

# An integrated Bayesian approach to shape representation and perceptual organization

Jacob Feldman, Manish Singh, Erica Briscoe, Vicky Froyen, Seha Kim, and John Wilder

**Abstract** We present a unified Bayesian approach to shape representation and related problems in perceptual organization, including part decomposition, shape similarity, figure/ground estimation, and 3D shape. The approach is based on the idea of estimating the skeletal structure most likely to have *generated* the observed shape via a process of stochastic “growth.” We survey the approach briefly and show how it can be extended in a principled way to solve a wide array of related problems.

## 1 Shape and perceptual organization

The visual representation of shape is a complex problem, requiring the reduction of an essentially infinite-dimensional object (the geometry of the shape) to a few perceptually meaningful dimensions. Human infants can recognize shape from line drawings without any prior experience [17], suggesting that the ability to abstract form from the bounding contour is innate. Much research in the study of shape has involved a quest for a set of shape descriptors that will allow just the right aspects of shape to be extracted—a representation that retains enough information to support recognition, shape similarity, and other key functions. Each of these techniques—geons [3], codons [37], medial axes [4], curvature extrema [18], Fourier descriptors [8], and so forth—has merits. Some have compelling mathematical motivations, while others (unfortunately not usually the same ones) have demonstrable agreement with human data. Still, broadly speaking, a complete computational characterization of human shape representation remains elusive.

---

Jacob Feldman, Manish Singh, Vicky Froyen, Seha Kim & John D. Wilder  
Rutgers University, New Brunswick, USA;  
e-mail: {jacob.feldman,manish.singh,vickyf,sehakim,jdwilder}@rutgers.edu

Erica Briscoe, Georgia Tech Research Institute; e-mail: erica.briscoe@gtri.gatech.edu

The approach we lay out below aims to address two inadequacies in the existing literature. First, many existing theories of shape lack a persuasive “theory of the computation,” in Marr’s influential phrase [34]—that is, an explanation of why, in principle, the proposed shape descriptors solve the shape problem better than alternatives. To provide such an account, one must adopt a particular definition of “the shape problem”—i.e., a model of what it is that we are actually trying to estimate when we describe a shape. Second, many shape theories have suffered from a lack of connection to other closely related problems in perceptual organization, including perceptual grouping and figure/ground. The shape literature in both psychology and computer science has generally focused on isolated shapes segregated from their backgrounds. But a great deal of evidence suggests that the problem of shape is, at least in the human visual system, intimately connected with the problems of figure/ground and perceptual organization more generally. The representation of a shape is in part determined by the factors that make it perceived as an integral, figural object in the first place, suggesting that shape and perceptual organization are intertwined.

In what follows we describe a framework that is both (a) principled, meaning that it stems from basic considerations of the nature of the shape inference, and (b) unified, in that it aims to approach a broad class of interrelated problems in a coherent way. We first briefly explain the principles of the Bayesian approach to shape representation, and then illustrate how it naturally gives rise to solutions to several related problems, including (i) shape similarity (ii) figure/ground, and (iii) 3D shape from line drawings.

## 2 Bayesian estimation of the shape skeleton

Skeletal or medial-axis representations were first introduced by Blum [4, 5]. Blum’s basic insight was that many aspects of contour shape are intuitively captured by a representation that extracts the *local symmetries* of the bounding contour. The medial axis transform (MAT), originally defined as the union of centers of inscribed circles, is highly suggestive of global shape structure, in that its branches often seem to correspond intuitively distinct shape parts such as limbs (and indeed Blum initially conceived it as a compact representation of animal morphology). Medial representations relate to many other problems in perceptual organization [23], and have both psychophysical correlates [25, 50] as well as known neural representations in brain areas V4 and IT [19, 26]. But as has long been recognized [5], the conventional MAT reflects global part structure very imperfectly; in particular, its branches do not reliably correspond to perceptually distinct shape parts. Many improvements on Blum’s original MAT have been developed (e.g. [21]), including some that represent mathematically deep generalizations of the “grassfire” procedure that underlies it [40]. But most contemporary medial axis models inherit the basic limitations of Blum’s approach, because (with some exceptions [54]) they share its essentially *determin-*

*istic* conception, which aims to define an information-preserving transformation of the shape, rather than an abstraction of the shape’s underlying structure.

In contrast, we view skeleton computation as a *probabilistic estimation* problem, the goal of which is to estimate the shape skeleton from which the shape is most likely to have been generated—that is, the skeleton that *best explains* the shape. Many natural shapes, especially biological ones, are effectively described as combinations of elongated parts [35]. We view the skeletal structure underlying such shapes as the “signal” which is combined with noisy local contour perturbations to yield the eventual shape. Specifically we conceive of the shape skeleton as the *generating source* of the contour, which then “extrudes” the shape via a partly stochastic process akin to growth (cf. [28]). We then adopt an inverse-probability framework, taking as our goal the recovery of the skeleton that gave rise to the observed shape. The problem then becomes a standard Bayesian inverse probability problem, with the goal being to estimate the skeleton with maximum posterior probability (called the MAP skeleton) as the best interpretation of the shape. The rest of the approach flows from this central conception: we define a model of the shape-generating process, and estimate the model.

