

PERFORMANCE, DEVELOPMENT, AND ANALYSIS
OF TACTILE VS. VISUAL RECEPTIVE FIELDS IN TEXTURE TASKS

A Dissertation

by

CHOON SEOG PARK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Computer Engineering

PERFORMANCE, DEVELOPMENT, AND ANALYSIS
OF TACTILE VS. VISUAL RECEPTIVE FIELD IN TEXTURE TASKS

A Dissertation

by

CHOON SEOG PARK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

| | |
|-------------------------|----------------------------------|
| Co-Chairs of Committee, | Yoonsuck Choe Dezhen Song |
| Committee Members, | Eun Jung Kim J. Maurice Rojas |
| Head of Department, | Valerie E. Taylor |

August 2009

Major Subject: Computer Engineering

ABSTRACT

Performance, Development, and Analysis
of Tactile vs. Visual Receptive Field in Texture Tasks. (August 2009)

Choon Seog Park, B.S, Korea Military Academy, Seoul, Korea;

M.S., Yonsei University, Seoul, Korea

Co-Chairs of Advisory Committee: Dr. Yoonsuck Choe
Dr. Dezhen Song

Texture segmentation is an effortless process in scene analysis, yet its neural mechanisms are not sufficiently understood. A common assumption in most current approaches is that texture segmentation is a vision problem. However, considering that texture is basically a surface property, this assumption can at times be misleading. One interesting possibility is that texture may be more intimately related with touch than with vision. Recent neurophysiological findings showed that receptive fields (RFs) for touch resemble that of vision, albeit with some subtle differences. To leverage on this, here I propose three ways to investigate the tactile receptive fields in the context of texture processing: (1) performance, (2) development, and (3) analysis.

For performance, I tested how such distinct properties in tactile receptive fields can affect texture segmentation performance, as compared to that of visual receptive fields. Preliminary results suggest that touch has an advantage over vision in texture segmentation. These results support the idea that texture is fundamentally a tactile (surface) property.

The next question is what drives the two types of RFs, visual and tactile, to become different during cortical development? I investigated the possibility that tactile RF and visual RF emerge based on the same cortical learning process, where the only difference is in the input type, natural-scene-like vs. texture-like. The main

result is that RFs trained on natural scenes develop RFs resembling visual RFs, while those trained on texture resemble tactile RFs. These results again suggest a tight link between texture and the tactile modality, from a developmental context.

To investigate further the functional properties of these RFs in texture processing, the response of tactile RFs and visual RFs were analyzed with manifold learning and with statistical approaches. The results showed that touch-based manifold seems more suitable for texture processing and desirable properties found in visual RF response can carry over to those in the tactile domain.

These results are expected to shed new light on the role of tactile perception of texture; help develop more powerful, biologically inspired texture segmentation algorithms; and further clarify the differences and similarities between touch and vision.

To my lovely wife, Insook Lee

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor and mentor, Dr. Yoonsuck Choe, for his guidance, support, and patience throughout my Ph.D. study. I appreciate his efforts and encouragement, numerous discussions, and valuable feedback in the last five years.

I would also like to thank all of my committee, Dr. Dezhen Song (co-chair) and Dr. Eun Jung Kim in the Department of Computer Science and Engineering, and Dr. Maurice Rojas (and Dr. Matthew Young) in the Department of Mathematics at Texas A&M University, for their contribution to my research and for their help, input, and motivation during the past few years.

I am also grateful to all the members of the Neural Intelligence Lab for the intriguing discussions which helped broaden my research insight. Special gratitude goes to Heeyoul Choi and Yoonho Bai for the numerous discussions and suggestions to my research and for their friendship as lab colleagues.

My special appreciation goes to my lovely wife, Insook Lee, and my sons, Yongtae Park and Yongjin Park. They helped me to concentrate on completing this dissertation and supported me mentally during the course of this work. Without their help and encouragement, this study would not have been completed. Finally, I would like to thank my mother and brother for their affection, support, and absolute confidence in me.

This research was supported in part by NIH/NINDS grant #R01-NS54252, and by the Korean Army

TABLE OF CONTENTS

| CHAPTER | | Page |
|---------|---|------|
| I | INTRODUCTION | 1 |
| | A. Motivation | 1 |
| | B. The Main Research Question: Tactile or Visual? | 2 |
| | C. Approach | 3 |
| | 1. Performance | 3 |
| | 2. Development | 4 |
| | 3. Statistical Analysis | 6 |
| | D. Outline of the Dissertation | 6 |
| II | BACKGROUND | 8 |
| | A. Nature of Texture | 8 |
| | B. Texture Analysis Approaches | 10 |
| | C. The Spatiotemporal Organization of Visual Receptive Fields | 13 |
| | D. The Cortical Mechanism of the Area 3b in the Primary Somatosensory Cortex | 18 |
| | E. Summary | 25 |
| III | PERFORMANCE: TACTILE VS. VISUAL RECEPTIVE FIELDS IN TEXTURE SEGMENTATION | 29 |
| | A. Visual or Tactile Perception for Texture | 29 |
| | B. Computational Model for Tactile RFs: The Three-Component Model | 30 |
| | C. Texture Boundary Detection Tasks with TRFs and VRFs . | 36 |
| | D. Discussion | 41 |
| IV | DEVELOPMENT: DIFFERENTIAL DEVELOPMENT OF TACTILE AND VISUAL RFS | 43 |
| | A. Input-driven Development of Receptive Fields | 43 |
| | B. The LISSOM and Self-Organization | 46 |
| | C. Spatiotemporal Emergence of TRFs and VRFs | 51 |
| | D. Discussion | 59 |
| V | MANIFOLD ANALYSIS OF FEATURE SPACE IN TAC- TILE VS. VISUAL RFS | 61 |

| CHAPTER | Page |
|---|------|
| A. Manifold Learning in Perception | 61 |
| B. Embedded Manifold in RF Response Feature Space | 64 |
| C. Texture Classification with Manifold | 69 |
| D. Texture Classification with Manifold of Integrated RF Responses | 75 |
| E. Discussion | 76 |
| VI STATISTICAL ANALYSIS OF RF RESPONSE | 79 |
| A. Power-law Distribution | 79 |
| B. Suspicious Coincidence | 82 |
| C. Experiments and Results | 86 |
| D. Discussion | 88 |
| VII DISCUSSION | 92 |
| A. Contribution | 92 |
| B. Discussion | 92 |
| 1. Potential Limitations of the Approach | 92 |
| 2. Potential Criticism | 95 |
| C. Future Directions | 98 |
| VIII CONCLUSION | 99 |
| REFERENCES | 101 |
| VITA | 111 |

LIST OF TABLES

| TABLE | Page |
|--|------|
| I Comparison of Texture Classification with Popular Classifiers. | 97 |

LIST OF FIGURES

| FIGURE | | Page |
|--------|--|------|
| 1 | Tactile Property in 3D Space. | 2 |
| 2 | Visual Receptive Field (VRF, left) and Tactile Receptive Field (TRF, right). | 3 |
| 3 | Architecture of the Basic LISSOM Model. | 5 |
| 4 | Sample Textures. | 8 |
| 5 | Textural Cue in the Natural Scene with Zebras. | 10 |
| 6 | Discrimination of Texture Pair with X, T, and L. | 11 |
| 7 | Human Visual Pathways (top view). | 13 |
| 8 | Receptive Field Types in Retina, LGN, and V1. | 15 |
| 9 | Gabor Filter Bank. | 16 |
| 10 | Spatiotemporal Receptive Fields in Cat. | 17 |
| 11 | Vertically Striped Tube Used in Selective Rearing Experiments. | 19 |
| 12 | Somatosensory Areas of the Cortex. | 21 |
| 13 | Columnar Organization of Area 3b in S1. | 22 |
| 14 | Drum Stimulator to Simulate Tactile Texture. | 24 |
| 15 | Three-component Gaussian Model of Tactile Receptive Field. | 26 |
| 16 | Tactile Receptive Fields in Monkey Area 3b. | 27 |
| 17 | Oriented Gabor Filters. | 31 |
| 18 | Estimated tactile receptive fields with 8 scanning directions. | 33 |

| FIGURE | Page |
|--------|--|
| 19 | Six Texture Sets. 35 |
| 20 | Sample Textures and Their Typical Response Profiles at Scanning Direction of 315° 36 |
| 21 | 3D Visualization of RF Responses. 37 |
| 22 | Texture Classification Experiment. 38 |
| 23 | Comparison of Average Classification Rate for Six Different Tex- ture Sets with TRFs and VRFs. 39 |
| 24 | Comparison of Classification Rate for Curvy and Linear Textures. . . 40 |
| 25 | Linear Discriminant (LD) Distributions of TRF and VRF Responses. 41 |
| 26 | Frequency Distribution of LDA of TRF and VRF Responses. 42 |
| 27 | Visual Receptive Fields (VRFs). 44 |
| 28 | Tactile Receptive Fields (TRFs). 45 |
| 29 | LISSOM Model of Orientation and Direction Selectivity. 48 |
| 30 | Neuron Activation Function $\sigma(s)$ 50 |
| 31 | Sample Input Patterns. 52 |
| 32 | Generation of Dynamic Input by Scanning the Gaze on an Image. . . 53 |
| 33 | Self-organization Process with LISSOM. 54 |
| 34 | Six Spatiotemporal RFs Resulting from Self-Organization on the Natural Scene Input Set. 55 |
| 35 | Detailed View of RFs Resulting from Self-Organization on the Natural Scene Input Set. 56 |
| 36 | Six Spatiotemporal RFs Resulting from Self-Organization on the Texture Input Set. 57 |

| FIGURE | Page |
|--------|---|
| 37 | Detailed View of RFs Resulting from Self-Organization on the Texture Input Set. 57 |
| 38 | Orientation Maps. 58 |
| 39 | Selectivity in Orientation Maps. 59 |
| 40 | The Swiss Roll Data Set Illustrating Nonlinear Dimensionality Reduction. 62 |
| 41 | Manifold Learning in Visual Perception. 63 |
| 42 | Manifold Learning in Tactile Perception. 65 |
| 43 | Eigenvalues Extracted from the Embedded Manifold of Designed Filters and Self-organized Receptive Fields. 68 |
| 44 | Projection of 4 Different Types of Filters Mapped onto the Embedded Manifold of Two Dimensions. 70 |
| 45 | Texture Classification Process with Visual and Tactile Manifolds. . . 71 |
| 46 | Feature Spaces for Texture. 73 |
| 47 | Feature Spaces for Natural Scene. 73 |
| 48 | Comparison of Texture Classification Rate between TRF-based Response and VRF-based Response with Texture-like Input. 74 |
| 49 | Comparison of Natural-scene Classification Rate between TRF-based Response and VRF-based Response with Natural-scene-like Input. 75 |
| 50 | Feature Spaces for Texture from the Manifold of the Integrated RF Responses. 76 |
| 51 | Comparison of Texture Classification Rate between TRF-based Response, VRF-based Response, and Integrated Response with Texture-like input. 77 |
| 52 | Gray-level Intensity Histogram of Natural Images. 80 |

| FIGURE | Page |
|--------|---|
| 53 | Orientation Energy (E) Distribution. 80 |
| 54 | Power-law vs. Baseline Gaussian Distribution. 81 |
| 55 | Visual Responding Thresholding in a Natural Image. 82 |
| 56 | Cortical Response on Input Images. 83 |
| 57 | Response Distribution of Input Images. 83 |
| 58 | Suspicious Coincidence between Image Pixels. 84 |
| 59 | Orientation Energy Distribution of an Image vs. Its Matching Normal Distribution. 86 |
| 60 | Process of Statistical Analysis. 87 |
| 61 | Comparison of the Distribution of RF Response to the Same Input Type used During Training. 88 |
| 62 | Comparison of the Distribution of RF Response to Opposite Input Type to the Training Input Type. 89 |
| 63 | Orientation Energy Threshold Selected by Human vs. $L2$ 90 |
| 64 | Threshold ($L2$) vs. Raw Second Moment (σ) of the Orientation Energy Distribution. 91 |
| 65 | A Somatotopic Map of the Body Surface in the Primary So- matosensory Cortex and the Sensory homunculus. 93 |
| 66 | Prenatal Orientation Maps in Animals and in HLISSOM. 94 |
| 67 | Emerging Pinwheel Organizations of Angular Detection Preference. . 96 |

CHAPTER I

INTRODUCTION

A. Motivation

Humans process sensory information from different specialized modalities (e.g., vision, touch, and hearing), yet relatively little is known about how specific input stimuli affect the cortical organization. Texture is an interesting class of stimulus. Textural patterns have been studied as important cues that help form the sensory cortex [1]. The human visual system also uses texture information in order to automatically or preattentively segregate parts of the visual scene [2]. Several theories and algorithms exist for texture discrimination based on vision [3, 4]. These models differ from one another in their algorithmic approaches in dealing with texture using the embedded spatial elements and their statistics. Even though there are differences among these approaches, they all begin from the fundamental assumption that texture segmentation is a visual task.

However, considering that texture is basically a surface property, this assumption can at times be misleading [5]. An interesting possibility is that since surface properties are most immediately accessible to touch, tactile perception may be more intimately associated with texture than with vision (it is known that tactile input can affect vision [6]). Fig. 1 shows that texture is basically a surface property in 3D space [5]. Textures on the 3D surfaces of the stone and the bark could be processed as textures in 2D when one of them occludes the other.

Interestingly, the basic organization of the tactile (somatosensory) system bears some analogy to that of the visual system [7]. Sensory neurons in the primary sensory

The dissertation follows the style of *IEEE Transactions on Automatic Control*.



Fig. 1. Tactile Property in 3D Space. Textures on the 3D surfaces of a stone and a piece of bark could be processed as textures in 2D when one of them occludes the other (see [5]).

cortices in the brain preferentially respond to specific patterns of input at a specific spatial location (this is the concept of “receptive field”). For example, neurons in the primary visual cortex (area 17, V1) have oriented Gabor-like receptive fields (RFs) [8]. In area 3b of the somatosensory cortex, neurons respond to tactile input from the finger tip, and just like in the visual cortex, they only respond to a specific pattern of input [9]. In particular, recent neurophysiological findings showed that receptive fields for touch resemble that of vision, albeit with some subtle differences [7, 9, 10]. Fig. 2 shows an example of a visual receptive field and a tactile receptive field [8], [9], respectively. The two receptive fields have common features, an excitatory component and an inhibitory component, while tactile receptive fields have one more component which is a dynamic inhibitory component.

B. The Main Research Question: Tactile or Visual?

Based on the above motivation, the main hypothesis of this dissertation is that touch may be a better, natural way for texture processing than vision. To test this hypothesis, we can ask three research questions: (1) Is touch-based performance better than vision-based performance for texture tasks? (2) What determines the differential development of tactile and visual receptive fields in the cortex? (3) Is tactile feature

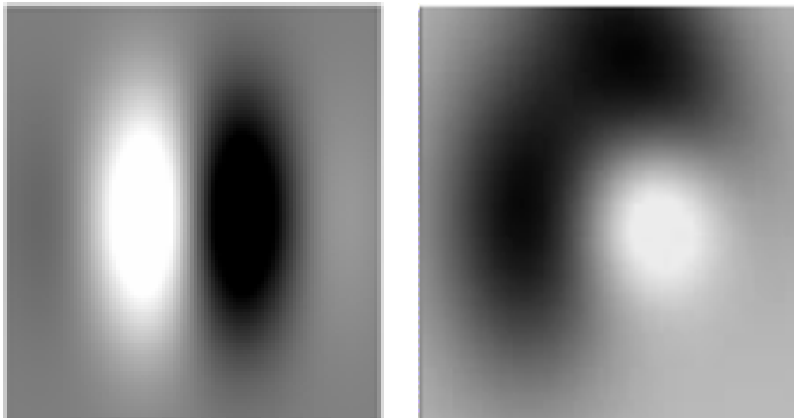


Fig. 2. Visual Receptive Field (VRF, left) and Tactile Receptive Field (TRF, right). Dark areas represents inhibitory region and bright areas excitatory region in each RF. The two receptive fields have common features, i.e., excitatory component (bright region) and inhibitory component (dark region), while the tactile receptive field has one more component (dark upper region) which is a dynamic inhibitory component.

representation more powerful than visual representation for texture analysis?

C. Approach

The approach of my research is as follows. (1) Performance: Test if tactile RF-based performance is better than visual RF-based performance in texture segmentation task. (2) Development: Test how tactile or visual stimulus characteristics (texture-like and natural-scene-like) can determine receptive field type in the cortex. (3) Statistical analysis: Analyze if tactile feature representation is more powerful than visual representation for texture analysis.

1. Performance

Even though tactile and visual RFs have some shared properties, there is a difference between the tactile receptive fields and the visual receptive fields. Instead of an ex-

citation/inhibition pair as in Gabor patterns, there is an extra third component that is inhibitory, where the position of that component dynamically changes depending on the direction of scan of the tactile patch. In chapter III I tested how such distinct properties in tactile receptive fields can affect texture segmentation performance, as compared to those of visual receptive fields. Even though visual RFs in the primary visual cortex (V1) have been a main inspiration in texture processing, tactile RFs in the primary somatosensory cortex (S1) could be more appropriate in texture tasks. I used two computational models: (1) Gabor-like filter for V1 and (2) the three-component model for S1. Statistical measures and standard back-propagation networks were used to test the performance of the RFs in texture processing.

2. Development

In chapter IV, the developmental process of the tactile receptive field in the cortex was examined to better understand perceptual learning of touch. I investigated how visual receptive fields (VRFs) and tactile receptive fields (TRFs) can self-organize and if the input type, (1) natural scene and (2) texture, play a key role. The main hypothesis is that TRFs can be self-organized using a visual cortical development model by simply exposing it to texture-like inputs. In order to test the hypothesis, I used the LISSOM (Laterally Interconnected Synergetically Self-Organizing Map) model which was originally developed to model the self-organization of the visual cortex [11] (Fig. 3 shows the basic architecture of the LISSOM model). However, the LISSOM model is actually a more general model of how the cortex in general organizes to represent correlations in the sensory input, regardless of the input modality, thus it could be used to model the development of other sensory cortices (e.g., see [12], for a model of the somatosensory cortex).

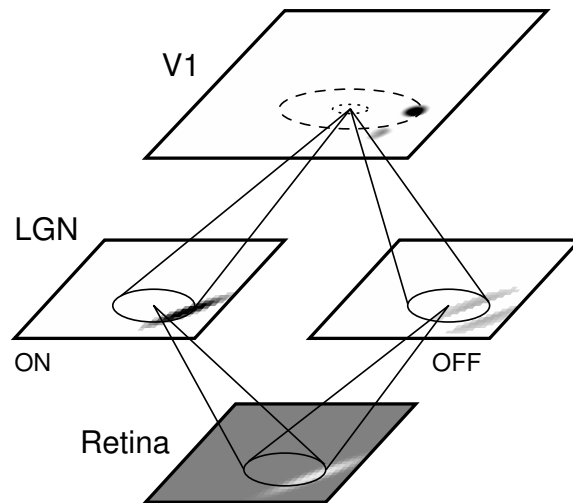


Fig. 3. Architecture of the Basic LISSOM Model. LISSOM has a hierarchical structure of two-dimensional sheets of neural units, including an array of retinal receptors, lateral geniculate nucleus (LGN) ON and OFF channels, and the primary visual cortex (V1). Gray scale input is used for all units that the LGN and V1 activation is shown in gray-scale coding from white to black (low to high) and the activity on the retina is presented like natural images: Light areas are strongly activated, dark areas are weakly activated, and medium gray represents background activation. The LGN afferent connections are formed from the retina and V1 afferent connections from each LGN sheet. V1 neurons also receive lateral excitatory and lateral inhibitory connections from nearby V1 neurons; these connections are shown as dotted and dashed circles around the V1 neuron, respectively. Adapted from [11].

3. Statistical Analysis

Based on results from the performance and the development of tactile vs. visual RFs, I will analyze the functional structure of the feature space of tactile and visual RF responses and the response distributions of these RFs to different input types. In chapter V, two manifold learning methods (kernel-Isomap [13] and kernel-FDA [14]) will be used to analyze the learned RFs from LISSOM and to experiment with the classification of textures. In chapter VI, the concept of power-law distribution [15] and suspicious coincidence [16] will be used to statistically analyze the response distribution. The results show that tactile manifold is more suitable for texture processing and the properties of tactile RF response distribution share common characteristics as those found in the visual domain.

D. Outline of the Dissertation

The rest of this dissertation is organized as follows.

Chapter II will present biological background and survey the empirical literature that form the basis for this research. The organization of the somatosensory system and the neural pathway of the tactile response in the cortex will be compared to those in the visual system.

In chapter III, I will show the relative advantage of tactile exploration of texture compared to that of vision. The three component model adapted from [9] which is the tactile counterpart of the V1 simple cell model was used in texture boundary detection tasks, and the results compared to those with visual receptive fields. In chapter IV I will investigate the possibility that tactile RF and visual RF emerge based on an identical cortical learning process, where the only difference is in the input type, natural-scene-like vs. texture-like. Finally, for the analysis part, (1) manifold learning

(chapter V) and (2) statistical analysis (chapter VI) of the receptive fields and their responses will be carried out.

In chapter VII, contributions and potential limitations of my work will be discussed, followed by future work.

Finally, I will conclude by highlighting the main novelties and contributions.

CHAPTER II

BACKGROUND

In this chapter, I will present some basic biological concepts from the empirical literature that form the basis of this investigation. The concept of texture and texture recognition will be presented first, followed by the spatiotemporal nature of visual receptive fields, and finally the cortical mechanism of the area 3b in the primary somatosensory cortex.

A. Nature of Texture

The natural world is rich in texture: any visible object is textured at a certain scale. A wealth of texture is observed on both artificial and natural objects such as wood, plants, materials, and skin. Fig. 4 shows a few natural and man-made textures, respectively, which could be encountered in daily life.

Although texture is an important research area in computer vision, there is no precise definition of texture. The main reason for this is that natural textures often display different yet contradicting properties, such as regularity versus randomness, uniformity versus distortion, which can hard to describe in a unified manner. Many



Fig. 4. Sample Textures. Natural textures (top row) and man-made textures (bottom row) are shown.

researchers have tried to define texture from a certain perspective.

Haralick considers texture as an “organized area phenomenon” which can be decomposed into ‘primitives’ having specific spatial distributions [17]. This definition, also known as the structural approach, comes directly from the human visual experience of textures. For instance, each texture in Fig. 4 is composed of particular texture elements, e.g., objects (windows), shapes (jigsaw pieces), or simply color patterns. Meanwhile, these primitives are organized in a particular spatial structure indicating certain underlying placement rules. Alternatively, as Cross and Jain suggested, texture is “a stochastic, possibly periodic, two-dimensional image field” [18]. This definition, which is also known as the stochastic approach, describes texture by a stochastic process that generates the texture. These different definitions usually lead to different computational approaches in texture analysis.

In a different sense, the word texture refers to surface characteristics and appearance of an object given by the size, shape, density, arrangement, and proportion of its elementary parts [19]. A texture is usually described as smooth or rough, soft or hard, coarse, or fine, matt or glossy, etc. Fig. 1 shows how an apparently 2D problem of texture segmentation could be caused by two separate 3D objects [5].

Thus, it seems that there are at least two ways to characterize texture: tactile and visual. Tactile textures refer to the immediate tangible quality of a surface. Visual textures refer to the visual impression relating to local spatial variations of simple stimuli like color, orientation and intensity in an image. This dissertation will evaluate both points of view.



Fig. 5. Textural Cue in the Natural Scene with Zebras. If two abutting image regions have different surface textures, this may lead to the detection of the intervening texture border like the border between adjacent zebras. Adapted from [21].

B. Texture Analysis Approaches

Texture perception has been studied using textures constructed by repeated pattern of texture elements [20]. Algorithms have been proposed to explain the process of texture discrimination and segmentation in terms of specific features within the macrotextures themselves. In the human perceptual process, textures provide important cues about surface property, scenic depth, surface orientation, etc. Amazingly, the human visual system utilizes the information effectively in interpreting the scene and performs efficient texture discrimination and segmentation. Fig. 5 shows a scene with zebras containing textural cue for the visual system [21].

Texture recognition in human vision is believed to be one of the earliest steps toward identifying objects and understanding the scene [22], [23]. For example, Julesz et al. conducted early studies on computational texture analysis [3], [22]. They empirically investigated the perceptual significance of various image statistics from texture

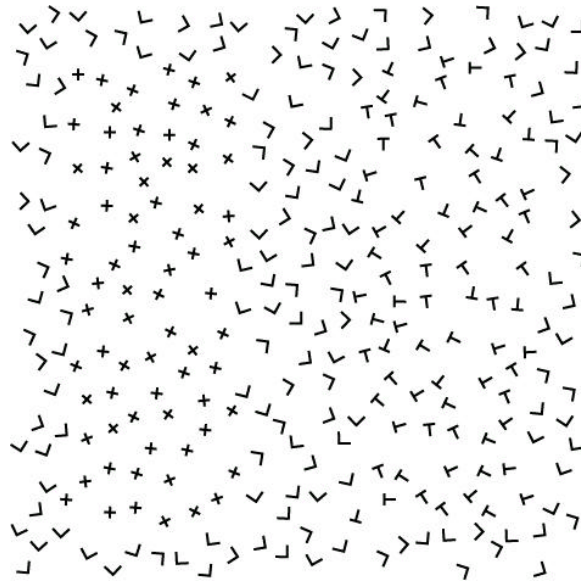


Fig. 6. Discrimination of Texture Pair with X, T, and L. T and L are hard to separate, while X stands out quite prominently. Adapted from [22].

patterns in order to determine how the human low-level visual system responds to the variation of a particular statistic property.

Much of the work on texture perception concerns the ability of observers to effortlessly discriminate certain texture pairs. For example, Fig. 6 shows rectangular regions of X's and T's on a background of L's. Observers can perceive effortlessly that there is a region of X's different from the background, that this region has smooth, continuous borders, and that these borders form a rectangular shape. This kind of perceived effect is referred to as the segregation of the figure from the ground or segmentation of the image into multiple, homogenous regions. Interestingly, none of these observations apply to the region of T's without effortful scrutiny on the individual texture elements one by one.

Texture and its effects on human visual perception have been the subject of interest in the vision research community, with extensive studies conducted in multiple disciplines including neuroscience, psychophysics, and computer science. In neuro-

science and psychophysics, texture research focuses on the neural processes involved in visual perception, with a great amount of effort on understanding the mechanisms of texture detection and segregation. In computer vision, texture studies focus on simulating human texture perception via computer algorithms and deriving appropriate mathematical representations of textures to facilitate computerized texture processing, classification, and segmentation. Note that texture research in computer vision is not an abstract exercise in mathematics, because it is motivated by studies in neuroscience and psychophysics, and inspires them in reverse.

However, as mentioned above, there is a yet unexplored approach involving tactile perception in texture analysis. Our knowledge of texture perception in the tactile domain and its neural mechanisms has changed dramatically in the last decade. A major step was the demonstration that tactile texture perception could be closely related to touch which involves two major dimensions: roughness and softness of surface property. Those dimensions are closely related to the tactile sense which occurs for example when a person swipes the finger tip across the surface of an object. Multidimensional scaling studies have shown that tactile texture perception includes soft-hard and smooth-rough as dominant perceptual dimensions, surface hardness and roughness can occur in almost any combination, and that they account for almost all of texture perception [24], [25]. In addition to whatever information may be gained about the shape and other geometrical properties of the object through touch, the observer also receives impressions related to the nature of the surface. Such an experience of surface texture is the subject of this dissertation. I will investigate the relationship between texture (in general) and the tactile sensory modality in rest of this chapter.

C. The Spatiotemporal Organization of Visual Receptive Fields

In neuroscience and psychophysics, the properties of visual cortical neurons are commonly characterized by their receptive fields. What is a receptive field? A “field” is a group of objects or properties that share a common characteristic. The concept of receptive fields can be extended to describe properties of the biological system. The receptive field of a neuron is the portion of the sensory field that affects the response of that neuron. That is, the receptive field is an area on the sensory surface where stimulation leads to response of a particular sensory neuron [26]. For instance, in Fig. 7, the entire area that the two eyes can see is called the visual field, and the patch of the visual field that any single neuron monitors is that neuron’s receptive field.

