This is to certify that the
thesis entitled

DEVELOPMENTAL STEREO: EMERGENCE OF DISPARITY PREFERENCE IN COMPUTATIONAL MODELS OF VISUAL CORTEX

presented by

Mojtaba Solgi

has been accepted towards fulfillment
of the requirements for the

Master of Science degree in Computer Science

Major Professor's Signature

5/8/2009

Date

Developmental Stereo: Emergence of Disparity Preference in Computational Models of Visual Cortex

By

Mojtaba Solgi

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Computer Science

2009

Abstract

Developmental Stereo: Emergence of Disparity Preference in Computational Models of Visual Cortex

By

Mojtaba Solgi

One of the major tasks carried out in our visual system is to create three dimensional representation of the visual world using the two-dimensional images reflected on the retinas. How do we estimate the relative depth of the objects in the visual field? It is known that stereoscopic cues for binocular disparity are one of the major ways for the brain to perceive three-dimensional objects. However, there is much unknown about how this complicated process takes place in the brain. This thesis proposes computational models to study the role of 6-layer architecture of the laminar cortex to detect the slight differences in the images on the left and right retina, *disparity*. Assuming the spatial continuity of the visual stimuli, we investigate how top-down signals can be used as temporal context information to guide recognition during the testing phase. The experimental results indicate that the use of top-down efferent signals in the form of supervision or temporal context signals not only greatly improves the performance of the networks, but also results in biologically compatible cortical maps – the representation of disparity selectivity is grouped, and changes gradually along the cortex. To our knowledge, this work is the first neuromorphic, end-to-end model of laminar cortex that integrates temporal context to develop internal representation, and generates accurate motor actions in the challenging problem of detecting disparity in natural images. The network reaches a sub-pixel error rate in the case of regression, and 0.90 recognition rate in the case of classification, given limited resources.

Dedication

I dedicate this thesis to my parents who have offered me unconditional love and care throughout the course of my life.

ACKNOWLEDGEMENTS

I would like to thank my adviser, Professor John Weng, for his insightful guidance throughout this thesis. My special gratitude goes to Dr. Danil Prokhorov from Toyota Research Institute for research assistantship support during the past two semesters. I thank my labmates Matthew Luciw for his helpful discussions, and Paul Cornwell for his generosity whenever I needed help. Last but not least, I would like to sincerely thank my roommate and dear friend, Rouhollah Jafari, for his true friendship and care over the past two years.

Keep away from people who try to belittle your ambitions. Small people always do that, but the really great make you feel that you, too, can become great.

Mark Twain

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

Humans and many other animals posses two eyes via which they perceive the visual world. Because the two eyes are placed in horizontally different positions in the skull, they receive two slightly different images of the visual scenes. This difference is referred to as *disparity*. Psychophysical studies indicate that disparity is one of the main cues for the emergence of three-dimensional representation of the world from two-dimensional retinal images [35].

Intensive amount of studies in computer vision community during the past few decades has proved that the challenges of stereo vision cannot be addressed without a thorough understanding of the biological visual systems. Recent studies of binocular depth perception in the physiological level has shed light onto many aspects of the role of stereoscopic cues in the perception of depth. However, there are much more unknown for a unified theory of how the actual mechanism takes place. The important role of computational models toward such theory should never be underestimated. It is via computational models that researchers can verify their theories based on experimental observations, and also predict the details of some mechanisms before any data is available for it. Depending on the matter of study, the proposed

models must be as biologically plausible as computaional tools allow, otherwise one cannot imply the biological analogy of the results.

One such computational model is MILN (Multilayer In-place Learning Networks), a cortex inspired learning network architecture that operates using LCA (Lobe Componenet Analysis). By in-place learning, we mean each neuron in the network learns on its own and by interacting with other neurons, without the need of any external controller. Lobe Component Analysis is a dual-optimal learning algorithm that atonomously derives representaion from input samples.

As an extention to the original MILN networks, we implemented a model of the 6-layer architecture of the laminar cortex within the same architecture. The main goal of the project was to investigate the mechanisms of top-down connections as supervision or context signals in the cortical architectures to the emergence of disparity preference in the modeled neurons.

1.2 Task Decomposition

The project was carried out in two main phases. In the first phase we utilized the default version of MILN, and investigated its abilities to detect disparities in a challenging setting of natural images. A new implementation of MILN was developed in C++ from scratch. The necessary modules were added to the basic MILN to handle binocular disparity data. After the preliminary study in the first phase, it was evident that MILN has the capability to operate on binocular data. The second phase involved designing a novel architecture of the newtworks to handle top-down context signals during testing. A graphical user interface was developed to visualize the internal states and operations of the network. Final results were convincing that the new architecture successfully utilizes context information to elevate the recognition abilities of the network, and demonstrates biologically plausible cortical maps.

1.3 Thesis Outline

The outline of the remainder of this thesis is as follows:

- *Chapter 2* introduces the fundamentals of biological visual systems required for understanding the biological terminology used in the later chapters. It also briefly presents an overview of the previous computational models of binocular disparity encoding and disparity detection.

- *Chapter 3* provides an overview of the specific problems addressed in this thesis.

- *Chapter 4* presents the structure of the different types of the network used in the thesis, along with the learning algorithms.

- *Chapter 5* analytically explains the mechanisms of the methods used, and provides reasons as to why we should expect such beahavior and outputs from the networks.

- *Chapter 6* presents the experiments done in this thesis along with the results obtained.

- *Chapter 7* concludes the thesis, and provides some predictions about the functionality of cortical regions.

Chapter 2

Background

This chapter presents the fundamentals of neurological knowledge required for understanding the biological binocular vision systems regarding disparity encoding and detection. Furthermore, the details of LCA (Lobe Component Analysis) and MILN (Multilayer In-place Learning Networks) are discussed and compared with other models of visual neural networks. At the end of the chapter, related works on disparity models are presented. Most material on biological visual systems is adapted from Kandel 2000 [24] and Ramtohul 2006 [39], and those about LCA and MILN are largely adapted from Weng & Luciw 2009 [49].

2.1 Basics of Human Visual System

The human visual system is one of the most remarkable biological systems in nature, formed and improved by millions of years of evolution. About the half of the human cerebral cortex is involved with vision, which indicates the computational complexity of the task. Neural pathways starting from the retina and continuing to V1 and the higher cortical areas form a complicated system that interprets the visible light projected on the retina to build a three dimensional representation of the world. In this chapter we provide background information about the human visual system

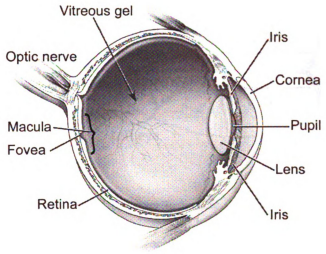


Figure 2.1: Anatomy of the human eye (reprinted from [33])

and the neural mechanisms involved during the development and operation of visual capabilities.

2.1.1 Eye

When visible light reaches the eye, it first gets refracted by the cornea. After passing through the cornea, it reaches the pupil. To control the amount of light entering the eye, the pupils size is regulated by the dilation and constriction of the iris muscles. Then the light goes through the lens, which focuses it onto the retina by proper adjustment of its shape.

2.1.2 Visual Pathway

The early visual processing involves the retina, the lateral geniculate nucleus of thalamus (LGN), and the primary visual cortex (V1). The visual signals then go through the higher visual areas, which include V2, V3, V4 and V5/MT. After initial

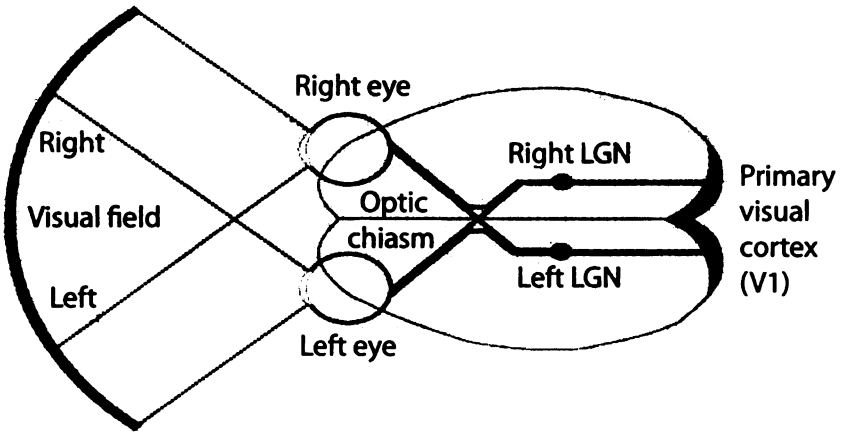


Figure 2.2: Visual pathway in human (reprinted from [31])

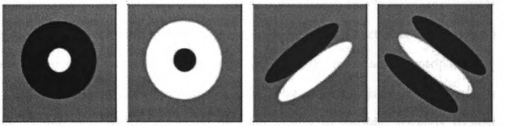
processing in the retina, output from each eye goes to LGN, at the base of the same side of the brain. LGN in turn does some processing on the signals and projects to the V1 of the opposite side of the brain. The optic nerves, going to opposite sides of the brain, cross at a region called the *optic chiasm*. V1 then feeds its output to higher visual cortices where further processing takes place. Fig. 2.2 presents a schematic overview of the visual pathway.

2.1.3 Retina

The retina is placed on the back surface of the eye ball. There is an array of special purpose cells on the retina, such as photoreceptors, that are responsible for converting the incident light into neural signals.

There are two types of light receptors on the retina: 1) rods that are responsible for vision in dim light 2) cones that are responsible for vision in bright light. The total number of rods is more than cones, however there are no rod cells in the center of retina. The central part of the retina is called the *fovea* which is the center of fixation. The density of the cone cells is high in the fovea, which enables this area to detect the fine details of retinal images.

For the first time, Stephen Kuffler recorded the responses of retinal ganglion cells



(a) ON cell in retina or LGN | **(b)** OFF cell in retina or LGN | **(c)** 2-lobe V1 simple cell | **(d)** 3-lobe V1 simple cell

Figure 2.3: Samples of the receptive fields shapes in human V1 (reprinted from [31])

to rays of light in a cat in 1953(Hubel, 1995). He discovered that it is possible to influence the firing rate of a retinal ganglion cell by projecting a ray of light to a specific spot on retina. This spot is called the *receptive field* (RF) of the cell. Below is a definition of receptive field from Livine & Shefner 1991:

"Area in which stimulation leads to a response of a particular sensory neuron"

In other words, for any neuron involved in the visual pathway, the receptive field is a part of the visual stimuli that influences the firing rate of the specific neuron. Fig. 2.3 shows a few examples of the shape of receptive fields in the visual pathway.

2.1.4 LGN

The LGN acts like a relay that gets signals from the retina and projects to the primary visual cortex (V1). It consists of neurons similar to retinal ganglion cells, however the role of these cells is not clear yet. The arrangement of the LGN neurons is *retinotopic*, meaning that the adjacent neurons have gradually changing, overlapping receptive fields. This phenomena is also called *topographic representation*. It is believed that the LGN cells perform edge detection on the input signals they receive from the retina.

2.1.5 Primary Visual Cortex

Located at the back side of the brain, the primary visual cortex is the first cortical area in the visual pathways. Similar to LGN, V1 neurons are reinotopic too. V1 is the lowest level of the visual system hierarchy in which there are *binocular* neurons. These neurons are identified by their ability to respond strongly to stimuli from either eye. These neurons also exhibit preference to specific features of the visual stimuli such as spatial frequency, orientation and direction of motion. It has been observed that some neurons in V1 show preference for particular disparities in binocular stimuli - stimuli with a certain disparity causes potential discharge in the neuron. V1 surface consists of columnar architecture where neurons in each column have more or less similar feature preference. In the columnar structure, feature preference changes smoothly across the cortex, meaning that nearby columns exhibit similar and overlapping feature preference while columns far from each other respond differently to the same stimuli. Overall, there is a smoothly varying map for each feature in which preferences repeat at regular intervals in any direction. Examples of such topographic maps include orientation maps, and disparity maps which are the subject of study in this thesis.

2.1.6 Disparity

It is known that the perception of depth arises from many different visual cues (Qian 1997 [37]) such as occlusion, relative size, motion parallax, perspective, shading, blur, and relative motion (DeAngelis 2000 [11], Gonzalez & Perez 1998 [18]). The cues mentioned were monocular. There are also binocular cues because of the stereo property of the human vision. Binocular disparity is one of the strongest binocular cues for the perception of depth. The existence of disparity is because the two eyes are laterally separated. The terms *stereo vision*, *binocular vision* and *stereospsis* are interchangeably used for the three-dimensional vision based on binocular disparity.

