Computational theory of occluded-occluding object perception in early visual cortex

Zaem Arif bin Zainal Abidin

Graduate School of Information Systems

The University of Electro-Communications

Tokyo, Japan

A thesis submitted for the degree of

Doctor of Engineering

March 2018

Computational theory of occluded-occluding object perception in early visual cortex

Examining Committee

Main Academic Adviser :   Assoc. Prof. Shunji Satoh

Academic Adviser :   Prof. Yutaka Sakaguchi

Academic Adviser :   Prof. Takashi Suehiro

Examiner :   Prof. Yoshiki Kashimori

Examiner :   Assoc. Prof. Tomohiro Ogawa

概要

外界には奥行きが異なる物体が多数存在し，これら物体の視覚情報は重なりっ

ている場合が多い．このとき「遮蔽物体と被遮蔽物体を隔てる境界（Border）は

遮蔽物体に帰属する」．Border Ownership（BO）問題は，物体の重なり順序を計

算するための基盤的問題である．本研究の具体的な成果として，境界の Owner が

存在する方向である BO 信号をベクトル場$E(x,y)$，重なり順序をスカラー場

$\phi(x,y)$として見なせば，電磁気学の電位と電場に関する定理（電場は電位の勾

配，$E(x,y) = \nabla\phi(x,y)$である）を用いることで問題の定式化ができることを発

見した．数値シミュレーション結果から，提案モデルは様々な遮蔽状況や形状の

変化に対して頑健に，BO 問題と重なり順序計算問題が解けることを見出した．

Abstract

Humans can distinguish the order of mutually overlapping objects in a visual scene. The border between an occluding object and the occluded object is "owned" by the occluding object. How the brain assigns these borders, or Border-ownership (BO) assignment, determines the perception of object depth order. Findings from physiological experiments reveal that some neurons in area V2 of the brain respond selectively when the object which "owns" the edge in its receptive field is located on a specific side. Several models have been proposed in existing studies to reproduce this phenomenon. However, these models are not based on a clear computational theory.

This study is the first to approach BO assignment from a computational viewpoint by treating it as a well-defined problem. I propose that the direction of BO assignment can be defined as a conservative vector field $\boldsymbol{E}(x, y)$ with arrowheads pointing towards the occluding object, and that information pertaining to depth order can be defined as its corresponding scalar field $\phi(x, y)$. By using a theorem in electromagnetics which states that the gradient of electric potential is its electric field $E(x, y) = \boldsymbol{\nabla}\phi(x, y)$, I demonstrate that the BO assignment problem can be solved by updating an initial vector field until its rotation, or "curl", is zero.

A model developed on this computational theory can simultaneously reproduce BO assignment and perceived depth order. Results of numerical simulations agree qualitatively with the response of object-side selective neurons in V2 to stimuli containing occlusion with simple geometry. Neural networks can be deduced from the update rule curl using only one parameter for adjusting the scale of neural connections. This study also presents new interpretations of existing models in addition to insight into a possible method for calculation of depth order.

Acknowledgements

I would like to acknowledge and thank everyone who has supported me throughout the production of this thesis.

Firstly, I am greatly thankful to my academic adviser, Associate Professor Shunji Satoh for his guidance, patience and enthusiasm. Without his support, writing this thesis would not have been possible.

I would also like to express my gratitude to Professor Yutaka Sakaguchi for his invaluable advice, as well as Professor Takashi Suehiro, Professor Yoshiki Kashimori and Associate Tomohiro Ogawa for their insightful opinions.

In addition, I would also like to thank to everyone in the Laboratory for Human Informatics for their support and advice.

Last but not least, I would like to thank everyone in my family and extended family who provided me with encouragement and support thought my years in Japan.

Table of Contents

1   Introduction

1.1 Background

### 1.1.1 Occlusion and object perception

Humans can identify and recognize objects with ease from a cluttered scene in their surrounding environment. Figure-ground organization, or the separation of an object from the rest of the image, is thought to be a preliminary process for cognition. **Figure 1.1a** shows an image referred to as "Rubin's vase", which illustrates the basis of figure-ground organization [1]. The image is perceived in two distinct ways. It is perceived as a vase if the light-grey region is treated as the "figure", and the dark-grey region is considered as the "ground". In contrast, the viewer perceives the image as two faces if the dark-grey region is considered as the "figure" and the light-grey region is considered as the "ground".

Evidence for figure-ground organization is most apparent for a single object occluding a background. However, the real world is filled with several objects of various shapes and sizes, many of which mutually overlap one another. **Figure 1.2** shows a photograph with several objects, some of which are occluded. There are several "figures" in this scene, and figure-ground organization only describes the separation of one figure at a time. To accurately grasp the relative location and nature of objects in a scene, one needs a sense of the order that obejcts are located in front of us: object "depth order". Figure-ground organization would involve segregating these objects one at a time, starting from the plastic bottle in the foreground occludes the laptop. In contrast, assigning larger depth order values to objects closer to the viewer would allow an instantaneous grasp of how objects are placed before us in a visual scene.

A process called "border-ownership" (BO) assignment is believed to be a strategy

adopted by the brain to approach problems in visual perception related to overlapping objects [2]. In a situation where one object occludes another, the border between them is "owned" by the occluding object. Let us revisit the example of Rubin's vase in **Figure 1.1** to see how BO assignment effects perception. The direction of BO assignment is expressed by arrows pointing towards the occluding object along the perimeter of the border. In the perception of a vase, the border is owned by the inner region, as indicated by the arrows. Similarly, outward-pointing arrows imply that the border is owned by the two faces. For a real-world example, take a look at the plastic bottle in the foreground in **Figure 1.2**. The border between the plastic bottle and the laptop computer is owned by the plastic bottle. This border describes the shape of the plastic bottle. Conversely, this border is not owned by the laptop computer; part of the laptop computer is occluded by the plastic bottle, and the border between them does not describe the shape of the laptop computer. It is clear that BO assignment is important for the perception of object shape, as observed by psychologists [2]. However, does the human brain actually carry out BO assignment?
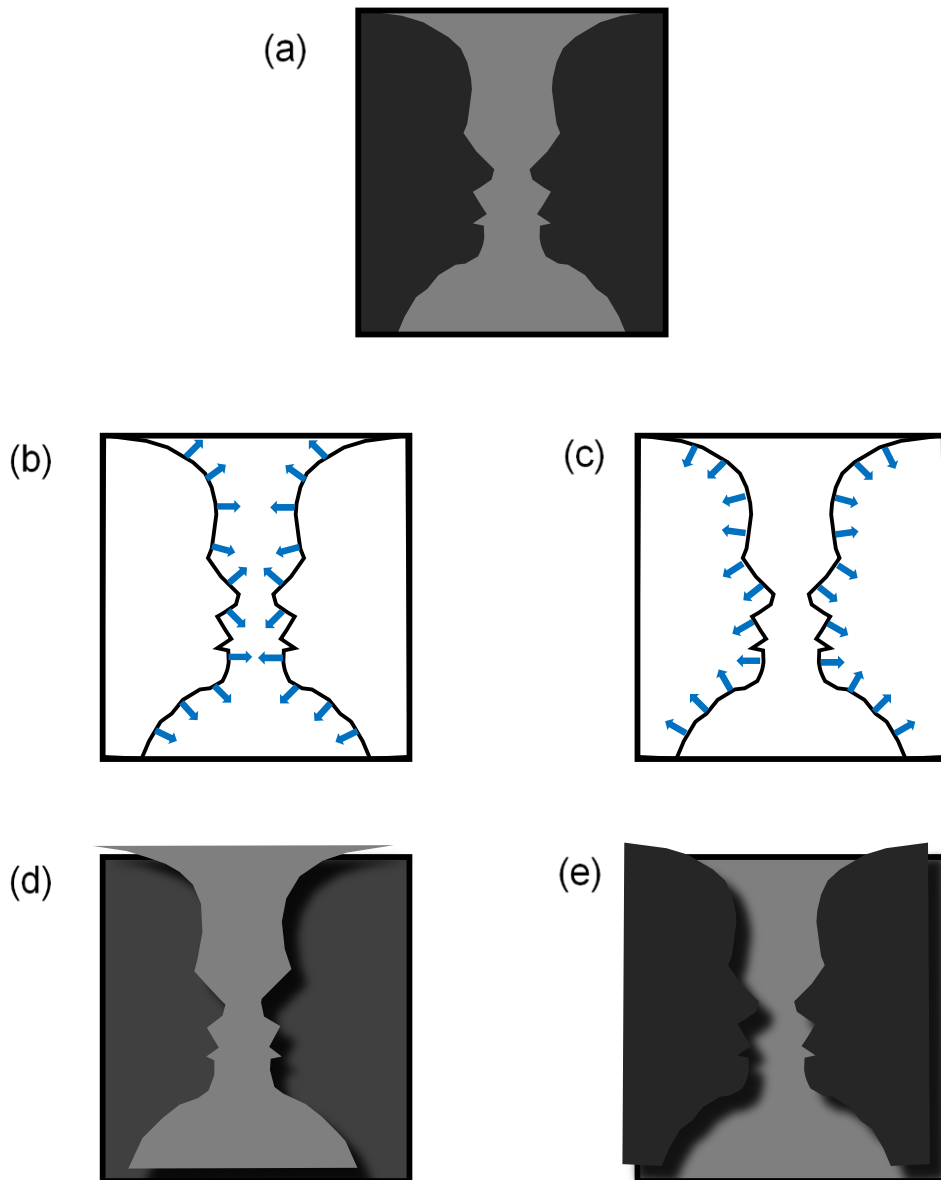
Figure 1.1 Bi-stable perception of Rubin's vase.

The image in (a) can either be perceived as a (d) vase over a dark grey background or (e) two faces over a light grey background. BO assignment is believed to account for these two percepts. The direction in which borders are assigned to occluding regions dictates how the image is perceived. Arrows along the border perimeters point either towards (b) the inner or (c) the outer region, resulting in the perception of (b) a vase or (e) two faces, respectively.

Figure 1.2 A real-world example, where several objects mutually overlap.

Figure-ground organization does not adequately explain the perception of a scene where multiple "figures" exist. For example, the plastic bottle occludes the laptop computer, which occludes a jar. The order in which these objects overlap, or "depth order", is important to gauge the nature of how objects are located in front of us. The border between the plastic bottle and the laptop computer, shown in red is owned by the plastic bottle. This means that the border describes the shape of the occluding plastic bottle, and not of the occluded laptop.

1.1.2 The human visual system

The human visual system processes information in a hierarchical manner. Local information such as luminance, color, and motion, are extracted from retinal images then subsequently integrated for a contextual understanding of the external world. A neuron in the visual cortex can only process information presented to it in a fixed range, located at a fixed location in the retinal image. This is the neuron's "receptive field". **Figure 1.3** shows the flow of information between areas in the visual cortex as compiled and organized from past findings [3]. Each area of the visual cortex carries out a different processing task. For example, neurons in area V1 respond selectively to discontinuities in luminance, at a specific orientation, placed in their receptive field [4]. This supports the assertion that one task of area V1 is to extract edges from an image, some of which coincide with object borders. More complex tasks are conducted higher up in the hierarchy; research shows that area IT is involved with recognition [5].

Physiological experiments revealed the existence of neurons in area V2 which respond strongly to an edge of an arbitrary orientation when an object is located on a specific of its receptive field: BO is coded in area V2 [6]. The emergence of BO signals were fast (30 ms after response onset), and latency was independent of object size [7]. Their experiments were conducted on monkeys using 2D images consisting of simple shapes such as single squares, C-shapes, and occluding rectangles. **Figure 1.4** shows data of the response of a neuron in area V2 with a left-side object selectivity. The pattern in its receptive field, a 90-degree vertical edge, is the same for all stimuli. This implies that visual information outside of the neuron's receptive field modulates its response. The response of a substantial subset of these BO-coding neurons were consistent with object-side even for complex natural scenes [8]. In addition, some BO-coding neurons also

respond selectively to three-dimensional (3D) objects in stereoscopic stimuli [9]. This implies that neurons in V2 utilize both global edge information and binocular depth information in BO coding to represent objects in a 3D space.

Area V4 may be the visual area where depth order is calculated. Some neurons have been reported to respond selectively to surfaces of various shapes [10]. Shape-selective neuron will respond selectively to an object of a specific shape if it is the occluding object [11]. Some neurons in area V4 have been shown to respond to subjective surfaces [12]. Lesions in area V4 also have been reported to impede the perception of color, brightness, shape, as well as subjective contours in monkeys [13], [14].
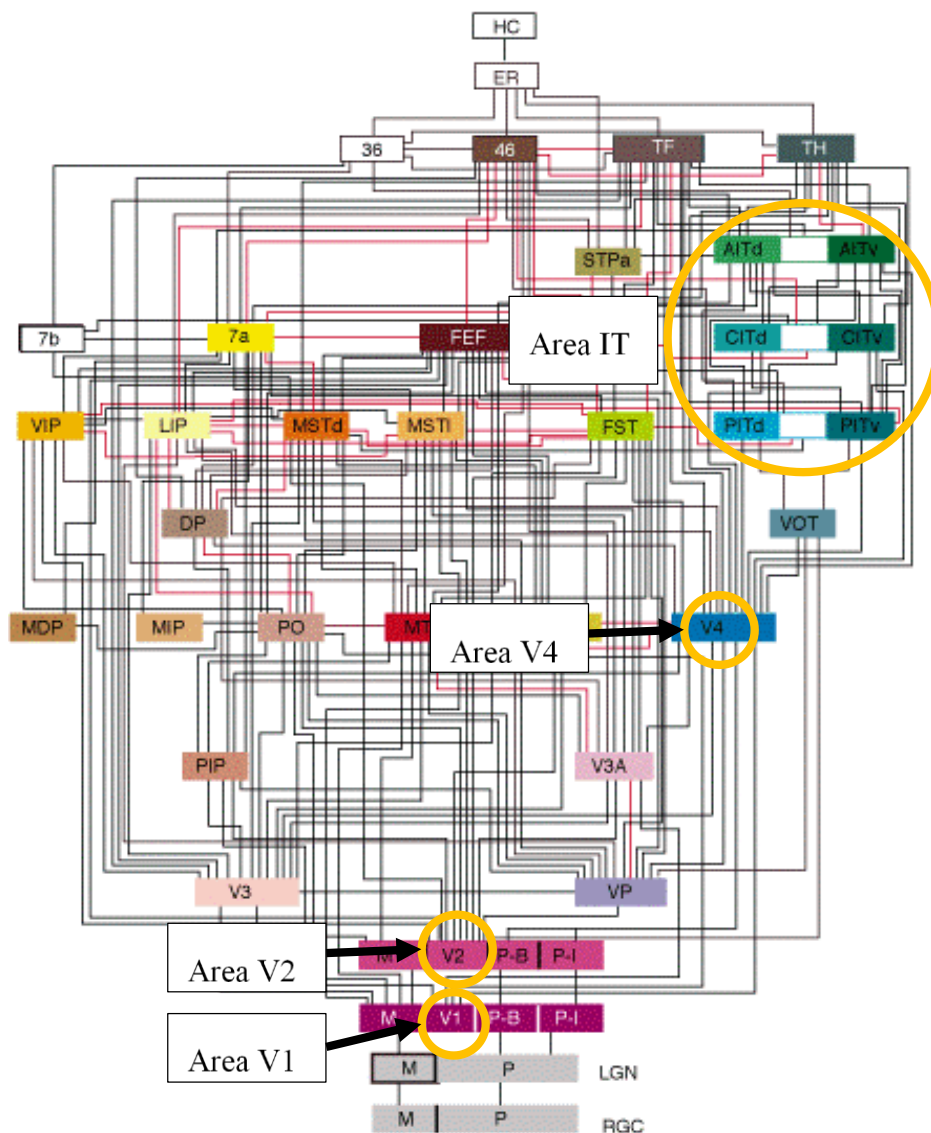
Figure 1.3 Hierarchical nature of visual processing in the brain.

Each area carries out a different visual processing task. Lower-level areas such as V1 extract local edges from retinal images for further processing. Area V2 is believed to be involved in BO assignment. Area V4 is involved with color, shape and surface perception. Area IT, located higher in the hierarchy is involved with object and face recognition.

Pattern inside receptive field
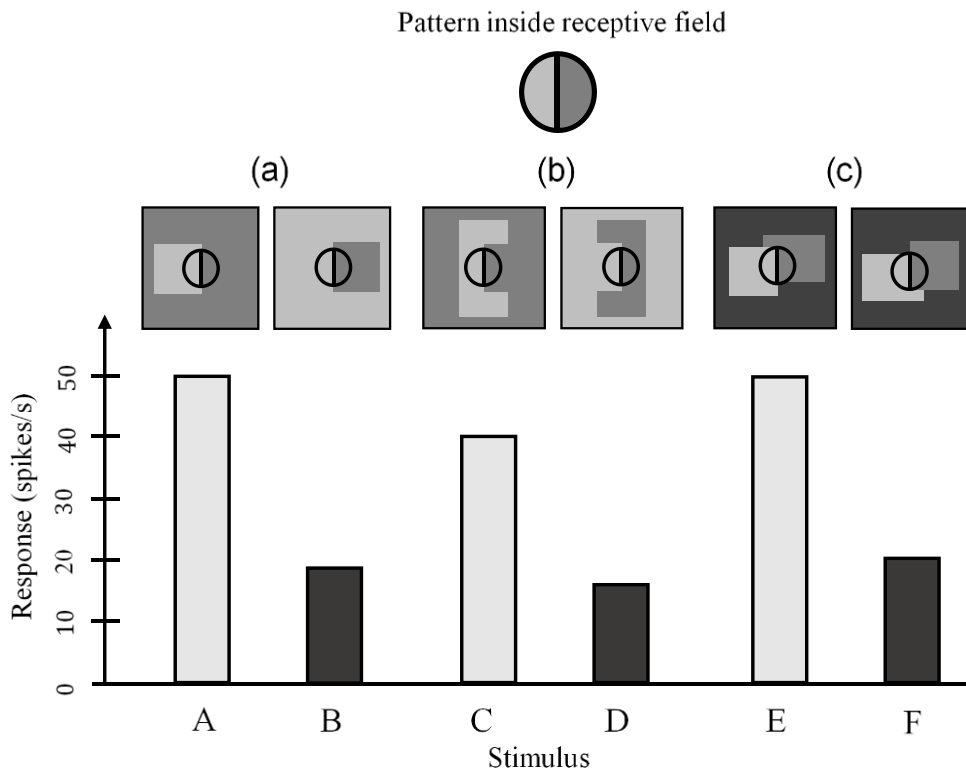
Figure 1.4 Response of a neuron in area V2 towards pairs of stimuli demonstrates its object-side selectivity.

The receptive field of the neuron is represented by a circle, enlarged for explanatory purposes. The pattern inside the receptive field is the same for all cases: a vertical (90 degree) edge. (a) This neuron responds stronger towards stimulus A, where an object exists on the left side of its receptive field. In contrast, an object does not exist on the left side of the neuron's receptive field in stimulus B. (b) Similarly, the neuron responds stronger to stimulus C than to stimulus D, and (c) stronger to stimulus E than to stimulus F. For all three cases, an occluding object exists on the left side of the neuron's receptive field. Data was adapted from Zhou et al. (2000) [6].

1.1.3 Marr's framework of visual processing

Marr (1982) proposed a framework of visual processing [15]. He argued that the processing of visual information is carried out in four distinct stages:

(1) Visual information from the external world is projected as two-dimensional images (2D) onto the retina

(2) A "Primal sketch" of low-level features such as edges and blobs is created.

(3) Information is represented as 2.5D sketches, where contour, depth, and orientation of surfaces is described.

(4) A 3D image of the external world is reconstructed based on visual information from the previous stages.

This framework is backed up by the hierarchical structure of the visual cortex outlined earlier. In particular, the reconstruction of surfaces using edge information might be what the visual system is ultimately trying to achieve through BO assignment.

1.2 Objective

Previous studies support the idea that BO assignment and assignment of depth order to surfaces are involved in visual processing in the brain. The objective of this study is to understand how the visual system processes information pertaining to object perception. The author will attempt to propose a clear computational theory for the BO-assignment problem based on the concept of a reconstruction of surface depth order.

1.3 Paper organization

The remainder of this paper is organized as follows. Section 2 outlines several existing models and the proposed approach taken by the author. Section 3 presents the computational theory underlying the model proposed in this paper. Section 4 implements the computational theory as a neural network model and presents simulation results for various problems in perception related to BO assignment. Section 5 contains deductions made from the computational theory and comparisons to existing models. Section 6 contains extensions of the proposed theory. Section 7 is the discussion and conclusion.

2 Related works and approach

2.1 Models of BO coding in area V2

The phenomenon of V2 neurons responding selectively to objects located on a specific side of their receptive field has been reproduced by existing models [16]–[18]. Li proposed horizontal connections as the medium for propagation of BO signals in area V2 [16]. **Figure 2.1a** shows a schematic diagram of excitatory and inhibitory connections between neurons with different receptive field locations (ovals) and object-side selectivity (arrows). Li considered cases in which objects likely exist on a certain side of the edge for either mutual excitation, or mutual inhibition. For example, neurons of similar object-side selectivity would mutually excite each other if collinearly placed (**Figure 2.1b**). Li also considered corners (**Figure 2.1c**) and T-junctions (**Figure 2.1d**). T-junctions are considered as cues for occlusion, where the top side of the "T" corresponds to the side of the occluding object [19]. These neural descriptions are complex, and they contain 23 free parameters that have to be determined properly for the model to function. Detailed descriptions of the neural connections can be seen in **Appendix A.1**.

Sakai et al. found that integrating edge using randomly generated regions of excitation and inhibition produced signals which resembled BO signals [17]. **Figure 2.2** shows an example of the random weights. Responses of V1 neurons towards edges are pooled and given an either excitatory or inhibitive feedforward effect on the response of V2 object-side selective neuron. For example, a neuron with a right object-side selectivity has an excitatory edge-pooling region on its right, and an inhibitory edge-pooling region on its left. Due to the random nature of these signals, the model successfully captures the diverse properties of neurons in V2. However, the computational theory underlying their model is unclear.

Craft et al. (2007) proposed feedforward-feedback with hypothetical "grouping cells" as the mechanism for determining BO signals [18]. A schematic diagram of the model is shown in **Figure 2.3**. A grouping cell would be activated if neurons with selectivity towards an object located at the center of the grouping cell's annular receptive field exist. This process is similar to calculating circles inscribed in an object's border. Annular shaped receptive fields of various sizes were used to account for objects of many sizes. It is not clear where, and if, grouping cells with annular receptive fields exists. Craft et al. (2007) also uses an existing model to detect T-junctions as occlusion cues [20].

Current models are only capable of reproducing the phenomenon of neurons responding selectively to objects located on a specific side of their receptive field: they are phenomenalistic in their approach. Thus, there have only been discussions regarding the plausibity of the neural mechanisms proposed as the medium of propagating BO signals (horizontal connections, feedforward, or feedback) when comparing the existing models [18], [21]. None of the existing models are based on a clear computational theory.
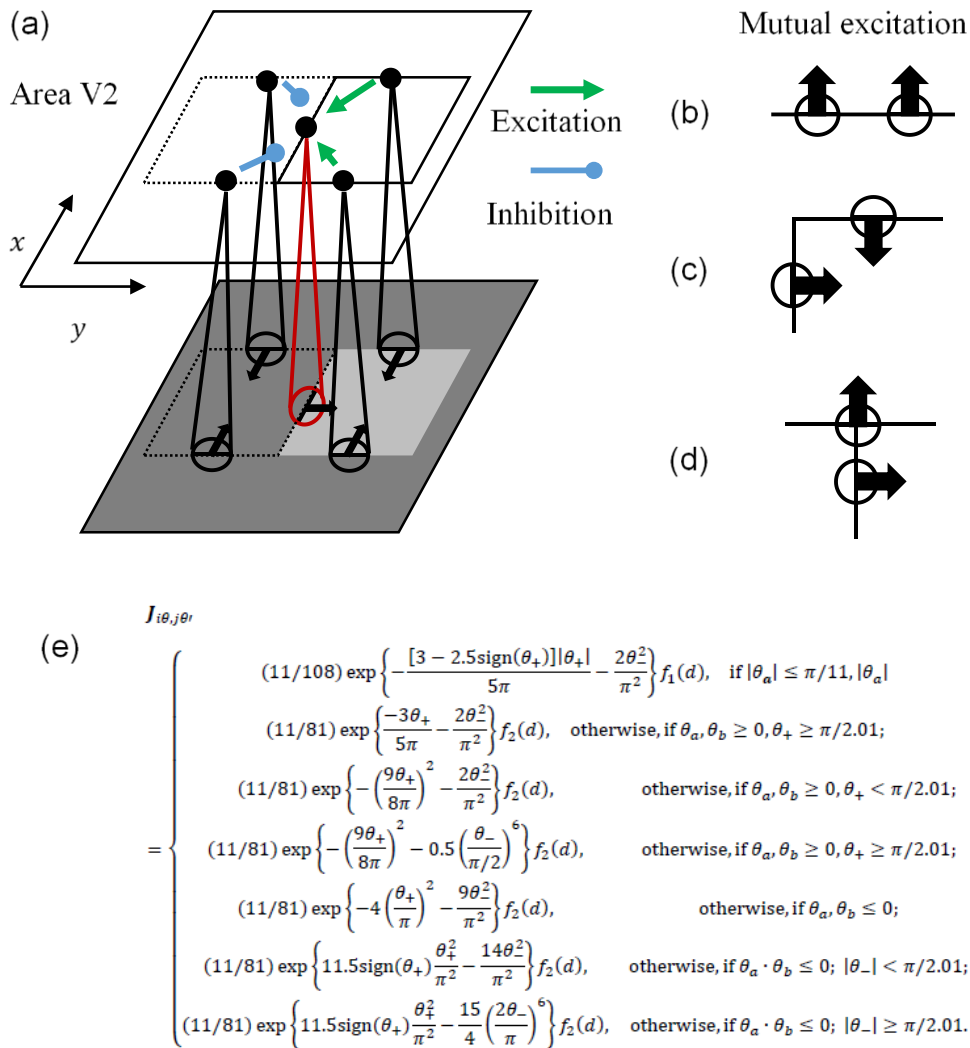
Figure 2.1 A diagram of Li's horizontal connection model and a description of the connections.

(a) Horizontal connections between neurons with different object-side selectivity. Ovals on the image represent receptive fields, while dots represent V2 neurons. Li considered cues for BO such as (b) collinearity, (c) corners and (d) T-junctions for describing connections for mutual excitation. Mutual inhibition was described in a similar manner. (e) An excerpt of the complex neural descriptions. The description for neural connections in Li's model contains 23 parameters.
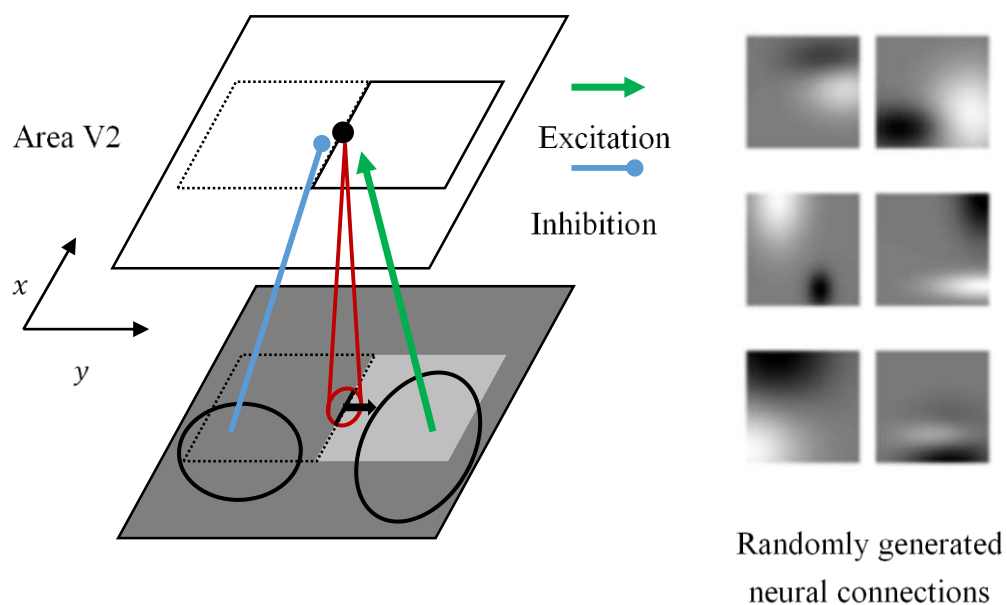
Figure 2.2 Diagram of the randomly generated feedforward connections of the model by Sakai et al. (2012).

The receptive field of a neuron with a right object-side selectivity is shown by the red oval. Responses of V1 neurons towards edges are pooled and given either excitatory or inhibitory properties based on randomly generated Gaussian weights. Some of the weights used are shown on the right. Black areas correspond to inhibition, while white areas correspond to excitation.
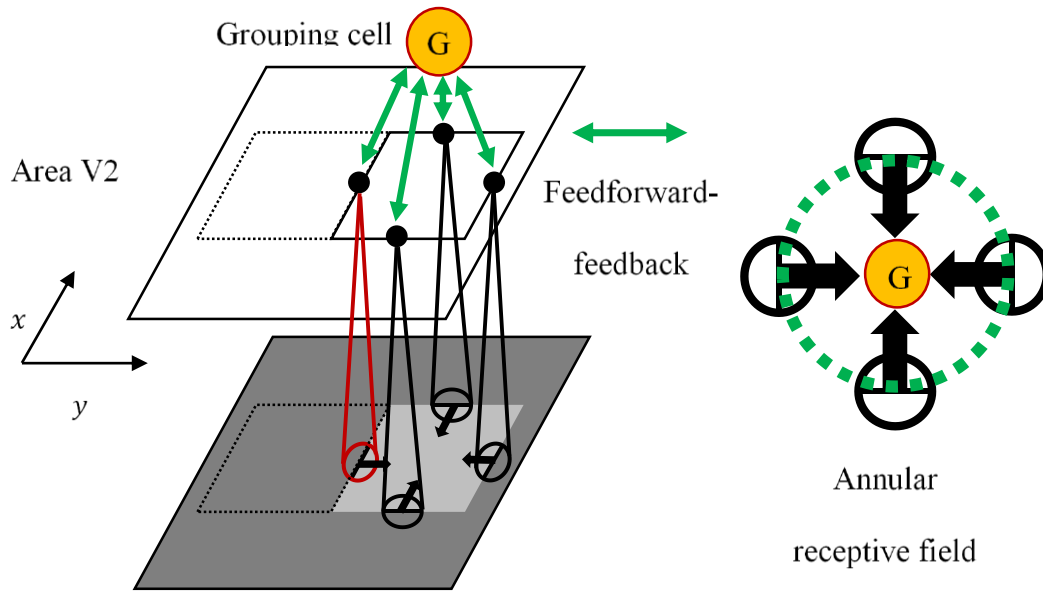
Figure 2.3 Diagram of hypothetical "grouping" cells in the feedforward-feedback model by Craft et al. (2007).

The receptive field of a neuron with a right object-side selectivity is shown by the red oval. Responses of V2 send feedforward signals to a "grouping cell" which has an annular receptive field. In effect, neurons with object-side selectivity (shown aby arrows) to the center of an annular receptive field are "grouped" together. The more annularly located V2 neurons exist, the stronger the feedback signal is.

2.2 Vision from a computational approach

In addition to the framework introduced earlier, Marr (1982) also proposed that visual processing in the brain can be expressed as three distinct stages [15]:

(1) Computational theory : Defining the problem and deriving a solution to solve a specific computational goal.

(2) Algorithms/software : Deducing algorithms that can carry out the computation task.

(3) Hardware implementation : Executing the deduced algorithms with physical means.

Marr's approach has been adopted in other studies of visual processing in the brain. For example, there are two rival computational theories for information processing in area V1: uncertainty principle and sparse coding. The Gabor function, which is the solution of minimizing the product of the uncertainty of spatial location and spatial frequency, was found to resemble the receptive fields of simple cells in area V1 [22], [23]. Increasing the sparseness of a set of filters required to encode natural image patches, resulted in filters which resembled the receptive fields of simple cells in area V1 [24]. Both studies are based on defining an optimization problem, where a suitable energy function is minimized.

As of this date, there are no other studies which approach studying area V2, specifically regarding to BO assignment, from the information processing standpoint outlined by Marr (1982). Such an approach has advantages over the existing phenomenalistic one, adopted by existing studies of BO assignment in V2. For example, deduced neural networks and algorithms might be able to provide a generalization of the various models outlined earlier. More specifically, it may be able to generate the neural connections by Li (2000) using fewer parameters.

3   Theory and algorithm

3.1 Computational theory

Figure 3.1 outlines the the central concept in this study based on the framework proposed by Marr (1982). This study proposes that visual information from the retina is used to extract edges in area V1. BO assignment in V2 is processed based on the output of area V1. Surface depth order is calculated in area V4 using the output of area V2. Interactions between these two visual areas will also be examined. This study will focus on BO assignment (area V2) and surface depth order (area V4).

Variables must be defined for BO and depth order to formulate the BO assignment problem. As demonstrated in the introduction, the direction of BO assignment for local edge segments can be expressed by arrows pointing to the occluding object. **Figure 3.2** shows an image that can be perceived as a light-grey rectangle occluding a dark-grey rectangle on its right. The occluded-occluding perception of objects can be expressed as a scalar field $\phi(x, y)$ taking either zero or positive values. Zero ϕ values correspond to background regions, whereas larger positive ϕ values correspond to regions viewed as located closer in the foreground. In this example, the occluding rectangle has a larger depth order value ($\phi = 2$) than the occluded rectangle ($\phi = 1$). Differentiating depth order produces a vector field in the Cartesian coordinate system $\boldsymbol{E}(x, y) = \left(E_x(x, y), E_y(x, y)\right)^{\mathrm{T}}$, where $E_x(x, y)$ and $E_y(x, y)$ are vector components in the $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ direction. This vector field resembles an array of BO signals. BO assignment problem can therefore be approached as a mathematical relationship between a vector field and a scalar field. The author proposes that the following electromagnetic theorem is an applicable analogy to express the mathematical relationship between BO and depth

order: the gradient of an electric potential field $\phi(x, y)$ is equal to its corresponding electric field $\boldsymbol{E}(x, y)$,

$$\boldsymbol{E}(x, y) = \nabla\phi(x, y). \tag{3.1}$$

Therein, operator ∇ is defined as partial differentiation in the $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ direction, $\nabla \overset{\text{def}}{=}$ $(\partial/\partial x, \partial/\partial y)^{\mathrm{T}}$. Both $\boldsymbol{E}(x, y)$ and $\phi(x, y)$ are unknown values prior to the presentation of a visual stimulus. This study attempts to solve the problem by assuming that the value for a two-dimensional (2D) vector field that evolves over time $\boldsymbol{E}(x, y, t)$ is given at $t = 0$. There is no guarantee that **Eq. (3.1)** is satisfied at this initial condition, and thus $\boldsymbol{E}(x, y, t)$ must be updated accordingly.

   Eq. (3.1) is satisfied when $\boldsymbol{E}(x, y)$ is a conservative vector field; line integration between two arbitrary points in a conservative vector field is independent of the integration path. A mathematical theorem states that the "curl", or rotation, is zero at all spatial points of a conservative vector field. Consider the following analogy to understand the concept of "curl". Calculating the curl at an arbitrary point in a vector field is similar to placing a windmill and observing its rotation. Vector field $\boldsymbol{E}(x, y)$ represents wind blowing on the windmill blades. A counter-clockwise (clockwise) rotation of the windmill coincides with a positive (negative) value of curl. The curl at $(0,0)$ is non-zero for examples in **Figure 3.3a-b**, because a windmill would rotate in a counter-clockwise (positive curl) or a clockwise (negative curl) direction, respectively. In contrast, the curl at $(0,0)$ for examples in **Figure 3.3c-d** are both zero. In mathematical terms, the curl of a 2D vector field $\boldsymbol{E}(x, y)$ is the difference of partial differentiation in the $\hat{\boldsymbol{y}}$ direction of $E_x(x, y)$ and in the $\hat{\boldsymbol{x}}$ direction of $E_y(x, y)$,

$$\mathrm{curl}(\boldsymbol{E}(x, y)) = \frac{\partial}{\partial x}E_y(x, y) - \frac{\partial}{\partial y}E_x(x, y). \tag{3.2}$$

This formula will be the basis for constructing a suitable energy function.
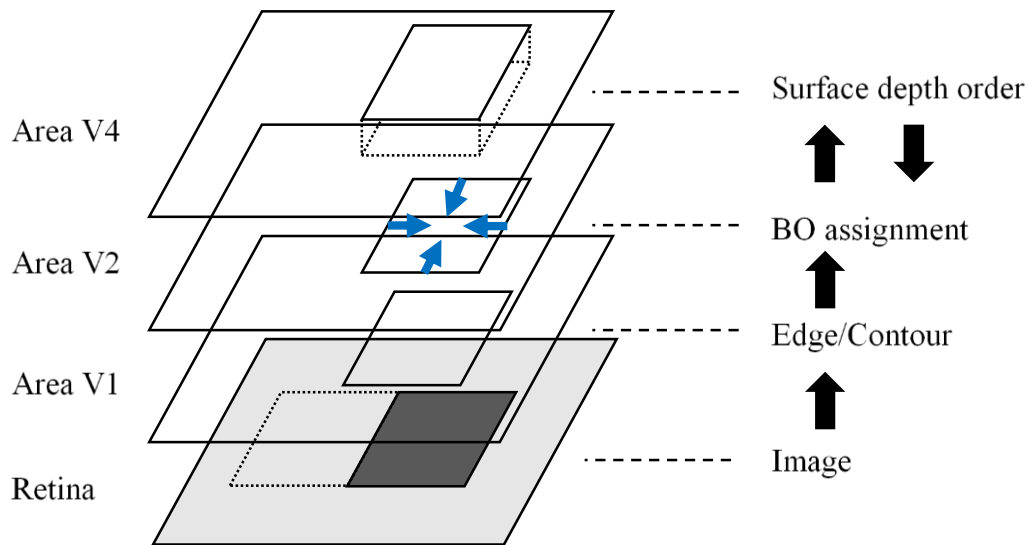
Figure 3.1 Hierarchical structure of visual information processing proposed in this study.

