

Form Metrics for Interactive Rendering via Figural Models of Perception

by
Robert A. Katz

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Department of Computer Science.

Chapel Hill

2002

Approved by:

Advisor: Prof. Stephen Pizer

Reader: Prof. Gary Bishop

Reader: Prof. Dinesh Manocha

Prof. Anselmo Lastra

Prof. Guido Gerig

© 2002

Robert A. Katz

ALL RIGHTS RESERVED

ABSTRACT

Robert A. Katz

Form Metrics for Interactive Rendering via Figural Models of Perception

(Under the direction of Professor Stephen M. Pizer)

This work presents a method for quantifying important form cues in 2D polygonal models. These cues are derived from multi-scale medial analysis, and they also draw upon several new form analysis methods developed in this work. Among these new methods are a means of using the Blum Medial Axis Transform for stable computing, a method to decompose objects into a set of parts that includes the ambiguity we find in human perception of objects, and the simultaneous application of both internal and external medial representations of objects. These techniques are combined to create a local saliency measure, a global saliency measure and a measure of object complexity.

This work also demonstrates a new approach to simplifying complex polygonal models in order to accelerate interactive rendering. I propose that simplifying a model's form, based on how it is perceived and processed by the human visual system, offers the potential for more effective simplifications. In this research, I suggest a means of understanding object simplification in these perceptual terms by creating a perceptually based scheduling of simplification operations as well as perceptual measures of the degree of simplification and the visual similarity between a simplified object and its original. A new simplification scheme is based on these measures, and then this perceptual scheme is compared via examples to a geometric simplification method.

To my grandmother

She fueled my curiosity and inspired me by graduating college at the age of 75, and she taught me to always see the goodness of people. May she celebrate this work with me, wherever she is.

ACKNOWLEDGEMENTS

Thanks to my advisor, Steve Pizer, for making my PhD journey so rich and valuable. In addition to being a challenging intellectual foil, he taught me many lessons about integrity through his honesty and insistence on excellence.

I appreciate the patience and partnership of my committee: Gary Bishop, Guido Gerig, Anselmo Lastra and Dinesh Manocha. Thanks also to the team of folks that I am honored to work with at Legacy Center for their unwavering support while shouldering my load so I could focus on being a PhD student. My brother-in-law Greg was invaluable in preparing this document as well as offering humor and support all through the course of this research.

My deepest gratitude goes to the people who have had such a significant impact on my life. My sister Nancy has been my constant champion and the best big sister ever. Lori has inspired me, encouraged me and offered unconditional love. Finally, my parents have always stood by me, and they have taught me everything I truly value as important. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES.....	x
CHAPTER 1 INTRODUCTION	1
1.1. Motivation	1
1.2. Thesis	4
1.3. Contributions of This Work	4
1.4. Overview of This Document	6
CHAPTER 2 BACKGROUND	7
2.1. Form Perception	7
2.1.1. Preattentive vs. Attentive Perception	8
2.1.2. Objects as a Collection of Figures	9
2.1.3. Perception as a Multi-Scale Process.....	10
2.1.4. Perception as a Medial Process.....	15
2.2. Medial Models of Form	16
2.2.1. Blum Medial Axis Transform.....	16
2.2.2. Other Medial Representations.....	19

2.2.3. Multiscale Medial Axes.....	20
2.2.4. Cores.....	22
2.3. Object Simplification.....	24
2.3.1. Metrics.....	25
2.3.2. Operations.....	27
2.3.3. Scheduling.....	29
CHAPTER 3 FORM AS SUBSTANCE AND CONNECTIONS.....	31
3.1. Introduction.....	31
3.2. Objects are Substance and Connections.....	33
3.3. Parts Hierarchies and Form Information in the Blum MAT.....	35
3.4. Calculating Substance and Connection.....	38
3.4.1. Visual Conductance.....	39
3.4.2. Part-end Adjustment.....	43
3.4.3. Connection Shadows.....	46
3.5. Results.....	51
3.6. Conclusions.....	58
CHAPTER 4 MEDIALY BASED FORM SALIENCY METRICS.....	59
4.1. Visual Saliency Measures.....	59
4.2. Purpose of This Work.....	60
4.3. Components of Form Saliency.....	61

4.3.1. Aperture-based Scale.....	61
4.3.2. Relationship-distance Scale	62
4.3.3. Bilateral Relationships.....	65
4.3.4. Scale-based Bilateral Relationships	67
4.3.5. Figures	67
4.3.6. Figural Components of Form Saliency.....	69
4.4. Perceptual Form Metrics	71
4.4.1. Visual Conductance.....	72
4.4.2. Visual Mass.....	74
4.4.3. Visual Length.....	75
4.4.4. Visual Potential for Single Figures	76
4.4.5. Visual Potential at Figural Ends.....	78
4.4.6. Visual Potential with Multiple Figures.....	80
4.4.7. Visual Significance	82
4.5. Results.....	82
4.6. Conclusions	90
CHAPTER 5 PERCEPTION-DRIVEN OBJECT SIMPLIFICATION.....	91
5.1. Perceptual Metrics for Object Simplification.....	91
5.2. Perception-based Simplification.....	92
5.2.1. Simplification Schedule.....	92

5.2.2. Visual Complexity and Visual Similarity.....	93
5.2.3. Simplification Operations.....	95
5.3. A Perception-based Simplification Method.....	95
5.3.1. Step 1: Identify Point with Lowest Visual Significance	96
5.3.2. Step 2: Apply Simplification Transform	96
5.3.3. Step 3: Iteratively Repeat Process.....	98
5.3.4. Step 4: Re-tile Simplified Objects into Economical Representations.	99
5.4. Results.....	100
5.5. Conclusions	113
CHAPTER 6 CONCLUSIONS.....	120
6.1. Research Summary	120
6.2. Future Work	124
6.3. New Research Areas to Investigate	126
BIBLIOGRAPHY	129

LIST OF FIGURES

Fig. 1.1 Object Simplification	3
Fig. 2.1 Example of Figures	10
Fig. 2.2 Scale in Visual Perception	13
Fig. 2.3 The Blum Medial Axis Transform (MAT).....	17
Fig. 2.4 Segment Types in the Blum MAT	19
Fig. 2.5 Instability in the Blum MAT	21
Fig. 2.6 Example of a Core.....	22
Fig. 2.7 Scale in Core Representation	23
Fig. 2.8 Example of Levels of Detail (LODs)	25
Fig. 3.1 Challenges with using the Blum MAT.....	33
Fig. 3.2 Examples of Medial Points.....	35
Fig. 3.3 Instabilities in Blum MAT-based Part Classification.....	36
Fig. 3.4 Ambiguous Part Classification.....	37
Fig. 3.5 Visual Vectors and Visual Conductance	40
Fig. 3.6 Part-end Adjustment to Visual Conductance.....	43
Fig. 3.7 Connection Shadow.....	47
Fig. 3.8 Substance Metric Results.....	55
Fig. 3.9 Stability of Substance Metric.....	56

Fig. 3.10 Ambiguity in Substance Metric and a Limitation	57
Fig. 4.1 Scale in Visual Perception	63
Fig. 4.2 Figural Elongation.....	69
Fig. 4.3 Internal and External Figures.....	71
Fig. 4.4 Figural Continuity	73
Fig. 4.5 Visual Mass.....	74
Fig. 4.6 Saliency at Figural Ends	80
Fig. 4.8 Visual Length Results for An Object with Many Protrusions	83
Fig. 4.9. Saliency Results for a Tube	84
Fig. 4.10 Saliency Results of An Object with Progressively Widening Protrusions	86
Fig. 4.11 Saliency Results of an object with Indentations	87
Fig. 4.12 Saliency Results of Two Pinched Objects	87
Fig. 4.13 Saliency Results of a Lizard Object	88
Fig. 5.1 Visual Significance Measured on External Figures	94
Fig. 5.2 Boundary Sampling Around Point of Least Visual Significance	97
Fig. 5.3. Geometric Simplification Method.....	100
Fig. 5.4. Perceptual Simplification of a Corner	101
Fig. 5.5. Perceptual Simplification of Object with a Protrusion.....	102
Fig. 5.6. Geometric Simplification of Object with Protrusion	103
Fig. 5.7 Comparison of Perceptual and Geometric Methods for Object with Protrusion	104

Fig. 5.8 Perceptual Simplification of Object with Indentation.....	106
Fig. 5.9. Geometric Simplification of Object with Indentation.....	107
Fig. 5.10. Comparison of Methods for Object with Indentation.....	107
Fig. 5.11 Comparison of Perceptual and Geometric Methods for Cat Object.....	108
Fig. 5.12 Comparison of Perceptual and Geometric Methods for Lizard Object.....	111
Fig. 5.13 Changing Topology in Simplifications	112
Fig. 5.14 Objects Tested with Perceptual Metrics.....	113
Fig. 5.15 Perceptual LODs of a Cat Object.....	116
Fig. 5.16 Geometric LODs of a Cat Object.....	117
Fig. 5.17 Perceptual LODs of a Lizard Object.....	118
Fig. 5.18 Geometric LODs of a Lizard Object.....	119

CHAPTER 1 INTRODUCTION

1.1. Motivation

Computer graphics addresses a variety of objectives. In general, the goal of computer graphics is to convey information to the computer user; this allows a computer to visually communicate with its users. Consider, for example, the field of virtual reality. Interactive 3D computer graphics has matured to the point where virtual reality (VR) systems successfully allow users to believe that they are part of the displayed scene rather than watching computer generated images. Virtual reality allows users to do such things as simulate walking through a building before it is ever constructed, viewing the walls, furniture and even wallpaper as if the users were in the actual building [Walk92]. Flight simulators train pilots without risking plane crashes [Zyda88], and other VR systems provide scientists with a method of interacting with data in ways that were until now impossible, such as walking on far-away planets [McGreevy91] and manipulating surfaces at an atomic level [Taylor93].

By definition, computer graphics is tightly linked with human perception. The purpose of computer graphics is to generate images for humans to interpret. The goal is to have viewers perceive the desired information or to respond in desired ways. Computers render information into images and display these to the viewers. Human viewers receive these images and turn them into higher level information that evokes thought and actions. Obviously, a successful graphics system will create computer generated images that allow for the most information to be extracted by viewers with the least amount of visual processing.

Virtual reality systems strive to present a continuous stream of images that viewers will perceive as reality. The goal of algorithms used in VR is to balance two conflicting re-

quirements in order to present images that viewers perceive as realistic scenes. There is a constant trade-off between richly detailed scenes and high frame rates. Ideally, a virtual reality system would display images that are indistinguishable from real, live scenes. This includes scenes filled with many objects, such as all the many items on a cluttered desk or all the plants, bugs, dirt and stones in a garden; this also includes all the minute texture and detail of each object, such as the dimples on an orange or the weave of a fabric. For some scenes, graphics algorithms do exist that can create images very similar to photographs. However, such images can require hours of computer time to create.

On the other hand, studies have shown that humans need up to 15 new computer images displayed every second, with a latency of less than 0.1 seconds between user input and data display, to create the illusion of smooth, realistic motion [Stanney02]. When a user is wearing a head-mounted display and quickly jerks around his head, significantly higher frame rates and lower latency may be necessary to maintain the illusion of reality. The demand for a high enough frame rate provides graphics systems with, at most, 0.1 seconds to generate each image. This severely limits the amount of time available to create realistic scenes.

Computer generated scenes start with a geometric description of each object in the scene; these descriptions are usually polyhedral models, and they form a "blueprint" that the computer renders into an image. The time required to render each polyhedral model is related to the number of polygons in the model. Models with fewer polygons generally render more quickly than those with more polygons.

One common and successful method of creating images in less time is to simplify the models in a scene before rendering them. This simplification involves creating new models that are similar to the original ones but that have fewer polygons. The challenge in designing a simplification scheme is to create simplified models that have removed as many polygons as possible while still retaining as much similarity as possible to the original model.

The goal of simplification methods is to create models that appear as similar as possible as the original model but render more quickly. In other words, the goal is to minimize the

perceptual deviation between the renderings of the two models. However, a very common method for measuring the similarity of simplified models to the original has been to use geometry-based metrics. These approaches minimize geometric deviation between models. In order to address the ultimate goal of conveying the most information to viewers, these methods should measure the perceived similarity between models. For this, perceptual metrics are needed.

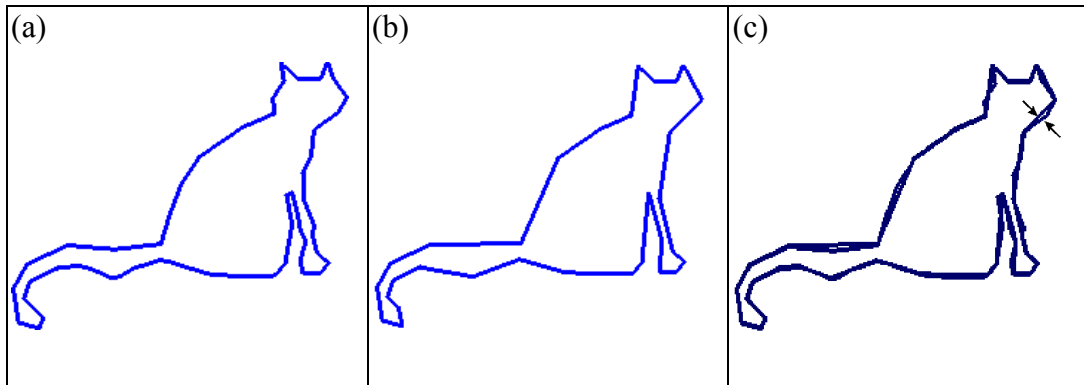


Fig. 1.1 Object Simplification

- (a) A 2D geometric model of an object.
- (b) A simplified version of the model.
- (c) Geometric measurement of the simplification.

With this motivation, this research explicitly addresses a link between computer graphics and human perception. The work explores how to apply our knowledge of the human visual system to solving graphics problems. Specifically, the task of object simplification is used as a testbed, and form perception is the targeted perceptual domain that is explored. The intent of this work is not to produce optimized or even practical algorithms. Instead, this research is an inquiry and investigation of possibilities that are available through applying a perceptual basis to task of graphics rendering.

1.2. Thesis

In this dissertation, I assert the following:

Medial form representations can provide a means to quantify perceptual form cues with measures that qualitatively satisfy perceptual effects over a variety of objects.

Using such perceptual measures for interactive rendering reveals desirable properties and provides significantly different results than non-perceptually based approaches.

Given the vast complexities of the human visual system, many choices were made in limiting the scope of this work to form perception and to medial form representations. All the algorithms were then designed to be consistent with these choices and their associated perceptual theories. The results of this work show that perception-based algorithms can be useful for interactive rendering.

A by-product of grounding this work in human vision theory is that the methods developed here are designed for 2D objects. All of our visual processing is applied to 2D retinal images; any 3D form information that we extract is derived from 2D cues in the images. Thus, this work is limited to 2D polygonal objects such as those that result from the projection of 3D models onto a viewing plane.

1.3. Contributions of This Work

This research contributes several new ideas and results to the fields of form analysis and computer graphics.

A. Stable Medial Analysis via Substance and Connection Measures

The first contribution is a method for classifying medial axes as the substance of an object or a connector between the parts of the object. The key to this method is that it uses a weighting scheme that allows for ambiguity in this classification; an axis point may be purely substance or purely connection, but it may also be a combination of the two.

This weighting scheme and its ability to allow ambiguity leads to a natural decomposition of objects into a set of simple parts, and this decomposition parallels the classifications and ambiguities generated by the human visual system. The weighting scheme also offers a solution to the long-standing instability problem with the Blum Medial Axis Transform that has vexed researchers for decades.

B. Combination of External and Internal Medial Axes

In this work, I also create medial axes that are distinctly internal or external to the object, but I then treat these axes in a very similar manner. An indentation into a figure is merely a protrusion from the figure's background, and I treat this "external" protrusion exactly the same way that I treat an internal protrusion. This combination and inter-relation of internal and external axes provides insights into form and new opportunities in medial form analysis.

C. Perceptual Measures

My research proposes several new methods and metrics that are useful for quantifying the cues that we use when perceiving objects. I develop a series of form metrics that lead to a form saliency measure for measuring the visual importance of every point on an object. These measures are visual conductance, visual length, visual mass, visual potential, and visual significance. Visual conductance measures the tendency of adjacent object parts to be perceived as a single part, and it directs the way we visually assemble the parts of an object into a visual hierarchy of "main" parts and "lesser" parts. Visual length measures the elongation of the object being evaluated, capturing the perceptual result that narrow figures appear longer than wide figures with the same Euclidean length. The visual mass of a figure reflects the visual impression of its size; figures with greater visual mass cover more of our field of view. Visual potential is a relationship between visual length and visual mass at a single point that captures the local significance; visual significance is the integration of this value in a region around a point, where the integration is affected by object information at that point. This visual significance is my metric for the perceptual importance of any point on an object when compared to every other point on the object.

D. Perception-driven Object Simplification

This work also offers a new approach to the problem of object simplification. The method simplifies the underlying object form before addressing an object's representation. It uses the metrics described above to rank the perceptual importance of every point on a model, and at each simplification iteration it identifies the point that is least important visually. The simplification transform itself also uses perceptual criteria to dictate how to simplify the object in a way that is least damaging perceptually.

The contributions here lie in the perceptual metrics used to measure simplification as well as the new insights into simplification that are generated with this approach. The metrics provide a perceptually-motivated simplification algorithm and a quantifiable form measure to compare the similarity of simplified objects. Using a perceptual approach, as opposed to the more common geometric measures, uncovers new opportunities for simplification, such as simplifying the regions of space immediately external to the object and using simplification operations that may drastically change an object's geometry but have relatively lower visual impact.

1.4. Overview of This Document

The rest of this dissertation is structured as follows. Chapter 2 presents background and previous research related to this work. The following three chapters describe the new advances in this research. Chapter 3 describes the method developed for analyzing objects as substance and connection and the benefits this provides. Chapter 4 presents the perceptual measures and the method of combining internal and external axes. Chapter 5 applies these developments to the problem of object simplification by demonstrating a perceptually driven shape simplification algorithm. Finally, chapter 6 summarizes this research and offers several ideas for follow-up work.

CHAPTER 2 BACKGROUND

2.1. Form Perception

Our understanding of how humans process visual form information is still very limited. We know that light passes through the pupil to create electrical signals in the photoreceptors of the retina. These signals are bundled together in groups called receptive fields, corresponding to patches of neighboring photoreceptors, and then passed along the optic pathway to the visual cortex. Here, in the visual cortex, is where visual cognition occurs, turning the electrical pulses from the retina into what we perceive as color, texture and form. And here, in the visual cortex, is where our understanding of the visual process becomes particularly inadequate; a variety of theories have been proposed to explain the workings of human form perception. In this section, I will survey some of the key theories, models, and experimental results in form perception and how they relate to the problem of shape simplification.

The first step in our perception of form occurs when rays of light reflected from objects in our 3D world project onto a 2D retina. Thus, from the very start, perception of 3D objects is converted into a 2D problem. This projection is much like the rendering step in computer graphics, where 3D models from world space are transformed onto a 2D image plane. Human perception of 3D objects is performed using 2D operations on 2D images.

There are many cues in 2D retinal images that our visual cortex uses to extract form information, including luminance boundaries, shadows and shading. Our visual cortex uses this information when parsing a scene to create the form relationships mentioned above. In this work, I ignore all cues except for an object's edge information, which is certainly among the most important. Furthermore, rather than render an image of a model and then

extract its edge information from the image, I derive the model's basic form relationships directly from its geometric description.

Using this representation of an object's edges, I then use a form perception theory called Object Representation by Cores [Burbeck95] to drive shape simplification. This theory suggests that our visual systems use multiscale medial processing to decompose objects into simple parts for further processing. My simplification method mimics the visual processing hypothesized in the Core model of form perception, and it uses the resulting form relationships to guide the simplifications.

The remainder of this section discusses the relevant background in form perception and computational issues when working with the form of objects. First, two fundamental types of visual processing, preattentive and attentive perception are discussed. Next, I discuss how objects are segmented into parts, and then the form relationships of scale-space and locality are presented. Finally, the medial relationships created in our visual system are reviewed.

2.1.1. Preattentive vs. Attentive Perception

A key distinction in visual perception is *preattentive* versus *attentive* perception. Objects that are visible in a scene without our attention focused on them are said to be preattentively processed [Wolfe97]. When we focus our attention on the objects, our higher-level attentive processing kicks in. For example, if we walk into a room, our preattentive processing tells us with only a single glance that there is a brown block-like object against the wall with a spindly object in front of it and that there are flat white rectangles and colored cylinders on the brown object. By focusing on the brown object and studying it further, our attentive processing identifies it as a desk with papers and pens on it, and a chair in front of it.

Our preattentive perception scans an entire scene and decomposes it into basic features; this is done within approximately 200 msec, too quickly for our eyes to move to and focus on more than a few parts of the scene. Our attentive perception focuses on a part of a

scene and performs much higher level processing leading to object recognition and cognitive tasks; this can take far more than 200 msec.

While the example above distinguishes preattentive and attentive perception, it blurs the fact that there are links and feedback paths in our brains between the two. In reality, the example would hold true if we were blindfolded, transported to a foreign place, and then opened our eyes; if we knew we were in an office building and walked into the room, our attentive visual system would probably assume that the brown block-like object was a desk covered with papers and pens without requiring focused attention on it.

Research suggests that preattentive form processing recognizes and decodes many features of form, such as size, curvature, orientation and length; however, there does not appear to be any global form processing in the preattentive stage [Wright93][Wolfe97]. An object that is preattentively processed will then “pop-out” to catch our attention and thus stimulate higher-level visual processing [Treisman80]. In the attentive processing stage, more complex form perception occurs such as recognizing, analyzing and evaluating objects. Attentive processing is highly context-dependent; the task a user is performing strongly affects attentive perception [Bruce96]. Although objects are not identified preattentively, many important features of objects are processed without attention, such as might be required for navigating virtual worlds. In order to avoid the limitations and complexities of task-dependent perception and to limit the scope of this research to a domain that can be investigated, my work focuses on the problem of navigating through virtual worlds and therefore on preattentive form processing. Within this limited domain, only certain aspects of form are considered, and these are detailed in Chapter 4.

2.1.2. Objects as a Collection of Figures

A crucial step in our perception of form appears to be segmenting and decomposing objects into simpler parts [Wolfe97][Hoffman84][Biederman87]. Intuitively, these parts can be thought of as a full object or a protrusion from or indentation into another part. Burbeck and Pizer formally define these parts as a spatial region that produces a Core [Burbeck95]. Parts defined in this way are called *figures* (see Fig. 2.1)..

Burbeck and Pizer go on further to show how figures encompass many of the phenomena shown in psychophysical studies. Figures are formed at the places of highest negative boundary curvature, which studies have shown are the places where we visually decompose objects [Hoffman84][Biederman87][Braunstein89]. The junctions of figures also have special importance, matching Biederman's work on the junctions of visual parts [Biederman87]. The ends of figures have also been shown to match work showing their special visual significance [Hubel77][Orban79][Leyton92]. [Rock93] shows that figures are captured preattentively by our visual systems.

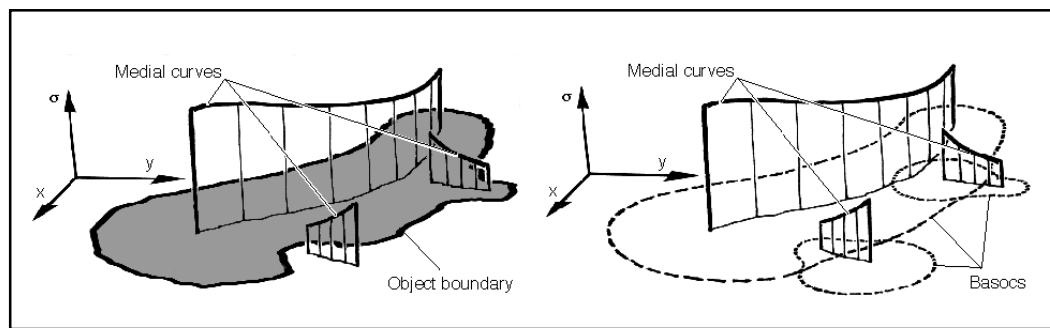


Fig. 2.1 Example of Figures

An object segmented into its three basic parts, known as *figures*.

2.1.3. Perception as a Multi-Scale Process

As Koenderink eloquently describes, all visual operations are performed over an aperture of finite size [Koenderink84]. While Mathematics can describe the world with infinite precision and with an infinitely small aperture, all practical applications of vision must account for an aperture of finite size. Finite apertures introduce blurring to vision systems, since details that project significantly smaller to the image plane than the aperture size are lost. This blurring process can be a drawback to a vision system or it can be used advantageously to remove unwanted details in order to focus on larger scene information. These finite apertures and their consequences can be described mathematically with the notion of scale-space.

The size of the aperture over which visual operations are performed is commonly called the *scale* of the operation. A second, distinct notion of scale is described below. Large apertures give coarsely detailed, larger-scale images. Small apertures give finely detailed, small-scale images. The same scene viewed with differently scaled apertures may appear very differently.

One way to visualize an object seen through different apertures is to imagine what it looks like from different distances. A maple tree viewed from far away looks like a cylindrical trunk with a bushy green mass on top; this large scale representation is commonly used by young children in their drawings. As we move closer to the tree and view it with a smaller aperture, some of the larger limbs are distinguished. This process continues until we are close enough to see individual leaves instead of a green mass; leaves are a smaller scale feature of the tree than the trunk is. The process can continue to reveal the smaller scale stems of the leaves, then the fine grain of the bark and so on.

At any scale, an image is clarified and more detailed compared to the same scene at a larger scale. If we treat aperture size as an independent parameter that can be infinitely varied, then we must attach a scale component to any measurement that is made through that aperture. We no longer think of a location (x, y) in an image; we must think of the location at a given scale σ , or (x, y, σ) .

Adding this scale component to our familiar Cartesian coordinates gives us a scale-space. Scale-spaces are distinguished by the filter that is applied to imaging operations. Young and Koenderink showed how this neurophysiological response could be modeled with linear combinations of offset Gaussians [Young86] [Young87][Koenderink90]. This strongly suggests that the human visual system operates over a scale space that has a Gaussian-like blurring filter, where the size of the Gaussian kernel is directly proportional to the size of the aperture.

In fact, the multi-scale nature of our perception is more than a conceptual model; it is created in the “wiring” of our visual systems. The receptive fields of the retina are patches of neighboring retinal photoreceptors; the signals from each of the photoreceptors are grouped together as they are passed to the visual cortex. Each photoreceptor feeds

many receptive fields, creating receptive fields of many sizes and fields that cover the entire retinal image [Bruce96]. The visual cortex performs its processing on these receptive fields, and the different size fields that contain the same scene elements lead to visual processing at different scales.

The second notion of scale reveals itself when we relate multiple entities on an object or across objects. These entities may be points on the boundary, a protruding figure and its parent figure, a medial point and a boundary point, two separate objects and so on. When relating such entities, the units for measuring distance between them must be determined. This is the second notion of scale. It can be thought of as the spacing of our ruler; is the distance measured in meters, inches or maybe some parameter of the object?

For example, consider Fig. 2.2a, where the two points can be thought of as two units apart, three units apart or ten units apart. The same entities can have different distance measurements depending on the scale used to perform the measurement. The choice of scale can be based on many factors such as the entities being related or the context of the problem being solved.

One natural method for setting the scale of interrelationship distances is to use an aperture-based ruler. [Burbeck96] shows how every medial location on an object has an associated natural aperture size for the object at that point, and such apertures are proportional to the width of the object at that point. The natural aperture size we use to view a car body is much larger than that to view a wheel, and that aperture is much larger than the natural one to view a lug nut. A scale-based ruler for distance measurements is based on the aperture size of the object at the points being related. Fig. 2.2b shows the same two points as in Fig. 2.2a with a width-based aperture of radius R . With this aperture the points are three units apart.

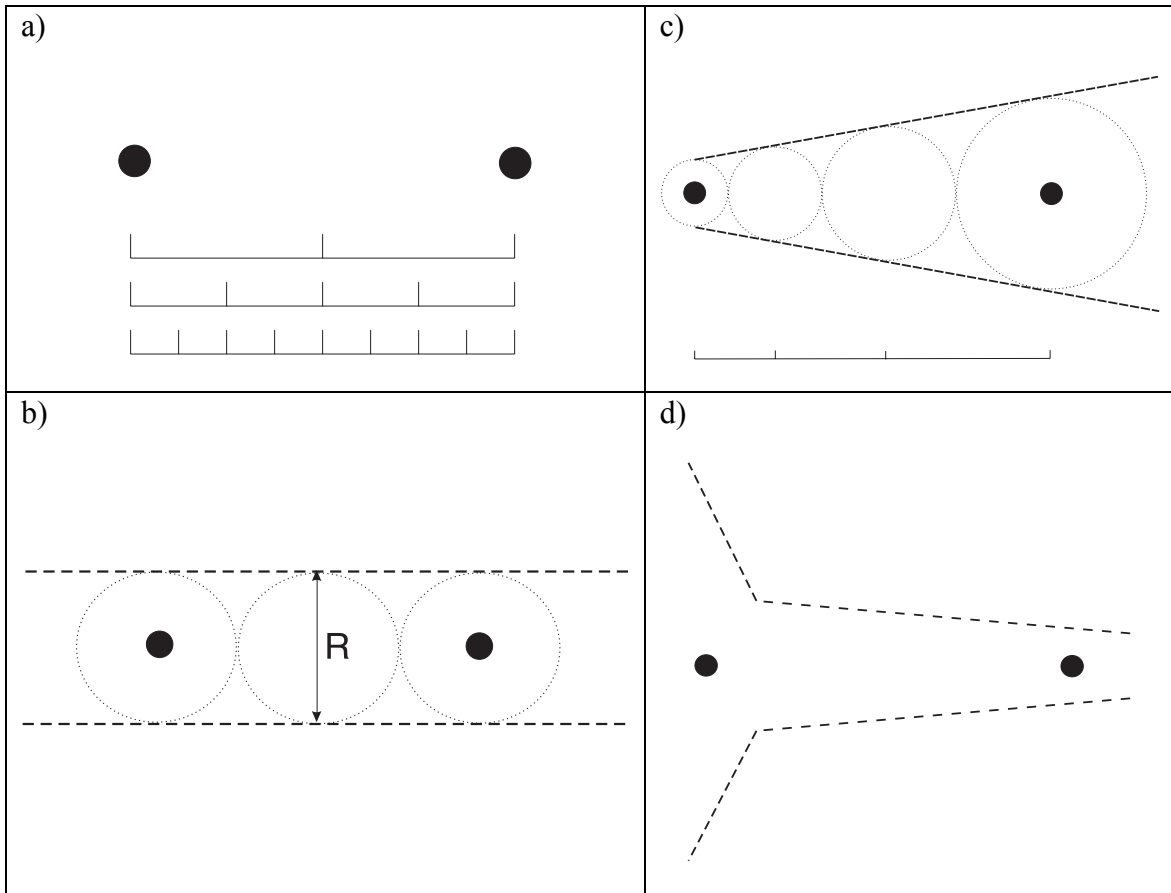


Fig. 2.2 Scale in Visual Perception

- (a) Two points with distance measured in different units.
- (b) Same two points with distance measured in width-based units.
- (c) Two points with distance measured with varying width-based units.
Gradual change in widths implies points are on same part of the object.
- (d) Same two points as in (c) but on separate object parts as distinguished by rapid change in width

There are also strong computational reasons for aperture-based distance measurements. In the process of understanding a stream of visual input, relationships are created among entities. For every relationship, we start with one entity and seek out a matching entity. This search spans all locations in all dimensions to find a match. If the search was always performed at the smallest possible scale, this would always require that an enormous number of locations be searched when attempting to create a relationship. The process would, at best, be inefficient and place extreme demands on the visual processing system. At worst, it would make the creation of some simple relationship intractable. By performing the search for related entities at the largest possible scale, aperture-based dis-

tance measurements allow a wide range of relationships to be made with a relatively constant amount of visual processing.

Combining the notion of aperture-based scale with the scale of interrelationship distances allows entities with different aperture widths to be related. Fig. 2.2c shows two points within a narrowing object. While the distance between these points, when measured by aperture widths, can be perceived as greater than the distance in Fig. 2.2b, the two points are still closely related. The points can be considered local to each other since they are local in the left-most point's frame of reference and the scale changes rather slowly as it progresses to the right-most point.

The same two points in Fig. 2.2d have a very different scale-based relationship. The two points have the same object widths as in Fig 2.2c, but there is a very rapid change in scale along the line connecting the points. The locality of the left-most point does not extend to the right-most one. This leads to our perception that the points are on different parts of the object instead of the same part as in Fig. 2.2c. Also, we perceive the right-most point to be on a part that protrudes from the larger scale "parent" to the left.

In general, scale-based analysis more closely relates points that have progressively similar scales. An example of this is the points along our arm, where our forearm progressively widens from our wrist to our elbow but the change in width is gradual enough that they are considered the same part of our body. On the other hand, the widths along our upper arm drastically change at our shoulders where suddenly the torso has a much larger width. This large and rapid change in scale leads to our perception of an arm as distinct from our torso. We think of our arms as "attached-to" but not "part-of" our torso.

