

1.04 Visual Information Processing

Michael R Ibbotson^{a,b} and Hamish Meffin^{a,c}, ^a National Vision Research Institute, Australian College of Optometry, Carlton, VIC, Australia; ^b Department of Optometry and Vision Sciences, University of Melbourne, Parkville, VIC, Australia; and ^c Department of Biomedical Engineering, Melbourne School of Engineering, University of Melbourne, Parkville, VIC, Australia

© 2020 Elsevier Inc. All rights reserved.

1.04.1	Abstract	36
1.04.2	Introduction: Visual Pathways	36
1.04.3	What Does Visual Processing Mean?	38
1.04.3.1	Encoding, Representation and Processing	38
1.04.3.2	Information Processing Streams	39
1.04.3.3	Receptive Fields, Image Filters and Neural Processing	39
1.04.3.4	Selectivity for Image Features	40
1.04.3.5	Invariance	41
1.04.3.6	Processing as a Change of Coordinate Systems in Image Space	41
1.04.3.7	Why Are Some Representations Chosen over Others?	42
1.04.4	High-Level Feedforward Processing	43
1.04.4.1	Primate Ventral Pathway: Object Processing	44
1.04.4.1.1	Form Processing: V1 – Emergence of Orientation Selectivity	44
1.04.4.1.2	Form Processing: V2 – Sensitivity to Illusory Contours and Border Ownership	46
1.04.4.1.3	Form Processing: V4 - Contour and Primitive Shape Selectivity	46
1.04.4.1.4	Form Processing: IT - Object Selectivity	48
1.04.4.1.5	Form Processing: Color and Texture Processing Stream	48
1.04.4.2	Primate Dorsal Pathway: Coding Motion and Position	49
1.04.5	Role of Feedback	50
1.04.6	Summary	51
References		51

1.04.1 Abstract

This article discusses the core visual processing principles in biological visual systems. We start by briefly introducing the primate visual system because it provides a useful representative framework to explain multiple principles in visual processing. Throughout the article we emphasize the importance of hierarchical processing, function-specific processing streams, and feedback (two-way) pathways. Hierarchical processing allows each processing step to be introduced in a maximally efficient manner using neural hardware, while multiple streams allow the neural circuits to deal with the specific problems associated with given visual problems. We introduce the concept of visual receptive fields, discuss how visual information is encoded and represented in single cells and in cell populations, and how researchers conceptualize the same information in coordinate systems. We show how the coordinate systems need to be different at different levels in the processing hierarchy. An essential principle discussed throughout the article is how the system generates selectivity for given visual features but at the same time incorporates invariance to changes in those and other features. This is a major guiding principle in understanding the limitations to the structure of all visual pathways. The final essential principle is the need for visual systems to incorporate feedforward, feedback and lateral connections to generate the essential framework for image analysis. With the basic principles outlined, we then show how they are manifest in several important visual functions, namely orientation selectivity, sensitivity to illusory contours and border ownership, contour and primitive shape selectivity, object selectivity, color and texture processing, and coding motion and position.

1.04.2 Introduction: Visual Pathways

In this chapter, we introduce many of the core visual processing principles in biological visual systems, using the primate model as a framework to explain these doctrines. We show how visual processing involves changes in the representation of information along a hierarchy of distinct visual areas from basic image elements to the perception of complex objects. There are multiple processing streams that emphasize form, color, depth and motion. We show how the changes in representation are conceptualized as changes in the coordinate systems of an “image space”. The descriptions use real examples from primates to explain why certain types of visual processing are necessary and what mechanisms are at work. Perhaps most important of all, we show for every increase in selectivity to particular image features, there is a parallel increase in the amount of tolerance to changes in that feature. Put simply,

increased selectivity goes hand in hand with increased invariance. Understanding this conundrum in visual processing is perhaps one of the greatest challenges for future research.

A brief primer on the anatomy of a representative visual system is essential before talking about visual processing (Fig. 1). We have chosen to focus on the primate visual system as it provides a model that has general applicability when discussing visual processing. Perhaps the most important features of the primate system are that processing is hierarchical and divides into multiple pathways depending on the property of the image being processed (Felleman and Van Essen, 1991; Van Essen and Maunsell, 1983; Markov et al., 2010). A hierarchical design allows each step to introduce a processing step that can be performed using local neural circuits. The importance of multiple streams is that each can deal with a specific visual processing task in its own hierarchical sequence.

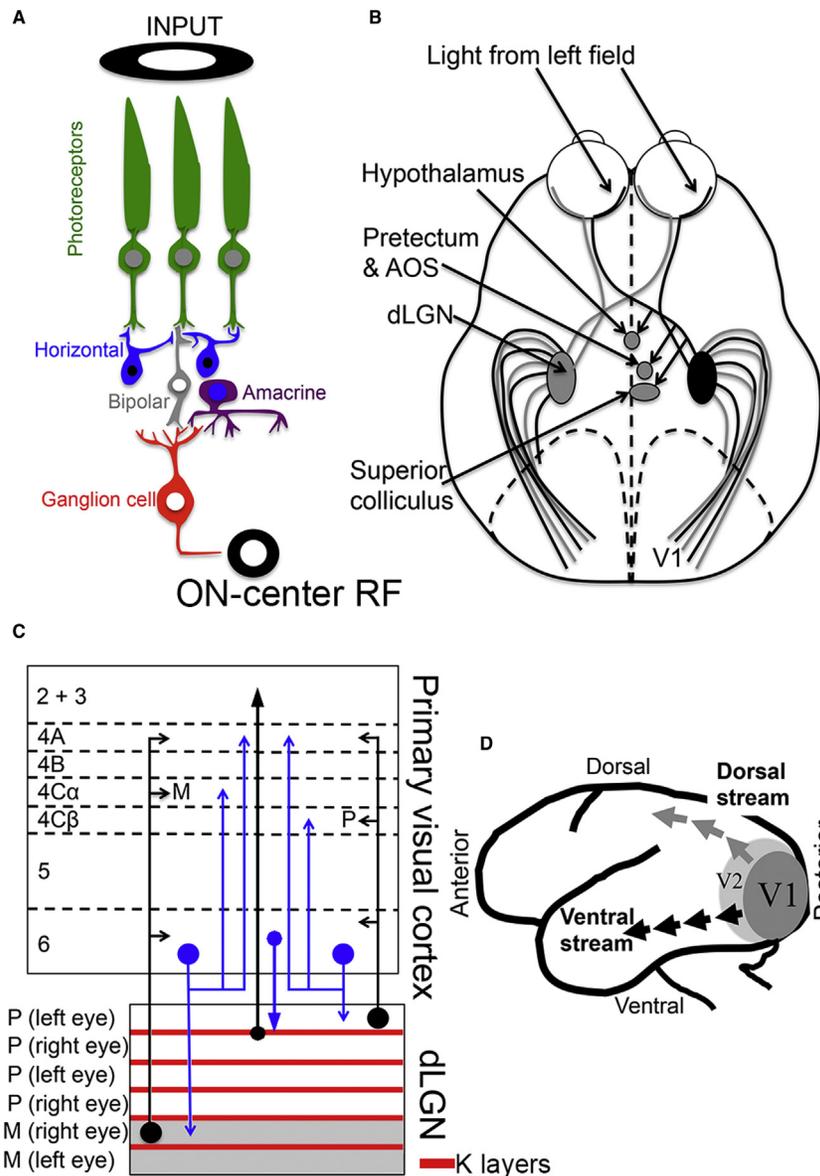


Figure 1 (A) Connections from photoreceptors (green), to a bipolar cell (gray) to a retinal ganglion cell (red) show the information flow through the retina. Horizontal (blue) and amacrine (purple) cells interconnect cells laterally. The optimal stimulus for this circuit is bright in the center and dark in the surround, as shown by the INPUT. (B) Pathways from the retina to the brain are dominated in primates by the pathway to the dorsal lateral geniculate nucleus (dLGN) and then to the primary visual cortex (V1). Information from the left visual field goes to the right half of the retina (black). (C) Layer 2–6 of the primary visual cortex, plus the sub-layers. The dLGN has six main layers comprising magnocellular (M - gray) and parvocellular (P - white) layers. The red bands between layers are the koniocellular layers (K - red). Black cells show feedforward connections and blue cells show feedback connections to the M, P and K layers. (D) The location of V1 and V2 in the primate brain. Black and gray arrows show the ventral and dorsal streams, respectively.

The retina contains photoreceptors that transform light into electrical signals (Nathans, 1999). They pass on their signals to bipolar cells, which then connect to retinal ganglion cells (RGCs) (Fig. 1A). The retina also contains horizontal and amacrine cells that interconnect cells laterally across the retina and thus introduce additional retinal processing (Fig. 1A) (Wässle, 2004). The RGCs send their axons along the optic nerve into several subcortical nuclei (Fig. 1B) (Ibbotson and Dreher, 2005). In primates the nucleus that receives axons from most cells is the dorsal lateral geniculate nucleus (dLGN) (Fig. 1B) (Rodieck and Watanabe, 1993). The dLGN is divided into six main layers, the dorsal four being Parvocellular (P) layers and the ventral two being Magnocellular (M) layers (Fig. 1C). There are also Koniocellular (K) layers that are sandwiched between each of the six primary layers (shown as red bands in Fig. 1C; Kaas et al., 1978). Each of the layers in the M and P sections of the dLGN are eye-specific, with cell responses dominated by stimuli presented in one eye (Fig. 1C). The next stage, the primary visual cortex (V1), has six layers (Fig. 1C). Parvocellular layer cells in the dLGN innervate layers 4A, 4C β and upper layer 6 (Fig. 1C). Magnocellular cells from the dLGN innervate 4A, 4C α and upper layer 6 (Fig. 1C). The K-cells innervate layer 3 in the cortex (Fig. 1C).

In V1, cells in layers 2 and 3 send axons to higher visual cortical areas such as V2, V3, V4 and V5 while cells in layers 5 and 6, respectively, send signals to the superior colliculus and back to the dLGN (Fig. 1C) (Casagrande and Kaas, 1994). Some of the axons that descend from layer 6 to the dLGN branch into several axon collaterals (blue cells in Fig. 1C). These not only travel to the dLGN but also travel within the primary visual cortex back into the primary input layers. These feedback circuits are very important for processing, as addressed at the end of this article.

Beyond visual areas V1 and V2, there are two primary visual pathways in the primate brain: ventral and dorsal (Fig. 1D) (Milner and Goodale, 1995, 2008). The ventral pathway primarily codes object related information, ending with cells that code specifically for faces and complex objects. The majority of the input to the ventral pathway comes from Parvocellular cells. The dorsal pathway deals more with the locations of objects in the visual field, image motion and the coordination of eye movements. Magnocellular cells provide a large fraction of the input to the dorsal stream.

