Vision Research 126 (2016) 308–317

Contents lists available at ScienceDirect

Vision Research

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

Figure–ground segregation: A fully nonlocal approach

Mariella Dimiccoli *

Barcelona Perception Computing Lab (BCNPCL), Computer Vision Center (CVC) and University of Barcelona (UB), Barcelona, Spain

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form 15 March 2015
Available online 28 March 2015

Keywords:
Figure–ground segregation
Nonlocal approach
Directional linear voting
Nonlinear diffusion

ABSTRACT

We present a computational model that computes and integrates in a nonlocal fashion several configural cues for automatic figure–ground segregation. Our working hypothesis is that the figural status of each pixel is a nonlocal function of several geometric shape properties and it can be estimated without explicitly relying on object boundaries. The methodology is grounded on two elements: multi-directional linear voting and nonlinear diffusion. A first estimation of the figural status of each pixel is obtained as a result of a voting process, in which several differently oriented line-shaped neighborhoods vote to express their belief about the figural status of the pixel. A nonlinear diffusion process is then applied to enforce the coherence of figural status estimates among perceptually homogeneous regions. Computer simulations fit human perception and match the experimental evidence that several cues cooperate in defining figure–ground segregation. The results of this work suggest that figure–ground segregation involves feedback from cells with larger receptive fields in higher visual cortical areas.

1. Introduction

Figure–ground is a particular kind of organizational phenomenon that determines the interpretation of a visual scene into figures (object-like regions) and grounds (background-like regions), thus enabling higher-level processing such as the perception of surfaces, shapes and objects. Understanding the laws underlying figure–ground organization has been a major focus of attention for Gestalt Psychologists: Rubin (1921), followed by Bahnsen (1928), Metzger and Wolfgang (1975), Kanizsa and Gerbino (1976) demonstrated that figure–ground perception is governed by several generic shape properties, such as region size, surroundness, symmetry, parallelism and convexity among others (see Fig. 1).

More recently, a few figure–ground principles that also apply to static, homogeneously colored regions have been discovered (e.g. lower region Vecera, Vogel, & Woodman (2002), and top-bottom polarity Hulleman & Humphreys (2004)) and evidence against the innate nature of Gestalt laws has accumulated both from direct reports (Gibson & Peterson, 1994; Peterson & Gibson, 1994; Peterson, Harvey, & Weidenbacher, 1991) and indirect measures (Driver & Baylis, 1996; Peterson & Skow, 2008; Vecera & Farah, 1997). Furthermore, the idea of Brunswik and Kamiya (1953), following which Gestalt cues reflect the statistics of the natural world in which the visual system evolved, has been validated to some extent by studying and quantifying the correlation between some statistical configural properties and depth in natural images (Burge, Fowlkes, & Banks, 2010; Fowlkes, 2007). Taken together, these findings are consistent with the claims that figure–ground segregation can occur preattentively, but it can also be affected by attention and that past experience can exert an influence on several aspects of figure–ground perception. Classic configural cues are not innate but are the results of a sophisticated learning mechanism that has evolved to allow humans to extract the statistical properties of the environment in which they live.

Despite the advances of the last century, the computational mechanisms underlying figure–ground perception are still poorly understood and of interest to both neuroscientists (Domijan & Šetić, 2008; Heitger & von der Heydt, 1993; Pao, Geiger, & Rubin, 1999) and computer vision researchers (Calderero & Caselles, 2013; Dimiccoli, 2009; Ren, Malik, & Fowlkes, 2005; Ren, Fowlkes, & Malik, 2006). Indeed, with the goal of providing input to high-level tasks such as shape recognition and 3D recovery, computer vision researchers are placing an increasing emphasis on the specific problem of figure–ground segregation and to the more general problem of monocular depth estimation in natural images. As outlined in Rubin (2001), one of the central issues concerns the way multiple configural cues yield a unitary percept. Similarly to the laws of perceptual grouping, all figure–ground principles tend to be treated as qualitative ceteris paribus rules, in which a given factor has a stated effect when all others are neutralized. As a consequence, they are unable to predict the outcome

* Addresses: Computer Vision Center, Edifici O, 08193 Bellaterra (Cerdanyola del Vallés), Barcelona, Spain. Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain.
E-mail address: mdimiccoli@cvc.uab.es

http://dx.doi.org/10.1016/j.visres.2015.03.007
0042-6989/

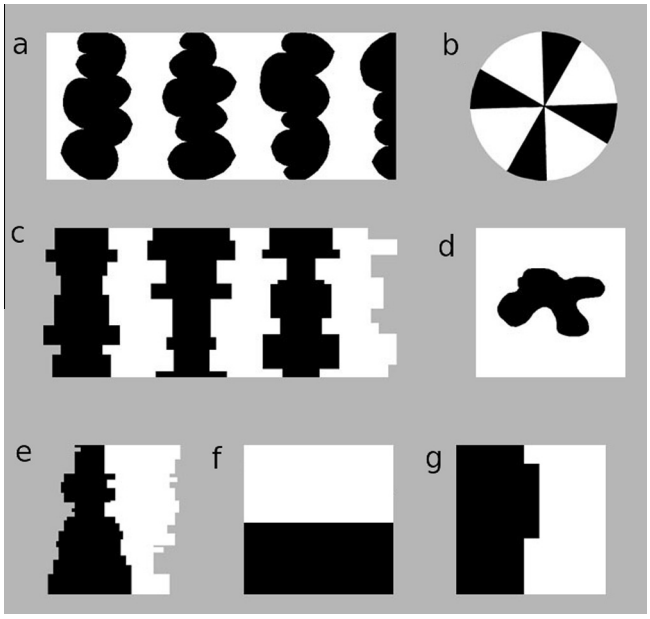

Fig. 1. Adapted from Kubovy Perception Lab website (2009). (a) Convexity: the black shapes are perceived as figures because, contrary to white regions, they are limited by piecewise convex boundaries. (b) Smallness: smaller areas (in black) tend to be seen as figures against a larger background. (c) Symmetry: the black shapes are perceived as figure surrounded by the white one because of theur symmetry. (d) Surroundedness: the surrounded area (in black) tends to be perceived as figure. (e) Bottom-up polarity: the black region is perceived as figure beacause its base is wider than the contiguous white region. (f) Lower region: the lower region in the visual field (in black) is perceived as figure. (g) Protrusion: the black region on the left protrudes into the white region and it is perceived as figure.

when several conflicting factors are at work in the same display. Kanizsa posed the problem in terms of competitive interactions formulated in descriptive terms and demonstrated that convexity has a stronger influence than other global shape properties, such as symmetry and contrast polarity (Kanizsa, 1979).

