Chapter 4
An Integrated Bayesian Approach to Shape Representation and Perceptual Organization

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

4.1 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 agree-

J. Feldman (✉) · M. Singh · V. Froyen · S. Kim · J. Wilder
Department of Psychology, Rutgers University, New Brunswick, USA
e-mail: jacob.feldman@rutgers.edu

M. Singh
e-mail: manish.singh@rutgers.edu

V. Froyen
e-mail: vickyf@rutgers.edu

S. Kim
e-mail: sehakim@rutgers.edu

J. Wilder
e-mail: jdwilder@rutgers.edu

E. Briscoe
Aerospace, Transportation and Advanced Systems Laboratory, Georgia Tech Research Institute, Atlanta, GA, USA
e-mail: erica.briscoe@gtri.gatech.edu

S. Dickinson, Z. Pizlo (eds.), *Shape Perception in Human and Computer Vision*,
Advances in Computer Vision and Pattern Recognition,
DOI 10.1007/978-1-4471-5195-1_4,

Book ID: 302364_1_En, Date: 2013-05-21, Proof No: 1, UNCORRECTED PROOF

ment with human data. Still, broadly speaking, a complete computational characterization of human shape representation remains elusive.

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.

4.2 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 [53]) they share its essentially *deterministic* 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.

4.2.1 Sketch of the Theory

In our formalism, a shape $\text{SHAPE} = \{(x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n)\}$ is a set of edges each of which is defined by a location x_i and a tangent vector t_i. A *skeleton* $\text{SKEL} = \{A_1, A_2, \ldots\}$ 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. 4.1). This prior is a simple hierarchical extension of our prior for smooth contours, which has been validated in a number of empirical settings (see [12–14, 43, 44]). Specifically, we assume that each of the N component axes A_i contains a series of points $A_i = \{a_{i,1}, a_{i,2}, \ldots\}$, which defines a sequence of turning angles $\alpha_{i,j}$ (e.g., $\alpha_{1,2}$ 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.

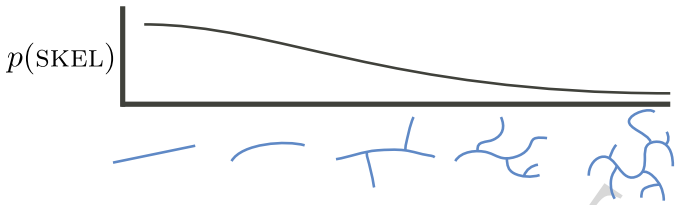

Fig. 4.1 "Vanilla" prior $p(\text{SKEL})$ for skeletons, favoring skeletons with fewer, straighter axes (*left*) and penalizing those with more numerous and curvier branches (*right*)