1.04.3 What Does Visual Processing Mean?

1.04.3.1 Encoding, Representation and Processing

Visual images observed by the eyes are encoded in the neural responses of the retina and subsequent downstream visual brain areas. The visual information is represented as a population code, with each neuron in the population carrying information in the form of spike rates, spike times or membrane potentials. The representation can be considered a “neural code” for the stimulus (Rieke et al., 1999). For example, the sheet of photoreceptors in the retina represents the image in spatiotopic coordinates, which is similar to the way that pixels are organized in a digital camera. Each photoreceptor transduces the incoming light into a membrane potential that represents the amount of light (luminance) in a small region of the visual field. In the primate retina, different cone photoreceptors are sensitive to long, medium and short wavelength light (L, M, S) (Wässle, 2004), corresponding roughly to the red, green and blue pixels of a liquid crystal display screen. This representation changes at each subsequent stage of the visual pathway, to emphasize some aspects of the visual stimulus over other aspects according to the particular visual area. The changes in the representations are carried out by neurons that receive inputs from cells at a previous processing stage, as well as local connections within an area, and inputs from other, more distant, brain areas. Together these changes in representation constitute the visual processing carried out by the brain.

A useful way to visualize a representation is as the coordinates in an image space (Fig. 2). For example, a pixelated grayscale image can be represented using coordinates of the grayscale value of each pixel, from 0 (= black) to 1 (= white). This set of pixel grayscale values ($p_1, p_2, p_3, \dots, p_n$) represents the coordinates of a location in an image space with the dimension of the space equal to the number of pixels (n) (Fig. 2A). Fig. 2Ai shows a typical scene, Fig. 2Aii shows a pixelated portion of the scene and Fig. 2Aiii shows the 3-D coordinate representation. In this case the coordinate system represents the grayscale values of 3 pixels in a vertical line (p_1, p_2, p_3). The three axes, respectively, present the luminance of the upper, middle and lower pixel. Any particular combination of pixel grayscale is plotted in a unique 3-D location within the image space (green arrow, Fig. 2A). Another example of coordinates

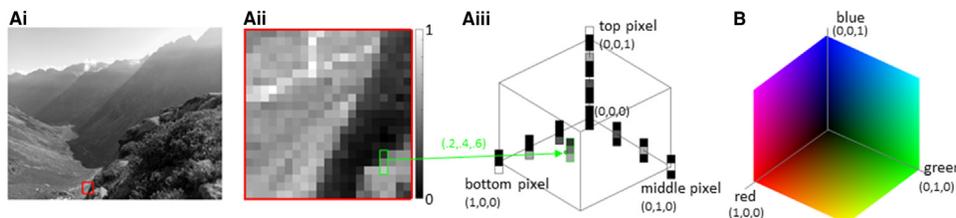


Figure 2 (Ai) A grayscale version of a natural scene. (Aii) A portion of the image (Ai) highlighted by the red square emphasizing the pixelated representation. (Aiii) A visualization of a 3-pixel image space in a 3-D plot. Each coordinate axis indicates the grayscale value of a pixel in the 3-pixel column as shown along each respective axis. The position within this image space of the 3-pixel column in (Aii) is indicated. (B) A 3-D representation of color along three axes: red, green and blue, with corresponding colors on three faces of the cube shown.

in an image space is a color representation of a single pixel corresponding to axes for the red, green and blue values (from 0 = black to 1 = brightest red, green or blue), which forms a three-dimensional space (Fig. 2B). Any color can be represented as a point in this image space. In most neural representations, the coordinates can be considered as the spike rates of neurons that encode the image.

1.04.3.2 Information Processing Streams

The image representation at the level of populations of photoreceptors is extremely detailed - it occurs at a very fine spatial and temporal resolution. This fine detail constitutes a massive inflow of information, much of which is of limited perceptual value. The activities of individual photoreceptors have little direct correlation with perceptions of the visual world because they vary in seemingly erratic ways over time and at neighboring points in space (by analogy, imagine trying to interpret a movie by viewing a single pixel!). A major task of processing in the visual pathway is to reduce the amount of information that is represented at higher levels of the pathway and transform this complex photoreceptor response into a representation that is behaviorally and perceptually relevant. Representations higher in the visual pathway emphasize more explicitly visual features that are closer to visual perception, e.g. form (shape), depth, color, texture, motion and object identity (Hubel and Livingston, 1987; Lu and Roe, 2007; Tanigawa et al., 2010; Conway, 2018; see Kandel et al., 2013 for an overview). The visual system organizes processing of many of these different qualities into different streams that emerge at various levels of the pathway and become increasingly explicit in their representation of these perceptual qualities as they ascend the pathway. Mathematically these transformations of the image representations can be understood as changes in coordinate systems in the image space, which frequently require non-linear processing. This idea will be illustrated with several examples in the following sections.

1.04.3.3 Receptive Fields, Image Filters and Neural Processing

A fundamental concept in understanding visual processing is the receptive field (RF). The RF of a neuron is the region of the visual field that generates responses when a suitable visual stimulus appears within it (Hartline, 1938). A modern and more elaborate concept of the RF is to think of it as a filter, i.e. a spatial filter of bounded extent that limits the region of the visual field to which the neuron is sensitive (Sharpee, 2013). Similarly, a temporal filter determines how far back in time a visual stimulus can affect a neuron's response, giving rise to the concept of the temporal RF. In general, RF filters process the incoming signal so the neuron becomes sensitive to a particular visual feature that is important for subsequent visual processing.

A linear filter can be implemented in a biologically plausible way by a single neuron at the first stage of the processing it performs on its inputs. In this first stage, input to the neuron arrives, typically as spikes from other neurons via synapses that modulate the strength of each input. The neuron then integrates its various synaptic inputs to generate a membrane potential that triggers spikes. This process usually occurs at a specific site in the neuron called the axon initial segment (Kole and Stuart, 2012). The membrane potential can be approximated as a weighted sum of the input spike rates, where the weights correspond to the strength of the synapses. This weighted sum of inputs can be represented mathematically as a filter that acts on the inputs (Chichilnisky, 2001).

Nonlinear processing is crucial to the brain's ability to perform sophisticated processing. Importantly, the second stage of processing performed by a single neuron is nonlinear. This stage involves converting the membrane potential into a spike rate. A common way to capture this process in a model is with a thresholding nonlinearity, whereby an output spike is triggered with a certain probability if the cell's membrane potential exceeds a threshold (Chichilnisky, 2001). The more the membrane potential exceeds this threshold the higher the probability of a spike. For a membrane potential below the threshold, the spike probability is at a spontaneous level. This form of nonlinear processing means that only information in the suprathreshold potential is passed on by this neuron for further processing, while information in the subthreshold potential is discarded. While this is a simple form of processing at the single neuron level, when it is applied iteratively through the cortical hierarchy and across large populations of neurons, it can lead to powerful computational capabilities. For example, in a two-stage hierarchy, a neuron in the second stage can be thought to have multiple filters because it combines the inputs from several neurons in the layer below, each of which has a single filter (McFarland et al., 2013). By combining these inputs in a nonlinear way, the neuron becomes sensitive to all the visual features corresponding to the filters of its inputs. In contrast, combining filters in a linear way would result in just a single filter corresponding to a weighted sum of the input filters. The result of nonlinear processing is that RFs become elaborate and sensitive to many feature dimensions.

There are usually two components to a visual RF. The most obvious is the classical RF. This is the region that, if stimulated visually *drives* a response (usually spikes). There are also often regions surrounding the classical RF, known as the extra-classical RF that do not generate spikes when stimulated (Spillmann et al., 2015). However, presenting stimuli in the extra-classical RF can *modulate* the spike rate being driven by the stimulus in the classical RF. This is necessarily a nonlinear form of processing, as the response to stimuli presented in both the classical and extra-classical RFs is not the same as the sum of the responses to each alone according to the definition of modulation. These extra-classical RFs are an integral component of the visual processing involved in contextual processing, i.e. understanding one stimulus in the context of other stimuli.

Classical RFs are formed through feedforward inputs from the previous areas in the visual hierarchy as well as local lateral connections within the cortical area. Extra-classical RFs are formed by lateral connections and feedback connections from higher cortical areas. Some types of neuron are capable of a range of nonlinear processing beyond that described here that can further enhance their computational capacities (Stuart et al., 2016).

As we will see below, as signals ascend the visual pathways, RFs become larger, more complex and more varied. This reflects the transformations in the representations of the visual world through the ascending visual system from a fine scale representation at the level of photoreceptors, which is closely related to the attributes of the light entering the eye, to a broader scale representation that is closer to the perceptual or behaviorally relevant attributes of visual objects.

1.04.3.4 Selectivity for Image Features

The RF filtering performed by neurons imparts response selectivity (tuning) for a particular visual feature. Through this filtering, selectivity is altered throughout the visual pathway to change the representation of the visual input (Wässle, 2004). This begins at the first synapse between photoreceptors and bipolar cells in the retina. Photoreceptors have relatively narrow RFs and are activated by the luminance level within that narrow RF (Fig. 3A). It does not matter from which object the light originates (e.g. a light source or reflections); each photoreceptor simply responds based on the luminance level in its RF. As such, photoreceptors are not very selective for perceptual attributes, but are highly selective for luminance changes in their RFs. As photoreceptors make synapses onto bipolar and horizontal cells in the retina, the system moves away from coding luminance to coding contrast. Contrast is the difference in luminance between two neighboring regions within a RF, e.g. as commonly created by edges between objects. Already, even at the first synaptic contact in the nervous system the selectivity of the visual system has transformed.

To make bipolar or RGCs respond vigorously, the image now needs to contain a stimulus that is brighter or darker in the central region of the RF compared to the surrounding region (e.g. Fig. 3B). To build a spatial filter that imparts selectivity of this sort to the bipolar cells, it forms a precise circuit with photoreceptors as the input, e.g. one or more photoreceptors in the center connect to a postsynaptic bipolar cell (Fig. 3C). Photoreceptors are inhibited by light, so in this case sign-inverting synapses lead to activation

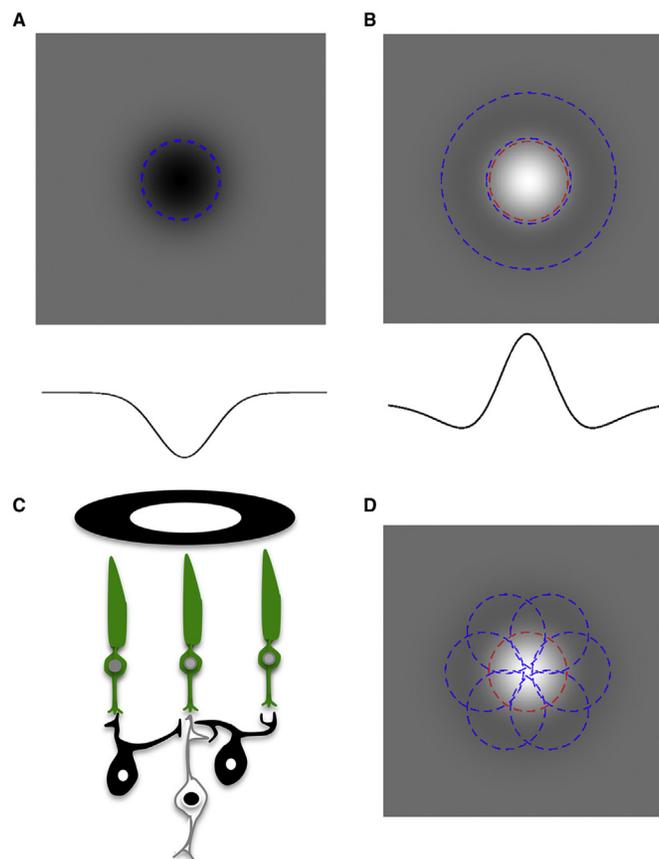


Figure 3 (A) The receptive field of a centrally located photoreceptor. Photoreceptors are excited as the light becomes darker within their RF (OFF response) (B) The ideal stimulus to excite an ON-sensitive bipolar cell. Due to the connectivity shown in (C) it is maximally excited by a bright central stimulus and a dark surround. (C) Schematic representation of the neural circuit in one plane that leads to an ON-center bipolar cell. The central bipolar cell is white, representing its response to ON via a sign inverting synapse from the central photoreceptor (green). The horizontal cells (black) provide sign-conserving signals to the central bipolar cell so that the latter is activated when the surrounding photoreceptors (also green) are activated by OFF. (D) The receptive fields of the contributing photoreceptors to the ON-center, OFF-surround response of the bipolar cell. It matches closely the ideal stimulus shown in (B).

of the central bipolar cell, creating an ON-center bipolar cell (Fig. 3D). Surrounding photoreceptors are activated by dark stimuli and pass on this signal via sign-conserving synapses to horizontal cells, which pass it onto the central bipolar cell. By doing this, the central bipolar cell becomes selective – it needs luminance increments (ON) in its RF center and/or luminance decrements (OFF) in its surround to respond. Other bipolar cells have the reverse circuitry (i.e. OFF center-ON surround).