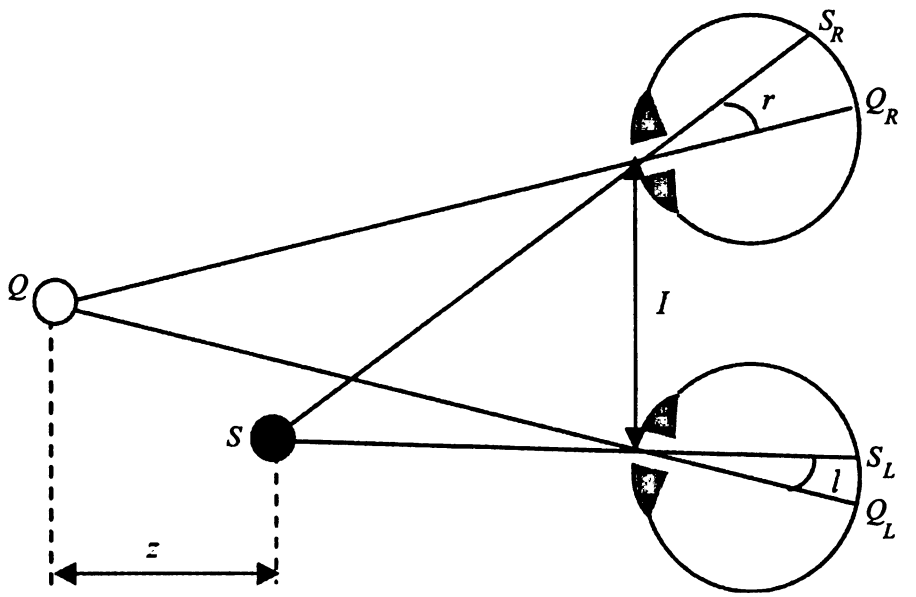


Figure 2.4: The geometry of stereospsis (reprinted from [40])

2.1.7 Geometry of Binocular Vision

Fig. 2.4 illustrates the geometry of the stereo vision. Suppose that the eyes are focused(fixated) at the point Q. The images of the fixation point falls on the fovea, Q_L and Q_R on the left and right eyes, respectively. These two points are called *corresponding* points on the retina, since they both get the reflection of the same area of the visual field (fixation point in this example). The filled circle S is closer to the eyes and its image reflects on different spots on the two retinas, which are called *non-corresponding* points. This lack of correspondence is referred to as disparity. The relative depth of the point S, distance z from the fixation point, can be easily calculated given the retinal disparity $\delta = r - l$, and the interocular distance (the distance between the two eyes), I. Since this kind of disparity is caused by the location of the objects on the horizontal plane, it is known as *horizontal disparity.*

It can be proven that all the points that are at the same disparity as the fixation point lie on a semi-sphere in the three-dimensional space. This semi-sphere is referred to as the *horopter.* Points on the horopter, inside and outside of the horopter have

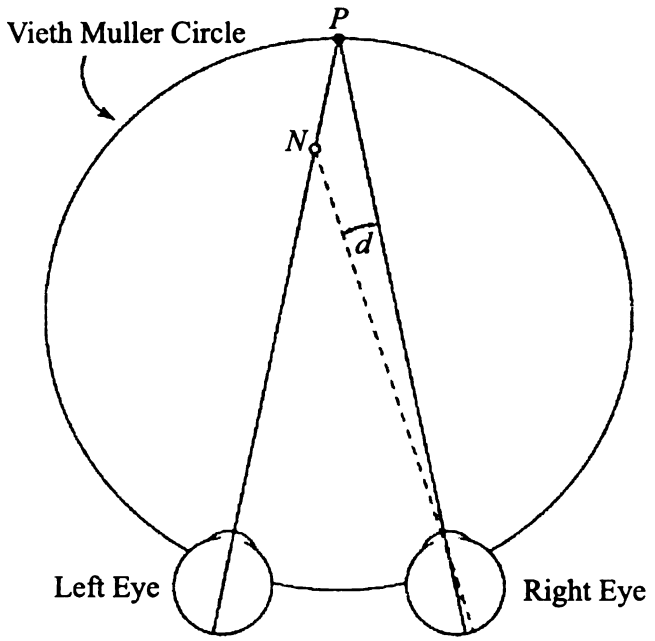


Figure 2.5: Horizontal Disparity and the Vieth-Muller circle(reprinted from [11])

zero, negative and positive disparities, respectively. The projection of the horopter on the horizontal plane crossing the eyes (at the eyes level) is the *Vieth-Muller* circle.

It is known that another type of disparity, called *vertical disparity*, plays some role in the perception of depth, however, it has not been studied as intensively as horizontal disparity. The vertical disparity occurs when an object is considerably closer to one eye than the other. According to Bishop 1989 [3], such vertical disparities occur when objects are located relatively close to eyes and are above or below the horizontal visual plane, but do not reside on the median plane, the vertical plane that divides the human body into left and right halves. Fig. 2.6 simply illustrates vertical disparity. Point P is above the visual plane and to the right of the median plane, which makes it closer to the right eye. It can be seen that the relation $\beta_2 > \beta_1$ holds between two angles β_1 and β_2. The vertical disparity, denoted by v, is the difference between these two angles, $v = \beta_2 - \beta_1$ [3].

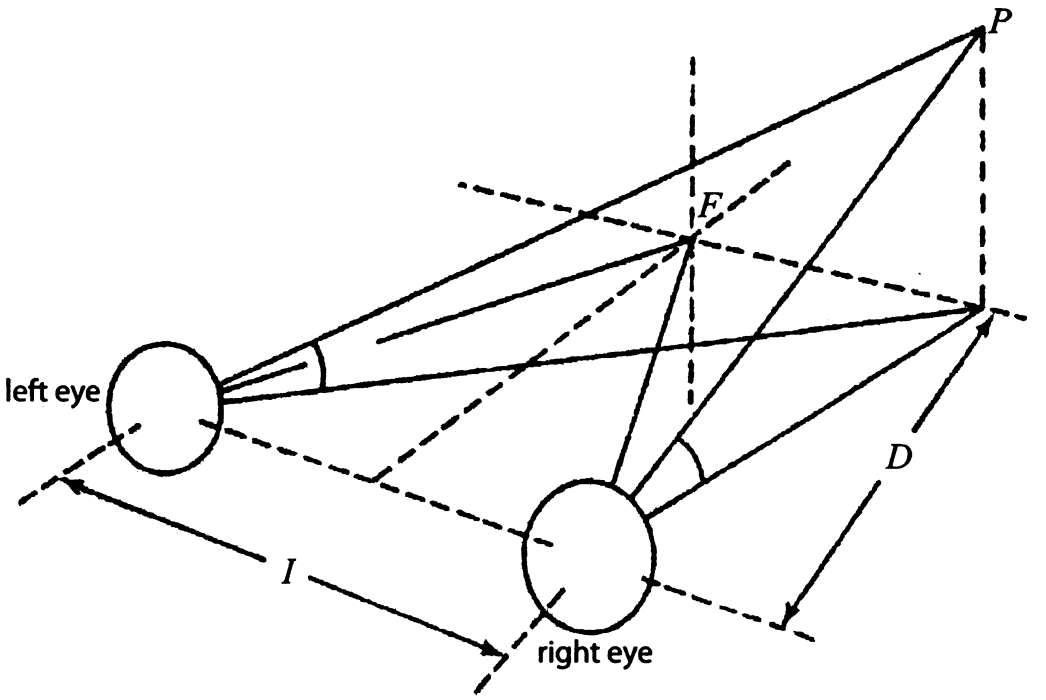


Figure 2.6: Vertical Disparity (reprinted from [3])

2.1.8 Encoding of Binocular Disparity

There are several ways that binocular disparities can be described. One can encode disparity as the retinal positions of visual features (such as edges) corresponding to the same spots in the visual field, or formulate the images as a set of sine waves using Fourier analysis, and encode disparity as the phase difference between the sine waves at the same retinal position. The former is referred to as *position disparity* and the latter is *phase disparity*. There is evidence supporting the existence of the both of disparities in biological visual systems [9]. These two possibilities are illustrated in Fig. 2.7.

Disparity selective cell type	Placement of stimuli
Tuned-excitatory	Stimuli at zero disparity
Tuned-inhibitory	Stimuli at all disparities except those near zero disparity
Near	Stimuli at negative disparity
Far	Stimuli at positive disparity

Table 2.1: Four basic types of disparity selective neurons.

2.2 Existing Work in Computational Modeling of Binocular Vision

Perhaps the first remarkable study of the neural mechanisms underlying binocular vision dates back to the 1960's by Barlow et. al. [2]. They discovered that neurons in the cat striate cortex respond selectively to the objects with different binocular depth. In 1997 Poggio and Fischer [16] did a similar experiment with an awake macaque monkey that confirmed the previous evidence by Barlow et. al. [2]. Since the visual system of these animals to a great extent resembles that of human, researchers believe that there are disparity-selective neurons in the human visual cortex as well. Poggio & Fischer [16] used solid bars as visual stimuli to identify and categorize the disparity selective neurons. Table 2.2 contains the naming they used to categorize the cell types.

Julesz 1971 [22] invented random dot stereogram (RDS), which was a great contribution to the field. A random dot stereogram consists of two images filled with dots randomly black or white, where the two images are identical except a patch of one image that is horizontally shifted in the other (Fig. 2.8).

When a human subject fixates eyes on a plane farther or closer to the plane on which RDS lies, due to the binocular fusion in the cortex, the shifted region jumps out (seems to be at a different depth from the rest of the image). Experiments based on RDS contributed to strengthen the theory of 4 categories of disparity selective

neurons [18]. Later experiments revealed the existence of two additional categories, named *tuned near* and *tuned far* [36]. Fig. 2.9 depicts the 6 categories identified by Poggio et. al. 1988 [36].

Despite neurophysiological data and thrilling discoveries in binocular vision, a computational model was missing until 1990 when Ohzawa et. al. [34] published their outstanding article in Science journal. They introduced a model called the *disparity energy model*. Later some results from physiological studies did not match the predictions made by energy model. Read et. al. 2002 [42] proposed a modified version of the original energy model. In the following sections, we present an overview of the two different versions of the important work of the energy model.

2.2.1 Energy Model

Ohzawa-DeAngelis-Freeman (ODF) 1990 [34] studied the details of binocular disparity encoding and detection in the brain, and tried to devise a computational model compatible with the biological studies of binocular vision. They argued that at least two more points need to be taken into account before one can devise a plausible model of the binocular vision.

1. Complex cells must have much finer receptive fields compared to what was reported by Nikara et. al. [32]

2. Disparity sensitivity must be irrelevant to the position of the stimulus within the receptive field.

Considering the limitations of the previous works and inspired by their own predictions, Ohzawa et. al. presented the *Energy Model* for disparity selective neurons. Fig. 2.10 schematically shows their model. There are 4 binocular Simple Cells (denoted by S) each receiving input from both eyes. The receptive field profile of the simple cells is depicted in small boxes. The output of the simple cells then

goes through a half-wave rectification followed by a squaring function. A complex cell (denoted by Cx in Fig. 2.10) then adds up the output of the 4 subunits $S1$, $S2$, $S3$ and $S4$ to generate the final output of the network.

Read et. al. [42] completed the previous energy model by Ohzawa et. al. [34]. They added monocular simple cells to the model that performs a half-wave rectification on the inputs from each eye before feeding them to the binocular simple cells. The authors claimed that the modification in the Energy Model results in the neurons exhibiting behavior close to real neuronal behavior when the input is anti-correlated binocular stimuli. Fig. 2.11 shows the modified Energy Model.

2.2.2 Wiemer et. al. 2000

Wiemer et. al. [21] used SOM as their model to exhibit self-organization for disparity preference. Their work was intriguing as for the first time it demonstrated the development of modeled binocular neurons. They took stereo images form three-dimensional scenes, and then built a binocular representation of each pair of stereo images by attaching corresponding stripes from the left and right images. They then selectively chose patches from the binocular representation to create their input to the network. An example of this pre-processing is shown in Fig. 2.12.

After self-organization they obtained *disparity maps* that exhibited some of the characteristics observed in the visual cortex. Fig. 2.13 shows one exmaple of the maps they reported.

2.2.3 Works based on LLISOM

Laterally Interconnected Synergetically Self-Organizing Maps by Mikkulainen et. al. [31] is a computational model of the self-organizing visual cortex that has been extensively studied over the past years. It emphasized the role of the lateral connections in such self-organization. Mikkulainen et. al. [31] point out three important

findings based on their models:

1. Self-organization is driven by bottom-up input to shape the cortical structure
2. Internally generated input (caused by genetic characteristics of the organism) also plays an important role in Self-organization of the visual cortex.
3. Perceptual grouping is accomplished by interaction between bottom-up and lateral connections.

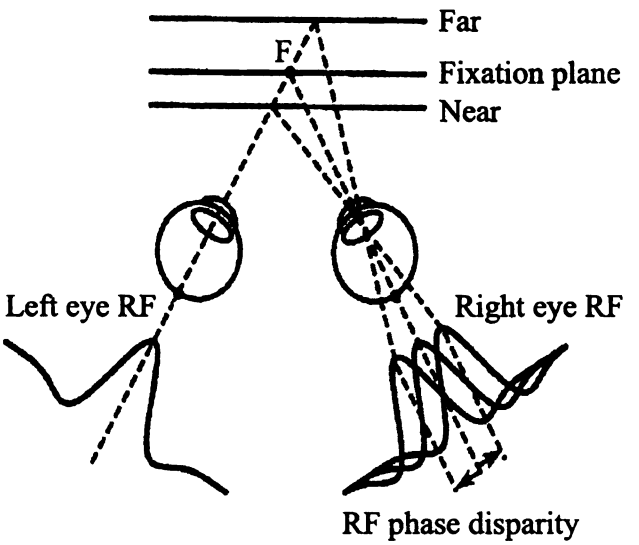
Although LLISOM was an important work that shed some light on the self-organization in the visual cortex, they failed to model an important part of the signals received at the visual cortex, namely *top-down* connections, and the role of this top-down connections in perception and recognition.

Fig. 2.14 shows an overall structure of the LLISOM. It consists of retina, LGN-ON and LGN-OFF sheets, and V1 sheet. Unlike SOM, in LLISOM each neuron is locally connected to a number of neurons in its lower-level sheet. Also, neurons are laterally connected to their neighbors. The strength of the connection between neurons is adapted during learning based on Hebbian learning rule. The process of learning connection weights is called *self-organization.* Thanks to lateral connections, LLISOM gains finer self-organized maps than SOM.

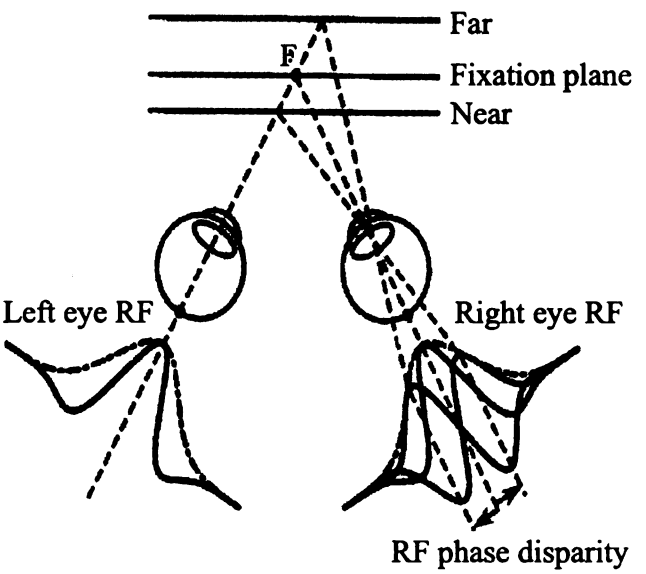
Fig. 2.15 presents an example of the self-organizing maps using LLISOM.

Ramtohul 2006 [39] studied the self-organization of disparity using LLISOM. He extended the basic architecture of LLISOM to handle two eyes, and the new architecture *two eye model* for disparity selectivity. Fig. 2.16 shows a schematic diagram of his model. He then provided the network with patches of natural images as input to investigate the emergence of disparity maps. The network successfully developed topographic disparity maps as a result of input-driven self-organization using LLISOM. However, this work did not provide any performance measurement

report, since the motor/action layer was absent in the model. Fig. 2.17 shows an example of the topographic disparity maps reported by Ramtohul 2006 [39].