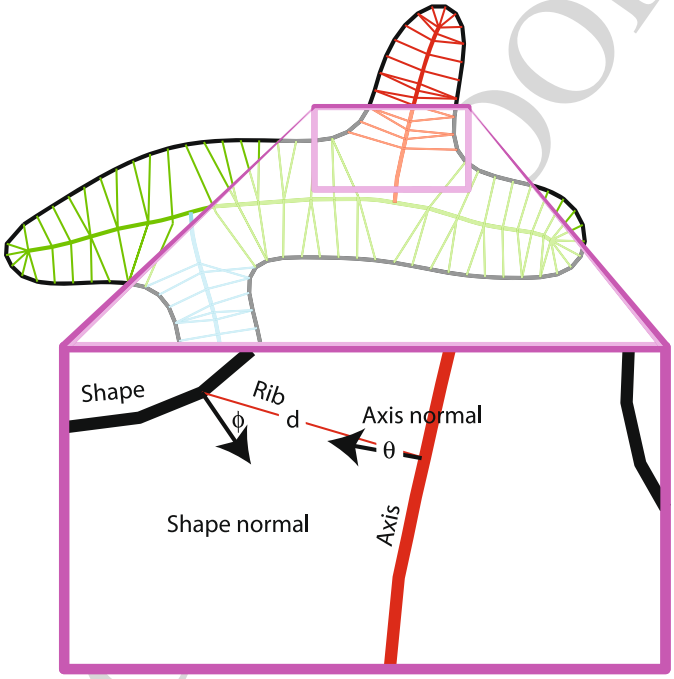

Fig. 4.2 Generative (likelihood) model $p(\text{SHAPE}|\text{SKEL})$, modeling a process of stochastic lateral growth. Random deviates ("ribs") sprout bilaterally from each axis, terminating in edges that taken together constitute the shape contour. The sprouting direction θ and edge orientation ϕ are 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 ($d \sim N(\hat{d}, \sigma_d^2)$)

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 αs, see Fig. 4.1).

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. 4.2). 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

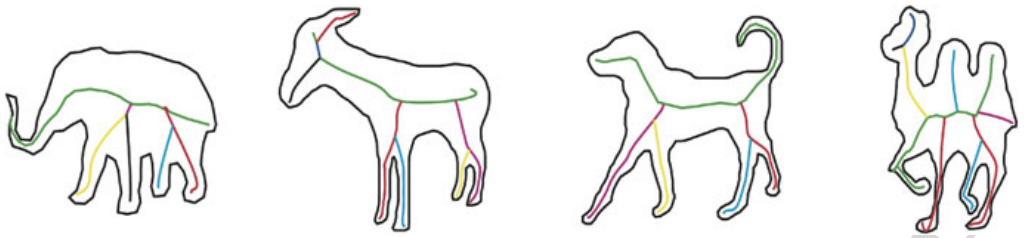

Fig. 4.3 Examples of the MAP skeleton for simple animal shapes (ribs not shown)

skeleton, so the likelihood of the shape is simply the product of the likelihoods of all its component points,

$$p(\text{SHAPE}|\text{SKEL}) = \prod_{i=1}^{n} p\big((x_i, t_i)|\text{SKEL}\big). \tag{4.1}$$

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,

$$p(\text{SKEL}|\text{SHAPE}) \propto p(\text{SKEL})p(\text{SHAPE}|\text{SKEL}). \tag{4.2}$$

To select a single best explanation of the shape, we estimate the skeleton with maximum posterior probability, referred to as the MAP skeleton (Fig. 4.3). 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).

4.3 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.

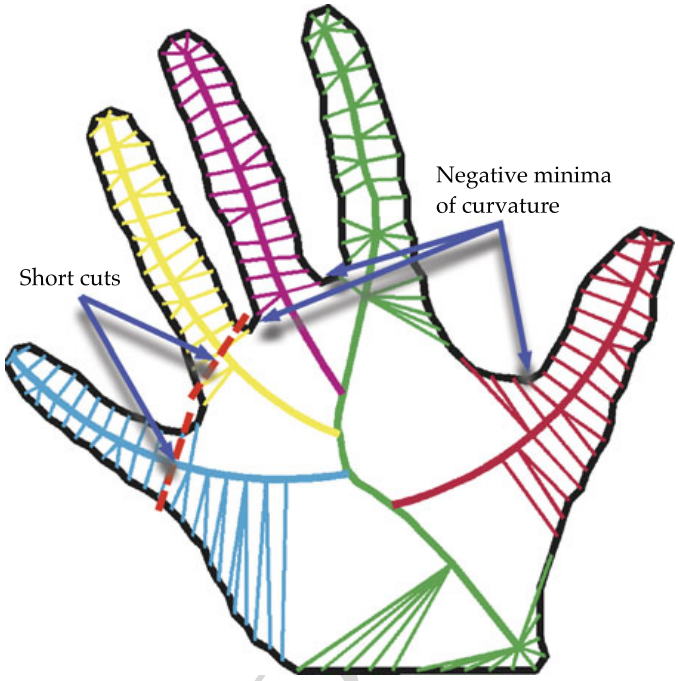

Fig. 4.4 The estimated skeleton divides the shape into sections "owned" by distinct axes (*color coding*). The entailed part boundaries tend to correspond to negative curvature minima and correspond to short part cuts, suggesting that skeleton estimation can subsume these principles

4.3.1 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 [1, 3]. 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. 4.4 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.

