

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Models of Information Processing in the Visual Cortex

---

Vincent de Ladurantaye, Jean Rouat and Jacques Vanden-Abeelee

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/50616>

---

## 1. Introduction

In the most general meaning of the word, a model is a way of representing something of interest. If we ask a biologist to model the visual system, he will probably talk about neurons, dendrites and synapses. On the other hand, if we ask a mathematician to model the visual system, he will probably talk about variables, probabilities and differential equations.

There are an almost infinite number of ways to model any given system, and this is particularly true if we are interested in something as complex as the human brain, including the visual cortex. It is impossible to review every type of model present in the literature. This chapter does not, by any means, claim to be an accurate review of all the models of vision published so far. Instead, we try to regroup the different types of models into small categories, to give the reader a good overview of what is possible to achieve in terms of modeling biological vision. Each section of the chapter concentrates on one of these categories, begins with an outline of the general properties and goals of the global model category, and then explores different implementation methods (i.e. specific models).

### 1.1. Why models?

There are many reasons why one might want to elaborate and use a model. Different kinds of models achieve different goals, however they all serve a common purpose: a tool to learn and better understand our world. Models are used as a simplification of reality. Many models only represent a specific aspect of a complex system and make arbitrary assumptions about the rest of the system which may be unrelated, judged negligible or simply unknown. A good example is the fact that current neurobiological models of the brain mainly focus on modeling neurons, neglecting interneurons and glial cells. Current models ignore glial cells, probably because they are not fully understood. Some models might purposefully ignore some parts of a system to isolate a specific structure and/or function that we want to understand.

Because of the extreme complexity of the cortex, simplification is unavoidable and even productive. Modeling is an efficient approach for simplifying reality as a way to gain a better

understanding of the system in question. In fact, one should strive to design the simplest possible model that could be complexified if necessary.

For a more philosophical discussion on the role of models in science, the work by Frigg and Hartmann [15] can be recommended. They describe two fundamental functions of models. First, models can represent a selected part of the world, and second, models can represent a theory. The two notions are not mutually exclusive, but they provide a good distinction between more practical models trying to reproduce a certain system, and more theoretical models constrained to a certain set of laws. As we don't have yet an established theory as how the brain works, this chapter focuses on the former.

Moreover, models can also be categorized into models of data and models of phenomena. This is an important distinction because they consist of two drastically different approaches to the same problem that, in our case, is understanding the visual cortex. Models of data try to reproduce data we observe in the brain, and models of phenomena try to reproduce the “miracle” of the visual cortex, which is vision. This chapter emphasizes practical models and the functioning of the visual cortex from an information-processing point of view.

## 1.2. Modeling the visual cortex

As established in previous section, we concentrate more on models of phenomena to understand how the visual system works, instead of models of data that are trying to be biologically accurate. We do so, because science is yet to provide a full understanding of the brain, thus it is not possible to propose accurate overall models of data.

The chapter rather gives an overview of different kinds of models of vision present in the literature, whether they are biologically accurate or not. Modeling with artificial neural networks is not the only way. There are many other models based on mathematical representations, statistics, physics principles that can represent behaviors observed in the visual system. Any model able to represent some aspects of the vision process can provide valuable insights as to how the visual cortex really works.

Because of the complexity, there exists no model in the literature that can represent visual cortex as a whole. For the sake of simplicity, when designing a model, only specific functions of the visual system are usually considered. Of course, there are models more encompassing than others with more functionalities. However, most of them deal with only one of the two main visual functions, that is, localization or recognition. Ironically, this fits the often used ventral vs. dorsal stream model of the visual cortex, where the ventral stream explains the “what?” and the dorsal stream the “where?”. This chapter is thus divided into three main sections:

1. A short section concerning models trying to be biologically accurate.
2. A section concerning models associated with the dorsal pathway, which deals with movement and localization.
3. A section on models related to the ventral pathway, which mainly deals with object recognition.

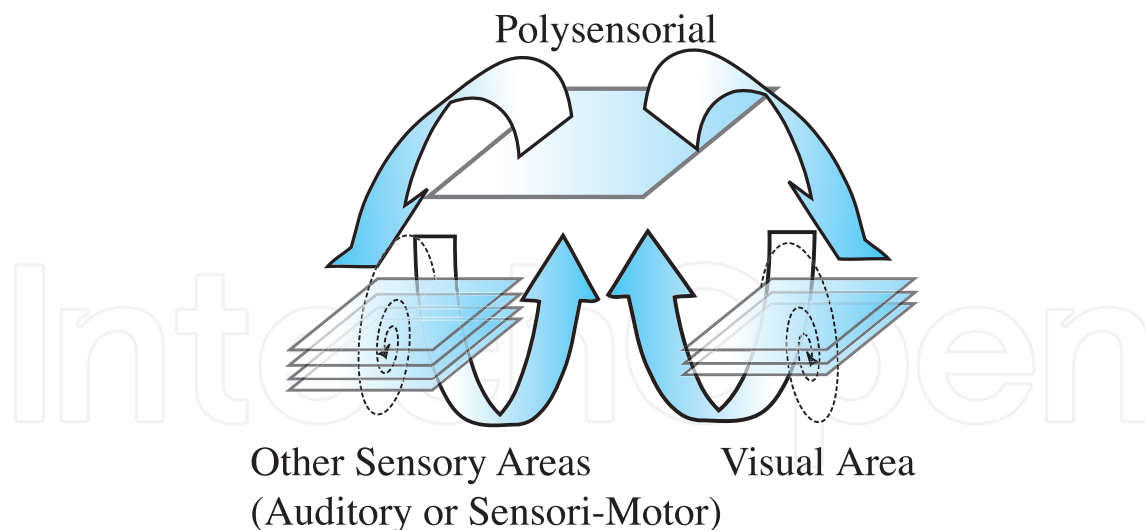
Before doing so, we first discuss the point that the visual cortex is embedded in a huge neural network, interacting and cooperating with other cortical and sensorial areas.

## 2. Visual cortex: integrated into a cooperative network

### 2.1. Integration into a collaborative network

Even if the organization of the brain is considered to be specific to the task, the various modalities seem to use the same independent processing mechanism [57]. Specialization of the cortices appears with the stimulation through specific sensory pathways. The visual cortex is interconnected with other sensory modalities and is therefore multisensorial [6]. For example, the visual cortex is also involved in the perception of tactile orientations as observed by Zangaladze et al. [68] and other authors. Many models of interactions between the visual cortex and other modalities are presented in the literature. For example, the recent work by Ohshiro et al. [43] in which a model of multisensory neural layer is proposed. From an engineering perspective, new applications that integrate multisensory perception can be designed. Wysoski et al. [64] give an example of a multimedia application in which models of audition and vision are combined through an implementation of a network of spiking neurons for the identification of 35 persons.

Figure 1 is our sketchy representation summarizing our integrated and general view of how functional multimodal information processing is done in the cortex. By no means, are we attempting to reproduce the biological characteristics of the cortex in relation to this figure. Emphasis is put on the main streams of information processing in relation with the visual cortex. We symbolize the cortical polysensoriality with a dedicated neural cortical layer, while there is no physiological agreement on the existence of such dedicated layer. In this section, interpretation of the functioning of the visual cortex is based on information and signal processing techniques to facilitate model designs without taking into account all physiological complexities.