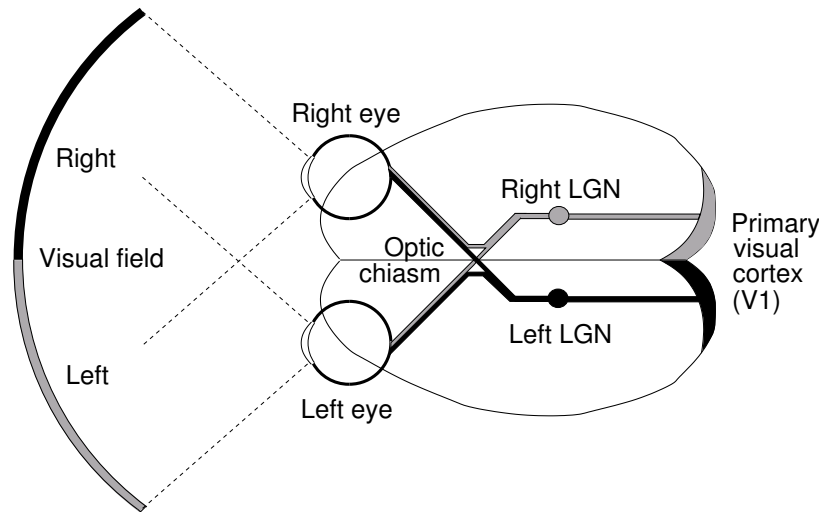


Fig. 7. Human Visual Pathways (top view). Visual information travel in separate pathways for each half of the visual field. For example, signals entering the right eye reach the left half of the retina, on the rear surface of each eye. The inputs join at the optic chiasm, and move to the LGN of the left thalamus, then to primary visual cortex (V1) of the left hemisphere. Information from each eye is kept segregated into different neural layers in the LGN, and is combined in V1. Adapted from [11].

Neurons at different stages in the visual system have receptive fields that differ

not only in size but also in structure. Fig. 8 illustrates the receptive field types in the retina, the lateral geniculate nucleus (LGN), and the primary visual cortex (V1). Most neurons in the retina and LGN have small receptive fields that resemble two concentric circles, ON-center and OFF-center (Fig. 8a, b). This concentric receptive field structure is usually known as center-surround organization, a term that was originally coined by [27]. ON-center receptive fields respond most strongly to light spots surrounded by dark backgrounds like the moon on a dark night. OFF-center receptive fields prefer dark spots surrounded by light backgrounds like a black dot on a white board. The relative size of the spot and background determines the spatial frequency preference of the neuron; neurons preferring large spots have a low preferred spatial frequency, and vice versa.

In V1, receptive fields are much more diverse and more complicated than those in the retina and the LGN. Hubel and Wiesel advanced the idea that there are cortical receptive fields that respond best when the stimulus was of a certain shape, had a given orientation, and/or moved in a given direction [28–31]. Some cortical cells respond to light and dark spots in different subregions of the receptive field and the arrangement of these subregions can be used to predict the response of the cell to visual stimuli such as lines, bars or squared shapes. Figs. 8c-e show examples of typical receptive fields of V1 neurons for static and moving stimuli. These neurons are simple cells, i.e. neurons whose ON and OFF regions are located at specific areas of the retinal field.

Many researchers have demonstrated that visual cortical receptive fields resemble Gabor patterns [8, 29, 30, 32], with specific tuning for orientation, phase, and spatial frequency. Fig. 9 shows examples of Gabor-like RF models of varying orientation and phase. In addition to these static features, visual cortical neurons are also sensitive to dynamic features such as the direction of motion of the stimuli [33–36]. Fig. 8e

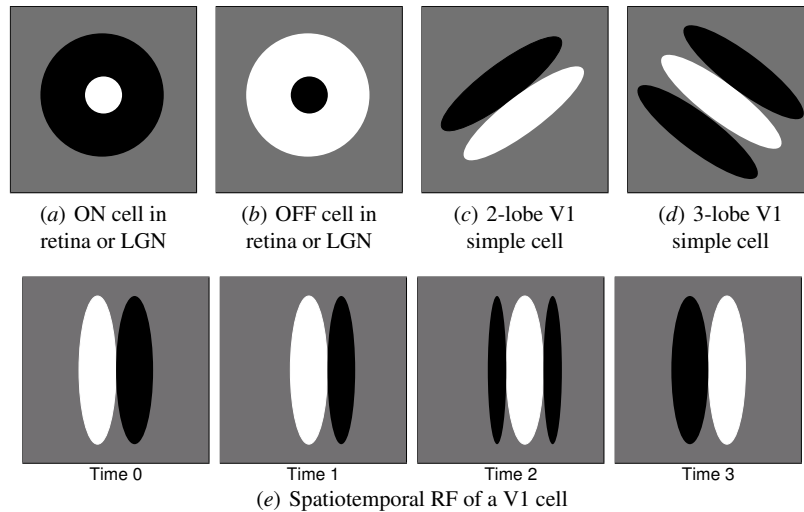


Fig. 8. Receptive Field Types in Retina, LGN, and V1. The diagrams in the two top left panels show a receptive field on the retina or LGN for one neuron: (a) ON cells prefer light areas surrounded by dark. (b) OFF cells respond strongly to dark areas surrounded by light. The diagrams in the two top right panels show V1 RFs which can be classified into a few basic spatial types, of which the two most common are shown above: (c) a two-lobe arrangement, preferring a 45° edge with dark in the upper left and light in the lower right, and (d) a three-lobe pattern, preferring a 135° white line against a dark background. Both types of RF are often represented with Gabor functions [8]. Many neurons in V1 are also sensitive to the direction of movement: (e) a spatiotemporal RF in V1 shown as a vertical light bar moving to the right. Adapted from [11].

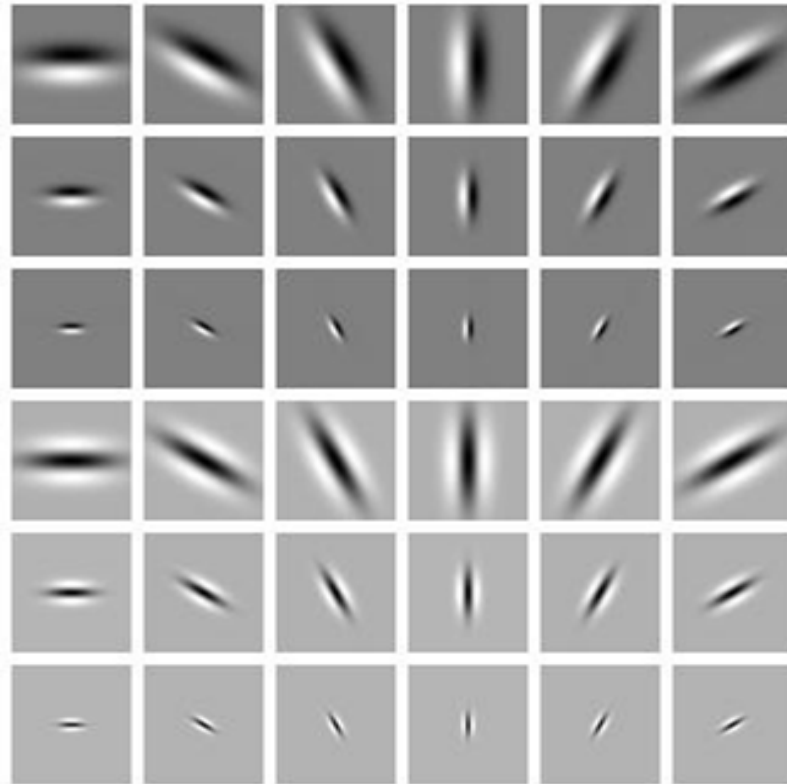


Fig. 9. Gabor Filter Bank. Gabor filter bank parameterized by 6 orientations (columns), 3 scales (rows), and phase (top 3 rows vs. bottom 3 rows) shown.

shows an example of such a spatiotemporal RF.

Fig. 10 shows an example of a direction-selective RF in the cat primary visual cortex [33]. Instantaneous RFs in the two-dimensional visual space at times 20, 60, 100, and 120 ms are shown on top, and a continuous integration of the RFs along the vertical is drawn in the bottom plane. The neuron's spatial preference changes systematically over time, giving it a spatiotemporal preference for a dark vertical line moving horizontally to the right.

The developmental origin of how these complicated yet orderly structures arise in the cortex has been studied since the 1960s. Hubel et al. first showed that if a kitten's vision is impaired by suturing the eyes shut, the visual cortex becomes disorganized,

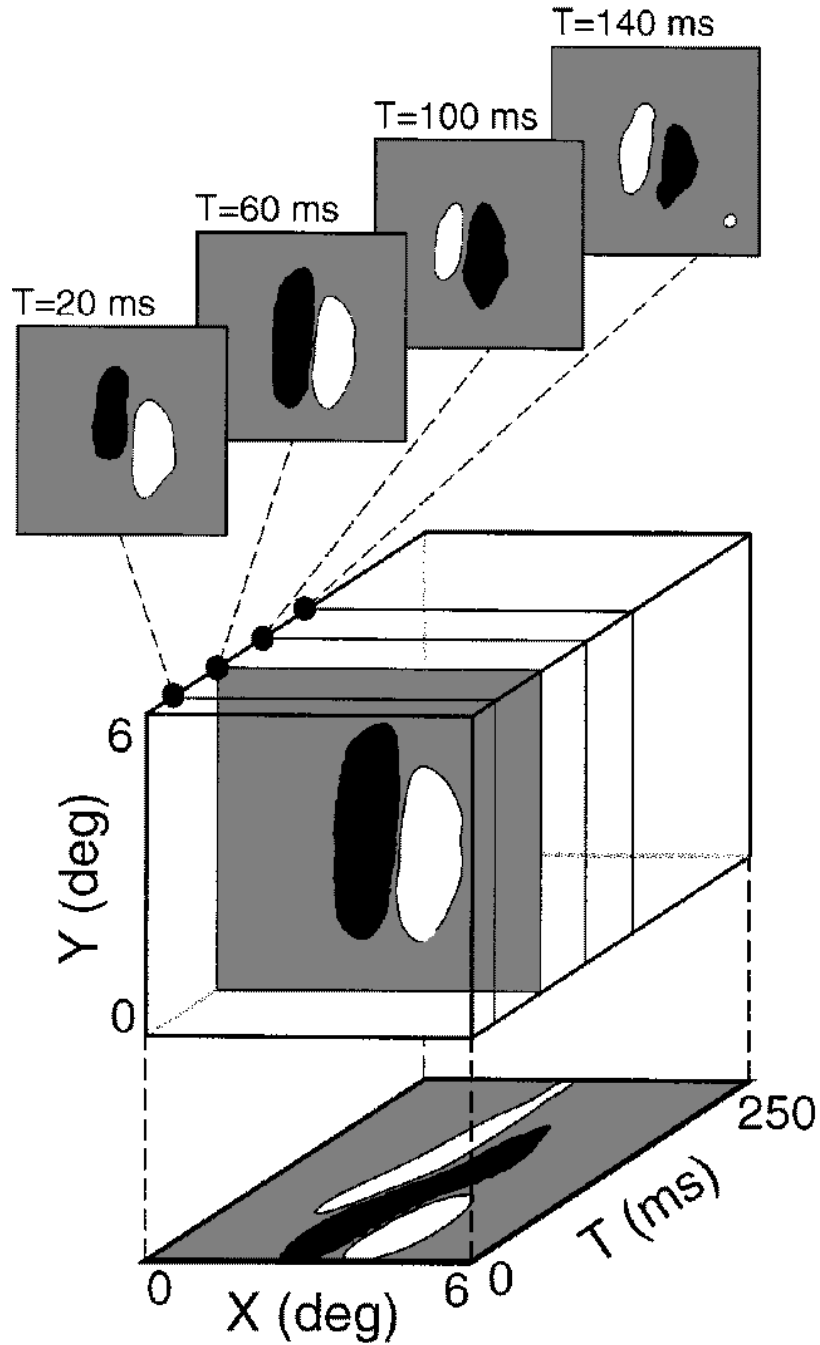


Fig. 10. Spatiotemporal Receptive Fields in Cat. Primary visual cortical neurons show direction (of motion) selectivity in addition to orientation selectivity. The figure shows the spatiotemporal pattern that optimally stimulates a visual cortical neuron in a cat. Here, we can see the dark region of a vertical Gabor pattern moving to the right. Adapted from [33] (as rendered in [11]).

lacking orientation selectivity and ocular dominance patches [29], [37], [38]. White et al. showed that if the animal (e.g. a ferret) is reared in the dark instead of suturing the eyes shut, the visual system becomes similarly impaired, although to a lesser extent [39], suggesting that abnormal visual stimulation through the closed eyelids is more harmful than receiving none at all. These results show how important normal visual stimuli are during the critical period in ensuring that the visual system develops normally.

Development of visual ability has been shown to depend on input in several more specific experiments as well. For example, kittens can be raised in an environment with only vertical or horizontal features, and as a result, they are unable to respond well to other orientations [40], [41], [42]. Fig. 11 shows the experiment with a kitten reared in a vertically striped tube.

These experimental results convincingly demonstrate that the synaptic connections firing rise to the visual receptive fields are shaped by environmental input. One of the aims of this dissertation is to understand the developmental origin of tactile RFs, showing that input-driven self-organization is able to construct the spatiotemporal structures of tactile receptive fields even from an initially uniform, unordered starting point, based on suitable input, especially texture-like input. A method to test this hypothesis is proposed in chapter IV showing how environmentally driven self-organization can account for many of the observed phenomena in tactile development, and in texture processing.

D. The Cortical Mechanism of the Area 3b in the Primary Somatosensory Cortex

The somatosensory system which is spread through all major parts of a vertebrate's body is a diverse sensory system comprising the receptors and processing center to



Fig. 11. Vertically Striped Tube Used in Selective Rearing Experiments. The cat is wearing a ruff to prevent it from turning the vertical stimuli into horizontal by tilting its head. Adapted from [40].

produce the sensory modalities such as touch (feel), temperature (hot or cold), body position, and pain (ache). As with all other sensory modalities, the most complex levels of tactile processing occur in the cerebral cortex. Most of the cortex concerned with the tactile sensory system is in the parietal lobe. Fig. 12 shows the somatosensory area of the cortex. The primary somatosensory cortex (S1) is easy to find in humans because it occupies an exposed cortical strip called the postcentral gyrus. Structurally, S1 consists of four distinct cortical areas, Brodmann's areas 3a, 3b, 1, and 2, counting from the central sulcus back. A secondary somatosensory cortex (S2) at the lateral end of S1 is revealed by pulling back the temporal lobe and peeking over the auditory cortex at the lower part of the parietal lobe. Finally, the posterior parietal cortex, consisting of areas 5 and 7, sits just posterior to S1 [43].

Especially, S1 is the primary somatosensory cortex because (1) it receives dense inputs from the thalamus, (2) its neurons are very responsive to somatosensory stimuli (but not to other sensory stimuli), (3) lesions in S1 impair somatic sensation, and (4) when electrically stimulated, it evokes somatosensory experience.

The different areas of S1 have different functions. Area 3b is concerned mainly with the texture, size, and shape of objects. Its projection to area 1 sends mainly texture information, while its projections to area 2 emphasize size and shape.

The somatosensory cortex, like other areas of the neocortex, has a layered structure. As is the case for the visual and the auditory cortex, the thalamic inputs to S1 terminate mainly in layer IV [44]. The neurons of layer IV project, in turn, to the cells in the other layers. Another important similarity with other regions of cortex is that S1 neurons with similar inputs and responses are stacked vertically into columns that extend across the cortical layers (Fig. 13).

Neurophysiological studies have shown that area 3b has smaller receptive fields, and a higher proportion of cells there respond to static skin indentation than other

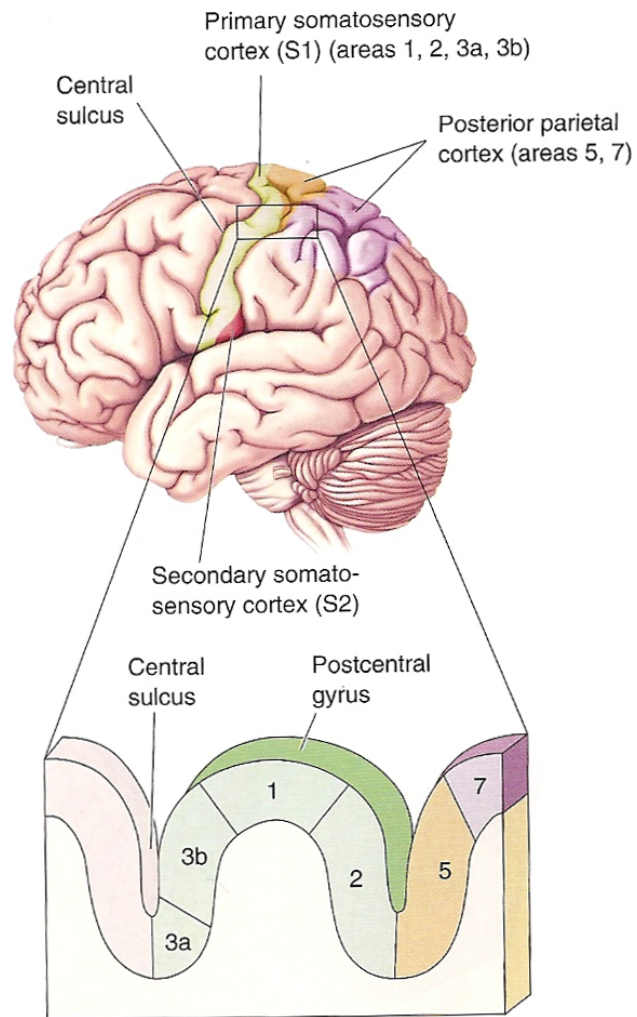


Fig. 12. Somatosensory Areas of the Cortex. All of the illustrated areas of the somatosensory cortex lie in the parietal lobe. The lower sketch shows that the postcentral gyrus contains S1, which consists of four cortical areas. Adapted from [43].

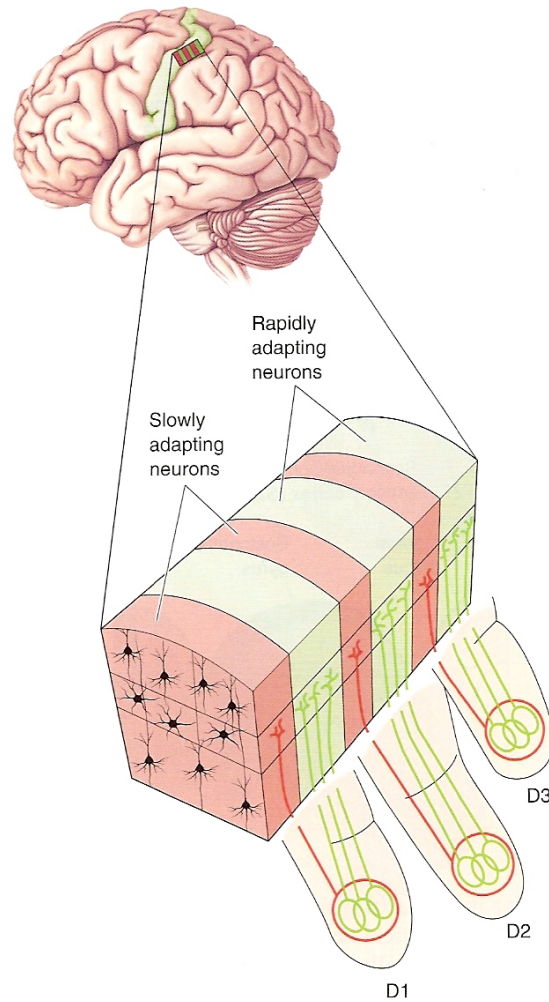


Fig. 13. Columnar Organization of Area 3b in S1. Each finger (D1-D3) is represented by an adjacent area in the cortex. Within the area representing each finger, there are alternating columns of cells with rapidly adapting (green) and slowly adapting (red) sensory responses. Adapted from [44].

cortical area, but those observations are not especially indicative of a role in spatial information processing [45, 46]. The smaller receptive fields in area 3b are almost certainly a consequence of the fact that area 3b lies at an earlier stage of processing within pathways leading to distributed representations of spatial form [47, 48].

A continued response to sustained, steady indentation implies input from slowly adapting afferent but the converse is not true, a transient response may result from the lack of slowly adapting input or, just as likely, delayed inhibition that shuts off the response to a sustained input. Ablation studies shed little light on the function of area 3b. Removal of area 3b results in profound behavioral deficits in all somatosensory tasks, while removal of other S1 areas appears to produce more specific deficits in the tactile discrimination of textures (area 1) and three-dimensional forms (area 2) [49].

A recent series of quantitative studies with controlled, scanned stimuli has confirmed that there is little or no directional selectivity in area 3b and that area 3b neuronal discharge rates are affected only mildly by changes in stimulus velocity across the skin [9, 50, 51]. But it was also shown that all neurons in area 3b are selectively responsive to particular spatial patterns of stimulation, that they are sensitive to the orientation of these patterns, and that this selectivity is shaped as much by inhibition as excitation. Fig. 14 shows how DiCarlo et al. measured the receptive fields of area 3b neurons in the monkey [51]. The stimulus pattern consisted of a field of randomly distributed, raised dots on a plastic surface, mounted on the surface of a drum which rotated around an axis and shift along its axis of rotation.

Ninety-five percent of area 3b neuronal receptive fields had three components; (1) a single, central excitatory region of short duration (10 *ms* at most), (2) one or more inhibitory regions that are adjacent to and synchronous with the excitation, and (3) a larger inhibitory region that overlaps the excitation partially or totally and delayed with respect to the first two components (by 30 *ms* on average). The remaining 5

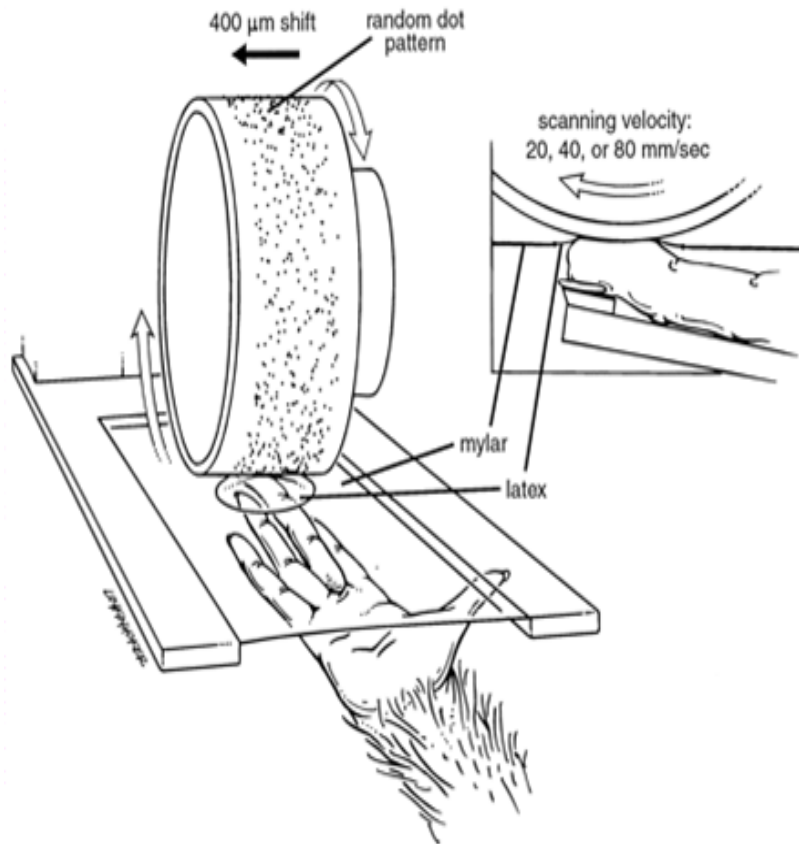


Fig. 14. Drum Stimulator to Simulate Tactile Texture. The stimulus pattern consists of randomly distributed, raised dots on a plastic drum surface: 28 mm wide, 3175 mm long, and 320 mm in circumference. The dot pattern stimulates the skin through a thin latex sheet positioned over the distal fingerpad that mapped to the neural RF. The intermediate latex sheet which is tethered to a circular aperture in a Mylar sheet supported by a Plexiglas frame is used to minimize lateral skin movement caused by tangential, frictional forces between the surface and the skin. The hand and finger are held fixed from below and the intermediate contacted the fingerpad with a force of 10 gm. The drum rotates with a controlled normal force (30 gm), producing a surface pattern motion from proximal to distal over the fingerpad. The scanning velocity is fixed at 20, 40, or 80 mm/sec for each scan through the surface pattern. After three drum rotations (one at each scanning velocity), the drum was translated by 400 mm along its axis of rotation. The data entering into the RF estimates were derived, on average, from 25 scans at each velocity, which corresponded to 10 mm of translation. Adapted from [51].

percent had two or more regions of excitation.

As briefly discussed in previous sections, tactile RFs have similar properties as the visual counterpart (Fig. 2), since they are based on the on-off Gabor pattern. However, there is a dynamic component that curiously depends on the direction of scan of the skin patch (e.g., from the finger tip). Fig. 15 depicts the three component model of the tactile receptive field, proposed by DiCarlo et al. The RF developed by the computational model in each scanning direction (i.e., each panel) is the sum of these three Gaussian functions.

Fig. 16 shows examples of the tactile RFs estimated from the cortical area 3b of an alert monkey compared to the model predictions [9]. In these plots, dark represents an inhibitory region and bright an excitatory region. Each row in the figure shows the RF estimated from the raw data (left) measured through microelectrode recording in area 3b of the alert monkey, the RF predicted by a three-component model (middle), and the positions of the Gaussian components in the model (right). Three ellipses in the rightmost column represent fixed excitatory (thick oval), fixed inhibitory (dashed oval), and lagged inhibitory (thin oval) lobe moving in the opposite direction of the scanning direction (arrow). Depending on the direction of scan, the dark blob (lagged inhibitory region) moves around, altering the final shape of the RF. The resulting RFs, even though they are based on a Gabor pattern, show patterns distinct from visual RFs, e.g., like a donut or a curve.

E. Summary

In this chapter, the nature of texture and texture analysis approaches were introduced. The spatiotemporal organization of visual receptive fields and the cortical mechanism of the area 3b in the primary somatosensory cortex were also presented to put my

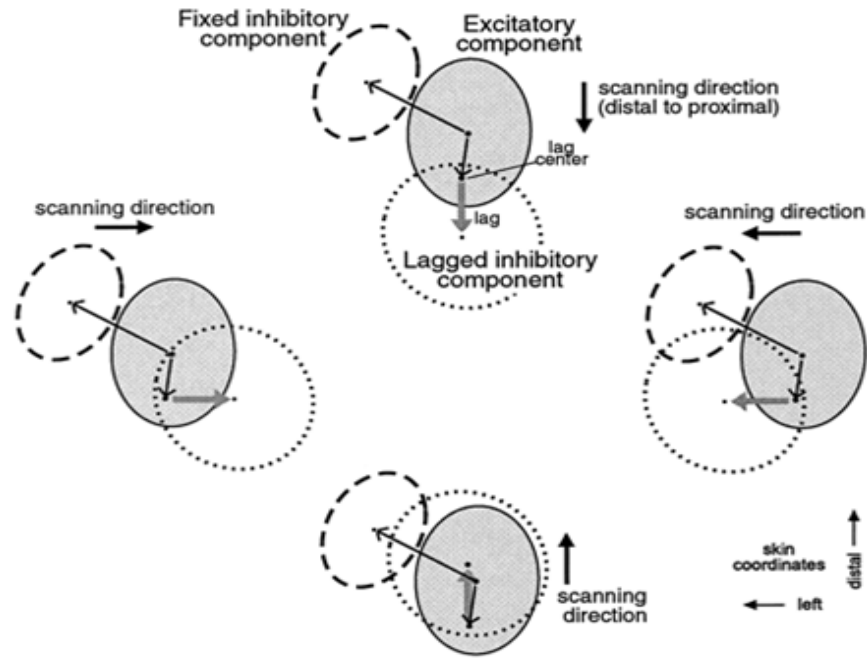


Fig. 15. Three-component Gaussian Model of Tactile Receptive Field. Three ellipses in each panel represent isoamplitude contours around Gaussian functions describing three RF components; fixed excitatory (solid circle), fixed inhibitory (dashed circle), and lagged inhibitory (dotted circle). The RF developed by the model in each scanning direction (i.e., each panel) is the sum of these three Gaussian functions. Not like the fixed component, the lagged inhibitory component only relocate according to the scanning direction of finger on the surface. The locations of the fixed inhibitory center and the lag center in relation to the fixed excitatory component are identified by the two thin arrows originating from the center of the fixed excitatory component. The displacement of the lagged inhibitory component from the lag center is indicated by the thick, gray arrow. The tail of the gray arrow is at the lag center; the arrow direction corresponds to the stimulus direction across the RF (i.e., scanning direction). The tip of the gray arrow specifies the apparent location of the lagged inhibitory center. Adapted from [9].