Later, we will show how these center-surround RFs feed into more sophisticated circuitry to generate more selective feature selectivity.

1.04.3.5 Invariance

The description so far shows that selectivity for perceptually important visual qualities increases as the information ascends a hierarchy. However, increased selectivity has a significant drawback - building a linear RF filter that acts as a highly selective object template renders the system vulnerable to deviations away from that template. For example, if a cell were tuned to detect hands that had a specific position and orientation (Fig. 4A), even a slight sideways shift or rotational deviation would render the cell incapable of hand recognition (Fig. 4B and C). This is because there is no longer sufficient overlap between the image and the template filter. To overcome this problem, visual processing has mechanisms to add invariance, making the system tolerant to change. Building invariance into the system is non-trivial – it requires both high selectivity (to recognize the hand) and invariance to many parameters such as translational shifts, rotations or size.

Just as selectivity to different features is built incrementally through the visual hierarchy, so too is invariance. Again, this is apparent at the first synapse in the visual pathway. That is, the flipside of bipolar cells becoming selective for contrast is that they also become invariant to the overall luminance in their RFs.

1.04.3.6 Processing as a Change of Coordinate Systems in Image Space

The transformation in representation from luminance to contrast (Selectivity for Image Features section) can be illustrated as a change in coordinates of the image space. To aid visualization it is easiest to consider this in 2-dimensions by considering the activity of the center photoreceptor and surrounding photoreceptors as two groups. In this limited space the range of all possible illumination conditions falls within the square-dashed box shown in Fig. 5. Uniform illumination of both center and surround corresponds to the point in the bottom left corner, since light inhibits photoreceptors. No illumination corresponds to the top right corner (dark center and surround). Full illumination of either center or surround alone corresponds to one of the remaining corners, depending on the location of the photoreceptor group. The transformation carried out by bipolar cells to represent contrast involves taking the difference between the activities of the two groups of photoreceptors. Many illumination conditions lead to the same contrast level, i.e. all those conditions lying on the same diagonal blue dashed lines. All these points are projected to the same point on the contrast axis (solid blue axis) that is orthogonal to the total luminance axis (Fig. 5). This visualization of the transformation in image space illustrates two things. First, the bipolar cell response becomes selective to a new dimension in the image space, namely the contrast axis, and invariant along the axis of total luminance in the orthogonal direction. Second, there is a reduction in the overall dimensionality of the representation from 2 dimensions to 1 because the dimension of total luminance is no longer represented.

More generally, across the whole retina this transformation from photoreceptor to bipolar cells occurs on a high dimensional image space equal to the number of photoreceptors. Although this makes it difficult to visualize the space, the transformation can be understood in terms of the vectors representing the RF filters because the transformation is linear. Each RF filter consists of a set of weights $w=[w_1, \dots, w_n]$, that are used to calculate the output of the filter, which is the weighted sum of pixel values. The magnitude of the response of each neuron corresponds to a value along a particular coordinate axis in image space, with the direction of the axis given by the weight vector of the neuron's RF filter.

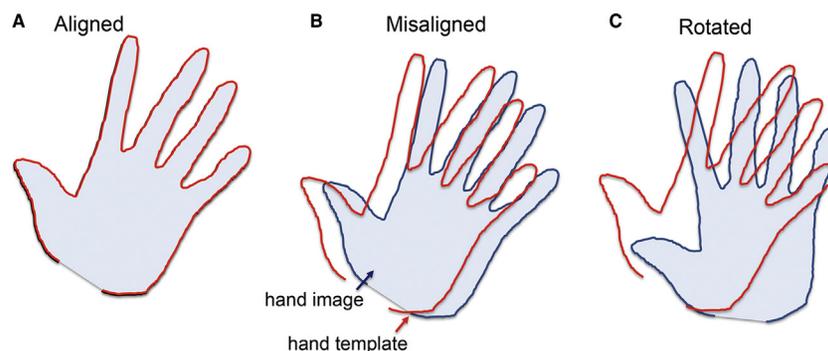


Figure 4 (A) The image of a hand (pale blue) aligned perfectly with a theoretical hand template (*red outline*) in the visual system. (B) A mismatch between the hand image and the template caused by a lateral misalignment. (C) A mismatch created by the rotation of the hand image.

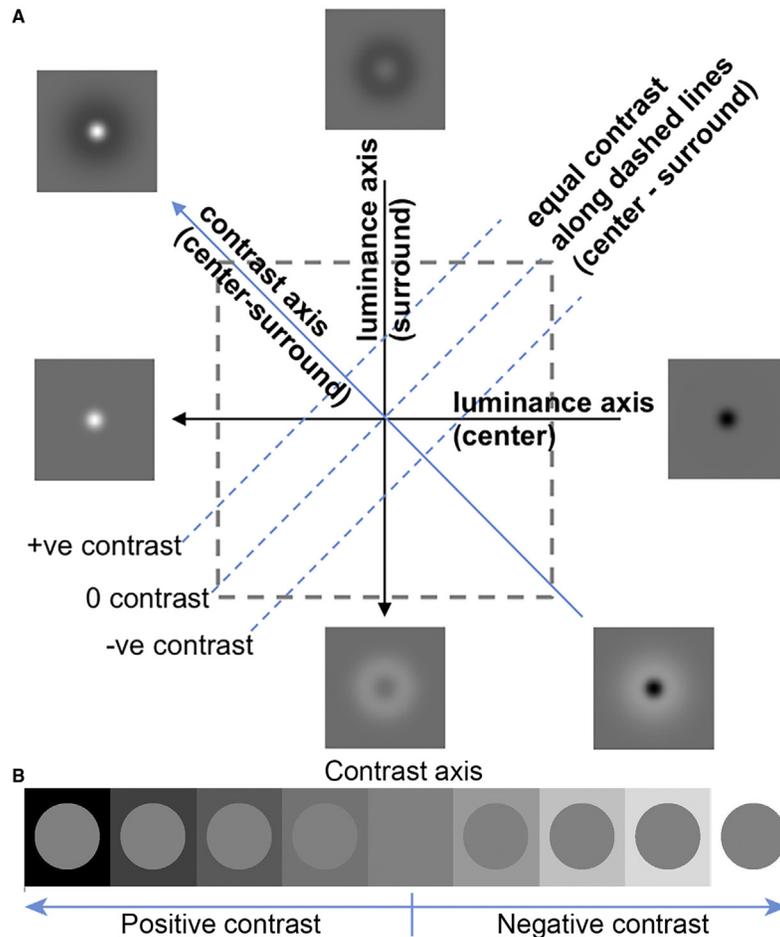


Figure 5 (A) Coordinate system for representing image space in the simplified two-dimensional example of a center-surround organization. The luminance axes are driven by the center (horizontal) and surround (vertical). Through subtraction of the center from the surround, a contrast axis is formed that is rotated by 45° (blue solid line). Multiple combinations of center and surround activation can generate the same contrast. Therefore, equal contrasts occur at all points along the dashed blue lines. Positive (+ve) contrasts are represented up and to the left, while negative (-ve) contrasts are represented down and to the right. The inset images show the optimal stimuli for the various axes. (B) Images show representative examples of stimuli along the positive and negative contrast axis.

1.04.3.7 Why Are Some Representations Chosen over Others?

Why did evolution select an explicit representation of contrast, while discounting luminance? An intuitive explanation is that contrast is a property of the object being viewed when light is reflected from bright or dark parts of the surface, while overall luminance is a property of the illumination conditions (sunny or overcast). Typically, the identification and analysis of objects is more important to the viewer than the lighting conditions. Further, information about the lighting conditions is highly redundant, as it is presented throughout the visual scene.

A more technical explanation is that bipolar cells are decorrelating or “whitening” the incoming image. Natural images have a $1/f^2$ power spectrum, where f is spatial frequency (Fig. 6A) (Field, 1987). Spatial frequency is $1/W$, where W is the spatial wavelength of Fourier components in an image. The result of this spectrum is that the amplitudes of low spatial frequency signals are very large, while the amplitudes of high spatial frequency information are very small. Therefore, the amplitude of the incoming signal is correlated with the spatial frequency (Fig. 6Bi). This is not advantageous for the visual system as detailed analysis of fine features is only possible if high spatial frequencies can be coded. Due to the power spectrum of natural images, low spatial frequencies have the capacity to swamp high spatial frequency information.

Center-surround RFs are very good at reducing the impact of low spatial frequencies because the surround differentially suppresses them (Fig. 6Bii). As a result, center-surround RFs can be thought to whiten (i.e. flatten) the spectrum by greatly suppressing low spatial frequency amplitudes but maintaining high spatial frequency amplitudes (Fig. 6Bii). Their power spectrum is given by f^2 . Thus, as a result of the whitening filters in the retina and dLGN (Atick and Redlich, 1992; Dan et al., 1996), the signals representing the visual image that arrives in V1 are decorrelated, i.e. the amplitude of the responses is approximately equal for the full spectrum of spatial frequencies (Fig. 6Biii). Therefore, the cortex processes signals that are not overwhelmed by low spatial frequencies (Fig. 6Biii).

