

A Computational Model for Amodal Completion

Maria Oliver¹, Gloria Haro¹, Mariella Dimiccoli², Baptiste Mazin³, and Coloma Ballester¹

¹Universitat Pompeu Fabra (UPF)
²Centre de Visió per Computador (CVC)
³DxO Labs

Abstract

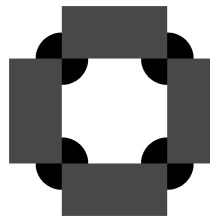
This paper presents a computational model to recover the most likely interpretation of the 3D scene structure from a planar image, where some objects may occlude others. The estimated scene interpretation is obtained by integrating some global and local cues and provides both the complete disoccluded objects that form the scene and their ordering according to depth. Our method first computes several distal scenes which are compatible with the proximal planar image. To compute these different hypothesized scenes, we propose a perceptually inspired object disocclusion method, which works by minimizing the Euler's elastica as well as by incorporating the relatability of partially occluded contours and the convexity of the disoccluded objects. Then, to estimate the preferred scene we rely on a Bayesian model and define probabilities taking into account the global complexity of the objects in the hypothesized scenes as well as the effort of bringing these objects in their relative position in the planar image, which is also measured by an Euler's elastica-based quantity. The model is illustrated with numerical experiments on, both, synthetic and real images showing the ability of our model to reconstruct the occluded objects and the preferred perceptual order among them. We also present results on images of the Berkeley dataset with provided figure-ground ground-truth labeling.

1 Introduction

Visual completion is a pervasive process in our daily life that works by hallucinating contours and surfaces in the scene when there is not a physical magnitude for them. Whenever we look at an image, our brain unconsciously reconstructs the 3D scene by completing partially occluded objects while inferring their relative depth order into the scene (see Figure 1). In Figure 1(b), for instance, our brain prefers to interpret the scene as four disks partially occluded by four rectangles instead of, e.g., the more straightforward description of eight quarters of a disk and four rectangles fitting together.



(a) Image source [34]



(b) Image source [26]

Figure 1: Some examples where our brain experiences visual completion.

In this paper, we are interested in computationally modeling this perceptual phenomenon, recovering what the brain infers about the structure and the relative depth of the objects composing the scene from a planar image. To simplify the analysis of our approach, we focus on scenes where objects appear at two different depths, ones occluding the others. The current approach can handle scenes with both partially occluded and fully visible objects. Our contribution is twofold: firstly, we propose a computational method relying on perceptual findings related to amodal visual completion to compute the disoccluded objects that form the possible 3D interpretations or configurations that arise from a planar image; and secondly, we propose a Bayesian probabilistic model which chooses between these possible interpretations of a planar image the most plausible one, justifying the visual completion human experience. The disocclusion method works by minimizing the Euler’s elastica and incorporates the concepts of relatibility of the occluded contours and convexity of the disoccluded objects. Roughly speaking, two contours are relatible if they can be connected with a smooth contour without inflection points [28] (see Figure 2(a)–(b)) . An equivalent and more precise definition is given in Section 3 (see Def. 1) but let us now notice that the relatibility property implies that two contours can be relatible no matter how far away their corresponding ending points are. Once the objects conforming the scene are disoccluded, we follow a Bayesian approach and give definitions for the prior probability and the likelihood, measured, respectively, by the object complexities and an elastica-based quantity. As a consequence, our probability model takes into account the shape of the objects in the hypothesized scenes as well as the effort of bringing these objects in their relative positions in the visual image.

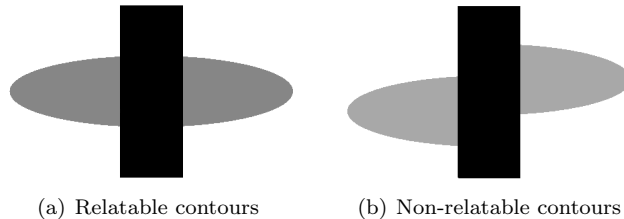


Figure 2: How many perceptual objects? (a) Because of the presence of two relatible contours, which are made of two pairs of relatible end-points (in the upper and lower part of the gray shape, respectively), we perceive an ellipse occluded by a rectangle. (b) There is no pair of relatible contours so we perceive three different shapes.

The structure of the paper is as follows. In Section 2 we review the related work and the fundamentals of visual completion. Section 3 is devoted to present the proposed approach: in particular, Subsection 3.1 presents the object disocclusion method, while Subsection 3.2 details the probabilistic model. Section 4 explains the numerical algorithm while Section 5 provides experimental results. Finally, we present our conclusions in Section 6.

2 Related research

The visual completion phenomenon has been intensively investigated during the past fifty years and it is still an active area of research [23, 25, 39, 17, 45, 42, 16, 7]. Nowadays, it is well acknowledged that occlusion patterns evoke both local and global completion processes, but how the final perception outcome is conveyed is still not well understood. Local completion has been related to the good continuation of visible contours [57], while global completion is driven by the simplicity principle [30, 24]: The assumption that the visual system favors interpretations characterized by phenomenal simplicity, such as symmetry, repetition, familiarity or context properties [48], and regularity, and typically leads toward the simplest completed shape, even though the good continuation principle may be violated. Figure 3(b) shows an example where two different amodal completions occur depending on whether a global cue as symmetry is incorporated or only more local cues, while in Figure 3(a), both interpretations coincide. Some authors (e.g., [40] and references therein) have noticed that features favoring completion through good continuation are read

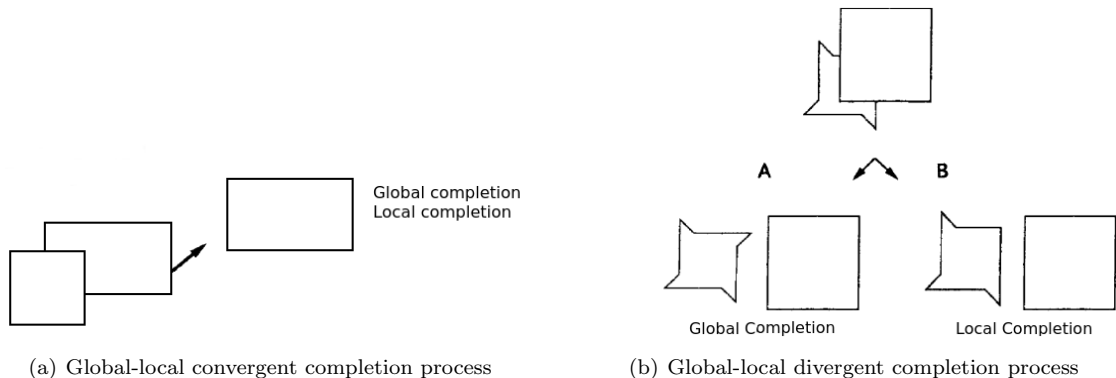


Figure 3: Examples of global-local completion processes. Images adapted from [55].

out more quickly (in the very first second) than are features favoring completion through symmetry (which are incorporated in the following 9 seconds). The incorporation of different cues was also studied by Rubin [46] who experimentally proved that local and global occlusion cues affect the perception of amodal completion at different stages of visual processing. As for global cues, the author focused in relatability and surface similarity, being cues that seem to be instantaneously used at first stages of occlusion perception. Relatability (see Figure 2) was first introduced by Kellman and Shipley [28], who noticed that it is a necessary global condition for completion to occur. Then, for the perception of amodal completion, Rubin proposed that the detection of local cues such as T-junctions (see Figure 4) generates a local pattern of activation which launches a process of propagation of the contour which is either enhanced or stopped depending on whether or not other global cues such as relatability or surface similarity hold.

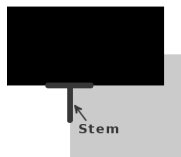


Figure 4: Example of a T-junction arising in presence of partial occlusion.

The evidence that the visual system generates multiple interpretations and that visual completion is the result of a competition between them, was discussed by van Lier et al. [55, 53]. In Figure3(b) is shown that sometimes global and local processes may diverge, and since it is not always the same process which is prevalent, a theory based on either local or global principles alone cannot hold. Later, van Lier et al. [54] proposed an integrative model of global and local aspects of occlusion. In the present work, the global cues that we use are relatability and convexity (which are instantaneously considered at first stages of occlusion perception). As discussed in the following Section, we leave the incorporation of other global properties such as global self-similarities for a future work.

In computer vision, the computational translation of the visual completion phenomenon is commonly referred to as disocclusion or inpainting. A pioneering contribution to the recovery of image plane geometry was given by Nitzberg, Mumford and Shiota [43]. The authors proposed a variational model for segmenting the image into objects which should be ordered according to their depth in the scene, providing the so-called image 2.1 sketch. The minimization of the functional should be able to find the occluding and the occluded objects, while finding the occluded boundaries. The energy functional is defined in terms of region simplification and completion of occluded contours. Contours completion is achieved by linking signatures of occlusion, the T-junctions (see Figure 4), with the Euler’s elastica, so that the completion tends to respect the principle of good continuation [24, 26]. Despite its theoretical importance, the complexity of

minimizing this energy makes the approach far from practical applications. On the other hand, it was one of the main sources of inspiration for the first inpainting algorithms. A first formulation was proposed by Masnou and Morel [36], who tried to interpolate the data into the uncomplete region by minimizing an energy functional based on the Elastica. A similar approach was followed in [3] and in the work of Chan and Shen [9]. Those methods belong to the so-called geometry-oriented methods where the images are modeled as functions with some degree of smoothness, expressed, for instance, in terms of the curvature of the level lines [36, 3, 9, 35, 11, 6] or the total variation of the image [8]. Binary inpainting tools for images are also used to disocclude shapes and thus can be considered as geometry-oriented methods. Binary inpainting can be based on diffusion processes followed by thresholding, named as thresholding dynamics (e.g. [38]). Thresholding dynamic interpolations also usually minimize a geometric functional, based either on the length, area, or curvature of the shape contours [38, 47, 19].

This work is focused on a computational model that, given an image of a 3D scene, it automatically outputs the preferred – according to human perception – interpretation of the scene in terms of depth configuration of the scene objects together with their completion in case of occlusions. A related and inspiring work in the literature is the proposal of van Lier et al. [54]. They proposed to choose the preferred scene interpretation based on the minimum complexity or description code, taking into account local and global aspects of occlusion. Their model assumes that the most likely interpretation is the one that minimizes the sum of the complexity of three components of the visual pattern: (i) The internal structure, related to each of the visible shapes separately, (ii) the external structure, related to the positional relation between these shapes, and (iii) the virtual structure, related to the occluded parts of the shapes. The perceptual complexity of each of these three components is expressed in terms of structural information theory (SIT) [31], a formal coding model that encodes complexity in terms of descriptive parameters. However, van Lier et al. do not automatically complete the occluded objects and the complexities are manually estimated from line drawings; thus their approach can not be directly applied to images in a computer vision task. The same authors noticed in [52] that the global minimum principle can be settled in a Bayesian framework (see [29] and references therein) by properly defining prior and conditional probabilities. In this paper, based on a Bayesian framework, we propose a fully automatic method that can be applied to any image decomposed in shapes.

3 The model

Our model is grounded in two elements: a disocclusion method that computes the different objects conforming different potential scenes that are compatible with the given planar image, and a probabilistic model that quantitatively justifies which scene configuration is the preferred. As we are considering two-depth images, the possible interpretations or hypothesis of the real 3D scene are three, namely: object A occluding object B, B occluding A, or A and B fitting together forming a mosaic (see Figure 6). We will denote them by H_1 , H_2 , and H_3 . Let us remark that sometimes the objects in the third interpretation H_3 (i.e., A and B fitting together) coincide with one of the others, both perceptually and using our algorithm (an example is shown in Figure 6(b)), or even the objects in all three hypothesis coincide (an example is shown in Table 3). Even if the objects forming the scene coincide in different hypothesis, the depth ordering is not the same in each hypothesis. In Section 5 we will provide some experiments analyzing this phenomenon which is related to the well-known optical illusion of relative depth perception of the objects (see Figure 5). In this work, when H_3 coincides with one of the other hypothesis, we assume that both objects appear at the same depth (see Figure 6(b)) and then the associated probabilities will decide which hypothesis has the highest likelihood. In the Perception community, the observed image is often called the *proximal stimulus* (e.g., the left image in Figure 6(a) and (b)), and each of the hypothesized interpretations H_i is called the *distal stimulus*.

