encode characteristics of images that would never show up. Additionally, brains are extremely expensive from an energetic viewpoint (Laughlin et al., 1998), and therefore it makes sense to allocate more resources where they are needed.

By understanding the structure and properties of natural images, it is possible to generate testable hypothesis about the preferences of neurons representing visual information (Barlow, 1972; Olshausen and Field, 1996; Simoncelli and Olshausen, 2001; Smith and Lewicki, 2006), a topic that we will come back to once we delve into the neural circuitry involved in processing visual information. Such specialization to represent the properties of natural images could arise as a consequence of evolution (Nature) or as a consequence of learning via visual exposure to the world (Nurture). As in other domains of the Nature versus Nurture dilemma, it seems quite likely that both are true. Certain aspects of the visual system are hard-wired, yet visual experience plays a central role in shaping neuronal tuning properties. We will come back to the question of what aspects of the neural circuitry are hard wired and which ones are plastic when we discuss visual cortex (Chapter 5); as an initial guideline for now, a reasonable conjecture is that the amount of plasticity increases as we move up the visual system. According to this conjecture, a lot of the initial processing of visual information discussed in this Chapter is mostly hard wired.

2.3. The visual world is slow
The visual properties of nearby locations in the natural world are correlated. In addition to those spatial correlations, there are also strong temporal constraints in the natural world. Expanding on the collection of natural world photographs, imagine that you go back to the same locations and now collect short videos while keeping the camera still. Because the camera is not allowed to move, the only changes across frames will be dictated by the movement of objects in the natural world. Assuming that you use a camera that captures about 30 frames per second, in most cases, adjacent frames in those videos will look extremely similar. With some exceptions, objects in the world move rather slowly. Consider a cheetah, or a car, moving at a rather impressive speed of 50 miles per hour. Assuming that we have a camera capturing about 40 yards in 2000 pixels, the cheetah will move approximately 30 pixels from one frame to the next. Most objects move at slower speeds. Therefore, the temporal power spectrum of the natural world also shows a peak at low frequencies spanning tens to hundreds of milliseconds. The visual world is slow and mostly continuous.

Several computational models have taken advantage of the continuity of the visual input under natural viewing conditions in order to develop algorithms that can learn about objects and their transformations (Foldiak, 1991; Stringer et al., 2006; Wiskott and Sejnowski, 2002), a theme that we will revisit when discussing computational accounts of learning in the visual system. The notion of using temporal continuity as a constraint for learning is often referred to as the “slowness” principle.

2.4. We are continuously moving our eyes

The assumption that the camera is perfectly still in the previous section is not quite right when considering real brains. To begin with, we can move our heads, therefore changing the information impinging on the eyes. Yet, head movements are also rather sparse and relatively slow. Even with our heads perfectly still, it turns out that humans and other primates are essentially moving their eyes all the time. The observation that the eyes are in almost continuous motion might seem rather counterintuitive. Unless you have reflected rather seriously about this, or spent time scrutinizing another person’s eye movements, introspection might suggest that the visual world around us does not change at all in the absence of external movements or head movements. However, it is dangerous to accept concepts derived from introspection without questioning our assumptions and testing them via experimental measurements.

Nowadays, it is relatively straightforward to measure eye positions quite precisely and rapidly in a laboratory but this was not always the case and physicists built ingenious contraptions to capture these rapid eye movements (Yarbus, 1967 #7771). Figure II-3 shows an example of a sequence of eye movements during presentation of a static image. The eyes typically stay in one location, and then rapidly jump to another location, exploring the new location briefly, before adventuring yet again into a new target. These rapid jumps are denominated visual saccades and typically take a few tens of milliseconds to...
execute from initial position to final position. The positions in between saccades are called fixations.

The pattern of fixations depends on the image, temporal history, and current goals. The characteristics of the image influence eye movements: for example, high contrast regions are more salient and tend to attract eye movements (Itti and Koch, 2001). The temporal history of previous fixations is also relevant: on average, subjects tend to avoid returning to a location they recently fixated on, a phenomenon known as inhibition of return. Current goals also play an important role as well: if you are looking for your car in the parking lot, you will probably be making more fixations on cars, and nearby objects of the same color as your car (Zhang et al., 2018).