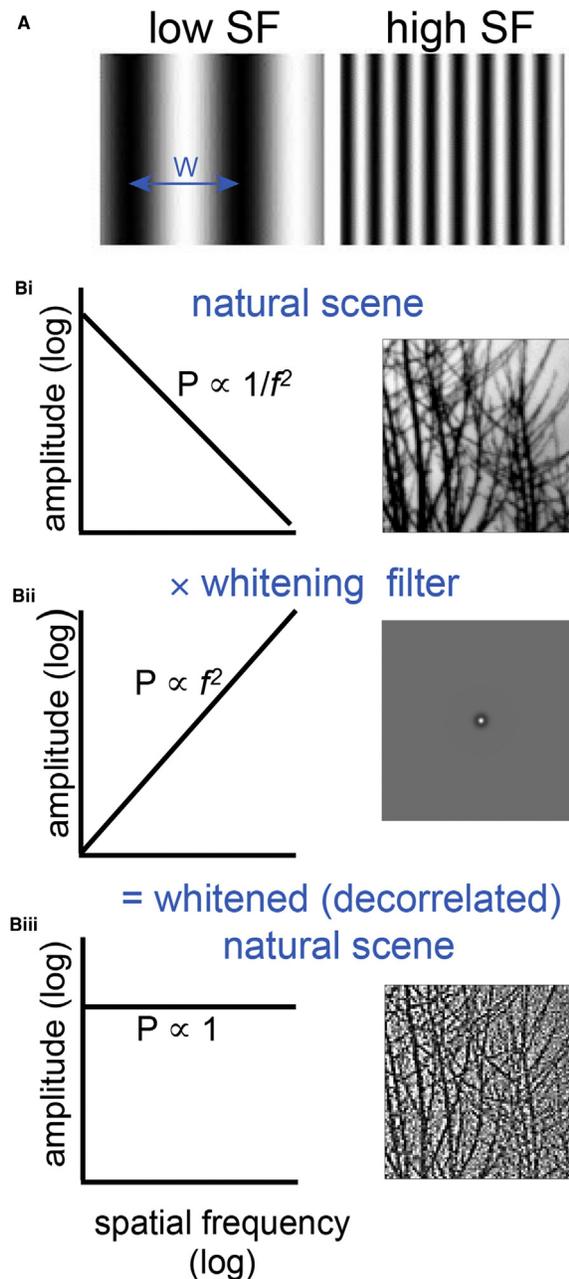


Figure 6 (A) Examples of low and high spatial frequency (SF) gratings. W is the spatial wavelength. Spatial frequency $f = 1/W$. (Bi) The power spectrum of a natural scene. Note the log-log scale on the axes here and below. This results in a straight line with slope -2 corresponding to the exponent of f . (Bii) The power spectrum of a center-surround filter, with the inverse relationship to that of the natural scenes shown in (Bi), i.e. a slope of $+2$. (Biii) A flat (i.e. whitened). power-spectrum obtained by multiplying the power spectra of the natural scene with that of the center-surround filter. This results in the filter output decorrelating the natural scene input.

This example illustrates a general principle in visual processing: each stage of processing produces a new representation that more explicitly encodes behaviorally relevant attributes of the visual scene. At the same time, it reduces redundancy and increases the efficiency with which behaviorally relevant information is represented.

1.04.4 High-Level Feedforward Processing

Having outlined some key concepts, we will now explain how they are manifest in high-level visual processing.

1.04.4.1 Primate Ventral Pathway: Object Processing

Processing in the ventral pathway passes from V1, V2, V4 and then inferotemporal cortex (IT), ultimately leading to visual representations that underlie object recognition and classification. The ventral pathway contains processing streams for form, color, depth and texture. Some controversy exists over the extent to which these processing streams occur in parallel, via segregated populations on the cortical sheet, or are processed together by the same population of neurons. For example, early studies found evidence for a segregation of form and color streams in V1 (Hubel and Livingstone, 1987; Lu and Roe, 2007; Tanigawa et al., 2010). However, other studies have found evidence for a more integrated processing of form and color (Friedman et al., 2003). A recent study that allowed large scale, unbiased sampling of neural populations by using calcium imaging showed that many neurons encode information about form and color jointly in primate V1 (Garg et al., 2019). They observed only moderate negative correlation between the strength of neural tuning for form and the strength of tuning for color across the population. This indicates a gradient, rather than segregation, of neurons that primarily code form through to color. In areas higher in the ventral pathway, beyond V1, previous studies using intrinsic optical imaging support a partial segregation of color and form streams. However, this imaging method only allows for coarse (i.e. mm scale) resolution (Lu and Roe, 2007; Tanigawa et al., 2010; Conway, 2018). Further studies with large scale unbiased sampling methods are required to assess the degree to which this is true at a single neuron level.

In the inferotemporal cortex, recent work indicates a partial segregation of streams related to a bias toward form or color across a constellation of four areas located in posterior, central, anterior and anterior-medial IT (Komatsu et al., 1992; Tootell et al., 2004; Matsumora et al., 2008; Lafer-Sousa and Conway, 2013). It remains unclear the extent to which this constellation of areas forms a strict hierarchy. The form-biased stream dominates in regions of each area associated with central vision, while the color-biased stream dominates in regions associated with more peripheral vision (Lafa-Sousa and Conway, 2013; Verhoef et al., 2015; Lafer-Sousa et al., 2016). RF size increases markedly along the ascending ventral pathway, with V1 to V4 maintaining a retinotopic organization, while in IT this retinotopic organization becomes indistinct.

1.04.4.1.1 Form Processing: V1 – Emergence of Orientation Selectivity

Form refers to object shapes as defined by their boundaries compared with the image background. At a fine spatial scale, the boundaries of objects are composed of oriented edges typically defined by contrast. Selectivity for oriented edges emerges in primate V1 (Hubel and Wiesel, 1968), with cells in the interblob regions exhibiting markedly greater orientation selectivity than those in blobs (Livingstone and Hubel, 1984; Lu and Roe, 2007). At the previous stage of the visual pathway, in dLGN, neurons have a center-surround RF, similar to RGCs, that does not code the orientation of the image features. However, if the RF of an ON-center cell is aligned appropriately with an OFF-center cell, such that the center of one is superimposed on the surround of another and combined linearly, it is possible to make an edge detector filter that is orientation-selective (Fig. 7Ai–iii). Orientation-selectivity can be made tighter by combining multiple ON- and OFF- center RFs (Fig. 7Aiv). Simple cells in the primary visual cortex have RFs with this arrangement and are selective for edge orientation (Fig. 7Bi–iii). They are also selective for displacements of the edge because their responses decline significantly if a correctly oriented edge is nudged into either the ON or OFF regions (Fig. 7iv). This displacement of the edge within the RF is known as spatial phase. As the phase of the stimulus affects the spike rate, simple cells are phase-selective.

A second type of orientation-selective neurons in V1, called complex cells, have nonlinear summation properties that allow them to respond in a way that markedly increases their invariance to spatial phase (Hubel and Wiesel, 1962, 1968; Movshon et al., 1978). Orientation-selective units can be formed in the same way as described above. However, in complex cells the response is then passed through an expansive nonlinearity, which means that the spike rates increase faster than predicted by linear summation (Fig. 7C). After the responses pass through nonlinearities, the outputs of two edge detectors are summed (Fig. 7C). If those two edge-detector filters are 90° phase shifted, so that one can be modeled as a sine and the other as a cosine, this arrangement creates some interesting processing opportunities. Let us consider the case where the nonlinearity is a squaring operation (Fig. 7C). In this case we have the responses from the two edge-detector filters (E_1 and E_2) being squared to give E_1^2 and E_2^2 . As sine-squared plus cosine-squared is one, the same spike rate will be generated by a complex cell wherever an optimally oriented edge is positioned in the RF. This model of processing is known as the energy model for complex cells and generates phase invariance (Adelson and Bergen, 1985).

This is different from what occurs in simple cells, which are highly phase-selective. Simple cells need the oriented edge to be precisely located in their RF while complex cells are just as selective for orientation but are tolerant to changes in phase. This is a very good example of the visual system using nonlinear processing to incorporate new forms of invariance into the representation. The transformation in representation performed by simple cells from their dLGN inputs amounts to a change from one rectilinear coordinate system to another, using linear spatial filters. This change in coordinate systems is similar in type to the transformation from photoreceptors to bipolar cells. However, in a subpopulation of complex cells that are perfectly phase invariant, the change in representation involves a more sophisticated transformation from a rectilinear to a polar coordinate system. For such a complex cell, the image space to which it is sensitive can be visualized as a two-dimensional plane in image space whose axes are defined by the two phase-shifted filters of the energy model (Fig. 7D). The spike rate is constant along circular contours, for which only the angular polar coordinate changes, but varies along spokes pointing out from the origin, for which the radial polar coordinate changes. This radial coordinate measures the contrast of oriented features sharing the same orientation but differing in spatial phase (see inset features in Fig. 7D). Thus the neuron is selective to orientation but invariant to phase.

