Neurocomputing 149 (2015) 546–558

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Neural construction of 3D medial axis from the binocular fusion of 2D MAs

Wei Qiu, Yasuhiro Hatori, Ko Sakai *

Department of Computer Science, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8347, Japan

A R T I C L E I N F O

Article history:
Received 12 July 2013
Received in revised form
29 April 2014
Accepted 4 August 2014
Communicated by Paolo Remagnino
Available online 21 August 2014

Keywords:
Vision
Shape perception
Cortical representation
Medial axis
Stereo

A B S T R A C T

The perceptual constancy of shape, including view invariance, is an amazing property of the visual system. Cortical representation by the medial axis (MA) is an attractive candidate for maintaining the constancy of a wide range of arbitrary shapes. Recent physiological studies have reported that neurons in the primary visual cortex (V1) show a response to two-dimensional (2D) MAs, and those in the inferior temporal cortex (IT) are selective to three-dimensional (3D) MAs. However, little is known about the neural mechanisms underlying the transformation of 2D to 3D MAs. As a first step toward investigating the cortical mechanism, we have proposed as a hypothesis that a pair of monocular 2D MAs is fused to generate a 3D MA. We examined the computational plausibility of the hypothesis; specifically, whether an energy-based fusion model is capable of generating 3D MAs. We generated blob-like, physiologically plausible 2D MAs, and used a standard energy model to detect the disparity between a pair of 2D MAs. The model successfully generated 3D MAs for a variety of objects that included typical shape characteristics. A reconstruction test showed that the computed 3D MAs captured the essential structure of the objects with reasonable accuracy and view invariance. These results indicate that the fusion of monocular blob-like 2D MAs is capable of generating a reasonable 3D MA within the framework of the energy model.

1. Introduction

Robust perception of the shape of objects is an amazing property of the visual system. Although the view and the size of an object on a retinal image change dramatically as we see the object from different directions and distances, our visual system perceives a stable, invariant shape for the object. The representation of shape in the visual cortex should play a crucial role in realizing such invariance in shape perception. An object-centered representation that describes shape as a spatial arrangement of parts has been supported widely by psychological and physiological studies [1–3], as it has the ability to establish the perceptual constancy of shape, including view and distance invariance. The medial axis (MA) is considered suitable for a parts-based representation among theorists [4,5]. MA representation encodes each part of the object with a medial line that is derived from the local symmetry of the part. This representation, based on an object-centered coordinate, is independent of view and capable of describing shape efficiently using two types of parameters: the spatial arrangement and relative length of the axes corresponding to the parts [4,6,7]. MA is an attractive candidate for the cortical representation of shape, as a robust and efficient coding scheme [8].

Recently, Hung et al. showed that a number of neurons in the inferior temporal cortex (IT) encode three-dimensional (3D) MA configurations, supporting the idea that the MA plays a critical role in the representation of shape in the ventral pathway [9]. IT has been reported to encode the 3D structure of shape [10,11], but little was known about the representation scheme for 3D shapes. The selectivity for 3D MA configurations reported recently in IT has provided crucial direct evidence to support MA coding for the cortical representation of shape. A recent fMRI study has also reported the cortical representation of MA structure in the ventral stream [12]. However, the computational processes that constitute the 3D MA along the ventral pathway remain unknown. One of the keys to understanding these processes lies in the lower cortex: cells in the primary visual cortex (V1) show strong responses to the MA of a textured figure [13,14]. Computational studies have shown that the MA response in V1 can be generated by simultaneous arrival of traveling spikes that are initiated by nearby V1 cells [14], or from onset synchronization of border-ownership (BO)-selective cells in V2 [15,16]. These computational studies have also reported that the generated MA encodes arbitrary two-dimensional (2D) shapes. These studies note that the MAs were not like thin skeletons as previous studies have assumed, but rather, the MAs were elongated blobs with spatial extent. This

---

* Corresponding author. Tel.: +81 29 853 5348.
E-mail address: sakai@cs.tsukuba.ac.jp (K. Sakai).

http://dx.doi.org/10.1016/j.neucom.2014.08.019
0925-2312/© 2014 Elsevier B.V. All rights reserved.

blob-like MA is expected to be robust for 3D construction. Because the structure of skeletons is sensitive to the direction of view (binocular difference) and noise in the contours, small changes in view and contour dramatically alter the structure of skeleton-like MAs, leading to erroneous stereo-matching. However, blob-like MAs are expected to be insensitive to such changes [17]. Investigating the fusion of blob-like MAs rather than conventional skeleton-like MAs is essential. The intermediate areas of the ventral visual pathway such as V4 are known to play a crucial role in the binocular fusion of object shapes [18,19]. A certain translation function that takes place along the ventral pathway may contribute to the construction of the 3D MA observed in IT from the 2D MAs observed in V1.

