Published in final edited form as:

Psychol Sci. 2014 April 1; 25(4): 851–860. doi:10.1177/0956797613512662.

Non-accidental properties underlie human categorization of complex natural scenes

Dirk B. Walther and

Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, Ohio, 43210, USA, Phone: 614-688-3923

Dandan Shen

Department of Electrical and Computer Engineering, The Ohio State University

Dirk B. Walther: bernhardt-walther.1@osu.edu; Dandan Shen: dandans.1102@gmail.com

Abstract

Humans can categorize complex natural scenes quickly and accurately. Which scene properties enable us to do this with such apparent ease? We extracted structural properties of contours (orientation, length, curvature) and contour junctions (types and angles) from line drawings of natural scenes. All of these properties contain information about scene category that can be exploited computationally. But, when comparing error patterns from computational scene categorization with those from a six-alternative forced-choice scene categorization experiment, we found that only junctions and curvature made significant contributions to human behavior. To further test the critical role of these properties we perturbed junctions in line drawings by randomly shifting contours. As predicted, we found a significant decrease in human categorization accuracy. We conclude that scene categorization by humans relies on curvature as well as the same non-accidental junction properties used for object recognition. These properties correspond to the visual features represented in area V2.

Keywords

Natural scenes; Scene categorization; Non-accidental properties; Line drawings; Structural description

We are able to recognize our natural environment as, say, a forest or an office (Tversky & Hemenway, 1983), and we succeed at categorizing scenes even with extremely short presentation times (Potter & Levy, 1969; Fei-Fei, Iyer, Koch, & Perona, 2007; Greene & Oliva, 2009; Loschky, Hansen, Sethi, & Pydimarri, 2010) and possibly in the absence of attention (Li, VanRullen, Koch, & Perona, 2002; but see Evans & Treisman, 2005; Cohen, Alvarez, & Nakayama, 2011). What are the properties of natural scenes that allow us to accomplish this impressive feat?

Over the last decades several featural representations of scenes have been proposed. Marr thought of scenes as collections of surfaces in three-dimensional space (a 2½-dimensional sketch of the scene), whose orientations and spatial relations would allow us to infer scene content (Marr, 1982). Others found statistics of color (Oliva & Schyns, 2000; Goffaux et al., 2005) or oriented image energy as represented in the spatial frequency spectrum (Oliva & Torralba, 2001; Torralba & Oliva, 2003) to be indicative of scene category. We here argue

that contour junctions, which are invariant to slight changes of viewing position (non-accidental properties; Biederman, 1987), and curvature play an essential part in human scene categorization.

Computational analysis of natural scenes has so far largely focused on the discovery of statistical regularities of the luminance patterns of photographs (in pixel space or frequency space) that allow for computational scene categorization and, in some cases, relate to human categorization performance (Oliva & Torralba, 2001; Torralba & Oliva, 2003; Renninger & Malik, 2004; Fei-Fei & Perona, 2005). The image patches preferred by the resulting filters tend to look like contrast edges or straight contour lines and, occasionally, like corners or curved contours (Renninger & Malik, 2004; Fei-Fei & Perona, 2005). The same types of filters were derived as an efficient encoding of natural images when enforcing sparsity (Olshausen & Field, 1996) or statistical independence (Bell & Sejnowski, 1997) of the ensuing code. However, establishing explicit descriptions of the structure of scenes from these luminance patterns has proven to be challenging.

Human artists are able to emphasize lines that describe object boundaries over those created by shadows or irrelevant textures (Sayim & Cavanagh, 2011). We can thus sidestep the issue of detecting contours from photographic images by extracting structural features from line drawings of natural scenes, which were created by trained artists. As we have shown recently (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011), line drawings of natural scenes elicit the same category-specific activation patterns as color photographs in the scene-selective parahippocampal place area (PPA) and retrosplenial cortex (RSC). With neural activity in the PPA tightly linked to behavioral scene categorization (Walther, Caddigan, Fei-Fei, & Beck, 2009), this suggests that the scene structure preserved in line drawings plays an important role in scene categorization.