## 2.1 Sketch of the theory

In our formalism, a shape  $\text{SHAPE} = \{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\}$  is a set of edges each of which is defined by a location  $x_i$  and a tangent vector  $t_i$ . A *skeleton*  $\text{SKEL} = \{A_1, A_2, \dots\}$  is a set of hierarchically connected axial curves, with a root axis, child axes, grandchildren, etc., branching off from it. Skeletons have a prior probability  $p(\text{SKEL})$  and generate shapes stochastically via the likelihood model  $p(\text{SHAPE}|\text{SKEL})$  explained below. Our computational goal is to find the best “explanation” of SHAPE by estimating the skeleton SKEL that is *most likely* to have generated it, i.e. that maximizes the product  $p(\text{SKEL})p(\text{SHAPE}|\text{SKEL})$  of prior and likelihood.

We begin by adopting a prior  $p(\text{SKEL})$ . As in all Bayesian approaches, the prior encodes our assumptions about which models (here, skeletons) are more and less likely to be encountered in the environment—assumptions that can then be easily modified to reflect different contexts or knowledge. In [15] we adopted a simple “vanilla” prior that assigns higher probability to simpler skeletons and lower probability to more complex ones, meaning ones with *more* axes or *more curved* axes (Fig. 1). This prior is a simple hierarchical extension of our prior for smooth contours, which has been validated in a number of empirical settings (see [12, 13, 14, 43, 44]). Specifically, we assume that each of the  $N$  component axes  $A_i$  contains a series of points  $A_i = \{a_{i,1}, a_{i,2}, \dots\}$ , which defines a sequence of turning angles  $\alpha_{i,j}$  (e.g.  $\alpha_{i,2}$  is the angle between the vector  $a_{i,3} - a_{i,2}$  and the vector  $a_{i,2} - a_{i,1}$ ) with the turning angles following a von Mises distribution  $p(\alpha_{i,j}) \propto \exp \beta \cos \alpha$  (the analog of the normal for circular variables, see [33]) and assumed independent. So the prior for axis  $A_i$  is the product of the probabilities of

its component turning angles,  $p(A_i) = \prod_j p(\alpha_{i,j})$ . Each axis is “born” with fixed probability  $p_A$ , leading to an overall prior  $p(\text{SKEL}) = p_A^N \prod_{i=1}^N \prod_j p(\alpha_{i,j})$ . This prior favors skeletons with fewer axes (low  $N$ ) and relatively straight axes (small  $\alpha$ s, see Fig. 1).

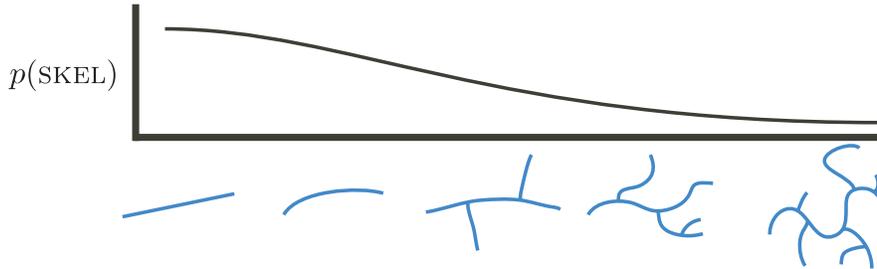


Fig. 1: “Vanilla” prior  $p(\text{SKEL})$  for skeletons, favoring skeletons with fewer, straighter axes (left) and penalizing those with more numerous and curvier branches (right).

The next component is the likelihood model  $p(\text{SHAPE}|\text{SKEL})$ , which quantifies how likely each shape is given a hypothesized skeleton. Our likelihood model expresses the idea of shape “growth:” contour points sprout laterally from each axial segment. The growth process is formalized via a set of random lateral vectors that sprout from both sides of a skeletal axis, referred to as “ribs” (Fig. 2). The ribs point in a stochastically chosen direction (we use a von Mises distribution, centered on perpendicular to the axis) and have a stochastically chosen length (we use a normal distribution, centered on an expected shape-part half-width whose value varies continuously over the length of the axis). The ribs thus represent *correspondences* between contour points and axial points that explain them—i.e., are interpreted as having generated them. This notion of “explanation” is central to the framework: the skeleton is understood as a hypothesis that explains the data, i.e. the observed contour points. We assume conditional independence of contour points given the skeleton, so the likelihood of the shape is simply the product of the likelihoods of all its component points,

$$p(\text{SHAPE}|\text{SKEL}) = \prod_{i=1}^n p((x_i, t_i)|\text{SKEL}). \quad (1)$$

The likelihood quantifies the degree of fit between a shape and a hypothetical skeleton that might explain it.