**Figure 1.** Functional view of the information processing going on in the cortex and between various sensorial areas of the cortex. Layers at bottom illustrate the hierarchical structure of feature extraction, corresponding to visual, auditory, somato-sensorial, sensori-motor, etc. The upper layer represents the multisensorial areas of the cortex. Large arrows represent the feedforward flow of information going from the sensorial areas to the polysensorial “layer” and feedbacks from the upper layer to the sensorial (and hierarchical feature) layers. Dotted arrows represent the iterative process of feature extraction which is occurring inside the sensory layers.

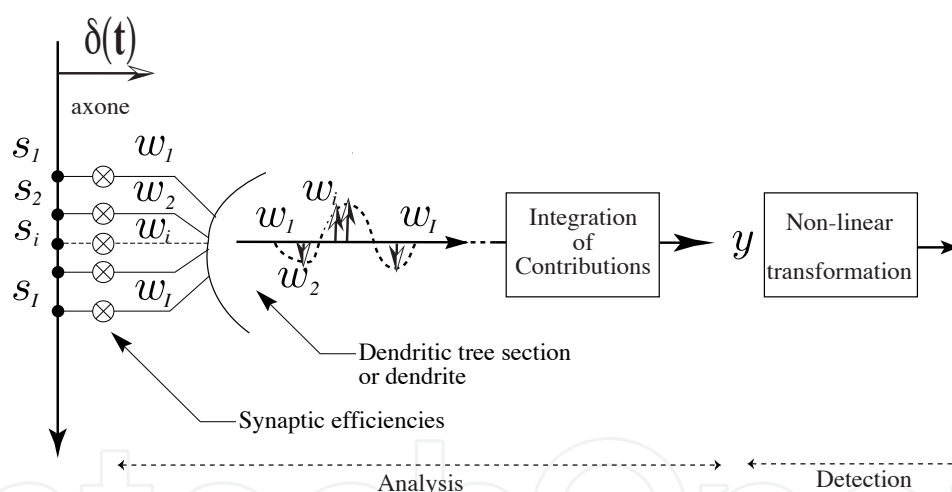
According to the common view, we represent sensing areas of the cortex in figure 1 as a hierarchy of layers and the “polysensorial” layer represents the connectivity between cortical areas. We know that connectivities between audition, vision and the motor system at lower levels of the senses exist (e.g. in the colliculus [6]), but these interactions are not discussed in this chapter.

Even if the perspective presented here is very schematic, it is still of interest for research and a better understanding of the cortex. Next subsections introduce models of feature-extraction through the visual sensory layers.

## 2.2. Hierarchical and sparse low level feature extractions

Low level features and characteristics are hypothesized to be hierarchically obtained through a succession of feedforward and feedback loops. This process is commonly modeled as a cascade of layers made of filters that mimic the receptive fields of neurons. Each layer output is subject to nonlinear transformations. Layers are interconnected into hierarchical structures that can be feedforward or include feedbacks as illustrated at the bottom of figure 1. We introduce here two types of models based on signal processing techniques or on artificial neural networks.

### 2.2.1. Signal processing models



**Figure 2.** Illustration of the receptive field of a neuron representing an elementary analysis function and the nonlinear output being associated to a decision function. In this illustration, a hypothetical presynaptic axon is sketched and it synapses with  $I$  contacts on a postsynaptic neuron.

In these approaches, neurons are assumed to be made of two simple modules: analysis and then detection. The analysis is modeled as a filter which characterizes to some extent the receptive field of a neuron. Detection is modeled as a threshold or a nonlinear function. When the input to the neuron matches its receptive field, the detection module has a strong response.

There is an intense research work going on the automatic finding and generation of receptive fields to be used in computational models or artificial vision models and systems. Signal processing methods estimating the receptive fields of models are briefly summarized here.

These models assume that any stimulus  $s(t)$  (denoted as a vector) to a neuron is transformed by this neuron into a new representation by projection on an elementary function  $w(t)$

(also denoted as vector). Each elementary function  $\mathbf{w}(t)$  (or “basis function”, in the signal processing jargon) represents the receptive field of a particular neuron (or group of neurons). The result of the projection of  $\mathbf{s}(t)$  on the specific basis function  $\mathbf{w}(t)$  is a scalar number  $y(t)$  that is added to the transmembrane potential of the neuron. It corresponds to the contribution of the stimulus input  $\mathbf{s}(t)$  to the neuron-activity. It is summarized below in mathematical terms.

$$\begin{aligned} y(t) &= \mathbf{w}(t) \cdot \mathbf{s}(t) \\ \text{or} \\ y(t) &= \sum_{i=1}^I w_i(t) s_i(t) \end{aligned} \quad (1)$$

with

$$\mathbf{w}(t) = \begin{bmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_I(t) \end{bmatrix} \quad \text{and} \quad \mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_I(t) \end{bmatrix}$$

$I$  is the dimension of the stimulus and receptive field (assuming that the neuron has  $I$  synapses).  $w_i$  is the efficiency at time  $t$  of synapse  $i$  from the neuron. In other words, at time  $t$ ,  $y(t)$  (equation 1) is the degree of similarity between the input stimulus  $\mathbf{s}(t)$  and the receptive field  $\mathbf{w}(t)$  of the neuron. In signal processing jargon,  $y(t)$  is the basis coefficient.

To some extent, a set of  $I$  neurons can be assimilated into a bank of  $I$  finite impulse response filters (FIR) which impulse response is equal to a set of  $I$  receptive fields  $\mathbf{w}$ . Depending on the number of neurons, the characteristics of the receptive field functions  $\mathbf{w}$ , and the constraints on the coefficients  $y$ , the analysis operation performed by a set of neurons can be equivalent to an independent component analysis (ICA) decomposition [27], a wavelet decomposition [33], or an overcomplete and sparse decomposition [39, 56].

- ICA [24, 27] constraints the output coefficients  $y_i(t)$  to be independent for a set of dependent input receptive fields  $\mathbf{w}$ . Therefore, a layer of postsynaptic neurons can uncouple the features coming from a presynaptic layer in which receptive fields are dependent. This is particularly efficient for hierarchical analysis systems. Applications in image processing commonly use ICA to find basis functions to decompose images into features [26].
- Wavelet decomposition [34] of signals is a well established field in signal processing. A layer of neurons can approximate a wavelet decomposition as long as the receptive fields  $\mathbf{w}$  respect some mathematical constraints such as limited support, bounded energy, and constraints on the Fourier transform of the receptive fields [17].
- Sparsity in the activity of neurons in the visual cortex can also be taken into account with equation 1. Sparsity in the activity means that very few neurons are active, that is only a few  $y_i$  are sufficiently high to generate a response in the neuron.  $\mathbf{w}$  is called the dictionary on which the presynaptic signals are projected on. To be sparse, the size of the dictionary should be sufficient, that is, the number of neurons is sufficiently high so that the basis is overcomplete.



Equation 1 is a simple model of a layer of neurons (at least for the receptive fields). Bank of neurons can be combined into hierarchical layers of filters that follow equation 1. Different combinations are possible. Depending on the architecture and on the connectivity, different hierarchical structures can be obtained.

