Bio-Inspired Computer Vision: Setting the Basis for a New Departure

N V Kartheek Medathati, Heiko Neumann, Guillaume S. Masson, Pierre Kornprobst

Bio-Inspired Computer Vision: Setting the Basis for a New Departure

N. V. Kartheek Medathati, Heiko Neumann, Guillaume S. Masson, Pierre Kornprobst

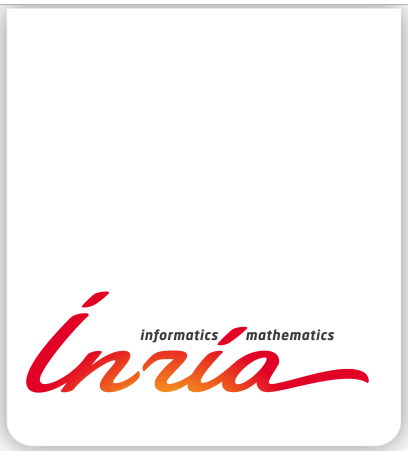
RESEARCH REPORT

N° 8698

March 2015

Project-Team Neuromathcomp

ISSN 0249-6399 ISRN INRIA/RR--8698--FR+ENG



Bio-Inspired Computer Vision: Setting the Basis for a New Departure

N. V. Kartheek Medathati*, Heiko Neumann†, Guillaume S. Masson‡, Pierre Kornprobst*

Project-Team Neuromathcomp

Research Report n° 8698 — March 2015 — 54 pages

Abstract: Studies in biological vision have always been a great source of inspiration for design of computer vision algorithms. In the past, several successful methods were designed with varying degrees of correspondence with biological vision studies, ranging from purely functional inspiration to methods that utilize models that were primarily developed for explaining biological observations. Even though it seems well recognized that computational models of visual cortex can help in design of computer vision algorithms, it is a non-trivial exercise for a computer vision researcher to mine relevant information from biological literature as very few studies in biology are organized at a task level. This has lead to a widening gap between biological vision and computer vision research. In this paper we aim to bridge this gap by provide insights and methodology to envision a new departure in this area. Not only we revisit some of the main features of biological vision and discuss the foundations of existing computational studies modeling biological vision, but also we revisit three classical computer vision tasks from a biological perspective: image sensing, segmentation and optical flow. Using this task-centric approach, we discuss well-known biological functional principles and compare them with approaches taken by computer vision. Based on this comparative analysis of computer and biological vision, we will present some recent promising approaches in modelling biological vision and we highlight a few novel ideas that we think are promising for future investigations in computer vision. To this extent, this papers provides new insights for the design of biology-based computer vision algorithms and pave a way for much needed interaction between the two communities.

Key-words: Bio-inspired vision, sensing, segmentation, optical flow, retina, dorsal stream, ventral stream

* Inria, Neuromathcomp project team, Sophia Antipolis, France
† Ulm University, Ulm, Germany
‡ Institut de Neurosciences de la Timone, CNRS, Marseille, France

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Vision Bio-Inspirée: Poser les Bases pour un Nouveau Départ

Résumé : L'étude de la vision biologique a toujours été une grande source d'inspiration pour la conception d'algorithmes de vision par ordinateur. Dans le passé, plusieurs méthodes ont été conçues avec succès, avec des degrés variables de correspondance avec les études de la vision biologique, allant de l'inspiration purement fonctionnelle à des procédés qui utilisent des modèles développés principalement pour comprendre les observations biologiques. Même s'il semble bien reconnu que les modèles inspirés du cortex visuel peuvent aider dans la conception d'algorithmes de vision par ordinateur, un exercice non trivial pour un chercheur en vision par ordinateur est de savoir extraire les informations pertinentes de la littérature biologique qui ne s'intéresse que très rarement à la résolution de tâche. Ceci a conduit à un élargissement du fossé entre les recherches menées en vision biologique et en vision par ordinateur. Dans cet article, nous visons à combler cette lacune en procédant à une présentation de la littérature récente en vision biologique orientée vers la résolution de tâches et en fournissant les pointeurs sur des découvertes récentes décrivant les processus sous-jacents. Non seulement nous revisitons certaines des principales caractéristiques de la vision biologique et discutons du fondements des études computationelles modélisant la vision biologique, mais aussi nous revisitons trois tâches classiques en vision par ordinateur avec un point de vue biologique: l'acquisition d'images, la segmentation et le flot optique. En utilisant cette approche orientée vers la résolution des tâches, nous discutons des principes fonctionnels biologiques connus pour les comparer avec les approches proposées en vision par ordinateur. Sur la base de cette analyse comparative entre vision biologique et artificielle, nous présentons des approches prometteuses récentes en modélisation de la vision biologique et nous soulignons des idées nouvelles qui nous paraissent prometteuses pour les recherches futures en vision par ordinateur. En ce sens, ce papier offre de nouvelles perspectives pour la conception d'algorithmes de vision inspirés de la biologie et il ouvre une voie à une interaction indispensable entre les modélisateurs des deux communautés.

Mots-clés : Vision bio-inspirée, acquisition, segmentation, flot optique, retina, voie dorsale, voir ventrale

Contents

1 Introduction

Biological vision systems are remarkable at extracting and analysing information essential for major functional needs such as spotting food [81], evading predators [158], identifying mates [146] and navigating through complex environments [217]. Given the functional importance of vision in a vast majority of animals, it is fascinating to watch these complex biological systems to solve to extract a huge amount of semantic information (e.g., scene category, presence absence or location of an object) from only a glimpse at very brief presentations of the visual scene (20ms–150ms) while an increasing amount of details will be then gathered as the time of presentation increases [306, 174, 156, 92]. However, the ease and speed at which humans, for instance capture these behavioural-relevant information from a cluttered, dynamical natural scene might have deceptively hidden the complexity of the underlying versatile biological vision systems. Such computational complexity is certainly much better acknowledged when we consider that, in primates, visual cortices fill up about half of the neocortex [93] and that, for instance, the macaque area V1 is made of nearly 1.5 billions of neurones. The fact that such a dense network of about 30 cortical areas is fed by barely one million geniculo-cortical axons representing only five percent of the synapses in the primary visual cortex further give a hint on the complexity of the biological machinery!

It is also remarkable that biological visual systems perform these tasks with both high sensitivity and strong reliability given the fact that their inputs are highly noisy, variable and ambiguous. Biological systems can efficiently and quickly solve many difficult computational problems that are still challenging for artificial systems such as the aperture and correspondence problems, image segmentation/integration, 3D reconstruction or the interpretation of complex biological objects or movements. All these aspects have been intensively investigated in human psychophysics (compare for instance [217, 225], two majors reviews on visual motion separated by 25 years)

and the neuronal underpinnings of visual performance have been scrutinised over a wide range of temporal and spatial scales, from single cell to large cortical networks (see [58] for an encyclopaedic review) so that visual systems are certainly the best-known of all neural systems.

Pioneers in computer vision dreamt of building machines that could match and perhaps outperform biological vision [222]. Indeed, several computer vision algorithms have demonstrated very high performance levels, in particular for object or face recognition [70]. Recently, deep learning approaches have been proved to bring these performances to an unprecedented level [133] and have attracted attention, and interest from far outside the research community as potential applications are tremendous [28, 34].Given these successes, one might ask whether computer vision, and more generally machine learning still needs anything from neuroscience. There are several reasons to temper this enthusiasm [311, 70]. First, the gap between humans and machines is still great, in particular when one consider the limits of the machine learning algorithms such as the size of the dataset, the close-set problems or the nature of representations. Second, the hallmark of human vision is its generality. Therefore, one shall not directly compare uni and "multitaskers" vision systems. Third, biological vision systems deal easily with dynamical inputs while computer vision systems have still difficulties to recover information from movies. Fourth, biological processes appear to adapt themselves dynamically while computer solutions are most often fixed and static. Fifth, biological vision systems do not work in isolation but rather are integrated in larger networks concerned with different tasks from the behavioural repertoire. There are constantly interacting with memory, attention or decision systems. All these aspects highlight the fact that computer vision would still need, and may be more than ever, to maintain strong links with visual and computational neurosciences.

It would seem natural to expect that the two fields of biological and computer vision would interact continuously since they target the same goals at task level: extracting and representing meaningful visual information. However, historically the strength of these interactions has continuously declined since the pioneering work of David Marr [193] who attempted to bring together the fields of neurobiology, visual psychophysics and computer vision. The unifying idea presented in his influential book entitled Vision was to articulate these fields around computational problems faced by both biological and artificial systems rather than on their implementation. This problem-based approach has inspired a decade of experimental and computational studies on, for a few instances, visual motion [132, 41, 200], object recognition [261, 79] or biological motion [106]. However, and despite these efforts, the two fields drifted apart. We believe that such disunion can be attributed to three main factors. The first factor is that, for decades, only a limited range of tools were available to understand how the brain process visual information. To give a short historical perspective, one can cite two hallmarks on the development of visual neurosciences. In 1967, Granit and Hartline shared the Nobel prize for the first electrical recording of light responses from individual ganglion cells of the vertebrate retina. This was the foundation of the receptive field doctrine: how a single cell can extract and represent visual information from the light pattern falling in a particular retinal location. Moreover, the theoretical understanding of the receptive field concept was largely borrowed from computer vision [273, 157] so that biological vision lost his role in inspiring algorithmic solution to engineers. In 1981, Hubel and Wiesel were awarded with the Nobel prize for their work establishing how these neurones specifically encode low-level visual attributes such as orientation, direction or binocular disparity and how they form functional sub-populations paving the different dimensions of visual information. Characterising cells as independent processors now seems to be an heroic effort to quantify input-output transfer functions and this approach deems to fail as we move along the cortical hierarchies. Several pitfalls have been identified since then, as illustrated by highly disputed debates such as

the use of natural versus synthetic images or the respective role of feedforward, feedback and intra-cortical connectivities [91, 55]. The second explanation comes from the limited computational power which did not allow complex simulations to model and simulate bio-inspired systems with the desired speed and accuracy. A large class of models were designed with arbitrary rules that were neither inspired by biological data nor biologically realistic. The dynamical relationships between network architectures, population activity and behavioural performance was largely overlooked, due to several limits in computational power as well as in the availability of relevant biological data. Moreover, many models of visual perception shared a common theoretical ground (e.g., the divisive normalisation) but each one is designed to tackle only a sub-set of problems, leading to the "one problem-one model" fragmentation of both fields. On the other hand, current attempts to build detailed models of biological mechanisms face the tantalising problem of the huge heterogeneity of biological data collection so that the search for common principles was largely lost despite the attempts to propose canonical circuits of cortical processing [82] and large-scale architectures [93].

The third reason can be explained by the market needs for task level solutions fuelled by automation of various industrial activities that must deal with an ever increasing amount of image content. This has led to fragmented solutions addressing a variety of problems and to practices such as sharing code [42, 323, 60] and consistent benchmarking of algorithms using publicly available task specific datasets [195, 9, 90, 337, 165, 175, 14, 254].

The field of computer vision has overcome many other challenges to meet the needs of variety of applications and fostered entirely new scientific and technological areas such as content-based image and video retrieval, object recognition and data mining, computer graphics, and medical image analysis [340]. However, in spite of the extraordinary advancements made over the last two decades, there remain many open challenges in computer vision and for many tasks automatic vision systems are still far from human performance. Only to mention, robustness and system level integration of various algorithms still remains a major issue [340] and the original, ambitious goal of building machines that could outperform human vision has not been reached outside of some very specific tasks such as face recognition.

In the context of building robust systems, biological vision could serve again as a rich source for designing principles [242, 100, 311, 70]. There are several reasons for this new wave. First, multiple scales functional analysis and connectomics have been exploding in brain sciences, and visual systems are upfront on this move [84]. Nowadays, it has become possible to identify selective neuronal populations and dissect out their circuitry at synaptic level by combining functional imaging and serial electron microscopy. Interestingly, the first series of studies have focused on visual circuits, at retinal [128] and cortical [33] levels. At a wider scale, a quantitative description of the connectivity patterns between cortical areas is now becoming available and, again the study of visual cortical networks is pioneering [191]. As a consequence, detailed small and large scales models of visual processing are now becoming available [155, 249]. More discoveries lie ahead with the recent announcements of the BRAIN (Brain Research through Advancing Innovative Neurotechnologies) and HBP (Human Brain Project) projects. Second, the progress achieved in computer architectures make it now possible to simulate large-scale models, something that was not even possible to dream of a few years ago. For example, the advent of multi-core architectures [85], parallel computing on clusters [245], GPU computing [17, 244] and availability of neuromorphic hardware [301] promises to facilitates the exploration of truly bio-inspired systems. Here again, visual tasks are once again upfront [209]. Last, and not the least, the theoretical difficulties encountered by each field call for a new, interdisciplinary approach for understanding how we process, represent and use visual information. For instance, it is still unclear how the dense network of cortical areas analyse the structure of the external world and part of the problem may come from

using a bad range of framing questions about what high-level vision is for [69]. In short, we cannot see the forest (representing the external world) for the trees (e.g., face and object recognition).

The general purpose of this review article is to show how novel approaches could be developed from biological insights. It is our conviction that the current knowledge about biological vision, new simulation technologies and identification of some ill-posed problems have reached a point to envision a new departure for a truly interdisciplinary range of approaches. Here we have three main objectives. The first one is to give to the computer vision community a comprehensive view of the major properties of biological vision that one should share. This modern view goes much beyond the classical view of the brain as a hierarchical feedforward system [164]. We will highlight an essential set of recent experimental and review articles that understand the primate visual system as a dynamical network where feedforward but also feedback and intra-cortical inputs all play a role. Our second goal is to re-appraise the task-oriented approach which is the natural way of thinking in both computer and biological vision community, by isolating key computational elements and architectures together. The third objective is then to discuss how these tasks could be solved not only at conceptual level but also at computational level, showing how neural computations can be modelled thus allowing to provide new testable ideas to the computer vision community. To do so, we will focus on three key tasks for both biological and artificial systems: sensing the optic array, segmenting figures from ground and estimating the optical flow.

This article is composed as follows. In Sec. 2, we revisit the classical view of the brain as a hierarchical feedforward system, pointing out its limitations. In Sec. 3, we discuss the frameworks used to study biological vision from a computational perspective and we emphasize the importance of putting always the task as the main motivation to look into biology. In Sec. 4, in order to relate studies in biological vision to computer vision, we focus on three tasks: sensing, segmentation and motion. These three tasks have been selected for their variability in the nature of the computational challenges and the involvement of different cortical processing streams. For each task, we will start by highlighting some of the key mechanisms that have been identified in biological vision. We will give a structural view of these mechanisms, relate these structural principles to prototypical models from both biological and computer vision and, finally we will detail potential insights and perspectives for rooting new approaches on the strength of both fields. Finally, in Sec. 5, based on the exemple tasks reviewed throughout this paper, we will discuss some possibilities for the future of bio-inspired vision as a re-emerging discipline.

2 Deep cortical hierarchies?

2.1 A classical view of biological vision: hierarchical and feedforward

The classical view of biological vision processing which is conveyed to the computer vision community by visual neurosciences is that of an ensemble of deep cortical hierarchies. Interestingly, this idea was proposed in computer vision by David Marr [193] even before the anatomical hierarchy was detailed in different species. Nowadays, there is a general agreement about this hierarchical organisation and its division into parallel streams, as supported by a large body of anatomical and physiological evidences (see [93, 319, 321, 320] for reviews).

The parallel streams of hierarchically organised cortical areas in human and non-human primates are illustrated in Fig. 1(a). In this canonical representation, information flows from the retina to the primary visual cortex (area V1) through two parallel retino-geniculo-cortical pathways [183]. The magnocellular (M) pathway conveys coarse, luminance-based spatial inputs with a strong temporal sensitivity towards Layer 4Cα of area V1 where a characteristic population of cells, called stellate neurones, immediately transmit the infor-

mation to higher cortical areas. A slower, parvocellular (P) pathway conveys retino-thalamo-cortical inputs with high spatial resolution but low temporal sensitivity, entering area V1 through the layer 4Cβ. Such color-sensitive input flows more slowly within the different layers of V1 and then to cortical area V2 and a network of cortical areas involved in form processing [183, 77].

Such dichotomy in the retino-thalamo-cortical processing pathway resonated with neuropsychological studies investigating the effects of parietal and temporal cortex lesions [318], leading to the popular, but schematic, two visual systems theory [318, 110, 319, 211]. As illustrated in Fig. 1(b), visual information processing is split between a dorsal and a ventral stream. The dorsal stream is specialised in motion perception and the analysis of the spatial structure of the visual scene and involved a occipito-parietal network including areas in the median and median temporal sulcus (areas MT and MST), the intra-parietal sulcus (VIP, LIP) and the posterior part of the parietal cortex. The ventral stream on the other hand is specialised in form perception, including object and face recognition, and relies on a temporal-occipital network of areas (areas TEO, AIT and TE in particular).

This hierarchical view was reinforced by the linear systems approach to understand visual processing (see [87, 138] for an historical perspective). As illustrated in Fig. 1(c), neurones in the primary visual system have small receptive fields paving a high resolution retinotopic map. The spatiotemporal structure of the receptive field corresponds to a processing unit that can locally filter a given property of the image. In V1, low-level features such as orientation, direction, color or disparity are encoded in different sub-populations forming a sparse and overcomplete representation of local feature dimensions [55]. These representations feed several, parallel cascades of converging influences so that, as one moves along the hierarchy, the receptive fields become larger and larger and encode for features of increasing complexities and conjunctions thereof (see [77, 269] for reviews). Causal relationships between neuronal selectivities and local motion or optic flow perception have been demonstrated at these different stages, from low (MT) to high (parietal cortex) stages along the hierarchy. For instance, along the motion pathway, V1 neurones are weakly direction-selective but converge onto MT cells where direction and speed are encoded in a form-independent manner. These cells project to MST neurones where receptive fields cover a large portion of the visual field and encode basic optic flow patterns such as rotation, translation or expansion. More complex flow fields can be decoded by parietal neurones when integrating these informations and be integrated with extra-retinal signals about eye movements or self-motion [41, 232]. The same logic flows along the form pathway, where V1 neurones encode the orientation of local edges. Through a cascade of convergence, units with receptive fields sensitive to more and more complex geometrical features are generated so that IT neurones are able to encode objects or face in a viewpoint invariant manner (see Fig. 1).

Global motion, as well as object or face recognition are prototypical examples where the canonical view of hierarchical, feed-forward processing nearly perfectly mixes anatomical, physiological and perceptual data. This synergy has resulted in realistic, computational models of receptive fields where converging outputs from linear filters are nonlinearly combined from one step to the subsequent one (see [276, 212, 52, 218] for a few recent examples in motion and object processing). It has also inspired feed-forward models working at task levels, as for instance global motion perception [171, 106] or rapid object categorisation [285, 284] as illustrated in Fig. 1(d), prominent machine learning solution for object recognition follow the same feedforward, hierarchical architecture where linear and nonlinear stages are cascaded between multiple layers representing more and more complex features [133, 70].