Finally the degree of belief in a given skeleton—that is, the degree to which the system ought to adopt that skeleton as an explanation for the given shape—is given by the posterior  $p(\text{SKEL}|\text{SHAPE})$ , which is proportional to the product of the prior and the likelihood,

$$p(\text{SKEL}|\text{SHAPE}) \propto p(\text{SKEL})p(\text{SHAPE}|\text{SKEL}). \quad (2)$$

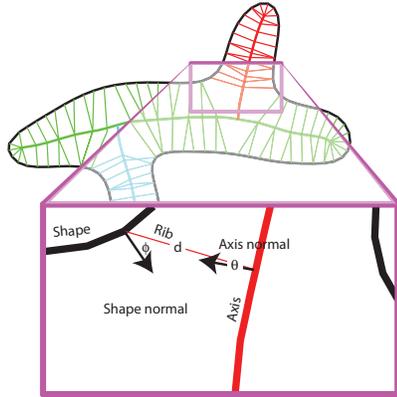


Fig. 2: Generative (likelihood) model  $p(\text{SHAPE}|\text{SKEL})$ , modeling a process of stochastic lateral growth. Random deviates (“ribs”) sprout bilaterally from each axis, terminating in edges that taken together constitute the shape contour. The sprouting direction  $\theta$  and edge orientation  $\phi$  are each von Mises distributed (respectively  $\theta \sim \exp[\beta_\theta \cos \theta]$  and  $\phi \sim \exp[\beta_\phi \cos \phi]$ ), and the rib length  $d$  is assumed Gaussian with a mean and variance estimated from the data ( $d \sim N(\bar{d}, \sigma_d^2)$ ).

To select a single best explanation of the shape, we estimate the skeleton with maximum posterior probability, referred to as the MAP skeleton (Fig. 3). The MAP skeleton represents the optimal skeletal interpretation of the shape, meaning that—given the assumptions captured by the prior and likelihood model—it identifies the single skeleton most likely to have generated the shape. Critically, the choice of the MAP involves a tradeoff between the prior, which favors simple skeletons, and the likelihood, which favors more complex skeletons that can fit the shape better. The axes that are included in the MAP skeleton, i.e. those whose contribution to the likelihood outweighs their penalization in the prior, represent statistically meaningful parts of the shape. That is, each distinct axis in the MAP skeleton represents what the procedure interprets as a distinct part of the shape (depicted with different colors in the figures).

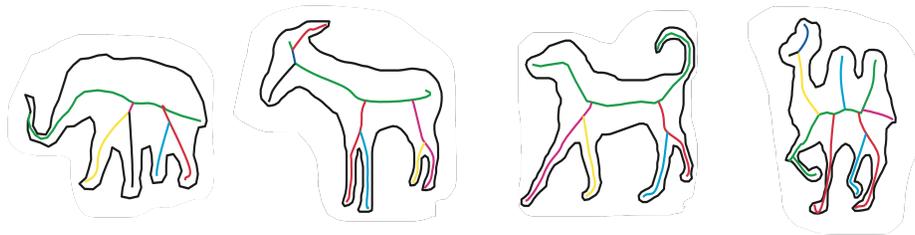


Fig. 3: Examples of the MAP skeleton for simple animal shapes (ribs not shown).

### 3 Applications and extensions

We next describe preliminary work extending the basic shape theory to key shape problems. Each of these applications grows directly out of the basic theory, illustrating the fecundity of the approach.

#### 3.1 Decomposing shapes into parts

Notwithstanding the success of appearance-based recognition models (e.g. [30]), there is substantial evidence that human object recognition uses structural representations based on combinations of shape parts [3, 1]. But though many factors are known to influence the decomposition of shapes into parts [9, 41, 45, 46], we still lack a comprehensive account of part decomposition. A simple and principled account of part decomposition is directly entailed by the Bayesian approach to shape representation, encompassing several well-known part-decomposition rules as side-effects. The MAP skeleton implies a part decomposition, because the shape contour naturally decomposes into regions that are “owned” by distinct component axes. For example in Fig. 4 contour sections indicated with different-colored ribs are owned (explained) by distinct axes. As explained above, the axial makeup of the winning skeleton reflects a Bayesian decision about which branches benefit the posterior; the MAP includes only those axes whose contribution to the likelihood outweighs their penalization by the prior.

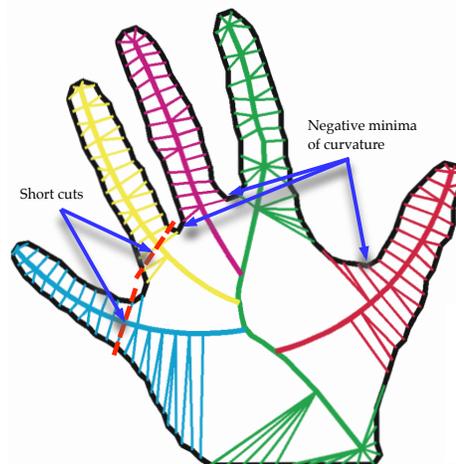


Fig. 4: The estimated skeleton divides the shape into sections “owned” by distinct axes (color coding). The entailed part boundaries tend to correspond to negative curvature minima and correspond to short part cuts, suggesting that skeleton estimation can subsume these principles.

The skeleton-based decomposition of shapes into component parts concurs with, and arguably subsumes, certain rules of part decomposition obeyed by the human vi-

sual system [42]. For example transitions between axial ownership (e.g. the boundary between red and green ribs in the hand in Fig. 4) tend to occur at within deeply concave sections of the contour, near negative curvature extrema, in accordance with the well-known minima rule [18] (even though curvature plays no overt role in skeleton estimation). In this sense skeleton-based shape decomposition explains both the successes of the minima rule, i.e. the fact that part boundaries tend to occur near minima, and also its failures, e.g. part boundaries that occur where there are no curvature minima and curvature minima that are not perceived as part boundaries (see [45]). Similarly, regions of common axial ownership tend to be relatively convex, subsuming the part convexity principle [38], and part cuts tend to be relatively short, subsuming the short-cut rule [46]. All of these known characteristics of intuitive part decomposition arise naturally from skeletal estimation, rather requiring additional assumptions, leading to a more principled and unified account than was previously possible.

