Contents lists available at ScienceDirect

Cognition

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

Original Articles

Local contour symmetry facilitates scene categorization

John Wilder^{a,*}, Morteza Rezanejad^b, Sven Dickinson^a, Kaleem Siddiqi^b, Allan Jepson^a, Dirk B. Walther^a

^a University of Toronto, Toronto, Canada ^b McGill University, Montreal, Canada

A R T I C L E I N F O A B S T R A C T Keywords: Scene perception Symmetry Perceptual grouping Scene categorization A B S T R A C T Human visual system that are applicable to line drawings. Previous work (Wilder, Dickinson, Jepson, & Walther, 2018) showed that standard local features of individual contours, or junctions between contours, do not account for this rapid classification ability but, rather, the relative placement of these contours appeared to be important. Here we provide strong support for this observation by demonstrating that local ribbon symmetry between neighboring pairs of contours facilitates the categorization of complex real-world environments. To this end, we introduce a novel computational approach, based on the medial axis transform, for measuring the degree of local ribbon symmetry in a line drawing. We use this measure to senarate the contour nivels for a given scene into the

introduce a novel computational approach, based on the medial axis transform, for measuring the degree of local ribbon symmetry in a line drawing. We use this measure to separate the contour pixels for a given scene into the most ribbon symmetric half and the least ribbon symmetric half. We then show human observers the resulting half-images in a rapid-categorization experiment. Our results demonstrate that local ribbon symmetry facilitates the categorization of complex real-world environments. This is the first study of the role of local symmetry in inter-contour grouping for human scene classification. We conclude that local ribbon symmetry appears to play an important role in jump-starting the grouping of image content into meaningful units, even in flashed presentations.

Vision feels natural and effortless: light hits our eyes, and we appear to understand our real-world environment almost instantly. Yet, the neural computations underlying visual perception are not yet fully understood. Great progress has been made in identifying the basic visual features that are extracted from the visual input, such as oriented edges (Hubel & Wiesel, 1963), corners (Link & Zucker, 1988), spatial frequency content (Oliva & Torralba, 2001; Sachs, Nachmias, & Robson, 1971), and disparity (Blakemore & Hague, 1972). At the other end of the processing pipeline, brain areas have been identified that are dedicated to the processing of objects (Malach et al., 1995), faces (Kanwisher, McDermott, & Chun, 1997), or places (Epstein & Kanwisher, 1998). The processes involved in the intermediate-level grouping of visual features is less well understood. Gestalt grouping rules, such as good continuation, symmetry, or similarity, were proposed as a qualitative account for how edge segments or shape parts are grouped into larger structures (Koffka, 1922, 1935; Kohler, 1947; Wertheimer, 1938). However, there is so far no mechanistic, quantitative model of how Gestalt rules are implemented and used to facilitated visual perception of complex real-world scenes.

Following its postulation as one of the Gestalt laws of perceptual organization, symmetry has been investigated as a grouping principle in both human and computer vision (Feldman & Singh, 2006; Kanizsa and Gerbino, 1976; Koffka, 1935; Liu, Hel-Or, Kaplan, & Van Gool, 2010; Pizlo, 2014; Stahl & Wang, 2008; Wagemans, 1993; Wagemans et al., 2012). We define symmetry as a redundancy in the shape of an object or its projection onto the image plane due to a similarity between subpieces of a larger part. In the context of an image, this can include mirror symmetry, where part of the image is reflected across an axis, rotational symmetry, where a section of the image is a copy of another section but at a different orientation, as well as translational symmetry, where a section of an image is a translated copy of another section. These forms of symmetry can either apply to part of an image (local symmetries) or to the entire image (global symmetry). Local symmetries do not need to apply to an entire object. In fact, a single part of an object may be locally symmetric. For example, consider a building with Greek columns. If the building is viewed from an oblique angle, the projection of the building onto the image plane does not necessarily result in a symmetric image. However, the projection of a single pillar

* Corresponding author.

E-mail address: jdwilder@cs.toronto.edu (J. Wilder).

https://doi.org/10.1016/j.cognition.2018.09.014







Received 3 May 2018; Received in revised form 20 September 2018; Accepted 22 September 2018 0010-0277/@ 2018 Elsevier B.V. All rights reserved.

in this view may still be locally symmetric.

We can consider many different types of local symmetry. Medial symmetry applies to those types of local symmetry that are the result of a reflection across a curved axis. This is a type of mirror symmetry on a local scale. The medial axis transform provides a way to capture medial symmetry (Blum, 1973). The intuitive idea behind medial symmetry is that the boundary of a shape can be formed by sweeping a disk along a suitable path (the medial axis).¹ As a special case of medial symmetry, ribbon symmetry occurs when the radii of the medial disks remain constant along the axis. Parallel lines are one example of ribbon symmetry. Another example is a river of constant width, which winds through a field. In this paper we constrain our discussion to ribbon symmetry.

Since first theorized as a grouping principle by the Gestalt psychologists, there has been a long history of research on global symmetry and its influence on human vision (for reviews, see Wagemans (1997) and Wagemans et al. (2012)). In addition to global symmetry, local symmetry influences several aspects of human vision (Burbeck & Pizer, 1995; Chan, Stevenson, Li, & Pizlo, 2006; Feldman & Singh, 2006; Firestone & Scholl, 2014; Li & Pizlo, 2008; Kovacs & Julesz, 1994; Marr & Nishihara, 1978). Machilsen, Pauwels, and Wagemans (2009) showed that mirror symmetric shapes are easier to detect than asymmetric shapes when embedded in a noise field. Wilder, Feldman, and Singh (2016) showed that medial symmetry also plays a role in object detection, where shapes with a simpler medial axis structure were more easily detected. The medial axis also helps explain human shape categorization (Wilder, Feldman, & Singh, 2011). This work is important for the current study as it demonstrates that symmetry has an influence on visual processing both prior to object detection and post detection during classification. As images of real-world scenes are rarely globally symmetric, we will focus on local symmetry in the current study. Specifically we will base our approach on the medial axis. In addition to the behavioral studies mentioned previously, there is also neural evidence of the importance of the medial axis for human vision. Some neurons in visual cortex have been found that respond highly when their receptive fields are centered on a medial axis (Lee, 1996). Additionally, the axial structure of a shape is represented in visual cortex (Lescroart & Biederman, 2013).