2.2 Going beyond the hierarchical feedforward model

Despite its success in explaining rapid motion perception or object recognition, for instance

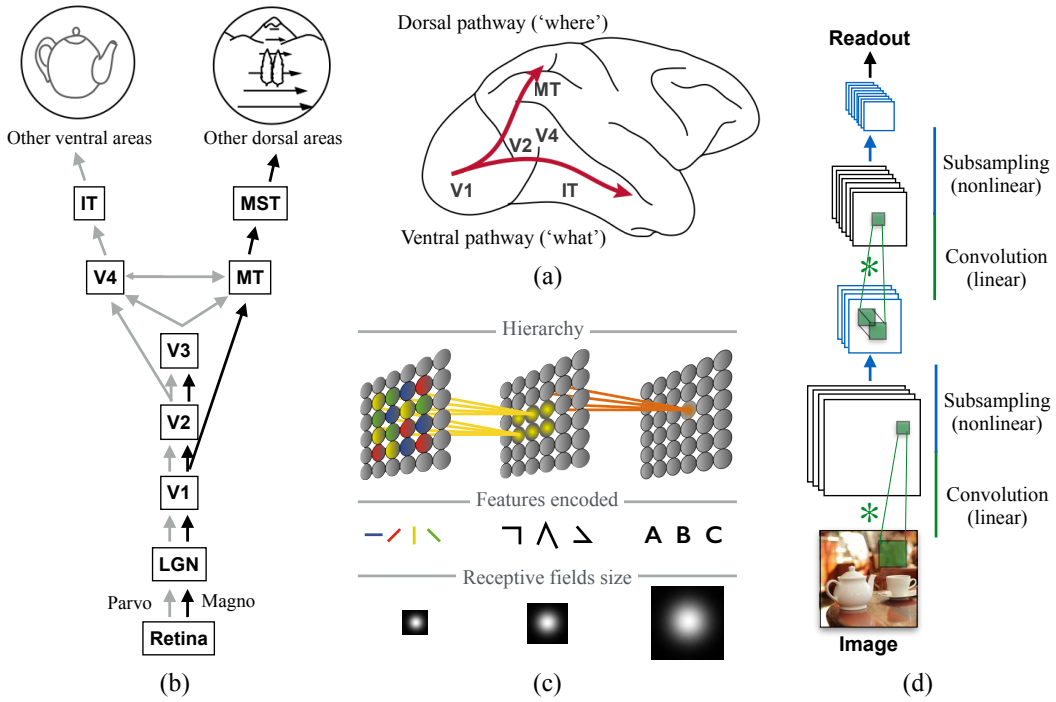


Figure 1: The classical view of hierarchical feedforward processing. (a) The two visual pathways theory states that primate visual cortex can be splitter between dorsal and ventral streams originating from the primary visual cortex (V1). The dorsal pathway runs towards the parietal cortex, through motion areas MT and MST. The ventral pathway propagates through area V4 along the temporal cortex, reaching area IT. (b) These ventral and dorsal pathways are fed by parallel retino-thalamo-cortical inputs to V1, known as the Magno (M) and Parvocellular pathways (P). (c) The hierarchy consists in a cascade of neurones encoding more and more complex features through convergent information. By consequence, their receptive field integrate visual information over larger and larger receptive fields. (d) A machine learning algorithm following the same hierarchical processing where a simple feedforward convolutional network implements two bracketed pairs of convolution operator followed by a pooling layer. Adapted from [70].

the hierarchical feedforward model of the parallel cortical streams remains highly schematic. By consequences, many aspects of biological visual processing, from anatomy to behaviour, do not fit in this cartoon-like framing. Moreover, bio-inspired models would requires certain level of abstraction, resulting in loss of many details, or fact that were first wrongly considered as details. In reality neural processing is much more complex than the hierarchical feedforward abstraction and very important connectivity patterns such as lateral and feedback interactions need to taken into account to overcome several pitfalls in understanding and modelling biological vision. In this section, we highlight some of these key features. We postulate that, when they are carefully considered, they would greatly influence computational models of visual processing and shall be bear in mind when claiming for bio-inspiration. We also believe that identifying some of these problems could help in reunifying natural and artificial vision.

Vision processing starts at the retina and LGN levels First, thinking about biological vision only focusing on the cortical hierarchy would be ignoring the two crucial steps happening before, namely at the retina and lateral geniculate nucleus (LGN) levels. This statement may appear to be obvious but, in fact the role played by these two structures seems have been largely underestimated, when not put aside voluntarily. Indeed, most current models take images as inputs rather than their retina-LGN transforms. Thus, by ignoring what is being processed at these levels, one could easily miss some key properties to understand the system efficiency.

At the retina level, the incoming light is transformed into electrical signals. This transformation was originally dissected using the linear systems approach [87]. The retinal network was seen as a smart and adaptive filter for luminance and color information. Recent works have changed this view and several cortex-like computations have been identified in the retina of different vertebrates. For instance, in rodents, direction and orientation selectivity of some V1 neurones may be in fact partly inherited from specific classes of retinal ganglion cells [72]. Moreover,it is now recognise that retinal circuits implement complex processing such as efficient coding of natural scenes, prediction and features detection (see [109, 152] for reviews, and more details in Sec. 4.1). What is already done at retinal level and what is delegated to cortical stages is still highly disputed with clear differences between species, as for instance the existence of direction selectivity in lower mammals but not in primates. Still, the fact that retinal and cortical levels share similar computational principles, albeit working at different spatial and temporal scales is an important point to consider when designing models of biological vision [152].

In the same perspective, the LGN and other visual thalamic nuclei (e.g., pulvinar) are no longer seen as a pure relay on the route from retina to cortex. For instance, cat pulvinar neurones exhibit some properties classically attributed to cortical neurones, as such pattern motion selectivity [208]. Moreover, there are strong center-surround interactions in LGN neurons and these interactions are under the control of feedback cortico-thalamic connections [148]. These strong feedback connections to the LGN from multiple cortical visual areas, in lower mammals as well as in non-human primates, point to the fact that retino-thalamo-cortical pathways must be seen as highly adaptive, dynamical systems rather than mere passive transmission cables and relays [214, 73, 44]. These two properties (center-surround interactions and feedback modulation) shape the dynamical properties of the cortical inputs, such as for instance the temporal precision of thalamic firing responses [5].

Overall, recent advances in visual neurosciences leave us with three important messages. First, we should not oversimplify the amount of processing done before visual inputs reach the cortex. Rather, we should consider how the cortex takes advantage of them when processing naturalistic images. Second, some of the computational and mechanistic rules designed for predictive-coding or feature extraction can be much more generic than previously thought and the retina-LGN processing hierarchy may become again a rich source of inspiration for computer vision. Third, the exact implementation (what is being done and where) may be not so important as it varies from one specie to another but the cascade of basic computational steps may be an important principles to retain from biological vision.

Functional and anatomical hierarchies are not always the same The deep cortical hierarchy is first rooted on anatomical, connectivity rules [341]. It has a functional counterpart where there is an increasing complexity of processing and information content as we go deeper in the anatomical hierarchy. However, if functional complexity of visual processing do tend to increase from striate to extra-striate and associative cortices, this observation does not imply that a feedforward hierarchy is the ultimate solution used by biological systems. A quick glance at the actual cortical connectivity pattern in non-human primates would be sufficient to eradicate this textbook view of how

the visual brain works [126].

For instance, the classical view that primary visual cortex represents luminance-based edges whereas illusory contours would be represented at the next processing stages such as areas V2 and V4 [239] was defeated. Recent studies have shown that illusory contours, and border ownerships are also represented in macaque area V1 [343, 173]. Similarly, the hierarchy of shape representation seems nowadays to be more opaque than previously thought [127]. Another example comes from motion processing. It has been believed for decades that, contrary to MT neurones, V1 cells cannot encode motion speed independently of the image spatiotemporal frequencies content [264]. However, more recent studies have shown that some V1 complex cells are speed tuned [252]. Moreover, while there is plenty of evidence for local motion analysis in area MT, thus often called the cortical motion area, recent experiments have proposed a more complex view where global motion is not encoded in area MT [124] but would also imply cortical areas along the ventral stream [344]. Lastly, the temporal hierarchy appears not to be a carbon copy of the anatomical hierarchy depicted by Felleman and Van Essen. The onset of a visual stimulus trigger fast and slow waves of activation travelling throughout the different cortical areas. The fast activation in particular by-passes several major steps along both dorsal and ventral pathways to reach frontal areas even before area V2 were fully activated (for a review, see [169]).

These different examples show that a more complex view of the functional hierarchy is emerging. Various aspects of dynamics result from interactions of various streams operating at different speeds and intra-cortical feedback needs to be taken into account when designing better vision algorithms. Yet, a modern theory of the primate visual systems is still awaited, going beyond the classical view based on static, feedforward connectivity rules.

Dorsal/ventral separation is a simplification A possible limitation of starting from an anatomical view is that its complexity usually requires undesired simplifications to present a coherent view. Such reduction introduce however the risk to base models on a caricatural description of the system. One striking example is the common dorsal/ventral separation which is, in fact not so strict. For instance, it has long been known from motion psychophysics that there is a strong influence of form signals onto motion processing (see [202] for a recent review). The interplay between form and motion cues seems to exist at different levels of the hierarchy, from V1 to STS, bringing to the visual system computational advantages that an artificial system could also benefit from (see [232] for a review). Just to give another exemple, it appears nowadays that there are strong interactions between color and motion information, through mutual interactions between cortical areas V4 and MT [302]. It is interesting to note that these two particular areas were previously attributed to the ventral and dorsal pathways, respectively [183, 77]. These interactions can be tracked to the primary visual cortex where detailed analysis of the layer 4 connectivity have shown that both Magno and Parvocellular signals can be intermixed and propagated to areas V2 and V3 and the subsequent ventral stream [338]. This could explain why fast and coarse visual signals can rapidly tune the most ventral areas along the temporal cortex and therefore shape face recognition mechanisms [106].

Thus, if the coarse division between ventral and dorsal streams remains valid, a closer look at functional interactions highlight the existence of multiple passerelles, occurring at many levels along the hierarchy. Moreover, it seems that each stream is traversed by successive waves of fast/coarse and slow/precise inputs so that visual processing is gradually shaped. Lastly, and this will be further discussed below, it now seems indispensable to consider the intricate networks of intra and inter-cortical interactions to capture the dynamics of biological vision. We still do not understand how the two visual systems can work hand-to-hand to resolve many of the input ambiguities and elaborate an integrated view point. Clearly, a new theoretical perspective on the cortical

functional architecture would be highly beneficial to both biological [192] and artificial vision research [106, 70].

A hierarchy embedded within a dynamical recurrent system The hierarchical view conceals the dynamic nature of information processing carried out by the brain. We have recalled above that the static and temporal hierarchy does not necessarily coincide as information flows can bypass some cortical areas [169]. This has led to the idea that fast inputs, carried by the Magnocellular stream, can travel across the cortical networks to shape each processing stage before being reached by more detailed information carried by the Parvocellular retino-thalamo-cortical pathway [49]. This dynamical view is still consistent with the feedforward hierarchical processing framework adopted by most computational models and could explain fast, automatic recognition mechanisms underlying categorisation (see [274, 305] for reviews).

One important aspect of visual information is that it is highly ambiguous because it comes from projecting 3D scenes onto one, or two 2D retinal images. Therefore, a given image or sequence of images can lead to many different interpretations, as evidenced by perceptual multi-stability. How the brain solves these ambiguities remains largely a mystery, but several theoretical works have proposed that its highly recurrent connectivity might play a crucial role [112, 214, 215, 83, 315, 172]. For instance, the model proposed by Lee and Mumford [172] reconsiders the hierarchical framework by proposing that concatenated feedforward/feedback loops in the cortex could serve to integrate top-down prior knowledge with bottom-up observations. This architecture generates a cascade of optimal inference along the hierarchy [274, 305, 172]. Several computational models have used such recurrent computation for motion integration [23, 46, 308] or figure-ground segmentation [270].

Interestingly, new experimental evidence for the role of feedback in these tasks and its mechanisms begin to be available in awake, behaving monkeys. Feedback from the first extrastriate stages of visual processing (e.g., areas V2, V3, MT) have been shown to modulate center-surround interactions in the primary visual cortex. This could impact suppressive surround mechanisms [221] and spatial normalisation [220], with functional implications for texture segmentation [140], figure-ground segregation [282], surface reconstruction and border-ownership [145]. Moreover, feedback signals can also carry attention-based modulation of low and mid-level visual processing such as curve tracing or object segmentation from higher-order visual areas (e.g., V4) or frontal cortex areas (e.g., FEF, PFC; see [247, 10]). Across different scales, there is now a growing body of evidence for different physiological signatures (e.g., single and multiple neurones activity, LFPs, BOLD responses) of these top-down effects, so that the temporal dynamics of visual processing can be dissected out with an unprecedented precision (see [99] for a review). Overall, early visual processing appears now to be strongly influenced by different top-down signals about attention, working memory or even reward mechanisms, opening the door for a more broader and integrative perspective on visual perception where both sensory inputs and context-related information must be taken into account when, for instance, modelling object segregation and selection (see [296, 169] for reviews). Despite these, still sparse successes, investigating the role of feedback and its detailed implementation, have been obstructed by strong technical difficulties. Thee emergence of genetically-encoded optogenetic probes targeting the feedback pathways in mice cortex are opening a new era of intense research about the role of feedforward and feedback circuits (see [141, 186] for recent reviews).

The role of lateral connectivity in information diffusion One last facet of the cortical dynamics is the contextual modulation of information processing through short- and long-range intra-cortical interactions. Processing of a local feature is always influenced by its immediate surrounding in the image. Feedback is one potential mechanisms for implementing such context-dependent processing but

its spatial scale is rather large, corresponding to far-surround modulation [7, 6]. Visual cortical areas, and in particular area V1, are characterised by dense short- and long-range intra-cortical interactions. Short-range connectivities are involved in proximal center-surround interactions and their dynamics fits with contextual modulation of local visual processing [7]. This connectivity pattern has been overly simplified as overlapping, circular excitatory and inhibitory areas of the non-classical receptive field. In area V1, these sub-populations were described as being tuned for orthogonal orientations corresponding to excitatory input from iso-oriented domains and inhibitory input from cross-oriented ones. In higher areas, similar simple schemes have been proposed with for instance opposite direction tuning of center and surround areas of MT and MST receptive fields [36]. Moreover, these surround inputs implement basic computation such as normalisation or gain control [56].

Nowadays, a more complex picture of the center-surround interactions have emerged where non-classical receptive fields are highly diverse in terms of shapes or features selectivity [57, 329, 336]. This complexity could result from more complex underlying connectivity patterns where, for instance, different orientation- or direction-selective neurones can be interconnected. Interestingly, in area V1 for instance, the connectivity pattern becomes less and less specific with farther distances from the recording sites [61, 72]. Moreover, further distant points of the image can also interact through the long-range interactions which have been demonstrated in area V1 of many species (see, e.g., [263, 262, 38] in monkeys). Horizontal connections extend over millimetres of cortex and propagate activity at a much lower speed than feedforward and feedback connections [49]. The functional role of these long-range connections is still unclear. They most probably support the waves of activity that travel across the V1 cortex [213, 339], in both the presence or absence of visual stimuli (see [278] for a review). They can also implement the spread of cortical activity underlying contrast normalisation [260], the spatial integration of motion and contour signals [260, 107] or the shaping of low-level percepts [143, 63].

3 Computational studies of biological vision

3.1 The Marr Legacy

At a conceptual level much of the current computational understanding of biological vision is based on the seminal theoretical framework defined by Marr [193]. A key message from his book was that complex systems, like brains or computers, must be studied and understood at various levels of description, namely, the computational task carried out by the system resulting in the observed behaviour, the instance of the algorithm used by the system to solve the computational task and the implementation that it used by a given system to execute the algorithm. Thus, once a functional framework is defined, the computational and implementation problems can be distinguished, so that in principle a given solution can be embedded into different biological, or artificial physical systems. This approach has inspired many experimental and theoretical works in the field of vision. The cost of this clear distinction between levels of description is that many many existing models have only a weak relationship with the actual architecture of the visual system or even with a specific algorithmic strategy used by the system. Such dichtomy contrast with the fact that it has became more and more evident over the same period that understanding cortical algorithms and networks are deeply coupled, as illustrated for instance by visual motion processing [132]. The risk of ignoring the structure-function dilemna is that computational principles can drift away from biology, becoming more and more metaphorical as illustrated by the fate of the Gestalt theory.

3.2 From circuits to behavior

Given the multiple scales of the nervous system organisation that we presented in Sec. 2, how shall we rephrase the question of understand-

ing visual computation in both biological and artificial systems? Following others, we state that a key milestone is to understand how neural circuits lead to behaviour. Carandini [54] argued that the gap between circuits and behaviour is too wide without the help of an intermediate level of description, that of neuronal computation. But how can we escape from the dualism between computational algorithm and implementation as introduced by Marr's approach? The solution depicted by Carandini, that is well representative of many current theoretical approaches in neuroscience, is based on three principles. First, some levels of description might not be useful to understand functional problems. In particular sub cellular and network levels are decoupled. It might thus not be important to consider the molecular details of synaptic mechanisms to understand, and model, face recognition. Second, the level of neuronal computation can be divided into building blocks forming a core set of canonical neural computations such as linear filtering, divisive normalization, recurrent amplification, coincidence detection, cognitive maps and so on. These standard neural computations are widespread across sensory systems. Third, these canonical computations occur in the activity of individual neurons and especially of population of neurons. In many instances, they can be related to stereotyped circuits such as feedforward inhibition, recurrent excitation-inhibition or the canonical cortical microcircuit for signal amplification (see [288] for a series of reviews). Thus, understanding the computations carried out at the level of individual neurons and neural populations would be the key for unlocking the algorithmic strategies used by neural systems. This appears to be essential to capture both the dynamics and the versatility of biological vision. Along this perspective, computational vision would regain its critical role when mapping circuits to behaviors and could regain interest in the field of computer vision not only by highlighting the limits of existing algorithms or hardwares but also by providing new ideas. At this cost, visual and computational neurosciences would be again a source of inspiration for computer vision. Figure 2 illustrates this approach to understand biological vision and is of relevance for the computer vision community as it might lead to architecture independent algorithmic principles.

3.3 Functional importance of the task

Biological systems exist to solve functional tasks so that an organism can survive. Thus, many biologists consider the brain as a "bag of tricks that passed evolutionary selection", even though some tricks can be usable in different systems or contexts. This biological perspective highlights the fact that understanding biological systems is tightly related to understanding the functional importance of the task at hands. For example, there is in the mouse retina a cell type able to detect small moving objects in the presence of a featureless or stationary background. These neurons could serve as elementary detectors of potential predators arriving from the sky [342]. In the same vein, it has been recently found that outputs of retinal direction-selective cells are segregrated from the other retino-thalamo-cortical pathways to directly influence specific target neurons in mouse V1 [72]. These two examples illustrate how evolution can shape the nervous systems to isolate neural circuits when computation and architecture are intrinsically coupled to find an optimal solution to the escape behaviours. Another well-known example is the different color sensitivity expressed in different non-human primate species as well as between male and female in some species (see [136, 142] for recent reviews). On the other hand, there is evidence that evolution has also selected neural microcircuits implementing more generic computations such as the recurrent amplification, divisive normalisation or the excitation-inhibition balance that play a key role in the emergence of low-level neuronal selectivities such as orientation or direction. Feedforward-feedback connectivity rules of canonical microcircuits for predictive coding have been also identified [21] and applied in the context of visual motion processing for instance [80]. As another example, divisive nor-

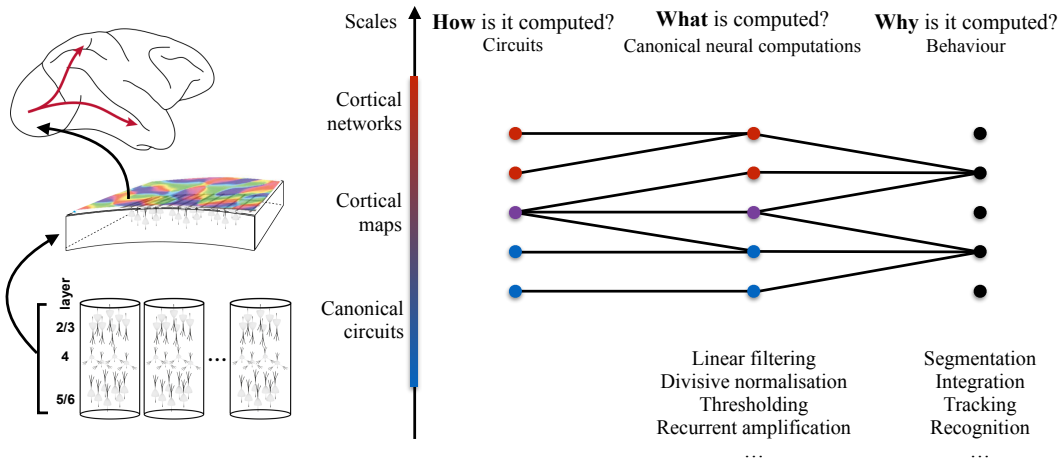


Figure 2: Between circuits and behaviour: reinvigorating the Marr approach. The nervous system can be described at different scales of organisation, from small networks of neurones forming cortical columns, cortical maps and large-scale networks. These levels of organisation can be mapped to three computational problems: how, what and why. All three aspects involve a theoretical description rooted on anatomical, physiological and behaviour data. These different levels are organised around computational blocks that can be combined to solve a particular task, corresponding to the versatility and efficiency of biological systems.

malisation has been a powerful explanation for many aspects of visual perception such as, contrast detection or attention-based modulation, at different neuronal levels as well as for human performance [56]. These examples are extrema lying on the continuum of biological structure-function solutions, from the more specific to the more generic. This diversity stresses the needs to clarify the functional context of the different computational rules and their performance dynamics so that fruitful comparisons can be made between living and artificial systems. This can lead to a clarification about which knowledge from biology is useful for computer vision.