During scene perception, subjects typically make saccades of approximately 4 degrees of visual angle. Degrees of visual angle are the most relevant and common unit to measure sizes and positions in the visual field. One degree of visual angle approximately corresponds to the size of your thumb at arm’s length. Under natural scene perception, subjects tend to make saccades approximately every 300 ms (Rayner, 1998).
The intuition that our eyes are mostly still is simply wrong. Why is it that the world does not appear to be jumping from one fixation to the next, several times per second? Watching a movie where the camera moves in a ballistic fashion 3-4 times a second can be quite irksome. The brain takes those retinal inputs that change a couple of times per second and creates the illusion of stability. This is one of many examples that show that our subjective perception of the world is a construct; perception constitutes an interpretation built by our brain based on the incoming sensory information, combined with expectations and with our general knowledge of the world. What we see is not a mere copy of what the eyes dictate.

If we zoom in and play close attention to Figure II-3, we observe that, in addition to the ballistic eye movements spanning several degrees of visual angle and occurring every hundreds of milliseconds, there are also many other smaller and faster eye movements. These involuntary eye movements are called microsaccades and typically span a fraction of a visual degree. Together with other fixational eye movements, these small shifts in eye position may play a critical role in preventing adaptation. In the absence of any type of external movement, head movement or eye movement, neurons quickly adapt to the inputs by reducing their activity. In fact, surprising experiments have shown that if the image on the retina is perfectly stabilized – through an apparatus that is capable of slightly moving the image to account for small eye movements – then the image eventually fades from perception (Coppola and Purves, 1996). In other words, without constant eye movements, we would not be able to see anything except for transient changes due to external movements or head movements.

2.5. The retina extracts information from light

The adventure of visual processing in the brain begins with the conversion of photons into electrical signals in the retina (diminutive form of the word net, in Latin). Due to its accessibility, the retina is the most studied part of the visual system. The conversion of light into electrical signals, combined with the precise retinal circuitry, can well be considered one of the great achievements of evolution. In fact, the ability to read light patterns into information and the structure of the eye made Darwin ponder whether such a feat could be achieved via random selection. Extensive and elegant biochemical work has characterized the signal transduction cascade responsible for capturing
photons, and for using the photon’s energy to trigger a set of chemical reactions that lead to voltage changes in photoreceptor cells (Yau, 1994).

Light information reaches the eye through a lens. When the light reaches the focal plane, the retina, the image is inverted (upside down and left/right, Figure II-4). This basic fact of Optics sometimes puzzles those who reflect about perception for the first time. Why don’t we see everything upside down? Because visual perception constitutes our brain’s construction of the outside world based on the pattern of activity from neurons in the retina. Since the day we are born, our brains learn that a certain pattern of activation in the retina is the way things are in the world. The brain does not know about what is right side up, it’s all electrical signals. In fact, it is possible to teach the brain to adapt to images with different rules, for example, by wearing glasses that invert the image (Stratton, 1896). It is not easy and takes a lot of dedication, but people can learn to ride a bicycle wearing glasses containing lenses that shift the world upside down or glasses that shift the image left and right. After adapting to these new rules, taking the glasses off becomes quite confusing and subjects need to learn again to navigate without those inversions. Upon taking these nasty glasses off, re-learning to adjust to the natural world is much faster than the initial brain training with the reversed world.
The net of neurons in the retina is a particularly beautiful structure that has mesmerized Neuroscientists for more than a century. The history of retinal studies is intimately connected to the history of Neuroscience and commences with the drawings of the famous Santiago Ramon y Cajal. Santiago Ramon y Cajal, considered to be the father of Neuroscience, had a skillful hand for drawing and wanted to become an artist. However, his parents had other plans; Ramon y Cajal ended up following their advice and becoming a medical doctor. After obtaining his medical degree, he studied the techniques to stain neural tissue from the great Camilo Golgi, with whom he would engage in a ferocious scientific dispute about the fundamental structure of brain tissue, and with whom he would share the Nobel Prize in 1906.