Line drawings give us ready access to several aspects of scene structure: orientation, length, and curvature of contours as well as type and angle of contour junctions. To determine the differential contributions of these structural properties to the representation of scene categories, we tested their efficacy for computational scene categorization as well as scene categorization by human subjects. Scene categorization by a computational model based on line orientations yielded the highest accuracy among all the properties. However, human subjects' patterns of errors in a behavioral scene categorization task were best matched by a computational model based on summary statistics of curvature and junctions, but not orientation or length. Importantly, this finding generalizes from line drawings to photographs of natural scenes, thereby validating that people rely on the same structural properties when categorizing scenes depicted in photographs. To determine if contour junctions have a causal role in scene categorization, we repeated the behavioral experiments using line drawings with an altered distribution of junctions, while keeping all other properties unchanged. If junctions are causally involved in scene categorization, then accuracy should be lower for the manipulated than for intact line drawings. We did indeed find such a decrease in categorization accuracy, confirming such a causal relationship between junction properties and scene categorization by humans.

Methods

Participants

22 subjects (12 male, 10 female; ages: 18–21) participated in the first, and a separate group of 27 subjects (8 male, 19 female; ages: 18–23) participated in the second experiment for partial course credit. All subjects had normal or corrected-to-normal vision and gave written informed consent. The experiments were approved by the Institutional Review Board of The Ohio State University. For each of the two experiments the data of four subjects were

excluded from the analysis, because these participants did not complete the entire experiment or did not follow the instructions.

Stimuli

We used 432 color photographs of six real-world scene categories (3 natural: beaches, forests, mountains; 3 man-made: city streets, highways, offices). These images were chosen from a set of 4025 images downloaded from the internet as the best exemplars of their categories according to ratings by 137 observers per image on average (Torralbo et al., 2013). Images were resized to a resolution of 800×600 pixels.

Line drawings were produced by trained artists at the Lotus Hill Research Institute (Wuhan, Hubei Province, PR of China) by tracing outlines in the color photographs via a custom graphical user interface. Artists were given the following instructions: *“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.”* Each contour was traced as a series of straight lines, connecting individual anchor points, thus closely approximating the contours in the images with as few anchor points as possible. Line drawings were rendered at a resolution of 800×600 pixels by connecting the anchor points with black straight lines on a white background.

Contour-shifted line drawings were generated automatically by randomly translating each contour a random distance, while ensuring that the entire contour would fit within the image boundaries. Horizontal and vertical translation distances were drawn from independent uniform distributions of allowable translations. Contour-shifted line drawing images were rendered the same way as the intact line drawings at a resolution of 800×600 pixels.

Color masks were generated by synthesizing random, scene-like textures, separately for the RGB color channels, using the texture synthesis method of Portilla and Simoncelli (2000) with color photographs from our stimulus set as input (see Loschky, et al., 2010 for a rationale for using such masks). Six random textures, each synthesized from a randomly chosen photograph from a different scene category, were averaged to avoid category bias. We pre-computed 100 such masks and selected from them at random for each trial. The same set of masks was used for color photographs and line drawings. See fig. S1 for examples of scene stimuli and masks.

Procedure

Images were presented on a CRT monitor (1024×768 pixels, 150Hz) at a resolution of 800×600 pixels (approximately 23°×18°, depending on exact viewing distance), centered on a 50% gray background. At the start of each trial, a fixation cross was presented for 500ms, followed by a brief presentation of an image (initially 250ms). The image was replaced by a perceptual mask for 500ms, and finally a blank screen appeared for 2000ms. Subjects performed six-alternative forced-choice (6AFC) categorization of the images by pressing one of six buttons (S, D, F, J, K, L) on a computer keyboard. The mapping of categories to buttons was randomized for each subject, and subjects practiced the mapping until they achieved 90% accuracy. Following practice, the stimulus onset asynchrony (SOA) was adjusted for each subject with a staircasing procedure to 65% accuracy. A randomly selected subset of 12 color photographs from each category was used for practice and staircasing, leaving 60 images per category for testing. A tone alerted participants to incorrect responses.