The combination of aperture-based scale and interrelationship-distance scale thus directs the segmentation of objects into a set of simple parts. Neighboring points at progressively similar scales are on the same part. Neighboring points at drastically different scales are on distinct parts that have a parent/child relationship. A body's different scales tells us that arms and legs are sub-parts of the torso and that fingers are sub-parts of hands, but a torso can not be directly related to fingers; such a relationship occurs through the inter-

mediate parts at intermediate scale levels. This analysis also tells us to tie an arm to its torso's coordinate system rather than the other way around.

One result of the multi-scale processing in preattentive perception is that the perception of an object's form at a given point is determined only by the local region surrounding that point, not by distant features of the object. The perception of our nose's form is not influenced directly by our shoulders, although it may be affected by the form of our eye sockets or our cheek bones. Since all visual operations are done in a scale-space, the size of the local region is determined by the scale of the operation. The perception of our head's form may be influenced by our shoulders, since the head is a larger-scale feature and thus is influenced by a proportionately larger region around it.

2.1.4. Perception as a Medial Process

For many years, researchers believed that our visual systems inferred form by detecting the boundaries of objects, and then performed further visual form processing by operating on representations of those boundaries. More recent work, however, offers evidence that an important aspect of human form perception functions by focusing on medial information: the middle and width of an object that occurs when opposing edges of the object are paired. Although many researchers still offer boundary representations as models for human visual processing [Grossberg94], Burbeck and Pizer argue that our brains do not have enough neurons to actually use boundary representations in any multiscale processing [Burbeck95]; they go on to show that processing models based on medial information, however, are physiologically possible and psychophysically supported.

Blum was the first to propose the full potential of using medial information [Blum67], and Kovacs & Julesz [Kovacs94], Psothka [Psothka78], Frome [Frome72] and Burbeck & Pizer [Burbeck95][Burbeck96] have offered psychophysical evidence of our vision system's sensitivity to medial properties. Although Blum's work itself does not adequately account for this evidence, he inspired work that does hold up to the psychophysical [Burbeck95][Burbeck96] and physiological studies [Lee95].

Central to any medial model of visual processing is the relationship between a boundary point and its related partner on the object's opposite side; each boundary point is a medial *involute* of the other. The neurophysiological and psychophysical support for medial properties suggest that involutes are important in our visual systems. A pair of involutes ties together boundary points that are usually widely separated when measured across the surface. This relation between distant points gives medial models some global form aspects that are not easily accessed in boundary-based models.

Medial object representations are based on a mathematical locus of central points between involutes. This locus of locations with their associated widths form a scale-space that is based on the scale at which the loci are extracted from the scene information. Burbeck and Pizer define a multiscale medial representation called "Cores" [Burbeck95], and they provide a complete mathematical description [Pizer98]. In [Burbeck96] they show how Cores closely predict certain phenomena of the human visual system.

Given the current body of knowledge, my premise in this work is that human form perception relies upon multiscale medial processing. Aperture size can be derived directly from the distance between medial involutes, while locality is based on scale and extends to include relations between involutes. Figures can be directly segmented from medial representations, and the intersections and ends of figures match known perceptual phenomena. The following section discusses these medial models of form.

2.2. Medial Models of Form

All medial models represent an object as a set of middle points centered in the object and a measure of the object's width at each middle point. Such a locus of medial points is often called a *skeleton* because it is centered in the object much like a skeleton would be. Medial models are differentiated mainly by the method used to find the medial locus and how the object's width is measured at each medial point.

2.2.1. Blum Medial Axis Transform

The first important medial model was proposed by Blum [Blum67]. He described his Medial Axis Transform (MAT) using a "prairie-grass analogy". Imagine an object as a

patch of prairie-grass with the entire perimeter set on fire simultaneously. As the fire uniformly eats into the patch of grass, fire-fronts from opposing sides meet and quench. Each of these quench points is a medial point in Blum's medial axis, and the time it takes for the fires to quench is directly proportional to the object's width at that point.

Mathematically, Blum's MAT can be represented as the set of centers of maximal circles and the radius of each circle. Maximal circles are those that are not contained in any other circle that also lies within the object. The locus of centers is a tree-like axis that is well defined and unique for any given object. For objects with holes, the axis will have loops and will produce a cyclic graph. The axis and the circle radius at each axis point together make up the Blum MAT.

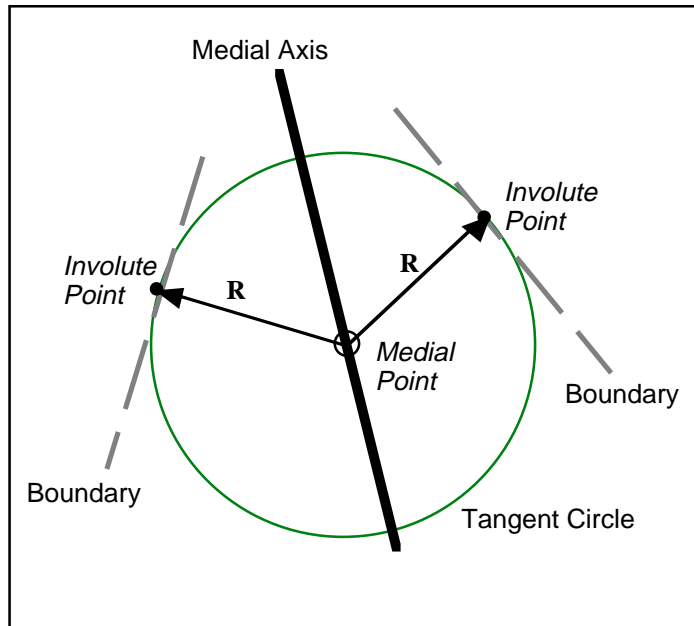


Fig. 2.3 The Blum Medial Axis Transform (MAT)

A point on the a Blum medial axis along with its maximal circle.

A different description is created by looking in from all boundary points in their normal directions. When two of these rays intersect and the intersection point is equidistant from the two originating boundary points, the intersection point is a locus on the medial axis. When I discuss Cores, I will invert this perspective and look out from potential axis

points; when the boundary is found at equal distances from the candidate point and at orientations that are orthogonal to the view directions, then it is a point on the Core.

Boundary points that are tangent points of the same maximal circle are the prairie-grass points that would quench at that circle's medial point. These tangent boundary points are *involutes*. In general, a point on the medial axis has two involutes; such a point is called a *normal* medial point. Some special medial axis points do not have a two boundary tangencies. This occurs at either medial axis *branch* points or at medial axis *end* points. Branch points have generically three involutes, with the exact number of circle tangencies equal to the number of axis segments that meet at the medial point. At axis end points, the two normal involutes coincide at a single point. Equivalently, a circle with radius equal to the boundary point's radius of curvature osculates the boundary at the end's involute point. For corners in the boundary, the axis ends right at the boundary; here, the maximal circle can be considered to have zero radius, and the involutes coincide at the corner point as well.

The Blum MAT is a transform; it contains exactly the same information as the boundary from which it is generated. Thus, the MAT is merely a different representation of the boundary. What makes it useful is that boundary information as well as involute information can be cheaply and directly derived from the medial axis representation. It is computationally expensive to derive involute information from boundary representations.

One way of computing the Blum MAT of an object involves mimicking the "fire-fronts" from the boundary to find their quench points. For polyhedra, the Blum MAT is merely a subset of the boundary's Voronoi diagram. The Voronoi diagram for 2D polygons can be computed exactly, and there are successful algorithms for computing the Blum MAT of 2D polyhedra [Lee82], [Seel94] etc.

There are three kinds of segments of the Blum MAT of 2D polygons. Given a boundary consisting of line segments and the concave vertices of the boundary, each combination of line segment and concave vertex gives a different kind of medial segment. The medial axis between two line segments is a line, and the width function along the line varies linearly. The medial axis between a concave vertex and a line segment is a parabolic

segment whose width function varies non-linearly. Finally, the axis between two concave vertices is a line segment with a non-linear width function.

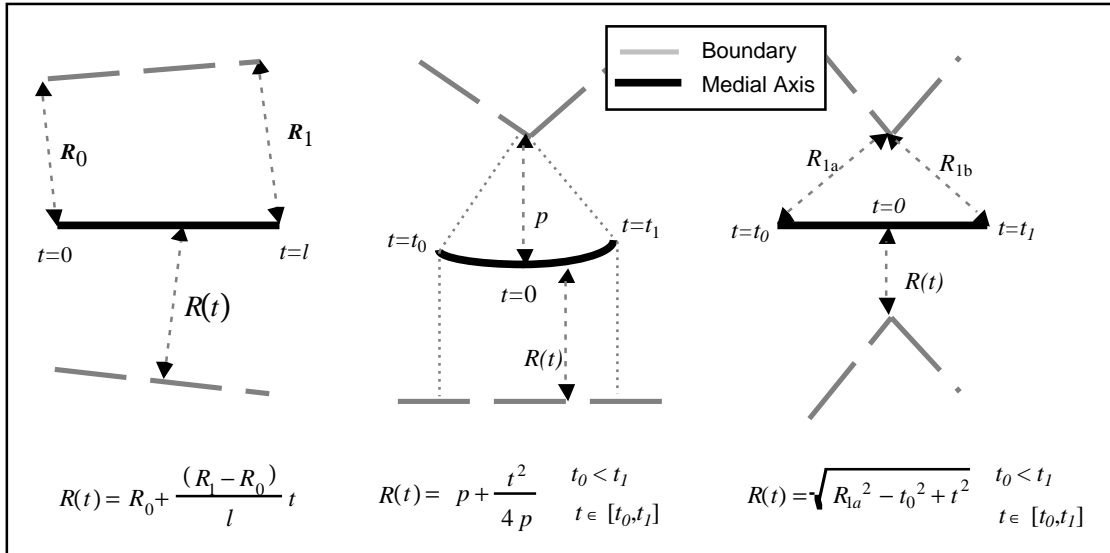


Fig. 2.4 Segment Types in the Blum MAT

The three types of Blum MAT segments. A line segment between two boundary line segments, with linear width function $R(t)$. A parabolic segment between a boundary line segment and a boundary concave vertex, with quadratic width function $R(t)$. And a line segment between two boundary concave vertices, with non-linear width function $R(t)$.

2.2.2. Other Medial Representations

Many researchers have followed Blum's original MAT and developed other medial representations. Some methods are variations of the Blum MAT, differing in the way that medial points are derived from a bitangent circle. Smoothed Local Symmetries (SLS) defines the axis points to be the center of the chord between the boundary points associated with the tangencies [Brady84]. Symmetry Sets is similar to the Blum MAT but it broadens Blum's maximal circle constraint to include all bitangent circles [Bruce85]. This definition allows axes to go outside of object's interior, and it also allows a boundary point to be associated with multiple medial axes. These extensions encode additional form information compared to the Blum MAT. Process-Infering Symmetry Analysis (PISA) uses the midpoint of the maximal circle's arc between the two involutes as loci of

the medial axes [Leyton88]. This approach is shown to be beneficial for a process-oriented description of form.

While there have been many papers that compute the exact Blum MAT and Blum-like medial axes, [Siddiqi99] produces a different medial axis that follows from a derivation that is different than using bitangent circles. This method simulates the grassfire analogy by starting with the boundary and letting it evolve through a partial differential equation. The medial locus is identified as singularities resulting from shocks of the PDE solution, and the width of the object is derived from the time of the shock's occurrence. With the proper parameter, this method produces the Blum MAT, but with other values for this parameter it provides a medial description that is more stable than the Blum MAT against small boundary perturbations.

2.2.3. Multiscale Medial Axes

One of the most severe limitations of the Blum MAT is that it is sensitive to small boundary perturbations; a minor change to the boundary causes a significant change in the MAT. This problem arises because Blum treats all scales identically. A small-scale perturbation in the boundary can affect the large-scale axes. As shown in Fig. 2.5, a small protrusion from a large scale rectangle creates an entire axis segment from the tip of the protrusion to the large-scale axis segment.

To solve this problem, multiscale analogs to the Blum MAT have been created. These solutions use the visual notion of aperture-based scale and regularize the medial extraction process. For example, [Ogniewicz92] has developed a method of regularizing medial axes by successively performing operations that remove boundary perturbations. A different representation is obtained in simulating the grassfire analogy using a partial differential equation [Kimia96]. [Siddiqi99] regularizes this process, where the PDE shocks corresponding to the quenching of grassfire fronts are detected with the use of two different apertures. The first stage of the process – finding the distance from the initial boundary – is smoothed using one aperture, and the shocks are found using a second aperture. See [Pizer02] for a detailed discussion of these methods as regularizing via apertures and a comparison of the properties of the methods.

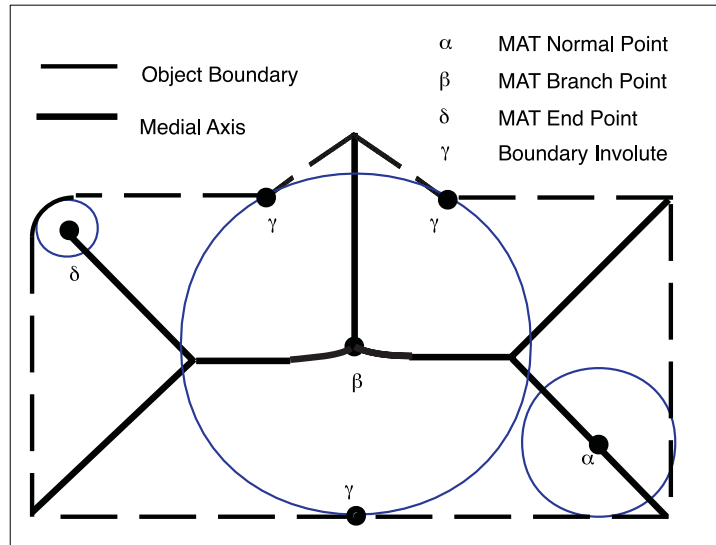


Fig. 2.5 Instability in the Blum MAT

An object and its medial axis. The small bump in the boundary at the top of the rectangle produces a significant addition to the MAT representation. (By Dave Chen, University of North Carolina.)

Another general approach to creating multiscale medial representations is to compute medial loci from a height ridge of medialness. Such methods identify all potential medial points and send out equal length rays to search for the boundary. At the end of the ray, a special detector is used to measure how “boundary-like” the region is at the end of the ray. This detector is a convolution kernel that is oriented to detect boundaries perpendicular to the ray and whose size is proportional to the ray’s length. The “boundariness” values of these detectors are in a 5-space $(x, y, \sigma, \theta_1, \theta_2)$, where σ is the ray length or *scale*, and θ_1 and θ_2 measure the orientation of the two rays relative to horizontal.

Summing the pair of boundariness values that stem from a potential medial point gives the “medialness” of the point, $M(x, y, \sigma, \theta_1, \theta_2)$. This measures the prospect that the location is a middle to the object, at that scale. When the two involutes both have a high boundariness value, it reinforces the medialness measure at the medial point. This approach explicitly ties scale-based apertures to an object’s medial width.’

As a specific example, Cores are 1D ridges in this 5D medialness space. This medialness function imposes a graph on the Cores of a complex object, from which a hierarchy of Cores can be created. See [Eberly93] for a complete description of the mathematics of ridges, and [Pizer98] for an in-depth discussion of the mathematics of scale-space and Cores.

[Székely96] reviews many multiscale medial representations, and [Pizer02] gives a comprehensive comparison of these three general approaches.

2.2.4. Cores

Cores represent objects by using a medial axis and a width at every point along the axis. Complex objects are represented with multiple Cores. A significant difference from the MAT, however, is that Cores are measured over a finite aperture. Rather than looking for circle tangent points on the boundary, Cores integrate a directional boundariness measure within a region. The size of the aperture used for this integration is proportional to the width of the object at that point. This width is a core's *scale*; it corresponds with the radius of Blum's MAT. A special measure of boundariness gives the "endness" of Core points; high endness values mark an end of a Core axis and correspond to the osculating end points of the Blum MAT.

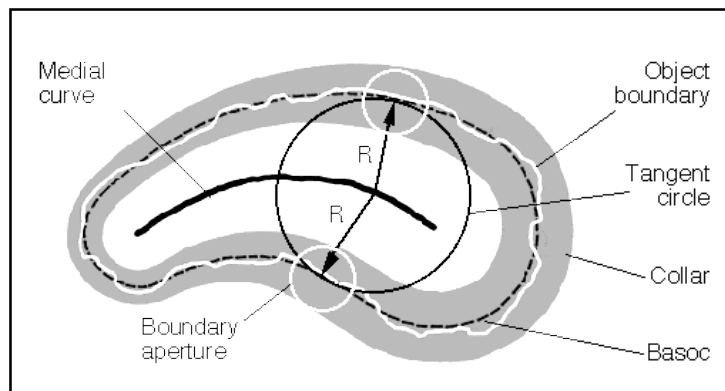


Fig. 2.6 Example of a Core

A Core point and its involutes. The object's boundary is measured with an aperture, and the size of the aperture determines the width of the boundary collar. The boundary at the scale of the Core (BASOC) is located within the collar. (By Andre State, University of North Carolina.)

There is an important consequence of the finite apertures used to create Cores. Since the size of the aperture used to determine involutes is proportional to the scale at their medial locus, the width of a Core determines the accuracy of its boundary measurement. Thus, when Cores are created, their boundaries are not exact; rather, the boundary is known to lie within a bounded region. This region is called the *collar*. Unlike the Blum MAT, where every axis point has exact involute points on the boundary, Cores have only a boundary with precision relative to the scale of the core point. Selecting an exact boundary surface within the collar region is an arbitrary process. While we will know exactly the object boundaries for polygonal simplification, this collar region is vital in bounding the locality of simplifying transformations.

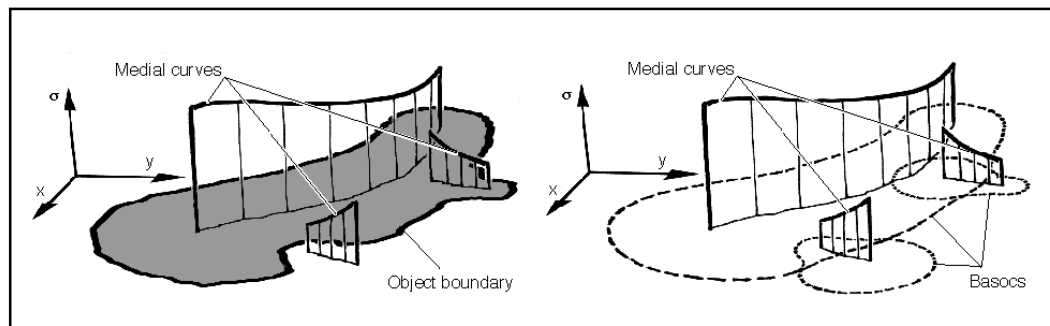


Fig. 2.7 Scale in Core Representation

Multiscale medial model of a 2D object. Height above the object indicates a medial point's scale. The left diagram shows the object and its medial curves. The right diagram shows the medial curves and representative boundaries. (By Andrei State, University of North Carolina.)

Cores represent form in a multiscale manner, but their real power is that they are based on a model of perception that embodies many important perceptual phenomena. Cores naturally reflect the medial and scale information of objects. Local operations are enforced through the finite measuring apertures and by restricting operations to a region proportional to a Core's collar. Symmetry around the medial loci does tie together non-local boundary points, but in a specific way that can be considered local to the medial locus. Finally, Cores segment complex objects into simple parts. Each individual Core defines a part of an object; these parts are called *figures*. The relative scales of two Cores defines their hierarchical relationship; smaller scale Cores are children of the larger-scale ones.

For all their perceptual and representational advantages, Cores do have drawbacks. They are expensive to compute, there is no information about how neighboring figures are attached, and Core-finding methods can fail when two figures are very close together. In this research, I use the Blum MAT as my underlying object representation but perform operations on it that mimic the figural properties of Cores.

2.3. Object Simplification

The idea of simplifying polyhedral models to render them more quickly has been in the graphics community for over two decades [Clark76]. Many different algorithms have been developed to approximate such models with fewer polygons, and two distinct applications of this idea have emerged. The first one is used for over-sampled objects, such as those produced by laser scanners or the Marching Cubes algorithm. Simplifying these models reduces to the problem of creating a new polygonal mesh that is a very close geometric approximation to the original mesh but requires far fewer polygons, since the original model was so heavily sampled.

The second application is interactive rendering such as in virtual reality, where the primary goal is to produce models that can be rendered at interactive frame rates; within that constraint, the simplified model must also look as similar as possible to the original. In this application, if a model looks similar to the original, it can have significant geometric deviations from the original and still be perfectly suitable for rendering. My work focuses on this second problem, where the important measure of success is how closely we can match a user's perception of a simplified model compared to the perception of its fully-detailed original.

For interactive rendering, fully-detailed models are typically transformed into a series of progressively simplified models, each with less detail and fewer polygons. Each of these models in the series is called a *Level of Detail (LOD)* and represents the best possible approximation to the original object for that given number of polygons. As an object moves away from the viewpoint during rendering, it becomes smaller and covers fewer pixels. At this point, an LOD with fewer polygons can be substituted for the more richly detailed one with little if any visual impact. There are times, too, when the scene contains too

many objects with too many polygons, and the only way to maintain interactive frame rates is to introduce lower-detailed LODs even if the change is noticeable. Rendering systems detect when to switch LODs and implement the blending between them to minimize any discernible “popping” when the new LOD is introduced [Funkhouser92] [Luebke97] [Hoppe97].

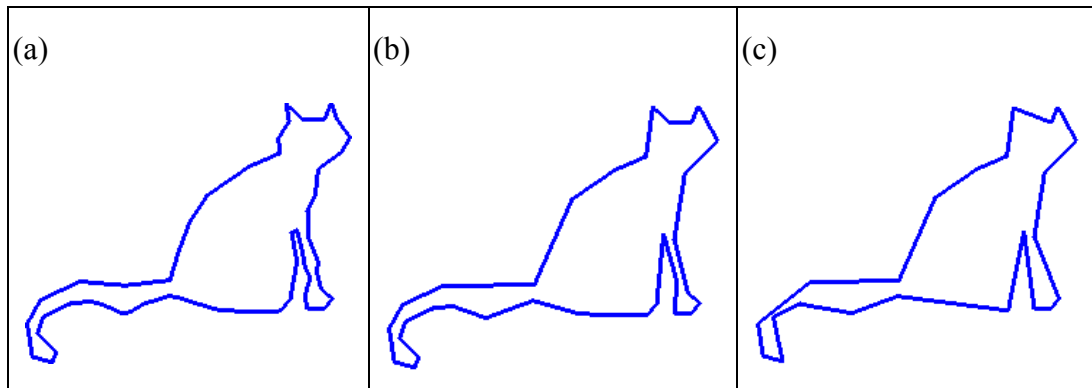


Fig. 2.8 Example of Levels of Detail (LODs)
A 2D object and two simplified levels of detail.

There are several important measurements that simplification algorithms use, as well as several operations employed to simplify polygonal meshes. Also important are the methods they use to order the simplifying operations, since this ordering determines which form cues are present in which levels of detail.

2.3.1. Metrics

The purpose of creating LODs is to reduce the number of polygons that must be rendered while maintaining the object’s appearance as much as possible; in order to do this, algorithms need a way to measure the amount of simplification as well as the degree of similarity between models. The measure of simplification that is used universally is the number of rendering primitives in a model. Rendering speed is almost always directly proportional to the number of polygons being displayed, so counting polygons to measure the degree of simplification is an effective metric. My work focuses on simplifying the perceptual form of objects rather than their representation. The simplification metrics described in Chapter 5 are perceptual metrics that are unrelated to polygon count, but ulti-

mately the simplified objects are converted to a simplified polygonal form where polygons are counted.

A measure of the degree of similarity between an original model and a simplified LOD is needed to preserve the look of the model at each successive approximation. Many metrics have been used for this measure. A very common means of measuring similarity is to use the geometric error between the original model and the approximating LOD. One way to do this is to measure the distance of each vertex in the new LOD to the original surface; the largest of these is the maximum error in the approximation [DeHaemer91]. A refinement of this approach is to find the maximum deviation of any point in the new LOD from the original surface. This has been implemented as a quadric measurement in [Garland97], a geometric norm in [Schroeder92][Varshney94] and [Cohen96] as well as an energy term incorporating such geometric terms in [Hoppe93], [Hoppe96] and [Popovic97].

The problem with using a geometric measure to monitor similarities among models is that the actual goal for LODs is to produce simplified models that *look* similar to higher-detailed models. Measures of geometric deviations do not necessarily capture the visual similarities between two objects. While two models with very small geometric differences generally will look very similar, a highly simplified model with large geometric errors may still contain all the relevant perceptual information from the original object.

Many algorithms have addressed this issue by using geometric measures to either estimate the visual similarity or to guide the simplification in a way that preserves the important similarities. [Schroeder92] identifies sharp edges and preserves those as important perceptually. [Rossignac93] preserves points that may likely lie on a silhouette and also tries to remove points that lie in mostly flat regions. [Turk92] suggests that regions with higher curvature are more important perceptually and uses more polygons to represent those areas. [Hoppe96] and [Popovic97] have added analysis of non-geometric shading information such as color and rendering parameters across a surface. They identify discontinuity curves across an object where either there is a boundary or there is a discontinuity in shading information, and their algorithms preserve these curves as much as pos-

sible. [Cohen98] also preserves shading information, minimizing deviations in surface attributes such as normals and color.

All the metrics described so far have a geometric basis, but there are some researchers beginning to apply perceptual work to object simplification. Reddy describes a visual metric that describes the degree of visual detail that a user can perceive in a computer generated image, based on the images spatial frequencies [Reddy96a], and in [Reddy96b] he describes a simplification method that is designed to remove high-frequency components of an object before small scale ones and to limit the region of simplification based on scale. [Rushmeier95] explores several perceptual metrics for images, and [Ferwerda96] [Bolin98] and [Pattanaik98] describe detailed metrics for image quality. [Ramasubramanian99] present a metric for predicting the perceptual threshold for detecting artifacts in scene features. This measure is based on models of our sensitivity to background illumination, spatial frequencies and contrast levels. [Watson01] performed user studies measuring the visual fidelity of LODs produced by two simplification methods, and then compared the predictions of three computed metrics against these results.

[Lindstrom00] offers a comprehensive simplification algorithm using a perceptual measure of similarity. In this work, images are rendered for each potential simplification operation and these images are compared against images of the original model. Using an image-based similarity metric, simplification operations are chosen to have the least visual impact. The method in [Lindstrom00] captures visual deviations as measured in the image plane. It can be adapted to use any image based measure.

My research differs from these works on perceptual metrics. While all the previous work is rooted in image-based measures, in this work I take a form-based perceptual approach throughout the entire simplification process. I propose new form-based perceptual metrics that can be measured directly from the underlying models.

2.3.2. Operations

There are four classes of simplification operations that have been used to simplify polygonal models. Many algorithms use more than one type of operation. The first class,

adaptive subdivision, was proposed by [Clark76] and is used in [DeHaemer91], [DeRose94] and [Certain96] Adaptive subdivision is actually the inverse of a simplifying operation, since each step actually adds detail to an approximation. In this approach, a very simple, coarse model (even a single polygon) is used to approximate the object. The model is then subdivided, with each smaller part fitted more closely to the original surface. This subdivision continues recursively until the object is represented to within the desired tolerance in each subdivided region.

Resampling is another approach to simplification, where a new set of vertices is selected to represent the object, and these new points are tessellated to create a new polygon mesh. Some algorithms select the new vertices all at the same time, as in [Turk92] and [Hoppe93]. Others select new vertices iteratively, as in [Rossignac93], where several old vertices are replaced by a single new one. The second method of [Cohen96] also adds new vertices by iteratively adding new triangles.

One of the original methods of object simplification employed image-based resampling to create texture LODs [Schachter81]. In recent work, He et al. perform image-based sampling to create new geometric LODs [He95]. In their work, they sample a geometric model into a 3D volume representation, low-pass filter the volume into mult-resolution volume buffers and then recreate a geometric model from the volume representations.

A very popular method for performing the simplifying operations is to iteratively remove geometry from a model. Some algorithms select a vertex to remove from a list of vertices, as in the first method of [Cohen96] and in the second step of [Turk92]. This ends up removing a vertex and all of its attached edges, leaving a hole to retessellate. Other methods collapse the two vertices of an edge [Hoppe93][Hoppe96][Popovic97] or any two vertices [Garland97] into a single vertex, effectively removing a vertex and several edges. [Erikson00] extends this to allow merging the geometry of different objects in order to more efficiently simplify scenes with many objects.

Simplification operations in the final class convert a polygonal mesh to a multiresolution representation. In these representations, models are stored as a very coarse approximation along with a series of detail-adding perturbations. [Certain96] and [DeRose94] used a

global multiresolution analysis to convert models into a wavelet representation, while [Hoppe96] designed a new representation called Progressive Meshes that was generalized in [Popovic97]. They used geometry-removing operations to create their representations. The advantage of these multiresolution representations is that many different LODs can be quickly generated from and compactly stored in a single representation, compared to traditional methods in which each LOD is stored separately as an entire model.

This work presents another class of simplifying operations. The operations that are presented here are transformations that alter the perceived form rather than the polygonal format of the original representation. These shape transformations deform the object in a way that minimizes a certain perceptual complexity measure while preserving an overall saliency measure. The perceptual effect they have is to smooth figures, combine them or completely remove them from the object. Geometric operations are used to implement the shape transformations, but it is the perceptual aspects of the operations that are important. The geometric operations are used only to realize the shape alterations. The objective here is not to directly produce a more efficient representation, and in fact the perceptual simplifications greatly increase the number of primitives required to represent an object. Instead, my goal is to produce objects that are simpler perceptually, with the hypothesis that these can then be converted to a representation with a reduced number of primitives.

2.3.3. Scheduling

The final important aspect of simplification methods is how they schedule their simplifying operations. One scheduling method is to randomly choose the ordering of operations. This approach is used in many methods [Turk92] [Schroeder92] [Hoppe93] and the local, vertex-removal method of [Cohen96]. Other methods create a prioritized ordering of either the operations to perform [Garland97] [Hoppe96] [Popovic97] or of the geometry on which to perform them [Cohen96] [Rossignac93]. The wavelet analysis of [DeRose94] and [Certain96] offers a global analysis that essentially orders the operations by their location in the multiresolution hierarchy. My research provides a global analysis and ordering of simplifying operations by associating a measure of perceptual importance

with every point on an object, and then applying simplifying operations at the points with the lowest ranking.

This work creates a new metric to determine the similarity between LODs. It is a perceptually-based measure of visual significance that is used to rank the visual importance of every point on the object. Simplifying operations are scheduled according to this visual significance; areas on the object with the lowest visual significance are simplified first.

The simplifying operations are actually shape transformations. Instead of creating a new and simplified object representation that has fewer primitives, my simplifying operations remove perceptual cues from the object.

CHAPTER 3 FORM AS SUBSTANCE AND CONNECTIONS¹

3.1. Introduction

The lure of a form representation that inherently reflects the perceptual qualities of form has long attracted scientists. Blum proposed his Medial Axis Transform (MAT) as a representation that embodies the skeleton of an object as well as the width of the object at every point on the skeleton [Blum67]. This work has spawned a tremendous amount of research into the use of the Blum MAT and other skeleton representations [Blum78] [Brady84] [Bruce85] [Ogniewicz92] [Pizer87] [Pizer98] [Székely96]. A goal of much of this work has been to create a form representation that defines a natural decomposition of an object into a set of basic parts that mirrors the object parts we perceive. At the same time, these representations are intended to allow easy access to the full form information about each part and about the object as a whole to support form analysis and computation.

A driving consideration faced by all skeleton methods is the condition put forth by Marr and Nishihara that small changes to a boundary's form should cause only small changes to the form's representation [Marr78]. This rule insists that skeletons remain stable under boundary change, and considerable effort has been spent on modifying Blum's MAT to exhibit this kind of stability [August99][Székely96]. Complicating this rule is the problem that some boundary perturbations actually are important; the growth of a new tail or limb from a body, for instance, can be a small boundary change that has a large percep-

¹This chapter has been submitted for publication as *Untangling the Blum Medial Axis*, by Robert A. Katz and Stephen M. Pizer. In order to provide a self-contained description of the material, this chapter has not been edited to remove redundancies with other chapters.

tual importance. Significant boundary perturbations correspond to the parts of an object; examples are the fingers on a hand, legs on a table and limbs on a body. In order to decompose an object into its parts, any general form description needs to be able to make a stable and clear distinction between an insignificant bump on the boundary and a separate and significant part of the object.

Used as originally proposed, the Blum MAT falls short of each of these goals (see Fig. 3.1). It is notoriously unstable since even tiny perturbations can cause the addition of very long segments with prominent radius values to the MAT. In fact, instabilities in the MAT can arise even under boundary smoothing [August99]. Furthermore, the MAT's network of branching axis segments quickly grows as object complexity grows, causing difficulty in creating a MAT hierarchy that reflects an object's perceptual parts. Without a clear hierarchy, there is no distinction between object parts and insignificant object protrusions.