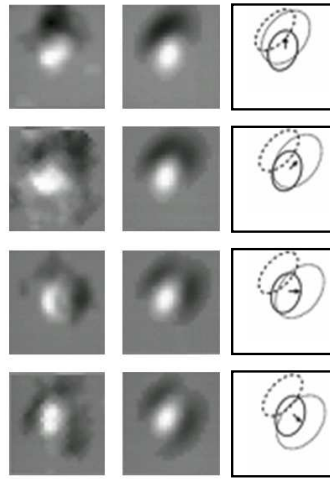


Fig. 16. Tactile Receptive Fields in Monkey Area 3b. RFs resulting from four different scanning directions on the finger tip of a monkey is shown. Each row shows, from the left to the right, (1) the actual measured RF (bright=excitation, dark=inhibition), (2) the three-component model by [9], and (3) the outline of the three-component model. Given the same excitation–inhibition pair, a third (inhibitory) component shows up, and the center of that component shifts its position in the opposite direction of the scan. For example, the top row corresponds to a downward scan, thus the third component shifts up. See page 45 for a detailed explanation. Adapted from [9] (grayscale was inverted to show excitation in white and inhibition in black).

research in a proper biological context. As I mentioned above, visual and tactile RFs share common features but they also have subtle differences. The similarity may be driven by the fact that the visual and tactile sensory surfaces have the same underlying structure (i.e., a 2D sheet), and that they are stimulated by the same underlying spatial environment. In the following chapters, I will build upon the computational model of the tactile receptive field model, show their performance in texture tasks comparing to that of visual approach, and find the developmental origin of tactile receptive fields.

CHAPTER III

PERFORMANCE: TACTILE VS. VISUAL RECEPTIVE FIELDS IN TEXTURE SEGMENTATION

As discussed above, one interesting possibility is that texture may be more intimately related with touch than with vision. Recent neurophysiological findings showed that receptive fields for touch resemble that of vision, albeit with some subtle differences. To leverage on this, I tested how such distinct properties in tactile receptive fields can affect texture segmentation performance, as compared to that of visual receptive fields.

A. Visual or Tactile Perception for Texture

Visual perception starts from segregation of scenes based on cues related to luminance, color, contours and texture of object surfaces. Furthermore, the human visual system uses texture information in order to automatically or preattentively segregate parts of the visual scene [2]. Several theories and algorithms exist for texture discrimination based on vision [3,4]. These models differ from each other in their algorithmic approaches in processing texture imagery based on the spatial elements and their statistics. Even though there are differences among these approaches, they all begin from the same fundamental assumption that texture segmentation is a visual task.

However, considering that texture is basically a surface property, this assumption can at times be misleading [5]. An interesting possibility is that since surface properties are most immediately accessible to touch, tactile perception may be more intimately associated with texture than with vision (it is known that tactile input can affect vision [6]).

Interestingly, the basic organization of the tactile (somatosensory) system bears

some analogy to that of the visual system [7]. In particular, recent neurophysiological findings showed that receptive fields for touch resemble that of vision, albeit with some subtle differences [7, 9, 10]. To leverage on this, I tested how such distinct properties in tactile receptive fields can affect texture segmentation performance, as compared to that of visual receptive fields.

The results suggest that touch has advantages over vision in texture processing. I expect these findings to shed new light on the role of tactile perception of texture and its interaction with vision, and help develop more powerful, biologically inspired texture segmentation algorithms.

The rest of this chapter is organized as follows. First, a computational model for tactile RFs will be presented (Sec. B), followed by introduction of texture sets that I used in this experiment. Second, performance in texture boundary detection tasks for TRFs and VRFs will be compared, and experimental results presented (Sec. C). Finally, I will discuss issues regarding our results (Sec. D).

B. Computational Model for Tactile RFs: The Three-Component Model

The most widely used feature generator for texture segmentation is the computational model of the visual receptive field (of V1 simple cell), the Gabor filter [8]. When generating Gabor features, typically, an input image $I(x, y)$, $(x, y) \in \Omega$ (Ω is the set of pixel locations) is convolved with a 2D Gabor function $G(x, y)$ as follows [52]:

$$G_{\lambda, \theta, \varphi}(x, y) = \exp\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cdot \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right) \quad (3.1)$$

$$x' = x \cdot \cos \theta + y \cdot \sin \theta, y' = -x \cdot \sin \theta + y \cdot \cos \theta \quad (3.2)$$

where λ is the wavelength ($1.5 \times$ window size), θ is the orientation preference, φ is the symmetry (phase), γ is the aspect ratio, and σ is the standard deviation of the

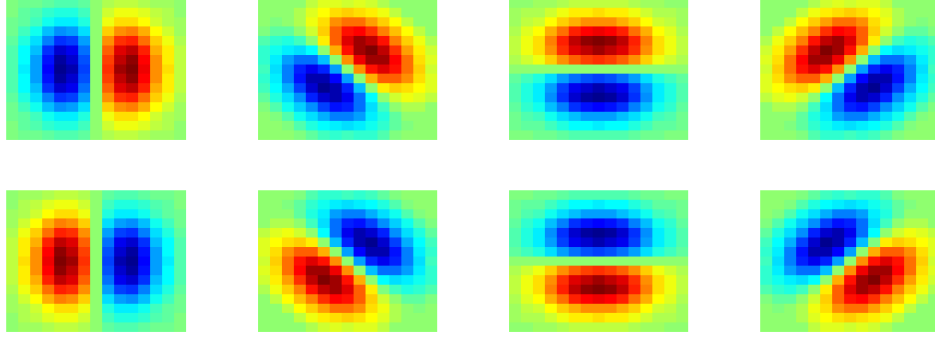


Fig. 17. Oriented Gabor Filters. Visual receptive fields modeled as Gabor filters with 8 orientations: 0° , 45° , 90° , 135° , 180° , 225° , 270° , 315° . The blue blob is the inhibitory component and the red blob is the excitatory component.

Gaussian. In the experiments, below these values were set to $\sigma = 0.56$, $\gamma = 1.0$, and $\varphi = 0.5\pi$, and a bank of Gabor filters with eight preferred orientations, $\theta = k \cdot \frac{\pi}{8}$, ($k = 0, 1, \dots, 7$) were constructed. Fig. 17 shows the visual receptive fields used in this experiment.

The tactile counterpart of the V1 simple cell model is the receptive field (RF) of neurons in the somatosensory area 3b [9]. To the best of my knowledge, tactile RFs have not been incorporated in any texture segmentation or computer vision related algorithms. DiCarlo and Johnson estimated the tactile RF model by recording area 3b neural responses to dot patterns using reverse correlation. The main structure of the RFs consists of three Gaussian subfields: central excitatory region accompanied by an inhibitory lobe and a temporally, dynamically lagging inhibitory lobe with respect to the excitation center [9]. Each subfield can be expressed as:

$$G(x, y) = a \cdot \exp\left(-\frac{1}{2}L^T S^{-1}L\right), \quad (3.3)$$

$$L = \begin{bmatrix} x - \mu_x - \mathbf{v}_x \tau \\ y - \mu_y - \mathbf{v}_y \tau \end{bmatrix}, S = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^w \end{bmatrix} \quad (3.4)$$

where (μ_x, μ_y) represents the center of the subfield, $(\mathbf{v}_x, \mathbf{v}_y)$ represents the stimulus velocity vector, and τ represents the delay of the peak excitation or inhibition with respect to skin stimulation. The parameters a , σ_x , σ_y , and ρ specify the amplitude, spread, orientation, and elongation (aspect ratio) of the excitatory ($a > 0$) or inhibitory ($a < 0$) component represented by the Gaussian function. The center of excitation was fixed to stay at the middle of all tactile models while the complementary inhibition and lagging inhibition centers varied with respect to the excitatory center.

Finally, the three Gaussian subfields are linearly summed to represent the tactile model. Fig. 18 shows the tactile receptive fields constructed for this experiment, based on [9]. The outline in the rightmost panel in each row depicts the three-component Gaussian model. The arrows represent the scanning directions of the fingertip. For each scan, the resulting RF is illustrated through three diagrams: (1) The excitatory and fixed inhibitory lobes are outlined in gray ellipses and the lagging component is illustrated as dotted (before scanning) and black (after scanning) ellipses; (2) the lagging inhibitory lobe (displaced in the opposite direction of the scan); (3) and the linear summation of arrows listed as fixed orientation components. The leftmost panel represents the computational model constructed for this experiment, and the middle panel shows the 3D profile of the tactile receptive field.

This extra degree of freedom of the TRF model (the lagging component) affects the level of occlusion of the excitatory lobe that ultimately determines orientation preference. As with the VRFs, a bank of eight TRF models with equidistant orientations was made.

Given the computational models of tactile and visual modalities, the performance of texture segregation was measured through means of detecting texture-defined boundaries from natural and synthetic texture image inputs. I used 18 textures from

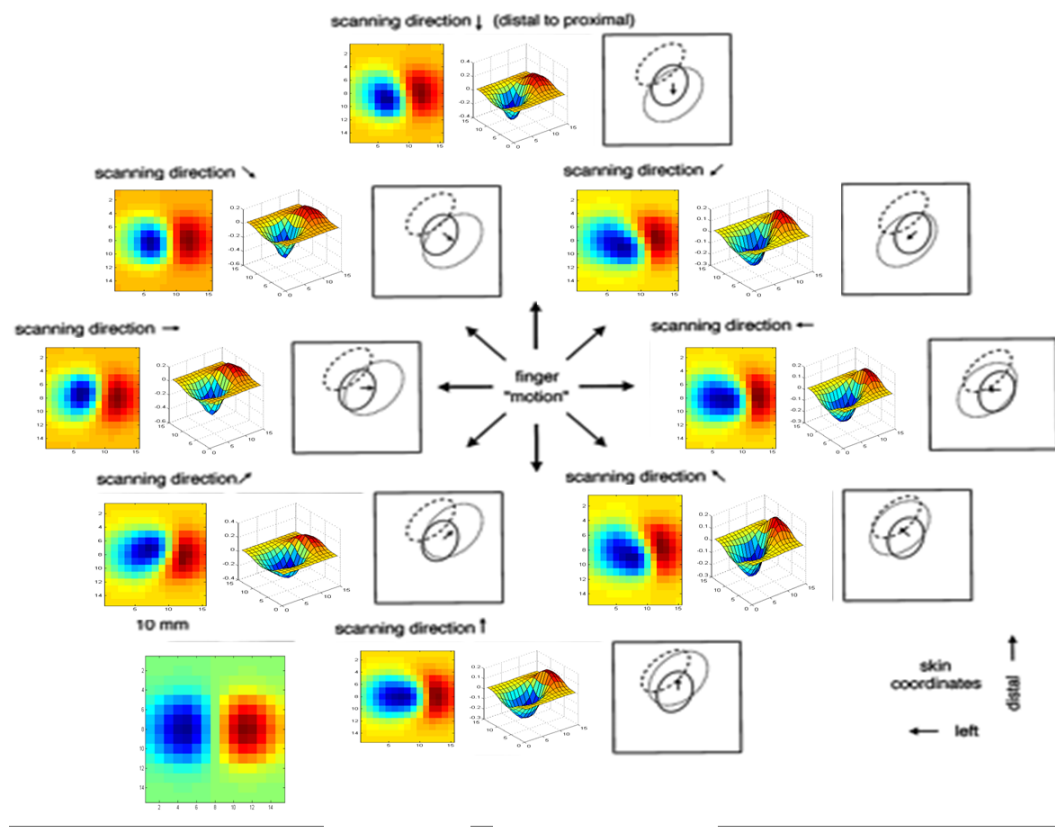


Fig. 18. Estimated tactile receptive fields with 8 scanning directions. In each row, the rightmost panel depicts the three-component Gaussian model which is the isoamplitude contours around Gaussian functions describing three RF components (excitatory, fixed inhibitory, and lagged inhibitory, adapted from [9]), the leftmost panel represents the model constructed for this experiment, and the middle panel shows the 3D profile of the receptive field.

the widely cited Brodatz texture collection [53] and 18 textures with boundary simply synthesized by appending two textures. For the experiments, I made six different texture sets so that each set contains three non-boundary textures and three boundary textures. Fig. 19 shows six exemplary texture sets containing three non-boundary textures and three boundary textures in each set.

In order to extract the RF response for the given textures, each texture was preprocessed by a Laplacian of Gaussian (LoG) filter, a popular choice for edge detection, followed by a transformation of the edge into detectable discontinuities [54]. The LoG filter is defined as below

$$G_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \left(-\frac{x^2 + y^2}{2\sigma^2} \right), \quad (3.5)$$

$$LoG = \Delta G_\sigma(x, y) = \frac{\delta^2}{\delta x^2} G_\sigma(x, y) + \frac{\delta^2}{\delta y^2} G_\sigma(x, y) \quad (3.6)$$

where σ is the standard deviation (width) of the Gaussian envelope and is set to 0.5 in the experiment.

In order to simulate the stimuli from the tactile sensation from a finger, I examined a certain number of consecutive window patches (frames) sliding across a predefined scanning direction inside the input image. The pixel intensity in the image played the role of surface height in texture surfaces. In this experiment, I used 12 frames with a window size of 15 (or 17) pixels. TRF and VRF filter banks constructed with 8 oriented RFs having identical size with the window patches (frames) were applied by vectorized dot product with the individual window patches, producing a vector consisting of 12 response values.

Specifically, I examined multiple scans from every input image with different scanning directions to accommodate all possible ways of encountering the texture boundaries. Fig. 20 shows a group of typical response profiles extracted from random

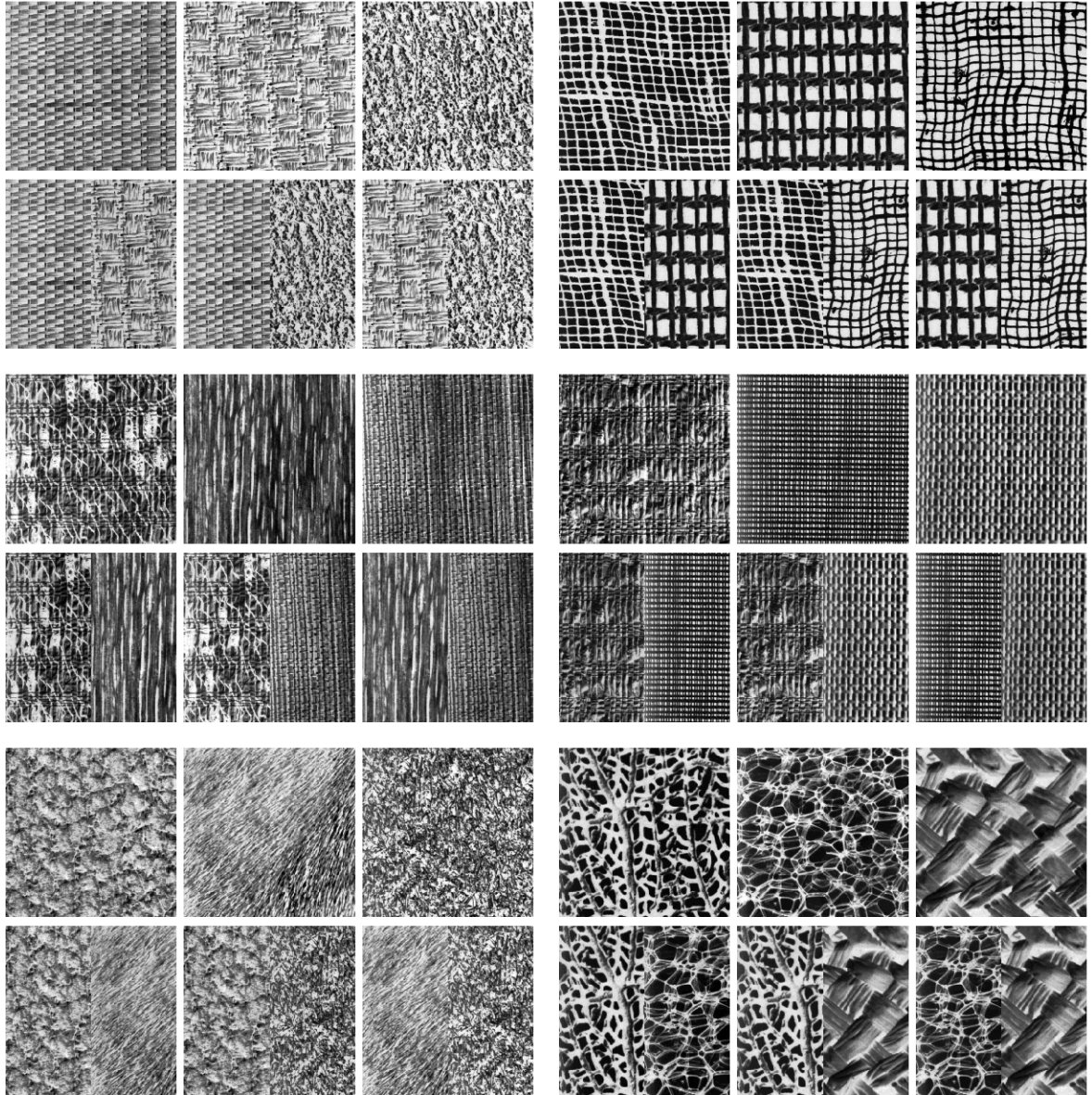


Fig. 19. Six Texture Sets. In each set, the top row shows non-boundary and the bottom row boundary textures. Boundary textures are constructed by attaching two non-boundary texture in each set.

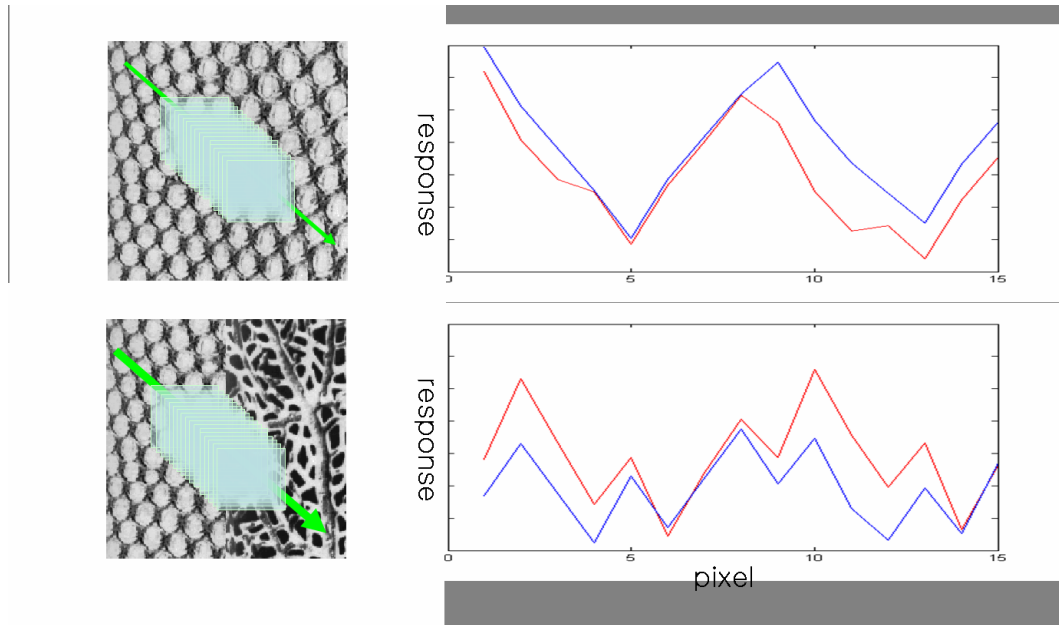


Fig. 20. Sample Textures and Their Typical Response Profiles at Scanning Direction of 315° . Upper part is for non-boundary texture and lower part for boundary texture. Non-boundary texture shows regular curve but boundary texture shows irregular curve with both tactile and visual RFs.

natural texture images scanned at various directions. Note that the profiles of TRF and VRF have similar characteristics, and the profiles without a boundary are more symmetric about the center compared to those with a boundary.

Fig. 21 shows a 3D visualization of RF responses (adapted from [55]). Texture boundaries were located in the middle of the input image made up of two distinct textures. As shown, TRF responses show higher amplitudes along the texture boundary, compared to those of VRFs.

C. Texture Boundary Detection Tasks with TRFs and VRFs

In order to test the performance of TRFs as compared to VRFs, I compared the accuracy of the two types of RF responses (TRFs and VRFs) in texture boundary detection tasks (boundary vs. non-boundary), and also analyzed the separability in

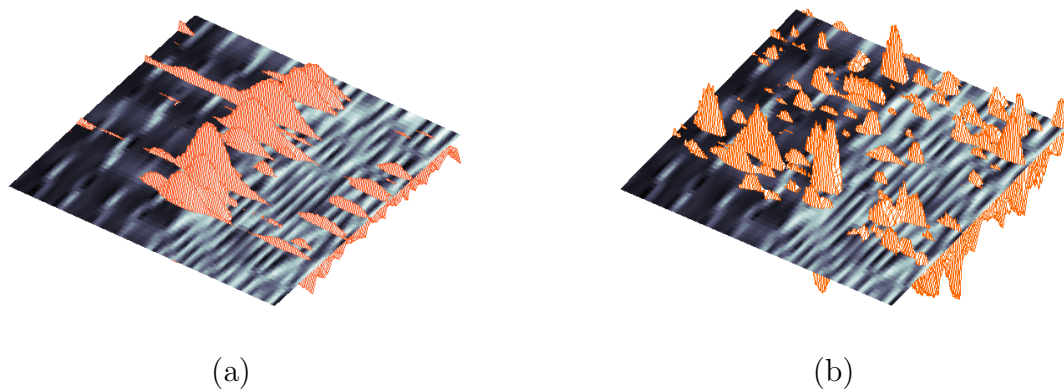


Fig. 21. 3D Visualization of RF Responses. (a) A sample of visual feature vectors from various scans are superimposed on top of an example input image with a texture boundary in the middle. (b) A sample of tactile feature vectors from various scans are superimposed onto an example input image with a texture boundary in the middle. The response features shown here were generated from various images in the sample set and not from the single example input image. The background image was only used to depict where the texture boundary lies throughout the images in the sample set. Note that tactile feature vectors respond abruptly across the texture-defined boundary while visual feature vectors are less sensitive to texture boundaries. This figure implies that touch-based representations can be more effective in to texture boundary detection than vision-based representations. Adapted from previously reported results in [55].

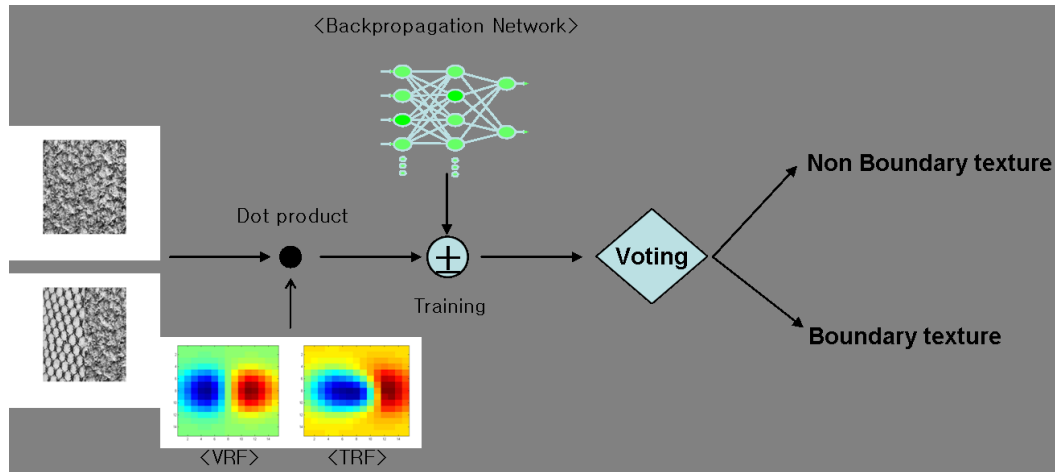


Fig. 22. Texture Classification Experiment. (1) Dot product between texture input and receptive fields generates the response vectors, (2) train a backpropagation networks to classify the texture type (boundary vs. non-boundary), (3) and vote to get the final classification.

the two representations.

I used 18 random textures from the Brodatz collection and constructed two sets of input images: target-present (texture-defined boundary) images versus target-absent (boundary-absent) images as depicted in Fig. 19. On the condition of the same combination of scanning direction, input textures (without any boundaries) and identical parameterizations, I have collected the same amount of data for the boundary case and for the non-boundary case. Fig. 22 shows the procedure for the texture classification experiment.

I trained a standard back-propagation network (120 input unit, 10 hidden units, 2 output units) for 200 epochs at learning rate $\eta = 0.5$ to discriminate texture boundary responses from non-boundary responses. On top of training a standard back-propagation, the final decision of detecting a texture boundary was based on voting [56]. I selected five neighboring response vectors from the same region of the input, and applied the majority rule based on the network output to finally deter-

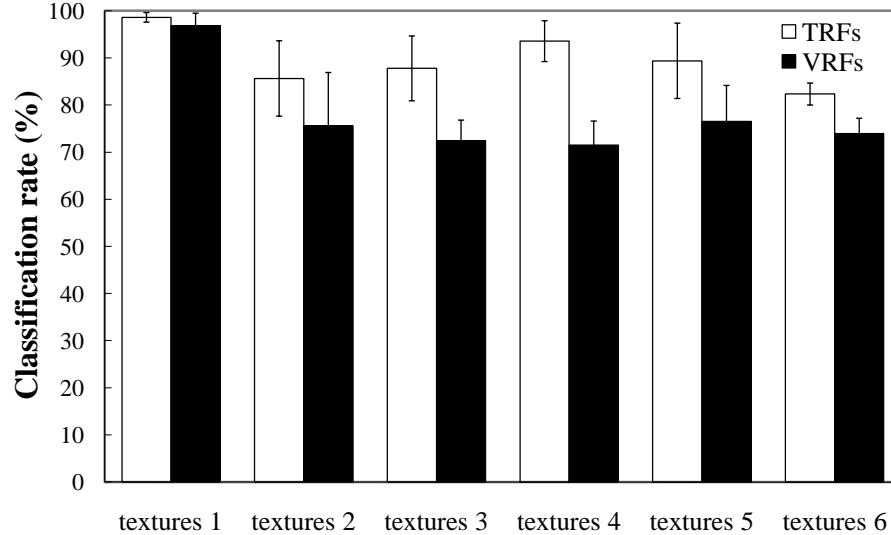


Fig. 23. Comparison of Average Classification Rate for Six Different Texture Sets with TRFs and VRFs. In almost all cases, the classification rate with TRFs is significantly better than the one with VRFs (errorbars indicate standard deviation, $p < 0.03$, $n = 200$). Adapted from previously reported results in [55].

mine whether the five outputs indicated a texture boundary or non-boundary texture input. The experiment with voting has higher classification rate than the non-voting case. In either case, the tactile representation showed better performance.

Fig. 23 shows the resulting classification rate of voted texture boundaries for the 6 texture sets. The TRF performance was significantly superior to that of the VRF (t-test: $n = 200$, $p < 0.03$) except for texture set 1 ($p = 0.27$).

Why are TRFs better than VRFs for texture segmentation? One possible reason is that the non-linear structure of the TRF is more ecologically suited for the representation of surface texture than the linear structure of the VRF because most textures are composed of more non-linear features than linear features.

