



Available online at www.sciencedirect.com



Procedia Computer Science 00 (2014) 1–23

Procedia
Computer
Science

www.elsevier.com/locate/procedia

Analysis of Visual Cues and a Computational Model for Border Ownership

Mehmet Akif Akkuş, Buğra Özkan, Gaye Topuz, and Sinan Kalkan

Abstract

Border Ownership is the problem of identifying which image regions the contours in an image belong to. This information is valuable and essential for several high-level vision problems such as depth perception, object segmentation and recognition, motion perception, etc. Existing computational approaches to Border Ownership estimation either use artificial images or limited type and number of real images. In this study, we propose a new comprehensive database composed of 500 indoor and 500 outdoor images, whose Border Ownership information is labeled by human participants. Moreover, using the database, we investigate the Border Ownership estimation capability of several visual cues such as curvature, T-junction, L-junction, lower region and contrast. We found that, in indoor and outdoor images, the consistency of labeling and the predictive capability of visual cues differ. The database and the analyses of the cues provided in the article can be used for (i) performing further experiments on the Border Ownership problem and other low-level vision problems, (ii) developing better computational models as well as making better comparisons of previously developed Border Ownership models. Moreover, we propose a computational model for estimating Border Ownership by adapting the Tensor Voting framework, which has been previously applied to many perceptual organization problems.

© 2011 Published by Elsevier Ltd.

Keywords: Border Ownership, Visual Cues for Border Ownership, Image Database, Tensor Voting

1. Introduction

Any artificial or biological vision system needs to tackle missing or ambiguous visual information. There is evidence that surfaces with uniform intensities may not activate neurons in the visual cortex and the retina since there is no significant *change* in their receptive fields (see, e.g., [1, 2]). Similarly, lack of visual structure in an image makes estimation of visual motion and disparity information ambiguous in artificial vision systems. Completion of missing information and rectification of ambiguities inside such uniform surfaces, or regions, can be achieved by filling-in the reliable information at the boundaries of regions (see, e.g., [3, 4, 5, 6, 7, 8, 9, 10, 11]).

Such filling-in activities require that the boundaries of regions be known beforehand; *i.e.*, each boundary, or border, in the scene has to be assigned *Border Ownership* (BO) information, linking the regions and their borders so that the relevant borders might provide the required reliable visual information to be filled in, which, in turn, requires that only one region can own a border [12] (though, in certain circumstances, the assignment might not be unique - see Figure 1). Apart from facilitating filling-in activities, this information is crucial for object segmentation, object recognition, occlusion reasoning and hence, depth estimation, figure-ground segregation etc. - see, e.g., [14].

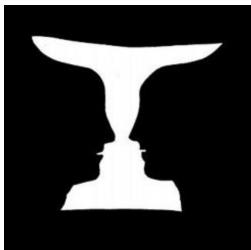


Fig. 1: Rubin's figure-ground vase. The figures in the image (*i.e.*, vase or human silhouette) emerge in the human visual perception as the border between black and white regions is perceived to belong to only one of the two regions. This process occurs as the direction of attention is changed. Vase figure is seen as the border is perceived to belong to the white region whereas human silhouette figure appears when the border belongs to the black region. (Source: [20])

In this article, we contribute to the study of the Border Ownership problem (i) by introducing a comprehensive dataset composed of 500 indoor and 500 outdoor images which are labeled by humans, (ii) by analyzing the Border Ownership estimation capability of several visual cues, namely, curvature, T-junction, L-junction, lower region and contrast in detail, and (iii) by proposing a computational model based on Tensor¹ Voting, which has been previously applied to many perceptual organization problems (see, *e.g.*, [15]).

The present study proposes a new dataset for the following purposes: (i) Although roughly at which stage of the visual processing Border Ownership in human visual system is known [16], any other detail about how and exactly where it is computed is not clear yet. We believe that our comprehensive dataset can be used for analyzing the ‘how’- and ‘where’-parts of the biological processing behind BO using, *e.g.*, eye trackers or neural activity capturing devices such as fMRI. (ii) A subset of the images in the Berkeley Segmentation Database has Border Ownership information; however, the size of this set is small and limited only to outdoor images. (ii) There are many computational models for estimating BO. Most of these models either use artificial images [17, 18] or a limited number of real images [19]. Our dataset can be used for developing better computational models, and it can serve as a benchmark for evaluating such models.

Our computational model for estimating Border Ownership, called Iterative Vector Voting, is a method based on the Tensor Voting Framework [15]. Simply, in this algorithm, BO decision of each border pixel is propagated to each other via voting to obtain the most reliable decisions, with their locations, directions and magnitudes. The algorithm is fast, local, data-driven and uses the visual cues analyzed in the first part of the article.

2. Literature

In this section, we review what is known about the Border Ownership problem and its estimation in Psychology, Neuroscience and Computer Vision.

2.1. Border Ownership in Neuroscience

It is known that there are Border Ownership selective neurons in human vision system though their discovery was only recent: In 2000, Zhou *et al.* [16] found out that 18% of the cells in V1 and more than 50% of the cells in V2 and V4 (along the ventral pathway) respond or code according to the direction of the owner of the boundary. V2 and V4 areas, having significant BO selectivity compared to V1 area (59% of cells in V2 and 53% of cells in V4), are claimed to be involved in determining the direction of the owner [21].

¹A tensor can be defined as the superclass of a vector: A vector represents a magnitude and a direction whereas a tensor can represent more directions.



Fig. 2: BO cell with the same contrast values in the left and right halves of its receptive field indicates the direction of the figure. The V2 cells show larger activation when the light gray patch on a dark ground is attached to the left of the figure (A) compared to the dark gray patch appeared on the right (B).

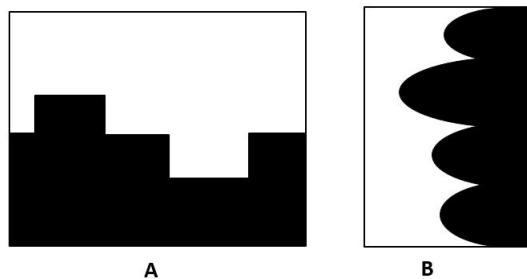


Fig. 3: (A) Lower region and (B) Convexity Gestalt cues for figure-ground assignment used. [Adapted from [25] with permission].

Although underlying mechanisms of how Border Ownership is estimated in the human vision system is largely unknown, it has been shown that visual cues such as contrast [22], depth order [23], curvature [16, 24] play a role in the estimation process. A majority of the neurons (48%) in V1 area of monkey visual cortex are found to show contrast polarity selectivity while a minority of cells (18%) showed border ownership modularity compared to cells in V2 and V4 areas (see Figure 2) [16]. Contrast selective BO cells respond when the right part of its receptive field has higher intensity - Figure 2. The V2 cells show larger activation when the light gray patch on a dark ground is attached to the left of the figure (A) compared to the dark grey patch appeared on the right (B) [22].

Another important cue that seems to be used for Border Ownership estimation is depth order. Qui and Heydt showed that neurons which are selective to the corners of 2D images in V2 regions of macaque monkeys' visual cortex are also selective to the depth order of 3D images and activated more by the 'near' side of the figures [23]. This study points out that human visual system processes BO information in 2D images the same as BO in 3D images as the gestalt cues modulate stereoscopic responses.

Among the visual cues, as also demonstrated in this article, curvature seems to be the most informative one. It is shown that BO selective neurons in V2 and V4 regions in the monkey visual cortex get activated when the contours of figures are convex [16]. In fact, especially V4 neurons whose receptive fields are large, curved and on-surround [24] are found to be sensitive to convex and concave figures. Therefore, V4 might be involved in curvature-based estimation of the Border Ownership information.

The fact that border ownership sensitive neurons differentiate the direction of the owner 10–25 ms after the onset of the response [16] and that border ownership sensitivity emerges as early as V1 (although to a lesser extent) suggests that border ownership can be determined using local cues that can be integrated by lateral long-range interactions along a boundary. This has already inspired computational studies and has been used as a constraint in computational models [26]. However, it is possible that BO computation mechanism is influenced by hierarchical feedback-feedforward dynamic connections [27, 28]. Information coming from the low-level activity of BO sensitive neurons might be sent to the higher level areas such prefrontal cortex. After that, the high level information about object form coming from those areas might be

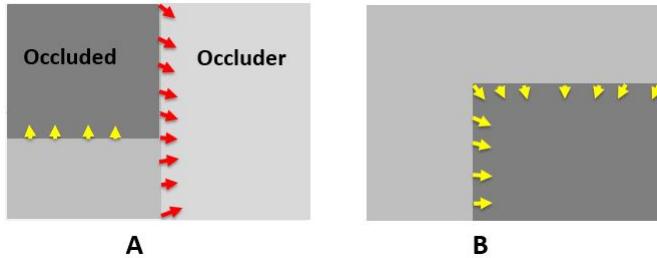


Fig. 4: (A) T-junction is a primary cue for occlusion. There is a discontinuity between occluded and occluding boundaries. Direction of occluding figure is pointed towards the occluding boundary that is the head of the T and the direction of occluded figure is pointed towards occluded boundary that is the tail of the T. (B) L-junction appears at the corners of the figure and the direction of figure is pointed inwards. (Adapted from [22])

involved in BO computation mechanism [29, 30, 31]. For example, high level properties such as familiarity and attention as well as emotion are shown to be involved in figure-ground as well as object recognition as well as BO assignment processes. In fact, as shown recently by Fang *et al.* [32], the process might also be modulated or affected from higher-level cortical areas with attention.

2.2. Border Ownership in Psychology

Psychophysical studies have also investigate the Border Ownership problem and shed light on the underlying mechanisms. For example, it has been shown that human subjects show a strong preference for perceiving convex visual regions as being closer to the viewer (called *figure*) compared to concave visual regions, which are known to be perceived as ground [33] - see Figure 3B.