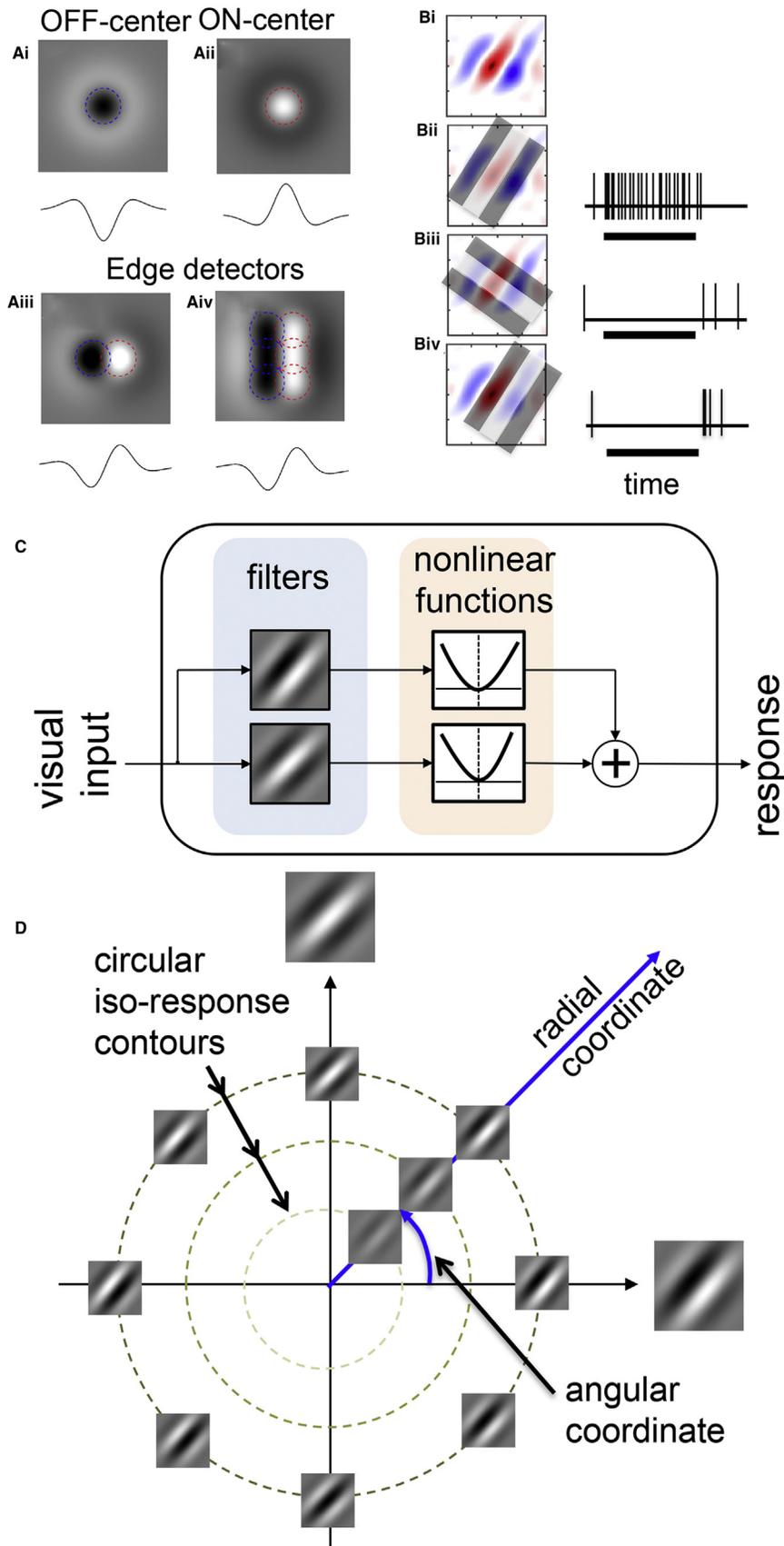


Figure 7 Receptive fields of an OFF-center (Ai) and ON-center (Aii) cell in the retina or dLGN. (Aiii) A basic edge detector formed by combining the responses of an ON- and OFF-center RF. (Aiv) An edge detector with far tighter orientation tuning, obtained by combining inputs from more center-surround RFs. (Bi) A simple cell RF recorded in cat cortex (red: ON-selective; blue: OFF-selective). (Bii) The optimal stimulus for the cell: correct spatial frequency and orientation. The spikes produced are shown to the right with duration of the stimulus presentation indicated by the black

An appreciable percentage of V1 complex cells have an intermediate level of phase sensitivity, which perhaps implies intermediate processing steps (Hietanen et al., 2013). These V1 complex cells exhibit properties more like those of simple cells when they are stimulated at low contrasts (Crowder et al., 2007; Cloherty and Ibbotson, 2015; Meffin et al., 2015). For this subset of complex cells, phase selectivity is dependent on image contrast, which can be considered as input strength at this stage in the hierarchy. Therefore, their phase invariance is not invariant to changes in input strength. Conversely, for the majority of complex cells in V1 and virtually all complex cells in V2 (Cloherty and Ibbotson, 2015), phase invariance is also contrast invariant.

Simple and complex cells in V1 are representing a key aspect of vision related to form processing, which is the presence of oriented features in the image that often form the boundaries of objects. In doing so, they bring the representation of objects a step closer to perception. Furthermore, this representation of oriented features is more efficient and sparser than the representation of contrast at the level of dLGN. This transformation to a more efficient (less redundant) representation (Olshausen and Field, 1996; Bell and Sejnowski, 1997) continues the trend we noted in retina involving a transformation from luminance to contrast. In both cases the visual system is utilizing the statistical dependence that naturally occurs in images to represent information more efficiently.

1.04.4.1.2 Form Processing: V2 – Sensitivity to Illusory Contours and Border Ownership

The transformation in representation in the form-processing stream from V1 to V2 is comparatively subtle compared to those in preceding and following stages. Neurons in V2 pale stripes show selectivity for orientation with most exhibiting invariance to spatial phase. One additional form of selectivity and invariance that is observed more often in V2 is cue-invariant selectivity to “illusory contours”. The simple cells described above in V1 detect a difference in luminance across an oriented edge. However, edges can also be defined by differences in other higher order cues such as contrast or texture, which occurs frequently in nature (Fig. 8A). Some neurons in V2 pale stripes exhibit selectivity to the orientation of contours defined by these higher order cues (An et al., 2014; Li et al., 2014). Moreover, many of these neurons have orientation tuning that is invariant to the type of cue that defines the contour, be it luminance, contrast or texture (Li et al., 2014). Thus, these neurons are invariant to the ways in which the boundaries are formed.

In addition to this feature extraction, the segregation of objects from the background of the visual scene begins to influence neural responses in V2 through the phenomenon of border-ownership. This is a form of contextual modulation of the responses of orientation selective neurons whereby the location of an oriented edge within a larger object can influence neural responses (Zhou et al., 2000; von der Heydt, 2015). For example, a neuron tuned to vertical orientations can respond vigorously when the left-hand edge of a dark square on a light background is placed in its RF (Fig. 8Bi). However, when an identical edge is placed in its RF as part of the right-hand edge of a light square on a dark background, the response is suppressed (Fig. 8Biv). Such a cell can also differentiate the reverse polarity condition as shown in Fig. 8Bii and Biii. Thus, regardless of brightness polarity, the neural response is enhanced when the square objects forming the edge are placed to the right of its classical RF and suppressed when they are placed to the left (Fig. 8B). This processing requires the ability to segregate figure from ground (i.e. object from background) and must contain information about the location of the figure. A major source of these signals is likely to be via feedback pathways from areas higher in the ventral pathway, and possibly connections within V2 from outside the cells classical RF.

1.04.4.1.3 Form Processing: V4 - Contour and Primitive Shape Selectivity

Many neurons in the form-processing stream in V4 are selective for curved contours or angles, demonstrating tuning to curvature ranging from a preference for acute angles though to oblique angles or straight lines (Pasupathy and Connor, 1999). In one population study the neural responses were shown to provide a representation of object shape across a limited set of primitive two-dimensional silhouetted shapes devised by the researchers (Fig. 8C; Pasupathy and Connor, 2002). This representation can be described by a coordinate system in which the boundary of each object is first decomposed into a set of contiguous contours. The shape of the object can be represented in a feature space using coordinates of the curvature of each contour, together with its angle with respect to the object’s center (Fig. 8C). The individual neural tuning contours and population representations of the shapes show some invariance to translations of the objects relative to the RFs (Pasupathy and Connor, 2002; Rust and DiCarlo,

horizontal bar. (Biii) A stimulus with an orthogonal orientation - no response is generated as dark and light regions of the stimulus do not align with OFF and ON regions of the RF, respectively, instead coinciding with both regions equally. (Biv) A phase shifted version of the stimulus. As it is 180° phase shifted there is a perfect mismatch between the stimulus and the requirements of the cell’s filter. Therefore the cell is strongly suppressed and there is no response. (C) Schematic of the energy model of a complex cell. It has two spatial filters identical except that they are phase shifted by 90°. The responses from the filters are squared such that an expansive nonlinearity is generated and both positive and negative contrasts generate spikes (full-wave rectification). These outputs are then summed leading to a phase invariant response. (D) The transformation in representation from simple to complex cells corresponds to a change in coordinate system in image space from rectilinear to polar coordinate. In the rectilinear coordinate system, the axes (*black lines*) correspond to the outputs of the two simple cell-like filters (large insets adjacent to axes *arrowheads*). In the polar coordinate system, the contrast of the oriented image features is represented as the radial coordinate, phase is represented as the angular coordinate. Small insets show how the contrast or spatial phase of the oriented image features changes in the image space along the radial and circular axes, respectively. For a complex cell described by the energy model, spike rates are constant on circular iso-response contours, and the contrast of features to which the cell is sensitive is represented by changes in spike rate along the radial axis. These features have the same orientation and spatial frequency, but can have arbitrary spatial phase.

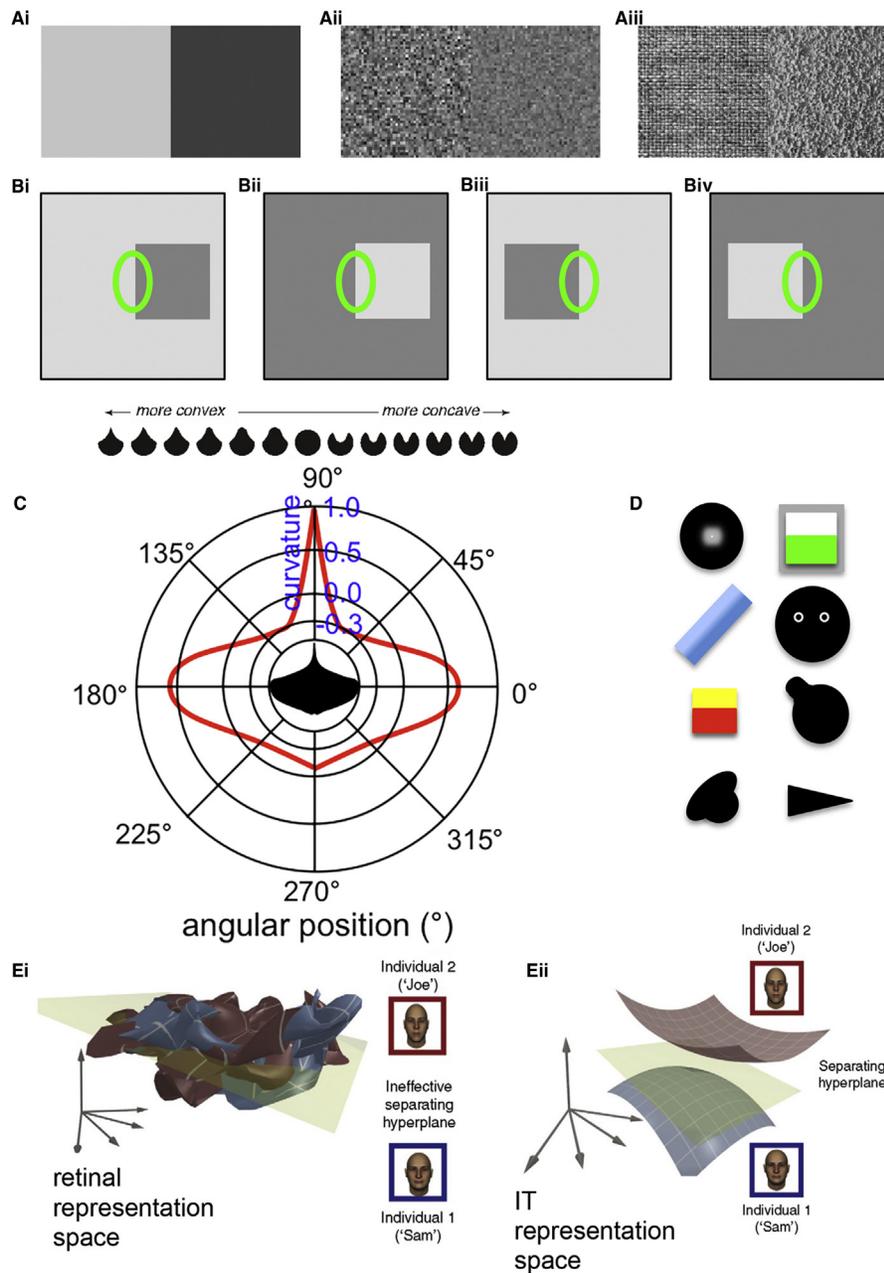


Figure 8 (A) Three examples of how to create an edge: a luminance border; a contrast border; a texture border. (B) Four examples of a box stimulus in relation to a cell's receptive field (*green line*), illustrating the concept of border ownership. The cell's response is enhanced when the square is on the right (Bi and Bii) but suppressed when it is on the left (Biii and Biv). This is despite the fact the edges within the classical RF are identical for pairs (Bi and Biv) and (Bii and Biii). (C) Population response for the shape shown in black on a white background at the center of the polar plot. The *red line* shows a coordinate representation of the shape's boundary curvature as a function of angular position. Individual neurons respond to segments of a boundary with the particular combination of curvature and angular position to which they are tuned. Thus, a neuron's response can be visualized as elevated spike rates at local spots in this polar coordinate system. Across the population, the shape is represented by elevated rates at points of the *red curve* that are sampled by neurons that respond to those particular boundary segments. It is in polar format centered around the shape's middle, which highlights the similarity between the representation and the actual boundary features. The row of figures at the top of the plot shows examples of shapes used in the study varying along a dimension such that the top segment of the shape changes continuously from convex to concave. (D) Eight example features for which IT cells were sensitive (Tanaka, 1996). (Ei) A multi-dimensional representation of retinal responses to Sam and Joe's faces (outlined as red or blue). The multiple *arrows* schematically represent the dimensions. The surface represents the two different faces under various identity preserving transforms. The surfaces are highly 'entangled' and cannot be separated by any planar (or other simple) boundary (light green) (Eii) A schematic representation of how IT untangles the retinal representation through multiple layers of processing to give a representation of the faces fully segregated by a planar boundary in multi-dimensional image space. (E) Adapted with permission from Di Carlo and Cox, 2007.