(a) Position Difference Model



(b) Phase Difference Model

Figure 2.7: Two models of disparity encoding (reprinted from [1])

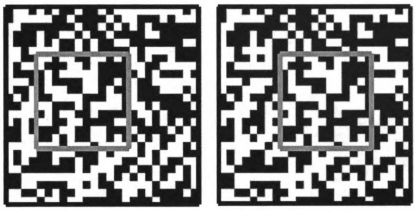


Figure 2.8: An example of random dot stereogram (reprinted from [40])

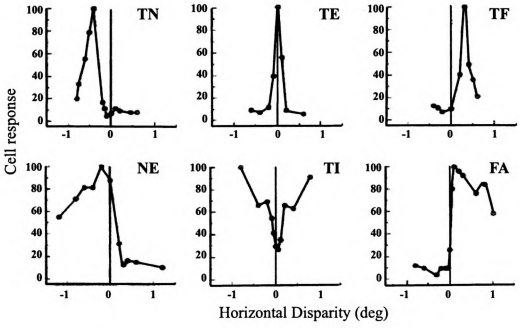


Figure 2.9: Disparity tuning curves for the 6 categories of disparity selective neurons. TN: tuned near, TE: tuned excitatory, TF: tuned far, NE: near, TI: tuned inhibitory, FA: far (reprinted from [18])

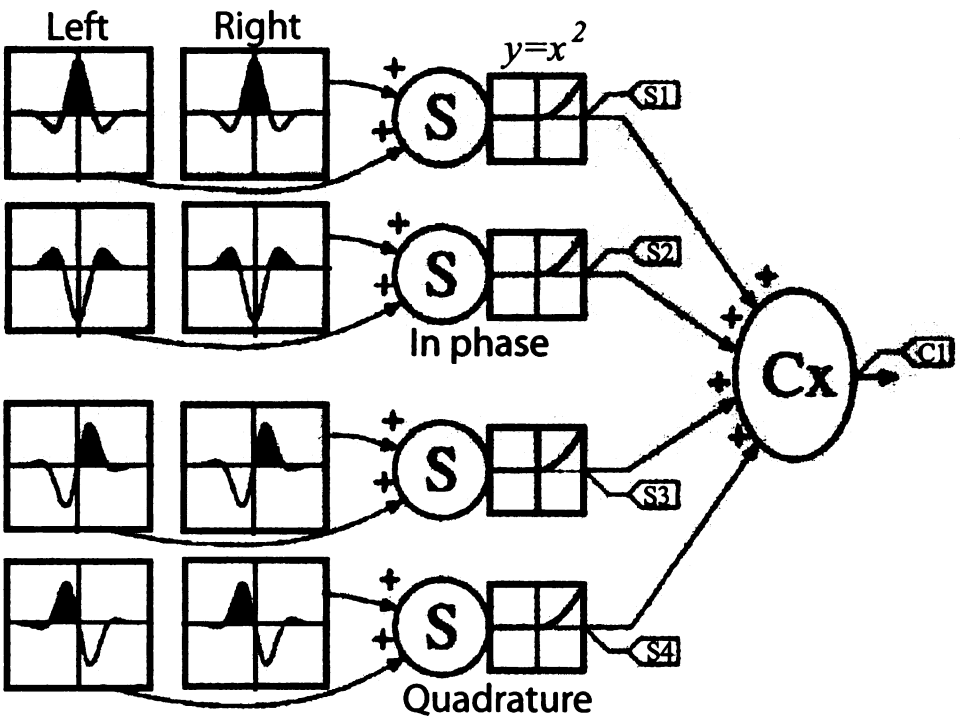


Figure 2.10: Energy Model by Ohzawa et. al. [34] (reprinted from [34])

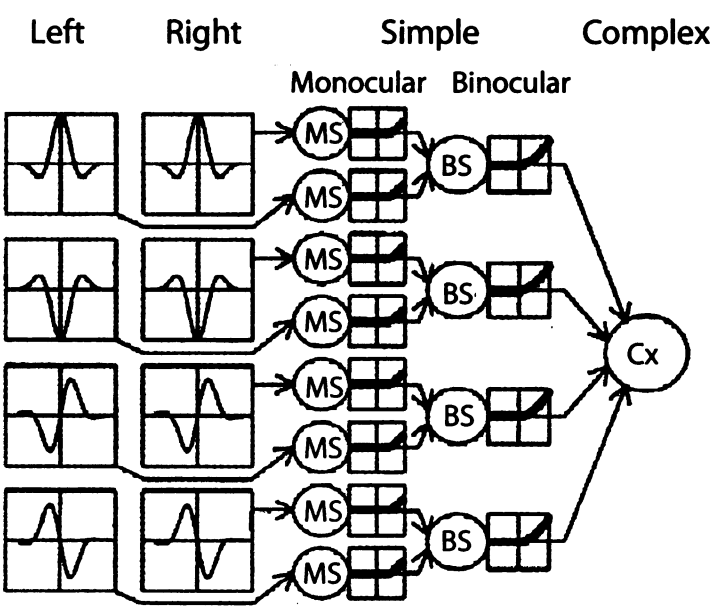


Figure 2.11: Modified Energy Model by Read et. al. [42] (reprinted from [42])

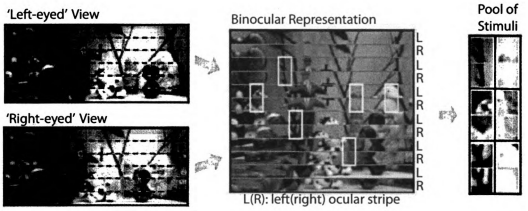


Figure 2.12: Pre-processing to create a pool of stimuli by Wimer et. al. [21] (reprinted from [21])

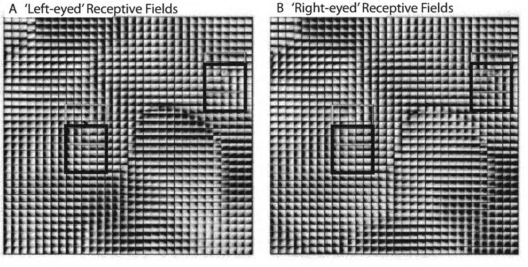


Figure 2.13: Self-organized maps of left and right eye receptive fields (reprinted from [21])

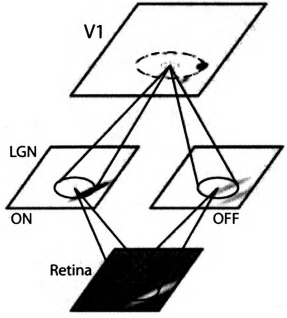


Figure 2.14: Schematic of the architecture for basic LLISOM (reprinted from [31])

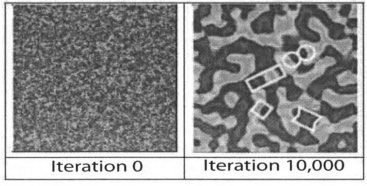


Figure 2.15: Self-organized orientation map in LLISOM (reprinted from [31])

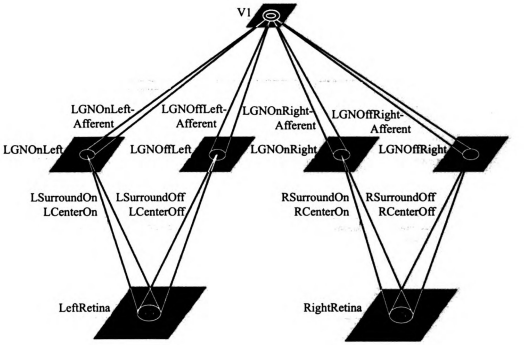


Figure 2.16: Two eye model for self organization of disparity maps in LLISOM (reprinted from [39])

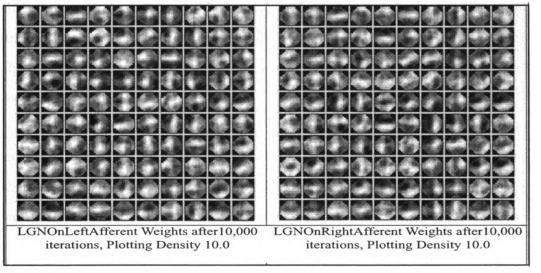


Figure 2.17: Topographic disparity maps generated by LLISOM (reprinted from [39])

Chapter 3

Overview of the Project

The past few decades of engineering efforts to solve the problem of stereo vision proves that the computational challenges of binocular disparity are far from trivial. In particular, the correspondence problem is extremely challenging considering difficulties such as featureless areas, occlusion, etc. Further, the existing engineering methods for binocular matching are not only computationally expensive, but are also hard to integrate with other visual cues to help the perception of depth. It is important to look at the problem from a different angle – How the brain solves the problem of binocular vision? In particular, what are the computational mechanisms that regulate the development of the visual nervous system, and what are the roles of gene-regulated cortical architecture and spatiotemporal aspects of such mechanisms?

In the real world, objects do not come into and disappear from the field of view randomly, but rather, they typically move continuously across the field of view, given their motion is not too fast for the brain to respond. At the pixel level, however, values are very discontinuous as image patches sweep across the field of view. Our model assumes that visual stimuli are largely spatially continuous. Motivated by the cerebral cortex, it utilizes the temporal context in the later cortical areas, including the intermediate areas and motor output area, to guide the development of earlier

areas. These later areas are more *"abstract"* than the pixel level, and thus provide needed information as temporal context. However, how to use such emergent information is a great challenge.

Existing methods for stereo disparity detection fall into three categories:

1. **Explicit matching:** Approaches in this category detect discrete features and explicitly match them across two views. Well-known work in this category include [19], [12] and [55].

2. **Hand-designed filters:** Filters are designed to compute profile-sensitive values (e.g. Gabor filters [53], [41], and phase information [14], [47]) from images and then utilize these continuous values for feature matching.

3. **Network learning models:** These models develop disparity-selective filters (i.e. neurons) from experience, without doing explicit matching, and map the responses to disparity outputs (e.g. [26], [27], [21], [15]).

Categories (1) and (2) employ explicit left and right match through either an explicit search or implicit gradient-based search. They are generally called explicit matching approaches.

Among the different stages of the explicit matching approaches, the *correspondence problem* is believed to be the most challenging step; i.e. the problem of matching each pixel of one image to a pixel in the other [30]. Solutions to the correspondence problem have been explored using area-, feature-, pixel- and phase-based, as well as Bayesian approaches [12]. While those approaches have obtained limited success in special problems, it is becoming increasingly clear that they are not robust against wide variations in object surface properties and lighting conditions [14].

The network learning approaches in category (3) do not require a match between the left and right elements. Instead, the binocular stimuli with a specific disparity are matched with binocular neurons in the form of neuronal responses. Different neurons

have developed different preferred patterns of weights, each pattern indicating the spatial pattern of the left and right receptive fields. Thus, the response of a neuron indicates a degree of match of two receptive fields, left and right. In other words, both texture and binocular disparity are measured by a neuronal response - a great advantage for integration of binocular disparity and spatial pattern recognition.

However, existing networks that have been applied to binocular stimuli are either bottom-up SOM type or error-back propagation type. There has been no biological evidence to support error back-propagation, but the Hebbian type of learning has been supported by the Spike-Time Dependent Plasticity (STDP) [10]. SOM type of networks that use both top-down and bottom-up inputs has not be studied until recently [44, 45, 48, 50]. In this paper we show that top-down connections that carry supervisory disparity information (e.g. when a monkey reaches an apple) enable neurons to self-organize according to not only bottom-up input, but also supervised disparity information. Consequently, the neurons that are tuned to similar disparities are grouped in nearby areas in the neural plane, forming what is called topographic class maps, a concept first discovered in 2007 [29]. Further, we experimentally show that such a disparity based internal topographic grouping leads to improved disparity classification.

Neurophysiological studies (e.g. [17] and [6]) have shown that the primary visual cortex in macaque monkeys and cats has a laminar structure with a local circuitry similar to our model in Fig. 4.3. However, a computational model that explains how this laminar architecture contributes to classification and regression was unknown. LAMINART [38] presented a schematic model of the 6-layer circuitry, accompanied with simulation results that explained how top-down attentional enhancement in V1 can laterally propagate along a traced curve, and also how contrast-sensitive perceptual grouping is carried out in V1. Weng et. al. 2007 [20] reported performance of the laminar cortical architecture for classification and recognition, and

Weng et. al. 2008 [50] reported the performance advantages of the laminar architecture (paired layers) over a uniform neural area. Franz & Triesch 2007 [15] studied the development of disparity tuning in toy objects data using an artificial neural network based on back-propagation and reinforcement learning. They reported a 90% correct recognition rate for 11 classes of disparity. In Solgi & Weng 2008 [46], a multilayer in-place learning network was used to detect binocular disparities that were discretized into classes of 4 pixels intervals from image rows of 20 pixels wide. This classification scheme does not fit well for higher accuracy needs, as a misclassification between disparity class -1 and class 0 is very different from that between a class -1 and class 4. The work presented here also investigates the more challenging problem of regression with sub-pixel precision, in contrast with the prior scheme of classification in Solgi & Weng 2008 [46].

For the first time, we present a spatio-temporal regression model of the laminar architecture of the cortex for stereo that is able to perform competitively on the difficult task of stereo disparity detection in natural images with sub-pixel precision. The model of the inter-cortical connections we present here was informed by the work of Felleman & Van Essen [13] and that for the intra-cortical connections was informed by the work of Callaway [5] and Wiser & Callaway [54] as well as others.

Luciw & Weng 2008 [28] presented a model for top-down context signals in spatio-temporal object recognition problems. Similar to their work, in this paper the emergent recursive top-down context is provided from the response pattern of the motor cortex at the previous time to the feature detection cortex at the current time. Biologically plausible networks (using Hebbian learning instead of error back-propagation) that use both bottom-up and top-down inputs with engineering-grade performance evaluation have not been studied until recently [20, 46, 50].

It has been known that orientation preference usually changes smoothly along the cortex [4]. Chen et. al. [7] has recently discovered that the same pattern applies

to the disparity selectivity maps in monkey $V2$. Our model shows that defining disparity detection as a regression problem (as opposed to classification) helps to form similar patterns in topographic maps; disparity selectivity of neurons changes smoothly along the neural plane.

In summary, the work here is novel in the following aspects: (1) The first laminar model (paired layers in each area) for stereo. (2) The first utilization of temporal signals in a laminar model for stereo (3) The first sub-pixel precision among the network learning models for stereo. Applying the novelties mentioned in (1) and (2) showed surprisingly drastic accuracy differences in performance. (4) The first study of smoothly-changing disparity sensitivity maps (5) Theoretical analysis that supports and provides insights into such performance differences.

Chapter 4

Network Architecture and Operations

The networks applied in this paper are extentions of the previous models of Multilayer In-place Learning Network (MILN) [50]. To comply with the principles of Autonomous Mental Development (AMD) [51], these networks autonomously develop features of the presented input, and no hand-designed feature detection is needed.

To investigate the effects of supervisory top-down projections, temporal context, and laminar architecture, we study two types of networks: 1) Single-layer architecture for classification and 2) 6-layer architecture for regression. An overall sketch of the networks is illustrated in Fig. 4.1. In this particular study, we deal with networks consisting of a sensory array (marked as *Input* in Fig. 4.1), a stereo feature-detection cortex, which may be a single layer of neurons or have a 6-layer architecture inspired by the laminar architecture of human cortex, and a motor cortex that functions as a regressor or a classifier.