Our method first computes several distal interpretations of the scene which are compatible with the proximal planar image. Rubín [46] studied the role of T-junctions, relatability and surface-similarity in the amodal completion phenomenon and illusory contour perception. The author proposed that T-junctions, being a local cue for occlusion, are used to launch the completion

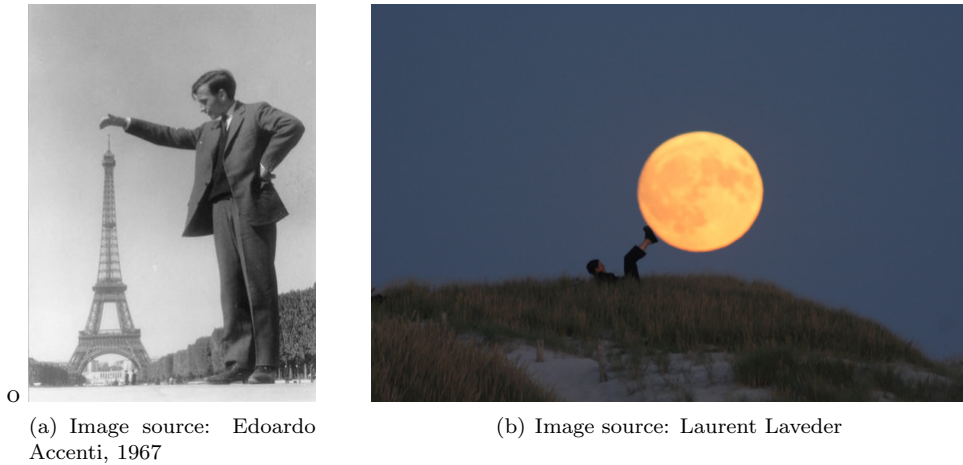


Figure 5: Two well-known optical illusions where the actual depth order of the scene objects is ambiguous or undetermined.

process when contours are relatable [28]. Then, the Gestalt law of good continuation plays an important role. This motivates us to use the Euler’s elastica in order to smoothly continue the contours. Let us recall that, given two T-junctions at points x_1 and x_2 , with tangents τ_{x_1} and τ_{x_2} to the respective terminating stems (also called T-stems, see Figure 4), Euler solved in 1744 [41] the problem of joining them with a smooth continuation curve minimizing

$$\int_{\gamma} (\kappa^2(s) + \beta) ds, \quad (1)$$

where $\beta > 0$ and the minimum is taken among all the curves γ joining x_1 and x_2 with tangents τ_{x_1} and τ_{x_2} , respectively, $\kappa(s)$ denotes the curvature of γ , ds its arc length, and β is a positive constant. The parameter β plays a geometric role by settling the expected underlying a priori regularity. In this sense, with a larger β , the energy favors the completion with straight lines (minimal length). Otherwise, smooth curves of low curvature are favored even if their length is increased. Figure 7 illustrates the effect of the parameter β ; as β decreases the disoccluded shape converges to a disk; in the limit case $\beta = 0$, the energy to be minimized is the Willmore energy (with the boundary constraints on the tangents). The elastica energy is not lower semicontinuous [4] and some relaxed versions have been proposed [4, 36, 3], which are compatible with Kanisza’s amodal completion theory [26]. It has been frequently used to solve different computer vision problems (e.g. [43, 41, 36, 51, 32] among others). In a recent work that proposes a computational method for modal completion [22], the elastica is a key ingredient to obtain illusory contours. A method for both modal and amodal completion which uses geodesics in the group of rotations and translations was proposed in [11]. In this work, the elastica is used in two ways. We propose in Section 3.1 an elastica-based object disocclusion method which incorporates the relatability of partially occluded contours and the convexity of the disoccluded objects. On the other hand, the elastica is also used in Section 3.2 to select the most probable disoccluded scene.

3.1 Elastica-based object disocclusion

For disoccluding the objects, we focus on the completion that takes place in the first time instants of observation [40] or when the local processes dominate due to limited regularities in the object or low saliency of the symmetry cues versus the good continuation [48]. In the perception community, this completion is usually called local completion. The disocclusion method we propose integrates global and local cues: Global cues such as relatability and convexity are incorporated in the initial step of our algorithm, followed by local ones such as smooth continuation.

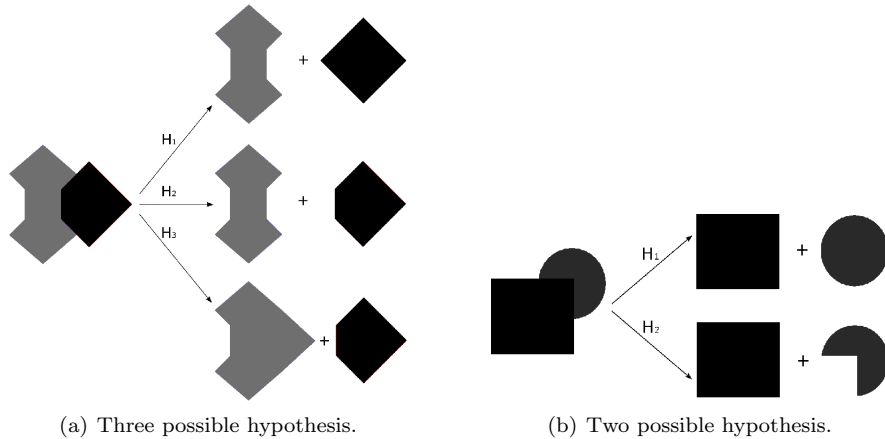


Figure 6: Two examples with its corresponding hypothesis, H_i , for describing the 3D scene structure that gives rise to the observed image (the left-one in each example). In Figure 6(b), where $H_3 = H_2$, the depth of the three-quarters-of-a-disk shape and the square is not the same: in H_3 both objects appear at the same depth and the real relative size is (exactly) the one reflected in the proximal image, while in H_2 the three-quarters-of-a-disk shape is closer to the observer and the square could be of real bigger size but farther away

We propose to disocclude partially occluded objects by a binary inpainting algorithm that simulates the minimization of the elastica (1). Disocclusion, also known as image completion or inpainting, is the recovery of missing or corrupted parts of an image in a given region so that the reconstructed image looks natural. Most available methods for inpainting can be divided into two groups: geometry [36, 3, 9, 8, 35, 47, 6] and texture-oriented methods [13, 12, 58, 27, 2, 1, 33]. The synthesis of methods of these two types is still an open question [6]. Since we are interested in recovering objects or shapes, we will focus on geometry-oriented methods, where images are usually modeled as functions with some degree of smoothness, expressed, for instance, in terms of the curvature of the level lines or the total variation of the image. Taking advantage of this structure, these methods interpolate the inpainting domain by continuing the geometric structure of the image (its level lines or edges), usually as the solution of a (geometric) variational problem or by means of a partial differential equation.

In this paper, we are concerned solely by the shape of the objects. Thus, we work with segmented objects and we perform a geometric inpainting of the binary images that represent these objects. More precisely, we disocclude each object in each hypothesis by separately considering the hypothesized occluding object as the inpainting mask. The object is automatically completed in such a way that its boundary minimizes a relaxed version of the elastica (1). For that, the object to be completed is represented in a binary image (given by the object segmentation) and its completion is performed through a threshold dynamics algorithm which mainly consists in a diffusion process followed by a thresholding. In our case, the minimization algorithm iteratively alternates one step of the Grzybowski-Heintz scheme [19] that decreases $\int_{\gamma} \kappa^2 ds$, one step of the standard Merriman-Bence-Osher scheme [38] that decreases $\beta \int_{\gamma} ds$, and a thresholding step, as proposed by Esedoglu et al. in [47]. We present the pseudo-code and more details in Algorithm 2. Figure 7 shows an example illustrating how the parameter β affects the disoccluded shape. When β is big, more weight is given to the length of the curve and then straight lines are favored. When β decreases the disoccluded shape converges to a disk avoiding singularities of the curvature no matter if it produces a bigger length. On the other hand, depending on the resolution of the proximal stimulus, which translates into a smaller or bigger curvature of curvy boundaries, the parameter β needs to be adapted to obtain the same underlying shape regularity. An example is shown in Figure 8: circles with larger radius need a larger value of β in order to obtain the same regularity of the disoccluded shape. The reason is the following: the curvature of smooth

plane curves is defined as the inverse of the radius of the osculating circle (the unique circle which most closely approximates the curve near the point). Therefore, there is a relationship between the numerical curvature of the disoccluded objects and the a priori regularity imposed through the parameter β : the larger the β , the larger the expected radius of the osculating circle.

In Section 3.2 we define prior probabilities and likelihoods which take into account global and local properties of the shape. As a consequence, the Bayesian approach is able to choose the more likely amodal completion not only between the different hypothesis on the scene configuration for a fixed disocclusion parameter β (as shown in Sect. 3.2), but also between several disocclusions associated to different parameters β , and therefore to integrate some global completion properties such as symmetry or repetitions. For instance, Figure 7 presents the different hypothesis certainty, denoted by \tilde{p} in Sect 3.2, for different values of β . However, for the experimental results in Section 5, where the non-occluded part of the shapes have no constant curvature, the parameter β has been fixed; an efficient method to compute the best β goes out of the scope of the present work.

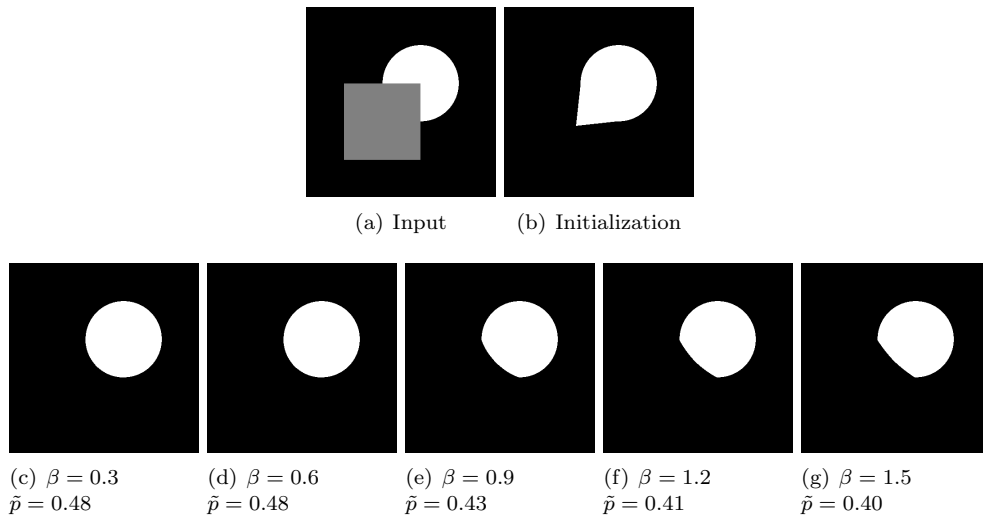


Figure 7: Disocclusion results for different values of β and its associated certainty \tilde{p} . When β grows, more weight is given to the length of the disoccluded boundary γ and then disoccluded straight lines are favored no matter if curvature singularities appear at the endpoints of γ . On the other hand, when $\beta > 0$ is small, more weight is given to the square of the curvature and thus the algorithm avoids singularities of the curvature no matter if it produces a bigger length (as β decreases the disoccluded shape converges to a disk; in the limit case $\beta = 0$, the energy to be minimized is the Willmore energy).

3.1.1 Initialization of the inpainting mask

Since the elastica energy (1) is not convex, the inpainting result depends on the initial condition inside the inpainting mask. Let us illustrate it with a simple example. In Figure 9 we show the inpainting results (shown in the second row) obtained by minimizing the elastica with different initializations, namely, initializing the mask with white, black, random (black and white chosen randomly from a uniform distribution) or with our proposal, which is explained in the remainder of this section. Notice how the proposed initialization gives a better result (according to the Gestalt laws of perception) and produces a completion that maintains the tangents at the endpoints of the disoccluded boundary.

In order to automatically compute an initialization of the inpainting problem sufficiently close to what humans perceive as disoccluded objects by amodal completion, we incorporate perceptual cues such as relatability of object contours [28] and convexity of the disoccluded objects.

The notion of relatability (see Figure 2) was introduced by Kellman and Shipley [28] in the

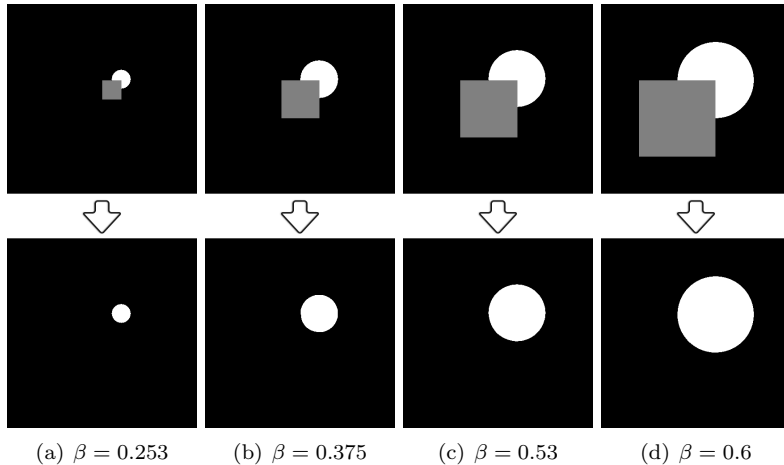


Figure 8: Depending on the resolution of the proximal stimulus, the parameter β needs to be adapted to obtain the same underlying shape regularity (in this case, a constant curvature of the completed shape). For smaller circles, a smaller value of β is needed to obtain a perfect disk (more details are given in the text).

attempt of defining under which conditions visual completion occurs. Let us recall the definition of relatability.

Definition 1 (Relatability [28, 50]). Let E_1 and E_2 be two (non-closed) edges and let x_1 and x_2 be one of their respective end-points. Let τ_{x_1} and τ_{x_2} be the tangents at these points, pointing to the direction along which the interpolated contour must continue. Consider the semi-lines $s_1 = \{x_1 + \lambda\tau_{x_1}, \lambda \geq 0\}$ and $s_2 = \{x_2 + \lambda\tau_{x_2}, \lambda \geq 0\}$. Then, E_1 and E_2 are relatable if: (a) s_1 and s_2 intersect, and (b) their outer angle of intersection (i.e. from τ_{x_1} to $-\tau_{x_2}$) is acute or 90° .

In [50], the authors showed that this definition is equivalent to the existence of a smooth contour without inflection points connecting x_1 and x_2 , and that the interpolating curve does not turn through a total angle of more than $\frac{\pi}{2}$.

Since (non-occluded) objects in the world tend to be convex [5], we favor the convexity of the disoccluded object by taking advantage of the following well-known property of convex sets.