We introduce in the following subsection 3 types of such structures.

### 2.2.2. Hierarchical neural network structures

Previous signal processing modules can be combined into hierarchical structures of neural networks. We list below some of the known models.

- Convolutional neural networks like the one proposed by Lecun et al. [30] have a mostly feedforward architecture;
- Deep learning machines [22] are layered neural networks that can be trained layer by layer. These layers are placed in a hierarchy to solve complex problems like image or speech recognition. First layers extract features and last layers perform a kind of classification by filtering the activity of the incoming layers. Feedbacks are established between layers.
- An example of a Bayesian neural network is given by Hawkins et al. [16, 20]. The connectivity between neurons inside a layer reflects partially the architecture of the cortex. Neurons are nodes of a statistical graphical model and feedbacks between layers are also established.

## 2.3. Summary

Simple models of the visual cortex have been presented. The present section has opened the way to a more integrated view in which peripheral and other senses interact with the visual cortex.

## 3. Biological models

This section reviews some models trying to reproduce as accurately as possible the biological visual system. The ultimate goal of those researches is having models able to represent the data we observe in the brain.

One of the earliest models is the Hodgkin & Huxley neuron model [23]. It is popular in the field of computational neuroscience and uses 4 differential equations to represent a biologically realistic neuron. Such models are important as they are one of the corner stones to the simulators that are used for a better understanding of the visual system. They can help us understand more precisely the different cortical mechanisms. Nowadays, computer simulations are very important in neuroscience and to the understanding of the visual cortex. Hundreds of computational models (along with their reference papers) are available to the computational neuroscientist. Many different models can be downloaded from the modelDB database [21]. Compartment, channel, spiking, continuous or discrete models cover a wide range of simulators and are available in different languages and simulators.

One of the biggest challenge when implementing biologically accurate models of the visual cortex is the amount of calculations required for the simulations. Even with massively parallel computer architectures, we cannot effectively rival the level of parallelism and computing power of the brain. As such, works claiming to represent the visual cortex in a biologically

accurate fashion have to cut corners by modeling only isolated sections of the cortex and sometimes also by simplifying the neuron model.

The section presents biologically accurate models, that is, models trying to fit the data observed in the brain. And then, some models are reviewed that are not exactly biologically accurate, but that are striving to reproduce biology.

### 3.1. Biologically accurate models

Since the primary visual cortex is the most studied and best understood, most models try to reproduce parts of the early visual system. In the work of McLaughlin et al. [36], a small local patch ( $1 \text{ mm}^2$ ) of V1 is reproduced. They use a relatively complex, and thus accurate, neuron model. The work focuses upon orientation preference and selectivity, and upon the spatial distribution of neuronal responses across the cortical layer. Furthermore, they discuss how large-scale scientific computation can provide significant understanding about the possible cortical mechanisms.

Similarly, Basalyga and Wennekers [4] are able to model a simplified version of the visual pathway of a cat's eye using three connected subsystems: the retina, the Thalamus and V1. They model a patch of about  $1.9 \text{ mm}^2$  of cortical surface using a Hodgkin-Huxley neuron model. They are able to reproduce the orientation preference and direction selectivity of cortical cells.

While biologically accurate neuron models are closer to reality, they are also computationally very expensive. By using a simpler, "integrate-and-fire" neuron model, Rangan et al. [47] are able to model a patch of  $25 \text{ mm}^2$ . They focus on reproducing V1 orientation preference maps with hypercolumns in a pinwheel organization. Models such as these give insights into the neuronal dynamics in V1.

### 3.2. Biologically realist models

Not all models are trying to directly reproduce biology. Most vision models use a more functional approach, trying to mimic the architecture and the behavior instead of the exact dynamics. Many distinct visual areas have been identified in the cerebral cortex, and functional organization of these areas has been proposed [62]. Such models do not generally use large-scale simulations, but are still trying to reproduce the biological architecture. As such, they are able to model a much larger part of the visual cortex.

A good example of such a model is based on the Adaptive Resonance Theory (ART) proposed by Grossberg. More specifically, the LAMINART model [19]. This model is a relatively complex system that is based on bottom-up, top-down and lateral interactions. The model integrates a "what" and a "where" stream, and as such, offers a very complete model of the visual cortex.

Another interesting work is a computational model of the human visual cortex based on the simulation of arrays of cortical columns [1]. The model suggests an architecture by which the brain can transform the external world into internal interpretable events. Furthermore, they argue that the model could be a good start for reverse engineering the brain. Modern super-computers have enough computational capacity to simulate enough cortical columns to achieve such a goal. However, the problem resides in the communication between the cortical columns. Communication between processing units in super-computer is a major



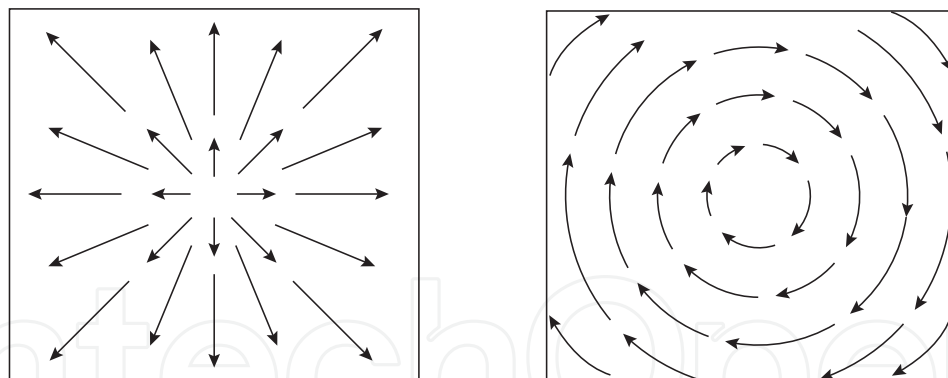
bottleneck. There is also the question of how the model should connect those cortical columns between them. How many must communicate with each other? How much information do they communicate? Nevertheless, reverse engineering of the brain is an objective that is becoming more and more realistic.

## 4. Visual localization and planning

In this section we present models that are related to the “where?” pathway of the visual cortex, that is, models having to do with movement and localization. However, another important function of our visual system is to anticipate trajectories of moving objects. In computer vision, this is commonly called target tracking. Many of these algorithms are based on probabilistic approaches which we will present in the first part of this section. In the second part, we will discuss the importance of vision in motion control, which uses the concepts of localization and planning to control actions.

### 4.1. Movement detection and prediction

Movement detection is a crucial element of our visual system. Moving objects attract more attention than stationary objects. Optical flow models implement the capture of movement in a sequence of images. Such models compute the local displacement gradient for each pixel in a sequence of images. The result is a series of gradient fields such as illustrated in figure 3. These methods can be used for a multitude of applications such as detecting motion, segmenting moving objects and analyzing stereo disparity to extract 3D information out of 2 cameras. Further information about optical flow models can be found in the work of Fleet and Weiss [14].



**Figure 3.** Example of optical flow gradient. In the left figure, a scene of forward movement is illustrated, thus, gradients are growing from the center toward the edges. In the right figure, the scene is rotating, thus forming a rotational pattern. In both scenes, the center of the image is moving slower than the rest, thus the arrows are shorter.