The retina soon became a persistent passion for Ramon y Cajal. The retina is located at the back of the eye; it has a thickness of approximately 250 µm and in humans it encompasses the surface area of a little more than half a sphere of 2 cm diameter. The retina is part of the central nervous system: it originates from the same embryonic structures that give rise to the rest of the brain and it has a blood barrier similar to the one in the rest of the brain.

A schematic diagram of the retina is shown in Figure II-5, illustrating the stereotypical connectivity composed of three main cellular layers (Dowling, 2012). In vertebrates, light has to traverse through all the other types of cells to get to the photoreceptors, shown at the top in Figure II-5. Photoreceptors come in two main varieties: rods and cones. There are about $10^8$ rods; these cells are very sensitive and they are particularly specialized for capturing photons under low-light conditions. Night vision depends on rods. Rods are so sensitive that they are able to capture and transmit a single photon, which constitutes about $10^{-19}$ Joules of energy in the visible portion of the spectrum. In fact, recent experiments suggest that sometimes humans are able to detect single photons above chance, which (Tinsley et al., 2016).

In addition to the rod cells, there are about $10^6$ cone cells specialized for vision under bright light conditions. Most people have three types of cones: long-wavelength sensitive peaking at ~564 nm, medium-wavelength sensitive peaking at ~533 nm, and short-wavelength sensitive peaking at ~437 nm. Color vision relies on the activity of cones. About 8 percent of males and less than 1 percent of females show some variation of color blindness, in most cases due to deficiencies or even absence of one of these types of cones, in rare cases absence of more than one type of cone. Even with only two types of cones, people can still see different hues. For example, if a person is missing the short-wavelength cones, they can still distinguish light of 400 nm versus 500 nm wavelength because of the differential responses triggered in the long and medium wavelength sensitive cones. Color blindness is therefore a misnomer and should be reserved only for people who see in grayscale, that is, they are only sensitive to intensity without any color sensation. A condition known as achromatopsia, caused by damage in the brain, not in the eye, can lead to
complete color blindness, as related brilliantly by Oliver Sacks in one of his books \cite{Sacks1997}. People missing one type of coin have specific confusion points, that is, certain combinations of wavelengths that they cannot distinguish. To be able to demonstrate this, it is critical to equalize light intensity. Under natural conditions, colors are often mixed in with different intensities and therefore people with cone deficiencies may use those intensity cues to circumvent their reduced resolution in the color spectrum. The Ishihara test is a common way of assessing color deficiencies and there are plenty of such tests available online. Many people are surprised when they take these tests and find out that they cannot distinguish certain color combinations.

Rods and cones are not uniformly distributed throughout the retina. In particular, there is a special part of the retina, called the fovea, which is specialized for high acuity. This \(\sim300\ \mu \text{m}\) region of the retina contains no rods and a high density of cones, leading to fine sampling of the visual field, thereby providing subjects with higher resolution at the point of fixation. For example, our ability to read depends on the fovea (try fixating on the letter “R” here, make sure that you do not move your eyes, and try to read a word that is 5 words away and two lines above or below the “R”). Cellular density decreases with eccentricity, that is, with distance from the fovea. Because of this non-uniform sampling, we only see in high resolution in the fovea (Figure II-6). Therefore, saccadic eye movements bring the center of fixation into sharp focus to obtain detailed information. People with macular degeneration show progressively more damage in the foveal area, leading to a deterioration of the quality of high-resolution information, eventually perceiving noise or a blurry version of the image in the center Figure II-6.
Despite the fact that locations that are far from the fovea have much coarser sampling, we have the illusion of perceiving approximately equal resolution throughout the visual field. Eye movements are partly responsible for this illusion: every time we move our eyes, we fixate on a new location, which appears in high resolution. We naturally assume that the whole visual field has the same resolution. Additionally, there is probably a stored representation of previous fixations. When we move our eyes to a new location, the old fixation now appears in the periphery, with lower resolution, but it could be combined with a version stored in working memory based on the previous high-resolution fixation.