Lemma 1. *Every closed convex set in R^n is the intersection of the closed half-spaces that contain it.*

The automatic initialization of the binary image inside the inpainting mask is illustrated in Figure 10. In practice, our algorithm considers all the end-points of the object contours (given in this case by the level lines) arriving to the inpainting mask together with their tangents (illustrated in Figure 10(b) by a line passing through them), and computes all the possible pairs of relatable contours (shown in Figures 10(c) and 10(h)). In order to compute these tangents we use the Line Segment Detector [56]. Then, for each pair of relatable contours, for the end-point x_i and tangent τ_{x_i} we consider the half-space $\{x \in \mathcal{R}^2 : \langle \tau_{x_i}^\perp, x \rangle - \langle \tau_{x_i}^\perp, x_i \rangle \geq 0\}$ (or ≤ 0 , depending on which half-space the object is), and we assign a vote to the half-space on which the known object is. Figure 10(d) and Figure 10(i) displays the image gathering these votes in the inpainting mask, for the shapes shown in Figure 10(a) and Figure 10(f), respectively (the shape to disocclude is shown in white and the inpainting mask is shown in gray, respectively). Let us remark that, in order to better illustrate our perceptually inspired initialization, in Figure 10(d) and Figure 10(e) (respectively, in Figure 10(i) and Figure 10(j)) we only show the computed values inside the inpainting mask. Finally, we binarize the image containing the votes with a threshold based on a rank order filter of these votes. We order the votes in increasing order and start with a threshold with the value ranked at percentile 75th. If no new connected components appear in the initialization with this

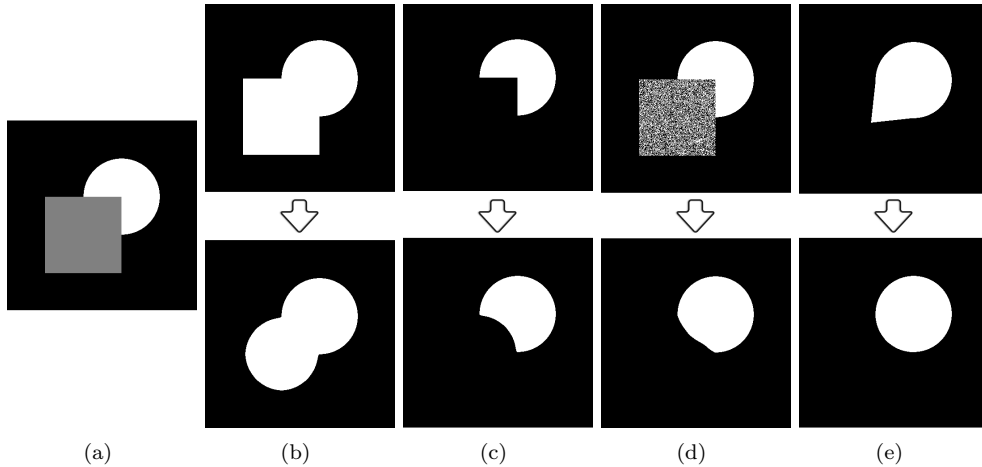


Figure 9: Each column contains an inpainting experiment with the same input image (a) where the inpainting mask is given by the square. The first row shows several initializations, namely, initializing the mask with white (b), black (c), random (d) or with our proposal (e). The second row shows the inpainted result of the corresponding image above. Notice how the proposed initialization gives a better result (according to the Gestalt laws of perception) and produces a completion that maintains the tangents at the endpoints of the disoccluded boundary.

threshold we keep it. Otherwise we decrease the threshold (taking the preceding ordered value) and repeat the process until no new connected components appear. Two different examples of this binary image are shown in Figure 10(e) and 10(j), they are the initialization of the binary inpainting algorithm. Figure 10(e) shows an example where the threshold on the votes correspond to the 75th percentile while in Figure 10(j) the threshold was automatically decreased to the 65th percentile in order to obtain an initialization with a single connected component.

3.2 Elastica-based probabilistic model

In this section, we follow a Bayesian approach [29] in order to choose among all possible interpretations of the scene, the most plausible one. We propose definitions for the prior and the conditional probabilities which take into account the global complexity of the objects in the hypothesized scenes as well as the effort of bringing these objects in their relative positions in the visual image. As a consequence, the result of this probability model indicates that the most simple interpretation is the one that more likely results from the amodal completion process, which was also suggested by [52].

Inspired by the work of van Lier et al. [54], our probabilistic model takes into account the complexity of the objects. Each hypothesized scene is formed both by completely visible objects and disoccluded objects (computed using the method described in previous Section 3.1); their respective global complexities are taken into account to define the prior probability of the hypothesized scene under analysis. The likelihood, i.e. the conditional probability of the given image (proximal stimulus) given a certain hypothesis (distal stimulus) is defined through an Euler’s elastica-based quantity that measures two attributes: the effort of bringing these objects in their relative positions given in the image and the smoothness of the disoccluded boundaries. Our probabilistic model provides a formalization allowing to computationally verify directly on images the proposal of van der Helm [54, 52] (based on manually estimating complexity from line drawings) giving a probabilistic interpretation of the visual completion process.

We propose to predict and justify the preferred interpretation by maximizing the responsibility

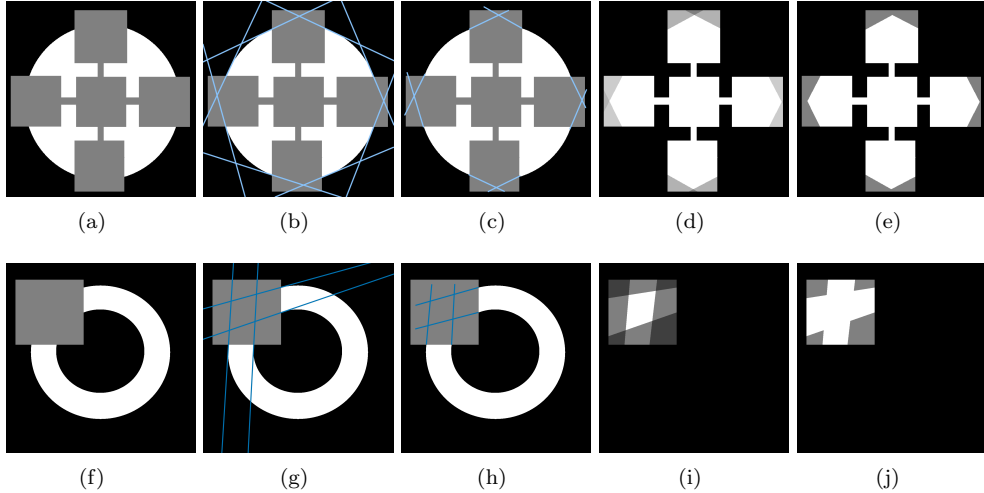


Figure 10: Two examples of reliability- and convexity-based initialization of the inpainting mask. (a), (f): shapes to disocclude (in white) and inpainting mask (in gray); (b), (g): extension by semi-lines (in blue, see Def. 1) of the tangents of the level-lines arriving to the inpainting mask; (c), (h): reliable contours (in blue); (d), (i): votes on the different half-spaces defined by any pair of reliable contours (darker colors mean less votes); (e), (j): final initialization of the mask, after thresholding.

or a posterior probability, given by the Bayes' rule as

$$p(H_i/I) = \frac{p(I/H_i)p(H_i)}{p(I)} \quad (2)$$

over the hypothesized interpretations H_i , where I is the proximal stimulus or given image. As the quotient $p(I)$ remains the same for all hypothesis H_i in the maximization process, we propose to select the preferred hypothesis H_P by

$$H_P = \arg \max_i p(H_i/I) = \arg \max_i \tilde{p}(I/H_i) \tilde{p}(H_i). \quad (3)$$

Given the underlying hypothesis H_i and the proximal image I , we define the conditional probability $p(I/H_i)$ as

$$p(I/H_i) \propto \tilde{p}(I/H_i) = e^{-\omega_1 \int_{\mathcal{B}_i^c} (\kappa^2 + \beta) ds} e^{-\omega_1 \int_{\mathcal{B}_i^d} (\kappa^2 + \beta) ds}, \quad (4)$$

where \mathcal{B}_i^c and \mathcal{B}_i^d stand for common and disoccluded boundaries, respectively (see Figure 11 for an example with two hypothesis) and ω_1 is a normalization constant. Formula (4) measures the responsibility that hypothesis H_i takes for explaining the proximal stimulus I as well as the deviation of I from H_i .

With the first integral in (4) we compute the difficulty of bringing the two objects together in order to get the perceived image taking into account only the known boundary of the objects; for example, it is easier to obtain configuration 11(a) than 11(b) as in the first case only two points need to coincide, independently of the two coinciding points we will perceive the same image, and in the other case, H_2 , a larger boundary needs to coincide in order to perceive exactly that configuration. The second integral takes into account the regularity of the occluded boundary of the shape to define the probability of obtaining a particular stimulus; for example in Figure 12(a) we can move the disk at many different positions behind the square to obtain the same image we are observing, but in Figure 12(b) the movements we can do are more limited, as the perceived image will change drastically. Let us remark that due to the way we disocclude the objects the resulting disoccluded boundaries are always smooth; if we had different models of disocclusion this

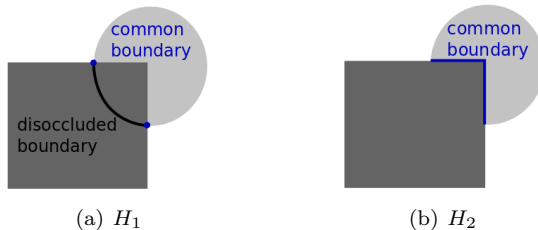


Figure 11: Boundaries \mathcal{B}_i^c and \mathcal{B}_i^d ($i = 1, 2$) for hypothesis H_1 and H_2 from Figure 6(b). (a) H_1 : square in front of circle; when the circle is disoccluded the common boundary \mathcal{B}_1^c among both objects is formed by the 2 T-junction points (in blue), while the disoccluded boundary \mathcal{B}_1^d is made of all the boundary of the circle that was behind the square (in black) (b) H_2 : square at the same depth than the circle-part; the common boundary \mathcal{B}_2^c is shown in blue while the disoccluded boundary coincides with it, $\mathcal{B}_1^d = \mathcal{B}_1^c$, due to the fact that we consider closed objects (see details in the text).

term would help to distinguish among them (in addition to the prior term). For instance, with our disocclusion model based on the elastica we are not able to recover the occluded object in Figure 12(b) or the objects A in Figure 3(b). The probability distribution in (4) also appeared in Mumford [41] and Williams and Jacobs [59], who characterized the probability distribution of the shape of boundary completions based on the paths followed by a particle undergoing a stochastic motion, a directional random walk. It turns out that the elastica has the interpretation of being the mode of the probability distribution underlying this stochastic process restricted to curves with prescribed boundary behavior (the maximum likelihood curve with which to reconstruct hidden contours).

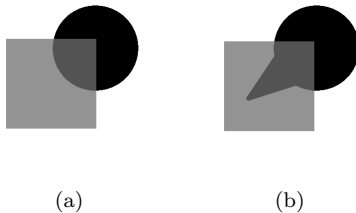


Figure 12: Example of two different disocclusions and how they affect the likelihood. In the example (a), where the reconstructed object is a disk, we can move the disk at many different positions behind the square to obtain the same kind of image (stimulus) we are observing. While in example (b) there is no so much freedom to move the position of the occluded object without changing the appearance of the observed image.

The prior probabilities are defined as

$$p(H_i) \propto \tilde{p}(H_i) = e^{-\omega_2 \text{compl}(O_i^1)} e^{-\omega_2 \text{compl}(O_i^2)}, \quad (5)$$

where ω_2 is a normalizing constant, and O_i^1 and O_i^2 are the (disoccluded) objects in the hypothesized interpretation H_i . The factor $\text{compl}(O_i^j)$ denotes the complexity of the object or shape O_i^j at depth j . In the case that the object at one depth is formed by more than one connected component the complexity is computed separately for each connected component and their sum constitutes the complexity of O_i^j . We use the definition of complexity of a shape defined by Chen and Sundaram in [10],

$$C = (1 + R) (0.6 \cdot \min(C_{\text{dist}}, C_{\text{angle}}) + 0.07 \max(C_{\text{dist}}, C_{\text{angle}}) + 0.33P), \quad (6)$$

which takes into account global properties of the shape such the global distance entropy (C_{dist}), the local angle entropy (C_{angle}), the perceptual smoothness (P), and a measure of shape randomness (R). The global distance is defined in [10] as the distance of boundary points to the centroid of the shape. The local angle is the angle formed by the two segments joining three consecutive boundary points. The perceptual smoothness is computed using the local angle (as closer to π the angle, the smoother the shape). Finally, the shape randomness is the maximum difference between two random traces obtained from the two more distant points of the boundary. Therefore, the prior probability considers global properties such as shape contour symmetries and repetitions.

Let us notice that with these definitions our whole model for amodal completion is able to choose, not only between the different hypothesis for a fixed disocclusion parameter β but also between several disocclusions associated to different parameters β , and therefore to take into account global completion properties such as symmetry or repetitions. In Figure 7 there is an example illustrating this computational ability, where the different probabilities associated to the different disocclusion results depending on β are given.

Both normalization constants, ω_1 and ω_2 , are defined, respectively, by the inverse of the maximum value, over all the hypothesis H_i , of the elastica and the object complexity.

Let us comment on the term $e^{-\omega_1 \int_{\mathcal{B}^c} (\kappa^2 + \beta) ds}$ in our definition (4). When visual completion occurs while propagating the stem, (e.g., hypothesis H_1 in Figure 6(b); also, hypothesis H_1 in Figure 11(a)), the common boundaries \mathcal{B}^c between the objects are reduced to the T-junctions. In this case: $\int_{\mathcal{B}^c} (\kappa^2 + \beta) ds = 0$ and thus $e^{-\omega_1 \int_{\mathcal{B}^c} (\kappa^2 + \beta) ds} = 1$. Let us notice that in the distal stimulus, since we are considering closed objects, \mathcal{B}^c belongs to both objects. Therefore, in the hypothesis where the objects are interpreted as being fit-together (e.g., hypothesis H_2 in Figure 6(b); also, hypothesis H_2 in Figure 11(b)), a disoccluded boundary \mathcal{B}^d appears which coincides with \mathcal{B}^c (i.e., $\mathcal{B}^d = \mathcal{B}^c$). Let us also comment on the effect of the regularity of \mathcal{B}^c . Figure 13 presents three different proximal stimuli or images. The numerical computation of the term $e^{-\omega_1 \int_{\mathcal{B}^c} (\kappa^2 + \beta) ds}$ associated to each of the three images will decrease from left to right in the fit-together (or mosaic) interpretation, the same behavior applies to the complexity-related terms $e^{-\omega_2 \text{compl}(O_1^i)}$ and $e^{-\omega_2 \text{compl}(O_2^i)}$. Therefore, the visual completion will become more and more evident and the interpretation of two complex pieces fitting together will become perceptually less favorable.

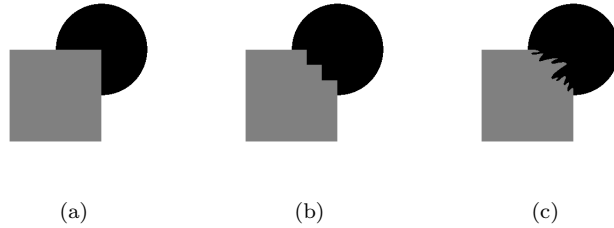


Figure 13: Three different proximal stimulus or images. From left to right, visual completion will become more and more evident than the interpretation of two pieces fitting together, both perceptually and quantitatively, with probability (4).