While some previous work computes medial axes only on closed object silhouettes, we here apply our consideration of local symmetry beyond individual objects through a computational approach which is applicable to entire scenes. We took inspiration from work that computed medial axes for entire images (Lee, Fidler, & Dickinson, 2013; Levinshtein, Sminchisescu, & Dickinson, 2013; Narayanan & Kimia, 2012; Tamrakar & Kimia, 2004). These methods are influenced by color and texture. Since we wish to focus on shape properties, namely local symmetry, we describe a method for measuring symmetry in line drawings of natural scenes. Other previous work that computes medial axes for entire images, specifically focusing on shape properties, is reviewed in Leymarie and Kimia (2008). Some of this work even directly analyzed scene layout (Bruck, Gao, & Jiang, 2007; Van Tonder, Lyons, & Ejima, 2002), however, using scenes from an overhead view, as opposed to frontally viewed scenes in our work. We restrict our attention to ribbon symmetry, which, recall, when a pair of contours has constant separation. As perfect symmetry in real-world images is rare, especially when discritized in a pixel grid, we develop a measure of the degree to which two contours have roughly constant separation. Specifically, we investigate whether or not pairs of contours that are ribbon symmetric play a significant role in human scene classification in the challenging context of briefly flashed presentations.

Humans have the ability to classify a photograph of a natural scene after a brief exposure (Potter & Levy, 1969; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). In spite of a sizable literature studying this phenomenon (Delorme, Richard, & Fabre-Thorpe, 2000; Oliva & Schyns, 2000; Torralba & Oliva, 2003; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010), there is no consensus on what accounts for this ability. Rapid scene perception does not require color photographs; observers can rapidly classify line drawings of real-world scenes (Biederman, Mezzanotte, & Rabinowitz, 1982; Walther & Shen, 2014). Furthermore, recent work demonstrates that scene content is carried primarily in the high spatial frequencies (Berman, Golomb, & Walther, 2017). In fact, the high-pass images used in the latter study closely resemble line drawings. Additionally, Walther, Chai, Caddigan, Beck, and Fei-Fei (2011) found that photographs and line drawings result in similar neural patterns, showing that the underlying category-specific representations are similar. Thus, we choose to use line drawings, without fear of loss of generality of our results, in order to allow us to study the influence of shape alone, without the confounds introduced by color and texture.

Various properties of the contours that are preserved in line drawings have been assessed for their role in scene perception. Walther et al. (2011) identified contour length as a critical feature, while Walther and Shen (2014) showed that curvature and contour junctions underlie categorization of line drawings and photographs of scenes. Indeed, contour junctions have been shown to play an important role in object recognition (Attneave, 1954; Biederman, 1987). In a recent study, using an experimental design similar to the present study, we directly tested the role of contour junctions in rapid scene classification (Wilder, Dickinson, Jepson, & Walther, 2018). The results showed that scenes from which the junctions were removed were more easily classified than scenes form which the middle sections, between junctions, were removed. The local relationships between elongated sections play at least as important of a role in scene perception, as opposed to intersection between contours, hinting at the importance of local symmetry relationships. Since we did not directly measure or manipulate symmetry in Wilder et al. (2018), we were unable to draw any conclusions about its importance for scene classification. Here we explicitly measure local ribbon symmetry in complex, real-world scenes and test for its importance for scene categorization.

We measure local ribbon symmetry by extracting the symmetric axes from line drawings of entire real-world scenes. Symmetric axis representations are often defined mainly for individual object silhouettes (Blum, 1973; Siddiqi, Shokoufandeh, Dickinson, & Zucker, 1999, for example). Here we apply the concept of a symmetric axis to entire scenes. We use the symmetric axis to assign a symmetry score to each contour pixel in an image. We then split the image into two half-images, one containing the high- and one the low-symmetry half of the pixels. The two half-images have no contour pixels in common and, when combined, result in the original, intact line drawing. We use both versions, along with the intact line drawings, in a categorization experiment. If symmetry is indeed a strong cue for scene processing, then the symmetric half-images should be more easily classified than the asymmetric half-images.

1. Methods

1.1. Scoring symmetry

In order to measure the degree of ribbon symmetry present along individual contours in a line drawing of a scene, we have devised a measure of symmetry based upon a modern formulation of the medial axis (Siddiqi, Bouix, Tannenbaum, & Zucker, 2002). Details of the geometry of the medial axis are reviewed in Appendix A. The code for computing the skeletons is available at https://github.com/ mrezanejad/AOFSkeletons.

To begin, we describe a method for computing a flux-based skeleton. Our examples will all be based upon the line drawing in Fig. 1 (top right). We consider the line drawing of a scene as a binary image

¹ A closely related representation in 3-D is Binford's *generalized cylinder* which sweeps a 2-D cross section along a 3-D space curve (Binford, 1971).



Fig. 1. A photograph of an office scene (A), along with its artist-traced line drawing (B), outward distance transform (C), average outward flux (AOF) map (D), flux skeletons (E), and symmetry score at each contour pixel (F). In C, color represents distance to nearest contour: blue is a small distance, and red is a large distance. In E, color simply denotes skeletons for different closed regions. In F, blue represents a weaker symmetry score and red represents a stronger symmetry score. Note that pairs of long smooth parallel contours, such as down the side of the chair, receive a large symmetry score. Non-parallel regions receive low scores. Square regions also receive low scores, because the medial axis is influenced by all four sides of the region, not just two parallel sides. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Fig. 2. (A) Using a portion of the office scene in Fig. 1, around the back of the chair, (B) we illustrate the manner in which a contour point *P* is given a symmetry rating. The boundary point is associated with two skeletal points on either side, *Q* and *R*. In the vicinity of each such skeletal point, the variation of the radius function is used to assign a symmetry score, as described in Algorithm 1. The grey circles depict the maximal inscribed disks along the interval under consideration around *Q*. The point *P* receives the larger of its two symmetry scores.

with contour (black) pixels and non-contour (white) pixels. Each noncontour point in the scene is assigned its Euclidean distance to the closest contour point (see Fig. 1, middle left). Once we have the Euclidean distance function, we compute its gradient. We then compute the outward flux of the gradient, through a shrinking disk placed at each non-contour point, and normalized by the perimeter of the disk. This construction of the average outward flux (AOF) is presented in greater detail in the appendix. Next, the AOF at each point is thresholded (Fig. 1, middle right). The result is a set of skeletal branches for each region in the image (Fig. 1, bottom left).