Edges are extracted in area V1 from a retinal image. Edge information is relayed to area V2, where BO assignment is processed, and then to area V4 where depth order is calculated (surface reconstruction). Top-down interaction from area V4 to area V2 will also be examined.
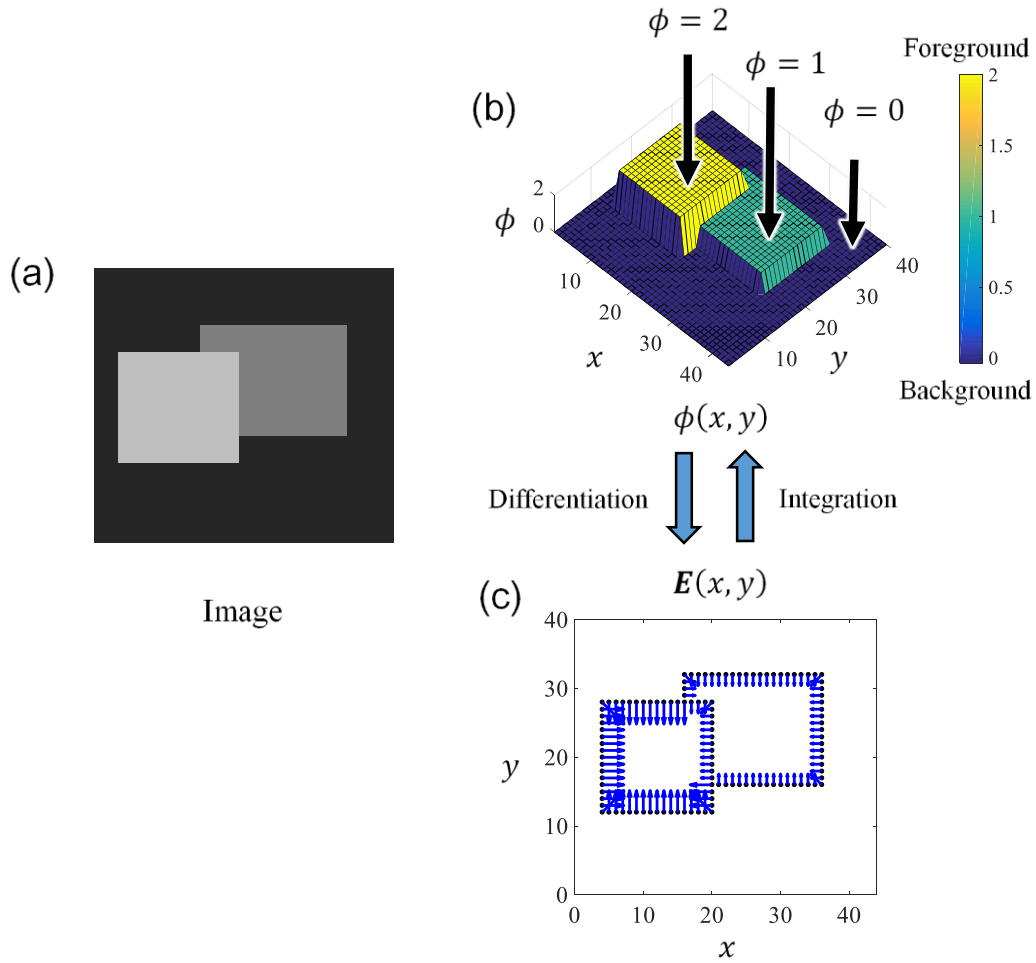
Figure 3.2 Proposed differentiation-integration relationship between depth order and BO.

(a) An image that can be interpreted as a light-grey rectangle occluding a dark-grey rectangle on its right. (b) Perceived depth order of the image expressed as a depth order scalar field $\phi(x,y)$. The occluding rectangle is in the foreground and thus has a higher depth order value ($\phi = 2$) than the occluded rectangle ($\phi = 1$). (c) Differentiating depth order in (b) produces a vector field $E(x,y)$ with arrowheads pointing to the object that "owns" the border. These vectors are analogous to BO signals, thus leading to the proposition that depth order and BO signals are in an integration-differentiation relationship, similar to the relationship of electric potential and its corresponding electric field $E(x,y) = \nabla\phi(x,y)$.
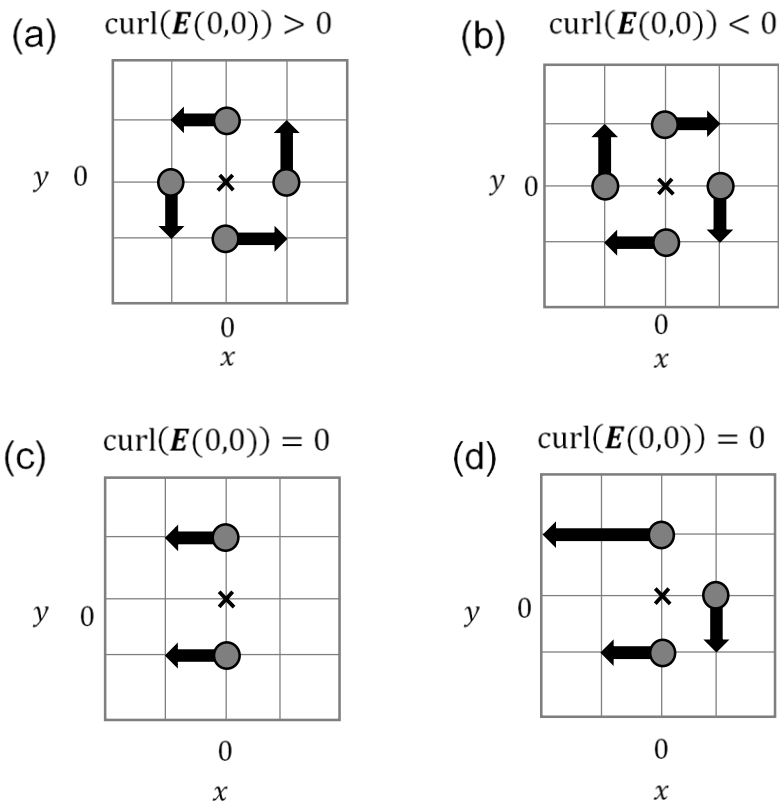
Figure 3.3 The "curl" operator calculates local rotation in a vector field.

An imaginary windmill located at the cross mark at $(0,0)$ will rotate depending on the size and direction of adjacent wind vectors in $E(x, y)$. A positive curl at $(0,0)$ coincides with (a) the windmill rotating in a counter-clockwise direction, whereas a negative curl at $(0,0)$ coincides with (b) the windmill rotating in a clockwise direction. A windmill placed at $(0,0)$ in (c) and (d) will not rotate, and thus a zero curl exists there. Every spatial location in a conservative field has zero curl, i.e. an imaginary windmill will not rotate regardless where it is placed.

3.2 Energy function and update rule

An energy function for minimizing curl is constructed below,

$$J[\boldsymbol{E}(x,y)] = \frac{1}{2} \iint (\mathrm{curl}(\boldsymbol{E}(x,y)))^2 \, dx dy \to \min. \tag{3.3}$$

Integration is applied on the entire image region. An update rule for $\boldsymbol{E}(x,y,t)$ is then

derived by applying the steepest descent method on energy function $J[\boldsymbol{E}(x,y)]$,

$$\frac{\partial}{\partial t}\boldsymbol{E}(x,y,t) = \begin{pmatrix} \frac{\partial}{\partial t}E_x(x,y,t) \\ \frac{\partial}{\partial t}E_y(x,y,t) \end{pmatrix} \propto \begin{pmatrix} \frac{\partial^2}{\partial y^2}E_x(x,y,t) - \frac{\partial^2}{\partial x \partial y}E_y(x,y,t) \\ -\frac{\partial^2}{\partial x \partial y}E_x(x,y,t) + \frac{\partial^2}{\partial x^2}E_y(x,y,t) \end{pmatrix}. \tag{3.4}$$

This update rule is also expressible in a concise equation using the curl operator and the

perpendicular of the gradient operator $\boldsymbol{\nabla}^\perp \overset{\text{def}}{=} (-\partial/\partial y, \partial/\partial x)^{\mathrm{T}}$,

$$\frac{\partial}{\partial t}\boldsymbol{E}(x,y,t) \begin{pmatrix} -\frac{\partial}{\partial y}\mathrm{curl}\big(\boldsymbol{E}(x,y,t)\big) \\ \frac{\partial}{\partial x}\mathrm{curl}\big(\boldsymbol{E}(x,y,t)\big) \end{pmatrix} = \boldsymbol{\nabla}^\perp \mathrm{curl}\big(\boldsymbol{E}(x,y,t)\big). \tag{3.5}$$

3.3 Model details

3.3.4 Production of BO signals from an input image

Borders were first calculated from greyscale images. Consider a 2D discretized image,

where x and y are set as unit length one. For an artificial image $I(x,y)$ with uniformly

filled figures, edges can be treated as figure borders. Figure border region $B(x,y) \in$

$\{0,1\}$ was calculated by applying Sobel edge filters on image $I(x,y)$ and thinning the

result. $B(x,y)$ is simply a binary map that carries a value of 1 at figure borders and 0 at

all other spatial locations.

A scheme for implementing the update rule derived in **Eq. (3.4)** will be outlined. Partial

differentiation in **Eq. (3.4)** can be substituted by convolution with Gaussian derivative

filters [25]. Convolution between two arbitrary functions $K(x, y)$ and $L(x, y)$ using the convolution operator $*$ is defined as

$$K(x, y) * L(x, y) \stackrel{\text{def}}{=} \sum_n \sum_m K(x - m, y - n) L(m, n). \tag{3.6}$$

In this study, derivatives of normalized two-dimensional Gaussian function of standard deviation σ,

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \tag{3.7}$$

was convolved with $E_x(x, y, t)$ and $E_y(x, y, t)$ to form

$$\begin{pmatrix} \frac{\partial}{\partial t} E_x(x, y, t) \\ \frac{\partial}{\partial t} E_y(x, y, t) \end{pmatrix}$$
$$\propto \begin{pmatrix} \frac{\partial^2}{\partial y^2} G_\sigma(x, y) * E_x(x, y, t) - \frac{\partial^2}{\partial x \partial y} G_\sigma(x, y) * E_y(x, y, t) \\ -\frac{\partial^2}{\partial x \partial y} G_\sigma(x, y) * E_x(x, y, t) + \frac{\partial^2}{\partial x^2} G_\sigma(x, y) * E_y(x, y, t) \end{pmatrix}. \tag{3.8}$$

Update rules for the $\hat{x}$ (horizontal direction) and $\hat{y}$ (vertical direction) vector components above were applied on the border region $B(x, y)$ until energy function $J[E(x, y, t)]$ was sufficiently minimized. In this model, standard deviation for the filters was set as $\sigma = 1$. The discrete filters used are shown in **Figure 3.4.**

Model dynamics were derived by constructing and energy function in **Eq.(3.3)** and minimizing it using the steepest descent method. A possible shape of the energy function curve is shown in **Figure 3.5**. Setting initial vectors is analogous to placing an object on the energy curve. This object is pulled down (in the negative gradient direction) to a local minimum, thus the name "steepest descent method". The energy function curve has two local minima where $E(x, y)$ is at a stable state. For example, if initial vectors are set at

either (a) or (b), updating vectors will result in stable state 1. Conversely, if initial vectors are set at either (c) or (d), updating vectors will result in stable state 2. In natural viewing conditions, border ownership (and depth order) converges at a single stable state. Thus, determining suitable initial vectors that will converge at a stable state consistent with perception is important. The model's behavior and the effect of initial vectors on its output will be examined in the next section.

$$\frac{\partial^2}{\partial y^2} G_\sigma(x, y)$$
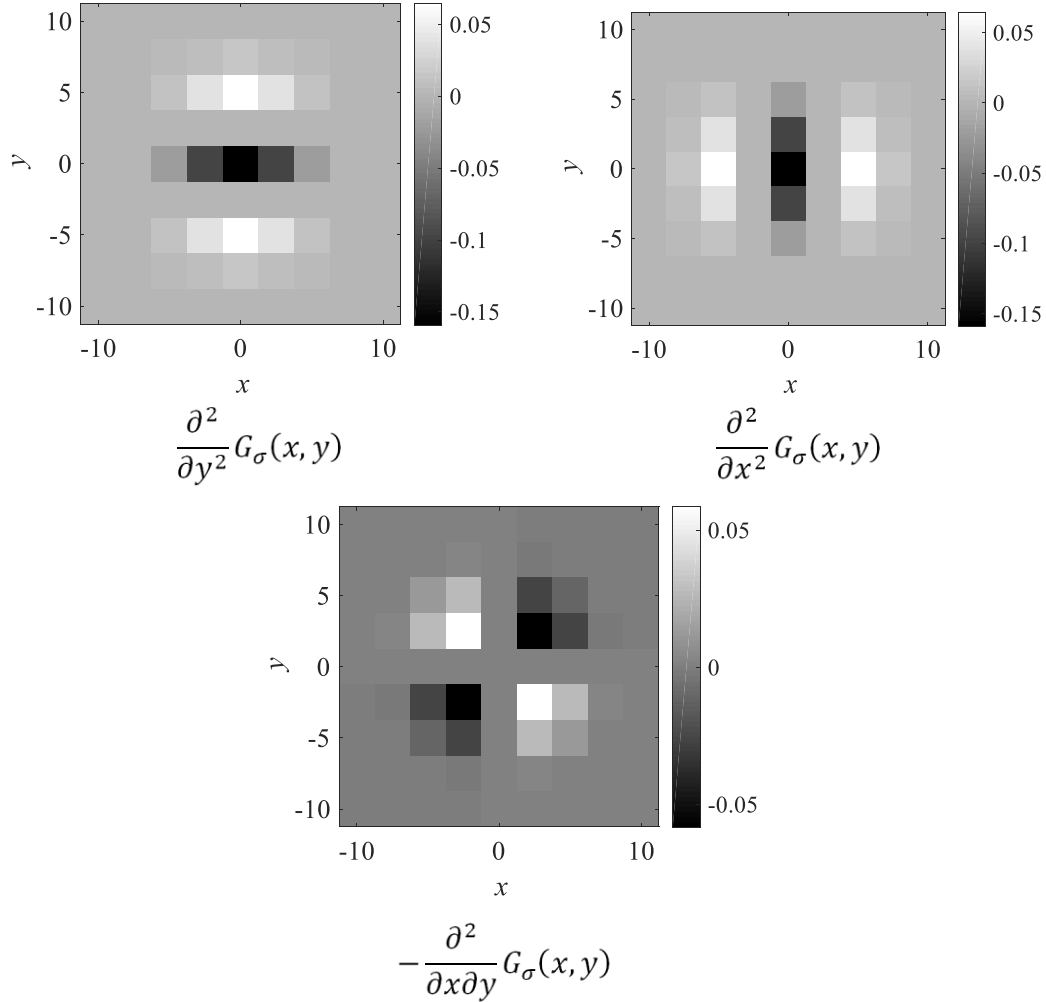
$$\frac{\partial^2}{\partial x^2} G_\sigma(x, y)$$

$$-\frac{\partial^2}{\partial x \partial y} G_\sigma(x, y)$$

Figure 3.4 Discrete Gaussian derivative filters used in simulations for standard derivative σ = 1.

Filters for $\frac{\partial^2}{\partial y^2} G_\sigma(x, y)$, $\frac{\partial^2}{\partial x^2} G_\sigma(x, y)$ and $-\frac{\partial^2}{\partial x \partial y} G_\sigma(x, y)$ are shown. White pixels coincide with positive values, whereas black pixels coincide with negative values. These filters were convoluted with $E_x(x, y, t)$ and $E_y(x, y, t)$ vector components in accordance to the update rule in **Eq. (3.8)**.
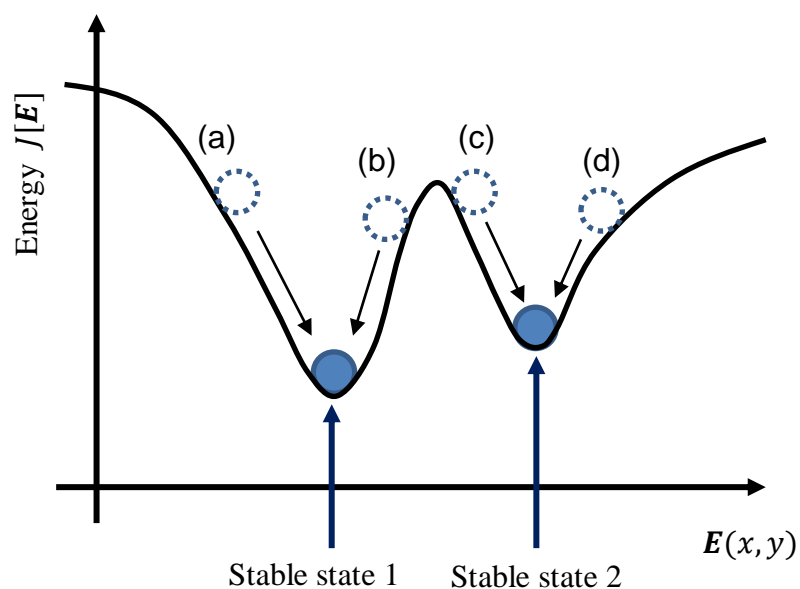
Figure 3.5 An example of an energy function with two local minima (stable state 1 and stable state 2).

The steepest descent method is analogous to an object pulled down by gravity towards the local minimum. Initial conditions (a) and (b) will result in stable state 1. Conversely, initial conditions (c) and (d) will result in stable state 2.

3.3.5 Production of depth order surface from BO signals

In theory, the result of line integration between any two points in a conservative field is the same regardless of the path taken. However, during the updating process line integration is path dependent. One method to estimate depth order as updating is conducted, is to sum line integration from multiple paths and rescale the value. Consider an open path C starting from an arbitrary position and ending at (x_0, y_0), $C(s) = (x(s), y(s))$, where s is the arc length. When the tangent of $C(s)$ is defined as $\hat{t}(s) = dC(s)/ds$, line integral of vectors $E(s)$ is calculated using $\sum_s E(s) \cdot \hat{t}(s)$. At a spatial position (x_0, y_0), the depth order value $\phi(x_0, y_0, t)$ was calculated using line integration of $E(x, y, t)$. First, consider integral paths of horizontal and vertical lines passing through (x_0, y_0),

$$\phi(x_0, y_0, t) = \sum_x E_x(x, y_0, t)\mathrm{sign}(x_0 - x) + \sum_y E_y(x_0, y, t)\mathrm{sign}(y_0 - y). \quad (3.9)$$

Therein, $\mathrm{sign}(x) = 1$ for $x > 0$ and $\mathrm{sign}(x) = -1$ for $x < 0$. Summation was operated on the image domain excluding its own position, (x_0, y_0). Depth order was normalized from 0 to 1 unless specified. The first component in **Eq. (3.9)** describes integration on the horizontal direction, whereas the second describes integration on the vertical path. **Figure 3.6** shows an example of calculation in the horizontal direction.

This method of calculation is similar to weights resembling a Heaviside function with $+1$ weights in from the $+\hat{x}$ ($+\hat{y}$) direction, and -1 weights in from the $-\hat{x}$ ($-\hat{y}$) direction. Line integration with horizontal paths is qualitatively similar to filtering the E_x component with a vertically shifted negative Heaviside step function $H(x) - 1$, where

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}. \quad (3.10)$$
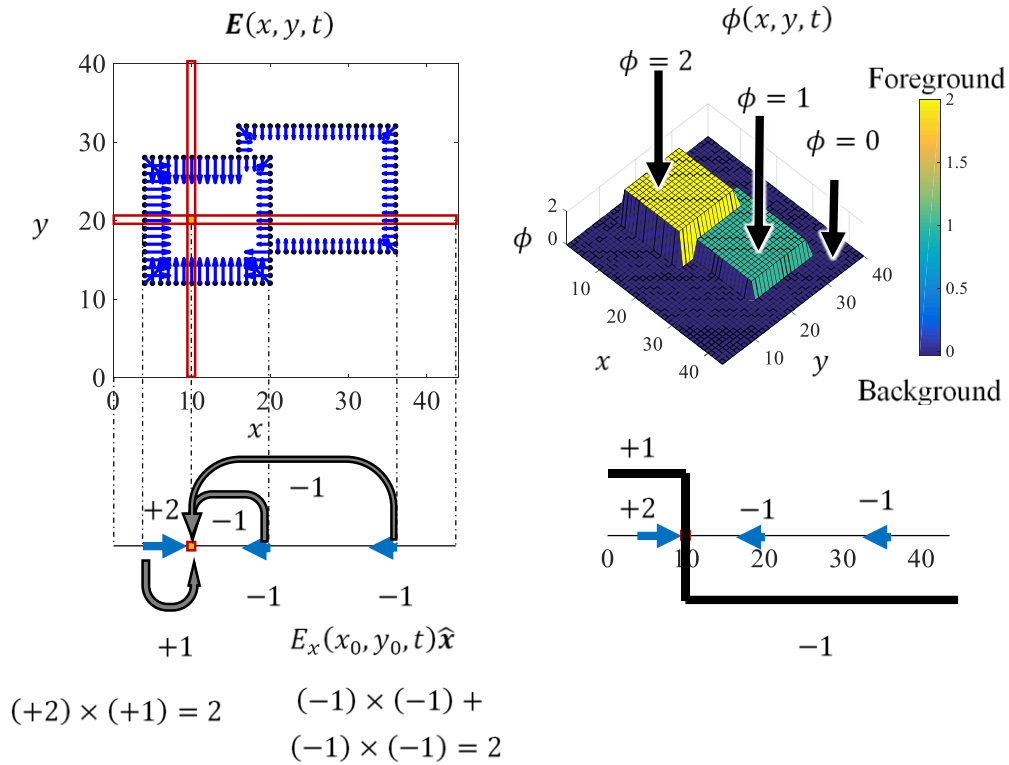
Figure 3.6 An example of calculating depth order for an arbitrary point (x_0, y_0) in a BO vector field.

For a conservative field, the result of line integration between any two points is independent of the path taken. Line integration for a path that forms a horizontal (or vertical) straight line through (x_0, y_0) is similar to applying a filter with weights resembling a Heaviside function on vector field $E(x, y)$.

4   Neural network implementation

4.1   Model behavior and initial vector dependence

To examine the model's behavior, initial vectors were assigned randomly at the borders of the three benchmark stimuli used in physiological experiments by Zhou et al. (2000). The three stimuli used were a square (**Figure 4.12a**), a C-shape (**Figure 4.12b**), and a square overlapping a rectangle on its right (**Figure 4.12c**). Vectors were set as (a) $+\hat{x}$, (b) **0** or (c) $-\hat{x}$ at horizontal edges and (a) $+\hat{y}$, (b) **0** or (c) $-\hat{y}$ at vertical edges. Similarly, vectors of magnitude 1 facing inwards (acute side), zero vectors, or vectors of magnitude 1 facing outwards (obtuse side) at L-junctions were also assigned. Vectors facing the acute side of the L-junction was calculated based on the partial derivatives of border $B(x, y)$ at L-junctions by central difference approximation,

$$E(x, y, 0) = \begin{pmatrix} \dfrac{\partial}{\partial x} B(x, y) \\ \dfrac{\partial}{\partial y} B(x, y) \end{pmatrix} \approx \frac{1}{2} \begin{pmatrix} B(x + 1, y) - B(x - 1, y) \\ B(x, y + 1) - B(x, y - 1) \end{pmatrix}. \tag{4.1}$$

Simulations were carried out for $t = 2000$ with 1000 sets of randomly assigned initial vectors for the three figures.

Model outputs were categorized according to the direction of the BO signal located at the position of the neuron's receptive field, represented by the red circle in **Figure 4.1**. The neuron responds selectively to an object placed on the left of its receptive field. This coincides with a left BO signal. The model outputs left-facing vectors at chance-level for all three stimuli.

Although this method of assigning initial vectors fails to reproduce the data from physiological experiments at a significant rate, it is interesting to note that surfaces coinciding with various depth order perception was reproduced. For example, **Figure 4.2**

and **Figure 4.3** shows the two types of surfaces, an object or a hole, reproduced by the model. In addition, for an occluding square, a left (or right) BO signal at the location of the neuron's receptive field coincides with three different depth order percepts, respectively (**Figure 4.4** or **Figure 4.5**).
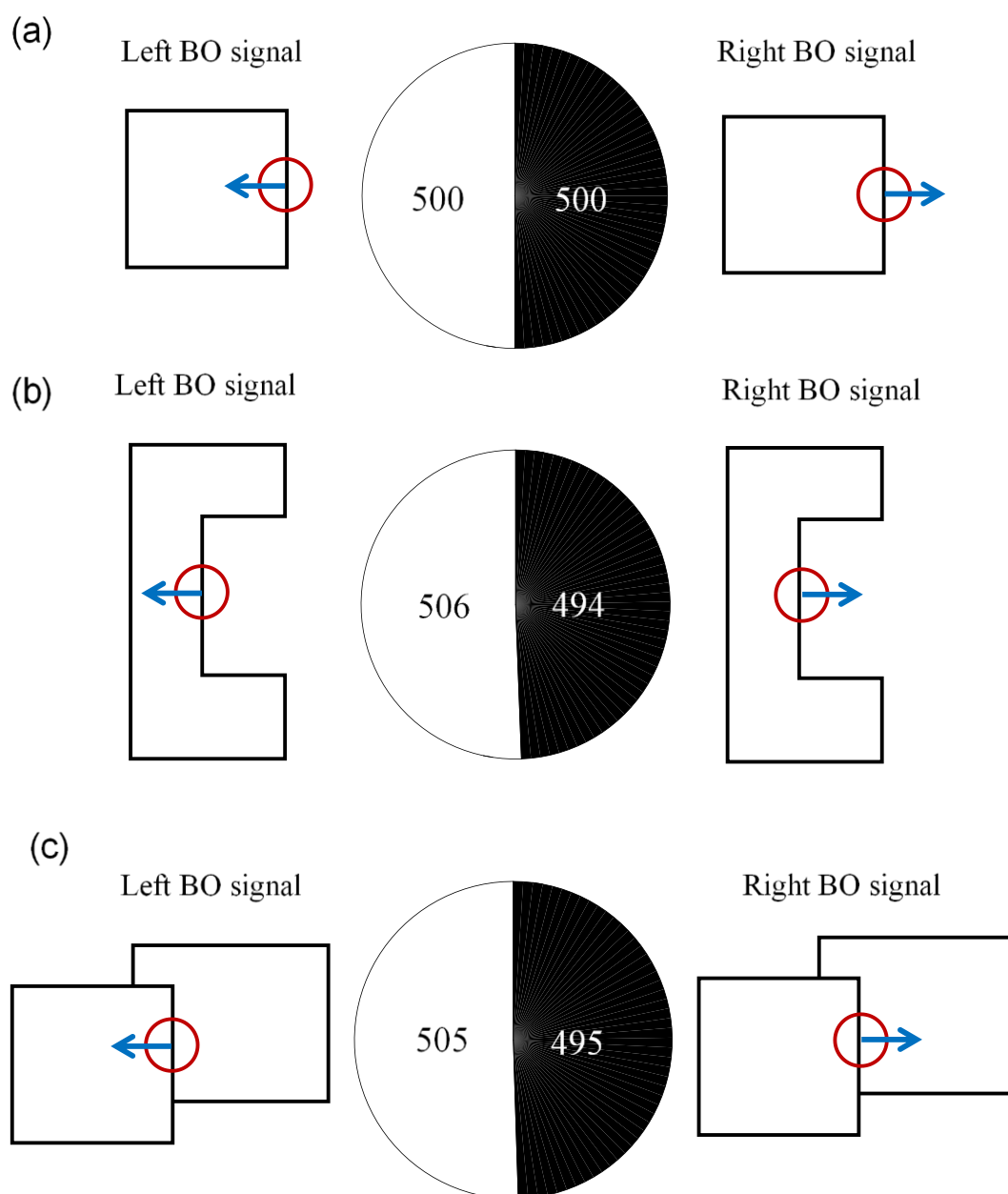
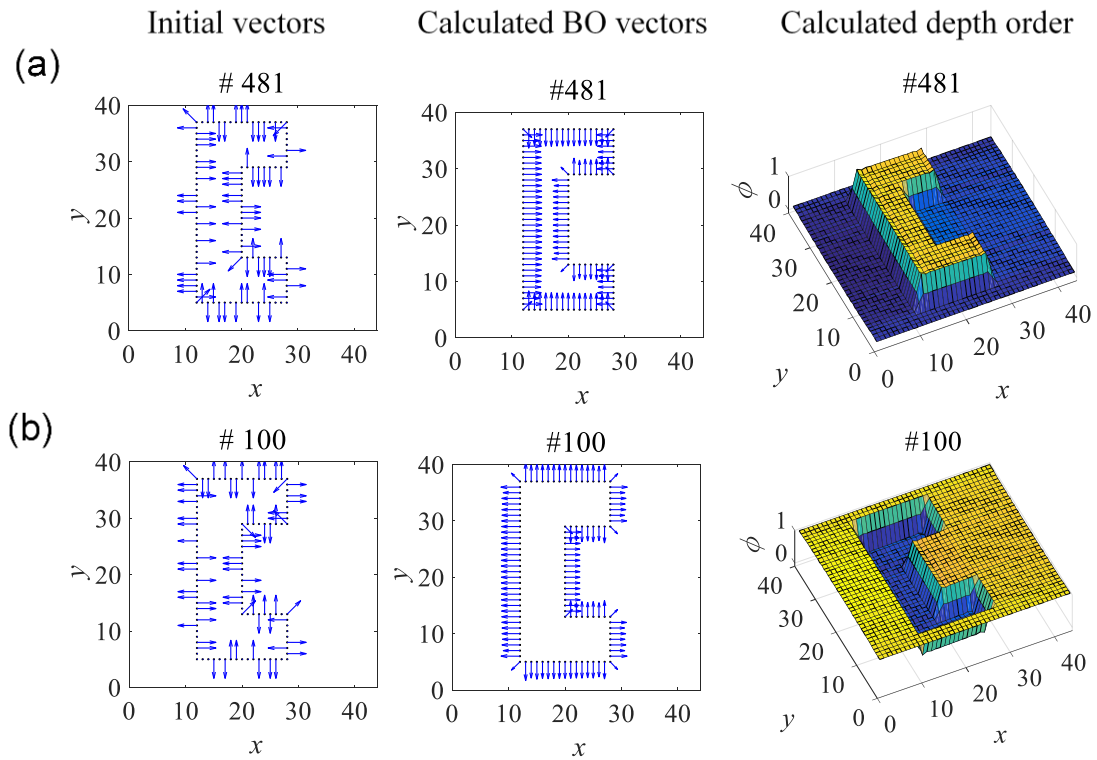**Figure 4.1 Categorization of calculated BO signals from randomly assigned initial vectors.**

The red circle represents a neuron's receptive field located at the border of (a) a square, (b) a C-shaped figure and (c) an overlapping square. These figures are similar to those used in physiological experiments outlined in **Figure 1.4**. Assigning initial vectors randomly (1000 combinations) results in BO signals in the left or right direction at chance-level. Random initial vectors cannot reproduce physiological data consistently.
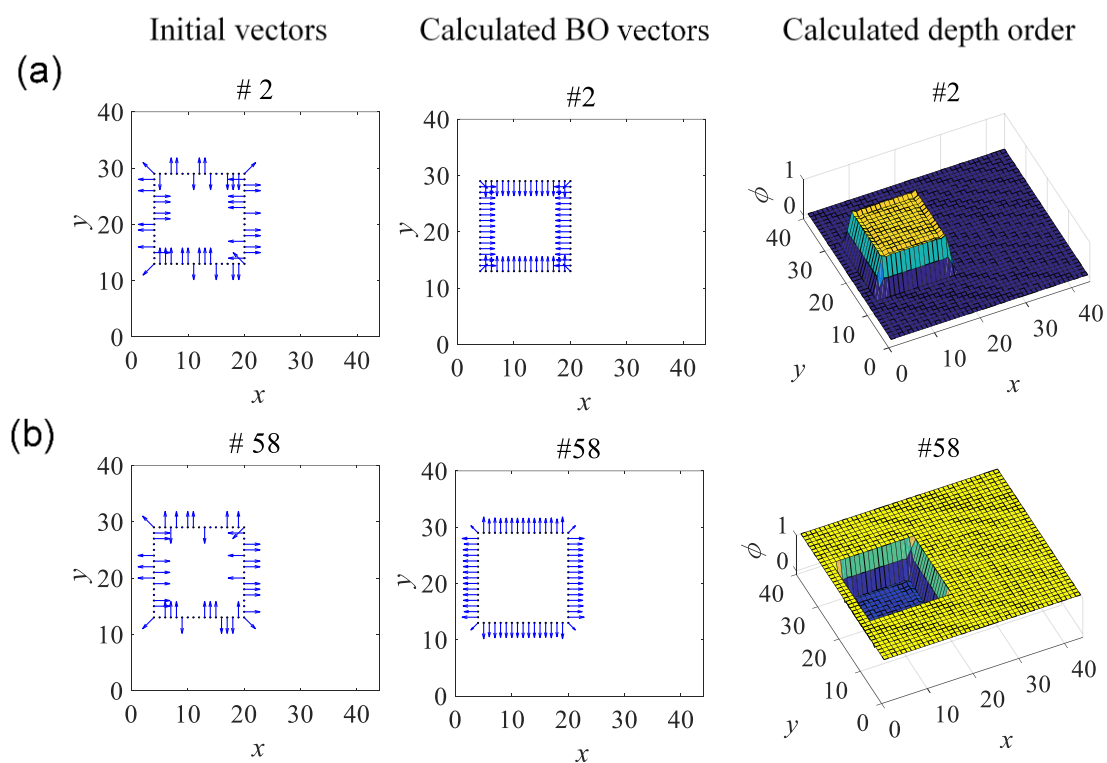
Figure 4.2 Calculated BO vectors and depth order for a square with randomly assigned initial vectors.

Example of the different percepts categorized as (a) a left BO signal and (b) a right BO signal in **Figure 4.1a**. A left BO signal coincides with the percept of a square object, whereas a right BO signal coincides with a square hole. The element number in the set of 1000 randomly assigned initial vectors is shown above the respective results.

Figure 4.3 Calculated BO vectors and depth order for a C-shaped figure with randomly assigned initial vectors.