4.1 Single-layer architecture

In the single-layer architecture, the feature-detection cortex simply consists of a grid of neurons that is globally connected to both the motor cortex and input. It performs the following 5 steps to develop binocular receptive fields:

1. Fetching input in $L1$ and imposing supervision signals (if any) in motor cortex – When the network is being trained, $\mathbf{z}^{(M)}$ is imposed originating from outside (e.g., by a teacher). In a classification problem, there are c motor cortex neurons and c possible disparity classes. The true class being viewed is known by the teacher, who communicates this to the system. Through an internal process, the firing rate of the neuron corresponding to the true class is set to one, and all others set to zero.

2. Pre-response – Neuron n_i on the feature-detection cortex computes its pre-competitive response $\hat{z}_i^{(L1)}$ – called *pre-response*, linearly from the bottom-up part and top-down part

$$\begin{aligned} \hat{z}_i^{(L1)}(t) &= (1-\alpha) \cdot \frac{\vec{b}^{(L1)}(t) \cdot \vec{w}_{b,i}^{(L1)}(t)}{\|\vec{b}^{(L1)}(t)\| \, \|\vec{w}_{b,i}^{(L1)}(t)\|} \\ &+ \alpha \cdot \frac{\vec{z}^{(M)}(t) \cdot \vec{w}_{e,i}^{(L1)}(t)}{\|\vec{z}^{(M)}(t)\| \|\vec{w}_{e,i}^{(L1)}(t)\|} \end{aligned} \tag{4.1}$$

where $\vec{w}_{b,i}^{(L1)}(t)$ and $\vec{w}_{e,i}^{(L1)}(t)$ are this neuron's bottom-up and top-down weight vectors, respectively, and $\vec{z}^{(M)}(t)$ is the firing rates of motor cortex neurons (supervised during training, and not active during testing). The relative top-down coefficient α is discussed in detail later. We do not utilize linear or non-linear function g, such as a sigmoid, on firing rate in this paper.

3. Competition via Lateral Inhibition – A neuron's pre-response is used for intra-level competition. k neurons with the highest pre-response win, and the others are inhibited. If $r_i = \text{rank}(\hat{z}_i^{(L1)}(t))$ is the ranking of the pre-response of the i'th neuron (with the highest active neuron ranked as 0), we have $z_i^{(L1)}(t) = s(r_i)\hat{z}_i^{(L1)}(t)$, where

$$\text{s}(r_i) = \begin{cases} \frac{k-r_i}{k} & \text{if } 0 \leq r_i < k \\ 0 & \text{if } r_i \geq k \end{cases} \tag{4.2}$$

4. Smoothing via Lateral Excitation – Lateral excitation means that when a neuron fires, the nearby neurons in its local area are more likely to fire. This leads to a smoother representational map. The topographic map can be realized by not only considering a nonzero-responding neuron i as a winner, but also its 3×3 neighbors, which are the neurons with the shortest distances from i (less than two).

5. Hebbian Updating with LCA – After inhibition, the top-winner neuron and its 3×3 neighbors are allowed to fire and update their synapses. We use an updating technique called lobe component analysis [52]. See Appendix A for details.

The motor cortex neurons develop using the same five steps as the above, but there is not top-down input, so Eq. 4.1 does not have a top-down part. The response $\vec{z}^{(M)}$ is computed in the same way otherwise, with its own parameter k controlling the number of non-inhibited neurons.

4.2 6-layer architecture

The architecture of the feature-detection cortex of the 6-layer architecture is sketched in Fig. 4.3. Layer $L1$ is connected to the sensory input in a *one-to-one* fashion; there is one neuron matched with each pixel, and the activation level of each neuron is equal to the intensity of the corresponding pixel (i.e. $\vec{z}^{(L1)}(t) = \vec{I}(t)$). We use no hand-designed feature detector (e.g. Laplacian of Gaussian, Gabor filters, etc.), as it

would be against the paradigm of AMD [51]. The other four layers[1] are matched in functional-assistant pairs (referred as feedforward-feedback pairs in [6]). $L6$ assists $L4$ (called assistant layer for $L4$) and $L5$ assists $L2/3$.

Layer $L4$ is globally connected to $L1$, meaning that each neuron in $L4$ has a connection to every neuron in $L1$. All the two-way connections between $L4$ and $L6$, and between $L2/3$ and $L5$, and also all the one-way connections from $L4$ to $L2/3$ are *one-to-one* and consant. In other words, each neuron in one layer is connected to only one neuron in the other layer at the same position in neural plane coordinates, and the weight of the connections is fixed to 1. Finally, neurons in the motor cortex are globally and bidirectionally connected to those in $L4$. There are no connections from $L2/3$ to $L4$.

The stereo feature-detection cortex takes a pair of stereo rows from the sensory input array. Then it runs the following developmental algorithm.

1. Fetching input in $L1$ and imposing supervision signals (if any) in motor cortex – $L1$ is a retina-like grid of neurons which captures the input and sends signals to $L4$ proportional to pixel intensities, without any further processing. During developmental training phase, an external teacher mechanism sets the activation levels of the motor cortex according to the input. If n_i is the neuron representative for the disparity of the currently presented input, then the activation level of n_i and its neighbors are set according to a triangular kernel centered on n_i. The activation level of all the other neurons is set to zero:

$$z_j^{(M)}(t) = \begin{cases} 1 - \frac{d(i,j)}{\kappa} & \text{if } d(i,j) < \kappa \\ 0 & \text{if } d(i,j) \geq \kappa \end{cases} \tag{4.3}$$

where $d(i,j)$ is the distance between neuron n_i and neuron n_j in the neural plane, and κ is the radius of the triangular kernel.

[1] $L2$ and $L3$ are counted as one layer ($L2/3$)

Then the activation level of motor neurons from the previous time step, $z_j^{(M)}(t-1)$, is projected onto $L2/3$ neurons via top-down connections.

2. Pre-response in $L4$ and $L2/3$ – Neurons in $L4(L2/3)$ compute their *pre-response* (response prior to competition) solely based on their bottom-up(top-down) input. They use the same equation as in Eq. 4.1, except $L4$ only has bottom-up and $L2/3$ only has top-down.

$$\hat{z}_i^{(L4)}(t) = \frac{\vec{b}^{(L4)}(t) \cdot \vec{w}_{b,i}^{(L4)}(t)}{\|\vec{b}^{(L4)}(t)\| \|\vec{w}_{b,i}^{(L4)}(t)\|} \tag{4.4}$$

and

$$\hat{z}_i^{(L2/3)}(t) = \frac{\vec{e}^{(L2/3)}(t) \cdot \vec{w}_{e,i}^{(L2/3)}(t)}{\|\vec{e}^{(L2/3)}(t)\| \|\vec{w}_{e,i}^{(L2/3)}(t)\|} \tag{4.5}$$

3. $L6$ and $L5$ provide modulatory signals to $L4$ and $L2/3$ – $L6$ and $L5$ receive the firing pattern of $L4$ and $L2/3$, respectively, via their one-to-one connections. Then they send modulatory signals back to their paired layers, which will enable the functional layers to do long-range lateral inhibition in the next step.

4. Response in $L4$ and second pre-response in $L2/3$ – Provided by feedback signals from $L6$, the neurons in $L4$ internally compete via lateral inhibition. The mechanism for inhibition is the same as described in Step 4 of single-layer architecture. The same mechanism concurrently happens in $L2/3$ assisted by $L5$, except the output of $L2/3$ is called the second pre-response (denoted by $\mathring{z}_i^{(L2/3)}(t)$).

5. Response in $L2/3$ – Each neuron, n_i in $L2/3$ receives its bottom-up input from one-to-one connection with the corresponding neuron in $L4$ (i.e. $b_i^{(L2/3)}(t) = z_i^{(L4)}(t)$). Then it applies the following formula to merge bottom-up and top-down

information and compute its response.

$$z_i^{(L2/3)}(t) = (1-\alpha) \cdot b_i^{(L2/3)}(t) + \alpha \cdot \dot{z}_i^{(L2/3)}(t) \tag{4.6}$$

where α is the relative top-down coefficient. We will discuss the effect of this parameter in detail in Section 6.2.1.

6a. Response of motor Neurons in Testing – The activation level of the motor neurons is not imposed during testing, rather it is computed utilizing the output of feature-detection cortex, and used as context information in the next time step. The neurons take their input from $L2/3$ (i.e. $\vec{b}_i^{(M)}(t) = \vec{z}^{(L2/3)}(t)$). Then, they compute their response using the same equation as in Eq. 4.4, and laterally compete. The response of the winner neurons is scaled using the same algorithm as in Eq. 4.2 (with a different k for the motor layer), and the response of the rest of the neurons will be suppressed to zero. The output of the motor layer is the response weighted average of the disparity of the winner neurons:

$$disparity = \frac{\sum_{n_i \text{ is winner}} d_i \times z_i^{(M)}(t)}{\sum_{n_i \text{ is winner}} z_i^{(M)}(t)} \tag{4.7}$$

where d_i is the disparity level that the winner neuron n_i is representative for.

6b. Hebbian Updating with LCA in Training – The top winner neurons in $L4$ and motor cortex and also their neighbors in neural plane (excited by 3×3 short-range lateral excitatory connections) update their bottom-up connection weights. Lobe component analysis (LCA) [52] is used as the updating rule. See Appendix A for details.

Afterwards, the motor cortex bottom-up weights are directly copied to $L4$ top-down weights.

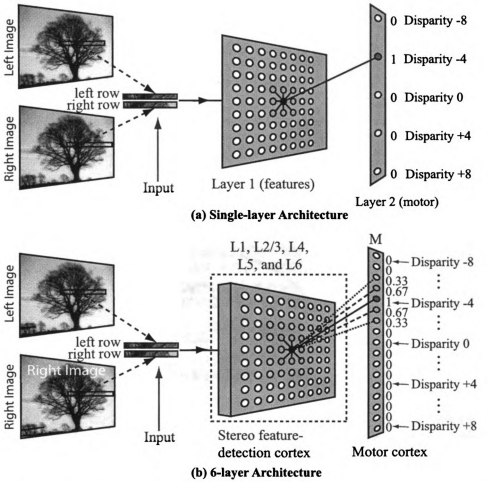


Figure 4.1: **(a)**. The binocular network single-layer architecture for classification. **(b)**. The binocular network 6-layer architecture for regression. Two image patches are extracted from the same image position in the left and right image planes. Feature-detection cortex neurons self-organize from bottom-up and top-down signals. Each motor neuron is marked by the disparity it is representative for (ranging from -8 to +8). Each circle is a neuron. Activation level of the neurons is shown by the darkness of the circles: the higher the activation, the darker the neurons are depicted. The diagram shows an instance of the network during training phase when the disparity of the presented input is -4. In **(a)** the stereo feature-detection cortex is a single layer of LCA neurons. A rectangular kernel sets the activation of only Disparity -4 neuron to 1 and all the others to 0. In **(b)**, the stereo feature-detection cortex has a 6-layer laminar architecture (see Fig. 4.3). A triangular kernel, centered at the neuron of Disparity -4, imposes the activation level of Disparity -4 neuron and four of its neighbors to positive values and all the others to 0.

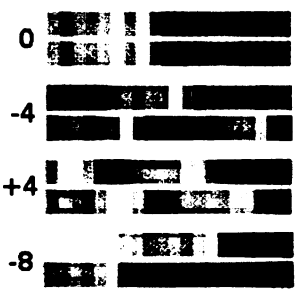


Figure 4.2: Examples of input, which consists of two rows of 20 pixels each. The top row is from the left view and the bottom row is from the right view. The numbers on the left side of the bars exhibit the amount of shift/disparity.

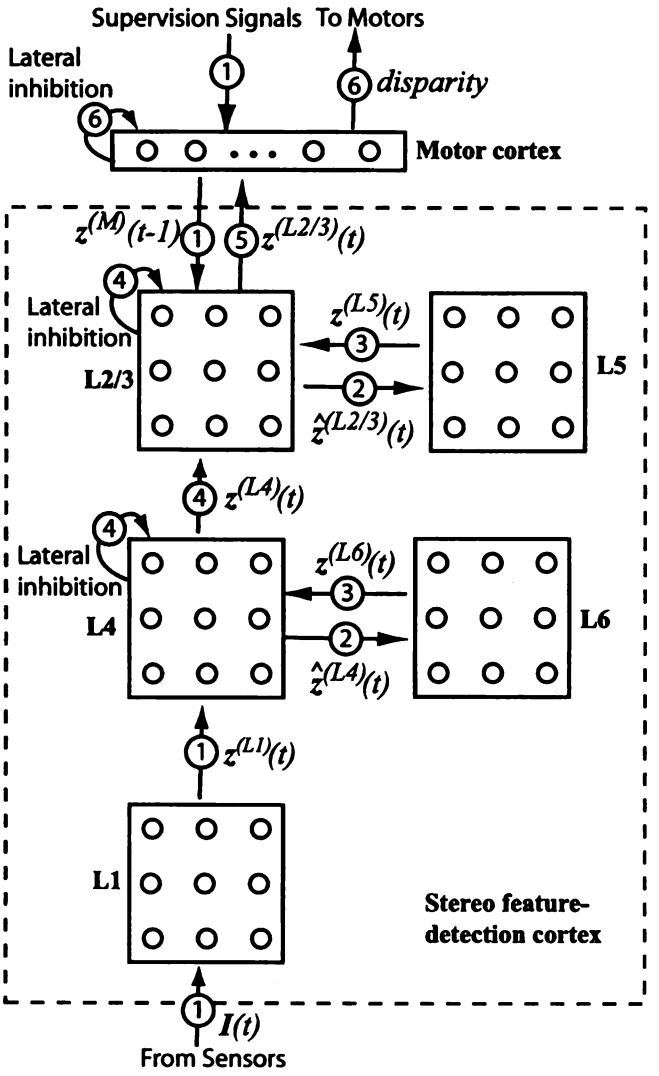


Figure 4.3: Architecture diagram of the 6-layer laminar cortex studied in this paper, which also introduces some notation. The numbers in circles are the steps of the algorithm described in Section 4. See the text for notations.

Chapter 5

Analysis

5.1 Elongated Input Fields Using Top-down

The neighborhood of the input space to which a neuron n_i is tuned (the neuron wins given input from that neighborhood) is called the spatial input field[1] of that neuron, denoted by $\Omega_i \subset \mathbb{R}^n$. We assume that for each neuron n_i the subspace Ω_i has a uniform distribution[2] along any direction (axis) d with mean value $\mu_{i,d}$ and standard deviation $\sigma_{i,d}$. The d'th element of the input vector $\vec{x}$ is denoted by x_d.