### 3.2 *Tuning the shape model to the environment*

The results given above are based on a very simple “vanilla” skeleton prior and likelihood, but both the prior and likelihood model can be modified to accommodate more realistic models of natural shape classes. Some shape classes tend systematically to have more axial branches, or fewer; or more curved branches, or straighter; or smoother contours (smaller variance in rib lengths), or rougher; and so forth, all suggesting modifications to the generative model. To illustrate the approach, we estimated the skeletal parameters of the shapes in several large databases of natural shapes [52], including one of animals and one of leaves (Fig. 5a). Tabulations of skeletal parameters show substantial differences between the two shape classes. For example the distribution of number of branches show not only different means but also qualitatively different distributional forms (Fig. 5b): Gaussian for animals (with a mean near 5, about the number of intuitively distinct parts in the typical animal body plan) but exponential for leaves (suggesting a recursively branching process). Such differences show how the skeletal generative model can be “tuned” to natural shape classes.

We have also found that human subjects’ classification of novel shapes can be predicted from their skeletal representations. We showed subjects composite shapes created by morphing animals and leaves in controlled proportions, and asked them to classify them into animal or leaf categories. (There is no correct answer since the shapes are actually novel composites.) Their responses closely match Bayesian classifications based on skeletal parameters, but disregard or even contradict predictions based on more conventional shape parameters such as aspect ratio or compactness. This suggests that human observers do indeed extract skeletal parameters and use category-specific probabilistic knowledge to classify novel shapes.

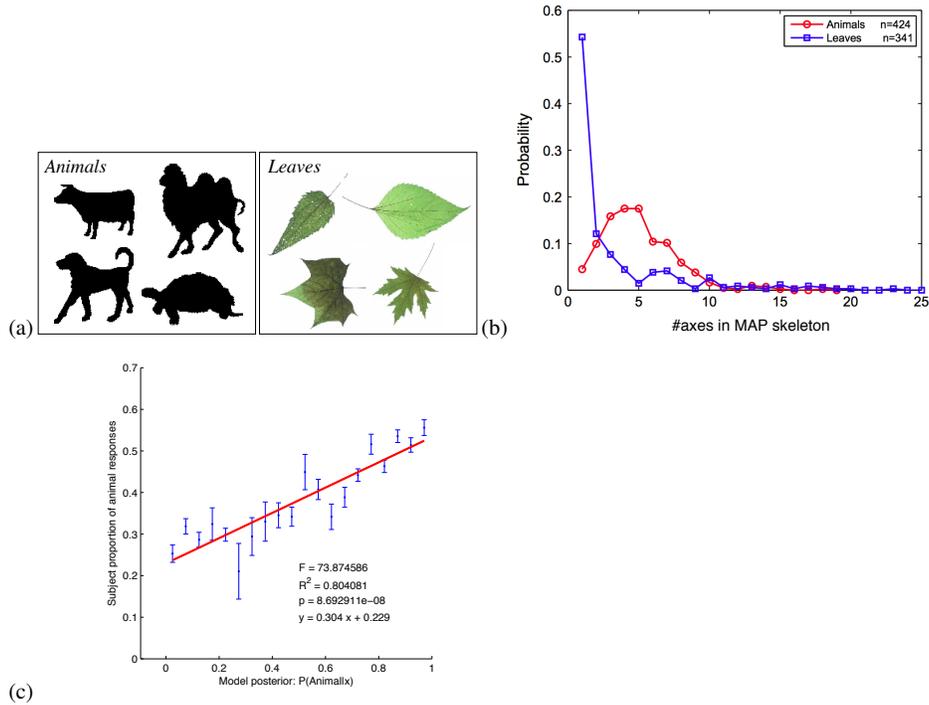


Fig. 5: (a) Samples of the shapes from which skeletal statistics were drawn. (b) Animals and leaves show systematic statistical differences, such as in the distribution of the number of axial branches. (c) A classifier based on these differences predicts human subjects’ classifications of morphed (composite) shapes.

### 3.3 Shape similarity

An essential application of shape representation is the evaluation of shape similarity. Measures of shape matching abound in the computational literature, where they form the basis of shape recognition [10], including some with properties suggestive of human intuitions, like robustness to part articulation [29]. But though similarity based on skeletal representations has been found effective [39], few if any algorithmic similarity measures have been validated against human similarity judgments.

Our skeletal representation provides a natural measure of shape similarity [6]. Because each skeletal estimate represents a “model” of the observed shape, it is natural to ask how well this model explains *another* shape. Specifically, given two shapes  $\text{SHAPE}_1$  and  $\text{SHAPE}_2$ , with associated skeletal estimates  $\text{SKEL}_1$  and  $\text{SKEL}_2$ , we define the similarity of  $\text{SHAPE}_1$  to  $\text{SHAPE}_2$  as the likelihood

$$\text{sim}(\text{SHAPE}_1, \text{SHAPE}_2) = p(\text{SHAPE}_1 | \text{SKEL}_2), \quad (3)$$

that is, the probability that shape  $SHAPE_1$  would “grow” from skeleton  $SKEL_2$ . This gives an asymmetric assessment of the first shape’s fit to the second shape’s representation (potentially accommodating the asymmetric similarity judgments that are well-known in the psychological literature). A symmetric similarity measure can be defined by taking the average

$$d(SHAPE_1, SHAPE_2) = \frac{1}{2} [\text{sim}(SHAPE_1, SHAPE_2) + \text{sim}(SHAPE_2, SHAPE_1)]. \quad (4)$$