In general, relatively little work has been done to understand the theoretical and neural basis underlying these interactions. A relevant attempt has been made by Kienker et al. (1986), who proposed an interesting neural network architecture but without systematic evaluation against human perception in terms of figure–ground rules and their combinations. Stanley and Rubin (2003) showed through functional Magnetic Resonance Imaging (MRI) that illusory contours elicit responses in the Lateral Occipital Complex (LOC), whose neurons pool information from large portions of the visual field. They also showed that salient regions (figure) activated the LOC but they did not go into the question of which cues are used by the brain to detect them and how these image cues are computed. Domijan and Šetić (2008) proposed a neural model based on the interaction between the dorsal and the ventral streams. The ventral stream computes object boundaries which are used to construct surface representation, whereas the dorsal stream computes saliency based on a blurred version of the boundary signals. Their model can account for how classic and recently discovered principles of figure–ground assignment influence the perception of the figure. However, their results contradict physiological findings about border ownership responses in V2 (Zhou, Friedman, & Von Der Heydt, 2000).

Most computational models by computer vision researchers have focused on how to combine local and global information more than how to robustly detect configural cues. The general approach is first to compute several local cues and than to enforce global consistency using different frameworks (Calderero & Caselles, 2013; Dimiccoli, 2009; Ren et al., 2005, 2006). An exception is the work of Calderero and Caselles (2013), which does not deal with cue detection explicitly but in which occlusion arises naturally with the image model leading to local estimations of border ownership.

This manuscript proposes a new computational model for figure–ground segregation, in which several geometrical shape properties, namely convexity, size, surroundness and lower region, are estimated in a nonlocal fashion, without explicitely relying on previously computed image boundaries. This leads to very robust and independent estimations of different configural cues, from which unitary figure–ground percepts can be inferred through a very simple integration mechanism.

2. Material and methods

Sampled images contain a finite number of values on a grid: in the biological case the elements of the grid are hexagonal cells with growing sizes from the fovea, whereas in the digital case the grid corresponds to a Charge Coupled Device (CCD) matrix. In the latter case, the elements of the grid together with their intensity values, which encode the number of photon hits during a fixed exposure time, are called pixels. A digital image is usually modeled as a real valued discrete function defined on a rectangular domain: $u : \Omega \subset \mathcal{Z}^2 \rightarrow \mathcal{R}$, where $\mathcal{R}$ is the set of real numbers, $\mathcal{Z}$ is the set of integer numbers, $\Omega = [0, N] \times [0, M]$ is a rectangle, $\Gamma = \partial\Omega$ is its boundary. In this section, we will focus on binary digital images, say $u : \Omega \subset \mathcal{Z}^2 \rightarrow \{0, 1\}$, corresponding to F/G displays. Following Mathematical Morphology (Serra, 1983), a digital binary image is fully characterized by its upper level set $\mathcal{X}_\lambda = \{x \in \Omega : u(x) \geqslant \lambda\},\ \lambda = 1$, that is the set of pixels with positive value.

A level set may be composed by one or by the union of several connected components (*cc*), usually referred to in Mathematical Morphology (Serra, 1983) as the *shapes* of the image: $\mathcal{X}_\lambda = \cup_{i=1}^{n} cc_i(\mathcal{X}_\lambda)$. We are interested in measuring the properties of these *image shapes* related to figure–ground assignment. To this goal, we first recall the definition of convexity of a set in $\mathcal{R}^2$, which is as follows.

Definition 1 (*Convex sets*). A set $A \subset \mathcal{R}^2$ is convex if for any $x, y \in A$ and any $\lambda \in [0, 1], \lambda \in \mathcal{R} : \{(1 - \lambda)x + \lambda y\} \in A$

Fig. 2 is an example of not convex set: the segment having as extremes the points x and y of A partially lies outside A. The following proposition shows that only boundary points matter as far as determining the convexity of a non-empty closed set.

Proposition 1 (*Characterization of convex sets*). *Let A be a non-empty closed subset of $\mathcal{R}^2$ and let intA and ∂A be its interior and its boundary respectively. A is convex if and only if for any $x, y \in \partial A$ and any $\lambda \in [0, 1], \lambda \in \mathcal{R} : \{(1 - \lambda)x + \lambda y\} \setminus \{x, y\}$ is either in ∂A or intA.*

This characterization, whose proof is provided in Beltagy and Shenawy (2013, corollary 5), can be used to evaluate the convexity of morphological *shapes* on a binary digital image. To do that

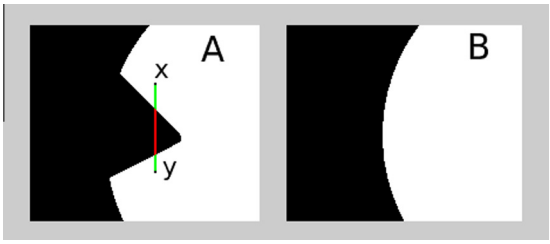

Fig. 2. In a strict mathematical sense, A is a not convex shape (left) and B is a convex shape (right).

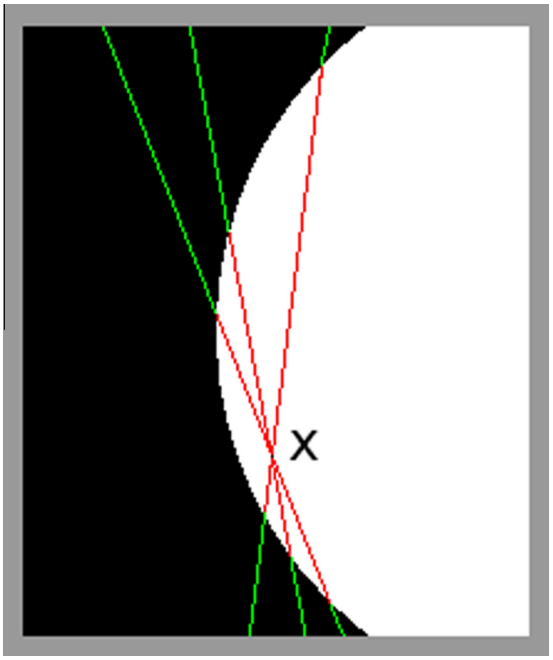

Fig. 3. Example of line-shaped neighborhoods of a pixel x partitioned into three segments, each composed of homogeneous pixels. In this display, points beloging to the central red segment can be understood as beloging to a convex shape and its extremes as points belonging to the boundary of a convex shape. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

following strictly the definition, it would be needed to first compute, on the bilinear interpolated image, the boundaries of the level set, the so called *level lines*. However, we will show how the convexity of a shape can be evaluated without directly relying on level set boundaries.