The next step is to assign a score of symmetry to each point on the line drawing. This is done by first scoring points on the medial axes, and then transferring these scores to the line pixels. A medial axis point is given a symmetry score related to the degree of local parallelism between the contours on either side of the medial point. The specific score we use is equal to the fraction of medial points in a local neighborhood for which the derivative of the radius function is below a set threshold. This process is illustrated for the medial point Q shown in Fig. 2, where neighboring medial points are illustrated with lightly shaded circles. The radii of these circles are a slowly varying function of the position along this medial branch and therefore (depending on the threshold used for the derivative) we might expect Q to have a high symmetry score.

Once the scores have been computed for all medial axis points, we then map these scores to points on the boundary contours by noting that each point on the boundary is associated with two skeletal points, one on each side of the contour. This is illustrated at the boundary point *P* in Fig. 2, which is associated with the two medial axis points *Q* and *R* on either side (that is, the boundary point P provides an active constraint on the size of the disks centered on the medial axis points Q and R). The score at P is then defined to be the maximum of the scores at the two associated medial axis points Q and R. Taking the maximum makes intuitive sense because the boundaries belonging to an object are nonaccidentally related and are more likely to be in a ribbon symmetric relationship than are the boundaries of that object with other structures. An object boundary and a boundary in the background (or the boundary of a different object) are only parallel if they are accidentally aligned. In our example, we would expect the score at Q to be larger than the score at R, since the radius function at R is changing more rapidly, and this process would assign the symmetry score at Q to the boundary point P. This provides a measure of the local parallelism of the boundary in the neighborhood of P with neighboring boundary points on one or the other side of that contour fragment. This procedure is detailed in Algorithm 1.

Algorithm 1. Scoring Symmetry

- 1: procedure Symmetryscoreforskeletalpoint(Skeletal Point Q)
- 2: Consider a window of 2K + 1 skeletal points centered at *Q*.
- 3: Let us assume these 2K + 1 points are Q_{-K} , ..., Q_K , where

$$Q_0 = Q$$

4: Assign the score of symmetry as:

4: Assign the score of symmetry as:
5:
$$\mathscr{S}(Q) = \frac{\#\{Q_l \mid \forall_{l=\{-K,\dots,K-1\}} \text{ where } \frac{\|\mathscr{R}(Q_l) - \mathscr{R}(Q_{l+1})\|}{\max(\mathcal{R}(Q_l),\mathcal{R}(Q_{l+1}))} <=\tau\}}{2K}$$

6: where $\mathscr{R}(Q_i)$ represents the radius value of the maximal inscribed disk centered at Q_i and τ is a marginal threshold. Intuitively, this represents the fraction of differences between of adjacent disk radii that are small, i.e., their boundaries on either side are close to parallel.

7: return $\mathscr{S}(Q)$

8: Procedure symmetryscoreforalllinedrawingpoints

- 9: **for** each line drawing point *P* **do**
- 10: Find the centers of the inscribed disks that touch point *P*

11: Let us assume these centers are called *R* and *Q*

12: $\mathscr{S}(P) = \max(\mathscr{S}(Q), \mathscr{S}(R))$

13: >In the generic case two maximal disks touch point *P*, one disk from each side of *P*, however there are cases in which more than two disks touch *P*. In those cases, take the max score over all disks.

14: return $\mathscr{S}(P)$

Having designed a method for scoring symmetry of line drawings, we can apply it to our database of lines drawings. The line drawings we used were first described in Walther et al. (2011). Each line drawing was obtained by having a photograph traced by an artist, who was given the instruction: "For every image, please annotate all important and salient lines, including closed loops (e.g., boundary of a monitor) and open lines (e.g., boundaries of a road). Our requirement is that, by looking only at the annotated line drawings, a human observer can recognize the scene and salient objects within the image." We used 72 images from each of the six categories (beaches, forests, mountains, city streets, highways, and offices). After scoring the symmetry in each image, we analyzed the distribution of symmetry scores by category. For each category the distribution is skewed toward low-symmetry pixels. Some categories (e.g., beaches) have relatively more low



Fig. 3. Distributions of average symmetry scores. Each distribution is composed of the mean symmetry score for each of the 72 images in that category. The distributions shown are fit using a log-normal distribution. The means are shown as the '*' symbol. Two distributions, Mountain and Highway, overlap, which is why it may appear as if there are only five distributions in the figure. To assess which distributions are different, we performed two-sample Kolmogorov-Smirnoff tests on each pair, using Bonferroni correction for multiple comparisons, resulting in an alpha level of 0.0033. Cities are significantly different from all others (all p < 0.001) except for forests (p = 0.048). The remaining pairs are not significantly different from one another (all p > 0.07).

symmetry scores than others (e.g., mountains), which have few low symmetry scores. Category differences are most easily apparent in the distributions of mean symmetry scores, see Fig. 3. Here, we compute the average symmetry score for an image and record the distributions of averages for each category. These distributions reveal that the symmetry ratings do differ by category, and thus the symmetry values could be potentially used in the categorization of a scene. Cities and offices have the lowest average mean symmetry score (0.0725 and 0.102). Next are forests, with an average of 0.129. Then, mountains and highways both have means of 0.141. Finally, beaches have the highest average symmetry (0.166). This may seem surprising, given that we think of human-made buildings and objects to be symmetric, but recall that we are specifically measuring ribbon symmetry. Our measure gives high scores to elongated regions. For example, in a beach scene, as waves roll in they tend to create pairs of ribbon symmetric lines. Many objects in an office or city, while symmetric in the real world, are not ribbon symmetric in a 2D image due to perspective foreshortening.

1.2. Stimuli

Having established a new method for scoring ribbon symmetry, we selectively removed either the most or the least symmetric contour pixels of line drawings of natural scenes. The line drawings were the same drawings used in the study by Walther et al. (2011), who obtained them by having artists trace photographs. We either showed the original line drawing (the intact condition), a line drawing with the most symmetric 50% of the contour pixels retained (the symmetric condition), or a line drawing with the least symmetric 50% of the contour pixels retained (the asymmetric condition). Example stimuli are shown in Fig. 4. If ribbon symmetry is influential in scene processing, then we should expect performance to be better in the symmetric than the asymmetric condition.