2010). Overall this change of representation emphasizes object form more explicitly and marks a shift away from retinotopic-based coordinates to object-based coordinates.

1.04.4.1.4 Form Processing: IT - Object Selectivity

Many neurons in IT are selective to complex features that represent proto-objects. Perhaps the most famous regions in IT are the face-selective brain areas. Cells within the various face-selective areas are specialized for face recognition but also show tolerance to changes in facial position, rotation and size (Tanaka, 1996; Zoccolan et al., 2007). Tanaka (1996) determined the critical features that generated spiking responses in IT cells and found most cells required quite complex features such as the 8 examples in Fig. 8D. The critical features were far more sophisticated than orientation, color, size and texture, as extracted by V1. Some of the features require exact combinations of shape, color, texture and the correct brightness polarity (Fig. 8D). At a population level the identify of faces and other classes of objects undergoing identity-preserving transformations can be decoded with good accuracy from the neural representation using a linear classifier (Fig. 8D; Hung et al., 2005; Rust and DiCarlo, 2010). This means that when the representation of the image as the spike rates of cells in a population are plotted as coordinates in a multi-dimensional space, the points in the space corresponding to different objects can be separated with straight planar boundaries, while the points corresponding to the same object under different viewing conditions lie on the same side of the planar boundaries. The importance of this is that such linear classification can be plausibly implemented by single neurons as a linear filter via synaptic integration to “read out” this neural code at a subsequent stage of processing. The population of neurons in IT performs better with natural object recognition compared to cell populations in V4, under various identity-preserving transformations. This form of explicit and invariant representation of objects in IT is more robust than that found in V4 and vastly different to that in V1 (Rust and DiCarlo, 2010). DiCarlo and Cox (2007) have argued that the transformation from V1 to IT amounts to an ‘untangling’ of the representation of objects (Fig. 8E). In V1 the representation of different objects is entangled because the points in the space representing the same object under different viewing conditions are widely dispersed and intermixed with those of other objects (Fig. 8Ei). The processing along the ventral stream of the visual pathway leads to a representation of objects in IT that allows easy separation of the points corresponding to one object using planar boundaries (Fig. 8Eii).

1.04.4.1.5 Form Processing: Color and Texture Processing Stream

Rather than giving a detailed account of the transformations and representations occurring in other visual processing streams, we will highlight a number of well studied examples that serve to emphasize the common principles at work throughout the visual cortex that have already been described.

We have introduced the concept of contrast-selective cells in the retina. The central part of the primate retina, responsible for high acuity vision, uses this same center-surround principle to code color (Fig. 9A). In this case cones sensitive to long wavelengths (red) might stimulate the center of the RF while cones sensitive to medium wavelengths (green) stimulate the surround (Fig. 9Ai and Aii). In this way the system generates color opponency: some cells are red-center and green-surround and others have the reverse organization. These cells are known as red-green opponent cells. Other cells in retina receive input from short wavelength (blue) cones in the center and a mix of red and green cones (collectively called yellow) in the surround or vice versa (Fig. 9 Aiii–Aiv): blue-yellow opponent. In the peripheral primate retina, color opponency appears much less frequently in RGCs, with a variety of degrees of opponency from cells that are fully opponent to cells that are not. The observed variety is consistent with nonselective wiring from red and green bipolar cell inputs to their target RGCs (Solomon et al., 2005; Wool et al., 2018).

The R-G and B-Y representation of color does not correspond very closely to its perception, which is aligned to axes of hue (similar to colors of the rainbow), saturation (pale to deep shades) and value (dim to bright) (Fig. 9Bi). The representation in

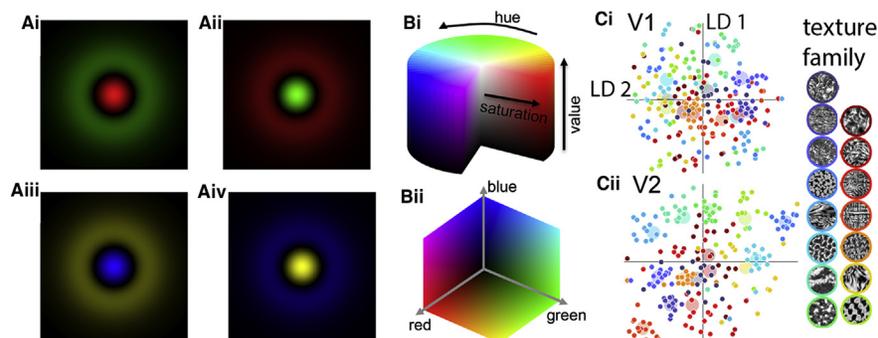


Figure 9 Retinal receptive fields of red-center (Ai), green-center (Aii), blue-center (Aiii) and yellow-center (Aiv) cells. (Bi) Depiction of color space using the coordinate system of hue, saturation and value. (Bii) Depiction of color space using the red, green, blue coordinate system. (Ci) Responses of V1 neurons to the various texture families shown to the right. Each texture family is color coded in the 2-D plot of latent dimensions (LD1 and LD2). In V1 the texture families are mixed. (Cii) In V2 the texture families are segregated within the latent dimension plot. Latent dimensions were obtained using t-SNE, which is an algorithm for visualizing high-dimensional spaces in 2 or 3 dimensions (Ziemba et al., 2016). Panel C is adapted with permission from Ziemba et al., 2016.

the next stage of the ventral pathway, in the thin stripes of V2 is more closely aligned to the perceptually relevant coordinates, with many neurons tuned to hue (Fig. 9Bi). An important aspect of color perception is color-constancy, in which the color of an object appears the same across different illumination conditions that can dramatically alter the spectral content of reflected light (Roe et al., 2009). This is a form of perceptual invariance that relies on contextual cues apparent in the visual scene surrounding the object. Neurons in V4 but not V2 demonstrate color-constancy by shifting their hue-tuning to compensate for the color of the illumination (Conway, 2014).

Another perceptual surface property that is represented more explicitly in V2 for the first time in the visual hierarchy is natural textures. We have placed the discussion of texture selectivity here because this type of texture processing is related to the perception of the surface properties of objects, while form vision is related to the boundaries of objects. Some neurons in V2 show selectivity to specific natural textures (e.g. bark, leaves, stone), while this is far less evident in V1 (Fig. 9C; Freeman et al., 2013). Moreover, at the population level neurons selective to particular textures show invariance in their classification of entire texture families that change the local details of the texture but leave key statistical properties of the image unchanged, leading to similar perceptual grouping in human observers (Freeman et al., 2013; Ziemba et al., 2016). It remains unclear whether these neurons are located in V2 thin stripes, corresponding to a stream for processing surface properties, or whether there is any topographic representation across axes of texture representation (e.g. smooth to rough).

1.04.4.2 Primate Dorsal Pathway: Coding Motion and Position

We have seen that V1 in the primate brain marks the emergence of pronounced tuning of neural responses for oriented features. Studies with drifting sinusoidal gratings have shown that many V1 neurons are also direction-selective, responding more to one direction of motion (perpendicular to their preferred orientation) than the opposite direction (De Valois et al., 1982). Direction-selective responses can be created by an elaboration of the “energy model” (e.g. Fig. 7C) (Clifford and Ibbotson, 2003). In this elaboration, a temporal delay to the spatial filters is added to give a pair of “direction-selective” filters in space and time (Adelson and Bergen, 1985; Emerson et al., 1992).

For these types of motion detecting V1 neurons, any motion in the direction parallel to the edge cannot be detected. This leads to a phenomenon known as the aperture problem: viewed through the comparatively narrow window of a V1 RF, the edge of a moving object will appear to move only perpendicular to the edge, although the object may be moving in a variety of directions (Fig. 10A).

Area MT (middle temporal area), also sometimes referred to as V5, is the first cortical area lying distinctly within the dorsal visual pathway, and receives major projections from both layer 4B of V1 and the thick stripes of V2 (Movshon and Newsome, 1996; DeYoe and Van Essen, 1985). Neurons in MT integrate local cues in a variety of ways to represent the motion of objects more explicitly, rather than the motion of local features, as coded in V1. This is analogous to the transformation seen in V4 that produces a more global, perceptually relevant, object centered representation compared to V1. There are several examples of this more elaborate type of motion selectivity in MT.

One example is the selectivity to pattern motion evident in plaids (Movshon et al., 1986). Plaids are composed of a sum of two sinusoidal drifting gratings moving in different directions (Fig. 10B). Humans perceive a unified movement of a single pattern, i.e. the intersection points of the gratings move coherently in one direction, which differs from either of the directions of the components or their vector sum/average (green arrow, Fig. 10B). Neurons in V1 and some MT cells respond to plaids in terms of the motion of separate sinusoidal components (component cells). A subpopulation of MT neurons is selective for the overall motion of plaid patterns (pattern cells), which match human perception. As yet another example of invariance being incorporated into the visual system, the responses of these neurons are invariant to changes in the angle between the component gratings that leave the pattern direction the same (Fig. 10B).

A second example is the barber’s pole illusion, which exemplifies a class of stimuli that generate a unified perception of motion, despite containing a mixture of ambiguous 1D motion cues (oblique contours), and 2D features (bar terminators) that are moving in two different directions (Pack et al., 2004). For example, the vertically elongated barber’s pole shown in Fig. 10C is perceived to have predominantly upward motion consistent with the direction of the terminators on the long vertical edges. This perception arises despite the presence of terminators moving to the right on the short horizontal edges and the oblique motion of contours (up and right). Pack et al. (2004) showed that many MT neurons can selectively integrate the information about the motion of terminators, while ignoring ambiguous information about the motion of 1D features (oriented contours).