Let us suppose that we are given the image u and we want to automatically estimate if each pixel $x \in \Omega$ belongs to a convex shape of u or not. In view of Proposition 1, this is equivalent to evaluating whether each pixel $x \in \Omega$ belongs to a segment that lies on the boundary or on the interior of a convex shape. What could we do in this case? If we consider a line l passing through x with slope θ, say $l_x(\theta)$, we could easily partition this line into L segments, say $s_x(\theta)^i, i = 1, \ldots, L$, by grouping similar adjacent pixels (see Fig. 3). The extremes of the segment including x, can be thought as points of the hypothetical contour of the region to which the segment belongs. Hence, the fact that this segment is between two others, that by construction have different pixel values, can be related through Proposition 1 to the belief that x is inside a convex shape. Therefore, a positive vote is given to x which reflects the belief for the statement *"x is part of a convex shape"* of its line-shaped neigborhood $l_x(\theta)$. Of course, we can iterate the same reasoning for a discrete number of directions by varying θ and sum their contributions to the estimation of the figural status of x. If we do the same thing for each image pixel, only points inside the convex shape will cast positive votes.

In Fig. 4(b) (first row), the result of the multi-directional linear voting is visualized through a gray level image, say $z : \Omega \rightarrow \mathcal{R}$, where the intensity of each pixel encodes the degree of belief about the figural status of all its line-shaped neighborhoods, with higher beliefs being visualized through higher intensity values. To reduce the computational burden and avoid redundancy, instead of parsing all image pixels and considering multiple lines through each of them, we consider only the lines through the pixels of the image boundary: when a segment is between two others on a line, a positive vote is assigned to all pixels of the segment.

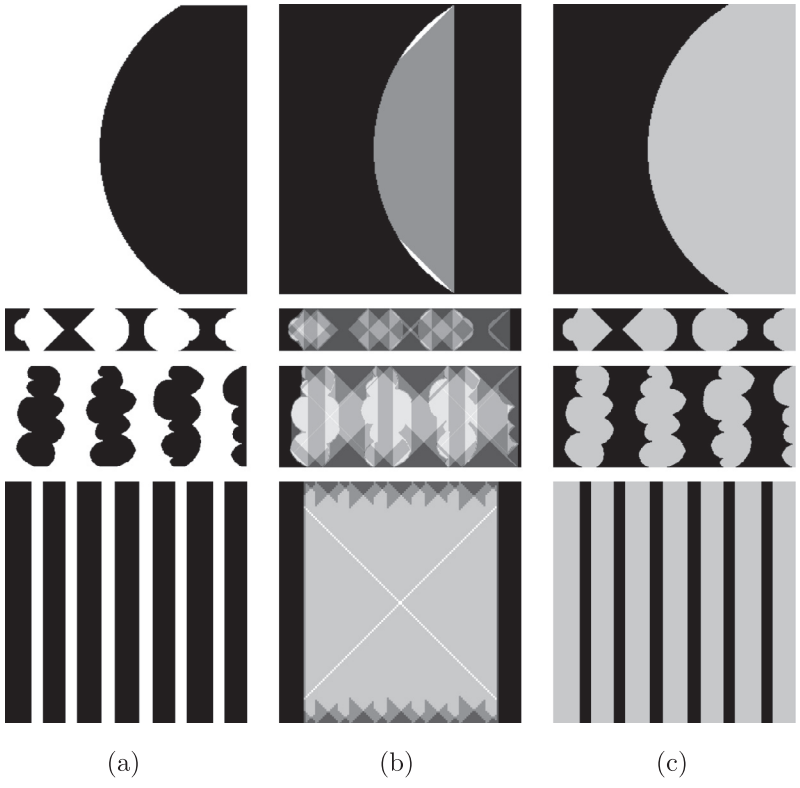

Fig. 4. Modeling convexity: (a) original image. (b) Image whose intensity values correspond to the beliefs computed through multi-directional linear voting taking into account only convexity. Pixels with high intensity values have an high belief to be part of a foreground region. (c) Normalized probability of foreground after averaging on homogeneous regions.

What happens if the shape is perceived as globally convex even if it is not convex in the strict mathematical sense? As outlined by Pao et al. (1999), human perception behaves in a "continuous" manner, perceiving as figure the *more convex* region. Therefore, a model of figure–ground segregation based on convexity should offer a continuous measure of convexity. In our case, some points outside the global convex shape will cast positive votes (see Fig. 4(b) second and third rows). Since we are interested in measuring the convexity of image *shapes*, it makes sense to average the belief of pixels which belong to the same level set of u. This can be achieved through a neighborhood filter, which performs an average of the values of pixels in the belief image z which are similar in gray level value in the original image u. As in the NL-means (Buades, Coll, & Morel, 2006), the average is computed in a fully nonlocal way on all image domain. Denoting by $\mathcal{X}_\lambda$ the unique upper level set of the binary image u, the neighborhood filter computes for each image pixel a weighted average:

$$\bar{z}(x) = \frac{1}{C(x)} \sum_{y \in \Omega} z(y) \exp\left(-\frac{d_E(u(x) - u(y))}{h}\right), \tag{1}$$

where h is a filtering parameter, d_E is the Euclidean distance and $C(x) = \sum_{y \in \Omega} \exp\left(-\frac{d_E(u(x)-u(y))}{h}\right)$ is the normalizing factor. Note that, since the image is binary the exponential term is equal to one for similar pixels. Hence, this is equivalent to simply averaging the votes on the upper level set and on its complement. For more details about neighborhood filters and their applications to the recovering of 3D information, the reader is referred to Digne et al. (2011).

In the following, we will show how multi-directional linear voting can be generalized as a strategy to measure in nonlocal way many geometrical shape properties, not just convexity. Consider the leftmost image on the fourth row of Fig. 4: in this case, the only cue for figural organization is small size. Taking into account sizeness requires favoring small shapes. This can be achieved by involving the length of the segments: if the segment $s_x^i(\theta)$ has a length $\mathcal{L}(s_x^i(\theta))$ smaller than those of its adjacent segments, that is if $\mathcal{L}(s_x^i(\theta)) < \min\{\mathcal{L}(s_x^{i-1}(\theta)), \mathcal{L}(s_x^{i+1}(\theta))\}$, than all pixels of $s_x^i(\theta)$ will receive a positive vote. To remove border effects, the intervals intersecting the image borders are considered to have infinite length. In Fig. 5, are shown the results of simultaneously taking into account convexity and size: for all image display, the interpretation given by our model matches reports on human observers.

Fig. 6 shows an image display where convexity, size and surroundness are acting at the same time. In this case, when taking into account only convexity (see first row), the interpretation agrees with human perception, whereas when taking into account convexity and size simultaneously, it is no longer the case (see second row). This because the role of surroundedness, which is crucial in this display, has been neglected. To model surroundedness through the multi-directional linear voting framework, we proceed as follows (see Fig. 7): let $cc_x^i(\theta)$ be the connected component of $s_x^i(\theta)$ and be $(cc_x^i(\theta))^C = \Omega - cc_x^i(\theta)$ its complement. The intersection of the line $l_x(\theta)$ with the set $(cc_x^i(\theta))^C$ is the set of segments $\left\{s_x^j(\theta)\right\}$, $j = 1, \ldots, M$. Given $l_x(\theta)$ and the segment $s_x^i(\theta)$ such that $x \in s_x^i(\theta)$ lying on it, we compare the length of $s_x^i(\theta)$ to those of its adjacent segments in the set $\left\{s_x^j(\theta)\right\}$, $j = 1, \ldots, M$. As it is shown in the last row of Fig. 6, this leads to a human-like interpretation.