Let us finally remark that we are not considering *all* possible configurations [18] but only the ones favored by relatability, convexity, and good continuation. On the other hand, even if global cues such as symmetry or repetitions are taken into account in our probability model, we do not incorporate them in the disocclusion algorithm. In the future, we plan to integrate it with other disocclusion strategies (such as, e.g., exemplar-based methods [2, 1] or [21]) allowing to model these global properties and obtain, e.g., the objects A in Figure 3(b).

4 Algorithm and implementation details

Algorithm 1 shows the steps of the whole numerical algorithm. Let us detail it: Our algorithm needs a decomposition of the given image into objects and object parts which are interpreted as projections of real 3D objects on the image plane. This decomposition can be given either from the classical decomposition in level sets, in bi-level sets or segmenting the image from a criterion. In this paper, for the synthetic images, we use the decomposition in bi-level sets, which are defined as usual by $X^{(\lambda_n, \lambda_{n+1})}I = \{x \in \Omega : \lambda_n \leq I(x) < \lambda_{n+1}\}$, where Ω is the image domain and $\{\lambda_n\} \subset \mathcal{R}$ is a finite strictly increasing sequence; and for the real images, we use the segmented shapes from the Berkeley segmentation dataset [34]. In this way we obtain the objects that appear in the image; these objects will be denoted by X_1 and X_2 . From X_1 and X_2 the three hypothesis will be considered by the algorithm: X_1 occluding the distal object D_2 (corresponding to the proximal X_2), X_2 occluding the distal object D_1 (corresponding to the proximal X_1), and X_1 and X_2 fit together. Now, by applying the disocclusion method of Section 3.1 where X_1 and X_2 are, respectively, the inpainting mask, we compute the *complete* hypothesis $H_1 = X_1 \cup D_2$ and $H_2 = X_2 \cup D_1$, respectively. Then, to this two hypothesis, we always add the additional hypothesis $H_3 = X_1 \cup X_2$ of the *mosaic* interpretation (which is obtained when we do not apply the disocclusion algorithm). For each H_i we compute the probabilities $\tilde{p}(I/H_i)$ and $\tilde{p}(H_i)$ from the definitions in Sect. 3.2. Finally, we compute the perceptually preferred hypothesis H_P by (3).

Input : An image I (the proximal stimulus) with objects X_1 and X_2 at two depths.

Output: The set of distal hypothesis H_1, H_2, H_3 , each one made of complete objects at two depths, and the preferred one H_P (with $P \in \{1, 2, 3\}$).

```

for  $i \in \{1, 2, 3\}$  do
  if  $i \neq 3$  then
    • Consider  $X_i$  as inpainting mask and initialize the inpainting mask using the
      perceptual method described in Sect. 3.1.1
    • Disocclude object  $X_j$ , with  $j \neq i$ , using Algorithm 2 implementing the
      elastica-based method of Sect. 3.1. From it, we obtain the disoccluded object  $D_j$  and
      the completed hypothesis  $H_i = X_i \cup D_j$ .
  else
    | • Set  $H_3 = X_1 \cup X_2$ 
  end
  • Compute the probabilities  $\tilde{p}(I/H_i)$  with Algorithm 3 and  $\tilde{p}(H_i)$  (equation (6)) from
  the definitions in Sect. 3.2.
end
Set  $H_P = \arg \max_i \tilde{p}(I/H_i) \tilde{p}(H_i)$ .

```

Algorithm 1: Pseudo-code summarizing the proposal.

In Algorithm 2 we describe the threshold dynamics method we use for disocclusion, and in

Algorithm 3 we present the algorithm for computing the conditional probability $\tilde{p}(I/H_i)$, $i = 1, 2, 3$.

Input : A binary image I containing a region without information (the inpainting region \tilde{M}), the inpainting region $\tilde{M} \subset \Omega$, and a elastica parameter $\beta > 0$.

Output: Disoccluded object D (given by an inpainted binary image \bar{I})

- Set $\alpha = 0.99$ and $\delta t = 12$.
- Set the initial shape $\Sigma_0 = \{x : I(x) = 1\}$.
- Set $n = 0$ and $\Sigma_1 = \Omega$.

while $\|\Sigma_{n+1} - \Sigma_n\| > 10^{-3}$ **do**

1. A step of Grzibovskis-Heintz algorithm. Set:

$$\Gamma_1 = \left\{ x : 2\alpha G_{\sqrt{\delta t}} * \mathbb{1}_{\Sigma_n}(x) - 2G_{\alpha^2 \sqrt{\delta t}} * \mathbb{1}_{\Sigma_n}(x) \leq \alpha - 1 \right\}.$$

2. A step of standard Merriman-Bence-Osher algorithm. Set:

$$\Gamma_2 = \left\{ x : G_{\beta \delta t} * \mathbb{1}_{\Gamma_1}(x) \geq \frac{1}{2} \right\}.$$

3. Fidelity step. Set

$$\Sigma_{n+1} = \left(\Gamma_2 \cap \tilde{M} \right) \cup \left(\Omega \setminus \tilde{M} \right).$$

$$n = n + 1.$$

end

- $\bar{I} = \mathbb{1}_{\Sigma_n}$ and $D = \Sigma_n$.

Algorithm 2: Pseudo-code of the disocclusion algorithm.

Let us add some details regarding Algorithm 2. The Gaussian convolution has been computed using the Lindeberg's discrete scale-space method and its implementation described in [44], that is, we use that the Gaussian convolution $v : t \mapsto G_{\sqrt{\delta t}} * u$ is the solution of the heat equation $\frac{\partial v}{\partial t} = \Delta v$ for a diffusion time δt (set to 12 in our experiments, to guarantee the prescribed upper and lower bounds depending on the curvature of the visible shape [38]) so we only need to discretize partial derivatives. We refer to [44] for more details on the discretization. Parameter α needs to be close but less than 1 [47, 14].

Input : Inpainting masks X_1, X_2 , disoccluded objects D_1, D_2 , elastica parameter β .

Output: Conditional probability of each hypothesis H_1, H_2, H_3 .

- Compute the boundaries $\partial X_1, \partial X_2, \partial D_1, \partial D_2$ of X_1, X_2, D_1, D_2 , respectively.
- Set $\mathcal{B}_i^c = \partial X_1 \cap \partial X_2$, $i = 1, 2, 3$.

for $i = 1, 2$ **do**

$$\quad | \quad \mathcal{B}_i^d = \partial D_i \setminus \partial X_i$$

end

- Set $\mathcal{B}_3^d = \mathcal{B}_3^c$

for $\{i = 1, 2, 3\}$ **do**

$$\quad | \quad E_{B_i} = \sum_{x \in \mathcal{B}_i^c} (\kappa^2(x) + \beta) + \sum_{x \in \mathcal{B}_i^d} (\kappa^2(x) + \beta)$$

end

- Set $\omega_1 = \max\{E_{B_1}, E_{B_2}, E_{B_3}\}$

for $i = 1, 2, 3$ **do**

$$\quad | \quad \tilde{p}(I/H_i) = \exp\{-\omega_1 E_{B_i}\}$$

end

Algorithm 3: Pseudo-code of the algorithm computing $\tilde{p}(I/H_i)$ for $i = 1, 2, 3$.

Regarding Algorithm 3, the discrete boundaries of each shape are computed as external boundaries and using 4-connectivity. On the other hand, in order to compute the curvature of a discrete

curve (or boundary) γ , we use the method of [49] and compute

$$\kappa(x) = \operatorname{div} \left(\frac{\nabla u(x)}{|\nabla u(x)|} \right), \quad (7)$$

where u is the signed distance function to the boundary γ . We use forward derivatives to compute the gradient, backward derivatives for the divergence. The discrete signed distance function u is computed using the algorithm explained in [37].

Finally, the prior probability is computed using (5) with the complexity measure given by (6). We consider as boundary points for computing (6) all the pixels that form the boundary of an object. In case of an object formed by more than one connected component we compute the complexity (6) of every connected component and the final complexity measure is the addition of the individual complexities. For details about how to compute $R, C_{\text{dist}}, C_{\text{angle}}$ and P we refer to Section 3.2 and to [10].

5 Experimental results

The proposed method has been tested with both synthetic and real images. Parameter β , which sets the underlying a priori regularity (see comments on its role in Section 3.1) has been fixed to 0.6 for all the experiments in order to have an algorithm as general as possible. There are two exceptions, namely, Proximal 2 of Table 1 and Example 4 of Table 11, where β is fixed to 1.2 and 1.7, respectively, due to the biggest size of the circular shapes. As explained in Section 3.1, there is a relationship between the numerical curvature of the disoccluded objects and the a priori regularity imposed through the parameter β : the larger the β , the larger the expected radius of the osculating circle locally approximating the curve.

The experiments of this Section are organized as follows: In Sections 5.1, 5.2 and 5.3 we introduce the experiments which agree with our perception, while in Section 5.4 we show and discuss the experiments that failed. The synthetic experiments, which are described in the following Section 5.1, are shown in Tables 1, 2 and 3 while the experiments on real images (described in Section 5.2) are shown in Tables 4, 5, 6 and 7. Table 5 shows our results on images of the Berkeley dataset with provided figure-ground ground-truth labeling [15]. Table 9 in Section 5.3 shows the ability of our method to also decide on (perceptually) fully visible objects over a background. Finally, Tables 10, 11 and 12 present the synthetic and real results where our method did not agree with human perception. For each row in each table we show a complete experiment.

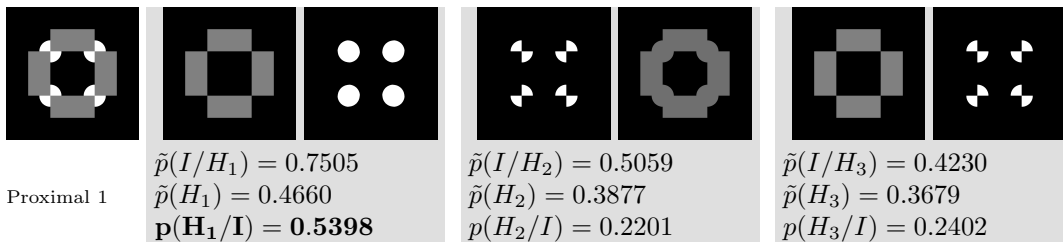
Let us recall that our method assumes the proximal stimulus to be decomposed into objects and object parts (which can be interpreted as projections of real 3D objects on the image plane). As in the synthetic experiments, the images are formed by objects with a single and unique color, this already gives a segmentation and we apply our algorithm directly. For the real experiments, we use a segmentation of the image. In particular, we have taken segmented images from the Berkeley segmentation dataset [34] and from [20].

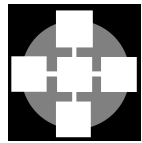
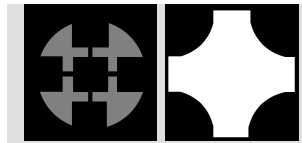
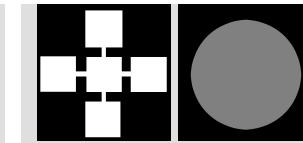
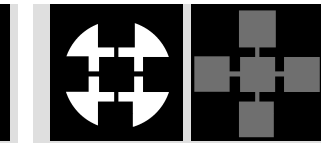
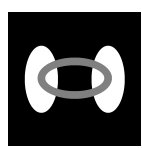
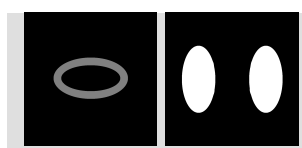
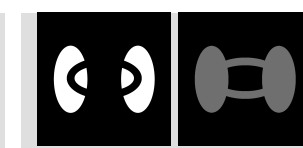
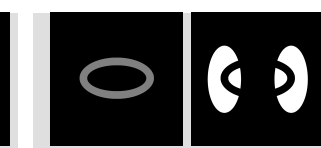

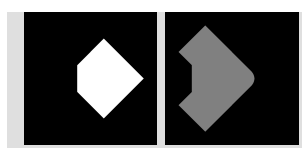
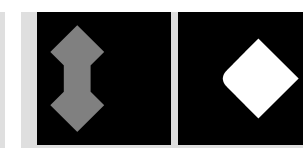
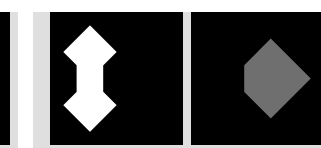
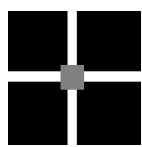
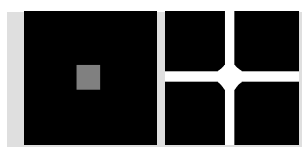
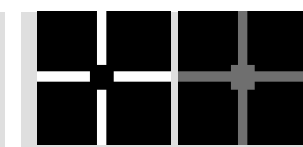
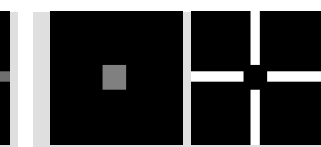
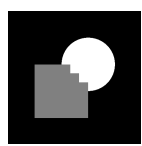
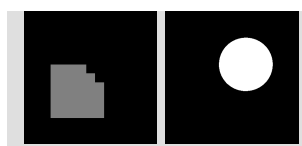
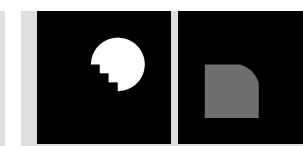
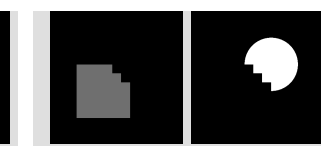

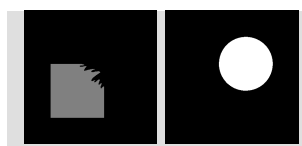
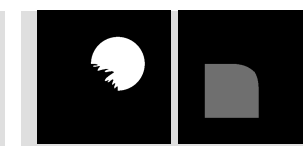
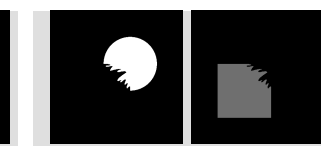
5.1 Synthetic images

Tables 1, 2 and 3 show some experiments on synthetic images. For each row in each table, a complete experiment is shown. We first present the proximal image (piecewise constant) I on the left, followed by the three hypothesis H_i (each one separated by a gray box), together with the values $\tilde{p}(I/H_i)$ and $\tilde{p}(H_i)$ proportional to the conditional probability and the prior probability, respectively, and the probability value $p(H_i/I)$. Let us remark that we have normalized the probabilities in such a way that $p(H_1/I) + p(H_2/I) + p(H_3/I) = 1$. The probability value of the preferred hypothesis H_P is highlighted in boldface. For the first two hypothesis, H_1 and H_2 , we display the objects at depth 1 on the left, and the disoccluded objects (at depth 2) on the right. Let us recall that the objects at depth 1 are considered the inpainting mask for disoccluding the objects at depth 2. Finally the last column is the hypothesis H_3 where the two objects are fitting together at the same depth.