Example of the different percepts categorized as (a) a left BO signal and (b) a right BO signal in **Figure 4.1b**. A left BO signal coincides with the percept of a C-shaped object, whereas a right BO signal coincides with a C-shaped hole. The element number in the set of 1000 randomly assigned initial vectors is shown above the respective results.
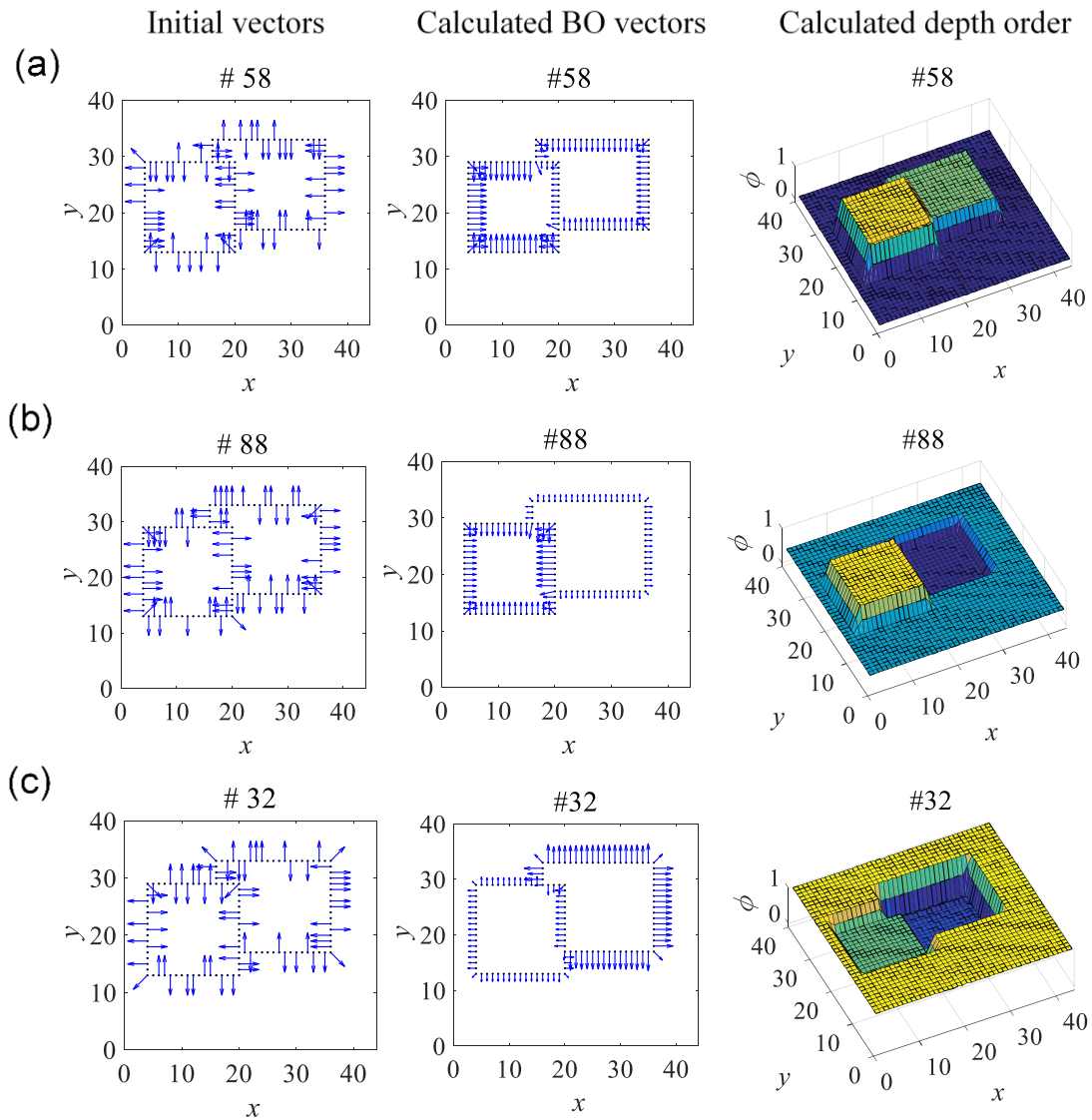
Figure 4.4 Calculated BO vectors and depth order for an overlapping square with randomly assigned initial vectors.

Example of the different percepts categorized as a left BO signal in **Figure 4.1c.** Unlike the square and C-shaped figure, a left BO signal coincides with three distinct depth order percepts (a~c). The element number in the set of 1000 randomly assigned initial vectors is shown above the respective results.
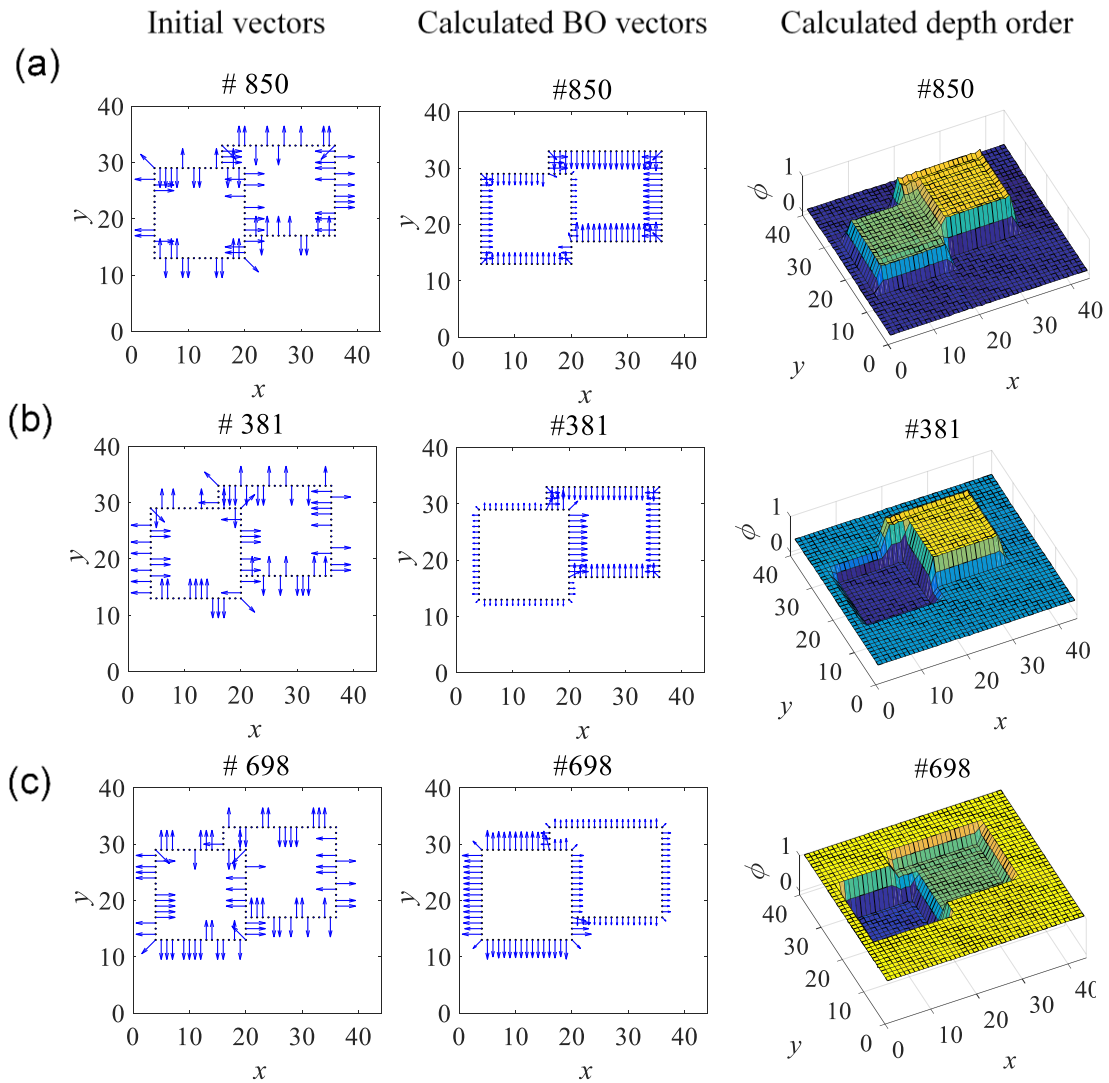
Figure 4.5 Calculated BO vectors and depth order for an overlapping square with randomly assigned initial vectors.

Example of the different percepts categorized as a right BO signal in **Figure 4.1c.** Unlike the square and C-shaped figure, a left BO signal coincides with three distinct depth order percepts (a~c). The element number in the set of 1000 randomly assigned initial vectors is shown above the respective results.

4.2 Determining suitable initial vectors

To reproduce the data from physiological experiments, suitable initial vectors must be assigned. There are two types of junctions that might serve as useful cues: (a) L-junction and (b) T-junctions.

First, let us consider L-junctions. The contour of an occluding object is essentially a simple closed curve. For a closed curve $C(s)$ described by its arc length s, the total curvature along the curve is a positive value $\int_s \kappa(s)ds > 0$. In simple terms, there are more acute angles than obtuse angles in the contour of an occluding object. This is demonstrated in **Figure 4.6** for the three figures used. Vectors facing the inner-side of the L-junctions might serve as suitable initial values. Physiological experiments reveal that some neurons in area V2 respond selectively to bent lines, especially to acute-angled lines, placed in their receptive fields [26], [27]. Gestalt psychologists observed that the convex side of curves are likely to be perceived as the inner-side of an object [28].

Next, consider T-junctions. There are several possible depth-order combinations at around the junction, as illustrated in **Figure 4.7**. Initial BO vectors facing the surface with the largest depth order value. Since there are several possible BO signals, T-junctions may not be a reliable cue for the determination of initial vectors. In summary, initial vectors were assigned facing the inner side of L-junctions using **Eq. (4.1)**, whereas zero vectors were assigned at straight-line segments and T-junctions for the three figures.

Simulations were conducted for $t = 10000$ to confirm the suitability of these initial vectors. The time course from $t = 0$ to $t = 10000$ of $E(x, y, t)$ and $\phi(x, y, t)$ is shown for the three stimuli (**Figure 4.8**, **Figure 4.9** and **Figure 4.10**). he red circle represents the approximate location of the receptive field for the neuron examined in the physiological experiment by Zhou et al. (2000). The neuron responds selectively to an

object on the left side of its receptive field. BO signals (vectors) located in the red circle point towards the left, qualitatively matching experimental results. Vectors around the concave region of the C-shaped figure were corrected to face the occluding object, and not the background. After sufficient updating of vectors reflect, relative depth order for all three stimuli how they are actually perceived. Notably, the occluding rectangle has a larger ϕ value than the occluded rectangle and vector magnitudes are adjusted automatically to reflect this.

These results show that setting initial vectors facing the acute side of L-junctions is an appropriate method to reproduce the left object-side selectivity of the neuron used in physiological experiments. The next subsection will compare the response of the V2 neuron with model outputs.
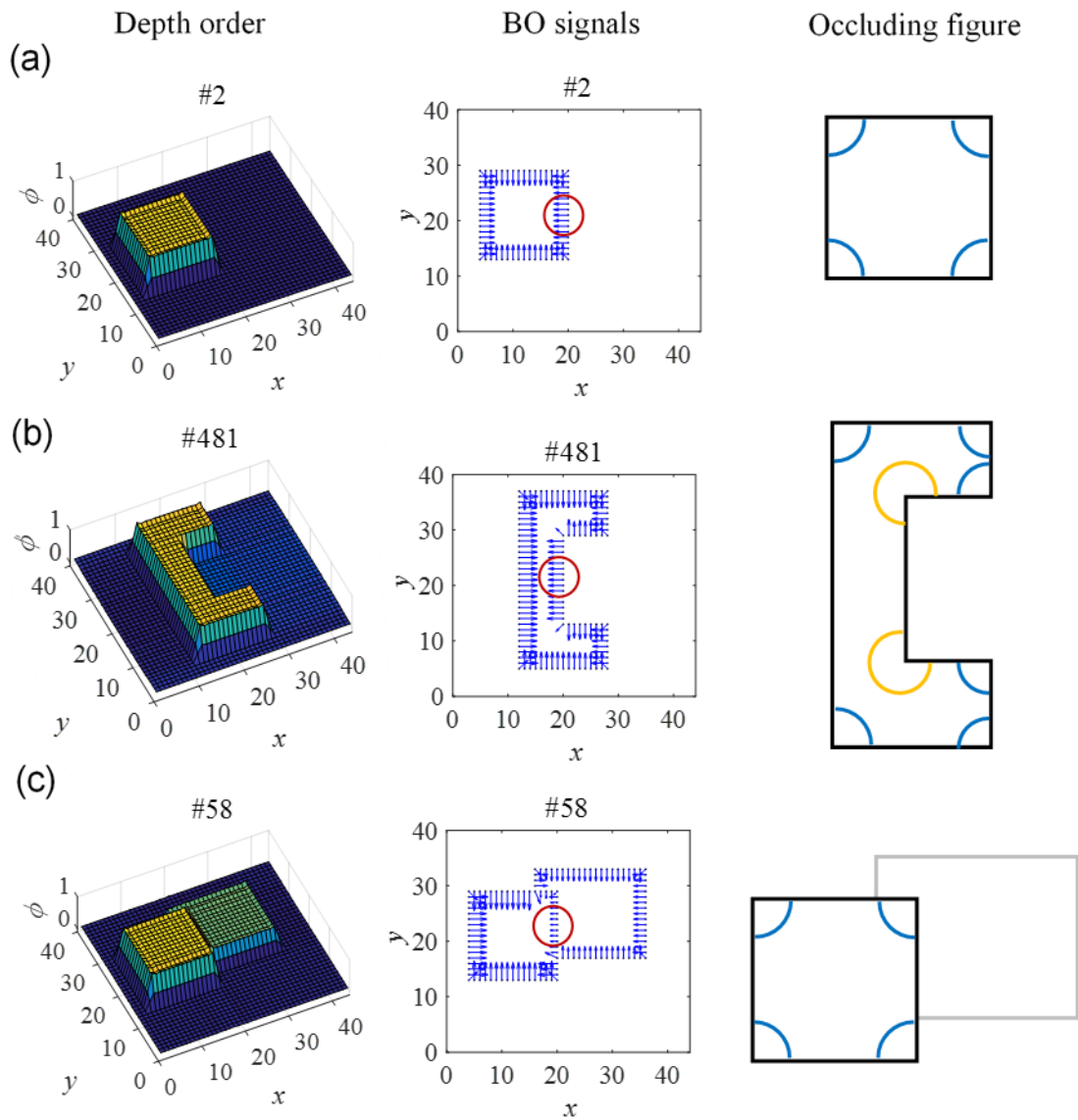
Figure 4.6 The contour of an occluding object has more acute angles than obtuse angles.

Calculated depth order and BO signals which coincide with perception of (a) a square, (b) a C-shaped figure and (c) an overlapping square. For all three examples, the BO signal in the red circle (receptive field) point in the left direction. This agrees qualitatively with the response of the neuron in **Figure 1.4**. The outline of an occluding figure is a simple closed curve. This means that there are more acute angles (blue) than obtuse angles (yellow). Acute angles might be a reliable cue to reproduce data from physiological experiments.
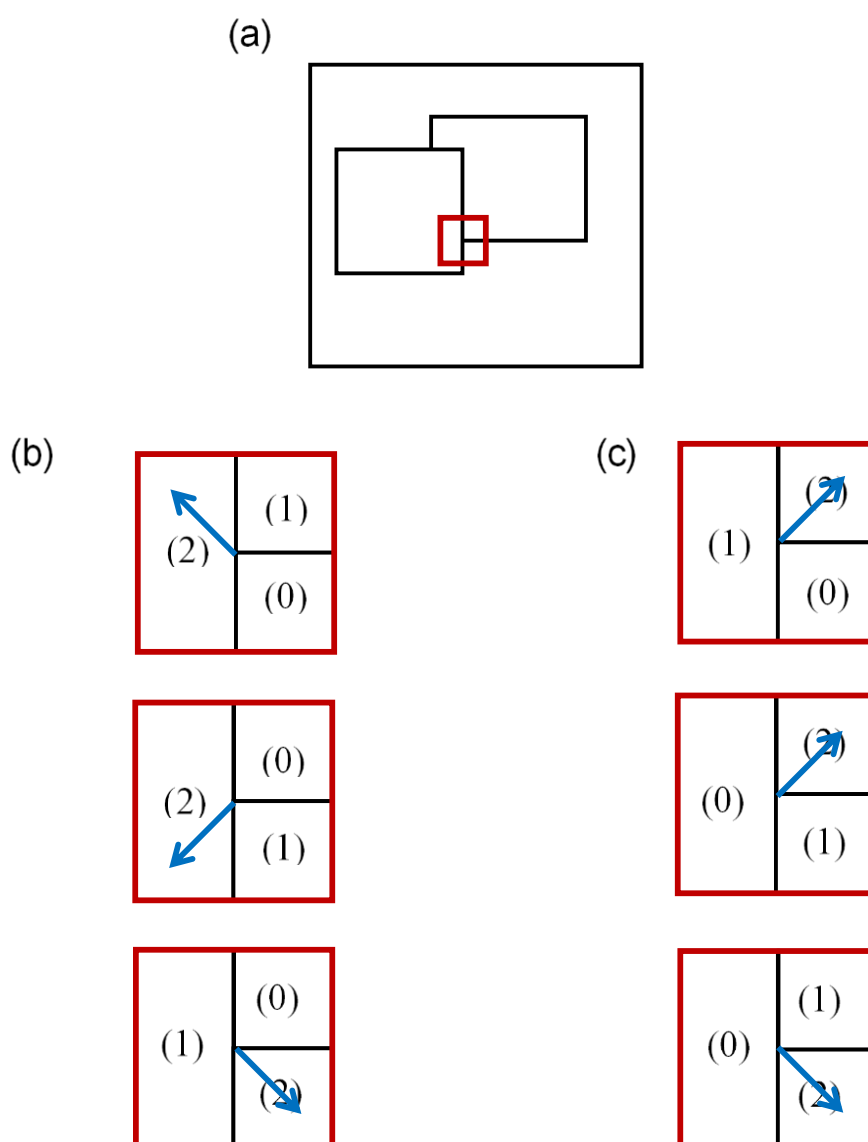
Figure 4.7 Several possibilities for initial vectors exist at T-junctions.

Examples of possible initial vectors at the T-junction in the figure of (a) an overlapping square. The T-junction at the red square is examined. Numbers in the brackets represent the depth order value of the surface. The blue arrow points towards the surface with the highest depth order value. Initial vectors (blue arrows) coinciding with the three respective percepts in (b) **Figure 4.4** and (c) **Figure 4.5**.
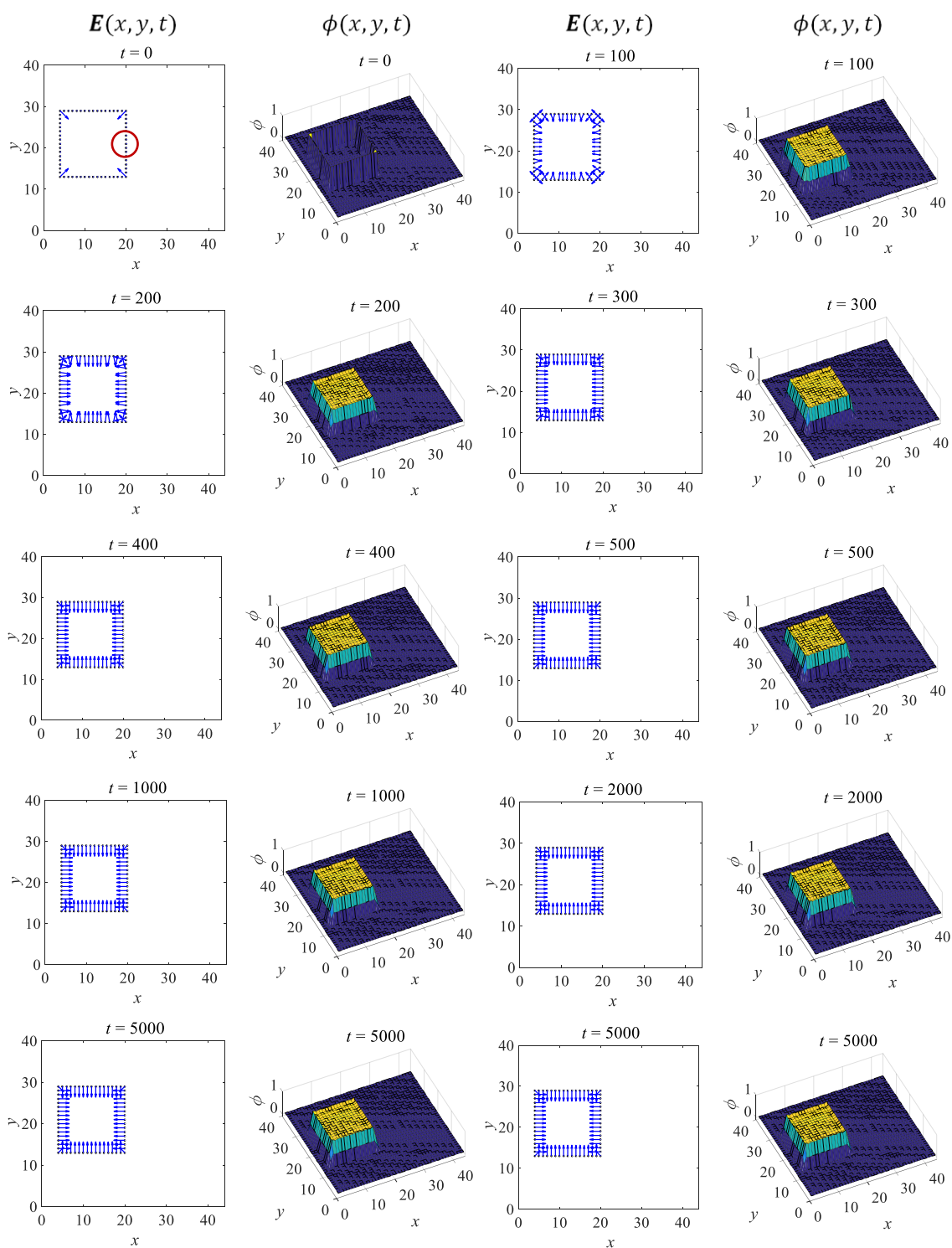
Figure 4.8 Time course of vector field and scalar field for a square.

Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 10000$ for a square. The red circle represents the location of the receptive field.
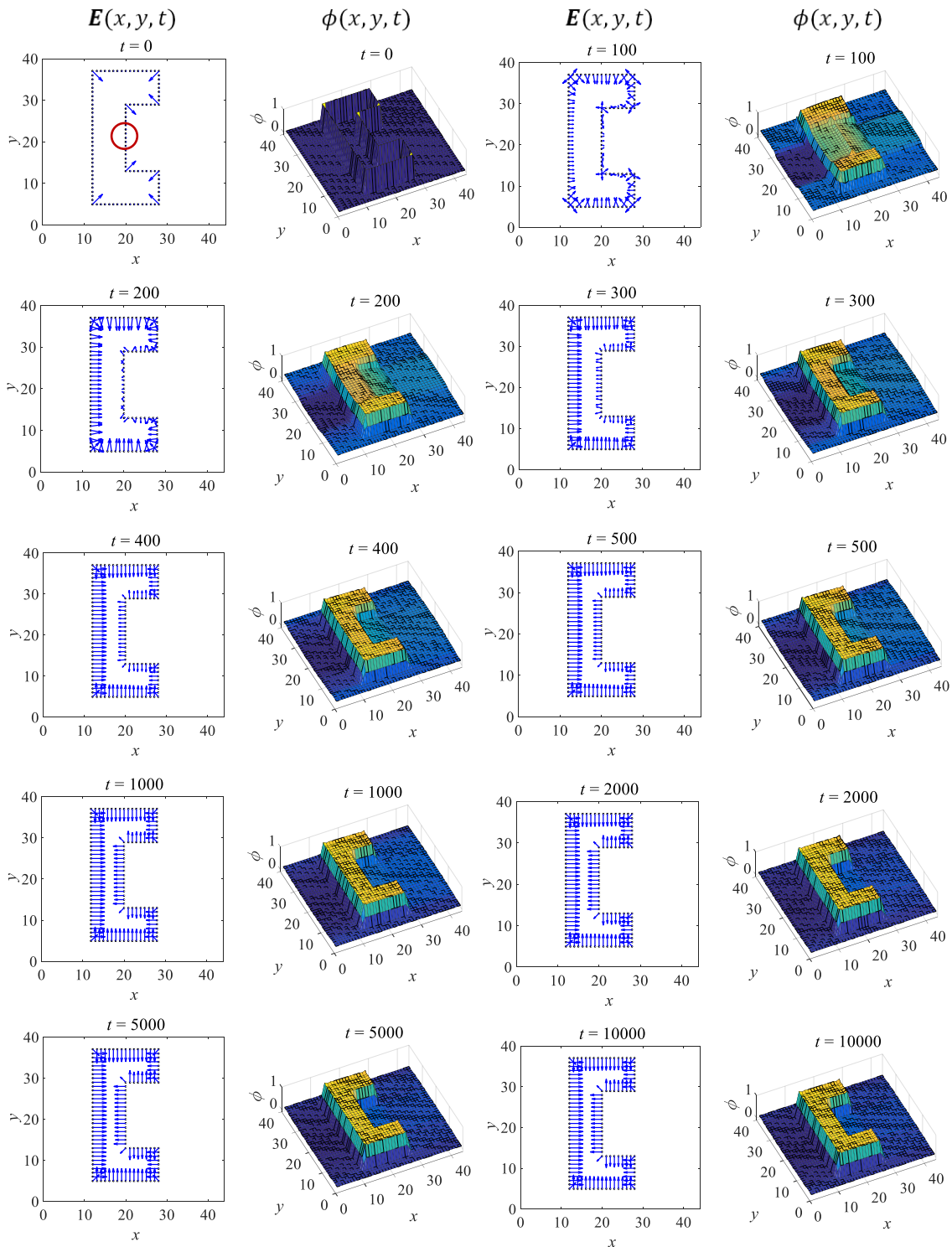
Figure 4.9 Time course of vector field and scalar field for a C-shaped figure.

Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 10000$ for a C-shaped figure. The red circle represents the location of the receptive field.
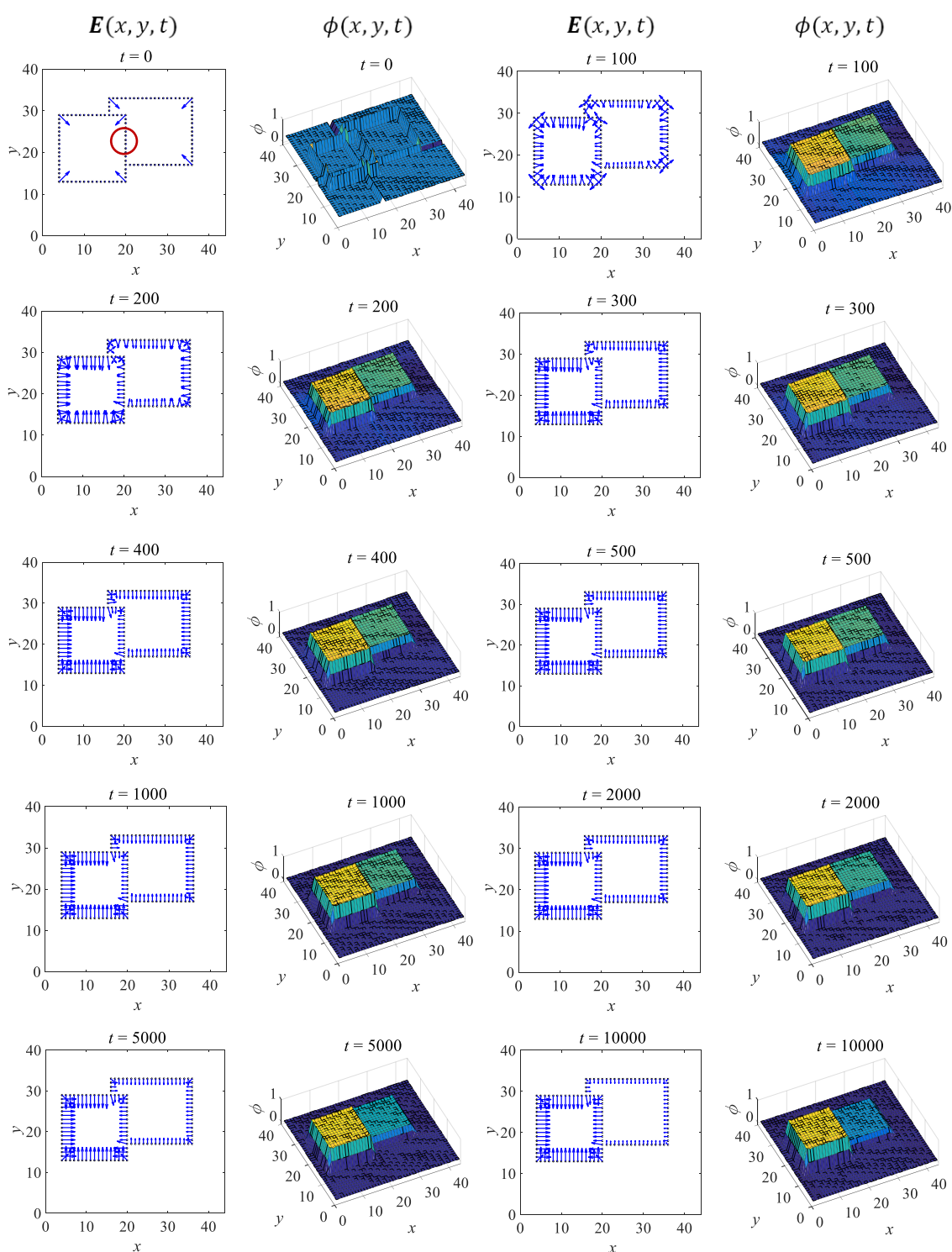
Figure 4.10 Time course of vector field and scalar field for an overlapping square.
Evolution of $E(x,y,t)$ and $\phi(x,y,t)$ over time $t = 0 \sim 10000$ for an overlapping square. The red circle represents the location of the receptive field.

4.3 Reproduction of experimental findings

4.3.1 Modeling V2 neuron responses

The relationship between BO signals and responses of object-side selective neurons in V2 must be modeled to make a direct comparison between model results and neuronal responses. The response of a neuron with right object-side (left object-side) selectivity is termed as $V2_{right}$ ($V2_{left}$). Outputs of these neurons can be modeled as the sum of the neuron's response to a border in its receptive field and ownership information,

$$V2_{right} = Border + Owner_{right}, \tag{4.2}$$

$$V2_{left} = Border + Owner_{left}. \tag{4.3}$$

Therein, Border is positive when a border is located in the neuron's receptive field; $Owner_{right}$ and $Owner_{left}$ are either positive or zero. Based on the model proposed above, a BO signal in the horizontal direction E_x can be defined as the difference in outputs of pairs of neurons with object-side selectivity in opposite directions,

$$E_x = V2_{right} - V2_{left} = Owner_{right} - Owner_{left} \tag{4.4}$$

In **Figure 4.11**, a vertical edge is located in the receptive field, invoking a positive Border response. An object is located to the left of the receptive field. Ownership information pertaining to an object on the left is thus a positive value ($Owner_{left} > 0$), whereas ownership information pertaining to an object on the right is zero ($Owner_{right} = 0$). The resulting BO signal in the horizontal direction is negative, $E_x = -Owner_{left} < 0$, indicating the existence of an object on the left. The same approach is taken to calculate BO signals in the vertical direction E_y.

Model outputs were found to qualitatively match the relative magnitude of neuron responses to the three stimuli up to $t = 2000$; neuron responses are largest for the square and smallest for the C-shaped object. The absolute value of the $\hat{x}$ component at the center

of the red circle $|E_x(20,21,2000)|$ was used to qualitatively compare model responses with physiological data. Notably, for the overlapping rectangle in **Figure 4.12**, vectors were scaled to a magnitude of 2 to account for the existence of two objects of different depth order. $|E_x(20,21,2000)|$ for the three figures are 0.92, 0.78 and 0.89, respectively. **Figure 4.13** shows relative model response and neuronal response for the three stimuli. Values for relative model response were scaled to a maximum of 1, based on the model response to the square. Model responses for the C-shape and the overlapping square are lower than that for the square. This matches V2 neuronal responses for similar stimuli used in Zhou et al. (2000). The proposed model captures the qualitative nature of the response of the neuron towards the three benchmark stimuli better than previous models [17], [18], [21]. The method of recording neuron responses is a possible explanation why physiological data and model outputs qualitatively match. Responses of neurons were only recorded for the first 200 ms after stimulus onset. Neuron recordings for more than 200 ms might present a different property. Craft et al. (2007) also specify a time frame for their model output.

Note that in theory, vector magnitude should be of equal value after updating. **Figure 4.14** shows relative model responses after 500, 1000, 2000, and 10000 iterations. Relative model response qualitatively match physiological data for the first 2000 iterations. Model responses at 10000 iterations show that the model does not update vectors for a C-shaped figure after a certain threshold of iterations. This might be due to the initial vectors facing outwards at the concave region. The relative response to the occluded figure exceeds the relative response to the square at 10000 iterations. Observations were made that the cause of this is due to the magnitude of occluded-object vectors decreasing as time progresses (**Figure 4.10**).
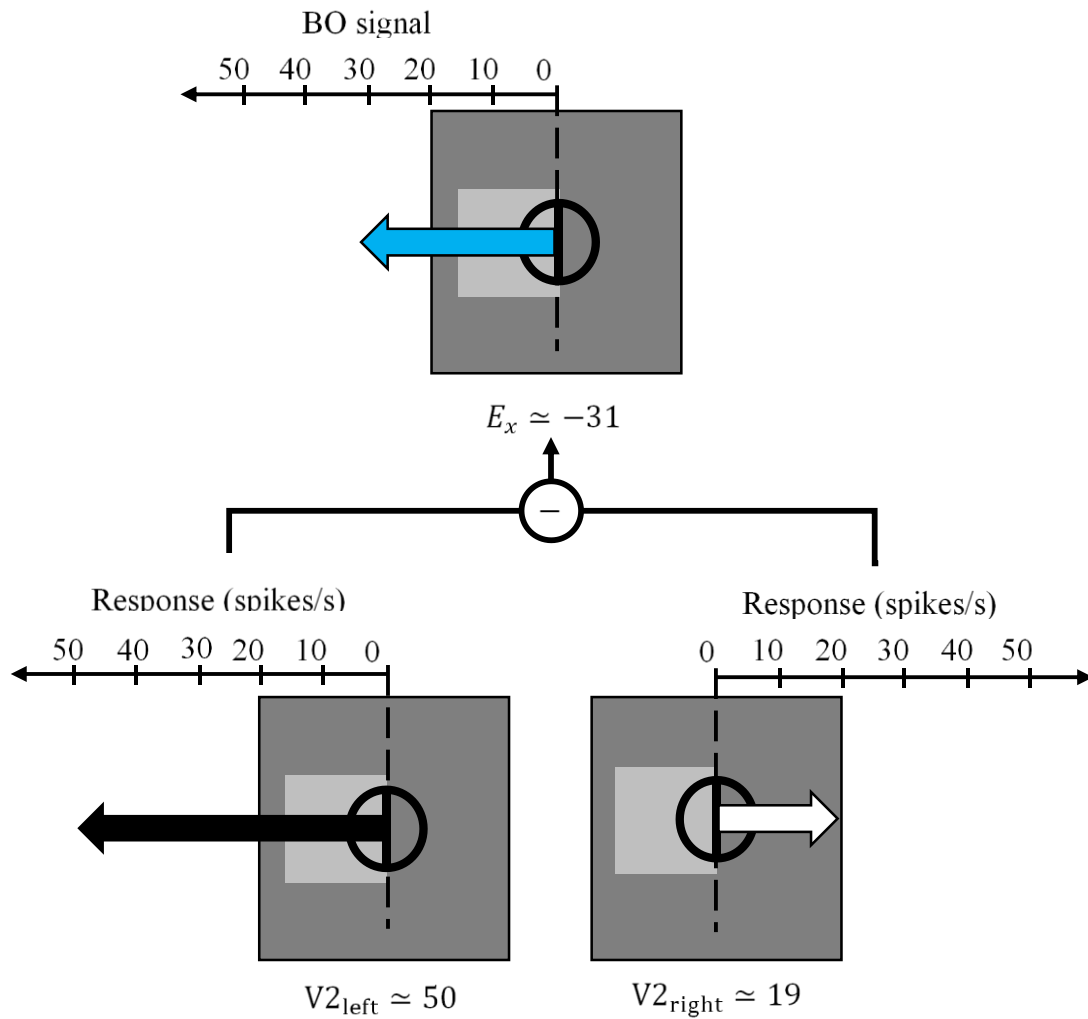
Figure 4.11 Data adapted from Zhou et al. (2000) is used to illustrate how to compare BO signals with V2 neuron responses.