1.3. Participants

Twenty-six undergraduate students in an introductory psychology course at the University of Toronto (19 female, 7 male, mean age 18.3) participated in the experiment for course credit. Five participants' data was excluded from the analysis due to floor performance. The number of participants was chosen based upon a previous study with a similar stimulus set and design (Walther & Shen, 2014). All participants gave written informed consent prior to the experiment, and all procedures were approved by the University of Toronto Research Ethics Board and adhere to the tenets of the Declaration of Helsinki.

1.4. Design and procedure

Participants were seated approximately 57 cm away from the monitor. The head position was not constrained. The experiment room was dark for the duration of the experiment. The experiment had three phases: Training, Ramping, and Testing.

On each trial, regardless of phase, participants were shown a line drawing of a scene. They were asked to respond with the category of the scene. The key mapping was randomized for each participant. At the start of each phase, participants were shown which key was mapped to which category. The possible keys were s, d, f, j, k, and l. The mapping was identical in the three phases, but was shown at the beginning of each phase as a reminder.

Each trial started with the presentation of a scene image; the duration was dependent upon the current phase (see below). Immediately following the scene, a perceptual mask was displayed for 500 ms. The mask was a line drawing image composed of contour segments which are randomly drawn from the pool of all contours, from all scenes, from all categories. After the mask disappeared, the screen was blank until the participant responded with a key-press. After the response, the next trial began.

The training phase lasted until the participant responded correctly in 17 of the last 18 trials or 72 trials in total, whichever came first. Scene images were presented for 233 ms. On an incorrect trial, a low tone was played to provide feedback. All stimuli were intact line drawings.

The ramping phase (54 trials) started with four trials of 200 ms, followed by a linear decrease in stimulus duration from 200 ms to 33 ms. As in the training phase, participants received feedback, and only intact line drawings were shown.

The testing phase lasted for 360 trials (20 line drawings per category \times 3 conditions \times 6 categories). No feedback was given after the participant's response. The stimulus duration was fixed to 53 ms, which led to a performance of 70% for intact line drawings in a pilot experiment with a different set of participants. This would result in some errors in the intact case, which allows for a comparison of the error patterns between intact and the other conditions. Each scene was only shown in one condition, and none of the test scenes were used in the previous phases, so that scenes were novel on each presentation. Participants could pause between trials if they needed a break. A schematic of the test phase of the experiment can be seen in Fig. 5.

2. Results

Participants most easily classified the intact line drawings (76.6 percent correct). Symmetric scenes were classified at 60.9 percent correct, and asymmetric scenes were classified at 53.3 percent correct. All conditions were well above chance performance of 16.7 percent (Fig. 6 A). Removing any image content clearly hindered performance; performance in the symmetric condition was significantly worse than in the intact condition (paired-sample t-test, t(20) = 13.80, $p = 1.10 \times 10^{-11}$). Categorization of the intact scenes was also significantly better than asymmetric scenes (paired-sample t-test, t(20) = 22.13, $p = 1.56 \times 10^{-15}$). Crucially, the performance for



Fig. 4. Examples for each category and condition. Rows denote category (Beaches, Forests, Mountains, Cities, Highways, and Offices), and columns denote image condition (Intact, Symmetric, Asymmetric). Note that for scenes with many contour pixels participating in strong local symmetries (e.g., the forest scene in the second row above), even the least symmetric 50% of the contour pixels can include pixels with relatively large symmetry scores.



Fig. 5. A schematic of the experiment. Stimuli (intact, symmetric, or asymmetric versions of a line drawing) were presented for 53 ms, followed by a perceptual mask for 500 ms. A blank screen was displayed until the participant responded. A total of 360 trials were presented in the testing phase.



Fig. 6. Proportion correct for each image condition. The boxplots are centered at the mean, with a line at the median. The box extends to the 25th and 75th percentiles. The lines extending from the box show the extent of all the data points. Intact categorization performance was better than either symmetric or asymmetric categorization performance (p < 0.001). Symmetric scenes were categorized more easily than asymmetric scenes (p < 0.001).

symmetric scenes was significantly better than for asymmetric scenes, even though both versions of the stimuli contained the same number of contour pixels (paired-sample t-test, t(20) = 6.21, $p = 4.56 \times 10^{-6}$).

We further break down performance into the different categories (see the confusion matrices in Fig. 7). The row labels of the confusion matrix denote the ground truth response, and the column labels denote the response of the observers. Each cell shows the proportion of observer responses for the given ground truth category, and each row sums to 1. The diagonal elements are correct answers, and off-diagonal elements are errors. We computed correlations between the off-diagonal elements of the confusion matrices (only off-diagonal elements were used so that the overall proportion correct does not affect the correlation). The confusion matrices do not show any obvious difference in the pattern of errors in the different image conditions; all correlations between error patterns were significant: intact vs symmetric (r = 0.57, p < 0.018), intact vs asymmetric (r = 0.74, $p < 1.0 \times 10^{-5}$), and symmetric vs asymmetric (r = 0.65, $p < 1.0 \times 10^{-5}$).

Comparing performance separately by category reveals variations in the performance for symmetric and asymmetric images. Four of the six categories showed better performance in the symmetric condition than in the asymmetric condition, leading to better average performance across all conditions. For office scenes, for instance, participants performed considerably better when seeing symmetric than asymmetric images (repeated-measures t-test, t(20) = 4.21, $p = 4.08 \times 10^{-4}$). For mountain scenes, on the other hand, performance is equal in the symmetric and asymmetric conditions (t(20) = -0.08, p = 0.93).

Presumably this is due to different types of contour relationships present in the mountain scenes than in other scenes where the symmetry effect is present. For example, some objects tend to be symmetric, but very few are present in our mountain scenes. Additionally, symmetries between scene elements, such as the symmetry present between neighboring tree trunks (in a forest) or between the windows in a building (in a city) are not present in a mountain scene. Since mountain scenes lack these sorts of symmetries, removing symmetric contours leads to different distortions in mountain scenes than for other categories. Highway scenes also showed a significant performance difference $(t(20) = 4.28, p = 3.6 \times 10^{-4})$. Forest scenes showed a large performance difference between symmetric and asymmetric $(t(20) = 3.86, p = 9.78 \times 10^{-4})$. Tree trunks, with their high degree of ribbon symmetry, are prone to being distorted disproportionately when symmetric content is removed, whereas the highly irregular foliage will be present, but less recognizable, in the asymmetric images. Beach scenes also showed a modest effect (t(20) = 2.86, p = 0.0097), slightly smaller than that present in human-made scenes.