Other configural cues such as lower region and bottom-up polarity can be easily integrated into this framework giving a special emphasis to the vertical orientation. This is consistent with psychophysical reports showing better performance in the lower visual field for a number of visual tasks (Levine & McAnany, 2005; McAnany & Levine, 2007; Previc, 1990; Rubin, Nakayama, & Shapley, 1996). Since the distinction between the lower and the

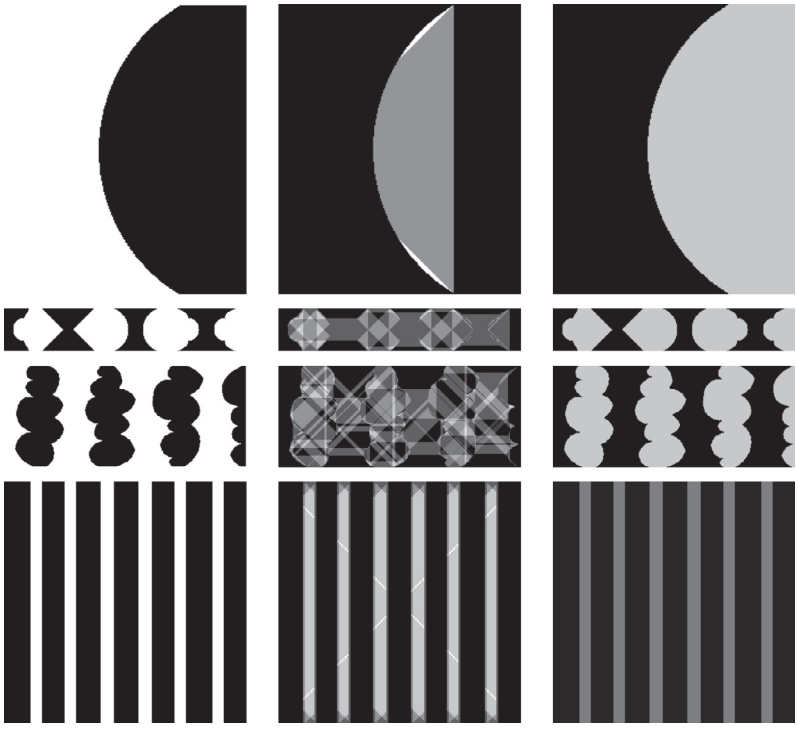

Fig. 5. Modeling convexity and size: (a) original image. (b) Image whose intensity values correspond to the beliefs computed through multi-directional linear voting taking into account convexity and small size. Pixels with high intensity values have an high belief to be part of a foreground region. (c) Normalized probability of foreground after averaging on homogeneous regions.

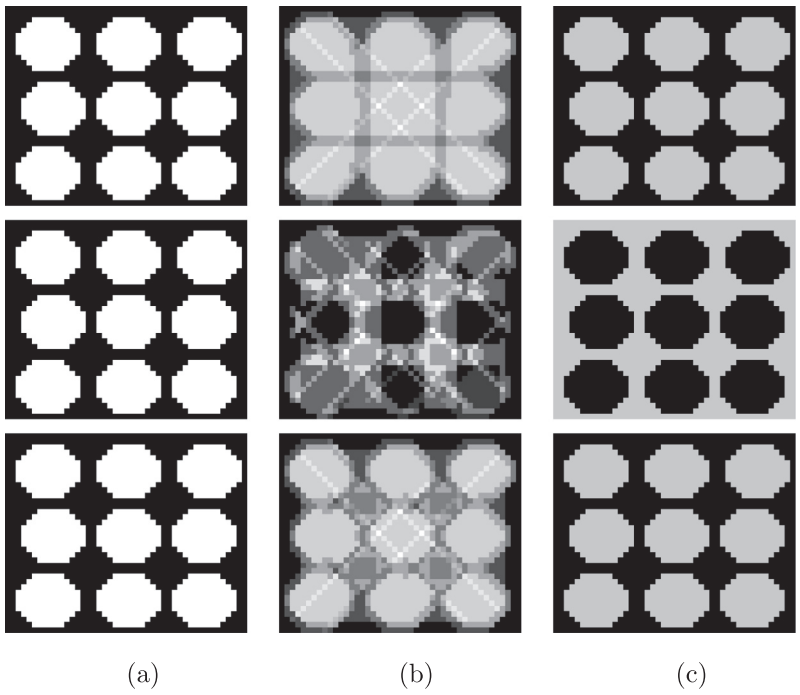

Fig. 6. (a) Original image. (b) Image whose intensity values correspond to the beliefs computed through multi-directional linear voting taking into account convexity (first row), convexity and small size (second row), convexity, size and surroundness (third row). Pixels with high intensity values have an high belief to be part of a foreground region. (c) Normalized probability of foreground after averaging on homogeneous regions.

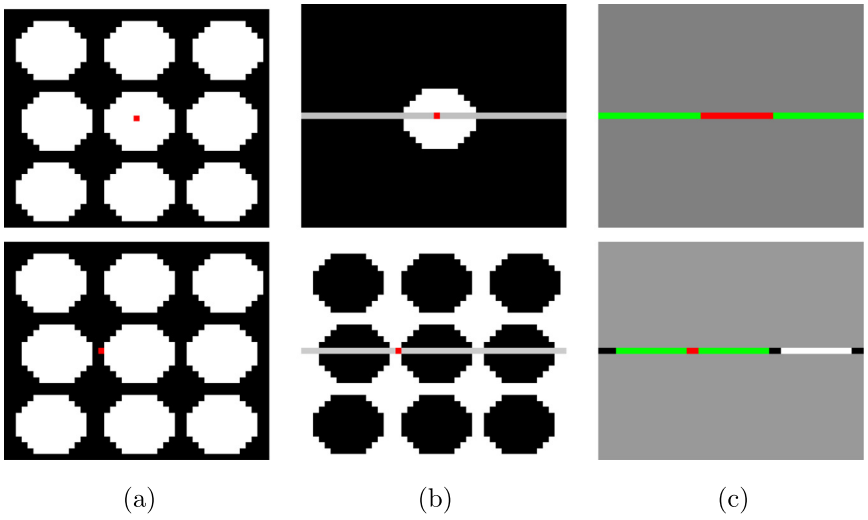

Fig. 7. Modeling surroundedness: (a) original image with a pixel, say x, marked in red. (b) The horizontal line through x, say $l_x(\theta)$ with $\theta = 0$, is depicted in gray; the connected component of x, say cc_x^i is depicted in white and its complementary $(cc_x^i)^C$ is in black. (c) The segment s_x^i to which x belongs is in red: its length is compared to the length of the two adjacent green segments, given by the intersection of $l_x(\theta)$ with $(cc_x^i)^C$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

higher part of an image is most clear in the vertical direction, we consider only vertical line-shaped neighborhoods whose intersection with the image domain give rise to just two segments. In this case, to favor regions in the lower part of the image, a positive vote is given to all pixels in the segment that cross the lower image border.