In the testing phase of the first experiment, half of the 360 test images were shown as photographs and half as line drawings, randomly intermixed. Each test image was shown

only once, either as a photograph or as a line drawing. Trials were grouped into 18 blocks of 20 images. The SOA was fixed to the final SOA achieved in the staircasing phase. No feedback was given during the testing phase. Responses were recorded in a 6×6 confusion matrix, separately for photographs and line drawings. Accuracy was computed as the mean over the diagonal elements of the confusion matrix.

The second experiment followed the same procedure, except that line drawings were used for practice and staircasing, and original and contour-shifted line drawings were used for testing.

Computational image analysis

We computed properties of individual contours (length, orientation and curvature) as well as junctions of contours (type and angle) from the line drawings (fig. 1A). These properties were computed directly from the geometrical information of the line drawings. They are closely related to the structure present in the corresponding scene photographs. Details on the computation of properties from line drawings are provided in the Supplemental Material available online. We then constructed histograms of these properties in order to quantify their distribution for a given line drawing (fig. 1B).

For each scene we concatenated the histograms of all properties into a 36-dimensional vector and performed ten-fold cross validation classification of scene categories with a linear support-vector machine (SVM) classifier. Predictions of scene categories by the classifier were recorded in a confusion matrix (fig. 2A).

We repeated the analysis for each of the five properties separately. Contour-shifted images were classified the same way, except that the classifier was trained on original and tested on contour-shifted line drawings to mimic the behavioral experiment.

Agreement between behavioral and computational results was assessed by computing the Pearson correlation coefficient between the off-diagonal elements of the two confusion matrices, thus quantifying the match of error patterns. Significance of correlations was determined non-parametrically by repeating the analysis with all 720 permutations of the six category labels. Results with $p < 0.05$ (fewer than 36 of the 720 permutations resulting in correlations larger than the correct order) were accepted as statistically significant.

Results

Image properties

We determined three types of properties of the contours in the line drawings of natural scenes: length, orientation, and curvature. Furthermore, we detected junctions at the intersections of lines and describe them by their type (X, T, Y, and arrow) as well as their angle. See fig. 1A for illustration.

To determine if these five properties contain sufficient information to infer scene category we computed summary statistics of all five properties over the line drawings. Fig. 1B shows the average histograms of the properties for each scene category. In these histograms we observe that, unsurprisingly, man-made scene categories (city streets, highways, offices) are dominated by horizontal and vertical orientations, while the distribution of orientations among natural scene categories (beaches, forests, mountains) are more nuanced. Beaches are dominated by horizontal and forests by vertical orientation, while the distribution of orientations is more balanced for mountains, with slightly more horizontal and near-horizontal orientations. These differences in orientations between scene categories are consistent with results found by Torralba and Oliva (2003). Furthermore, we see that T

junctions are more prevalent in man-made than natural scenes, which is also reflected in the dominance of 90-degree intersection angles. Contour curvature and length do not show an obvious distinction between the categories, although both reflect the total line length of the line drawings in their overall histogram magnitude.

Computational analysis

Can we use these regularities to discriminate among natural scene categories? To answer this question we trained a support vector machine classifier to predict the category of a scene based on the distribution of its structural properties. Predictions for scenes previously unseen by the classifier were correct for 83.6% of the test images (chance level: 16.7%), suggesting that the properties that were derived from the line drawings contain sufficient information to determine their category membership in most cases. Classifier errors were recorded in the off-diagonal elements of a confusion matrix (fig. 2A).