Other visual cues shown to be related to the BO problem are *lower region*, T-junctions and L-junctions. Vecera *et al.* [25] investigated the relationship between some gestalt figure-ground cues. They showed that human subjects were more likely to perceive lower regions as figure than upper regions. Moreover, figures are perceived as occluders as being closer to viewer than the background (see Figure 3A). Occlusion appears as the three luminance values available in the overlapping figures merge bringing about T-junctions in natural and artificial images (see Figure 4B) whereas L-junctions stem from merging of two luminance values do not constitute the presence of occlusion as seen in Figure 4A [34]. Therefore, T-junctions as occlusion features are important for the Border Ownership problem.

Although utilization of several visual cues for Border Ownership has been identified in Psychology, more controlled experiments [13, 25, 28, 33, 34] are needed for those cues mentioned above as well as further important visual cues for determining border ownership.

2.3. Computational Models of Border Ownership

Although Border Ownership is beneficial for many vision tasks, and it was already identified as an important problem by Koffka [13] in 1935 (although with a different term, “belongingness”), its estimation was largely neglected in computational studies until recently. Early attempts at modeling BO were neural network based [35, 36], who used curvature and L-junctions as visual cues, and modeled propagation and competition mechanisms between several neural networks. Similarly, Nishimura and Sakai [17] developed a physiologically plausible network-model based on surrounding contrast context. This was achieved by an early stage of simple cells which are selective to a certain contrast, which were later combined with inhibitory and excitatory units to arrive at a final decision. In these early attempts, only artificial images were used for testing the models.

For a related problem, Chen *et al.* [37] used semantic, compactness, position (*i.e.*, lower region), junction as well as convexity cues to predict the occlusion relationships between 200 rural, 250 artificial images as well as 645 outdoor images. By using the occlusion information, they also inferred layer sequence of the

image scene. Their layer inference method included the steps of developing a semantic label map from the groundtruth of human semantic labeling, computation of occlusion cues, training an occlusion classifier and lastly obtaining the prediction of occlusion relations by using Adaboost. Their analysis of the occlusion cues and the interaction between those cues showed that semantic cue was the most powerful among those five cues for rural images followed by position, junction and compactness as well as boundary cues whereas position cue was higher for the artificial images followed by semantic, junction, boundary and compactness cues. Moreover, a combination of occlusion cues yielded that semantic-position cues were better than the other cues for the images from both of two categories.

Fowlkes *et al.* [19] tested the cues of region size, lower region and convexity for the prediction of figure-ground in 200 nature images. According to their analysis, size is the most powerful one as it has 68% accuracy only by itself, followed by lower region and convexity cues. They also combined multiple cues fitting a logistic regression to their training data. The combination of lower region and size cues were more powerful compared to the combination of size and convexity cues. These cues are often closely related to each other as locally smaller regions are also locally convex omitting the cues emerge from relations between objects in natural images such as junctions which are important for occlusion and depth order, thus weakening the predictive power.

A second model by Fowlkes *et al.* [38] applies the concept of familiar configuration on the border ownership problem both through local and global aspects in order. A logistic classifier is developed to locally predict FG labels, based on the shapeme representation. After local findings, a global figure-ground model using conditional random field is used to enforce global consistency by learning T-junction frequency and continuity. Inference on this model is handled by loopy belief propagation. They show that their shapeme-based classifier outperform when compared to a baseline model using cues of size & convexity. On a set of 100 labeled Berkeley images, they report 78.3% accuracy [39].

Leichter and Lindenbaum [40] also propose a model for estimating Border Ownership using Conditional Random Fields. However, in their model, they use the ordinal depth of segments as a high-level cue. With such an informative cue about border ownership, they achieve 82% accuracy on a subset of Berkeley images [39].

2.4. Summary and Contributions

From these findings, we deduce the following:

1. Although there are many findings from Psychology and Neuroscience regarding which visual cues are used for the BO problem and where, in the brain, there are BO selective neurons, exact details of the underlying mechanisms of how human vision system estimates BO is largely unknown. Especially, since there are many cues used for the BO problem, more experiments need to be performed on the interaction of the visual cues, especially when they are in conflict. For this reason, the current article is important since (i) the comprehensive database can be used for detailed analysis of visual cues in relation to different labeling data from humans, and (ii) it does analyze the conflicting cases of visual cues for indoor and outdoor data.

Analysis of the cues for the Border Ownership problem has been attempted previously on the Berkely dataset [38]. Although these results have been enlightening for the Border Ownership problem, it was based on only 200 outdoor images. For a more coherent analysis, more images with different characteristics should be analyzed.

2. Although there are datasets with Border Ownership labels, they are not comprehensive in the sense that (i) they do not have outdoor images, or (ii) the images in those datasets are labeled only by one subject. For this reason, the dataset we propose in this article is an important benefit for the BO studies as well as the vision community.
3. There are promising computational studies for estimating BO based on neural networks, conditional random fields, etc. The current article proposes a novel method based on Tensor Voting for the BO problem.

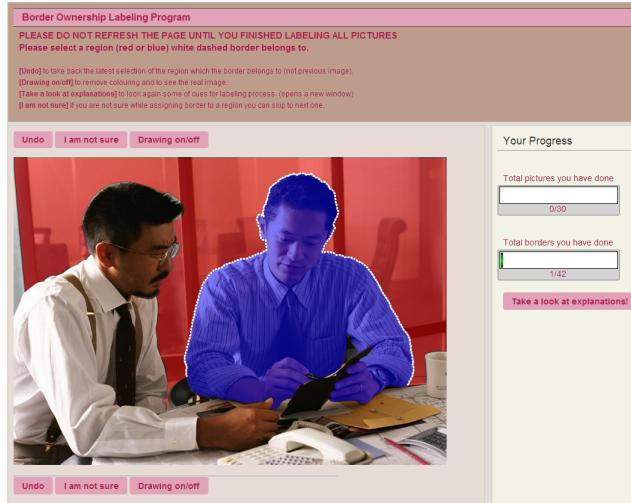


Fig. 5: The web page for collecting BO labels [42]. [Best viewed in color]

3. Data Collection

In this section, we describe the details of how border ownership labels were collected from humans.

3.1. Data to Be Labeled

In our dataset, there are 1000 images, 500 of which are indoor and 500 of which are outdoor images. The 500 outdoor images are taken from the Berkeley Segmentation Dataset [39]; 323 indoor copy-free images are taken from the web arbitrarily and segmented manually by us, and the remaining indoor images are from the LHI segmentation dataset [41]. Each image is segmented into its image regions by humans beforehand. Using these regions, borders and neighboring regions of borders are easily obtained. During this process, regions smaller than 4% of the image width are eliminated since they are hard to visualize and select as the owner of a border.

3.2. Software for Data Collection

For labeling the images, we developed a web page [42] where participants are shown a border and its neighboring regions at a time, and then asked to select the region owning the border. Figure 5 shows a sample labeling page where an original image and a border with white dashes and its neighboring regions with red and blue transparent regions are displayed.

3.3. Participants

BO labeling was realized by 151 participants who were computer engineering students and researchers (115 male, 36 female) whose ages varied from 17 to 34. The participants were given information about the purpose of the experiment and participated a tutorial session on the software (see Section 3.4).

3.4. Labeling Procedure

Before labeling, users were shown an information page explaining what BO is and how to use the labeling tool (Figure 6). The information page includes three types of images with sample borders and explains what is meant by border ownership. After this page, there is a tutorial page including five border ownership questions from an artificial image to make sure the concept of BO is understood well. Figure 6 shows the tutorial page which includes a rectangular, a circle and a pentagon with occlusions between them. If the participant clicks to the wrong side of a border, the participant is informed about the which region owns the border and why. After this stage is completed, the participant starts labeling the real indoor (including



(a) Participants are expected to click on one of the regions for the border in given white
 (b) Participant clicked wrong. Feedback given
 (c) Participant clicked correctly. Feedback given

Fig. 6: Tutorial page on online labeling tool. For a set of simple borders, users are informed whether they labeled right or wrong. [Best viewed in color]

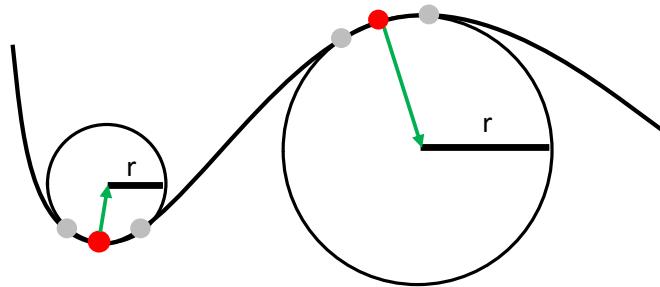


Fig. 7: Calculation of a curvature on a certain point.

office, living room) and the outdoor (including nature, animals) images. During labeling a border, the participant may also choose an “I am not sure” option to pass a border, and we believe that this information (*i.e.*, where participants are more sure and unsure) may be informative for the Border Ownership problem as it is. After labeling, for each image, we have three labels for the borders.

4. Cues for Border Ownership

This section describes the visual cues that we investigate for determining border ownership and use in our computational model. The list of cues we investigate is far from being complete, and includes the widely-used cues for border ownership: **lower region**, **curvature**, **contrast**, **T-junction** and **L-junction**.

4.1. Lower Region

Especially in outdoor images, objects with positions lower in the image tend to be closer to the camera. This observation has been used for scene layering [19, 25, 37] and border ownership [40].