This paper offers a new method for using the Blum MAT that does not suffer from these problems. I argue that objects are an assemblage of solid parts and that form representations should explicitly separate the substance of each part from the connections between parts. The method proposed here adds a component to the MAT that describes for every point on the MAT the degree to which the point embodies object substance as opposed to object connection. I call this measure the *substance* value. The measure easily reveals the natural distinction between the parts of an object and mere bumps, and it is a continuous measure that reflects the naturally ambiguous part classification found in many objects. The new method also reduces small boundary perturbations to small substance changes in the new weighted representation, maintaining a stable form description even with changes to the boundary. The new method permits a substance-weighted application of the MAT where the MAT pieces that cause instabilities in the tradition Blum representation are weighted by small values of this new substance measure. In this way the method presented here maintains a stable form description even with changes to the boundary.

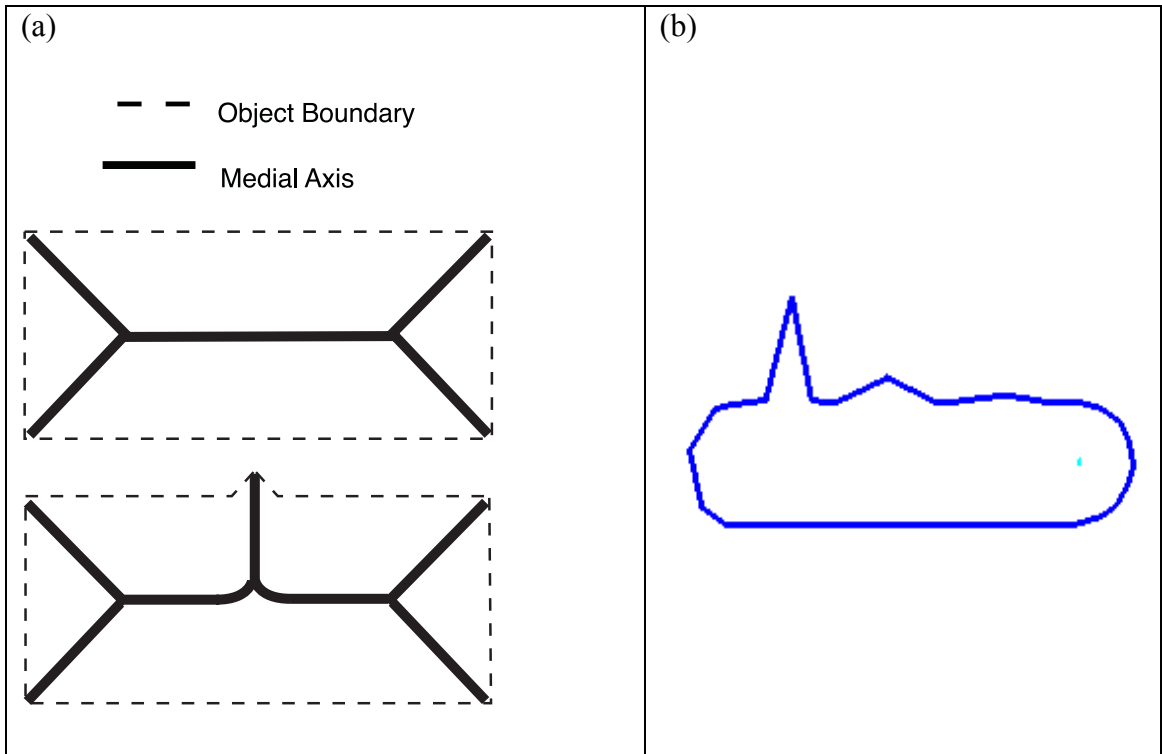


Fig. 3.1 Challenges with using the Blum MAT

- (a) **MAT instabilities: A tiny change in the boundary produces a large change in the MAT.**
- (b) **Part/Protrusion ambiguity: Classifying a protrusion as a bump or separate part can be ambiguous.**

While this chapter describes how to instrument the Blum MAT to separate substance and connection, I anticipate that these ideas are applicable to the general class of skeletons and medial form representations.

3.2. Objects are Substance and Connections

Objects in our physical world are solid, tangible entities. They have mass and fill space. Physical objects are real substance and not mathematical abstractions. Complex objects are perceived as collections of many solid parts. A boulder is a round-ish blob, and hands are a blob (the region between the wrist and the fingers) with five elongated protrusions (the fingers).

For multi-part objects, we perceive not only the substance of each part but also the connections between parts. A hand is a hand not just because it has a palm and five fingers

but because of the way that the parts are connected. The same five parts can be connected in ways that are not at all hand-like. Clearly, both the substance information in each part of an object and the connection information between parts are crucial components of form description.

Another type of connectivity is also implicit in skeleton methods. If we consider each part of an object as a series of tiny slices, then each slice is connected to its neighboring slices to form the continuous substance of the part. This *continuity* information is usually built into the representation used for a skeleton; for the piecewise linear boundaries used in this work, Blum's MAT uses line segments and parabolic segments to represent the skeleton, and the continuity along each of these segments is implicitly defined. The distinction between continuity along a single part and connections between parts is exploited in this paper.

Analyzing and computing with form requires access to substance information and connection information. Form researchers want to be able to work explicitly with the substance information for calculations involving the perceptual properties of an object such as saliency, and they want to work explicitly with the connection information for analyzing the decomposition of objects into parts and analyzing the relationships between parts. Medial form representations naturally include substance information. By representing objects as "middle" points and the "widths" at those points, medial representations define an object from its inside outward thus implicitly defining a solid object. Medial representations also can include connection information. In the same way that we can trace our own bones to know that our legs are connected to the bottom of our torso, medial representations can naturally define the connectivity of an object.

While medial representations are natural carriers of substance and connection information, many implementations do not harness the full potential of the two components. For example, medial methods such as SLS [Brady84] and Cores [Pizer87] provide substance information but have no explicit connection information. Fig.3.2 shows how the Blum MAT contains both substance and connection information, but they are not easily separable for creating parts-hierarchies or for form computation. The next section describes

what it is about the standard approaches to using the Blum MAT that fails to discriminate between substance and connection, and section 3.4 presents a method that successfully separates the two form components.

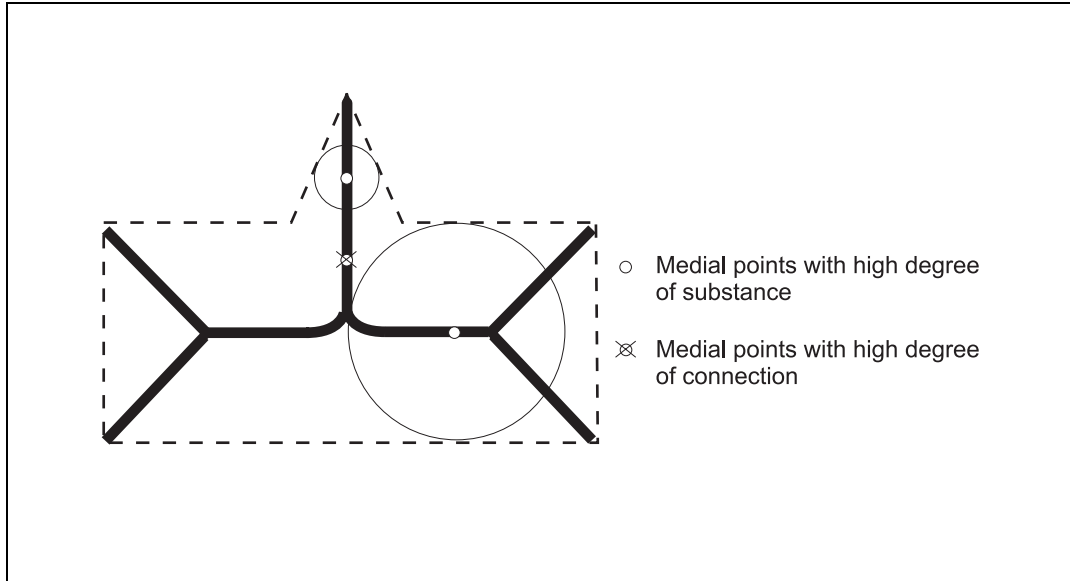


Fig. 3.2 Examples of Medial Points

Examples of medial points with high degree of substance information and with high degree of connection information. Also shown is a perceptual aperture around the substance points, showing the aperture within which computations are performed.

3.3. Parts Hierarchies and Form Information in the Blum MAT

In order to solve the problems inherent in the Blum MAT, we must understand their causes. A complex object with many parts and many bumps generates a MAT with a complex web of axes. There are far more branching axis segments than the number of parts in the object, and creating a parts-hierarchy requires deciding which axis branches correspond to part substance and which to part connections. Previous hierarchy methods have attempted to impose a binary hierarchy, examining every axis point to determine if it belongs to an object part, either with its neighbors in a single part or on a new object part that is different from its neighbors, or if it should be pruned away and ignored as merely connection information. Shaked and Bruckstein give a good overview of pruning methods [Shaked98]. Fig. 3.1a shows how using binary decisions to create parts-

hierarchies is highly unstable. A small change to the boundary can cause a drastic change in the topology of the hierarchy, and, with binary classifications, insignificant boundary bumps can change to significant parts as a consequence of the slightest boundary perturbation.

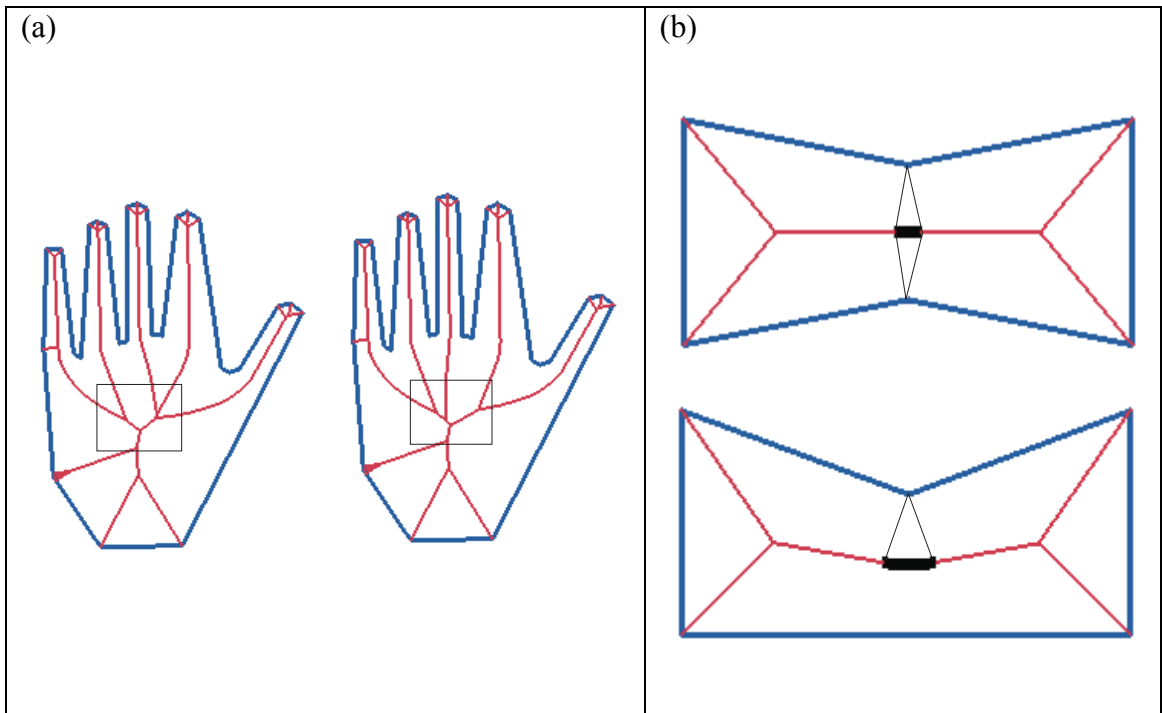


Fig. 3.3 Instabilities in Blum MAT-based Part Classification.

Unstable part-hierarchies caused by binary hierarchy decisions. A slight change to the boundary between the index and middle fingers causes a drastic change to the MAT topology in the middle of the palm (within the rectangles). The thick lines are examples of ligature (top) and semi-ligature (bottom).

The instabilities in the Blum MAT are due to the fact that the MAT is not a one-to-one correspondence. As Fig. 3.3b shows, a single point on the boundary can generate many medial axis points. Thus, changing a single boundary point can cause many axis points to change, creating the instability. These many-to-one axis points were identified by Blum as *semi-ligature* and *ligature* (see Fig. 3.3). August et al. nicely describe the instabilities caused by ligature [August99]. Their work reveals that some ligature points are purely connection information and some contain substance information as well. They use a tertiary classification based on the geometric heat equation to identify axis points with substance-like information, and then they cull the ligature that roughly corresponds to what I call connection information; they propose that the culled ligature are the points that cause

instabilities. However, their classification still requires explicit decisions to be made at every point, and in this way they encounter the same problems as a binary part-hierarchy.

Furthermore, some objects have a parts-hierarchy that is naturally ambiguous. Fig. 3.4 shows progressive deformations of an object that starts with one clearly defined bending part and ends with three clearly defined parts. However, there is no exact point in the deformation where the object changes from one part to three. Instead, as the deformation progresses, the single main part exhibits less and less continuity in its substance and more and more connection among distinct parts. The actual part classification is perceptually ambiguous for the middle configurations of the deformation. This example demonstrates the perceptual continuum from substance to connection that must be addressed in any part-classification approach.

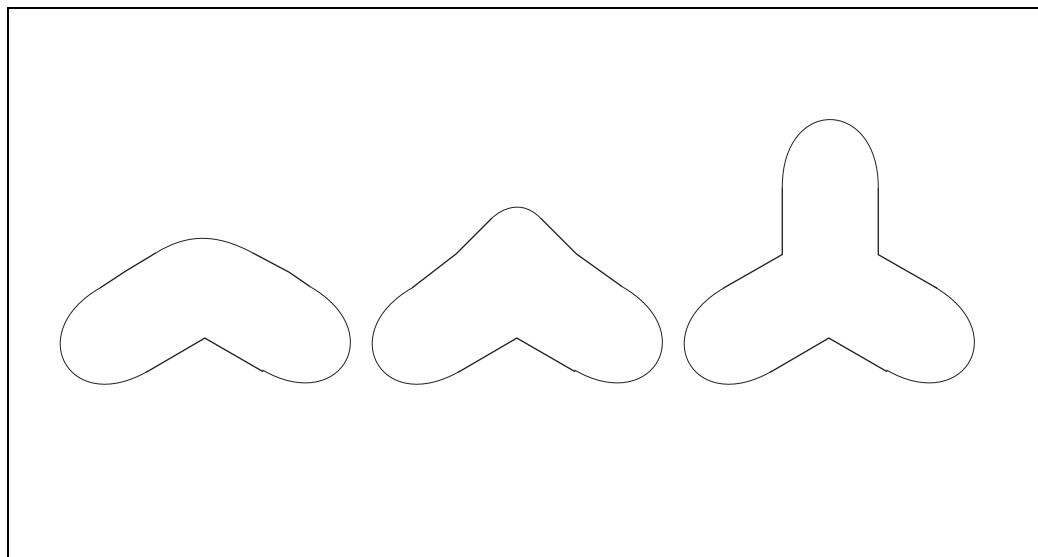


Fig. 3.4 Ambiguous Part Classification

Part classification that is ambiguous under deformation

Based on this line of reasoning, the creation of perceptual parts-hierarchies can be resolved by first extracting explicit substance and connection information from the MAT and then using these components to create a fuzzy parts-classification that allows for ambiguity. This allows the inherent form information to be used for parts-classification and

form computation. By using these measures of substance and connection, the underlying instabilities of the MAT are not changed but their effects are eliminated.

3.4. Calculating Substance and Connection

In order to create a natural parts-decomposition, this method defines a measure that separates substance and connection components for the Blum MAT. Every axis point will be considered to have these two components, and a measure between 0 and 1 is calculated for each. Neighboring axis points that have a substance component that is much greater than their connection component are continuous points on the same part of the object. Axis points with a high connection component and low substance component are connection points that separate different parts of the object. Axis points with similar substance and connection measures represent regions of the object that are perceived to have an ambiguous part classification.

In this work, the two components are defined to be directly related; if ω is the connection measure, then substance measure ψ is $1 - \omega$. In this way every point in the object is a blend of substance and connections, and the degree of each component can be represented with a single value. In the remainder of this section, I will develop the calculation for the connection measure ω , but I will refer to both the substance and the connection components of an object. The end result will be a new component added to the Blum MAT representation. Instead of every medial point having just location and scale components, (x, y, σ) , every point will be a 4-tuple, (x, y, σ, ψ) .

As described in Chapter 4, perceptually based form calculations at a medial point should be integrated over an aperture centered at that point. This idea captures the effects of the receptive fields in our visual systems that operate at a scale that depends on the medial width of the object at the point of interest [Burbeck96] [Kovacs94]. According to this model, the size of the aperture is directly related to the radius of the maximal circle at the medial point, and calculations within the aperture are weighted with a Gaussian fall-off to give points closer to the center point a higher influence than more distal points. All cal-

culations to produce the connection measures for medial points are integrated over and performed within such an aperture. Fig. 3.2 shows a perceptual aperture with RMS width that is equal to width of the maximal circle around a medial point.

The key to calculating the connection measure for the Blum MAT is noticing that branches are the crucial points in determining the connection and substance components of an object. A single axis representing a simple object would be fully substance with no connection information. This reflects the intra-part continuity along the axis; there are no other connected parts and thus there is no connection information. However, when there is a branch point with three incident axes, we must face the question: “which two axes combine to form the ‘main body’ of the object, and which remaining axis reflects a bump on the object or an entirely new part?” Or, in other words, “what are the connection and the substance components of the axes around this branch point?”

Any answer to this question must reflect the notion that our visual systems will tend to follow the main medial path of an object while ignoring less significant off-shoots. This continuity of a visual path is determined by many factors; two of the most important factors are the continuity of the direction of the medial axes and the continuity of scale along the axes. In this work, *visual conductance* is developed as a measure of how likely any two of the axes at a branch point will be perceived as the single, main perceptual path. In the remainder of this section, the properties required for useful perceptual metrics such as visual conductance are presented, and then functions, sometimes ad-hoc ones, are offered that satisfy these properties.

3.4.1. Visual Conductance

To measure the likelihood that we will perceptually connect two axes at a branch point, visual conductance is defined to be a relative measure that compares, at a branch point, each pair of the branch’s axes. When one pair has a high conductance while the other two pairs are low, the first pair is perceived as the main visual path through the branch, and therefore that path is perceived as mostly substance while the third axis is perceived as mostly connection information. When two or all three pairs have similar values of visual conductance through the branch point, there is perceptual ambiguity about which

branches are the main visual path, and therefore all branches may reflect a significant amount of both substance and connection information.

Visual conductance should capture the visual continuity provided by the continuity of direction along two branching axes, and it should capture the visual continuity provided by similar medial widths across the medial axes. In order to calculate visual continuity through a branch according to these notions, a “visual vector” is first calculated for each axis emanating from the branch (see Fig. 3.5), and then these vectors are compared to produce the visual conductance measures at the branch. Visual conductance of any axis pair through a branch point is defined to be a value between 0.0 and 1.0, and the conductance at all non-branch points is defined to be 1.0. The following develops further this conductance measure for axis pairs through branch points.

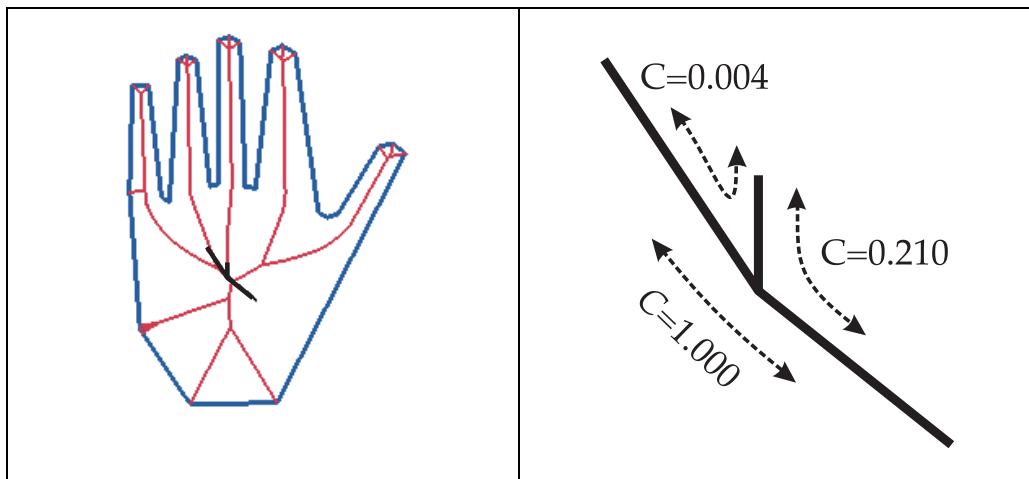


Fig. 3.5 Visual Vectors and Visual Conductance

- (a) Hand object with its Blum MAT and visual vectors at the branch where the middle finger's axis joins the other axes. Longer vectors represent greater accumulated visual scale in that direction.
- (b) Visual conductance values for each pair of axes represented by the visual vectors in (a)

The visual vector for an axis emanating from a branch point is defined to be a weighted directional average of the tangent of the medial axes A_i starting at the branch point and following that branch axis outward through any neighboring branches. To be consistent with the scale-based medial view, this directional average is weighted by the width $\sigma(s)$ of every axis point s on the averaged path. In this work, the medial width $\sigma(s)$ is defined

to be the Blum MAT medial radius R , that is, the radius of the maximal disk centered at the medial point s . The final magnitude of the vector reflects the accumulated scale of the object along the vector. The calculation is performed within a perceptual aperture of radius $\sigma(s_b)$ around the initial branch point b , and closer axis points are given a higher influence than more distant ones using a Gaussian fall-off $G(s)$ with standard deviation $\sigma(s_b)$. Finally, since the scale of all axis points around a branch point are nearly the same (and are exactly equal at the branch point), a term $\Phi(s)$ is added to lessen the influence of axis points very close to the branch.

For a generic branch point with three intersecting axes, there is a visual vector associated with each of the three axes. In summary, each vector includes in its calculation all the axes that can be reached by traversing the MAT, starting along axis A_i and emanating away from the branch point b , traveling along the axes in a depth-first manner up to a distance of $\sigma(s_b)$.

To incorporate all of these requirements, the visual vector $\bar{\mathbf{v}}$ along axis A_i emanating from branch point b is defined as follows. Here, the functions are parameterized as the arc-length distance from branch point b .

$\bar{\mathbf{u}}(s)$ = tangent vector at a medial point s

$\sigma(s)$ = radius of maximal circle at a medial point s

s_b = arc length parameter at branch point b

$\sigma(s_b)$ = radius of maximal circle at branch point b

A_i = the i 'th axis emanating from the branch point b

$$G(s) = e^{-\frac{1}{2}\left(\frac{t}{\sigma(s_b)}\right)^2}$$

$$\Phi(s) = 1 - e^{-\left(\frac{t}{\sigma(s_b)}\right)^2}$$

$$\bar{\mathbf{v}} = \int_{A_i}^{t=\sigma(s_b)} \bar{\mathbf{u}}(t)\sigma(t)[G(t)][\Phi(t)]dt$$

With visual vectors defined for each axis emanating from a branch, the vectors can then be analyzed in pairs to determine the visual conductance between each pair of branch

axes. Our perceptual systems use a combination of factors to determine how we perceive an object at its medial branching points. Visually, we will tend to follow along medial axes that provide a straight path rather than a curved path; at the same time will tend to follow axes that have the most similar scales. For a branch point b , the visual conductance measure incorporates these phenomena; visual conductance C between branching axes A_i and A_j at branch b is defined as follows.

$$C_b^{continuity}(A_i, A_j) = \left[1 - \left(\frac{\bar{v}_i \cdot \bar{v}_j}{|\bar{v}_i| |\bar{v}_j|} \right) \right]^2 \left[\min \left(\frac{|\bar{v}_i|}{|\bar{v}_j|}, \frac{|\bar{v}_j|}{|\bar{v}_i|} \right) \right]$$

The first term captures the continuity of direction using the dot product function, and the value is squared as an ad-hoc means to emphasize the continuity along straight paths and decrease continuity when the path bends. The second term captures the continuity of scale by comparing the magnitude of the two branches under consideration and then keeping the comparison value less than one.

There are three measurements $C_b^{continuity}(A_i, A_j)$ for every generic branch point, with one for each pair of the three branching axes. If I had been required to deal with the non-generic case of more than three axes meeting at a branch point, I could have handled that by dividing the branch into two or more very nearby generic branches.

To use this measure, the visual conductances for each pair of axes at a branch point are compared. The two axes that have the highest conductance are chosen as the “main visual path”, and their conductance is scaled to be 1. The conductance of the other pairs are then scaled by the same amount to maintain their relative magnitudes but to have a value less than or equal to 1. In this way, the main visual path through a branch point fully “conducts” perceptually while other paths with conductances less than 1 “attenuate” visual flow.

For example, in Fig. 3.5, the conductance between the left-most and right-most vectors is the highest, so it is scaled to be equal to 1. The conductance between the left-most and middle vector and the conductance between the right-most and the middle vector are both

scaled by the same amount. As Fig. 3.5 shows, while there is full visual conductance between the axes represented by the left-most and right-most vectors, the conductance is attenuated between each of those axes and the axis represented by the middle vector.

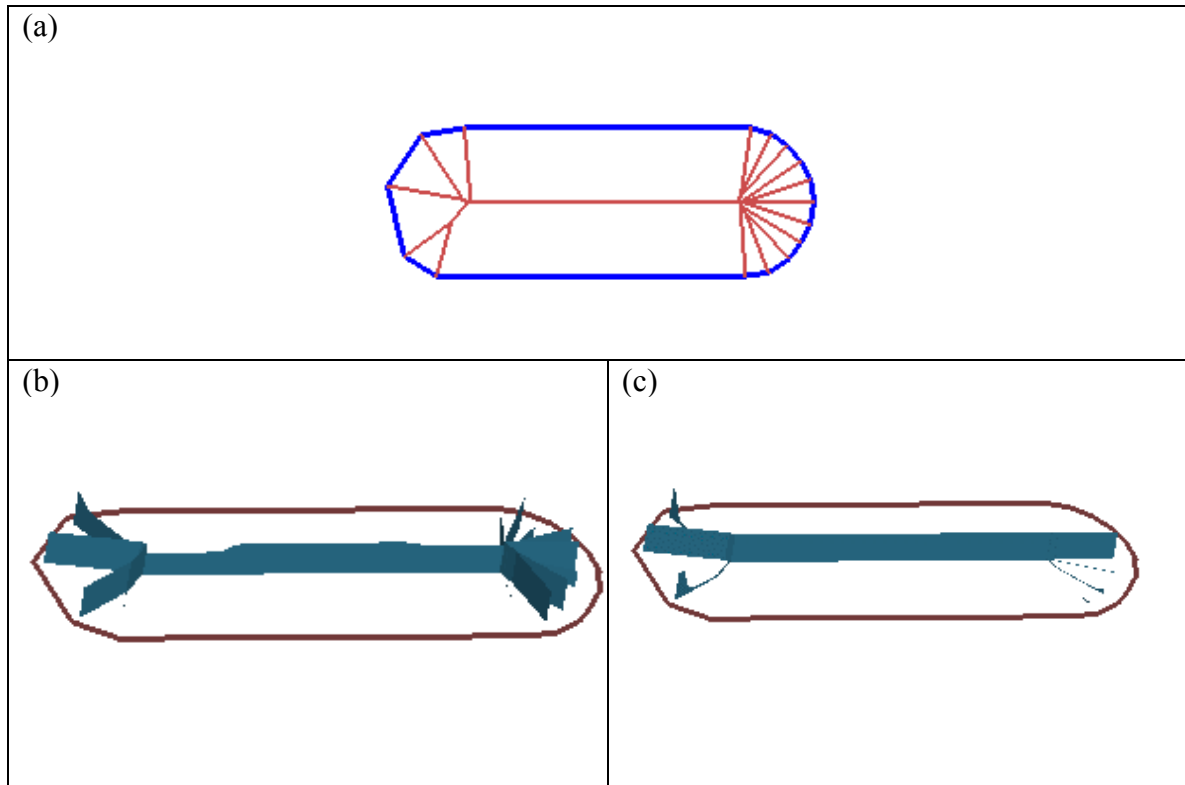


Fig. 3.6 Part-end Adjustment to Visual Conductance

- (a) Rounded-end object with its Blum MAT.
- (b) Substance measure of an object with only continuity component of visual conductance. Height at each medial point is the substance measure at that point.
- (c) Substance measure with endness adjustment added to visual conductance in order to capture the visual effects at the tip of a part.

3.4.2. Part-end Adjustment

While this formulation for visual conductance applies to general branch points, it fails when an object part ends in a curved tip (see Fig. 3.6). The MAT in these regions consists of a large scale axis segment branching into segments with much smaller scales. As defined so far, visual conductance would recognize the small scale segments at such

branches as most similar in scale and connect them as the main visual path unless they meet at a very sharp angle. Clearly, this is not perceptually correct. Visually, at such a branch point the large scale segment appears to connect with one of the small scale ones to become the main visual path through the branch, and the remaining segment is seen as a connecting bump.

In order to implement this, I leverage the understanding in human vision that ends of object parts are identified by end-stopped cells independently from the identification of other features [Orban79a] [Orban79b]. In this work, an endness factor is calculated separately for each axis at every branch point. *Endness* measures the discontinuity in scale between one axis and its other two branch axes; in other words, the endness factor represents how much larger in scale each axis is compared to the other two neighboring axes. The largest endness factor of the three axes becomes $endness_b$, the degree to which the branch point b will be perceived as an ending of an object part.

Endness is used in formulating visual conductance to signal branch points in which an axis with a high degree of endness should be kept as part of the main visual path through the branch even when the previous conductance calculation indicates that it should not be. When this is the case, the large scale axis is combined with one of the other two smaller scale axes to form the visual path, and is given an endness-conductance $C_b^{endness}$ that is equal to 1. Then the remaining axis is considered an off-shoot of that path and is given endness-conductances that are calculated and scaled as in the non-endness case.

If the degree of endness is very low at a branch, the visual conductances at the branch should not be changed by the endness calculations; if endness is very high, the branch's conductances should become these alternative end-condition conductances. For cases in between where there is perceptual ambiguity about whether a part is ending at that medial point, the visual conductance should be a blended combination of the two conductance values.

The *endness* measure and final visual conductance are calculated at branch point b , following the path between branch axes A_i and A_j , as follows. The new variables and parameters are described in the next paragraph.

$$\lambda = \max \left[\min \left(\frac{|\bar{v}_i|}{|\bar{v}_j|}, \frac{|\bar{v}_i|}{|\bar{v}_k|} \right), \min \left(\frac{|\bar{v}_j|}{|\bar{v}_i|}, \frac{|\bar{v}_j|}{|\bar{v}_k|} \right), \min \left(\frac{|\bar{v}_k|}{|\bar{v}_i|}, \frac{|\bar{v}_k|}{|\bar{v}_j|} \right) \right]$$

$$endness_b = \left(1 - e^{-(\lambda)^p} \right)^n$$

$$C_b(A_i, A_j) = endness_b C_b^{endness}(A_i, A_j) + (1 - endness_b) C_b^{continuity}(A_i, A_j)$$

Recall that the endness adjustment is intended to modify the visual conductances at the MAT branches at an end of a part. The initial factor λ is used to detect the part-end situation by comparing branch-axis scales at a branch to determine the greatest mismatch. The ad-hoc formulation for the final $endness_b$ factor is designed to create a high threshold that the end condition must meet before it influences visual conductance. The values of p and n are used to adjust this threshold, with p typically set to 4 and n typically set to 2. When a part-end situation is indicated by a non-zero value of the $endness$ measure, the alternative visual conductance $C_b^{endness}$, which is always 1.0, is factored in to the total visual conductance between two branch axes.

In addition to the endness adjustment, the measures developed here must also account for multiple branches, as developed in Section 3.4.3. For a simple object with a single branch, if a point lies close to the branch and is on the main visual path through the branch, then that point has a high substance component. Similarly, if the point lies on an axis whose conductance to the other two axes is much less than the conductance of the main path, then that point has a high connection component. However, this classification fails when multiple branches are close together. Each branch point may influence the substance/connection classification of neighboring axis points. For example, in Fig. 3.7c the axes labeled A_2 and A_4 around branch b_2 would have a substance measure of 1 if b_2 were the only branch in the object. However, because of b_2 's proximity to branch b_1 , A_2 and A_4 are low in substance and are mostly connection information as shown in Fig 3.7b. When many branches are close together, each branch casts a shadow of influence around itself that affects the calculation of nearby connection measures.