There is a region in the back of each eye that contains no photoreceptors. This region is where the axons of the retinal output cells, the retinal ganglion cells, exit the eye. People cannot detect light that is focused on exactly this region, which is thus denominated the blind spot. The easiest way to detect the blind spot is to close one eye, fixate on a given distant spot, and slowly move your index finger from the center to the periphery until part of it disappears from view (but not in its entirety which would imply that you moved your finger completely outside of your visual field). There are many nice demos online to help you detect your blind spot. Legend has it that King Charles II of England was fascinated with the blind spot and used to entertain himself by placing the head of a prisoner in his blind spot to mentally picture him headless before the actual decapitation. Under normal circumstances, we are not aware of the blind spot, i.e., we have the subjective feeling that we can see the entire field in front of us (even with one eye closed). Given that we do not normally perceive the blind spot, one may surmise that it is actually rather small. However, you can fit the projection of nine full moons in the sky into the blind spot. How is it possible to be so completely oblivious to such a large empty region of the visual field? We are not normally aware of the blind spot because the brain fills in and compensates for the lack of receptors in the blind spot. This fill-in process emphasizes one more time the notion that our visual percepts are not a literal reflection of reality but rather a reconstruction concocted by our brains. We will return to the notion of vision as a subjective construction when we discuss visual consciousness.

Information from the photoreceptors is conveyed to a second cellular layer consisting of horizontal cells, bipolar cells and amacrine cells, and finally to retinal ganglion cells. The human retina contains approximately 100 million photoreceptors (~4.5 million cone cells, and ~92 million photoreceptors), and about one million retinal ganglion cells. Thus, on average there is a convergence of about 100 photoreceptors to one ganglion cell but these numbers change a lot depending on the location in the retina. As noted above, convergence is much smaller in the fovea and much larger in the distant periphery.

Figure II-5 shows a schematic of the connectivity and information flow in the retina from photoreceptors to horizontal and bipolar cells, then onto to amacrine cells and ganglion cells. Molecular and anatomical markers have
helped define multiple types of horizontal cells, of bipolar cells, and even more types of amacrine cells and ganglion cells, each of which is involved in specific computations to capture different aspects of the incoming images (Masland, 2012). Furthermore, serial electron microscopy is beginning to elucidate the retinal connectome, that is, the precise pattern of synaptic connections in the retina (Helmstaedter et al., 2013). In the not too distant future, it is conceivable that we may have access to a rather complete anatomical map of the retina.

2.6. It takes time for information to reach the optic nerve

At first glance, vision may seem to be instantaneous. We open our eyes, and the world emerges rapidly in all its glory. However, there is no such thing as instantaneous signal propagation. It takes time for the cascade of processes that converts incoming photons into the spiking activity of retinal ganglion cells. The latency of retinal ganglion cell responses to a stimulus flash depends on multiple factors including previous history of visual stimulation, the intensity of the stimulus flash, its size, its color, etc. On average, it takes approximately 40 ms for spikes to emerge from the optic nerve and propagate down to the rest of the brain (Levick, 1973). These latencies are further combined with the computational time required to interpret the information in the brain, to be elaborated upon in Chapters 5 and 6. Because of these delays, and despite our erroneous intuitions, what we see reflects what transpired in the world in the past. The delays are sufficiently short to trick our perception and allow us to get a rapid assessment of what happens in the world.

2.7. Visual neurons respond to a specific region within the visual field

Figure II-7. Neurons have localized receptive fields. A. A light stimulus is flashed in a restricted location while recording the activity of a neuron in a fixating animal (“X” denotes the fixation location). B. The procedure is repeated in multiple different locations. The small vertical ticks denote neuronal activity. The location of maximum activity (blue) denotes the neuron’s receptive field. The stimulus size is also changed to map the boundaries of the receptive field. The neuron also shows a low spontaneous rate at other locations.
Retinal ganglion cells provide the output of the retina; their axons form the optic nerve that carries visual signals to the rest of the brain. The functional properties of ganglion cells have been extensively examined by electrophysiological recordings that go back to the prominent work of Hartline (Hartline, 1938), Barlow (Barlow, 1953), and Kuffler (Kuffler, 1953). Retinal neurons (as well as most neurons in visual cortex) respond most strongly to a circumscribed region of the visual field called the receptive field (Figure II-7). The receptive field is mapped by flashing a stimulus at different locations and different sizes to locate the areas that trigger maximum neuronal activation. Neurons tend to also fire spontaneously at a certain rate, so that there can be small neuronal responses even when the stimulus is very far from the receptive field (i.e., neuronal firing rates are graded and they are not completely zero in the absence of visual stimulation inside the receptive field). These receptive fields tile the entire visual field. Without moving your eyes, any location in the visual field where you can see anything implies that there is a retinal ganglion cell with a receptive field that encompasses that location. The receptive fields of retinal ganglion cells are topographically organized, that is, nearby retinal ganglion cells in the retina represent nearby visual field locations. The non-uniform distribution of neurons from the fovea to the periphery means that there is a consistent eccentricity dependence in the size of the receptive fields. In the fovea there is a one-to-one mapping between cones and retinal ganglion cells. Receptive fields near the fovea are smallest and grow in size as an approximately linear function of eccentricity. The large receptive fields in the periphery are one of the main reasons why we have less resolution outside of the fovea.

