Unique contributions of medial axis structure in human object recognition

Vladislav Ayzenberg (vayzenb@emory.edu)
Department of Psychology, Emory University
36 Eagle Row Atlanta, GA 30322, USA

Stella F. Lourenco (stella.lourenco@emory.edu)
Department of Psychology, Emory University
36 Eagle Row Atlanta, GA 30322, USA

Abstract:

Despite great strides in computer vision, computational models of object recognition remain unable to wholly account for organisms' ease at identifying objects across changes in view or their ability to categorize novel exemplars. One important feature that is frequently overlooked in these models is the role that an object's spatial configuration, or 'structure', plays in determining object identity and category. Across three experiments, we tested whether a model of structure, based on an object's medial axis, characterizes human object recognition independently of other models of vision. We found that human perceptual similarity judgments for novel three-dimensional (3D) objects was predicted by the medial axis similarity between objects, and participants preferentially categorized objects by their medial axes across changes to the object's surface form. Importantly, we found that this pattern of responses could not be accounted for by other visual properties. These results suggest that human object representations incorporate the medial axis and that the medial axis may play a crucial role in object recognition.

Keywords: Object recognition; Structure; Medial Axis; Vision, computation

Introduction

Object structure is an important, but poorly understood, property of objects. Consider an everyday object such as a chair. The image of a single chair on the retina varies across viewpoints, and the specific components of a chair (e.g., seat, legs, and back) vary in shape across different exemplars. Because of the variability in a chair's appearance, an optimal model of vision might instead identify 'chair' by a common structure (i.e., two perpendicular surfaces above some posts) instead of its image-level or component properties. Indeed, changing an object's structure without changing its components, disrupts object recognition more than changing the components while leaving the structure intact (Cave & Kosslyn, 1993). Moreover, work from patients with integrative agnosia suggests that their inability to identify objects is rooted in a deficit in perceiving the structure of objects, rather than the image-level or component properties (Behrmann, Peterson, Moscovitch, & Suzuki, 2006). Although these studies support the importance of structure in object recognition, structure is rarely incorporated into models of object recognition. In models that do incorporate structure, it is merely a means by which contours, or components, are configured, instead of a diagnostic cue to identity.

One model of structure that is both biologically plausible and may support object recognition is based on the medial axis of objects (Blum, 1973). The medial axis is a model that describes the structure of an object via its internal skeleton. It provides a low-dimensional, but precise, description of an object's overall shape by specifying the relations between the contours and components of an object. Although accumulating evidence suggests that humans extract the medial axis of shapes (e.g., Firestone & Scholl, 2014), and that the medial axis is represented neurally in the visual system (e.g., Hung, Carlson, & Connor, 2012), few studies have tested whether humans rely on the medial axis for object recognition while controlling for other visual properties.

Here we examined whether humans represent object structure via the medial axis by testing whether the medial axis serves as a diagnostic cue of object identity. To this end, we tested whether human object similarity judgments scaled with the medial axis of objects (Experiments 1 and 2) and whether humans preferentially categorized objects by their medial axis across surface-level changes (Experiment 3). Importantly, to isolate the unique contributions of the medial axis in recognition, we tested it against models of low-level image similarity (i.e., Gabor-jet model), mid-level visual similarity (i.e., GIST), high-level object similarity (top fully connected layer of alexNet), as well as other models of structure (i.e., coarse spatial relations).

Experiment 1

If the medial axis is diagnostic of object identity, then the prediction is that, as the medial axes of two objects become more similar, participants will also perceive the objects as being more similar to one another. Because the medial axis is a formalized model of object structure, it is possible to measure whether human judgments of object similarity scale with 3D medial axis similarity. To test this hypothesis, we generated a novel set of 3D objects with unique medial axes and tested whether human perceptual similarity judgments of these objects were predicted by the medial axis independently of other models of vision.

Methods

One-hundred and fifty novel 3D objects consisting of 30 medial axes were generated (see Figure 1a). Participants (n = 42) were administered a discrimination task in which they were shown images of two objects presented simultaneously in one of three depth orientations (-30°, 0°, +30°), with either the same or different medial axes. Participants were instructed to decide whether the two images showed the same or different object.