To determine the contribution of each of the five properties we repeated the computational analysis for each property separately. In figure 2C we see that line orientation appears to be the most informative property, allowing for correct classification of 71.6% of all line drawings. The other properties allowed for correct classification of between 47.4% and 56.9% of all images.

Experiment 1

To address how these computational results relate to human behavior we performed a 6AFC categorization experiment. Staircasing resulted in SOAs between 16.7 ms and 86.7 ms ($M=26.5$ ms). Subjects' responses were recorded in a confusion matrix, separately for line drawings and photographs (fig. 2B). Categorization accuracy (mean of the diagonal of the confusion matrix; chance: 16.7%) was 66.2% (SEM 3.1%) for line drawings and 77.3% (SEM 3.0%) for photographs, averaged over 18 subjects (fig. 2C). The difference in accuracy between the two image types was significant at $p = 4.9 \cdot 10^{-6}$ ($t(17) = 6.6$, paired t-test). Confusion matrices are shown in fig. 2B. There was no significant difference in response times.

To assess the similarity in error patterns, we correlated the off-diagonal elements (confusions) of the behavioral confusion matrices with those obtained from the computational analysis. For line drawings, we observed the highest correlation of behavior with junction angles ($r = 0.730$, $p < 0.0014$), junction types ($r = 0.513$, $p = 0.0083$), and contour curvature ($r = 0.583$, $p = 0.0069$), whereas line orientation ($r = 0.091$, $p = 0.33$) and contour length ($r = 0.169$, $p = 0.20$) were not significantly correlated with human behavior (fig. 2D).

It is conceivable that this finding may be an artifact of human participants viewing line drawings of natural scenes, where line intersections may be more readily accessible than in photographs under natural lighting conditions. To address this possible confound we correlated the patterns of errors made by human subjects when categorizing photographs with the classification errors made in the computational analysis of the corresponding line drawings. We again found significant correlation of error patterns for junction angles ($r = 0.658$, $p = 0.0014$), junction types ($r = 0.576$, $p = 0.0097$), and curvature ($r = 0.511$, $p = 0.0097$), but not for line orientation ($r = 0.078$, $p = 0.34$) or contour length ($r = 0.023$, $p = 0.41$; fig. 2D). Since we observed the same pattern of correlations for photographs as for line drawings, we conclude that the importance of junctions and curvature for human behavior transcends the specifics of the low-level representation of scenes.

Experiment 2

If humans rely so heavily on intersection-related properties for categorizing scenes, then perturbing these properties should have a detrimental effect on categorization performance. To test this prediction we randomly translated the contours of the line drawings within the image bounds, thus preserving all properties that pertain to the contours themselves (length, orientation, curvature), but altering junction angles and counts of junction types (fig. 3A).

By design, computational classification of line drawings using length, orientation or curvature remained unaffected, because these properties of individual contours do not change when the contours are shifted. Classification using counts of junction types, on the other hand, dropped from 47.4% to 35.2% and classification using junction angles from 53.7% to 39.7%, leading to a decrease from 83.6% to 73.4% when using the combination of all five properties (fig. 3B). To test the effect of contour-shifting on behavioral performance we repeated the 6AFC experiment as before, but using intact line drawings for practice and staircasing and intact and contour-shifted line drawings randomly intermixed for testing. Stimuli were presented for 13.3 ms to 133.3 ms ($M=33.9\text{ms}$; staircased for each participant to 65% accuracy) and were followed by a perceptual mask. Performance for the contour-shifted line drawings (31.9%) was significantly lower ($p = 4.8 \cdot 10^{-9}$, $t(22) = 9.26$; paired t -test) than for the original line drawings (53.8%), confirming both our ad-hoc prediction and the computational analysis when using intersection-related properties (fig. 3B). The relatively large variation in SOA, caused by individual differences in subject performance during staircasing, did not affect the outcome of the experiment adversely. Participants with longer SOAs showed higher overall accuracy during testing, but accuracy was lower for contour shifted than intact line drawings for all participants.