The pseudocode and the source code are available as Supplementary Material.

3. Results

This work has proposed a novel computational model able to compute and integrate, under a common framework, several ceteris paribus rules for figure–ground segregation. The proposed model has a feature that differs from the state of the art: configural cues are quantified without explicitly relying on object boundaries but are based on global relations between image regions. This approach leads to very robust estimations for which a simple mechanism, namely a linear summation followed by a nonlinear diffusion, can account for how these cues interact.

Fig. 8 shows the results of computer simulations performed on image displays involving a lower region (first three rows), bottom up polarity (fourth row) and protrusion (last row). Although protrusion and bottom-up polarity are not explicitly modeled, they seem to be a consequence of the general sensitivity to global convexity and lower-region respectively.

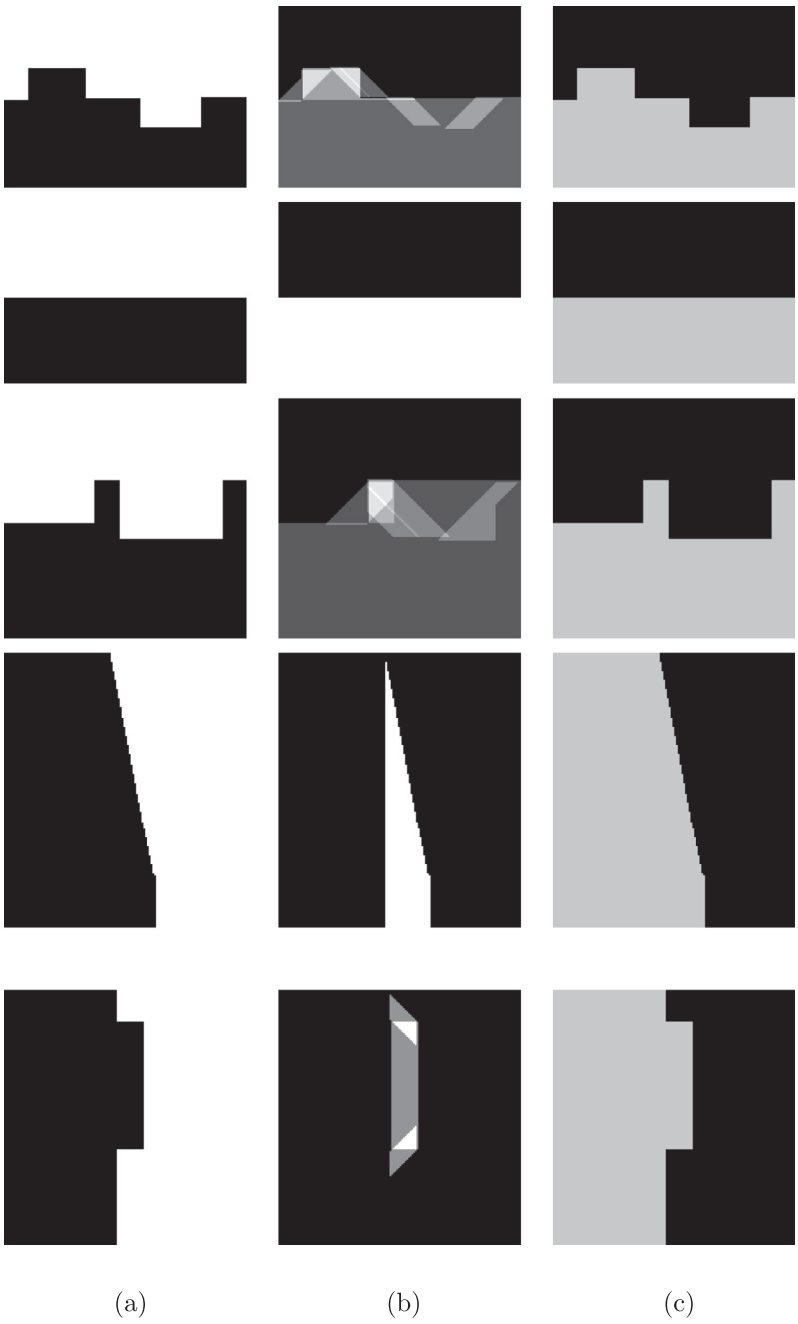

Fig. 8. Lower-region, bottop-up polarity and protrusion: (a) original image. (b) Image whose intensity values correspond to the beliefs computed through multi-directional linear voting: pixels with high intensity values have an high belief to be part of a foreground region. (c) Probability of foreground averaged on homogeneous regions and normalized for visualization.

The computer simulations in Figs. 9 and 10 model the cues of convexity, size, surroundness and lower region. These simulations illustrate the model ability to account for all these cues as well as for their interactions. On the last row of Fig. 9 there is an image display with a not simply connected component (because of the presence of the hole): also in this case, averaging the results of the multi-directional linear voting on similar regions leads to an human-agreed interpretation.

The computer simulations in Fig. 11 illustrate the model's ability to assign figural status to the occluding surface. Since these image displays are not binary, we first decompose the image into bi-level sets ($\mathcal{X}_{\lambda_1\lambda_2} = \{x|\lambda_1 \leqslant u(x) \leqslant \lambda_2\}$), than we apply the proposed method to each bi-level set separately and finally we linearly sum the results. On the first row of Fig. 11, the small circular surface is placed on the large circular surface and our model gives the correct interpretation. On the second row, we presented two overlapping squares: in this case, the model correctly selects the occluding surface. In fact, when we inspect the result of the multi-directional linear voting (see Fig. 11(f) second row), we see high values at the location of the missing (or amodal) corner of the occluded surface, whose presence is implicated by the boundaries of the occluded surface.