Lastly, these computational building blocks are embedded into a living organism and low-to-high vision levels are constantly interacting with many other aspects of animal cognition [4]. Just to give some example, the way an object is examined (i.e., the way its image is processed) shall depend on its behavioural context, whether it is going to be manipulated or only scrutinised to identify it. A single face can be analyse in different ways depending upon the social or emotional context. Thus, we must consider these contextual influence of "why" a task is being carried out when integrating information (and data) from biology [331]. All these above observations stress the difficulty of understanding biological vision as an highly adapted, plastic and versatile cognitive system where circuits and computation are like Janus face.

3.4 Matching multi-scale connectivity rules and computational problems

In Sec. 2, we have only given a brief and partial glimpse of the enormous literature on the intricate networks underlying biological vision. Focusing on primate low-level vision, we have illustrated both the richness, the spatial and temporal heterogeneity and the versatility of these connections. This is illustrated in Fig. 3 for a simple case, the segmentation of two moving surfaces. Figure 3(a) illustrates the different sub-cortical and cortical stages needed for a minimal model of surface segmen-

tation. Local information is transmitted upstream through the feedforward pathway. In the classical scheme, V1 is seen as a router filtering and sending the relevant information along the ventral (V2, V4) or dorsal (MT, MST) pathways. We have depicted above how information flows also backward within each pathway as well as across pathways (e.g., connections between V2/V4 and MT in Fig. 3). A consequence of these cross-over is that, for instance MT neurones are able to use both motion and color information. We have also recalled that area V1 endorses a more active role where the thalamo-cortical feedforward inputs and the multiple feedback signals interact to implement contextual modulations over different spatial and temporal scales. These local processing are modulated by short and long-range intra-cortical interactions such as visual features located far from the non-classical receptive field (or along a trajectory) can influence them. Each cortical stage implements these interactions although with different spatial and temporal scales and through different domains. This aspect is illustrated for V1 where different maps (e.g., retinotopic and orientation-tuning maps in Fig. 3) are implemented but other maps are found at each cortical stages as well. At the single neurone level, these intricate networks result in a large diversity of receptive field structures and in complex, dynamical non-linearities. It is now possible to collect physiological signatures of these networks at multiple scales, from single-neurons to local networks and networks-of-networks such that connectivity patterns can be dissected out. In the near future, it will become possible to manipulate specific cell subtype and therefore change the functional role and the weight of these different connectivities.

How these connectivity patterns would relate to information processing? In plot b, we sketch the key computational steps underlying surface segmentation. Traditionally, each computational step has been attributed to a particular area and to a specific type of receptive fields. For instance, local motion computation is done at the level of the small receptive fields of V1 neurones. Motion boundaries detectors have been found in area V2 while different subpopulation of MT and MST neurones are responsible for motion integration at multiple scales. However, each of these receptive fields are highly context-dependent, as expected from the dense interactions between all these areas. Matching the complex connectivity patterns illustrated in plot a with the computational dynamics illustrated in plot b is one of the major challenges in computational neurosciences. It can also be a fruitful source of inspiration for computer vision if we were able to draw rules and numbers by which the visual system is organised at different scales. Only a few computational studies have taken into account this richness and its ability to adaptively encode and predict sensory inputs from natural scenes. In the next section we will summarise some of key papers from biological vision literature in a task centric manner in order show how critical information gathered at different scales, and different context can be used to design innovative and performing algorithms.

4 Solving vision tasks with a biological perspective

So far in this paper, we have revisited some of the main features of biological vision and we have discussed the foundations of existing computational studies modelling biological vision. A key idea is the functional importance of the task at hand when exploring or simulating the brain. Thus, one should always focus on the task as the main motivation to look into biology. Following this idea, this section revisits three classical computer vision tasks from a biological perspective: image sensing, segmentation and optical flow. To meet our objective to illustrate some well-known biological functional principles and to compare them with several approaches taken by computer vision, our discussion of each task with be organised as follow:

Task definition We start with a definition of the visual processing task of interest.

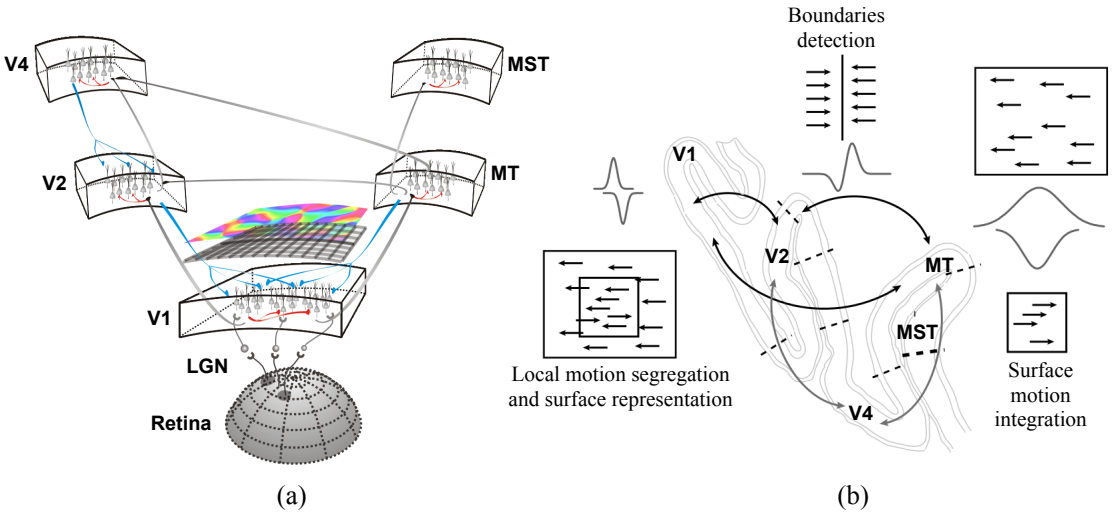


Figure 3: Matching multi-scale connectivity rules and computational problems for the segmentation of two moving surfaces. (a) A schematic view of the early visual stages with their different connectivity patterns: feedforward (grey), feedback (blue) and lateral (red). (b) A sketch of moving surface segmentation and its potential implementation in the primate visual cortex. The key processing elements are illustrated as computational problems (e.g., local segregation, surface cues, motion boundaries, motion integration) and corresponding receptive field structures. These receptive fields are highly adaptive and reconfigurable, thanks to the dense interconnections between the different stages/areas

Core challenges We briefly summarise its physical, algorithmic or temporal constraints albeit and how they impact the processing that should be carried on images or sequences of images.

Biological vision solution We briefly review some of the novel biological facts about the neuronal dynamics and circuitry underlying the biological solutions for these tasks (the How question), stressing the canonical computing elements being implemented (the What question) in some recent computational models.

Comparison with computer vision solutions We discuss some of the current approaches in computer vision with the objective to outline some of their limits and challenges. Contrasting these challenges with known mechanisms in biological vision would be to foresee which aspects are essential for computer vision and which ones are not.

Promising bio-inspired solutions Based on this comparative analysis of computer and biological vision, we will discuss some recent approaches in modelling biological vision and we highlight a few novel ideas that we think are promising for future investigations in computer vision.

4.1 Sensing the environment

Task definition Sensing is the process of capturing patterns of light from the environment so that all the visual information that will be needed downstream to cater the computational/functional needs of the biological system could be faithfully extracted. This definition does not necessarily mean that its goal is to construct a veridical, pixel-based representation of the environment by passively transforming the light the sensor receives.

Core challenges From a functional point of view, the process of sensing (i.e., transducing, transforming and transmitting) light patterns encounters multiple challenges because visual environments are highly cluttered, noisy and diverse. First, illumination levels can vary over several range of magnitudes. Second, image formation onto the sensor is sensitive to different sources of noise and distorsions due to the optical properties of the eye. Third, transducing photons into electronic signals is constrained by the intrinsic dynamics of the photosensitive device, being either biological or artificial. Fourth, transmitting luminance levels on a pixel basis is highly inefficient. Therefore, information must be (pre-)processed so that only the most relevant and reliable features are extracted and transmitted upstream in order to overcome the limited bandpass properties of the optic nerve. At the end of all these different stages, the sensory representation of the external world must still be both energy and computationally very efficient. All these aforementioned aspects raise some fundamental questions that are highly relevant for both modelling biological vision and improving artificial systems.

Herein, we will focus on four main computational problems (what is computed) that are illustrative about how biological solutions can inspire a better design of computer vision algorithms. The first problem is called *adaptation* and explains how retinal processing is adapted to the huge local and global variations in luminance levels from natural images in order to maintain high visual sensitivity. The second problem is *feature extraction*. Retinal processing extracts information about the structure of the image rather than mere pixels. What are the most important features that sensors should extract and how they are extracted are pivotal questions that must be solved to subserve an optimal processing in downstream networks. Third is the *sparseness* of information coding. Since the amount of information that can be transmitted from the front-end sensor (the retina) to the central processing unit (area V1) is very limited, a key question is to understand how spatial and temporal information can be optimally encoded, using context dependency and predictive coding. The last selected problem is called *precision* of the coding, in particular what is the temporal precision of the transmitted signals that would best represent

the seaming-less sequence of images.

Biological vision solution The retina is one of the most developed sensing devices (see [109, 197, 198] for reviews). It transforms the incoming light into a set of electrical impulses, called spikes, which are sent asynchronously to higher level structures through the optic nerve. In mammals,it is sub-divided into five layers of cells (namely, photoreceptors, horizontal, bipolar, amacrine and ganglion cells) that forms a complex recurrent neural network with feedforward (from photoreceptors to ganglion cells), but also lateral (i.e., within bipolar and ganglion cells layers) and feedback connections. The complete connectomics of some invertebrate and vertebrate retinas now begin to be available [190].

Regarding information processing, an humongous amount of studies have shown that the mammalian retina can tackle the four challenges introduced above using *adaptation*, *feature detection*, *sparse coding* and *temporal precision* [152]. *Adaptation* is absolutely crucial since retinas must maintain high contrast sensitivity over a very broad range of luminance, from starlight to direct sunlight. Adaptation is both global through neuromodulatory feedback loops and local through adaptive gain control mechanisms so that retinal networks can be adapted to the whole scene illuminance level while maintaining high contrast sensitivity in different regions of the image, despite their considerable differences in luminance (see [287, 304] for reviews).

It has long been known that retinal ganglion cells extract local luminance profiles. However, we have now a more complex view of retinal form processing. The retina of higher mammals sample each point in the images with about 20 distinct ganglion cells [197, 198] associated to different *features*. This is best illustrated in Fig. 4, showing how the retina can gather information about the structure of the visual scene with four exemple cell types tilling the image. They differ one from the others by the size of their receptive field and their spatial and temporal selectivities. These spatiotemporal differences are related to the different sub-populations of ganglion cells which have been identified. Parvocellular (P) cells are the most numerous are the P-cells (80%). They have a small receptive size and a slow response time resulting in a high spatial resolution and a low temporal sensitivity. They process information about color and details. Magnocellular cells have a large receptive field and a low response time resulting in a high temporal resolution and a low spatial sensitivity, and can therefore convey information about visual motion [286]. Thus visual information is split into parallel stream extracting different domains of the image spatiotemporal frequency space. This was taken at a first evidence for feature extractions at retinal level. More recent studies have shown that, in many species, retinal networks are much smarter than originally thought. In particular, they can extract more complex features such as basic static or moving shapes and can predict incoming events, or adapt to temporal changes of events, thus exhibiting some of the major signatures of predictive coding [109, 197, 198].

A striking aspect of retinal output is its *high temporal precision* and *sparseness*. Massive in vitro recordings provide spiking patterns collected from large neuronal assembles so that it becomes possible to decipher the retinal encoding of complex images [243]. Modelling the spiking output of the ganglion cell populations have shown high temporal precision of the spike trains and a strong reliability across trials. These coding properties are essential for upstream processing what will extract higher order features but also will have to maintain such high precision. In brief, the retina appears to be a dense neural network where specific sub-populations adaptively extract local information in a context-dependent manner in order to produce an output that is both adaptive, sparse, overcomplete and of high temporal precision.

There is a large, and expanding literature proposing models of retinal processing. We attempted to classify them and isolated three main classes of models. A first class regroups the linear-nonlinear-poisson (LNP) models [229]. In its simplest form, a LNP model

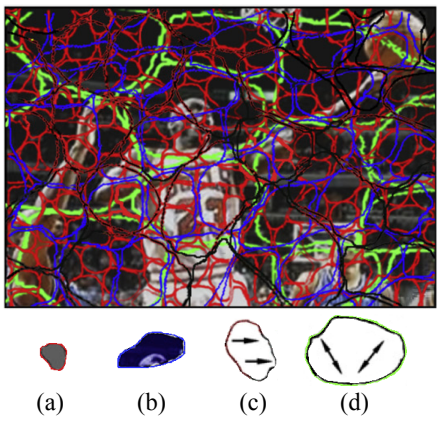


Figure 4: How retinal ganglion cells tile a scene extracting a variety of features. This illustrates the tiling of space of a subset of four cell types. Each tile covers completely the visual image independently from other types. The four cell types shown here correspond to (a) cell with small receptive fields and center-surround characteristics extracting intensity contrasts, (b) color coded cells, (c) motion direction selective cells with a relatively large receptive field, (d) cells with large receptive fields reporting that something is moving (adapted from [198], with permissions).

is a convolution with a spatio-temporal kernel followed by a static nonlinearity and stochastic (Poisson-like) mechanisms of spikes generation. These functional model are widely used by experimentalists to characterise the cells that they record, map their receptive field and characterise their spatiotemporal feature selectivities [64]. LNP models can simulate the spiking activity of ganglion cells (and of cortical cells) in response to synthetic or natural images [55] but they voluntarily the neuronal mechanisms and the details of the inner retinal layers that transform the image into a continuous input to the ganglion cell (or any type of cell) stages.Moreover, they implement static non-linearities, ignoring many existing non-linearities. Applied to computer vision, they however provide some inspiring computational blocks for contrast enhancement, edge detection or texture filtering.

The second class of models has been developed to serve as a front-end for subsequent computer vision task. They succeeded in providing efficient and fast bio-inspired modules for low level image processing. One interesting example is given by [30] (see also [129, 130, 25]), where the model includes parvocellular and magnocellular pathways using different non-separable spatio-temporal filter that are optimal for form or motion detection (this model is also available in OpenCV).

Spiking retinal models make the third class. They focus on the internal and detailed retinal circuitry, in order to reproduce and predict the individual or collective response measured at the ganglion cells level. Virtual Retina [334] is one example of such spiking retina model. This models enables large scale simulations (up to 100,000 neurones) in reasonable processing times while keeping a strong biological plausibility. These models are expanded to explore several aspects of retinal image processing such as (i) understanding how to reproduce accurately the statistics of the spiking activity at the population level [219] (ii) reconciling connectomics and simple computational rules for visual motion detection [155] and (iii) investigating how such canonical microcircuits can implement the different retinal processing modules cited above (feature extraction, predictive coding...) [109]. It is worth also mentioning herein a series of retina simulators designed to future visual prosthetic devices as they provide other solutions to obtain simulated spike train but in real-time [236, 184, 196].

Comparison with computer vision solutions Most computer vision systems are rooted on a sensing device based on CMOS technology to acquire images in a frame based manner. Each frames if obtained from sensors representing the environment as a set of pixels whose values indicate the intensity of light. Pixels pave homogeneously the image domain and their number defines the resolution of images. Dynamical inputs, corresponding to videos are represented as a set of frames, each one representing the environment at a different time, sampled at a constant time step defining

the frame rate.

To make an analogy between the retina and typical image sensors, the dense pixels which respond slowly and capture high resolution color images are at best comparable to P-Cells in the retina. Traditionally in computer vision, the major technological breakthroughs for sensing devices have aimed at improving the density of the pixels, as best illustrated by the ever improving resolution of the images we capture daily with cameras. Focusing of how videos are captured, ones can see that a dynamical input is not more that a series of images sampled at regular intervals. Significant progress have been achieved recently in improving the temporal resolution with advent of computational photography but at a very high computational cost [181]. This kind of sensing for videos introduces a lot of limitations and the amount of data that has to be managed is high.

However, there are two main differences between the retina and a typical image sensor such as a camera. First, as stated above, the retina is not simply sending an intensity information but it is already extracting features from the scene. Second, the retina asynchronously processes the incoming information, transforming it as a continuous succession of spikes at the level of ganglion cells, which mostly encode changes in the environment: retina is very active when intensity is changing, but its activity becomes quickly very low with a purely static stimulation. These observations show that the notion of representing static frames does not exist in biological vision, drastically reducing the amount of data that is required to represent temporally varying content.

Promising bio-inspired solutions Analysing the sensing task from a biological perspective has potential for bringing new insights and solutions to tackle the four challenges outlined in this section. The first step is to use what is known from the retinal circuitry and to model it in order to optimally implement the four generic computational problems identified above, namely adaptation, feature detection, sparseness and temporal precision. Some steps in this direction have already been taken by the computer vision community but this shall be amplified.

Retinal contrast *adaptation* mechanisms have been used for tone mapping as in [95] using empirical laws such as the Naka-Rushton equation. Using more realistic models of luminance and contrast adaptation as in [334], we can go beyond and process videos [29, 210]. Retinal spike coding has been used to investigate new compression schemes [199] or to propose rank-order coding as an efficient strategy to perform ultra-fast categorization [322]. Finally, retina has also inspired neuromorphic models of low-level processing as in [30, 129]. In terms of *feature detection*, retinal geometry has inspired image descriptors such as FREAK [3] but ideas are still far from the known retinal properties in primates. Taking into account how the different types of cells pave the visual field to efficiently capture the useful information could be a source of inspiration for future improvements of this technology. These few examples show how retinal principles have already inspired to some computer vision design, and how progress could be made in each case by further taking into account the latest discoveries in retinal processing.

In terms of *sparse coding* a promising idea is to consider event-based vision sensors where pixels autonomously communicate the change and grayscale events to the readout. The Dynamic Vision Sensor (DVS) [182, 178] and the Asynchronous time-based image sensor (ATIS) [248] are two examples of such sensor using Address-Event Representation (AER) circuits. The main principle is that pixels signal only significant events. More precisely, an event is sent when the log intensity has changed by some threshold amount since the last event (see Fig. 5). Their major advantage is that these sensors provide an output corresponding to pixels that register a change in the scene, thus allowing extremely high temporal resolution to describe changes in the scene while discarding all the redundant information. Immediate applications areas of event-based sensors have been in object tracking [224] but it is also possible to revisit some classical computer vision

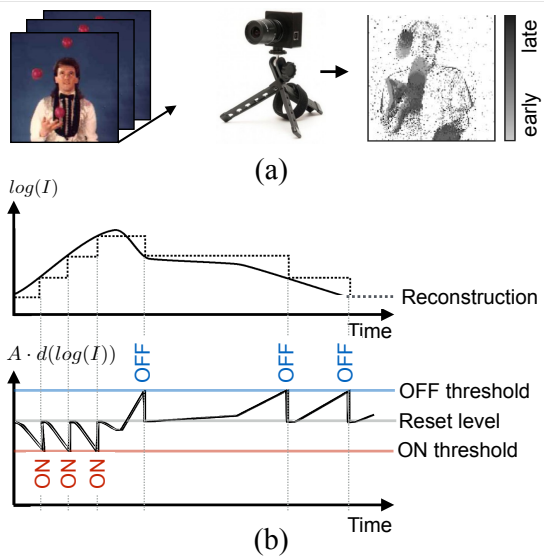


Figure 5: How DVS sensor generate spikes. (a) Example of video with fast motions (a juggling scene), DVS camera and DVS output: Events are rendered using a grayscale colormap corresponding to events times (black = young, gray = old, white = no events). (b) DVS principle: Positive and negative events are generated depending on the variations of $log(I)$ which are indicated as ON and OFF events along temporal axis. Adapted from [178], with permissions.

problems with this frame-free representation. For example, event-based sensor have been used for stereo [272] and for optical flow [310, 31] at a cost that is approximately 25 times lower than does a frame-based method for 128×128 frames at 60 fps.