Let us comment on the results in Tables 2 and 3. In Table 2, the third hypothesis is not shown because it coincides with H_2 due to the fact that the disocclusion algorithm does not change the objects being disoccluded. Besides, in Table 3 there is shown a synthetic experiment where the three hypothesis coincide on account of the obtained disoccluded objects: The disocclusion algorithm applied in the first two hypothesis does not change the objects and thus $H_1 = H_2 = H_3$, and the posterior probability is the same for all three hypothesis. Let us remark that, even if the objects forming the scene coincide in different hypothesis, the depth ordering is not the same in each hypothesis. Let us singularize and explain one example in Figure 6(b), where our method produces $H_3 = H_2$. However, as for depth order, H_3 is interpreted as two objects at the same depth (and having the real relative size which is observed in the proximal image) while H_2 can be interpreted as three quarters of a disk which is closer to the observer, plus a square which can be of bigger size but farther away from the three-quarters-of-a-disk shape and whose boundary partially coincides with part of the boundary of the three-quarters-of-a-disk shape. Notice that this situation is related to the well-known ambiguity in depth of some proximal stimulus, sometimes causing optical illusion of relative depth perception as those in the images displayed in Figure 5.

These experiments show that our method agrees with human perception. The perception literature acknowledges that, in a T-junction, the occluder is the surface on the T-head side while the surfaces on the T-stem side continue behind the occluder [46]. This phenomenon is validated by our method: The preferred hypothesis, highlighted in boldface, is the one that is obtained by continuation of the T-stems. Let us comment on the results corresponding to Proximal 6 and 7 of Table 1, which include quite similar shapes with equal occlusion signatures but different common boundaries among the shapes. In Proximal 7 and 8 (and also in Proximal 9 of Table 2), the local perception cue at the T-junctions indicates that there is an occluded disk which continues behind an incomplete square (the occluder). Our method is able to choose the corresponding preferred hypothesis as is shown by the probability values $p(H_1/I)$. In Proximal 12 on Table 2, according to the T-junctions we should prefer H_1 (as the method chooses), but for symmetries most of us perceptually prefer H_2 . In this case, according to the prior H_2 should be preferred (as symmetries are valued positively), but as the disoccluded and common boundary are so large in H_2 , the value $\tilde{p}(I/H_1)$ is much bigger, and H_1 is selected as preferred. Finally, let us comment on Proximal 4 which is a well-known and perceptually controversial example. The preferred hypothesis for Proximal 4 is the one that agrees with the T-junction cues and the one reported in [54] to be the most preferred by the subjects participating in their psychophysics experiments. However, in this case the posterior probabilities for H_2 and H_1 are quite close and, according to our personal experience (e.g., by incorporating our knowledge about the world and objects of similar shapes), some people prefer the interpretation according to H_2 . Both experiments, Proximal 4 and 12, are examples where hypothesis H_1 and H_2 correspond, respectively, to a local and global completion of the occluded object. In both cases, our algorithm favors local completion, that is, a completion that produces good continuation instead of the global one which produces a more symmetric object (notice that the local completion in both cases produces a symmetric object with respect to one axis). As commented in Section 2, both kind of completions interplay and the prevalence of one of them depends on the observation time [40] and on the saliency of the good-continuation versus symmetry in the completion [48].



 <p>Proximal 2</p>	 <p> $\tilde{p}(I/H_1) = 0.7557$ $\tilde{p}(H_1) = 0.3679$ $p(H_1/I) = 0.2366$ </p>	 <p> $\tilde{p}(I/H_2) = 0.8388$ $\tilde{p}(H_2) = 0.8793$ $p(H_2/I) = \mathbf{0.6279}$ </p>	 <p> $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4327$ $p(H_3/I) = 0.1355$ </p>
 <p>Proximal 3</p>	 <p> $\tilde{p}(I/H_1) = 0.9343$ $\tilde{p}(H_1) = 0.5544$ $p(H_1/I) = \mathbf{0.6391}$ </p>	 <p> $\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.3738$ $p(H_2/I) = 0.1696$ </p>	 <p> $\tilde{p}(I/H_3) = 0.4214$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1813$ </p>
 <p>Proximal 4</p>	 <p> $\tilde{p}(I/H_1) = 0.7466$ $\tilde{p}(H_1) = 0.3910$ $p(H_1/I) = \mathbf{0.3943}$ </p>	 <p> $\tilde{p}(I/H_2) = 0.7445$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.3700$ </p>	 <p> $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4743$ $p(H_3/I) = 0.2357$ </p>
 <p>Proximal 5</p>	 <p> $\tilde{p}(I/H_1) = 8823$ $\tilde{p}(H_1) = 0.6020$ $p(H_1/I) = \mathbf{0.5265}$ </p>	 <p> $\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.3879$ $p(H_2/I) = 0.1414$ </p>	 <p> $\tilde{p}(I/H_3) = 0.9108$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.3321$ </p>
 <p>Proximal 6</p>	 <p> $\tilde{p}(I/H_1) = 0.4409$ $\tilde{p}(H_1) = 0.7056$ $p(H_1/I) = \mathbf{0.4618}$ </p>	 <p> $\tilde{p}(I/H_2) = 0.5562$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.3037$ </p>	 <p> $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4295$ $p(H_3/I) = 0.2345$ </p>
 <p>Proximal 7</p>	 <p> $\tilde{p}(I/H_1) = 0.7343$ $\tilde{p}(H_1) = 0.5995$ $p(H_1/I) = \mathbf{0.4917}$ </p>	 <p> $\tilde{p}(I/H_2) = 0.7823$ $\tilde{p}(H_2) = 0.4087$ $p(H_2/I) = 0.3572$ </p>	 <p> $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1512$ </p>


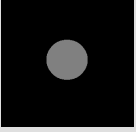
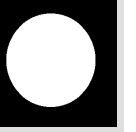
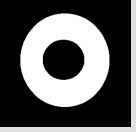
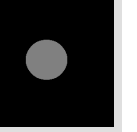
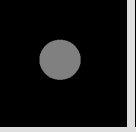
 Proximal 8					
	$\tilde{p}(I/H_1) = 1$ $\tilde{p}(H_1) = 0.7087$ $\mathbf{p(H_1/I) = 0.7223}$	$\tilde{p}(I/H_2) = 0.3729$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.1398$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1379$		

Table 1: Synthetic experiments. Each row shows a different experiment: the original image (proximal stimulus) is shown on the left and it is followed by the three different hypothesis (each one separated by a gray box). For the first two hypothesis, H_i for $i = 1, 2$, we show: the object at depth 1 (left) and the disoccluded object at depth 2 (right). Notice that the object at depth 1 acts as a mask for disoccluding the object at depth 2. In the case of the third hypothesis, H_3 , both objects are considered to be at the same depth and completely visible in the original image (no disocclusion is applied). In the lower part of each hypothesis we show the values $\tilde{p}(I/H_i)$, $\tilde{p}(H_i)$ (proportional, respectively, to the likelihood and prior probabilities), and the posterior probability $p(H_i/I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface.



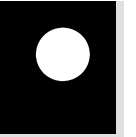

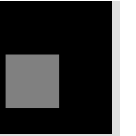
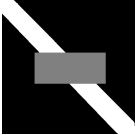

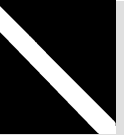
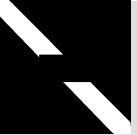
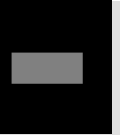
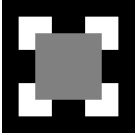
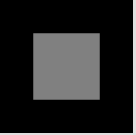
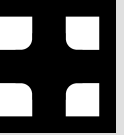
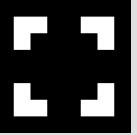
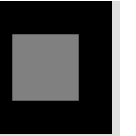


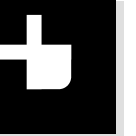
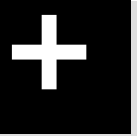

 Proximal 9				
	$\tilde{p}(I/H_1) = 0.6848$ $\tilde{p}(H_1) = 0.6487$ $\mathbf{p(H_1/I) = 0.6214}$	$\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.1893$		
 Proximal 10				
	$\tilde{p}(I/H_1) = 0.3877$ $\tilde{p}(H_1) = 0.3804$ $\mathbf{p(H_1/I) = 0.3527}$	$\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.3237$		
 Proximal 11				
	$\tilde{p}(I/H_1) = 0.5436$ $\tilde{p}(H_1) = 0.4051$ $\mathbf{p(H_1/I) = 0.4486}$	$\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.2757$		
 Proximal 12				
	$\tilde{p}(I/H_1) = 0.6172$ $\tilde{p}(H_1) = 0.3679$ $\mathbf{p(H_1/I) = 0.4165}$	$\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.4324$ $p(H_2/I) = 0.2918$		

Table 2: Synthetic experiments. Each row shows a different experiment: the original image (proximal stimulus) is shown on the left and it is followed by two different hypothesis (each one separated by a gray box). For each hypothesis H_i we show: the object at depth 1 (left) and the disoccluded object at depth 2 (right). Notice that the object at depth 1 acts as a mask for disoccluding the object at depth 2. In the lower part of each hypothesis we show the values $\tilde{p}(I/H_i)$, $\tilde{p}(H_2)$ (proportional, respectively, to the likelihood and prior probabilities), and the posterior probability $p(H_i/I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface. The third hypothesis, H_3 , where both objects are considered to be at the same depth and completely visible in the original image (no disocclusion is applied) is not shown here because it coincides with H_2 (due to the fact that the disocclusion algorithm does not change the objects being disoccluded in H_2). More details are given in the text.

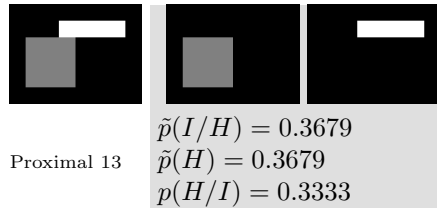


Table 3: Synthetic experiment where the three hypothesis coincide (since the disocclusion applied in the first two hypothesis does not change the objects). Thus, we have $H = H_1 = H_2 = H_3$, and the posterior probability is the same for all three hypothesis. As for depth order, H_3 is interpreted as two objects at the same depth (and having the real relative size which is observed in the proximal image) while H_1 can be interpreted as a gray square which is closer to the observer, plus a white rectangle which can be of bigger size but farther away from the square and whose boundary partially coincides with part of the boundary of the square. Finally, H_2 can be interpreted as a white rectangle which is closer to the observer, plus a gray square which can be of bigger size but farther away from the rectangle and whose boundary partially coincides with part of the boundary of the rectangle.

5.2 Real images

In this section we show some results on real images from the Berkeley dataset [34] and the dataset provided in [20]. For all experiments, we present the real image, the proximal stimulus I which is a segmentation of the real image (one of the segmentations provided in the databases), followed by the three hypothesis H_i (each one separated by a gray box), together with the values $\tilde{p}(I/H_i)$ and $\tilde{p}(H_i)$ proportional to the conditional probability and the prior probability, respectively, and the probability value $p(H_i|I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface. For the first two hypothesis, H_1 and H_2 , we display the objects at depth 1 on the left, and the disoccluded objects (at depth 2) on the right. Let us recall that the objects at depth 1 are considered the inpainting mask for disoccluding the objects at depth 2. Finally the last column is the hypothesis H_3 where the two objects are fitting together at the same depth.

We start illustrating that our method is robust to different segmentations of the same image. Table 4 shows a real image with a bear holding a branch and two different segmentations (representing the proximal stimuli). Both segmentations are from the ground truth available in [34]. Segmentation 1 reflects that some flowers are partially occluding the bear and increasing the complexity of the bear shape; the flowers do not appear in segmentation 2 and thus the bear shape has a lower complexity (its complexity is 0.53, while in the previous case, Segmentation 1, was 1.34). Notice that the values \tilde{p} are not comparable among the two experiments (only among different hypothesis within the same experiment) because they use a different normalizing constant ω_2 (see Section 3.2 for further details). Finally, the most preferred interpretation of the image coincides using the two different segmentations, i.e., it is a branch partially occluding a bear for

both stimulus.






























							
Segmentation 1							
	$\tilde{p}(I/H_1) = 0.7818$		$\tilde{p}(I/H_2) = 0.7590$		$\tilde{p}(I/H_3) = 0.3679$		
$\tilde{p}(H_1) = 0.6996$		$\tilde{p}(H_2) = 0.3717$		$\tilde{p}(H_3) = 0.3679$			
$\mathbf{p(H_1/I) = 0.4184}$		$\mathbf{p(H_2/I) = 0.3966}$		$\mathbf{p(H_3/I) = 0.1850}$			
Segmentation 2							
	$\tilde{p}(I/H_1) = 0.6739$		$\tilde{p}(I/H_2) = 0.6676$		$\tilde{p}(I/H_3) = 0.3679$		
$\tilde{p}(H_1) = 0.5265$		$\tilde{p}(H_2) = 0.3679$		$\tilde{p}(H_3) = 0.3751$			
$\mathbf{p(H_1/I) = 0.4805}$		$\mathbf{p(H_2/I) = 0.3326}$		$\mathbf{p(H_3/I) = 0.1869}$			

Table 4: Experiments with real images from [34]. The left-most image is a segmentation of the original image (shown in the first row). Both experiments correspond to the same image but considering different segmentations (both extracted from the ground truth segmentation available in [34]). The most preferred interpretation of the image coincides in both experiments, i.e. a branch partially occluding a bear. Notice that the values \tilde{p} are not comparable among the two experiments (only among different hypothesis within the same experiment) because they use a different normalizing constant ω_2 (see Section 3.2 for further details).