The retinal ganglion cell schematically illustrated in Figure II-7 increases its firing rate with increased luminance inside the receptive field. This type of cell is referred to as an on-center cell. There are also other retinal ganglion cells, off-center cells, which increase their firing rate when there is a decrease in luminance in the center of their receptive fields.

2.8. The difference-of-Gaussians operator extracts salient information and discards uniform surfaces

Even when the center of an on-center cell is bombarded with a high-luminance flash, its response will be modulated by what is outside of the receptive field. In particular, a perfectly uniform high-luminance white wall will not trigger high activation for retinal ganglion cells. Consider the following experiment: a small uniform white circle is shown in the center of the receptive field and the neuron fires above baseline levels. The circle is enlarged and the neuron shows a higher firing rate. At some point, however, making the circle larger leads to a reduction in firing rate. This phenomenon is known as surround inhibition. Surround inhibition is observed not only for retinal ganglion cells but it is actually prevalent throughout the entire visual system. On-center neurons are particularly interested in spatial changes, i.e., increased luminance within the
receptive field combined with decreased luminance outside the receptive field. The converse is true for off-center neurons.

This form of spatial context-dependent response pattern is known as center-surround receptive fields and is typically modeled as a difference of two Gaussian curves (Figure II-8). Considering an on-center cell, and assuming that the center of the receptive field is at location $x=0$, $y=0$, the response at position $x$, $y$ will be driven by an excitatory component proportional to $\left(\frac{1}{2\pi \sigma_{cen}^2}\right)e^{-\frac{(x^2+y^2)}{2\sigma_{cen}^2}}$, where $\sigma_{cen}$ reflects the spatial extent of the excitatory driving force (red curve in Figure II-8). This excitation is counterbalanced by the surround inhibitory component given by $\left(\frac{1}{2\pi \sigma_{sur}^2}\right)e^{-\frac{(x^2+y^2)}{2\sigma_{sur}^2}}$, where $\sigma_{sur}$ reflects the spatial extent of the inhibitory driving force (blue curve in Figure II-8). The difference-of-Gaussians operator is used to describe the receptive field structure of retinal ganglion cells:

$$D(x,y) = \pm \frac{1}{2\pi \sigma_{cen}^2} e^{-\frac{x^2+y^2}{2\sigma_{cen}^2}} - \frac{B}{2\pi \sigma_{sur}^2} e^{-\frac{x^2+y^2}{2\sigma_{sur}^2}}$$

where the scaling factor $B<1$ controls the relative strength of excitation and inhibition, where $\sigma_{sur} > \sigma_{cen}$, and where the ± corresponds to the on-center and off-center cells, respectively. The difference between the two terms yields a "Mexican-hat" structure with a peak in the center and an inhibitory dip in the surround.

2.9. Visual neurons show transient responses

In the same way that a large spatially uniform stimulus does not elicit strong activation because neurons are tuned to detect spatial changes, a temporally constant stimulus leads to transient responses. Neurons respond to the onset of the stimulus as well as to the offset of the stimulus, but their responses tend to rapidly adapt when the stimulus remains constant and without any other changes.
(in the absence of head movements or eye movements). Some neurons maintain a tonic response above baseline during the duration of the stimulus after the initial transient while in other neurons the firing rate decreases to baseline after the initial transient. Visual neurons are therefore sensitive not only to spatial context but also temporal context. The incorporation of contextual information allows neurons to efficiently encode transient changes without spending abundant spikes to indicate that the stimulus is constant in space or time.