Another promising idea can be seen as a combination of the two previous strategies: building smart, bio-inspired sensors. This has already been explored for the design of CMOS sensors [206, 328]. To go beyond, the next step would be to build smart-sensors incorporating more of these elementary computing blocks in order to enrich sensors output. A first step in this direction has already been made in [184] where the authors have combined the ATIS sensor with a model pulling nonlinear subunits to reproduce the spatial and temporal properties of the majority of ganglion cells. With such sensors, computer vision tasks could be revisited to exploit the benefits from such a new generation of sensors, with high temporal precision and complex feature-based response.

4.2 Segmentation and figure-ground segregation

Task definition The task of segmenting a visual scene is to generate a meaningful partitioning of the input feature representation into surface-related components. The segregation of an input stimulus into prototypical parts, characteristic of surfaces or objects, is guided by a coherence or homogeneity property that region elements share [101, 234]. Homogeneities are defined upon feature domains such as, e.g., color, motion, depth, statistics of luminance items (texture), etc., or combinations of them. In addition, the goal of segmentation might be extended in regard to eventually single out a target item, or object, from its background in order to recognise it or to track its motion [195]. The specificity of the behavioral task, e.g., grasping an object, distinguishing two object identities, or avoiding collisions during navigation, may influence the required detail of segmentation [18, 123].

Core challenges The segmentation of a spatio-temporal visual image into regions that correspond to prototypical surfaces or objects faces several challenges which derive from distinct interrelated subject matters. The following themes refer to issues of *representation*. First, the feature domain or domains need to be identified which constitute the coherence or homogeneity properties relevant for the segregation task. Feature combinations as well as the nested structure of their appearance of coherent surfaces or objects introduces apparent feature hierarchies [159, 160]. Second, the segmentation process might focus on the analysis of homogeneities that constitute the coherent components within a region or, alternatively, on the discontinuities between regions of homogeneous appearances. Approaches belonging to the first group focus on the segregation of parts into meaningful prototypical

regions utilising an agglomeration (clustering) principle. Approaches belonging to the second group focus on the detection of discontinuous changes in feature space (along different dimensions) [227] and group them into contours and boundaries. Note that we make a distinction here to refer to a contour as a grouping of oriented edge or line contrast elements whereas a boundary already relates to a surface border in the scene. Regarding the boundaries of any segment, the segmentation task itself might incorporate an explicit assignment of a border ownership (BOwn) direction label which implies the separation of figural shape from background by a surface that occludes other scenic parts [240, 162]. The variabilities in the image acquisition process caused by, e.g., illumination conditions, shape and texture distortions, might speak in favor of a boundary oriented process. On the other hand, the complexity of the background structure increases the effort to segregate a target object from the background, which argues in favour of region oriented mechanisms. It should be noted, however, that the region vs boundary distinction might not appear as binary as in the way outlined above. Considering real world scenes the space-time relationships of perceptual elements (defined over different levels of resolution) are often defined by statistically meaningful structural relations to determine segmentation homogeneities [333]. Here, an important distinction has been made between structure that might be influenced by meaning and primitive structure that is perceived even without a particular interpretation.

While the previous challenges were defined by representations, the following themes refer to the process characteristic of segmentation. First, the partitioning process may yield different results given changing view-points or different noise sources during the sensing process. Thus, segmentation imposes an inference problem that is mathematically ill-posed [246]. The challenge is how a reliability, or confidence, measure is defined that characterises meaningful decompositions relating to reasonable interpretations. To illustrate this, Fig. 6 shows segmentation results as drawn by different human observers. Second, figural configurations may impose different efforts for mechanisms of perceptual organisation to decide upon the segregation of an object from the background and/or the assignment of figure and ground direction of surface boundaries. A time dependence that correlates with the structural complexity of the background has in fact been observed to influence the time needed in visual search tasks [335].

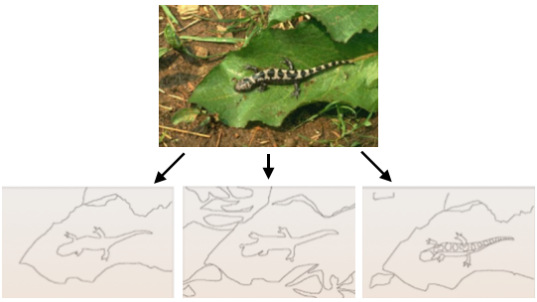


Figure 6: Example of possible segmentation results for a static image drawn by different human observers. Lower images shows segmentations happening at different levels of detail but consistent with each other. Adapted from [9].

Biological vision solution Evidence from neuroscience suggests that the visual system uses segmentation strategies based on identifying discontinuities and grouping them into contours and boundaries. Such processes operate mainly in a feedforward fashion and automatic, utilising early and intermediate-level stages in visual cortex. In a nutshell, contrast and contour detection is quickly accomplished and is already represented at early stages in the visual cortical hierarchy, namely areas V1 and V2. The assignment of task-relevant segments happens to occur after a slight temporal delay and involves a recurrent flow of lateral and feedback processes [281, 268, 266]. Such dynamical process, called re-entry [83], recursively links representations distributed over different levels. Mechanisms of lateral integration, although slower in processing speed, seem to further support intra-cortical grouping [150, 151, 108]. In addition, surface segregation is reflected in a later temporal processing phase but is also evi-

dent in low levels of the cortical hierarchy, suggesting that recurrent processing between different cortical stages is involved in generating neural surface representations. Once boundary groupings are established surface related mechanisms "paint", or tag, task-relevant elements within bounded regions. The feature dimensions used in such grouping operations are, e.g., local contour orientations defined by luminance contrasts, direction and speed of motion, color hue contrasts, or texture orientation gradients. Sketched above, counter-stream interactive signal flow [315] imposes a temporal signature on responses in which after a delay a late amplification signal serves to tag those local responses that belong to a region (surrounded by contrasts) which has been selected as a figure [166] (see also [271]).

The grouping of visual elements into contours appears to follow the Gestalt rules of perceptual organisation [161]. Grouping has also been studied in accordance to the ecological validity of such rules as they appear to be embedded in the statistics of natural scenes [48]. Mechanisms that entail contour groupings are implemented in the structure of supragranular horizontal connections in area V1 in which oriented cells preferentially contact like-oriented cells that are located along the orientation axes defined by a selected target neurone [150, 38]. Such long-range connections form the basis for the Gestalt concept of good continuation and might reflect the physiological substrate of the association field, a figure-eight shaped zone of facilitatory coupling of orientation selective input and perceptual integration into contour segments [96, 103, 116]. Recent evidence suggests that the perceptual performance of visual contour grouping can be improved by mechanisms of perceptual learning [176].

Once contours have been formed they need to be labeled in accordance to their scene properties. In case of a surface partially occluding more distant scenic parts the *border ownership (BOwn) or surface belongingness* can be assigned to the boundary [161]. A neural correlate of such a mechanism has been identified at different cortical stages along the ventral pathway, such as V1, V2 and V4 areas [15, 343, 230]. The dynamics of the generation of the BOwn signals may be explained by feedforward, recurrent lateral and feedback mechanisms (see [332] for a review). The time course of the neuronal responses encoding invariance against different figural sizes argues for a dominant role of feedback signals when dynamically establishing the proper BOwn assignment. Grouping cells have been postulated that integrate (undirected) boundary signals over a given radius and enhance those configurations that define locally convex shape fragments. Such fragments are in turn enhanced via a recurrent feedback cycle so that closed shape representations can be established rapidly through the convexity in closed bounding contours [343]. Neural representations of localized features composed of multiple orientations may further influence this integration process, although this is not firmly established yet [8]. BOwn assignment serves as a prerequisite of figure-ground segregation. The temporal dynamics of cell responses at early cortical stages suggest that mechanisms exist that (i) decide about ownership direction and (ii) subsequently enhance regions (at the interior of the outline boundaries) by spreading a neural tagging, or labelling, signal that is initiated by the region boundary [270] (compare the discussion in [332]). Such a late enhancement through response modulation of region components occurs for different features, such as oriented texture [167] or motion signals [271], and is mediated by recurrent processes of feedback from higher levels in the cortical hierarchy. It is, however, not clear whether a spreading process for region tagging is a basis for generating invariant neural surface representations in all cases. All experimental investigations have been conducted for input that leads to significant initial stimulus responses while structureless homogeneous regions (e.g., a homogeneous coloured wall) may lead to void spaces in the neuronal representation that may not be filled explicitly by the cortical processing (compare the discussion in [238]).

Yet another level of visual segmentation operates upon the initial grouping representations, those base groupings that happen to be processed effortlessly as outlined above. How-

ever, the analysis of complex relationships surpasses the capacities of the human visual processor which necessitates serial staging of some higher-level grouping and segmentation mechanisms to form incremental task-related groupings. In this mainly sequential operational mode visual routines establish properties and relations of particular scene items [314]. The elemental operations of such routines are, e.g., shifting the processing focus (related to attentional selection), indexing (to select a target location), coloring (to label homogeneous region elements), and boundary tracing (determining whether a contour is open or closed and whether items belong to the contour). For example, in boundary tracing tasks a neural signal propagates along prototypical groupings to tag those elements that define task-relevant segment. Such tagging is implemented by a lateral spreading mechanism as evident by monotonically increasing delays of such modulatory labelings with increasing distances of elements along the perceptual entity (curve or boundary) [147, 266]. Maintenance operations then interface such elemental operations into sequences to compose visual routines for solving more complex tasks, like in a sequential computer program. Such cognitive operations are implemented in cortex by networks of neurons that span several cortical areas [267]. The execution time of visual cortical routines reflects the sequential composition of such task-specific elemental neural operations tracing the signature of neural responses to a stimulus [169, 267].

Comparison with computer vision solutions Segmentation as an intermediate level process in computational vision is often characterised as one of agglomerating, or clustering, picture elements to arrive at an abstract description of the regions in a scene [234]. It can also be viewed as a preprocessing step for object detection/recognition. It is not very surprising to see that even in computer vision earlier attempts were drawn towards single aspects of the segmentation like edge detection [53, 194, 179] or grouping homogeneous regions by clustering [66, 67]. The performance limitations of both these approaches independently have led to the emergence of solutions that reconsidered at the problem as a juxtaposition of both edge detection and homogeneous region grouping with implicit consideration for scale. The review paper by [98] presents various approaches that attempted in merging edge based information and clustering based information in a sequential or parallel manner. The state of the art techniques that are successful in formulating the combined approach are variants of graph cuts [289], active contours [59], and level sets. At the bottom of all such approaches is the definition of an optimisation scheme that seeks to find a solution under constraints such as, e.g., smoothness or minimising a measure of total energy. These approaches are much better in terms of meeting human defined ground truth compared to simpler variants involving discontinuity detection or clustering alone.

A major challenge is still how to compare the validity and the quality of segmentation approaches. Recent attempts emphasise to compare the computational results - from operations on different scales - with the results of hand-drawn segmentations by human subjects [9, 97]. These approaches suggest possible measures in judging the quality of automatic segmentation given that ground truth data is missing. However, the human segmentation data does not elucidate the mechanisms underlying the processes to arrive at such partitions. Instead of a global partitioning of the visual scene, the visual system seems to adopt different strategies of computation to arrive at a meaningful segmentation of figural items. The grouping of elements into coherent form is instantiated by selectively enhancing the activity of neurones that represent the target region via a modulatory input from higher cortical stages [166, 168]. The notion of feedback to contribute in the segmentation of visual scenes has been elucidated above. Recent computer vision algorithms begin to make use of such recurrent mechanisms as well. For example, since bottom-up data-driven segmentation is usually incomplete and ambiguous the use of higher-level representations might help to validate initial instances and further stabilise their repre-

sentation [316, 35]. Along this line, top-down signalling applies previously acquired information about object shape (e.g., through learning), making use of the discriminative power of fragments of intermediate size, and combines this information with a hierarchy of initial segments [317]. Combined contour and region processing mechanisms have also been suggested to guide the segmentation. In [9], multi-scale boundaries are extracted which later prune the contours in a watershed region-filling algorithm. Algorithms of figure-ground segregation and border-ownership computation have been developed for computer vision applications to operate under realistic imaging conditions [297, 300]. These were designed to solve tasks like shape detection against structured background and for video editing.

Promising bio-inspired solutions Numerous models that account for mechanisms of contour grouping have been proposed to linking orientation selective cells [116, 118, 177]. The rules of mutual support utilize a similarity metric in the space-orientation domain giving rise to a compatibility, or reliability measure [153] (see [223] for a review of generic principles and a taxonomy). Such principles migrated into computer vision approaches [235, 121, 163] and, in turn, provided new challenges for experimental investigations [290, 27]. Note that the investigation of structural connectivities in high dimensional feature spaces and their mapping onto a low-dimensional manifold lead to define a "neurogeometry" and the basic underlying mathematical principles of such structural principles [43, 241].

As outlined above, figure-ground segregation in biological vision segments an image or temporal sequence by boundary detection and integration followed by assigning border ownership direction and then tagging the figural component in the interior of a circumscribed region. Evidence suggests that region segmentation by tagging the items which belong to extended regions involves feedback processing from higher stages in the cortical hierarchy [281]. Grossberg and colleagues proposed the FACADE theory (form-and-color-and-depth, [116, 113, 114]) to account for a large body of experimental data, including figure-ground segregation and 3D surface perception. In a nutshell, the model architecture consists of mutually coupled subsystems, each one operating in a complementary fashion [115]. A boundary contour system (BCS) for edge grouping is complemented by a feature contour system (FCS) which supplements edge grouping by allowing feature qualities, such as brightness, color, or depth, to spread within bounded compartments generated by the BCS. Such a mechanism is proposed to also facilitate the assignment of figural sides to boundaries. BOwn computation has been incorporated in computer vision algorithms to segregate figure and background regions in natural images or scenes [259, 134, 300]. Such approaches use local configurations of familiar shapes and integrate these via global probabilistic models to enforce consistency of contour and junction configurations [259] of learning of templates from ensembles of image cues to depth and occlusion [134].

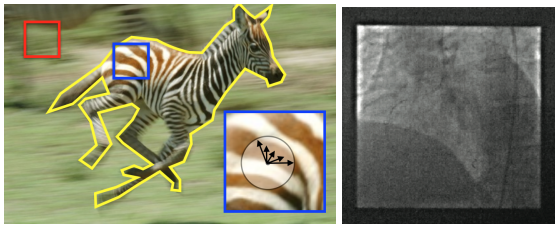
Feedback mechanisms as they are discussed above allow to build robust boundary representations such that junctions may be reinterpreted based on global context information. The hierarchical processing of shape from curvature information in contour configurations [265] can be combined with evidence for semi-global convex fragments or global convex configurations [71]. Such activity is fed back to earlier stages of representation to propagate contextual evidences and quickly build robust object representations separated from the background. A first step towards combining such stage-wise processing capacities and integrating them with feedback that modulates activities in distributed representations at earlier stages of processing has been suggested in [309]. The step towards processing complex scenes from unconstrained camera images, however, still needs to be further investigated.

4.3 Optical flow

Task definition Estimating optical flow refers to assignment of 2-D vectors at sample locations of the visual image in order to de-

scribe their displacements within the sensor's frame of reference using the change of structured light in the retinal or camera images. This displacement vector field constitutes the image flow representing apparent 2-D motions resultant of their 3-D velocities being projected onto the sensor [324, 325]. Such 2-D motions are observable only at intensity variations (and are thus contrast dependent) as a consequence of the relative change between an observer (eye or camera) and the surfaces or objects in a visual scene.

Core challenges Achieving a robust estimation of optical flow faces several challenges. First of all, visual system has to establish form based correspondences across temporal domain despite motion induced geometric and photometric distortions. Second, velocity space has to be optimally sampled and represented to achieve robust and energy efficient estimation. Third, the accuracy and reliability of the velocity estimation is dependent on the local structure/form but the visual system must achieve a form independent velocity estimation. Difficulties arise from the fact that any local motion computation faces different sources of noise and ambiguities, such as for instance the aperture problem. Therefore, estimating optical flow requires to resolve these local ambiguities from the estimation process by integrating different local motion signals while still maintaining segregated those that belong to different surfaces or objects of the visual scene (see Fig. 7(a)). In other words, image motion computation faces two opposite goals when computing objet global motion, integration and segmentation [40]. As already emphasised in Sec. 4.2, any computational machinery should be able to keep segregated the different surface/object motions as one goal of motion processing is to estimate accurately the speed and direction of each of them to be able to track, capture or avoid one or several of them. Fourth, the visual system has to be able to deal with complex scenes full of occlusions, transparencies or non-rigid motion. This is well illustrated by the transparency case. Since optical flow is a projection of 3D displacements in the world, some situations yield to perceptual (semi-) transparency [205]. In videos, several causes have been identified, such as reflections, phantom special effects, dissolve effects for a gradual shot change and medical imaging such as X-rays (for example see Fig. 7(b)). All of these exemple raise serious problems to current computer vision algorithms.



(a) (b)

Figure 7: Core challenges in motion estimation: (a) This snapshot of a moving scenes illustrates several ideas discussed in the text: inset with the blue box shows the local ambiguity of motion estimation while the yellow boundary shows how segmentation and motion estimation are related. (b) Transparent motion example coming from an image of X-ray exam yielding a situation of transparency (from [12]).

Herein, we will focus on four main computational strategies used by biological systems for dealing with aforementioned problems and could inspire the design of computer vision algorithms. First is *Motion energy estimation* by which the visual system estimates a contrast dependent measure of translations in order to indirectly establish correspondences. Second is *Local velocity estimation*. These contrast dependent motion energy features must be combined to achieve a contrast invariant local velocity estimation after denoising the dynamical inputs and resolving local ambiguities by integrating local form and motion cues. The third challenge concerns the *global motion estimation* of each independent object, regardless its shape or appearance. Fourth, *distributed multiplexed representations* must be used by both natural and artificial systems to segment cluttered scenes, handle multiple/transparent surfaces, and encode depth ordering to achieve 3D

motion perception and goal-oriented decoding.

Biological vision solution In Fig. 3, we have already sketched the backbone of the cortical motion stream and its recurrent interactions with both area V1 and the form stream. This figure illustrates the advantages and limits of the deep hierarchical model. We will discuss some recent data about the neuronal dynamics of these areas in regards with the four challenges of optic flow processing.