Results

Medial axis similarity was a significant predictor of participants' judgments, $r = .31$, $p < 0.001$. As the similarity of medial axes between objects increased, participants were more likely to judge the objects as the same (See Figure 1b). These results suggest that participants extract the 3D medial axis structure of novel objects, even when they are presented as still images. However, one important question is whether this result could be explained by another model of vision that does not incorporate structural information. For instance, it is possible that objects with similar medial axes also have similar image-level properties, such that the degree of image-similarity could account for the effect between medial axis similarity and human performance. To test this possibility, we compared participants' judgments to models of visual dissimilarity as computed by the Gabor-jet model, the GIST model, as well as the best performing layer of alexNet, a convolutional neural net pre-trained to classify objects. Although these models were independently predictive to varying degrees (βs = -0.23 to 0.42), when all of the models were entered into a regression analysis, the medial axis explained unique variance in participants' judgments, $\beta = .26$, $p < 0.001$. These findings suggest that the medial axis plays a unique role in determining object similarity.

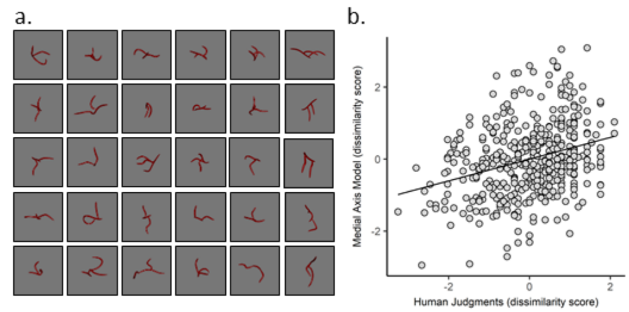

Figure 1. (a) A subset of procedurally generated objects with different medial axes. (b) Scatterplot showing the relation between medial axis similarity of the tested objects and human perceptual judgments of these objects (units presented as standard scores).

However, a question raised by this experiment is whether participants represented object structure by their medial axes or simply by the coarse spatial relations between object parts. In this stimulus set, objects with similar spatial relations would also have more similar medial axes than objects with different spatial relations. Thus, the relation between medial axis similarity and human perceptual judgments could reflect the co-variation between the medial axis and an object's coarse spatial relations. If so, this would suggest that participants do not represent object structure with the precision of the medial axis, but instead, they do so at a coarser level that is only sensitive to qualitative changes in component positions.

Experiment 2

A model based on coarse spatial relations suggests that object structure is represented by the approximate arrangement of component parts (e.g., two components below a third; see Figure 2a), which does not include precise information about part position within an arrangement (Biederman & Gerhardstein, 1993). Here we tested whether human perceptual similarity judgments could be accounted for by a coarse spatial-relations model. Based on this model, only qualitative changes to the overall spatial arrangement of an object's component parts (e.g., moving a component from the bottom to the side) would influence recognition. By contrast, if participants represent structure by the medial axis, then a proportional change to the medial axis would elicit a proportional decrease in recognition, even holding coarse spatial relations constant.

Methods

A subset of objects from Experiment 1 were generated to have the same coarse spatial relations but to differ by six increments of medial axis similarity (0%, 10%, 20%, 30%, 40%, 50% difference; see Figure 2a). Participants (n = 40) completed a discrimination task where they were shown two objects presented simultaneously in one of three depth orientations (30°, 60°, 90°) and, as in the previous experiment, were instructed to judge whether objects were the same or different in their coarse spatial relations. Participants were given training to ignore the precise positions of object parts so as to judge similarity on the basis of coarse spatial relations.

Results

Participants performed significantly above chance (0.50) at every level of medial axis change demonstrating that they followed the task instructions to identify objects by their coarse spatial relations. However, performance less accurate as a function of medial axis change (*ps* < 0.001; see Figure2b). Importantly, mixed-effect models confirmed that the medial axis remained predictive even when accounting for other models $\chi^2(1) = 24.01$, *ps* < 0.001 (Figure 2b). Thus, although the coarse spatial relations between the object's parts remained constant, and participants were trained to ignore the precise position of object parts, performance was nevertheless modulated by the degree of medial axis change.

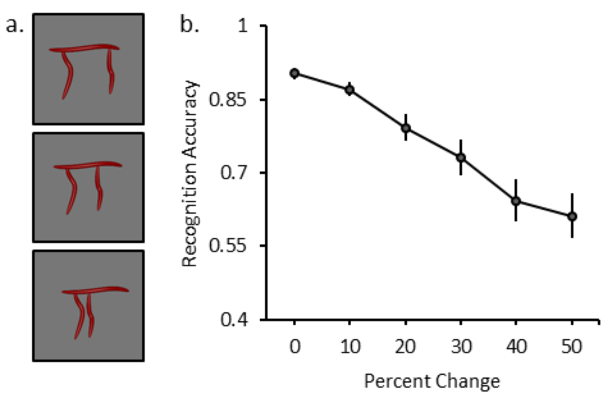

Figure 2. (a) Objects with the same coarse spatial relations but different medial axes. (b) Participants' recognition accuracy (proportion correct) as a function of medial axis change.