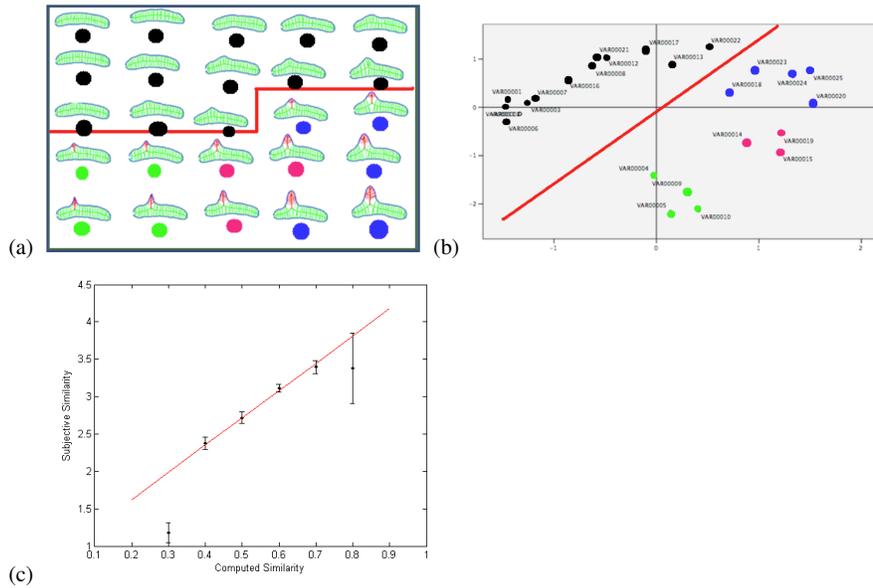


Fig. 6: Shape similarity model. (a) Shapes tested, showing MAP skeletons including ribs and entailed part decomposition. The red border divides the shapes estimated to have 1 part from those estimated to have 2 parts. (b) Results of multidimensional scaling based on human similarity ratings; the red border here corresponds to the border in (a), showing the exaggerated psychological distance between 1- and 2-part shapes. (c) Plot showing approximately linear relationship between human and computed similarity ratings.

In [6] we tested the psychological validity of this shape similarity measure by asking subjects to rate similarity of all pairs drawn from several collections of shapes. For example Fig. 6a shows a set of shapes generated from a 2-axis skeleton, in which the length of the secondary axis was modulated from very small to large. The red border shows the boundary between shapes whose MAP skeletons contain one axis (that is, in which the second part was too small to be included in the estimated skeleton) and those that contain two distinct axes. Fig. 6b shows the similarity space of the same shapes, computed via multidimensional scaling from subjects’ similarity judgments. The exaggerated division (marked in red) between

shapes perceived to have one part and those perceived to have two parts is plainly visible, and as can be seen in the figure corresponds exactly to the division between 1-axis and 2-axis MAP skeletons. Finally Fig. 6c shows the very close linear relationship between judged similarity and similarity computed via Eq. 4. Experiments with several other classes of shapes also show close matches between computed and perceived shape similarity [6].

### 3.4 Figure and ground

Figure/ground (f/g) assignment is intrinsically intertwined with the representation of shape, in part because figural polarity (border ownership) determines the sign of curvature, which plays a central role in shape representation [18]. Indeed because figural regions “own” the border [2, 11], only figural regions’ shapes are overtly represented [36], with ground regions perceived as extending indefinitely behind. F/g assignment is known to be influenced by a number of shape factors, including region size [24] convexity [20] and symmetry [20]. But nonetheless theoretical connections between shape and f/g remain largely unexplored.

Our approach to contour interpretation can easily be extended to encompass geometric factors on f/g assignment in a simple but principled way. Above, we have assumed that f/g assignment along the contour is known, and that shapes have to be explained from their interiors—that is, by skeletons in their interiors. Instead, we now (a) relax the assumption that border ownership is known, and instead treat it as a parameter to be estimated; and (b) relax the assumption that there is a single shape to be explained by a single skeleton, and instead attempt to explain the entire set of image contours by an ensemble of skeletons. More specifically, we no longer assume that the sign of each tangent vector  $t_i$  (which defines which side of each contour is figural) is given as part of the data, and instead treat it as an unknown parameter to be estimated. The formal problem of image explanation now reduces to the estimation of the ensemble of skeletons that, collectively, best explain the observed image structure. That is, we seek the set of skeletons that best explains all the observed edges, with each edge being explained from *whichever side* provides the best overall posterior, which determines its perceived border ownership.

In this expanded view of the problem, the MAP interpretation assigns (skeletal models) conjointly with f/g assignments over the entire ensemble of edges. The winning interpretation explains as much the contour as possible “from within”—that is, with each contour owned by a skeleton in what is perceived as its interior—while also maximizing everything else that the Bayesian model maximizes, such as the simplicity of the skeletons and the fit between the skeletons and the contours.