As said above, the classical view of the cortical motion pathway is a feedforward cascade of cortical areas spanning from the occipital (V1) to the parietal (VIP, area 7) lobes, forming the skeleton of the dorsal stream. Areas MT and MST are located in the deep of the superior temporal sulcus and are seen as a pivotal hub for both object motion and self-motion (see, e.g., [232, 41, 233] for reviews). This motion pathway is extremely fast, with the information flowing in less that 20ms from the primary visual area to the frontal cortices and brainstem structures underlying visuomotor transformations (see [169, 49, 201, 180] for reviews). Such high processing speed originates from the Magnocellular retino-geniculo-cortical input to area V1. This input carries low spatial and high temporal frequencies luminance information with high contrast sensitivity (i.e., high contrast gain). This cortical input to layer 4β projects directly to the extra striate area MT, also called the cortical motion area. The fact that this feedforward stream by-passes the classical recurrent circuit between area V1 cortical layers is attractive for several reasons. First, it implements a fast, feedforward hierarchy fitting the classical two-stage motion processing model [217, 132] where V1 neurones sense local motion direction and MT cells integrate them to extract pattern motion speed and direction. The next stage, area MST can then extracts object-motion through small receptive fields (area MSTl) or optic flow pattern (e.g., rotation, expansion...) covering some very large receptive fields of MSTd neurones (see [232] for an exhaustive review). Second, it illustrates the fact that built-in, fast and highly specific modules of visual information have conserved through evolution to subserve automatic, behaviour-oriented visual processing (see, e.g. [201, 78, 37] for reviews). Third, this scheme is a good example of a canonical circuit that implements a sequence of basic computations such as spatio-temporal filtering, gain control and normalisation at increasing spatial scales [276]. The final stages of these bio-inspired models consist in population of neurones broadly selective for translation speed and direction [291, 237] as well as more complex optical flow patterns (see e.g., [117, 171] for recent examples).

This feedforward architecture, and the physiological properties of motion selective cells at each stage have been highly influential when promoting the dominant hierarchical framework of biological and computer vision. For instance, direction-selective cells in area V1 are best described as spatio-temporal filters that can extract motion energy along the direction orthogonal to the luminance gradient [86, 68, 189]. These filters are integrated to compute local motion direction and speed by MT neurons. Such integration allows extracting motion signals embedded in noise with high precision, normalizing its through centre-surround interactions and solving the aperture problem. Thus, speed and motion direction selectivities at single-cell and population levels become largely independent upon the contrast or the shape of the moving inputs [41, 36, 232]. Further integrating MT output within the same coarse retinotopic representation allows to build template motion detectors and to solve difficult problems such as optic flow computation, object motion tracking or heading estimation [232].

One could therefore think that biology has indeed provided computational neuroscience and computer vision with a robust set of algorithms and implementation rules so that bio-inspired algorithms can match the best computer vision solutions for estimating motion. However, the dominant feedforward, two-stage motion framework has lead to several oversimplification that can hinder the emergence of more versatile and robust solutions. We briefly review some novel biological facts that per-

tain to each of our four computational problems. First, the *motion energy estimation* through a set of spatio-temporal filters have been recently re-evaluated to account for the neuronal responses to complex dynamical textures and natural images. When presented with rich, naturalistic inputs, responses of both V1 complex cells and MT pattern-motion neurones become more selective (i.e., their tuning is sharper) [251, 105], sparser [326] and more precise [22]. These improvements could be explain by a more complex sampling of inputs, through a set of adaptive and weighted filters that optimally sample the spatiotemporal frequency plane [226]. The diversity of centre-surround interactions in both areas V1 and MT would most certainly also contribute to these nonlinearities [74]. The diversity of excitatory and inhibitory inputs can also explains how the aperture problem is dynamically solved by MT neurones presented not only with plaid patterns [276] but also with single elongated bars or barber poles [312]). Thus, *local velocity estimation* in V1 or MT cells that are selective to 2D motion direction would rely on a complex, context-dependent integration of motion signals extracted at multiple spatial and temporal scales. Moreover, the role of non-luminance based cues, such as local 2D features or second-order motion signals shall most certainly be taken into account to further refine speed and direction estimates of a single object [312]. Finally, the role of feedback in this context-dependent integration of local motion has been demonstrated by computational studies [24, 23] and is now addressed at physiological level despite the considerable technical difficulties (see [73] for a review). Overall, several computational studies have shown the importance of the adaptive normalisation of spatiotemporal filters for motion perception; see [292] illustrating how a generic computation (normalisation) can be adaptively tuned to match the requirement of different behaviours.

Recent physiological works have also shown how the feedforward backbone of *global motion integration* must be enriched to better explain several of each features. Figure 3 illustrate these different aspects, from a functional connectivity and a computational perspective. A first aspect is the richness of centre-surround interactions, in both retinotopic, direction and speed domains. These properties are important to encode optic flow patterns [212] and biological motion [89, 88]. They involve both the classical convergence of projections from one step to the next but also the dense network of lateral interactions within V1 as well as within each extra-striate areas. These lateral interactions implement long-distance normalisation, seen as centre-surround interactions at population level [260] as well as feature grouping between distant elements [107]. These intra- et inter-cortical areas interactions can support a second important aspect of motion integration: motion diffusion. In particular, anisotropic diffusion of local motion information can play a critical role in global motion integration by propagating reliable local motion signals within the retinotopic map [308]. The exact neural implementation of these mechanisms is yet unknown but modern tools will soon allow to image, and manipulate, the dynamics of these lateral interactions.

Global motion integration is only one side of the coin. As pointed out by Braddick [40], motion integration and segmentation works hand-to-hand to selectively group the local motion signals that belong to different surfaces. For instance, some MT neurones integrate motion signals within their receptive field only if they belong to the same contour [137] or surface [298]. They can also filter out motion within the receptive when it does not belong to the same surface [294, 298], a first step for representing motion transparency or structure-from-motion in area MT [120]. The fact that MT neurons for instance can adaptively integrate local motion signals, and explain away other is strongly related to the fact that motion sensitive cells are most often embedded in *distributed multiplexed representations*. Indeed, most direction-selective cells are also sensitive to binocular disparity [76, 170, 253, 293], eye/head motion [216] and dynamical perspective cues [154] in order to filter out motion signals from outside the plane of fixation or to disambiguate motion parallax. Thus, ones want to

consider depth and motion processing as two intricate problems allowing the brain to compute objects motion in 3D space rather than in 2D space.

Depth-motion interaction is only one example that the motion pathway receives and integrates visual cues from different modules [231]. This is illustrated in Fig. 3, where form cues can be extracted in areas V2 and V4 and fetch to area MT. Information about the spatial organisation of the scene using boundaries, colours, shapes might then be used to further refine the fast and coarse estimate of the optic flow that emerges from the V1-MT-MST backbone of the hierarchy. Ones could illustrate this with a classical pitfalls of the feedforward model where information is gathered over larger and larger receptive fields. One penalty is that object boundaries and form are strongly blurred. Thus, the large receptive fields of MT and MST neurones can be useful for tracking large objects with the eyes, or avoiding approaching ones, but they certainly lower the spatial resolution of the estimated optic flow field, in contrast with the sharp, crystal-like perception that we have of the moving scene. Mixing different spatial scales through recurrent connectivity between cortical areas is one solution [73]. Constraining the diffusion of motion information along edges or within surface boundaries in certainly another as shown for texture-ground segmentation [282]. Such form-based representations play a significant role in disambiguation of motion information [104, 204, 202, 131]. It could also play a role in setting the balance between motion integration and segmentation dynamics, as illustrated in Fig.3(b).

We have mentioned above how a large family of bio-inspired models have been proposed to deal with these separate aspects of optic flow estimation. A first step is to achieve a form-independent representation of velocity from the spatio-temporal responses arriving from V1. A dominant computational model was proposed by Heeger and Simoncelli [291], where a linear combination of afferent inputs from V1 is followed by a non linear operation known as untuned divisive normalisation. This model, and it subsequent developments [276, 226, 292] were shown to replicate a variety of observations from physiology to psychophysics using stimuli such as drifting grating and plaids. Even though these models partially solve the problem of estimating velocity locally when 2D cues are present, they cannot resolve ambiguities in regions without any 2D cues because of the absence of diffusion mechanisms. Moreover, normalization and integration properties are static, explaining in part that they do not perform well on dynamical natural movies. Feedback signals from and to MT (the latter from higher stages, such as MST) could play a key role in reducing these ambiguities and one such model describing these interactions was proposed by [23]. An extended model of V1-MT-MST interaction that uses center-surround competition in velocity space was later presented by [257], showing good optic flow computations in the presence of transparent motion. These feedback and lateral interactions primarily play the role of context dependent diffusion operators that spread the most reliable information throughout ambiguous regions. Such diffusion mechanism can be gated to generate anisotropic propagation, taking advantage of local form information [308, 26]. An attempt at utilising these distributed representation for integrating both optic flow estimation and segmentation was proposed in [228]. The same model explored the role of learning in establishing the best V1 representation of motion information, although this approach was largely ignored in optic flow models contrary to object categorisation for instance.

Comparison with computer vision solutions The vast majority of computer vision solutions for optical flow estimation can be split into four major computational approaches (see [299] for a recent review). First, a constancy assumption deals with correspondence problem, assuming that brightness or color is constant across adjacent frames and assigning a cost function in case of deviation. Second, the reliability of the matching assumptions optimised using priors or a regularisation to deal with aperture problem. Both of these solutions pose the problems as an energy func-

tion and optical flow itself is treated as a energy minimisation. Interestingly, a lot of recent research has been done in this area, always pushing further the limits of the state-of-the-art. This research field has put a strong emphasis on performance as a criterion to select novel approaches and sophisticated benchmarks have been developed. Since the early initiatives along this framework [19], current benchmarks now cover a much wider variety of problems at hand. Popular examples are the Middleburry flow evaluation [14] and, more recently the Sintel flow evaluation [51]. The later has important features which are not present in the Middlebury benchmark: long sequences, large motions, specular reflections, motion blur, defocus blur, and atmospheric effects.

Initial motion detection is an example where biological and computer vision research have already converged. The correlation detector proposed by Hassenstein and Reichardt [122] serves as a reference for a velocity sensitive mechanisms to find correspondences of visual structure at image locations in consecutive temporal samples. Formal equivalence of correlation detection with a multi-stage motion energy filtering has been demonstrated [1]. There are now several examples of spatiotemporal filtering models that are used to extract motion energy across different scales. Initial motion detection is ambiguous since motion can locally be measured only orthogonal to an extended contrast (aperture problem), while this ambiguity can be resolved at localised image features, such as corners or junctions from non-occluding configurations. In computer vision, the formulation of the aperture problem largely depends on the local motion estimation. For example in the local gradient-based formulation, it becomes ill-posed as it any given location, one has to estimate two variables based on the single equation and in spatiotemporal energy based methods, all the spatiotemporal samples lie on a straight line in frequency space where the task it to identify a plane that passes through all of them [41]. Computer vision has dealt with the problem in two ways: by imposing local constraints [185] or by posing smoothness constrains through penalty terms [135]. More recent approaches are attempted to fuse the two formulations [47]. The penalty term plays a key role as a diffusion operator can act isotropically or anisotropically [11, 32, 280]. A variety of diffusion mechanisms has been proposed so that, e.g., optical flow discontinuities could be preserved depending on velocity field variations or image structures. All these mechanisms have demonstrated powerful results regarding the successful operation in complex scenes. Computational neurosciences models also tend to rely on diffusion mechanisms too, but they differ in their formulation. A first difference stems from the fact that local motion estimation is primarily based on the spatio-temporal energy estimation. Second, the representation is distributed, allowing multiple velocities at the same location, thus dealing with layered/transparent motion. Diffusion operator is also gated based on the local form cues also relying on the uncertainty estimate which could possibly be computed using the distributed representation [228].

Promising bio-inspired solutions Bio-inspired models of motion integration tend to use more form-motion interactions for disambiguating information and this should be further exploited in computer vision models. Future research will have to integrate the growing knowledge about how diffusion process, form-motion interaction and multiplexing of different cues are implemented and impact global motion computation [312, 256, 149]. Despite the similarities in the biological and artificial approaches to solve optical flow computation, it is important to note that there is only little interaction happening between computer vision engineers and biological vision modellers. One reason might be that biological models have not been rigorously tested on regular computer vision datasets and are therefore considered as specifically confined to laboratory conditions only. It would thus be very interesting to evaluate models such as [307, 45, 24, 291] to identify complementary strengths and weaknesses in order to find converging lines of research investigations. Figure 8 illustrates some recent work initiated in this direction where two bio-inspired models have been evaluated

on the Middlebury optical flow dataset [14]. Results show that the AMPD model performs better than the FFV1MT although some improvements need to be done to reach the state-of-the-art, for example by integrating feedback to better recover true velocities around motion boundaries.

Some elements of the mechanisms discussed above have already been incorporated in the specification of computer vision models, for example at early motion detection stage [125]. The solution proposed by [330] uses a regularization scheme that considers different temporal scales, namely a regular motion mechanism (using short exposure frames) as well as a slowly integrating representation (using long exposure frames), the latter resembling the form pathway in the primate visual system [283]. The goal there was to reduce inherent uncertainty in the input [188].

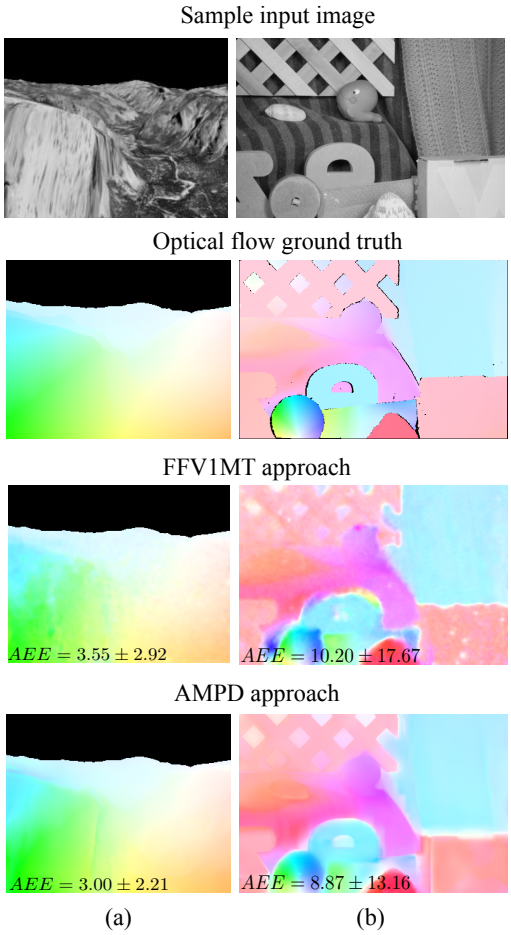


Figure 8: Example of biological models tested on regular computer vision datasets. Results using two approaches are shown: FFV1MT approach [295], where the authors have revisited the seminal work by Heeger and Simoncelli [291] using spatio-temporal filters to estimate optical flow from V1-MT feedforward interactions; AMPD approach [207], where the authors have extended the Heeger and Simoncelli model with adaptive processing by focussing on the role of local context indicative of the local velocity estimates reliability. Results on two videos are shown: (a) Yosemite and (b) Rubberwhale from Middlebury dataset [14]. Optical flow is represented using the colorcode from Middlebury dataset.

5 Discussion

In Sec. 4 we revisited three computer vision tasks and discussed strategies that seemed to be used by biological vision systems to solve these tasks. Tables 1 and 2 provide a concise summary of models and biological data respectively for each task. From this analysis, we identified three major ways to identify which studies from biological vision could be leveraged to advance computer vision algorithms.

5.1 Structural principles that relate to function

Studies in biological vision reveal structural regularities in various regions of the visual cortex, such as hierarchical organisation of processing with response selectivities becoming more and more selective with the levels in the hierarchy. At the same time, the convergence or input integration from spatial positions is getting larger. The potential for using deep feedforward architectures for computer vision has recently been discussed by [164]. However, such principles of bottom-up cascading should be combined with those of lateral integration of

	Reference	Model	Application	Code
Sensing	Vanrullen et.al., 2002 [322]	Spatial model based on difference-of-Gaussian kernels at different scales	Object recognition using the idea of latency coding	○
	Benoit et. al., 2010 [30]	Spatio-temporal model of retinal parvocellular and magnocellular pathways (also includes a V1 model)	Low level image processing	●
	Wohrer et.al., 2009 [334]	Spiking retina model with contrast gain control (*Virtual Retina*)	Comparisons to single cell recordings and large scale simulations	●
	Lorach et.al., 2012, [184]	Retina-inspired sensor combining an asynchronous event-based light sensor (DVS) with a model pulling nonlinear subunits to reproduce the parallel filtering and temporal coding of the majority of ganglion cell types	Target artificial visual systems and visual prosthetic devices	○
	Martinez et.al., 2013, [196]	Compiler-based framework with an ad hoc language allowing to produce accelerated versions of the models compatible with COTS microprocessors, FPGAs or GPUs (*Retina Studio*)	Target visual prosthetic devices	○
Segmentation	Parent et.al., 1989 [235]	Model of curve detection and boundary grouping using tangent orientation and local curvature information	Tested on artificial noisy images for curve evaluation and natural images from different domains	○
	Ren et.al., 2006 [259]	Figure-ground assignment to contours in natural images based on mid-level visual shapes (so-called shapemes) and global consistency enforcement for contour junctions	Bottom-up figure-ground label assignment in still images of large data bases with human ground truth labelings	○
	Bornstein et.al., 2008 [35]	Model for image segmentation combining bottom-up processing (to create hierarchies of segmented uniform regions) with top-down processing (to employ shape knowledge from prior learning of image fragments)	Tested on data sets with four classes of objects to demonstrate improved segmentation and recognition performance	○
	Rodriguez et.al., 2012 [265]	Computational model of mid-level 2D shape representation utilizing hierarchical processing with end-stopping and curvature selective cells	Tested on artificial shape configurations to replicate experimental findings from neurophysiology	○
	Azzopardi et.al., 2012 [13]	Computational model of center-surround and orientation selective filtering with nonlinear context-dependent suppressive modulation and cross-orientation inhibition	Tested on two public data sets of natural images with contour ground truth labelings	○
	Tschechne, 2014 [309]	Recurrent network architecture for distributed multi-scale shape feature representation, boundary grouping, and border-ownership direction assignment	Tested on a selection of stimuli from public data sets	○
Optical flow	Heeger, 1988 [125]	Feedforward model based on spatio-temporal motion energy filters	Used to simulate psychophysical data and Yosemite sequence	○
	Nolan et.al., 1994 [228]	Model based on spatio-temporal motion energy filters with a selection mechanism to deal with occlusions and transparency	Optical flow estimation, tested on synthetic images only	○
	Grossberg et.al., 2001 [119]	Dynamical model representative of interactions between V1, V2, MT and MST areas	Grouping and optical flow estimation, tested on synthetic images only	○
	Bayerl et.al., 2007 [24]	Recurrent model of V1-MT with modulatory feedbacks and a sparse coding framework for neural motion activity patterns	Optical flow estimation, tested using several real world classical videos	○
	Tlapale et.al., 2010 [308]	Dynamical model representative of V1-MT interactions and luminosity based motion information diffusion	Optical flow estimation, tested on synthetic images only	○
	Perrone et.al., 2012 [237]	Model explaining the speed tuning properties of MST neurons by afferent pooling from MT	Optical flow estimation, tested on synthetic and two natural sequences	○
	Tschechne et.al., 2014 [310]	Model of cortical mechanisms of motion detection using an asynchronous event-based light sensor (DVS)	Motion estimation with limited testing for action recognition	○
	Solari et.al., 2015 [295]	Multi-scale implementation of a feedforward model based on spatio-temporal motion energy filters inspired by [125]	Dense optical flow estimation, evaluated on Middlebury bechmark	●

Table 1: Highlight on models for each of the three tasks considered in Sec. 4.

	Biological mechanism	Experimental paper	Models
Sensing	Visual adaptation	[152, 304, 287]	[129, 334]
	Feature detection	[152]	[129]
	Sparse coding	[243]	[184]
	Precision	[243]	[184]
	Surveys	[197, 198]	–
Segmentation	Contrast enhancement and shape representation	[103]	[13, 265]
	Feature integration and segmentation	[16, 38, 48, 96, 107, 151, 176, 239, 290, 335]	[9, 27, 35, 52, 116, 118, 195, 223]
	Border ownership and figure-ground segregation	[140, 145, 166, 168, 240, 282, 339, 343]	[71, 97, 113, 134, 259, 309]
	Continuation and visual routines	[147, 153, 162, 247]	[123, 258]
	Surveys	-	[30, 70, 130]
Optical flow	Motion energy estimation	[86, 68, 189, 277]	[291, 1, 125]
	Local velocity estimation	[226, 41, 276, 303, 252]	[295, 226]
	Global motion integration	[137]	[237, 308, 24, 119, 228]
	Distributed multiplexed representations	[203, 139, 293, 216, 76, 20]	[253, 50, 253, 170, 94, 231]
	Surveys	[233, 217]	[39]

Table 2: Summary of the strategies highlighted in the text to solve the different task, showing where to find more details about the biological mechanisms and which models are using these strategies.

activations and massive feedback from higher stages in the processing hierarchy. In this section we would like to emphasise the role of lateral interaction and feedback processing, which have been less explored in computer vision.