A general principle of visual motion tracking is that it requires two opposing processes: integration and segmentation. Integration involves combining motion cues belonging to the same object to obtain unambiguous and accurate estimates of the motion. Segmentation involves distinguishing between the motion of different objects so that they can be tracked separately. Huang et al. (2007; 2008) studied these effects using large drifting box stimuli that cover both the centers and surrounds of MT neurons (Fig. 10D). They found that many individual MT neurons can shift between integration and segregation in a dynamic way that depends on the stimulus. When the motion cues in the RF center were weak and ambiguous (i.e. 1D contours), modulation by the surround had an integrative effect. However, when the central motion cues were unambiguous and strong (i.e. 2D features) the modulation by the surround was consistent with segmentation. Modeling investigations have demonstrated the ability to incorporate all of the response properties of MT listed above using the building blocks present in the prior components of the hierarchical visual system (Eskikand et al., 2016, 2018; 2019).

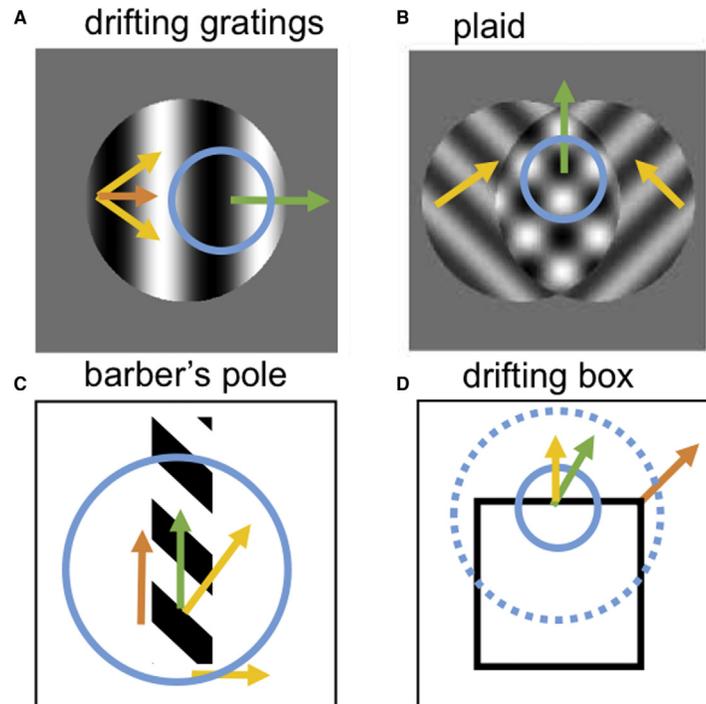


Figure 10 (A) A drifting grating stimulating the RF of a cell (*blue circle*). Regardless of the actual direction of motion (*orange or yellow arrows*), the cell always sees the grating move to the right (*green arrow*). This is the aperture problem. (B) Summation of two drifting gratings create a plaid that covers the RF (*blue circle*). Humans perceive upward motion, as do MT pattern cells (*green arrow*). MT component cells code the directions of motion of the individual gratings (*yellow arrows*). (C) The barber's pole stimulus. Humans perceive upward motion (*green arrow*), which is provided by the upward moving terminators (*orange arrow*). Humans ignore the rightward moving terminators (*yellow horizontal arrow*) and the 1-D signal provided by the edge that signals upward-right motion (*yellow slanted arrow*). (D) A drifting box stimulus. The classical RF is shown as a *solid blue circle*. It is confounded by the aperture problem because it only contains the upper horizontal 1-D edge of the box (*yellow arrow*). However, the extra-classical RF can see the box terminator moving upward and to the right (*orange arrow*), which allows the extra-classical RF to influence the cell to signal the correct direction of motion (*green arrow*).

1.04.5 Role of Feedback

Feedback is perhaps one of the most important engineering principles and is ubiquitous within biology (homeostasis). Feedback allows areas higher in the cortical hierarchy to modulate the input to the system so that neural processing can be optimized. There is very little feedback from the brain to the retina. However, as if to make up for this physical limitation of the system (i.e. there is not enough space in the optic nerve), the number of synapses to the dLGN (Fig. 1B) from V1 is far greater than the number of feedforward synapses (Guillery, 1969; Erisir et al., 1997a,b). Recall that there is an entire layer of the primary visual cortex (i.e. 1/6th of the cortical thickness) devoted to sending signals back to the dLGN.

Two possible roles for feedback to dLGN are to sharpen/broaden the tuning of RFs and to enhance signal transmission (Briggs and Usrey, 2008). Sharpening/broadening appears to occur primarily by changing the strength of extra-classical RFs (Murphy and Sillito, 1987). This could improve the capacity for dLGN to adjust the filtering of the incoming visual signal that decorrelates or “whitens” the response of neighboring neurons, as discussed in Fig. 6. As this whitening relies on the relative strength and spatial extent of the surround and center of the RF, adjusting the strength of the extraclassical surround could allow this whitening to be optimized to match the degree of spatial correlation in the prevailing visual input. In terms of enhancing signal transmission, feedback increases the gain of dLGN responses in a contrast-independent manner (Przybylski et al., 2000). Most of what we know about dLGN comes from anesthetized animals. It may well be that feedback has yet more profound influences in awake animals where feedback is stronger (Briggs and Usrey, 2011).

As an example of the importance of higher-level feedback in vision, consider the example shown in Fig. 11. In this model, V1 cells with small RFs feed into a high-level cell with a larger RF. As the stimulus moves to the right, the higher order neuron is stimulated first by cell 1. The higher order neuron then sends a signal back to all of the early cells, including cell 2. This feedback signal increases the sensitivity of the V1 cell by reducing its threshold for activation. Therefore, when the stimulus reaches cell 2, it responds more strongly. The feedback has allowed the system to enhance future responses and therefore make the system more sensitive, and it has given the system some basic predictive capabilities. This is a type of predictive coding and is very useful in many scenarios such as enhancing responses to stimuli with low salience (e.g. objects that are camouflaged) (Hupé et al., 1998).

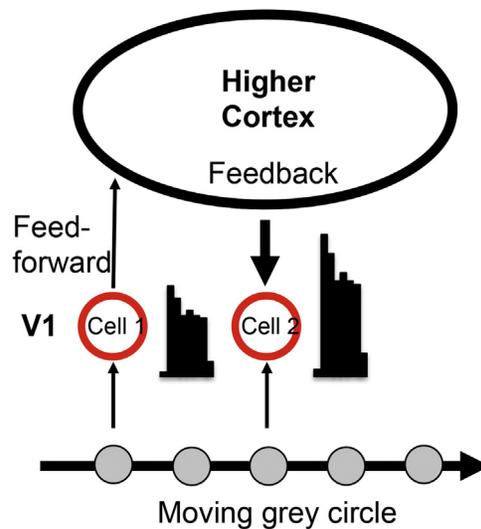


Figure 11 A higher cortical neuron with a large RF sends a signal back to V1 neurons. Cell 2 responds more strongly because the feedback has reduced its threshold for activation. The black histograms show the response amplitudes of cell 1 and cell 2 relative to time. This is a simple form of predictive coding.

Predictive coding is a theory for generating predictions of lower-level sensory inputs at higher levels of the visual hierarchy, which uses these “generative models” to filter sensory inputs arising from lower levels (Rao and Ballard, 1999; Friston and Kiebel, 2009; Huang and Rao, 2011). A comparison is performed between the predictions, known as priors, and the sensory input, known as the likelihood, which generates a prediction error. If the predicted error is larger than the signal expected from noise, the generative model is updated. If there is no predictive error the match is perfect and the system is maximally efficient, not having to re-process the incoming signal from scratch. The updating process allows the generative model to “learn” and better predict future sensory signals, thus reducing the burden on the sensory input system.

1.04.6 Summary

We have introduced many of the core visual processing principles in biological visual systems. We used the primate visual system as a representative framework to explain multiple principles as it is well studied and has many segregated brain areas that are each selective for particular types of processing. We have emphasized that visual processing involves changes in the representation of information in stages through a hierarchy of visual areas. These changes in representation begin with raw sensory input in the retina that reflects the physical transduction of light signals by photoreceptors into neural signals. Each subsequent stage of processing in the ascending hierarchy transforms the representation to bring it closer to perception. This occurs across multiple processing streams including form, color, depth and motion. These changes in representation can be conceptualized as changes in the coordinate system of an “image space”. It is hoped that through a series of examples, we have explained why certain types of visual processing are necessary and given an introduction to the mechanisms involved. An important theme throughout was to highlight the constant conflict between the need for selectivity and invariance to visual features.

References

- Adelson, E.H., Bergen, J.R., 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2 (2), 284–299.
- An, X., Gong, H., Yin, J., Wang, X., Pan, Y., Zhang, X., 2014. Orientation-cue invariant population responses to contrast-modulated and phase-reversed contour stimuli in macaque V1 and V2. *PLoS One* 9 (9), e106753.
- Atick, J.J., Redlich, A.N., 1992. What does the retina know about natural scenes? *Neural Comput.* 4 (2), 196–210.
- Bell, A.J., Sejnowski, T.J., 1997. The “independent components” of natural scenes are edge filters. *Vis. Res.* 37 (23), 3327–3338.
- Briggs, F., Usrey, W.M., 2008. Emerging views of corticothalamic function. *Curr. Opin. Neurobiol.* 18 (4), 403–407.
- Briggs, F., Usrey, W.M., 2011 Jan 1. Corticogeniculate feedback and visual processing in the primate. *J. Physiol.* 589, 33–40.
- Casagrande, V.A., Kaas, J.H., 1994. The afferent, intrinsic, and efferent connections of primary visual cortex in primates. In: *Primary Visual Cortex in Primates*. Springer, Boston, MA, pp. 201–259.
- Chichilnisky, E.J., 2001. A simple white noise analysis of neuronal light responses. *Netw. Comput. Neural Syst.* 12 (2), 199–213.
- Clifford, C.W.G., Ibbotson, M.R., 2003. Fundamental mechanisms of visual motion detection: models, cells and functions. *Prog. Neurobiol.* 68, 409–437.
- Cloherly, S.L., Ibbotson, M.R., 2015. Contrast-dependent phase sensitivity in V1 but not V2 of macaque visual cortex. *J. Neurophysiol.* 113, 434–444.
- Conway, B.R., 2014. Color signals through dorsal and ventral visual pathways. *Vis. Neurosci.* 31, 197–209.
- Conway, B.R., 2018. The organization and operation of the inferior temporal cortex. *Annu. Rev. Vis. Sci.* 4, 381–402.