Because figural surfaces are perceived as closer, the induced figural assignment induces depth differences among skeletal axes: the axis that “wins” a given edge is interpreted as closer. This in turn allows the 3D relations among objects in the scene to be estimated. More specifically, each contour point  $(x, t)$  can be explained by skeletons on either side of it, and whichever skeleton assigns it a higher poste-

rior will be interpreted as “owning” the point, thus determining the direction of  $t$ , the polarity of local  $f/g$ , and the relative (qualitative) depth (Fig. 7a). Recall that the direction of the normal at the contour point influences the posterior in part because the likelihood function penalizes contour normals that point “away” from the generating skeleton (see Fig. 2). Fig. 7b shows results of a Bayesian belief network that implements a version of this computation [16]. The belief network estimates border ownership at each contour point, propagating the  $f/g$  estimate along each contour in a manner similar to previous  $f/g$  belief networks [51], but here including the skeleton-based likelihood function as a determinant of  $f/g$  status. As can be seen in the illustration, the procedure assigns border ownership to the perceived interiors of both overlapping shapes, critically including assigning the common boundary to the region human observers judge to be in front.

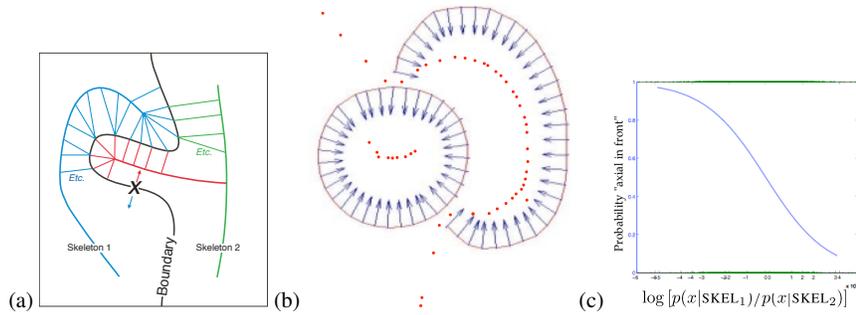


Fig. 7: (a) A contour point  $x$  can be “explained” by the skeleton on one side or the other, determining its border ownership. (b) Belief estimation of border ownership from skeletons, showing the medial structure present (on both sides of all contours) and estimated border ownership (arrows point towards figure). (c) Human judgments of figural status depend on the log posterior ratio (relative axiality) of the skeletons on either side of a boundary, favoring the side with the stronger posterior.

An empirical prediction derived directly from this framework is that more “axial” regions, that is, regions with stronger skeletal posteriors, are more likely to be perceived as figural. We tested this by constructing displays in which a symmetric region abutted a more “axial” one, and asked subjects which side appeared to own their common boundary [22], while manipulating the shape of the axial side so as to modulate the skeletal posterior. In the Bayesian model, ownership of a point  $(x, t)$  along the common boundary should follow the posterior ratio  $p((x, t)|SKEL_1)/p((x, t)|SKEL_2)$ , where  $SKEL_1$  and  $SKEL_2$  are MAP skeletons on respectively the axial and symmetric sides. Fig. 7c gives a representative plot showing the observed decrease in  $f/g$  responses as a function of (log) posterior ratio, confirming the basic claim that axiality under the Bayesian model tends to “draw” figural status.

### 3.5 3D shape

Our approach to skeleton estimation, like virtually all medial approaches, is based on two-dimensional silhouettes, which do not generally give rise to strong 3D interpretations. But much richer 3D interpretations arise from images that include interior contours, including T-junctions stemming from self-occlusions. Even without texture, shading, or other surface cues, human subjects can interpret 3D shape from such line drawings about as well as from natural images [7]. Yet interpretation of such figures remains a virtually unsolved problem. The extensive early literature on line and junction labeling [31, 32, 49] largely failed to solve it due to reliance on hard-and-fast junction classification rules. Our framework replaces these deterministic rules with a probabilistic inference in which the goal is to estimate the 3D skeleton that best explains the ensemble of contours and junctions in the image.

The first step is to extend the generative (likelihood) model to 3D. A direct 3D generalization of the conventional MAT [27] results in a complex and psychologically implausible combination of space curves and 2D medial “scaffolds.” In contrast, the skeletal generative model generalizes to 3D in an intuitive way. The key idea is simply to assume that the ribs (random deviates), rather than being generated laterally on both sides of the skeleton as in the 2D model, are instead generated in all directions in the plane orthogonal to the skeleton, thus “inflating” the skeleton into a 3D shape (Fig. 8a,b) (cf. [48]). The inverse problem is to estimate the 3D skeleton that is most likely, when inflated, to yield the observed contour when projected, including both outer silhouette, internal contours, and other contour features such as T-junctions. The resulting estimate can be substantially non-planar (Fig. 8c). A suitable estimation procedure for this model is of course a difficult problem, but in principle one that can be solved by conventional Bayesian techniques.

## 4 Discussion and conclusion

We have described a principled probabilistic conception of shape representation, which provides natural approaches to part decomposition, shape similarity, and figure/ground, and can be extended in a conceptually simple way to 3D. The main idea is to view contour elements as data generated by a skeletal growth process, and then estimate the structure of the skeleton. The best representation of a shape is the skeleton that best explains it; the similarity of two shapes is the degree to which one shape’s skeletal interpretation explains the other shape; and the best interpretation of multiple shapes is the collection of skeletons that best explains the ensemble of contours, thus inducing estimates of *f/g* and depth relations.