Neural systems are found to exhibit very rich dynamic response behaviour which can be attributed to contextual modulation of individual responses. It has been known long before that the visual system's information processing changes adaptively depending on the task. Lateral interactions are one of the core principles at the roots of such context-driven changes in the response gain. For example, the lateral integration of responses to oriented contrast input (which is most prominent in the superficial layers in cortical area V1 [38]) has been extensively investigated from the perspective of contour grouping and integration of locally related fragments [150]. Another related principle that might be realised in part by lateral interaction of neural responses is the nonlinear suppression of activities by local context. For example, the response gain of a single feature selective cell is reduced when other cells in the surround outside the receptive field of the target cell [75, 57, 65] are simulated in parallel. This observation led Carandini and Heeger to suggest normalization as a canonical principle in cortical computation, not only for vision but also for attention, multi-sensory integration, and decision-making [56]. Several authors have studied such normalization process from a theoretical, e.g., statistical whitening, source separation [187] or from an application-oriented perspective, e.g., performance increase in recognition schemes [144]. The neural computational mechanisms that can account for such normalization mechanism are potentially rooted in a competitive interaction between cells that are responsive to target features and a large pool defined over a space-feature domain that acts

divisively on the target activation.

In this survey, we also observed that forward processing of sensory activity is accompanied by massive feedback projections of signal pathways from higher stages that innervate the representations at earlier stages. Along such feedback connections different higher order representations of activities can be delivered, namely context that changes the bias of sensory inputs or attention that aims at biasing the current feature processing by expectations to increase the relative importance of features that cohere with the top-down activations, and generic knowledge (short-term or long-term) can be incorporated in the current processing. Even in its simplest definition, such feedback my help to improve the detectability of shapes and objects in noisy and cluttered environments [250]. Feedback mechanisms seamlessly fit into the generic processing of bottom-up and top-down streams such that the emphasis in a representation can be dynamically shifted and could be used to allocate the computational resources in a task- and goal-directed fashion. The functional role feedback signals still remains a matter of controversy. Different proposals on how feedback signals affect the driving feedforward stream are discussed. Two main theoretical frameworks are the *predictor theory* and the *bias-competition theory*, both of which received different support from the experimental literature [73, 192]. In the *predictor theory*, the main processing goal of the feedback stream is to reduce the residual error between the bottom-up input features and the top-down sensory prediction generated by higher levels of processing [315]. This view is supported by the theoretical framework of Bayesian inference and the Kalman filter realisation of prediction and correction mechanisms in feedback processing [255]. Conversely, in the *bias-competition theory* the main processing goal of the feedback stream is to amplify the gain for those features that match the top-down prediction. In a nutshell, bottom-up signal representations and top-down predictions would achieve a resonant state in which the context re-enters the earlier stages of representation to emphasise their relevance in a larger context [111, 83]. In combination with the above-mentioned mechanisms of response normalization based on non-linear divisive processes the sensory signals that are enhanced by feedback receive a competitive advantage yielding to enhanced response patterns. A corresponding canonical circuit can be found for example in [46]. Variants of such computational elements have been utilized in models by [23, 308] and tested on a set of real-world images and sequences. Further extensions that generate context or attention information and related representations will further improve the selectivity of such top-down signals and, therefore, the resulting enhancements of the sensory feature representation.

5.2 Data encoding and representation

Biological systems are found to use several strategies such as event based sensory processing, distributed multiplexed representation of sensory inputs and active sensory adaptation to the environmental conditions in order to possibly operate in a robust and energy efficient manner.

Traditionally, video input is captured by cameras that operate to generate sequences of frames at a defined rate. The consequence is that the stream of spatio-temporal scene structure is regularly sampled at fixed time steps regardless of the spatio-temporal structure. In other words, the plenoptic function [2] is sliced in sheets of image-like representations. The result of such a strategy is a highly redundant representation of any constant features in the scene along the temporal axis. In contrast, the brain encodes and transmits information through discrete sparse events and this spiking encoding appears at the very beginning of visual information processing, i.e., at the retina level. As discussed in Sec. 4.1, ganglion cells transmit a sparse asynchronous encoding of the time varying visual information to LGN and then cortical areas. This sparse event-based encoding inspired development of new type of camera sensors. Some events are registered whenever changes occur in the spatio-temporal luminance functions which are represented in

a stream of events, with a location and time stamp [178, 182, 248]. Apart from the decrease in redundancy, the processing speed is no longer restricted to the frame-rate of the sensor. Rather, events can be delivered at a rate that is only limited by the refractory period of the sensor elements. Using these sensors brings massive improvements in terms of efficiency of scene encoding and computer vision approaches could benefit from such an alternative representation as demonstrated already on some isolated tasks.

In terms of representation, examining the richness of receptive fields of cells from retina of the visual cortex (such as in V1, MT and MST) shows that the visual system is almost always using a distributed representation for the sensory inputs. Distributed representation helps the system in a multiplicity of ways: It allows for an inherent representation for the uncertainty, it allows for task specific modulation and it could also be useful for representing the multiplicity of properties such as transparent/layered motion. Another important observation to be made is that the visual features that are encoded since the early stages could be optimal for carrying out computations related to multiplicity of tasks in the higher areas. Thus, taking into account this richness of representations could be very helpful to design systems that could deal with an ensemble of tasks simultaneously instead of subserving a single task at a time.

5.3 Psychophysics and human perceptual performance data

Psychophysical laws and principles which can explain large amounts of empirical observations should be further explored and exploited for designing robust vision algorithms. Weber law descriptor proposed in [62] could be cited as a good example for such an instance. Human vision dynamically adjusts decision-boundaries related to changes observed in the environment. It has been demonstrated that this adaptation can be achieved dynamically by non-linear network properties that incorporate activation transfer functions of sigmoidal shape [111]. In [62], such a principle has been adopted to define a robust image descriptor that adjusts its sensitivity to the overall signal energy, similar to human sensitivity shifts.

Most of the problems in computer vision are ill-posed, as observable data is insufficient in terms of variables to be estimated. In order to overcome this limitation, biological systems exploit statistical regularities. The data from human performance studies either on highly controlled stimuli with careful variations in specific attributes or large amounts of unstructured data can be used to identify the statistical regularities, particularly significant for identifying operational parameter regimes for computer vision algorithms. This strategy is already being explored in computer vision and is getting more popular with the introduction of larges scale internet based labelling tools such as [275, 327, 313]. Classic examples for this approach in the case of scene segmentation are exploration of human marked ground truth data for static scenes [195] and dynamic scenes [102].

We therefore argue that further investigation into the front-end interfaces to learning functions of decision-making or separation boundaries for classifiers might improve the performance level of already developed algorithms or new augmented versions of it. Emerging work such as [279] indicates the potential in this direction. [279] use the human performance errors and difficulties for the task of face detection to bias the cost function of the svm to get closer to the strategies that we might be adapting or trade-offs that our visual systems are banking on.

6 Conclusion

Computationals models of biological vision aim at identifying and understanding the strategies used by visual systems to solve ill posed problems which are often the same as the ones encountered in computer vision. As a consequence, these models which are actively emerging could not only shed light into functioning of biological vision but also provide solutions to engineering problems tackled by computer

vision. In the past, largely due to the scale at which brain has been examined, these models were limited to capture observations at a scale not directly relevant to solve tasks of interest for the computer vision community which took a more task centric approach. Since then, a great deal of progress has been made by both communities: biological vision is slowly moving towards systems level understanding while computer vision has developed a great deal of task centric algorithms and datasets enabling rapid evaluation. At this juncture task centric algorithmic evaluation remains one of the major stumbling block for cross fertilisation of ideas between both the domains. Computer vision engineers often ignore ideas that are not thoroughly evaluated on established datasets and modellers often limit themselves to evaluating highly selected set of stimuli. Reasons why we are not witnessing studies of emerging computational models of biological vision on standard datasets might be that computer vision is not the primary target of the modeler or that they are not surpassing the state of art performance in terms of metrics on computer vision datasets and thus might be facing rejection on the grounds of performance. This has detrimental effect on both sides: computer vision engineers might miss out efficient strategies developed by biological vision systems and modellers might miss out on limitations of their models on complex real world stimuli. To bridge this gap, a necessary preliminary step is to conduct a systematic task-based evaluation of biological models on datasets regularly used by computer vision community which would enable us to build on strengths of ideas coming from both sides of investigation. There is high need for such kind of studies to be published keeping aside few performance draw backs of models to set a strong basis for a new departure of bio-inspired computer vision.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318723 (MATHEMACS) and grant agreement no. 269921 (BrainScaleS).

References

[1] E. Adelson and J. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2:284–299, 1985.

[2] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[3] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak: Fast retina keypoint. *CVPR*, pages 510—517, 2012.

[4] T. Albright. On the perception of probable things: neural subtrates of associative memory, imagery and perception. *Neuron*, 74:227–245, 2012.

[5] I. Andolina, H. Jones, W. Wang, and A. Sillito. Corticothalamic feedback enhances stimulus response precision in the visual system. *Proceedings of the National Academy of Sciences*, 104(1685–1690), 2007.

[6] A. Angelucci and P. C. Bressloff. Contribution of feedforward, lateral and feedback connections ot the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research*, 154:93–120, 2006.

[7] A. Angelucci and J. Bullier. Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *Journal of Physiology - Paris*, 97(2–3):141–154, 2003.

[8] A. Anzai, X. Peng, and D. Van Essen. Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, 10(10):1313–1321, 2007.

[9] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011.

[10] K. Armstrong, J. Fitzgerald, and T. Moore. Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, 50:791–798, 2006.

[11] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics*, 60(1):156–182, 1999.

[12] V. Auvray, P. Bouthemy, and J. Liénard. Joint Motion Estimation and Layer Segmentation in Transparent Image Sequences—Application to Noise Reduction in X-Ray Image Sequences. *EURASIP Journal on Advances in Signal Processing*, 2009, 2009.

[13] G. Azzopardi and N. Petkov. A corf computational model of a simple cell that relies on lgn input outperforms the gabor function model. *Biological Cybernetics*, 106(3):177–189, 2012.

[14] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[15] J. Bakin, K. Nakayama, and C. Gilbert. Visual responses in monkey areas v1 and v2 to three-dimensional surface configurations. *Journal of Neuroscience*, 20(8188-8198), 2000.

[16] J. S. Bakin, K. Nakayama, and C. D. Gilbert. Visual responses in monkey areas v1 and v2 to three-dimensional surface configurations. *The Journal of Neuroscience*, 20(21):8188–8198, 2000.

[17] J. Baladron, D. Fasoli, and O. Faugeras. Three applications of GPU computing in neuroscience. *Computing in Science and Engineering*, 14(3):40–47, 2012.

[18] D. Ballard, M. Hayhoe, G. Salgian, and H. Shinoda. Spatio-temporal organization of behavior. *Spatial Vision*, 13(2-3):321–333, 2000.

[19] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994.

[20] A. Basole, L. White, and D. Fitzpatrick. Mapping multiple features in the population response of visual cortex. *Nature*, 423:986–990, 2003.

[21] A. Bastos, W. Usrey, R. Adams, G. Mangun, P. Fries, and K. Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4):695 – 711, 2012.

[22] P. Baudot, M. Levy, O. Marre, M. Pananceau, and Y. Fregnac. Animation of natural scene by virtual eye-movements evoke high precision and low noise in v1 neurons. *Frontiers in Neural Circuits*, 7:206, 2013.

[23] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16(10):2041–2066, 2004.

[24] P. Bayerl and H. Neumann. Disambiguating visual motion by form–motion interaction – a computational model. *International Journal of Computer Vision*, 72(1):27–45, 2007.

[25] W. Beaudot. *The neural information processing in the vertebrate retina: A melting pot of ideas for artificial vision.* PhD thesis, PhD Thesisin Computer Science, INPG (France), Dec. 1994.

[26] C. Beck and H. Neumann. Interactions of motion and form in visual cortex – a neural model. *Journal of Physiology - Paris*, 104:61–70, 2010.

[27] O. Ben-Shahar and S. Zucker. Geometrical computations explain projection patterns of long-range horizontal connections

in visual cortex. *Neural Computation*, 16(3):445—476, 2004.

[28] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009. Also published as a book. Now Publishers, 2009.

[29] A. Benoit, D. Alleysson, J. Hérault, and P. Le Callet. Spatio-temporal tone mapping operator based on a retina model. In *Computational Color Imaging Workshop*, 2009.

[30] A. Benoit, A. Caplier, B. Durette, and J. Herault. Using human visual system modeling for bio-inspired low level image processing. *Computer Vision and Image Understanding*, 114(7):758 – 773, 2010.

[31] R. Benosman, S.-H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan. Asynchronous frameless event-based optical flow. *Neural Networks*, 27:32–37, 2011.

[32] M. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. Imag. Proc.*, 7(3):421–432, 1998. Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.

[33] D. Bock, W.-C. A. Lee, A. Kerlin, M. Andermann, G. Hood, A. Wetzel, S. Yurgenson, E. Soucy, H. Kim, and R. Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(177-182), 2011.

[34] J. Bohannon. Helping robots see the big picture. *Science*, 346:186–187, 2014.

[35] E. Borenstein and S. Ullman. Combined top-down/bottom-up segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2109–2125, 2008.

[36] R. Born and D. Bradley. Structure and function of visual area MT. *Annu. Rev. Neurosci*, 28:157–189, 2005.

[37] A. Borst. Fly visual course control: behaviour, algorithms and circuits. *Nature Reviews Neuroscience*, 15:590–599, 2014.

[38] W. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, 17(6):2112–2127, 1997.

[39] J. Bouecke, E. Tlapale, P. Kornprobst, and H. Neumann. Neural mechanisms of motion detection, integration, and segregation: From biology to artificial image processing systems. *EURASIP Journal on Advances in Signal Processing*, 2011, 2011. special issue on Biologically inspired signal processing: Analysis, algorithms, and applications.

[40] O. Braddick. Segmentation versus integration in visual motion processing. *Trends in neurosciences*, 16(7):263–268, 1993.

[41] D. Bradley and M. Goyal. Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, 9(9):686–695, 2008.

[42] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[43] P. C. Bressloff and J. D. Cowan. A spherical model for orientation and spatial frequency tuning in a cortical hypercolumn. *Philosophical Transactions of the Royal Society B*, 2003.

[44] F. Briggs and W. M. Usrey. Emerging views of corticothalamic function. *Current Opinion in Neurobiology*, 18(4):403–407, Aug. 2008.

[45] R. S. A. Brinkworth and D. C. O'Carroll. Robust models for optic flow coding in natural scenes inspired by insect biology. *PLoS Comput Biol*, 5(11):e1000555, 2009.

[46] T. Brosch and H. Neumann. Interaction of feedforward and feedback streams in visual cortex in a firing-rate model of columnar computations. *Neural Networks*, 54(0):11 – 16, 2014.

[47] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61:211–231, 2005.

[48] E. Brunswik and J. Kamiya. Ecological cue-validity of 'proximity' and of other gestalt factors. *The American Journal of Psychology*, 66(1):20–32, 1953.

[49] J. Bullier. Integrated model of visual processing. *Brain Res. Reviews*, 36:96–107, 2001.

[50] G. T. Buracas and T. D. Albright. Contribution of area MT to perception of three-dimensional shape: a computational study. *Vision Res*, 36(6):869–87, 1996.

[51] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VI*, ECCV'12, pages 611–625, Berlin, Heidelberg, 2012. Springer-Verlag.

[52] C. Cadieu, M. Kouh, A. Pasupathy, C. Connor, M. Riesenhuber, and T. Poggio. A model of V4 shape selectivity and invariance. *Journal of Neurophysiology*, 98(1733-1750), 2007.

[53] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):769–798, Nov. 1986.

[54] M. Carandini. From circuits to behavior: a bridge too far? *Nature Publishing Group*, 15(4):507–509, Apr. 2012.

[55] M. Carandini, J. B. Demb, V. Mante, D. J. Tollhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–10597, Nov. 2005.

[56] M. Carandini and D. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2011.

[57] J. R. Cavanaugh, W. Bair, and J. A. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of Neurophysiology*, 88(5):2530–2546, 2002.

[58] L. M. Chalupa and J. Werner, editors. *The visual neurosciences*. MIT Press, 2004. Two volumes.

[59] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb. 2001.

[60] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[61] F. Chavane, D. Sharon, D. Jancke, O. Marre, Y. Frégnac, and A. Grinvald. Lateral spread of orientation selectivity in v1 is controlled by intracortical cooperativity. *Frontiers in Systems Neuroscience*, 5, 2011.

[62] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. WLD: a robust local image descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1705–1720, Sept. 2010.

[63] Y. Chen, W. Geisler, and E. Seidemann. Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neuroscience*, 9(11):1412–1420, 2006.

[64] E. J. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Comput. Neural Syst.*, 12:199–213, 2001.

[65] R. Cipriani and C. Pack. Direction selectivity of center-surround interactions in macaque mt. *Journal of Vision*, 10(7):933, 2010.

[66] G. Coleman and H. C. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.

[67] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.

[68] B. Conway and M. Livingstone. Space-time maps and two-bar interactions of different classes of direction-selective cells in macaque V1. *Journal of Neurophysiology*, 89:2726–2742, 2003.

[69] D. Cox. Do we understand high-level vision? *Current Opinion in Neurobiology*, 25(187-193), 2014.

[70] D. D. Cox and T. Dean. Neural networks and neuroscience-inspired computer vision. *Current Biology*, 24(18):921–929, 2014.

[71] E. Craft, H. Schutze, E. Niebur, and R. von der Heydt. A neural model of figure-ground organization. *J. Neurophysiology*, 97:4310–4326, 2007.

[72] A. Cruz-Martin, R. El-Danaf, F. Osakada, B. Sriram, O. Dhande, P. Nguyen, E. Callaway, A. Ghosh, and A. Huberman. A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex. *Nature*, 507:358–361, 2014.

[73] J. Cudeiro and A. M. Sillito. Looking back: corticothalamic feedback and early visual processing. *Trends in Neurosciences*, 29(6):298–306, June 2006.

[74] Y. Cui, L. Liu, F. Khawaja, C. Pack, and D. Butts. Diverse suppressive influences in area mt and selectivity to complex motion features. *Journal of Neuroscience*, 33(42):16715–16728, 2013.

[75] D. Dacey, O. S. Packer, L. Diller, D. Brainard, B. Peterson, and B. Lee. Center surround receptive field structure of cone bipolar cells in primate retina. *Vision Research*, 40:1801–1811, 2000.

[76] G. DeAngelis and W. Newsome. Organization of disparity-selective neurons in macaque area mt. *The Journal of neuroscience*, 19(4):1398, 1999.

[77] E. DeYoe and D. C. Van Essen. Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, 11(219-226), 1988.

[78] O. Dhande and A. Huberman. Retinal ganglion cell maps in the brain: implications for visual processing. *Current Opinion in Neurobiology*, 24:133–142, 2014.

[79] J. J. DiCarlo, D. Zoccolan, and N. C. Rust. How does the brain solve visual object recognition? In *Neuron*, pages 415–434. Cell Press,, Jan. 2012.

[80] K. Dimova and M. Denham. A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive bayesian estimation. *Biological Cybernetics*, 100(3):185–201, 2009.

[81] N. J. Dominy and P. W. Lucas. Ecological importance of trichromatic vision to primates. *Nature*, 410(6826):363–366, 2001.

[82] R. Douglas and K. A. C. Martin. Neuronal circuit of the neocortex. *Ann. Rev. Neuroscience*, 27:419, 2004.