As we can see in Fig. 18, a three-component model has curvy lobes between excitatory and inhibitory components because the lagged inhibitory component affects

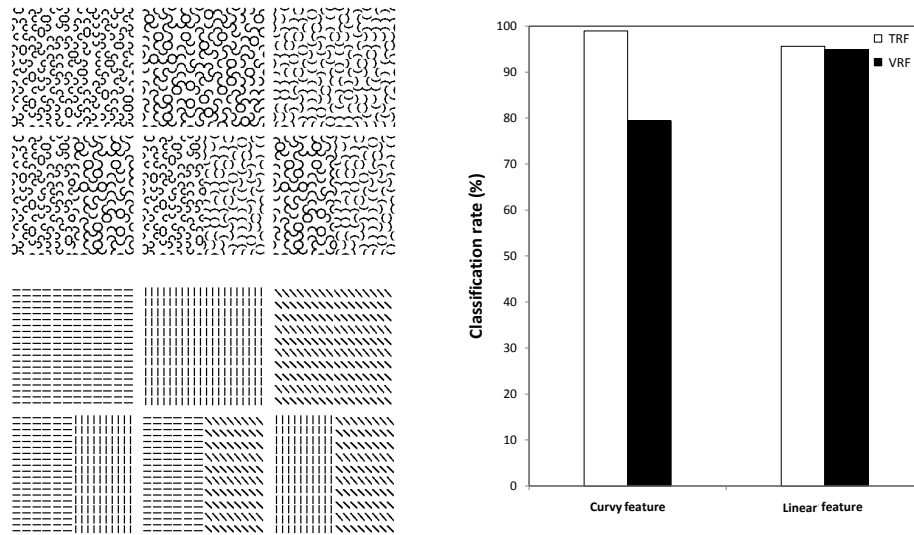


Fig. 24. Comparison of Classification Rate for Curvy and Linear Textures. Left: textures with curvy or linear features are shown. Right: performance of TRF vs. VRF on curvy/linear textures are shown. Adapted from previously reported results in [55].

the two fixed components whereas a Gabor filter has linear division between the excitatory and the inhibitory components. Hence, the bank of three-component models may easily extract more nonlinear features in the surface texture than a Gabor filter bank.

To validate this idea, I tested boundary detection with curvy textures and linear textures. Fig. 24 shows two types of texture and the comparison of the classification rate. Curvy textures without boundary were synthesized with many segments of circles at different curvatures (0.333, 0.2, and 0.143) and linear textures without boundary were synthesized with lines at different orientations (horizontal, vertical, and diagonal). As we can see, TRFs show a higher competitive edge on curvy textures.

In order to test the representational power of TRFs as compared to VRFs, I used Fisher Linear Discriminant Analysis (LDA). Fig. 25 shows the scatter plot of 2-D projections (from LDA) of the tactile and the visual feature spaces and Fig. 26

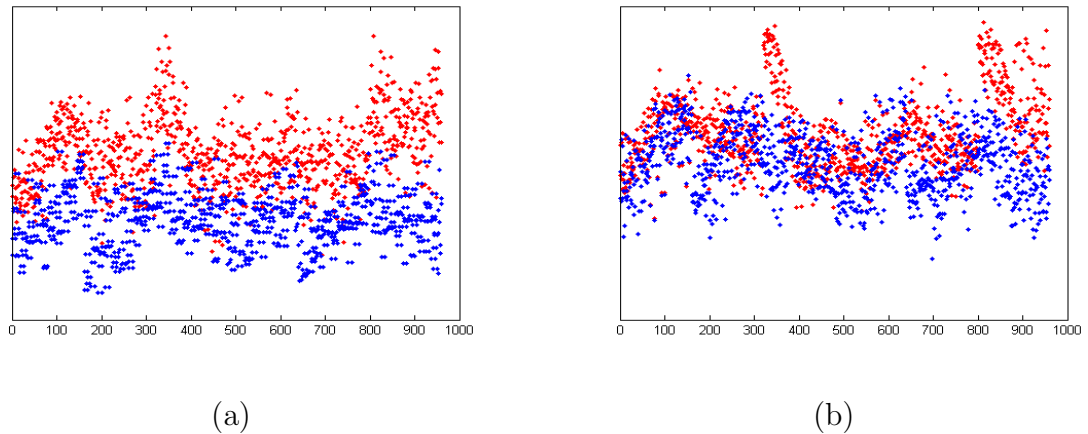


Fig. 25. Linear Discriminant (LD) Distributions of TRF and VRF Responses. For the scatter plot, the x-axis depicts the sample index of each (a) tactile and (b) visual response vector, and the y-axis indicates the linear discriminant values. In each plot, red dots and blue dots represent boundary and non-boundary features, respectively.

shows the probability density of the LDA distribution extracted from a set of input images. In each case, data from the non-boundary and the boundary cases are shown as two separate classes. The plots show that the TRF response feature distribution is more separable than that from the VRF.

D. Discussion

The primary aim of this part of the study was to explore and compare texture segregation performance based on tactile receptive fields (TRFs) and visual receptive fields (VRFs). The results suggest that touch-based texture representation contains more discriminative information than vision-based local spatial features. Statistical measures and classification (boundary vs. non-boundary) performances were used to evaluate this characteristic as well as providing insight for further analysis of the TRF representation. It was found that due to the extra degree of freedom and component

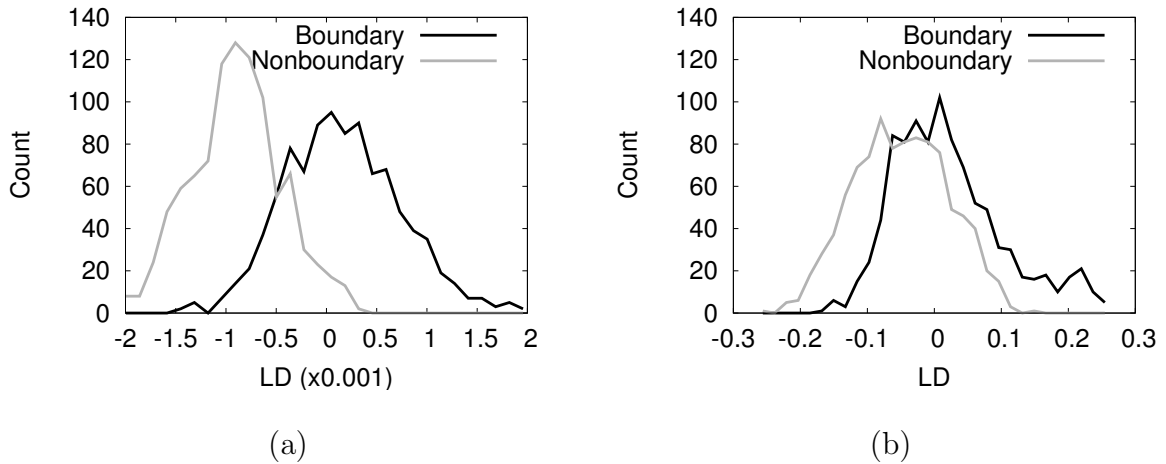


Fig. 26. Frequency Distribution of LDA of TRF and VRF Responses. Visualization of the distribution of the (a) tactile and (b) visual RF responses' LD is shown. In each plot, the red curve and the blue curve represent boundary and non-boundary features, respectively. The TRF response feature distribution is more separable than the VRF distribution.

in the RF structure, TRFs can accommodate more complex spatial properties, e.g., curvature, than VRFs.

The main novelty and contribution of the work reported in this chapter is in the use of tactile receptive field responses for texture segmentation. Furthermore, I showed that touch-based representation is superior to its vision-based counterpart when used in texture boundary detection.

Tactile representations were also found to be more discriminable (LDA). I expect these results to help better understand the nature of texture perception and help build more powerful texture processing systems.

CHAPTER IV

DEVELOPMENT: DIFFERENTIAL DEVELOPMENT OF TACTILE AND
VISUAL RFS

From an information processing point of view, touch and vision are very similar (i.e., two-dimensional sensory surface). As we have seen in the previous chapter, that feature space induced by tactile receptive fields (TRFs) are better than that by visual receptive fields (VRFs) in texture boundary detection tasks. This suggests that TRFs could be intimately associated with texture-like input. Then, the question is what drives the two types of RFs to become different in the first place? In this chapter, I investigated the possibility that TRF and VRF emerge based on an identical cortical learning process, where the only difference is in the input type: natural-scene-like vs. texture-like. The main hypothesis is that TRFs can be self-organized using the same cortical development mechanism found in the visual cortex, simply by exposing it to texture-like inputs, as opposed to natural-scene-like inputs. To test this hypothesis, the LISSOM (Lateral Interconnected Synergetically Self-organizing Map) model of visual cortical development was used. The main results show that texture-like inputs lead to the self-organization of TRFs while natural-scene-like inputs lead to VRFs. These results suggest that TRFs can better represent texture than VRFs.

A. Input-driven Development of Receptive Fields

Sensory neurons in the primary sensory cortices in the brain preferentially respond to specific patterns of input. For example, neurons in the primary visual cortex (area 17, V1) have oriented Gabor-like receptive fields (RFs) [8] (Fig. 27). Interestingly, neurons in the somatosensory area 3b exhibit similar RF properties, with a subtle difference [7, 9, 10]. In area 3b, neurons respond to tactile input from the finger

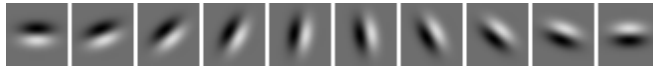


Fig. 27. Visual Receptive Fields (VRFs). Visual cortical receptive fields have a Gabor-like pattern, with different orientation, phase, and spatial frequency. Shown here are oriented Gabor patterns with the same phase and spatial frequency (dark area represents inhibitory region and bright area excitatory region).

tip, and just like in the visual cortex, they only respond to a specific pattern of input. However, there is a difference between the tactile receptive fields and visual receptive fields. Instead of an excitation/inhibition pair as in Gabor patterns, there is an extra third component that is inhibitory, where the position of that component dynamically changes depending on the direction of scan of the tactile patch (Fig. 28). Given that the two sensory modalities (vision and touch) have the same basic spatial organization (i.e., a 2D sensory surface), and that the cortex is a fairly uniform medium, it is curious as to why the two RF types show this kind of difference.

One obvious reason could be that the types of input stimulating the two modalities differ in their statistical characteristics. Vision is exposed more to natural scenes containing various objects and backgrounds that do not repeat over space (on a large scale), while touch is exposed more to surface texture with a regular repetition of pattern in all directions. In other words, visual RFs could have adapted to deal with natural scenes, while tactile RFs adapted to handle textures.

Texture is basically a surface property, so it may be more intimately related to touch, thus tactile RFs would be better for texture processing than visual RFs. The experiment on texture boundary detection in chapter III indicates that this could be the case, i.e., preprocessing with tactile RFs gives better texture boundary detection performance compared to visual RFs.

In this chapter, we will explore the possibility that tactile RFs and visual RFs emerge based on an identical cortical learning process, where the only difference is

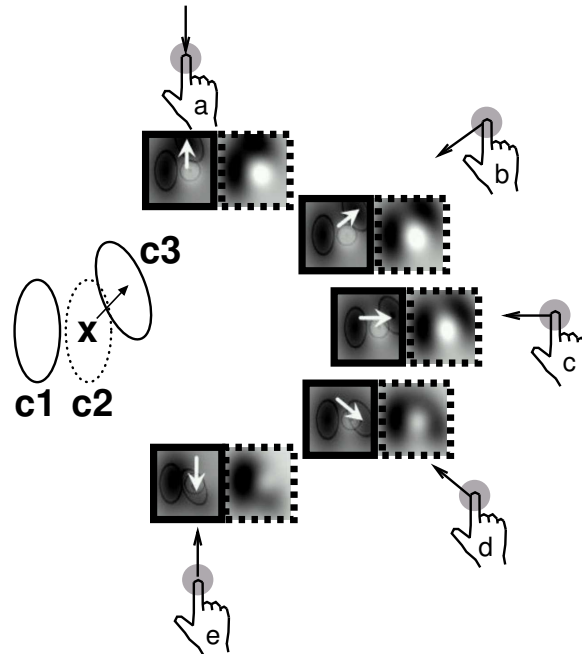


Fig. 28. Tactile Receptive Fields (TRFs). Tactile receptive fields are similar to visual receptive fields (marked C1 and C2, representing inhibitory and excitatory blobs) but there is an added dynamic inhibitory component (marked C3). An interesting feature of this extra inhibitory component is that its position relative to the fixed components C1 and C2 change, centered at “X”, depending on the direction of scan of the tactile surface (e.g., the tip of the index finger). The dynamic component’s shift in position is in the opposite direction of the scan direction. The five groups of figures to the right show how scan direction alters the tactile receptive field property. In each group, the arrow on the finger tip shows the scan direction; the box with a solid outline shows how the dynamic inhibitory component is shifted (white arrow) in the opposite direction of the scan; and the box with the dotted outline shows the resulting tactile receptive field shape. Adapted from [55] (also see [9]).

in the input type: natural scene vs. texture. I trained a self-organizing map model of the cortex (the LISSOM model [11]) on two different kinds of input, (1) natural scene and (2) texture, and compared the resulting RFs. The main result is that RFs trained on natural scenes have RFs resembling visual RFs, while those trained on texture resemble tactile RFs. These results suggest that the type of input most commonly stimulating the sensory modality (natural scene for vision and texture for touch), not the intrinsic organization or developmental process, determine the RF property in the primary sensory cortices, and that textural inputs are implicated in the developments of TRFs.

The rest of this chapter is organized as follows. First, I will present the LISSOM self-organizing map algorithm (Sec. B), explain in detail the experimental design, and present the results (Sec. C). Finally, I will talk about some interesting perspectives and issues regarding the results (Sec. D).

B. The LISSOM and Self-Organization

In order to investigate the possibility that tactile RFs and visual RFs emerge based on an identical learning process, I trained LISSOM (Laterally Interconnected Synergetically Self-Organizing Map), a self-organizing map model of the visual cortex [11].

LISSOM was originally developed to model the visual cortex, but it is actually a more general model of how the cortex organizes to represent correlations in the sensory input. Thus, LISSOM should work equally well in modeling the development of non-visual sensory modalities, as demonstrated by [12], where somatosensory cortical development (of the barrel cortex in rodents) was successfully modeled using LISSOM.

Since tactile RFs have a dynamic component, I adopted a variant of LISSOM that can handle dynamically changing input, i.e., LISSOM model of combined orien-

tation and direction map formation. The resulting RFs in the model would have a spatiotemporal pattern.

Fig. 29 shows the LISSOM architecture for orientation and direction selectivity. The description below closely follows [11]. I mainly used the Topographica neural map simulator package for the experiments (<http://topographica.org>), developed by Bednar et al. [11]. The model is similar to a general LISSOM model consisting of two-dimensional sheet of neural units, roughly corresponding to the retina at the input level, ON- and OFF-LGN (lateral geniculate nucleus) channels at the intermediate level, and V1 neurons at the cortical level. LGN units have four sheets with different time delays for each ON and OFF channel so that V1 neurons can use these time-varying inputs to develop spatiotemporal receptive fields.

An input consists of four sequential frames of an image, moving across the retina at a certain location and direction. At each time step t , the frame t is presented on the retina, and the activities of two LGN ON/OFF cells with time t are calculated. The fixed weights for the LGN ON units are computed as:

$$L_{xy,ab} = \frac{\exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{\sigma_c^2}\right)}{\sum_{uv} \exp\left(-\frac{(u-x_c)^2+(v-y_c)^2}{\sigma_c^2}\right)} - \frac{\exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{\sigma_s^2}\right)}{\sum_{uv} \exp\left(-\frac{(u-x_c)^2+(v-y_c)^2}{\sigma_s^2}\right)}, \quad (4.1)$$

where $L_{xy,ab}$ is the weight from the retinal receptor (x, y) in the receptive field to an LGN ON neuron (a, b) with center (x_c, y_c) , and σ_c defines the width of the central Gaussian and σ_s the width of the surround Gaussian. The size of the central Gaussian was set to 0.07385 and the size of the surround Gaussian to 0.29540. The OFF neuron weights are the negative of the ON weights.

After receiving input from the retina, the lateral geniculate nucleus (LGN) units compute their responses as a squashed weighted sum of the total received activation:

$$\xi_{ab} = \sigma \left(\gamma_L \sum_{xy} \chi_{xy} L_{xy,ab} \right), \quad (4.2)$$

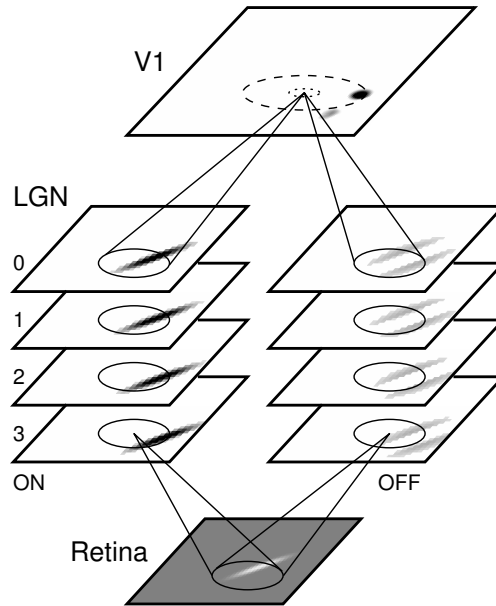


Fig. 29. LISSOM Model of Orientation and Direction Selectivity. Moving input patterns are drawn on the retina in discrete time steps, like frames of a movie. At each time step, the input pattern (or the gaze) is moved slightly on the retina and lateral geniculate nucleus (LGN) cells with time step index 3, 2, 1, and 0 each compute their activity with varying delay from the retina. Once all LGN cells have been activated, the initial V1 response is computed based on the responses on the eight LGN sheets. The activity then spreads laterally within V1 through excitatory (small dotted circle in V1) and inhibitory (large dashed circle in V1) connections. Adapted from [11].

where ξ_{ab} is the response of the LGN ON/OFF unit (a, b) , χ_{xy} is the activation of retinal unit (x, y) within the receptive field of (a, b) , $L_{xy,ab}$ is the afferent weight from (x, y) to (a, b) , and γ_L is a constant scaling factor for LGN's afferent weight. To produce activity for low-contrast inputs of images, we set γ_L to 4.7 which is double that of the Gaussian width. Here, σ is a piecewise-linear approximation of the sigmoid activation function:

$$\sigma(s) = \begin{cases} 0, & s \leq \theta_l \\ (s - \theta_l)/(\theta_u - \theta_l), & \theta_l < s < \theta_u \\ 1, & s \geq \theta_u \end{cases}, \quad (4.3)$$

where s is the activation level of the neuron, θ_l is the lower bound and θ_u is the upper bound. While the default value for the initial upper bound ($\theta_u = 0.038$) in the Topographica package was used, the initial sigmoid lower bound was set to a lower value ($\theta_l = 0.076$) than the default value to allow responses to low-contrast stimuli. Both of these parameters were gradually adjusted during self-organization to enhance the self-organization results. Fig. 30 shows the piecewise linear approximation of the sigmoid activation function (Eq. 4.3).

Fig. 31 shows typical inputs used for training, and Fig. 32 shows how the inputs are sampled, and subsequently fed into the LGN layers. Moving input patterns following a scanning direction are projected on the retina in discrete time steps, like frames of a movie. At each time step, LGN cells compute their activities based on the moving input pattern on the retina.

After all four frames are drawn on the retina, one after another, the LGN sheets are activated in sequence, with the prescribed time delay. Then, each V1 neuron computes its initial response projected from the activation on all eight LGN ON/OFF sheets. The initial response of V1 neurons is computed as a weighted sum of activation

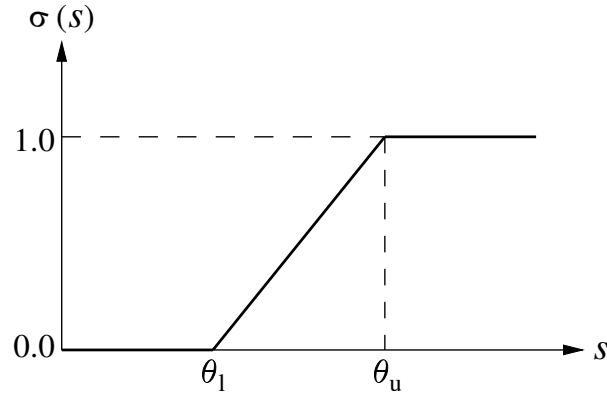


Fig. 30. Neuron Activation Function $\sigma(s)$. An computationally efficient approximation of the sigmoid function is shown with upper threshold θ_l and lower threshold θ_u . The output activation values of a neuron are limited from 0 to 1. Adapted from [11].

received from the LGN and then passed through a sigmoid activation function:

$$s_{ij} = \gamma_A \left(\sum_{ab \in ON} \xi_{ab} A_{ab,ij} + \sum_{ab \in OFF} \xi_{ab} A_{ab,ij} \right) \quad (4.4)$$

$$\eta_{ij}(0) = \sigma(s_{ij}), \quad (4.5)$$

where $\eta_{ij}(0)$ is the initial response of V1 neuron (i, j) , s_{ij} is the afferent activation of V1 neuron (i, j) , ξ_{ab} is the activation of LGN ON/OFF neuron (a, b) in the receptive field of V1 neuron (i, j) , $A_{ab,ij}$ is the afferent weight, and γ_A is a constant scaling factor for the afferent weight. We set the γ_A to 1 which is a default value for LISSOM simulations in Topographica.

After the initial computation, V1 calculates lateral excitatory and inhibitory contributions to settle the activity:

$$\eta_{ij}^{\text{new}}(t) = \sigma(s_{ij} + \gamma_E \sum_{kl} \eta_{kl}^{\text{pre}}(t-1) E_{kl,ij} - \gamma_I \sum_{kl} \eta_{kl}^{\text{pre}}(t-1) I_{kl,ij}), \quad (4.6)$$

where $\eta_{kl}^{\text{pre}}(t-1)$ is the activity of the neighbor the V1 neuron (k, l) in the previous time step, $E_{kl,ij}$ is the excitatory lateral connection weight connecting neuron (i, j)

and (k, l) , $I_{kl,ij}$ is the inhibitory lateral connection weight, and γ_E and γ_I are scaling factors that determine the strength of excitatory and inhibitory lateral interactions.

Because image patterns have significant long-range correlations, and inhibitory weights spread over a larger area, the lateral interaction factor γ_E and γ_I were set to 0.9 and -0.9 to keep the balance between excitatory and inhibitory lateral weights approximately constant.

After the activity settles, the afferent and lateral connection weights of V1 neurons are modified according to the Hebbian learning rule:

$$W_{pq,ij}^{\text{new}} = \frac{w_{pq,ij}^{\text{cur}} + \alpha X_{pq} \eta_{ij}}{\sum_{uv} (w_{uv,ij}^{\text{cur}} + \alpha X_{uv} \eta_{ij})}, \quad (4.7)$$

where $w_{pq,ij}^{\text{cur}}$ is the current connection weight from neuron (p, q) to (i, j) , $w_{pq,ij}^{\text{new}}$ is the new connection weight, α is the learning rate for each type of connection, X_{pq} is the presynaptic activity after settling, and η_{ij} is the activity of neuron (i, j) after settling.

To enhance the resulting self-organization of the lateral inhibitory weights into long range regions, the lateral inhibitory learning rate α was updated over time; 0.090365 at first, 0.090365×2 at 1,000 iterations, 0.090365×3 at 2,000 iterations, and 0.090365×5 at 5,000 iterations (see [11] for detailed justification).

C. Spatiotemporal Emergence of TRFs and VRFs

The purpose of the main experiment in this chapter was to test the possibility that tactile RFs and visual RFs emerge based on an identical learning process, where the only difference is in the input environment, natural scene vs. texture. The self-organizing map model of the cortex (the LISSOM model) was used on two different kinds of input, (1) natural scene and (2) texture, and the spatiotemporal properties of the resulting RFs were analyzed. Fig. 31 shows the two types of input patterns.



Fig. 31. Sample Input Patterns. The top row shows natural scenes and the bottom row textures used in the experiments. Note that the texture set has texture elements at varying scales. Adapted from previously reported results in [57].

For these experiments, a variant of the LISSOM was used, that can learn both the orientation and the direction of the visual stimuli. Given an image, I randomly picked an initial location and moved the gaze window in a random direction along a straight line at a fixed interval as shown in Fig. 32. Each gaze location gave a 48×48 image patch that was the same size as the retina in the LISSOM simulation. All four images in each image type were used to generate the input sequence for each simulation. All the LGN sheets were 24×24 in size, and the single V1 sheet was 48×48 in size.

All simulations in this work were based on the same set of default parameters in the Topographica package, with small modifications described in the previous section. Specifically, I set the speed of the input pattern (i.e., the number of retinal units the pattern moves between time steps) to 4 pixels/time step. This value was determined experimentally, as neurons showed the best selectivity for direction of motion at that speed. After training for several thousand iterations (usually between 10,000 to 20,000 iterations), the network developed patterned afferent and lateral connections for retina to LGN, LGN to V1 and V1 to V1. I focused on the projection from LGN

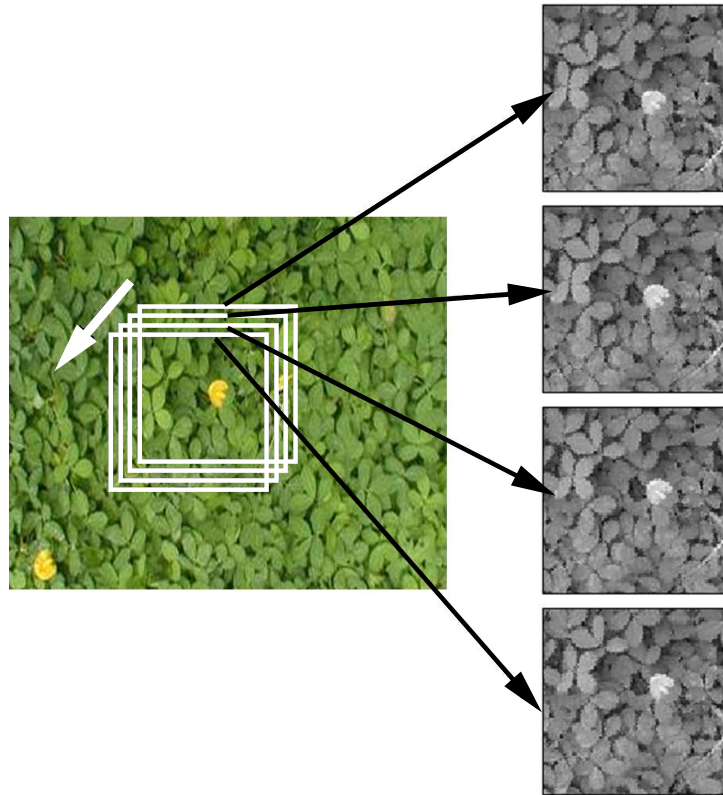


Fig. 32. Generation of Dynamic Input by Scanning the Gaze on an Image. Given a large image, a small region the size of the retina in the LISSOM model is attended to. Motion of the gaze window results in a sequence of inputs being generated on the LISSOM retina, which in turn activates the LGN ON/OFF sheets, one by one depending on the sheet's built-in delay. Adapted from previously reported results in [57].

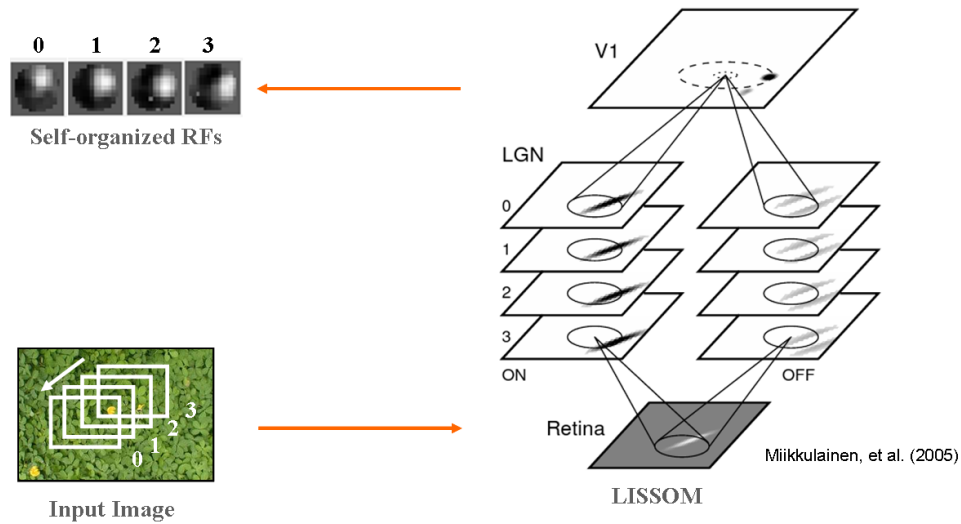


Fig. 33. Self-organization Process with LISSOM. Given a large image, motion of the gaze window results in a sequence of inputs being generated on the LISSOM retina, which in turn activates the lateral geniculate nucleus (LGN) ON/OFF sheets, one by one, depending on the sheet's built-in delay. After projecting the activities from the LGN ON/OFF sheets, V1 (the primary visual cortex) self-organizes its RFs and lateral connections (excitatory and inhibitory). LISSOM figure adapted from [11].

to V1 because they represent the spatiotemporal character of the RF. Fig. 33 shows the experimental process I followed to develop self-organized RFs. I generated input stimulus that are natural-scene-like or texture-like, while sampling across the input image with the retina.