With junction properties no longer available for categorizing scenes, which properties did participants resort to in order to make their decision? To address this question we correlated the errors (off-diagonal entries in the confusion matrix) of behavioral performance when viewing contour-shifted line drawings with the computational classification errors using properties of intact line drawings (fig. 3C). As we should expect, there is no longer any correlation with junction properties (types: $r=0.160$, $p=0.20$; angles: $r=0.141$, $p=0.18$), but there is highly significant correlation now with contour length ($r=0.725$, $p<0.0014$) and orientation ($r=0.412$, $p=0.0069$), with marginal error correlation with curvature ($r=0.310$; $p=0.057$) and significant correlation using all properties ($r=0.352$, $p=0.017$).

Discussion

We have demonstrated that structural properties describing junctions and curvature in line drawings of natural scenes lead to computational scene categorization results that are more similar to human behavior than properties relating to orientation and length of contours. This similarity is not limited to viewing line drawings; the effect is just as pronounced for color photographs, suggesting that people readily extract these properties from photographs when perceiving scenes. Beyond mere correlation we have shown that contour junctions are causally involved in natural scene categorization by demonstrating that human categorization performance decreased significantly when the distribution of junction properties was altered. When deprived of junction properties, observers resorted to contour length and line orientation in order to categorize scenes.

These junction properties coincide with some of the non-accidental properties identified as critical for object recognition by Biederman (1987): Y and arrow junctions indicate the orientation of corners in three-dimensional space, and T junctions occur along the contour of one surface occluding another. Note that our set of properties is bound to be incomplete. For practical reasons, we did not take into account L junctions, which indicate corners, or curved

Y junctions, which indicate curved, cylinder-like objects. We did not distinguish between internal object contours and contours at the boundary between objects as would be necessary for an explicit representation of shape (Marr & Nishahara, 1978). Even more importantly, although the junction properties are localized, our summary statistics ignored their locations and spatial relations, which are believed to be critical for object and scene recognition (Biederman, 1987; Kim & Biederman, 2011) as well as the segregation of surfaces (Rubin, 2001). In light of these gross simplifications it is even more remarkable that simple summary statistics of junctions and curvature have such a clear relation to natural scene categorization by humans. One possible explanation is that summary statistics of the image properties may suffice for fast gist recognition, whereas spatial locations and relations may be more important for more detailed inspection of a scene. This interpretation is consistent with findings that scene categorization does not require focused attention (Li, et al., 2002; but see Evans & Treisman, 2005; Cohen, et al., 2011).

Although line orientation and length are outweighed by junctions and curvature in their importance for human performance, they nevertheless allowed for fairly accurate categorization of natural scenes by computational algorithms (fig. 2C). These statistical regularities can be exploited successfully in computational scene analysis, for instance in the spatial frequency domain (Oliva & Torralba, 2001). Why then do people rely predominantly on non-accidental junction properties? There may be two, somewhat related reasons. First, non-accidental properties afford stable representations under moderate three-dimensional transformations. In fact, the neural representation of a scene in the retrosplenial cortex generalizes over different views of the same scene (Park & Chun, 2009). Such limited viewpoint invariance may form the basis for interpreting scenes as arrangements of surfaces in three dimensions (Marr, 1982). Second, it is likely to be more efficient for people to make use of the same visual features for scene perception as for a somewhat simpler visual task, object recognition, namely junctions and curvature (Biederman, 1987). Such convergence dispenses with the need for a separate scene-processing pathway, suggesting instead a feature representation that is shared between scene and object perception.