Many aspects of our framework are present in other approaches, including stochastic estimation of skeletal structure [47, 53], belief propagation for *f/g* [51], and inflation of 2D skeletal representations into 3D shape [48]. But the main attraction of our approach is its simplicity, comprehensiveness, and coherence: all the applications derive from the central conception of shape as a rational inference

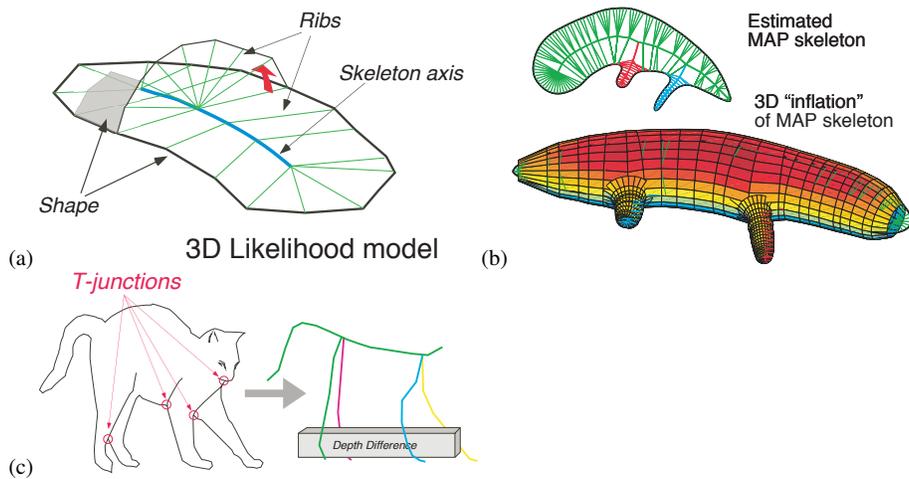


Fig. 8: (a) 3D likelihood model, yielding (b) inflation of the skeleton into a 3D shape. (c) When T-junctions and internal contours are included, the estimated skeleton can be non-planar.

problem. Broadly speaking, the aim is to make some assumptions about shape-generating processes in the environment; express these assumptions as a probability model; and estimate the model. As mentioned, the probability model can then be tuned to natural shape statistics, used to model shape similarity, extended to multiple shapes in a way that yields  $f/g$  estimates and depth relations, and easily extended to 3D. None of these extensions require elaborate new hacks, nor indeed any change to the basic principles. The psychological literature attests a wealth of connections among these different aspects of perceptual organization, and we would argue that our approach integrates them in a way that properly respects their interconnections.

It is important to understand that our approach does not intrinsically require mediality; axial forms are simply a reasonable model for many natural shapes. For others, alternative (non-medial) generative models might be adopted without changing the essentials of the approach. All the extensions we have presented derive from the central probabilistic estimation problem, not from specific aspects of medial geometry or local symmetry. The ultimate goal of this work is thus not to deepen our understanding of medial representations specifically, but rather to “probabilize” shape and related problems of perceptual organization, thus unifying them with the growing literature on probabilistic and Bayesian approaches to visual perception.

## References

1. E. Barenholtz and M. J. Tarr. Visual judgment of similarity across shape transformations: evidence for a compositional model of articulated objects. *Acta Psychologica*, 128(2):331–338, 2008.

2. G. Baylis and J. Driver. Obligatory edge assignment in vision: the role of figure and part segmentation in symmetry detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1323–1342, 1995.
3. I. Biederman. Recognition by components: a theory of human image understanding. *Psychological Review*, 94:115–147, 1987.
4. H. Blum. Biological shape and visual science (part i). *Journal of Theoretical Biology*, 38:205–287, 1973.
5. H. Blum and R. N. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, 10:167–180, 1978.
6. E. Briscoe. *Shape skeletons and shape similarity*. PhD thesis, Rutgers University, 2008.
7. F. Cole, K. Sanik, A. F. D. DeCarlo, T. Funkhouser, S. Rusinkiewicz, and M. Singh. How well do line drawings depict shape? In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, volume 28, 2009.
8. J. M. Cortese and B. P. Dyre. Perceptual similarity of shapes generated from fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1):133–143, 1996.
9. J. de Winter and J. Wagemans. Segmentation of object outlines into parts: A large-scale integrative study. *Cognition*, 99(3):275–325, 2006.
10. F. Demirci, A. Shokoufandeh, Y. Keselman, L. Bretzner, and S. Dickinson. Object recognition as many-to-many feature matching. *International Journal of Computer Vision*, 69(2):203–222, 2006.
11. J. Driver and G. C. Baylis. Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognitive Psychology*, 31:248–206, 1996.
12. J. Feldman. Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, 37(20):2835–2848, 1997.
13. J. Feldman. Bayesian contour integration. *Perception & Psychophysics*, 63(7):1171–1182, 2001.
14. J. Feldman and M. Singh. Information along contours and object boundaries. *Psychological Review*, 112(1):243–252, 2005.
15. J. Feldman and M. Singh. Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Science*, 103(47):18014–18019, 2006.
16. V. Froyen, J. Feldman, and M. Singh. A Bayesian framework for figure-ground interpretation. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 631–639. 2010.
17. J. Hochberg and V. Brooks. Pictorial recognition as an unlearned ability: A study of one child’s performance. *American Journal of Psychology*, 75(4):624–628, 1962.
18. D. D. Hoffman and W. A. Richards. Parts of recognition. *Cognition*, 18:65–96, 1984.
19. C. C. Hung, E. T. Carlson, and C. E. Connor. Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, 74(6):1099–1113, 2012.
20. G. Kanizsa and W. Gerbino. Convexity and symmetry in figure-ground organization. In M. Henle, editor, *Vision and artifact*. Springer, New York, 1976.
21. R. A. Katz and S. M. Pizer. Untangling the Blum medial axis transform. *International Journal of Computer Vision*, 55(2/3):139–153, 2003.
22. S. Kim. The influence of axiality on figure/ground assignment. Master’s thesis, Rutgers University, 2011.
23. B. B. Kimia. One the role of medial geometry in human vision. *Journal of Physiology Paris*, 97:155–190, 2003.
24. K. Koffka. *Principles of Gestalt Psychology*. Harcourt, New York, 1935.
25. I. Kovács, A. Fehér, and B. Julesz. Medial-point description of shape: a representation for action coding and its psychophysical correlates. *Vision Research*, 38:2323–2333, 1970.
26. M. D. Lescroart and I. Biederman. Cortical representation of medial axis structure. *Cerebral Cortex*, 2012.
27. F. F. Leymarie and B. B. Kimia. The medial scaffold of 3d unorganised point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):313–330, 2007.