3.4.3. Connection Shadows

Since the receptive fields in our visual systems process visual input from a region of an object and not at a single point, there may be many medial branches within this aperture of viewing, and each branch can affect the determination of connection and substance components at the focal point. A branch that is close to the point of interest may indicate that the point is on a high conductance visual path, but a branch farther upstream may indicate a much lower conductance path at the same point. More branch points may each indicate a different conductance along the point's path. To combine the effects of many branch points, there needs to be a way to propagate the influence of a branch through other branches and combine the many influences into a single measure. At the same time, any solution to this problem must insulate the main visual paths through an object from branches that are on a path with low visual conductance to a main path. The solution presented in this work is to allow branch points to cast a shadow that covers the local region around the branch. These *connection shadows* propagate the substance-attenuating effect of their branch, where this attenuation results in increased connection measure (and reduced substance measure) at the points their shadow affects. The large circle in Fig. 3.7c illustrates a shadow at branch b_2 , and the following paragraphs describe how the attenuation is calculated.

To implement connection shadows, the attenuating effect cast by a branch along one of its branch axes is 1 minus the largest of the conductances of that axis connecting with any of the other branch axes. In this way, a branch will cast a highly attenuating shadow along its branch axes that are not on a main visual path, and it will cast no attenuation along its main visual axes. For the example in Fig. 3.7, axis A_3 has low conductance with both A_2 and A_4 , so the shadow from b_2 creates a high connection measure on A_3 . At the same time, since the conductance between A_2 and A_4 is 1, there is no substance attenuation cast along those axes by b_2 . The shadow is adjusted by a fall-off from the branch point casting the shadow, with the aperture of the fall-off determined by the scale at the branch point. This provides a high shadowing effect within most of the shadow with a rapid fall-off at the edges of the shadow.

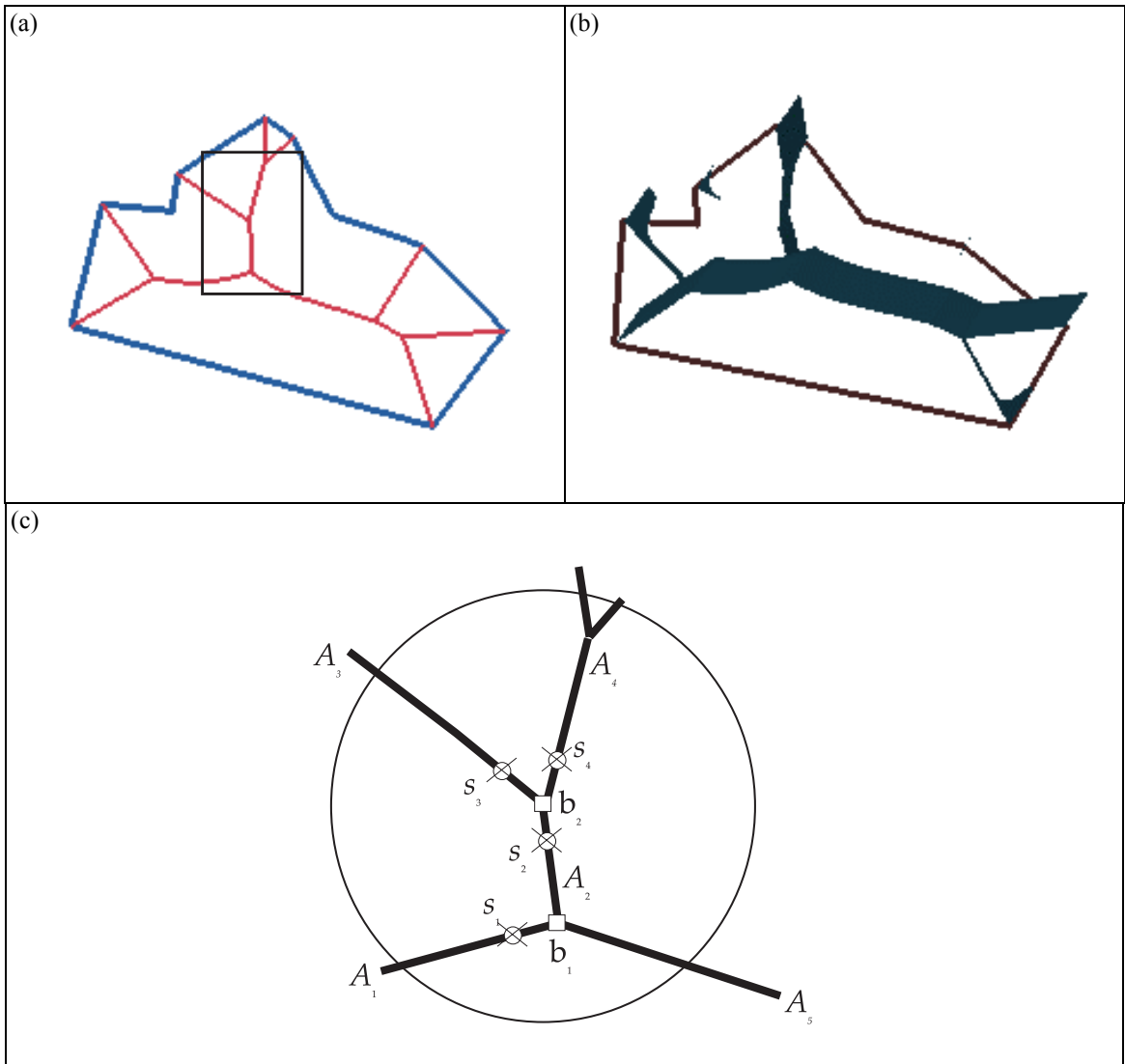


Fig. 3.7 Connection Shadow

- (a) Object with its Blum MAT.
- (b) The substance measure of the object.
- (c) A blow-up of the highlighted region from (a) showing a connection shadow and the axes that it covers. See the text for more details.

The connection shadow is calculated as follows, where n is a parameter used to adjust the rate of fall-off from the branch.

If s is a point on axis A_i , where A_i ends at branch points b_1 and b_2 , and b_j of A_i has neighboring axes $A_{j,1}$ and $A_{j,2}$, then the shadow from a branch point b cast on an axis point s that lies on axis A_i is

$$S_{i,b}(s) = \left[2 \left(\operatorname{erf} \left[\frac{|s - s_b|}{\sigma(s_b)} \right]^n - 0.5 \right) \right] \left[1 - \max_{j=1,2} C_b(A_i, A_{b,j}) \right]$$

This value is clamped to zero if the result becomes negative.

Following the principle that our visual processing occurs within an aperture and not at a single point, connection shadows are propagated out from a branch along each branch axis. A high value for the shadow, indicating high connection and low substance components, will flow along the axis and along all branching axes downstream from the initial branch point. This in effect casts a shadow along those axes, and the shadow is used to determine the connection measure at points along the axes that it covers. With this construction, a branch's effect is felt by all points within its shadow even when separated by other branches.

However, as a connection shadow flows through other branch points its effect should be attenuated by every branch that it flows through. This can be understood by remembering that visually there are the main medial paths through an object and there are sub-parts that branch from the main paths. These sub-parts are separated from the main parts by axes with a high connection measure, and the medial configuration of the sub-parts have little visual impact on the main part. For example, in Fig. 3.7c, point s_l is on the main visual path of the object and its substance measure should not be attenuated by the shadow shown that is cast from b_2 . In general, every branch propagates a shadow according to the conductance of the path the shadow follows through the branch. In this way, branches that are located on a path that is already highly connection information will not cast influence on neighboring branches that are on the main visual path attached to the connection path. In other words, main visual paths through a branch are insulated from connection shadows on attached low conductance paths.

In order to compute shadow propagation, the accumulated conductance $V_{i,b}$ between a branch point b and a point s on axis A_i must be determined. This is the product of conductances encountered through each branch point on a path starting at branch point b and

traversing the medial axes to reach axis A_i . These conductances potentially attenuate the effects of branch point b 's shadow on the points on a distant axis.

Accumulated conductance $V_{i,b}$ is computed as follows.

$\{p_{i,b,m}\}$ for a fixed axis A_i and branch point b is the sequential ordered list of branch points numbered $1..m_{max}$ on the path from axis A_i to branch point b .

$\{k_{i,b,m}\}$ is the sequential ordered list of axes numbered $1..m_{max}$ on the path from axis A_i to branch point b that is associated with $\{p_{i,b,m}\}$.

The *accumulated conductance* $V_{i,b}$ between a point on axis A_i and a branch point b through the path of branch points $\{p_{i,b,m}\}$ and axes $\{k_{i,b,m}\}$ is the product of conductances encountered at each of the m_{max} branch points along the path,

$$V_{i,b} = \prod_m C_m(k_{i,b,n}, k_{i,b,n+1})$$

Finally, the shadowing effect of a single branch point b onto a point s on axis A_i is the attenuation of the connection shadow multiplied by the accumulated conductance between the point and the axis: $V_{i,b}S_{i,b}(s)$.

Combining all of these developments, the final measure $\omega(s)$ of the connection component (and the related substance measure $\psi(s)$) at every point on a Blum medial axis is the combined effects of all branch points b on the point of interest, s . While all branch points are considered, only those whose shadow overlaps the point of interest will have any effect on s . This is defined as follows, with a summary of the terms involved.

$$\text{Substance measure : } \psi(s) = 1 - \omega(s)$$

$$\text{Connection measure : } \omega(s) = \sum_b V_{i,b}S_{i,b}(s)$$

$$\text{Accumulated conductance : } V_{i,b} = \prod_m C_m(k_{i,b,n}, k_{i,b,n+1})$$

$\{p_{i,b,m}\}$ for a fixed axis A_i and branch point b is the sequential ordered list of branch points numbered $1..m_{max}$ on the path from axis A_i to branch point b .

$\{k_{i,b,m}\}$ is the sequential ordered list of axes numbered $1..m_{max}$ on the path from axis A_i to branch point b that is associated with $\{p_{i,b,m}\}$.

The accumulated conductance $V_{i,b}$ between a point on axis A_i and a branch point b through the path of branch points $\{p_{i,b,m}\}$ and axes $\{k_{i,b,m}\}$ is the product of conductances encountered at each of the m_{max} branch points along the path.

$$\text{Connection Shadow} : S_{i,b}(s) = \left[2 \left(\text{erf} \left[\frac{|s - s_b|}{\sigma(s_b)} \right]^n - 0.5 \right) \right] \left[1 - \max_{j=1,2} C_b(A_i, A_{b,j}) \right]$$

This is the connection shadow cast by branch point b on a point s on axis A_i , where A_i ends at branch points b_1 and b_2 , and b_j of A_i has neighboring axes $A_{j,1}$ and $A_{j,2}$.

$$\begin{aligned} \text{Visual Conductance} : C_b(A_i, A_j) &= \text{endness}_b C_b^{\text{endness}}(A_i, A_j) + \\ & (1 - \text{endness}_b) C_b^{\text{continuity}}(A_i, A_j) \end{aligned}$$

$$\lambda = \max \left[\min \left(\frac{|\bar{v}_i|}{|\bar{v}_j|}, \frac{|\bar{v}_i|}{|\bar{v}_k|} \right), \min \left(\frac{|\bar{v}_j|}{|\bar{v}_i|}, \frac{|\bar{v}_j|}{|\bar{v}_k|} \right), \min \left(\frac{|\bar{v}_k|}{|\bar{v}_i|}, \frac{|\bar{v}_k|}{|\bar{v}_j|} \right) \right]$$

$$\text{endness}_b = \left(1 - e^{-(\lambda)^p} \right)^n$$

$$C_b^{\text{continuity}}(A_i, A_j) = \left[1 - \left(\frac{\frac{\bar{v}_i \cdot \bar{v}_j}{|\bar{v}_i| |\bar{v}_j|}}{2} \right) \right]^2 \left[\min \left(\frac{|\bar{v}_i|}{|\bar{v}_j|}, \frac{|\bar{v}_j|}{|\bar{v}_i|} \right) \right]$$

The *endness* measure and final visual conductance are calculated at branch point b , following the path between branch axes A_i and A_j .

$$\text{Visual Vector} : \bar{v} = \int_{A_i}^{t=\sigma(s_b)} \bar{\mathbf{u}}(t) \sigma(t) [G(t)] [\Phi(t)] dt$$

$$\begin{aligned}
\bar{\mathbf{u}}(s) &= \text{tangent vector at a medial point } s \\
\sigma(s) &= \text{radius of maximal circle at a medial point } s \\
s_b &= \text{arc length parameter at branch point } b \\
\sigma(s_b) &= \text{radius of maximal circle at branch point } b \\
A_i &= \text{the } i\text{'th axis emanating from the branch point } b \\
G(s) &= e^{-\frac{1}{2}\left(\frac{t}{\sigma(s_b)}\right)^2} \\
\Phi(s) &= 1 - e^{-\left(\frac{t}{\sigma(s_b)}\right)^2}
\end{aligned}$$

In the example given in Fig. 3.7c, the shadow cast from b_1 has no effect on s_1 because A_1 and A_5 constitute the main pathway through b_1 . The shadow does cast its attenuation on s_2 , and this is propagated to s_3 . The shadow is blocked by b_2 from affecting s_4 because A_2 and A_4 constitute the main path through that branch. The shadow cast from b_2 has no effect on s_2 or s_4 because they are on the mail pathway through b_2 , and there is no effect on s_1 because it is on the main visual path from A_2 to A_1 . The shadow from b_2 does cast its full attenuating effect on s_3 . Summing all of these effects produces no substance attenuation at s_1 , but much attenuation at s_2 , s_3 , and s_4 . The results are seen in Fig. 3.7b, showing the final substance measure $\psi(s)$ for the object. This measure is added to the Blum MAT representation, giving a value ψ at every medial point and leading to an augmented representation (x, y, σ, ψ) at every medial point.

The Blum MAT produces $O(n)$ axes for n lines on a polygonal boundary. This means that the calculation of the substance measure is an $O(n^2)$ process. A first pass over all branches is required to compute the visual vectors, visual conductances and connection shadows. In a second pass over all the axes, the substance measure is actually computed.

3.5. Results

Figs. 3.8, 3.9 and 3.10 show the substance measure of several objects. The medial points with a high substance measure correspond to the parts of the objects. Points where the substance measure is low correspond to axes that serve mainly to connect the object

parts. Points where the substance measure is between 0.0 and 1.0 reflect regions of the object where a parts-classification is perceptually ambiguous.

These results show how extraneous axes are removed and an object's parts are naturally extracted. In the cat example, the most salient parts of the cat are signaled by the visual path that runs from the head through the body and down to the end of the tail. The ears and front paws are clearly distinct perceptual parts with the connecting axes exhibiting very low substance measure (and therefore a high connection measure). In a similar manner, the lizard's legs are perceptually distinct from its body, and the stem of the leaf is distinct from the body of the leaf.

The examples also demonstrate how perceptual ambiguity about an object's parts is reflected in the substance measure. The five points of the maple leaf have no clear hierarchy of parts, and substance measures in the region reflect this. Here, the branching axes reflect the medial axis filling a region of similar saliency rather than connecting parts.

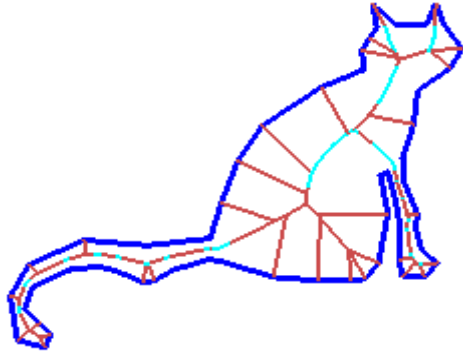
The object with protrusions in Fig. 3.10 shows how the substance measure captures a broad range of perceptual ambiguity. Some protrusions are clearly separate parts while others have some tendency to be seen as an extension of the main object. The bottom-most protrusions demonstrate this; the lower-right one is clearly an independent part, a property highlighted by the high connection measure of its connecting axis. On the other hand, the lower-left protrusion can be perceived as an extension of the central body, and the lower-middle one can be perceived as an extension of the upper-left protrusion. Perceptually there is no clear hierarchy in this region, and the substance and connection measures reflect this.

A subtle effect of performing perceptual calculations within a local aperture can be seen in the object with protrusions example of Fig. 3.10, with the right-most protrusion and the smaller one underneath it. While the larger, right-most protrusion can be perceived clearly as being a more salient part in a global sense, at the medial junction of the parts there is no clear part hierarchy. Within the local region around those branch points, this can be seen to be valid. The substance and connection measures are not intended to be used directly for high-level global object interpretation.

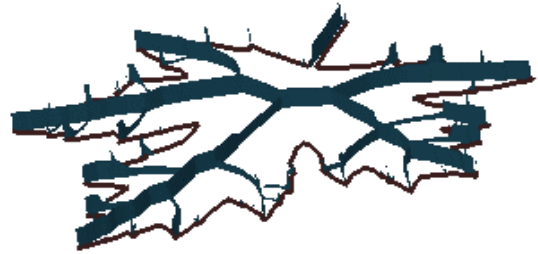
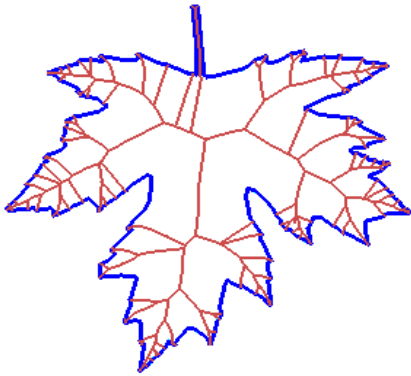
Fig. 3.9 shows a protruding part of an object and two deformed versions created with tiny perturbations to the object's boundary. The MAT is significantly different in each of the three versions; yet the substance measures remain very similar. Only the most distal edges of the MAT pieces created by deforming the object show any change in substance. This example demonstrates how the substance-weighted MAT reflects only the minor changes to the boundary even when the underlying MAT shows drastic instabilities.

Finally, Fig. 3.10 shows how the adjustment for ends of figures works and where it fails. At the tip of each protrusion, only one of the small-scale branches remains and it is connected to the main visual path that passes through the MAT to the tip; the other small scale branches are identified as connection information with only their very ends possibly showing any substance. However, Fig. 3.10b shows where the ad-hoc endness function is applied in a way that generates a substance measure that has no perceptual basis. This object reflects the case of a large-scale body of an object with the two small wings off of one end. Clearly, both of the wings are separate parts and neither is a continuation of the main body. In this implementation, however, two of the axes at a branch are always connected and given a visual conductance of 1.0. While this choice was suitable for using these ideas to develop a visual saliency measure and an associated object simplification algorithm (see Chapters 4 & 5), it probably will not support other perceptual analysis. For such work, the previous endness function could still be used to identify MAT points where an adjustment should be applied and a new adjustment created that does not connect any axes at that point. Alternatively, a new endness function also could be developed for these cases.

(a) Cat



(b) Maple Leaf



(c) Lizard

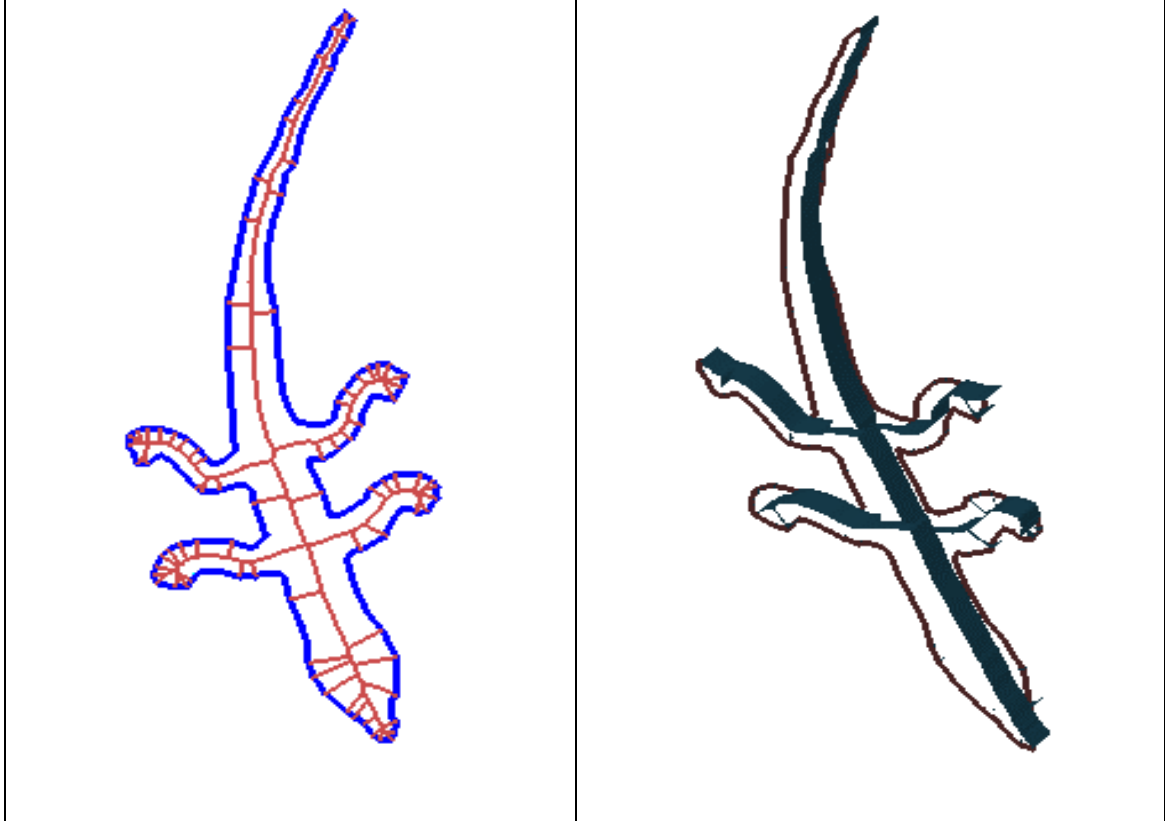


Fig. 3.8 Substance Metric Results

The MAT weighted by the substance measure of various objects. Substance measure is the height of each medial point.

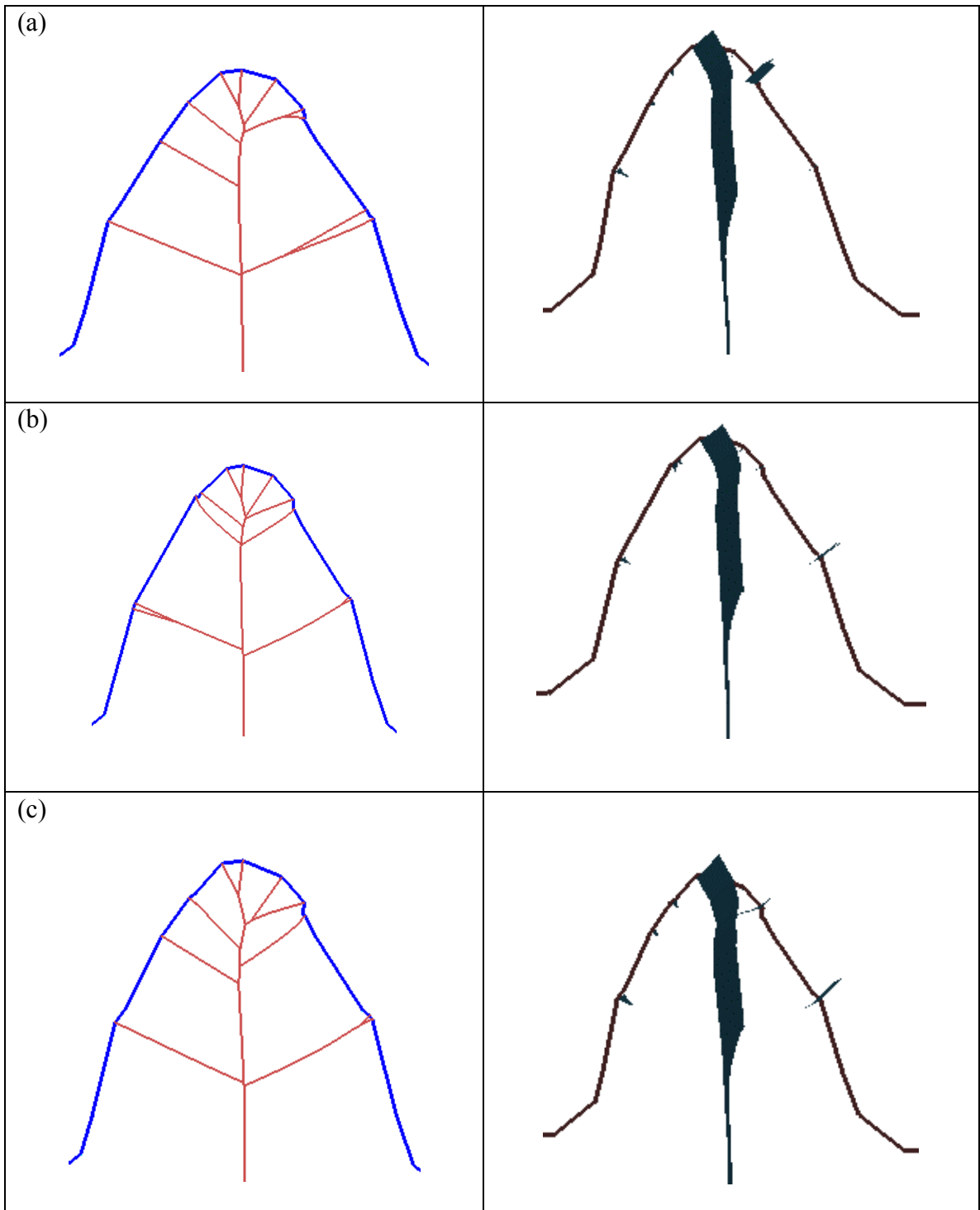


Fig. 3.9 Stability of Substance Metric

(a) The MAT and substance-weighted MAT of an object's protrusion. (b) and (c) Two deformed versions of the object, showing significantly different MATs but very similar substance-weighted MATs.

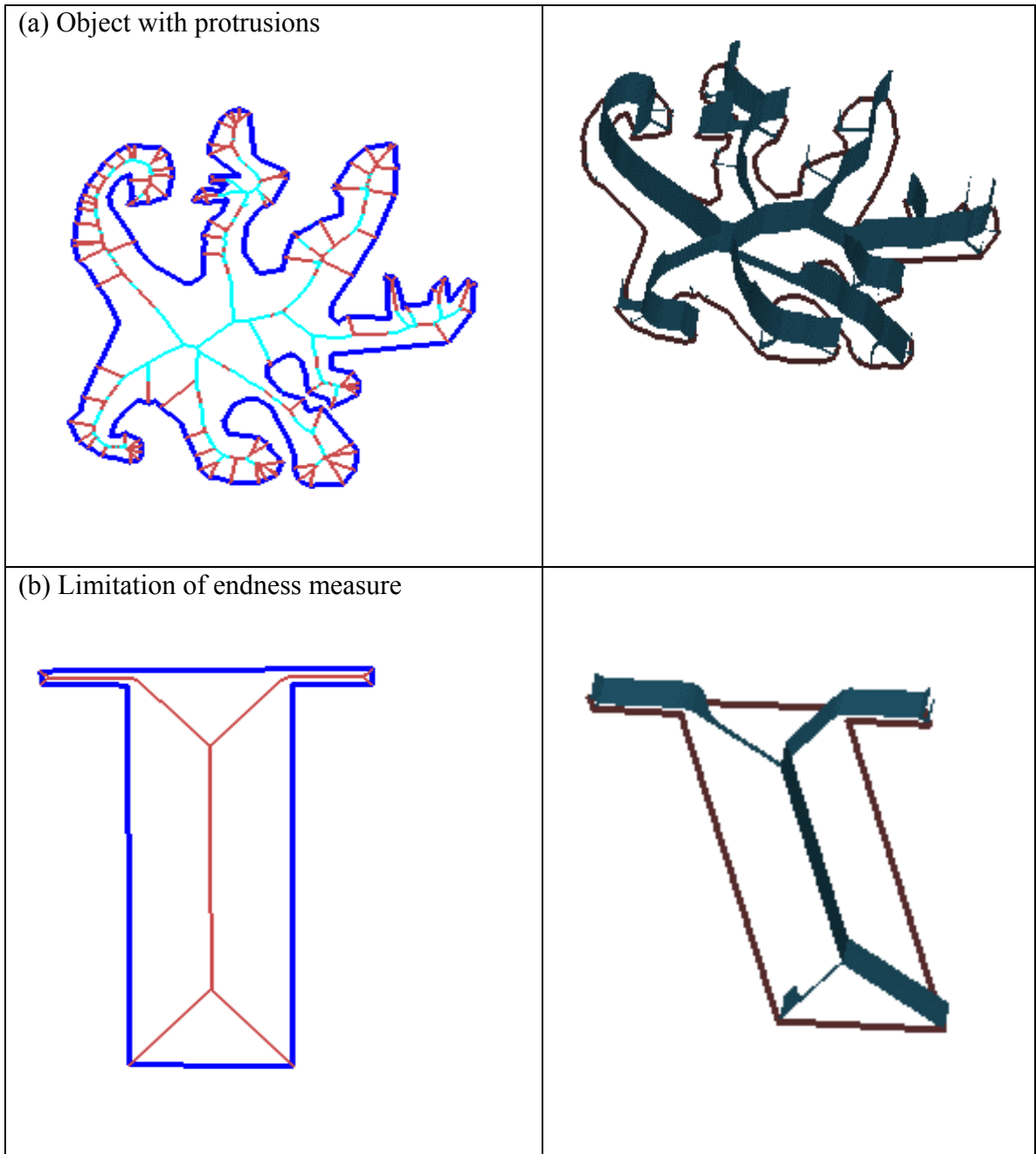


Fig. 3.10 Ambiguity in Substance Metric and a Limitation

Objects and their substance-weighted MATs. (a) The protrusions show how perceptual uncertainty about the part hierarchy is reflected in the substance measure. (b) This example shows how the implementation of the endness adjustment gives unintuitive results.

3.6. Conclusions

This paper describes a method for instrumenting the Blum MAT to perform perceptual part-decompositions of objects and to support stable and robust form calculations. The method considers objects to be a collection of solid parts and a set of connections among those parts. A measure is created at every point on the medial axes to grade the amount of substance information at each medial point and the amount of connection information. This measure is then used to decompose the object into a set of perceptually defined parts. A key aspect of this classification is that this metric gives a fuzzy measure of “part-ness”, and the parts of an object are not forced into a binary classification scheme.

Using the substance measure and complementary measure of connection as well as the resulting fuzzy classification of parts, form calculations remain stable under MAT instabilities caused by continuously deforming boundaries. This stability has been demonstrated by extensive tests and is detailed in Chapter 5. Even when drastic changes to the MAT result from tiny changes to the boundary, these instabilities are confined to the connection component of the object and the substance-weighted MAT shows only the small deformation. In this way, form analysis can track the minor boundary changes and remain immune to the sensitivities of the underlying MAT.

CHAPTER 4 MEDIALY BASED FORM SALIENCY METRICS²

4.1. Visual Saliency Measures

When we perceive objects, we perceive the inherent properties that define them. There are many ideas and theories on what properties define form and what properties our brains actually perceive. One little studied notion is the relative contribution of different form properties to an object's overall saliency. What makes one object more prominent or noticeable than another? What makes one part of an object stand out more than another? This paper develops a theoretical foundation for the important components of a saliency measure and introduces a measure to quantify saliency for relative comparisons among objects.

Saliency metrics can fall roughly into two classes: form saliency and image saliency. Form captures an object's geometric configuration, representing properties such as size, symmetry, elongation and the spatial relationship among regions on an object. Form saliency measures visual importance with metrics that are derived from geometric attributes. Image saliency, on the other hand, measures visual importance using non-geometric attributes, such as intensity, color, lighting and shadows. Image saliency is derived from these attributes or from image-based techniques such as Fourier analysis. Form saliency measures were chosen for this work to support the objective of an object simplification method driven by form-perception.

² This chapter will be submitted for publication as *Medial Measurement of Shape Saliency*, by Robert A. Katz and Stephen M. Pizer. In order to provide a self-contained description of the material, this chapter has not been edited to remove redundancies with other chapters.

4.2. Purpose of This Work

The purpose of this work is to develop a perceptually sound basis for a practical and useful form saliency metric. This is intended to be a theoretical study to produce a metric based on perceptual theories and studies. Although thorough testing remains to be done, the ideas in this chapter have been demonstrated in an object simplification algorithm for computer graphics.

In this work, a saliency metric is developed to support the process of object simplification. An active and pertinent area of research in computer graphics is the problem of simplifying object models to render them more quickly while still maintaining as much visual fidelity as possible compared to the original, unsimplified model. Saliency metrics support the approach of simplifying objects at points that are least important visually. Such metrics can indicate how much the overall perception of an object will change if the object is simplified at a given point. Also needed in such an approach to object simplification is a global saliency measure that can be used to create simplified models that maintain important perceptual characteristics of the original. The overall goal of the project is to explore what new techniques and results follow from a perceptually principled approach to object simplification.

In addition to focusing the scope of this work on a few properties of form perception, this work is further limited to 2D objects. All of our visual processing is applied to 2D retinal images, and any 3D form information is derived from 2D cues in the image. Further, with the driving goal of this work to develop new algorithms for computer graphics where polygonal models dominate, the sample objects in this work are polygonal as well.

Finally, this work concentrates on saliency in preattentive perception. Objects that are visible in a scene without our attention focused on them are said to be preattentively processed. In the attentive processing stage, more complex form perception occurs such as recognizing, analyzing and evaluating objects. Attentive processing is highly context-dependent; the task a user is performing strongly affects attentive perception [Bruce96]. To explore the basic attributes of form-based saliency, high-level context-driven perception is avoided here.