A horizontal BO signal E_x can be obtained from the difference in response of a pair of V2 neurons with opposite object-side selectivity ($V2_{left} \simeq 50$ and $V2_{right} \simeq 19$). Neuron receptive fields are represented by circles, whereas arrows represent the neurons' object-side selectivity and response strength. The difference in response of these neurons is a negative BO signal in the horizontal direction $E_x = V2_{right} - V2_{left} \simeq -31$, signifying an object to the left of the receptive field.
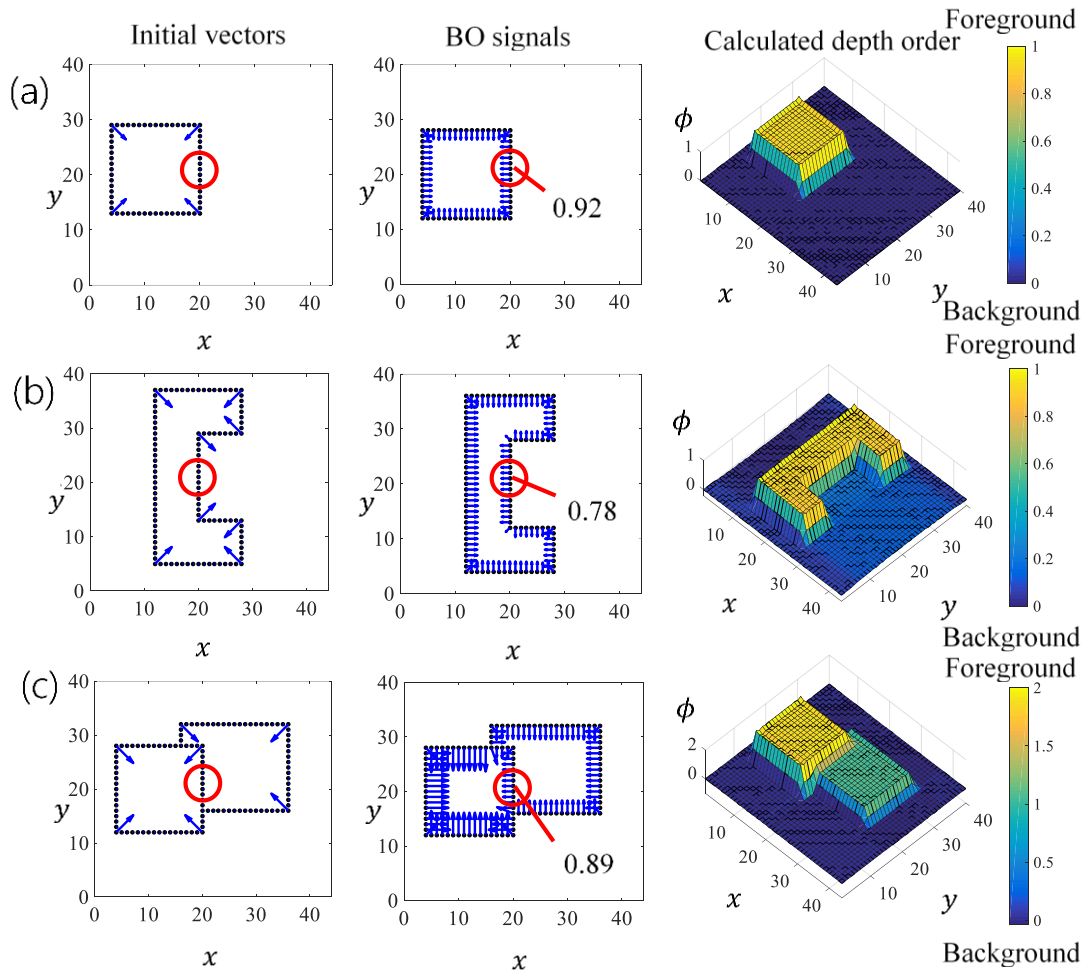
Figure 4.12 Initial vectors, calculated BO signals and depth order for figures similar to the stimuli used in physiological experiments for a V2 neuron with selectivity for an object to the left of its receptive field.

Respective depth order scalar field agrees with the perception of (a) a square, (b) a C-shape, and (c) an overlapping square. Red circles indicate the approximate location of the neuron's receptive field in respect to the figure. Calculated BO signals face to the left, agreeing qualitatively with the neuron's object-side selectivity. The absolute value for $\hat{x}$ component at the center of the red circle, $|E_x(20,21,2000)|$, is displayed on the respective figures (a. 0.92; b. 0.78; c. 0.89).
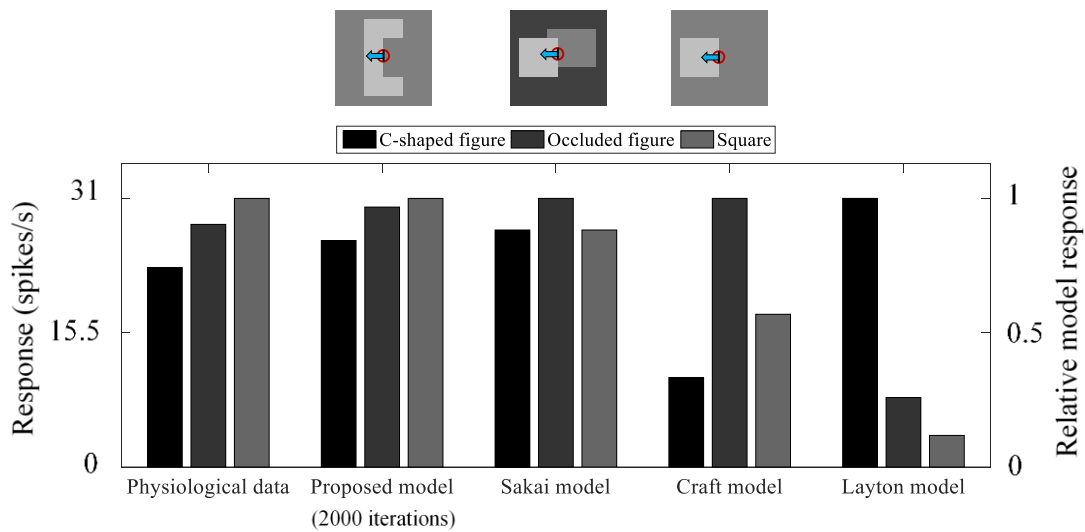
Figure 4.13 BO signals calculated from experimental data produced by Zhou et al. (2000) and relative model response from the proposed model in comparison with existing models.

Neuronal responses are shown for three benchmark stimuli (a C-shaped figure, an occluded figure, and a square). Responses to the stimuli are arranged in ascending order of response strength. Proposed model responses are based on absolute values for $\hat{x}$ component at the center of the red circle, $|E_x(20,21,2000)|$, in **Figure 4.12**. The red circle in the respective figures represent the neuron's receptive field, and the arrow represents the direction of the BO signal. The proposed model qualitatively reproduces the difference in response of the neuron to the three stimuli for $t = 2000$. In comparison, existing models are incapable of reproducing this phenomenon [17], [18], [21].
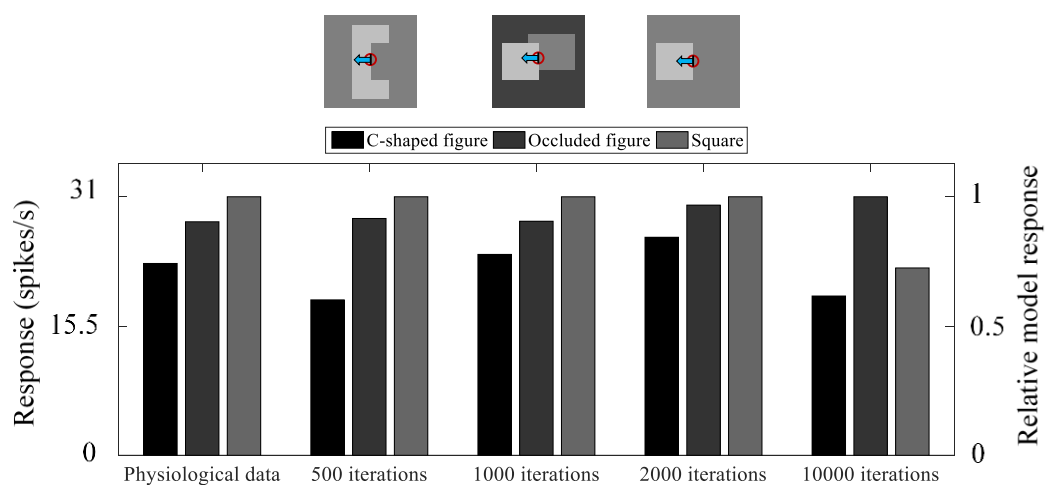
Figure 4.14 BO signals calculated from experimental data produced by Zhou et al. (2000) and relative model response for 500, 1000, 2000, and 10000 iterations.

In general, relative model responses converge to 1 as the number of iterations increase. However, the response of the model to the C-shaped figure ceases to increase, and the relative response to the occluded figure exceeds the relative response to the square (10000 iterations).

4.3.2 Transparent overlay

Some neurons in V2 which possess object-side selectivity also respond selectively to "transparent" occluding objects [29]. For example, the image in **Figure 4.15b** can be perceived as either a light-grey rectangle overlapping a dark-grey rectangle, or a dark-grey rectangle overlapping a light-grey rectangle. Both rectangles are perceived to be "transparent". Consider a neuron with a receptive field represented by the red circle, located at a vertical border. An occluding object is located on the left of the receptive field. This agrees with perception. **Figure 4.15a** is perceived as an opaque light-grey rectangle overlapping an opaque dark-grey rectangle. Similar to **Figure 4.15b**, the occluding object is located on the left-side of the receptive field. In contrast, the occluding object is located on the right-side of the receptive field in **Figure 4.15c**. This is because the image is perceived as four separate squares. The pattern and luminance in all three examples are the same. These three figures are similar to those used in physiological experiments by Qiu and von der Heydt (2007) [29].

Calculated BO signals after $t = 1000$ iterations and the enlarged region around $(20,17)$ is shown for the three figures in **Figure 4.15**. The direction of the circled vectors in the enlarged region for **Figure 4.15a** and **Figure 4.15b** point to the left. In contrast, the direction of the circled vector in **Figure 4.15c** faces the right. The modulation (direction) of these vectors agree with BO signals from actual physiological experiments by Qiu and von der Heydt (2007) [29]. The model is capable of qualitatively reproducing BO signals despite the absence of luminance information as an input.
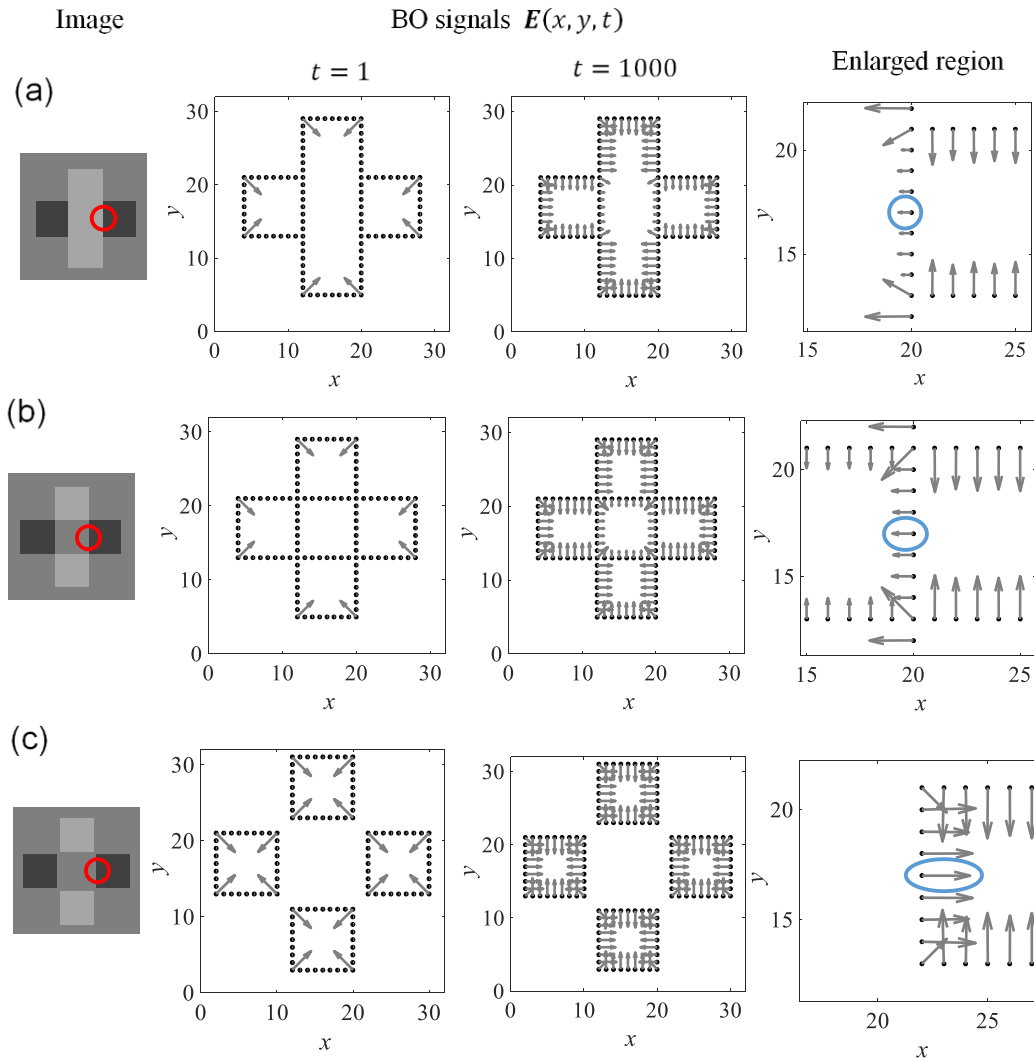
Figure 4.15 BO signals calculated for borders of (a) overlapping opaque rectangles, (b) overlapping transparent rectangle, and (c) four separated squares.

Only borders, void of luminance information, was used as model inputs. The direction of vectors at $(20,17)$ agree with BO signal modulation for a neuron with a receptive field represented by the red circles on the images [29].

4.3.3 Zero BO signal

Physiological experiments have demonstrated that for specific figures, the BO signal at a border between two objects can take a value of zero [30]. **Figure 4.16** shows a figure where the occluding object and the occluded object is unclear. The image can be perceived as two rectangles placed side-to-side. This figure is similar to the one used in the physiological experiment. Calculated BO signals (vectors) at the red circle, which represents the location of a receptive field, are zero. This agrees qualitatively with experimental data. Depth order for both objects are the same despite the existence of a border.
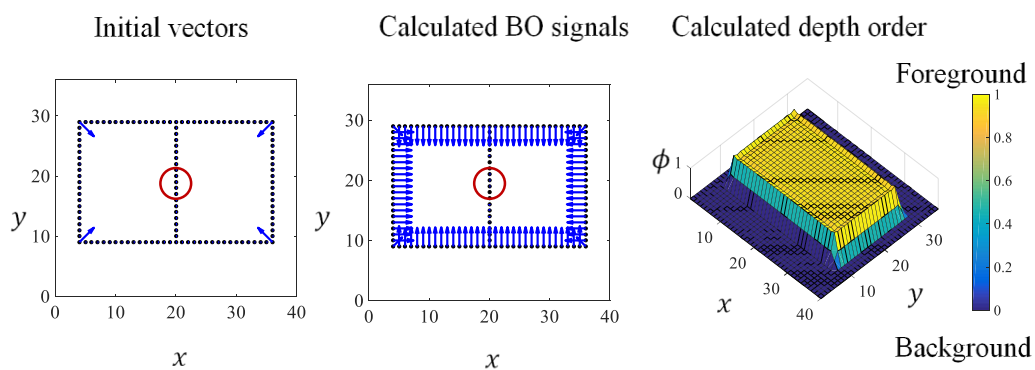
Figure 4.16 BO signals and depth order for a figure with a zero BO signal at the border.

Zero vectors are assigned along the central border. The regions on both sides have equally valued depth order. The neuron used in physiological experiments shares the same location for the receptive field represented by the red circle [30]. BO signals from experiments were found to be zero, as well.

4.4 Complex patterns

Several complex patterns which do not have corresponding data from physiological experiments were also tested to examine the robustness of the model towards shape and occlusion (Figure 4.17). These include (a) a complex Y-shaped figure, (b) an O-shaped figure, (c) C-shaped figure occluding a square, (d) interlocking L-shapes and (e) four overlapping cards. The evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time for the respective figures are shown in **Figure 4.18** to **Figure 4.22**.

Figure 4.17a shows a figure with several concave regions. Initial vectors point towards these concave regions in **Figure 4.18**. As time progresses, these vectors are corrected to point to the object. However, the model is incapable of reproducing the perception of an O-shaped given the initial vectors set to face the acute side of the L-junctions at the perimeter of the hole (**Figure 4.19**). **Figure 4.17b** shows a figure with both local concavity and occlusion. The model is capable of correcting vectors facing the locally concave region and resolve depth order for the two objects simultaneously (**Figure 4.20**).

Figure 4.17d shows an image perceivable as two interlocking L-shapes. **Figure 4.17e** is a similar figure. It can be perceived as four overlapping cards. As iteration t increases, calculated depth order converges at a flat surface; the depth order of the two (or four) objects carries the same value. However, it is interesting to note that at $t = 100$, depth order resembles perception.
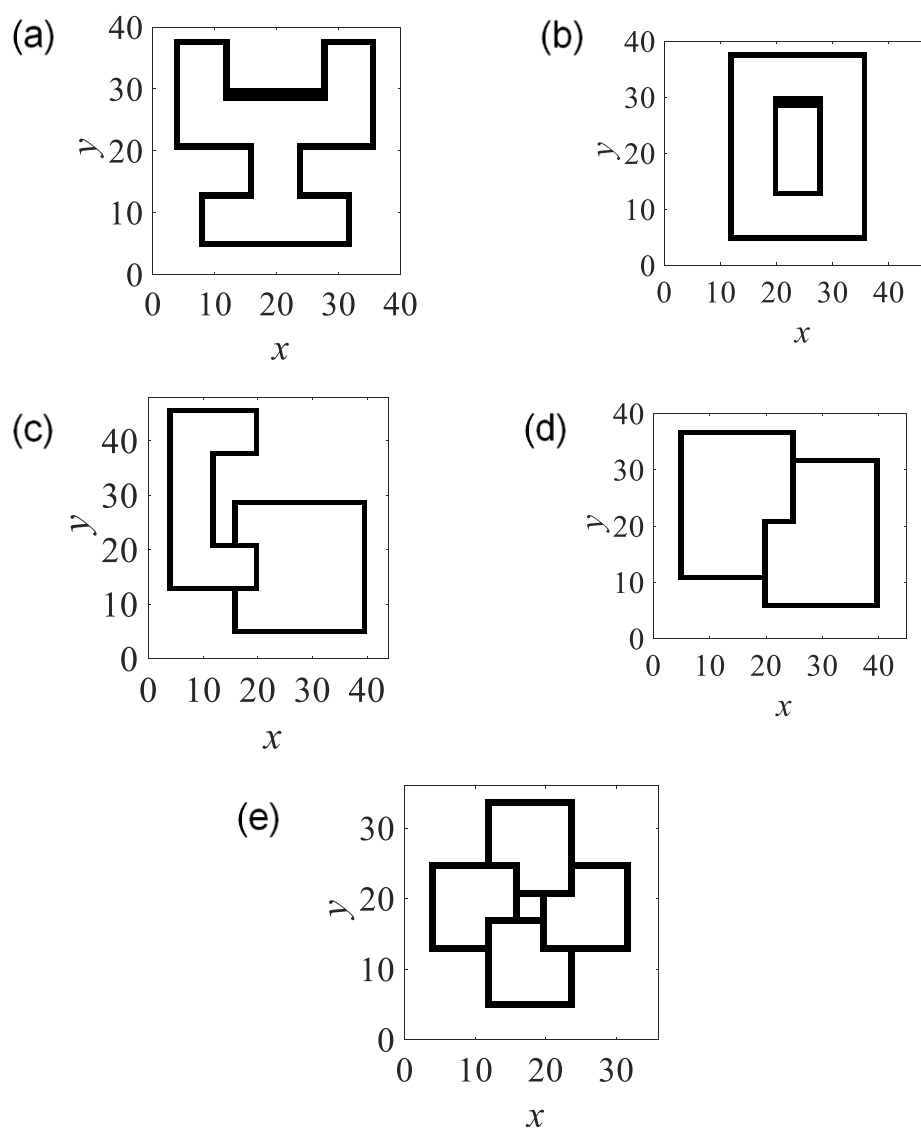
**Figure 4.17 Five figures with problems of shape, occlusion, or both.**

Simulations were conducted for (a) a complex Y-shaped figure, (b) an O-shaped figure, (c) a C-shaped figure occluding a square, (d) interlocking L-shapes and (e) four overlapping cards.
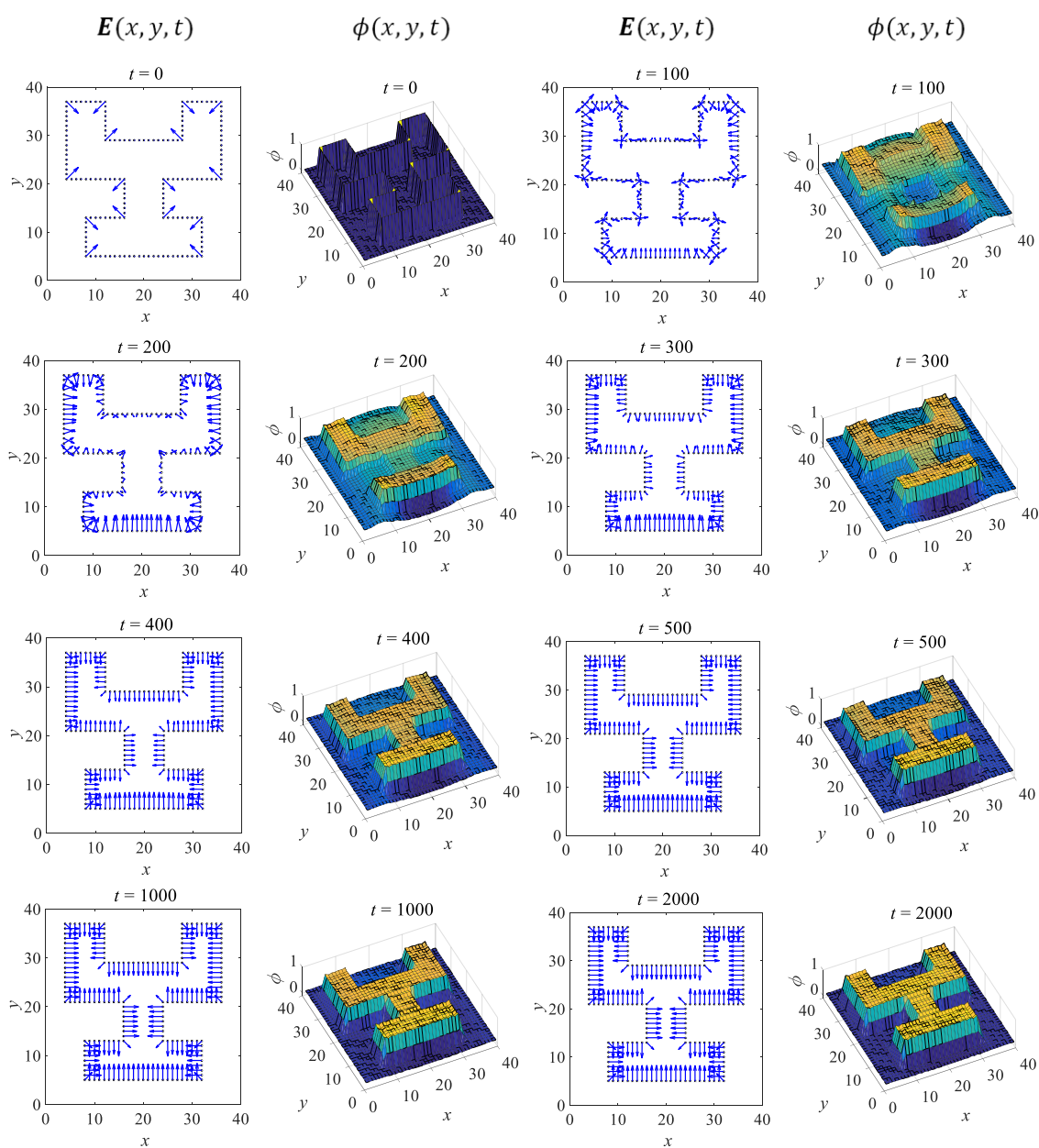
Figure 4.18 Evolution of BO signals and depth order over time for a figure perceivable as a Y-shaped object.
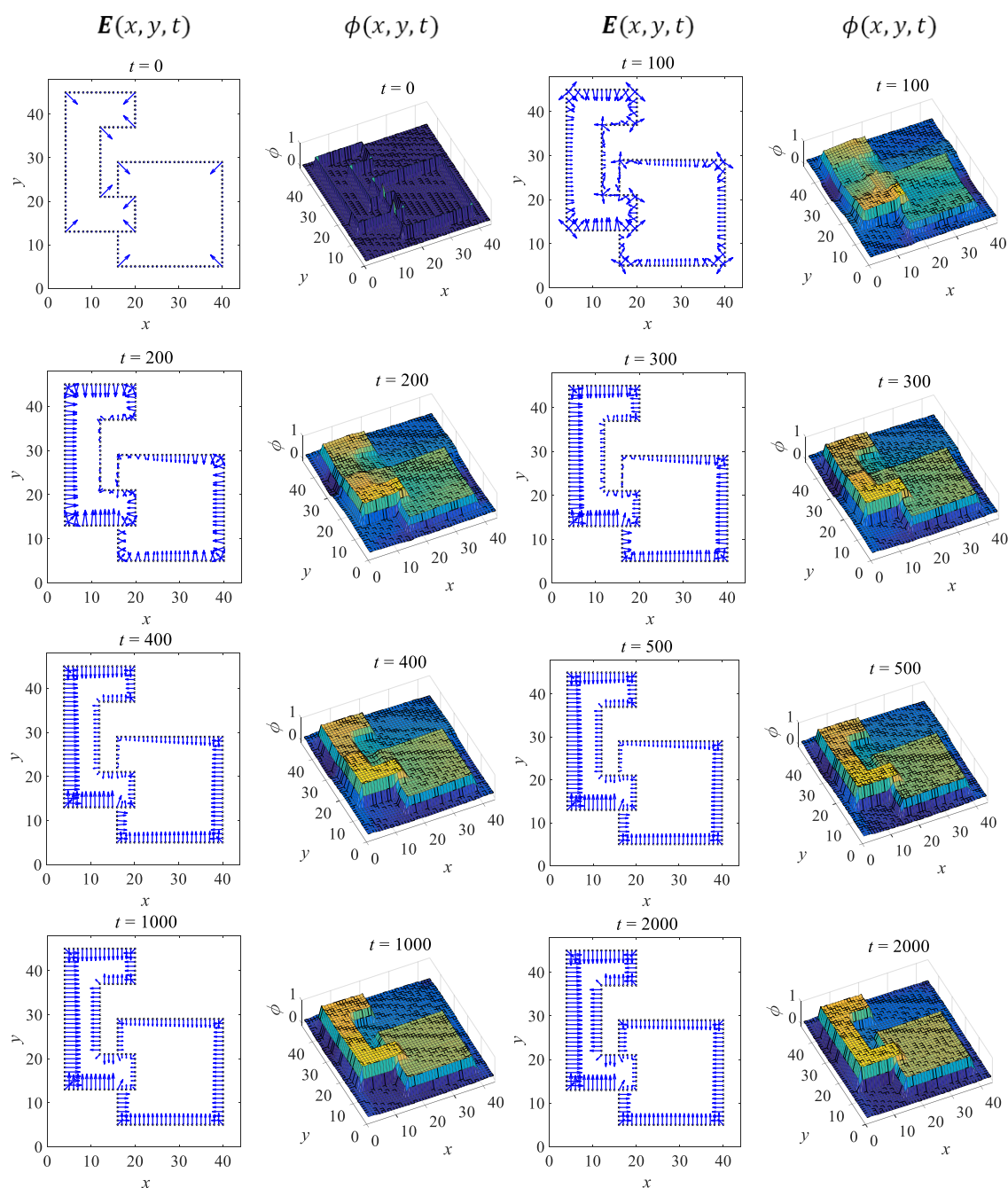
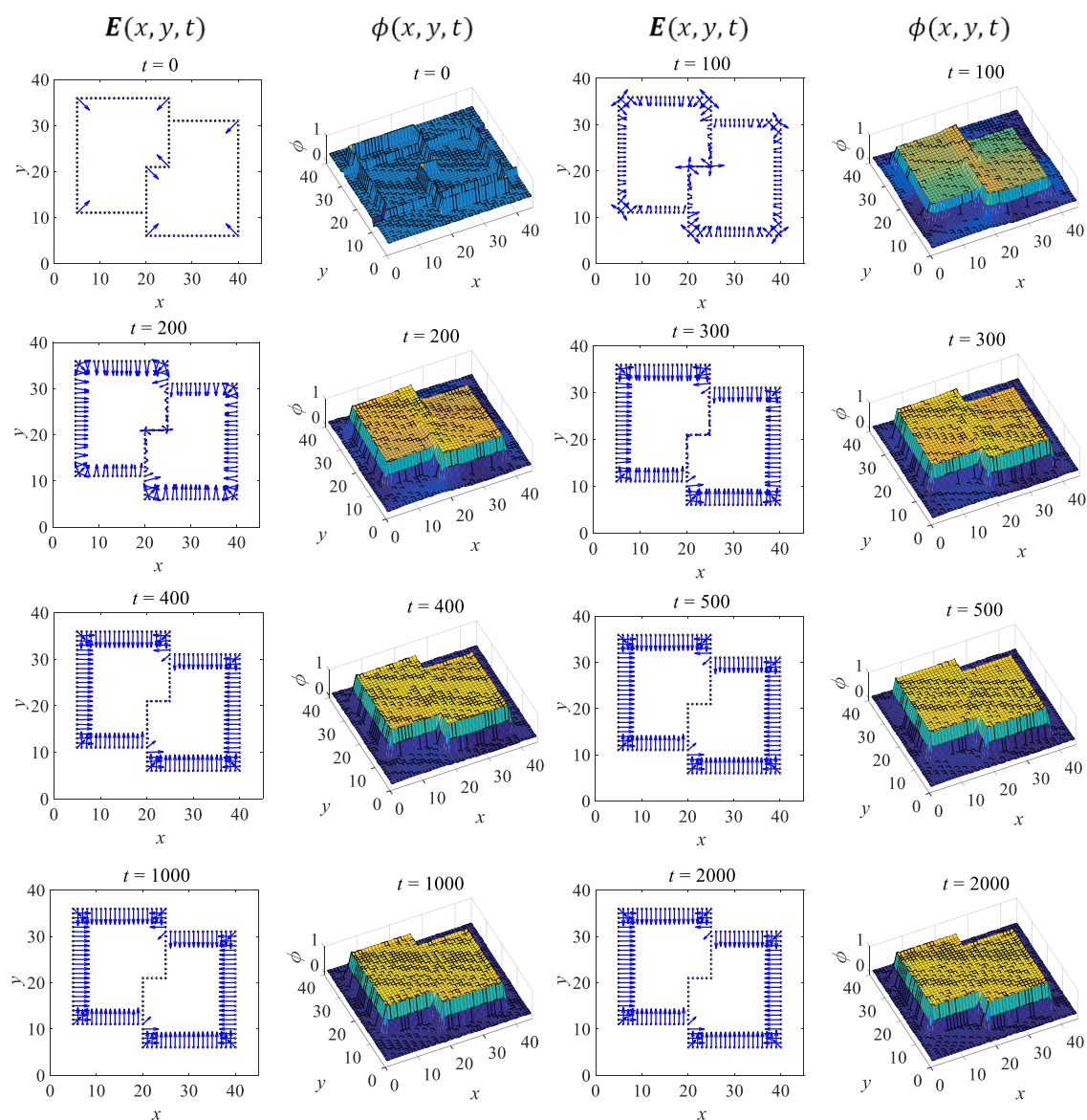Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 2000$.
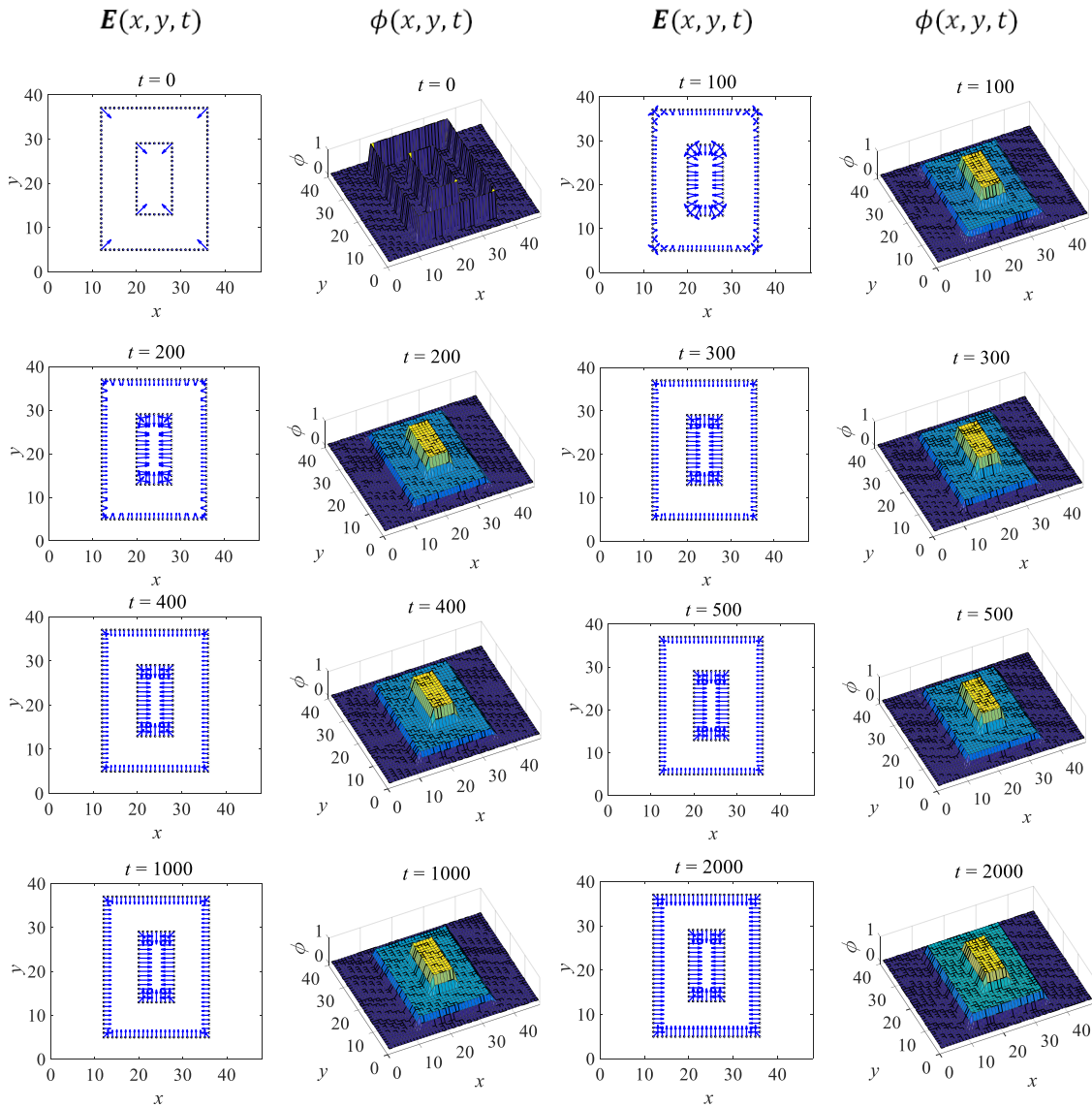
Figure 4.19 Evolution of BO signals and depth order over time for a figure perceivable as an O-shape.

Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 2000$. Model outputs do not agree with the perception of an O-shaped figure.
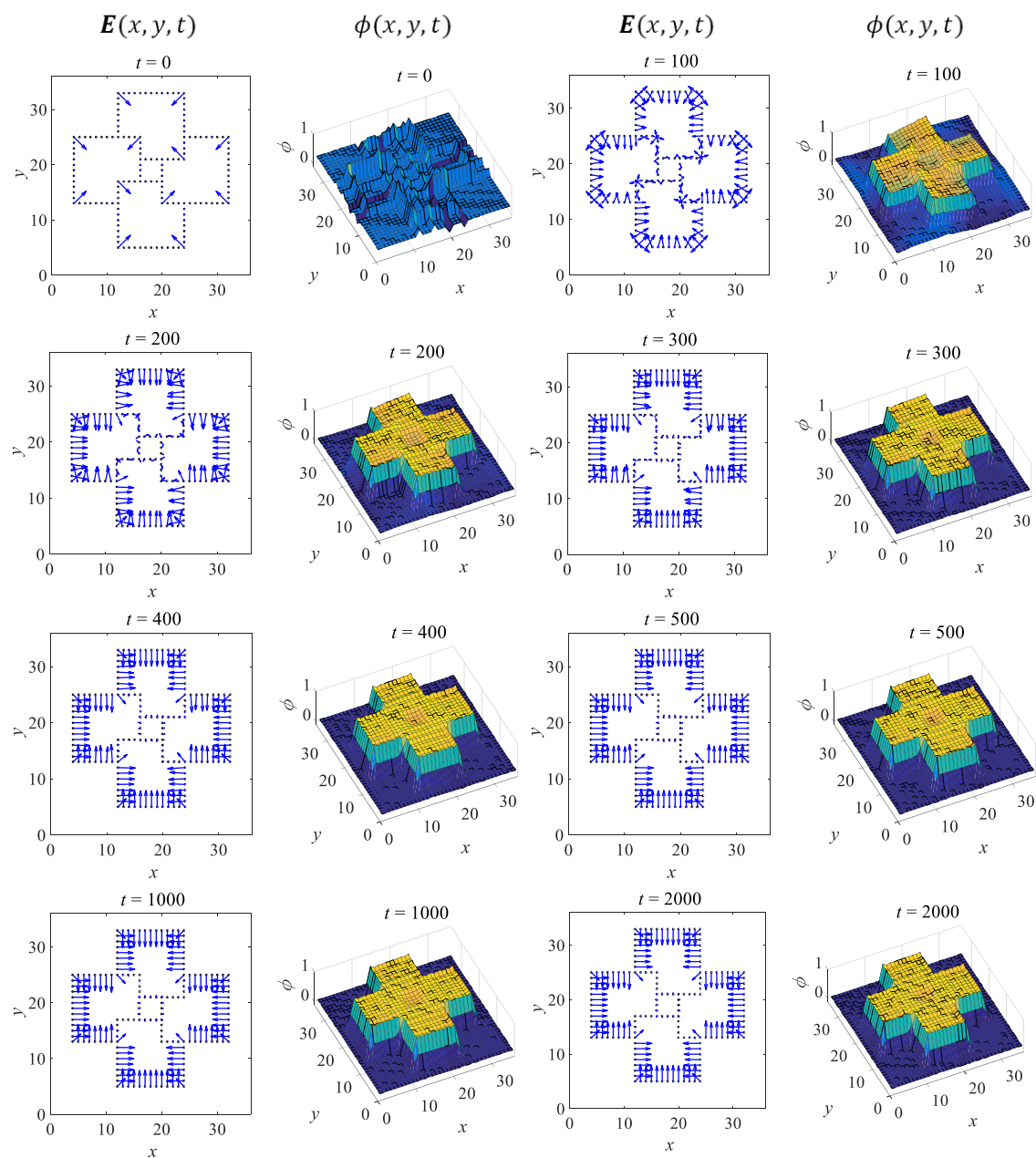
Figure 4.20 Evolution of BO signals and depth order over time for a figure perceivable as a C-shape occluding a square.

Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 2000$.

**Figure 4.21 Evolution of BO signals and depth order over time for a figure perceivable as a interlocking L-shapes.**

Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 2000$.

Figure 4.22 Evolution of BO signals and depth order over time for a figure perceivable as overlapping cards.

Evolution of $E(x, y, t)$ and $\phi(x, y, t)$ over time $t = 0 \sim 2000$.

4.5 Evaluation of results

Simulation results demonstrate that the model was capable of assigning BO correctly for figures with occlusion without the use of a specific T-junction detector. Although a T-junctions may serve as an important occlusion cue in some cases, there are also situations where T-junctions are not formed by an intersection of three surfaces of differing depth order. **Figure 4.23** shows four ways in which a local T-junction can be interpreted as. For all four cases, the rotation at the T-junction is zero. In particular, the depth order in **Figure 4.23c-d** can be seen at T-junctions in figures containing ambiguous borders (**Figure 4.16**). The proposed model uses global contextual information to determine a unique solution for various perceptual problems involving depth order.

The proposed model simultaneously produced BO signals and depth order that match perception. Simulation results demonstrate that the model is highly dependent on initial vectors. To reproduce data from physiological experiments, initial vectors were provided in a deterministic manner through L-junction detectors. Spatial propagation of contextual information using the update rule automatically determines BO signals regardless of local concavity. Calculating curvature might serve as a replacement to L-junction detectors. This would allow simulations of figures with more complex shapes to be carried out.

Although initial vectors based on L-junctions alone are useful, they are insufficient to reproduce depth order perception for figures such as an O-shaped figure. Reversing the vectors at the center, as shown in **Figure 4.24** demonstrates the importance of initial vectors. A more global method of initial vector determination is required to reproduce human perception.
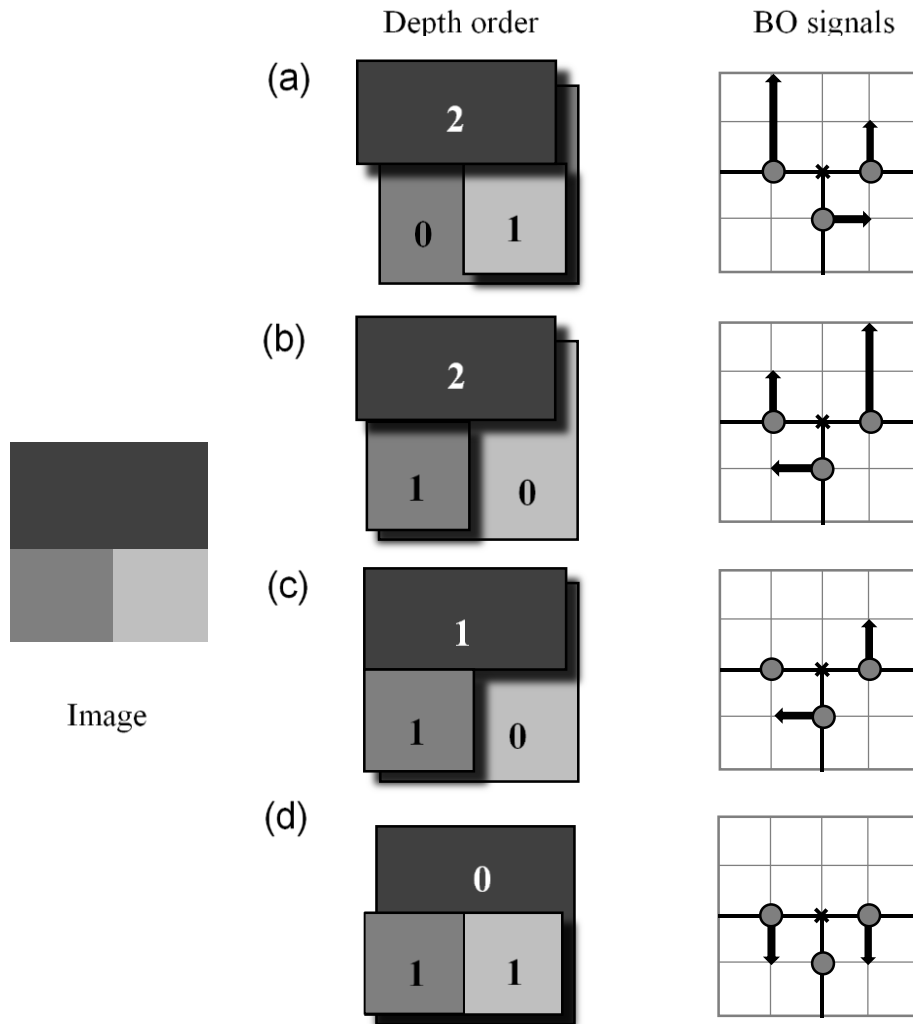
Figure 4.23 Image of a T-junction, four possible perceptions of depth order and their corresponding BO signals.

Numbers on the surfaces represent depth order values. Short vectors are of length one, while long vectors are of length two. (a) and (b) are T-junctions where the surface to the top of the T-junction has the highest depth order value of the three. However, adjoined surfaces of equal depth order value, such as in (c) and (d) may also form T-junctions. Notably, the local curl at the cross-mark for all four examples is zero.
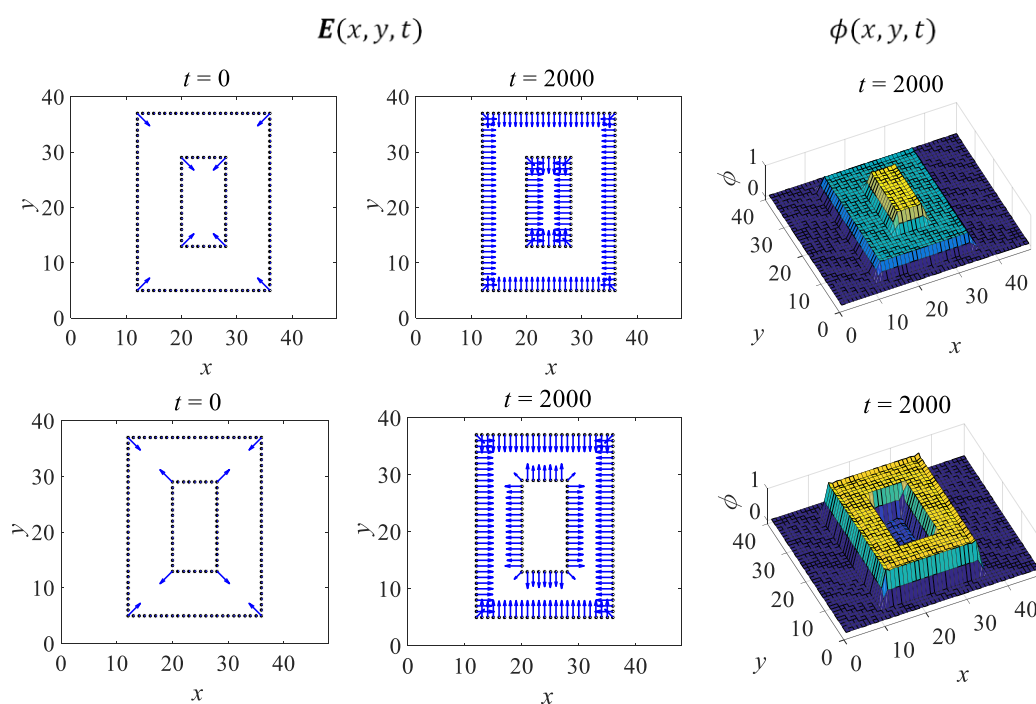
Figure 4.24 Reversing initial vectors at the inner border produces depth order consistent with perception of an O-shaped figure.

L-junctions are vital cues to occluded object perception, but are insufficient to reproduce the perception of an O-shaped figure. For example, reversing initial vectors around the border perceived as the hole reproduces this percept.

5   Deductions and model comparisons

5.1   Mathematical foundation for present models

5.1.1 Spatial distribution of intra-cortical connections in the model by Li

Deductions can be made about the spatial distribution of intra-cortical connections from the update rule in **Eq. (3.4)**. For example, the $\hat{x}$ component of the update rule $\frac{\partial}{\partial t} E_x(x, y, t)$ reveals neural connections for a neuron at (x, y) with either a right object-side ($E_x(x, y, t) > 0$), or a left object-side (($E_x(x, y, t) < 0$) selectivity. A closer look at the right hand side of this update rule shows that this neuron at (x, y) has connections to other neurons with either left or right object-side selectivity $\frac{\partial^2}{\partial y^2} E_x(x, y, t)$ and neurons either up or down object-side selectivity $-\frac{\partial^2}{\partial x \partial y} E_y(x, y, t)$. The differentiation operators dictate the spatial distribution of these connections. **Figure 5.1** illustrates possible intra-cortical connection for a neuron with a right object-side selectivity, with a receptive field represented by a red oval on the image. As seen in **Figure 5.1a**, connections with left or right object-side selective neurons E_x is equivalent to diffusion in the horizontal y-direction. **Figure 5.1b**, illustrates connections with left or right object-side selective neurons E_y positioned diagonally.

The structure of these connections is similar to that of the model by Li (2000). A method to qualitatively compare the two is proposed. Spatial filters for E_x and E_y in **Eq. (3.8)** can be expressed as weights $w_x(x, y) \overset{\text{def}}{=} \frac{\partial^2}{\partial y^2} G_\sigma(-x-, y)$ and $w_y(x, y) \overset{\text{def}}{=} \frac{-\partial^2}{\partial xy} G_\sigma(-x, -y)$. Discrete weights for a neuron positioned at $(0,0)$ with a left or right object side selectivity can thus be expressed as

$$\frac{\partial}{\partial t} E_x(0,0,t) = \sum_i \sum_j w_x(i,j) E_x(i,j,t) + \sum_i \sum_j w_y(i,j) E_y(i,j,t). \qquad (5.1)$$

Weights can then be expressed as vectors $\boldsymbol{w} = (w_x, w_y)^{\mathrm{T}}$ of magnitude $\|\boldsymbol{w}\| = \sqrt{w_x{}^2 + w_y{}^2}$. Weights for $E_x(x,y,t)$ are visualized in **Figure 5.2a** using weights consisting of Gaussian derivative kernels with standard deviation $\sigma = 1$. Each vector illustrates the direction of $\boldsymbol{w}$, whereas pixel intensity represents vector magnitude $\|\boldsymbol{w}\|$. Vector magnitudes were normalized to 1.

Horizontal connections proposed by Li (2000) were visualized in a similar manner using weights for a right object-side selective neurons. In particular, the weights which describe connections to four types of other object-side selective neurons were used (left, right, up, down). These weights are visualized in **Figure 5.2b**. The nature of deduced intra-cortical connections from the proposed model and those from the model by Li (2000) were found to be qualitatively similar. However, the results cannot be treated as ground truths since the nature of horizontal connections in area V2 are currently unknown.

$$(a) \frac{\partial^2}{\partial y^2} E_x \qquad\qquad (b) -\frac{\partial^2}{\partial x \partial y} E_y$$
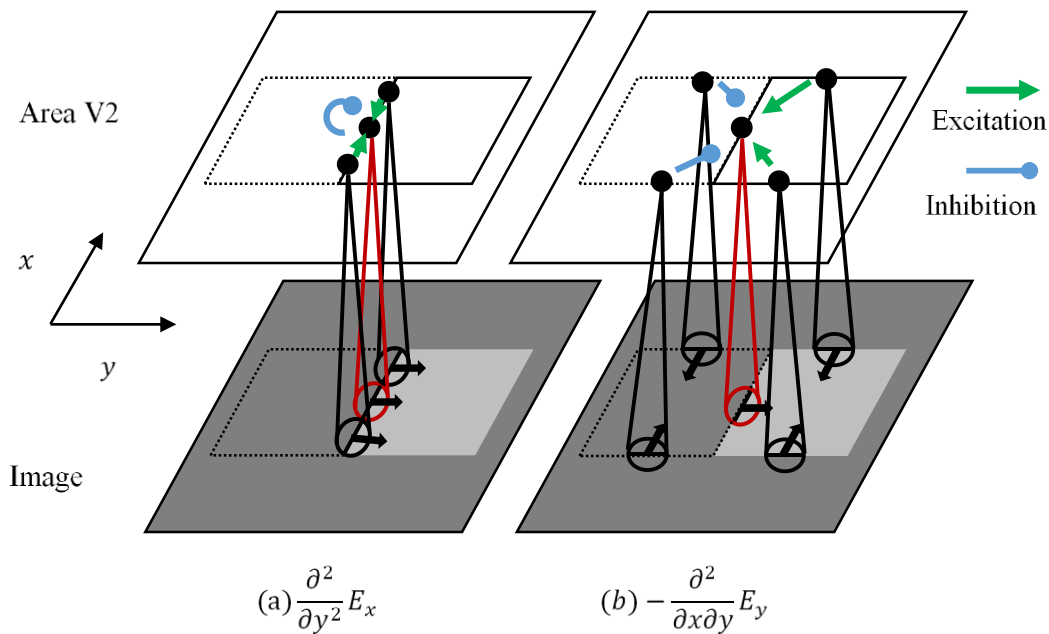
Figure 5.1 Plausible neural network for a neuron that responds selectively to a vertical border and an object to the right of its receptive field.

Neuron receptive fields are represented by ovals; arrows represent their object-side selectivity. The neuron focused on has a red oval as a receptive field. Black circles in area V2 represent neurons with receptive fields on the image. Neurons are either connected by excitatory or inhibitive neural connections, according to their relative location and object-side selectivity. Connections with (a) right (or left) object side selective neurons were derived from the E_x component; (b) up (or down) object side selective neurons were derived from the E_y component in in **Eq. (3.4)**.
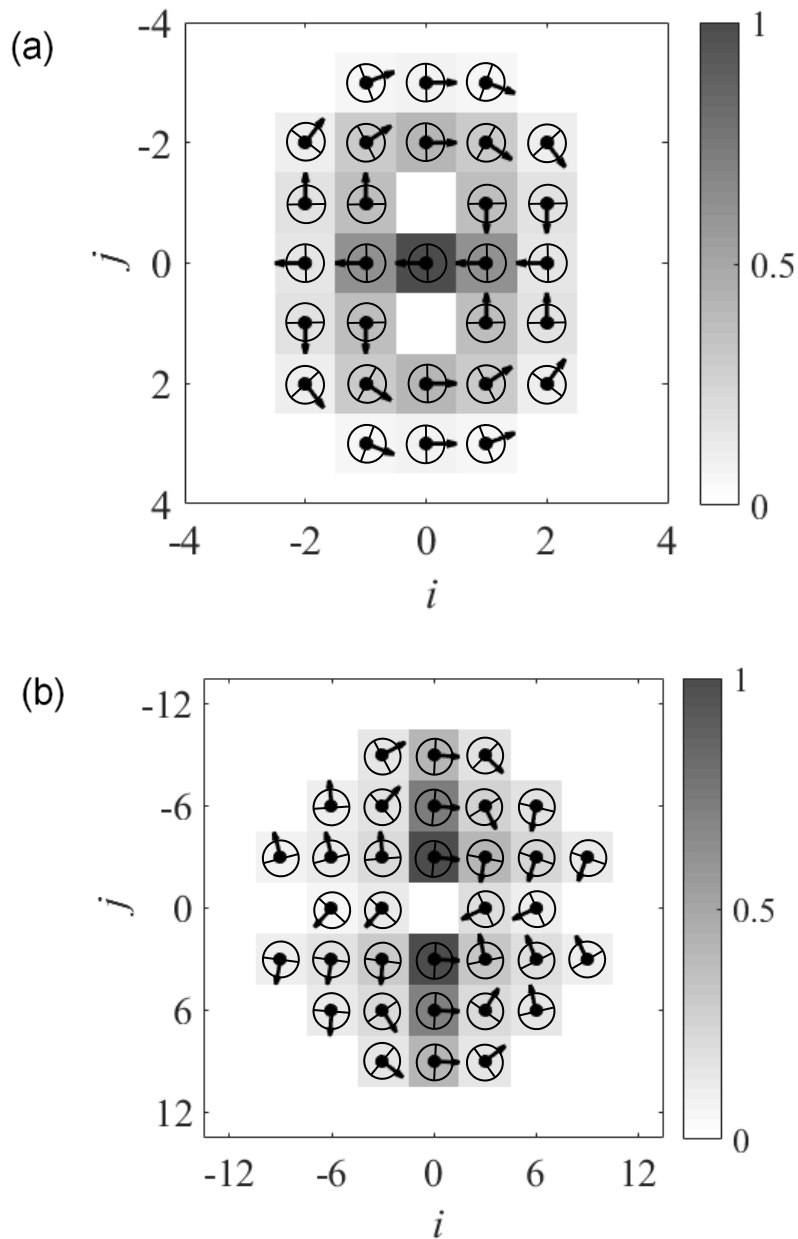
Figure 5.2 Charactertistics of neural weights w for a neuron at spatial position $(0,0)$ with a right object-side selectivity, visualized using vectors.

Vector magnitudes $\|w\|$ are represented by pixel intensity. The characteristics of neural weights for (a) the proposed model are qualitatively similar to those of (b) the model by Li (2000).

5.1.2 Annular receptive fields in the model by Craft et al.

Craft et al. (2007) proposed that the responses of object-side selective neurons in V2 are modulated by hypothetical "grouping cells" with annular-shaped receptive fields. They base their assertion on findings in psychological experiments which suggested the existence of mechanisms in the visual cortex to integrate concentric circular patterns, as well as neurophysiological findings on the response of neurons in area V4 to concentric gratings [31], [32]. Weights of annular receptive fields of different radii were created in an ad-hoc manner; spatially discrete circles were convolved with 2D Gaussian filters. This study proposes that these ad-hoc weights can be reproduced by generating vector-to-scalar integration kernels at different scales. Details will be outlined next.

As proposed in **Eq. (3.1)**, integrating BO signals should produce depth order. Although **Eq. (3.9)** outlines a way to estimate depth order from line integration with horizontal and vertical paths, no evidence exists that this calculation is feasible by the brain, especially since it requires calculation across the entire image. The author discovered that the sum of normalized one-dimensional (1D) Gaussian first derivatives is qualitatively similar to a Heaviside step function. The details how this was achieved will be explained next.

First, a 1D Gaussian with standard deviation σ is defined as

$$g_\sigma(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{5.2}$$

This was combined with the octave concept, where standard deviation was replaced with powers of two 2^n,

$$g_n(x) = \exp\left(-\frac{x^2}{2^{2n+1}}\right). \tag{5.3}$$

An example of the normalized sum of the first derivative

$$\sum_n 2^n \left(\frac{d}{dx} g_n(x) \right) = -\sum_n x \exp\left(-\frac{x^2}{2^{2n+1}} \right) \tag{5.4}$$

for integers between -10 and 10 is shown at the bottom of **Figure 5.3**. Notably, for a narrow spatial range, it is qualitatively similar to a Heaviside step function.

Next, a quantitative comparison will be made by differentiating the two and comparing the result in Fourier space. The differentiation of the negative Heaviside function $H(x)$ is a negative delta function $-\delta(x)$. Fourier transformation on a function $f(x)$ is

$$\mathcal{F}(\omega) \overset{\text{def}}{=} \int_{-\infty}^{\infty} f(x) \exp(-j\omega x)\, dx. \tag{5.5}$$

Therein, j is the imaginary unit, where $j^2 = -1$. The Fourier transformation of negative delta function $-\delta(x)$ is a constant $-1/\sqrt{2\pi}$. Therefore, if it can be demonstrated that the differentiation of the normalized sum of the first Gaussian derivative $\sum_n 2^n \left(\frac{d}{dx} g_n(x) \right)$ seen in Fourier space is a constant, one can say that it is qualitatively similar to a Heaviside function $H(x)$. The Fourier transformation of the differentiation of $g_n(x)$ was calculated using Mathematica,

$$-2^{\frac{1}{2}+2n} \exp(-2^{-1+2n}) \sqrt{\pi}\omega^2. \tag{5.6}$$

Integration of n between $-\infty$ and ∞ of **Eq. (5.6)** resulted in a constant,

$$\int_{-\infty}^{\infty} -2^{\frac{1}{2}+2n} \exp(-2^{-1+2n}) \sqrt{\pi}\omega^2\, dn = \frac{\sqrt{2\pi}}{\log 2}, \tag{5.7}$$

demonstrating the quantitative similarity of the two functions.

This theory can be expanded to a 2D filter that calculates line integration through paths of all orientations passing through (x, y). This would be similar to integrating vectors in an annular direction. An estimate of depth order $\phi_{\text{est}}(x, y, t)$ can be calculated by convolution of the proposed filter with the $\hat{x}$ and $\hat{y}$ components,

$$\phi_{\text{est}}(x, y, t) = 2^n \sum_n \left(\frac{\partial}{\partial x} G_n(x, y) \right.$$

$$\left. * E_x(x, y, t) \right) + 2^n \sum_n \left(\frac{\partial}{\partial y} G_n(x, y) * E_y(x, y, t) \right)$$

(5.8)

Therein, $G_n(x, y)$ is the normalized 2D Gaussian function with its standard derivative replaced with powers of two 2^n,

$$G_n(x, y) \stackrel{\text{def}}{=} \frac{1}{2^{n+1}\pi} \exp(-\frac{x^2 + y^2}{2^{n+1}})$$

(5.9)

Comparisons of this method of calculating depth order with the existing method is shown in **Figure 5.4a** using values of n from -1 to 5 at 0.1 intervals. Values were scaled to a maximum of 1. The proposed method produces smoother depth order surfaces compared to the existing method. The respective filters can be expressed as weights $w_x(x, y; n) \stackrel{\text{def}}{=} 2^n \frac{\partial}{\partial x} G_n(-x, -y)$ and $w_y(x, y; n) \stackrel{\text{def}}{=} 2^n \frac{\partial}{\partial y} G_n(-x, -y)$. Discrete weights at $(0,0)$ is thus rewritten as

$$\phi_{est}(0,0) = \sum_i \sum_j \sum_n 2^n \left(w_x(i, j; n)E_x(i, j) + w_y(i, j; n)E_y(i, j) \right)$$

(5.10)

The magnitude of these weights $\|w(i, j; n)\| = \sqrt{w_x(i, j; n)^2 + w_y(i, j; n)^2}$ at various n was found to resemble the annular receptive fields proposed by Craft et al. (2007). **Figure 5.4b** shows individual filters for $n = 1,2,3$. Notably, parameters for annular receptive fields proposed by Craft et al. (2007) were not based on either a specific computational theory or physiological data. A qualitative comparison of the connection weights will be made next.

Craft et al. (2007) start by defining a discrete approximation to a circle of radius r centered at (x_0, y_0),

$$I_r(x,y) = \begin{cases} 1 & \text{if } \left|\sqrt{(x-x_0)^2 + (y-y_0)^2} - r\right| \le 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

The annular weight $I_r(x,y)$ is then normalized so that the total weight adds up to 1, regardless of radius r,

$$\hat{I}_r(x,y) = \frac{I_r(x,y)}{\sum_i \sum_j I_r(i,j)}. \tag{5.12}$$

Connection weights $K_r(x,y)$ was obtained by blurring $\hat{I}_r(x,y)$ by means of convolution,

$$K_r(x,y) = \hat{I}_r(x,y) * \hat{h}(x,y) \tag{5.13}$$

where $\hat{h}(x,y)$ is a discrete, 2D normalized Gaussian function of standard deviation σ,

$$\hat{h}(x,y) = \frac{\exp(-(x^2 + y^2)/2\sigma^2)}{\sum_x \sum_y \exp(-(x^2 + y^2)/2\sigma^2)}. \tag{5.14}$$

Radius r is proportionate to standard deviation σ where $\sigma = 2.5r$. Examples of annular receptive fields of different r and σ, as well as receptive connection weights for all the annular receptive fields used is shown in **Figure 5.5a-b**. Connection weights are shown for $x > 0$ of the $y = 0$ cross-section of "grouping cell" receptive fields centered at spatial position $(0,0)$.

As demonstrated in **Figure 5.4**, individual integration kernels required to reconstruct a scalar surface from a vector field resembles the annular receptive fields proposed by Craft et al. (2007), shown in **Figure 5.5a**.

$$w_{\text{study}}(x) = \frac{x}{2^{2n+1}\pi} \exp\left(-\frac{x^2}{2^{n+1}}\right). \tag{5.15}$$

The above equation was derived from **Eq.(5.10)** by substituting $y = 0$ and calculating the Gaussian derivatives expressed in terms of x and n. The connection weights proposed by Craft with the integration kernels in this study at four scales ($n = 1,2,3,4$) are qualitatively similar (**Figure 5.5b**).

(a) $\quad 2^n \dfrac{d}{dx} g_n(x)$

- n=-2
- n=-1
- n=0
- n=1
- n=2

(b) $\quad 1 - H(x)$

$$\dfrac{d}{dx} \implies -\delta(x) \implies \xrightarrow{\mathcal{F}} -\dfrac{1}{\sqrt{2\pi}}$$

$$\sum_n 2^n \left(\dfrac{d}{dx} g_n(x) \right)$$

$$\{n \mid n \in \mathbb{Z}, -10 \le n \le 10\}$$

$$\dfrac{d}{dx} \implies \boxed{???} \xrightarrow{\mathcal{F}} \boxed{???}$$
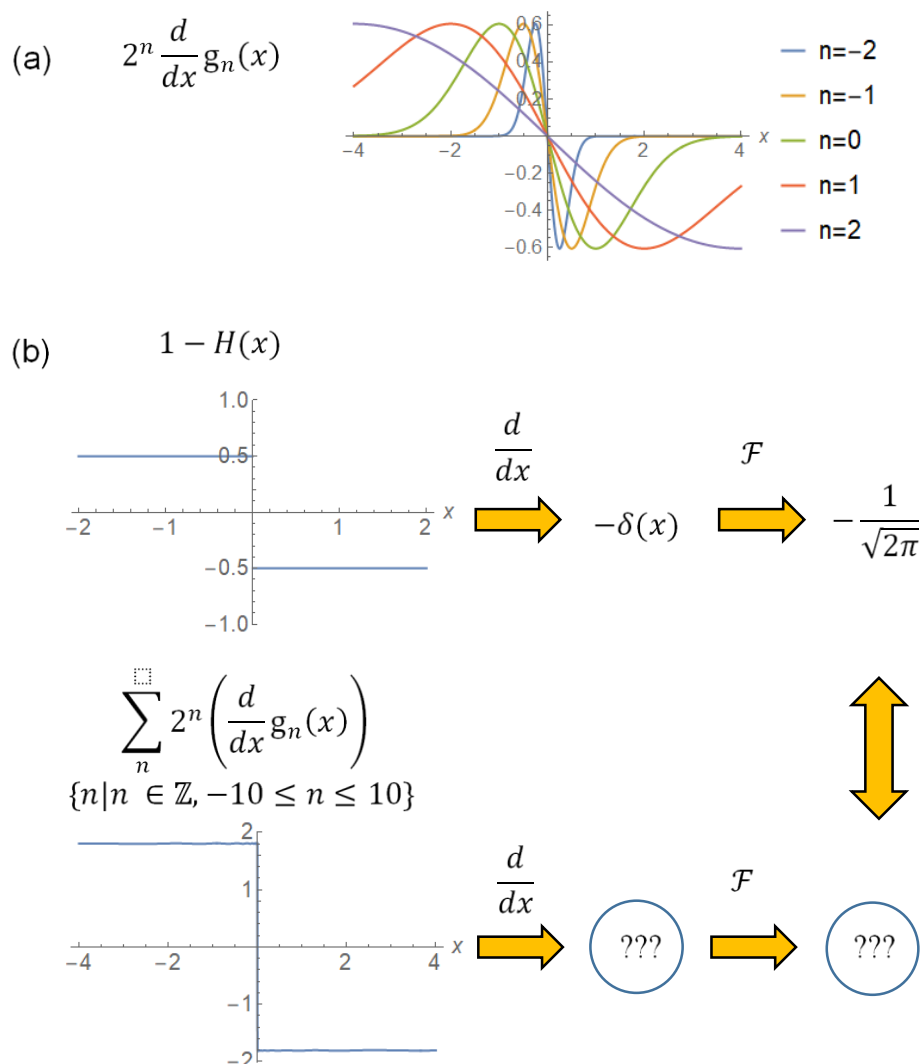
Figure 5.3 A schematic diagram showing the qualitative similarity between the sum of normalized 1D Gaussian first derivatives and a Heaviside function.

(a) 1D Gaussian first derivatives of various standard deviations 2^n. (b) Strategy to quantitatively compare both methods by differentiating and applying the Fourier transformation.
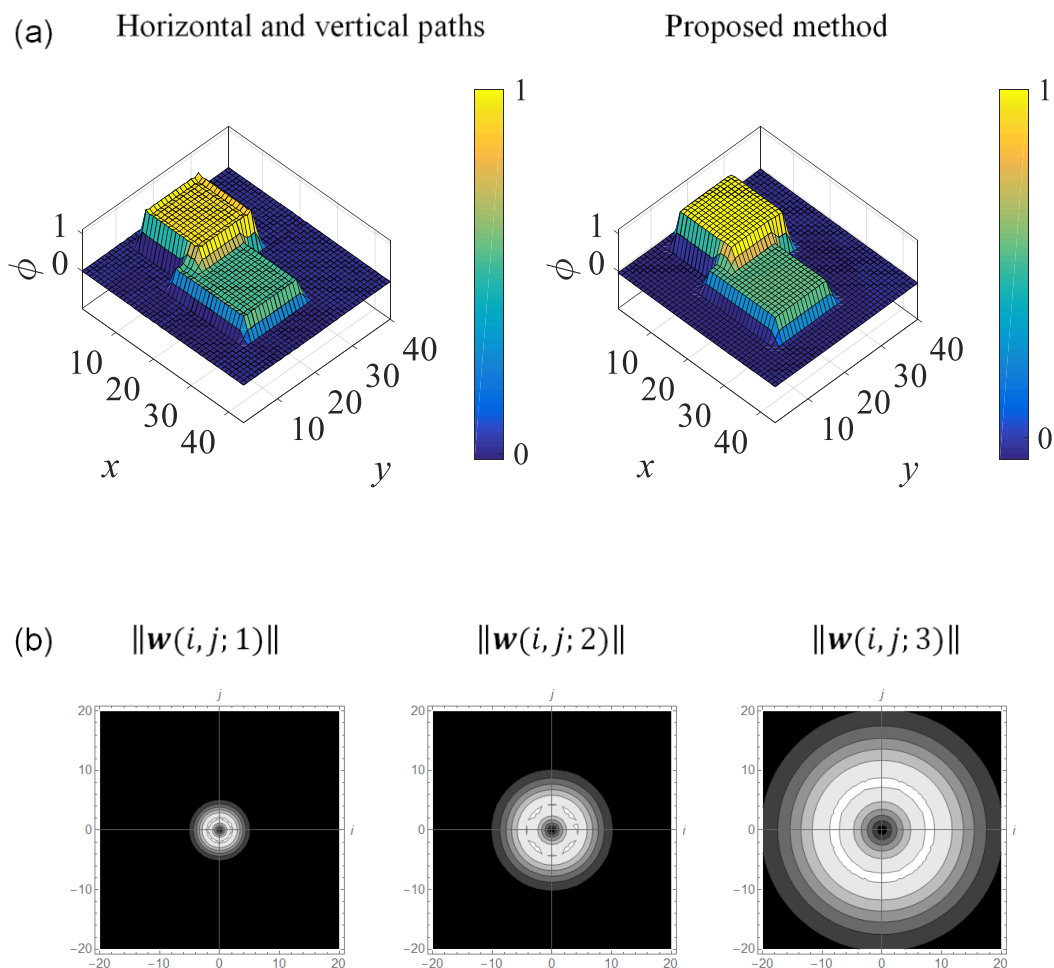
Figure 5.4 Analysis of the proposed line integration method.

(a) Comparison of integration results using horizontal and vertical paths with integration results using paths from all directions implemented using annular weights. (b) Magnitude of the individual weights resemble receptive fields of "grouping cells" proposed by Craft et al. (2007) in their model.
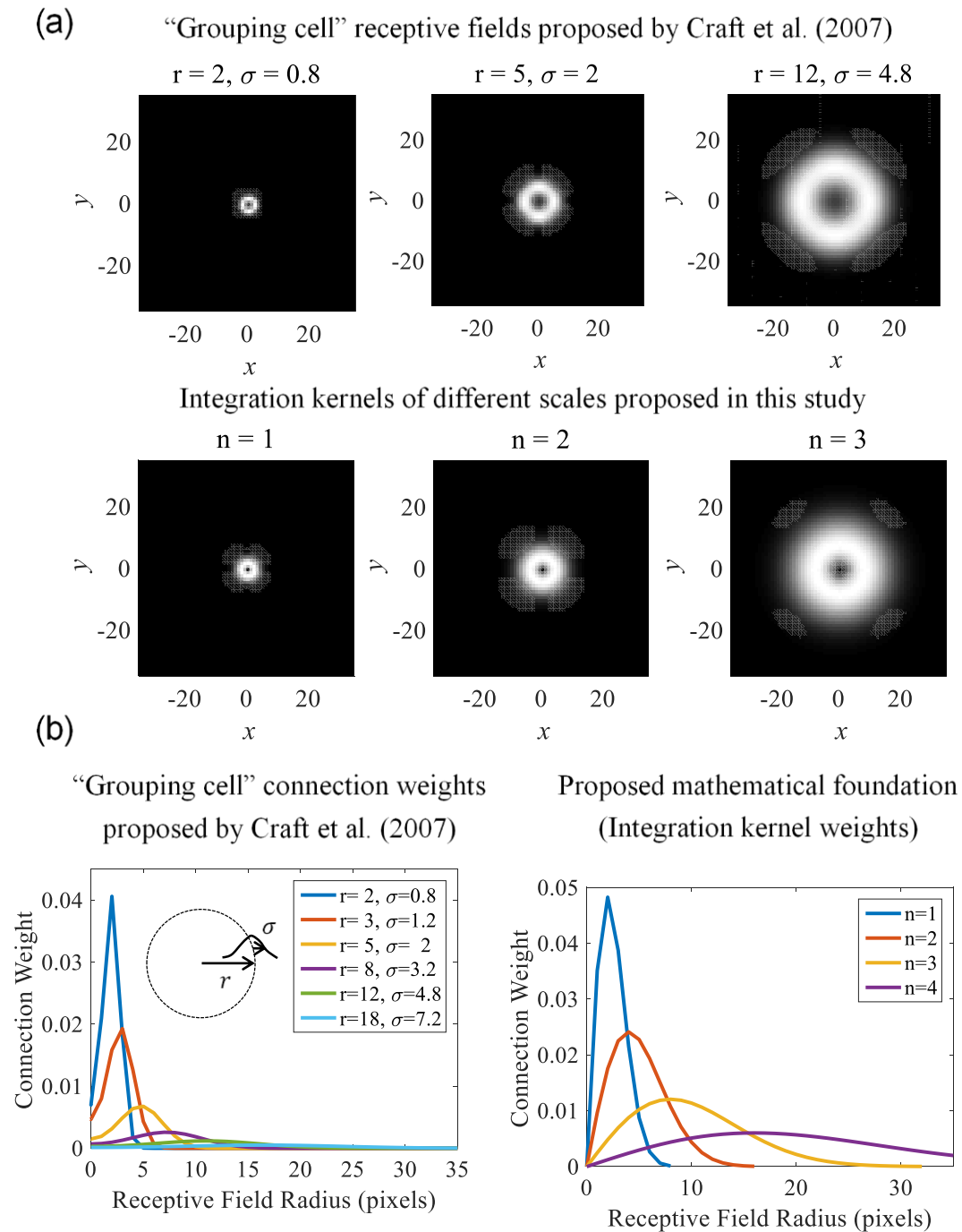
Figure 5.5 Receptive fields of "grouping cells" proposed by Craft et al. (2007) compared to integration kernel weights proposed in this study.