The skeleton-based decomposition of shapes into component parts concurs with, and arguably subsumes, certain rules of part decomposition obeyed by the human visual system [42]. For example, transitions between axial ownership (e.g., the boundary between red and green ribs in the hand in Fig. 4.4) 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, for example, 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.

4.3.2 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. 4.5a). 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. 4.5b): 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.

4.3.3 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

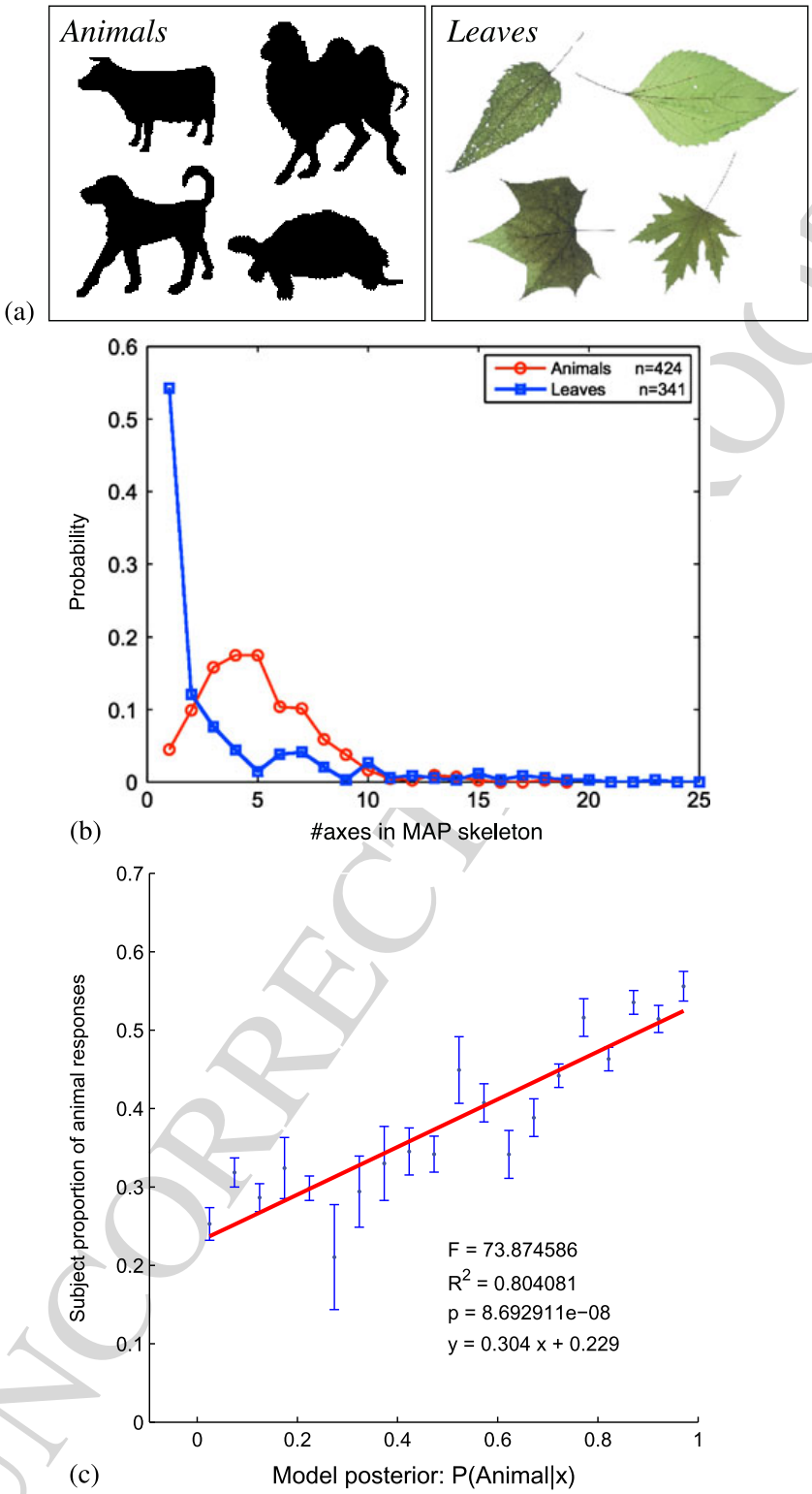

Fig. 4.5 (**a**) Samples of the shapes from which skeletal statistics were drawn. (**b**) Animals and leaves show systematic statistical differences, such as in the distribution of the number of axial branches. (**c**) A classifier based on these differences predicts human subjects' classifications of morphed (composite) shapes

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 SHAPE_1 and SHAPE_2, with associated skeletal estimates SKEL_1 and SKEL_2, we define the similarity of SHAPE_1 to SHAPE_2 as the likelihood

$$\text{sim}(\text{SHAPE}_1, \text{SHAPE}_2) = p(\text{SHAPE}_1|\text{SKEL}_2), \tag{4.3}$$

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

$$d(\text{SHAPE}_1, \text{SHAPE}_2) = \frac{1}{2}\big[\text{sim}(\text{SHAPE}_1, \text{SHAPE}_2) + \text{sim}(\text{SHAPE}_2, \text{SHAPE}_1)\big]. \tag{4.4}$$

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. 4.6a 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. Figure 4.6b 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. 4.6c shows the very close linear relationship between judged similarity and similarity computed via Eq. (4.4). Experiments with several other classes of shapes also show close matches between computed and perceived shape similarity [6].

4.3.4 *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.