Equation 2.1 can be expanded to provide a quantitative description of the dynamic responses of retinal ganglion cells to a constant stimulus:

$$D(x,y,t) = \pm \left( \frac{D_{\text{cen}}(t)}{2\pi\sigma_{\text{cen}}^2} \exp \left[ -\frac{x^2 + y^2}{2\sigma_{\text{cen}}^2} \right] - \frac{BD_{\text{sur}}(t)}{2\pi\sigma_{\text{sur}}^2} \exp \left[ -\frac{x^2 + y^2}{2\sigma_{\text{sur}}^2} \right] \right)$$

Equation 2.2

where $D_{\text{cen}}(t) = \alpha_{\text{cen}}^2 t e^{-\alpha_{\text{cen}} t} - \beta_{\text{cen}}^2 t e^{-\beta_{\text{cen}} t}$ describes the dynamics of the center excitatory function and $D_{\text{sur}}(t) = \alpha_{\text{sur}}^2 t e^{-\alpha_{\text{sur}} t} - \beta_{\text{sur}}^2 t e^{-\beta_{\text{sur}} t}$ describes the dynamics of the surround inhibitory function (Dayan and Abbott, 2001; Wandell, 1995).

Equation 2.2 describes the internal dynamics of a retinal ganglion cell upon presentation of a stimulus flash that remains constant. In addition to this type of responses, some ganglion cells are also strongly activated by a stimulus that moves within the receptive field. One such type of cell is the on-off direction selective retinal ganglion cell, which shows enhanced responses when a stimulus within the receptive field is moving in a specific direction. Such direction-selective responses are also modulated by the context: neurons will respond most vigorously when there is a difference in direction of motion between the receptive field and the surround, such that an entire visual field moving in the same direction is a very poor stimulus for this cell. This contextual subtraction enables the neurons to distinguish the movement of external objects from self-motion. Motion sensitive retinal ganglion cells tend to have large dendritic arbors and are particularly abundant in the periphery. In fact, detecting an object in the periphery is much easier when it moves, an observation that you can easily test by fixating on any given letter here, extending your hand in the periphery and compare your perception of the hand when it is static versus when it is moving.

The conduction velocities of retinal ganglion cells have also been used to separate between magnocellular cells (M-type ganglion cells) and parvocellular cells (P-type ganglion cells, also called midget cells). M-type cells have large dendritic arbors, fast conduction velocity, respond to low-contrast stimuli, show transient responses, and have little sensitivity for colors. In contrast, P-type cells show small dendritic arbors, have color sensitivity, tend to exhibit more sustained responses, and low conduction velocities. There continues to be exciting research to elucidate all the different types of functional and structural specializations of retinal ganglion cells, current estimates suggest that there are...
at least tens of distinct ganglion cell types, depending on exactly how a "type" is defined. With the exception of the fovea, which is quite specialized, within the periphery different ganglion cell types are approximately distributed throughout so that the same external stimulus features can be captured throughout the visual field.

### 2.10. Onto the rest of the brain

The main output of retinal ganglion cells projects to a part of the thalamus called the lateral geniculate nucleus (LGN). The retina also projects to the suprachiasmatic nucleus and the superior colliculus, among other regions. The suprachiasmatic nucleus plays an important role in regulating circadian rhythms. The superior colliculus constitutes the main visual processing center for many species before the expansion of the cerebral cortex. Yet, primates can recognize objects after lesions to the superior colliculus but not after lesions to visual cortical regions (see (Gross, 1994) for a historical overview). The key visual pathway involves the LGN.

As we will discuss in Chapter 5, there are massive backprojections throughout the visual system. If area A projects to area B, then in most cases area B also projects back to area A. One of the few exceptions to this rule is the connection from the retina to the LGN. There are no connections from the LGN back to the retina.