Proposition 1: *The higher the variation of data along a direction in the input field of a neuron, the less is the contribution of that direction of input to the neuron's chance to win in lateral competition.*

According to the principles of LCA learning [49], after development each neuron n_i is tuned to the mean value of its input field[3], $\mu_{i,d}$, along any direction d. Therefore, the average deviation of input from the neurons tuned weight is $\sigma_{i,d}$ for any direction d. It is evident that the larger this deviation $\sigma_{i,d}$ is, the less it is

[1] "a plot of the relationship between position in the input field and neural response [9]. It is also referred to as *input field profile.*

[2] which is a reasonable assumption given the data is patches from natural images

[3] from now on, wherever we refer to "input field" we mean "input field profile" or equivalently "spatial input field"

statistically probable that the input *matches* with the neuron's tuned weight along that direction, which in turn implies that the less is the contribution of x_d on the neuron's final chance to win in lateral competition with other neurons in the same layer.

Proposition 2: *Top-down connections help neurons develop input fields with higher variation along the irrelevant dimensions of input (elongated input fields).*

Given uniform distribution in input data, the neurons always develop in such a way that input space is divided equally among their input fields, in a manner similar to Voronoi diagrams. In other words, they develop *semi-isomorphic* input fields. Therefore, we expect that

$$\sigma_{i,d_1} = \sigma_{i,d_2} \tag{5.1}$$

for any neuron n_i and directions d_1 and d_2 along the uniform distribution manifold. However, when the neurons develop using top-down input, the projection of their input field on the bottom up input space is not isomorphic anymore. Instead, the bottom-up input field of the neuron is *elongated* along the direction of irrelevant input (See Fig. 5.1). Assuming linear dependence of Z on X_R in Fig. 5.1), we have:

$$\sigma_{i,d_{ir}} = \lambda\beta\sigma_{i,d_{rel}} \tag{5.2}$$

where d_{ir} d_{rel} respectively represents any irrelevant and relevant dimensions of the bottom-up input, and β and λ are constants. According to the triangle similarity (see Fig. 5.1), when we project the input space onto bottom-up space, the constant λ is a function of the ratio of the range of top-down input, z_m, to the bottom-up input, x_m:

$$\lambda = \frac{\sqrt{x_m^2 + z_m^2}}{x_m^2} = \sqrt{1 + \left(\frac{z_m}{x_m}\right)^2} \tag{5.3}$$

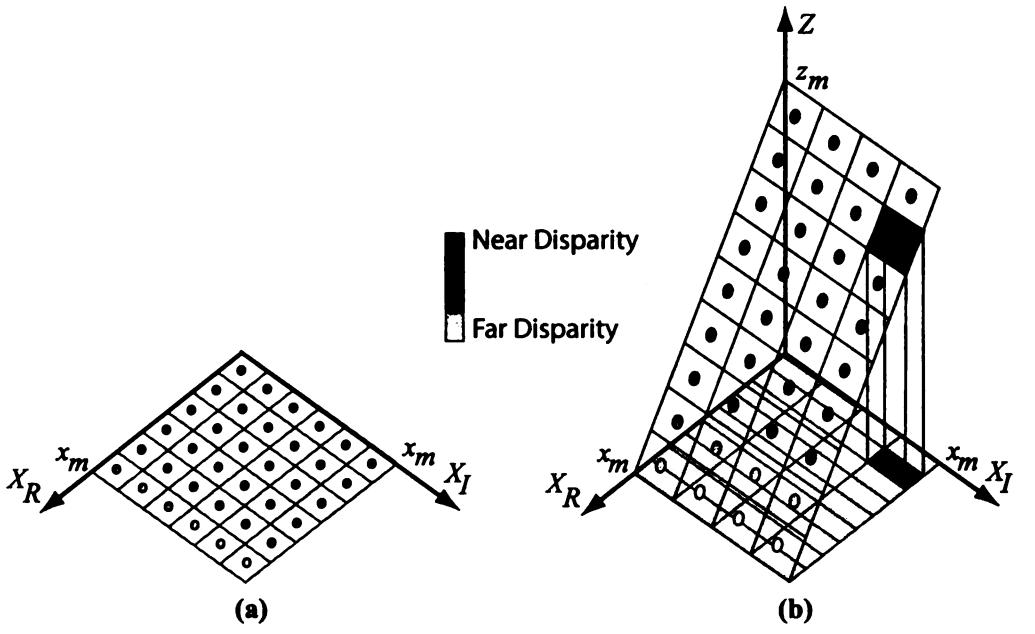


Figure 5.1: Each circle represents a neuron, and the shade of circles represents the degree of disparity the neuron is tuned to. The areas shown around neurons are the the input fields of neurons. **(a)** The quantization of input space by neurons without top-down input. The input fields of neurons has the same amount of variation in either of directions relevant and irrelevant input (shown as a square for the sake of visualization simplicity, should be Voronoi diagrams). **(b)** The quantization of input space by neurons with top-down input. For simplicity we assume the there is a linear relation between relevant part of bottom-up input, X_R, and the top-down input, Z. The input fields of the neurons are still isomorphic (shown as squares) on the input manifold. However, the projection of the input fields on the bottom-up space is no longer isomorphic, but elongated along the irrelevant axis.

where $x_d \in (0, x_m)$ and $z_d \in (0, z_m)$ for any direction d. Hence:

$$\lambda > 1 \tag{5.4}$$

[4]. The value of β is a function of relative top-down coefficient, α, in Eq. 4.1, and also the ratio of the number of relevant and irrelevant dimensions in input. In the settings we used in this paper, an estimation of β is as follows:

$$\beta \simeq \alpha \frac{dim(\vec{x})}{dim(\vec{z})} = 0.4 \times \frac{32}{8} = 1.6 \tag{5.5}$$

[4] $\lambda = \sqrt{2}$ given $z_m = x_m$

where $dim(\vec{x})$ and $dim(\vec{z})$ are the average[5] number of dimensions (number of elements) in the bottom-up and top-down input vectors. Therefore, the following inequality always holds:

$$\beta > 1 \tag{5.6}$$

Equations 5.2, 5.4 and 5.6 together imply that:

$$\sigma_{i,d_{ir}} > \sigma_{i,d_{rel}} \tag{5.7}$$

which is the variation of input fields of the neurons is higher along the irrelevant dimensions, and the reasoning is complete.

Combining Proposition 1 and Proposition 2, we conclude that:

Theorem 1: *As a result of top-down connections, neurons autonomously develop input fields in which they are relatively less sensitive to irrelevant parts of the input.*

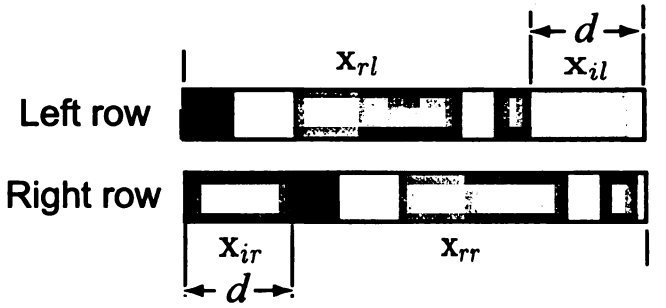


Figure 5.2: Top-down connections enable neurons to pick up relevant receptive fields. If a neuron is supervised by the top-down connections to detect a particular disparity d, the irrelevant subspace includes those areas where object images do not overlap, i.e. $\vec{x}_{il}$ and $\vec{x}_{ir}$. The first subindex indicates whether it is the irrelevant or relevant part of the input space (i and r respectively), and the second subindex shows whether it is from the left view or right view (l and r respectively).

[5]dimensions change according to degree of disparity (See Fig. 5.2)

5.2 Top-down Connections Help Recruit Neurons More Efficiently

According to the rules of in-place learning [48], neurons don't know whether their inputs are from bottom-up or top-down, neither do they know where they are in the cortical architecture. Each neuron can be thought as an autonomous agent that learns on its own without the help of any controlling mechanism from outside. Adding top-down connections to a neuron increases its input dimensionality from X to $X \times Z$ where

$$U = X \times Z = \{(x, z) | x \in X, z \in Z\} \tag{5.8}$$

where $\times$ is the *Cartesian product* operator meaning that the new space $X \times Z$ includes inputs from both bottom-up and top-down input spaces. X and Z are respectively bottom-up and top-down input spaces, defined as the following:

$$X = \{x = \vec{b}_i | \vec{b}_i \text{ is the bottom-up weights of any neuron } n_i\} \tag{5.9}$$

$$Z = \{z = \vec{e}_i | \vec{e}_i \text{ is the top-down weights of any neuron } n_i\} \tag{5.10}$$

In general, bottom-up input space X of each neuron in a cortical area is composed of the relevant subspace R, the space that is related to motor output, and irrelevant subspace I, the part of input that is not related to the output:

$$X = R \times I \tag{5.11}$$

It is evident that the top-down input from the space Z is relevant to the output. Thus, we write:

$$U = X \times Z = (I \times R) \times Z = I \times (R \times Z) \tag{5.12}$$

representing that when top-down input is present the new relevant subspace consists of both subspaces R and Z. Besides, the top-down inputs are relatively very variant compared to bottom-up input, since during supervision each value is set to either zero or a non-zero value. Therefore, the following property holds:

Property 1: *Adding top-down signals to a neuron increases the dimensionality and variance of its relevant input subspace.*

Furthermore, the following property is true given any distribution of input

Property 2: *Neurons are more recruited along the direction of higher variation in input space.*

A rigorous mathematical proof of this property is beyond the scope of this paper, however, an intuitive illustration is given in Fig. 5.3.

Combining Properties 1 and 2, we conclude that:

Property 3: *Adding top-down connections to neurons results in the recruitment of the neurons more along the direction of relevant input subspace and hence improves the performance of the network.*

Even if the top-down signals are not available during testing (in case we don't use context signals during testing), they have already helped neurons tune along the direction of relevant input subspace.

To sum up, we argued that the top-down signals help improve the network performance by increasing the variance of the input space along the direction of relevant input space.

5.3 Why use 6-layer Architecture?

In this section, we analytically investigate why and how the 6-layer laminar architecture outperforms the single-layer architecture model. Fig. 5.4 compares the algorithms by which the activation level of the neurons in single-layer and 6-layers architectures is computed. In single-layer architecture (the top row in Fig. 5.4),

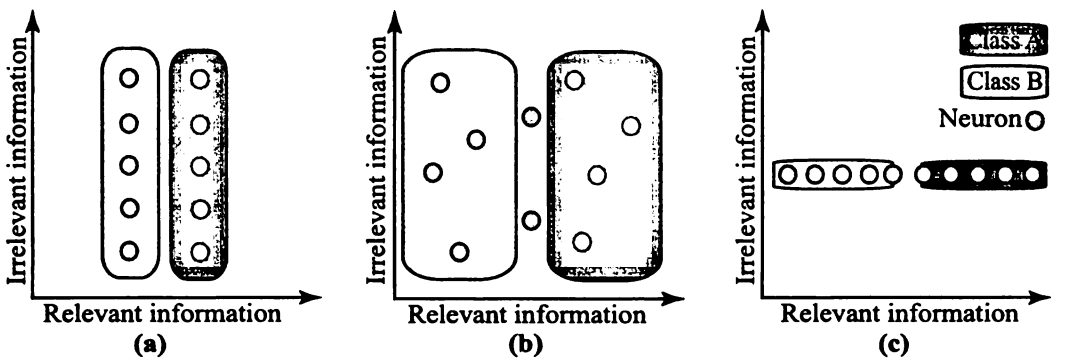


Figure 5.3: The deviation of samples along any direction in the input space recruits neurons along this direction. **(a)** The subspace of relevant information has smaller variance than the irrelevant information. Neurons spread more along the direction of irrelevant subspace. In other words, more neurons characterize the values in the irrelevant space (e.g., 5 neurons per unit distance versus 2 per unit distance). **(b)** Scale the relevant input by a factor of 2, increasing the standard deviation by a factor of two. Then, neurons spread in both direction with similar densities. **(c)** Further scale down the irrelevant input, enabling neurons to spread exclusively along the relevant direction (i.e., invariant to irrelevant direction).

the top-down and bottom-up energies are first computed and proportionally added according to Eq. 5.13.

$$z_i = (1 - \alpha) \cdot E_{b,i} + \alpha \cdot E_{e,i} \tag{5.13}$$

$$E_{b,i} = \frac{\vec{b}_i \cdot \vec{w}_{b,i}}{\|\vec{b}_i\| \|\vec{w}_{b,i}\|}, E_{e,i} = \frac{\vec{e}_i \cdot \vec{w}_{e,i}}{\|\vec{e}_i\| \|\vec{w}_{e,i}\|} \tag{5.14}$$

The notation here is consistent with those in Equations 4.4, 4.5 and 4.6 [6]. In most real world sensory data, such as stereo pairs in our case, the bottom-up sensory vector ($\vec{b}_i$ in Eq. 5.14) is significantly more uniform than the top-down supervision/context vector [7]. In the case of binocular disparity detection, the input pair of images is often featureless with similar intensities for the majority of pixels, while the top-

[6] Except we dropped the time and layer ID components, for the sake of simplicity.

[7] Variance of the elements of the bottom-up sensory vector ($\vec{b}_i$ in Eq. 5.14) is significantly lower than variance of the elements of the top-down supervision/context vector ($\vec{e}_i$ in Eq. 5.14)

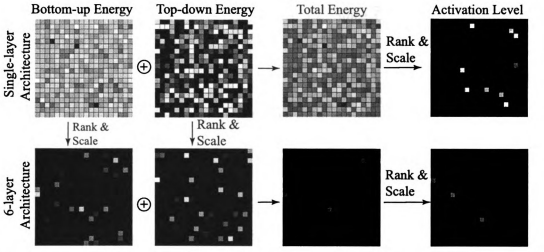


Figure 5.4: The mechanisms of neuron winner selection (via lateral inhibition) in single-layer and 6-layer architectures. The maps are taken from a snap-shot of the 20×20 neurons in the networks performing on real data. Each small square projects the value for a neuron in that particular position (black(white): minimum(maximum) values). The top row shows the steps in the single-layer architecture, and the bottom row shows the steps for the 6-layer architecture (which shares some steps with the single-layer architecture). $\oplus$ represents the operation of taking weighted average of two vectors (similar to Eq. 4.6).

down context/supervision vector is relatively more variant. As a result we have

$$var(E_b) \ll var(E_e) \tag{5.15}$$

where E_b and E_e are two random variables that can get any of the values $E_{b,i}$ and $E_{e,i}$, respectively. Here, we show that as a result of the lack of variation in bottom-up stimuli in such a single-layer architecture, activation level of the feature detection neurons is mostly determined by only top-down energy and the bottom-up energy is almost discarded. Obviously, this greatly reduces the performance of the network, as the top-down context signals are misleading when the input to the network at time t is considerably different from the input at time $t-1$. We call this effect "*hallucination*".