Fig. 7. Confusion matrices for the different conditions. Rows are the true category labels, and the columns are the subject responses. Correct answers lie on the diagonal, so a strong diagonal represents good performance.

While there was a large effect in all human-made scenes, the

direction of the effect was reversed for city scenes, relative to the direction found for all other scene categories (t(20) = -3.31, p = 0.004). With build environments exhibiting a high degree of structural symmetry, smaller, isolated objects, such as people and cars, frequently show comparably weaker 2D ribbon symmetry than buildings, even thought they are 3D symmetric. As a result, such objects are almost entirely contained in the asymmetric image and may be a strong cue to scene category. Finally, the scale of symmetry we measure may not match the scale of the symmetry that exists in a city. For example, neighboring contours may not be object boundaries but instead boundaries of regions/parts inside a single object. We will consider this in more detail when we discuss possible limitations of our symmetry scoring method.

3. Discussion

What drives the large difference in performance between the ribbon symmetric and ribbon asymmetric scenes? One possibility is that participants use local symmetry content as a summary statistic, either computing a single symmetry summary score for each scene or the entire distribution of symmetry content. Wilder et al. (2018) demonstrated that contour cotermination at junctions had a weaker influence on scene perception than what appeared to be a longer-ranged relationship. Here we concretely measure and control parallelism in the image, and we demonstrate that local ribbon symmetry does indeed influence scene perception. Even though the symmetry measured here represents a relationship between contours, once the symmetry is measured, this information could be ignored and only the distribution of symmetry in the image used for classification. We believe this is unlikely. The distributions for a given scene category are very different in the three different conditions. Thus, participants would need to learn the distribution for each condition without prior experience with these manipulations and in the absence of feedback. Moreover, a participant's visual system would need to know which condition they are seeing in order to accurately use this information. If this were the case, we would expect different error patterns in the asymmetric condition (without strong symmetry content) than in either the intact or symmetric conditions (with strong symmetry content), but the confusion matrices show no obvious difference in error patterns (Fig. 7). We hypothesize that the visual system uses symmetry to jump-start the grouping of image information into meaningful units. Performance was lower in the asymmetric condition because the process could not be jump-started to



Fig. 8. Histogram showing the length of contours in the symmetric and asymmetric images for all line-drawings in the data-set. Note that the x-axis is on a log scale.

the same extent, resulting in poorer grouping.

Walther et al. (2011) suggested that longer contours are more useful for scene classification. When selecting the most/least symmetric contour pixels we did not control the length of contiguous sets of contour pixels. Thus, the average length of a contiguous contour segments is not necessarily equal in the two half-images. While this could play a role in performance, we argue that it is not responsible for our large performance difference.

In order to estimate the length of each segment, we selected a set of connected black pixels, and counted the number of pixels in that set. Fig. 8 shows histograms of contour length for the symmetric (turquoise) and asymmetric (red) images combined across all categories. The average length is shorter in asymmetric than symmetric images. Note, however, that the variance in contour length is larger for asymmetric images, as they also tend to contain many of the longest contours. If, as Walther et al. (2011) showed, longest contours convey the most information about scene content, the asymmetric condition should benefit, as these most informative contours live in the asymmetric images. Thus, contour length does not appear to drive the behavioral difference between symmetric and asymmetric images in our data.

Additionally, Walther et al. (2011) only found a performance difference between long and short contours when 75% of the pixels were removed. When they removed 50% of the pixels, on the same line drawing used here, performance for the long and short contour versions was statistically equal. Therefore, we should not expect a performance difference here based on line length alone.

Previous work, where portions of contours were deleted, has argued that the important contour pieces were junctions (Biederman, 1987), or the straight portions between junctions (Kennedy & Domander, 1985). Our previous work (Wilder et al., 2018) found that for scenes, middle segments between junctions were more useful for determining scene category than were junctions. The current study finds that scene perception is aided by segments participating in a symmetry relationship with another segment rather than by middle segments in general.

How does ribbon symmetry facilitate grouping image content into meaningful surface or object parts? Most objects are not mirror symmetric in the image plane, and many objects that are locally symmetric will have low symmetry scores assigned to portions of their boundary contours. As a consequence, pixels from a single object can be assigned to different half-images, splitting the object into pieces. Additionally, some objects entirely fall into the asymmetric image. Many objects that are 3D mirror symmetric in the real world are not 2D mirror symmetric when projected onto the image plane. However, local symmetry between the boundaries of a part of an object can persist, and we conjecture that this property is used by the visual system to help group these image elements into surfaces and objects. Symmetry has been the basis of many prominent parts-based representations (Biederman, 1987; Hoffman & Richards, 1984), and symmetry has been shown to be involved in image segmentation (Machilsen et al., 2009). Symmetry may assist in perceptual grouping, as ribbon symmetry has been shown to attract attention during scene viewing (Damiano, Wilder, & Walther, 2018). Our data suggests that when ribbon symmetric contours are removed, the image is much harder to understand. We hypothesize this is because the observers are unable to segment the scene into meaningful objects and parts. In the asymmetric image, the symmetric contour portions required for grouping contours into objects and object parts are missing. Thus, scene categorization performance will deteriorate, if it relies on object classification.

There are limitations to our results. Our symmetry score measures the extent to which there is a constant distance between a pair of contours (i.e., ribbon symmetry). Consider a rectangle; the contours along the main axis score highly, but due to the nature of the medial axis, a lower score is assigned to the contours on the short sides. Additionally, our score is based upon the medial axis, which captures the relationship between contours that bound the same region in the scene, and will not capture the symmetry relationships between contours separated by intervening contours. Comparing all possible symmetry relationships between all contours in the image is intractable, but some consideration of longer range symmetries is worth pursuing.

We are not claiming that the visual system relies only on symmetry when rapidly classifying a scene. Here we have only considered local ribbon symmetry in order to understand its power in isolation. Other features, such as contour junctions, may also contribute to scene categorization (Walther & Shen, 2014). Combining these features with local symmetry and other longer-range symmetry relationships could provide a more complete explanation of human scene categorization.