On the last row is illustrated a limitation of our model: in the presence of objects in partial occlusion, the small partially occluded regions appear in the foreground, independent of the depth order suggested by the T-junctions (see Fig. 11). This is not

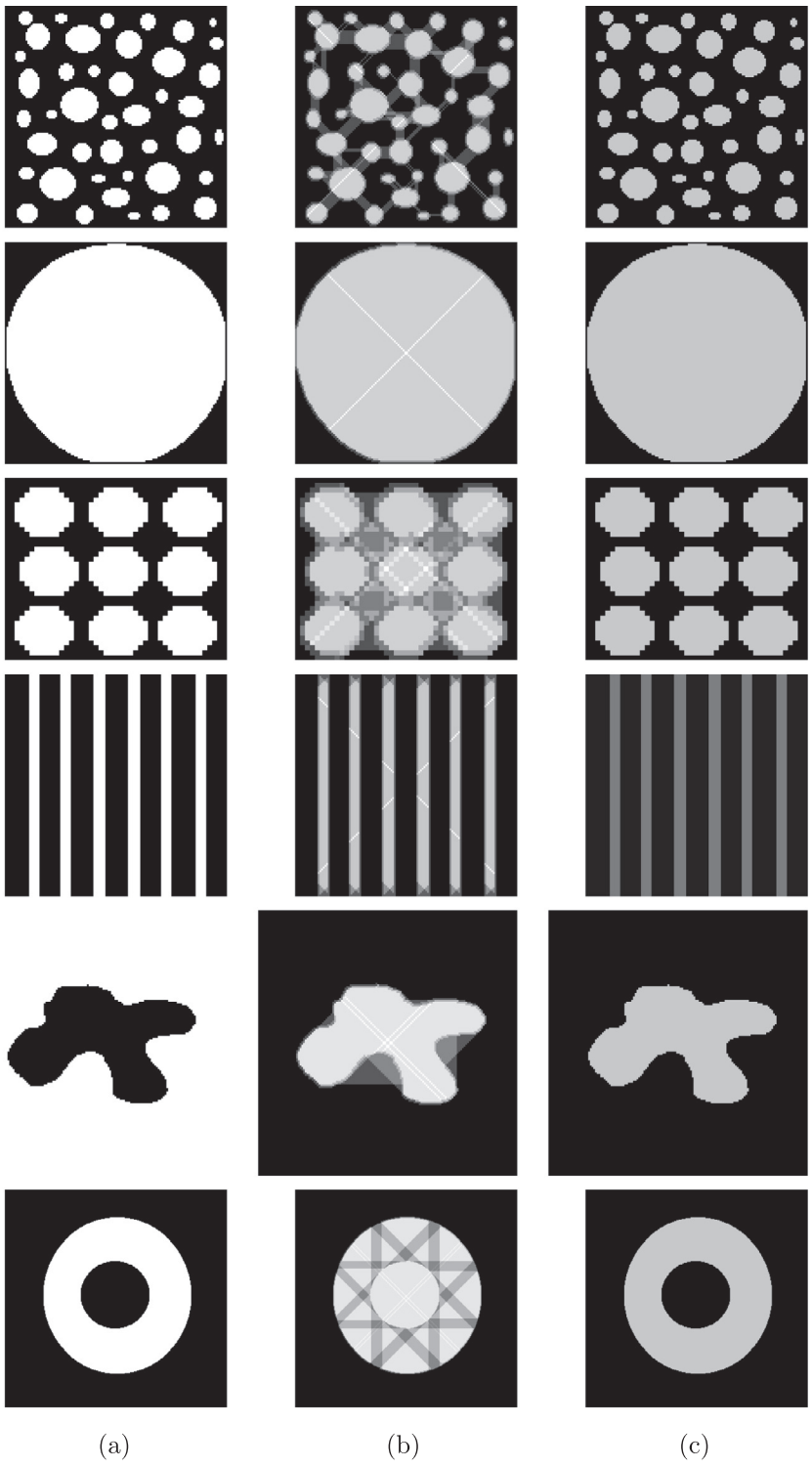

Fig. 9. (a) Original image. (b) Image whose intensity values correspond to the beliefs computed through multi-directional linear voting: pixels with high intensity values have an high belief to be part of a foreground region. (c) Probability of foreground averaged on homogeneous regions and normalized for visualization.

surprising considering that our model is based on image shape properties and does not take into account the principle of amodal completion induced by T-junctions.

Hovewer, if it were possible to model amodal completion and reconstruct the partially occluded object following global theories of occlusion perception (Boselie, 1994), linear summation would still be a good strategy to extract the three levels of depth (see Fig. 12).

Finally, an extension of our model to natural images would only require a method to extract meaningful bi-levels sets from images. This could for instance be done by taking into account color and texture properties.

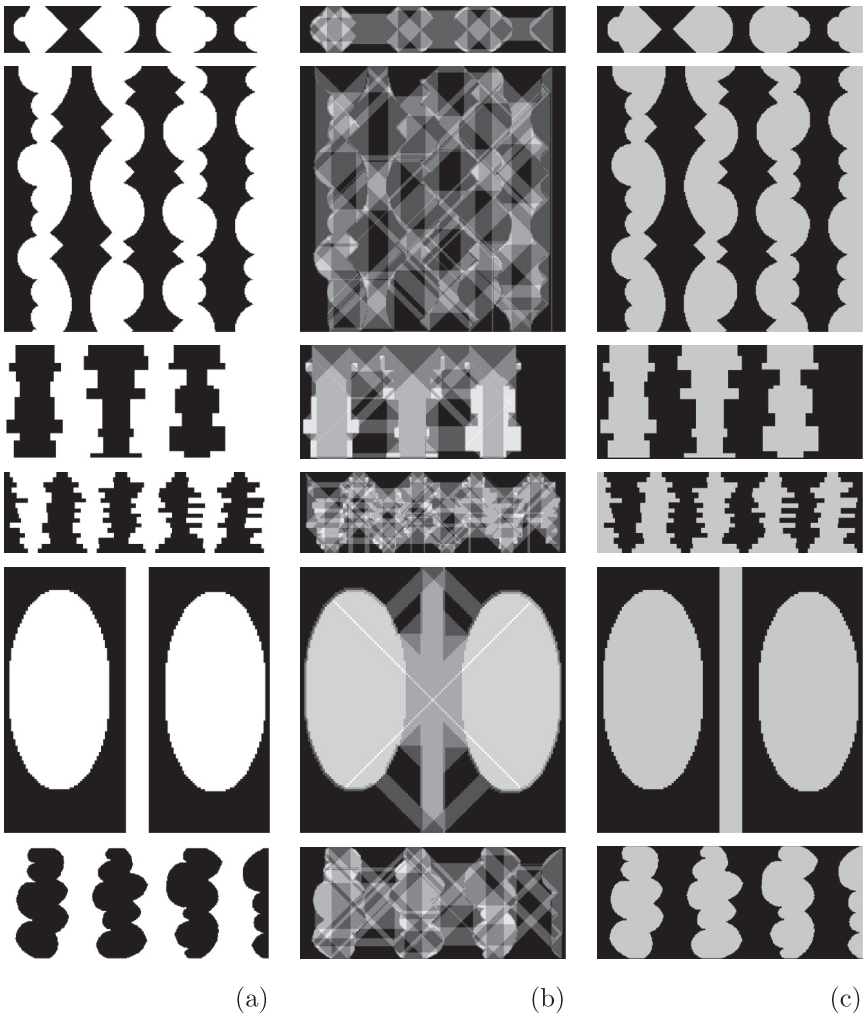

Fig. 10. (a) Original image. (b) Image whose intensity values correspond to the beliefs computed through multi-directional linear voting: pixels with high intensity values have an high belief to be part of a foreground region. (c) Probability of foreground averaged on homogeneous regions and normalized for a better visualization.