For a border b with two neighboring regions r_1 and r_2 , the lower region cue predicts the owning border $\hat{r}(b)$ of b as follows:

$$\hat{r}_l(b) = \arg \min_{r \in \{r_1, r_2\}} \arg \min_{(x,y) \in r} y. \quad (1)$$

4.2. Curvature

A curved border in an image implies mostly a curved surface in 3D, and the border is likely to be belonging to the surface or region on the convex side of the border [19, 37]. The more a border is curved, the more likely the convex region owns the border - see Figure 7.



Fig. 8: Sample *T-junctions* and ownership directions on indoor and outdoor images (dashed arrows show ownership direction). [Best viewed in color]



Fig. 9: Sample *L-junctions* and ownership directions on indoor and outdoor images (dashed arrows show ownership direction). [Best viewed in color]

The curvature at a point is formally defined as the inverse of the radius of a circle fitted at that point. For fitting a circle o at a point p_i , along a border b composed of P points, we use three consecutive points $(p_{i-12}, p_i, p_{i+12})$. As shown in Figure 7, red denotes p_i and two more are selected nearby point p_i . The direction $s(p_i)$ of the curvature is the direction of the vector connecting the middle point p_i to the center $m(o_{p_i})$ of the circle o_{p_i} :

$$s(p) = m(o_p) - p. \quad (2)$$

Then, the region owning the border is determined using curvature as follows:

$$\hat{r}_c(b) = \arg \max_{r \in \{r_1, r_2\}} \sum_{p \in b} (s(o_p) \rightarrow r), \quad (3)$$

where $s(o_p) \rightarrow r$ is 1 if the direction of curvature is towards region r , and 0 otherwise.

4.3. Contrast

Another widely used cue for border ownership is contrast. The object that is brighter is more likely to be closer to the camera, and hence, more like to own the border. The region \hat{r}_{co} owning a border b is determined as follows:

$$\hat{r}_{co}(b) = \arg \max_{r \in \{r_1, r_2\}} \sum_{p \in r} \frac{1}{N} I(p), \quad (4)$$

where $I(p)$ is the intensity of pixel p .

4.4. T-junction

Junctions are known to be non-accidental features that are highly informative about the 3D structure (see, e.g., [43]). T-junctions for example imply with high confidence that one of the surfaces occludes the others (see Figure 8 for some examples). This has been utilized in several studies on occlusion detection and figure-ground segregation [44, 45].

In our study, we first run a curvature-based corner detector, proposed by He & Yung [46], on the boundaries between regions. With this step, we acquire a list of junctions j_1, \dots, j_J . Then, we extract the line segments l_1, \dots, l_L meeting at a junction j and compute the angles $\theta_1, \dots, \theta_L$ between the neighboring line segments to determine the type of the junction. If a junction j has three line segments with one of the angles $\theta_1, \theta_2, \theta_3$ being close to 180° , then we take j as a T-junction.

The region \hat{r}_T owning the border b at a junction j is determined to be the region on the side of the biggest angle. Figure 8 shows sample T-junctions on real images and the direction of ownership.

4.5. L-junction

L-junction is another cue that is useful for BO. Region lying on the inward part of L-junction is more likely to own the border (see Figure 4).

Unlike T-junctions, a L-junction has two line segments and two angles. The region \hat{r}_L owning a border b at a junction is then the one that the small angle faces.

4.6. Entropy

Image entropy is a concept borrowed from information theory to quantity the information content of an image region. It effectively describes the textured-ness, or energy of an image region. A low entropy image region does not have high amount of texture, or a high number of sudden intensity changes or high energy. There exist various ways of entropy calculation, as energy or texture does not have a unique representation; however, mostly, entropy is calculated as follows:

$$E(r) = - \sum_j P_j \log_2 P_j, \quad (5)$$

where, P_j is the probability that the difference between two adjacent pixels is equal to j . Probability P_j is simply calculated by the histogram counts.

Using the entropy cue, the region \hat{r}_e owning a border b is determined to be the one having more entropy [38, 47]:

$$\hat{r}_e(b) = \arg \max_{r \in \{r_1, r_2\}} E(r). \quad (6)$$

5. A Computational Model for Border Ownership

In this section, we first describe the Tensor Voting framework of Medioni *et al.* (*e.g.*, [15, 48]) and then present *Iterative Vector Voting*, our modification for Tensor Voting applied to the Border Ownership problem. We have *adopted* Tensor Voting because: (i) it is local yet a global result can be computed using local interactions; (ii) it is computationally inexpensive, making it especially suitable on large data or images; (iii) it is data driven, and hence, it does not require training data; (iv) it is based on the lateral interactions in the early stages of perceptual organization. We have *adapted* Tensor Voting because: (a) In Tensor Voting, tensors are bi-directional, whereas for the Border Ownership problem, we need a uni-directional tensor. (b) The strength of a vote decays with distance in Tensor Voting, whereas, in Border Ownership, we want the vote to decay when there are abrupt changes in the curvature or the orientation of a border.

5.1. Tensor Voting (TV)

Medioni *et al.* [15, 48] proposed Tensor Voting for inferring perceptual structures from sparse low-level visual features, which are either noisy, incomplete, oriented or non-oriented. They have successfully applied Tensor Voting to many vision problems related to curves and junctions in 2D as well as their counterparts in 3D. In Tensor Voting, for data representation, tensors are utilized and all the calculations are handled through linear voting [49]. In fact, Tensor Voting is composed of two main parts: data representation and voting.

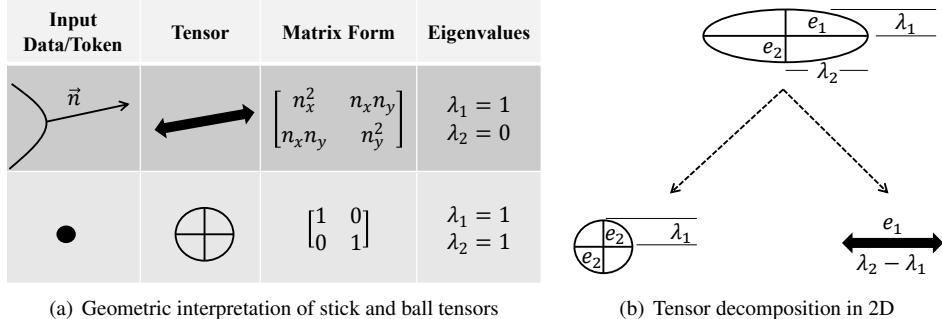


Fig. 10: Stick and ball tensors (a), and an ellipse tensor composed of a ball and a stick tensor (b). [Adapted from [50]]

5.1.1. Data Representation

Data representation in the case of an image² is in the form of a 2nd order, symmetric, non-negative definite tensor, which will be called a *2D tensor* in the article. A tensor can represent both a saliency value and a preferred orientation of perceptual information or structure related to a curve, a junction or a region that the information, or the token, belongs to: For example, a point is represented using a ball tensor whereas a curve is represented using a stick tensor where the direction information can be used for denoting the tangent of the curve or its normal (see Figure 10(a)). A 2D tensor can be considered as a 2x2 matrix, or an ellipse in 2D, and decomposed as:

$$T = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T \quad (7)$$

$$= (\lambda_1 - \lambda_2) e_1 e_1^T + \lambda_2 (e_1 e_1^T + e_2 e_2^T), \quad (8)$$

where λ and e are the eigenvalues and eigenvectors of the 2x2 matrix, respectively. The first part of Equation 8, *i.e.*, $(\lambda_1 - \lambda_2) e_1 e_1^T$, corresponds to the stick component of the tensor. If T equals to this component, then the tensor is called a **stick tensor**. It has a degenerate elongated ellipsoid structure, as seen in Figure 10(a). Presence of a stick tensor can signify, *e.g.*, that a piece of curve exists at that position with e_1 being the curve tangent. Then, the size of the stick component, which is $(\lambda_1 - \lambda_2)$, would indicate the curve saliency. On the other hand, the second part of the sum in Equation 8, *i.e.*, $\lambda_2 (e_1 e_1^T + e_2 e_2^T)$, corresponds to the ball component of the tensor, *i.e.*, the **ball tensor**. It can be used to represent a perceptual structure having no orientation, or multiple orientations which neutralizes each other at this point. The size of the tensor, again, provides the certainty of information at that position. Clearly, geometrically, a ball tensor is in the shape of a circular disk, as shown in Figure 10(a).

Figure 10(b) shows how a generic tensor can be decomposed into stick and ball components. As the tensor has parameters $\lambda_1, \lambda_2, e_1, e_2$ as its eigenvalues and eigenvectors, this simply means that the orientation of its normal is at the direction of \hat{e} , and the saliency of the curve is measured by $\lambda_1 - \lambda_2$.

5.1.2. Voting

After representing data with tensors, a two-phase tensor voting procedure is applied: sparse voting and dense voting. Both stages can be considered as a tensor convolution with separate voting kernels. The main difference between the two is the voting domain.

In sparse voting, tensors vote on other tensors that are located in their neighborhood. A generic tensor is generated at each token location, which equals the *tensor sum* of all votes at that location. Sparse voting deduces the most preferred orientation by refining the initial one for each token. Thus, sparse voting is also called **token refinement**.

²In the rest of the article, we assume that a tensor is used for the purposes of a vision problem, and that it is 2D.

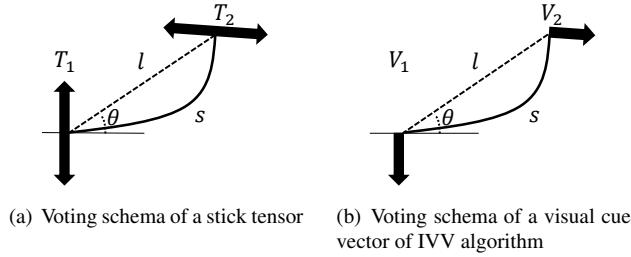


Fig. 11: Voting schema of stick tensor (used in Tensor Voting) and visual cue vector (used in Iterative Vector Voting).