Fig. 34 shows the self-organized RFs of six representative neurons trained with natural scenes. Nearly all neurons developed spatiotemporal RFs strongly selective for both direction and orientation. That is, each neuron is highly responsive to a line with a particular orientation moving in a direction perpendicular to that orientation. The receptive fields consist of white (excitatory) and black (inhibitory) lobes according to the preferred orientations and direction of the neuron, showing spatiotemporal preference. Such properties of the receptive fields are similar to those of the receptive fields of neurons found experimentally in the visual cortex [33] (cf. Fig. 10).

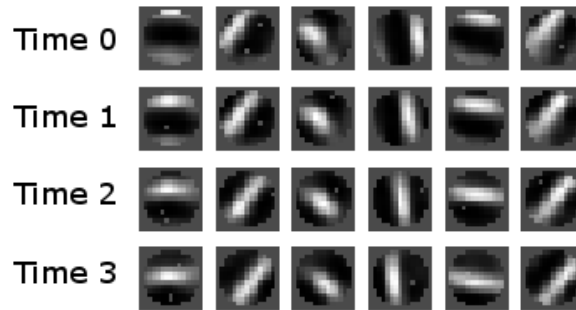


Fig. 34. Six Spatiotemporal RFs Resulting from Self-Organization on the Natural Scene Input Set. Six spatiotemporal RFs from the natural scene experiment are shown. Each column corresponds to an individual neuron’s RF, and each row represents the different time-lag. Within each column, we can see that the pattern moves in a direction perpendicular to the orientation preference. Adapted from previously reported results in [57].

The overall layout of the RFs developed in this simulation, trained with four natural scenes, are shown in Fig. 35 by plotting roughly every 3rd neuron horizontally and vertically. A number of two-lobed RFs can be seen with strong orientation preferences except some neurons which have nonlinear-shape and respond to all directions. This figure only shows the first frame among the total of four (note that these are spatiotemporal RFs).

The self-organized RFs produced from LISSOM after 20,000 training iterations with the texture input set are visualized in Fig. 36. The neurons developed spatiotemporal RFs strongly resembling tactile RFs found in the experimental literature (cf. Fig. 16). Excitatory and inhibitory components of each neuron consists of “ring- and blob-like” features as found in [9]. For example, the first three columns in Fig. 36 closely resemble the RFs in Fig. 16. Note that these RF shapes arise not because circular texture elements dominate the texture input set used in the training. The texture input set show texture elements at varying scales, and also the size of the receptive field (15×15) is usually smaller than the round or oval features in the

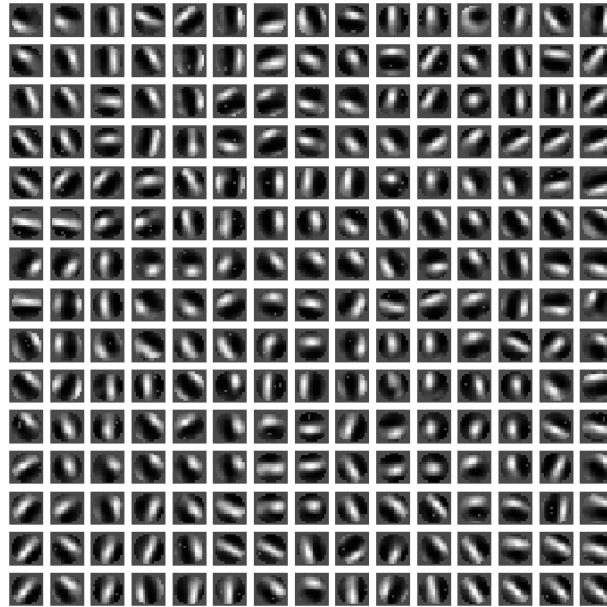


Fig. 35. Detailed View of RFs Resulting from Self-Organization on the Natural Scene Input Set. From the 48×48 cortex, only 15×15 are plotted (roughly every 3rd RF) for a detailed view of the RFs. The RFs mostly resemble visual RFs. Adapted from previously reported results in [57].

texture input set. There are interesting variations (last three columns in Fig. 36) where the polarity is reversed, i.e., instead of an excitatory region in the middle and inhibitory region in the surround, these RFs have an inhibitory region in the middle and the excitatory region in the surround. Fig. 37 shows the overall organization of the cortical map (roughly every 3rd neuron's RF is shown).

The results show that exposure to different kinds of inputs can drive an identical underlying cortical learning model to develop two different kinds of RFs, tactile or visual. RFs trained on natural scenes resemble visual RFs, while those trained on texture resemble tactile RFs. These results suggest that the type of input most commonly stimulating the sensory modality (natural scene for vision and texture for touch), and not the intrinsic organization or developmental process, determine the RF property. Furthermore, the results further support the idea that texture and the

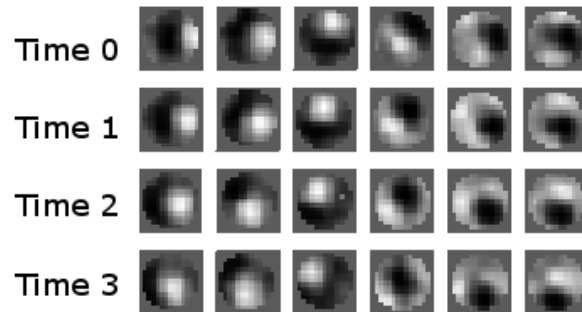


Fig. 36. Six Spatiotemporal RFs Resulting from Self-Organization on the Texture Input Set. Six spatiotemporal RFs from the texture experiment are shown. Each column corresponds to an individual neuron's RF, and each row represents the different time-lag. The RF shapes resemble the ring-like shape of tactile RFs found in the experimental literature [9]. Adapted from previously reported results in [57].

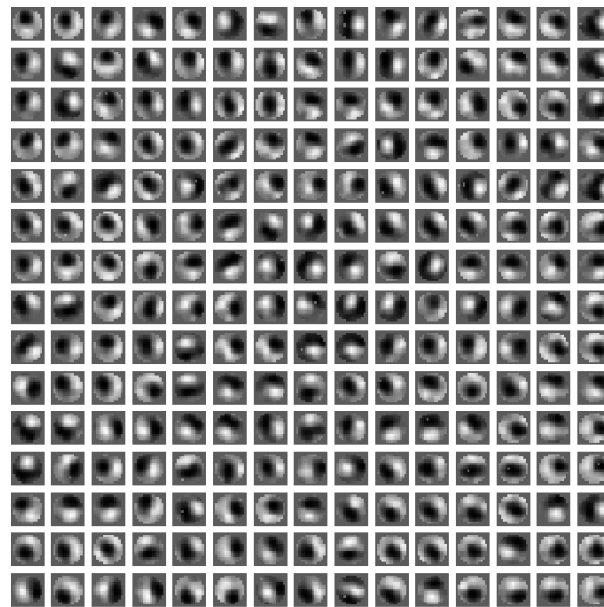


Fig. 37. Detailed View of RFs Resulting from Self-Organization on the Texture Input Set. From the 48×48 cortex, only 15×15 are plotted (roughly every 3rd RF) for a detailed view of the RFs. The RFs mostly resemble tactile RFs. Adapted from previously reported results in [57].

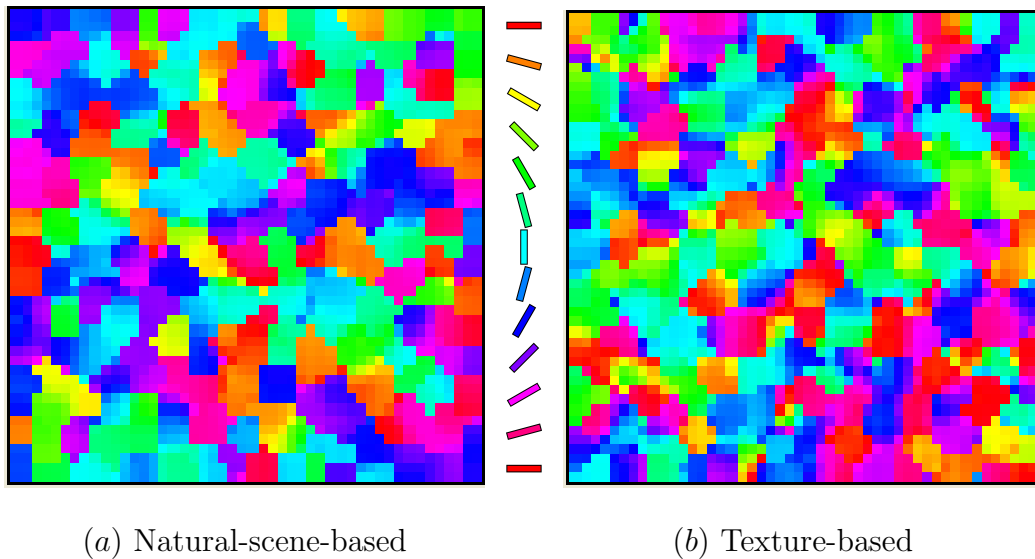


Fig. 38. Orientation Maps. The orientation maps resulting from (a) the natural scene experiment and (b) the texture experiment are shown. The color-key for the orientation is shown in the middle.

tactile modality (and TRFs) are closely related.

The experiments also turned up interesting additional results. The global organization of the RFs can be visualized just like for visual cortical orientation maps, by labeling each neuron by its preferred angle. Fig. 38 shows the overall organization of the two maps. The orientation map developed with natural scene input shows similar characteristics as those found in the visual cortex, with smooth transition across neighboring orientation domains. The map also has more red and cyan than other colors, which means more representation of near horizontal and vertical orientations, which is also observed in experiments [11].

It is interesting to note that the map trained on textures (i.e., those that develop tactile RFs) also has a rough orientation map. Note that in the tactile map case, the concept of orientation is less well defined than the visual map case, since not all RFs have a clear orientation preference. Fig. 39 shows the orientation selectivity

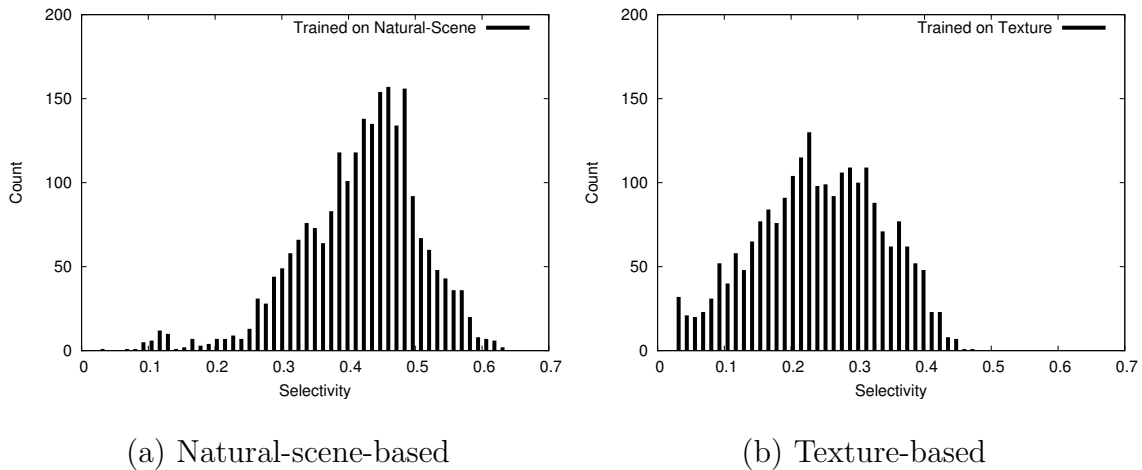


Fig. 39. Selectivity in Orientation Maps. The orientation selectivity histogram are shown for the two 48×48 V1 sheets (maps): (a) natural-scene-based map, and (b) texture-based map. As we can already see from Fig. 38, the map trained with natural scenes show much higher selectivity (peak near 0.45), compared to the case with textures (peak near 0.25). Note that higher selectivity means that RFs are more sharply tuned to one specific orientation (i.e., RFs are more slender).

histograms for the two maps: texture-based and natural-scene-based. The natural-scene based map (i.e., the “visual” map) shows a much higher orientation selectivity.

D. Discussion

The main contribution of this chapter is to have shown what drives a virtually identical learning medium (the cerebral cortex) to specialize and diverge, to represent different sensory modalities, visual vs. tactile. By comparing two structurally similar modalities of touch and vision, along with the assumption that texture (whether visual or tactile) is basically a surface property, it was shown that it is the input type (natural-scene-like or texture-like) that determines the learned RF type, not the direct sensory modality. These and results in earlier chapters [5, 55, 58] suggest an intricate relationship among touch, texture, and surface property in 3D.

The main finding of this chapter is that a common cortical development framework (LISSOM) can develop two different RF types, just based on the type of input presented during training. The results suggest that texture in general, whether it is tactile or visual in origin, contributes to the emergence of the unique properties similar to those observed in tactile RFs in area 3b of the somatosensory cortex. This is an interesting result that helps us better understand the intimate relationship among texture, surface, and touch, and further strengthens our earlier finding that tactile RFs can outperform visual RFs in texture segmentation tasks. I expect the new tactile-oriented approaches to texture segmentation to complement the traditional visually oriented approach, and help us better understand the nature of texture.

CHAPTER V

MANIFOLD ANALYSIS OF FEATURE SPACE IN TACTILE VS. VISUAL RFS

The responses of RFs occupy a high-dimensional feature space, and can be difficult to interpret. One interesting approach to address the difficulty is to assume that the response distribution lies on a non-linear low-dimensional manifold embedded within the higher dimensional space. For examples, a curve embedded in a three-dimensional space [59] is a one-dimensional manifold, whereas a sphere in the same space is a two-dimensional manifold. In order to find the representational power of the RF responses, I analyzed the feature space of the RF response with manifold learning methods.

A. Manifold Learning in Perception

The human brain distinguishes one image from another based on the similarity or dissimilarity between images. That is, the perceptual process is based on similarity rather than absolute coordinates. Then, what is similar or dissimilar? How does the brain perceive constancy even though its raw sensory inputs are in flux? One more challenge that the human brain confront in everyday perception is to find meaningful organizational structure hidden in large volumes of high-dimensional data [60].

Scientists in many fields facing the problem of simplifying high-dimensional data utilize manifold learning to find low-dimensional structure in the data. Recently, various methods (for example see [60–63]) have been developed in the machine learning community and their applications started to draw attention in pattern recognition, signal processing, and robotics. The approaches all map a given set of high-dimensional data points into a surrogate low-dimensional space. First of all, a pre-processing step decides for each data point which of the other data points should be

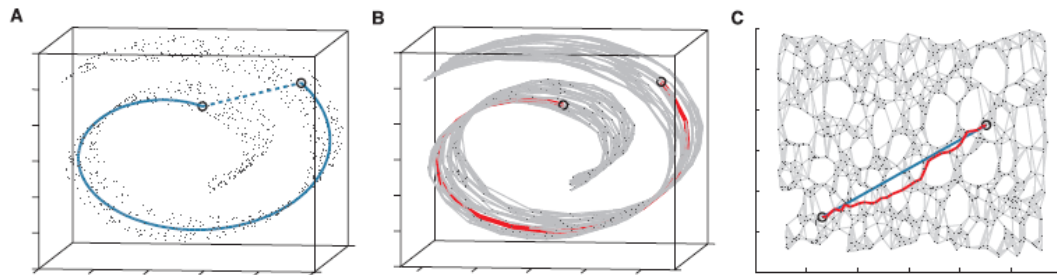


Fig. 40. The Swiss Roll Data Set Illustrating Nonlinear Dimensionality Reduction. (A) Euclidean distance (dashed blue line) of two points (circles) in the high-dimensional input space may not accurately characterize their intrinsic similarity (solid blue line). (B) Isomap compute the geodesic path which is locally Euclidean distance (red segmentations) to find the true short path. (C) The two-dimensional embedding recovered by Isomap, which best preserves the shortest path distances. Straight lines in the embedding (blue line) now represent clearer approximations to the true geodesic paths than do the corresponding graph paths (red line). Adapted from [60].

considered its neighbors. Then measures of the local geometry of the manifold are computed, after which the original data points are no longer needed.

Fig. 40 illustrates the swiss roll data set illustrating nonlinear dimensionality reduction. While the distance of two points measured by straight-line Euclidean distance may appear deceptively close in the high-dimensional input space (Fig. 40A, dotted line), the geodesic distances (Fig. 40A, B, solid curve) reflect the true low-dimensional geometry of the manifold.

An image can be regarded as a large volume of data set, such as light intensity at an image pixel and the Cartesian coordinates of a point with respect to a set of axes. Therefore, any image can be distinguished as a point in an abstract, high-dimensional image space.

Fig. 41 shows a simple example of perceptual manifold defined on visual photoreceptor responses [64]. The two sets of facial images are generated by varying the orientation of the two faces. These sets are represented by continuous curves in the

abstract space, because the images vary smoothly as the face is rotated. Each curve is one dimensional because it is generated by changing single parameter, the angle. In other words, each set is intrinsically one-dimensional, although it is embedded in image space, which has a high dimensionality equal to the number of image pixels. If we were to allow more degree of freedom, such as scaling and luminance, then the dimensionality of the manifold would increase.

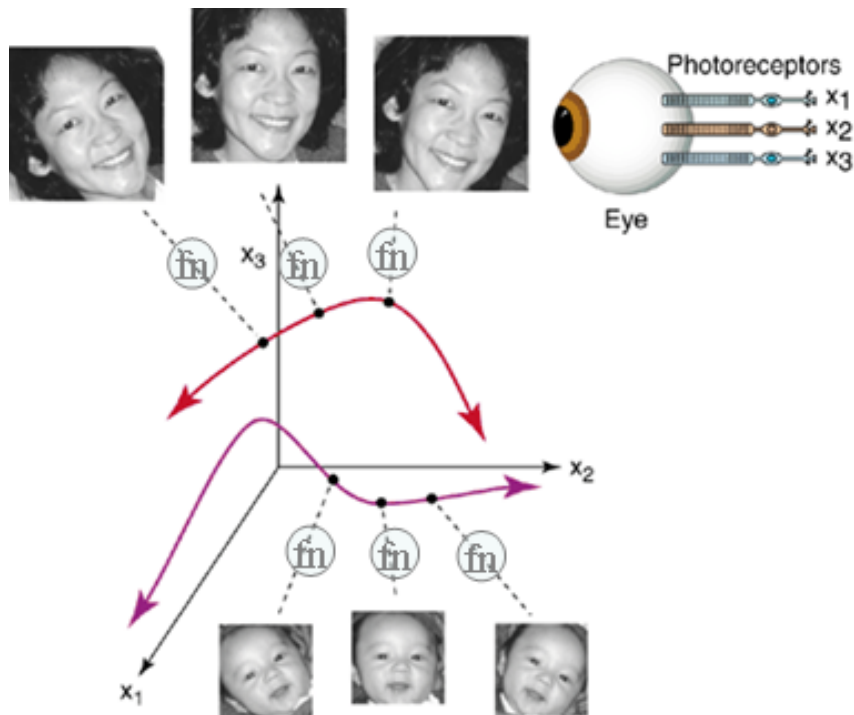


Fig. 41. Manifold Learning in Visual Perception. The retinal image is a collection of signals from the photoreceptor cells. If the number of these cells is taken to be the coordinates in an abstract image space, then an image can be represented by a single point. To discriminate faces, the brain must treat all images from the same manifold as being the same, based on the manifold reconstructed from observed data up to that point. Adapted from [64].

In Fig. 41 the image space is depicted as a three dimensional space, but actually the dimensionality is equal to the number of photoreceptor cells. As the faces are rotated, they trace out nonlinear curves embedded in the image space. If changes in

rotation of head, direction of light, distance to see, and other sources of continuous variability are also included, then the images would lie on manifolds with dimension higher than one but still lower than that of the image space. In this generalized case, the set is said to be a manifold embedded in the image space. For examples, a curve represents a one-dimensional manifold, whereas a sphere represents a two-dimensional manifold, embedded in a three-dimensional space [59].

Although only the visual manifold has been discussed above, manifolds can be also relevant to other types of perception, such as touch, hearing, and taste, because the brain must have some way of representing all types of perception which is very similar way of visual system [64]. Hence, manifold learning could be crucial for understanding how perceptual information arises from the activity in large populations of neurons in the brain.

The rest of this chapter is organized as follows. Section B describes the embedded manifolds in RF space, recorded using manifold learning methods. In section C and D, the experiment of texture classification was illustrated. Section E discusses issues arising from this work, followed by some results.

B. Embedded Manifold in RF Response Feature Space

Based on “manifold ways of perception”, a question arises as to how to find the manifold of RF responses and interpret its representation. Visual or tactile RF responses could be a collection of signals from the receptor cells. If this collection is taken to be coordinates in an abstract space, then an image can be represented by a point. Fig. 42 shows how tactile responses can form a manifold, analogous to the visual counterpart shown in Fig. 41. Each curve is an example of a one-dimensional manifold that represents one texture object in the feature space.

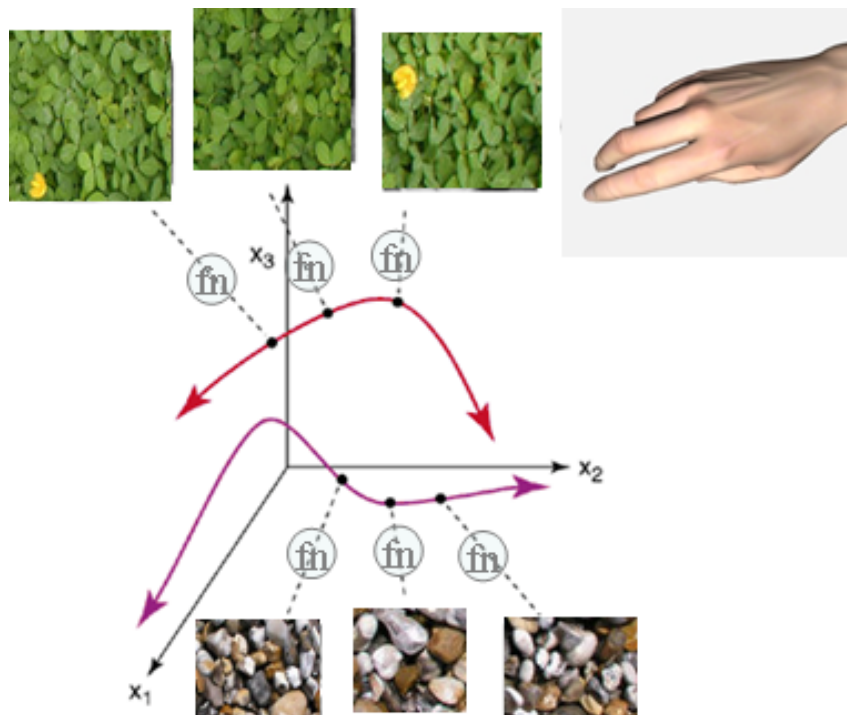


Fig. 42. Manifold Learning in Tactile Perception. The tactile texture is a collection of signals from the finger tip. If these numbers are taken to be coordinates in an abstract tactile space, then a single tactile impression can be represented by a point. To discriminate textures, the brain must equate all impressions from the same texture based on the learned manifold. Adapted from [64].

I used a non-linear manifold learning method, kernel Isomap [13] to project response data of the tactile receptive fields (TRFs) to a low dimensional space, and compared the result to that based on visual receptive field (VRF) responses.

Below is the kernel Isomap algorithm [13] I used in the analysis of the feature space of RF responses. Given N objects with each object being represented by an m -dimensional vector \mathbf{x}_i , $i = 1, \dots, N$, the kernel Isomap algorithm finds an implicit mapping which places the N points in a low-dimensional space. Kernel Isomap mainly exploits the additive constant problem, the goal of which is to find an appropriate constant to be added to all dissimilarities (or distances), apart from the self-dissimilarities, that makes the kernel matrix to be positive semidefinite.

Given a distance matrix, we calculate Dijkstra's geodesic distance (shortest path) \mathbf{D} and the doubly centered kernel matrix as below

$$\mathbf{K} = -\frac{1}{2}\mathbf{H}\mathbf{D}^2\mathbf{H}, \quad (5.1)$$

where $\mathbf{D}^2 = [D_{ij}^2]$ means the element-wise square of the geodesic distance matrix $\mathbf{D} = [D_{ij}]$, \mathbf{H} is the centering matrix, given by $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{e}_N\mathbf{e}_N^\top$ for $\mathbf{e}_N = [1 \dots 1]^\top \in \mathbb{R}^N$. Then, we can make the kernel matrix positive definite by adding a constant, c .

$$\widetilde{\mathbf{K}} = \mathbf{K}(\mathbf{D}^2) + 2c\mathbf{K}(\mathbf{D}) + \frac{1}{2}c^2\mathbf{H}, \quad (5.2)$$

where c is the largest eigenvalue of the matrix

$$\begin{bmatrix} \mathbf{0} & 2\mathbf{K}(\mathbf{D}^2) \\ -\mathbf{I} & -4\mathbf{K}(\mathbf{D}) \end{bmatrix}. \quad (5.3)$$

Eq. 5.2 implies substituting $\widetilde{\mathbf{D}}$ for \mathbf{D} in Eq. 5.1, which is given by

$$\widetilde{D}_{ij} = D_{ij} + c(1 - \delta_{ij}), \quad (5.4)$$

which makes the matrix \mathbf{K} to be positive semi-definite. The term δ_{ij} is the Kronecker delta. Finally, projection mapping \mathbf{Y} is given by Eq. 5.5 after eigen-decomposition, $\widetilde{\mathbf{K}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$.

$$\mathbf{Y} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}. \quad (5.5)$$

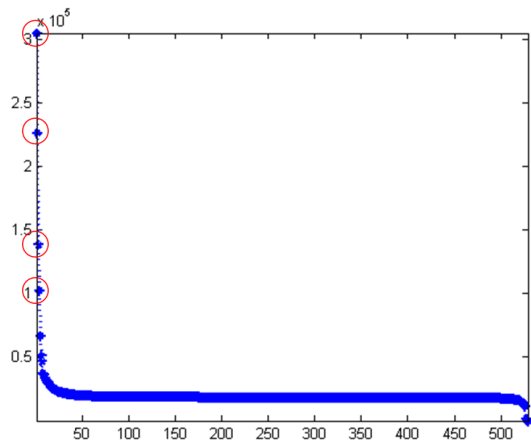
The projection for the test data points is similar to kernel PCA described in [13].

In order to understand the properties of the RFs, I checked the eigenvalues of the embedded manifolds in the RF space as well as the projection of the RFs onto the the dominant two-dimensional manifold.

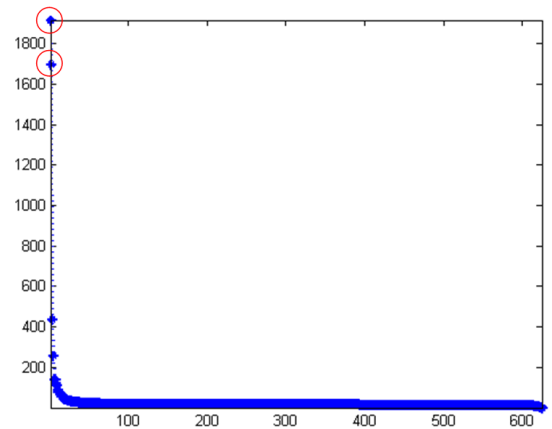
Fig. 43 shows the eigenvalues of the embedded manifolds for the designed RFs: Gabor-like filters which are widely used as a model of VRFs, and the three-component model developed by DiCarlo and Johnson [9] as a computational model for TRFs. Basically, Gabor filters have more dominant factors than the three-component model, (i.e., intrinsic manifold has a higher dimension). Also, the noise terms in the Gabor filters have more relatively high positive values than the three-component model. The three-component model has two dominant factors with the others having relatively small eigenvalues.