The next section of this chapter discusses key attributes of form saliency. It argues that scale is an important part of any form saliency measure, and it presents the idea of *figures* as the basic primitives for form. Section 4.4 develops a saliency metric based on these properties. Among these is a measure of visual complexity that is both a component of the saliency metric as well as an important measure itself. Then results of the saliency and complexity metrics are shown and discussed.

4.3. Components of Form Saliency

I argue in this section that three of the fundamental aspects of form perception are aperture-based scale, relationship distances and bilateral relationships. Furthermore, when these are combined they produce *figures*, the basic building blocks of objects. This section describes each of these ideas and ends with a discussion of figural properties that are important to form saliency.

4.3.1. Aperture-based Scale

As Koenderink eloquently describes, all visual operations are performed over an aperture of finite size [Koenderink84]. While mathematics can describe the world with infinite precision and with an infinitely small aperture, all practical visual systems must account for an aperture of finite size. The size of the aperture over which visual operations are performed is commonly called the *scale* of the operation. A second, distinct notion of scale is described below. Large apertures give coarsely detailed, larger-scale images. Small apertures give finely detailed, small-scale images. The same scene viewed with differently scaled apertures may appear very differently.

At any scale, an image is clarified and more detailed compared to the same scene at a larger scale. If we treat aperture size as an independent parameter that can be infinitely varied, we must attach a scale component to any measurement that is made through that aperture. We no longer think of location (x, y) of an object; we must think of location at a given scale σ , or (x, y, σ) . At small scales, objects are detailed. At larger scales objects are seen in a summarized manner which may have significantly different properties than the fully detailed objects. Adding this scale component to our familiar Cartesian coordi-

nates gives us a scale-space. Scale-spaces are distinguished by the operators that are applied to image features to create the larger scale representations.

The multi-scale nature of our perception is more than a conceptual model; it is created in the “wiring” of our visual systems. The receptive fields of the retina are patches of neighboring retinal photoreceptors, and the signals from the photoreceptors are grouped together as they are passed to the visual cortex. Each photoreceptor feeds many receptive fields, creating receptive fields of many sizes and fields that cover the entire retinal image [Bruce96]. The visual cortex performs its processing on these receptive fields, and the different size fields that contain the same scene elements lead to visual processing at different scales.

Young and Koenderink showed how this neurophysiological response could be modeled with linear combinations of offset Gaussians [Young86] [Young87][Koenderink90]. This suggests that the human visual system operates over a scale space that has a Gaussian-like blurring filter, where the size of the Gaussian kernel is directly proportional to the size of the aperture.

4.3.2. Relationship-distance Scale

The second notion of scale reveals itself when we relate multiple entities on an object or across objects. These entities may be points on the boundary, a protruding figure and its parent figure, a medial point and a boundary point, two separate objects, and so on. When relating such entities, the units for measuring distance between them must be determined. This is the second notion of scale. It can be thought of as the spacing of our visual ruler; is the distance measured in meters, inches or maybe some parameter of the object?

For example, consider Fig. 4.1a, where the two points can be thought to be two units apart, four units apart, or eight units apart. The same entities can have different distance measurements depending on the scale used to perform the measurement. The choice of scale can be based on many factors such as the entities being related or the context of the problem being solved.

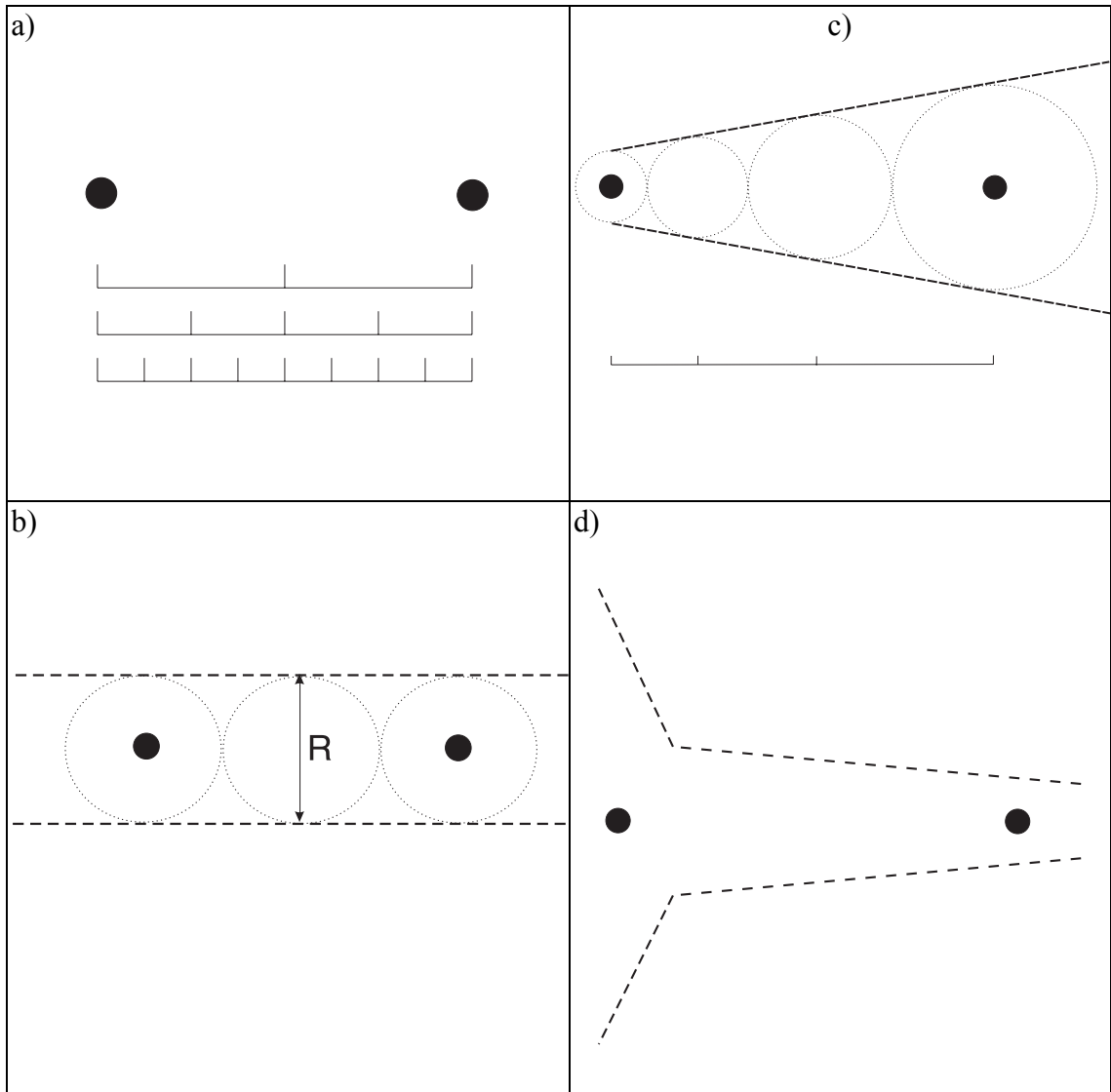


Fig. 4.1 Scale in Visual Perception

(a) Two points with distance measured in different units. (b) Same two points with distance measured in width-based units. (c) Two points with distance measured with varying width-based units. Gradual change in widths implies points are on same part of the object. (d) Same two points as in (c) but on separate object parts as distinguished by rapid change in width

A natural method for setting the scale of inter-relationship distances is to use an aperture-based ruler. The natural aperture size we use to view a car body is much larger than that to view a wheel, and that aperture is much larger than the natural one to view a lug nut. A scale-based ruler for distance measurements is based on the aperture size of the object at the points being related. Fig. 4.1b shows the same two points as in Fig. 4.1a with a width-

based aperture of radius $R/2$. With this aperture, the points are two units of R apart. In this way, relationship-distance scale is proportional to aperture size.

There are strong computational reasons for aperture-based distance measurements. In the process of understanding a stream of visual input, relationships are created among entities. For every relationship, we start with one entity and seek out a matching entity. This search spans all locations in all dimensions to find a match. A search that was always performed at the smallest possible scale would require that a combinatorially large number of location pairs be searched when attempting to create a relationship. The process would, at best, be inefficient and place extreme demands on the visual processing system. At worst, it would make intractable the creation of some simple relationships.

When performing the search for related entities with an aperture, features are sensed with a spatial tolerance on the order of the aperture width. Search steps are in units of the tolerance, giving all the accuracy in relationship distance measurements that is possible at that scale. This allows a wide range of relationships to be made with a limited amount of visual processing. In other words, processing requirements can be considered to be proportional to the number of aperture-widths that need to be processed or to the square of that number.

There is psychophysical evidence for aperture-based distance measurements. [Shipley92] found that perceived boundary clarity of a square defined by four $3/4$ disks forming the corners of a square remained constant under varying lengths of specified edges and length of gaps in boundary (non-specified edges) when the ratio of specified length to total length remained constant. This can also be interpreted as aperture-based distance measures; when the aperture is scaled relative to total edge length, the gap is perceived as the same number of units long.

Combining the notion of aperture-based scale with the scale of relationship distances allows entities with different aperture widths to be related. Fig. 4.1c shows two points within a narrowing object. While the distance between these points, when measured by aperture widths, can be perceived as greater than the distance in Fig. 4.1b, the two points are still closely related. The points can be considered local to each other since they are

local in the left-most point's frame of reference and the scale changes rather slowly as it progresses to the right-most point.

The same two points in Fig. 4.1d have a very different scale-based relationship. The two points are separated by the same Euclidean distance as in 4.1c, but there is a very rapid change in scale along the medial line connecting the points. The locality of the left-most point does not extend to the right-most one. This leads to our perception that the points are on different parts of the object instead of the same part as in 4.1c. Also, we perceive the right-most point to be on a part that protrudes from the larger scale "parent" region on the left.

One result of the multi-scale processing in preattentive perception is that the perception of an object's form at a given point is determined only by the local region surrounding that point, not by distant features of the object. The perception of our nose's form is not influenced directly by our shoulders, although it may be affected by the form of our eye sockets or our cheek bones. Since all visual operations are done in a scale-space, the size of the local region is determined by the scale of the operation. The perception of our head's form may be influenced by our shoulders, since the head is a larger-scale feature than our nose and thus is influenced by a proportionately larger region around it.

4.3.3. Bilateral Relationships

Another critical component of form saliency is the relationship of points on opposing sides of an object. These points can be thought of as being reflections of each other through a point centered in the object. In other words, every point on an object has a bilateral relationship with a point on the opposite boundary, and these two points are related through a point in the middle of the object. The two related boundary points are called *involutes*.

[Blum67] first introduced this idea with his Medial Axis Transform. As Blum and many others have described [Brady84] [Leyton88] [Székely96], bilateral relationships are conveniently represented with a medial object representation that consists of the middle

points of an object and the width of the object at each middle point. This creates a skeleton of the object in the same spirit as an animal skeleton, with the addition of knowing how much flesh is affixed at each point on the skeleton. Except for a few isolated special points, every skeleton point corresponds to two involutes: the opposing boundary points. This describes an object from the inside out, and it offers the advantage of representing the substance of the object and not merely the boundary. Many people have harnessed this representational advantage in computations involving form [Ogniewicz92] [Székely92].

Beyond the computational advantages of bilateral form relationships and their medial representations, there is evidence that our visual systems process these relationships directly and use this information to perceive form at a basic level. [Lee95] presents physiological evidence of neurons that respond to bilateral stimulus when the associated medial location falls within the neurons' receptive field. There is also psychophysical evidence for the importance of medial sensitivity in our perception. [Frome72] performed a study with subjects aligning forms to reference lines. Ellipses and curved ellipsoids were used as forms, and the study tested whether the length of the ellipse's major axis or the length of the ellipse's medial axis best matched the reference line. The study found that alignment accuracy correlated significantly better with medial axis length than with major axis length. In [Psothka78], subjects were asked to place a single dot inside outlined objects, choosing a location that best represents the object. The positions chosen fell on the medial axes of the object. [Vos93] found a similar tendency to choose a centered point to represent objects. Finally, [Kovacs94] found that when measuring contrast sensitivity for a target enclosed by a boundary, maximal sensitivity is along the boundary's medial axis.

[Burbeck97] performed a study testing orientation discrimination of rectangular objects with sinusoidally modulated long edges, with the modulation adjusted in both amplitude and phase. Strong evidence was found for the importance of across-object relationships in determining perceived orientation. A similar conclusion can be drawn from the work of [Kovacs93]. They found that detection of fragmented contours composed of Gabor patches is greatly enhanced when the patches form a closed contour, thus providing two

opposing boundaries. Further, detection of a target probe is enhanced when it is located inside a closed circle.

4.3.4. Scale-based Bilateral Relationships

Beyond the importance of scale and bilateral relationships as independent aspects of form perception, there is growing evidence that the relationship of medial involutes is deeply related to scale. Opposite points are related at a scale that is proportional to the medial width that separates them. Burbeck and Pizer capture this in their theory of Cores [Burbeck95]. Cores are loci of medialness that link opposing boundary points that are sensed through an aperture. Since the size of the aperture used to determine involutes is proportional to the scale at their medial locus, the width of a Core determines the accuracy of its boundary measurement. Thus, when Cores are created, their boundaries are not exact; rather, the boundary is known to lie within a bounded region. In [Burbeck96], when subjects were asked to bisect rectangles with two long sinusoidally modulated sides, the perceived center points were determined by the relationship of their corresponding boundary points. Moreover, the effect of the edge modulation corresponded with the object width, strongly suggesting that same mechanism determines both the location of the perceived center as well as its standard deviation.

There is other evidence in the literature for the approximate proportionality between aperture size and the ruler used to determine bilateral relationships. In [Vos93], where the point that subjects chose to represent an object tracked the object's center of gravity, they also found that the standard deviation of the chosen point versus the actual center of gravity grew in proportion to the width of the object. [Kovacs98] describes a multi-scale medial representation that is similar to medial-axis type representations but uses a small set of medial points that each represent a maximal amount of boundary information; they found that their model predicted the results of several contrast sensitivity studies.

4.3.5. Figures

A crucial step in our perception of form is the segmentation and decomposition of objects into simpler parts. Intuitively, these parts can be thought of as a simple object or a protru-

sion from or indentation into another part. [Biederman87] proposed a set of geometrically simple components as the basic parts of an object. Each of his components can be seen as a configuration with a simple medial axis. [Hoffman84] proposes that our visual systems decompose objects into parts at boundary points that are concave as viewed from the inside of the object and are a curvature maximum, and [Braunstein89] confirms this in a psychophysical study. These geometric properties hold at the ends of parts, i.e., what we define below as *figures*. [Wolfe94] and [Wolfe97] show that part-decomposition happens preattentively.

The three aspects of form perception discussed above, aperture-based scale, the scale of relationship distances, and bilateral relationships, can be combined to describe our visual system's ability to perceive *figures*, the basic building blocks of form. A figure can be thought of as the part of an object represented by a single medial axis, independent of other protrusions and indentations. Each protrusion or indentation is a separate figure, where the term *figure* is defined as either a whole object or the main object part of a protrusion or indentation in another figure.

[Siddiqi96] reaches a similar conclusion. They decompose limb-based parts where there is a pair of negative curvature minima with “good continuation” connecting those boundary points through the limb, and they also define neck-based parts at narrowings in the object. This theory is supported with a series of experiments that found perceived parts do correspond to the parts of this model. While they do not describe their work in terms of a medial model, the effects can be explained easily and directly with scale-based bilateral relationships.

In perceiving figures, the size of the aperture used to perceive boundary involutes is directly related to the width of their common medial point. [Burbeck96] shows that medial width determines whether objects with the same wiggly boundaries but different widths are perceived as straight or wiggly. [Siddiqi96] shows how the distance between opposing points influences the response strength of perceived parts. As seen also in the contrast between Fig. 4.1c and Fig. 4.1d, medial width determines the boundary sensing aperture, and this contributes to the decomposition of objects into figures.

4.3.6. Figural Components of Form Saliency

Using figures as the primitive components of form, figural attributes become the components of form saliency. To measure the saliency of an object, any metric must account for an object's figural properties. This work proposes the following as the primary figural attributes: mass of figures, elongation of figures, continuation of figures, ends of figures, and the relationship between internal and external figures.

Figural mass accounts for the general notion that “bigger is more prominent”; it is defined as the area enclosed by a figure. In general, as a figure grows larger, it will stimulate more of the retina which makes it more noticeable to the viewer.

The elongation of figures, i.e., the ratio of length to width, captures the notion that thick, blobby objects are more prominent than long, thin ones. A dumpling has more saliency than a piece of spaghetti with the same amount of dough. Blobby objects have similar widths and lengths, but elongated objects have a much greater length than width. When perceiving an object through a width-based aperture, more of an object's mass is concentrated within one aperture for rounder objects than for elongated ones; one aperture's width contains more of the object. Fig. 4.2 demonstrates this effect. [Michels65] provides an overview of this parameter of form and studies of its effects.

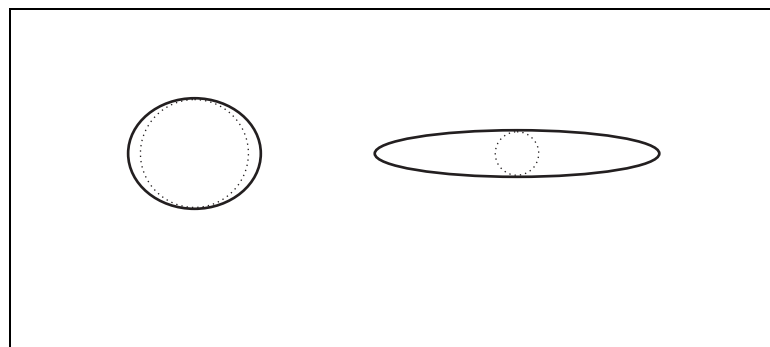


Fig. 4.2 Figural Elongation

For two figures with the same total area, the more rounded figure has a greater area within its width-based medial apertures than the more elongated figure.

For figures that are elongated, those that are straight are more prominent than those that are curved or bent. This is captured in the continuation of figures and is reflected in their medial axes. Figures whose medial axes are straighter carry more prominence than those with bends. While there are no reports directly studying the continuity of medial axes, there is strong psychophysical and physiological evidence that our visual systems follow straight contours even with large gaps, and we integrate straighter curves more easily than those more sharply bent [Pettet99][Polat93][Polat94][Nelson85]. I hypothesize that similar results will hold for medial curves, where the curvature of a medial axis is determined by the curvature of its related boundaries and also by the rate of change of the figure's width along the axis.

The ends of figures are special locations visually, and they carry added visual significance. For example, the tips of a hot dog have more visual information than the regions in the middle. [Leyton87] highlights these points in his Symmetry-Curvature Duality, which states: "Any segment of a smooth planar curve, bounded by two consecutive curvature extrema of the same type, has a unique symmetry axis, and the axis terminates at the curvature extremem of the opposite type." He further proves that these points correspond to the ends of medial axes and argues that they are especially salient, carrying a maximum of boundary information. There is also strong physiological evidence for these kinds of ends. [Hubel77] and [Orban79] have found cells in the visual cortex that respond to the ends of lines, acting as specialized "endness detectors".

Finally, the relationships between the internal figures in an object to its external figures are also visually important. A simple indentation into an object has no consistent internal figural representation; it may show up as a narrowing of the interior or as a segmenting into parts (see Fig. 4.3a). This same indentation, however, can be directly represented with a single external figure that protrudes into the object (see Fig. 4.3b). Indentations into the interior are simply protrusions of the exterior regions, and likewise, protrusions from the interior are indentations into the exterior.

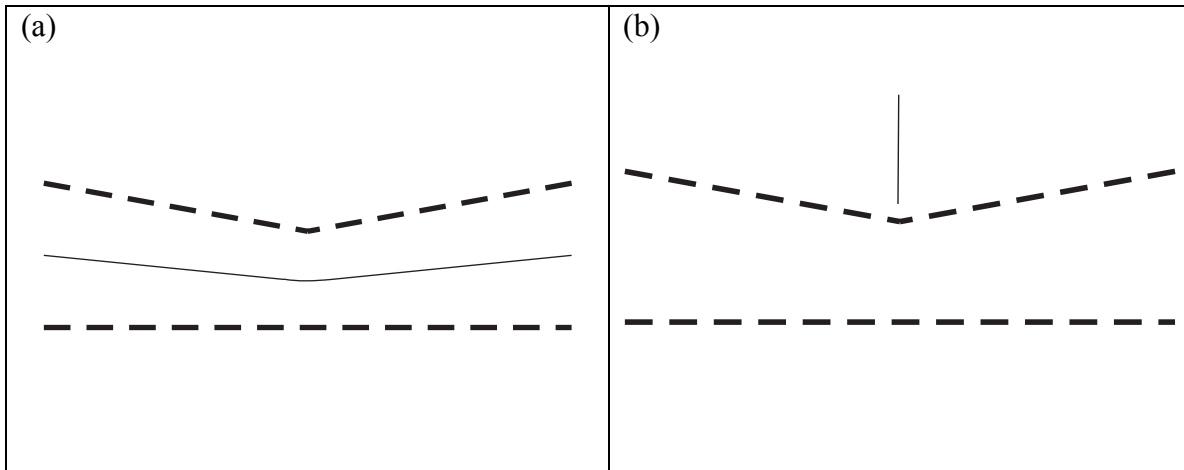


Fig. 4.3 Internal and External Figures

- (a) The indentation in the top boundary has no direct representation in the internal figure.**
(b) Using external figures, the same indentation corresponds to a distinct figure.

In his work that identifies the special significance of figure ends, [Leyton87] includes both internal and external figures, making no distinction between the two types other than the flavor of boundary curvature extrema that creates them. [Siddiqi96] suggests a role for medial-based parts in figure/ground segregation. While they never explicitly mention medial models, they clearly rely on bilateral relationships as the basis of their theory. In [Kovacs93], Kovacs and Julesz suggest that figure/ground segmentation is facilitated by the same closed contours that caused the enhanced sensitivity response along points on the medial axis, again suggesting a role for figures and bilateral relationships in figure/ground separation.

4.4. Perceptual Form Metrics

Perceptual metrics contributing to a form saliency metric are developed in this section. They are based on the preceding development of figures and the figural components of form saliency. This theory proposes that the bilateral relationships of an object, as represented by medial axes and the object width along every point on the axes, provide a basis for form perception. Furthermore, the theory adds that form perception is performed

within finite size apertures that can vary along an object, creating a multi-scale medial approach to form processing.

This work has been implemented using the Blum Medial Axis Transform (MAT) as a medial representation from which to partition figures and calculate metrics. There are reliable algorithms to extract the MAT from 2D objects, and the MAT can be calculated exactly for polygons. Using the MAT, however, introduces several new problems. The MAT is extremely sensitive to small boundary perturbations. Complex objects, especially those represented as here by a polygon, contain an overwhelming number of axes that do not directly represent the object's boundaries but instead serve to connect the various parts of the objects. These problems are detailed and solved in Chapter 3.

The remainder of this section details the metrics contributing to a form saliency metric as well as the saliency metric itself. These include measures for the continuity of figures, the mass of figures, elongation, the saliency in a local axis interval around a medial point, and finally the overall saliency of an object. *Visual conductance* captures continuity of a figure and the degree of continuity between neighboring figures. *Visual mass* measures the area enclosed by a figure, reflecting its overall size. *Visual length* is a measure of length per medial width that reflects a figure's elongation. *Visual potential* captures the local saliency in an axis interval about a medial point, and *visual significance* measures the saliency over a region or an entire object.

4.4.1. Visual Conductance

In these metrics, as detailed in Chapter 3, the continuity of figures at a branch point b is measured by *visual conductance* C_b , with $0 \leq C_b \leq 1$. The metric described here is applied at medial axis branch points to determine how to combine axis segments into figures. This assumes that conductance $C_b=1$ along axes between branch points. In an alternative development of the metric, C_b at non-branch points could be lowered in proportion to the curvature of the axis.

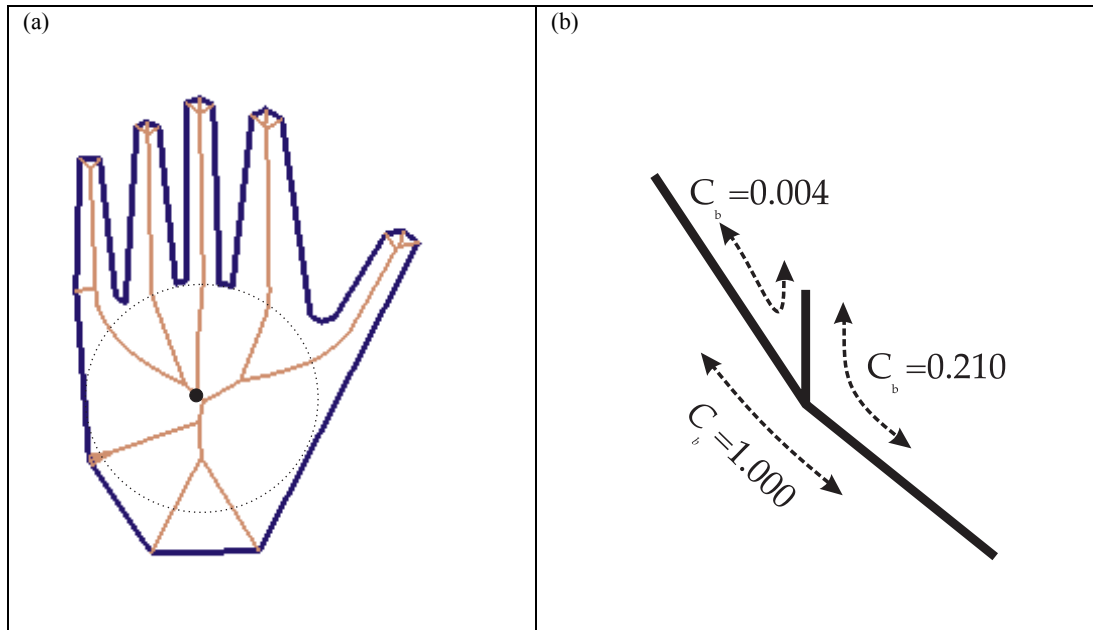


Fig. 4.4 Figural Continuity

(a) Hand object with its Blum MAT and width-based aperture at the branch where the middle finger's axis joins the other axes (b) Weighted vectors for each axis emanating from the branch, with longer vectors representing greater accumulated visual scale in that direction. Also shown are the visual conductance values for each pair of axes represented by the weighted vectors.

For every medial branch point, three values of C_b are calculated, one for each of the three pairs of axes emanating from the branch. These values of C_b are based on the widths and directions of the axes in a medially based aperture centered at the branch point. To calculate C_b for the axis pairs through b , three weighted vectors are calculated that correspond to each of the three branch axes. Each vector summarizes the region of object that falls within an aperture around the starting branch and that is represented by its corresponding axis emanating from the branch plus any branches traversed within the aperture while traveling away from the starting branch. The direction of the vector captures a weighted average of the medial direction for the selected axes, and the magnitude of the vector reflects the weighted average of the medial width of the selected axes. Fig. 4.4 demonstrates this. For a more detailed description of this measure, see Chapter 3. C_b then captures the length and direction consistency of axis pairs through a branch, as follows. At each branch point, the two axes whose vectors have the highest correspondence of direction and scale are combined into a single extended axis with $C_b=1$. The remaining axis becomes a distinct figure, and the conductances between it and each of the other two axes

emanating from the branch are set to a value less than 1, according to the agreements in length and direction of their vectors. In this way, visual conductance is used to segment, in a fuzzy manner, a Blum MAT representation into figures. Using visual conductance to determine which medial axes should be combined into the same figure and which should be considered to be separate figures, the axis segments of the Blum MAT are combined into an approximation of perceptual figures. All of the remaining metrics are based on this set of figures

4.4.2. Visual Mass

Visual mass, M_v , captures the overall size of a figure, and it is calculated as the Euclidean area. At any medial point s of an object, a differential slice of visual mass, $dM_v(s)$, can be calculated. $dM_v(s)$ measures the differential area defined by a medial point s and its two involutes (see Fig. 4.5). The visual mass of an entire object is the integration of each differential slice along all the internal axes of an object. Thus, total visual mass

$$M_v = \int_s dM_v(s).$$

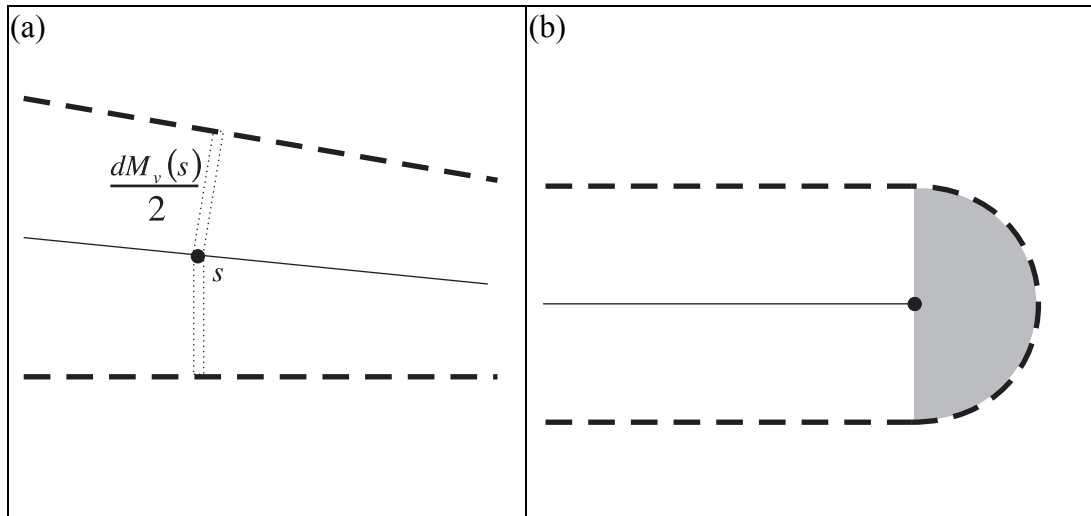


Fig. 4.5 Visual Mass

- (a) The differential slice of mass $dM_v(s)$ for a generic middle point on a figure.
- (b) Increased amount of visual mass associated with a figure end point.

4.4.3. Visual Length

Visual length, L_v , reflects the elongation of figures by measuring Euclidean length per medial width. Fig. 4.2 shows how round figures capture more of their area within each medial width-based aperture than elongated figures. The medial width along a figure provides the step size of the length-measuring ruler as described above in the relationship-distance notion of scale (see Fig. 4.1). Visual length uses this to quantify the visual notion of narrow figures versus thick figures.

Visual length is naturally invariant to magnification; as an object is magnified, the radius at every medial point grows but so does the length of each medial axis. In other words, when an object is zoomed larger or smaller, both its width and its length change proportionately. Any comparison of length versus width will give the same result regardless of the size of the object. A thick figure protruding from an object will always be perceived as thick, *relative to the rest of the object*, regardless of the object's size. Likewise, thin figures remain thin at any scale because the width of the object along the figure is small compared to the figure's length.

Visual length is a scale-space arc length measuring the distance between two points along a path in scale-space. In scale-space, the measuring unit varies according to the medial width of the object. As a simple example, an object six inches long and one inch wide might be measured in one-inch units based on the one-inch width; this would give a visual length equal to six units. Another object that is six inches long but two inches wide would be measured in two-inch units, giving a visual length equal to only 3 units. The wider object has a shorter visual length than the narrow object, even though both objects have the same Euclidean length.

Since the width of an object can vary continuously, visual length is measured by first defining the differential visual length at every medial point s along a figure, $dL_v(s)$.

[Eberly94] has defined this as $dL_v(s) = \frac{\sqrt{\dot{x}^2(s) + \dot{\sigma}^2(s)}}{\sigma(s)} ds$, where $\sigma(s)$ is the scale at the

point s . I use $\sigma(s) = R(s)$, where $R(s)$ is the medial width at point s . Again, the total visual length of an object is obtained by integrating each differential slice along each figure,

$$L_v = \int_s dL_v(s).$$

Visual length is used as part of other perceptual metrics, but it is also a measure of visual complexity. Visual length is a measure of elongation, and Gestalt views on perception have long held that elongation is a simple indicator of visual complexity. Visual length is the metric I use in Chapter 5 as a relative measure for the degree of an object's simplification. The simplification method described in Chapter 5 "blobbifies" objects, making them thicker and more circular. In the limit, all objects are simplified into one or more circular blobs which have a visual length of zero. Comparing the visual length of an original object to the visual length of a simplified version measures how much simplification has occurred.

4.4.4. Visual Potential for Single Figures

Visual potential, Pot_v , is a quantity that reflects the visual importance of small regions of an object, i.e., when integrated, it measures how prominent any small, local medially-based region is compared to others. This metric embodies in a single number the density of saliency of any point on an object or, when accumulated over many points, the saliency of an entire region. The following discussion will incrementally build a formula for visual potential to show how a single metric can reflect and satisfy all of the form cues and relationships that have been proposed as desirable in a visual metric for shape simplification. First, the metric will be developed for a single figure, and then it will be extended to multiple, branching figures.