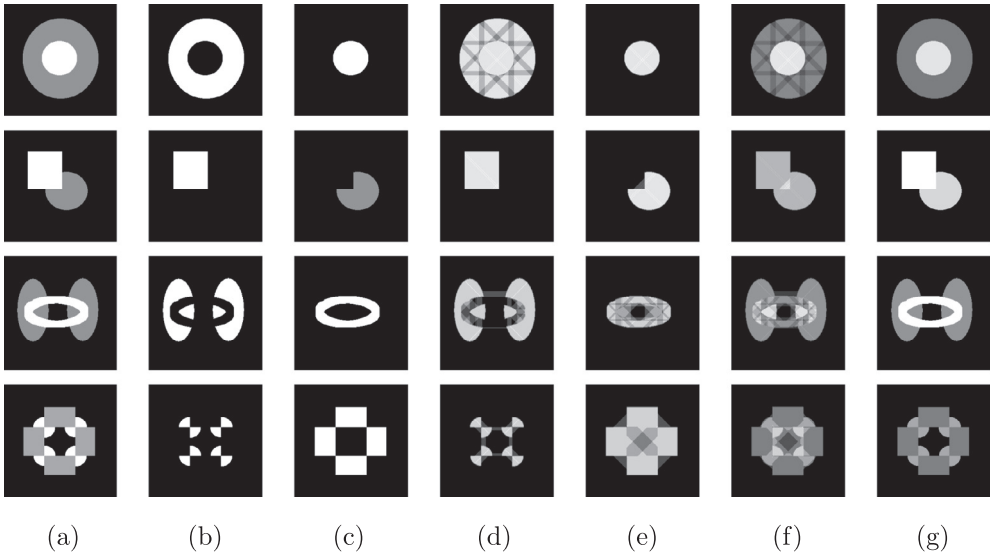

Fig. 11. (a) Original image. (b) and (c) are the binary images correspoding to the level sets of the original image. (d) and (e) are the beliefs computed on (b) and (c) respectively. (f) Sum of the beliefs on each level set. (g) Result of averaging the belief on homogeneous regions.

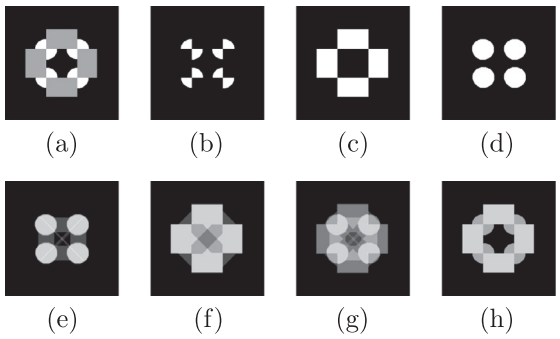

Fig. 12. (a) Original image. (b) and (c) are the binary images correspoding to the level sets of the original image. (d) Completion of image (b) taking (c) as hole. (e) and (f) are the beliefs computed on (d) and (c) respectively. (g) Sum of the the beliefs on each level set. (h) Result of averaging the beliefs on homogeneous regions.

4. Discussion

The proposed model does not pretend to have a biological basis, rather it uses interesting computational mechanims to emulate the outcome of the visual system.

Configural cues are classically computed by relying on the activity of cells in the primary visual cortex. These cells are characterized by small receptive fields, from which only elementary features such as local luminance or chrominance can be extracted. In our model, the computation of configural cues is performed non-locally: the figural status of each pixel relying on a specific shape property is estimated by analyzing line-shaped neighborhoods in multiple directions and by linearly summing their contributions. Performing such computations would require cells with large complex receptive fields, formed from input by cells at a lower level in the visual system, hence supporting the finding that figure–ground organization involves neurons in the higher cortical visual areas. Cortical hypercomplex cells, which are sensitive to direction, orientation and length are good candidates to be involved in this task. Actually, the estimation of the figural status at a given point of the retina, would require collecting and integrating nonlinear information from several hypercomplex cells.

Despite the simplicity of its formulation, the multi-directional voting strategy appears flexible enough to model much of the figure–ground information available from global shapes: convexity, sizeness, surroundedness and lower region. Interestingly, the effects of protrusion and bottom-up polarity emerged via global convexity and lower regions. Actually, it seems that these local cues are redundant since they are just different manifestations of a more general global cue. The issue of which cues would be needed seems to be related to their ecological validity. Fowlkes (2007) quantified the relative power of local configural cues in natural images and showed that convexity, size and lower region are the most important from an ecological point of view. The authors did not include in their study surroundness since they were focusing on local configural cues. However, intuitively, it could be argued that surroundness has a strong ecological validity since figural regions, being the projection of the objects closer to the viewpoint, tend to be surrounded by objects farther from the viewpoint.

How the visual system has adapted over time to these statistical properties of the world? As argued in van der Helm (2011), this evolution could have been guided either by the likelihood principle von Helmholtz (1962), that aims at ensuring external veridicality, or by the simplicity principle (Koffka, 1935), a modern information-theoretic translation of the law of Pragnanz that promotes internal efficiency. Following the likelihood principle, evolution may have selected a natural-statistics special-purpose system, that is, a system which is highly adapted to one specific environment. Following the simplicity principle, evolution may have selected an innate general-purpose system, that is, a system which is fairly adapted to many different environments but still sufficiently veridical in everyday perception. This is currently an ongoing issue of debate.

The integration of configural cues has frequently been addressed in computational and human vision. Since studies of neurophysiology (Zhou et al., 2000) suggest the presence of figure–ground processing in V2 as soon as 25 ms after response onset, figure–ground segregation is commonly thought to start from local cues, which can be available at this time. On the computational side, integrating local cues into an unified percept implies a more difficult integration process since local cues are inherently ambiguous. Contrary to state-of-the-art models, in this work the integration is performed by first linearly summing the beliefs based on different shape properties (Kastner & McMains, 2007) and then averaging their values on similar image regions through a nonlinear diffusion process. Since our model is based on very simple operations such as linear summation and nonlinear diffusion, it can be easily parallelized, suggesting that figure–ground segregation can be performed in visual cortex by a large number of neurons working together in parallel as proposed in (Kienker et al., 1986). In addition, the fact that surface segregation can be computed without relying on image boundaries supports the hypothesis of Palmer and Rock (1994) following which grouping and parsing depend on shape properties which are well defined only after boundaries have been assigned (Palmer & Rock, 1994).

If we consider the beliefs as activation levels of neurons on the image grid, the proposed model seems to predict that V1 neurons respond more strongly to figure than background regions (Li, 2003) and that low-level cues suffice to explain figure–ground segregation on image displays involving multiple cues, without invoking top-down feedback. Therefore our model seems to suggest the plausibility of the hypothesis that top-down mechanisms only play a modulatory role in the perception of border ownership.