Note that in Fig. 43a, even though the number of parameters used to generate the Gabor filters were two (orientation and phase), they have more than two dominant eigenvalues, which means the images are really sensitive to the parameters. When the parameters change a little, the extracted manifold changes a lot. On the other hand, the three-component model is also generated using two parameters (orientation and scanning direction) and the embedded manifold is also described by two dominant eigenvalues.

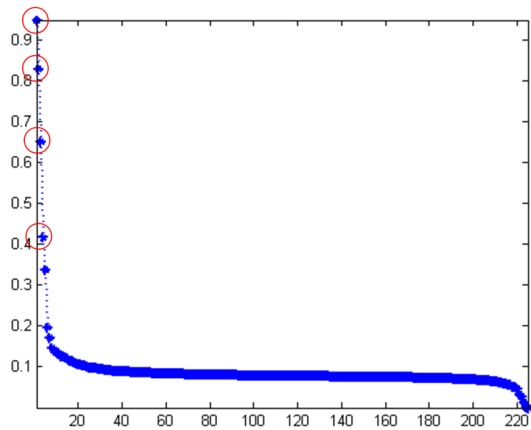
I also checked the two learned receptive field types from the LISSOM simulation (trained with natural scene images and texture images). I expected the RFs from



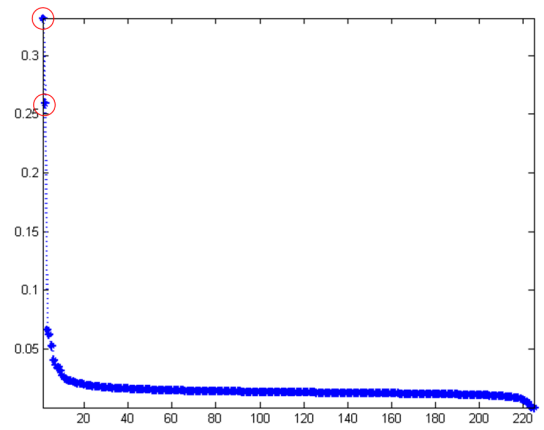
(a)



(b)



(c)



(d)

Fig. 43. Eigenvalues Extracted from the Embedded Manifold of Designed Filters and Self-organized Receptive Fields. (a) Gabor filters, (b) three-component model, (c) self-organized VRFs, and (d) self-organized TRFs. Red circles represent the dominant eigenvalues of the embedded manifold.

natural scene images to be like Gabor filters and the RFs from texture images to be like the three-component model. Figs. 43*c* and *d* show two plots of the eigenvalues for VRFs and TRFs learned from natural scenes and textures. Figs. 43*c* has a property similar to Fig. 43*a* with 4 to 5 dominant eigenvalues and relatively high noise terms, while Fig. 43*d* looks like the graph of the eigenvalues of three-component model (Fig. 43*b*) which had two dominant factors with relatively small noise terms.

Furthermore, the RFs were projected to a two-dimensional space (their manifolds). In Fig. 44, each point means one RF (or filter). The two top figures are projections of the designed filters, Gabor filters and three-component model, respectively. The two bottom figures are projections of the self-organized RFs (VRFs and TRFs), trained with natural scene images and texture images, respectively. While (b) and (d) are pretty much alike, which I expected, (a) and (c) are not so much alike. This may be due to the lack of certain information for the VRFs in this limited space.

Based on the eigenvalues and the two-dimensional projections, we can see that the VRFs from natural scene have properties similar to the Gabor filters widely used as the computational model of V1 simple cells [8], and the TRFs from texture have properties similar to the three component model by DiCarlo and Johnson [9].

C. Texture Classification with Manifold

Based on previous results on the performance of tactile vs. visual RFs in texture tasks, interesting questions arise whether tactile manifolds are better for classifying textures than visual manifolds. I tested texture classification performance between tactile vs. visual manifolds. Fig. 45 shows the experimental process. Given texture inputs, I generated TRF and VRF responses, calculated their manifold, and classified

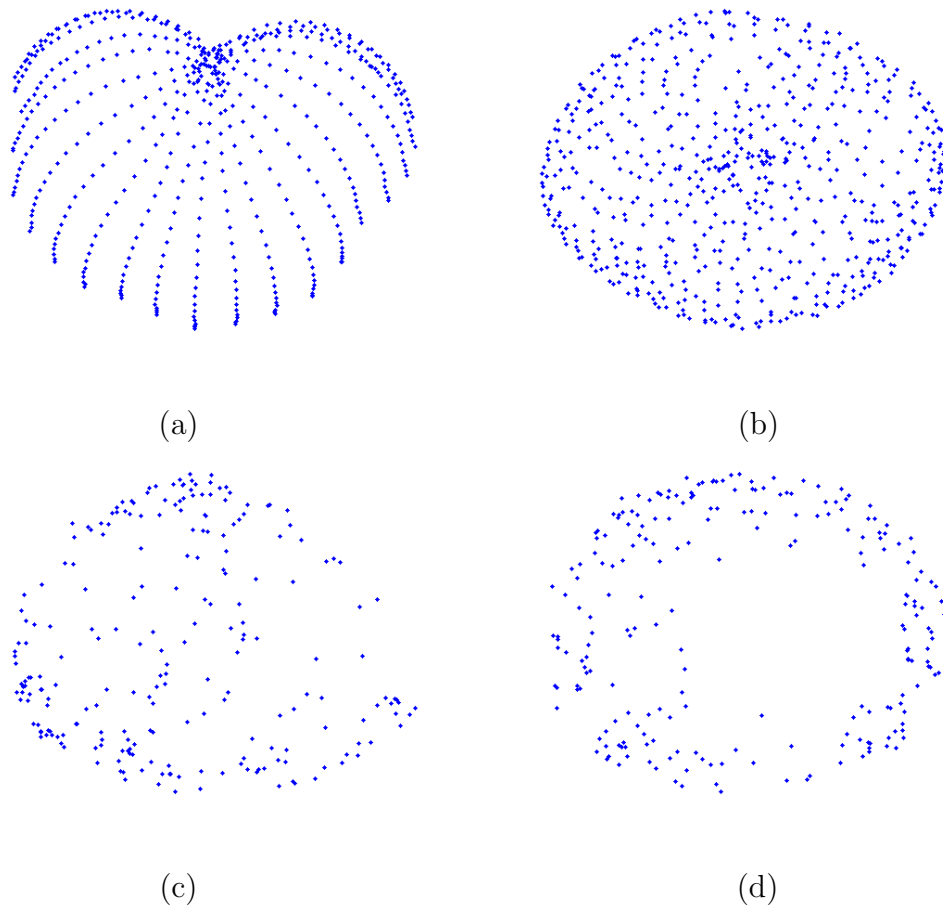


Fig. 44. Projection of 4 Different Types of Filters Mapped onto the Embedded Manifold of Two Dimensions. (a) Gabor filters, (b) three-component model, (c) VRFs trained with natural scene images, and (d) TRFs trained with texture images.

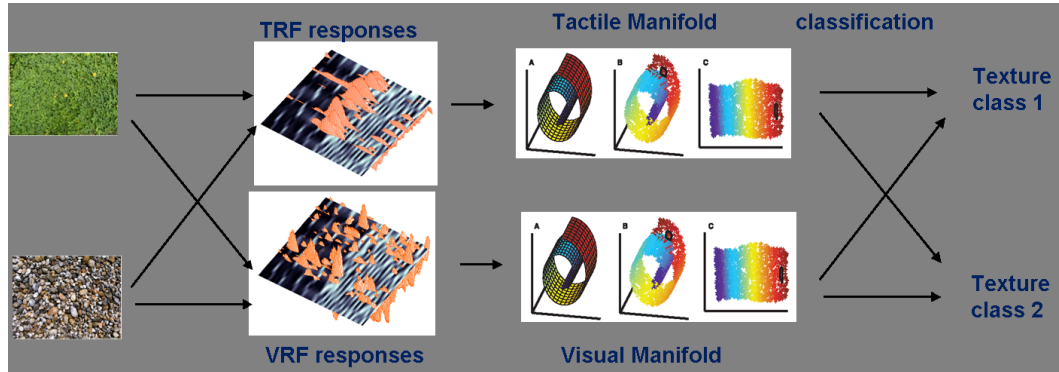


Fig. 45. Texture Classification Process with Visual and Tactile Manifolds. The overall procedure was as follows: (1) generate texture response with TRF and VRF as in chapter III, (2) calculate their manifold, and (3) classify the projection onto the manifold into different texture classes.

input textures into different classes based on the projection of the input images onto those manifolds.

In order to test the representational power of TRF responses as compared to VRF responses, I generated RF responses from 3 different textures, the same ones used in the previous chapter, and used kernel Fisher Discriminant (KFD) [65]. Here, I briefly review KFD.

KFD is a generalized version of Fisher Linear Discriminant Analysis (LDA) with an added kernel trick as in support vector machine or kernel principal component analysis [66]. The basis function in the feature space can be obtained by maximizing the ratio of the within-class scatter matrix in the feature space to the between-class scatter matrix in the feature space as in LDA. Let $\chi_i = \{x_1^i, x_2^i, \dots, x_{l_i}^i\}$, ($i = 1, \dots, C$) be samples from C classes and $\chi = \bigcup_i^C \chi_i$. Suppose $\Phi(\cdot)$ is a nonlinear mapping function to feature space, then the within-class scatter matrix in feature space \mathbf{S}_W^Φ is given by

$$\mathbf{S}_W^\Phi = \sum_{i=1}^C \sum_{x \in \chi_i} (\Phi(x) - m_i^\Phi)(\Phi(x) - m_i^\Phi)^T \quad (5.6)$$

where $m_i^\Phi = \frac{1}{l_i} \sum_{j=1}^{l_i} \Phi(x_j^i)$. The between-class scatter matrix in feature space is given by

$$\mathbf{S}_B^\Phi = \mathbf{S}_T^\Phi - \mathbf{S}_W^\Phi \quad (5.7)$$

where the total scatter matrix in feature space \mathbf{S}_T^Φ is given by

$$\mathbf{S}_T^\Phi = \sum_{x \in \chi} (\Phi(x) - m^\Phi)(\Phi(x) - m^\Phi)^T \quad (5.8)$$

where $m^\Phi = \frac{1}{|\chi|} \sum_{i=1}^C l_i m_i^\Phi$ and $|\chi|$ is the sample size.

I applied KFD to the responses of both TRF and VRF. Fig. 46 shows the two different embedded manifolds based on textures in two-dimensional space. To see more clear separability, I used ellipses that are equidistant traces from each class' center (mean) at 1.5 times standard deviation. I used a polynomial function as the kernel function for both cases. The figure shows that the responses of the TRFs are more clearly clustered than those of the VRFs. In order to compare textural feature and natural feature, I also generated responses of 3 different natural scenes. Fig. 47 shows the two different embedded manifolds based on natural scenes in two-dimensional space.

For the classification task, the experiments were run 30 times and for each experiment 50% of the data set was randomly chosen as training data and the rest as testing data. As the classifier, k-nearest neighbor (kNN) was used. Fig. 48 shows the boxplot of the classification rate based on both RFs, TRF and VRF, with texture-like input. The averages are 89.15% and 85.61%, respectively. We can see that TRF is better than VRF in the texture classification task. Note that the LDA result (which is based on a linear manifold) for TRF is 78.90% which means that the responses of TRF lie on a curved manifold. Another interesting thing is that the standard deviation in TRF (0.054) is smaller than that in VRF (0.059), which means that the

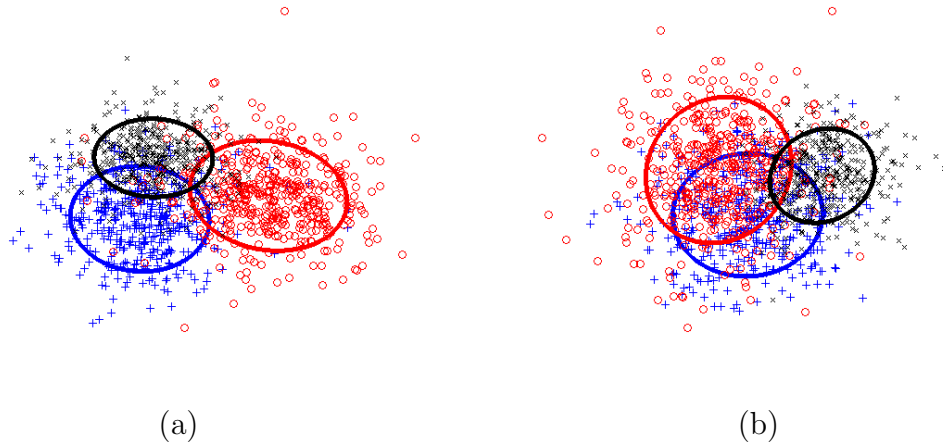


Fig. 46. Feature Spaces for Texture. KFD analysis of (a) TRF and (b) VRF responses to texture-like input are shown. In each plot, response samples from three different textures, projected on the 1st and 2nd KFD axes, are shown. The ellipses show the $1.5 \times \sigma$ equidistance trace from the class centers. We can see that the classes in (a) are more separable than those in (b). Adapted from previously reported results in [67].

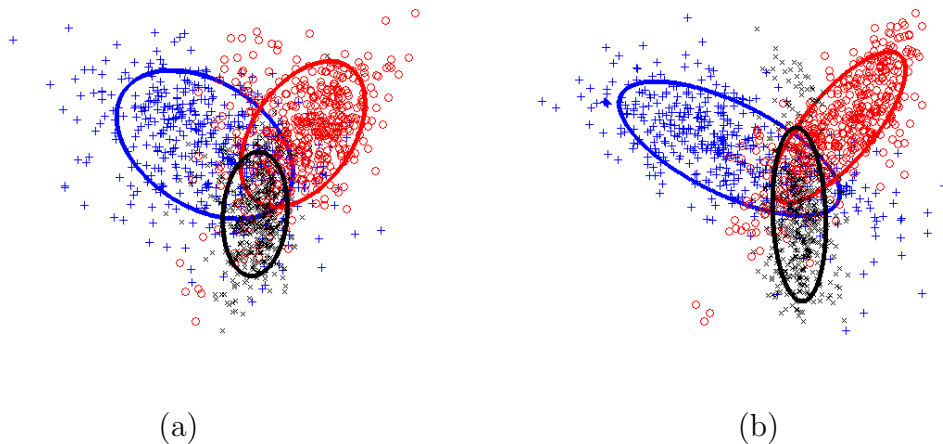


Fig. 47. Feature Spaces for Natural Scene. KFD analysis of (a) TRF and (b) VRF responses to natural-scene-like input are shown. In each plot, response samples from three different natural-scene, projected on the 1st and 2nd KFD axes, are shown. The ellipses show the $1.5 \times \sigma$ equidistance trace from the class centers. Overlap regions of (a) and (b) are similar, but the clusters in (b) are more stretched out away from the overlap region.

performance of TRF is more stable than VRF's.

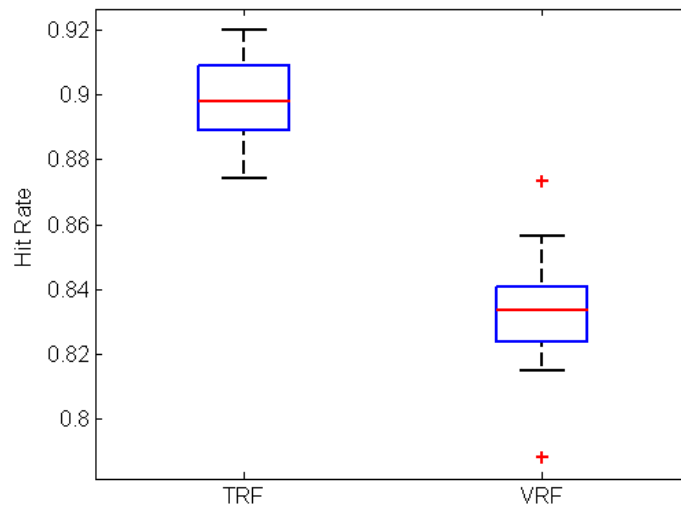


Fig. 48. Comparison of Texture Classification Rate between TRF-based Response and VRF-based Response with Texture-like Input. The band near the middle of each box is the the median, the bottom and top of the box are the 25th and 75th percentile of the data, the ends of the whiskers represent the minimum and maximum of all the data, and “+” marks outliers. Adapted from previously reported results in [67].

A similar experiment with natural-scene-like inputs show a an analogous result: VRFs are better suited for natural-scene-like inputs. Fig. 49 shows the boxplot of the classification rate for both RFs with natural-scene-like input. The averages are 76.5% and 80.42%, respectively. We can see VRF is better than TRF in natural scene classification task. By comparing classification results of two structurally similar modalities of touch and vision, I have shown that tactile RFs are more ecologically suited for texture tasks while visual RFs more for natural scene tasks.

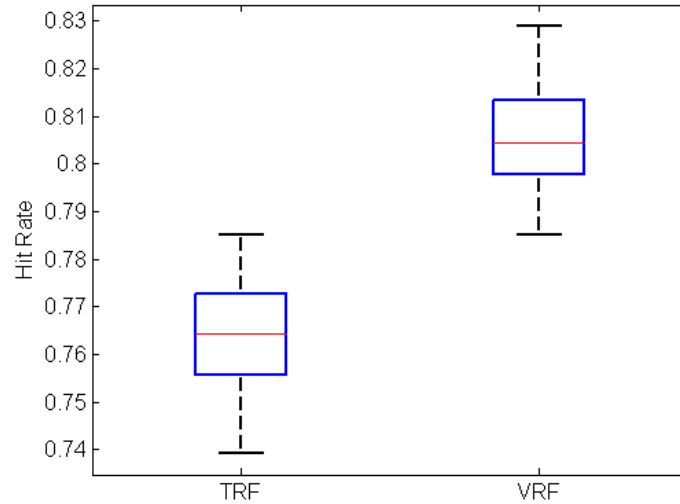


Fig. 49. Comparison of Natural-scene Classification Rate between TRF-based Response and VRF-based Response with Natural-scene-like Input. The band near the middle of the box is the the median, the bottom and top of the box are the 25th and 75th percentile of the data, the ends of the whiskers represent the minimum and maximum of all the data, and “+” marks outliers.

D. Texture Classification with Manifold of Integrated RF Responses

So far, only one sensory modality was used for each task to get comparable results to the other sensory modality: touch vs. vision. However, humans perceive the environment through a variety of senses. Many tasks can be carried out by combining several sources of information. The brain integrates this redundant information to come up with the most reliable (unbiased) estimate. For example, an object’s size and shape can be judged based on visual as well as haptic cues. It has been shown recently that human observers integrate visual and haptic size information in a statistically optimal fashion, in the sense that the integrated estimate is the most reliable one [68].

Interestingly, the manifold of integrated response from tactile RFs and visual RFs shows higher classification rates than those based on tactile only or visual only manifold (see [69] for details on manifold integration.). Fig. 50 shows the embedded

manifold of the integrated response over textures in two-dimensional space, which shows more clearly separated clusters than those based on TRF or VRF alone. The boxplot in Fig. 51 shows the classification rate after 30 repeated experiments. The results show that the integrated manifold’s performance exceeds that of the individual manifolds based on TRF or VRF responses.

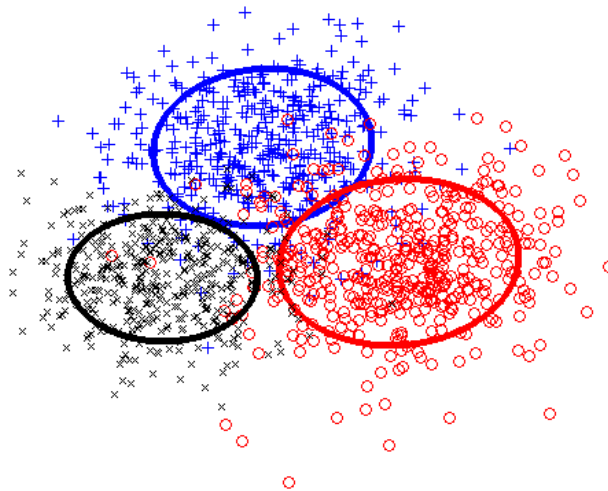


Fig. 50. Feature Spaces for Texture from the Manifold of the Integrated RF Responses. KFD analysis of integrated responses of TRF and VRF responses to three different textures are shown. Response samples, projected onto the 1st and the 2nd KFD axes, are shown. The ellipses show the $1.5 \times \sigma$ equidistance trace from the class centers. We can see that the classes are clearly separated, even more so than in Fig. 46 or Fig. 47.

E. Discussion

The main contribution of this chapter is to have analyzed the representational power of tactile RFs and visual RFs with manifold learning methods. I compared the embedded manifolds in the RF to those based on designed TRFs and VRFs that were used in chapter III [55]. I used two manifold learning methods, kernel Isomap [13], and

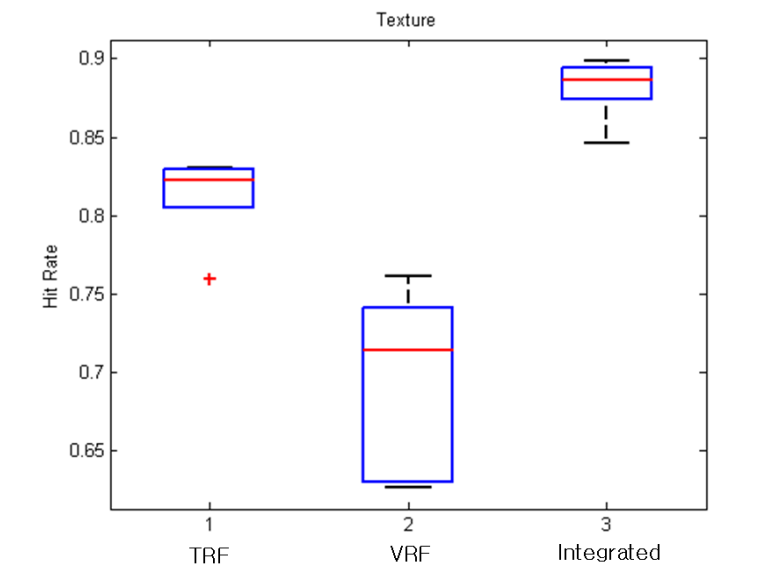


Fig. 51. Comparison of Texture Classification Rate between TRF-based Response, VRF-based Response, and Integrated Response with Texture-like input. The band near the middle of each box is the the median, the bottom and top of the box are the 25th and 75th percentile of the data, the ends of the whiskers represent the minimum and maximum of all the data, and “+” marks outliers. The manifold of integrated responses outperforms the individual manifolds.

kernel Fisher discriminant analysis (KFD) [66]. The kernel Isomap mainly exploits the additive constant trick, the goal of which is to find an appropriate constant to be added to all dissimilarities (or distances), apart from the self-dissimilarities, that makes the kernel matrix to be positive semidefinite. KFD is a generalized version of Fisher's linear discriminant analysis (LDA) with the added kernel trick as in support vector machine or kernel principal component analysis [66].

CHAPTER VI

STATISTICAL ANALYSIS OF RF RESPONSE

In this chapter, I interpreted the response distribution of input image type under the concept of *Power-law distribution* [15] and *Suspicious Coincidence* [16] because the response distribution of the tactile RFs shows similar properties as that of the visual RFs.

A. Power-law Distribution

I am interested in how the human tactile system perceives patterns as compared to the human visual system. Lee and Choe, among others, showed that visual response histograms of natural images have a fairly uniform shape, unlike the gray-level intensity histograms which greatly vary across images [70]. Fig. 52 depicts gray-level intensity histograms of six natural images. Fig. 53 shows the orientation energy distribution (simulating the visual cortical response distribution) for the same six sample images. We can see that the orientation energy distributions are similar to each other, and they share a unique feature, that of a power law (i.e., $p(x) = 1/x^a$, where a is the fractal exponent). Thus when plotted in log-log scale, they show up as straight lines.

I raised two questions in the previous chapters III and IV, regarding (1) the functional role and (2) the developmental origin of tactile RFs. There are some more interesting properties of the tactile RF that can link to its functionality. Sarma and Choe showed that the power-law-like response distribution in a visual cortical neuron model can help subsequent stages in the cortical processing to extract salient features in the input, such as edges and contours [15,70] (cf. [71]). The basic idea was that the heavy-tail part in the power-law distribution, compared to a Gaussian baseline, can be used to accurately predict the response threshold. Power-law distributions have

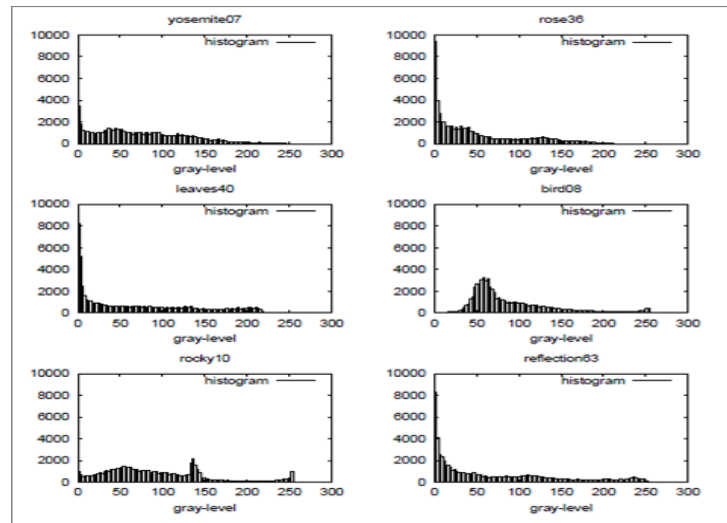


Fig. 52. Gray-level Intensity Histogram of Natural Images. Intensity histogram for six natural images are shown. The x-axis is the gray-level intensity (0 to 255), and the y-axis the frequency of the intensity value. The histograms vary across images. Adapted from [70].

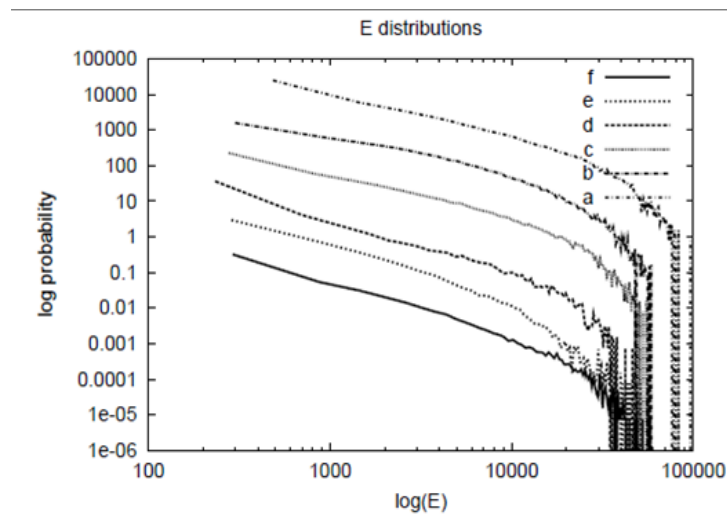


Fig. 53. Orientation Energy (E) Distribution. The E -distribution of six natural images are shown in log-log plot. The same images from Fig. 52 were used to calculate the E values. The curves are scaled by a factor of 10 successively (from a to f) to separate the curves for easier comparison. All curves have a similar slope with mostly straight line, demonstrating power-law property. Note that the high energy area (toward the right) has a lot of noise due to the scarcity of samples. Adapted from [70].