By making visual potential proportional to visual mass, it will reflect the perceptual result that larger or thicker figures are more visually prominent. Since a large visual mass makes a figure more prominent, $Pot_v(s)$ should grow as $M_v(s)$ grows. Similarly, figures with a large visual length are less prominent than figures with a small visual length but

the same visual mass. By making $Pot_v(s)$ inversely proportional to visual length, $Pot_v(s)$ lessens as $L_v(s)$ grows. So far,

$$Pot_v(s)ds \propto \frac{(dM_v(s))^a}{(dL_v(s))^b}.$$

For the simple case of an infinitely long object with constant width,

$$\begin{aligned} dM_v(s) &= 2Rds, \quad dL_v(s) = ds/R \\ Pot_v(s)ds &\propto \frac{(2Rds)^a}{(ds/R)^b} = R^{a+b} ds^{a-b} \end{aligned}$$

This constrains $a-b=1$, and since $Pot_v(s)$ is designed to lessen as $L_v(s)$ grows, $b > 0$.

Picking the most simple solution to these constraints, $a=2$ and $b=1$.

$$Pot_v(s)ds \propto \frac{(dM_v(s))^2}{dL_v(s)}$$

This gives $Pot_v(s)$ as a density function with terms for visual mass and visual length.

Adding to this, we must address the effects of scale and the special saliency at figure end points. Scale effects are discussed in the next paragraph. For figure end saliency, the terms $\alpha(s)$ and $A^{end}(s)$ are added as follows. These terms are discussed in the next section.

$$Pot_v(s)ds \propto \alpha(t)A^{end}(t)ds + (1 - \alpha(t))\frac{dM_v(t)^2}{dL_v(t)}$$

Visual potential becomes a scale-space metric by averaging the measurement in an aperture around the point of interest s_0 and by tying the size of the averaging aperture to the scale σ at s_0 . To mimic the visual processing observed in studies of the human visual system, the scale-space used is a Gaussian filter whose kernel is proportional to the size of the measuring aperture. Integrations in the averaging aperture are weighted by a Gaussian function with the weight of the integrand decreasing as the point of integration moves away from the aperture's center. For single figures, this gives

$$Pot_v(s)ds \propto \int_{s-k\sigma(s)}^{s+k\sigma(s)} e^{-\frac{1}{2}\left(\frac{t-s}{\sigma(s)}\right)^2} \left(\alpha(t)A^{end}(t)ds + (1-\alpha(t))\frac{dM_v(t)^2}{dL_v(t)} \right) dt.$$

The constant k is set to a value of 3 as developed below. For the scale term σ , I know of no experimental or theoretical work to suggest the parameters of this weighting; the following discussion develops one. To detect a boundary that is a distance R from its medial point s , our visual system must link two boundary detecting neurons that are measuring the object over a distance of at least $2R$. The boundary location is known only to within a tolerance that is proportional to its scale R . Making the assumption that this tolerance is $0.5R$, the edge of the boundary-detecting aperture in our visual system is a distance of $1.5R$ from s . Thus, with involutes that are a distance R from their corresponding medial point s , I assume the visual system's Gaussian filtering has a standard deviation of $\sigma(s) = R(s)/2$. This same distance is used here to bound the integrating axis along medial axes.

The constant k in the Gaussian weighting determines the actual limits of the aperture as measured in arc length along the axes as they emanate from the center of the aperture. The preceding argument suggests a value of $k = 3$, to create an aperture with radius $1.5R$. As further support for this value of k , note that we want the Gaussian weight to drop to zero at the edge of the integrating aperture. An aperture of radius 3σ , or $k = 3$ in the equation above, provides near-zero values at its outer limits. While these values for k and for σ are based on some broad assumptions, they have proven to be useful in this work.

This development of $Pot_v(s)$ addresses points in the interior regions of single figures. The next section details the effect of figural ends on visual potential; in the section after that, the metric is extended to include multiple figures.

4.4.5. Visual Potential at Figural Ends

To match perceptual results, it is desired that ends of figures show an increased visual potential. At the end of a figure, such as the tip of a finger, there is a much larger amount visual mass associated with the single end point than the differential amount of visual

mass associated with a generic medial point in the middle of a figure (see Fig. 4.5). Considering this from another perspective, a generic medial point has two boundary involutes. but a figural end point relates to an entire arc of boundary; at end points, it is this correspondence that generates increased saliency. The term $\alpha(s)$ in the formula developed thus far for $Pot_v(s)$ determines to what degree a medial point should be considered an end point, and the term $A^{end}(s)$ reflects the increased visual potential due to this end response.

[Kovacs98] shows that there is an increase in saliency at the ends of figures, and they provide some psychophysical data on this response. However, there is no clear indication of the actual form of this function. Here, I create the perceptual effect by including in $Pot_v(s)$ the endness factor $A^{end}(s)$ that increases visual potential at end points. This is implemented as a scaled delta function that is zero at figural middle points and then an impulse at end point s_{end} .

$$A^{end}(s) = c_{end} R(s_{end})^3 \delta(s - s_{end})$$

The amplitude of the delta function in $A^{end}(s)$ increases with the medial width at the end point. To match the dimension of the “internal axis response” term dM_v^2 / dL_v , which has units of $R^2 ds / (1/R)$, the amplitude is set to R^3 . An additional ad-hoc variable c_{end} is added to adjust the increase in visual potential at end points. A value of 10.0 for c_{end} produces good results.

$\alpha(s)$ determines the proportional strength of the endness response at a medial point as compared to the “internal axis response” term. As Fig. 4.6 shows, $\alpha(s)$ is chosen to be proportional to the number of radians of the aperture around medial point s that encloses part of the boundary. This aperture has a width that is slightly larger than the medial width at s ; in practice, a width of $1.1R$ is used. In order to emphasize the endness response caused by boundary points directly in line with the continuation of the medial axis, the contribution of each boundary point is weighted by the angular deviation from the tangent of the axis. This also removes the endness response from boundary involute points and their local neighbors. Fig. 4.6b details this weighting scheme. The endness

response is defined to be maximal when a figure ends in a semi-circular end cap; this is the case when π radians of boundary are enclosed. Thus, $\alpha(s)$ is clamped at a maximum value of 1.0 when π radians are enclosed, and it is given as follows.

$$\alpha(s) = \min \left(\sum_{\substack{\text{boundary segments} \\ \text{within endness-} \\ \text{aperture around } s}} \frac{w(\hat{v}_0, \hat{v}_1) \varphi(\hat{v}_0, \hat{v}_1)}{\pi}, 1.0 \right)$$

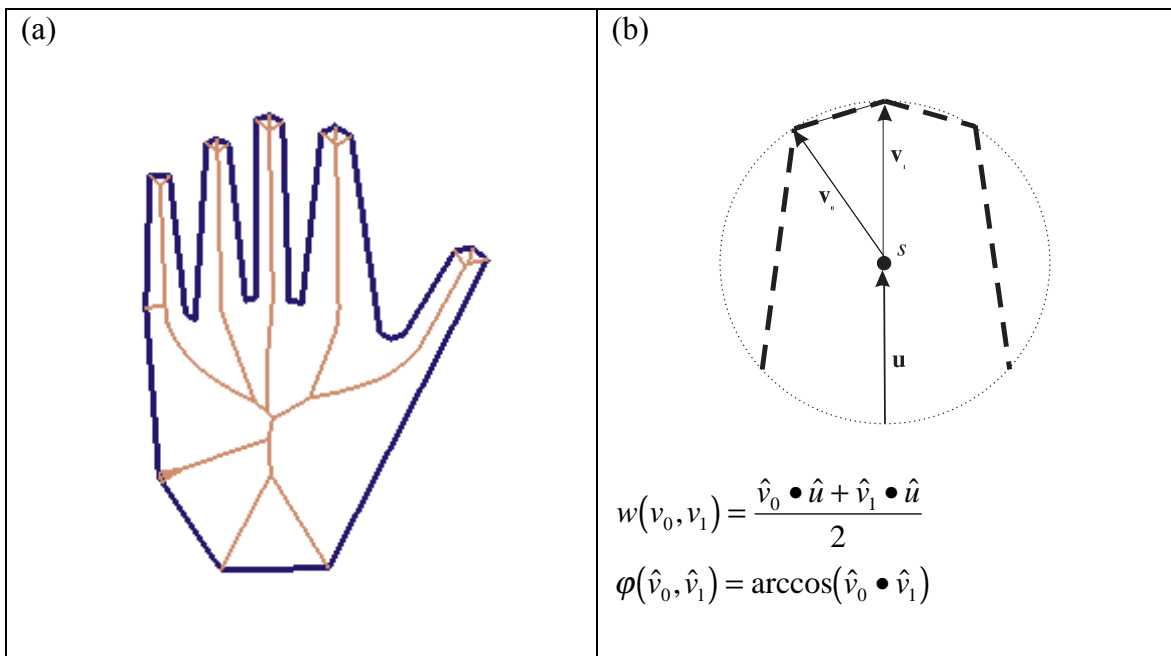


Fig. 4.6 Saliency at Figural Ends

- (a) An endness-capture aperture and its boundary points enclosed by the aperture around medial point s .
- (b) The weighting scheme used to emphasize the endness contribution of boundary points directly in line with the continuation of the medial axis with direction u . This is shown for a single boundary segment; the total endness measure sums this value for each boundary segment enclosed by the aperture.

4.4.6. Visual Potential with Multiple Figures

In order to extend visual potential to multi-figure objects, the measure must include all medial axes that represent figures falling within the integration aperture discussed above.

Where figures meet, there is an intersection of axes in the medial representation. When the integration within an aperture encounters such an intersection, the integration follows one and then the other to include values along both of the intersecting figures. The Gaussian weighting within the aperture is based on the distance measured along the medial axes from the central point through any branch points.

Including the factor of the *substance measure* $\psi^{branch}(s)$ at branches in visual potential addresses the fact that figures differ in their degree of connectedness where they branch. The final expression for visual potential is as follows.

$$Pot_v(s)ds = \int_{s-k\sigma(s)}^{s+k\sigma(s)} e^{-\frac{1}{2}\left(\frac{t-s}{\sigma(s)}\right)^2} \psi^{branch}(t) \left(\alpha(t)A^{end}(t)ds + (1 - \alpha(t))\frac{dM_v(t)^2}{dL_v(t)} \right) dt$$

The substance measure $\psi^{branch}(s)$, discussed below and detailed in Chapter 3, measures the degree to which medial points reflect the substance of a figure and or act as a connection between figures. The saliency of axes that serve principally to connect figures is downgraded. However, rather than a binary indicator for each axis, the substance measure allows for a continuous grading between 0.0 and 1.0.

$\psi^{branch}(s)$ determines the perceptual relationship of neighboring figures. For an object with a tiny pimple on a large figure, the saliency of the pimple itself is influenced very little by the larger, parent figure. Similarly, the saliency of the parent figure is affected very little by the pimple. This independence is reflected in a low value for $\psi^{branch}(s)$ at the branch where these figures meet. However, in a ‘Y’-shaped object with three figures of similar scales, each figure affects the saliency of its neighbors. This, in turn, is reflected with a high value of $\psi^{branch}(s)$ at the corresponding branch.

The gating of the influence of one figure on another is computed from the visual conductance at each branch. For every non-branch point s along a figure, $\psi^{branch}(s)=1$. In the neighborhood of every branch point, the substance measure is proportional to the visual conductance between figures at the branch, and this value is then multiplied by a fall-off factor that regulates the effect within an aperture around the branch point. The substance

weight also includes a component based on the endness response at the branch point. Beyond the fall-off aperture, $\psi^{branch}(s)=1$ again. The effects of multiple branches that fall within the a weighting aperture are multiplicatively cumulative. See Chapter 3 for a detailed discussion.

4.4.7. Visual Significance

Over a region or an entire object, the total visual saliency is called *visual significance*, Sig_v . This is simply the visual potential integrated over all points s on all figures A_i , giving the following:

$$\begin{aligned}
 \text{Visual Significance} &= \sum_{A_i} \int_s Pot_v(s) ds \\
 &= \sum_{A_i} \int_s \left(\int_{s-k\sigma(s)}^{s+k\sigma(s)} \psi^{branch}(t) e^{-\frac{1}{2} \left(\frac{t-s}{\sigma(s)} \right)^2} \left(\alpha(t) A^{end}(t) ds + (1 - \alpha(t)) \frac{dM_v(t)^2}{dL_v(t)} \right) dt \right) ds
 \end{aligned}$$

This can be measured over a small, local medially-based region as $Pot_v \Delta s$ to give a measure of local saliency, or it can be measured over many figures to give a global measure of saliency. Visual significance has been used in a perception-based shape simplification algorithm by calculating the visual potential of an object on all of the internal and external figures. The least salient point, either internally or externally, becomes the center of a simplification transformation. The combined visual length over internal and external figures is used as a measure of simplification; as the object is simplified its visual length becomes smaller.

4.5. Results

Figs. 4.7 and 4.8 show objects and their corresponding visual lengths. More rounded objects, such as the tube in Fig. 4.7 and the blobby object in Fig. 4.8 are more circular than their counterparts and have lower visual lengths; thus they have lower visual complexity. As these examples show, objects with similar extents and visual mass can have very different levels of visual complexity. In fact, the object on the right in Fig. 4.8 would have

even lower visual length if its corners had been rounded, accentuating even further the difference in complexity.

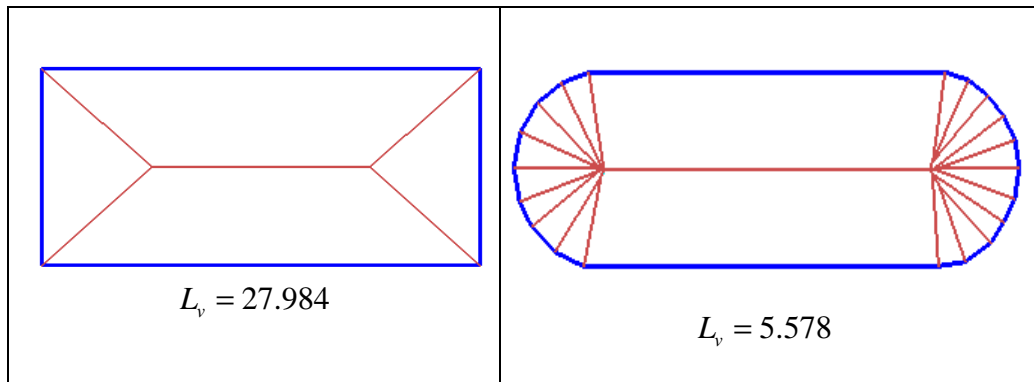


Fig. 4.7 Visual Length Results for a Rectangle and a Tube

A rectangle and a tube with rounded ends. The visual length L_v for the tube is much less than for the rectangle.

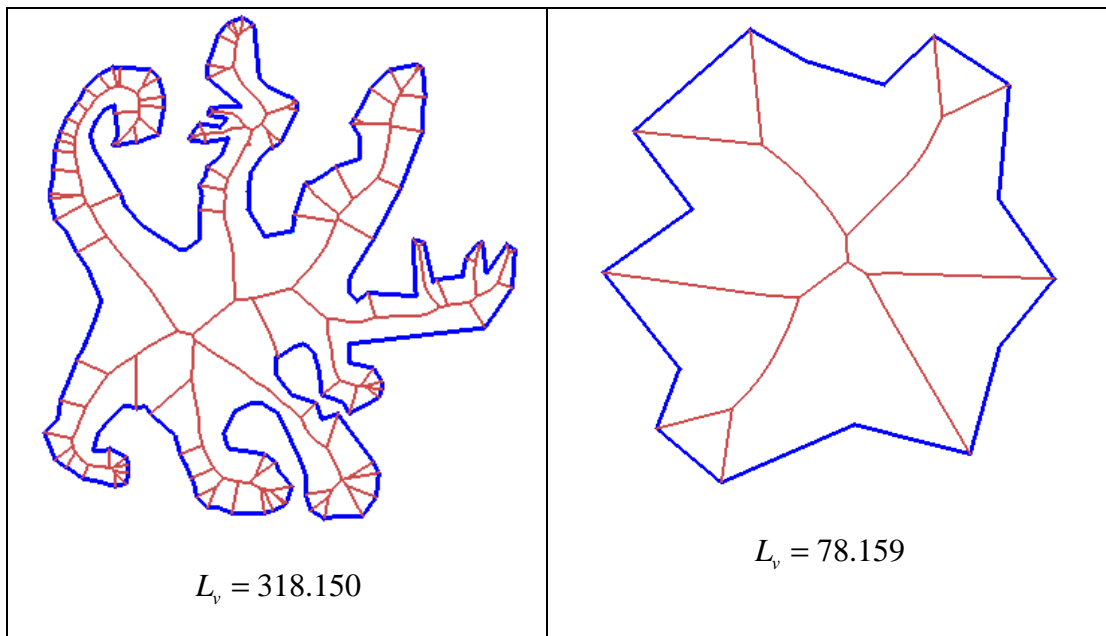


Fig. 4.8 Visual Length Results for An Object with Many Protrusions

An object with many protrusions and a bloby object. The visual length for the bloby object is significantly lower.

Fig. 4.9 shows the local visual significance of a tubular object. Notice how the saliency increases at the ends of the figure, as suggested by [Kovacs98]. The saliency drops abruptly at the end of the figure, however, in contrast to the tapered drop-off that was

measured in [Kovacs98]. This is due to the fact that all measurements in this work are constrained to be on the underlying Blum MAT. The MAT around figure ends is a complex network of axes, and the significant culling and masking of axes that is described in Chapter 3 is necessary in order to perform stable computations.



Fig. 4.9. Saliency Results for a Tube

Local visual significance of a tube object, showing increased saliency at figural ends.

The local saliency for several objects with a protrusion is shown in Fig. 4.10. As the protrusions grow in width, the visual potential of the main figure grows near the protrusion. This demonstrates the interaction among figures as their relative scales become progressively similar. In the middle steps of the progression, the parts of the object have similar scale and there is no clear parent/child relationship; in these configuration, all of the figures affect the saliency of each other. As the original protrusion continues to grow in width, it becomes the main figure and the original parent figure becomes two subordinate figures.

Fig. 4.11 presents two objects with indentations of different widths. Perceptually, the first object is a thick tube with a narrow indentation and the second is a much thinner, curved tube. The underlying visual potential imposed on the internal MAT varies greatly as the indentation grows in width. Again, the topological complexity of the underlying Blum MAT causes these problems. For indentations, the internal MAT undergoes a drastic transformation. The external MAT would provide a different picture of indentations, however. There is a single external figure that represents the indentation, and the width of this figure, along with its local saliency, are easily tracked. A more suitable medial repre-

sentation for indentations would be a multi-scale representation that maintained a single, straight tubular figure at a large scale and then added a varying scale indentation.

Objects with multiple indentations are shown in Fig. 4.12. The object in Fig. 4.12c with wide indentations is likely to be perceived as two blobby figures connected by a thin bridge, and this is borne out by its map of local visual significance. The object in Fig. 4.12a, on the other hand, is perceived as a single tubular figure with two indentations. Its local visual significance map, however, would indicate that it is really two blobby figures like in Fig. 4.12c. Again, this is a limitation of this method that could be addressed with a multi-scale medial model instead of the Blum MAT.

Figs. 4.13 and 4.14 show more complex objects and their local visual significance. In general, regions with larger width show a higher saliency, such as the head and body of the lizard and cat and the central region of the maple leaf. Distinct figures such as the legs of the lizard the the cat's front leg are clearly separate, while axes in the more ambiguous regions such as the rear legs of the cat and the middle of the maple leaf all display saliency. These reflect the substance and connections described in Chapter 3.

Fig. 4.13c shows how the ends of the lizard's leg figures have increased saliency. The cat's body, at the junction of the axes to the rear leg and the tail, also has an endness response that is reflected in the decreased saliency of the tail and leg axes. This same effect can also be seen where the cat's head connects to the neck. Figures that have more pointed ends instead of rounded ends do have higher saliency, and this is reflected in regions like the points of the maple leaf in Fig. 4.14c.

In this initial implementation, the running time for the lizard with 59 boundary segments is 21 seconds on a Sun Ultra 10 workstation. This is to compute the local visual significance that is shown in Fig. 4.13b on all internal and external figures. For a cat object with 57 boundary segments, the same computations take 34 seconds. For a more finely tiled version of the lizard with 152 boundary segments, the running time to compute internal and external saliency is 3 minutes 28 seconds. These timings reflect a prototype system with no optimizations.

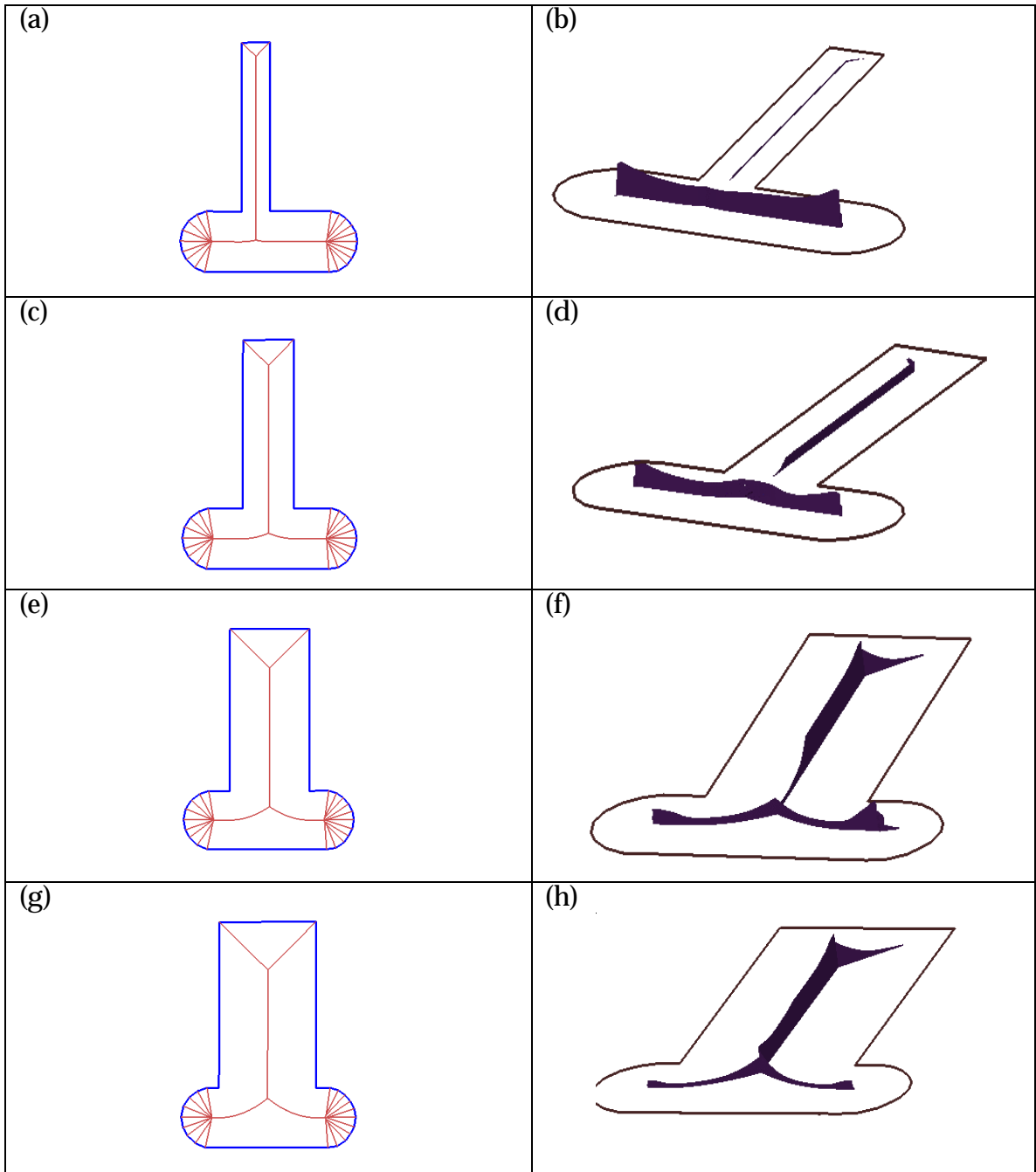


Fig. 4.10 Saliency Results of An Object with Progressively Widening Protrusions

An object with a progressively widening protrusion, along with the local visual significance of each member in the series. In (a), the protrusion is clearly a subordinate figure to the main body of the object, and in (g) the “protrusion” has grown to become the main figure itself. In (c) and (e), there is perceptual ambiguity about which the parent and which are the child parts of the object.

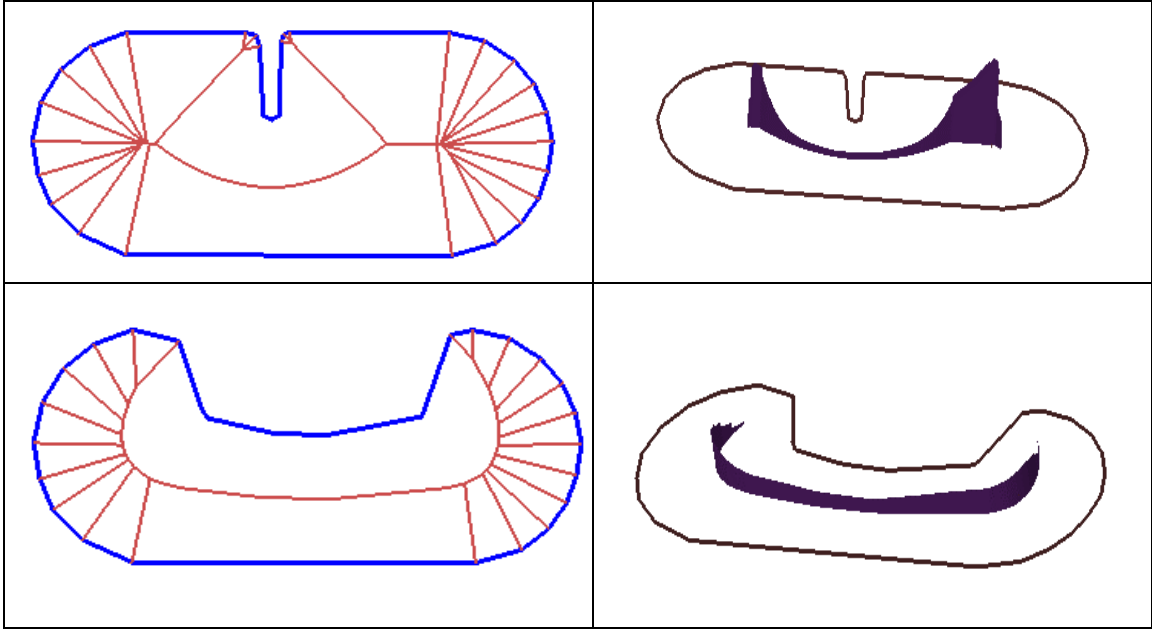


Fig. 4.11 Saliency Results of an object with Indentations

An object with a narrow and a wide indentation, along with the local visual significance of each.

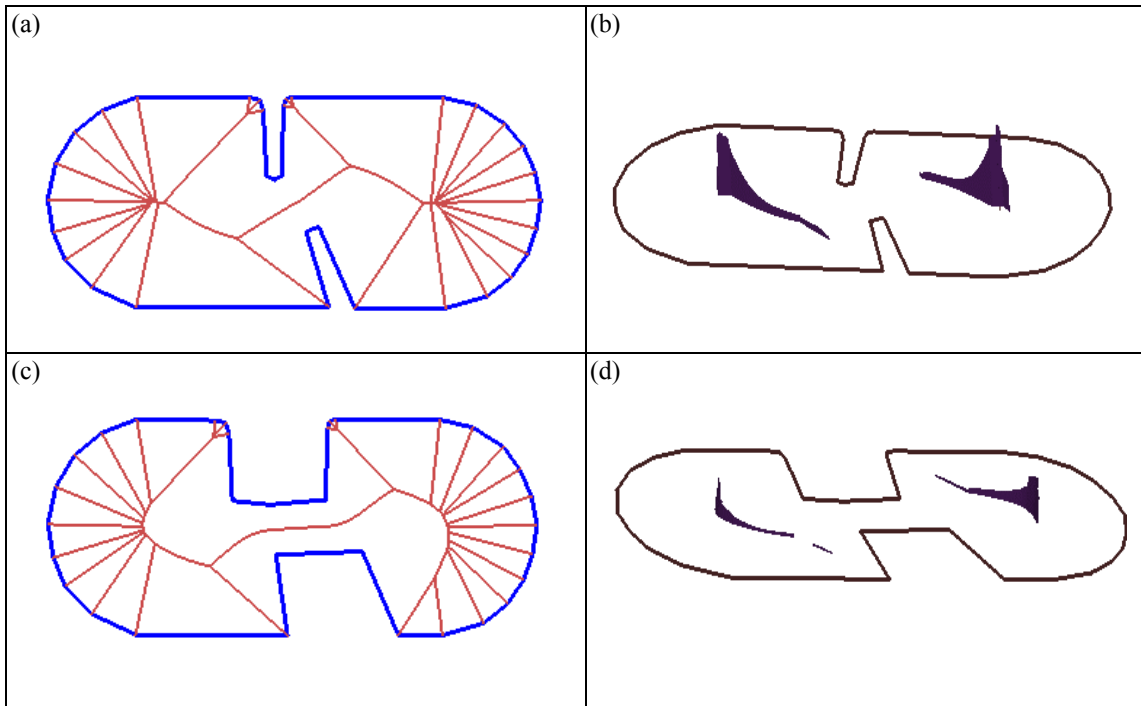


Fig. 4.12 Saliency Results of Two Pinched Objects

Two pinched objects with opposing indentations, and their local visual significance.

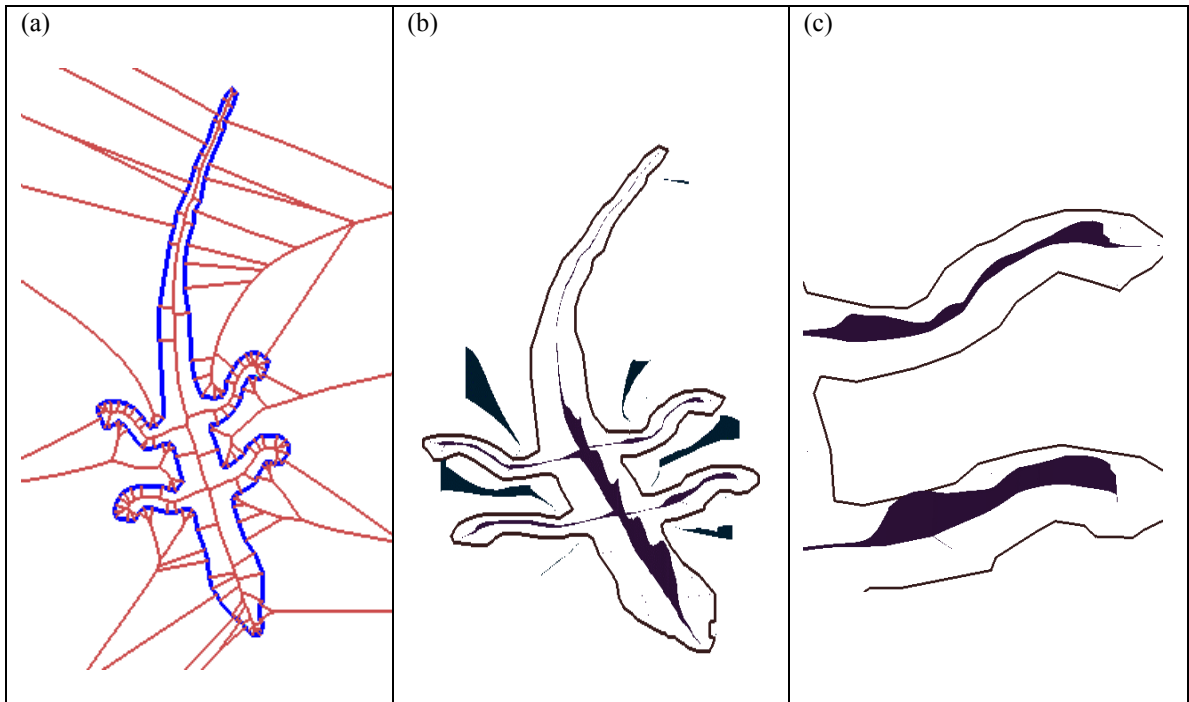


Fig. 4.13 Saliency Results of a Lizard Object

(a) A lizard object, and (b) it's local visual significance. (c) shows a magnified view of the legs demonstrating the increased saliency at the ends of the figures.