In Table 5 we present results on images of the Berkeley dataset with provided figure-ground ground-truth labeled by humans. Then, Table 6 shows experimental results on real images from [20] and Table 7 shows results on images from the Berkeley Segmentation database [34]. Each row shows a different experiment: the two left-most images are, respectively, the original image and a segmentation of it, they are followed by the three different hypothesis (each one separated by a gray box). For the images in Table 5, superimposed on the original image, we display the provided figure-ground ground-truth [15] as a boundary in two colors, namely, black and white. The black side of the border indicates the object that is behind, while the white region indicates the frontal object.

Image 1							
	$\tilde{p}(I/H_1) = 0.8665$		$\tilde{p}(I/H_2) = 0.6820$		$\tilde{p}(I/H_3) = 0.3679$		
$\tilde{p}(H_1) = 0.5212$		$\tilde{p}(H_2) = 0.3701$		$\tilde{p}(H_3) = 0.3679$			
$\mathbf{p(H_1/I) = 0.5380}$		$\mathbf{p(H_2/I) = 0.3007}$		$\mathbf{p(H_3/I) = 0.1612}$			
Image 2							
	$\tilde{p}(I/H_1) = 1$		$\tilde{p}(I/H_2) = 0.8952$		$\tilde{p}(I/H_3) = 0.3679$		
$\tilde{p}(H_1) = 0.3688$		$\tilde{p}(H_2) = 0.3724$		$\tilde{p}(H_3) = 0.3692$			
$\mathbf{p(H_1/I) = 0.4410}$		$\mathbf{p(H_2/I) = 0.3972}$		$\mathbf{p(H_3/I) = 0.1618}$			

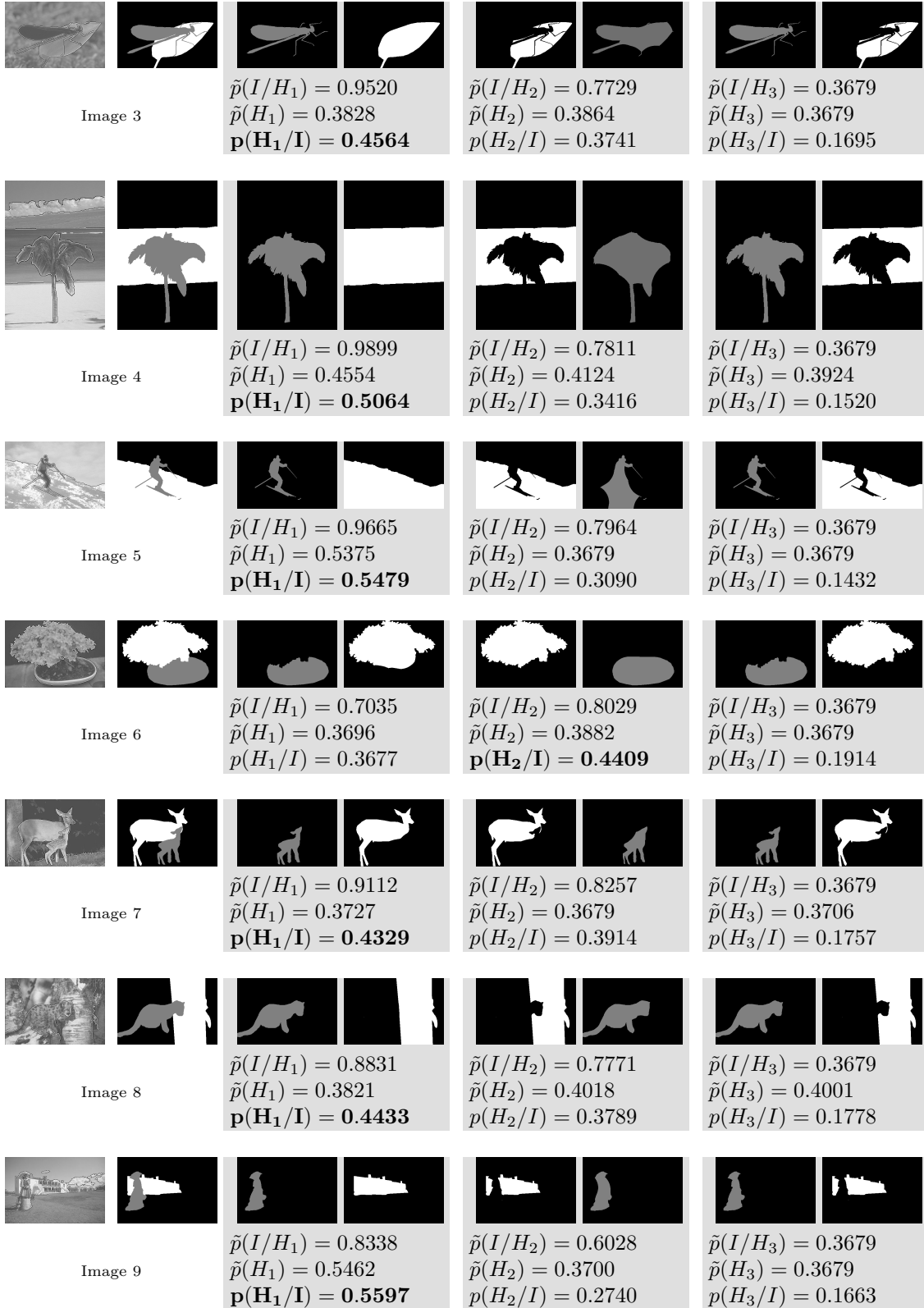


Image 10	$\tilde{p}(I/H_1) = 0.8935$ $\tilde{p}(H_1) = 0.3679$ $p(H_1/I) = 0.3390$		$\tilde{p}(I/H_2) = 0.9504$ $\tilde{p}(H_2) = 0.5218$ $p(H_2/I) = \mathbf{0.5115}$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.5219$ $p(H_3/I) = 0.1495$		
Image 11	$\tilde{p}(I/H_1) = 0.8059$ $\tilde{p}(H_1) = 0.5367$ $p(H_1/I) = \mathbf{0.5002}$		$\tilde{p}(I/H_2) = 0.8029$ $\tilde{p}(H_2) = 0.3696$ $p(H_2/I) = 0.3432$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1565$		
Image 12	$\tilde{p}(I/H_1) = 0.8940$ $\tilde{p}(H_1) = 0.3764$ $p(H_1/I) = 0.3620$		$\tilde{p}(I/H_2) = 0.9192$ $\tilde{p}(H_2) = 0.4979$ $p(H_2/I) = \mathbf{0.4924}$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1456$		
Image 13	$\tilde{p}(I/H_1) = 0.7809$ $\tilde{p}(H_1) = 0.3712$ $p(H_1/I) = 0.3293$		$\tilde{p}(I/H_2) = 0.9201$ $\tilde{p}(H_2) = 0.4947$ $p(H_2/I) = \mathbf{0.5170}$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1537$		

Table 5: Experiments with real images from [34]. Each row shows a different experiment: the two left-most images are, respectively, the original image and a segmentation of it. They are followed by the three different hypothesis (each one separated by a gray box). The lines superimposed in the original image indicate the figure/ground ground-truth labels (from [15]) for each boundary in the segmentation: the white boundary indicates the figure side and the black one the ground side. For the first two hypothesis, H_i for $i = 1, 2$, we show: the object at depth 1 (left) and the disoccluded object at depth 2 (right). Notice that the object at depth 1 acts as a mask for disoccluding the object at depth 2. In the case of the third hypothesis, H_3 , both objects are considered to be at the same depth and completely visible in the original image (no disocclusion is applied). In the lower part of each hypothesis we show the values $\tilde{p}(I/H_i)$, $\tilde{p}(H_i)$ (proportional, respectively, to the likelihood and prior probabilities), and the posterior probability $p(H_i/I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface.



























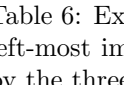
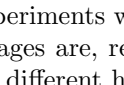
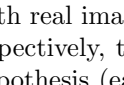
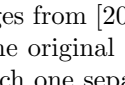
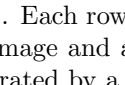
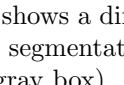
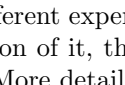
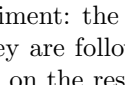
Image 14									$\tilde{p}(I/H_1) = 0.8526$ $\tilde{p}(H_1) = 0.3701$ $p(\mathbf{H}_1/I) = \mathbf{0.4356}$	$\tilde{p}(I/H_2) = 0.7416$ $\tilde{p}(H_2) = 0.3688$ $p(H_2/I) = 0.3776$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4472$ $p(H_3/I) = 0.1868$
Image 15									$\tilde{p}(I/H_1) = 0.8796$ $\tilde{p}(H_1) = 0.3750$ $p(\mathbf{H}_1/I) = \mathbf{0.4539}$	$\tilde{p}(I/H_2) = 0.7074$ $\tilde{p}(H_2) = 0.3680$ $p(H_2/I) = 0.3582$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3729$ $p(H_3/I) = 0.1879$
Image 16									$\tilde{p}(I/H_1) = 0.9112$ $\tilde{p}(H_1) = 0.4034$ $p(\mathbf{H}_1/I) = \mathbf{0.5058}$	$\tilde{p}(I/H_2) = 0.6035$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.3055$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1887$
Image 17									$\tilde{p}(I/H_1) = 0.8833$ $\tilde{p}(H_1) = 0.3718$ $p(\mathbf{H}_1/I) = \mathbf{0.4263}$	$\tilde{p}(I/H_2) = 0.8279$ $\tilde{p}(H_2) = 0.6354$ $p(H_2/I) = 0.3980$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.6334$ $p(H_3/I) = 0.1757$

Table 6: Experiments with real images from [20]. Each row shows a different experiment: the two left-most images are, respectively, the original image and a segmentation of it, they are followed by the three different hypothesis (each one separated by a gray box). More details on the results shown for each hypothesis in Table 5.

























Image 18									$\tilde{p}(I/H_1) = 0.7913$ $\tilde{p}(H_1) = 0.3685$ $p(H_1/I) = 0.2990$	$\tilde{p}(I/H_2) = 0.8845$ $\tilde{p}(H_2) = 0.6163$ $p(\mathbf{H}_2/I) = \mathbf{0.5608}$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1392$
Image 19									$\tilde{p}(I/H_1) = 0.6789$ $\tilde{p}(H_1) = 0.5380$ $p(H_1/I) = \mathbf{0.4808}$	$\tilde{p}(I/H_2) = 0.7031$ $\tilde{p}(H_2) = 0.4302$ $p(H_2/I) = 0.3405$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4313$ $p(H_3/I) = 0.1787$
Image 20									$\tilde{p}(I/H_1) = 0.8933$ $\tilde{p}(H_1) = 0.5970$ $p(\mathbf{H}_1/I) = \mathbf{0.5533}$	$\tilde{p}(I/H_2) = 0.7496$ $\tilde{p}(H_2) = 0.3715$ $p(H_2/I) = 0.3063$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1404$

Image 21	$\tilde{p}(I/H_1) = 0.8579$ $\tilde{p}(H_1) = 0.4983$ $p(\mathbf{H}_1/I) = \mathbf{0.4820}$		$\tilde{p}(I/H_2) = 0.7293$ $\tilde{p}(H_2) = 0.4976$ $p(H_2/I) = 0.3441$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4976$ $p(H_3/I) = 0.1739$		
Image 22	$\tilde{p}(I/H_1) = 0.7574$ $\tilde{p}(H_1) = 0.3679$ $p(H_1/I) = 0.2904$		$\tilde{p}(I/H_2) = 0.9523$ $\tilde{p}(H_2) = 0.5717$ $p(\mathbf{H}_2/I) = \mathbf{0.5674}$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.03709$ $p(H_3/I) = 0.1422$		
Image 23	$\tilde{p}(I/H_1) = 0.8382$ $\tilde{p}(H_1) = 0.3722$ $p(H_1/I) = 0.3911$		$\tilde{p}(I/H_2) = 0.9034$ $\tilde{p}(H_2) = 0.3877$ $p(\mathbf{H}_2/I) = \mathbf{0.4392}$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1697$		
Image 24	$\tilde{p}(I/H_1) = 0.8392$ $\tilde{p}(H_1) = 0.4028$ $p(\mathbf{H}_1/I) = \mathbf{0.4467}$		$\tilde{p}(I/H_2) = 0.7695$ $\tilde{p}(H_2) = 0.7417$ $p(H_2/I) = 0.3742$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.7414$ $p(H_3/I) = 0.1788$		

Table 7: Experiments with real images from [34]. Each row shows a different experiment: the two left-most images are, respectively, the original image and a segmentation of it, they are followed by the three different hypothesis (each one separated by a gray box). More details on the results shown for each hypothesis in Table 5.

We present in Table 8 two experiments with real images from [34] where there is an ambiguity in the depth ordering (there are conflicting local depth cues). This situation can appear when the proximal image is made of objects that are not fronto-parallel to the camera or when their relative order changes due to, for example, mutual occlusions as in these examples. In other words, an object does not appear at a single depth layer. In this situation our algorithm chooses the object that is more occluded as being behind but let us remark how the posterior probabilities of the two first hypothesis are very close; in fact, these two hypothesis correspond to the two different depth orderings indicated by the local depth cues and figure/ground ground-truth labels superimposed on the original image.