Let us define $\vec{\tilde{E}}_b = \vec{E}_b - \bar{\vec{E}}_b \vec{I}$ where $\bar{\vec{E}}_b$ is the mean value of the elements in $\vec{E}_b$ (scalar value) and $\vec{I}$ is the unit matrix of the same size as $\vec{E}_b$. Also, $\vec{\tilde{E}}_e = \vec{E}_e - \bar{\vec{E}}_e \vec{I}$

in the same manner, and $\tilde{\vec{z}} = (1-\alpha) \cdot \vec{\tilde{E}}_b + \alpha \cdot \vec{\tilde{E}}_e$. Since $\tilde{\vec{z}}$ is only a constant term different from $\vec{z}$, we have

$$\text{rank}(z_i) = \text{rank}(\tilde{z}_i) \tag{5.16}$$

which is, the rank of each element z_i in $\vec{z}$ is the same as the rank of the corresponding element $\tilde{z}_i$ in $\tilde{\vec{z}}$. In addition, the rank of each element $\tilde{z}_i = (1-\alpha) \cdot \tilde{E}_{b,i} + \alpha \cdot \tilde{E}_{t,i}$ is mostly determined by its top-down component, $\tilde{E}_{t,i}$. The reason is because Eq. 5.15 induces the absolute value of the top-down component for most of the neurons is much greater than the absolute value of the bottom-up component, i.e. $|\tilde{E}_{t,i}| \gg |\tilde{E}_{b,i}|$. Hence, the ranking of neurons' activation is largely effected only by their top-down component, and the reasoning is complete.

On the other hand, in the case of 6-layer architecture (the bottom row in Fig. 5.4), the bottom-up and top-down energies are ranked separately in $L4$ and $L2/3$, respectively, before they get mixed and compete again to decide the winner neurons in $L2/3$. Therefore, as a result of separation of bottom-up and top-down energies in different laminar layers, the 6-layer architecture manages to out-perform the single-layer architecture, specially when the imperfect context top-down signals are active (as opposed to supervision top-down signals which are always perfect).

5.4 Recovery from Hallucination

Fig. 5.5 is an intuitive illustration of how ranking top-down and bottom-up energy separately, as done in the 6-layer laminar architecture, will lead to recovery from a *hallucination* state, while the single layer architecture cannot recover. This analysis is consistent with the results presented in Fig. 6.5.

In Fig. 5.5, the input space of neurons is shown on the two axes; top-down input is represented by the horizontal axis, and bottom-up input is represented by the vertical axis. The input signals to the networks are depicted in filled curves along

the axes. Distribution of the two classes A and B are shown in rounded rectangles which are wider along the direction of the top-down input since, as discussed earlier in Section 5.3, top-down input is more variant than the bottom-up which results in recruitment of neurons more along the top-down direction according to Property 2. The two classes are shown to be linearly separable [8] along the direction of top-down input, but not along the bottom-up input, because top-down signals are always relevant during training. We assume that only top 2 neurons fire (e.g. $k = 2$).

In a single-layer architecture (Fig. 5.5a), given an input with wrong top-down component of class B while the input actually belongs to class A (e.g. when context is unrelated to the bottom-up input), the network will be trapped in a hallucination state, because the high variation of the top-down signal leaves a very small chance for the input to lie close to neurons in class A. Fig. 5.5a illustrates that having a similar bottom-up input at time $t + 1$ (according to spatial continuity of the input) will not change the situation.

On the other hand, in a 6-layer architecture, the neurons compete for top-down energy (in $L2/3$) and bottom-up energy (in $L4$) separately. In the first row, far left plot of Fig. 5.5b two neurons in class B have high pre-responses because of the wrong (misleading) top-down input, and two other neurons in class A have high pre-responses because of the right (correct) bottom-up input. As a result, there is a high chance that there are winners among the class A neurons. As the new sample comes in at time $t + 1$ (with the same or very similar bottom-up component due to spatial continuity of input), it is expected that only neurons in the correct class A win as both their bottom-up and top-down component are closer to the input. Finally the network *recovers* in the far right plot in Fig. 5.5b as both the winner neurons are from the correct class A, and the top-down input will be right from then on.

[8] shown linearly separable only for the sake of illustration simplicity in the figures

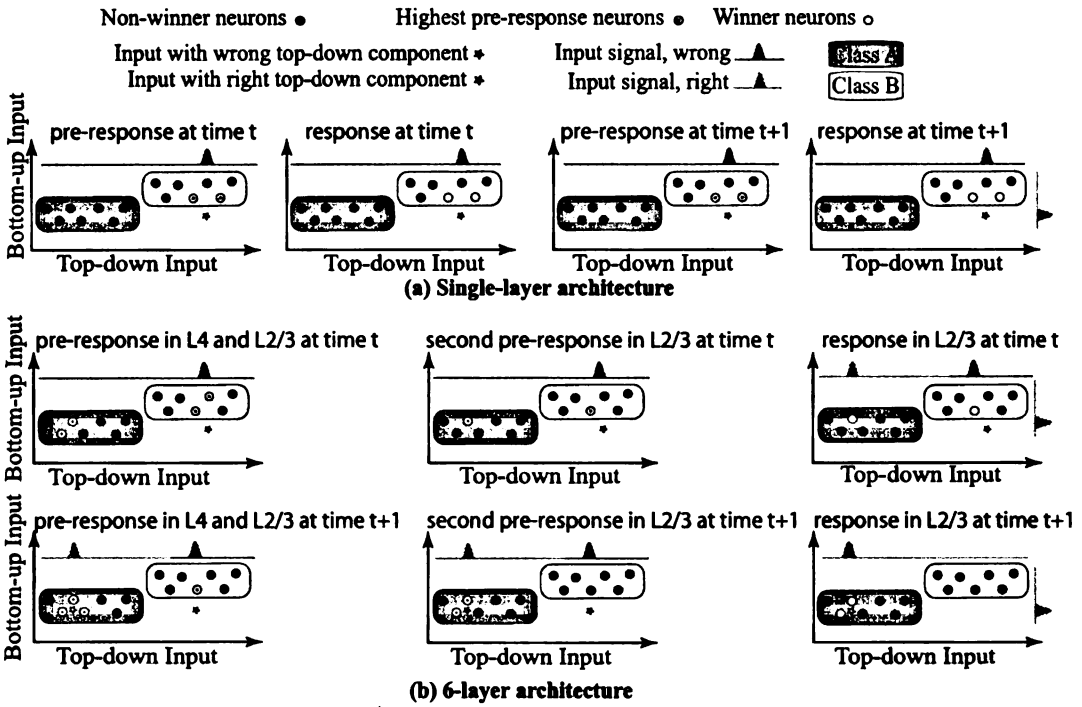


Figure 5.5: Schematic illustration of how 6-layer architecture, as opposed to single-layer architecture, makes recovery possible. A sample from class A is given to the network during testing (after the network is developed) while the context top-down signals are related to class B (wrong top-down signals depicted in red(darker) in the figure) . This causes the input to the neurons to be considered as a malicious (wrong) input (denoted by red(darker) stars) and lie out of the input distributions. This figure illustrates the state of the networks after receiving such an input. **(a)** Single-layer architecture. At time t, two closest neurons to the input have the highest pre-responses ($k = 2$). They win and fire. The winner neurons cause the top-down context input to slightly change/adapt to their top-down values. However, this change is not beneficial as the top-down component is still wrong. Therefore, at time $t + 1$ the input will still be classified as class B, which is wrong. **(b)** In a 6-layer architecture, neurons in $L4$ compete for bottom-up energy and two vertically closest neurons to the input have the highest pre-response and win. In the same manner, two horizontally closest neurons to the input in $L2/3$ have the highest pre-response and win. Then when the pre-response of neurons in $L2/3$ is computed it is very probable that some neurons from the correct class A have high preresponses and win in the next step (1st row of (b) far right graph). As a result, top-down input will have a *right* component as well. Because of this right component of the top-down signal, at the next time step $t + 1$, the network receives a right input (shown by light star in the 2nd row of (b) far left graph) besides the wrong input. Therefore, we see that one of the final winner neurons is in the correct class A. At the next time step $t + 1$ the network recovers to the state where the top-down signals are right again.

Chapter 6

Experiments and Results

The results of the experiments carried out using the models discussed in the previous chapters are presented here. The binocular disparity detection was formulated once as a classification problem, and then as a regression problem.

6.1 Classification

The input to the network is a pair of left and right rows, each 20 pixels wide. The image-rows were extracted randomly from 13 natural images (available from http://www.cis.hut.fi/projects/ica/imageica/). The right-view row position is shifted by -8, -4, 0, 4, 8 pixels, respectively, from the left-view row, resulting in 5 disparity classes. Fig. 4 shows some sample inputs. There were some image regions where texture is weak, which may cause difficulties in disparity classification, but we did not exclude them. During training the network was randomly fed with samples from different classes of disparity. The developed filters in Layer 2 are shown in Fig. 6.1.

6.1.1 The Effect of Top-Down Projection

As we see in Fig. 6.2, adding top-down projection signals improves the classification rate significantly. It can be seen that when $k = 50$ for the top-k updating rule,

the correct classification rate is higher early on. This is expected as no feature detector can match the input vector perfectly. With more neurons allowed to fire, each input is projected onto more feature detectors. The population coding gives richer information about the input, and thus, also the disparity. When more training samples are learned, the top-1 method catches up with the top-50 method.

6.1.2 Topographic Class Maps

As we see in Fig. 6.3, supervisory information conveyed by top-down connections resulted in topographically class-partitioned feature detectors in the neuronal space, similar to the network trained for object recognition [29]. Since the input to a neuron in Layer 1 has two parts, the iconic input $\vec{x}_b$ and the abstract (e.g. class) input $\vec{x}_t$, the resulting internal representation in Layer 1 is *iconic-abstract.* It is grossly organized by class regions, but within region it is organized by iconic input information. However, these two types of information are not isolated - they are considered jointly by neuronal self-organization.

To measure the purity of the neurons responding to different classes of disparity, we computed the entropy of the neurons as follows:

$$H = \sum_{i=1}^{N} -p(n, C_i) \log(p(n, C_i)) \tag{6.1}$$

where N is the number of classes and $p(n, C_i)$ is defined as:

$$p(n, C_i) = \frac{f(n, C_i)}{\sum_{j=0}^{m} f(n, C_j)} \tag{6.2}$$

where n is the neuron, C_i represents class i, and $f(n, C_i)$ is the frequency for the neuron n to respond to the class C_i.

Fig. 6.4 shows that the topographic representation enabled by the top-down projections generalizes better and increases the neurons' purity significantly during

training and testing.

6.2 Regression

From a set of natural images (available from http://www.cis.hut.fi/projects/ica/imageica/), 7 images were randomly selected, 5 of them were randomly chosen for training and 2 for testing. A pair of rows, each 20 pixels wide, were extracted from slightly different positions in the images. The right-view row was shifted by $-8, -7, -6, \ldots, 0, \ldots, +6, +7, +8$ pixels from the left-view row, resulting in 17 disparity degrees. In each training epoch, for each degree of disparity, 50 spatially continuous samples were taken from each of the 5 training images. Therefore, there was $5 \times 50 \times 17 = 4250$ training samples in each epoch. For testing, 100 spatially continuous samples were taken from each of the 2 testing images (disjoint test), resulting in $2 \times 100 \times 17 = 3400$ testing samples in each epoch.

We trained networks with 40×40 neurons in each of $L2/3$, $L4$, $L5$ and $L6$ layers of stereo feature-detection cortex (of course there were 2×20 neurons in $L1$, as there is a one-to-one correspondence between input and $L1$ neurons). The k parameter (the number of neurons allowed to fire in each layer) was set to 100 for the stereo feature-detection cortex, and 5 for the motor cortex. We set $\kappa = 5$ in Eq. 4.3 and $\alpha = 0.4$ in Eq. 4.6 for all of the experiments, unless otherwise is stated.

6.2.1 The Advantage of Spatio-temporal 6-layer Architecture

Fig. 6.5 shows that applying top-down context signals in single-layer architecture (traditional MILN networks [50]), increases the error rate up to over 5 pixels (we intentionally set the relative top-down coefficient, α, as low as 0.15 in this case, otherwise the error rate would be around chance level). As discussed in Section

5, this observation is due the absolute dominance of misleading top-down context signals provided complex input (natural images in this study). On the other hand, context signals reduce the error rate of the network to a sub-pixel level in 6-layer architecture networks. This result shows the important role of assistant layers (i.e. $L5$ and $L6$) in the laminar cortex to modulate the top-down and bottom-up energies received at the cortex before mixing them.

For comparison, we implemented two versions of Self-Organizing Maps updating rules, Euclidean SOM and dot-product SOM [25]. With the same amount of resources, the 6-layer architecture outperformed both versions of SOM by as much as at least 3 times lower error rate.

In another experiment, we studied the effect of relative top-down coefficient α. Different networks were trained with more than 40 thousand random training samples (as opposed to training with epochs). Fig. 6.6 shows the effect of context parameter, α, in disjoint testing. It can be seen that the root mean square error of disparity detection reaches to around 0.7 pixels when $\alpha = 0.4$. We believe that in natural visual systems, the ratio of contribution of top-down temporal signals (α in our model) is tuned by evolution.

6.2.2 Smoothly Changing Receptive Fields

In two separate experiments, we studied the topographic maps formed in $L2/3$.

Experiment A $-\ \kappa = 5$ As depicted in Fig. 6.7a, the disparity-probability vectors for neurons tuned to close-by disparities are similar; neurons tuned to close-by disparities are more likely to fire together. Equivalently, a neuron in the stereo feature-detection cortex is not tuned to only one exact disparity, but to a disparity range with a Gaussian-like probability for different disparities (e.g. neuron n_i could fire for disparities $+1, +2, +3, +4, +5$ with probabilities $0.1, 0.3, 0.7, 0.3, 0.1$,

respectively). This fuzziness in neuron's disparity sensitivity is caused by smoothly changing motor initiated top-down signals ($\kappa > 1$ in Eq. 4.3) during training. Fig. 6.7b shows this effect on topographic maps; having $\kappa = 5$ causes the regions sensitive to close-by disparities quite often reside next to each other and change gradually in neural plane (in many areas in Fig. 6.7b the colors change smoothly from dark blue to red).

Experiment B $-\ \kappa = 1$ However, if we define disparity detection as a classification problem, and set $\kappa = 1$ in Eq. 4.3 (only one neuron active in motor layer), then there is no smoothness in the change of the disparity sensitivity of neurons in the neural plane.

These observations are consistent with recent physiological discoveries about the smooth change of stimuli preference in topographic maps in the brain [8] and disparity maps in particular [7, 43].