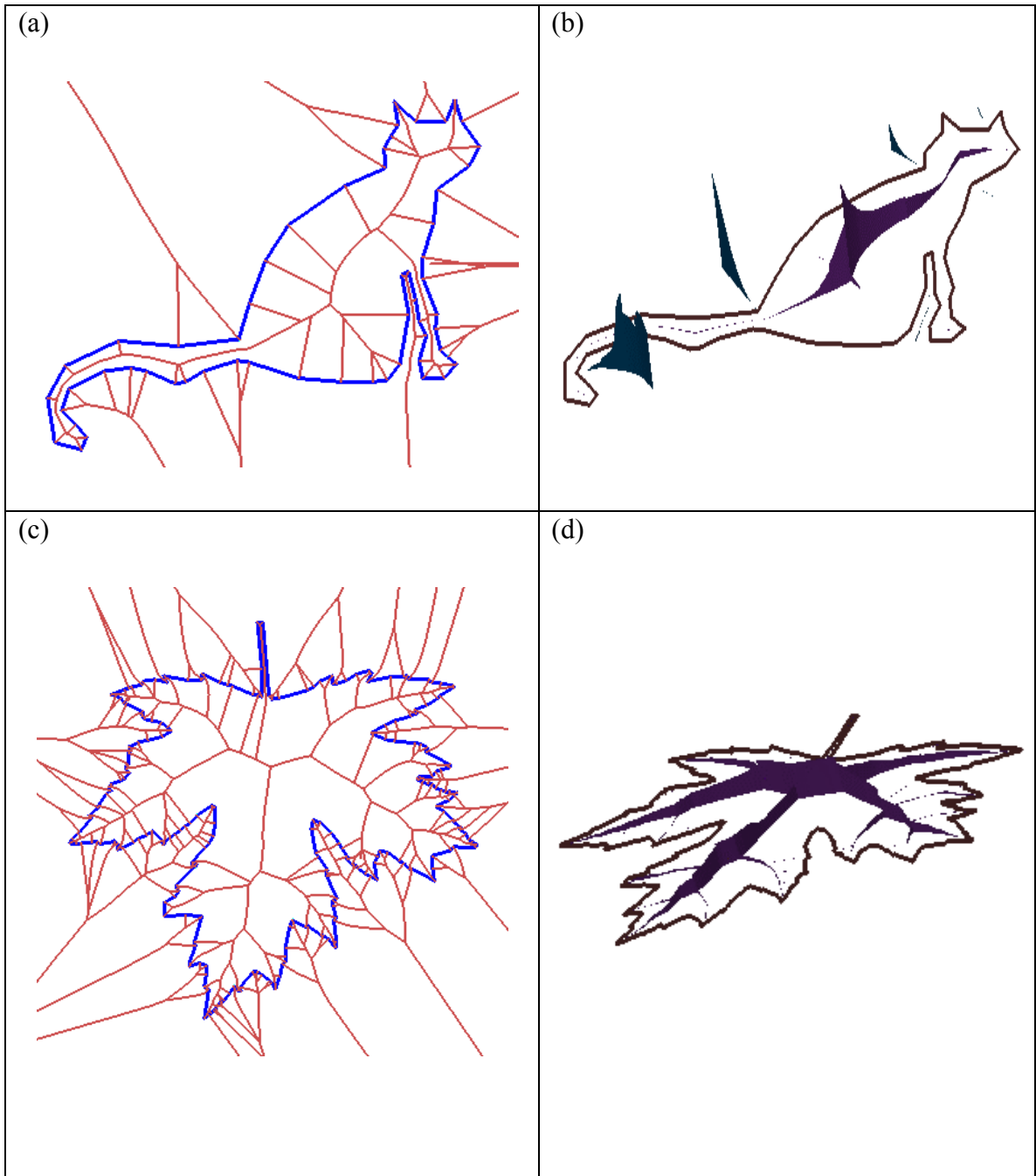


Fig. 4.14 Saliency Results of Cat and Maple Leaf Objects
Cat and maple leaf objects, and their local visual significance.

4.6. Conclusions

This work claims that the scale-based aperture size through which an object is perceived, the spatial relationship measures between features on an object, and the bilateral relationship between opposing boundary points are fundamental elements of form saliency.

These elements produce *figures*, which I propose are the basic visual building blocks of form.

Using physiological and psychophysical studies of the human visual system, metrics have been developed for figural properties such as the mass of a figure, its elongation, the visual continuity between figures and the end response of a figure. These metrics, as well as the relationship between internal and external figures, have been used to produce a saliency metric for each medial point of an object that reflects its visual impact relative to all other medial points on the object.

This research reflects certain aspects of form and is specifically focused on form in preattentive vision. While this work calls for further testing with user experiments, these saliency metrics capture form properties in ways that directly relate to properties demonstrated in psychophysical studies. This has been demonstrated on a variety of objects. Among the form properties worthy of further study are figural curvature, the relationships among groups of figures, object topology and the entire domain of attentive perception, all of which are intentionally unaddressed in this work.

CHAPTER 5 PERCEPTION-DRIVEN OBJECT SIMPLIFICATION

5.1. Perceptual Metrics for Object Simplification

Object simplification algorithms are used in graphics to create new objects that are visually similar to the original object but that are less complex and thus more economical to render. Saliency metrics have obvious possibilities for ensuring that simplifications are visually similar to an original. The goal of this research is to investigate the application of perceptual metrics to rendering algorithms, and object simplification is the driving problem behind the work.

For determining that simplified objects are less complex and thus more rapidly rendered, the hypothesis is that reducing the visual complexity of an object can lead to a lower representation cost and therefore more efficient rendering. That is, by reducing the proper perceptual measures of an object, it can be represented and rendered at a lower cost. Metrics have been developed in the preceding chapters that quantify important form properties, and these metrics have been used to form measures of complexity and of saliency. For the sake of this chapter, I will assume that these metrics do accurately measure complexity and saliency in ways that mirror our perceptual response. This leads to the following question: “how does a perception-motivated simplification approach lead to different simplifications than the classical geometrically motivated methods?” This chapter demonstrates that there are qualitatively different results produced by applying medially based form metrics to the problem.

In general, simplification algorithms are iterative processes that produce a series of LODs. Every algorithm must include a method to determine the ordering of simplification operations – what is the schedule for each primitive or region of the object to be simplified. Every algorithm must also specify its simplification operations; how are the

simplifications effected. Finally, every method also includes metrics for ensuring that the simplifying operations maintain as much similarity as possible to the original object and metrics for measuring the reduction in object complexity.

As detailed in Chapter 2, most of the research done until recently on simplification algorithms has used geometric measures to determine the degree of simplification and the reduction in object complexity. Nearly all scheduling of the simplification operations are also based on geometric measures, and the operations themselves are all motivated geometrically. The work that has been done on applying perceptual measures to object simplification has all focused on image-based metrics.

What this research provides is an investigation into how form-based metrics can be applied to object simplification and an investigation into the new possibilities that such an application may open. The next section describes an approach to perception-based simplification, and the following section describes the results produced by the method. These results are then visually compared to results produced by a geometric-based algorithm to examine the qualitative differences.

5.2. Perception-based Simplification

5.2.1. Simplification Schedule

The first task in designing a simplification algorithm is to decide how, at each iteration, a location will be selected to perform the simplification operations. This decision making process defines the schedule of operations. Taking a perceptual approach to the problem suggests that we want to simplify at the location on the object that will least affect the object's perception. What point on the object will allow modification to itself and its surrounding region with the minimal change in our perception of the object?

For this work, the point with the least local visual significance is used as the location for simplification. Since visual significance is a measure of saliency based on the object's

form, it is a natural choice for this purpose; applying simplification operations around this point will least affect the way we perceive the object.

As described in Chapter 4, indentations into an object are not easily represented with internal figures. The boundary points at the tip of a sharply ended indentation may have the lowest saliency and be ideal for a simplification location, but they can not be identified with internal figures. However, each indentation is directly represented by a single external figure. The saliency of the indentation is also directly reflected in the visual significance along the external figure. Thus, searching for the lowest saliency among points on all internal and external figures will provide a simplification point that will allow for simplification with the least visual impact.

5.2.2. Visual Complexity and Visual Similarity

Another important choice in designing a simplification method is the metrics that will be used to track the degree of simplification and for comparing the visual similarity of an LOD to the original. As discussed in Chapter 2, the number of primitives in an object's representation is used almost universally to measure the degree of simplification. In this case, the simplification process directly simplifies an object's *representation*.

Again, taking a perceptual approach has led to the idea of simplifying the object's underlying form. The hypothesis is that reducing an object's perceptual complexity will lead to representations that can be rendered more quickly. Perceptually, the elongation of an object is a measure of its complexity. Objects with many tentacles of protrusion and indentation have more visual complexity than more circular, blobby objects.

Taking a figural view of objects, each internal figure reflects a protrusion from an object, and each external figure corresponds to an indentation into an object. Each figure protruding from the a prominent part of an object, such as limbs from a body, adds visual complexity. Likewise, each figure extending *into* a prominent part of an object also adds complexity, as in the indentations that define a neck between head and body. Each of these figures adds complexity to the object that can be tracked by its elongation, and simplifying away these figures removes complexity and reduces total object elongation.

Visual length has been designed to measure object elongation. Integrating visual length along all the figures of an object, both internal and external, provides a measure of the total elongation of the object. This measure is used here to monitor the degree of simplification.

The goal of object simplification is to produce LODs that are perceived to be as similar as possible to the original, unsimplified object. The visual significance metric developed in the previous chapter is a simple and concise measure of the visual impact of an object. Visual significance is used in this simplification method to quantify the visual impact of each LOD, and the difference in visual significance between objects is used as the measure of their visual similarity. Maintaining a constant visual significance before and after applying the simplification operations maintains the visual impact of the object. As the simplification progresses and the object becomes significantly simplified, its impact on the user will remain close to the impact of the original object.

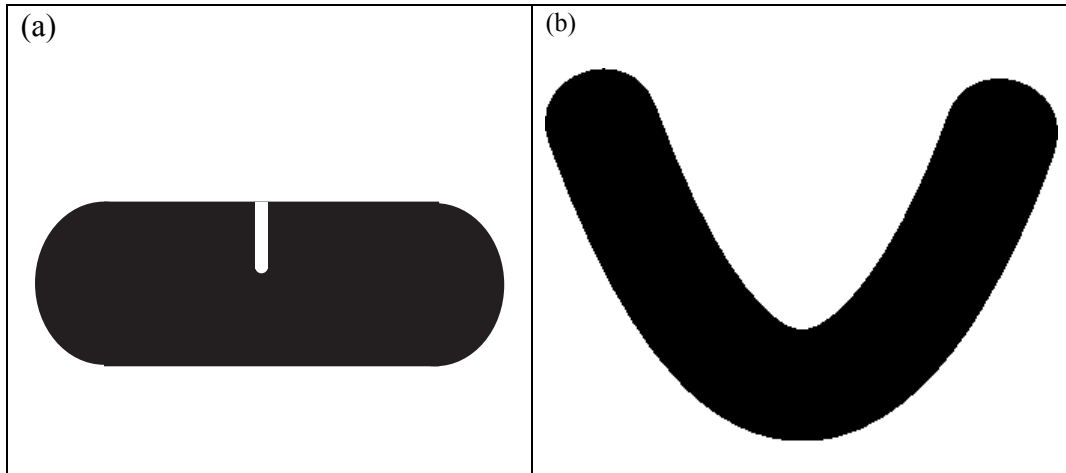


Fig. 5.1 Visual Significance Measured on External Figures

- (c) A narrow indentation into a wide tube would add a small amount of visual significance to the tube.
- (d) A wide indentation into a narrow, bent tube might add far more significance than the tube alone presents.

The visual significance of objects is measured over only internal figures. One might expect that the significance could be measured over both internal and external objects; for

example, a tube with a narrow indentation in the middle presents to a viewer both the visual impact of the tube plus the added visual impact of the indentation. However, a long, narrow tube bent into a wide “U” shape would then possess an indentation whose visual significance would be much greater than the tube itself (see Fig. 5.1). This problem was not addressed in developing the perceptual metrics of Chapter 4. For the simplification method developed in this research, it was found that calculating visual significance over only internal figures provided consistently better results than including both internal and external figures.

5.2.3. Simplification Operations

Once a location for simplification is identified, the operations used to effect simplification are, again, almost universally based on geometric transformations. Using a strictly perceptual approach leads to a different direction. The goal of the simplification process suggests a constrained minimization procedure. We want the successive LODs to have as little complexity as possible while maintaining visual similarity. Using the measures defined above, this translates to modifying the boundary around the selected simplification point in a way that minimizes the visual length with the constraint of maintaining a constant visual significance over the entire object.

5.3. A Perception-based Simplification Method

With the preceding motivation, a perceptual simplification method can be written as a local optimization algorithm. The first step is to calculate the locally integrated visual potential, $Pot_v \Delta s$, at every point on an object’s figures and then identify the point with lowest $Pot_v \Delta s$. This point becomes the focal point on the object, around which a simplifying transform is applied to produce a new, simplified object in the second step. The entire process is repeated on the new object, to produce another, even more simplified object. This can continue until a desired level of simplification is achieved. When the process is done, the simplified objects can be re-tiled in the final step into simplified representations that are more economical than the original object’s representation.

5.3.1. Step 1: Identify Point with Lowest Visual Significance

The first step of the simplification process is straightforward; calculate $Pot_v \Delta s$ for every point on every figure and find the point with the lowest value. This point becomes the focal point for the simplification transform. Points on both internal and external figures are considered here, since the external figures might identify a simplification location that is undetected by internal figures.

Step 1: Find s_0 s.t. $\forall u, Pot_v(s_0) \Delta s < Pot_v(u) \Delta s$.

5.3.2. Step 2: Apply Simplification Transform

The next step of the simplification process is to apply the simplifying transform to the object at the point of lowest visual significance. The process of simplifying an object is implemented by performing a constrained minimization around that point. The total visual length of the object is the objective function to minimize, the total object visual significance is the constraint function to maintain, and sample points on the boundary are the free variables to adjust in order to reach a minimum.

Recall that for the objective functions, object complexity is measured by the visual length of an object over all internal and all external figures, and for the constraint function, visual similarity among LODs is measured using visual significance measured over only internal figures.

To provide free variables for the minimization, the boundary is sampled within the collar around the point of lowest visual significance. The collar was initially defined in the discussion of Cores in Chapter 2, and the size of $1.5R$ was developed in Chapter 4. Sampling is performed by taking samples along the axes within the collar, and using the involutes related to each medial sample point as free variables. Samples are taken at a sampling rate of n samples per R , the radius at the minimization focal point, where n is a user parameter. The boundary points beyond the collar are also allowed to move toward or away from the focal point in order to allow shrinking or stretching of the object during minimization. These movements are also free variables during the minimization.

As can be seen Fig. 5.2, sampling objects in this way actually *increases* the number of

primitives in the object's representation. Recall that the goal of perceptual simplification is to lower the perceptual complexity of the perceived form and then to take the final, simplified form and re-tile that with a small number of primitives.

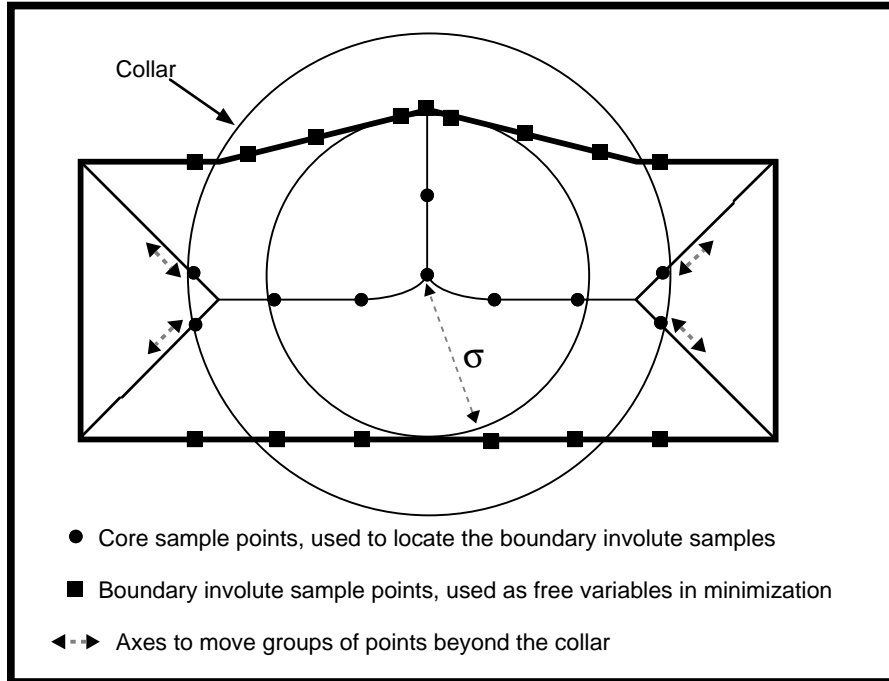


Fig. 5.2 Boundary Sampling Around Point of Least Visual Significance

For minimization, the free variables are sample points on the boundary within the focal point's collar plus the distances that each group of points beyond the collar moves towards or away from the collar. In this example, the smaller circle is the medial width at the focal point, the outer circle is the extent of the collar, and there are two samples per sigma.

The total visual significance of a simplified object must fall within some tolerance ϵ of the original object in order to satisfy the constraint function. This constraint tolerance is also a user parameter. In practice, a tolerance of 2% provides good results.

$$\text{Step 2: minimize } L_v^{total} \text{ s.t. } |Sig_v^{new} - Sig_v^{orig}| < \epsilon$$

Both a genetic algorithm-based minimizer as well as a non-linear optimizer based on sequential quadratic programming were used for this step, and both failed to produce reliable results. They regularly stopped in local minima that often yielded geometrically tiny and visibly imperceptible changes to the boundary.

In the end, to complete this research I manually performed the minimization for each simplification iteration. To speed the process, I identified, at each iteration, all the points whose local visual significance, $Pot_v(s_0)\Delta s$, was less than a threshold level. This level was, at most, 1% of the highest local visual significance found on the object, and it was often less than 0.1% of the highest value. Around each of these points, the boundary was sampled using the sampling method described above, and only the boundary points identified by the sampling were altered.

To perform the minimization, all eligible boundary sample points were manipulated to create a new boundary, and the total visual length and visual significance were calculated. At least five such boundaries were created for each iteration, and the boundary with the lowest visual length that maintained a visual significance within 2% of the original became the new LOD.

To create a new boundary, the eligible boundary points were not randomly manipulated. Instead, several guidelines were applied based on the experience of many trials with many different objects. The first operation performed was to round any polygonal corners that were identified by “lowest visual significance” axes. To minimize visual length in these regions, the boundary sample points were formed into a circular arc whose center was a point on the lowest-visual-significance axis that produced the sample points. A second operation was to widen or narrow matching involutes produced by the boundary sampling process. The third technique that was applied was to move groups of boundary points that were outside the sampling collar, as described in Fig. 5.2.

While this manual approach certainly did not produce a true optimization for each LOD, it did yield simplifications where each LOD measured lower in visual complexity while maintaining visual significance.

5.3.3. Step 3: Iteratively Repeat Process

The process described above represents the heart of this simplification scheme. By iteratively applying the first two steps, objects can be simplified to any degree desired, and many levels of detail can be produced. After each iteration, consisting of steps 1 and 2

above, a new object is created. Each point on the figures is ranked anew for visual significance. Finding the lowest value among those points provides a new focal point for the next simplification transformation.

5.3.4. Step 4: Re-tile Simplified Objects into Economical Representations

Using perceptual information to simplify shapes does not guarantee a simplified form *representation*. Without using the same means to represent an object as our brains use, a perceptually driven simplification process may end up with a simplified object that requires more primitives to represent it. Objects become, in general, rounded and blobby when simplified with this method. Using polygonal pieces to represent these smoother shapes requires more primitives rather than fewer. This weakness is caused by the same thing that gives this approach its advantages; this method simplifies an object's form rather than its representation.

Fortunately, there is a simple and effective way to solve this problem. Since most of the additional primitives are created to represent a smoothing of the object, a single, larger primitive can easily replace many small ones. Boundary regions that are gently curving, requiring many primitives, can be represented using one or a few primitives with very little loss of object information. In effect, the difference between the smooth representation and a more coarse one is only small scale information that can be safely removed.

This re-tiling of the simplified objects may appear to be identical to many existing simplification methods. A valid question can be asked, "Why not re-tile immediately and skip the expensive simplification process?" What shape simplification adds is the removal of less important visual information from the object so that re-tiling is performed only on the visually important areas. Areas with high geometric complexity but low visual importance are simplified away.

The re-tiling method used here is the classical Douglas-Peucker line approximation algorithm [Douglas73], as optimized by [Hershberger92]. In this algorithm, 2D polygonal curves are reduced to a minimal representation whose points are all within an error toler-

ance ϵ of the original curve (see Fig. 5.3). This algorithm is used later, in the next section, to produce geometric-based simplifications for comparison with the perceptual method.

Using a common geometric simplification method like Douglas-Peucker allows me to characterize the effectiveness of these perceptual simplifications in terms of the number of line primitives needed for the LODs. Line primitives are by far the most common primitives, and counting primitives continues to be the standard for measuring the degree of simplification. Re-tiling the perceptual LODs into a small number of line primitives allows a direct comparison of this method with a geometric method. However, it is not at all clear that line primitives are the best ones for effective representation and rendering. This is discussed further in Section 5.5.

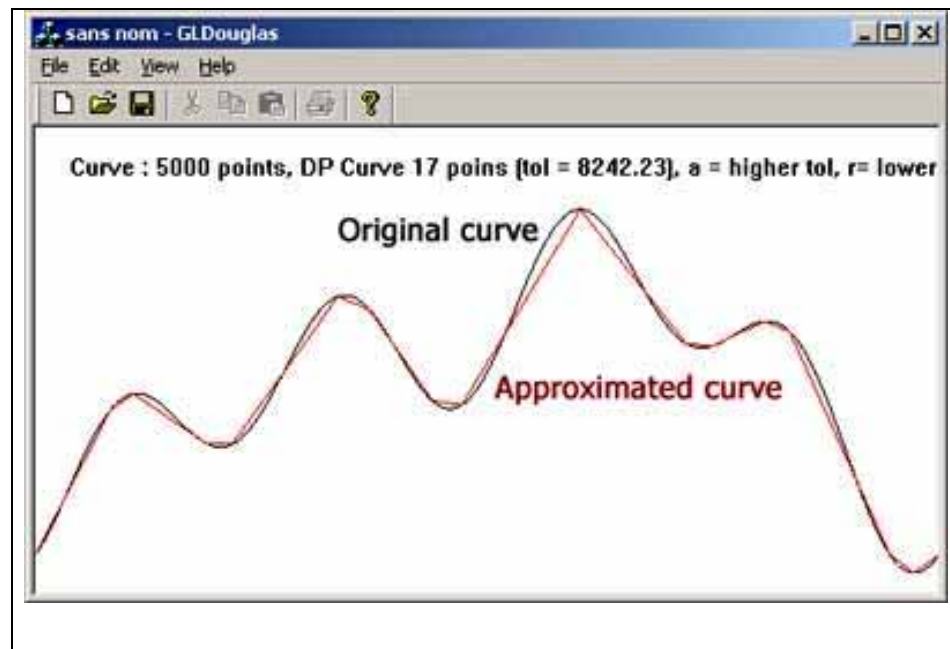


Fig. 5.3. Geometric Simplification Method

Example of original curve and a simplified version created by the Douglas-Peucker algorithm. (Fig. courtesy of Jonathan de Halleux.)

5.4. Results

When simplifying shapes using this approach, the regions scheduled first are, in general, sharp corners. As Fig. 5.4 shows, these corners become rounded by the simplification

process. As the simplification progresses, the corners round with an arc of greater radius. This rounding of corners removes the high frequency components from the object and mimics the effect of a low-pass filter. Carried to the limit, the four corners of a square would be simplified into a circle.

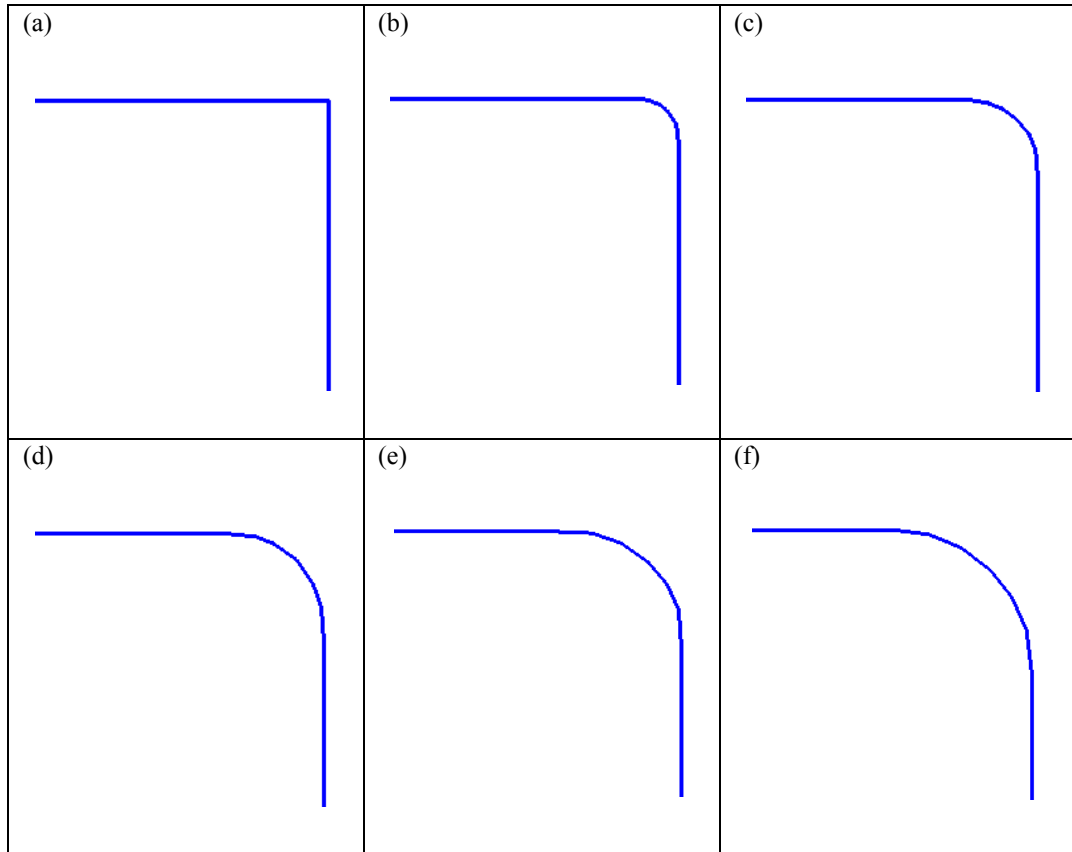


Fig. 5.4. Perceptual Simplification of a Corner

A sharp corner become rounded through the perceptual simplification. These corners are, in general, scheduled first.

Fig. 5.5a shows a tubular object with a protrusion, and Fig. 5.5b – Fig. 5.5h show selected LODs created with the perceptual simplification method applied to the protrusion. In the progression of LODs, the protruding figure can be seen to melt into the main figure of the object. As the protrusion shrinks, the main figure swells to maintain a constant visual significance. In the final LODs, there is no longer a distinct protruding figure because it has been absorbed into the main body of the object. The final simplifications show how the two-figure object becomes a single, bulging figure. Also notable in this example is

the fact that the main figure shrinks in extent as the two figures merge into one; this is necessary to maintain a constant visual significance for the object as the mass of the protrusion is added to the mass of the larger scale main figure.

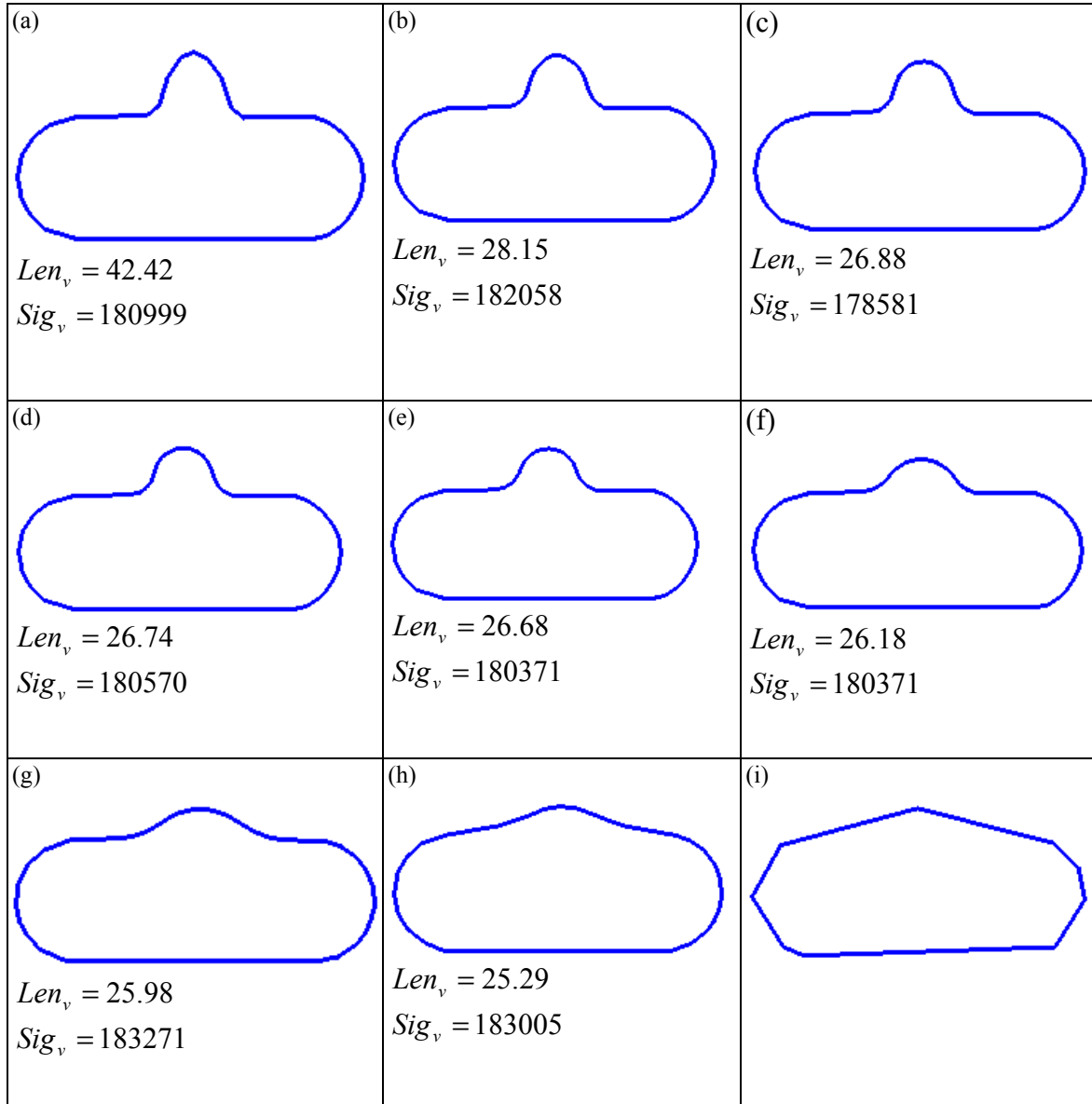


Fig. 5.5. Perceptual Simplification of Object with a Protrusion

(a) A tubular object with a protruding figure. (b- h) Selected LODs created with the perceptual simplification method. (i) A final representation of the simplest LOD in (h) created by applying the Douglas-Peucker algorithm with a tolerance of 10% of the object's maximum dimension.

In Fig. 5.6, the same protruding object is shown simplified using the Douglas-Peucker line simplification algorithm. Each LOD uses an increasingly larger error tolerance to allow the points in the LODs greater geometric deviation from the original object. The geometric algorithm schedules the rounded ends first where it can simplify with little error tolerance. After several iterations simplifying the ends, the protrusion is scheduled. The protrusion is simplified to a pointed pimple, and then the ends and main body are scheduled for the rest of the LODs.

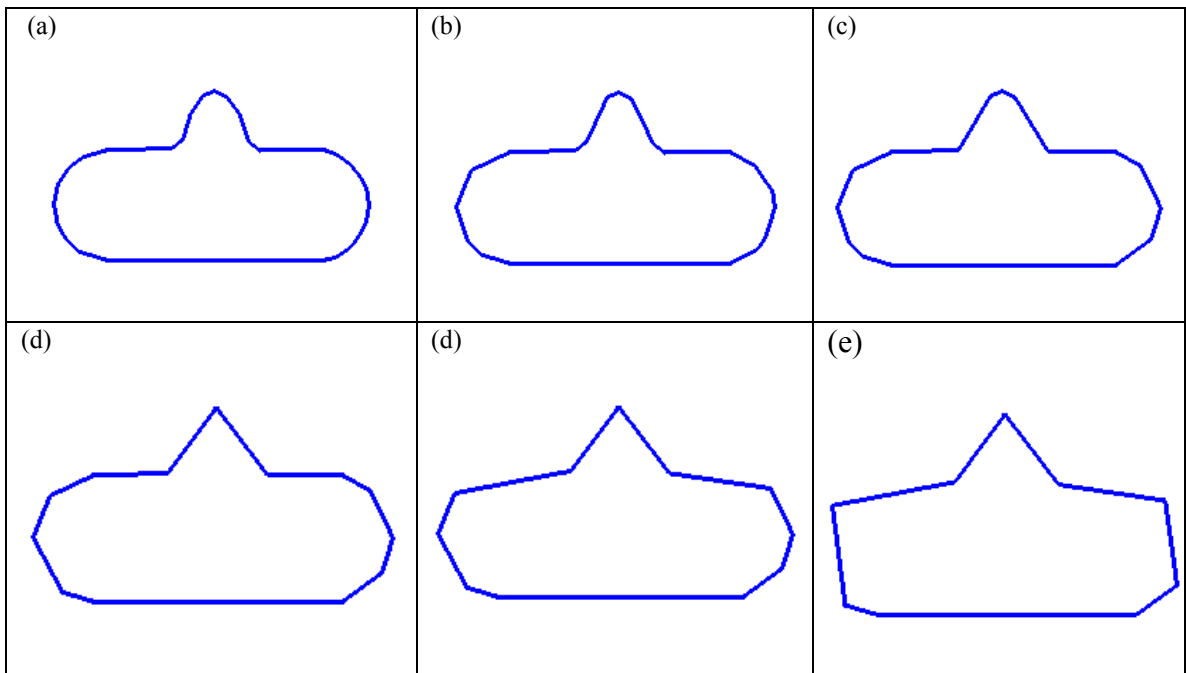


Fig. 5.6. Geometric Simplification of Object with Protrusion

LODs of the tubular object with a protruding figure, generated using the geometry-based Douglass-Peucker line simplification algorithm.