Integration kernels are qualitatively similar to the annular-shaped weights used to model receptive fields of "grouping cells". This suggests that "grouping cells" may in fact be "integration cells", involved in producing scalar surfaces from BO signals (vector fields).

5.2 Computational clarity

The proposed model is based on the computational theory of minimizing curl. This puts it above other models in term of computational clarity. It consists of simple Gaussian derivative filters, and only one parameter dictating kernel size: standard deviation σ. In comparison, the model by Li (2000) requires 23 free parameters that need to be adjusted accordingly to express intra-cortical connections. In addition, it also may provide clarity to the feedforward model by Sakai et. al (2012), which utilizes randomly generated neural connections [17]. This study also suggests that the "grouping cells" suggested by Craft et al. (2007) may actually be integration cells for reproducing depth order [18].

5.3 Model limitations

Some object-side selective neurons in area V2 are selective to object shape [6]. The model by Sakai et al. (2012) successfully captured this diversity of neurons through randomly generated neural connections. Deduced neural connections proposed earlier are not capable of accounting for such diversity. Furthermore, a fraction of object-side selective neurons display selectivity to the polarity of the pattern in their receptive fields. For example, a neuron might respond to a black object to the right of its receptive field, but not to a white object to the right of its receptive field. Currently, the proposed model does not use luminance as an input and cannot reproduce these characteristics.

Initial vectors are set in a deterministic manner for the proposed model. As a result, BO signals for only one percept. Randomly setting initial vectors show that the model is capable of reproducing many percepts, as shown in **Section 4.1**. Visual areas other than area V2 might be responsible for the various percepts. The model does not include this factor in setting initial vectors through L-junctions, alone.

6 Extention of the proposed theory

6.1 Surface coding and top-down interactions

This study agrees with the concept introduced by Marr (1982) that the objective of the visual system is to reconstruct a 3D interpretation of the surrounding environment using 2D retinal images. This objective persists even in the absence of binocular disparity (actual depth information), such as for 2D images. For objects closer to us, binocular disparity is large. Binocular disparity decreases the further away from us it is located. However, our visual system is not a perfect machine that can measure binocular disparity precisely. Similarly, there may be "missing" or "undetected" information in the image of the outside world retrieved by our eyes. Thus, the visual system has to account for the possibility that information is incomplete. The completion of a surface based on observable/detectable edges might be the cause of the perception of a "subjective surface".

The perception of a Kanizsa figure in **Figure 6.1** is one example of perception of a "subjective surface" [33]. A subjective surface, shaped like a square, can be perceived at the center of the image. This square is perceived to occlude four white circles. In addition, "subjective contours" of the square is perceived despite no real edges existing. The visual system performs surface reconstruction with the assumption that these "subjective contours" are undetectable. Performance of the proposed model towards the Kanizsa figure serves as a good evaluation criterion.

The proposed model is only conducted on real edges, and not subjective ones. Thus, the model needs to be expanded to account for subjective surfaces. An update rule for surface depth order $\phi(x, y)$ can be obtained in a similar manner to BO vector field $\boldsymbol{E}(x, y)$ by setting a suitable energy function to minimize. Ideally, depth order values at all spatial points on a closed surface should be of the same value; surfaces should be flat.

An existing model which uses level-set methods to obtain flat surfaces, and consequently reproduce subjective contours in Kanizsa figures was adapted to obtain update rule $\partial\phi(x, y, t)/\partial t$ [34]. The details of this model are summarized below. First, an edge detector was defined as

$$g(x, y) = \frac{1}{1 + (|\nabla G_\sigma(x, y) * I(x, y)|/\beta)^2} \tag{6.1}$$

where

$$G_\sigma(\xi) = \frac{\exp(-(\xi/\sigma)^2)}{\sigma\sqrt{\pi}}. \tag{6.2}$$

Values of g at the edges approach 0 at discontinuities in intensity (edges), and 1 for areas of homogenous intensity. The differential area of a surface $S: (x, y) \rightarrow (x, y, \phi)$ can be defined as

$$dA_E = \sqrt{1 + \phi_x^2 + \phi_y^2}\,dxdy. \tag{6.3}$$

Therein, ϕ_x and ϕ_y are partial derivatives of ϕ with respect to x and y. Differential area dA will effectively take a minimum value for flat surfaces since the gradients ϕ_x and ϕ_y will be 0. An energy function based on the edge detector $g(x, y)$ and surface area dA_E can be formulated as

$$J[\phi] = \iint g(x, y)\sqrt{1 + \phi_x^2 + \phi_y^2}\,dxdy. \tag{6.4}$$

At regions where the value of g approaches 0, partial derivatives ϕ_x and ϕ_y, can take large values and not effect energy $J[\phi]$. At regions where the value of g approaches 1, partial derivatives ϕ_x and ϕ_y, will have to take small values so energy $J[\phi]$ approaches a minimum.

Using the steepest descent method on **Eq. (6.4)** results in the following update rule,

$$\frac{\partial\phi}{\partial t} = g\frac{(1 + \phi_x^2)\phi_{yy} - 2\phi_x\phi_y\phi_{xy} + (1 + \phi_y^2)\phi_{xx}}{1 + \phi_x^2 + \phi_y^2} + (g_x\phi_x + g_y\phi_y) \tag{6.5}$$

The authors then consider a rectangular grid in space-time with the points $(t_n, x_i, y_j) = (n\Delta t, i\Delta x, j\Delta y)$ and denote ϕ_{ij}^n as the value of ϕ at (t_n, x_i, y_j). The scheme they propose to approximate **Eq. (6.6)** is

$$
\phi_{ij}^{n+1}
$$
$$
= \phi_{ij}^n
$$
$$
+ \Delta t \left\{ \left[g_{ij} \frac{\left(1 + D_{ij}^{0x^2}\right)D_{ij}^{0yy} - 2D_{ij}^{0x}D_{ij}^{0y}D_{ij}^{0xy} + \left(1 + D_{ij}^{0y^2}\right)D_{ij}^{0xx})}{1 + D_{ij}^{0x^2} + D_{ij}^{0y^2}} \right] \right. \tag{6.6}
$$
$$
- \left[\left[\max\left(g_{ij}^{0x}, 0\right)D_{ij}^{-x} + \min\left(g_{ij}^{0x}, 0\right)D_{ij}^{+x}\right] \right.
$$
$$
\left. \left. + \left[\max\left(g_{ij}^{0y}, 0\right)D_{ij}^{-y} + \min\left(g_{ij}^{0y}, 0\right)D_{ij}^{+y}\right]\right] \right\}
$$

where D is a finite difference operator on ϕ_{ij}^n, superscripts $\{-, 0, +\}$ indicate backward, central and forward differences respectively, and superscripts $\{x, y\}$ indicate the direction of differentiation. Locating the boundary of the surface in the foreground, or subjective surface can be achieved by choosing the level set $\phi = \{\max(\phi) - \varepsilon\}$. Simulations were carried out for a Kanizsa figure and the three benchmark stimuli. For simulations in this paper, $\beta = 0.01$, $\Delta t = 0.1$ and $n = 10000$, except for the occluded square where $n = 20000$. The subjective surface was extracted by $\varepsilon = 0.15$. ϕ was rescaled to 1 after every iteration. The initial surface ϕ_0 was a point of value 1 decaying with distance $1/d$ set at the center of the image. **Figure 6.2** shows the results of numerical simulations. The time course of ϕ for the four figures is shown in **Appendix A.2**. Edges were calculated using a Canny edge detector.

 Qualitatively, the model is capable of reproducing the subjective surface in the Kanizsa figure (**Figure 6.2a**). However, the edges do no describe the shape of the inducers (Packman-shaped figures). In addition, the model does not produce satisfactory results

for the benchmark stimuli, except for the square. From these results, it is clear that the model by Sarti et al. (2000) alone might not be adequate to describe the perception of these stimuli. A method to produce satisfactory results with interactions with the BO assignment model in this study is proposed.

The author proposes using a combination of real contours with subjective contours from the results in **Figure 6.2** as the border for the BO assignment model to be applied on. These subjective contours are regions where the gradient of $\phi(x, y)$ is large. Based on the established relationship, $\boldsymbol{E}(x, y) = \nabla\phi(x, y)$, $\boldsymbol{E}(x, y)$ also should carry a large value at these regions. Pixels containing subjective contours are thus suitable regions for updating $\boldsymbol{E}(x, y)$. This process, for example, might represent interactions between area V4 to area V2. The update rule was applied on (a) real contours, (b) subjective contours and (c) a combination of both real and subjective contours.

Initial vectors were calculated based on the partial derivatives of border $B(x, y)$ at L-junctions, similar to **Eq. (4.1)**, but with Sobel filters instead of central differentiation. Results are shown in **Figure 6.3** to **Figure 6.6**. Human perception tends to agree with either the result using (a) real or (c) a combination of real and subjective contours. A possible method to determine which is preferred over the other is to evaluate the number of real border pixels existing inside the subjective surface seen in (b). If this number crosses a certain threshold, the result using the real contour is adopted. This is particularly useful for the C-shaped figure in **Figure 6.5**, which has 24 real border pixels existing inside the subjective contour. For the combination of real and subjective contours, real edges exist within the subjective surface, and the result using real contours is chosen.
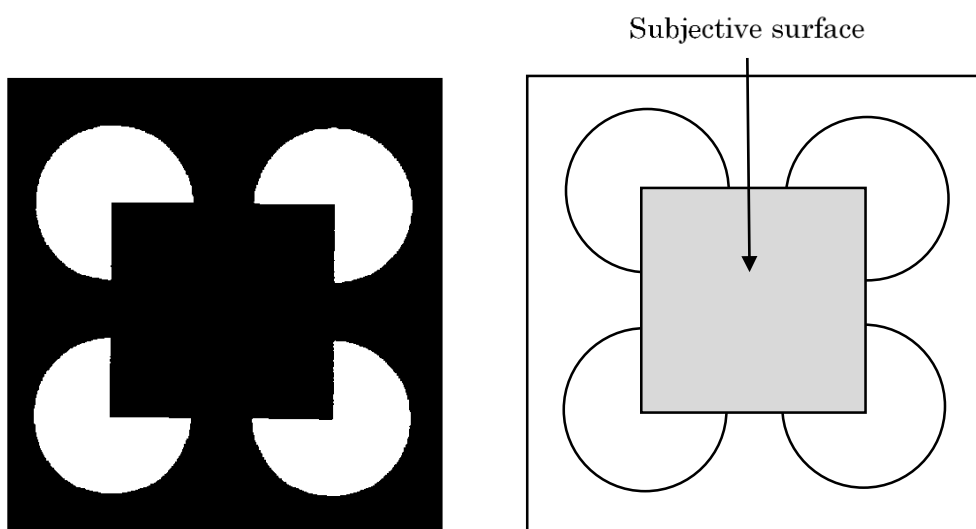
Figure 6.1 The Kanizsa figure demonstrates the perception of a subjective surface.
Based on the "real contours", there should be one interpretation of the image on the left:
four packman-shaped figures. However, humans can also interpret the image as a square
occluding four circles. As shown on the right, a "subjective surface" is perceived in this
instance. "Subjective contours" of this "subjective surface" are perceived.
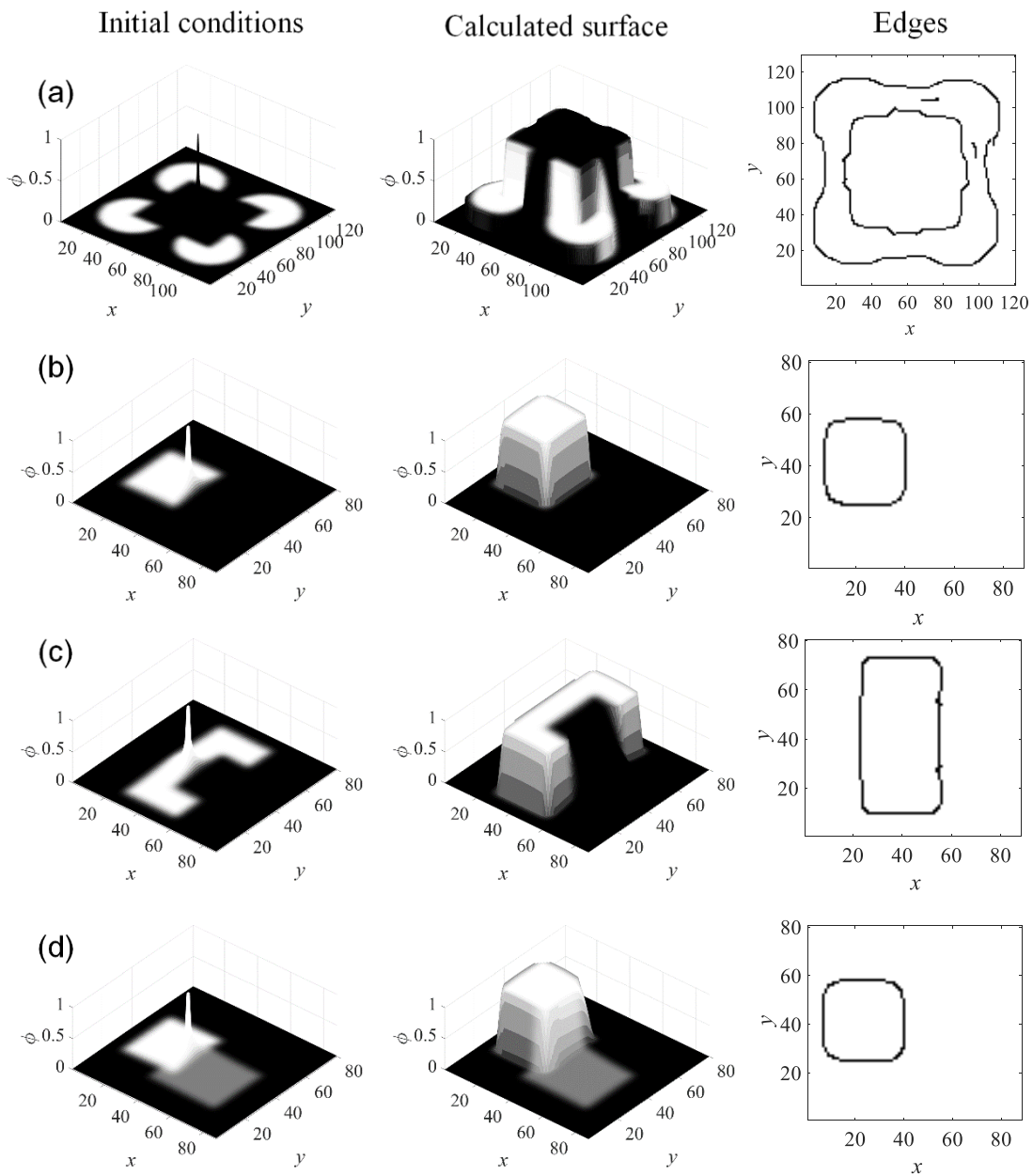
Figure 6.2 Numerical simulations for (a) Kanizsa figure, (b) a square, (c) a C-shaped figure, and (d) an occluded figure using the model by Sarti et al. (2000) and results of Canny edge detection.

Although the model reproduces the subjective surface, the edges do not adequately capture the shape of the occluded objects. The model does not perform well with the benchmark stimuli, other than the square.
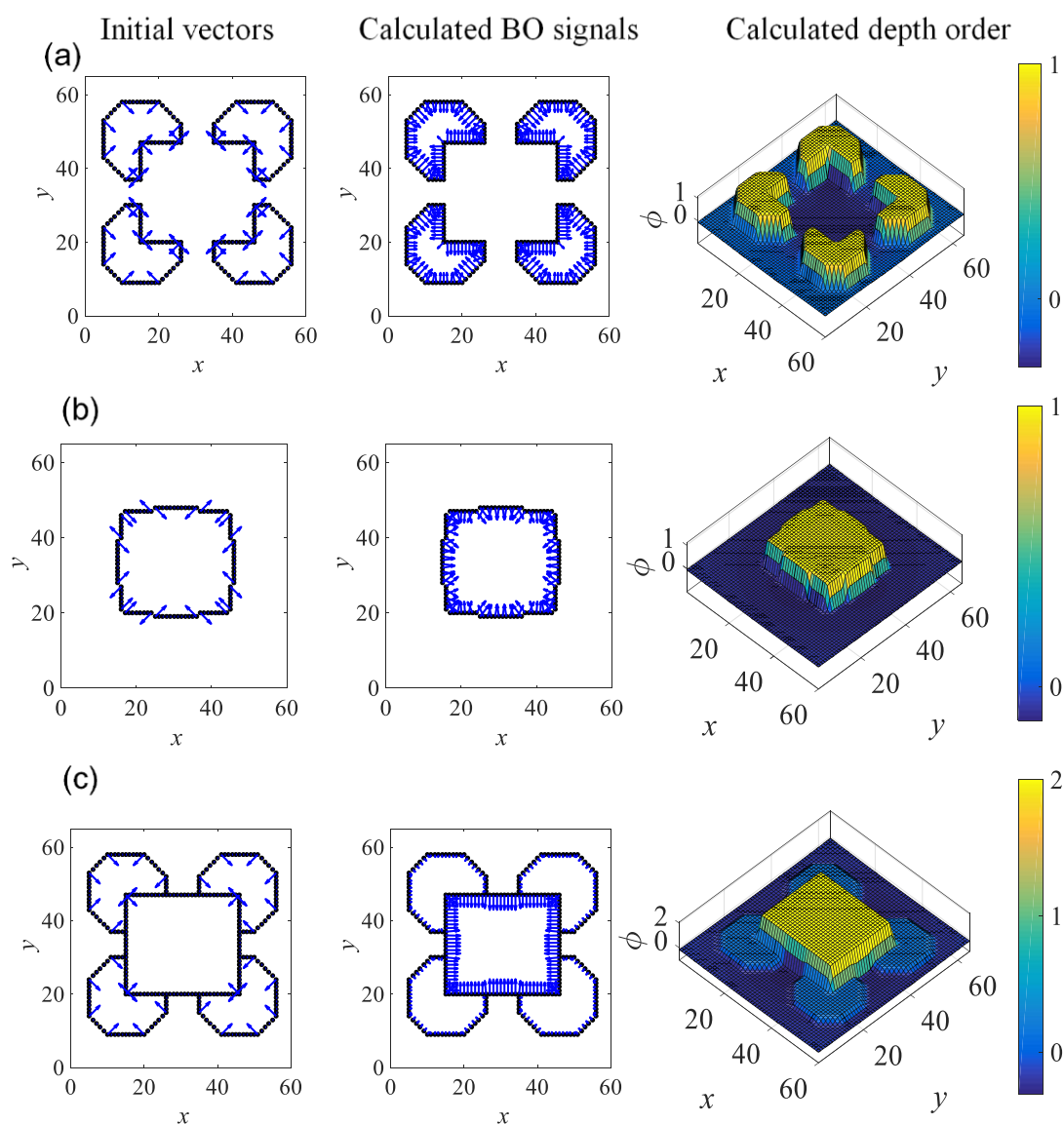
Figure 6.3 Numerical simulations of the BO assignment model for a Kanizsa figure using (a) real contours, (b) subjective contours and (c) a combination of both subjective and real contours.
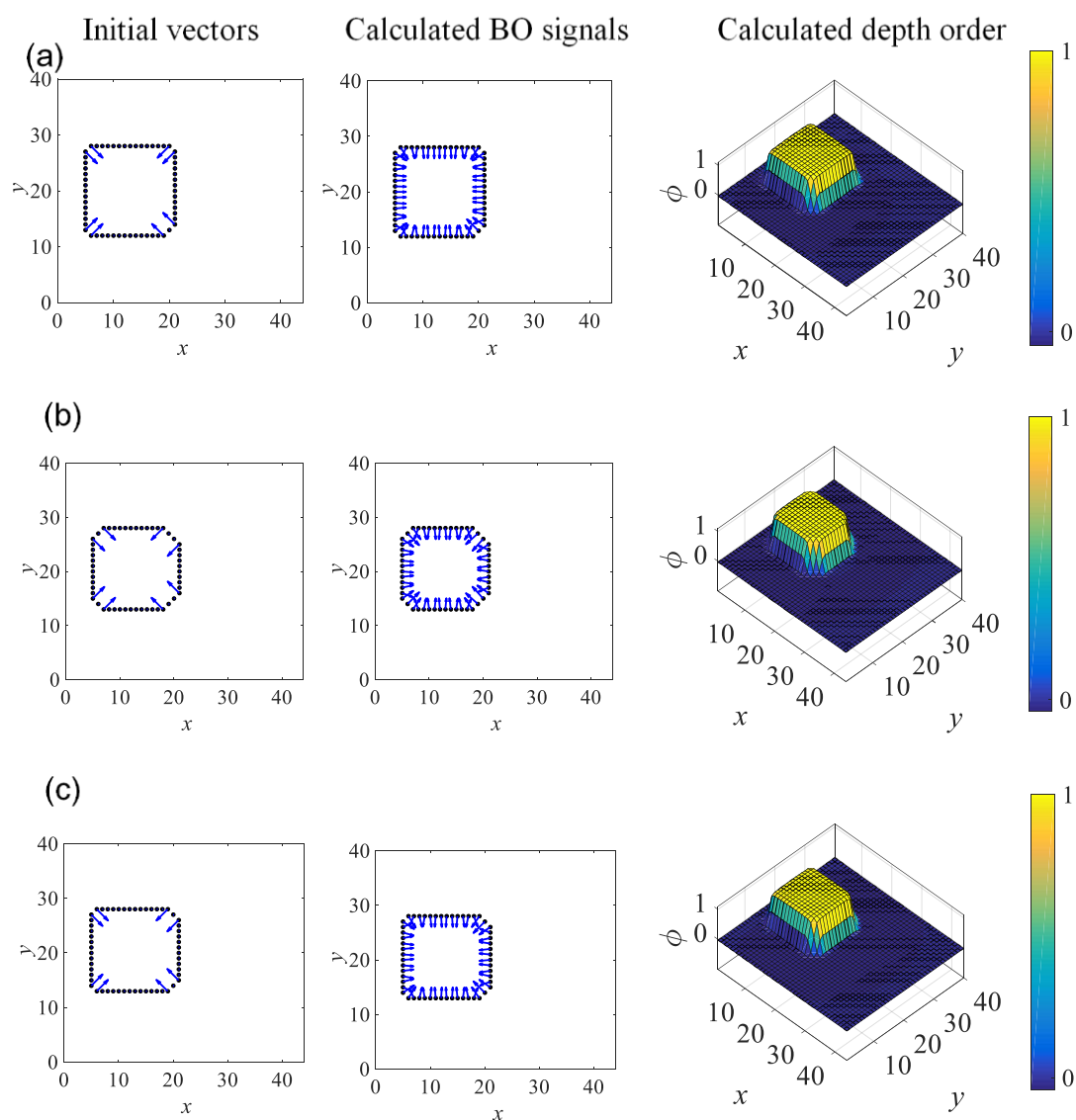
Figure 6.4 Numerical simulations of the BO assignment model for a square using (a) real contours, (b) subjective contours and (c) a combination of both subjective and real contours.
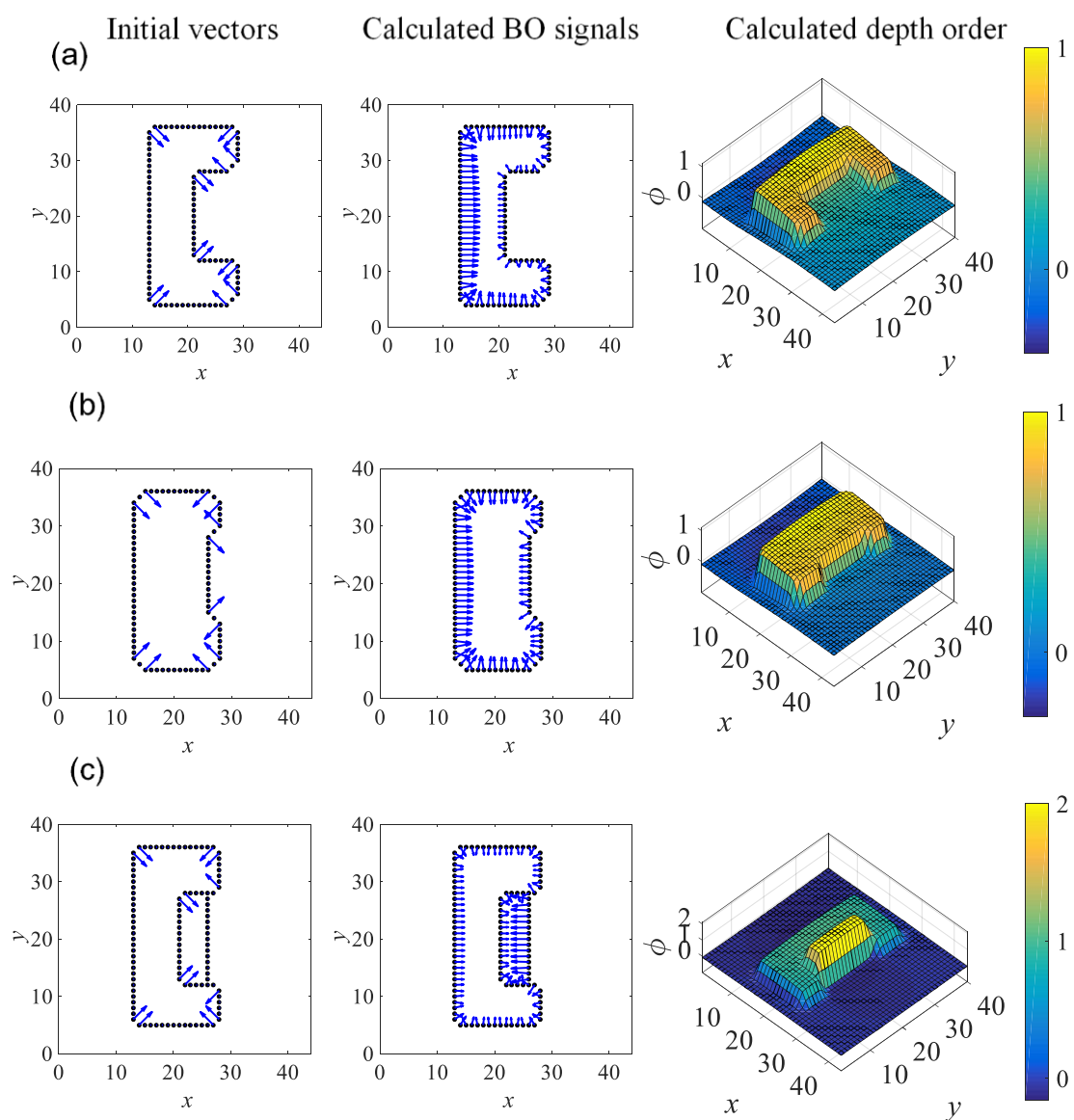
Figure 6.5 Numerical simulations of the BO assignment model for a C-shaped figure using (a) real contours, (b) subjective contours and (c) a combination of both subjective and real contours.
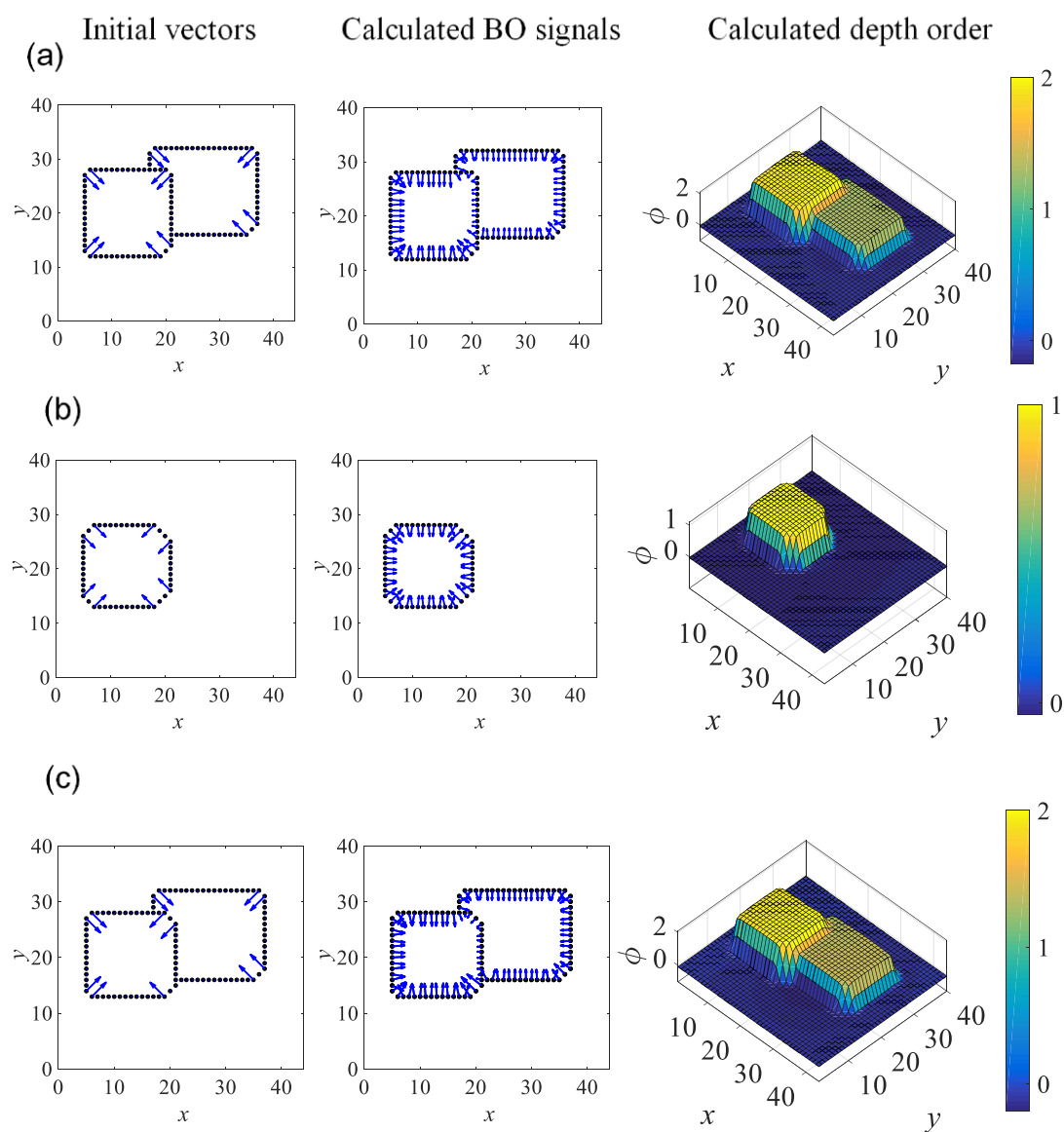
Figure 6.6 Numerical simulations of the BO assignment model for an occluded figure using (a) real contours, (b) subjective contours and (c) a combination of both subjective and real contours.

6.2 Scale-space selection and visual attention

Results of convolution for individual filters can be used for scale-space selection for the size of objects. Scale-space theory, as first proposed by Lindeberg (1994), is based on convolution of an image $I(x, y)$ with normalized Laplacian of Gaussian (LoG) filters [25]. An estimation of the location and size of an object can be made, in addition to an estimation of the medial axis of objects. Figure 6.7 shows an the squared results of convolution for individual weights proposed in **Eq. (5.8)** with BO calculated for the square in **Figure 4.12a**. Results shown for scales $n = 1, 2, 3, 4, 5$ were rescaled to 1. **Figure 6.7f** shows a graph of scale n against squared convolution at spatial location (12,20). The graph takes maximum value of 1 at $n = 2.7$, indicating that (12,20) is the center of an object with an approximate size of $2^{2.7}$. The location of this maximum value coincides with the center of an object closest to the viewer. This can be applied to simulating the effect of attention. Object-based saliency map have been shown to predict the location of visual attention [35].

The image of four overlapping cards in **Figure 6.8** is an example where local attention at different locations produces different perceptions of depth order. Notably, calculating BO using the method proposed in this study produces a flat surface when the number of iterations are large, and an "endless staircase" when $t = 100$ (**Figure 4.22**). Model results from adding local bias to intial vectors are summarized in **Figure 6.8**. For example, if one focuses on (a), the surface with the highest depth order in the vicinity is (b). When one focuses their attention on (b), the surface with the depth order in the vicinity is (c). This process continues until attention moves back to the starting point at (a).

Details of how the effect of local attention is seen in **Figure 6.8**. If we assume visual attention as a spotlight, initial vectors $\boldsymbol{E}(x, y, 0)$ can be changed with weights α taking

a maximum at the spatial location of visual attention. A possible weight α using a Gaussian function of standard deviation σ, located at (x_o, y_o) is

$$\alpha = (1 - \beta) \exp\left(-\frac{(x - x_o)^2 + (y - y_o)^2}{2\sigma^2}\right) + \beta \qquad (6.7)$$

Therein, $\beta = 0.4$ and $\sigma = 2^{2.5}$ for this study. The location of attention (x_o, y_o) was first set at $(27,19)$. The location (x_o, y_o) for the next weight is obtained by searching for the location where squared convolution takes a maximum value. First, let us examine the results in **Figure 6.10**. Attention towards the object on the right side $(27,9)$ of the image produces depth order, which resembles an ascending staircase. This agrees with perception if attention is given in the same manner. The next location of attention, calculated automatically using scale-space selection is shifted to the object with a higher depth order value : the object at the upper side $(18,28)$. This process is repeated, and the location of attention makes a full circle back to the initial position, thus reproducing the perception of an "endless staircase". **Figure 6.11** to **Figure 6.13** illustrates the different depth order (surface) perceptions based on the location to where attention is given.
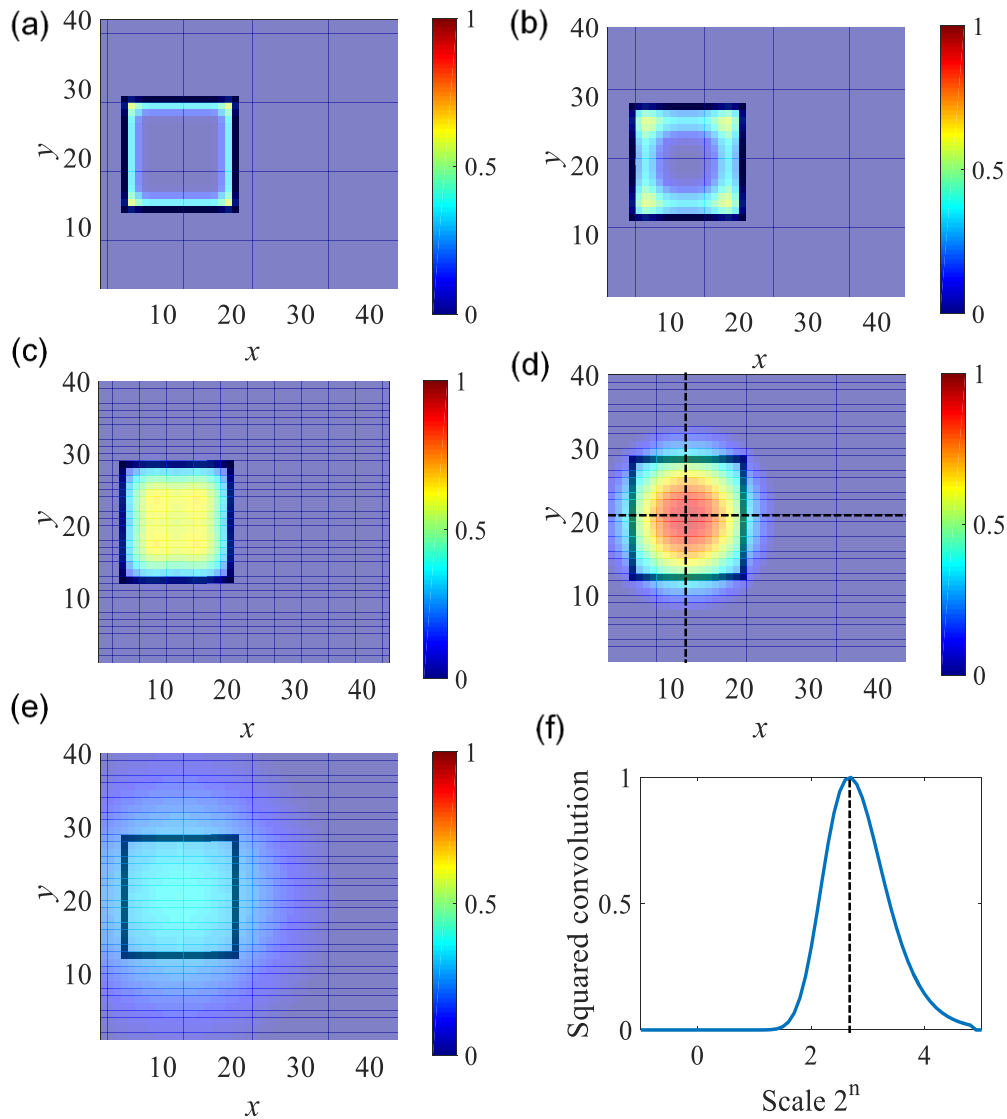
Figure 6.7 Squared convolution of individual integration kernels at different scales in Eq. (5.8).