Asymmetric images of cities were more easily categorized, which is the reverse effect of the other categories. This demonstrates the aforementioned limitations, and makes apparent another. The city scenes in our data-set contain many windows, and the two longer edges of a window will appear in the symmetric image if and only if they are sufficiently elongated; the opposing parallel sides are missed. Also, the scale of symmetry in our current measure may not be optimal for a city scene. While the symmetry of a single window may be important, the symmetry between the sides of a single building is also important, and our method does not look at symmetry relationships at larger spatial scales. In other categories, such as forests, long-range symmetry may be less important, since the symmetric pairs tend to be the boundaries of single objects (i.e., tree trunks), which is one reason why our manipulation resulted in much better performance for symmetric forest scenes than asymmetric ones.

4. Conclusions

The non-accidental relation of symmetry was noted by the Gestalt psychologists almost a century ago (Koffka, 1922; Kohler, 1947; Wertheimer, 1938) and reflects the ubiquity of symmetry in both our natural and human-made world. Given this regularity, it was not surprising that symmetry became a powerful basis for parts-based object representations in both human and computer vision (Agin & Binford, 1973; Biederman, 1987; Blum, 1973; Pentland, 1986). Symmetry has been carefully studied for object recognition in images containing single objects. Less attention has been paid to the role of symmetry in the perception of complex scenes which contain many objects and surfaces. The complexity of a cluttered scene has encouraged approaches that are global in nature, focusing on global scenes statistics

Appendix A. Geometry of medial representations

which, in turn, avoids the challenging problem of perceptual grouping. This study represents the first attempt to study the role of a quantitative measure of local ribbon symmetry in the categorization of line drawings of complex, natural scenes. We focused on the non-accidental property of symmetry, arguably the most powerful form of perceptual grouping for relating elements at a distance, and we introduced a novel scene statistic based on the medial axis. We demonstrated that in two subsets of a stimulus, each with the exact same number of black pixels, the subset with the stronger symmetry leads to significantly better scene perception.

The obvious question raised by our findings is "why does symmetry offer an advantage to scene perception?" Our hypothesis is that the importance of correct contour grouping is even more critical in a cluttered scene, in which any given contour may be proximal to many contours belonging to other objects. Under such conditions, where proximity leads to highly ambiguous groupings, adding symmetry cues can reduce ambiguity and lead to better grouping of contours into surfaces that comprise object parts and, in turn, the objects that make up a scene.

Our work shows that local ribbon symmetry is a key feature that allows for the rapid analysis of complex real-world scenes. This finding lends further support to our previous work on the importance of local details of the structure of a scene for rapid scene perception (Choo & Walther, 2016; Walther et al., 2011; Walther & Shen, 2014). Incorporating principles of perceptual organization originally proposed by the Gestalt psychologists in a computationally rigorous way is a promising avenue toward a more complete understanding of the computational processes that make vision appear so natural and effortless.

Supplementary material

Participant data is publicly available in the Open Science Framework data repository at: https://osf.io/cwbyq/.

Acknowledgements

The authors would like to acknowledge funding from the following sources: NSERC, CFI Project Number 32896, a Sony Faculty Research Award, and SSHRC Insight Development Grant 430-2017-01189.

Visual shape analysis is a fundamental problem in perception by man and by computer and allows for inferences about properties of objects and scenes in the physical world. If a 3-D real world object or a scene is projected onto an image plane, the resulting image could be simplified by a set of contours that separate the silhouette of that object from the background or look like the line drawing of a scene. One of the ways to perform visual analysis on such a set of contours is to find good representations for them. Medial representations are among the popular choices for problems of shape analysis and they have been used for different tasks of representations and recognition in the literature. In this report, we used medial axes to analyze the boundary contours of regions within natural scenes. In the following, we review the geometry of medial representations for shapes of objects and their boundary, and we utilize the same geometry which applies to the boundary contours of regions within natural scene.



Fig. A.1. LEFT: Iterations of the grassfire process. RIGHT: The resulting skeleton.

Medial axis based representations were first introduced by Blum (1967), along with the process for generating them based on a grassfire analogy. Here the boundary is set on fire and the front advances inward at a constant speed, and as fire fronts meet, skeletal points are created. (see Fig. A.1).

The application of the grassfire process to reveal its quench points along with their radius values is called the Medial Axis Transform (MAT). Since the grassfire process is applicable to all bounded shapes, as well as the regions outside of closed shapes, the MAT gives a comprehensive representation in visual shape problems. The geometry and methods of computing the medial axis based on a notion of average outward flux of an Euclidean distance function are discussed below.

The medial axis consists of a set of medial curves about which a contour, or pair of contours, is locally mirror symmetric. Generically, each point on the skeleton inbetween contours is the result of a collision of two distinct boundary points.

Definition. Assume an *n*-dimensional open object Ω , with its boundary given by $\partial \Omega \in \mathbb{R}^n$ such that $\overline{\Omega} = \Omega \cup \partial \Omega$. A closed disk $D \in \mathbb{R}^n$ is a maximal inscribed disk in $\overline{\Omega}$ if $D \subseteq \overline{\Omega}$ but for any disk D' such that $D \subset D'$, the relationship $D' \subseteq \overline{\Omega}$ does not hold.

Definition. The Blum interior medial locus or skeleton, denoted by $Sk(\Omega)$, is the locus of centers of all maximal inscribed disks in $\partial\Omega$.

Topologically $Sk(\Omega)$ consists of a set of branches that join at branch points to form the complete skeleton. A skeletal branch, denoted by χ , is a set of contiguous regular points from the skeleton that lie between a pair of junction points, a pair of end points or an end point and a junction point. As shown by Dimitrov, Damon, and Siddiqi (2003) these three classes of points can be analyzed by considering the behavior of the average outward flux (AOF) of the gradient of the Euclidean distance function to the boundary of a 2D object, given by $\frac{\int_{\partial R} (\dot{\mathbf{q}}, \mathbf{N}) ds}{\int_{\partial R} ds}$, through a shrinking disk, where $\dot{\mathbf{q}} = \nabla \mathbf{D}$ (Dimitrov et al., 2003), with **D** the Euclidean distance function to the object's boundary. In the following, we review this computation.