Experiment 3

In the previous experiments, it was found that humans extracted object structure according to the precise scale of the medial axis, not just the coarse spatial relations of the objects. However, it remains unclear whether the medial axis is a privileged source of information for identifying objects when compared to image-level or component properties. Specifically, if the medial axis is a diagnostic cue of object identity, then the prediction is that humans should preferentially categorize objects by their medial axes rather than by other types of visual properties. To test this hypothesis, the medial axis of the object was pitted against surface form changes that change the visible shape of the object without changing the medial axis.

Methods

A subset of objects from Experiment 1 were rendered with five surface forms (Figure 3a). Surface forms were independently validated to ensure that they scaled with image-level similarity and were comprised of unique component parts. Participants (n = 41) completed a match-to-sample task where they were required to choose which of two objects were most likely to be in the sample category as the sample object. The choices could match the sample in medial axis, surface form, or have no match. Participants were tested with each cue in isolation, as well as a conflict trial where one object would match in medial axis, but not surface form, and the other would match in surface form, but not medial axis. Objects could be presented in one of three depth orientations (30°, 60°, 90°).

Results

Participants successfully categorized objects by either their medial axes or surface forms, as indicated by their above chance (0.50) performance when each was presented in isolation (*ps* < 0.001; Figure 3b). Crucially, however, when medial axis conflicted with surface form, participants categorized objects by their medial axis, not surface form (*ps* < 0.001; see Figure 3b-c). Thus, although surface forms varied in both their image-level properties and component parts, these data suggest that participants prioritized the medial axis in object categorization.

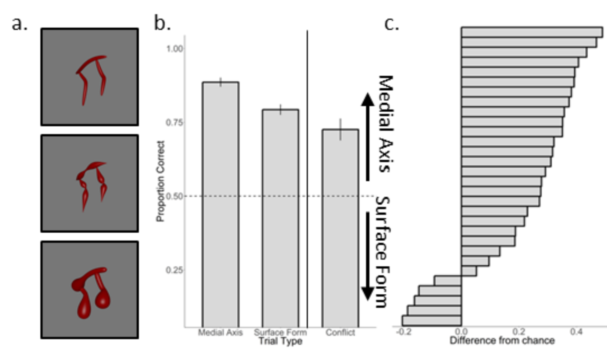

Figure 3. (a) Objects with the same medial axis but different surface forms. c) Participants categorization performance in medial axis match, surface form match, and conflict trials. (d) Histogram distribution of each participant's response on the conflict trial. A score greater than zero indicates greater medial axis weighting.

Discussion

Although structure has been described as a necessary element of object recognition models (Barenholtz & Tarr, 2006) and studies have hinted at its importance for recognition

(Behrmann et al., 2006), few studies have systematically investigated the role of structure in object recognition. Here we tested whether humans represent and recognize objects via a model of structure known as the medial axis, or 'shape skeleton'. We found that the medial axis was predictive of perceived object similarity and category membership, suggesting that it is used as a diagnostic cue of identity. Importantly, the predictive power of the medial axis was not explained by low-level or high-level models of vision, nor other models of structure, suggesting that it plays a unique role in object recognition. Together, these experiments shed light on a critical, but rarely studied, property of objects that has the potential to explain how organisms achieve rapid and invariant object recognition.

Acknowledgements

This work was partially supported by a National Institutes of Health (NIH) institutional training grant (T32 HD071845) and by the Emory PERS program awarded to S.F.L.

References

Barenholtz, E., & Tarr, M. J. (2006). Reconsidering the role of structure in vision. *Psychology of learning and motivation, 47*, 157-180.

Behrmann, M., Peterson, M. A., Moscovitch, M., & Suzuki, S. (2006). Independent representation of parts and the relations between them: evidence from integrative agnosia. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1169-1184.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*(6), 1162-1182.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of Theoretical Biology, 38*(2), 205-287.

Cave, C. B., & Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception, 22*(2), 229-248.

Firestone, C., & Scholl, B. J. (2014). "Please tap the shape, anywhere you like" shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science, 25*(2), 377-386.

Hung, C.-C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron, 74*(6), 1099-1113.