- Crowder, N.A., Van Kleef, J., Dreher, B., Ibbotson, M.R., 2007. Complex cells increase their phase-sensitivity at low contrasts and following adaptation. *J. Neurophysiol.* 98, 1155–1166.
- Dan, Y., Atick, J.J., Reid, R.C., 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* 16 (10), 3351–3362.
- De Valois, R.L., Yund, E.W., Hepler, N., 1982. The orientation and direction selectivity of cells in macaque visual cortex. *Vis. Res.* 22 (5), 531–544.
- DeYoe, E.A., Van Essen, D.C., 1985. Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature* 317 (6032), 58.
- DiCarlo, J.J., Cox, D.D., 2007. Untangling invariant object recognition. *Trends Cognit. Sci.* 11 (8), 333–341.
- Emerson, R.C., Bergen, J.R., Adelson, E.H., 1992. Directionally selective complex cells and the computation of motion energy. *Vis. Res.* 32 (2), 203–218.
- Erisir, A., Van Horn, S.C., Bickford, M.E., Sherman, S.M., 1997a. Immunocytochemistry and distribution of parabrachial terminals in the lateral geniculate nucleus of the cat: a comparison with corticogeniculate terminals. *J. Comp. Neurol.* 377, 535–549.
- Erisir, A., Van Horn, S.C., Sherman, S.M., 1997b. Relative numbers of cortical and brainstem inputs to the lateral geniculate nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 94, 1517–1520.
- Eskikand, P.Z., Kameneva, T., Ibbotson, M.R., Burkitt, A.N., Grayden, D.L., 2016. A possible role for end-stopped V1 neurons in the perception of motion: a computational model. *PLoS One* 11, e0164813.
- Eskikand, P.Z., Kameneva, T., Ibbotson, M.R., Burkitt, A.N., Grayden, D.L., 2018. A biologically based computational model of visual cortex that explains the X-junction illusion. *Neural Networks* 102, 10–20.
- Eskikand, P.Z., Kameneva, T., Burkitt, A.N., Grayden, D.B., Ibbotson, M.R., 21 June 2019. Pattern motion processing by MT neurons. *Front. Neural Circ.* 13, 43.
- Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebr. Cortex* 1 (1), 1–47.
- Field, D.J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4 (12), 2379–2394.
- Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., Movshon, J.A., 2013. A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* 16 (7), 974–981.
- Friedman, H.S., Zhou, H., von der Heydt, R., 2003. The coding of uniform colour figures in monkey visual cortex. *J. Physiol.* 548, 593–613.
- Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 364 (1521), 1211–1221.
- Garg, A.K., Li, P., Rashid, M.S., Callaway, E.M., 2019. Color and orientation are jointly coded and spatially organized in primate primary visual cortex. *Science* 364, 1275–1279.
- Guillery, R.W., 1969. The organization of synaptic interconnections in the laminae of the dorsal lateral geniculate nucleus of the cat. *Z. für Zellforsch. Mikrosk. Anat.* 96, 1–38.
- Hartline, H.K., 1938. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am. J. Physiol.-Legacy Content* 121 (2), 400–415.
- Hietanen, M.A., Cloherty, S.L., van Kleef, J.P., Wang, C., Dreher, B., Ibbotson, M.R., 2013. Phase-sensitivity of complex cells in primary visual cortex. *Neuroscience* 237, 19–28.
- Huang, Y., Rao, R.P., 2011. Predictive coding. *Wiley Interdiscipl. Rev. Cognit. Sci.* 2 (5), 580–593.
- Huang, X., Albright, T.D., Stoner, G.R., 2007. Adaptive surround modulation in cortical area MT. *Neuron* 53 (5), 761–770.
- Huang, X., Albright, T.D., Stoner, G.R., 2008. Stimulus dependency and mechanisms of surround modulation in cortical area MT. *J. Neurosci.* 28 (51), 13889–13906.
- Hubel, D.H., Livingstone, M.S., 1987. Segregation of form, color, and stereopsis in primate area 18. *J. Neurosci.* 7 (11), 3378–3415.
- Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160 (1), 106–154.
- Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243.
- Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J., 2005. Fast readout of object identity from macaque inferior temporal cortex. *Science* 310 (5749), 863–866.
- Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., Bullier, J., 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394 (6695), 784–787.
- Ibbotson, M.R., Dreher, B., 2005. Visual functions of the retinorecipient nuclei in the midbrain, pretectum and ventral thalamus of primates. In: Kremmer, J. (Ed.), *The Primate Visual System: A Comparative Approach*. Wiley, Chichester, pp. 213–266.
- Kaas, J.H., Huerta, M.F., Weber, J.T., Harting, J.K., 1978. Patterns of retinal terminations and laminar organization of the lateral geniculate nucleus of primates. *J. Comp. Neurol.* 182 (3), 517–553.
- Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., Hudspeth, A.J., Mack, S. (Eds.), 2013. *Principles of Neural Science*, fifth ed. McGraw-hill, New York.
- Kole, M.H., Stuart, G.J., 2012. Signal processing in the axon initial segment. *Neuron* 73 (2), 235–247.
- Komatsu, H., Ideura, Y., Kaji, S., Yamane, S., 1992. Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *J. Neurosci.* 12, 408–424.
- Lafer-Sousa, R., Conway, B.R., 2013. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat. Neurosci.* 16, 1870–1878.
- Lafer-Sousa, R., Conway, B.R., Kanwisher, N.G., 2016. Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *J. Neurosci.* 36, 1682–1697.
- Li, G., Yao, Z., Wang, Z., Yuan, N., Talebi, V., Tan, J., et al., 2014. Form-cue invariant second-order neuronal responses to contrast modulation in primate area V2. *J. Neurosci.* 34 (36), 12081–12092.
- Livingstone, M.S., Hubel, D.H., 1984. Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* 4 (1), 309–356.
- Lu, H.D., Roe, A.W., 2007. Functional organization of color domains in V1 and V2 of macaque monkey revealed by optical imaging. *Cerebr. Cortex* 18 (3), 516–533.
- Markov, N.T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., et al., 2010. Weight consistency specifies regularities of macaque cortical networks. *Cerebr. Cortex* 21 (6), 1254–1272.
- Matsumura, T., Koida, K., Komatsu, H., 2008. Relationship between color discrimination and neural responses in the inferior temporal cortex of the monkey. *J. Neurophysiol.* 100, 3361–3374.
- McFarland, J.M., Cui, Y., Butts, D.A., 2013. Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Comput. Biol.* 9 (7), e1003143.
- Meffin, H., Hietanen, M.A., Cloherty, S.L., Ibbotson, M.R., 2015. Spatial phase sensitivity of complex cells in primary visual cortex depends on stimulus contrast. *J. Neurophysiol.* 114, 3326–3338.
- Milner, A.D., Goodale, M.A., 1995. *The Visual Brain in Action*. Oxford University Press, Oxford.
- Milner, A.D., Goodale, M.A., 2008. Two visual systems re-viewed. *Neuropsychologia* 46 (3), 774–785.
- Movshon, J.A., Newsome, W.T., 1996. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* 16 (23), 7733–7741.
- Movshon, J.A., Thompson, I.D., Tolhurst, D.J., 1978. Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol.* 283 (1), 79–99.
- Movshon, J.A., Adelson, E.H., Gizzi, M.S., Newsome, W.T., 1986. The analysis of moving visual patterns. In: Chagas, C., Gattass, R., Gross, C. (Eds.), *Pattern Recognition Mechanisms*. Springer, New York, pp. 117–151.
- Murphy, P.C., Sillito, A.M., 1987. Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature* 329, 727–729.
- Nathans, J., 1999. The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron* 24 (2), 299–312.
- Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381 (6583), 607.
- Pack, C.C., Gartland, A.J., Born, R.T., 2004. Integration of contour and terminator signals in visual area MT of alert macaque. *J. Neurosci.* 24 (13), 3268–3280.
- Pasupathy, A., Connor, C.E., 1999. Responses to contour features in macaque area V4. *J. Neurophysiol.* 82 (5), 2490–2502.
- Pasupathy, A., Connor, C.E., 2002. Population coding of shape in area V4. *Nat. Neurosci.* 5 (12), 1332.
- Przybylski, A.W., Gaska, J.P., Foote, W.E., Pollen, D.A., 2000. Striate cortex increases contrast gain of macaque LGN neurons. *Vis. Neurosci.* 17 (4), 485–494.
- Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2 (1), 79.
- Rieke, F., Warland, D., Van Steveninck, R.D.R., Bialek, W.S., 1999. *Spikes: Exploring the Neural Code*, vol. 7. MIT Press, Cambridge, No. 1.
- Rodieck, R., Watanabe, M., 1993. Survey of the morphology of macaque retinal ganglion cells that project to the pretectum, superior colliculus, and parvocellular laminae of the lateral geniculate nucleus. *J. Comp. Neurol.* 338 (2), 289–303.

- Roe, A.W., Chen, G., Lu, H.D., 2009. Visual system: functional architecture of area V2. In: Squire, L. (Ed.), *Encyclopedia of Neuroscience*, vol. 10, pp. 333–349.
- Rust, N.C., DiCarlo, J.J., 2010. Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* 30 (39), 12978–12995.
- Sharpee, T.O., 2013. Computational identification of receptive fields. *Annu. Rev. Neurosci.* 36, 103–120.
- Solomon, S.G., Lee, B.B., White, A.J., Rüttiger, L., Martin, P.R., 2005. Chromatic organization of ganglion cell receptive fields in the peripheral retina. *J. Neurosci.* 25 (18), 4527–4539.
- Spillmann, L., Dresch-Langley, B., Tseng, C.H., 2015. Beyond the classical receptive field: the effect of contextual stimuli. *J. Vis.* 15 (9), 7–7.
- Stuart, G., Spruston, N., Häusser, M. (Eds.), 2016. *Dendrites*. Oxford University Press.
- Tanaka, K., 1996. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19 (1), 109–139.
- Tanigawa, H., Lu, H.D., Roe, A.W., 2010. Functional organization for color and orientation in macaque V4. *Nat. Neurosci.* 13, 1542–1548.
- Tootell, R.B., Nelissen, K., Vanduffel, W., Orban, G.A., 2004. Search for color ‘center(s)’ in macaque visual cortex. *Cerebr. Cortex* 14, 353–363.
- Van Essen, D.C., Maunsell, J.H., 1983. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci.* 6, 370–375.
- Verhoef, B.E., Bohon, K.S., Conway, B.R., 2015. Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. *J. Neurosci.* 35, 6952–6968.
- von der Heydt, R., 2015. Figure–ground organization and the emergence of proto-objects in the visual cortex. *Front. Psychol.* 6, 1695.
- Wässle, H., 2004. Parallel processing in the mammalian retina. *Nat. Rev. Neurosci.* 5, 1–11.
- Wool, L.E., Crook, J., Troy, J.B., Packer, O.S., Zaidi, Q., Dacey, D.M., 2018. Nonselective wiring accounts for red-green opponency in midget ganglion cells of the primate retina. *J. Neurosci.* 38 (6), 1520–1540.
- Zhou, H., Friedman, H.S., von der Heydt, R., 2000. Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20, 6594–6611.
- Ziomba, C.M., Freeman, J., Movshon, J.A., Simoncelli, E.P., 2016. Selectivity and tolerance for visual texture in macaque V2. *Proc. Natl. Acad. Sci. U.S.A.* 113 (22), E3140–E3149.
- Zoccolan, D., Kouh, M., Poggio, T., DiCarlo, J.J., 2007. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J. Neurosci.* 27 (45), 12292–12307.