The thalamus has been often succinctly called the gateway to cortex, modulating what type of sensory information reaches cortex. The receptive fields of LGN cells also display the center-surround structure depicted in Figure II-8 and can be approximated by Equations 2.1 and 2.2. Thalamic cells are often referred to – in a rather unfair fashion – as relay cells, advocating the idea that the thalamus merely copies and pastes the output of the retina. This is likely to be only an oversimplification and the picture will change dramatically as we understand more about the neuronal circuits and computations in the LGN.

Like the vast majority of brain structures, there are two copies of the LGN, one in each hemisphere. The right LGN receives input from both eyes, but only from the left hemifield while the converse is true for the left LGN. The right eye receives information from both hemifields, and sends right hemifield information to the LGN in the left hemisphere, and left hemifield information to the LGN in the right hemisphere. Six layers can be distinguished in the LGN. Layers 2, 3 and 5 receive ipsilateral input (i.e., information from the eye on the same side). Layers 1, 4 and 6 receive contralateral input (i.e., information from the eye on the opposite side). A single point in space is therefore represented in 6 different maps at the level of the LGN and information form the right eye and left eye does not merge in the LGN. Layers 1 and 2 are called magnocellular layers and receive input from M-type ganglion cells. Layers 3-6 are called parvocellular layers and receive input form P-type ganglion cells. There are about 1.5 million cells in the LGN. While we often think of the LGN predominantly in
terms of the input from retinal ganglion cells, there is a large number of back-projections, predominantly from primary visual cortex (Chapter 5), to the LGN (Douglas and Martin, 2004).

2.11. Digital cameras versus the eye

In Chapters 7-9, we will examine computational models of visual processing. By and large, state-of-the-art computer models start with the output of a regular digital camera that has captured a picture and represents it as a two-dimensional matrix of pixels, each one of which is coded in a 3-dimensional color world (such as red, green, and blue intensities). Yet, the eye lenses and the retina go through a pretty sophisticated series of computations that are still not quite matched by even the best digital cameras out there.

The angle of view of a digital camera depends on the type of length. For a focal length of 17mm (approximate distance from the optical center of the eye lens to the retina), the field of view is approximately 90 degrees, whereas the field of view for humans spans almost 180 degrees. The resolution of the human eye has been estimated to be on the order of 500 megapixels, still well above some of the fanciest commercially available digital cameras out there. Digital cameras are approximately uniform in their sensitivity to light, whereas the retina allocates more resources to dark areas. It is not easy to take digital pictures under very low light conditions. To circumvent these challenges, photographers may use contraptions such as tripods to stabilize the camera and leave the camera shutter open for many seconds, if not minutes or more. In contrast, the eye can convey rather accurate information and help us navigate in the forest with minimal moonlight. One of the tricks to achieve this is that the retina can adapt to low light condition and change its own gain to achieve higher sensitivity. The eye has to work under conditions of strong sunlight all the way to moonless nights, a difference of about 9 orders of magnitude in light intensity. Additionally, digital cameras typically lack many of the sophisticated motion detection and contextual correction mechanisms described above for retinal ganglion cells.

2.12. Summary

- Natural images are special: they are spatially smooth and change slowly in time.
- Positions and sizes in the visual field are measured in degrees of visual angle. One degree corresponds approximately to the size of your thumb at arm’s length.
- Humans and other primates make frequent eye movements denominated saccades, spanning multiple degrees of visual angle, and occurring 3-4 times a second.
- Two types of photoreceptors convert light into electrical signals: rods and cones. Rods are responsible for night vision. Cones are responsible for color.

- Retinal ganglion cells communicate the output of the retina to the rest of the brain.

- Retinal ganglion cells respond to a localized region of the visual field denominated the receptive field.

- The center of focus is projected onto the fovea, an area populated by cones, with higher cellular density and smaller receptive field sizes, providing high resolution.

- On-center retinal ganglion cells are excited by light within their receptive field and inhibited by light in the surrounding region. Their responses can be described by a difference-of-gaussians operator.

- Information from retinal ganglion cells is conveyed to the lateral geniculate nucleus.

- As a coarse approximation, the eye can be considered to be a fancy and sophisticated digital camera, though eyes are capable of many sophisticated tricks that current digital cameras cannot perform.

- Perception is a construct, an interpretation made by the brain inspired by sensory formation but not a literal reflection of the outside world.

### 2.13. References