Optical flow models can detect movement, however, they are unable to achieve tracking in a reliable fashion because they lack anticipating behavior. The human brain is very good at predicting short term events. In fact, it is believed that the brain anticipates almost everything in our environment [20]. Indeed, brain activity is greater when we encounter an unanticipated event [2]. When we see a moving object, we automatically anticipate its trajectory. To do so, our brain unconsciously estimates the object speed and direction, and “runs” these number in a “mental physic model”, which is acquired through experience”. This anticipation allows

us to perform actions such as catching a ball, by placing our hands in the predicted trajectory estimated by our brain.

Models that use such anticipation principle have applications in target tracking. Most of these models are based on a method known as Recursive Bayesian Estimation. To put things simply, the algorithm consists of a loop that constantly tracks the movement of the target based on its current position and speed. The loop estimates and corrects the previous estimation using new data acquired from the moving object. Similar principles can also apply to visual control tasks. A quick overview of the mathematical framework can be found in the work of Diard et al. [10].

One of the most popular implementation of recursive Bayesian estimation is the Kalman filter. Many applications have used this algorithm to achieve target tracking. The model proposed by Bai [3] is a good example that achieves tracking of not only the target position in space, but also its size and rotation. Many different methods can be used for tracking. For instance, an approach based on a low pass filter has been shown to give better precision than traditional Kalman filters [7]. For tracking targets with more unreliable movement patterns, particle filters can be used [8]. These approaches are not making predefined assumptions about the target, and are thus more polyvalent, but generally slower. However, optimization techniques such as the one proposed by Zhou et al. [69] can be used for faster processing.

## 4.2. Visual control

The predictive behavior mentioned previously is used by the visual system for planning and controlling many of our actions. This section is divided in two parts. The first section present studies discussing the importance of vision in visual control models. The second section presents different models using visual stimuli for controlling actions such as pointing, grasping and locomotion.

### 4.2.1. Importance of vision in motion control

A study from Kawato [29] analyses the internal models for trajectory planning and motor control. It is suggested that humans or animals use the combination of a kinematic and dynamic model for trajectory planning and control. To catch a ball, one must anticipate the trajectory of the ball and position the hand and orient adequately the palm. Fajen and Cramer [13] study the positioning of the hand in function of distance, the angle and the speed. They discuss the implications for predictive models of catching based on visual perception. However, the targeted object is not the only thing tracked by our visual system. The movement of the hand itself needs to be anticipated. Saunders and Knill [53] study the visual feedback to control movements of the hand. Alternatively, Tucker and Ellis [59] are interested in the cortical area of visuomotor integration. They discuss on the fact that motor involvement impacts visual representations. They set up 5 experiments in which they study the impact of grasping and touching objects on the speed of visual perception.

The use of vision for motion control is thus a very complex process using both feedforward and feedback visual information to achieve movement.

### 4.2.2. Visual control models

Now that we know vision plays an important role in movement, we present some models using vision to achieve motion control. Fagg and Arbib [12] developed the FARS model for

grasping. They study the interaction between anterior intra-parietal and premotor areas. They also make predictions on neural activity patterns at population and single unit levels.

Moving on to models with more concrete applications, Yoshimi and Allen [65] propose an integrated view with a robotic application. They describe a real-time computer vision system with a gripper. A closed feedback loop control is used between the vision system and the gripper. Visual primitives are used to assist in the grasping and manipulation. Böhme and Heinke [5] implement the Selective Attention for Action model (SAAM) by taking into account the physical properties of the hand including anatomy. Their model is based on the fact that visual attention is guided by physical characteristics of objects (like the handle of a cup). Mehta and Schaal [37] study the visuomotor control of an unstable dynamic system (the balancing of a pole on a finger) with Smith predictors, Kalman filters, tapped-delay lines, and delay-uncompensated control. After validation with human participants, they exclude these models and propose the existence of a forward model in the sensory preprocessing loop.

## 5. Object recognition

### 5.1. Overview

In this section, we present models that are related to the ventral pathway of the visual cortex, that is, models that have to do with form and object recognition. For many years, object recognition has been one of the most challenging problems in the field of artificial intelligence. However, for humans, this task is something so simple that we do it unconsciously within a fraction of a second. For this reason, the visual cortex has been used as an inspiration in many artificial vision systems.

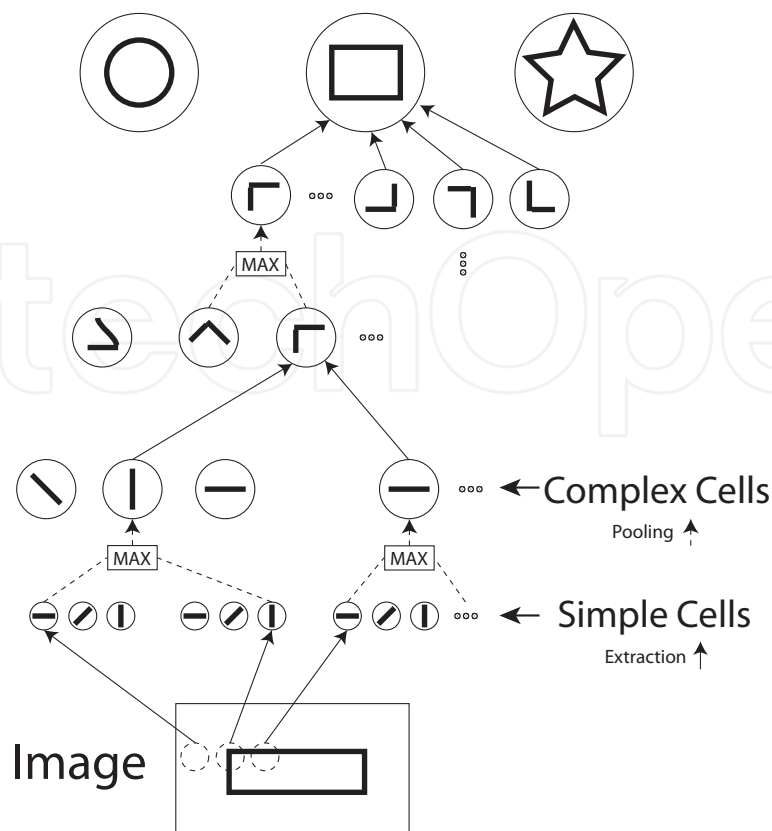
There are many ways to categorize the different object recognition models present in the literature. For the purpose of this chapter, we will discern two main categories: spiking and non-spiking models. The principal differences between those two approaches are that spiking models use bio-inspired spiking mechanism like STDP, synchrony and oscillations while non-spiking approaches tend to use statistical methods to achieve brain-like behavior. Spiking models are less popular in the literature because they tend to be computationally more expensive. However, they offer models that are more biologically plausible, and in specific cases more robust.

### 5.2. Non-spiking models

Since there are so many non spiking visual models in the literature, we further divide this section into three parts. We first review bottom-up models, which are strictly feed forward. We then review models incorporating top-down components, that is, models using feedback from previous or learned data to influence the input. We then take a quick look at models based on modeling visual attention.