We investigated the cortical mechanisms underlying the construction of 3D-shape representation, by focusing on blob-like 2D MAs and their fusion along the ventral pathway. Fusion of 2D MAs based on their disparity is a plausible candidate mechanism for filling the gap between the 2D MA in the primary cortex and the 3D MA in the higher cortex. It is conceivable that the 2D MAs resulting from the left and right retinal images are fused in an intermediate-level area by a process based on disparities in the 2D MAs, thereby establishing a 3D MA in IT. An alternative mechanism for the construction is that the MA responses in V1 are binocular with absolute disparity, and are thus “3D MA segments”. The 3D MA segments in V1 would then be integrated along the visual pathway to establish a global 3D MA with relative disparities in IT. Although a number of V1 cells are selective to the binocular disparity of contours, it is not at all certain whether cells responding to MAs are selective for the binocular disparity of the local MA. V1 cells could respond to the depth of contours, but not necessarily to that of the MA. Specifically, the depths of both sides of an object as well as its MA are generally different. This concept is illustrated by a cuboid with a different depth for each side of the object; for example, the left side is nearer and the right side is farther (see Fig. 1A). Although the depths of these sides can be determined correctly, the depth of the MA is inherently ambiguous; the MA could be located anywhere between the two sides and there is no way to determine its depth from the depth of the sides. On the other hand, in the former case involving 2D MAs, the local disparities between the 2D MAs could be integrated without ambiguity (see Fig. 1B). This idea appears to be consistent with the tuning of three-dimensional orientation in the macaque V4 [19]. In the present study, we focused on the fusion of monocular 2D MAs that are formed in V1, and are fused along the ventral pathway based on the disparities between the axes, to generate a 3D MA in IT.

Physiological evidence for the generation process of a 3D MA has not been available. As a first step toward investigating our hypothesis, we conducted computational studies to determine whether the fusion of monocular, blob-like (physiologically plausible) 2D MAs is capable of generating a 3D MA, and how accurately this method would work. Specifically, we constructed a fusion model based on a standard energy model [20] that is thought to capture the essential functions of physiological properties in early- to intermediate-level visual areas. We examined whether the model is capable of generating a correct 3D MA, and whether the computed 3D MA captures the essential structure that is sufficient for the reconstruction of a 3D shape. Our simulation results showed that the model was capable of generating 3D MAs for a variety of shapes including those of natural objects. The results also showed that the reconstruction of 3D shapes based on the computed 3D MAs was successful, with similar levels of accuracy for various shapes with different degrees of shape complexity, which is one of the most remarkable features of the visual system. Furthermore, we tested view invariance of the model in terms of the reconstruction error. Similar reconstruction errors were observed for images from different views, suggesting that the representation of a 3D MA from the fusion of 2D MAs has invariance to rotation. View invariance has been reported in MA-selective cells in IT [9]. Our results indicate that the energy-based fusion of monocular blob-like 2D MAs is capable of generating a 3D MA with robustness in terms of shape complexity and view invariance. Therefore, the generation of a 3D MA from the fusion of 2D MAs is a plausible candidate for the cortical mechanisms underlying the representation of 3D shape.

2. The model

To investigate whether the fusion of physiologically plausible 2D MAs is capable of generating a correct 3D MA, and whether the computed 3D MA captures the structure essential for the reconstruction of 3D shape, we constructed a computational model and conducted simulations. An outline of the model is illustrated in Fig. 1. The model is composed of two stages: (i) the detection of monocular 2D MAs based on the distances from surrounding contours, and (ii) the generation of a 3D MA from the disparities between the two 2D MAs (Fig. 1B). A unit in the first stage computes the distances between the unit and the points on the contours surrounding the unit, and evaluates how much the unit is similarly distant from the surrounding contours by taking pairwise differences between the distances. Units with small differences (similar distances) tend to be located around local symmetry axes, thus their locations are highly likely a part of the 2D MA. The second stage fuses a pair of 2D MAs using a standard energy model to generate a 3D MA. Note that the model includes neither the representation nor the reconstruction of a 3D object. We conducted the reconstruction in Section 3 solely for the evaluation of the computed 3D MA.

2.1. The detection of 2D MA

A computational study by Hatori and Sakai has shown that onset synchronization of BO-selective cells appears to generate V1 activities in response to 2D MAs [16,21]. BO-selective cells on the contour of a figure depolarize if the figure is located on their preferred side [22]. The spikes from BO-selective cells, which are initiated at the same time and travel at the same speed, reach the center of the figure at the same time. Temporal integration of the traveling spikes would result in strong responses of cells located at the center of the figure and along the axes of local symmetry, generating the V1 activity corresponding to the MA. The magnitude of the activity depends on how much the cell is similarly distant from the contours. Taking into account the essence of their idea, the present model computes the possibility of being a 2D MA based on distances from the surrounding contours. Although Hatori's model was capable of processing multiple objects, we limited our model to deal with a single object for the sake of simplicity. We computed an index that describes how much a cell is similarly distant from the contours. If the value of the index exceeds a certain threshold, we consider it as an indication of the MA.

The input to the model was a pair of stereo images with a spatial resolution of 200×200 pixels (considered as $5 \times 5^\circ$ of visual angle). To evaluate the similarity of distances from nearby contours, we measured the distance, $dist(p, q_i)$, between a point within a figure, p, and every point on the contour, q_i

$$dist(p, q_i) = \| p - q_i \|, \tag{1}$$

where $\| . - . \|$ represents the Euclidean distance between the two points. The distances between p and q_i were measured for every 5°