In dense voting, on the other hand, tensors vote on every point in their neighborhood. Initially, each generic tensor is decomposed into its ball and stick components. Ball and stick tensors have distinctive voting fields, and these fields define their neighborhoods. They broadcast their information to all discrete cell locations in these neighborhoods. Simply, dense voting extrapolates the information to the whole domain so as to extract features coherently. For this reason, this step is also called ***dense extrapolation***.

If we look at how two tensors interact in the original work by Mordohai and Medioni [50] for two tensors located on a contour, we see that the weight of a vote is affected by both *good continuation* and the distance (Figure 11(a)), as shown below:

$$DF(s, \kappa, \sigma) = \exp\left(-\frac{s^2 + c\kappa^2}{\sigma^2}\right), \quad (9)$$

where s is the length of the border connecting the tensors; c is a constant controlling the contribution of the curvature; σ is a constant that affects the scale of voting, determining the effective neighborhood size; and finally, κ is the curvature.

Having defined a *decay function* for votes, the votes of 2-D stick and ball votes, of which the parameters are shown in Figure 11(a), can be integrated as follows:

$$Vote_{Stick} = DF(s, \kappa, \sigma) [-\sin(2\theta), \cos(2\theta)]^T [-\sin(2\theta), \cos(2\theta)], \quad (10)$$

$$Vote_{Ball} = \int_0^{2\theta} R_\theta^{-1} Vote_{Stick}(R_\theta P) R_\theta^{-T} d\theta. \quad (11)$$

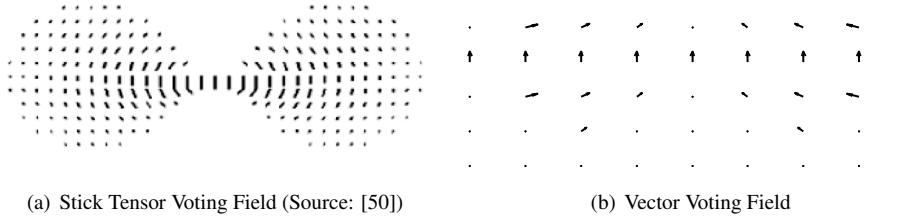
The interaction of tensors can be visualized using a voting field. A voting field can be considered as a map for the decay function affecting which directions (orientations) and magnitudes of all votes cast by the voter tensor. Votes are stored in these pre-computed voting fields, thus the expense of computation decreases during voting.

5.2. Iterative Vector Voting (IVV)

For the BO problem, the following needs to be represented: (i) The direction of the owner region of the boundary, and (ii) The magnitude, *i.e.*, a confidence level of how strongly the region owns the border.

Tensor voting framework needs to be modified since border ownership information does not diminish on smooth borders unless sudden orientation changes occur. Additionally, vectors are sufficient for border ownership problem. In the original algorithm, votes are aggregated by both distance and orientation as formulated in Equation 9. To attenuate votes by orientation and distance together for IVV, the distance parameter is removed, the scale of voting is decreased and voting is iterated. This way, distance does not affect Border Ownership information unless there is an orientation change. Therefore, the decay function of IVV becomes:

$$DF_{IVV}(\kappa, \sigma) = \exp\left(-\frac{c\kappa^2}{\sigma^2}\right). \quad (12)$$



(a) Stick Tensor Voting Field (Source: [50])

(b) Vector Voting Field

Fig. 12: Voting fields of Tensor Voting (a) and Iterative Vector Voting (b).

Voting algorithm of IVV is adapted from stick voting algorithm of TV. The only difference originates from the decay function which is revised for IVV. The reason for such an adaptation is that stick voting keeps the information of curve continuation. It emits the maximum vote to the boundary curve which is predicted to continue from these locations. Thus, our voting function, which depends on both orientation and saliency decay function, is defined as follows:

$$Vote(\theta, \sigma) = DF_{IVV}[-\sin(2\theta), \cos(2\theta)]^T [-\sin(2\theta), \cos(2\theta)]. \quad (13)$$

Another difference to the Tensor Voting framework is that, for the BO problem, unidirectional tensors, *i.e.*, vectors, are used, as shown in Figure 11(b). In this figure, vector V_1 and its vote at V_2 , which is a vector too, are visualized. Here orientations of both V_1 and V_2 show the same region, that represents BO, but their saliences are different due to vote attenuation. Due to the change in voting functions, voting fields of TV and IVV differ, as shown in Figure 12. With this voting field, tokens whose votes in the same direction can increase their votes even if they are far provided that there is no orientation discontinuity between them.

Moreover, for the Border Ownership information, there is no need for dense voting as the votes on the borders are enough to assign the BO information.

5.2.1. The Algorithm for IVV

The IVV algorithm consists of three main phases (Figure 13): (i) Cue extraction, (ii) Curvature extraction, (iii) Voting. The inputs of the algorithm are an image and the corresponding boundary image. Initially, both visual cues and curvature map are extracted from these. Visual cues are represented by vectors: their magnitudes show the saliency of cue, *i.e.*, the magnitude of border ownership information while their directions show the owner region of the border pixel.

In cue extraction, curvature map is used for determining *linels*³ Linels are simply extracted by cutting the border from points having curvature of local maxima. An example image of linels, which are colored differently, is shown in Figure 14.

After visual cue vectors and curvels are extracted, the voting process comes next. Visual cue vectors propagate their border ownership information iteratively to each other, in the limits of pre-defined saliency decay function.

After the voting procedure, all border pixels have more reliable, enriched BO information. However, as in the original TV framework, insignificant votes are to be eliminated. For such purposes, thresholding via local maxima points are used. Thresholding by local maxima means just holding the vectors of local maxima and eliminating all others. Reason for such an elimination is to choose pixels of salient features, to create an human-vision-system (HVS)-like model. It is argued that early stages of HVS initially use salient, attention-taking points instead of examining the whole boundary [51].

³*Linel* is not an existing term in English; it is used in the article by inspiration from the term *curvel*. A *curvel* is defined by Medioni [50] as “perfectly oriented point”. Thus, a curvel can be considered as the smallest, meaningful piece of curve. In the light of this definition, a linel is defined as “largest smooth piece of line”.

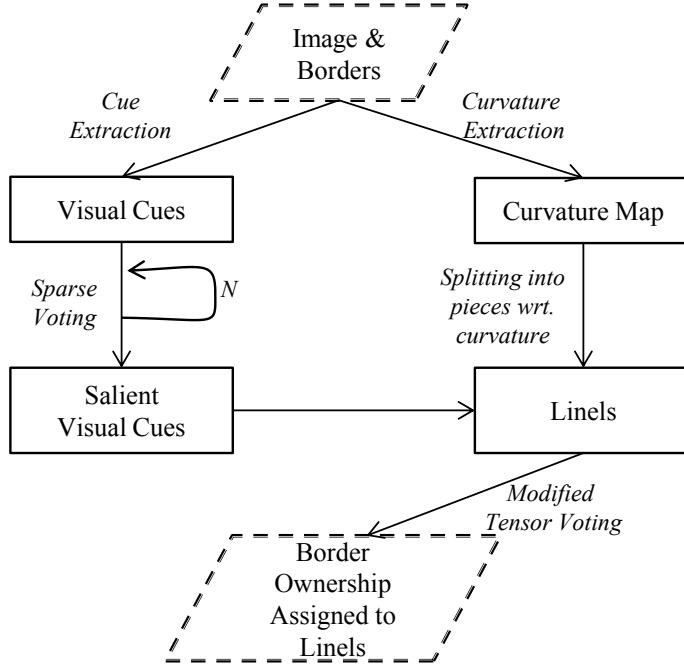


Fig. 13: Iterative Vector Voting algorithm overview.

Finally, after thresholding insignificant votes, the BO information is assigned to each linel simply by addition of all cue vectors belonging to the linel. Resulting vector of addition provides us the joint decision of salient cue vectors about border ownership. Details of each phase of the algorithm are provided below:

• Curvature

The method for calculating curvature is as described in Section 4.2.

The distance of neighborhood, which determines the right and left neighbors of the origin point, is learned through the observations. It is chosen as 12 with respect to these observations, which means that there should be 12 pixels on the border, between the origin and the neighbors on which the circle is placed.

• Visual Cues

The method consists of four pixel-based visual cues: T-junctions, curvature, entropy, contrast. The vote of each cue is represented by a vector for each border pixel. Each visual cue has its own map, where pixels are assessed with their cue values. The ratio of cue values on both sides of the border defines the magnitude of cue vector, as the vector perpendicular to the border yields the direction.

In Figure 15, as an example, how a visual cue is converted to a vector is visualized for the contrast cue. The neighbor pixel with value 40 is the owner of the seed pixel. Thus, the direction of the visual cue vector is perpendicular to the border, towards the region of this neighbor pixel, as shown in Figure 15. The magnitude on the other hand, which represents the information saliency, is measured by the ratio of gray scale values. Greater difference on cue values means higher saliency. The magnitude is scaled down to the range [0-1]. It means that a ratio higher than 10 is considered a vector with magnitude of 1, corresponding to the most reliable information.