Summarizing, based on the results of this work, configural cues such as convexity, size, surroundedness and lower region can be computed without taking image contours as input and their integration for figure–ground estimation can be performed by simple operations. A linear summation combines independent modulations based on different shape properties and a nonlinear diffusion enforces the coherence of figural status estimates among perceptually homogeneous regions. Therefore, instead of computing local configural cues which are inherently ambiguous and using a complex model to integrate them into a unitary percept, our model shows the possibility of computing configural cues globally and then performing integration in a simpler and effective way.

Acknowledgments

The author would like to thank Jean-Michel Morel for exciting and inspiring discussions had at the time of her PhD, Coloma Ballester for valuable comments on an early version of this manuscript and two anonimous reviewers for their insightful suggestions. Most of this work has been carried out during a stay as Visiting Professor in the Image Processing Group (GPI) at University Pompeu Fabra.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2015.03.007.

References

Bahnsen, P. (1928). Eine untersuchung über symmetrie und asymmetrie bei visuellen wahrnehmungen. *Zeitschrift für Psychologie, 108*, 129–154.

Beltagy, M. & Shenawy, S. (2013). On the boundary of closed convex sets in $\mathcal{E}^n$, ArXiv e-prints arXiv:1301.0688.

Boselie, F. (1994). Local and global factors in visual occlusion. *Perception, 23*, 517–527.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of proximity and of other Gestalt factors. *The American Journal of Psychology, 66*(1), 20–32.

Buades, A., Coll, B., & Morel, J.-M. (2006). The staircasing effect in neighborhood filters and its solution. *Image Processing, IEEE Transactions on, 15*(6), 1499–1505.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *The Journal of Neuroscience, 30*(21), 7269–7280.

Calderero, F., & Caselles, V. (2013). Recovering relative depth from low-level features without explicit t-junction detection and interpretation. *International Journal of Computer Vision, 104*(1), 38–68.

Digne, J., Dimiccoli, M., Salembier, P., & Sabater, N. (2011). Neighborhood filters and the recovery of 3d information. In *Handbook of mathematical methods in imaging* (pp. 1203–1229). Springer.

Dimiccoli, M. (2009). *Monocular depth estimation for image segmentation and filtering*, Dept. Signal Theory and Communications, Ph. D. dissertation (2009). Barcelona, Spain: Univ. Politécnica de Catalunya.

Domijan, D., & Šetić, M. (2008). A feedback model of figure–ground assignment. *Journal of Vision, 8*(7), 10.

Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology, 31*(3), 248–306.

Fowlkes, C. C., Martin, D. R., & Malik, J. (2007). Local figure–ground cues are valid for natural images. *Journal of Vision, 7*(8), 2.

Gibson, B. S., & Peterson, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 299–316.

Heitger, F. & von der Heydt, R. (1993). A computational model of neural contour processing: Figure–ground segregation and illusory contours. In *Computer Vision, 1993. Proceedings., Fourth International Conference on, IEEE, 1993* (pp. 32–40).

Hulleman, J., & Humphreys, G. W. (2004). A new cue to figure–ground coding: Top-bottom polarity. *Vision Research, 44*(24), 2779–2791.

Kanizsa, G. (1979). *Organization in vision*. Praeger.

Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure–ground organization. In M. Henle (Ed.), *Vision and artifact*. New York: Springer.

Kastner, S., & McMains, S. A. (2007). Out of the spotlight: Face to face with attention. *Nature neuroscience, 10*(11), 1344–1345.

Kienker, P. K., Sejnowski, T. J., Hinton, G. E., & Schumacher, L. E. (1986). Separating figure from ground with a parallel network. *Perception, 15*(2), 197–216.

Koffka, K. (1935). *Principles of Gestalt psychology*. Routledge & Kegan Paul.

Lab, K. P. (2009). Cues affecting figure-grouping perception. URL: http://faculty.virginia.edu/kubovylab/figureground.php.

Levine, M. W., & McAnany, J. J. (2005). The relative capabilities of the upper and lower visual hemifields. *Vision Research, 45*(21), 2820–2830.

Li, Z. (2003). V1 mechanisms and some figure–ground and border effects. *Journal of Physiology – Paris, 97*(4–6), 503–515.

McAnany, J. J., & Levine, M. W. (2007). Magnocellular and parvocellular visual pathway contributions to visual field anisotropies. *Vision Research, 47*(17), 2327–2336.

Metzger & Wolfgang (1975). *Gesetze des sehens*. Kramer.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1*(1), 29–55.

Pao, H.-K., Geiger, D., & Rubin, N. (1999). Measuring convexity for figure/ground separation. *The proceedings of the seventh IEEE international conference on* (Vol. 2, pp. 948–955). IEEE.

Peterson, M. A., & Gibson, B. S. (1994). Must figure–ground organization precede object recognition? *An Assumption in Peril, Psychological Science, 5*(5), 253–259.

Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure–ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 1075–1089.

Peterson, M. A., & Skow, E. (2008). Inhibitory competition between shape properties in figure–ground perception. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 251–267.

Previc, F. H. (1990). Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. *Behavioral and Brain Sciences, 13*(03), 519–542.

Ren, X., Malik, J., Fowlkes, C. C. (2005). Cue integration for figure/ground labeling. In *Advances in neural information processing systems* (pp. 1121–1128).

Ren, X., Fowlkes, C. C., & Malik, J. (2006). Figure/ground assignment in natural images. In *Computer vision—ECCV 2006* (pp. 614–627). Springer.

Rubin, E. (1921). *Visuell wahrgenommene Figuren: Studien in psychologischer analyse*. Copenhagen: Gyldendalske Boghandel.

Rubin, N. (2001). Figure and ground in the brain. *Nature Neuroscience, 4*, 857–858.

Rubin, N., Nakayama, K., & Shapley, R. (1996). Enhanced perception of illusory contours in the lower versus upper visual hemifields. *Science, 271*(5249), 651–653.

Serra, J. (1983). *Image analysis and mathematical morphology*. Orlando, FL, USA: Academic Press Inc.

Stanley, D. A., & Rubin, N. (2003). fmri activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron, 37*(2), 323–331.

van der Helm, P. A. (2011). Bayesian confusions surrounding simplicity and likelihood in perceptual organization. *Acta Psychologica, 138*(3), 337–346.

Vecera, S. P., & Farah, M. J. (1997). Is visual image segmentation a bottom-up or an interactive process? *Perception & Psychophysics, 59*(8), 1280–1296.

Vecera, S. P., Vogel, E. K., & Woodman, G. F. (2002). Lower region: A new cue for figure–ground assignment. *Journal of Experimental Psychology: General, 131*(2), 194–205.

von Helmholtz (1962). Treatise on physiological optics.

Zhou, H., Friedman, H. S., & Von Der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience, 20*(17), 6594–6611.