Image 25	$\tilde{p}(I/H_1) = 0.7757$ $\tilde{p}(H_1) = 0.3709$ $p(H_1/I) = 0.3986$		$\tilde{p}(I/H_2) = 0.7942$ $\tilde{p}(H_2) = 0.3763$ $p(H_2/I) = 0.5140$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1875$		

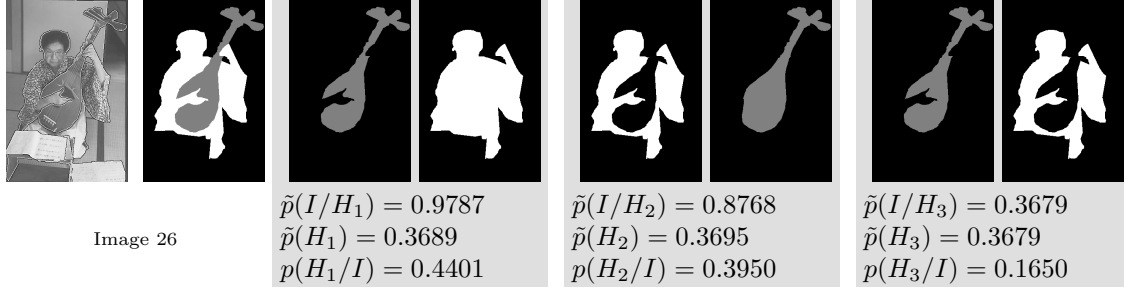
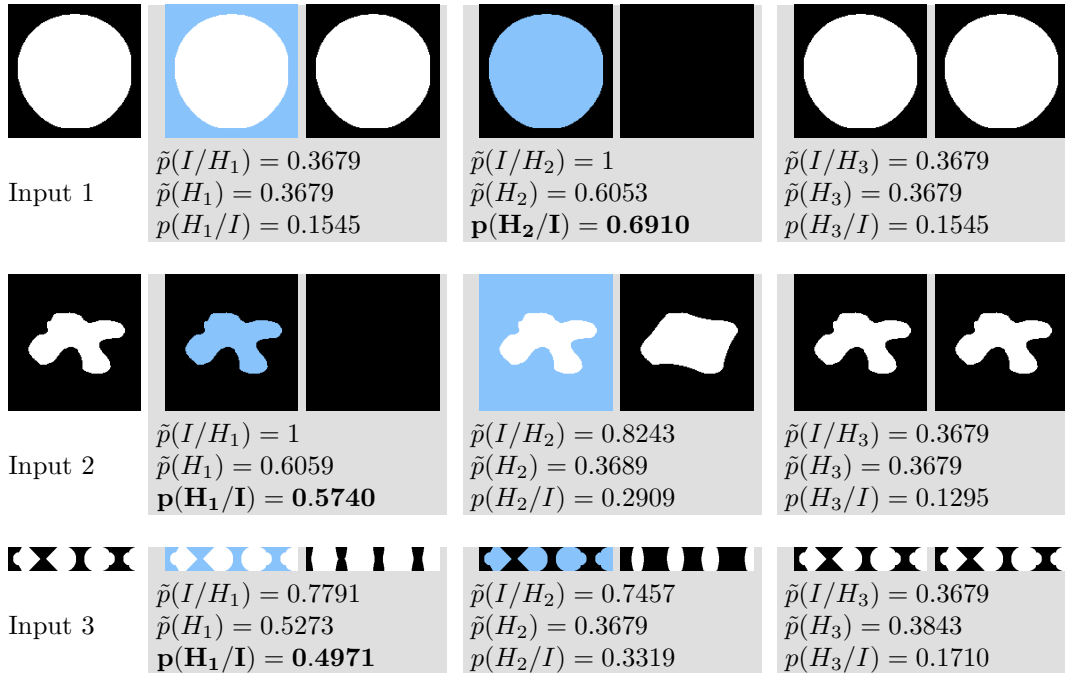


Table 8: Two experiments with real images from [34] where there is an ambiguity in the depth ordering (there are conflicting local depth cues). Notice how the posterior probabilities of the two first hypothesis are very close; in fact, these two hypothesis correspond to the two different depth orderings indicated by the local depth cues and figure/ground ground-truth labels superimposed on the original image. More details on the results shown in each row in Table 5.

5.3 Shapes in front of a background

Finally, we present some results showing the ability of our method to also decide on (perceptually) fully visible objects over a background. Table 9 displays several synthetic images of this type, where there are no T-junctions present and, according to human perception, the depth ordering is established by convexity cues [26] (in the perception literature, they are also called figure-ground images). For each row, the proximal stimulus is shown in the first column, and the following columns are the different hypothesis separated by a gray box. For hypothesis H_1 and H_2 , the object which is supposed to be at depth 1, i.e. the occluding object, is displayed in blue. In H_3 , where both objects are fitted together, the objects are presented with its original color. In these experiments, our method fails in Results 3, 4 and 5; in all of them the convexity cue is a stronger depth cue than symmetry, while the algorithm we are using for computing shape complexity favors symmetries. Let us also remark that Result 9 allows both interpretations: black in front of white and white in front of black, as they form the same shape but with different orientation. In this case our algorithm prefers H_1 but with a small difference with respect to H_2 .



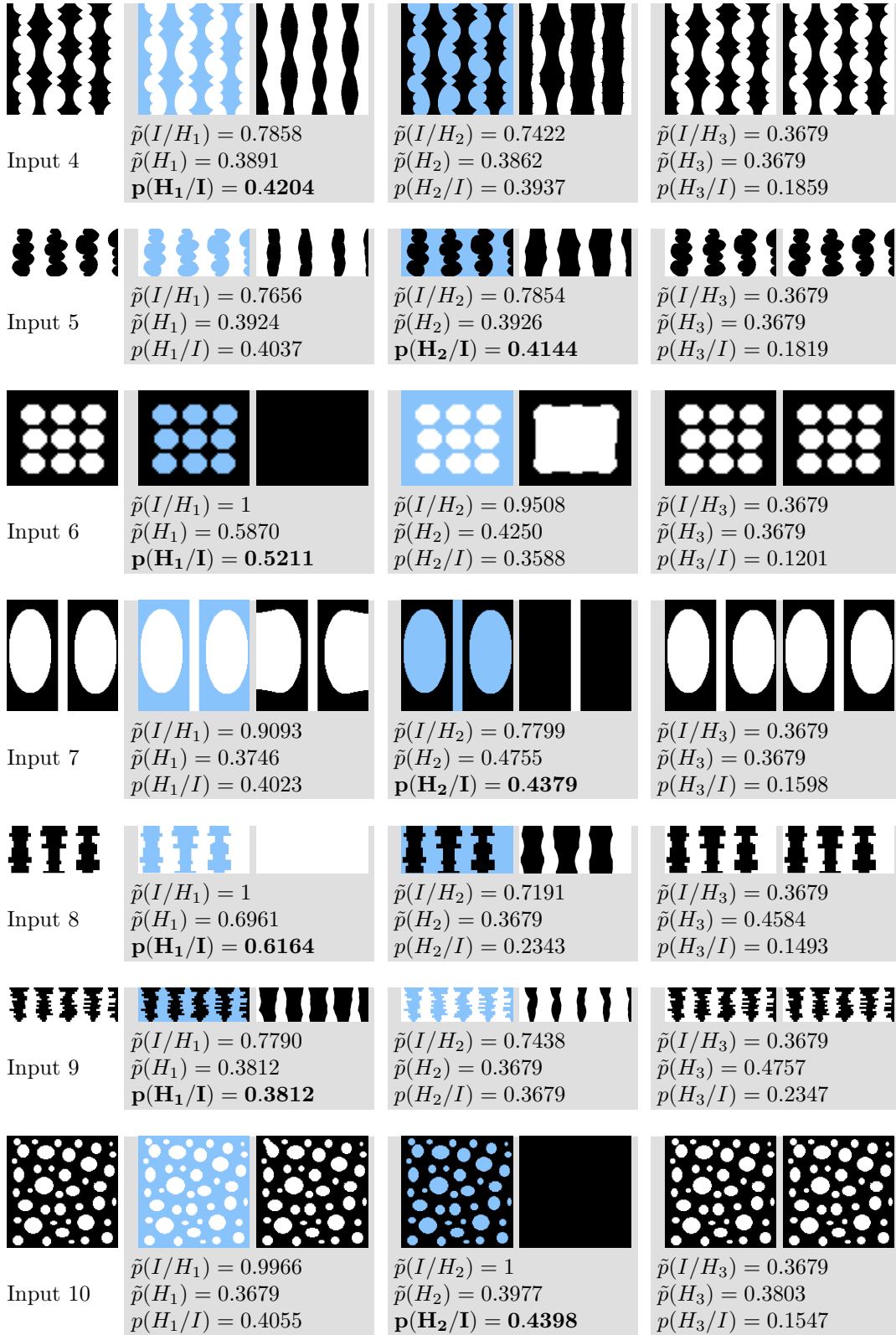


Table 9: Synthetic experiments with a shape in front of a background. Each row shows a different experiment: the original image (proximal stimulus) is shown on the left and it is followed by the three different hypothesis (each one separated by a gray box). For the first two hypothesis, H_i for $i = 1, 2$, we show: the object at depth 1 in blue (left) and the disoccluded object at depth 2 (right). In the case of the third hypothesis, H_3 , both objects are considered to be at the same depth and completely visible in the original image (no disocclusion is applied). In the lower part of each hypothesis we show the values $\tilde{p}(I/H_i)$, $\tilde{p}(H_2)$ (proportional, respectively, to the likelihood and prior probabilities), and the posterior probability $p(H_i/I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface.

5.4 Discussion of failure cases

This section is devoted to present and discuss situations where our method can fail. Let us first comment on the experiments on synthetic images. Example 1 and Example 2 of Table 10 do not agree with human perception. As commented before, in a T-junction, the occluder is the surface on the T-head side while the surfaces on the T-stem side continue behind the occluder. However, in these two results, the local occlusion signatures given by the T-junctions indicate that there is an occluded square which continues behind an incomplete disk (the occluder). Taking this into account, our method fails to give the hypothesis that agrees with the T-junction cues (which should be H_2). In Example 1, the likelihood of the hypothesis H_1 and H_2 are similar but the global complexity of the shapes in H_1 is smaller (thus higher prior) than the global complexity of the shape in H_2 . In particular, any of the two shapes present in H_1 is simpler than any of the two shapes in H_2 . Regarding Example 2, the highly irregular contour of the shapes makes difficult a straightforward analysis and the final chosen hypothesis is due to a balance among the corresponding complexities and likelihoods. In the examples of Table 11, according to the local cues given by the T-junctions, the preferred option should be H_1 . However, our method obtains H_2 . In Example 3, although a higher prior due to a smaller complexity of the objects in H_1 , the likelihood of H_2 is higher due to smaller elastica values in H_2 . Example 4 is the opposite: H_1 presents a higher complexity (thus smaller prior) and a higher likelihood.

Example 1	$\tilde{p}(I/H_1) = 0.6198$ $\tilde{p}(H_1) = 0.6784$ $p(H_1/I) = \mathbf{0.5587}$		$\tilde{p}(I/H_2) = 0.4949$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = 0.2420$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.4077$ $p(H_3/I) = 0.1993$	
Example 2	$\tilde{p}(I/H_1) = 0.6948$ $\tilde{p}(H_1) = 0.3876$ $p(H_1/I) = \mathbf{0.4026}$		$\tilde{p}(I/H_2) = 0.6535$ $\tilde{p}(H_2) = 0.4043$ $p(H_2/I) = 0.3951$		$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.2023$	

Table 10: Synthetic experiments that fail. Each row shows a different experiment: the original image (proximal stimulus) is shown on the left and it is followed by the three different hypothesis (each one separated by a gray box). For the first two hypothesis, H_i for $i = 1, 2$, we show: the object at depth 1 (left) and the disoccluded object at depth 2 (right). Notice that the object at depth 1 acts as a mask for disoccluding the object at depth 2. In the case of the third hypothesis, H_3 , both objects are considered to be at the same depth and completely visible in the original image (no disocclusion is applied). In the lower part of each hypothesis we show the values $\tilde{p}(I/H_i)$, $\tilde{p}(H_2)$ (proportional, respectively, to the likelihood and prior probabilities), and the posterior probability $p(H_i/I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface.



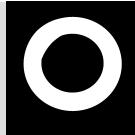


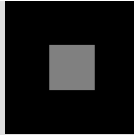









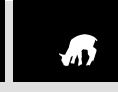
Example 3				
		$\tilde{p}(I/H_1) = 0.3679$ $\tilde{p}(H_1) = 0.3770$ $p(H_1/I) = 0.2951$		$\tilde{p}(I/H_2) = 0.4501$ $\tilde{p}(H_2) = 0.3679$ $p(H_2/I) = \mathbf{0.3524}$
Example 4				
		$\tilde{p}(I/H_1) = 0.4062$ $\tilde{p}(H_1) = 0.3679$ $p(H_1/I) = 0.2853$		$\tilde{p}(I/H_2) = 0.3679$ $\tilde{p}(H_2) = 0.5089$ $p(H_2/I) = \mathbf{0.3574}$

Table 11: Synthetic experiments that fail. Each row shows a different experiment: the original image (proximal stimulus) is shown on the left and it is followed by two different hypothesis (each one separated by a gray box). For each hypothesis H_i we show: the object at depth 1 (left) and the disoccluded object at depth 2 (right). Notice that the object at depth 1 acts as a mask for disoccluding the object at depth 2. In the lower part of each hypothesis we show the values $\tilde{p}(I/H_i)$, $\tilde{p}(H_2)$ (proportional, respectively, to the likelihood and prior probabilities), and the posterior probability $p(H_i/I)$. The probability value of the preferred hypothesis H_P is highlighted in boldface. The third hypothesis, H_3 , where both objects are considered to be at the same depth and completely visible in the original image (no disocclusion is applied) is not shown here because it coincides with H_2 (due to the fact that the disocclusion algorithm does not change the objects being disoccluded in H_2). More details are given in the text.