Entropy cue follows the same way with the contrast cue. A region with more textured structure is more probable owner of the border, compared to the coarse, plain one [47, 38]. More textured structure means higher energy, *i.e.*, higher entropy, thus entropy of 3x3 neighborhood is calculated for

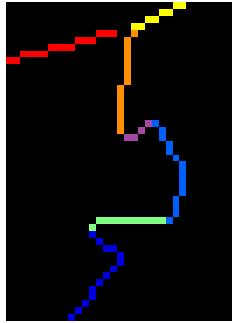


Fig. 14: A set of linels extracted from an image. Each linel is shown with a different color. [Best viewed in color]

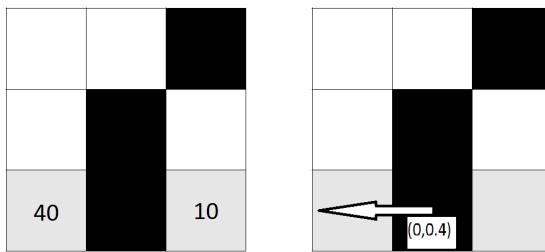


Fig. 15: Pixel-based contrast tensor

each neighbor pixel, as defined in Section 4.1. As T-junctions indicate very salient BO information, they are represented by vectors of magnitude 1, which is the maximum. Thus, they transmit very reliable BO decision to their neighbors.

As the curvature map is calculated earlier in the linel extraction procedure (Figure 13), curvature values and directions are directly used from this map when constructing curvature cue vectors. Lower-region cue can also be considered among successful pixel-based cues. However, as it provides discrete BO decision, lower region cue can not be utilized for voting.

• Voting

Four separate cue vector maps are constructed by assigning four different cue vectors to every border pixels. After these maps are obtained, five step voting algorithm are applied sequentially as follows:

1. Apply iterative voting algorithm for each cue map separately, each with a specific number of iterations, which is a parameter that is investigated in the article.
2. Apply local maxima thresholding to these votes, to acquire the most reliable updated votes.
3. Calculate the vector sum on each linel for each cue.
4. Apply majority voting on each linel to gather an associate decision.
5. Apply second majority voting on BO decisions of linels to assign a BO label to the whole boundary.

All steps of IVV are illustrated on an example in Figure 16. In Figures 16(b) and (c), a map of contrast vectors and the map of voted cue vectors are given respectively. Observe that the number of salient vectors increase around right shoulder of the child, as there exist two T-junctions and many contrast & entropy vectors exist and show the same region as the owner. Thresholded, associate BO vectors are shown in Figure 16(d). In Figure 16(e), correctly assigned borders are shown in green while wrongly labeled ones are shown in red.

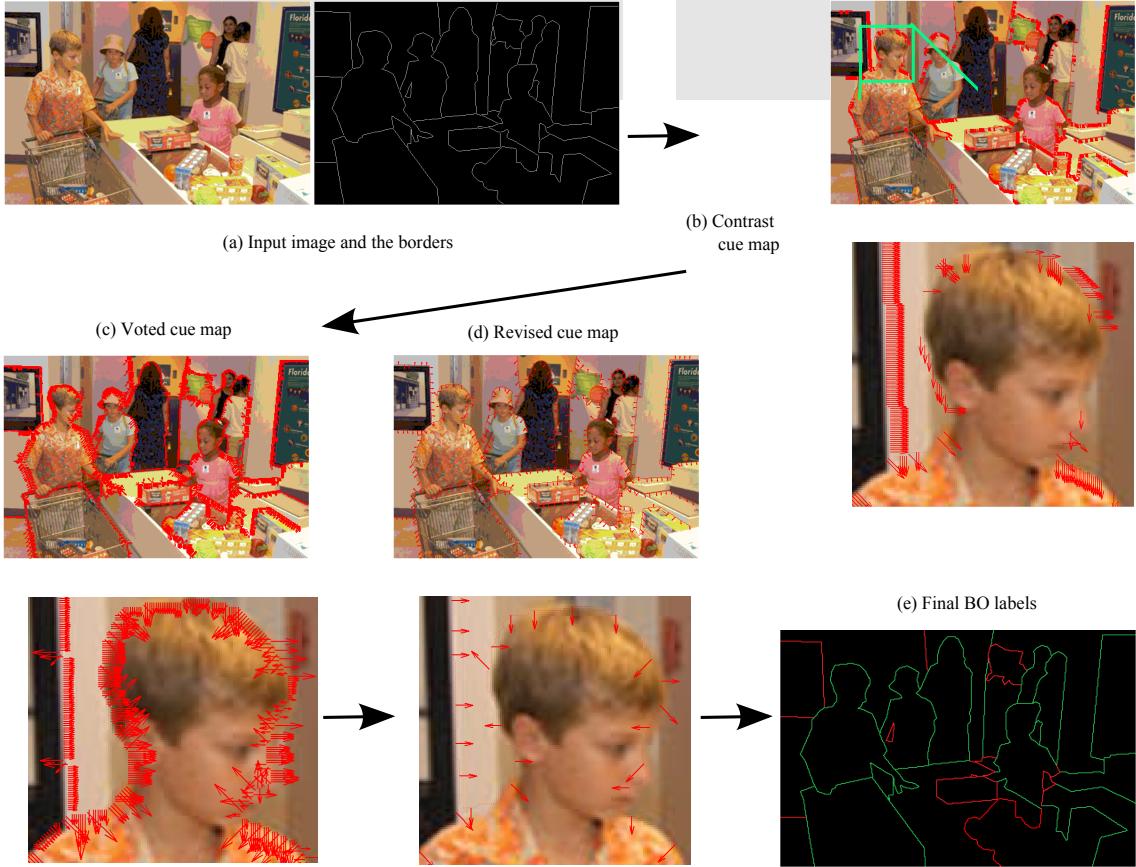


Fig. 16: Snapshots of the intermediate steps from the IVV method on a sample image. [Best viewed in color]

6. Results

In this section, we first analyze the visual cues and then, present the results of our computational model.

6.1. Analysis of the Dataset

The Border Ownership information acquired from hand-labeling by 36 female and 115 male participants where the same border was labeled by three participants had high ratios of consistency (75.2%) for indoor and (62.2%) for outdoor images. It should be noted that, during labeling, the participants had the option to pass a border if they were not sure about the owner of the border. In the analyses, such borders are not taken into consideration for consistency analysis.

Figure 17 shows the consistency ratios of Border Ownership information according to age and gender in percentages. There is a very minor decrease in the consistency ratios as the age increases, and slight differences exist between the consistency ratios of males and females for indoor and outdoor images. However, the age and the gender of the participants are not uniform to make the results in Figure 17 conclusive, and Border Ownership information is unlikely to be correlated with age or gender.

In Figure 18, we see the consistency of the labels for indoor and outdoor images. It can be seen that the consistency of indoor images is significantly higher than the outdoor images ($F = 14.29$, $M = 0.130031$, $P < 0.05$). The reason for this difference might be that indoor images are more structured than outdoor images, and that depth ordering is more clear in general compared to outdoor images since, for large segments in outdoors, depth ordering might be unclear especially when they are far.

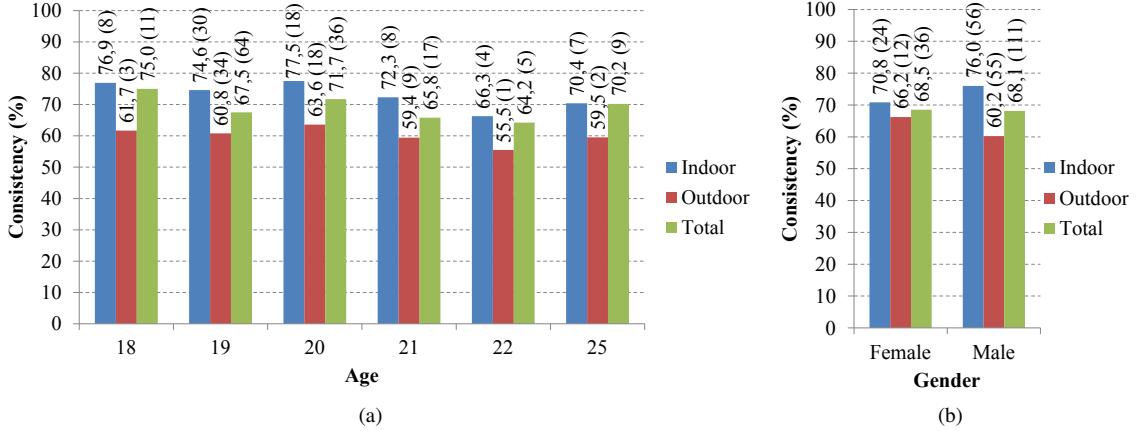


Fig. 17: Border Ownership consistency ratios in percentages according to age (a) and gender (b). Color legend of two types of images (*i.e.*, indoor and outdoor) as well as total images are listed on the right. The number of participants is shown in parenthesis. [Best viewed in color]

Table 1: Visual cues and their predictions compared against different consistencies of labels.

	“single set”		“2/3 set”		“3/3 set”	
	Indoor (%)	Outdoor (%)	Indoor (%)	Outdoor (%)	Indoor (%)	Outdoor (%)
Curvature	65.6	61.0	65.6	59.2	72.0	64.9
T-junction	77.0	66.0	79.0	63.3	82.0	66.9
L-junction	50.0	55.0	58.0	50.6	59.6	54.1
Lower Region	46.0	50.0	52.0	50.9	54.7	52.0
Contrast	44.0	45.8	44.7	47.4	44.1	47.4

6.2. Analysis of Visual Cues

Border Ownership might in certain cases be subjective (*e.g.*, see Figure 20), which might lead to different labels for a border. In our dataset, 65.5% of the borders are given the same label by three participants, 30.7% of the borders are given two different labels, while 3.8% of the borders are marked “I am not sure” - see Table 2.

We analyze the predictive capability of different visual cues for different labeling consistencies: the “1/3 set”, where one label that is different from the other two is considered; the “2/3 set”, where the two same labels that are different from the last one are considered; and, the “3/3 set”, consisting of the borders where all participants agreed upon. Such an analysis allows investigation of where participants disagree more about the Border Ownership and whether there is an underlying reason for that.