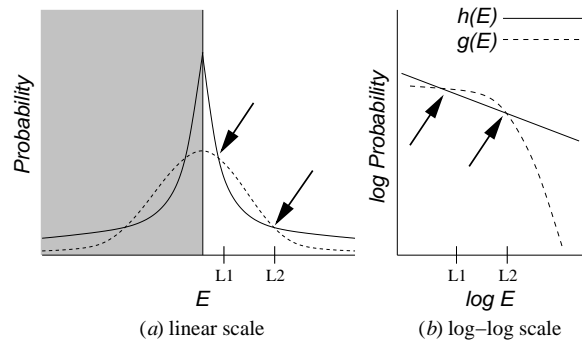


Fig. 54. Power-law vs. Baseline Gaussian Distribution. The response distribution $h(E)$, where E is the activation level, and the matching Gaussian distribution $g(E)$ that has the same standard deviation is illustrated. When the two curves intersect at the point marked $L2$, the probability of the response distribution becomes higher than the baseline Gaussian distribution. This point has been shown to play an important role in saliency thresholding [15, 70]. Note that the shaded part in (a) is only for illustrative purposes (to show the peakedness of $h(E)$), since the response E is always positive. Adapted from [57].

higher probability near extreme values (see Fig. 54; especially, E values greater than $L2$). Such a comparison can give us some clue on how to detect contours in images. In Fig. 55, an example of visual response thresholding is shown.

Interestingly, the response distribution of the tactile map also shows a power-law property, even though the spatiotemporal structure of the RF is different from the visual RFs (Figs. 56 and 57). However, we should note that not only the RFs but the input statistics was also different, natural-scene-like vs. texture-like. Thus, we can speculate that one goal of early sensory processing is to generate RF coding that maps the specific input distribution into a canonical response distribution that can be easily utilized by a similar (or even a common) second stage of processing, regardless of the modality. These observations raise the interesting possibility that later stages in multi-modal sensory processing may share a common, integrated mechanism, thanks to the customized encoding done at the early stages of processing. (In a similar

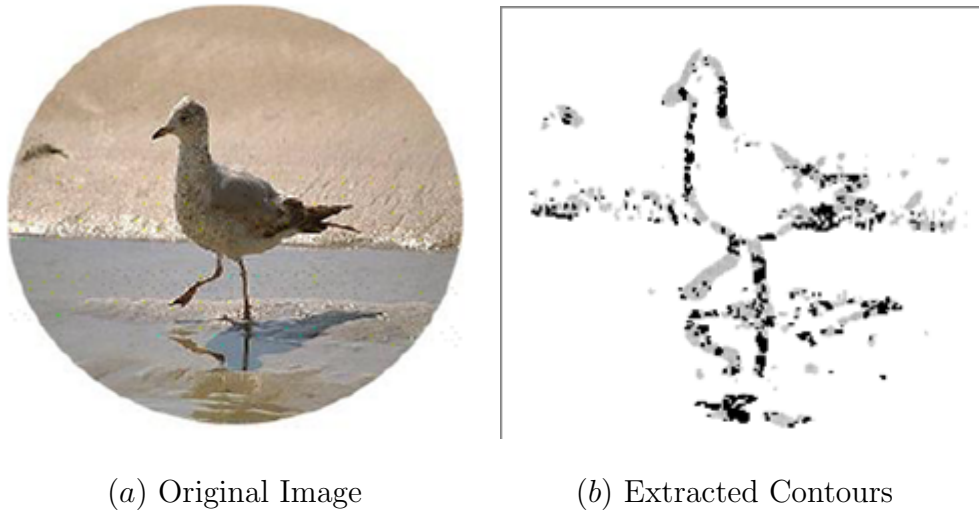


Fig. 55. Visual Responding Thresholding in a Natural Image. An example of thresholding approach to detect the edge orientation is shown: (a) the original natural image; (b) the extracted contours after thresholding shown. Adapted from [15].

vein, [72] and [73] showed how requirements in subsequent stages of cortical processing can dictate initial sensory encoding strategy.)

The observations above can lead to interesting research questions. For example, we can check if the response distribution of tactile maps responding to natural scene inputs maintain the power-law property. My prediction is that the power-law property will not be maintained in such a case. A dual experiment with visual maps can also be done. We can also extend the saliency thresholding approach developed in [15] to the tactile domain, exploiting the power-law distribution in TRF responses.

B. Suspicious Coincidence

The nature of the power-law response distribution can be investigated in the framework of suspicious coincidence [15]. Two statistical events A and B are said to be suspicious if they occur more often together than can be expected from their indi-

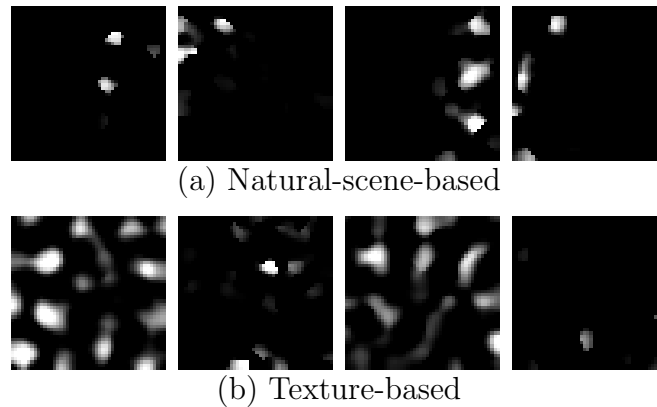


Fig. 56. Cortical Response on Input Images. Cortical responses of (a) the natural-scene-based cortex and (b) the texture-based cortex are shown (bright area represents high and dark area represents low activity). Both show a sparse activation profile. See Fig. 57 for the response histogram. Adapted from previously reported results in [57].

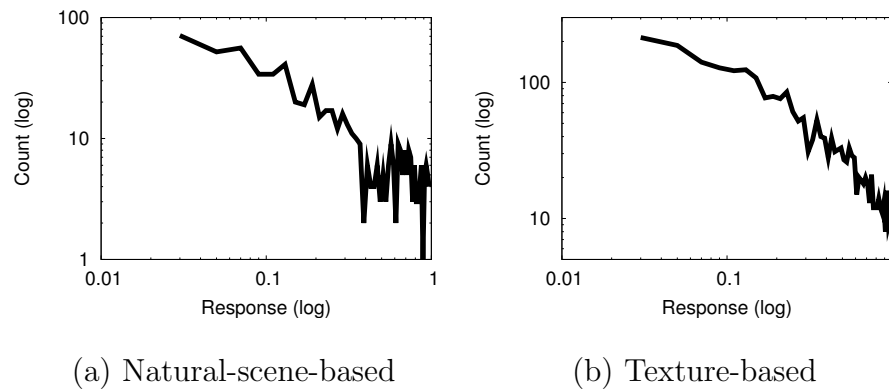


Fig. 57. Response Distribution of Input Images. The response histogram of (a) the natural-scene-based RFs and (b) the texture-based RFs are shown, in a log-log scale. The histograms were calculated from the response matrix in Fig. 56. Both show a power-law distribution. (Note that the first and the last bin are not plotted, because those values were artificially exaggerated due to the use of the piece-wise linear activation function that had a hard lower and upper bound [Eq. 4.3].)

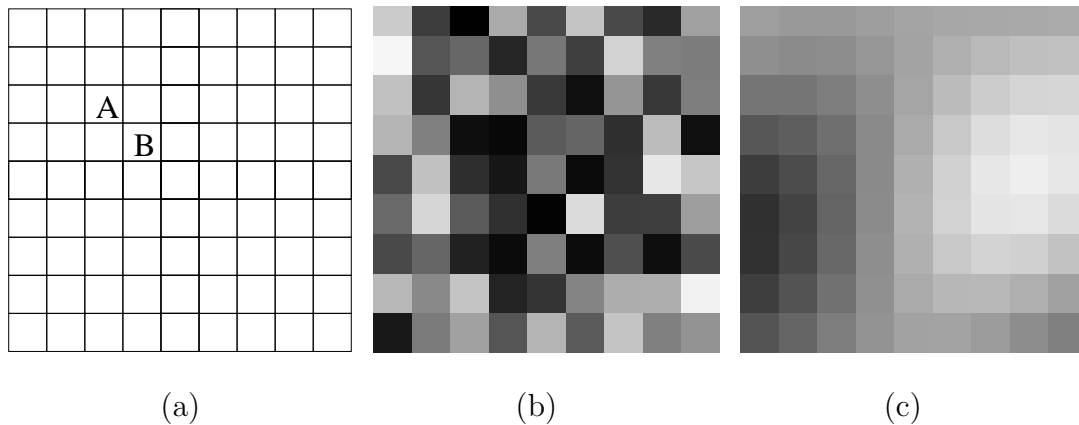


Fig. 58. Suspicious Coincidence between Image Pixels. A simple example of suspicious coincidence in the image is shown. (a) Each pixel in the image is treated as a random variable. (b) White noise image contain no suspicious coincidence. (c) Structured image contain suspicious coincidence.

vidual probabilities. In other words, the joint probability for the two events should exceed the product of their individual probabilities in order for them to be deemed suspicious:

$$P(A, B) > P(A)P(B). \quad (6.1)$$

Fig. 58 shows a simple example of suspicious coincidence in the image. Each pixel in the image is treated as a random variable. White noise images will show no suspicious coincidence between pixel pairs. Therefore, images with more structure will have higher suspicious coincidence between pixel pairs (e.g. contours, edges).

This approach is easily applied into the problem of image analysis, where each pixel is treated as a random variable. Suspiciousness in the image can then be determined by testing the inequality shown above where events correspond to pixels from different locations in the image [74].

Lee and Choe empirically derived the effective threshold for the detection of

salient contours, which was found to be linear to the orientation energy corresponding to the second point of intersection (L2) of the response distribution and its matching normal distribution (same mean, same variance) [70]. For example, Fig. 55 shows the effect of thresholding on the image of a bird using the threshold derived from L2. Only the salient edges in the image remain after the thresholding, and the resulting plot is very close to our perceived edginess in the image. Even though the thresholding criterion was effective as we have seen above, it does not tell us why the simple idea of comparing to a normal distribution has to be so effective. That is, why does a Gaussian distribution form a reasonable baseline for comparison?

Sarma and Choe linked the suspiciousness of the image contour to salience, i.e., more suspicious contour may have to be seen as more salient to the perceptual system [15]. An image where each pixel is independent from each other would be seen as containing no suspicious coincidence between any pair of pixels. Fig. 58a shows each pixel in the image treated as a random variable (for example, A and B). Thus as in Fig. 58b, a white-noise image can be said to have no suspicious coincidence in it, so that we cannot find any salient contour in it.

Fig. 59a shows the orientation energy distribution $h(E)$ from a white noise image compared with its corresponding normal distribution $g(E)$ to see if there is any similarity between the two. It turns out that the two distributions closely overlap as expected. Fig. 59b shows a typical orientation energy distribution $h(E)$ of a natural scene compared to its corresponding normal distribution $g(E)$. The two curves intersect at two points, the effective threshold for the detection of salient contour, which was linear to the orientation energy matching to the second point of intersection of the response distribution and its matching distribution.

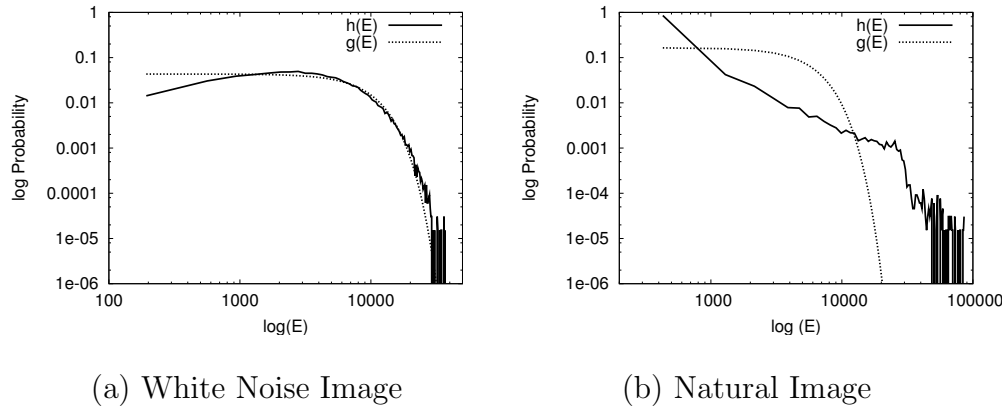


Fig. 59. Orientation Energy Distribution of an Image vs. Its Matching Normal Distribution. The orientation energy distributions $h(E)$ (solid line) for (a) white noise image and (b) natural image corresponding their normal distribution $g(E)$ (dashed line) with same variance are shown in log-log plot. The x-axis is the orientation energy E and the y-axis the probability. The two distributions in (a) show a close resemblance while the two in (b) intersect at two points. Adapted from [15] [75].

C. Experiments and Results

As I empirically observed and briefly discussed in the previous section, the response distribution of the tactile map also interestingly shows a power-law property, even though the spatiotemporal structure of the RF is different from that of the visual RFs (Figs. 56 and 57). Now I will check if the response distribution of tactile maps responding to natural scene inputs maintain the power-law property. A dual experiment with visual maps will also be done.

Fig. 60 shows the process followed to get the response distributions. TRF and VRF responses were generated from texture and natural-scene inputs as in chapter III. Then, I plotted the response distributions and compared the distributions with their matching normal distribution. To test if the tactile response distributions show the power-law property, I projected the same image types used during training of the

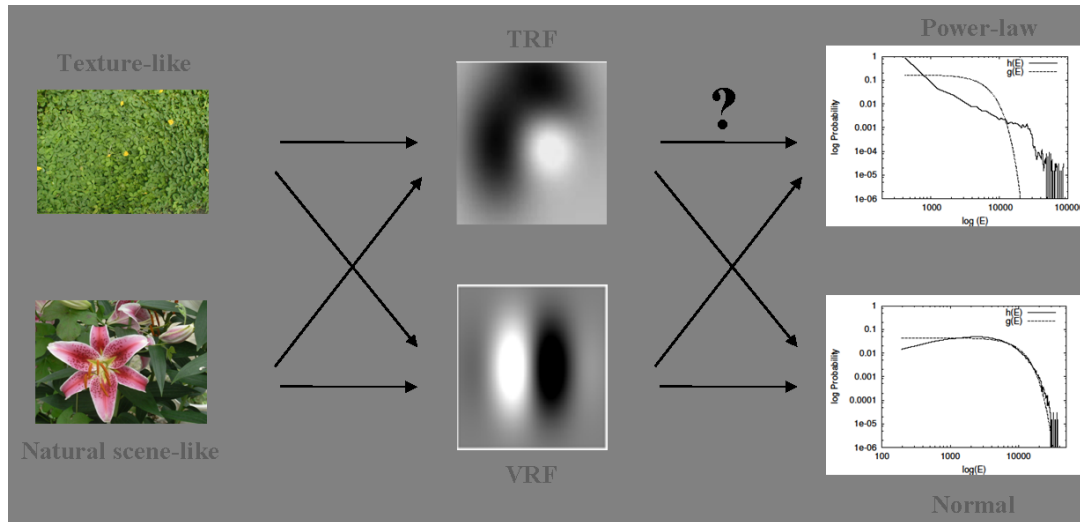


Fig. 60. Process of Statistical Analysis. The overall procedure was as follows: (1) generate TRF and VRF responses from texture and natural-scene inputs as in chapter III, (2) plot the distributions of the responses, and (3) compare the distributions with their matching normal distributions.

receptive fields in LISSOM: natural scene to visual receptive fields (VRFs), texture to tactile receptive fields (TRFs), natural scene to TRFs, and texture to VRFs.

Fig. 61 shows the log-log plot of the response distributions of the test image compared with the response distributions of the training images (the same images used in the development process). It turns out that the two distributions have closely overlapping lines where the straight declining slope characteristic of a power-law distribution is evident in both the visual and the tactile response distributions.

In the second experiment, I projected image type opposite of those used in training to the receptive fields developed by LISSOM: texture to VRFs and natural scene to TRFs. Fig. 62 shows the response distribution in log-log plot for both cases. The response distribution on opponent image type shows a more Gaussian shape (quick drop off) while the response distribution to the input images that were used in the developmental process shows a strong heavy-tail characteristic of a power-law. The two curves always intersect at a point. If we can analyze the two curves under the concept

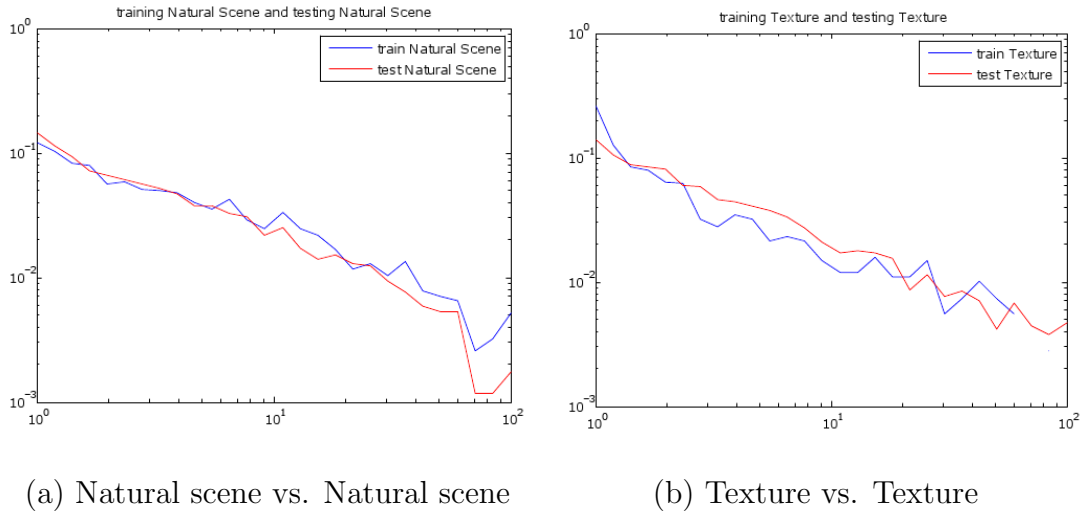


Fig. 61. Comparison of the Distribution of RF Response to the Same Input Type used During Training. The x-axis is the response and the y-axis the probability. The comparisons of response distributions for (a) natural-scene-based RF to natural-scene, and (b) texture-based RF to textures are shown. Each case shows close overlap.

of suspicious coincidence developed by [16], and find thresholding approach developed in [15], to discriminate natural scenes and textures, it would be the first systematic study of the relationship between input image type using a tactile approach.

D. Discussion

We can extend the saliency thresholding approach developed in [76] to the tactile domain, exploiting the power-law response distribution. Sarma and Choe conducted an experiment in which they generated new threshold values by comparing the orientation energy distribution with the white noise-based distribution. Then these values were compared to the orientation energy thresholds selected by humans. The results are shown in Fig. 63. It is clear that the new white noise-based L_2 values also have a strong linear relationship with the human-selected thresholds. Such psychophysi-

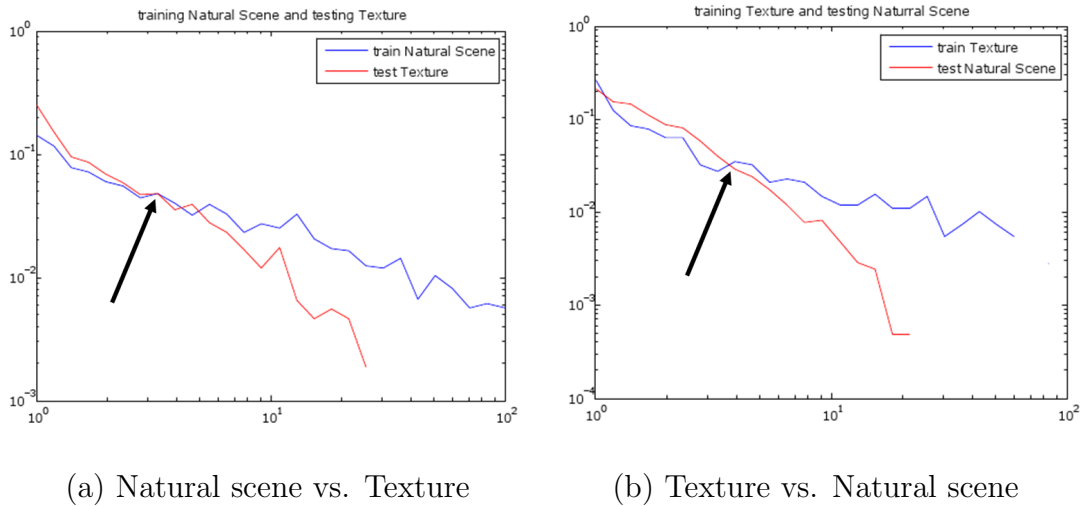


Fig. 62. Comparison of the Distribution of RF Response to Opposite Input Type to the Training Input Type. The x-axis is the response and the y-axis the probability. The comparisons of response distributions for (a) natural-scene-based RF to textures, and (b) texture-based RF to natural scene are shown. The two curves intersect at a mid point (marked by black arrows).

cal approach can help us better understand the relationship between input type and self-organized receptive fields.

As we can see in Figs. 61 and 62, the curves of the response distribution on opponent image type and the response distribution of the same input type always intersect at a point. Intersection point between two curves of the response distribution of natural-scene-like and texture-like could serve as an informative threshold to determine beyond which the input image's property becomes more texture-like or more natural-scene-like.

Sarma and Choe also derived a linear equation for the preferred orientation energy threshold as a function of the raw second moment (σ) of the orientation energy values, which showed a linear fit with the threshold ($L2$) derived from the orientation energy distribution $h(E)$ and the normal distribution $g(E)$ [76]. Fig. 64 shows the

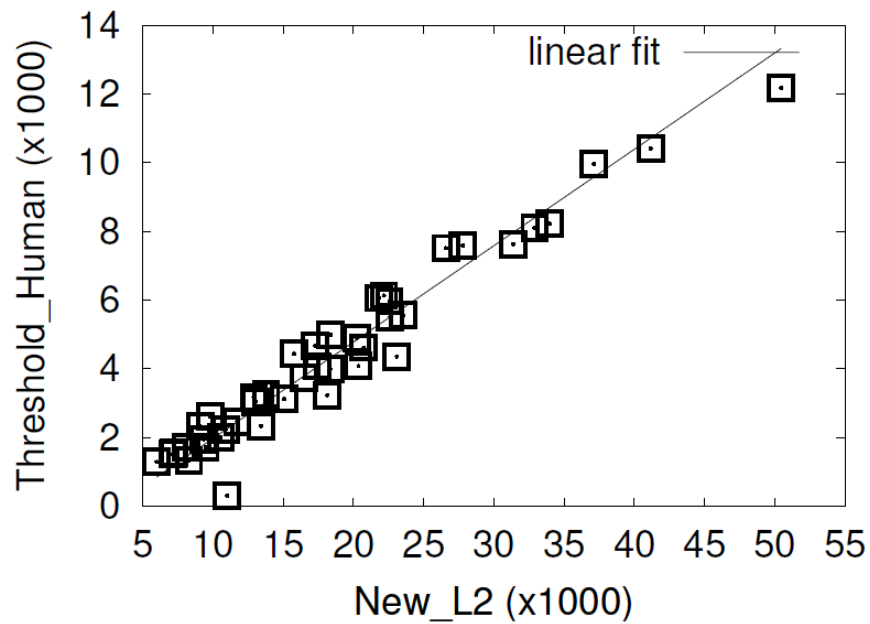


Fig. 63. Orientation Energy Threshold Selected by Human vs. $L2$. The manually chosen thresholds are compared to the $L2$ values for each image. Each point in the plot corresponds to one of 31 natural images used in the calculation. The straight line in the figure shows a linear fit to the data. Adapted from [76].

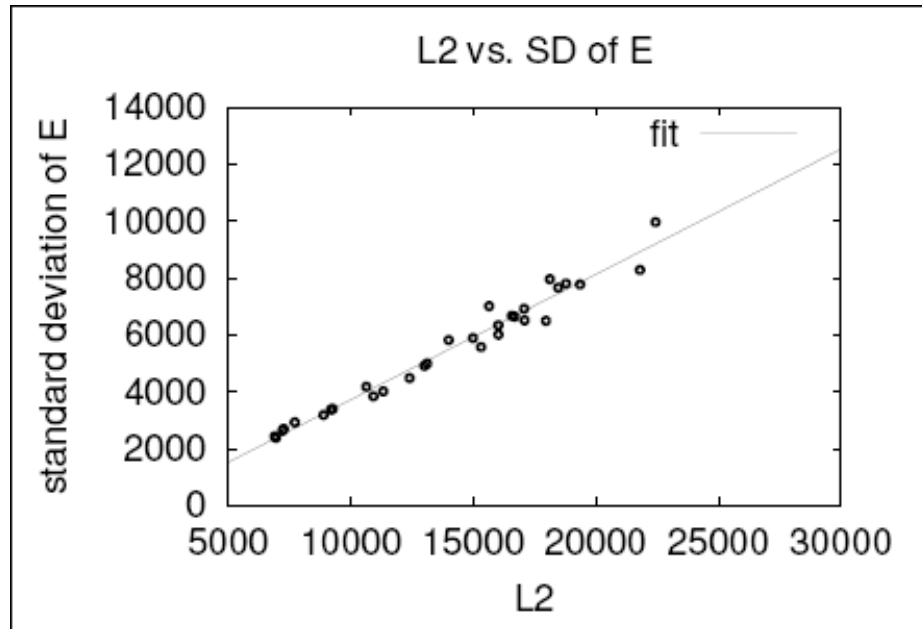


Fig. 64. Threshold ($L2$) vs. Raw Second Moment (σ) of the Orientation Energy Distribution. The $L2$ values obtained by comparing $h(E)$ and $g(E)$ are compared to the square root of the raw second moment (σ) of $g(E)$. Each point in the plot corresponds to one of 31 natural images and placed around the straight line. The straight line in the figure shows a linear fit to the data. Adapted from [76].

linearity between the threshold value ($L2$) of orientation energy distribution (OED) and the square root of the raw second moment (σ) of the orientation energy values. If we can similarly derive a linear equation of intersection of two curves as in [76], quantization of textural and natural components in an images could become plausible for image type.

CHAPTER VII

DISCUSSION

A. Contribution

In this dissertation, I showed that texture in general, whether it is tactile or visual in origin, contributes to the emergence of unique functional and developmental properties similar to those observed in tactile RFs in area 3b of the somatosensory cortex.

The main novelty of this work is the application of tactile approaches to texture tasks: (1) the use of tactile receptive field responses for texture segmentation, (2) a common cortical development framework (LISSOM) used for the development of two different RF types, just based on the type of input presented during training, natural-scene-like or texture-like, (3) touch-based manifold found to be more suitable for texture processing than vision-based, and (4) intersection point between the response distribution of natural scene-like and texture-like can be informative for thresholding.

In sum, the results reported here suggest an intimate relationship between texture and the tactile modality.

B. Discussion

1. Potential Limitations of the Approach

Tactile sensibility is not only for the finger: The sensation of touch begins at the skin. The skin provides our most direct contact with the world; indeed, skin is the largest sensory organ we have [77]. A touch to the skin is transduced into neural signals, these signals make their way to the brain, and the brain makes sense of them. Fig. 65 depicts somatotopic maps generated by stimulating and recording tactile receptive fields from the whole body, and the homunculus, a caricature from

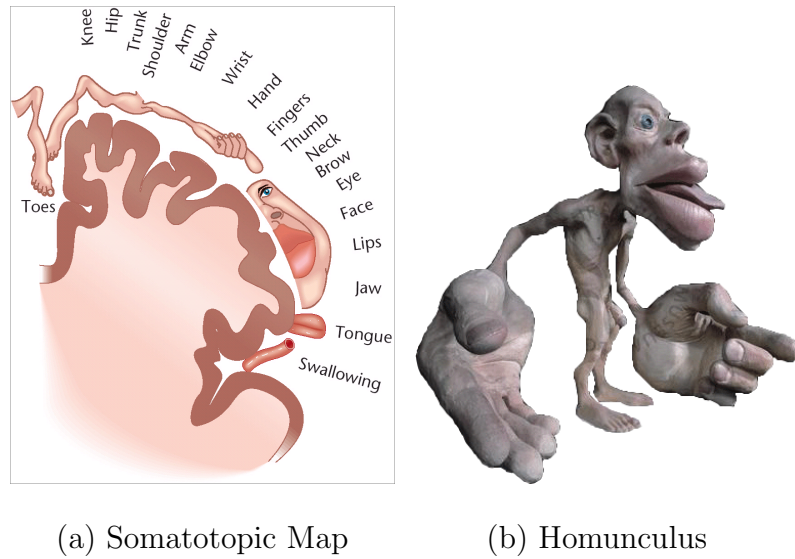


Fig. 65. A Somatotopic Map of the Body Surface in the Primary Somatosensory Cortex and the Sensory homunculus. The somatotopic map (a) is shown in a cross section through the postcentral gyrus. Neurons in each area are most responsive to the parts of the body illustrated directly above them. The Homunculus (b) is a caricature from (a). Adapted from [77].

the somatotopic map.