Both contour junctions and curvature are representatives of third-order image properties in the nomenclature of Koenderink and van Dorn (1987), with the zeroth order being luminance patterns, the first order luminance gradients, and the second order straight lines: junctions describe the interaction between lines, and curvature describes the change of line orientation along contours. With contours and their junctions clearly delineated, line drawings may serve as a useful stand-in for the analysis of third-order scene structure. However, line drawings do not capture all information that is present in color photographs of natural scenes, as is illustrated by their significantly lower categorization accuracy in experiment 1 (fig. 2C). Aside from color, photographs contain additional texture information. Regularities in the textures of natural scenes have been modeled successfully at the level of V1 (Olshausen & Field, 1996; Bell & Sejnowski, 1997; Portilla & Simoncelli, 2000; Karklin & Lewicki, 2009) as well as V2/V4 (Freeman & Simoncelli, 2011). In the latter study, V2-type features were found to be critical for human scene perception: observers were unable to distinguish between intact images and images with peripheral noise that mimicked the summary statistics of visual features at V2. Nevertheless, when human observers were asked to apply scene category labels to the textures of scenes produced by the Portilla & Simoncelli (2000), they performed only slightly above chance (Loschky, et al., 2010). Thus, more work is needed to determine how such texture information may be used by human observers for rapidly categorizing scenes.

An important question remains: How could our visual system determine the statistics of structural scene properties during such brief image presentations? First, sensitivity to the structural features found in the current study to be most important for human scene

categorization, junctions and curvature, has been identified in visual areas V2 (e.g., Peterhans & von der Heydt, 1989; Hegde & Van Essen, 2000) and V4 (e.g., Pasupathy & Connor, 2002). Thus, these properties are available relatively early in the visual processing stream. Second, various visual features can be summarized by the visual system efficiently, thus enabling extraction of statistical properties from a scene. Statistics of visual features as varied as average orientation (Baldassi & Burr, 2000; Choo, Levinthal, & Franconeri, 2012), average size (Ariely, 2001; Chong & Treisman, 2003), average color and brightness (Bauer, 2009), average spatial location of multiple objects (Alvarez & Oliva, 2008), or the count of objects (Burr & Ross, 2008) have been shown to be estimated efficiently by the visual system. Furthermore, it has been suggested that the human visual system may rely on the summary statistics of visual features for peripheral vision (Pelli, Palomares, & Majaj, 2004). Considering the significance of peripheral vision to scene categorization (Larson & Loschky, 2009) and crowding (Balas, Nakano, & Rosenholtz, 2009), statistical properties may play an important role in the perception of scenes.

Although our work does not address the mechanisms of natural scene categorization in the brain directly, we can make some conjectures. In previous work we have shown that human performance in essentially the same behavioral task used in this paper is tightly linked to decoding of scene category from brain activity in the parahippocampal place area (PPA) (Walther, et al., 2009). Here we show a close connection between behavior and junction properties. This could be interpreted to mean that activity in the PPA reflects the distribution of junctions. However, this issue will need to be addressed conclusively in a future neuroimaging study.

To summarize, we find that the summary statistics of contour curvature as well as contour junction types and angles are important features for natural scene categorization by humans. Statistics of line orientation and length are useful for computational scene categorization but appear to be less important to human observers. These findings cast doubt on the view that the visual system has a separate processing pathway for scene gist, which would rely, for instance, on the statistics of spatial frequencies (Bar, 2004). Our work suggests that non-accidental V2-type structural properties, which emphasize relations between contours, play an important role in the processing of scene content, just as they do in object recognition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Bart Larsen, Monica Craigmile, and Ryan McClincy for help with the running the experiments and two anonymous reviewers as well as Shaun Vecera, Wil Cunningham and members of the BWlab for insightful comments on the manuscript. This research was funded in part by the National Eye Insititute (CRCNS, NIH 1 R01 EY019429).