#### 5.2.1. Bottom-up models

One of the earliest models of the visual cortex originated in the work of Hubel and Wiesel[25]. They described a hierarchical organization of the primary visual cortex where the lower level cells were responsive to visual patterns containing bars (edges-like). These “simple cells”, as they called them, are selective to bars of specific orientation, location and phase. Higher up



**Figure 4.** Example of a hierarchical model [51], consisting of multiple layers organized hierarchically. Each level in the hierarchy is composed of a layer of simple cells followed by a layer of complex cells. The simple cells are doing an extraction operation, combining features of the lower level to form more complex features. Complex cells are doing a pooling operation, selecting only the strongest (MAX) features for passage to the next level.

in the hierarchy, they found what they called “complex cells”, which are insensitive to phase and location in the visual field, but are still sensitive to oriented bars. This is explained in their model by the complex cells having a larger receptive field than simple cells, thus integrating outputs of many simple cells.

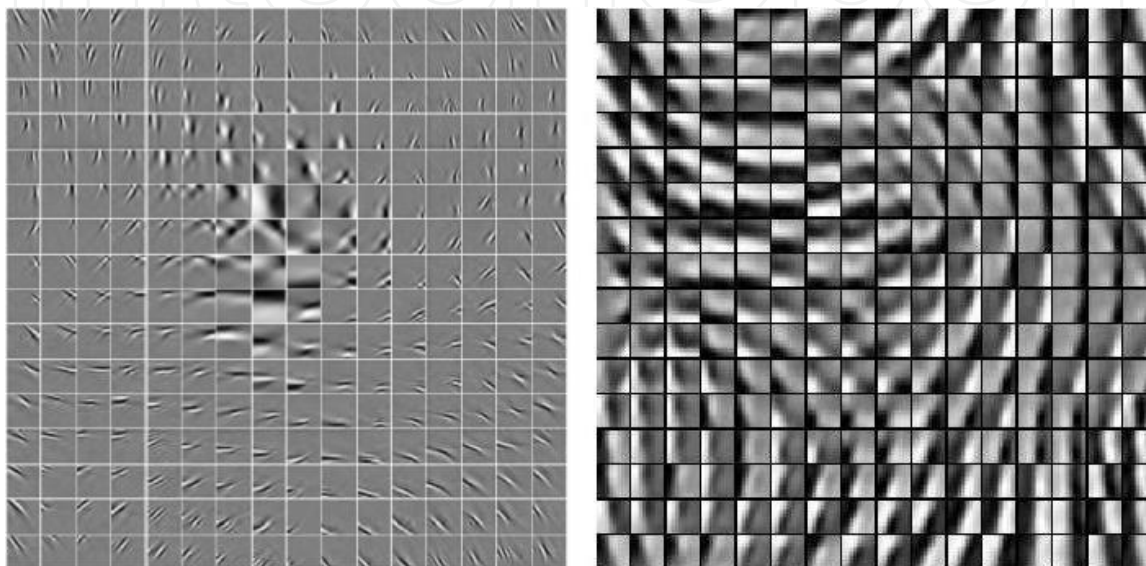
#### 5.2.1.1. Hierarchical models

Inspired by the hierarchical model of Hubel and Wiesel, convolutional neural networks were proposed. Good examples of such networks are LeNet-5[30] and HMAX[54]. As illustrated in figure 4, a typical hierarchical model is composed of multiple layers of simple and complex cells. As we go up in the hierarchy, lower level features are combined into more and more complex patterns.

Simple cells achieve feature extraction while complex cells realize a pooling operation. Feature extraction is done using the convolution between the input and the features, hence the name “convolutional neural networks”. The pooling operation can be achieved in many different ways, but the most popular is the one used in the HMAX model, that is, the MAX operation. The MAX operation selects the highest input on its receptive field, and thus only allows the strongest features to be propagated to the next level of the hierarchy.

### 5.2.1.2. Learning feature extractors

The biggest problem with convolutional networks is to find a way to learn good features to extract at each level. In our visual cortex, visual stimuli are perceived from birth and visual features are learned in function of what we see in our everyday life. In the primary visual cortex, these features happen to be, as discussed earlier, edge extractors. Interestingly enough, using natural image statistics, mathematical algorithms are able to reproduce this result. Figure 5 shows some of the features learned with different algorithms. Moreover, we can see that it is possible to learn these features in a topographical organization, just like cortical maps.



**Figure 5.** Visual features learned using image statistics. The features on the right are obtained using TICA (Topographic Independent Component Analysis) [26] and the left features are obtained using the IPSD algorithm [28]. In both cases, the features are learned in a pinwheel fashion, using unsupervised learning.

However, those features are only first level features, and learning good features for higher levels is a great challenge. Networks composed of multiple layers are typically referred to as “Deep Belief Networks”. Hinton et al. [22] were the first to be able to effectively learn multi-level features using his RBM (Restricted Boltzman Machine) approach. This learning method has later been extended to deep convolutional networks [48]. An approach combining RBM and a convolutional approach has also been proposed [31].

However, deep belief networks are not the only way to learn feature extractors. The SIFT features [32] are a very popular and for a long time have been considered state of the art. More recently, it has also been shown that the concepts behind SIFT features are biologically plausible [40].

### 5.2.2. Top-down models

As mentioned earlier in the planning section, the brain is very good at anticipating our environment. While this is obvious for applications such as target tracking, this behavior is also applied to object recognition. When we see an object, a top-down process based partly



on the context and the shape of the object helps the recognition process. This section presents some of the models trying to reproduce this prediction process.

One of the first models to include a top-down process involving prediction was an hierarchical neural network incorporating a Kalman Filter [49]. An improved version of this model has since been proposed [41], where sparse coding is used as a pretreatment. Another model using top-down prediction to influence inputs is Grossberg's ART model (Adaptive Resonance Theory)[18]. If complementarity is detected between ascendant (bottom-up) and descendent (top-down) data, the network will enter a resonance state, amplifying complementary data. Hawkins has proposed the HTM model (Hierarchical Temporal Memory) [20] [16], which is based on a hierarchical Bayesian network.

Another category of top-down models are the ones learning a compositional representation of objects. They learn object structures by combining objects parts, and can then use this representation in a top-down fashion to help recognition or fill in missing elements. One such approach is known as hierarchical recursive composition [70]. This method is able to form segments of objects and combine them in a hierarchical representation that can be used to segment and recognize objects. Another approach using a compositional method is based on a Bayesian network [45]. They demonstrate the power of their system by showing their ability to construct an image of an object using only their top-down process.

### 5.2.3. Visual attention models

Visual attention models try to reproduce the saccadic behavior of the eyes by analyzing specific regions of the image in sequence, instead of perceiving the whole image simultaneously. One of the first models of pattern recognition using the concept of visual attention has been proposed by Olshausen [44]. The visual attention process is driven by the interaction between a top-down process (memory) and bottom-up data coming from the input. In this section, we quickly review some of the most recent models for both bottom-up and top-down guidance of the visual glaze.

Bottom-up visual attention models mostly rely on visual saliency maps. Such maps use algorithms trying to reproduce the visual saliency properties observed in the biological visual cortex. Examples of such maps are given in figure . Many algorithms for computing such maps have been proposed. For a recent review of the most popular of these algorithms, refer to the work of Toet [58].