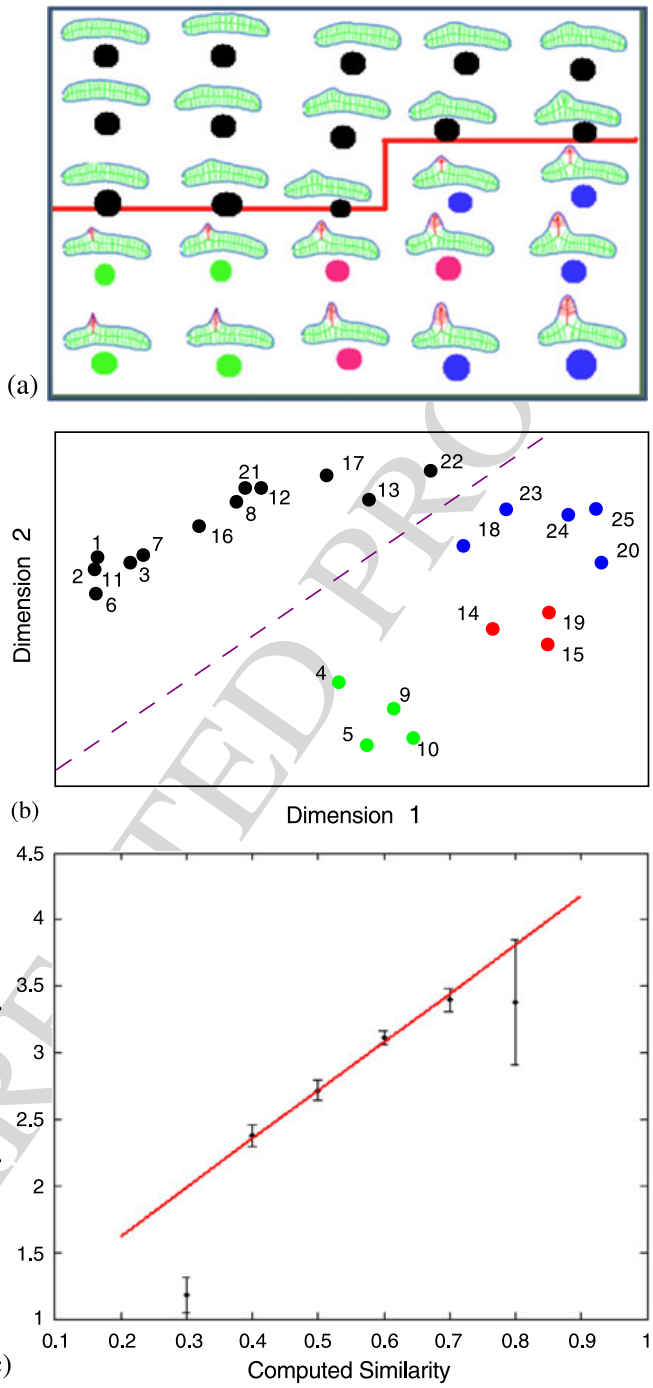

Fig. 4.6 Shape similarity model. (**a**) Shapes tested, showing MAP skeletons including ribs and entailed part decomposition. The *red border* divides the shapes estimated to have 1 part from those estimated to have 2 parts. (**b**) Results of multidimensional scaling based on human similarity ratings; 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

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

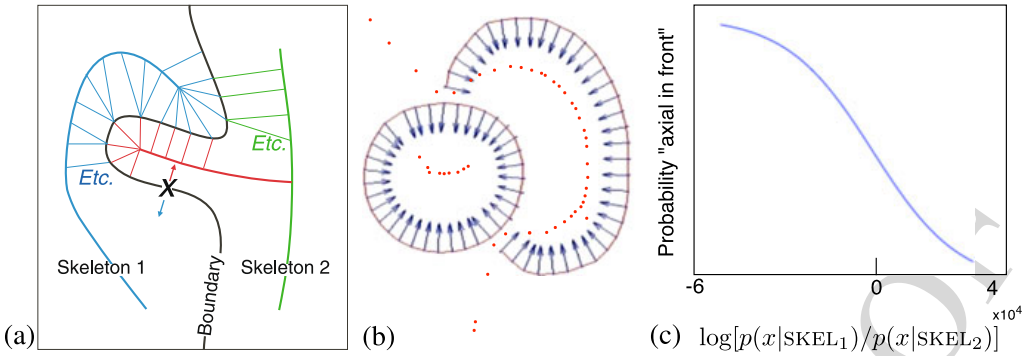

Fig. 4.7 (**a**) A contour point x can be "explained" by the skeleton on one side or the other, determining its border ownership. (**b**) Belief estimation of border ownership from skeletons, showing the medial structure present (on both sides of all contours) and estimated border ownership (*arrows* 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

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 posterior 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. 4.7a). 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. 4.2). Figure 4.7b 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.

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 $p((x,t)|\text{SKEL}_1)/p((x,t)|\text{SKEL}_2)$, where SKEL_1 and SKEL_2 are MAP skeletons on respectively the axial and symmetric sides. Figure 4.7c 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.

4.3.5 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. 4.8a, 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. 4.8c). 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.