Table 1 lists the accuracies of the predictions of the different visual cues for indoor and outdoor images. We see the following from the table: (i) Accuracy varies significantly with respect to the visual cues. This is expected since some of these visual cues are more predictive in certain settings. For example, lower-region

Table 2: The analysis of different labeling with image type.

Consistency Type	Image Type	Indoor	Outdoor	Total
Three participants labeled same (3/3)		69.4%	62.2%	65.4%
Two participants labeled same (2/3)		22.3%	37.7%	30.8%
Participant labeled “I am not sure”		8.3%	0.1%	3.8%

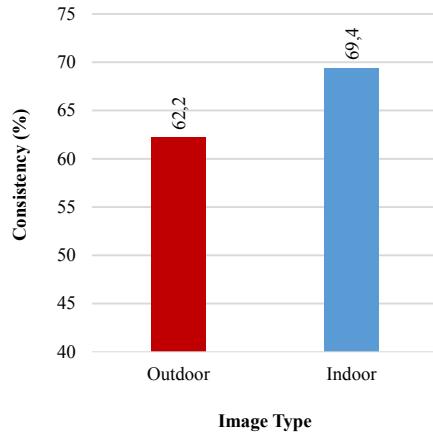


Fig. 18: Consistency ratios of two image types: Indoor and outdoor.

cue is more predictive in an outdoor setting, and T-junction, curvature are better indoors due to the nature of the objects found indoors. (ii) T-junction and curvature are strong predictors of Border Ownership. This is in line with the literature on Border Ownership and figure-ground segregation [16]. (iii) Indoor images generally have higher accuracy than outdoor images. This is probably due to having more regular structures in indoor images. This is also the reason why T-junction and L-junction cues are more accurate in indoor images.

Now, we analyze whether we can predict Border Ownership better by combining the visual cues. Figure 19 shows the analysis of the relative power of Border Ownership cue combinations in indoor (a) and outdoor (b) images. The majority rule was used to combine the multiple cues. In other words, we select the side that more cue has an agreement on that side. If a majority does not occur between the voting of cues, that is to say, equal number of cues have different ownership prediction, we select the side that total accuracies of cues is more. The effect of cue combinations are quantified according to the accuracies in BO prediction of individual cues. It can be seen that the best accuracy percentage among the all 26 combinations of five BO cues belongs to L-junction, T-junction, curvature and lower region cue combination in indoor images, and T-junction and curvature cue combination in outdoor images. T-junction cue was found to be more powerful than other cues in those combinations (see Figures 21(a) and 22(a) for cue accuracies in conflicting cases).

Furthermore, the second most powerful combination consists of L-junction and T-junction cues whose accuracy is the highest among combinations of three cues followed by L-junction, T-junction, curvature cue combination which performs the best among combinations of four cues in indoor images. Regarding the outdoor images, T-junction, curvature and lower region cue combination is the second most powerful of all combinations as well as of combinations of three cues. L-junction, T-junction, curvature, contrast cue combination has the highest accuracy among all combinations of four cues. Among a total of 26 combinations from our five BO cues, the combinations of four cues turned out to be more powerful than combinations of five, three and two cues respectively in both types of images.

Figures 21(a) and 22(a) show the matrix of conflicting Border Ownership cues in indoor and outdoor images of Border Ownership prediction with different combination of cues. The numbers in the middle of each cell shows the number of conflicting occurrences. The number of as well as the percentages of occurrences, where accuracy of each of the conflicting cues is accurate in BO prediction, are demonstrated on the upper right and down left corners of the each cell. The components of the cue combinations are demonstrated on the upper horizontal axis and left vertical axis. These conflict matrices are very valuable since they depict which cues provide conflicting BO information with each other.

Figures 21(a) shows that, in indoor images, T-junction, L-junction and curvature are quite in accordance with each other for predicting Border Ownership whereas lower-region and contrast cues conflict frequently.

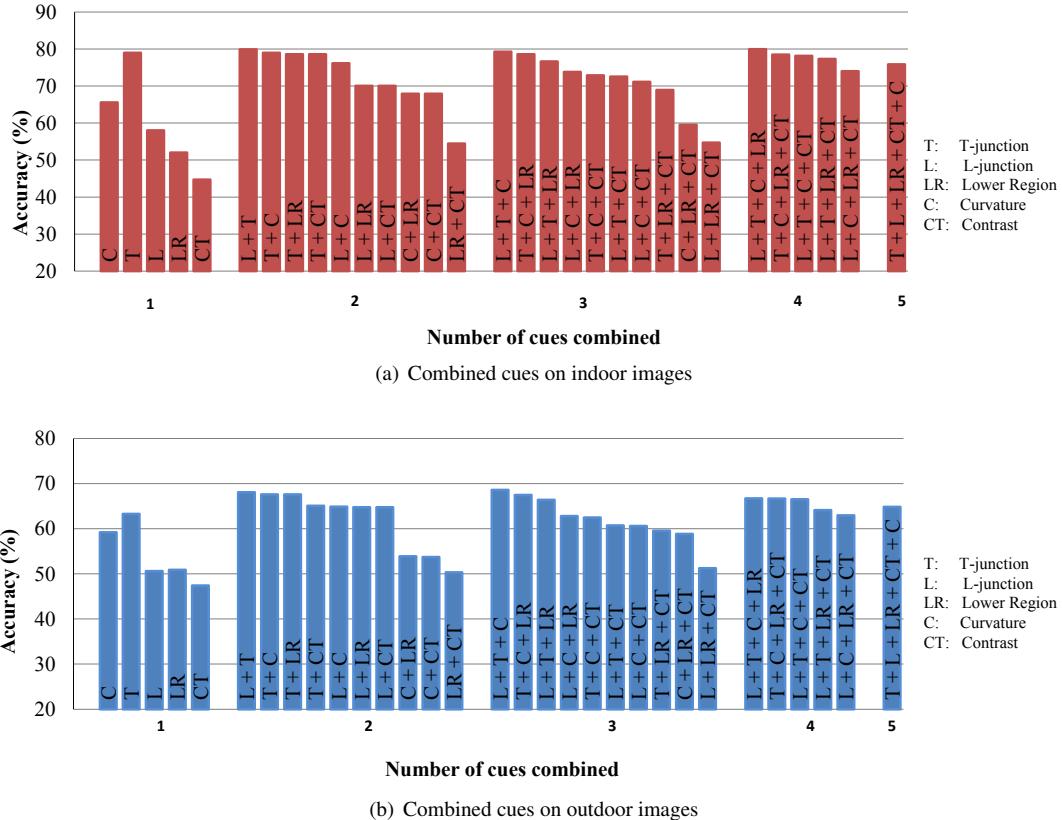


Fig. 19: The accuracy of Border Ownership prediction with different number of combined cues. The x-axis shows the number of combined cues and y-axis shows the accuracies in percentages. The components of the cue combinations are demonstrated inside the bars.

However, in outdoor images (Figure 22(a)), we see that the predictions of T-junction, L-junction and curvature disagree more compared to indoor images. This is in line with the observation that, in indoor images, there are more objects with corners, where occluding boundaries yield more T and L junctions.

6.3. Computational Results

In the IVV model, initially each cue gives a BO decision itself, then the decisions of all cues are combined to reach a consensus for the border via weighted majority voting. Thus it is easier to get separate accuracies for each cue of IVV, as listed in Table 3, where we also analyze the effect of the number of iterations on the performance. We see that T-junction, curvature, entropy and contrast provide better-than-chance (50%) accuracies, and that iterating over the votes improves the results. We conclude that, at three iterations, best performances are obtained for each cue except for curvature. Therefore, in the rest of the article, we will use three iterations for T-junction, entropy and contrast and five iterations for curvature.

In Table 4, overall performances of IVV algorithm for indoor, outdoor and all images are provided separately. We see that combining the cues improves the performance and overall, we can get 77% accuracy. As shown in the analysis of visual cues, the model also makes more mistakes in outdoor images. See also Figure 23 for some example detection results.

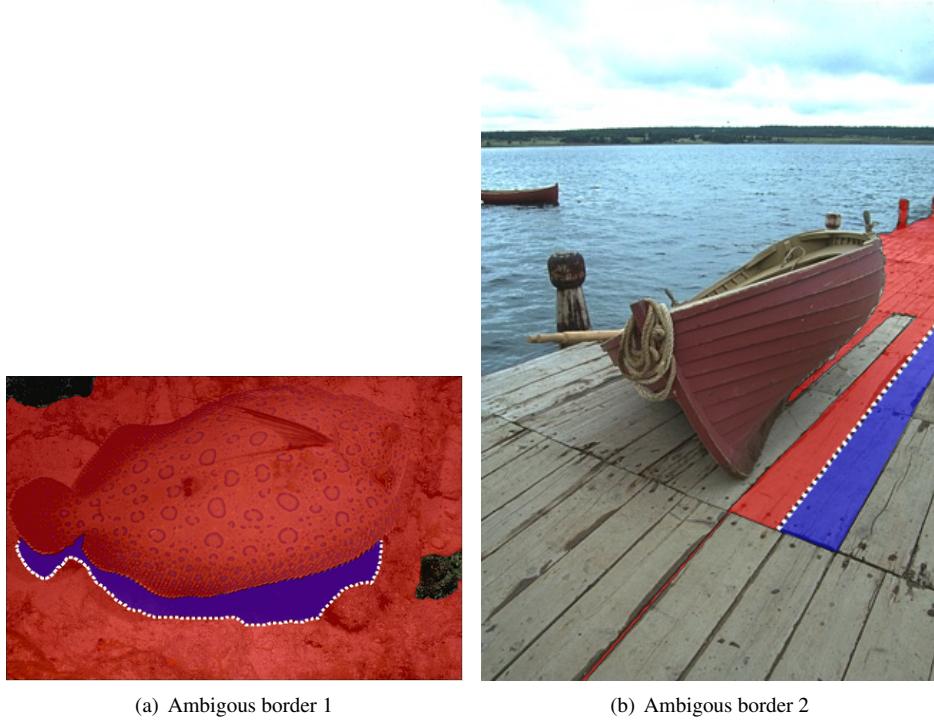


Fig. 20: For some borders, it may not be so obvious to determine ownership of the border. This might be due to wrong segmentation by humans (a) or that the borders might seem shared by two objects (b). [Best viewed in color]

Table 3: Cue Performances of the Iterative Vector Voting method.