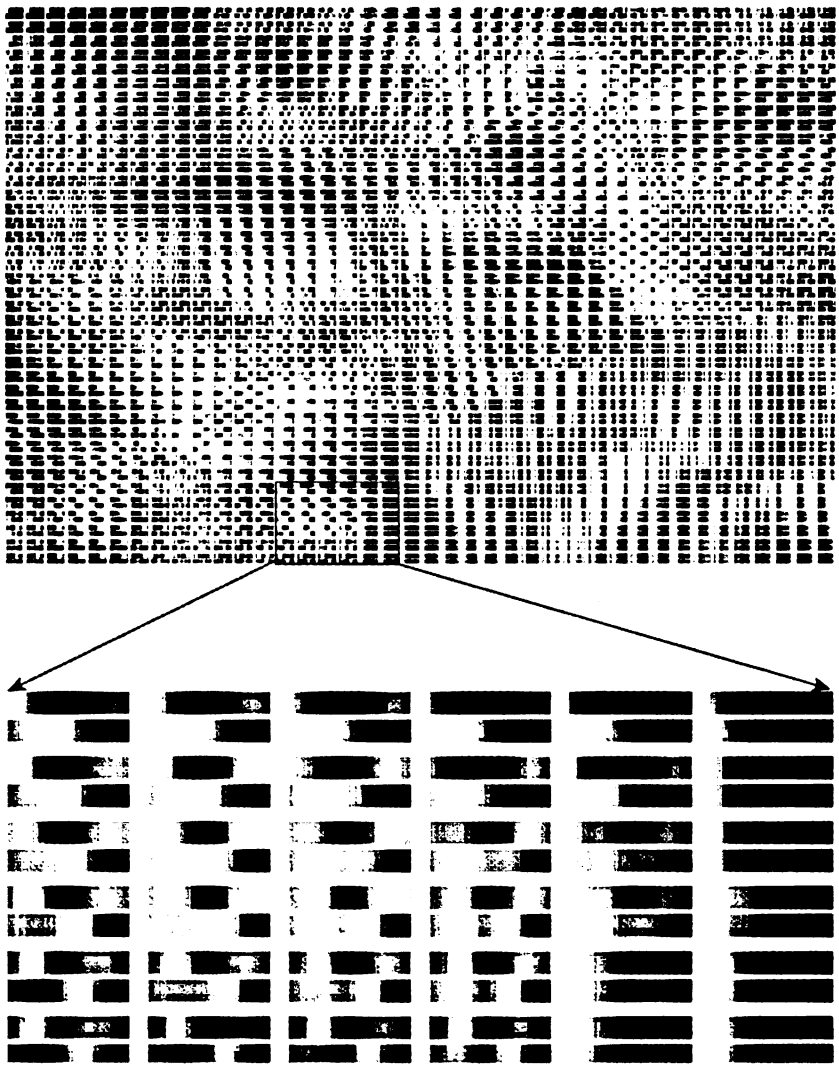


Figure 6.1: Bottom-up weights of 40 × 40 neurons in feature-detection cortex using top-down connections. Connections of each neurons are depicted in 2 rows of each 20 pixels wide. The top row shows the weight of connections to the left image, and the bottom row shows the weight of connections to the right image.

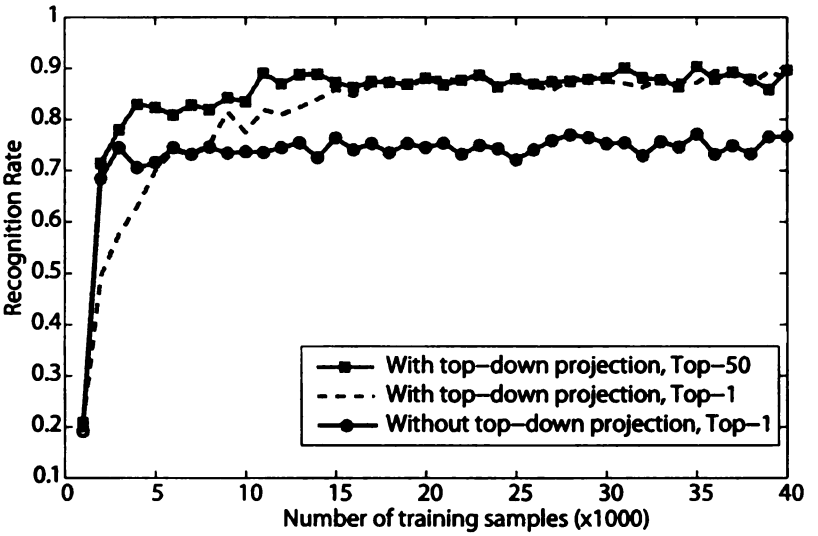


Figure 6.2: The recognition rate versus the number of training samples. The performance of the network was tested with 1000 testing inputs after each block of 1000 training samples.

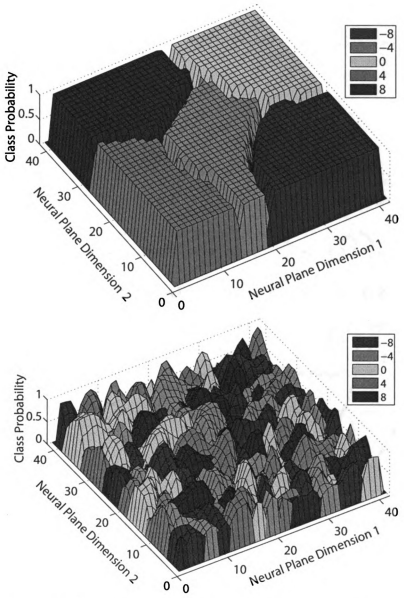


Figure 6.3: The class probability of the 40×40 neurons of the feature-detection cortex. (a) Top-down connections are active ($\alpha = 0.5$) during development. (b) Top-down connections are not active ($\alpha = 0$) during development.

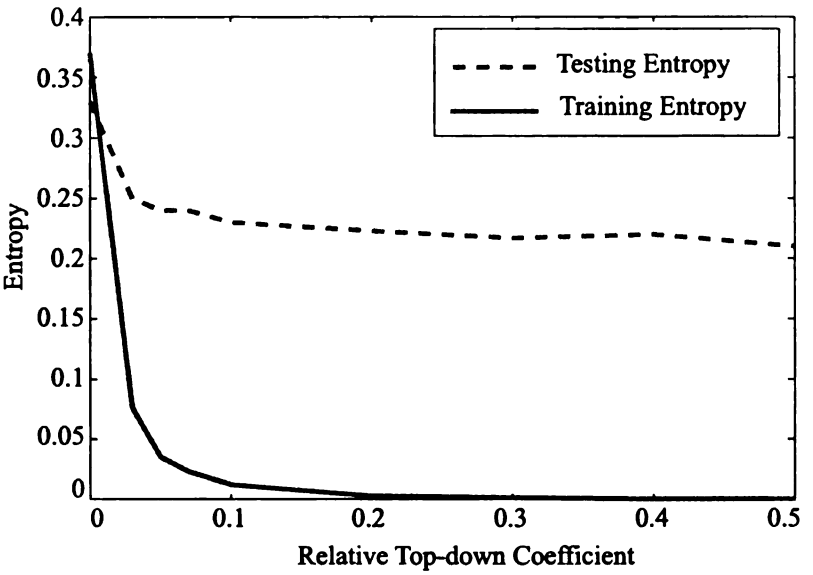


Figure 6.4: The effect of top-down projection on the purity of the neurons and the performance of the network. Increasing α in Eq. 4.1 results in purer neurons and better performance.

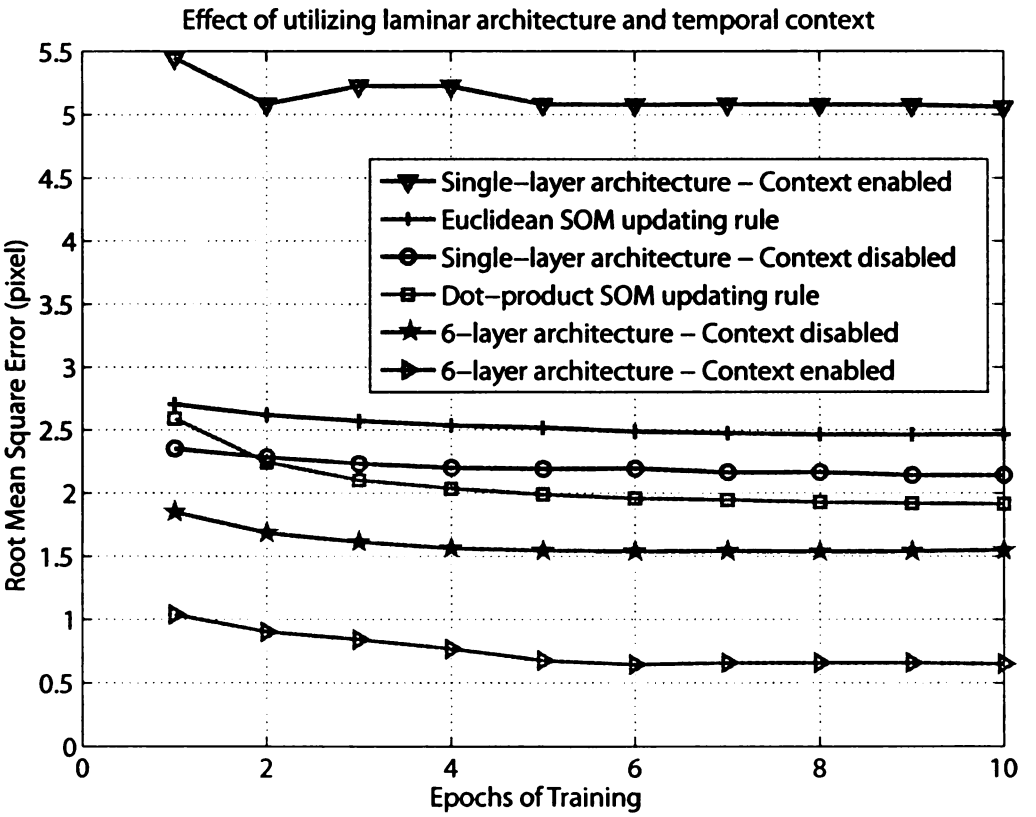


Figure 6.5: How temporal context signals and 6-layer architecture improve the performance.

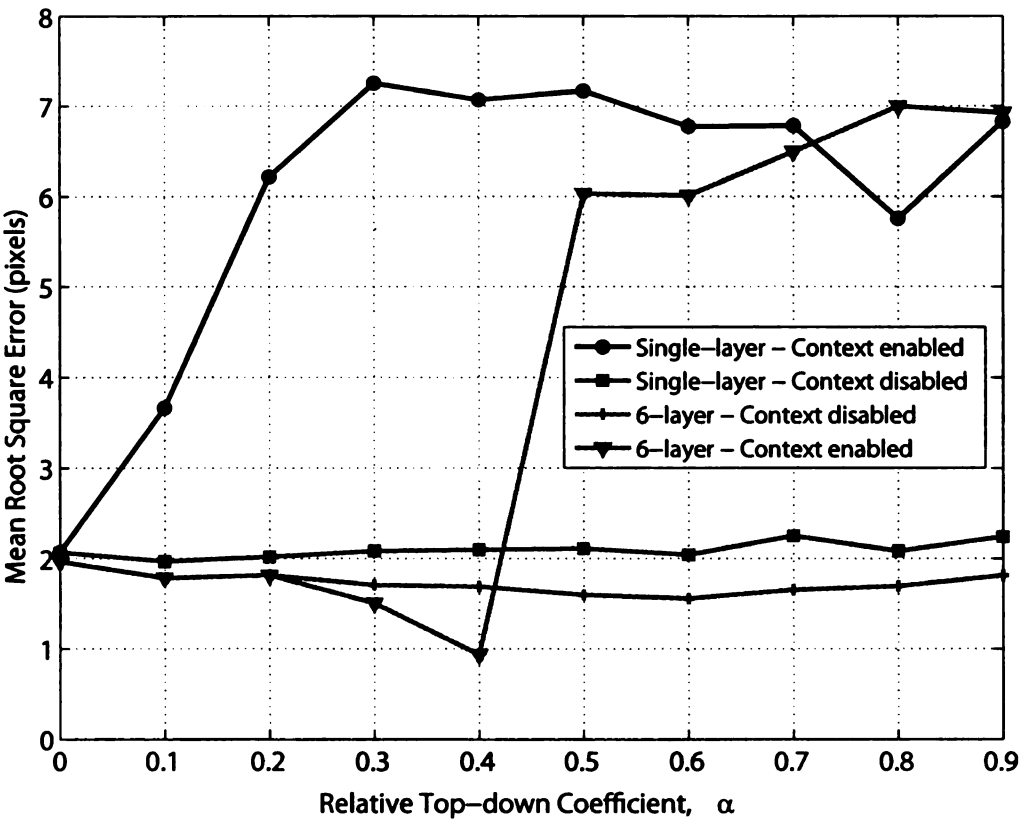


Figure 6.6: The effect of relative top-down coefficient, α, on performance in disjoint recognition test on randomly selected training data.

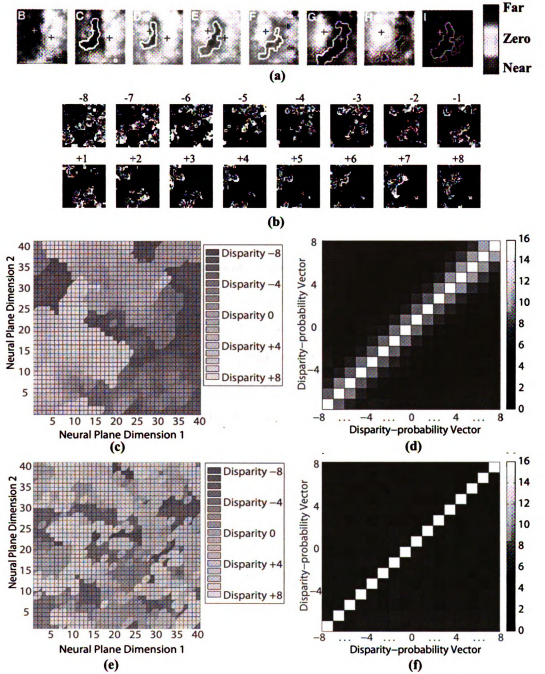


Figure 6.7: **(a)** Map of neurons in V2 of macaque monkeys evoked by stimuli with 7 different disparities. Adapted from Chen et. al. 2008 [7] **(b)** Disparity-probability vectors of $L2/3$ neurons for different disparities when $\kappa = 5$. **(c,e)**. Disparity-probability maps in $L2/3$ where $\kappa = 5$ in (c) and $\kappa = 1$ (e). **(d,f)**. Cross-correlation of disparity-probability where $\kappa = 5$ in (d) and $\kappa = 1$ in (f).