The Euclidean distance between two *n*-dimensional points $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ is the length of the line segment that connects points, and the Euclidean metric $d(P,Q): \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a function that represents this distance: two these $d(P, Q) = d(Q, P) = ||Q-P|| = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$. For each point P, and a given object Ω , a distance metric, $d_{\Omega}(P)$, can be defined as follows: $d_{\Omega}(P) = \inf_{Q \in \partial \Omega} d(P, Q)$. The distance transform of an object Ω is a signed distance function that specifies how close a given point P is to the boundary of that object $\partial\Omega$: $\mathscr{D}_{\Omega}(P) = d_{\Omega}(P)$ if *P* is inside Ω , $\mathscr{D}_{\Omega}(P) = 0$ if $P \in \partial\Omega$ which is the boundary of Ω , and $\mathscr{D}_{\Omega}(P) = -d_{\Omega}(P)$ if *P* is outside Ω . In this paper, we consider the interior of regions, thus we can assume an unsigned distance function per each region.

Let us define the projection $\Pi(\mathbf{p})$ as the set of closest points on the boundary $\partial\Omega$ to \mathbf{p} , i.e., $\Pi(\mathbf{p}) \triangleq \{\mathbf{p}' \in \partial\Omega: \|\mathbf{p}-\mathbf{p}'\| = \min\{\|\mathbf{p}-\mathbf{p}'\| \forall \mathbf{p}' \in \partial\Omega\}\}$. Assume that on the boundary $\partial\Omega$, there exists only one point *Q* of minimum distance to *P* ($\Pi_{\Omega}(P) = \{Q\}$). We then define the *distance function gradient vector* for point *P* as: $\dot{\mathbf{q}}_{\Omega}(P) = \frac{Q-P}{\|Q-P\|}$. In the case of $|\Pi_{\Omega}(P)| > 1$, one cannot define the closest boundary point uniquely, and therefore the distance function gradient vector is multivalued. Except for at skeletal points, $\dot{\mathbf{q}}$ is continuous everywhere on its domain and it satisfies the equation: $|\dot{\mathbf{q}}| = 1$. Exploiting the relationship of the integral of the divergence of a vector field within a simply-connected region to the outward flux of that vector field through the boundary of that region, using the divergence theorem, leads to characterization of skeletal points. Let R be a region where its boundary ∂R is a simple closed curve, and **N** be the outward normal at each point on the boundary ∂R .

Definition. The outward flux of $\dot{\mathbf{q}}$ through ∂R is defined as $\int_{\partial R} \langle \dot{\mathbf{q}}, \mathbf{N} \rangle ds$, and the average outward flux of $\dot{\mathbf{q}}$ through ∂R is defined as $AOF = \frac{\int_{\partial R} \langle \dot{\mathbf{q}}, \mathbf{N} \rangle ds}{\int_{\partial R} ds}$.

Using the divergence theorem, Dimitrov et al. Dimitrov et al. (2003) categorized points in the image into classes by considering the behavior of the average outward flux (AOF) through a shrinking disk of the gradient of the Euclidean distance function to the boundary of a 2D object. In particular, the limiting AOF value of all points not located on the skeleton is equal to zero.

Moreover, as discussed in Siddiqi and Pizer (2008), there are three classes of generic points that are located on the skeleton, where by generic we mean stable under arbitrarily small pertubations of the boundary. These are regular skeleton points, end points, and junction points. For any point p:

- 1. **p** is a *regular point* if the maximal inscribed disk at **p** touches the boundary at two corresponding boundary points such that $|\Pi_{\Omega}(\mathbf{p})| = 2$ (The projection $\Pi(\mathbf{p})$ is the set of the closest points on the boundary $\partial\Omega$ to \mathbf{p} , i.e., $\Pi(\mathbf{p}) \triangleq \{\mathbf{q} \in \partial\Omega: \|\mathbf{p}-\mathbf{q}\| = \min\{\|\mathbf{p}-\mathbf{q}\| \forall \mathbf{q} \in \partial\Omega\}\}$. The computed AOF at a regular point **p** is given by $\lim_{\epsilon \to 0} \frac{\mathbf{F}_{\epsilon}(\mathbf{p})}{2\pi \epsilon} = -\frac{2}{\pi} \sin \theta$. 2. **p** is an *end point* if there exists δ ($0 < \delta < r$) such that for any ϵ ($0 < \epsilon < \delta$) the circle centered at **p** with radius ϵ intersects $Sk(\Omega)$ just at a single
- point (*r* is the radius of the maximal inscribed disk at **p**). The computed AOF at an end point **p** is given by $\lim_{\epsilon \to 0} \frac{\mathbf{F}_{\epsilon}(P)}{2\pi \epsilon} = -\frac{1}{\pi} (\sin \theta_P \theta_P)$. 3. **p** is a *junction point* if $\Pi_{\Omega}(\mathbf{p})$ has three or more corresponding closest boundary points. Generically a junction point has degree 3. All other branch points are unstable. The computed AOF at a junction point **p** is given by $\lim_{\epsilon \to 0} \frac{\mathbf{F}_{\epsilon}(P)}{2\pi \epsilon} = -\frac{1}{\pi} \sum_{i=1}^{n} \sin \theta_i$.

References

- Agin, G. J., & Binford, T. O. (1973). Computer description of curved objects. Proceedings of the 3rd international joint conference on artificial intelligence (pp. 629-640).
- Attneave, R. (1954). Some informational aspects of visual perception. Psychological Review, 61(3), 183.
- Berman, D., Golomb, J. D., & Walther, D. B. (2017). Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. PloS One, 12(12), e0189828.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94(2), 115.
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology, 14(2), 143-177.
- Binford, I. (1971). Visual perception by computer. IEEE Conference of Systems and Control (pp. 1489-1496). .
- Blakemore, C., & Hague, B. (1972). Evidence for disparity detecting neurones in the human visual system. The Journal of Physiology, 225(2), 437-455.
- Blum, H. (1967). A transformation for extracting new descriptors of shape. Models for the Perception of Speech and Visual.
- Blum, H. (1973). Biological shape and visual science (part 1). Journal of Theoretical

Biology, 38, 205-287.