Visual Cue / No of Iter.	2	3	4	5
T-junction	72,2%	73,1%	73%	71,8%
Curvature	57,4%	58%	56,7%	59,4%
Entropy	66,7%	71%	65,4%	58,3%
Contrast	64%	68%	64,2%	60,7%

7. Conclusion

In this study, we contribute to the study of the Border Ownership problem and related vision problems by proposing a publicly available dataset [42]. The dataset is the most comprehensive one so far, including labeled 500 indoor and 500 outdoor images where each border is labeled by three subjects. We believe that this dataset can be used as a benchmark for computational models as well as for analyzing the underlying mechanisms of different vision problems, especially that of Border Ownership.

Moreover, we analyzed several widely-used visual cues that are shown to be related to Border Ownership or used by computational models for Border Ownership. We showed that the predictions of visual cues might be different for indoors and outdoors. In fact, by combining the visual cues and analyzing the cases where the predictions of visual cues conflict, we suggest that, for indoors and outdoors, different combinations of visual cues should be used for predicting Border Ownership.

Another important contribution of our study is a computational model that predicts Border Ownership. Our model is based on Tensor Voting, which has been previously successfully applied to many perceptual organization problems [15, 52, 53]. We show that our modified Tensor Voting based model can predict Border Ownership with high accuracy.

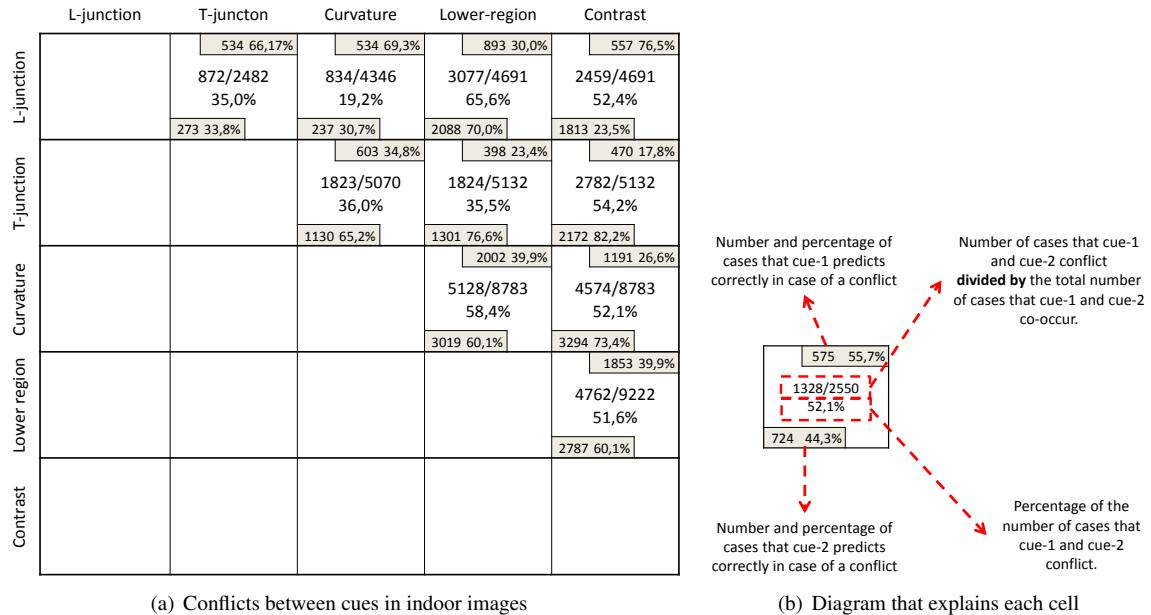


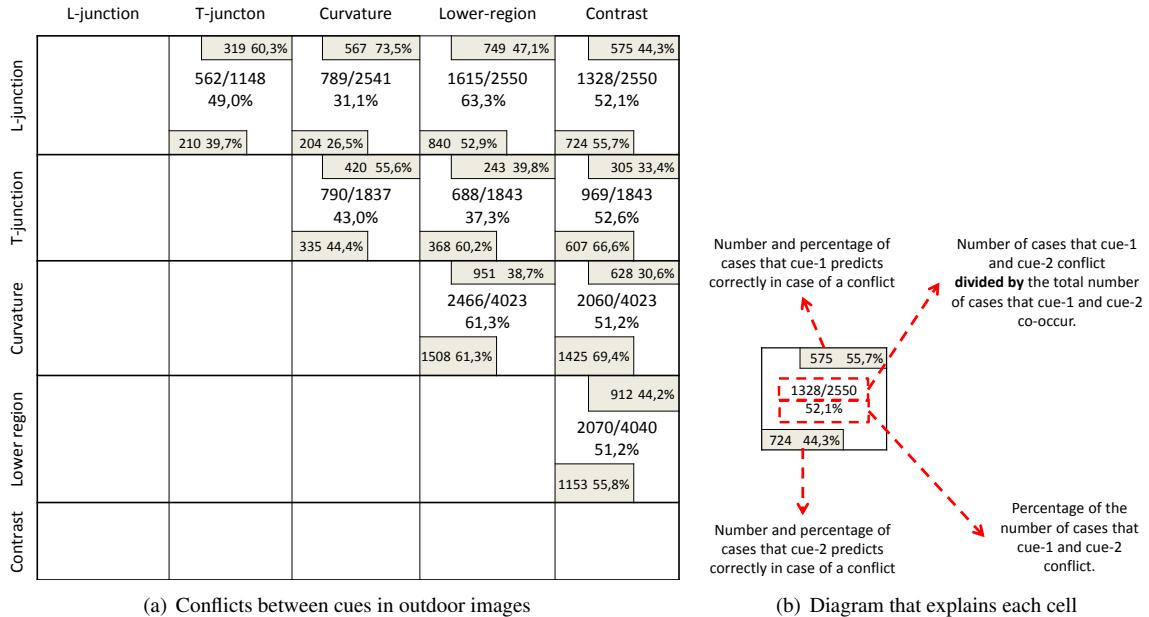
Fig. 21: Conflicting cues in indoors. (a) A matrix depicting which cues conflict, and the *winning cue* in case of a conflict. (b) Explanation of each cell in the matrix.

Table 4: Outdoor and Indoor Performances of the Iterative Vector Voting method.

Visual Cue / Performance Type	Indoor	Outdoor	All
T-junction	77,4%	70,2%	73,1%
Curvature	62%	57,1%	59,4%
Entropy	72,2%	69,8%	71%
Contrast	68,4%	67,6%	68%
All cues combined	78,3%	75,9%	77%

References

- [1] H. Neumann, A. Yazdanbakhsh, E. Mingolla, Seeing surfaces: The brain's vision of the world, *Physics of Life Reviews* 4 (3) (2007) 189–222.
- [2] D. H. Hubel, T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *The Journal of physiology* 195 (1) (1968) 215–243.
- [3] B. Anderson, M. Singh, R. Fleming, The interpolation of object and surface structure, *Cognitive Psychology* 44 (2) (2002) 148–190.
- [4] H. G. Barrow, J. M. Tenenbaum, Interpreting line drawings as three-dimensional surfaces, *Artificial Intelligence* 17 (1) (1981) 75–116.
- [5] T. S. Collett, Extrapolating and Interpolating Surfaces in Depth, *Royal Society of London Proceedings Series B* 224 (1985) 43–56.
- [6] W. E. L. Grimson, A Computational Theory of Visual Surface Interpolation, *Royal Society of London Philosophical Transactions Series B* 298 (1982) 395–427.
- [7] B. Julesz, *Foundations of Cyclopean Perception*, Univ. of Chicago Press, Chicago, IL, 1971.
- [8] H. Komatsu, The neural mechanisms of perceptual filling-in, *Nature Reviews Neuroscience* 7 (3) (2006) 220–231.
- [9] L. Pessoa, E. Thompson, A. Noe, Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception, *Behavioral and Brain Sciences* 21 (06) (1998) 723–748.
- [10] D. Terzopoulos, The computation of visible-surface representations, *IEEE Trans. Pattern Anal. Mach. Intell.* 10 (4) (1988) 417–438.
- [11] S. Treue, R. A. Andersen, H. Ando, E. C. Hildreth, Structure-from-motion: perceptual evidence for surface interpolation., *Vision Research* 35 (1) (1995) 139–48.



(a) Conflicts between cues in outdoor images

(b) Diagram that explains each cell

Fig. 22: Conflicting cues in outdoors. (a) A matrix depicting which cues conflict, and the *winning cue* in case of a conflict. (b) Explanation of each cell in the matrix.