Square convolution of the BO signals (vector field) of the square in **Figure 4.12a** at scales (a) 2^1, (b) 2^2, (c) 2^3, (d) 2^4 and (e) 2^5. (f) Squared results of convolution for scales between -1 to 5 at spatial location $(12,20)$. A maximum value 1 can be seen at $n = 2.7$.
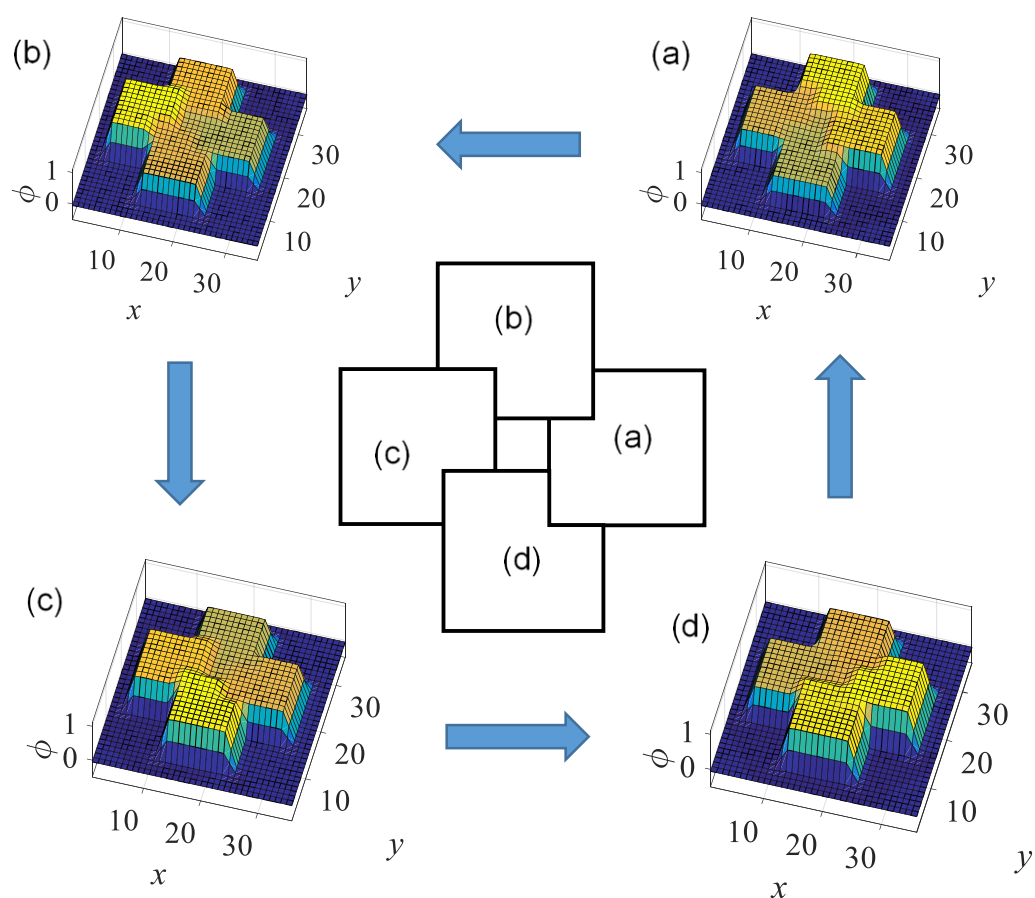
Figure 6.8 Example of focusing local visual attention of the perception of object depth order with model outputs.

When one focuses local visual attention on (a) at the center, the surface depth order of the vicinity, calculated by changing initial vectors, is the highest at surface (b). One then moves the focus of attention to occluding object (b). This is repeated until attention moves back to (a). The proposed model can carry out this process automatically by use of scale-space selection and adjusting initial vectors locally.
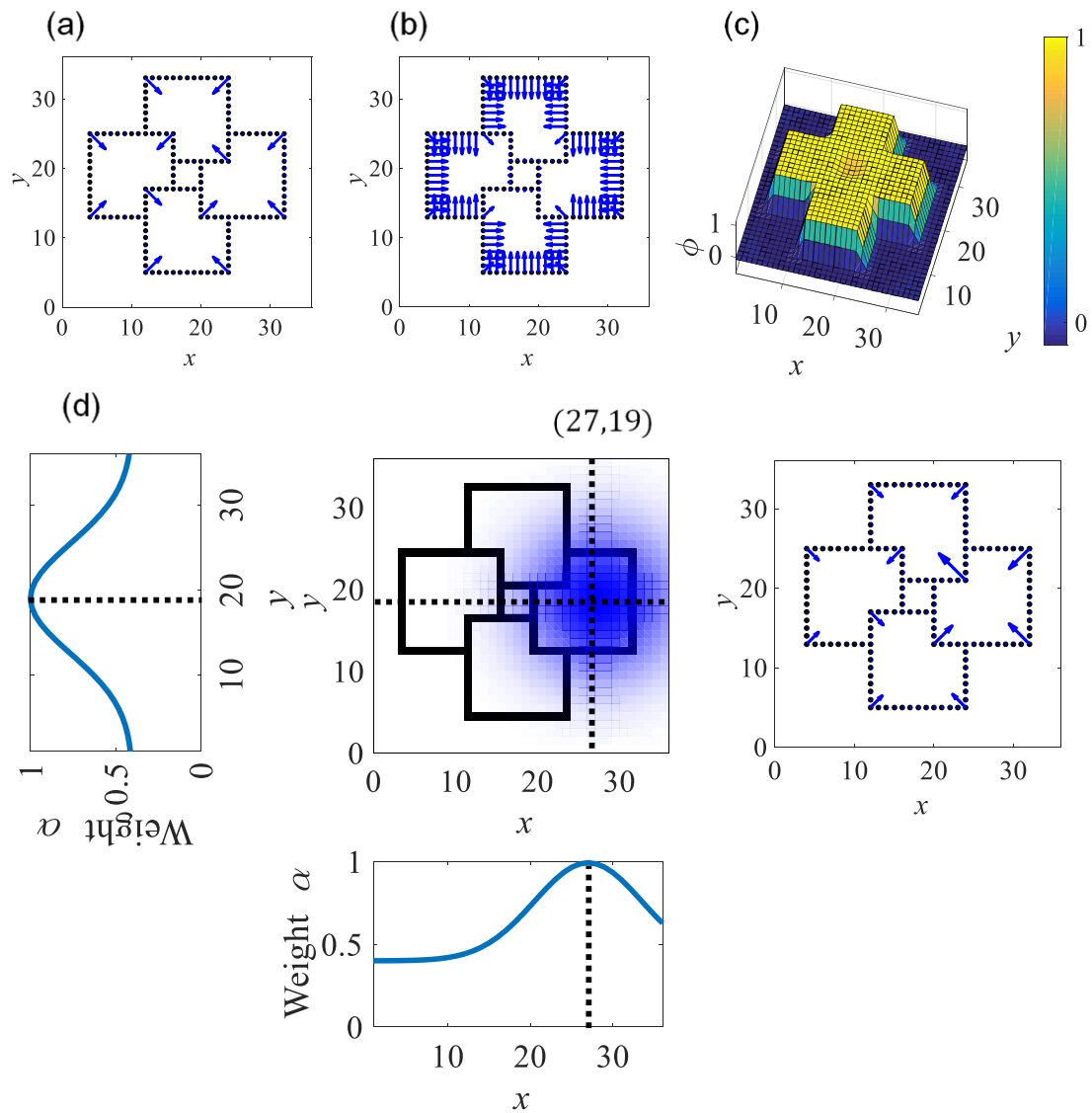
Figure 6.9 Model results for BO signals without the effect of attention and a possible weight for local processing.

(a) Initial vectors, (b) calculated BO signals and (c) calculated depth order without the effect of attention. (d) A possible weight to change initial vectors, caused by attention to the object on the right side of the image.
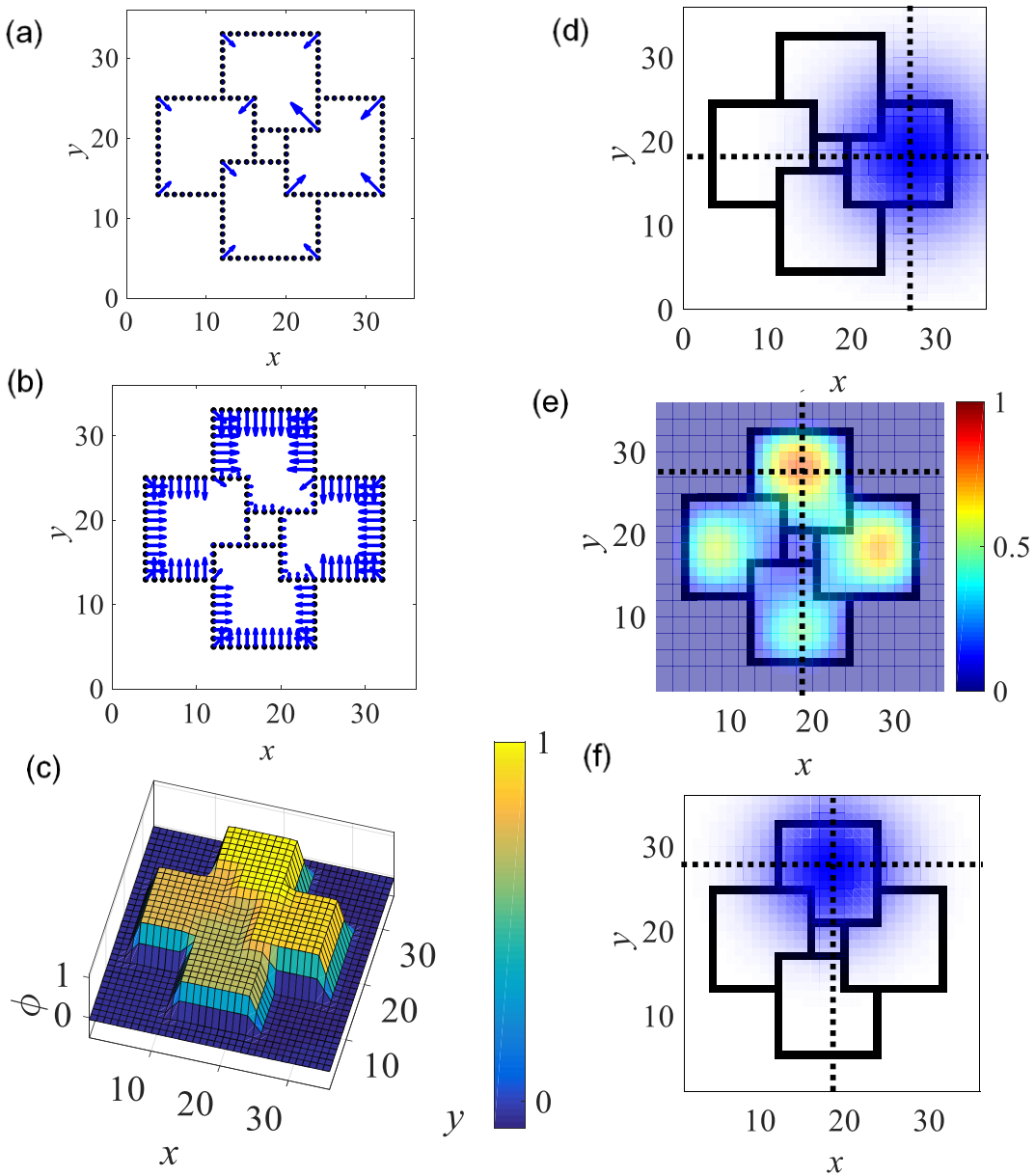
Figure 6.10 Simulation results with the effect of attention on the object on the right side of the image.

(a) Initial vectors, (b) calculated BO, and (c) depth order. (d) Location of attention that produced (a) initial vectors. (e) Squared convolution at scale $n = 2$, where a maximum value is located at $(18,28)$, (f) the next location of attention (object with highest depth order).
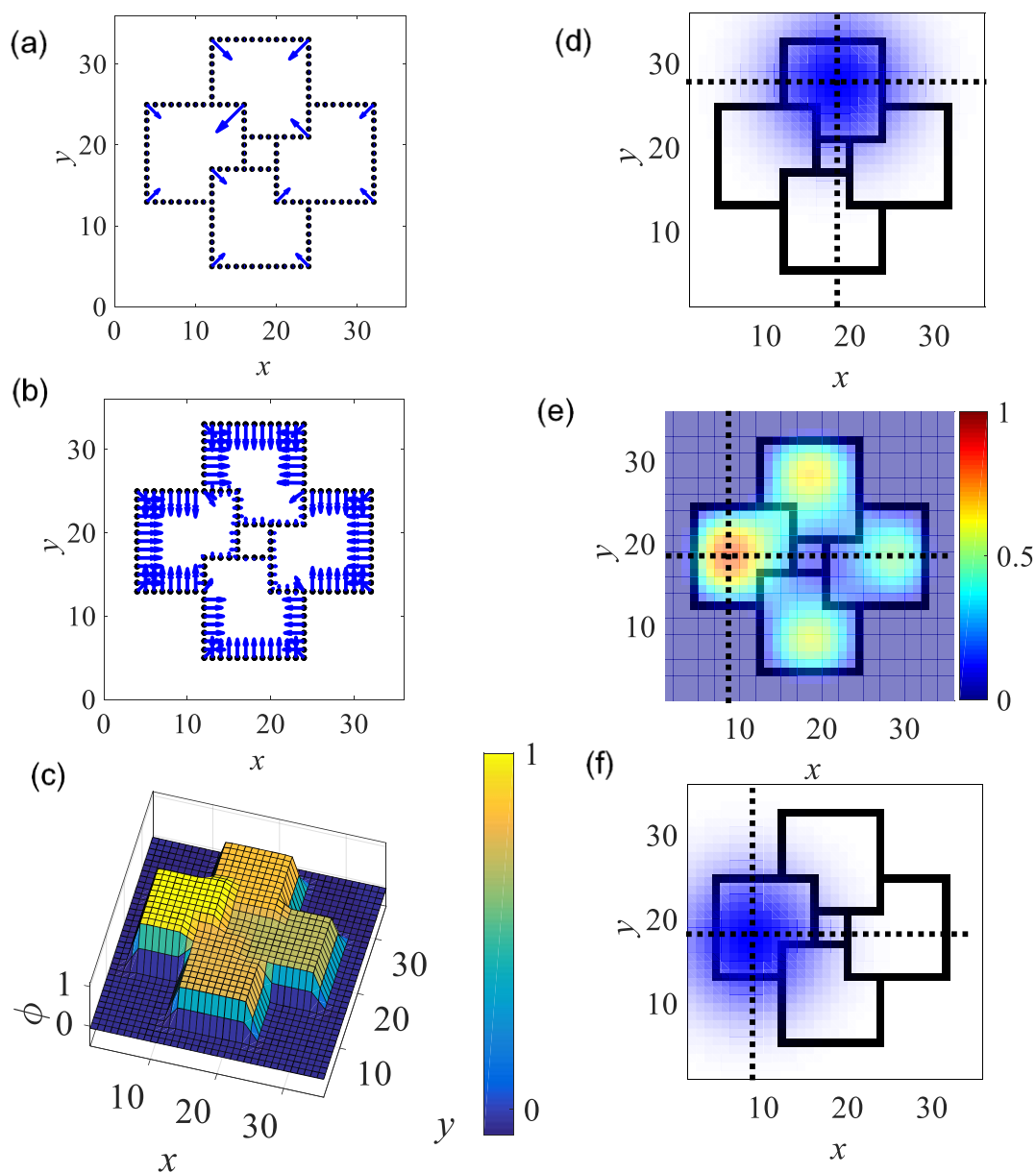
Figure 6.11 Simulation results with the effect of attention on the object on the upper side of the image.

(a) Initial vectors, (b) calculated BO, and (c) depth order. (d) Location of attention that produced (a) initial vectors. (e) Squared convolution at scale $n = 2$, where a maximum value is located at (8,18), (f) the next location of attention (object with highest depth order).
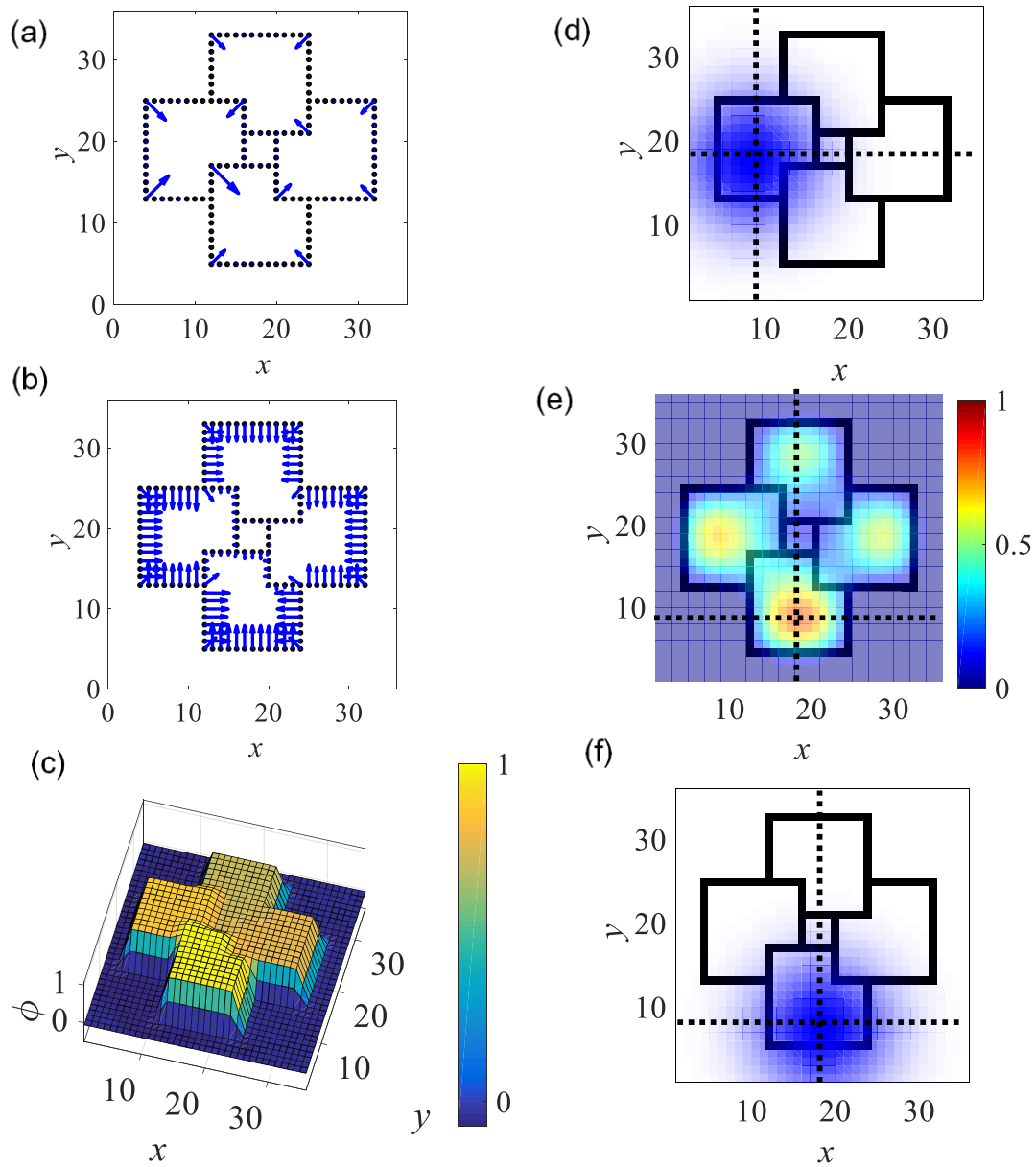
Figure 6.12 Simulation results with the effect of attention on the object on the left side of the image.

(a) Initial vectors, (b) calculated BO, and (c) depth order. (d) Location of attention that produced (a) initial vectors. (e) Squared convolution at scale $n = 2$, where a maximum value is located at $(18,8)$, (f) the next location of attention (object with highest depth order).
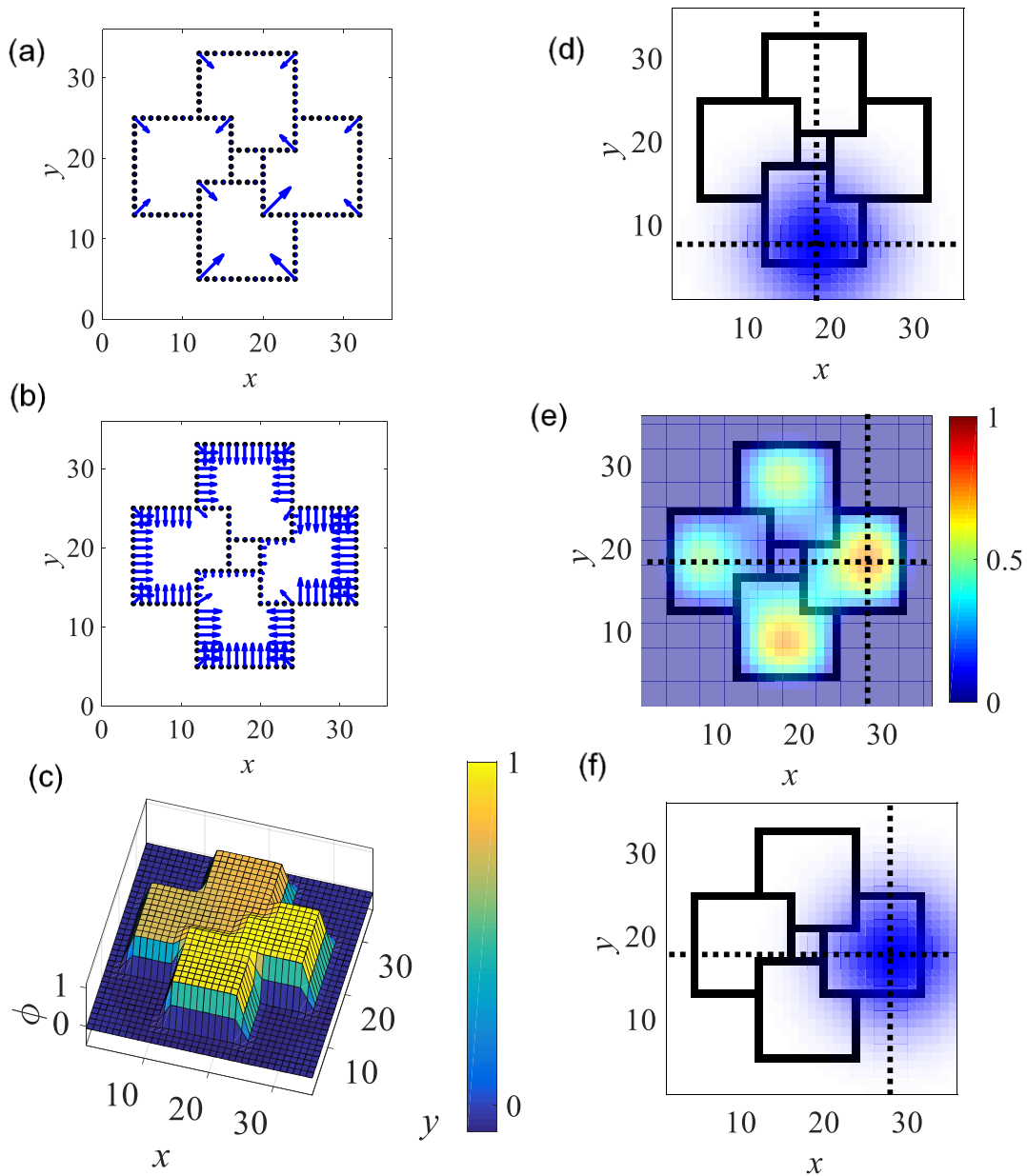
Figure 6.13 Simulation results with the effect of attention on the object on the lower side of the image.

(a) Initial vectors, (b) calculated BO, and (c) depth order. (d) Location of attention that produced (a) initial vectors. (e) Squared convolution at scale $n = 2,$ **where** a maximum value is located at (8,18), (f) the next location of attention (object with highest depth order). This is the approximate location of the first location of visual attention in **Figure 6.10**.

6.3 Brightness reconstruction from vector integration

In the previous section, integration of a BO vector field using multiple annular-shaped receptive fields was used to reconstruct surfaces and in turn reproduce depth order. This can be likened to processing from area V2 to area V4. One surprising discovery was the result of such integration on the gradient of an image $\nabla I(x,y) = \frac{\partial}{\partial x} I(x,y)\hat{x} + \frac{\partial}{\partial y} I(x,y)\hat{y}$. Such a process is analogous to processing from area V1 to area V4, or the reconstruction of brightness,

$$\hat{I}(x,y) = 2^n \sum_n \left(\frac{\partial}{\partial x} G_n(x,y) \right.$$

$$\left. * \frac{\partial}{\partial x} I(x,y) \right) + 2^n \sum_n \left(\frac{\partial}{\partial y} G_n(x,y) * \frac{\partial}{\partial y} I(x,y) \right). \tag{6.8}$$

Therein, partial derivatives of $I(x,y)$ were calculated with central differentiation.

Annular weights in **Eq. (5.8)** for n from -1 to 5 at 0.1 invervals were used to reconstruct brightness. **Figure 6.14** shows a bar that is uniformly grey over a dark to white gradient. The bar is perceived as light grey on the left, and dark grey on the right. The proposed integration method reproduces this perception. This is because the annular weights were limited to a scale of $n = 5$, meaning that vectors located further away are have a smaller weight. The integration method in this study also reproduces the white and black "bands" in the Chevreul illusion in **Figure 6.15**. The relative brightness of these four bands are qualitatively reproduced. "Bands" arise because integration of vectors is conducted in a radial manner.
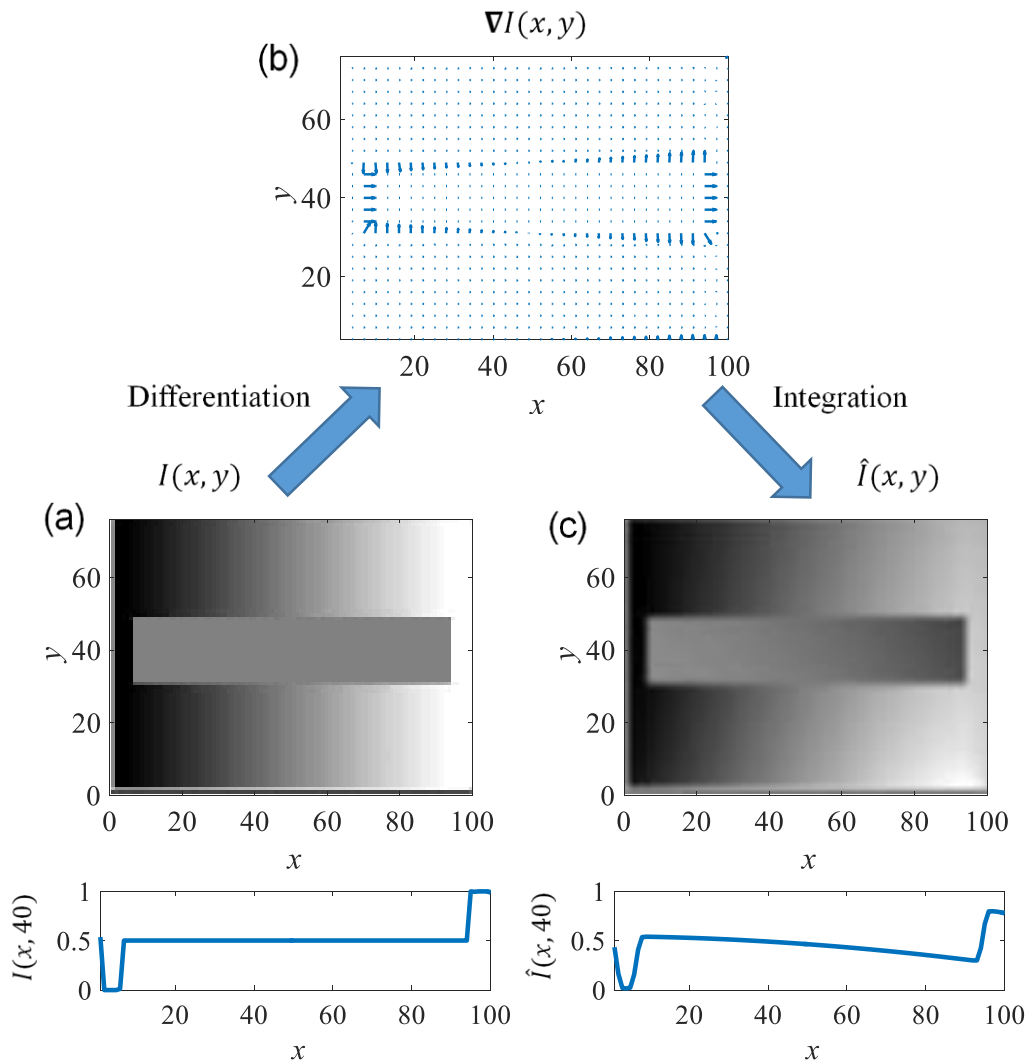
Figure 6.14 The method of brightness reconstruction proposed in this study reproduces the perception of a brightness gradient on a uniformly-filled bar.

The integration method proposed in this study reproduces the perception of a light to dark grey bar. (a) Input image $I(x, y)$, (b) gradient of image $\nabla I(x, y)$, and (c) integration of the image gradient using integration kernels proposed in this paper $\hat{I}(x, y)$. Contrast of the bar with its background produces the perception of a grey bar that has a gradient from light to dark, despite it having a uniform physical luminance.
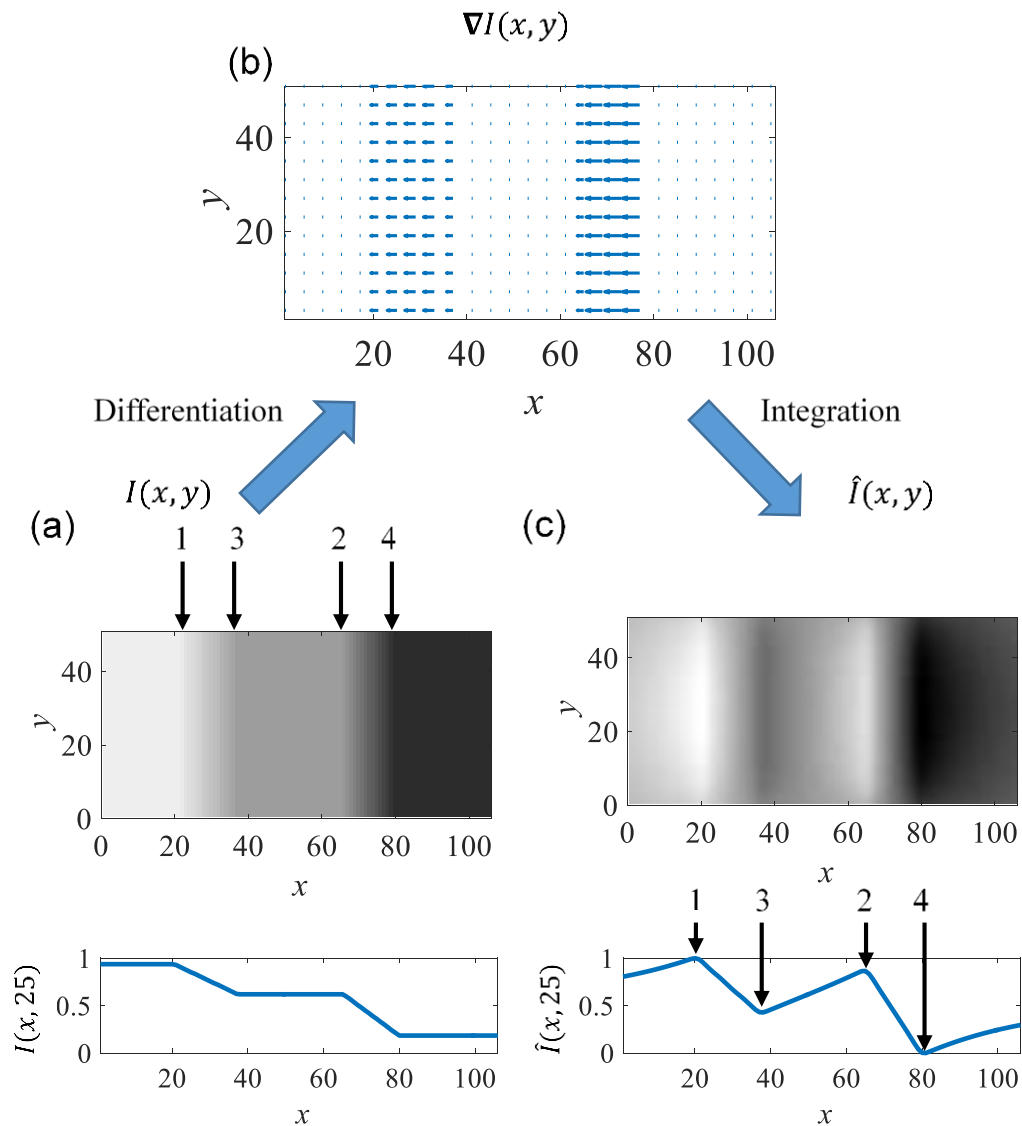
Figure 6.15 The method of brightness reconstruction proposed in this study reproduces the bands seen in the Chevreul illusion.

The integration method proposed in this study reproduces the perception of a light to dark grey bar. (a) Input image $I(x, y)$, (b) gradient of image $\nabla I(x, y)$, and (c) integration of the image gradient using integration kernels proposed in this paper $\hat{I}(x, y)$. Four light and dark "bands" are perceived at the ramps between regions of three different luminance values. Integration using the multiple annular-shaped integration kernels qualitatively produces the relative brightness of these bands, where 1 is the brightest and 4 is the darkest.

7   Discussion and conclusion

7.1 Overall discussion

In the previous section, perceptual alternation and dependence of simulation results on initial vectors was discussed briefly. Although the mechanics of such an initial vector assignment is unknown, the model is capable of reproducing the two percepts from a C-shaped contour, as seen in **Figure 4.3**. The energy is at a local minima and the system is its stable state for both percepts. **Figure 7.1** illustrates the nature of an energy function with two local minima, corresponding to the two stable percepts. For explanation purposes, suppose that perceptual alternation occurs in this case. This would require crossing the peak which separates the two local minima. Perceptual alternation was reported to to be caused by a discrete schotastic process in the brain [36]. Although the proposed model cannot produce the discrete states in the transition phase, it is capable of producing the stable states where energy is at a minimum.

Local concavity is a low-level perceptual cue for object-side. A possible mechanism that might explain various percepts is attention. Psychophysical research using stimuli with locally concave but globally convex images was used to demonstrate the effect of attention on the perception of depth order [37]. Drawing the viewer's attention to the locally concave area resulted in a reverse of the perception of depth order of the attended region. Physiological experiments support the effect of attention on depth order, as object-side selective neurons respond stronger when attention is given to the object [38]. It is also important to note that object familiarity plays a role in the perception of depth order in an image [39]. A combination of attention towards local concavity cues and top-down effects from object recognition might contribute to how an image is perceived.