- Bruck, J., Gao, J., & Jiang, A. (2007). MAP: Medial axis based geometric routing in sensor networks. Wireless Networks, 13(6), 835-853.
- Burbeck, C., & Pizer, S. (1995). Object representation by cores: Identifying and representing primitive spatial regions. Vision Research, 35, 1917-1930.
- Chan, M. W., Stevenson, A. K., Li, Y., & Pizlo, Z. (2006). Binocular shape constancy from novel views: The role of a priori constraints. Perception & Psychophysics, 68(7), 1124-1139.
- Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex. NeuroImage, 135, 32-44.
- Damiano, C., Wilder, J., & Walther, D. B. (2018). Mid-level feature contributions to category-specific gaze guidance. Attention, Perception, & Psychophysics, 1-12.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. Vision Research, 40(16), 2187-2200.
- Dimitrov, P., Damon, J. N., & Siddiqi, K. (2003). Flux invariants for shape. Computer vision and pattern recognition, 2003. Proceedings. 2003 IEEE computer society conference on: Vol. 1, (pp. I-835-I-841). ieeexplore.ieee.org.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment, Nature, 392(6676), 598.
- Feldman, J., & Singh, M. (2006), Bayesian estimation of the shape skeleton, Proceedings of the National Academy of Sciences, 103(47), 18014-18019.

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* 0956797613507584.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, 18(1), 65–96. Hubel, D. H., & Wiesel, T. (1963). Shape and arrangement of columns in cat's striate

- cortex. The Journal of Physiology, 165(3), 559–568.
 Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.). Vision and Artifact (pp. 25–32). New York: Springer.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302–4311.
- Kennedy, J. M., & Domander, R. (1985). Shape and contour: The points of maximum change are least useful for recognition. *Perception*, 14(3), 367–370.

Koffka, K. (1922). Perception: An introduction to the Gestalt-theorie. Psychological Bulletin, 19(10), 531–585.

Koffka, K. (1935). Principles of gestalt psychology, international library of psychology, philosophy and scientific method. New York: Harcourt Brace.

Kohler, W. (1947). Gestalt psychology (1929). New York, NY: Liveright.

- Kovacs, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370(6491), 644.
- Lee, T. S. (1996). Neurophysiological evidence for image segmentation and medial axis computation in primate v1. Computation and neural systems: Proceedings of the fourth annual computational neuroscience conference.
- Lee, T., Fidler, S., & Dickinson, S. (2013). Detecting curved symmetric parts using a deformable disc model. *Proceedings of the IEEE international conference on computer vision* (pp. 1753–1760).
- Lescroart, M. D., & Biederman, I. (2013). Cortical representation of medial axis structure. *Cerebral Cortex*, 23(3), 629–637.

Levinshtein, A., Sminchisescu, C., & Dickinson, S. (2013). Multiscale symmetric part detection and grouping. *International Journal of Computer Vision (IJCV)*, 104(2), 117–134.

Leymarie, F. F., & Kimia, B. B. (2008). From the infinitely large to the infinitely small. In K. Siddiqi, & S. Pizer (Eds.). Medial representations: Mathematics, algorithms and applications (pp. 327–351). Springer.

- Link, N. K., & Zucker, S. W. (1988). Corner detection in curvilinear dot grouping. Biological Cybernetics, 59(4–5), 247–256.
- Li, Y., & Pizlo, Z. (2008). Perception of 3d shapes from line drawings. Journal of Vision, 8(6), 452.
- Liu, Y., Hel-Or, H., Kaplan, C. S., & Van Gool, L. (2010). Computational symmetry in computer vision and computer graphics. Foundations and Trends? In Computer Graphics and Vision, 5(1-2), 1–195.
- Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, *9*(12), 11.
- Malach, R., Reppas, J., Benson, R., Kwong, K., Jiang, H., Kennedy, W., ... Tootell, R. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, 92*(18), 8135–8139.
- Marr, D., & Nishihara, H. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society of London B, 200, 269–294.

Narayanan, M., & Kimia, B. (2012). Bottom-up perceptual organization of images into object part hypotheses. *European conference on computer vision* (pp. 257–271).

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41(2), 176–210.

- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175.
- Pentland, A. P. (1986). Perceptual organization and the representation of natural form. Artificial Intelligence, 28(3), 293–331.
- Pizlo, Z. (2014). Making a machine that sees like us. UK: Oxford University Press.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, 81(1), 10.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. JOSA, 61(9), 1176–1186.

Siddiqi, K., Bouix, S., Tannenbaum, A. R., & Zucker, S. W. (2002). Hamilton-jacobi skeletons. International Journal of Computer Vision, 41(9), 215–231.

Siddiqi, K., & Pizer, S. (2008). Medial representations: Mathematics, algorithms and applications, Vol. 37. Springer Science & Business Media.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999). Shock graphs and shape matching. *International Journal of Computer Vision*, 35(1), 13–32.

- Stahl, J. S., & Wang, S. (2008). Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3), 395–411.
- Tamrakar, A., & Kimia, B. B. (2004). Medial visual fragments as an intermediate image representation for segmentation and perceptual grouping. *Computer vision and pattern* recognition workshop, 2004. CVPRW'04. Conference on (pp. 47).
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. Network: Computation in Neural Systems, 14(3), 391–412.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, 30(6), 655–668.

Van Tonder, G. J., Lyons, M. J., & Ejima, Y. (2002). Perception psychology: Visual structure of a Japanese Zen garden. *Nature*, 419(6905), 359.

- Wagemans, J. (1993). Skewed symmetry: A nonaccidental property used to perceive visual forms. Journal of Experimental Psychology: Human Perception and Performance, 19(2), 364.
- Wagemans, J. (1997). Characteristics and models of human symmetry detection. Trends in Cognitive Sciences, 1(9), 346–352.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I, Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172.
- Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings* of the National Academy of Sciences, 108(23), 9661–9666.
- Walther, D. B., & Shen, D. (2014). Nonaccidental properties underlie human categorization of complex natural scenes. *Psychological Science* 0956797613512662.
 Wertheimer, M. (1938). *Laws of organization in perceptual forms*. Kegan Paul, Trench,
- Trubner & Company.
- Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision*, 10(4), 6.

Wilder, J., Dickinson, S., Jepson, A., & Walther, D. B. (2018). Spatial relationships between contours impact rapid scene classification. *Journal of Vision*, 18(8), 1.

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, 119(3), 325–340.

Wilder, J., Feldman, J., & Singh, M. (2016). The role of shape complexity in the detection of closed contours. Vision Research, 126, 220–231.