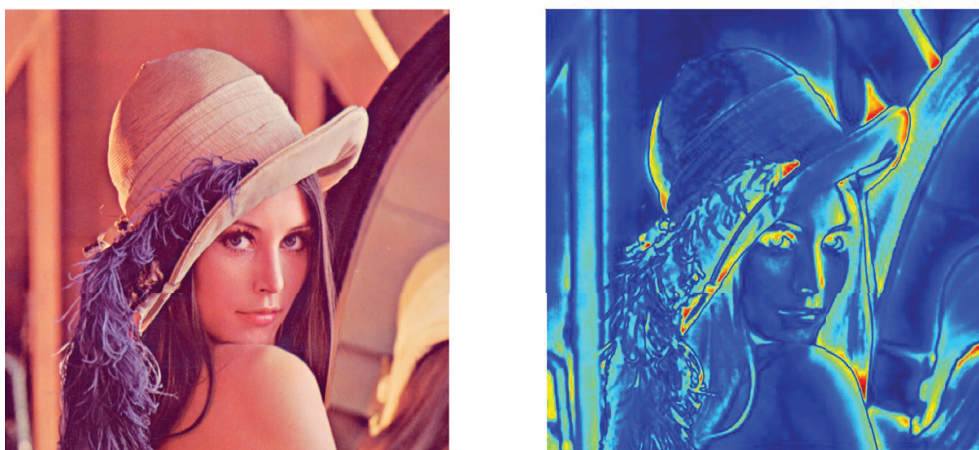
## 5.3. Spiking and synchronisation models

All the neural models presented until now in this section are using simpler neuron models that do not have a time component. They give a numerical value as output, that can be interpreted as a strong firing rate if the value is strong, or weak otherwise. However, they do not generate spikes as biological neurons do. By modeling spikes, we can model a phenomenon such as synchronization, oscillation and Hebbian learning.

### 5.3.1. Spiking models

This section presents general models using spiking neurons in a way that is not using synchrony or oscillation. Many models have been proposed to achieve spike-timing-





**Figure 6.** Example of saliency map. Using a JET color map, regions in red are considered to be more visually salient. The saliency was computed using the FastSUN algorithm [42].

dependent plasticity (STDP). Thorpe and Masquelier have proposed an implementation of HMAX using spiking neurons [35]. They show that such hierarchical models can be implemented and trained using STDP.

Another concept that can be implemented using spiking neurons is rank-order coding (ROC). In ROC, the activation order of neurons transports information. Neurons firing first represent more important information than neurons firing at a later time after presentation of stimuli. A recognition model based on this principle is the SpikeNet network [9].

Many spiking models are now proposed to achieve different vision tasks. For instance, a vision system achieving image recognition using spiking neurons [55] has been shown to be robust to rotation and occlusion.

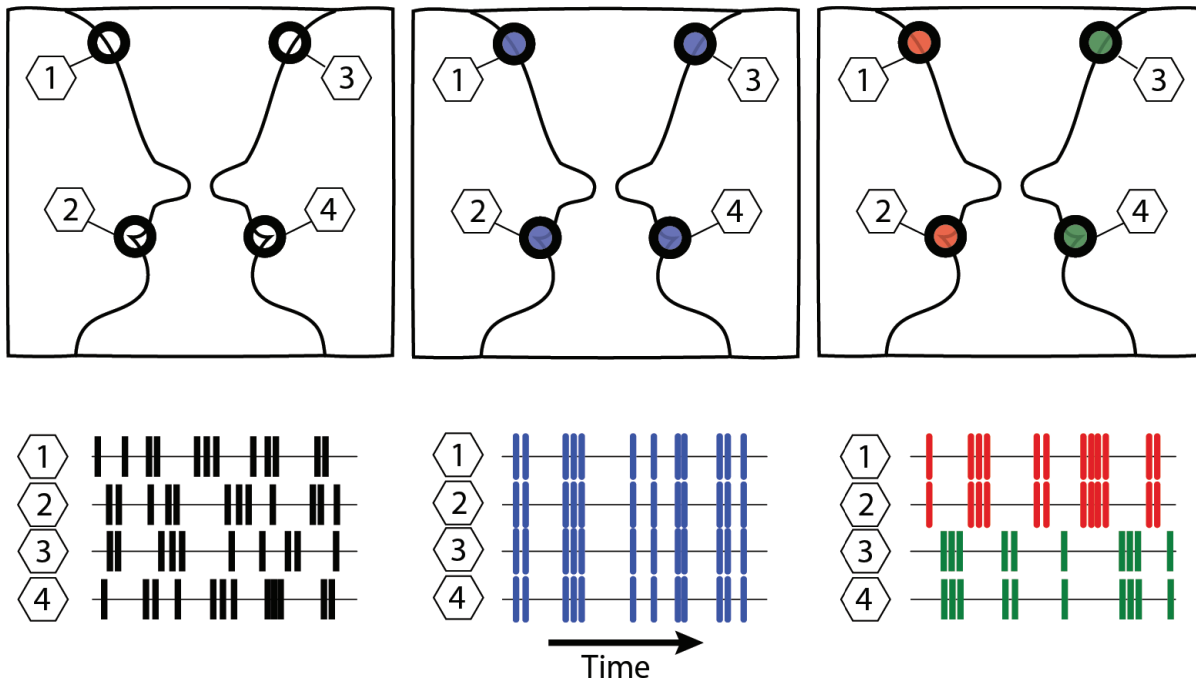
### 5.3.2. Synchrony models

Most of the vision models based on synchrony and oscillations are based on the binding problem and the binding-by-synchrony hypothesis. In this section, we first give a quick overview of what is the binding problem and we then describes some of the models implementing it.

#### 5.3.2.1. The binding problem

A fundamental dilemma in the brain, or in any recognition system, is the combinational problem. When a signal is perceived, it is broken down into a multitude of features, which then need to be recombined to achieve recognition. This has been defined by Malsburg [61] as the “Binding Problem”.

One solution proposed to solve the binding problem is the binding-by-synchrony hypothesis. According to Milner [38] and van der Malsburg [61], groups of neurons in the brain can be bound by synchrony when these groups belong to a same external entity. The validity of this hypothesis is however highly debated in the literature. The binding problem itself is highly controversial [50]. In fact, no real evidence has yet been found regarding the synchronization



**Figure 7.** The binding problem [11]. In the leftmost column, we have an image from which 4 features are responding. Each feature is represented by a neuron spiking through time in the lower part of the figure. The binding problem consists in grouping these features together to recognize objects. The binding-by-synchrony offers a solution to this problem by using rhythmic oscillations. Features belonging to a same object will oscillate in synchrony, as shown in the middle column, where all the features combine to form a vase. In the rightmost column, with the same 4 features responding, 2 distinct groups of oscillation occur, thus forming 2 faces looking at each other.

of groups of neurons encoding an extended object [60]. Most studies however agree that synchrony and oscillation plays a role in the functioning of the visual cortex. Mechanisms such as binding-by-synchrony and temporal binding could very well be occurring in higher areas of the brain which are harder to study and understand [60].

#### 5.3.2.2. Binding-by-synchrony models

One of the most popular models based on temporal binding is the LEGION network[63]. It was first developed to provide a neurocomputational foundation to the temporal binding theory. There are many applications for legion in scene analysis, particularly for image segmentation[67]. Pichevar and Rouat then proposed the ODLM network [46], which extends the LEGION model. ODLM is able to perform binding and matching in signal processing applications that are insensitive to affine transforms. Therefore, patterns can be recognized independently of their size, position and rotations. They have shown how binding can be used to segment and/or compare images based on pixel-values, and also to achieve sound source separation.