Chapter 7

Conclusion

The lack of computational experiments on real-world data in previous works has led to the oversight of the role of sparse coding in neural representation in the models of laminar cortex. Sparse coding of the input is computationally advantageous both for bottom-up and top-down input, specially when the input is complex. Therefore, we hypothesize that the cortical circuits probably have a mechanism to sparsely represent top-down and bottom-up input. Our model suggests that the brain computes a sparse representation of bottom-up and top-down input independently, before it integrates them to decide the output of the cortical region. Thus, we predict that:

Prediction 1: What is known as Layer 2/3 in cortical laminar architecture [1] has two functional roles:

1. Rank and scale the top-down energy received at the cortex (modulated by signals from $L5$)

2. Integrate the modulated bottom-up energy received from $L4$ to the modulated top-down energy received from higher cortical areas to determine the output signals of the cortex

[1] Marked as *Level2* , layers 2 through $4B$ in [5] Figure 2.

Neuroscientists have known for a long time that there are sublayers in the laminar cortex [23]. However, the functionality of these sublayers has not been modeled before. This is a step towards understanding the sublayer architecture of the laminar cortex. Our prediction breaks down the functionality of $L2/3$ to two separate tasks. This is different from the previous models (e.g. [5]), as they consider $L2/3$ as one functional layer.

Fig. 7.1 illustrates the result of an experiment in which we compared two models of $L2/3$. In the traditional model of $L2/3$, it is modeled as one functional layer that integrates the sparse coded signals received from $L4$ with the top-down energy, while in our novel model used in this thesis, $L2/3$ functions as 2 functional layers (see Prediction 1).

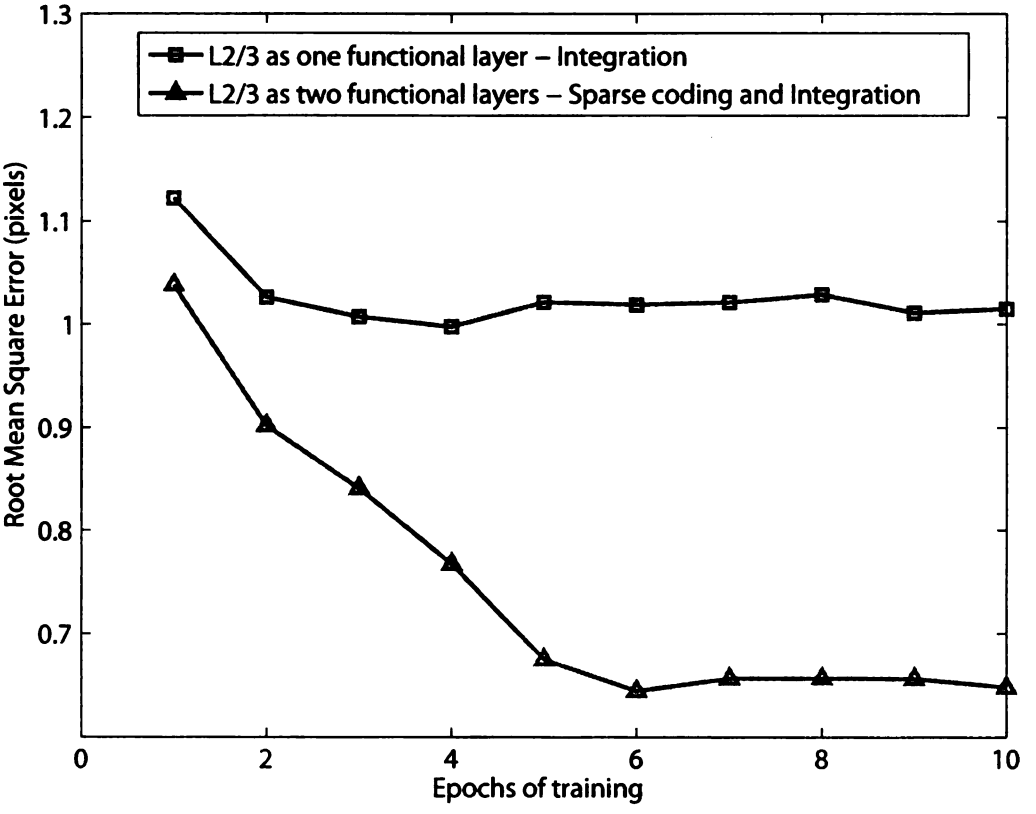


Figure 7.1: Comparison of our novel model of $L2/3$ where it performs both sparse coding and integration of top-down and bottom-up signals, with traditional models in which it only does integration.

Presented is the first spatio-temporal model of the 6-layer architecture of the cortex which incorporated temporal aspects of the stimuli in the form of top-down context signals. It outperformed simpler single-layer models of the cortex by a significant amount. Furthermore, defining the problem of binocular disparity detection as a regression problem by training a few nearby neurons to relate to the presented stimuli (as opposed to only one neuron in the case of classification), resulted in biologically-observed smoothly changing disparity sensitivity along the neural layers.

Since the brain generates actions through numerical signals (spikes) that drive muscles and other internal body effectors (e.g. glands), regression (output signals) seems closer to what the brain does, compared to many classification models that have been published in the literature. The regression extension of the MILN [50] has potentially a wide scope of application, from autonomous robots to machines that can learn to talk. A major open challenge is the complexity of the motor actions to be learned and autonomously generated.

As presented here, an emergent-representation based binocular system has shown disparity detection abilities with sub-pixel accuracy. In contrast with engineering methods that used explicit matching between the left and right search windows, a remarkable computational advantage of our work is the potential for integrated use of a variety of image information for tasks that require disparity as well as other visual cues.

Our model suggests a computational reason as to why there is no top-down connection from $L2/3$ to $L4$ in laminar cortex; to prevent the top-down and bottom-up energies received at the cortex from mixing before they internally compete to sort out winners. Hence, we predict that the thick layer $L2/3$ in laminar cortex carries out more functionality than what has been proposed in previous models - it provides sparse representation for top-down stimuli, combines the top-down and

bottom-up sparse representations and projects the output of the cortical region to higher cortices.

Utilization of more complex temporal aspects of the stimuli and using real-time stereo movies will be a part of our future work.

Appendix A

Neuronal Weight Updating

For a winner cell i, update the weights using the lobe component updating principle [52]. That reference also provides a theoretical perspective on the following. Each winner neuron updates using the neuron's own internal temporally scheduled plasticity as $\mathbf{w}_{b,i}(t) = \beta_1 \mathbf{w}_{b,i}(t-1) + \beta_2 z_i \mathbf{b}(t)$ where the scheduled plasticity is determined by its two age-dependent weights:

$$\beta_1 = \frac{m_i - 1 - \mu(m_i)}{m_i}, \beta_2 = \frac{1 + \mu(m_i)}{m_i}, \tag{A.1}$$

with $\beta_1 + \beta_2 \equiv 1$. Finally, the cell age (maturity) m_i for the winner neurons increments: $m_i \leftarrow m_i + 1$. All non-winners keep their ages and weight unchanged. In Eq. (A.1), $\mu(m_i)$ is the plasticity function depending on the maturity m_i of neuron i. The neuron maturity increments every time a neuron updates its weights, starting from zero. The plasticity function prevents learning rate from converging to zero. Details are presented in [52].

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Ohzawa I. Freeman R.D. Anzai, A. Neural mechanisms underlying binocular fusion and stereopsis: position v/s phase. In *Proc. Natl. Acad. Sci.*, pages 5438–5443, 1997.

[2] Blakemore C. Pettigrew J.D. Barlow, H.B. The neural mechanisms of binocular depth discrimination. *J. Physiol.*, 193:327342, 1967.

[3] P.O. Bishop. Vertical disparity, egocentric distance and stereoscopic depth constancy: a new interpretation. In *Proc. R. Soc. London Ser.*, pages 445–469, 1989.

[4] W. H. Bosking, Y. Zhang, B. Shoefield, and D. Fitzpatrick. Orientation selectivity and arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, 17:2112–2127, 1997.

[5] E. M. Callaway. Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21:47–74, 1998.

[6] Edward M. Callaway. Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw.*, 17(5-6):625–632, 2004.

[7] G. Chen, H. D. Lu, and A. W. Roe. A map for horizontal disparity in monkey v2. *Neuron*, 58(3):442–450, May 2008.

[8] D. B. Chklovskii and A. A. Koulakov. Maps in the brain: What can we learn from them? *Annual Review of Neuroscience*, 27:369–392, 2004.

[9] B. Cumming. Stereopsis: how the brain sees depth. *Current Biology*, 7(10):645–647, 1997.

[10] Y. Dan and M. Poo. Spike timing-dependent plasticity: From synapses to perception. *Physiological Review*, 86:1033–1048, 2006.

[11] G.C. DeAngelis. Seeing in three dimensions: the neurophysiology of stereopsis. *Trends in Cognitive Sciences*, 4(3), 2000.

[12] U. R. Dhond and J. K. Aggarwal. Structure from stereo - a review. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(6):1489–1510, Nov./Dec. 1989.

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

[14] D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin. Phase-based disparity measurement. In *CVGIP: Image Understand.*, volume 53, pages 198–210, 1991.

[15] A. Franz and J. Triesch. Emergence of disparity tuning during the development of vergence eye movements. In *International Conference on Development and Learning*, pages 31–36, 2007.

[16] Poggio GF and Fischer B. Binocular interaction and depth sensitivity of striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol*, 40:13921405, 1977.

[17] C.D. Gilbert and T.N. Wiesel. Microcircuitry of the visual cortex. *Annu. Rev. Neurosci.*, 6:217247, 1983.

[18] F. Gonzales and Perez R. Neural mechanisms underlying stereoscopic vision. *Progress in Neurobiology*, 55:191–224, 1998.

[19] W. E. L. Grimson. *From Images to Surfaces: A Computational Study of the Human Early Visual System.* MIT Press, 1981.

[20] T. Luwang J. Weng, H. Lu and X. Xue. A multilayer in-place learning network for development of general invariances. *International Journal of Humanoid Robotics*, 4(2), 2007.

[21] T. Burwick J. Wiemer and W. Seelen. Self-organizing maps for visual feature representation based on natural binocular stimuli. *Biological Cybernetics*, 82(2):97–110, 2000.

[22] B. Julesz. Foundations of cyclopean perception. 1971. University of Chicago Press: Chicago.

[23] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, editors. *Principles of Neural Science.* Appleton & Lange, Norwalk, Connecticut, 3rd edition, 1991.

[24] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, editors. *Principles of Neural Science.* McGraw-Hill, New York, 4th edition, 2000.

[25] T. Kohonen. *Self-Organizating Maps.* 1997.

[26] S. R. Lehky and T. J. Sejnowski. Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *The Journal of Neuroscience*, 70(7):2281–2299, July 1990.

[27] J. Lippert, D. J. Fleet, and H. Wagner. Disparity tuning as simulated by a neural net. *Journal of Biocybernetics and Biomedical Engineering*, 83:61–72, 2000.

[28] M. D. Luciw and J. Weng. Motor initiated expectation through top-down connections as abstract context in a physical world. In *Proc. 7th International Conference on Development and Learning (ICDL'08)*, 2008.

[29] M.D. Luciw and J. Weng. Topographic class grouping with applications to 3d object recognition. In *Proc. International Joint Conf. on Neural Networks*, Hong Kong, June 2008. accepted and to appear.

[30] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman, New York, 1982.

[31] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh. *Computational Maps in the Visual Cortex.* Springer, Berlin, 2005.

[32] Bishop P. O. Nikara, T. and J. D. Pettigrew. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp. Brain Res.*, 6:353372, 1968.

[33] National Eye Institute [NEI] of the U.S. National Institute of Health. `http://www.nei.nih.gov/photo` (first visited 04/24/09).

[34] DeAngelis G.C. Freeman R.D. Ohzawa, I. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249:10371041, 1990.

[35] A. J. Parker. Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, 8(5):379–391, 2007.

[36] G. F. Poggio, F. Gonzalez, and F. Krause. Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *JNeuSci*, 8:4531–4550, December 1988.

[37] N. Qian. Binocular disparity and the perception of depth. *Neuron*, 18:359368, 1997.

[38] R. D. Raizada and S. Grossberg. Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cereb Cortex*, 13(1):100–113, January 2003.

[39] T. Ramtohul. A self-organizing model of disparity maps in the primary visual cortex. Master's thesis, School of Informatics, University of Edinburgh, 2006.

[40] J.C.A. Read. Early computational processing in binocular vision and depth perception. *Progress in Biophysics and Molecular Biology 87*, pages 77–108, 2005.

[41] Jenny C A C. Read and Bruce G G. Cumming. Sensors for impossible stimuli may solve the stereo correspondence problem. *Nat Neurosci*, September 2007.

[42] Parker A.J. Cumming B.G. Read, J.C.A. A simple model accounts for the reduced response of disparity-tuned v1 neurons to anti-correlated images. *Vis. Neurosci.*, 19:735753, 2002.

[43] A. W. Roe, A. J. Parker, R. T. Born, and G. C. DeAngelis. Disparity channels in early vision. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(44):11820–11831, October 2007.

[44] P. R. Roelfsema and A. van Ooyen. Attention-gated reinforcement learning of internal representations for classification. *Journal of Neural Computation*, 17:2176–2214, 2005.

[45] Y. F. Sit and R. Miikkulainen. Self-organization of hierarchical visual maps with feedback connections. *Neurocomputing*, 69:1309–1312, 2006.

[46] M. Solgi and J. Weng. Developmental stereo: Topographic iconic-abstract map from top-down connection. In *Proc. the First of the Symposia Series New developments in Neural Networks (NNN'08)*, 2008.

[47] J. Weng. Image matching using the windowed Fourier phase. *International Journal of Computer Vision*, 11(3):211–236, 1993.

[48] J. Weng and M. D. Luciw. Optimal in-place self-organization for cortical development: Limited cells, sparse coding and cortical topography. In *Proc. 5th International Conference on Development and Learning (ICDL'06)*, Bloomington, IN, May 31 - June 3 2006.

[49] J. Weng and M. D. Luciw. Dually optimal neural layers: Lobe component analysis. *IEEE Transaction on Autonomous Mental Development*, 1, 2009.

[50] J. Weng, T. Luwang, H. Lu, and X. Xue. Multilayer in-place learning networks for modeling functional layers in the laminar cortex. *Neural Networks*, 21:150–159, 2008.

[51] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.

[52] J. Weng and N. Zhang. Optimal in-place learning and the lobe component analysis. In *Proc. World Congress on Computational Intelligence*, Vancouver, Canada, July 16-21 2006.

[53] Peter Werth, Stefan Scherer, and Axel Pinz. Subpixel stereo matching by robust estimation of local distortion using gabor filters. In *CAIP '99: Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns*, pages 641–648, London, UK, 1999. Springer-Verlag.

[54] A. K. Wiser and E. M. Callaway. Contributions of individual layer 6 pyramidal neurons to local circuitry in macaque primary visual cortex. *Journal of neuroscience*, 16:2724–2739, 1996.

[55] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684, Jul. 2000.