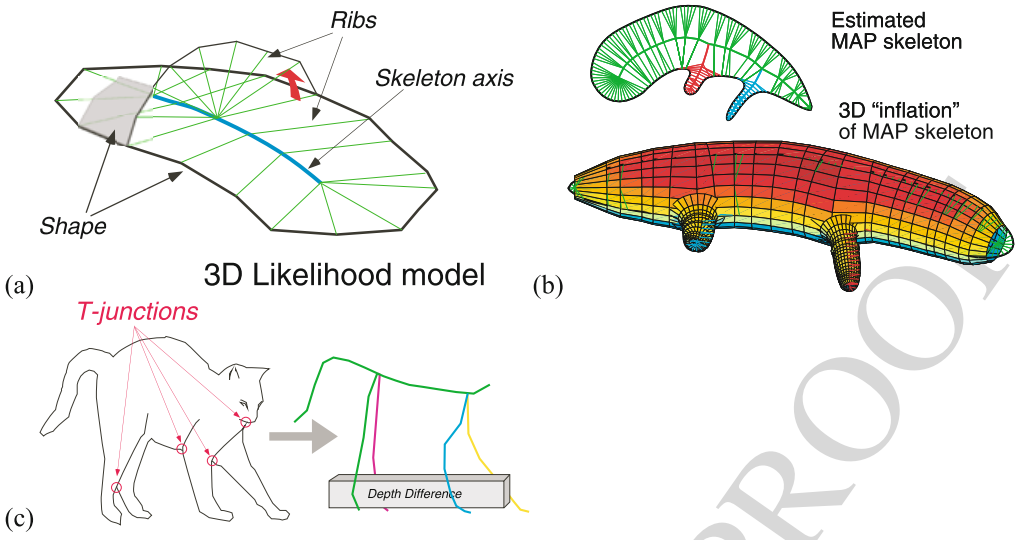

Fig. 4.8 (**a**) 3D likelihood model, yielding (**b**) inflation of the skeleton into a 3D shape. (**c**) When T-junctions and internal contours are included, the estimated skeleton can be non-planar

4.4 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 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 oth-

ers, 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. Barenholtz E, Tarr MJ (2008) Visual judgment of similarity across shape transformations: evidence for a compositional model of articulated objects. Acta Psychol 128(2):331–338
2. Baylis G, Driver J (1995) Obligatory edge assignment in vision: the role of figure and part segmentation in symmetry detection. J Exp Psychol Hum Percept Perform 21(6):1323–1342
3. Biederman I (1987) Recognition by components: a theory of human image understanding. Psychol Rev 94:115–147
4. Blum H (1973) Biological shape and visual science (part I). J Theor Biol 38:205–287
5. Blum H, Nagel RN (1978) Shape description using weighted symmetric axis features. Pattern Recognit 10:167–180
6. Briscoe E (2008) Shape skeletons and shape similarity. PhD thesis, Rutgers University
7. Cole F, Sanik K, DeCarlo AFD, Funkhouser T, Rusinkiewicz S, Singh M (2009) How well do line drawings depict shape? In: ACM transactions on graphics (Proc. SIGGRAPH), vol 28
8. Cortese JM, Dyre BP (1996) Perceptual similarity of shapes generated from Fourier descriptors. J Exp Psychol Hum Percept Perform 22(1):133–143
9. de Winter J, Wagemans J (2006) Segmentation of object outlines into parts: a large-scale integrative study. Cognition 99(3):275–325
10. Demirci F, Shokoufandeh A, Keselman Y, Bretzner L, Dickinson S (2006) Object recognition as many-to-many feature matching. Int J Comput Vis 69(2):203–222
11. Driver J, Baylis GC (1996) Edge-assignment and figure-ground segmentation in short-term visual matching. Cogn Psychol 31:248–306
12. Feldman J (1997) Curvilinearity, covariance, and regularity in perceptual groups. Vis Res 37(20):2835–2848
13. Feldman J (2001) Bayesian contour integration. Percept Psychophys 63(7):1171–1182
14. Feldman J, Singh M (2005) Information along contours and object boundaries. Psychol Rev 112(1):243–252
15. Feldman J, Singh M (2006) Bayesian estimation of the shape skeleton. Proc Natl Acad Sci 103(47):18014–18019
16. Froyen V, Feldman J, Singh M (2010) A Bayesian framework for figure-ground interpretation. In: Lafferty J, Williams CKI, Shawe-Taylor J, Zemel R, Culotta A (eds) Advances in neural information processing systems, vol 23, pp 631–639
17. Hochberg J, Brooks V (1962) Pictoral recognition as an unlearned ability: a study of one child's performance. Am J Psychol 75(4):624–628
18. Hoffman DD, Richards WA (1984) Parts of recognition. Cognition 18:65–96
19. Hung CC, Carlson ET, Connor CE (2012) Medial axis shape coding in macaque inferotemporal cortex. Neuron 74(6):1099–1113
20. Kanizsa G, Gerbino W (1976) Convexity and symmetry in figure-ground organization. In: Henle M (ed) Vision and artifact. Springer, New York
21. Katz RA, Pizer SM (2003) Untangling the Blum medial axis transform. Int J Comput Vis 55(2/3):139–153
22. Kim S (2011) The influence of axiality on figure/ground assignment. Master's thesis, Rutgers University