Slotine and his group have also proposed a model based on synchronization [66] that achieves visual grouping of orientation features. However, they do not use real images, but only orientation features. The only binding model that reproduces the vision process from real images is the Maplet model [52] from Malsburg's lab. Their model is computationally quite intensive, but is able to achieve face recognition.

## 6. Conclusion

This chapter presented different types of models of the visual cortex. We first discussed the idea that the visual cortex is in fact only a part of a more complex system, cooperating with other sensorial areas. However, because models are mainly used for simplifying reality, we then focused on vision specific models. We first presented models closely related to biology, with models directly trying to reproduce biological data, and others trying to reproduce the global architecture of the visual cortex. We then presented models related to movement detection and planning. Anticipation being an important mechanism in the brain we presented models using vision for planning and controlling movements. Finally, we glanced through vision models related to form and object recognition. We looked at feedforward and top-down models, as well as models based on neural synchrony and oscillations. Since all these models are very different, there is no consensus on how the visual cortex should be modeled. Also, all models do not have the same goals. The best possible model for a given situation is the simplest model that allows reaching the targeted goals.

We have only given a small overview of all the models of vision reported in the literature. This chapter was only intended to have a quick peek into the world of visual cortex models. We also have to keep in mind that our current understanding of the brain is far from complete. What we are doing right now is akin to trying to understand the mechanics of the ocean by studying one drop of water at a time. Considering the rapid evolution of knowledge about the brain, it is quite possible that some of the currently available models may become obsolete. But this is the basic characteristic of any scientific endeavor.

## Author details

Vincent de Ladurantaye, Jean Rouat and Jacques Vanden-Abeeel  
 NECOTIS, Université de Sherbrooke, Québec, Canada

## 7. References

- [1] Albus, J. [2008]. A computational model of the human visual cortex, *Proceedings of the SPIE - The International Society for Optical Engineering*, Vol. 6943, pp. 69430 – 1.
- [2] Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W. & Muckli, L. [2010]. Stimulus predictability reduces responses in primary visual cortex., *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(8): 2960–2966.
- [3] Bai, K.-J. [2010]. A new object tracking algorithm based on mean shift in 4-d state space and on-line feature selection, *Information and Computing (ICIC), 2010 Third International Conference on*, Vol. 1, pp. 39 –42.
- [4] Basalyga, G. & Wennekers, T. [2009]. Large-scale computational model of cat primary visual cortex, *BMC Neuroscience* 10: 358.
- [5] Böhme, C. & Heinke, D. [2009]. Modeling visual affordances: The selective attention for action model (saam), *Connectionist Models of Behaviour and Cognition II Proceedings of the 11th Neural Computation and Psychology Workshop* (April): 39–50.
- [6] Calvert, G., Spence, C. & Stein, B. E. [2004]. *The handbook of multisensory processes*, MIT Press.

- [7] Chen, K., Li, D. & Jhun, C. G. [2010]. Video motion predictive tracking quality: Kalman filter vs. lowpass filter, *2010 International Conference on Multimedia Technology (ICMT)*, pp. 1–4.
- [8] Czyz, J. [2006]. Object detection in video via particle filters, *Proceedings - International Conference on Pattern Recognition*, Vol. 1, Hong Kong, China, pp. 820 – 823.
- [9] Delorme, A. & Thorpe, S. [2001]. Face identification using one spike per neuron: resistance to image degradations, *Neural Networks* 14(6-7): 795 – 803.
- [10] Diard, J., Bessière, P. & Mazer, E. [2003]. A survey of probabilistic models, using the bayesian programming methodology as a unifying framework, *In The Second International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2003)*.
- [11] Engel, A. K., Fries, P. & Singer, W. [2001]. Dynamic predictions: oscillations and synchrony in top-down processing., *Nature Reviews Neuroscience* 2(10): 704–716.
- [12] Fagg, A. & Arbib, M. [1998]. Modeling parietal-premotor interactions in primate control of grasping., *Neural Networks* 11: 1277 – 1303.
- [13] Fajen, B. & Cramer, C. [2007]. Visual control of hand position and orientation during one-handed catching, *Journal of Vision* 7(9): 165–165.
- [14] Fleet, D. & Weiss, Y. [2005]. *Optical Flow Estimation*, Vol. 19, Springer, pp. 1–10.
- [15] Frigg, R. & Hartmann, S. [2012]. Models in science, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, spring 2012 edn.
- [16] George, D. & Hawkins, J. [2005]. A hierarchical bayesian model of invariant pattern recognition in the visual cortex, *In Proceedings of the International Joint Conference on Neural Networks. IEEE*, pp. 1812–1817.
- [17] Goupillaud, P., Grossmann, A. & Morlet, J. [1984]. Cycle-Octave and related transforms in seismic signal analysis, *Geoexploration* 23: 85–102.
- [18] Grossberg, S. [1999]. The link between brain learning, attention, and consciousness, *Consciousness and Cognition* 8(1): 1 – 44.
- [19] Grossberg, S., for Adaptive Systems, B. U. C., of Cognitive, B. U. D. & Systems, N. [2006]. *Towards a Unified Theory of Neocortex: Laminar Cortical Circuits for Vision and Cognition*, Technical report CAS/CNS, Boston University, Center for Adaptive Systems and Department of Cognitive and Neural Systems.
- [20] Hawkins, J. & Blakeslee, S. [2004]. *On Intelligence*, Times Books.
- [21] Hines, M., Morse, T., Migliore, M., Carnevale, N. & Shepherd, G. [2004]. Modeldb: A database to support computational neuroscience, *Journal of Computational Neuroscience* 17(1): 7–11.
- [22] Hinton, G. E., Osindero, S. & Teh, Y.-W. [2006]. A fast learning algorithm for deep belief nets, *Neural Comp.* 18(7): 1527–1554.
- [23] Hodgkin & Huxley [1952]. A quantitative description of membrane current and its application to conduction and excitation in nerve, *The Journal of Physiology* 117(4): 500–544.
- [24] Hoyer, P. O. & Hyvärinen, A. [2000]. Independent component analysis applied to feature extraction from colour and stereo images, *Network: Computation in Neural Systems* 11(3): 191–210.
- [25] Hubel, D. H. & Wiesel, T. N. [1965]. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat, *J Neurophysiol* 28(2): 229–289.