Example 5				
		$\tilde{p}(I/H_1) = 0.6858$ $\tilde{p}(H_1) = 0.3726$ $p(H_1/I) = \mathbf{0.3962}$	$\tilde{p}(I/H_2) = 0.03726$ $\tilde{p}(H_2) = 0.3704$ $p(H_2/I) = 0.3939$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.2099$
Example 6				
		$\tilde{p}(I/H_1) = 0.7469$ $\tilde{p}(H_1) = 0.3686$	$\tilde{p}(I/H_2) = 0.6515$ $\tilde{p}(H_2) = 0.3695$	$\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$







	$p(\mathbf{H}_1/I) = 0.4227$	$p(H_2/I) = 0.3696$	$p(H_3/I) = 0.2078$
 Example 7	 $\tilde{p}(I/H_1) = 0.8575$ $\tilde{p}(H_1) = 0.3709$ $p(H_1/I) = 0.3709$	 $\tilde{p}(I/H_2) = 0.7297$ $\tilde{p}(H_2) = 0.5332$ $\mathbf{p}(\mathbf{H}_2/I) = 0.5332$	 $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.3679$
 Example 8	 $\tilde{p}(I/H_1) = 0.8732$ $\tilde{p}(H_1) = 0.3679$ $p(H_1/I) = 0.4115$	 $\tilde{p}(I/H_2) = 0.8325$ $\tilde{p}(H_2) = 0.3792$ $\mathbf{p}(\mathbf{H}_2/I) = 0.4118$	 $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1767$
 Example 9	 $\tilde{p}(I/H_1) = 0.8744$ $\tilde{p}(H_1) = 0.3763$ $p(H_1/I) = 0.4082$	 $\tilde{p}(I/H_2) = 0.9118$ $\tilde{p}(H_2) = 0.3747$ $\mathbf{p}(\mathbf{H}_2/I) = 0.4239$	 $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3679$ $p(H_3/I) = 0.1679$
 Example 10	 $\tilde{p}(I/H_1) = 1$ $\tilde{p}(H_1) = 0.3679$ $\mathbf{p}(\mathbf{H}_1/I) = 0.4213$	 $\tilde{p}(I/H_2) = 0.8920$ $\tilde{p}(H_2) = 0.4117$ $p(H_2/I) = 0.4205$	 $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3758$ $p(H_3/I) = 0.1583$
 Example 11	 $\tilde{p}(I/H_1) = 0.7878$ $\tilde{p}(H_1) = 0.3679$ $p(H_1/I) = 0.3623$	 $\tilde{p}(I/H_2) = 1$ $\tilde{p}(H_2) = 0.3737$ $\mathbf{p}(\mathbf{H}_2/I) = 0.4672$	 $\tilde{p}(I/H_3) = 0.3679$ $\tilde{p}(H_3) = 0.3708$ $p(H_3/I) = 0.1705$

Table 12: Experiments with real images that fail . Each row shows a different experiment: the two left-most images are, respectively, the original image and a segmentation of it, they are followed by the three different hypothesis (each one separated by a gray box). More details on the results shown for each hypothesis in Table 5.

Let us comment on the images of Table 12. Examples 5 and 6 reflect the same situation; the inpainting method is unable to recover the leg of the older horse or sheep. In any case, the difference among the posterior probabilities of the first two hypothesis is very small. On the other hand, in Examples 7 and 8, although according to the likelihood the preferred hypothesis is the correct one (e.g., two ladybugs in front of two flowers in Example 7), the complexity of the objects in the second hypothesis (flowers in front of ladybugs) is smaller (higher prior) because of the simplified completed object and this second hypothesis wins. In Example 9, the prior probabilities of H_1 and H_2 are similar but the likelihood of H_2 is slightly higher. Finally, Examples 10 and 11 show the same situation, where there appears a window showing the sky, which is behind. Our method fails in these cases, which are interpreted as small convex shapes over a biggest shape which is behind. Other example would be the arches of a bridge, which are further away compared to the bridge itself, which would be interpreted as closest by our method.

Finally, let us observe that the hypothesis of several objects fitting together (called H_3 in this work) implies an ambiguity in depth. When H_3 coincides with one of the other hypothesis, we assume that both objects appear at the same depth, but this is not always true (although this proximal configuration would probably have smaller likelihood due to its oddity).

6 Conclusions

We have proposed a computational model of amodal completion that allows to compute the most preferred structure of a scene given a still image of it. As we are considering scenes where objects appear at two different depths, we take into account the three different hypothesis: Object A occluding object B, B occluding A, or A and B fitting together forming a mosaic. Our main contribution is a Bayesian probabilistic model based on the elastica and the global complexity of the hypothesized objects in order to choose the most preferred explanation of the image. This explanation includes both the disoccluded objects that form the scene and their ordering according to depth. Furthermore, we have proposed a disocclusion method, to compute the hypothesized objects, based on human visual completion, which is modeled by a binary inpainting method based on the Euler’s elastica and that takes into account perceptual findings related to amodal completion, such as relatability, convexity, and good continuation. Finally, we have shown the capability of our method with numerical experiments, both with real and synthetic images.

As future work, we plan to extend the approach to scenes with more than two depth layers. Furthermore, we plan to incorporate other disocclusion strategies (such as, e.g., exemplar-based methods [2, 1] or [21]) allowing to model global completions taking into account properties such as symmetries or repetitions. Last but not least, we are also interested in the extension of the model to video sequences.

Acknowledgements. The first, second and fifth authors acknowledge partial support by MICINN project, reference MTM2012-30772, and by GRC reference 2014 SGR 1301, Generalitat de Catalunya. The third author is supported by a Beatriu de Pinós fellowship (Marie-Curie COFUND action). The final publication is available at Springer via <http://dx.doi.org/10.1007/s10851-016-0652-x>

References

- [1] P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *International Journal of Computer Vision*, 93:319–347, 2011.
- [2] J.-F. Aujol, S. Ladjal, and S. Masnou. Exemplar-based inpainting from a variational point of view. *SIAM Journal on Mathematical Analysis*, 42(3):1246–1285, 2010.
- [3] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. on IP*, 10(8):1200–1211, 2001.
- [4] G. Bellettini, G. Dal Maso, and M. Paolini. Semicontinuity and relaxation properties of a curvature depending functional in 2d. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 20(2):247–297, 1993.
- [5] J. Burge, C. Fowlkes, and M. Banks. Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *The Journal of Neuroscience*, 30(21):7269–7280, 2010.
- [6] F. Cao, Y. Gousseau, S. Masnou, and P. Pérez. Geometrically guided exemplar-based inpainting. *SIAM Journal on Imaging Sciences*, 4(4):1143–1179, 2011.
- [7] S. Carrigan, E. Palmer, and P. Kellman. Differentiating local and global processes in amodal completion through dot localization. *Journal of vision*, 15(12):1123–1123, 2015.

- [8] T. Chan and J. H. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. App. Math.*, 62(3):1019–43, 2001.
- [9] T. Chan and J. H. Shen. Nontexture inpainting by curvature-driven diffusions. *Journal of Visual Communication and Image Representation*, 12(4):436–449, 2001.
- [10] Y. Chen and H. Sundaram. Estimating complexity of 2d shapes. In *IEEE 7th Workshop on Multimedia Signal Processing*, pages 1–4, 2005.
- [11] G. Citti and A. Sarti. A cortical based model of perceptual completion in the roto-translation space. *Journal of Mathematical Imaging and Vision*, 24(3):307–326, 2006.
- [12] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based inpainting. *IEEE Trans. on IP*, 13(9):1200–1212, 2004.
- [13] L. Demanet, B. Song, and T. Chan. Image inpainting by correspondence maps: a deterministic approach. *Applied and Computational Mathematics*, 1100:217–50, 2003.
- [14] S. Esedoglu, S. J. Ruuth, and R. Tsai. Threshold dynamics for high order geometric motions. *Interfaces and Free Boundaries*, 10(3):263–282, 2008.
- [15] C. Fowlkes, D. Martin, and J. Malik. Local figure/ground cues are valid for natural images. *Journal of Vision*, 7(8):1–9, 2007.
- [16] V. Froyen, N. Kogo, M. Singh, and J. Feldman. Modal and amodal shape completion. *Journal of vision*, 15(12):321–321, 2015.
- [17] W. Gerbino and D. Salmaso. The effect of amodal completion on visual matching. *Acta Psychologica*, 65(1):25–46, 1987.
- [18] R. Grompone von Gioi. Toward a computational theory of perception. *apcap 2009*, page 49, 2009.
- [19] R. Grzibovskis and A. Heintz. A convolution-thresholding scheme for the willmore flow. Technical Report 34.
- [20] T. N. J. Z. H. Li, J. Cai. A benchmark for semantic image segmentation. In *ICME*, 2013.
- [21] T. Hayashi and M. Sasaki. Contour completion of partly occluded skew-symmetry objects. In *Multimedia (ISM), 2014 IEEE International Symposium on*, pages 90–93, 2014.
- [22] S. H. Kang, W. Zhu, and J. Shen. Illusory shapes via corner fusion. *SIAM Journal on Imaging Sciences*, 7(4):1907–1936, 2014.
- [23] G. Kanizsa. Amodal completion and shrinking of visual fields. *Studia Psychologica*, 1972.
- [24] G. Kanizsa. *Organization in vision: essays on Gestalt perception*. NY, Praeger, 1979.
- [25] G. Kanizsa. Seeing and thinking. *Acta Psychologica*, 59(1):23–33, 1985.
- [26] G. Kanizsa. *Vedere et pensare*. Il Mulino, Bologna, 1991.
- [27] N. Kawai, T. Sato, and N. Yokoya. Image inpainting considering brightness change and spatial locality of textures and its evaluation. In *Advances in Image and Video Technology*, pages 271–282. 2009.
- [28] P. Kellman and T. Shipley. A theory of visual interpolation in object perception. *Cognitive psychology*, 23(2):141–221, 1991.
- [29] D. Knill, D. Kersten, and A. Yuille. Introduction: A bayesian formulation of visual perception. In D. Knill and W. Richards, editors, *Perception as Bayesian inference*. Cambridge University Press, 1996.

- [30] K. Koffka. *Principles of Gestalt psychology*. London: Routledge and Kegan Paul, 1935.
- [31] E. Leeuwenberg and P. A. van der Helm. *Structural Information Theory. The Simplicity of Visual Form*. Cambridge University Press, 2012.
- [32] T. Leung and J. Malik. Contour continuity in region based image segmentation. In *Computer Vision ECCV*, pages 544–559. Springer, 1998.
- [33] A. Mansfield, M. Prasad, C. Rother, T. Sharp, P. Kohli, and L. van Gool. Transforming image completion. In *Proceedings of BMVC*, pages 121.1–121.11, 2011.
- [34] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th ICCV*, volume 2, pages 416–423, July 2001.
- [35] S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Trans. on IP*, 11(2):68–76, 2002.
- [36] S. Masnou and J.-M. Morel. Level lines based disocclusion. In *Proc. of IEEE ICIP*, 1998.
- [37] A. Meijster, J. B. Roerdink, and W. H. Hesselink. A general algorithm for computing distance transforms in linear time. In *Mathematical Morphology and its applications to image and signal processing*, pages 331–340. Springer, 2002.
- [38] B. Merriman, J. Bence, and S. Osher. Diffusion generated motion by mean curvature. In *J.E. Taylor, editor, Computational Crystal Growers Workshop*, pages 73–83. American Mathematical Society, Providence, Rhode Island, 1992. Also available as UCLA CAM Report 92-18, April 1992.
- [39] A. Michotte, G. Thines, and G. Crabbé. Amodal completion of perceptual structures. *Michotte’s experimental phenomenology of perception*, pages 140–167, 1991.
- [40] L. Moravec and J. Beck. Amodal completion: Simplicity is not the explanation. *Bulletin of the Psychonomic Society*, 24(4):269–272, 1986.
- [41] D. Mumford. Elastica and computer vision. In *Algebraic Geometry and its Applications*, pages 491–506. Springer, New York, 1994.
- [42] M. Murray, D. Foxe, D. Javitt, and J. Foxe. Setting boundaries: brain dynamics of modal and amodal illusory shape completion in humans. *The Journal of neuroscience*, 24(31):6898–6903, 2004.
- [43] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, segmentation, and depth*, volume 662. Lecture notes in computer science, Springer, 1993.
- [44] I. R. Otero and M. Delbracio. Computing an Exact Gaussian Scale-Space. *Image Processing On Line*, 6:8–26, 2016.
- [45] D. Ringach and R. Shapley. Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision research*, 36(19):3037–3050, 1996.
- [46] N. Rubin. The role of junctions in surface completion and contour matching. *Perception*, 30(3):339–366, 2001.
- [47] S. R. S. Esedoglu and R. Tsai. Threshold dynamics for shape reconstruction and disocclusion. In *Proc of IEEE ICIP*, volume 2, pages 502–505, 2005.
- [48] A. B. Sekuler. Local and global minima in visual completion: effects of symmetry and orientation. *PERCEPTION-LONDON*, 23:529–529, 1994.

- [49] J. A. Sethian. Curvature and the evolution of fronts. *Communications in Mathematical Physics*, 101(4):487–499, 1985.
- [50] M. Singh and D. Hoffman. Completing visual contours: The relationship between relatability and minimizing inflections. *Perception & Psychophysics*, 61(5):943–951, 1999.
- [51] K. K. Thornber and L. R. Williams. Characterizing the distribution of completion shapes with corners using a mixture of random processes. *Pattern Recognition*, 33(4):543–553, 2000.
- [52] P. van der Helm. Bayesian confusions surrounding simplicity and likelihood in perceptual organization. *Acta psychologica*, 138(3):337–346, 2011.
- [53] R. Van Lier, E. Leeuwenberg, and P. Van der Helm. Multiple completions primed by occlusion patterns. *PERCEPTION-LONDON-*, 24:727–727, 1995.
- [54] R. van Lier, P. van der Helm, and E. Leeuwenberg. Integrating global and local aspects of visual occlusion. *Perception*, 23:883–903, 1994.
- [55] R. Van Lier, P. Van der Helm, and E. Leeuwenberg. Competing global and local completions in visual occlusion. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3):571, 1995.
- [56] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. Lsd: a line segment detector. *Image Processing On Line*, 2:35–55, 2012.
- [57] M. Wertheimer. *Untersuchungen zur Lehre von der Gestalt, II [Investigations in Gestalt Theory, II: Laws of organization in perceptual forms]*, pages 301–350. Psychologische Forschung, 1923.
- [58] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Transactions on PAMI*, 29(3):463–476, 2007.
- [59] L. R. Williams and D. W. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural computation*, 9(4):837–858, 1997.