References

- Alvarez GA, Oliva A. The representation of simple ensemble visual features outside the focus of attention. *Psychol Sci.* 2008; 19(4):392–398.10.1111/j.1467-9280.2008.02098.x [PubMed: 18399893]
- Ariely D. Seeing sets: representation by statistical properties. *Psychological Science.* 2001; 12:157–162. [PubMed: 11340926]
- Balas B, Nakano L, Rosenholtz R. A summary-statistic representation in peripheral vision explains visual crowding. *J Vis.* 2009; 9(12):13, 11–18.10.1167/9.12.13 [PubMed: 20053104]

- Baldassi S, Burr DC. Feature-based integration of orientation signals in visual search. *Vision Res.* 2000; 40(10–12):1293–1300. [PubMed: 10788640]
- Bar M. Visual objects in context. *Nat Rev Neurosci.* 2004; 5(8):617–629.10.1038/nrn1476 [PubMed: 15263892]
- Bauer B. Does Steven's power law for brightness extend to perceptual brightness averaging? *Psychological Record.* 2009; 59:171–186.
- Bell AJ, Sejnowski TJ. The “independent components” of natural scenes are edge filters. *Vision Res.* 1997; 37(23):3327–3338. S0042-6989(97)00121-1 [pii]. [PubMed: 9425547]
- Biederman I. Recognition-by-components: a theory of human image understanding. *Psychol Rev.* 1987; 94(2):115–147. [PubMed: 3575582]
- Burr D, Ross J. A visual sense of number. *Curr Biol.* 2008; 18(6):425–428.10.1016/j.cub.2008.02.052 [PubMed: 18342507]
- Chong SC, Treisman A. Representation of statistical properties. *Vision Res.* 2003; 43(4):393–404. [PubMed: 12535996]
- Choo H, Levinthal BR, Franconeri SL. Average orientation is more accessible through object boundaries than surface features. *J Exp Psychol Hum Percept Perform.* 2012; 38(3):585–588.10.1037/a0026284 [PubMed: 22082217]
- Cohen MA, Alvarez GA, Nakayama K. Natural-scene perception requires attention. *Psychol Sci.* 2011; 22(9):1165–1172.10.1177/0956797611419168 [PubMed: 21841149]
- Evans KK, Treisman A. Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform.* 2005; 31(6):1476–1492.10.1037/0096-1523.31.6.1476 [PubMed: 16366803]
- Fei-Fei L, Iyer A, Koch C, Perona P. What do we perceive in a glance of a real-world scene? *Journal of Vision.* 2007; 7(1):10, 11–29. [PubMed: 17461678]
- Fei-Fei, L.; Perona, P. A Bayesian hierarchical model for learning natural scene categories. Paper presented at the Computer Vision and Pattern Recognition; San Diego, CA. 2005.
- Freeman J, Simoncelli EP. Metamers of the ventral stream. *Nat Neurosci.* 2011; 14(9):1195–1201.10.1038/nn.2889 [PubMed: 21841776]
- Goffaux V, Jacques C, Mouraux A, Oliva A, Schyns PG, Rossion B. Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition.* 2005; 12(6):878–892.
- Greene MR, Oliva A. The briefest of glances: the time course of natural scene understanding. *Psychol Sci.* 2009; 20(4):464–472. [PubMed: 19399976]
- Hegde J, Van Essen DC. Selectivity for complex shapes in primate visual area V2. *J Neurosci.* 2000; 20(5):RC61. [PubMed: 10684908]
- Karklin Y, Lewicki MS. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature.* 2009; 457(7225):83–86. [PubMed: 19020501]
- Kim JG, Biederman I. Where do objects become scenes? *Cereb Cortex.* 2011; 21(8):1738–1746.10.1093/cercor/bhq240 [PubMed: 21148087]
- Koenderink JJ, van Doorn AJ. Representation of local geometry in the visual system. *Biol Cybern.* 1987; 55(6):367–375. [PubMed: 3567240]
- Larson AM, Loschky LC. The contributions of central versus peripheral vision to scene gist recognition. *J Vis.* 2009; 9(10):6, 1–16.10.1167/9.10.6 [PubMed: 19810787]
- Li FF, VanRullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America.* 2002; 99(14):9596–9601. [PubMed: 12077298]
- Loschky LC, Hansen BC, Sethi A, Pydimarri TN. The role of higher order image statistics in masking scene gist recognition. *Atten Percept Psychophys.* 2010; 72(2):427–444. [PubMed: 20139457]
- Marr, D. *Vision.* New York: W. H. Freeman and Company; 1982.
- Marr D, Nishahara HK. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc London B.* 1978; 200:269–294. [PubMed: 24223]
- Oliva A, Schyns PG. Diagnostic colors mediate scene recognition. *Cogn Psychol.* 2000; 41(2):176–210.10.1006/cogp.1999.0728 [PubMed: 10968925]

- Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*. 2001; 42:145–175.
- Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996; 381(6583):607–609. [PubMed: 8637596]
- Park S, Chun MM. Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*. 2009; 47(4):1747–1756.10.1016/j.neuroimage.2009.04.058 [PubMed: 19398014]
- Pasupathy A, Connor CE. Population coding of shape in area V4. *Nat Neurosci*. 2002; 5(12):1332–1338.10.1038/nn972 [PubMed: 12426571]
- Pelli DG, Palomares M, Majaj NJ. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J Vis*. 2004; 4(12):1136–1169. 10:1167/4.12.12. [PubMed: 15669917]
- Peterhans E, von der Heydt R. Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *J Neurosci*. 1989; 9(5):1749–1763. [PubMed: 2723748]
- Portilla J, Simoncelli EP. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*. 2000; 40(1):49–71.
- Potter MC, Levy EI. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*. 1969; 81(1):10–15. [PubMed: 5812164]
- Renninger LW, Malik J. When is scene identification just texture recognition? *Vision Res*. 2004; 44(19):2301–2311.10.1016/j.visres.2004.04.006 [PubMed: 15208015]
- Rubin N. The role of junctions in surface completion and contour matching. *Perception*. 2001; 30(3):339–366. [PubMed: 11374204]
- Sayim B, Cavanagh P. What line drawings reveal about the visual brain. *Front Hum Neurosci*. 2011; 5:118.10.3389/fnhum.2011.00118 [PubMed: 22065509]
- Torralba A, Oliva A. Statistics of natural image categories. *Network*. 2003; 14(3):391–412. [PubMed: 12938764]
- Torralbo A, Walther DB, Chai B, Caddigan E, Fei-Fei L, Beck DM. Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PLoS One*. 2013; 8(3):e58594.10.1371/journal.pone.0058594 [PubMed: 23555588]
- Tversky B, Hemenway K. Categories of Environmental Scenes. *Cognitive Psychology*. 1983; 15:121–149.
- Walther DB, Caddigan E, Fei-Fei L, Beck DM. Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci*. 2009; 29(34):10573–10581.10.1523/JNEUROSCI.0559-09.2009 [PubMed: 19710310]
- Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L. Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc Natl Acad Sci U S A*. 2011; 108(23):9661–9666.10.1073/pnas.1015666108 [PubMed: 21593417]

Figure 1.

A. Example photograph of a city street scene and the corresponding line drawing with contour and junction properties indicated by color. B. Average histograms over the five properties of line drawings for each of the six categories. Original photograph by flickr user aleph.pukk.

Figure 2.

A. Confusion matrices for the computational analysis. B. Confusion matrices for experiment 1. C. Categorization accuracy for the computational analysis of line drawings with different sets of features and behavioral accuracy for line drawings and photographs (N = 18). The dashed line denotes chance level (1/6). D. Correlation of computational with behavioral errors for line drawings and photographs. ** $p < 0.01$, *** $p < 0.001$

Figure 3.

A. Example of an intact line drawing (left) and the same line drawing with randomly shifted contours (right); B. Categorization accuracy of the computational analysis and behavior ($N = 23$) for intact and shifted line drawings. The three contour-related properties (orientation, length, curvature) have the exact same values for contour-shifted as for intact line drawings. The dashed line denotes chance level ($1/6$); C. Error correlation of human responses to contour-shifted line drawings with computational properties of intact line drawings. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, † $p = 0.057$.