[83] G. M. Edelman. Neural darwinism: Selection and reentrant signaling in higher brain function. *Neuron*, 10(2):115 – 125, 1993.

[84] Editorial. Focus on neurotechniques. *Nature Neuroscience*, 16(7):771–771, June 2013.

[85] H. Eichner, T. Klug, and A. Borst. Neural simulations on multi-core architectures. *Frontiers in Neuroinformatics*, 3(21), 2009.

[86] R. Emerson, J. Bergen, and E. Adelson. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, 32:203–218, 1992.

[87] C. Enroth-Cugell and J. Robson. Functional characteristics and diversity of cat retinal ganglion cells. basic characteristics and quantitative description. *Investigative Ophthalmology and Visual Science*, 25(250-257), 1984.

[88] M.-J. Escobar and P. Kornprobst. Action recognition via bio-inspired features: The richness of center-surround interaction. *Computer Vision and Image Understanding*, 116(5):593—605, 2012.

[89] M.-J. Escobar, G. S. Masson, T. Viéville, and P. Kornprobst. Action recognition using a bio-inspired feedforward spiking network. *International Journal of Computer Vision*, 82(3):284, 2009.

[90] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[91] A. Fairhall. The receptive field is dead. long life the receptive field? *Current Opinion in Neurobiology*, 25:9–12, 2014.

[92] L. Fei-Fei, A. Iyer, C. Koch, and P. Perona. What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1):–10, 2007.

[93] D. Felleman and D. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1:1–47, 1991.

[94] J. Fernandez, B. Watson, and N. Qian. Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research*, 42(7):863–898, 2002.

[95] S. Ferradans, M. Bertalmio, E. Provenzi, and V. Caselles. An Analysis of Visual Adaptation and Contrast Perception for Tone Mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2002–2012, Oct. 2011.

[96] D. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: evidence for a local "association field". *Vision Research*, 33(2):173–193, 1993.

[97] C. Fowlkes, D. Martin, and J. Malik. Local figure-ground cues are valid for natural images. *J. of Vision*, 7(8):1–9, 2007.

[98] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision — ECCV 2002*, volume 2352 of *Lecture Notes in Computer Science*, pages 408–422. Springer Berlin Heidelberg, 2002.

[99] P. Fries. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32:209–224, 2009.

[100] J. P. Frisby and J. V. Stone. *Seeing, Second Edition: The Computational Approach to Biological Vision.* The MIT Press, 2nd edition, 2010.

[101] K. Fu and J. Mui. A survey on image segmentation. *Pattern Recognition*, 13:3–16, 1981.

[102] F. Galasso, N. Nagaraja, T. Cardenas, T. Brox, and B.Schiele. A unified video segmentation benchmark: Annotation, metrics and analysis. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[103] W. Geisler, J. Perry, B. Super, and D. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.

[104] W. S. Geisler. Motion streaks provide a spatial code for motion direction. *Nature*, 400(6739):65–69, July 1999.

[105] S. Gharaei, C. Talby, S. Solomon, and S. S.G. Texture-dependent motion signals in primate middle temporal area. *Journal of Physiology*, 591:5671–5690, 2013.

[106] M. Giese and T. Poggio. Neural mechanisms for the recognition of biological movements and actions. *Nature Reviews Neuroscience*, 4:179–192, 2003.

[107] A. Gilad, E. Meirovithz, and H. Slovin. Population responses to contour integration: early encoding of discrete elements and late perceptual grouping. *Neuron*, 2(389–402), 2013.

[108] C. D. Gilbert and W. Li. Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5):350–363, 2013.

[109] T. Gollisch and M. Meister. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2):150–164, Jan. 2010.

[110] M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, Jan. 1992.

[111] S. Grossberg. How does the brain build a cognitive code? *Psychol. Science*, 87(1):1–51, 1980.

[112] S. Grossberg. How does a brain build a cognitive code? In *Studies of Mind and Brain*, volume 70 of *Boston Studies in the Philosophy of Science*, pages 1–52. Springer Netherlands, 1982.

[113] S. Grossberg. A solution of the figure-ground problem for biological vision. *Neural Networks*, 6:463–483, 1993.

[114] S. Grossberg. 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55(1):48–120, 1994.

[115] S. Grossberg. The complementary brain: unifying brain dynamics and modularity. *Trends in Cognitive Science*, 55(4):233–246, 2000.

[116] S. Grossberg and E. Mingolla. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review*, 92(2):173–211, 1985.

[117] S. Grossberg, E. Mingolla, and C. Pack. A neural model of motion processing and visual navigation by cortical area mst. *Cerebral Cortex*, 9(8):878–895, Dec. 1999.

[118] S. Grossberg, E. Mingolla, and W. D. Ross. Visual brain and visual perception: how does the cortex do perceptual grouping? *Trends in Neurosciences*, 20(3):106–111, 1997.

[119] S. Grossberg, E. Mingolla, and L. Viswanathan. Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41(19):2521–2553, 2001.

[120] A. Grunewald, D. Bradley, and R. Andersen. Neural correlates of structure-from-motion perception in macaque area V1 and MT. *Journal of Neuroscience*, 22(14):6195–6207, 2002.

[121] G. Guy and G. Medioni. Inference of surfaces, 3d curves, and junctions from sparse, noisy, 3d data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(11):1265–1277, 1997.

[122] B. Hassenstein and R. W. Systemtheoretische analyse der zeit, reihenfolgen und vorzeichenauswertung. *In The Bewegungsperzeption Des weevil Chlorophanus. Z. Naturforsch.*, 1956.

[123] M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188 – 194, 2005.

[124] J. H. Hedges, Y. Gartshteyn, A. Kohn, N. C. Rust, M. N. Shadlen, W. T. Newsome, and J. A. Movshon. Dissociation of neuronal and psychophysical responses to local and global motion. *Current Biology*, 21(23):2023–2028, 2011.

[125] D. Heeger. Optical flow using spatiotemporal filters. *The International Journal of Computer Vision*, 1(4):279–302, Jan. 1988.

[126] J. Hegdé and D. Felleman. Reappraising the Functional Implications of the Primate Visual Anatomical Hierarchy. *The Neuroscientist*, 13(5):416–421, Oct. 2007.

[127] J. Hegde and D. C. Van Essen. A comparative study of shape representation in macaque visual areas v2 and v4. *Cerebral Cortex*, 17(5):1100–1116, 2007.

[128] M. Helmstaedter, K. Briggman, S. Turaga, V. Jain, S. Seung, and W. Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500:168–174, 2013.

[129] J. Hérault. *Vision: Images, Signals and Neural Networks: Models of Neural Processing in Visual Perception*. World Scientific, 2010.

[130] J. Hérault and B. Durette. Modeling visual perception for image processing. In F. Sandoval, A. Prieto, J. Cabestany, and M. Grana, editors, *Computational and Ambient Intelligence : 9th International Work-Conference on Artificial Neural Networks, IWANN 2007*, 2007.

[131] D. Heslip, T. Ledgeway, and P. McGraw. The orientation tuning of motion streak mechanisms revealed by masking. *Journal of Vision*, 13(9):376, 2013.

[132] E. C. Hildreth and C. Koch. The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience*, 10(1):477–533, 1987. PMID: 3551763.

[133] G. E. Hinton and S. Osindero. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:2006, 2006.

[134] D. Hoiem, A. A. Efros, and M. Hebert. Recovering occlusion boundaries from an image. *Int. J. Comput. Vision*, 91(3):328–346, Feb. 2011.

[135] B. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[136] G. Horwitz. What studies of macaque monkeys have told us about human color vision. *Neuroscience*, in press, 2014.

[137] X. Huang, T. Albright, and G. Stoner. Adaptive surround modulation in cortical area MT. *Neuron*, 53:761–770, 2007.

[138] D. H. Hubel. Exploration of the primary visual cortex, 1955–78. *Nature*, 299(5883):515–524, 1982.

[139] A. C. Huk. Multiplexing in the primate motion pathway. *Vision Research*, 62(0):173 – 180, 2012.

[140] J. Hupé, A. James, B. Payne, S. Lomber, P. Girard, and J. Bullier. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394:784–791, 1998.

[141] J. Issacson and M. Scanziani. How inhibition shapes cortical activity. *Neuron*, 72:231–240, 2011.

[142] G. Jacobs. Primate color vision: a comparative perspective. *Visual Neuroscience*, 25(5-6):619–633, 2008.

[143] D. Jancke, F. Chavane, S. Naaman, and A. Grinvald. Imaging cortical correlates of illusion in early visual cortex. *Nature*, 428:423–426, 2004.

[144] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2146–2153, 2009.

[145] D. Jeurissen, M. W. Self, and P. R. Roelfsema. Surface reconstruction, figure-ground modulation, and border-ownership. *Cognitive Neuroscience*, 4(1):50–52, 2013. PMID: 24073702.

[146] V. S. Johnston. Mate choice decisions: the role of facial beauty. *Trends in Cognitive Sciences*, 10(1):9 – 13, 2006.

[147] P. Jolicoeur, S. Ullman, and M. Mackay. Curve tracing: A possible basic operation the perception of spatial relations. *Memory and Cognition*, 14(2):129–140, 1986.

[148] H. E. Jones, I. M. Andolina, B. Ahmed, S. D. Shipp, J. T. C. Clements, K. L. Grieve, J. Cudeiro, T. E. Salt, and A. M. Sillito. Differential feedback modulation of center and surround mechanisms in parvocellular cells in the visual thalamus. *The Journal of Neuroscience*, 32(45):15946–15951, 2012.

[149] M. JS, C. CW, S. SS, C. SC, and S. SG. Integration and segregation of multiple motion signals by neurons in area mt of primate. *J Neurophysiol.*, 2014.

[150] M. Kapadia, M. Ito, C. Gilbert, and G. Westheimer. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 50:35–41, 1995.

[151] M. K. Kapadia, G. Westheimer, and C. D. Gilbert. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84(4):2048–2062, 2000.

[152] D. B. Kastner and S. A. Baccus. Insights from the retina into the diverse and general computations of adaptation, detection, and prediction. *Current Opinion in Neurobiology*, 25:63–69, Apr. 2014.

[153] P. Kellman and T. Shipley. A theory of visual interpolation in object perception. *Cognitive Psychology*, 23:141–221, 1991.

[154] H. Kim, D. Angelaki, and G. DeAngelis. A novel role for visual perspective cues in the neural computation of depth. *Nature Neuroscience*, 18(1):129–137, 2015.

[155] J. S. Kim, M. J. Greene, A. Zlateski, K. Lee, M. Richardson, S. C. Turaga, M. Purcaro, M. Balkam, A. Robinson, B. F. Behabadi, M. Campos, W. Denk, H. S. Seung, and EyeWirers. Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509(331-336), 2014.

[156] H. Kirchner and S. J. Thorpe. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11):1762 – 1776, 2006.

[157] C. Koch, H. T. Wang, and B. Mathur. Computing motion in the primate's visual system. *Journal of Experimental Biology*, 146(1):115–139, 1989.

[158] U. Kodandaramaiah. The evolutionary significance of butterfly eyespots. *Behavioral Ecology*, 2011.

[159] J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[160] J. Koenderink, W. Richards, and A. van Doorn. Blow-up: a free lunch? *i-Perception*, 3(2):141–145, 2012.

[161] K. Koffka. *Principles of Gestalt psychology*. Routledge & Kegan Paul Ltd., London, 1935.

[162] N. Kogo and J. Wagemans. The "side" matters: How configurality is reflected in completion. *Cognitive Neuroscience*, 4(1):31–45, 2013. PMID: 24073697.

[163] P. Kornprobst and G. Médioni. Tracking segmented objects using tensor voting. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 118–125, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.

[164] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. J. Rodriguez-Sanchez, and L. Wiskott. Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1847–1871, Aug. 2013.

[165] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. Hmdb: A large video database for human motion recognition. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2556–2563, 2011.

[166] V. Lamme. The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience*, 15(2):1605–1615, 1995.

[167] V. Lamme, V. Rodriguez-Rodriguez, and H. Spekreijse. Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex*, 9:406–413, 1999.

[168] V. Lamme, K. Zipser, and H. Spekreijse. Figure-ground activity in primary visual cortex is suppressed by anesthesia. *PNAS*, 95:3263–3268, 1998.

[169] V. A. F. Lamme and P. R. Roelfsema. The distinct modes of vision offered by feedforward and recurrent proce ssing. *Trends in Neurosciences*, 23(11):571–579, 2000.

[170] M. Lappe. Functional consequences of an integration of motion and stereopsis in area MT of monkey extrastriate visual cortex. *Neural Comput.*, 8(7):1449–1461, 1996.

[171] O. Layton and N. Browning. A unified model of heading and path perception in primate mstd. *PLoS Comput Biol*, 10(2):e1003476, 2014.

[172] T. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, 20(7), 2003.

[173] T. Lee and M. Nguyen. Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, 98(4):1907, 2001.

[174] F. F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14):9596–9601, 2002.

[175] J. Li, Y. Tian, T. Huang, and W. Gao. A dataset and evaluation methodology for visual saliency in video. In *Proceedings of the 2009 IEEE international conference on Multimedia and Expo*, ICME'09, pages 442–445, Piscataway, NJ, USA, 2009. IEEE Press.

[176] W. Li, V. Piech, and C. Gilbert. Learning to link visual contours. *Neuron*, 57:442–451, 2008.

[177] Z. Li. The immersed interface method using a finite element formulation. *Applied Numerical Mathematics*, 27(3):253–267, 1998.

[178] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128× 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008.

[179] T. Lindeberg. Feature detection with automatic scale selection. *The International Journal of Computer Vision*, 30(2):77–116, 1998.

[180] S. Lisberger. Visual guidance of smooth-pursuit eye movements: sensation, action and what happens in between. *Neuron*, 66(4):477–491, 2010.

[181] D. Liu, J. Gu, Y. Hitomi, M. Gupta, T. Mitsunaga, and S. Nayar. Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(2):248–260, 2014.

[182] S.-C. Liu and T. Delbruck. Neuromorphic sensory systems. *Current Opinion in Neurobiology*, 20:1–8, 2010.

[183] M. Livingstone and D. H. Hubel. Segregation of form, color, movement and depth: anatomy, physiology and perception. *Science*, 240(740-749), 1988.

[184] H. Lorach, R. Benosman, O. Marre, S.-H. Ieng, J. A. Sahel, and S. Picaud. Artificial retina: the multichannel processing of the mammalian retina achieved with a neuromorphic asynchronous light acquisition device. *Journal of Neural Engineering*, 9(6):066004, Oct. 2012.

[185] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[186] L. Luo, E. Callaway, and K. Svoboda. Genetic dissection of neural circuits. *Neuron*, 57(5):634–660, 2008.

[187] S. Lyu and E. P. Simoncelli. Nonlinear extraction of independent components of natural images using radial gaussianization. *Neural Comput.*, 21(6):1485–1519, June 2009.

[188] O. Mac Aodha, A. Humayun, M. Pollefeys, and G. Brostow. Learning a confidence measure for optical flow. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1107–1120, 2013.

[189] V. Mante and M. Carandini. Mapping of stimulus energy in primary visual cortex. *Journal of Neurophysiology*, 94:788—798, Mar. 2005.

[190] R. Marc, B. Jones, C. Watt, J. Anderson, C. Sigulinsky, and S. Lauritzen. Retinal connectomics: towards complete, accurate networks. *Progress in Retinal and Eye Research*, 37:141–162, 2013.

[191] N. Markov, M. Ercsey-Ravasz, D. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy. Cortical high-density counterstream architectures. *Science*, 342(1238406-1-13), 2013.

[192] N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, P. Barone, C. Dehay, K. Knoblauch, and H. Kennedy. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, 522(1):225–259, 2014.

[193] D. Marr. *Vision*. W.H. Freeman and Co., 1982.

[194] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society London, B*, 207:187–217, 1980.

[195] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, 2001.

[196] A. Martinez-Alvarez, A. Olmedo-Payá, S. Cuenca-Asensi, J. M. Ferrandez, and E. Fernandez. RetinaStudio: A bioinspired framework to encode visual information. *Neurocomputing*, 114:45–53, Aug. 2013.

[197] R. H. Masland. Cell populations of the retina: The proctor lecture. *Investigative Ophthalmology and Visual Science*, 52(7):4581–4591, June 2011.

[198] R. H. Masland. The Neuronal Organization of the Retina. *Neuron*, 76(2):266–280, Oct. 2012.

[199] K. Masmoudi, M. Antonini, and P. Kornprobst. Streaming an image through the eye: The retina seen as a dithered scalable image coder. *Signal Processing-Image Communication*, 2012.

[200] G. Masson and U. Ilg, editors. *Dynamics of Visual Motion Processing*. Neuronal, Behavioral, and Computational Approaches. Springer Verlag, 1 edition, 2010.

[201] G. Masson and L. Perrinet. The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neurosciences and BioBehavioral Reviews*, 36(1):1–25, 2012.

[202] G. Mather, A. Pavan, R. M. Bellacosa, and C. Casco. Psychophysical evidence for interactions between visual motion and form processing at the level of motion integrating receptive fields. *Neuropsychologia*, 50(1):153 – 159, 2012.

[203] J. Maunsell and D. C. Van Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. ii. binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, 49:1148–1167, 1983.

[204] J. McCarthy, D. Cordeiro, and G. Caplovitz. Local form–motion interactions influence global form perception. *Attention, Perception, & Psychophysics*, 74(5):816–823, 2012.

[205] P. W. McOwan and A. Johnston. Motion transparency arises from perceptual grouping: evidence from luminance and contrast modulation motion displays. *Current Biology*, 6(10):1343 – 1346, 1996.

[206] C. Mead and M. Mahowald. A silicon model of early visual processing. *Neural Network*, 1:91–97, 1988.

[207] N. V. K. Medathati, M. Chessa, G. S. Masson, P. Kornprobst, and F. Solari. Adaptive motion pooling and diffusion for optical flow. Technical Report 8695, INRIA, Mar. 2015.

[208] L. Merabet, A. Desautels, K. Minville, and C. Casanova. Motion integration in a thalamic visual nucleus. *Nature*, 396(265–268), 1998.

[209] P. Merolla, J. Arthur, R. Alvarez-Icaza, A. Cassidy, J. Sawada, F. Akopyan, B. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. Esser, R. Appuswamy, B. Taba, A. Amir, M. Flickner, W. Risk, R. Manohar, and D. Modha. Artificial brains. a million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(668-673), 2014.

[210] L. Meylan, D. Alleysson, and S. Süsstrunk. Model of retinal local adaptation for the tone mapping of color filter array images. *J. Opt. Soc. Am. A*, 24(9):2807–2816, 2007.

[211] A. D. Milner and M. A. Goodale. Two visual systems re–viewed. *Neuropsychologia*, 46:774–785, 2008.

[212] P. Mineault, F. Khawaja, D. Butts, and C. Pack. Hierarchical processing of complex motion along the primate dorsal visual pathways. *Proceedings of the National Academy of Sciences*, 109(972-980), 2012.

[213] L. Muller, A. Reynaud, F. Chavane, and A. Destexhe. The stimulus-evoked population response in visual cortex of awake monkey is a propagating wave. *Nature Communications*, 5(3675), 2014.

[214] D. Mumford. On the computational architecture of the neocortex. I. the role of the thalamo-cortical loop. *Biological Cybernetics*, 65:135–145, 1991.

[215] D. Mumford. On the computational architecture of the neocortex. II. the role of the cortico-cortical loop. *Biological Cybernetics*, 66:241–251, 1992.

[216] J. Nadler, M. Nawrot, D. Angelaki, and G. DeAngelis. MT neurons combine visual motion with a smooth eye movement

signal to code depth-sign from motion parallax. *Neuron*, 63(4):523–532, 2009.

[217] K. Nakayama. Biological image motion processing: A review. *Vision Research*, 25:625–660, 1984.

[218] A. Nandy, T. Sharpee, J. Reynolds, and J. Mitchell. The fine structure of shape tuning in area V4. *Neuron*, 78(1102-1115), 2013.