Notice in Fig. 65 that the representation of the hand separates that of the face and the body, while the genitals are mapped onto the most hidden part of S1, somewhere below the toes. Size on the map is related to the importance of the sensory input from that part of the body; information from the index finger is more useful than that from the elbow. The importance of tactile information from our hands and fingers is quite obvious even though the importance of a body part can vary greatly among species. Moreover, when we want to feel textural patterns, we use our fingers rather than our mouth, face, or elbow. The question given the above is, does texture play a key role in these other parts of the tactile surface.

Single type of input pattern was used to develop RFs: The developmental process of receptive fields (RFs) described in this research was based on a single type of

input pattern per experiment: natural-scenes for visual RFs or texture for tactile RFs. The biological cortex, however, may be exposed to multiple input (and input type) of activity during development including spontaneous, internally generated patterns and externally evoked inputs. Biological development is thus likely to depend on a complex combination of such patterns. Fig. 66 shows the comparison between prenatal orientation maps in animals and in HLISSOM (Hierarchical LISSOM) which can process both genetically determined internal input as well as external input [78]. This result shows that even simple internally generated inputs can be responsible for the observed prenatal self-organization.

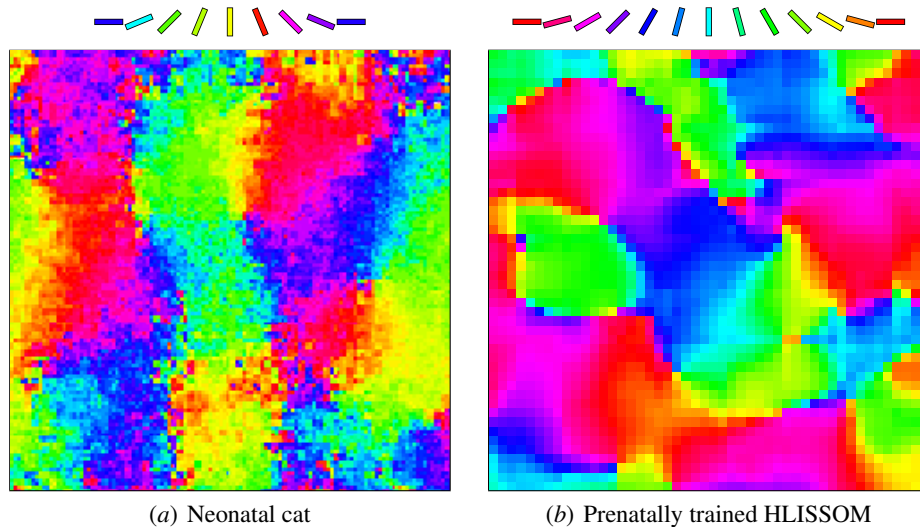


Fig. 66. Prenatal Orientation Maps in Animals and in HLISSOM. Comparison between (a) prenatal orientation maps in a 2-week-old binocularly deprived kitten, i.e. a kitten without prior visual experience (detail of a figure by [79]) and (b) the orientation map trained HLISSOM. The overall organization is very similar in the two maps, suggesting that prenatal training with internally generated patterns may be responsible for the observed maps at birth. Adapted from [11].

Computational modeling of both prenatal internal activities and postnatal input patterns allows the model to account for both the primitive orientation maps seen

at birth in animals and the more refined adult maps. More complex experiments as those are a plausible extension of the single input experiments presented in this dissertation, providing a potentially better mean for understanding how the primary sensory cortex (V1 or S1) develops in animals.

2. Potential Criticism

Performance of texture classification should be verified with benchmark tasks: Because an effective method of texture task is still far from being completely optimal for identifying surfaces, successful scene analysis, medical applications, and many others, benchmark test is needed to check several variants of a developed method and to carefully compare results with state-of-art in this area. However, since there is no available benchmark fully supporting texture task, I just used several popular classifiers to check the reliability of the performance experimented with standard backpropagation networks in chapter III, i.e., linear, quadratic, mahalanobis, k-nearlist KNN, and support vector machine (SVM). Table 1 shows the results of consistency in its texture classification task with previous experiment.

LISSOM can hardly be considered as a general model for the development of the cortex: LISSOM (Lateral Interconnected Synergetically Self-Organizing Model) was originally developed for the self-organizing map model of the visual cortex by Miikkulainen et al [11]. However, LISSOM is actually a more general model of how the cortex organizes to represent correlations in the sensory input. Thus, LISSOM should work equally well in modelling the development of non-visual sensory modalities, as demonstrated by [12], where the development of the barrel cortex (part of the somatosensory cortex) in rodents was successfully modeled using LISSOM, and by [80] for the auditory cortex. Fig. 67 shows an example where LISSOM was used for modeling the barrel cortex in rodents.

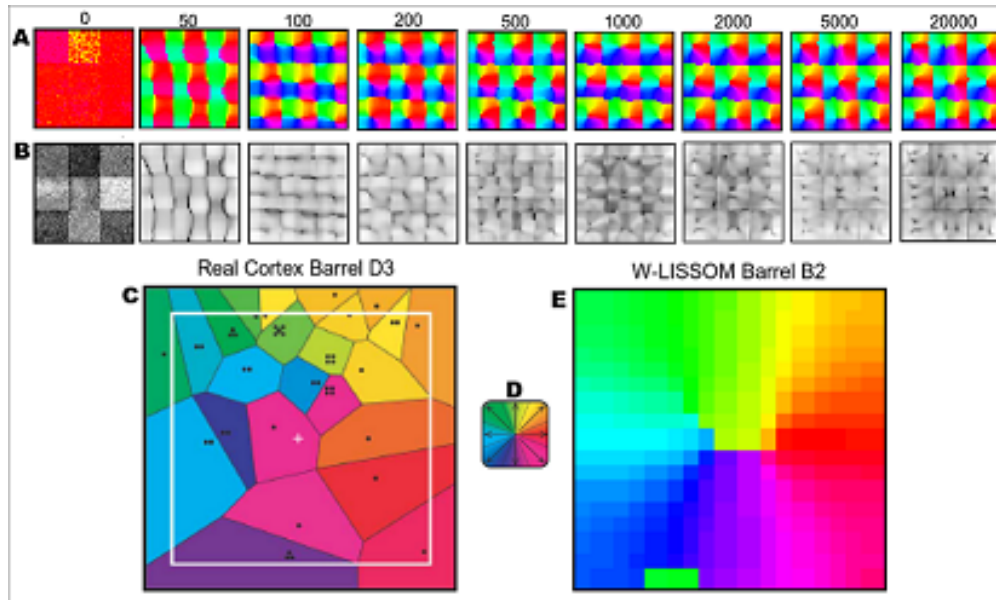


Fig. 67. Emerging Pinwheel Organizations of Angular Detection Preference. A: Outward-radiating complementary pinwheel maps emerged from the W-LISSOM over time (iterations). B: The gray scale diagrams of the corresponding selectivity of each cortical unit to angular deflections. C: The angular deflection mapping measured by [81]. D: Angular deflection color key (adapted from [81]). E: The angular deflection preference map in the model of B2 barrel field from the W-LISSOM. Plots C and E show similar preference, demonstrating that self-organization from can give rise to the angular detection map found in barrel cortex, which is a non-visual sensory modality. Adapted from [12].

Table I. Comparison of Texture Classification with Popular Classifiers. The classification rates verify that the performance from backpropagation networks is reliable.

| RFs | Classifier | texture1 | texture2 | texture3 | texture4 | texture5 | texture6 |
|-----------|-------------|----------|----------|----------|----------|----------|----------|
| TRF-based | Linear | 87.14801 | 64.9554 | 69.1592 | 75.2232 | 73.7351 | 85.6027 |
| | Quadratic | 98.6979 | 92.5967 | 88.6719 | 90.6064 | 94.3824 | 96.0937 |
| | Mahalanobis | 98.5491 | 93.3408 | 87.3326 | 90.4018 | 85.1935 | 94.3824 |
| | KNN | 88.239 | 82.882 | 80.145 | 83.332 | 81.819 | 85.392 |
| | SVM | 85.23 | 85.345 | 84.284 | 82.107 | 82.989 | 84.468 |
| VRF-based | Linear | 80.1339 | 67.2247 | 67.3363 | 75.7626 | 72.0982 | 87.0536 |
| | Quadratic | 98.8095 | 93.1176 | 88.3743 | 90.3274 | 94.5312 | 96.503 |
| | Mahalanobis | 98.5863 | 92.4851 | 86.4025 | 89.8065 | 85.4911 | 94.9405 |
| | KNN | 83.337 | 81.593 | 77.123 | 82.339 | 80.013 | 80.332 |
| | SVM | 90.232 | 80.567 | 83.272 | 80.503 | 79.318 | 85.783 |

Texture with non-linear elements may not need the curvature information: Texture is an equivalent class of images in 2D that share identical statistics [82] and texture segmentation has long been attributed to changes in the orientation of elementary features across images [3, 22, 83]. Identical curvature can also be one of the elementary features of textures. Ben-Shahar and Zucker showed in psychophysical experiments that the sensitivity to curvature strongly affects texture segmentation [84]. Hence, in chapter III, I experimented with the classification with two types of texture: Curvy textures with many segments of circles at different curvatures (0.333, 0.2, and 0.143) and linear textures with line segments at different orientations (horizontal, vertical, and diagonal). The results showed that TRFs can better handle curvature than VRFs.

Tactile approach need no scaling/normalization while vision has to be much more general: In image processing, scaling/normalization of the image can affect the response distribution. As the size of an image is increased, pixels

in the images become more visible but less tactile, making the pattern of features disappear. Conversely, reducing an image will tend to enhance its patterns and apparent sharpness, making them more tactile, but less visible. We can also note that touch is more sensitive to scaling/normalization of the pattern, while vision is less sensitive. Hence, in order to get constant results, I used a fixed size, but different scaling of the input types as in Fig. 31 for the developmental process in chapter IV. The tactile RFs result in varying shape depending on the scale, while the shape of visual RFs are more constant.

C. Future Directions

I expect the work presented here to be extended to computational, psychological, and neurophysiological research. I plan to continue this line of research in the following directions: (1) a more rigorous analysis and optimization of the proposed work; (2) validating the computational results against existing experimental literature; (3) extending the suspicious coincidence-based analysis of the response distributions of RFs to the quantization of the degree of input type in a mixed image.

This research is expected to result in a better understanding of the nature of tactile perception showing an intimate relationship among texture, surface, and touch. Such understanding may greatly help in designing and building more powerful methods for texture processing.

CHAPTER VIII

CONCLUSION

The primary aim of the research reported in chapter III was to explore and compare the texture segmentation performance based on Tactile Receptive Fields (TRFs) and Visual Receptive Fields (VRFs). The main finding suggests that touch-based texture representation contains more discriminative information than vision-based local spatial features. Statistical measures and classification performance were used to evaluate this characteristic as well as providing insight on analyzing the TRF representation.

The next topic in chapter IV focused on what drives a virtually identical learning medium (the cerebral cortex) to specialize and to diverge to represent different sensory modalities, visual vs. tactile. By comparing two structurally similar modalities of touch and vision, along with the assumption that texture (whether visual or tactile) is basically a surface property, I showed that it is the input statistics (natural-scene-like or texture-like) that determine the learned RF type, not the direct sensory modality, thus providing further support for the idea that texture and touch are intimately related.

In chapter V, I conducted a non-linear analysis of RFs based on manifold learning. Based on the eigenvalues and the projections on two-dimensional manifolds, we could observe that the VRFs learned from natural scenes have similar properties to those of the Gabor filters, and the TRFs learned from texture have similar properties to those of the three component model developed by DiCarlo and Johnson [9], and that tactile manifold is better than visual manifold for texture classification.

Finally in chapter VI, I observed that tactile response has a power-law property just like the visual counterpart, even though the spatiotemporal structure of the TRFs is different from that of the VRFs, and the response distributions of TRFs and VRFs

responding to opponent input types relative to the training input type show a more Gaussian shape. I analyzed the two distributions under the concept of suspicious coincidence developed by [16], and the thresholding approach developed in [15], to discriminate natural scenes and textures.

In summary, texture can be better understood as being tactile in nature. I showed that TRFs have better performance than VRFs in texture processing, textural input leads to the development of TRFs, and finally, manifold and statistical analysis further supports the link between texture and touch. I expect these results to help better understand the intimate relationship among texture, surface, and touch, and further strengthen our earlier finding that tactile-oriented approach to texture segmentation complement the traditional visually oriented approach.

REFERENCES

- [1] J. J. Knierim and D. C. van Essen, “Neuronal responses to static texture pattern in area v1 of the alert macaque monkey,” *Neurophysiology*, vol. 67, pp. 961–980, 1992.
- [2] A. Thielscher and H. Neuman, “A computational model to link psychophysics and cortical cell activation patterns in human texture processing,” *Journal of Computational Neuroscience*, vol. 3, no. 22, pp. 255–282, 2006.
- [3] B. Julesz, “Texton gradients: The texton theory revisited,” *Biological Cybernetics*, vol. 54, no. 4-5, pp. 245–251, 1986.
- [4] J. Malik and P. Perona, “Preattentive texture discrimination with early vision mechanisms,” *Journal of the Optical Society of America A*, vol. 7, no. 5, pp. 923–931, 1990.
- [5] S. Oh and Y. Choe, “Segmentation of textures defined on flat vs. layered surfaces using neural networks: Comparison of 2d vs. 3d representations,” *Neurocomputing*, vol. 70, pp. 2245–2255, 2007.
- [6] C. Spence, F. Pavani, and J. Driver, “Crossmodal links between vision and touch in covert endogenous spatial attention,” *Journal of Experimental Psychology, Human Perception and Performance*, vol. 26, no. 4, pp. 1298–1319, 2000.
- [7] E. Deibert, M. Kraut, S. Kremen, and J. Hart, “Neural pathways in tactile object recognition,” *Neurology*, vol. 52, pp. 1413–1417, 1999.
- [8] J. P. Jones and L. A. Palmer, “An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex,” *Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

- [9] J. J. DiCarlo and K. O. Johnson, “Spatial and temporal structure of receptive fields in primate somatosensory area 3b: Effects of stimulus scanning direction and orientation,” *Neuroscience*, vol. 20, no. 1, pp. 495–510, March 2000.
- [10] S. Bensimaia, P. Eenehev, J. D. Maass III, J. Craig, and S. Hsiao, “The representation of stimulus orientation in the early stages of somatosensory processing,” *Neuroscience*, vol. 28, pp. 776–786, 2008.
- [11] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh, *Computational Maps in the Visual Cortex*. New York: Springer, 2005.
- [12] S. Wilson, “Self-organisation can explain the mapping of angular whisker deflections in the barrel cortex,” Master’s thesis, The University of Edinburgh, Scotland, UK, 2007.
- [13] H. Choi and S. Choi, “Robust kernel Isomap,” *Pattern Recognition*, vol. 40, no. 3, pp. 853–862, March 2007.
- [14] P. Khurd, S. Baloch, R. Gur, C. Davatzikos, and R. Verma, “Manifold learning techniques in image analysis of high-dimensional diffusion tensor magnetic resonance images,” in *IEEE Conference on CVPR*, 2007.
- [15] S. Sarma and Y. Choe, “Salience in orientation-filter response measured as suspicious coincidence in natural images,” in *The 21st National Conference on Artificial Intelligence*, The Association for the Advancement of Artificial Intelligence. Boston, MA: AAAI Press, 2006, pp. 193–298.
- [16] H. B. Barlow, “Unsupervised learning,” *Neural Computation*, vol. 1, no. 3, pp. 295–311, 1989.

- [17] R. M. Haralick, “Statistical and structural approaches to texture,” in *IEEE*, vol. 67, no. 5, 1979, pp. 786–804.
- [18] G. R. Cross and A. K. Jain, “Markov random field texture models,” *Neuroscience*, vol. 5, pp. 25–39, 1983.
- [19] L. Urdang, *Random House Dictionary of the English Language*. New York: Random House, 1968.
- [20] J. R. Bergen and E. H. Adelson, “Early vision and texture perception,” *Nature*, vol. 333, pp. 363–364, 1988.
- [21] N. Houhou, J. P. Thiran, and X. Bresson, “Fast texture segmentation model based on the shape perator and active contour,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [22] B. Julesz, “Visual pattern discrimination,” *IRE Transactions on Information Theory*, vol. 8, pp. 84–92, 1962.
- [23] J. R. Bergen and M. S. Landy, *Computational Modeling of Visual Texture Segregation*. Boston: MIT Press, 1991.
- [24] M. Hollins, R. Faldowski, S. Rao, and F. Young, “Perceptual dimension of tactile surface texture: A multidimensional-scaling analysis,” *Perception and Psychophysics*, vol. 54, no. 6, pp. 697–705, 1993.
- [25] M. Hollins, S. J. Bensmaia, K. Karlof, and F. Young, “Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling,” *Perception and Psychophysics*, vol. 62, no. 4, pp. 1534–1544, 2000.
- [26] M. W. Levine and J. M. Shefner, *Fundamentals of Sensation and Perception*. Pacific Grove, CA: Cole, 1991.

- [27] S. W. Kuffler, “Discharge patterns and functional organization of mammalian retina,” *Neurophysiology*, vol. 167, pp. 37–68, 1953.
- [28] D. H. Hubel and T. N. Wiesel, “Receptive fields of single neurones in the cat’s striate cortex,” *Physiology*, vol. 148, pp. 574–591, 1959.
- [29] —, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *Physiology*, vol. 160, pp. 106–154, 1962.
- [30] —, “Receptive fields and functional architecture of monkey striate cortex,” *Physiology*, vol. 195, pp. 215–243, 1968.
- [31] D. L. Ringach, “Mapping receptive fields in primary visual cortex,” *Physiology*, vol. 558, pp. 717–728, 2004.
- [32] J. G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” *Visual Research*, vol. 20, pp. 847–856, 1980.
- [33] G. C. DeAngelis, G. M. Ghose, I. Ohzawa, and R. D. Freeman, “Functional micro-organization of primary visual cortex: Receptive-field analysis of nearby neurons,” *Neuroscience*, vol. 19, pp. 4046–4064, 1999.
- [34] G. C. DeAngelis, L. Ohzawa, and R. D. Freeman, “Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. 1. General characteristics and postnatal development,” *Neurophysiology*, vol. 69, no. 4, pp. 109–117, April 1993.
- [35] A. Shmuel and A. Grinvald, “Functional organization for direction of motion and its relationship to orientation maps in cat area 18,” *Neuroscience*, vol. 16, pp. 6945–6964, 1996.

- [36] M. Weliky, W. H. Bosking, and D. Fitzpatrick, “A systematic map of direction preference in primary visual cortex,” *Nature*, vol. 379, pp. 725–728, 1996.
- [37] D. H. Hubel and T. N. Wiesel, “Sequence regularity and geometry of orientation column in the monkey striate cortex,” *Comparative Neurology*, vol. 158, pp. 267–294, 1974.
- [38] D. H. Hubel, T. N. Wiesel, and S. LeVay, “Plasticity of ocular dominance column in monkey striate cortex,” *Philosophical Transactions of the Royal Society of London Series B*, vol. 278, pp. 377–409, 1977.
- [39] L. E. White, D. M. Coppola, and D. Fitzpatrick, “The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex,” *Nature*, vol. 41, pp. 1049–1052, 2001.
- [40] C. Blakemore and G. G. Cooper, “Development of the brain depends on the visual environment,” *Nature*, vol. 228, pp. 477–478, 1970.
- [41] C. Blakemore, “Innate and environmental factors in development of kittens visual-cortex,” *Physiology*, vol. 248, pp. 663–716, 1975.
- [42] H. V. B. Hirsch and D. Spinelli, “Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats,” *Science*, vol. 168, pp. 869–871, 1970.
- [43] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain 2e*. Baltimore, MD: Lippincott Williams and Wilkins, 2001.
- [44] J. H. Kaas, R. J. Nelson, M. Sur, and M. M. Merzenich, “Organization of somatosensory cortex in primates,” in *Proceedings of a Neuro-sciences Research*

- Colloquium*, F. Schmitt, F. G. Worden, G. Adelman, and S. G. Dennis, Eds. Cambridge, MA: MIT Press, 1981, pp. 237–262.
- [45] R. I. Paul, H. Goodman, and M. M. Merzenich, “Alterations in mechanoreceptor input to Brodmann’s areas 1 and 3 of the postcentral hand area of *Macaca Mulatta* after nerve section and regeneration,” *Brain Research*, vol. 39, no. 1, pp. 1–19, 1972.
- [46] M. Sur, M. M. Merzenich, and J. H. Kaas, “Magnification, receptive-field area, and hypercolumn size in areas 3b and 1 of somatosensory cortex in owl monkeys,” *Neurophysiology*, vol. 44, pp. 295–311, 1980.
- [47] K. O. Johnson, S. S. Hsiao, and I. A. Twombly, “Neural mechanisms of tactile form recognition in the somatosensory system,” in *The Cognitive Neurosciences*, M. S. Gazzaniga, Ed. Cambridge, MA: MIT Press, 1995, pp. 235–268.
- [48] I. N. Bankman, K. O. Johnson, and S. S. Hsiao, “Neural image transformation in the somatosensory system of the monkey: Comparison of neurophysiological observations with responses in a neural network model,” *Cold Spring Harb. Symp. Quant. Biol.*, vol. 55, pp. 611–620, 1990.
- [49] M. Randolph and J. Semmes, “Behavioral consequences of selective subtotal ablations in the postcentral gyrus of *Macaca mulatta*,” *Brain Research*, vol. 70, no. 1, pp. 55–70, 1974.
- [50] J. J. DiCarlo, K. O. Johnson, and S. S. Hsiao, “Structure of receptive fields in area 3b of primary somatosensory cortex in the alert monkey,” *Neuroscience*, vol. 18, pp. 2626–2645, 1998.
- [51] J. J. DiCarlo and K. O. Johnson, “Velocity invariance of receptive fields structure

- in somatosensory cortical area 3b of the alert monkey,” *Neuroscience*, vol. 19, no. 1, pp. 401–419, 1999.
- [52] S. Grigorescu, N. Petkov, and P. Kruizinga, “Comparison of texture features based on Gabor filters,” *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1160–1167, 2002.
- [53] P. Brodatz, *Textures: A Photographic Album for Artists and Designer*. New York: Dover Publication, 1966.
- [54] T. Randen and J. Husoy, “Multichannel filtering for image texture segmentation,” *Optical Engineering*, vol. 33, no. 88, pp. 2617–2625, 1994.
- [55] Y. H. Bai, C. Park, and Y. Choe, “Relative advantage of touch over vision in the exploration of texture,” in *Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL*, 2008.
- [56] L. Hansen and P. Salamon, “Pattern analysis and machine intelligence,” *Neural Network Ensembles*, vol. 12, no. 10, pp. 993–1001, 1990.
- [57] C. Park, Y. H. Bai, and Y. Choe, “Tactile or visual?: Stimulus characteristics determine RF type in a self-organizing map model of cortical development,” in *Proceedings of 2009 IEEE Symposium on Computational Intelligence for Multimedia Signal and Vision Processing*, Nashville, TN, pp. 6–13.
- [58] S. Oh and Y. Choe, “Texture segmentation in 2d vs. 3d: Did 3d developmentally precede 2d?” in *The 2004 International Conference on Development and Learning [electronic]*, USCD Institute for Neural Computation. J. Triesch and T. Jebara, Eds., 2004, pp. 175–182.

- [59] E. Devlin, *Mathematics: The Science of Patterns*. New York: Scientific American Library, 1997.
- [60] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, pp. 2319–2323, 2000.
- [61] S. T. Roweis and L. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, pp. 2323–2326, 2000.
- [62] L. Saul and S. T. Roweis, “Think globally, fit locally: Unsupervised learning of low dimensional manifolds,” *Journal of Machine Learning Research*, vol. 4, pp. 119–155, June 2003.
- [63] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computation*, vol. 15, pp. 1373–1396, 2003.
- [64] H. S. Seung and D. D. Lee, “The manifold ways of perception,” *Science*, vol. 290, pp. 2268–2269, December 2000.
- [65] S. Mika, G. Ratsch, J. Weston, B. Schölkopf, and K. Müller, “Fisher discriminant analysis with kernels,” in *Proceedings of IEEE Neural Networks for Signal Processing Workshop*, 1999, pp. 41–48.
- [66] B. Schölkopf and A. J. Smola, *Learning with Kernels*. Cambridge, MA: MIT Press, 2002.
- [67] C. Park, H. Choi, and Y. Choe, “Self-organization of tactile receptive fields: Exploring their textural origin and their representational properties,” in *Proceedings of 7th International Workshop on Self-Organizing Maps*, St. Augustine, FL, 2009.

- [68] M. O. Ernst and M. S. Banks, “Humans integrate visual and haptic information in a statistically optimal fashion,” *Nature*, vol. 415, pp. 429–433, 2002.
- [69] H. Choi, S. Choi, and Y. Choe, “Manifold integration with Markov random walks,” in *23rd Association for the Advanced of Artificial Intelligence (AAAI-08)*, Chicago, Illinois, 2008.
- [70] H.-C. Lee and Y. Choe, “Detectiong salient contours using orientation energy distribution,” in *The International Joint Conference on Neural Networks*. IEEE, 2003, pp. 206–211.
- [71] D. J. Field, “What is the goal of sensory coding?” *Neural Computation*, vol. 4, pp. 559–601, 1994.
- [72] E. Salinas, “How behavioral constraints may determine optional sensory representations,” *PLoS Biology*, vol. 4, pp. 2383–2392, 2006.
- [73] Y. Choe, H.-F. Yang, and N. Misra, “Motor system’s role in grounding receptive field development, and shape recognition,” in *The Seventh International Conference on Development and Learning, Nashville, TN*, 2008.
- [74] H. B. Barlow, “What is the computational goal of the neocortex?” in *Large Scale Neuronal Theories of the Brain*, C. Koch and J. Davis, Eds. Cambridge, MA: MIT Press, 1994, pp. 1–22.
- [75] Y. Choe and S. Sarma, “Relationship between suspicious coincidence in natural images and oriented filter response distributions,” Texas A&M Uninversity, Tech. Rep., 2003.
- [76] S. Sarma, “Relationship between suspicious coincidence in natural images and

contour-salience in oriented filter responses,” Master’s thesis, Texas A&M University, College Station, TX, 2003.

- [77] W. Penfield and T. Rasmussen, *The Cerebral Cortex of Man*. New York: MacMillan, 1972.
- [78] J. A. Bednar, “Learning to see: Genetic and environmental influences on visual development,” Ph.D. dissertation, The University of Texas at Austin, 2002.
- [79] M. C. Crair, D. C. Gillespie, and M. P. Stryker, “The role of visual experience in the development of columns in cat visual cortex,” *Science*, vol. 279, pp. 566–570, 1998.
- [80] A. Plebe, “A neural model of object naming,” in *Proceedings of World Academy of Science, Engineering and Technology*, Barcelona, Spain, 2007, pp. 26–28.
- [81] M. L. Andermann and C. I. Moore, “A somatotopic map of vibrissa motion direction within a barrel column,” *Nature Neuroscience*, vol. 4, no. 4, pp. 543–551, 2006.
- [82] S. C. Zhu, X. Liu, and Y. Wu., “Exploring texture ensembles by efficient Markov chain monte carlo - toward a trichromacy theory of texture,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 554–569, 2000.
- [83] M. S. Landy and J. R. Bergen, “Texture segregation and orientation gradient,” *Vision Research*, vol. 31, pp. 679–691, 1991.
- [84] O. Ben-Shahar and S. Zucker, “Sensitivity to curvatures revealed in orientation-based texture segmentation,” *Vision Research*, vol. 44, pp. 257–277, 2004.

VITA

Choon Seog Park was born in Seoul, Korea, the son of Bokman Park and Soonim Jung. He received his Bachelor of Science degree in mechanical engineering from Korea Military Academy in 1990 and Master of Science in electrical engineering from Yonsei University in 2000. Since 1990, he has worked at Korean Army in Seoul, Korea (now he is a lieutenant Colonel in Korean Army). He entered Texas A&M University for his Ph.D. in the fall of 2004 and graduated in August, 2009. His permanent address is TAMU 3112, College Station, TX 77843-3112.