28. M. Leyton. Inferring causal history from shape. *Cognitive Science*, 13:357–387, 1989.
29. H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007.
30. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
31. A. Mackworth. Interpreting pictures of polyhedral scenes. *Artificial Intelligence*, 4:121–137, 1973.
32. J. Malik. Interpreting line drawings of curved objects. *International Journal of Computer Vision*, 1:73–103, 1987.
33. K. V. Mardia. *Statistics of directional data*. Academic Press, London, 1972.
34. D. Marr. *Vision: a computational investigation into the human representation and processing of visual information*. Freeman, San Francisco, 1982.
35. D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B.*, 200:269–294, 1978.
36. S. Palmer, J. Davis, R. Nelson, and I. Rock. Figure-ground effects on shape memory for objects versus holes. *Perception*, 37(10):1569–1586, 2008.
37. W. Richards, B. Dawson, and D. Whittington. Encoding contour shape by curvature extrema. In *Natural computation*. M.I.T. Press, Cambridge, MA, 1988.
38. P. L. Rosin. Shape partitioning by convexity. *IEEE Trans. Systems, Man, and Cybernetics, part A*, 30:202–210, 2000.
39. T. B. Sebastian and B. B. Kimia. Curves vs. skeletons in object recognition. *Signal Processing*, 85:247–263, 2005.
40. K. Siddiqi, A. Shokoufandeh, S. Dickinson, and S. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, 30:1–24, 1999.
41. K. Siddiqi, K. J. Tresness, and B. B. Kimia. Parts of visual form: psychophysical aspects. *Perception*, 25:399–424, 1996.
42. M. Singh, V. Froyen, and J. Feldman. Unifying parts and skeletons: a Bayesian approach to part decomposition. Forthcoming., 2012.
43. M. Singh and J. M. Fulvio. Visual extrapolation of contour geometry. *PNAS*, 102(3):939–944, 2005.
44. M. Singh and J. M. Fulvio. Bayesian contour extrapolation: geometric determinates of good continuation. *Vision Research*, 47:783–798, 2007.
45. M. Singh and D. D. Hoffman. Part-based representations of visual shape and implications for visual cognition. In T. Shipley and P. Kellman, editors, *From fragments to objects: segmentation and grouping in vision, Advances in Psychology, Vol. 130*, pages 401–459. Elsevier, New York, 2001.
46. M. Singh, G. D. Seyranian, and D. D. Hoffman. Parsing silhouettes: the short-cut rule. *Perception & Psychophysics*, 61(4):636–660, 1999.
47. A. Telea, C. Sminchisescu, and S. Dickinson. Optimal inference for hierarchical skeleton abstraction. In *Proceedings IEEE International Conference on Pattern Recognition*, Cambridge, MA, 2004.
48. N. R. Twarog, M. F. Tappen, and E. H. Adelson. Playing with puffball: simple scale-invariant inflation for use in vision and graphics. In *Proceedings of the ACM Symposium on Applied Perception*, pages 47–54, 2012.
49. D. Waltz. Understanding line drawings of scenes with shadows. In P. H. Winston, editor, *The Psychology of Computer Vision*, pages 19–91. 1975.
50. X. Wang and C. A. Burbeck. Scaled medial axis representation: evidence from position discrimination task. *Vision research*, 38(13):1947–1959, 1998.
51. Y. Weiss. Interpreting images by propagating Bayesian beliefs. *Adv. in Neural Information Processing Systems*, pages 908–915, 1997.
52. J. Wilder, J. Feldman, and M. Singh. Superordinate shape classification using natural shape statistics. *Cognition*, 119:325–340, 2011.
53. S.-C. Zhu. Embedding gestalt laws in Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:1170–1187, 1999.
54. S.-C. Zhu. Stochastic jump-diffusion process for computing medial axes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1158–1169, 1999.