- [26] Hyvärinen, A., Hurri, J. & Hoyer, P. [2009]. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.*, Springer.
- [27] Hyvärinen, Karhunen & Oja [2001]. *Independent Component Analysis*, Wiley.
- [28] Kavukcuoglu, K., Ranzato, M., Fergus, R. & LeCun, Y. [2009]. Learning invariant features through topographic filter maps, *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR'09)*, IEEE.
- [29] Kawato, M. [1999]. Internal models for motor control and trajectory planning., *Current Opinion in Neurobiology* 9: 718–727.
- [30] Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. [1998]. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, pp. 2278–2324.
- [31] Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. [2009]. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, *Proceedings of the Twentieth-Sixth International Conference on Machine Learning* .
- [32] Lowe, D. G. [2004]. Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60: 91–110.
- [33] Mallat, S. [1996]. Wavelets for a vision, *Proceedings of the IEEE* 84(4): 604–614.
- [34] Mallat, S. [1997]. *A wavelet tour of signal processing*, ACADEMIC PRESS.
- [35] Masquelier, T. & Thorpe, S. J. [2007]. Unsupervised learning of visual features through spike timing dependent plasticity, *PLoS Computational Biology* 3(2): e31.
- [36] McLaughlin, D., Shapley, R. & Shelley, M. [2003]. Large-scale modeling of the primary visual cortex: influence of cortical architecture upon neuronal response., *Journal Of Physiology Paris* 97(2-3): 237–252.
- [37] Mehta, B. & Schaal, S. [2002]. Forward models in visuomotor control, *Journal of Neurophysiology* 88(2): 942–953.
- [38] Milner, P. [1974]. A model for visual shape recognition, *Psychological Review* 81: 521–535.
- [39] Molotchnikoff, S. & Rouat, J. [2012]. Brain at work: Time, sparseness and superposition principles, *Frontiers in Bioscience* 17(1): 583–606.
- [40] Muralidharan, K. & Vasconcelos, N. [2010]. A biologically plausible network for the computation of orientation dominance, in J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel & A. Culotta (eds), *NIPS*, pp. 1723–1731.
- [41] Murray, J. F. & Kreutz-Delgado, K. [2007]. Visual Recognition and Inference Using Dynamic Overcomplete Sparse Learning, *Neural Comp.* 19(9): 2301–2352.
- [42] Nicholas J. Butko, Lingyun Zhang, G. W. C. & Movellan, J. R. [2008]. Visual salience model for robot cameras, *Proceedings of the 2008 IEEE International Conference on Robotics and Automation (ICRA)*, p. 2398–2403.
- [43] Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. [2011]. A normalization model of multisensory integration, *Nat Neurosci* 14(6): 775–782.
- [44] Olshausen, B., Anderson, C. & Essen, D. V. [1993]. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *J. Neurosci.* 13(11): 4700–4719.
- [45] Ommer, B. & Buhmann, J. [2010]. Learning the compositional nature of visual object categories for recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 32: 501–516.
- [46] Pichevar, R., Rouat, J. & Tai, L. [2006]. The oscillatory dynamic link matcher for spiking-neuron-based pattern recognition, *Neurocomputing* 69(16-18): 1837–1849.

- [47] Rangan, A., Tao, L., Kovacic, G. & Cai, D. [2009]. Multiscale modeling of the primary visual cortex, *Engineering in Medicine and Biology Magazine, IEEE* 28(3): 19–24.
- [48] Ranzato, M., Poultney, C., Chopra, S. & Lecun, Y. [2006]. Efficient learning of sparse representations with an energy-based model, *Advances in Neural Information Processing Systems (NIPS)*.
- [49] Rao, R. P. N. & Ballard, D. H. [1997]. Dynamic Model of Visual Recognition Predicts Neural Response Properties in the Visual Cortex, *Neural Computation* 9(4): 721–763.
- [50] Riesenhuber, M. & Poggio, T. [1999a]. Are cortical models really bound by the "binding problem"? *Neuron* 24(1).
- [51] Riesenhuber, M. & Poggio, T. [1999b]. Hierarchical models of object recognition in cortex, *Nature neuroscience* 2(11): 1019–1025.
- [52] Sato, Y. D., Jitsev, J. & Malsburg, C. [2008]. A visual object recognition system invariant to scale and rotation, *Proceedings of the 18th international conference on Artificial Neural Networks, Part I*, Springer-Verlag, pp. 991–1000.
- [53] Saunders, J. & Knill, D. [2004]. Visual feedback control of hand movements, *The Journal of Neuroscience* 24(13): 3223–3234.
- [54] Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. & Poggio, T. [2007]. Robust object recognition with cortex-like mechanisms, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29(3): 411–426.
- [55] Shin, J.-H., Smith, D., Swiercz, W., Staley, K., Rickard, J., Montero, J., Kurgan, L. & Cios, K. [2010]. Recognition of partially occluded and rotated images with a network of spiking neurons, *Neural Networks, IEEE Transactions on* 21(11): 1697–1709.
- [56] Starck, J.-L., Murtagh, F. & Fadili, J. M. [2010]. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press.
- [57] Striem-Amit, E., Dakwar, O., Reich, L. & Amedi, A. [2011]. The large-scale organization of "visual" streams emerges without visual experience, *Cerebral Cortex*.
- [58] Toet, A. [2011]. Computational versus Psychophysical Image Saliency: A Comparative Evaluation Study, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 99(Preliminary).
- [59] Tucker, M. & Ellis, R. [2001]. The potentiation of grasp types during visual object categorization, *Visual Cognition* 8(6): 769–800.
- [60] Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolic, D. & Singer, W. [2009]. Neural synchrony in cortical networks: history, concept and current status, *Frontiers in Integrative Neuroscience* 3(00017).
- [61] v. d. Malsburg, C. [1981]. The correlation theory of brain function, *Technical Report Internal Report 81-2*, Max-Planck Institute for Biophysical Chemistry.
- [62] Van Essen, D. C. & Maunsell, J. H. [1983]. Hierarchical organization and functional streams in the visual cortex, *Trends in Neurosciences* 6(0): 370–375.
- [63] Wang, D. & Terman, D. [1995]. Locally excitatory globally inhibitory oscillator networks., *IEEE Transactions on Neural Networks* 6(1): 283–286.
- [64] Wysocki, S. G., Benuskova, L. & Kasabov, N. [2010]. Evolving spiking neural networks for audiovisual information processing, *Neural Networks* 23(7): 819–835.
- [65] Yoshimi, B. & Allen, P. [1998]. The confluence of vision and control, in D. Kriegman, G. Hager & A. Morse (eds), *Lecture Notes in Control and Information Sciences*, Vol. 237, Springer Berlin / Heidelberg, pp. 195–209.



- [66] Yu, G. & Slotine, J. [2009]. Visual grouping by oscillator networks, *I.E.E.E. Transactions on Neural Networks* 49: 4528–4538.
- [67] Yuan, J., Wang, D., Wu, B., Yan, L. & Li, R. [2009]. Legion-based automatic road extraction from satellite imagery, *IEEE Transactions on Geoscience and Remote Sensing* 49: 4528–4538.
- [68] Zangaladze, A., Epstein, C. M., Grafton, S. T. & Sathian, K. [1999]. Involvement of visual cortex in tactile discrimination of orientation, *Nature* 401(6753): 587–590.
- [69] Zhou, H., Zhang, X., Li, H. & Li, J. [2010]. Video object tracking based on swarm optimized particle filter, *Industrial Mechatronics and Automation (ICIMA), 2010 2nd International Conference on*, Vol. 2, pp. 702–706.
- [70] Zhu, L. L., Lin, C., Huang, H., Chen, Y. & Yuille, A. L. [2008]. Unsupervised structure learning: Hierarchical recursive composition, suspicious coincidence and competitive exclusion, *ECCV (2)*, Vol. 5303 of *Lecture Notes in Computer Science*, Springer, pp. 759–773.