In this study, an existing model which successfully reproduce subjective surfaces was

adopted [40], [41]. As demonstrated in **Section 4.2**, using real and subjective contours appropriately successfully reproduced human perception. It has been reported that there are sets of neurons which respond at equal strength towards both real and subjective contours, in addition to solely real or subjective ones (100 ms after stimulus onset) [42]. Neurons which respond towards subjective surfaces respond earlier, or simultaneously with those which respond towards subjective contours in area V4 [43]. This suggests area V4 might be the location where surface reconstruction and border completion is conducted. In simulations, initial surfaces were set manually. However, scale-space theory can be used to set the location for an initial surface in the model.

Experiments by Chen at al. (2014) on areas V1 and V4 in awake monkeys suggest that global contours are processed concurrently in a bidirectional loop [44]. In this study, a possible framework framework for interactions between for BO and depth order (area V2 and V4) was proposed based on real and subjective contours. Interactions in time between BO and depth order need to be addressed to produce a model which might reproduce the findings by Chen et al (2014). One possible dynamic to achieve this is to incorporate the update rule for depth order $\partial\phi(x, y, t)/\partial t$ in the update rule for BO $E(x, y, t)/\partial t$ and vice-versa,

$$\frac{\partial E}{\partial t} = \nabla^{\perp}\mathrm{curl}(E) + \alpha\nabla\frac{\partial\phi}{\partial t} \tag{7.1}$$

$$\frac{\partial\phi}{\partial t} = g\,\frac{(1 + \phi_x^2)\phi_{yy} - 2\phi_x\phi_y\phi_{xy} + (1 + \phi_y^2)\phi_{xx}}{1 + \phi_x^2 + \phi_y^2} + (g_x\phi_x + g_y\phi_y) \\ + \beta\iint\frac{\partial E}{\partial t}dxdy \tag{7.2}$$

where α and β are suitable parameters. The author hopes to address this in a future work.

The proposed model is based on the assumption that both BO $E(x, y)$ and depth order $\phi(x, y)$ are unknown at stimulus onset. This condition occurs for 2D images or when disparity is not large enough to estimate object depth order. As introduced earlier, the response of some BO-coding neurons in area V2 are effected by both Gestalt cues such as closure of contours, as well as actual stereoscopic depth [9]. In a condition where depth order $\phi(x, y)$ can be estimated by, for example, binocular disparity, calculating the gradient of depth order $\nabla\phi(x, y)$ will result in BO $E(x, y)$.

Amodal completion is an important process in object recognition. For example, Fukushima (2005) proposed a neural network model to restore partially occluded alphabets [45]. The occluding patterns were removed using predefined masks before the pattern restoration process. It might be possible to employ the proposed model to remove these occluding patterns automatically.
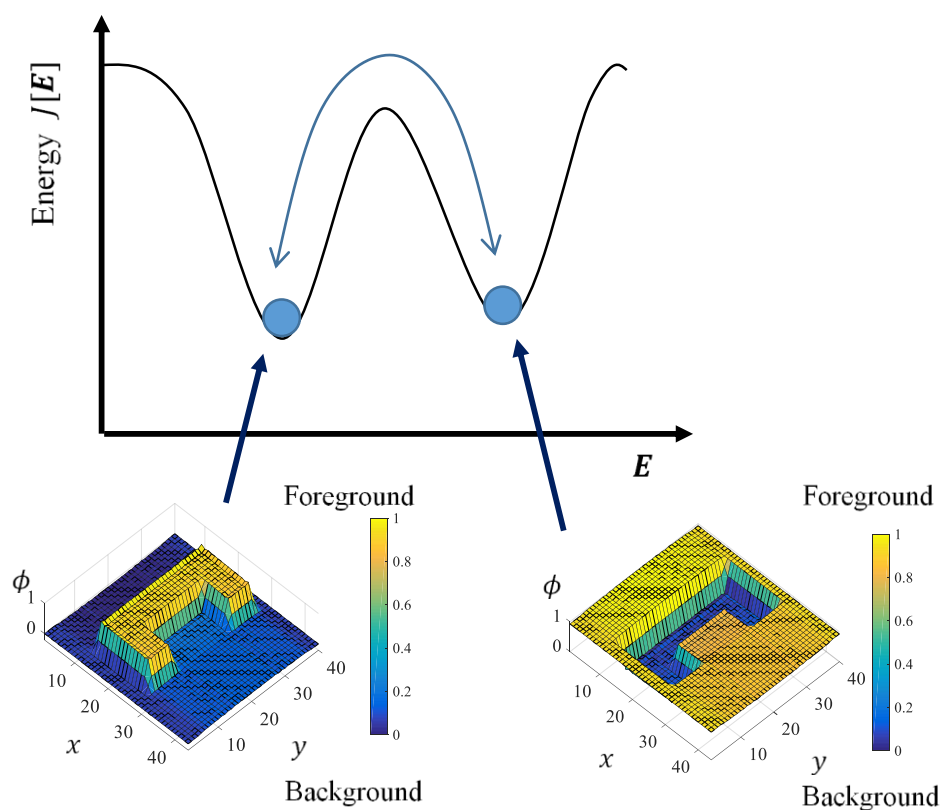
Figure 7.1 Characteristics of the energy function proposed in this study.

Two local minima exist, where the system is in a stable state. The proposed model is capable of producing these two percepts depending on the initial conditions. Supposing the image is perceptually bistable, and perceptual alternation occurs, the system will have to reach the local minimum on the opposing side of the peak between them. This is represented by the double-sided arrow.

7.2 Summary and conclusion

This study is the first to formulate the BO assignment in a well-defined manner. Treating BO signals as a vector field $E(x, y)$, and depth order as a scalar field $\phi(x, y)$, an analogy was made to the relationship between an electric field and its electric potential: the gradient of electric potential is an electric field ($E(x, y) = \nabla\phi(x, y)$). Using the theorem of zero curl for a conservative field, an update rule was formulated. Simulation results of the proposed model was capable of solving BO assignment and depth order simultaneously for various problems involving occlusion and shape. This study also provided a mathematical base for existing models (**Figure 7.2**). Characteristics of deduced neural networks was similar to the model by Li (2005), suggesting that its computational objective is to minimize curl. The annular receptive fields of grouping cells by Craft et al. (2007) was similar to those deduced from executing the task of reconstructing surfaces from BO signals, suggesting that its computational objective is to integrate vector fields. The reproduction of random neural connections Sakai et al. (2012) will be addressed in a future work. The addition of surface completion using an existing model showed how subjective surfaces and contours can be reproduced using top-down processing. Although the computational theory proposed is a simple one, the findings in this study might serve as a fundamental base for others studies in object perception.
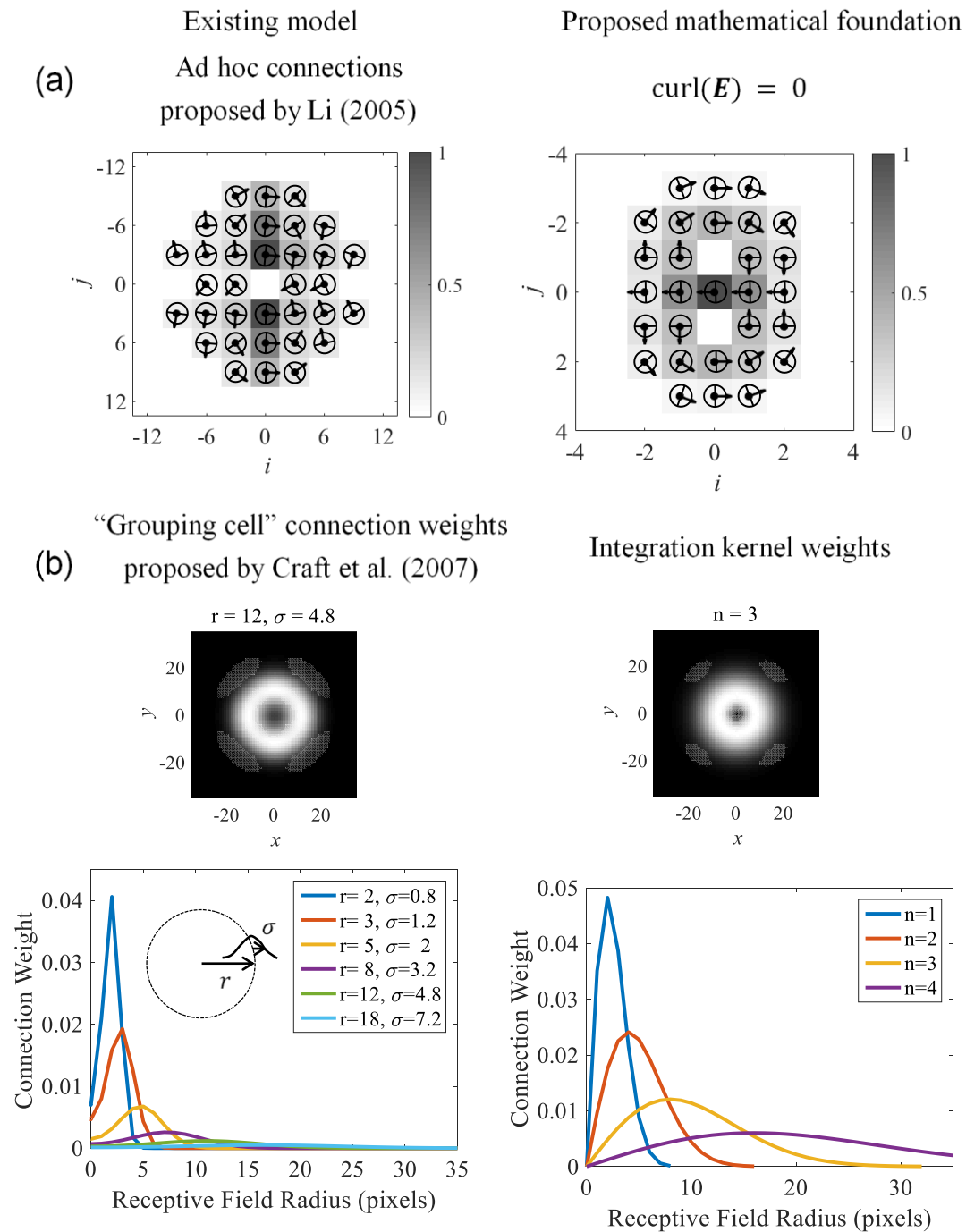
Figure 7.2 This study provides a mathematical foundation for existing models.
(a) This study suggests that the computational objective of ad hoc neural weights proposed by Li (2005) is to minimize the curl of a vector field. (b) This study suggests that the computational objective of annular weights based on circle detection proposed by Craft et al. (2007) is to integrate BO signals to reconstruct depth order scalar field.

References

[1]   E. Rubin, *Visuell wahrgenommene Figuren*. Copenhagen: Gyldendals, 1921.

[2]   K. Nakayama, S. Shimojo, and G. H. Silverman, "Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects," *Perception*, vol. 18, pp. 55–68, 1989.

[3]   D. J. Felleman and D. C. Van Essen, "Distributed Hierarchical Processing in the Primate Cerebral Cortex," *Cereb. Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[4]   T. N. Hubel and D. H. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *J. Physiol.*, vol. 148, no. 3, pp. 574–591, 1959.

[5]   D. Y. Tsao, W. A. Freiwald, T. A. Knutsen, J. B. Mandeville, and R. B. H. Tootell, "Faces and objects in macaque cerebral cortex," *Nat. Neurosci.*, vol. 6, no. 9, pp. 989–995, Sep. 2003.

[6]   H. Zhou, H. S. Friedman, and R. von der Heydt, "Coding of border ownership in monkey visual cortex.," *J. Neurosci.*, vol. 20, no. 17, pp. 6594–6611, 2000.

[7]   T. Sugihara, F. T. Qiu, and R. von der Heydt, "The speed of context integration in the visual cortex.," *J. Neurophysiol.*, vol. 106, no. 1, pp. 374–85, Jul. 2011.

[8]   J. R. Williford and R. von der Heydt, "Figure-Ground Organization in Visual Cortex for Natural Scenes," *eNeuro*, vol. 3, no. 6, p. ENEURO.0127-16.2016, 2016.

[9]   F. T. Qiu and R. von der Heydt, "Figure and Ground in the Visual Cortex: V2 Combines Stereoscopic Cues with Gestalt Rules," *Neuron*, vol. 47, no. 1, pp. 155–166, 2005.

[10]   A. Pasupathy and C. E. Connor, "Population coding of shape in area V4.," *Nat. Neurosci.*, vol. 5, no. 12, pp. 1332–1338, 2002.

[11]   B. N. Bushnell, P. J. Harding, Y. Kosai, and A. Pasupathy, "Partial Occlusion Modulates Contour-Based Shape Encoding in Primate Area V4," vol. 31, no. 11, pp. 4012–4024, 2011.

[12]   M. A. Cox and A. Maier, "Serial versus parallel processing in mid-level vision : filling-in the details of spatial interpolation," *Neurosci. Conscious.*, no. August, pp. 1–7, 2015.

[13]   P. De Weerd, R. Desimone, and L. G. Ungerleider, "Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques," *Vis. Neurosci.*, vol. 13, no. 3, p. 529, May 1996.

[14]   P. H. Schiller, "The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey," *Vis. Neurosci.*, vol. 10, no. 4, pp. 717–746, 1993.

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

[16]   Z. Li, "Border ownership from intracortical interactions in visual area V2," *Neuron*, vol. 47, no. 1, pp. 143–153, Jul. 2005.

[17]   K. Sakai, H. Nishimura, R. Shimizu, and K. Kondo, "Consistent and robust determination of border ownership based on asymmetric surrounding contrast," *Neural Networks*, vol. 33, pp. 257–274, Sep. 2012.

[18]   E. Craft, H. Schütze, E. Niebur, and R. von der Heydt, "A neural model of figure-ground organization," *J. Neurophysiol.*, vol. 97, no. 6, pp. 4310–4326, Jun. 2007.

[19]   A. Thielscher and H. Neumann, "Globally consistent depth sorting of overlapping 2D surfaces in a model using local recurrent interactions," *Biol.*

Cybern., vol. 98, no. 4, pp. 305–337, 2008.

[20]   F. Heitger, R. von der Heydt, E. Peterhans, L. Rosenthaler, and O. Kübler, "Simulation of neural contour mechanisms: representing anomalous contours," *Image and Vision Computing*, vol. 16, no. 6–7. pp. 407–421, 1998.

[21]   O. W. Layton, E. Mingolla, and A. Yazdanbakhsh, "Dynamic coding of border-ownership in visual cortex," *J. Vis.*, vol. 12, no. 13, p. 8, 2012.

[22]   D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Electr. Eng. - Part III Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, 1946.

[23]   S. Marĉelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Am.*, vol. 70, no. 11, pp. 1297–1300, Nov. 1980.

[24]   B. a Olshausen and D. J. Field, "Sparse coding with an incomplete basis set: a strategy employed by V1?," *Vision Research*, vol. 37, no. 23. pp. 3311–3325, 1997.

[25]   T. Lindeberg, "Scale-space theory: a basic tool for analyzing structures at different scales," *J. Appl. Stat.*, vol. 21, no. 1, pp. 225–270, 1994.

[26]   M. Ito and H. Komatsu, "Representation of angles embedded within contour stimuli in area V2 of macaque monkeys.," *J. Neurosci.*, vol. 24, no. 13, pp. 3313–24, Mar. 2004.

[27]   M. Ito and N. Goda, "Mechanisms underlying the representation of angles embedded within contour stimuli in area V2 of macaque monkeys," *Eur. J. Neurosci.*, vol. 33, no. 1, pp. 130–142, Jan. 2011.

[28]   M. Wertheimer, *Untersuchungen zur Lehre von der Gestalt II*, 4th ed. Psycologische Forschung, 1923.

[29] F. T. Qiu and R. von der Heydt, "Neural representation of transparent overlay.,"
      Nat. Neurosci., vol. 10, no. 3, pp. 283–284, 2007.

[30] P. O'Herron and R. von der Heydt, "Short-Term Memory for Figure-Ground
      Organization in the Visual Cortex," *Neuron*, vol. 61, no. 5, pp. 801–809, 2009.

[31] H. R. Wilson, F. Wilkinson, and W. Asaad, "Concentric Orientation Summation
      in Human Form Vision," *Vision Res.*, vol. 37, no. 17, pp. 2325–2330, 1997.

[32] J. Gallant, C. Connor, S. Rakshit, J. Lewis, and D. Van Essen, "Neural responses
      to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey.,"
      J Neurophysiol, vol. 76, no. 4, pp. 2718–39, 1996.

[33] N. Kogo, A. Drozdzewska, P. Zaenen, N. Alp, and J. Wagemans, "Depth
      perception of illusory surfaces," *Vision Res.*, vol. 96, pp. 53–64, 2014.

[34]   a Sarti, R. Malladi, and J. a Sethian, "Subjective surfaces: a method for
      completing missing boundaries.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 12,
      pp. 6258–6263, 2000.

[35] A. F. Russell, S. Mihalaş, R. von der Heydt, E. Niebur, and R. Etienne-
      Cummings, "A model of proto-object based saliency," *Vision Res.*, vol. 94, pp.
      1–15, 2014.

[36] T. Murata, T. Hamada, Y. Kakita, and T. Yanagida, "Meaning of gamma
      distribution in perceptual rivalry Tsutomu Murata," no. 29, 2004.

[37] S. Kim and J. Feldman, "Globally inconsistent figure/ground relations induced by
      a negative part," *J. Vis.*, vol. 9, no. 10, pp. 1–13, 2009.

[38] F. Fang, H. Boyaci, and D. Kersten, "Border ownership selectivity in human
      early visual cortex and its modulation by attention.," *J. Neurosci.*, vol. 29, no. 2,
      pp. 460–465, Jan. 2009.

[39] M. A. Peterson and B. S. Gibson, "Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours," *Percept. Psychophys.*, vol. 56, no. 5, pp. 551–564, 1994.

[40] A. Sarti, R. Malladi, and J. A. Sethian, "Subjective surfaces: A geometric model for boundary completion," *Int. J. Comput. Vis.*, vol. 46, no. 3, pp. 201–221, 2002.

[41] N. Kogo, C. Strecha, L. Van Gool, and J. Wagemans, "Surface construction by a 2-D differentiation-integration process: a neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures.," *Psychol. Rev.*, vol. 117, no. 2, pp. 406–439, 2010.

[42] Y. Pan, M. Chen, J. Yin, X. An, X. Zhang, Y. Lu, H. Gong, W. Li, and W. Wang, "Equivalent Representation of Real and Illusory Contours in Macaque V4," *J. Neurosci.*, vol. 32, no. 20, pp. 6760–6770, 2012.

[43] M. a Cox, M. C. Schmid, A. J. Peters, R. C. Saunders, D. a Leopold, and A. Maier, "Receptive field focus of visual area V4 neurons determines responses to illusory surfaces.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 42, pp. 17095–100, 2013.

[44] M. Chen, Y. Yan, X. Gong, C. D. Gilbert, H. Liang, and W. Li, "Incremental Integration of Global Contours through Interplay between Visual Cortical Areas," *Neuron*, vol. 82, no. 3, pp. 682–694, May 2014.

[45] K. Fukushima, "Restoring partly occluded patterns: A neural network model," *Neural Networks*, vol. 18, no. 1, pp. 33–43, 2005.

Appendix A.1

The following is a direct excerpt of the description of the connection weights in the model by Li (2005).

To describe the synaptic weights, we need some notation. Let β be the direction of the spatial displacement $j - i$ (spatial distance is in the unit of the grid) from one cell $i\theta$ to another $j\theta'$, $d = |i - j|$, and $0 \leq \theta, \theta' < 2\pi$. Let $\theta_1 = \varphi(\theta, \beta)$ and $\theta_2 = \varphi(\beta, \theta')$, where

$$\varphi(x, y) = \begin{cases} x - y & if \ -\pi < x - y \leq \pi \\ x - y + 2\pi & if \ x - y \leq \pi \\ x - y - 2\pi & if \ x - y > \pi \end{cases}$$

Denoting

$$sign(x) = \begin{cases} 1 & x > 0 \\ -1 & otherwise \end{cases}$$

Define $(\theta'_1, \theta'_2) = \big(sign(\theta_1)|\pi - |\theta_1||, sign(\theta_2)|\pi - |\theta_2||\big)$. Then,

$$(\theta_a, \theta_b) = \begin{cases} (\theta_1, \theta_2) & if \ |\theta_1| + |\theta_2| \leq |\theta_1'| + |\theta_2'| \\ (\theta'_a, \theta'_b) & otherwise \end{cases}$$

Now θ_a and θ_b describe the directional angle between the two border segments $(i\theta)$ and $(j\theta')$ and the spatial displacement $j - i$. The directional angles are positive or negative if a right turn or left turn of no more than half a cycle brings the border segments aligned with $j - i$ or $i - j$. Define $\theta'_\pm \equiv \theta_a \pm \theta_b$,

$$\theta_\pm = \begin{cases} \theta'_\pm & -\pi \leq \theta'_\pm \leq \pi \\ 2\pi - \theta'_\pm & \theta'_\pm > \pi \\ -2\pi - \theta'_\pm & \theta'_\pm < -\pi \end{cases}$$

$$J_{i\theta,j\theta'} = \begin{cases} \left(\dfrac{11}{108}\right) exp\left\{-\dfrac{[3-2.5sign(\theta_+)]|\theta_+|}{5\pi} - \dfrac{2\theta_-^2}{\pi^2}\right\} f_1(d), & if \ |\theta_a| \le \pi/11, |\theta_a| \\[2mm] (11/81) \, exp\left\{\dfrac{-3\theta_+}{5\pi} - \dfrac{2\theta_-^2}{\pi^2}\right\} f_2(d), & otherwise, if \ \theta_a, \theta_b \ge 0, \theta_+ \ge \pi/2.01; \\[2mm] (11/81) \, exp\left\{-\left(\dfrac{9\theta_+}{8\pi}\right)^2 - \dfrac{2\theta_-^2}{\pi^2}\right\} f_2(d), & otherwise, if \ \theta_a, \theta_b \ge 0, \theta_+ < \pi/2.01; \\[2mm] (11/81) \, exp\left\{-\left(\dfrac{9\theta_+}{8\pi}\right)^2 - 0.5\left(\dfrac{\theta_-}{\pi/2}\right)^6\right\} f_2(d), & otherwise, if \ \theta_a, \theta_b \ge 0, \theta_+ \ge \pi/2.01; \\[2mm] (11/81) \, exp\left\{-4\left(\dfrac{\theta_+}{\pi}\right)^2 - \dfrac{9\theta_-^2}{\pi^2}\right\} f_2(d), & otherwise, if \ \theta_a, \theta_b \le 0; \\[2mm] (11/81) \, exp\left\{11.5sign(\theta_+)\dfrac{\theta_+^2}{\pi^2} - \dfrac{14\theta_-^2}{\pi^2}\right\} f_2(d), & otherwise, if \ \theta_a \cdot \theta_b \le 0; |\theta_-| < \pi/2.01; \\[2mm] (11/81) \, exp\left\{11.5sign(\theta_+)\dfrac{\theta_+^2}{\pi^2} - \dfrac{15}{4}\left(\dfrac{2\theta_-}{\pi}\right)^6\right\} f_2(d), & otherwise, if \ \theta_a \cdot \theta_b \le 0; |\theta_-| \ge \pi/2.01. \end{cases}$$

where

$$f_1(d) = exp\left[-\left(\frac{d}{9}\right)^2\right],$$

$$f_2(d) = exp\left[-\frac{d}{5}\right],$$

$$f_1(d) = f_2(d) = 0 \ for \ d > 10 \ and \ d = 0.$$

This, though cumbersome, is no more than a piecewise parameterization of the lateral connections with changes in spatial configuration between the underlying border segments. Additionally, the connection strength decays with distance between linked cells, vanishes for distance larger than 10, and is a translation invariant quantity depending only on θ, θ', and the relative displacement $j - i$. Meanwhile, the connections onto the interneurons are

$$W_{i\theta,j\theta'} = c(J_{i(\theta+\pi)\%(2\pi),j\theta'} + J_{i\theta,j(\theta'+\pi)\%(2\pi)})/J_{i,0,i+1_x,0}$$

Where $x\%(2\pi) = x$ if $x < 2\pi$ and $x\%(2\pi) = x - 2\pi$ otherwise, $i + 1_x$ is the grid position one unit displaced from i horizontally, and $c = 0.02646$ usually, except when (θ_a, θ_b) as defined above for the two border segment $(i\theta)$ and $(j(\theta' + \pi)\%(2\pi))$ satisfy $|\theta_a|, |\theta_b| \le \pi/11$, in which case $c = 0.0147$.

Appendix A.2

The time course of the upate rule for $\phi(x, y)$ is shown below.

Table A.1

$n = 1$	$n = 1000$	$n = 2000$
$n = 3000$	$n = 4000$	$n = 5000$
$n = 6000$	$n = 7000$	$n = 8000$
$n = 9000$	$n = 10000$	
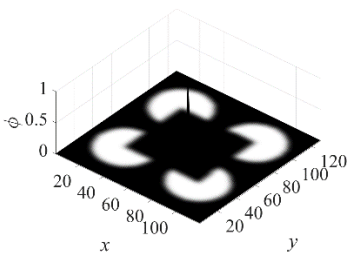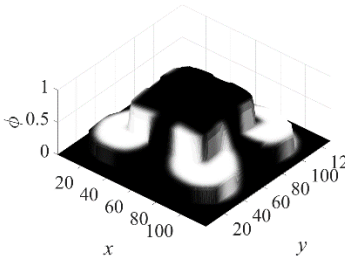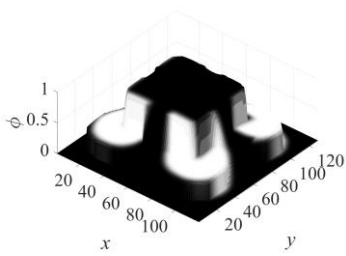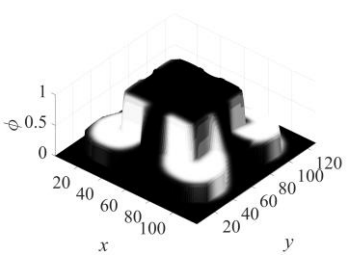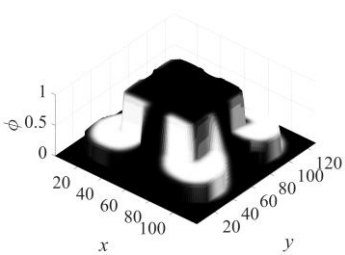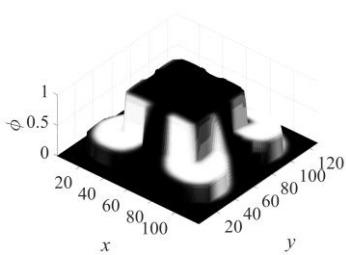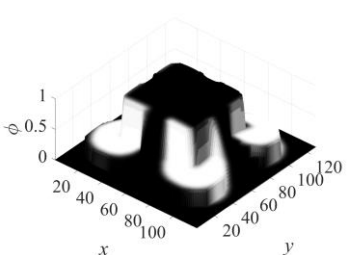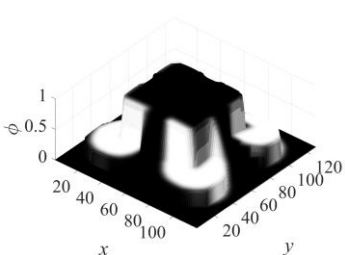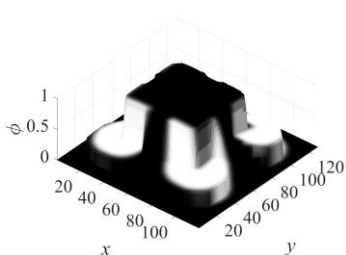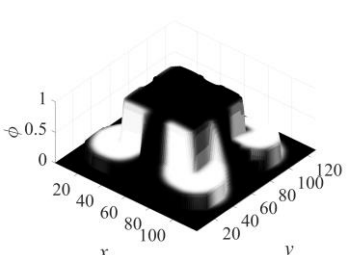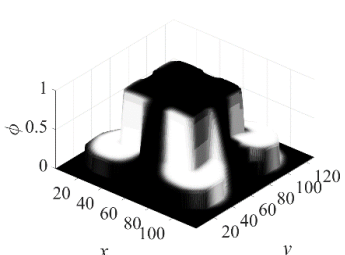
Table A.2

$n = 1$	$n = 1000$	$n = 2000$
$n = 3000$	$n = 4000$	$n = 5000$
$n = 6000$	$n = 7000$	$n = 8000$
$n = 9000$	$n = 10000$	
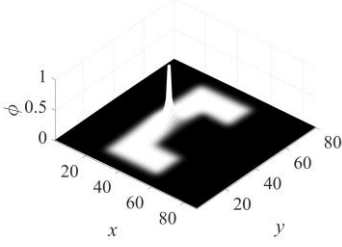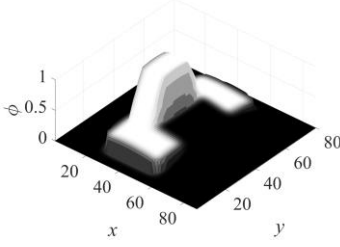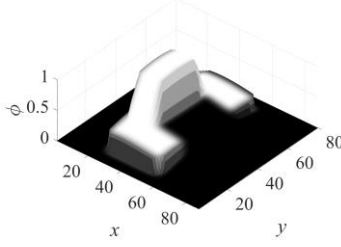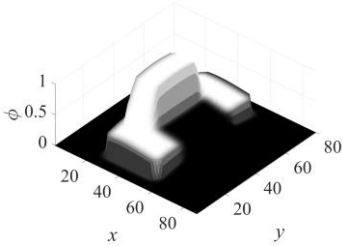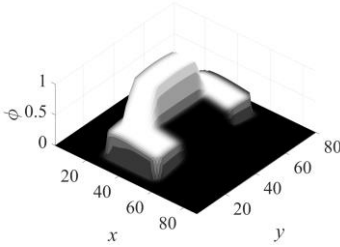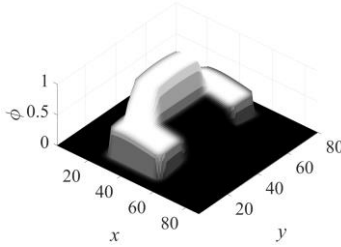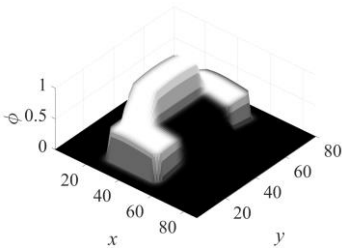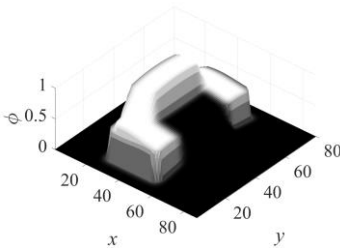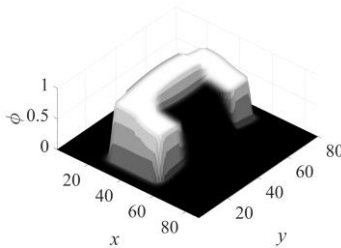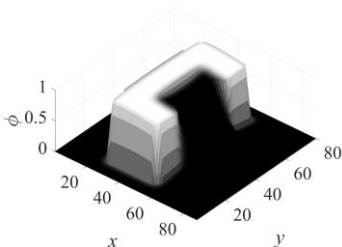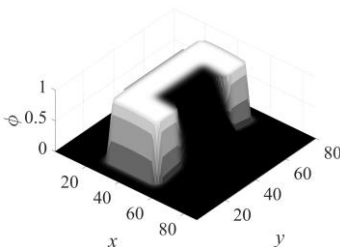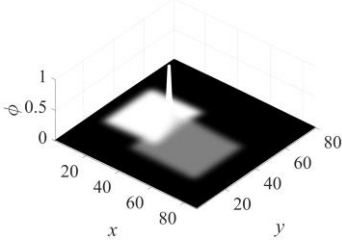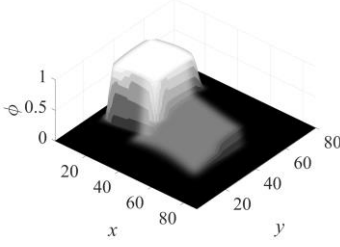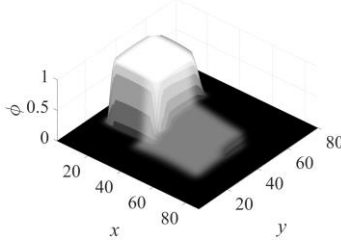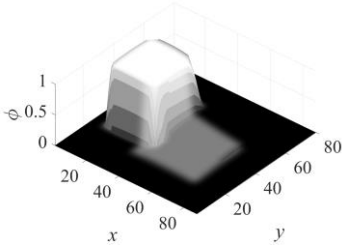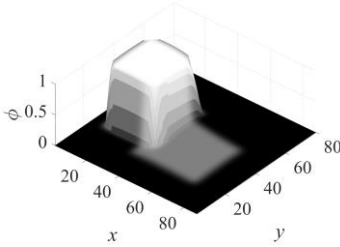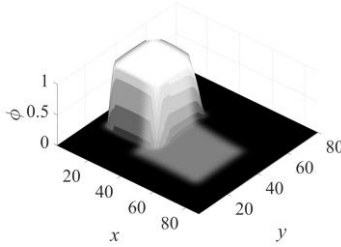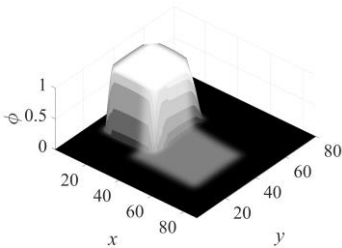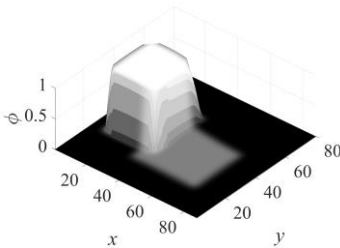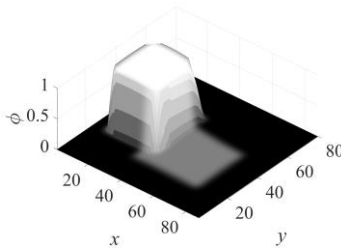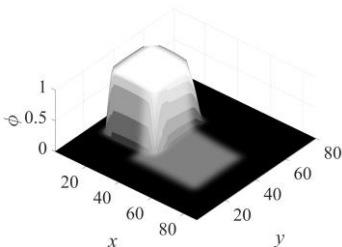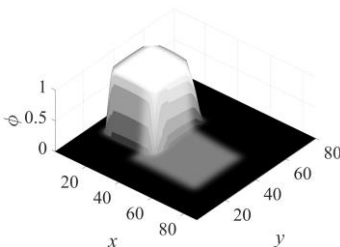
Table A.3

$n = 1$	$n = 1000$	$n = 2000$
$n = 3000$	$n = 4000$	$n = 5000$
$n = 6000$	$n = 7000$	$n = 8000$
$n = 9000$	$n = 10000$	

Table A.4

$n = 1$	$n = 1000$	$n = 2000$
$n = 3000$	$n = 4000$	$n = 5000$
$n = 6000$	$n = 7000$	$n = 8000$
$n = 9000$	$n = 10000$	

Publications and Achievements

Journals

1. <u>Zaem Arif Zainal</u>, Shunji Satoh : "A border-ownership model based on computational electromagnetism", Neural Networks, vol.99，pp.114-122, 2018.

International Conferences

1. <u>Zaem Arif Zainal</u>, Shunji Satoh : "Formulation of Border-Ownership Assignment in Area V2 as an Optimization Problem", Proceedings of the 24th International Conference On Neural Information Processing, Neural Information Processing, vol.3, pp. 856-866, 2017.

Domestic Conferences

1. <u>Zaem Arif Zainal</u>，佐藤俊治：＂中低次の視覚計算問題を電場と電位の定理を用いて解く＂，第 25 回日本神経回路学会全国大会，2015 年 9 月．
2. <u>Zaem Arif Zainal</u>，Shunji Satoh : "Applying an electromagnetic theorem to solve the Border-Ownership assignment problem"，ニューロコンピューティング研究会，2016年3月．
3. <u>Zaem Arif Zainal</u>，佐藤俊治：
"境界帰属方向と物体の重なり順序を同一の計算理論により再現する視覚モデル"，視覚学会2018年冬期大会，2018年1月．

Oral Presentations

(1) <u>Zaem Arif Zainal</u>, "Border ownership
計算問題を電場と電位の定理を用いて解く"，東北大学電気通信研究所共同プロジェクト（招待講演），2016年2月．
(2) <u>Zaem Arif Zainal</u>,
"中低次の視覚計算問題の定式化とモデル比較"，東北大学電気通信研究所共同プロジェクト（招待講演），2017年2月．