- [12] K. Nakayama, Z. J. He, S. Shimojyo, Visual surface representation: a critical link between low-level and higher-level vision, *An Invitation to Cognitive Science: Visual cognition* (1995) 1–70.
- [13] K. Koffka, *Principles of Gestalt Psychology*, Harcourt, Brace, 1935.
- [14] M. K. Albert, Cue interactions, border ownership and illusory contours, *Vision Research* 41 (22) (2001) 2827–2834.
- [15] G. Medioni, C. K. Tang, M. S. Lee, Tensor voting: Theory and applications, *Proceedings of RFIA*, Paris, France.
- [16] H. Zhou, H. Friedman, R. Von Der Heydt, Coding of border ownership in monkey visual cortex, *The Journal of Neuroscience* 20 (17) (2000) 6594–6611.
- [17] K. Sakai, H. Nishimura, Surrounding suppression and facilitation in the determination of border ownership, *Journal of Cognitive Neuroscience* 18 (4) (2006) 562–579.
- [18] L. Zhaoping, Border ownership from intracortical interactions in visual area v2, *Neuron* 47 (1) (2005) 143–153.
- [19] C. C. Fowlkes, D. R. Martin, J. Malik, Local figure-ground cues are valid for natural images, *Journal of Vision* 7 (8) (2007) 1–9.
- [20] E. Rubin, *Synsoplevede figurer: studier i psykologisk analyse*. 1. del, Gyldendalske Boghandel, Nordisk Forlag, 1915.
- [21] F. T. Qiu, T. Sugihara, R. von der Heydt, Figure-ground mechanisms provide structure for selective attention, *Nature Neuroscience* 10 (11) (2007) 1492–1499.
- [22] O. W. Layton, E. Mingolla, A. Yazdanbakhsh, Dynamic coding of border-ownership in visual cortex, *Journal of Vision* 12 (13) (2012) 1–21.
- [23] F. T. Qiu, R. Von Der Heydt, Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules, *Neuron* 47 (1) (2005) 155–166.
- [24] J. Hegdé, D. C. Van Essen, A comparative study of shape representation in macaque visual areas v2 and v4, *Cerebral Cortex* 17 (5) (2007) 1100–1116.
- [25] S. P. Vecera, E. K. Vogel, G. F. Woodman, Lower region: A new cue for figure-ground assignment, *Journal of Experimental Psychology-General* 131 (2) (2002) 194–205.
- [26] H. Nishimura, K. Sakai, Determination of border ownership based on the surround context of contrast, *Neurocomputing* 58–60 (2004) 843 – 848.
- [27] S. P. Vecera, R. C. O’reilly, Figure-ground organization and object recognition processes: an interactive account, *Journal of Experimental Psychology: Human Perception and Performance* 24 (2) (1998) 441–462.
- [28] M. A. Peterson, E. M. Harvey, H. J. Weidenbacher, Shape recognition contributions to figure-ground reversal: which route counts?, *Journal of Experimental Psychology: Human Perception and Performance* 17 (4) (1991) 1075–1089.
- [29] S. P. Vecera, R. C. O'Reilly, Graded effects in hierarchical figure-ground organization: reply to peterson (1999), *Journal of Experimental Psychology: Human Perception and Performance* 26 (3) (2000) 1221–1231.
- [30] M. A. Peterson, What's in a stage name? comment on vecera and o'reilly (1998), *Journal of Experimental Psychology: Human Perception and Performance* 25 (1) (1999) 276–286.
- [31] S. W. Zucker, Local field potentials and border ownership: a conjecture about computation in visual cortex, *Journal of Physiology-Paris* 106 (5) (2012) 297–315.

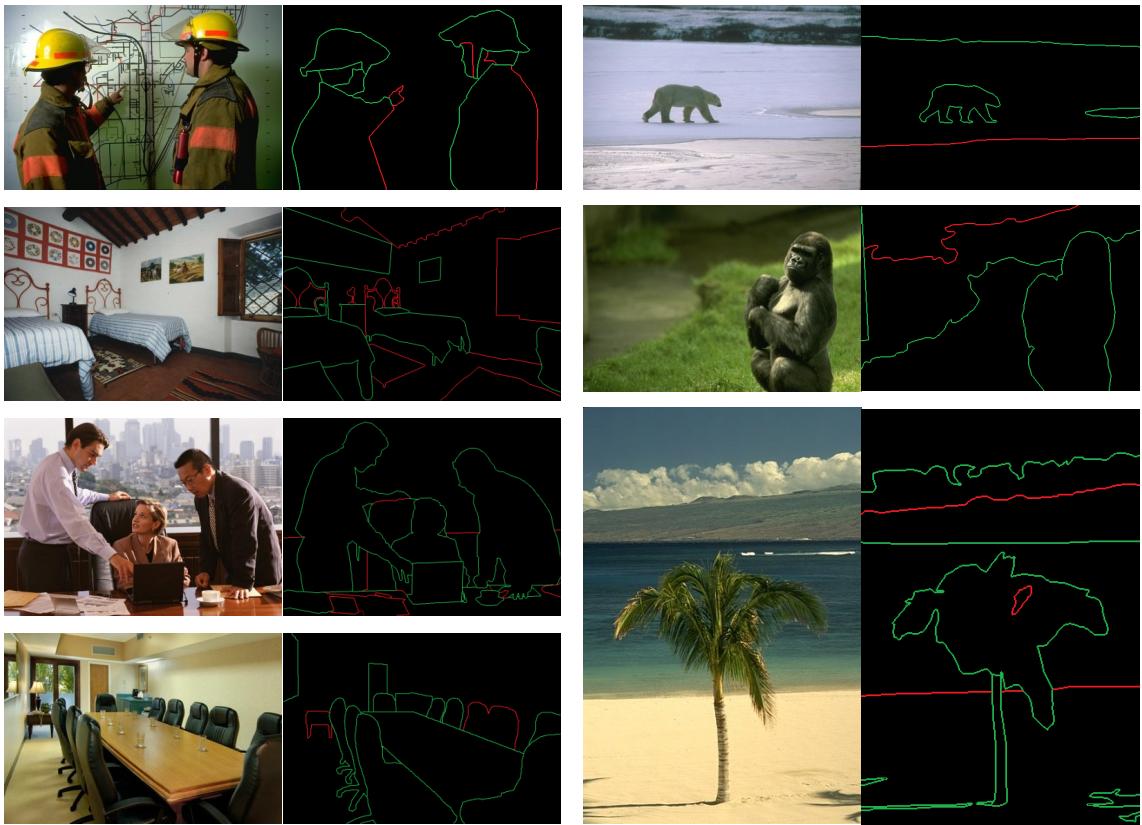


Fig. 23: Snapshots of the predicted Border Ownership information using Iterative Vector Voting. Left: indoor. Right: outdoor. Green shows the correct prediction whereas red shows the wrong ones. [Best viewed in color]

- [32] F. Fang, H. Boyaci, D. Kersten, Border ownership selectivity in human early visual cortex and its modulation by attention, *The Journal of Neuroscience* 29 (2) (2009) 460–465.
- [33] M. A. Peterson, E. Salvagio, Inhibitory competition in figure-ground perception: Context and convexity, *Journal of Vision* 8 (16) (2008) 1–13.
- [34] J. McDermott, Psychophysics with junctions in real images, *Perception* 33 (9) (2004) 1101–1128.
- [35] M. Kikuchi, Y. Akashi, A model of border-ownership coding in early vision, in: International Conference on Artificial Neural Networks, 2001, pp. 1069–1074.
- [36] M. Kikuchi, K. Fukushima, Assignment of figural side to contours based on symmetry, parallelism, and convexity, in: Knowledge-Based Intelligent Information and Engineering Systems, 2003, pp. 123–130.
- [37] X. Chen, Q. Li, D. Zhao, Q. Zhao, Occlusion cues for image scene layering, *Computer Vision and Image Understanding* 117 (1) (2013) 42 – 55.
- [38] X. Ren, C. C. Fowlkes, J. Malik, Figure/ground assignment in natural images, in: European Conference on Computer Vision–ECCV, Springer, 2006, pp. 614–627.
- [39] Berkeley segmentation dataset.
URL <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>
- [40] I. Leichter, M. Lindenbaum, Boundary ownership by lifting to 2.1d, in: IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 9–16.
- [41] Lhi dataset.
URL <http://www.imageparsing.com/>
- [42] Border ownership labelling programme.
URL <http://www.kovan.ceng.metu.edu.tr/bo/>
- [43] A. Jepson, W. Richards, What makes a good feature, *Spatial vision in humans and robots* (1993) 89–126.
- [44] D. Hoiem, A. N. Stein, A. A. Efros, M. Hebert, Recovering occlusion boundaries from a single image, in: IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8.
- [45] X. Ren, C. C. Fowlkes, J. Malik, Figure/Ground assignment in natural images, in: European Conference on Computer Vision, 2006, pp. 614–627.

- [46] X. C. He, N. H. Yung, Corner detector based on global and local curvature properties, Optical Engineering 47 (5) (2008) 057008–057008.
- [47] D. Hoiem, A. Efros, M. Hebert, Recovering occlusion boundaries from an image, International Journal of Computer Vision 91 (3) (2011) 328–346.
- [48] M. Lee, G. Medioni, Inferred descriptions in terms of curves, regions and junctions from sparse, noisy binary data, in: Proc. IEEE Int. Symp. Computer Vision, 1995, pp. 73–78.
- [49] G. Medioni, C.-K. Tang, M.-S. Lee, Tensor Voting: Theory and Applications, in: Proceedings of RFIA, 2000.
- [50] P. Mordohai, G. Medioni, Tensor voting: a perceptual organization approach to computer vision and machine learning, Synthesis Lectures on Image, Video, and Multimedia Processing 2 (1) (2006) 1–136.
- [51] J. B. Subirana-Vilanova, W. Richards, Perceptual organization, figure-ground, attention and saliency, A.I. Memo No 1218 (1991).
- [52] M. Niculescu, G. Medioni, Perceptual grouping from motion cues using tensor voting in 4-d, European Conference on Computer Vision - ECCV 2002.
- [53] E. Y. E. Kang, G. Medioni, Color image segmentation based on tensor voting, in: USC Computer Vision, 2001.