23. Kimia BB (2003) One the role of medial geometry in human vision. J Physiol (Paris) 97:155–190
24. Koffka K (1935) Principles of gestalt psychology. Harcourt, New York
25. Kovács I, Fehér A, Julesz B (1970) Medial-point description of shape: a representation for action coding and its psychophysical correlates. Vis Res 38:2323–2333
26. Lescroart MD, Biederman I (2012) Cortical representation of medial axis structure. In: Cerebral cortex
27. Leymarie FF, Kimia BB (2007) The medial scaffold of 3d unorganised point clouds. IEEE Trans Pattern Anal Mach Intell 29(2):313–330
28. Leyton M (1989) Inferring causal history from shape. Cogn Sci 13:357–387
29. Ling H, Jacobs DW (2007) Shape classification using the inner-distance. IEEE Trans Pattern Anal Mach Intell 29(2):286–299
30. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
31. Mackworth A (1973) Interpreting pictures of polyhedral scenes. Artif Intell 4:121–137
32. Malik J (1987) Interpreting line drawings of curved objects. Int J Comput Vis 1:73–103
33. Mardia KV (1972) Statistics of directional data. Academic Press, London
34. Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Freeman, San Francisco
35. Marr D, Nishihara HK (1978) Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc Lond B 200:269–294
36. Palmer S, Davis J, Nelson R, Rock I (2008) Figure-ground effects on shape memory for objects versus holes. Perception 37(10):1569–1586
37. Richards W, Dawson B, Whittington D (1988) Encoding contour shape by curvature extrema. In: Natural computation. MIT Press, Cambridge
38. Rosin PL (2000) Shape partitioning by convexity. IEEE Trans Syst Man Cybern, Part A, Syst Hum 30:202–210
39. Sebastian TB, Kimia BB (2005) Curves vs. skeletons in object recognition. Signal Process 85:247–263
40. Siddiqi K, Shokoufandeh A, Dickinson S, Zucker S (1999) Shock graphs and shape matching. Int J Comput Vis 30:1–24
41. Siddiqi K, Tresness KJ, Kimia BB (1996) Parts of visual form: psychophysical aspects. Perception 25:399–424
42. Singh M, Froyen V, Feldman J (2012, forthcoming) Unifying parts and skeletons: a Bayesian approach to part decomposition
43. Singh M, Fulvio JM (2005) Visual extrapolation of contour geometry. Proc Natl Acad Sci USA 102(3):939–944
44. Singh M, Fulvio JM (2007) Bayesian contour extrapolation: geometric determinates of good continuation. Vis Res 47:783–798
45. Singh M, Hoffman DD (2001) Part-based representations of visual shape and implications for visual cognition. In: Shipley T, Kellman P (eds) From fragments to objects: segmentation and grouping in vision, advances in psychology, vol 130. Elsevier, New York, pp 401–459
46. Singh M, Seyranian GD, Hoffman DD (1999) Parsing silhouettes: the short-cut rule. Percept Psychophys 61(4):636–660
47. Telea A, Sminchisescu C, Dickinson S (2004) Optimal inference for hierarchical skeleton abstraction. In: Proceedings IEEE international conference on pattern recognition, Cambridge
48. Twarog NR, Tappen MF, Adelson EH (2012) Playing with puffball: simple scale-invariant inflation for use in vision and graphics. In: Proceedings of the ACM symposium on applied perception, pp 47–54
49. Waltz D (1975) Understanding line drawings of scenes with shadows. In: Winston PH (ed) The psychology of computer vision, pp 19–91
50. Wang X, Burbeck CA (1998) Scaled medial axis representation: evidence from position discrimination task. Vis Res 38(13):1947–1959

51. Weiss Y (1997) Interpreting images by propagating Bayesian beliefs. In: Adv. in neural information processing systems, pp 908–915
52. Wilder J, Feldman J, Singh M (2011) Superordinate shape classification using natural shape statistics. Cognition 119:325–340
53. Zhu S-C (1999) Stochastic jump-diffusion process for computing medial axes. IEEE Trans Pattern Anal Mach Intell 21(11):1158–1169