[219] H. Nasser, S. Kraria, and B. Cessac. Enas: a new software for neural population analysis in large scale spiking networks. In *Springer Series in Computational Neuroscience*. Organization for Computational Neurosciences, July 2013.

[220] J. Nassi, C. Gomez-Laberge, G. Kreiman, and R. Born. Corticocortical feedback increases the spatial extent of normalization. *Frontiers in Systems Neuroscience*, 8:105, 2014.

[221] J. Nassi, S. Lomber, and R. Born. Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *Journal of Neuroscience*, 33(19):8504–8517, 2013.

[222] S. Negahdaripour and A. K. Jain. Challenges in computer vision research; future directions of research. In *Final Report*, Lahaina, Maui, Hawaii, June 1991. NSF Workshop.

[223] H. Neumann and E. Mingolla. Computational neural models of spatial integration in perceptual grouping. In T. . P. Kellman, editor, *From Fragments to Objects: Grouping and Segmentation in Vision*, pages 353–400. Amsterdam: Elsevier, 2001.

[224] Z. Ni, C. Pacoret, R. Benosman, S. Ieng, and S. Régnier. Asynchronous event-based high speed vision for microparticle tracking. *Journal of Microscopy*, 245(3):236–244, Nov. 2011.

[225] S. Nishida. Advancement of motion psychophysics: Review 2001-2010. *Journal of Vision*, 11(5):11, 1–53, 2011.

[226] S. Nishimoto and J. L. Gallant. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *The Journal of Neuroscience*, 31(41):14551–14564, 2011.

[227] H. Nothdurft. Texture segmentation and pop-out from orientation contrast. *Vision Research*, 31(6):1073–1078, 1991.

[228] S. Nowlan and T. Sejnowski. Filter selection model for motion segmentation and velocity integration. *J. Opt. Soc. Am. A*, 11(12):3177–3199, 1994.

[229] B. Odermatt, A. Nikolaev, and L. Lagnado. Encoding of luminance and contrast by linear and nonlinear synapses in the retina. *Neuron*, 73(4):758 – 773, 2012.

[230] P. O'Herron and R. von der Heydt. Representation of object continuity in the visual cortex. *J. of Vision*, 11(2):12, 1–9, 2011.

[231] T. Ohshiro, D. E. Angelaki, and G. C. DeAngelis. A normalization model of multisensory integration. *Nature Neuroscience*, 14(6):775–782, May 2011.

[232] G. A. Orban. Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews*, 88(1):59–89, 2008.

[233] C. Pack and R. Born. Cortical mechanisms for the integration of visual motion. In R. H. Masland, T. D. Albright, T. D. Albright, R. H. Masland, P. Dallos, D. Oertel, S. Firestein, G. K. Beauchamp, M. C. Bushnell, A. I. Basbaum, J. H. Kaas, and E. P. Gardner, editors, *The Senses: A Comprehensive Reference*, pages 189 – 218. Academic Press, New York, 2008.

[234] N. Pal and S. Pal. A review of image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[235] P. Parent and S. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):823–839, 1989.

[236] F. J. Pelayo, S. Romero, C. A. Morillas, A. Martinez, E. Ros, and E. Fernandez. Translating image sequences into spike patterns for cortical neuro-stimulation. *Neurocomputing*, 58–60:885–892, 2004.

[237] J. A. Perrone. A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs. *Journal of Vision*, 12(8), 2012.

[238] L. Pessoa, E. Thompson, and A. Noë. Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and brain sciences*, 21:723–802, 1998.

[239] E. Peterhans and R. von der Heydt. Subjective contours: bridging the gap between psychophysics and physiology. *Trends in Neurosciences*, 14(3):112–119, 1991.

[240] M. A. Peterson and E. Salvagio. Inhibitory competition in figure-ground perception: Context and convexity. *Journal of Vision*, 8(16), 2008.

[241] J. Petitot. An introduction to the Mumford-Shah segmentation model. *Journal of Physiology - Paris*, 97:335–342, 2003.

[242] M. Petrou and A. Bharat. *Next generation artificial vision systems: Reverse engineering the human visual system.* Artech House Series Bioinformatics & Biomedical Imaging, 2008.

[243] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[244] N. Pinto and D. Cox. *GPU Meta-Programming: A Case Study in Biologically-Inspired Machine Vision. GPU Computing Gems*, volume 2, chapter 33. Elsevier, 2012.

[245] H. Plesser, J. Eppler, A. Morrison, M. Diesmann, and M.-O. Gewaltig. Efficient parallel simulation of large-scale neuronal networks on clusters of multiprocessor computers. In A.-M. Kermarrec, L. Bougé, and T. Priol, editors, *Euro-Par 2007 Parallel Processing*, volume 4641 of *Lecture Notes in Computer Science*, pages 672–681. Springer Berlin Heidelberg, 2007.

[246] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(6035):314–319, Sept. 1985.

[247] J. Poort, F. Raudies, A. Wannig, V. Lamme, N. H., and P. Roelfsema. The role of attention in figure-ground segregation in areas v1 and v4 of the visual cortex. *Neuron*, 108(5):1392–1402, 2012.

[248] C. Posch, D. Matolin, and R. Wohlgenannt. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE Journal of Solid-State Circuits*, 46(1):259–275, 2011.

[249] T. Potjans and D. M. The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebral Cortex*, 24(785-806), 2014.

[250] M. S. Pratte, S. Ling, J. D. Swisher, and F. Tong. How attention extracts objects from noise. *Journal of Neurophysiology*, 110(6):1346–1356, 2013.

[251] N. Priebe, C. Cassanello, and S. Lisberger. The neural representation of speed in macaque area MT/V5. *Journal of Neuroscience*, 23(13):5650–5661, July 2003.

[252] N. Priebe, S. Lisberger, and A. Movshon. Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *The Journal of Neuroscience*, 26(11):2941—2950, 2006.

[253] N. Qian and R. A. Andersen. A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Research*, 37:1683–1698, 1997.

[254] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli, and T.-S. Chua. An eye fixation database for saliency detection in images. In *ECCV 2010*, Crete, Greece, 2010.

[255] R. Rao and D. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, 2(1):79–87, 1999.

[256] M. J. Rasch, M. Chen, S. Wu, H. D. Lu, and A. W. Roe. Quantitative inference of population response properties across eccentricity from motion-induced maps in macaque V1. *Journal of Neurophysiology*, 109(5):1233–1249, 2013.

[257] F. Raudies, E. Mingolla, and H. Neumann. A model of motion transparency processing with local center-surround interactions and feedback. *Neural Computation*, 23:2868–2914, 2011.

[258] F. Raudies and H. Neumann. A neural model of the temporal dynamics of figure–ground segregation in motion perception. *Neural Networks*, 23(2):160 – 176, 2010.

[259] X. Ren, C. Fowlkes, and J. Malik. Figure/ground assignment in natural images. In A. Leonardis, H. Bischof, and A. Pinz, editors, *Computer Vision – ECCV 2006*, volume 3952 of *Lecture Notes in Computer Science*, pages 614–627. Springer Berlin Heidelberg, 2006.

[260] A. Reynaud, G. Masson, and F. Chavane. Dynamics of local input normalization result from balanced short- and long-range intracortical interactions in area V1. *Journal of Neuroscience*, 32:12558–12569, 2012.

[261] M. Riesenhuber and T. Poggio. Computational models of object recognition in cortex: A review. Technical report, and 190, Artificial Intelligence Laboratory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 2000.

[262] K. Rockland. Collateral branching of long-distance cortical projections in monkey. *Journal of Comparative Neurology*, 521(18):4112–4223, 2013.

[263] K. Rockland and J. Lund. Widespread periodic intrinsic connections in the tree shrew visual cortex. *Science*, 215:1532–1534, 1982.

[264] H. Rodman and T. Albright. Coding of visual stimulus velocity in area mt of the macaque. *Vision Research*, 27(12):2035–2048, 1987.

[265] A. J. Rodriguez Sanchez and J. K. Tsotsos. The roles of endstopped and curvature tuned computations in a hierarchical representation of 2d shape. *PLoS ONE*, 7(8):e42058, 2012.

[266] P. Roelfsema and R. Houtkamp. Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics*, 73(8):2542–2572, 2011.

[267] P. R. Roelfsema. Elemental operations in vision. *Trends in Cognitive Sciences*, 9(5):226–233, 2005.

[268] P. R. Roelfsema. Cortical algorithms for perceptual grouping. *Annual Review*

of Neuroscience, 29(1):203–227, 2006. PMID: 16776584.

[269] P. R. Roelfsema, V. A. F. Lamme, and H. Spekreijse. The implementation of visual routines. *Vision Research*, 40(10–12):1385–1411, 2000.

[270] P. R. Roelfsema, V. A. F. Lamme, H. Spekreijse, and H. Bosch. Figure/ground segregation in a recurrent network architecture. *J. of Cognitive Neuroscience*, 14(4):525–537, 2002.

[271] P. R. Roelfsema, M. Tolboom, and P. S. Khayat. Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron*, 56(5):785 – 792, 2007.

[272] P. Rogister, R. Benosman, and S. H. Ieng. Asynchronous event-based binocular stereo matching. *IEEE Transactions in Neural Networks and Learning Systems*, pages 347–353, 2012.

[273] A. Rosenfeld. Computer vision: a source of models for biological visual processes? *Biomedical Engineering, IEEE Transactions on*, 36(1):93–96, 1989.

[274] G. Rousselet, S. Thorpe, and M. Fabre-Thorpe. How parallel is visual processing in the ventral path? *TRENDS in Cognitive Sciences*, 8(8):363–370, Aug. 2004.

[275] B. C. Russell, A. Torralba, K. Murphy, and W. Freeman. Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3):157–173, 2008.

[276] N. Rust, V. Mante, E. Simoncelli, and J. Movshon. How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, 9:1421–1431, 2006.

[277] N. Rust, O. Schwartz, J. Movshon, and E. Simoncelli. Spatiotemporal elements of macaque V1 receptive fields. *Neuron*, 46:945–956, 2005.

[278] T. Sato, I. Nauhaus, and M. Carandini. Traveling waves in visual cortex. *Neuron*, 75:218–229, 2012.

[279] W. Scheirer, S. Anthony, K. Nakayama, and D. Cox. Perceptual annotation: Measuring human vision to improve computer vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1679–1686, 2014.

[280] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, Feb. 2000.

[281] H. Scholte, J. Jolij, J. Fahrenfort, and V. Lamme. Feedforward and recurrent processing in scene segmentation: electroencephalography and functional magnetic resonance imaging. *J. of Cognitive Neuroscience*, 20(11):2097–2109, 2008.

[282] M. W. Self, T. van Kerkoerle, H. Super, and P. R. Roelfsema. Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Current Biology*, 23(21):2121 – 2129, 2013.

[283] A. Sellent, M. Eisemann, B. Goldlucke, D. Cremers, and M. Magnor. Motion field estimation from alternate exposure images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(8):1577–1589, 2011.

[284] T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences (PNAS)*, 104(15):6424–6429, 2007.

[285] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE PAMI*, 29(3):411–426, Mar. 2007.

[286] R. Shapley. Visual sensitivity and parallel retinocortical channels. *Annual Review of Psychology*, 41:635–658, 1990.

[287] R. Shapley and C. Enroth-Cugell. Visual adaptation and retinal gain controls. *Progress in retinal research*, 3:263–346, 1984.

[288] G. Sheperd and S. Grillner, editors. *Handbook of brain microcircuits*. Oxford University Press, 2010.

[289] J. S. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[290] A. Sigman, A. Cecchi, C. Gilbert, and M. Magnaso. On a common circle: Natural scenes and gestalt rules. *PNAS*, 98(4):1935–1940, 2001.

[291] E. Simoncelli and D. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38:743–761, 1998.

[292] C. Simoncini, L. Perrinet, A. Montagnini, P. Mamassian, and G. Masson. More is not always: adaptive gain control explains dissociation between perception and action. *Nature Neuroscience*, 15(11):1586–1603, 2012.

[293] A. Smolyanskaya, D. A. Ruff, and R. T. Born. Joint tuning for direction of motion and binocular disparity in macaque mt is largely separable. *Journal of Neurophysiology*, 2013.

[294] R. J. Snowden, S. Treue, R. G. Erickson, and R. A. Andersen. The response of area MT and V1 neurons to transparent motion. *The Journal of Neuroscience*, 11(9):2768–2785, 1991.

[295] F. Solari, M. Chessa, K. Medathati, and P. Kornprobst. What can we expect from a classical v1-mt feedforward architecture for optical flow estimation? *Signal Processing: Image Communication*, 2015.

[296] R. Squire, B. Noudoost, R. Schafer, and T. Moore. Prefrontal contributions to visual selective attention. *Annual Review of Neuroscience*, 36:451–466, 2013.

[297] A. N. Stein and M. Hebert. Occlusion boundaries from motion: Low-level detection and mid-level reasoning. *International Journal of Computer Vision*, 82(3):325–357, 2009.

[298] G. R. Stoner and T. D. Albright. Neural correlates of perceptual motion coherence. *Nature*, 358:412–414, 1992.

[299] D. Sun, S. Roth, T. Darmstadt, and M. Black. Secrets of optical flow estimation and their principles. *cvpr*, 2010.

[300] P. Sundberg, T. Brox, M. Maire, P. Arbeláez, and J. Malik. Occlusion boundary detection and figure/ground assignment from optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[301] O. Temam and R. Héliot. Implementation of signal processing tasks on neuromorphic hardware. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1120 –1125, 2011.

[302] A. Thiele, K. Dobkins, and T. Albright. Neural correlates of chromatic motion perception. *Neuron*, 32(351-358), 2001.

[303] A. Thiele and J. Perrone. Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nature Neuroscience*, 4(5):526–532, 2001.

[304] W. Thoreson and S. Mangel. Lateral interactions in the outer retina. *Progress in Retinal and Eye Research*, 31:407–441, 2012.

[305] S. Thorpe. The speed of categorization in the human visual system. *Neuron*, 62(2):168–170, 2009.

[306] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[307] E. Tlapale, P. Kornprobst, G. S. Masson, and O. Faugeras. A neural field model

for motion estimation. In S. Verlag, editor, *Mathematical Image Processing*, volume 5 of *Springer Proceedings in Mathematics*, pages 159–180, 2011.

[308] E. Tlapale, G. S. Masson, and P. Kornprobst. Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, 50(17):1676–1692, Aug. 2010.

[309] S. Tschechne and H. Neumann. Hierarchical representation of shapes in visual cortex - from localized features to figural shape segregation. *Frontiers in Computational Neuroscience*, 8(93), 2014.

[310] S. Tschechne, R. Sailer, and H. Neumann. Bio-inspired optic flow from event-based neuromorphic sensor input. *Artificial Neural Networks in Pattern Recognition*, 8774:171–182, 2014.

[311] J. K. Tsotsos. It's all about the constraints. *Current Biology*, 24(18):854–858, Sept. 2014.

[312] J. M. G. Tsui, J. N. Hunter, R. T. Born, and C. C. Pack. The role of V1 surround suppression in mt motion integration. *Journal of Neurophysiology*, 103(6):3123–3138, 2010.

[313] A. Turpin, D. J. Lawson, and A. M. McKendrick. Psypad: A platform for visual psychophysics on the ipad. *Journal of Vision*, 14(3), 2014.

[314] S. Ullman. Visual routines. *Cognition*, 18:97–159, 1984.

[315] S. Ullman. Sequence seeking and counter streams a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, 5:1–11, 1995.

[316] S. Ullman. Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Science*, 11(2):58–64, 2007.

[317] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7):682–687, 2002.

[318] L. Ungerleider and M. Mishkin. *Two cortical visual systems*, pages 549–586. MIT Press, 1982.

[319] L. G. Ungerleider and J. V. Haxby. 'what' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4(2):157–165, 1994.

[320] D. C. Van Essen. Organization of visual areas in macaque and human cerebral cortex. In L. Chapula and J. Werner, editors, *The Visual Neurosciences*. MIT Press, 2003.

[321] D. C. Van Essen and J. L. Gallant. Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13:1–10, July 1994.

[322] R. VanRullen and S. J. Thorpe. Surfing a spike wave down the ventral stream. *Vision Research*, 42:2593–2615, 2002.

[323] A. Vedaldi and B. Fulkerson. Vlfeat: an open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia*, MM '10, pages 1469–1472, New York, NY, USA, 2010. ACM.

[324] A. Verri and T. Poggio. Against quantitative optical flow. In *Proceedings First International Conference on Computer Vision*, pages 171–180. IEEE Computer Society, 1987.

[325] A. Verri and T. Poggio. Motion field and optical flow: qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):490–498, 1989.

[326] W. Vinje and J. L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287:1273–1276, 2000.

[327] C. Vondrick, D. Patterson, and D. Ramanan. Efficiently scaling up crowdsourced video annotation. *International Journal of Computer Vision*, 101(1):184–204, 2013.

[328] B. A. Wandell, A. El Gamal, and B. Girod. Common principles of image acquisition systems and biological vision. *Proceedings of the IEEE*, 90(1):5–17, 2002.

[329] B. Webb, C. Tinsley, N. Barraclough, A. Parker, and A. Derrington. Gain control from beyond the classical receptive field in primate visual cortex. *Visual Neuroscience*, 20(3):221–230, 2003.

[330] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1663–1668, 2009.

[331] R. M. Willems. Re-Appreciating the Why of Cognition: 35 Years after Marr and Poggio. *Frontiers in Psychology*, 2, 2011.

[332] J. Williford and R. von der Heydt. Border-ownership coding. *Scholarpedia*, 8(10):30040, 2013.

[333] A. Witkin and J. Tenenbaum. *Human and Machine Vision*, chapter On the role of structure in vision, pages 481–543. Academic Press, 1983.

[334] A. Wohrer and P. Kornprobst. Virtual Retina : A biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience*, 26(2):219, 2009. DOI 10.1007/s10827-008-0108-4.

[335] J. M. Wolfe, A. Oliva, T. S. Horowitz, S. J. Butcher, and A. Bompas. Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, 42(28):2985 – 3004, 2002.

[336] D. Xiao, S. Raiguel, V. Marcar, J. Koenderink, and G. A. Orban. Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proceedings of the National Academy of Sciences*, 92(24):11303–11306, 1995.

[337] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3485–3492, 2010.

[338] N. Yabuta, A. Sawatari, and E. Callaway. Two functional channels from primary visual cortex to dorsal visual cortex. *2001*, 297-301, 292.

[339] Z. Yang, D. Heeger, B. R., and E. Seidemann. Long-range traveling waves of activity triggered by local dichoptic stimulation in V1 of behaving monkeys. *Journal of Neurophysiology*, 2014.

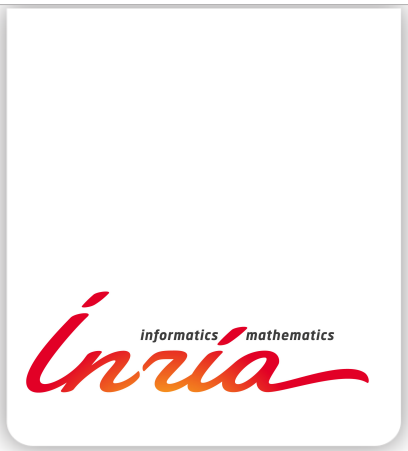
[340] A. Yuille and A. Oliva. Frontiers in computer vision: NSF white paper. In *Frontiers in Computer Vision Workshop*, 2010.

[341] S. Zeki. *A vision of the brain*. Blackwell Scientific Publications, 1993.

[342] Y. Zhang, I.-J. Kim, J. R. Sanes, and M. Meister. The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proceedings of the National Academy of Sciences*, 109(36):E2391–E2398, 2012.

[343] H. Zhou, H. S. Friedman, and R. von der Heydt. Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, 20(17):6594–6611, 2000.

[344] Y. Zhuo, T. Zhou, H. Rao, J. Wang, M. Meng, M. Chen, C. Zhou, and L. Chen. Contributions of the visual ventral pathway to long-range apparent motion. *Science*, 299(5605):417, 2003.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399