As discussed above, the perceptual simplification method schedules corners first and rounds them. Given the tubular object with rounded ends, this approach simplifies the protruding figure immediately in its sequence. This is in contrast to the geometric method, which schedules the ends first before the protrusion. In general, geometric methods like this will simplify rounded curves first, creating a cruder approximation that uses fewer primitives. In this way, the perceptual scheme and the geometric schemes effect completely different scheduling, with the perceptual method smoothing corners and the

geometric method creating corners from smooth curves. If the object in Fig. 5.5 and Fig. 5.6 had been rectangular, then the perceptual approach would have rounded the corners first before scheduling operations on the protrusion, while the Douglas-Peucker algorithm would have left the corners untouched and immediately scheduled operations on the protrusion.

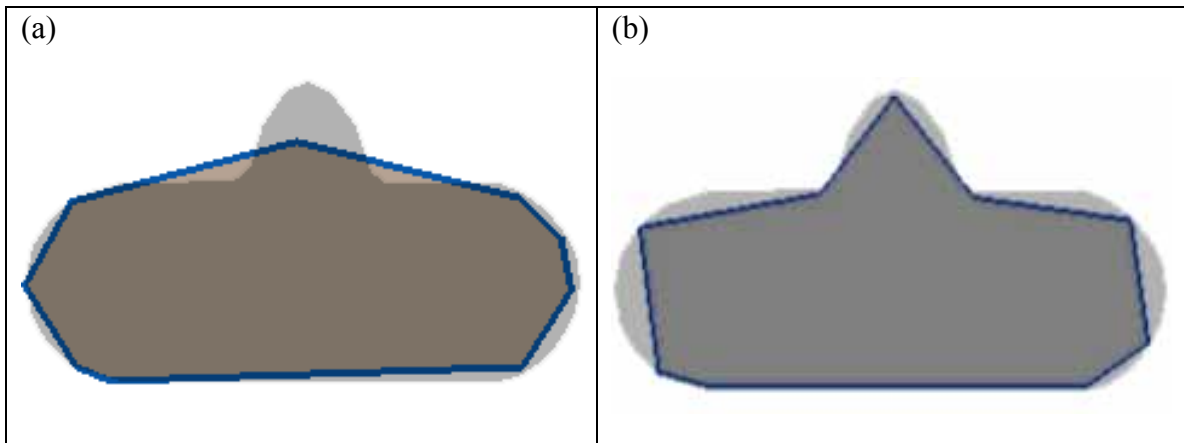


Fig. 5.7 Comparison of Perceptual and Geometric Methods for Object with Protrusion

Comparison of perceptual simplification and Douglas-Peucker geometric simplification, both with the same number of primitives. (a) Re-tiled perceptual simplification overlaid on top of original object with protrusion. (b) Douglas-Peucker simplification overlaid on top of original object with protrusion.

Fig. 5.7a shows the original object with protrusion compared to a perceptual LOD, and Fig. 5.7b shows it compared to a geometric LOD. Both LODs have the same number of primitives. The perceptual LOD has removed the protruding figure and compensated by creating a bulge in the main part of the object. The ends of the tubular body also retain some degree of roundedness. In the geometric LOD, the protrusion is retained but the rounded nature of the main body is degraded. This example demonstrates the tendency of the perceptual approach to create and preserve rounded areas while removing protruding figures, and it shows the geometric methods tendency to create corners in rounded areas and to follow the original boundary as closely as possible. The geometric deviation of the perceptual LOD in Fig. 5.7a is 12.78 distance units, and the deviation of the geometric LOD in Fig. 5.7b is 8.89 units. This shows how the perceptual method permits larger

changes in the underlying geometry of the object while maintaining a measure of visual similarity.

Fig. 5.8 shows a tubular object with an indentation and several of its LODs. This example shows how the simplification of an internal indentation can be considered to be a simplification of an external protrusion. The filling in of the internal indentation corresponds to the external protrusion melting into the exterior space around the object. The simplified versions that are later in the progression show how an object with indentation becomes a single figure object with a narrowing of the main figure replacing the separate indentation figure.

In Fig. 5.9, the same indented tubular object is shown simplified using the Douglas-Peucker algorithm. As in the case of the protrusion, the rounded ends are scheduled for simplification first. Then the indentation is simplified into a sharply pointed inverted pimple, and then the rest of the LODs affect the ends. Compare this to the perceptual simplification in Fig. 5.8i where the indentation is essentially removed. The perceptual LOD could be easily retiled into a rectangle with fewer primitives. While the geometric approach retains the indentation figure, the perceptual approach provides the opportunity to completely eliminate small scale figures. In this way, primitives can be reallocated from perceptually insignificant areas to those that are more important visually. This effect can be seen in Fig. 5.10, where the indentation between the cat's body and front leg is removed through the perceptual simplification.

Fig. 5.10 shows the original object with indentation Fig. 5.10a compared to a perceptual LOD and Fig. 5.10b compared to a geometric LOD. Here, the geometric methods goal of following the boundary as closely as possible retains the full length of the indentation figure. The perceptual approach, on the other hand, has filled in the indentation and leaves the main body with only a slight depression in the boundary remaining from the indentation. The geometric deviation of the perceptual LOD in Fig. 5.10a is 18.46 units, and the deviation of the geometric LOD in Fig. 5.10b is 3.16 units. Even with a large geometric error in the perceptual LOD, the visual difference from the original is arguably moderate.

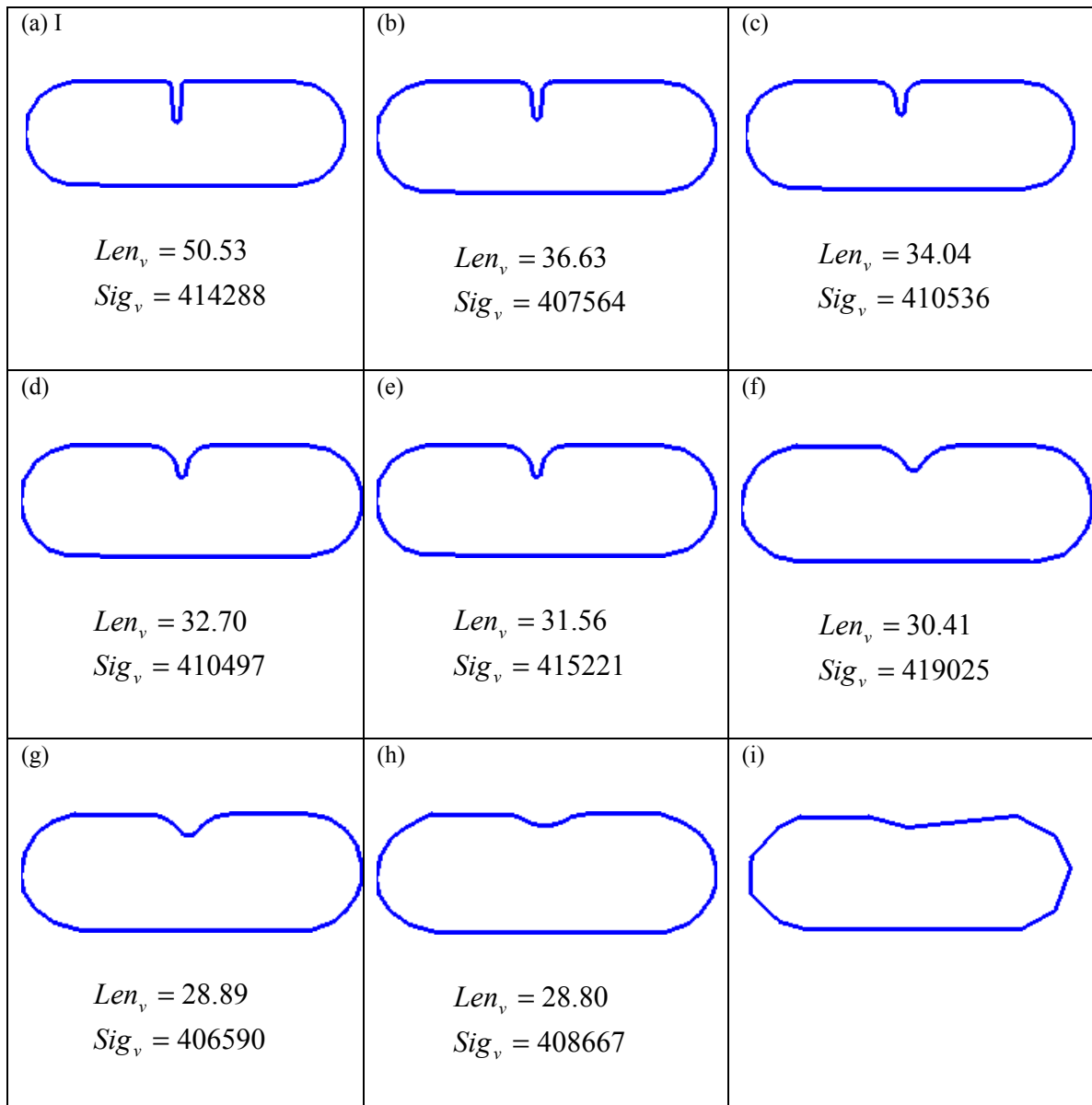


Fig. 5.8 Perceptual Simplification of Object with Indentation

(a) A tubular object with an indentation figure. (b - h) Selected LODs created with the perceptual simplification method. (i) A final representation of the simplest LOD in (h) created by applying the Douglas-Peucker algorithm with a tolerance of 13% of the object's maximum dimension.

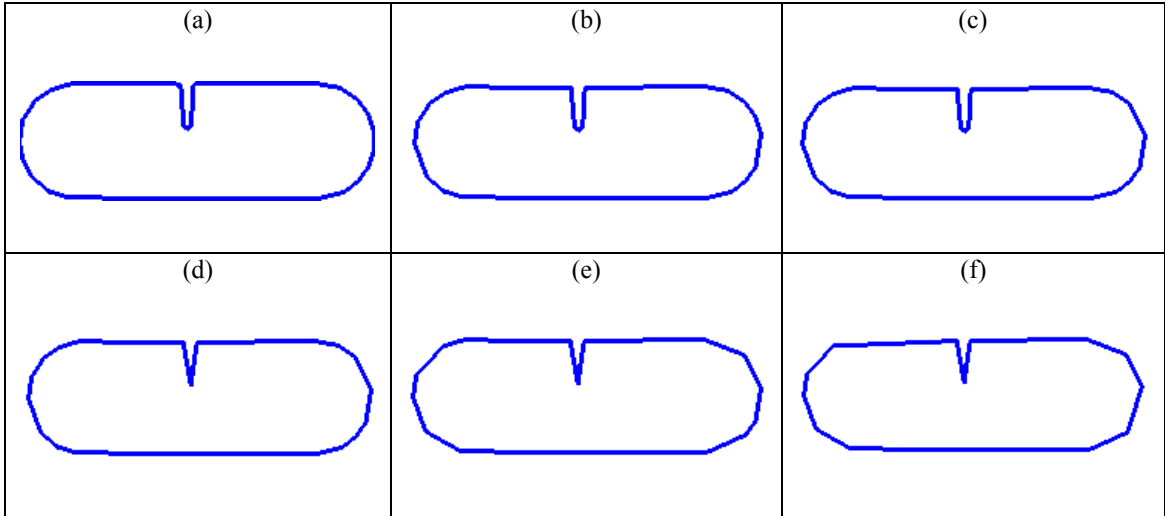


Fig. 5.9. Geometric Simplification of Object with Indentation

LODs of the tubular object with an indenting figure, generated using the geometry-based Douglas-Peucker line simplification algorithm

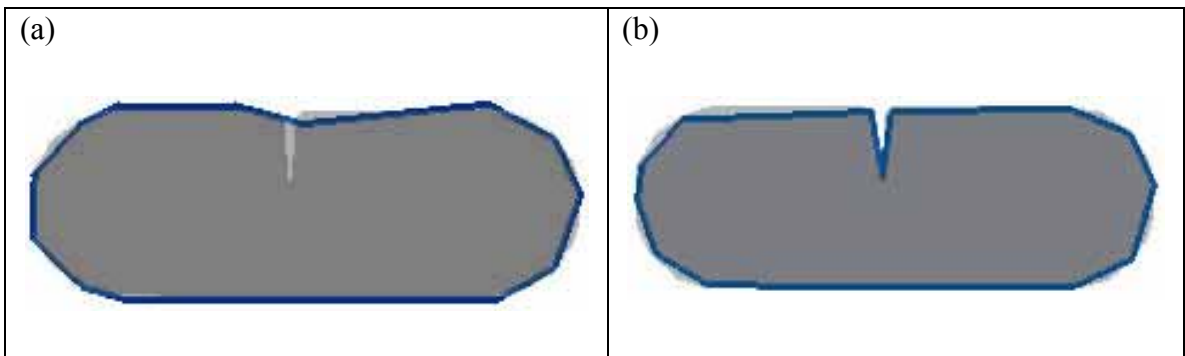


Fig. 5.10. Comparison of Methods for Object with Indentation

Comparison of perceptual simplification and Douglas-Peucker geometric simplification, both with the same number of primitives. (a) Re-tiled perceptual simplification overlaid on top of original object with indentation. (b) Douglas-Peucker simplification overlaid on top of original object with indentation.

Fig. 5.15, found at the end of this chapter, shows LODs of a cat object. In the first LODs of Fig. 5.15, the sharp corners of the object are rounded. Next, the tip of the tail and the front paw are scheduled. Once the object is sufficiently rounded, the least significant figures are simplified. As can be seen in the LODs in Fig. 5.10b and onward, the tail and the front legs are scheduled, but the indentation between the front legs and the body is also

modified. Each of these figures is shortened and, to maintain visual significance, they are also widened in the area where they are shortened. In this way, the tail shortens while the front leg is drawn into the main body of the cat. This process continues through the rest of the simplifications. As part of this process, the LODs in Fig. 5.10f and Fig. 5.10i show how the larger figure of the main body rounds and grows as the protruding figure of the front leg melts. Fig. 5.10j and Fig. 5.10k show how the head becomes a rounded blob. It, too, shrinks in size to maintain visual significance.

The Douglas-Peucker simplification of the same cat object is shown in Fig. 5.16 at the end of the chapter. This approach is similar to many geometric methods in that it replaces two or more primitives with a single one. This is demonstrated clearly in the cat's back over the series of LODs. The original curved back ends up represented by two primitives. Similarly, the detail along the bottom of the cat's body and tail becomes a single primitive. One aspect of most geometric approaches is that they attempt to maintain either original vertex points or new sample points on the original boundary. In this case, the tips of the cat's ears are retained as data points in all but the final LOD, even though the ears themselves are no longer discernible much earlier in the sequence.

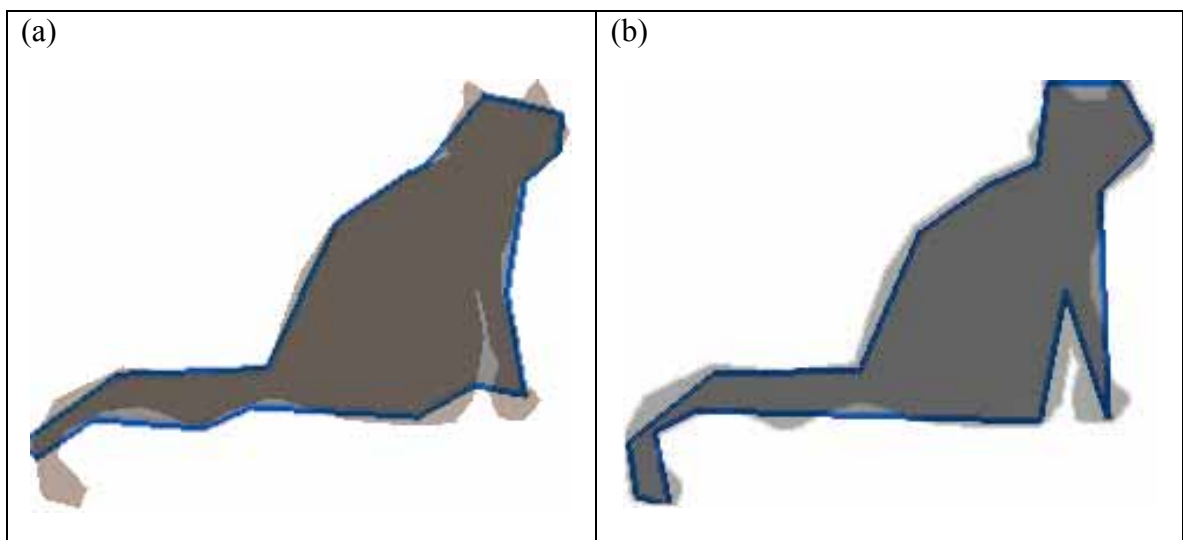


Fig. 5.11 Comparison of Perceptual and Geometric Methods for Cat Object

Comparison of perceptual simplification and Douglas-Peucker geometric simplification, both with the same number of primitives. (a) Re-tiled perceptual simplification overlaid on top of original cat object. (b) Douglas-Peucker simplification overlaid on top of original cat object.

The differences between the perceptual and geometric approaches are highlighted in Fig. 5.11, which compares an LOD created using the Douglas-Peucker algorithm directly from the original cat object to an LOD created by using the Douglas-Peucker algorithm to retile the perceptual LOD of Fig. 5.15k to have the same number of primitives as in the Douglas-Peucker LOD in Fig. 5.11b.

The perceptual LOD in Fig. 5.11a has eliminated the narrow indentation between the cat's body and its front leg, while the geometric LOD in Fig. 5.11b retains the indentation. The perceptual approach has used its primitives, instead, to better represent the curve of the front of the cat's body as well as the profile along the bottom of the body. Similarly, the curve of the cat's tail is crudely represented in Fig. 5.11b while the LOD in Fig. 5.11a removes the curve of the tail but retains more detail along the rest of the tail. Finally, the perceptual method represents the larger scale form of the cat's head minus the ears, while the geometric approach gives more of a bounding polygon around the head and ears together, resulting in an oversized head. In general, the perceptual simplification scheme leads to an LOD that ignores the small scale form information, which is considered to have least saliency. Instead, the scheme uses its primitives to represent the larger scale form information. In contrast, the geometric LOD is heavily influenced by some of the small scale regions of the object.

In Fig. 5.17, found at the end of the chapter, a lizard object is shown simplified using the perceptual method. In this progression, the first LODs show the rounding of the lizard along the entire boundary. Of special note is the rounding occurring where the legs meet the boundary. This rounding causes the body of the lizard to grow wider; to maintain overall saliency, the body is simultaneously shortened. This is particularly visible in the shortening and thickening of the lizard's neck. The tips of the legs are also rounded, and they are shortened as well to maintain the visual significance measure with the original object. Starting around frame (e) in the simplification schedule, the limbs and tail are identified as narrow protrusions and they are progressively simplified. This shows up as a shortening and widening of the figures.

The Douglas-Peucker simplification sequence is shown in Fig. 5.18, also at the end of this chapter. The Douglas-Peucker approach removes any rounding and replaces those areas with a more angular boundary that has fewer segments. The full extent of each protrusion is also maintained, with each LOD extending to the tip of each limb and the tail. The area enclosed by each of the LODs and by each figural part varies widely across the LODs. Because of this, all notion is lost of the “head” of the lizard being distinct from the body and of the feet being distinct from the legs.

The two approaches are compared side by side in Fig. 5.12, with LODs of equal number of primitives overlaid on top of the original object. In the perceptually simplified result of Fig. 5.12a, the visual impact of the object has been condensed into a shorter but thicker representation. However, with this low number of line primitives, the rounding of parts in the perceptual approach is lost, leaving artifacts like the pimple representing the left rear leg. The geometric approach maintains a sense of each limb and the tail, and a general size of the object. It introduces artifacts as well, resulting from attempting to match the zig-zag nature of the limbs. This is seen most in the hind legs, where the LOD’s creates a shape of the legs that does not follow from the original.

Fig. 5.12 is a good example of the general consequences of these two methods. The perceptual approach tends to smooth smaller scale details and condense visual information into a more compact form. The geometric method often introduces detail while following the original boundary as faithfully as possible. It tends to maintain a sense of the extent of an object even if it introduces artifacts into those protrusions or indentations.

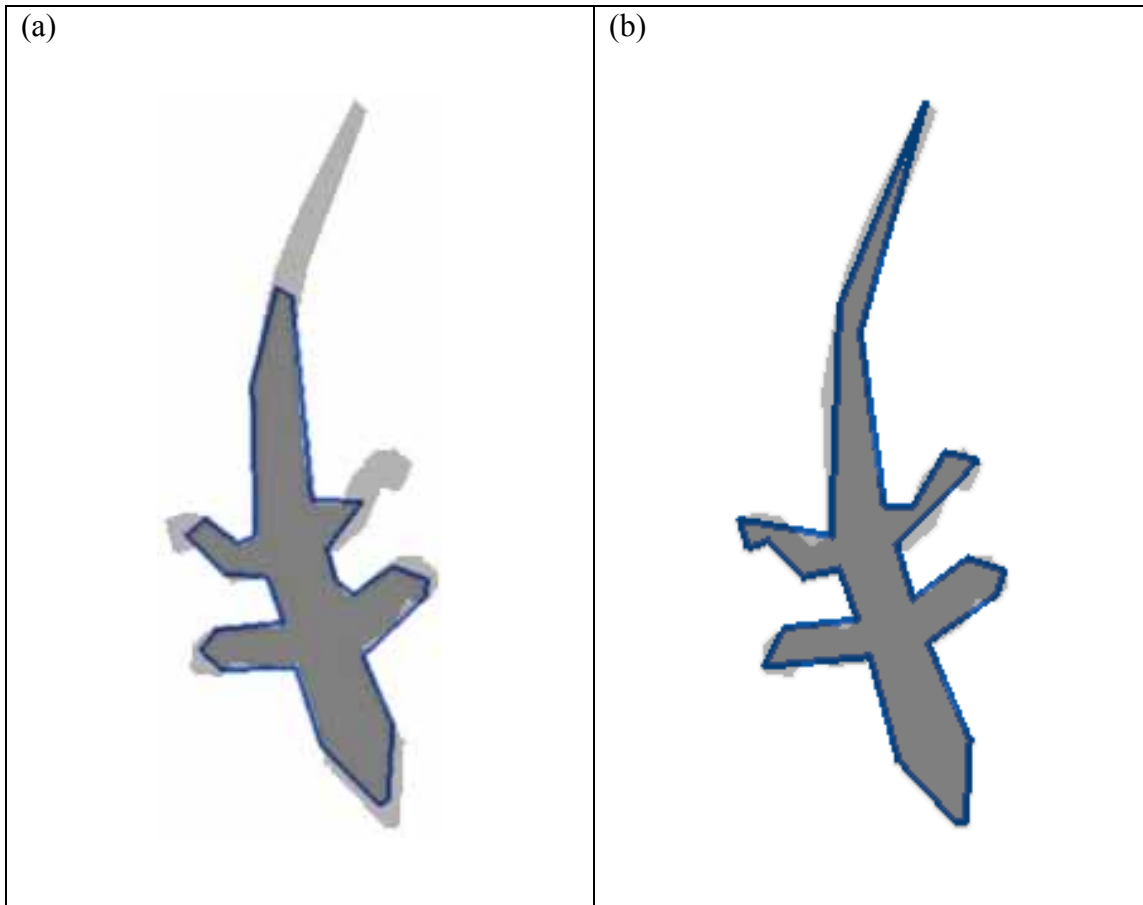


Fig. 5.12 Comparison of Perceptual and Geometric Methods for Lizard Object

Comparison of perceptual simplification and Douglas-Peucker geometric simplification, both with the same number of primitives. (a) Ret-tiled perceptual simplification overlaid on top of original lizard object. (b) Douglas-Peucker simplification overlaid on top of original lizard object.

The Douglas-Peucker algorithm maintains topology although there are many geometric simplification methods that do not. The perceptual simplification operations developed in this chapter were limited so as to not break objects into separate pieces or merge pieces into a single one; that is, they also maintain the topology of the original object. This was a decision intended to limit the scope of this research. Breaking objects into separate pieces is a conceivable operation for perceptual simplification. Fig. 5.13 shows an example where changing topology may be perceptually appropriate. In this case, breaking the object would lessen the measure for internal visual complexity but increase the external complexity. Different scientists will disagree on whether and for what tasks changing to-

pology is perceptually beneficial to object simplification, and this may be a fruitful area for perception researchers to explore.

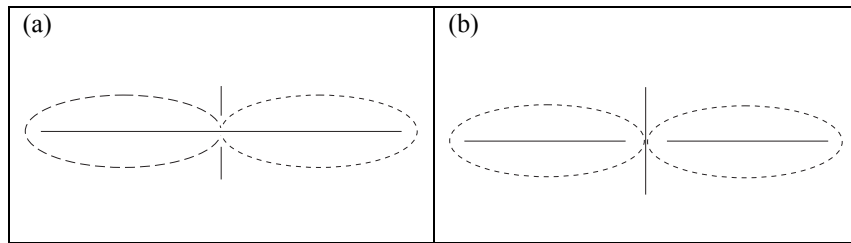


Fig. 5.13 Changing Topology in Simplifications

- (a) An object with a very narrow neck between two main masses, along with its medial axes.
- (b) The same object broken into two pieces, along with the new medial axes.

This example highlights the question of whether or not it would be beneficial to break objects into separate pieces during a perceptual simplification.

Finally, the calculations necessary for optimizing the perceptual LODs have demonstrated the stability of the substance and connection measure described in Chapter 3. In addition to simplifying a wide range of corners, protrusions and indentations, perceptual simplifications were calculated for the objects shown in Fig. 5.14. For each of these objects, the simplification optimization was performed at many locations using both the automatic optimizers and the manual approach. The automatic optimizers, in particular, generated hundreds of trials with small boundary changes. In total, I estimate that I performed tens of thousands of calculations of the local visual significance for the objects in Fig. 5.14. For small boundary perturbations there were always small changes in the metrics; large changes in the metrics were caused only by large changes to the boundary. This was an extensive exercise of the substance-weighted Blum MAT, and it provided consistently stable results.

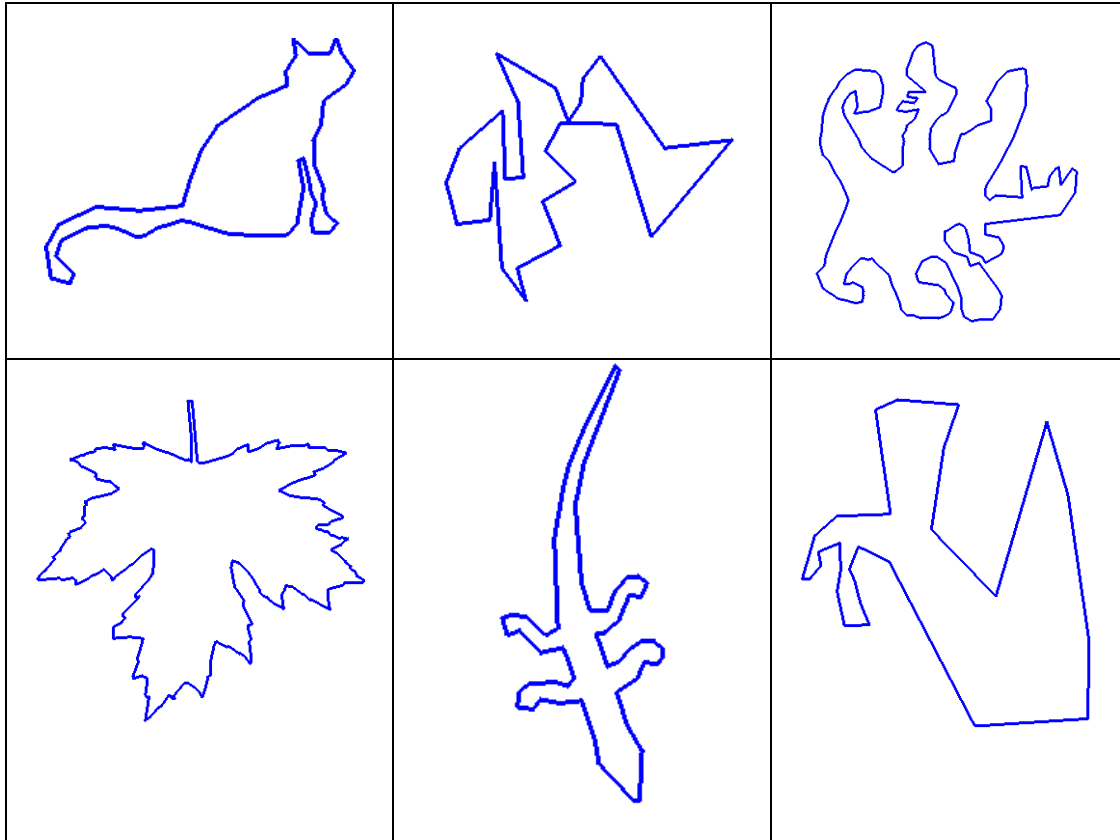


Fig. 5.14. Objects Tested with Perceptual Metrics

Extensive trials with a variety of objects demonstrated the stability of the perceptual metrics described in Chapter 4.

5.5. Conclusions

This chapter has presented a perceptually motivated approach for object simplification. The method uses perceptual saliency metrics for setting the simplification schedule as well as for comparing visual similarity between LODs. A perception-based measure is also used in measuring the level of simplification for each successive LOD. These measures are applied in a method that addresses object simplification as a constrained minimization problem: for each LOD, the method adjusts local regions of the object to minimize visual complexity while maintaining a consistent visual similarity.

The results of the perceptual simplification method have been compared to results from a geometric-based algorithm. The perceptual approach showed simplifications that are qualitatively different than those from the geometric approach. Unlike that geometric ap-

proach that operates on independent boundary points, the perceptual approach identifies protrusion and indentation figures and has the effect of simplifying these perceptual features as a whole. Also, successively simplifying a local region of an object using the perceptual method can affect the object globally. As figural features were removed in the simplification process, the main body of the object was altered to incorporate the saliency of those features. This contrasts with the geometric approach, where, in general, features were removed or modified without altering any of the remaining features of the object. Also, the geometric method had a lower likelihood of removing a protrusion or indentation. In fact, the extremal points of figures were often retained which caused the volume of the main body of the object to differ greatly from the original.

Other geometric simplification methods implicitly capture form perception effects. To what extent do they capture the effects displayed by this method? The Simplification Envelopes method of [Cohen96] maintains sharp corners rather than rounding boundaries, and its local simplification operations do not have any global effects. [Garland97] permits arbitrary vertex pairs to collapse together, and thus its operations can have global effects. However, it still preserves extremal points and so has the effect of preserving figures. The image-driven simplification method of [Lindstrom00] also retains sharp corners and is limited to local effects with each operation. As a general framework, however, it could be extended to use any operation. The results of [He95] most closely follow the perceptual simplifications produced by this research. Recall that their method samples objects into a voxel representation, performs low-pass filtering on the volume representation and then recreates a polygonal model from the filtered volume. Their results show the same rounding of corners as shown here, and there is some “melting” of protrusions and indentations. A detailed comparison of my approach with [He95] might yield useful insights.

The computing time required to analyze a candidate LOD of the cat object ranged from 34 seconds to more than 4 minutes, and sometimes dozens of candidates were examined during the optimization process. This work was intended as an exploration of how to quantify perceptual form cues and apply such metrics to object simplification. The methods described here were merely vehicles for this process.

Finally, this research suggests the need for modeling and rendering primitives other than lines or polygons. This work suggests that using curved surfaces as primitives may match more closely the representations that provide the most important cues in our perception of form. As accelerated rendering support for curved surfaces becomes more widely available, this may be a fruitful direction of exploration. Further, the basis of this work is that objects are substance and not just an outline or boundary. Given this, space-filling primitives such as circles are a more natural means of rendering. The volume “splats” described in [Westover90] provide such a primitive. The results of this research suggest that alternative rendering primitives like this may ultimately provide the most efficient and effective means of conveying form information through computer graphics.

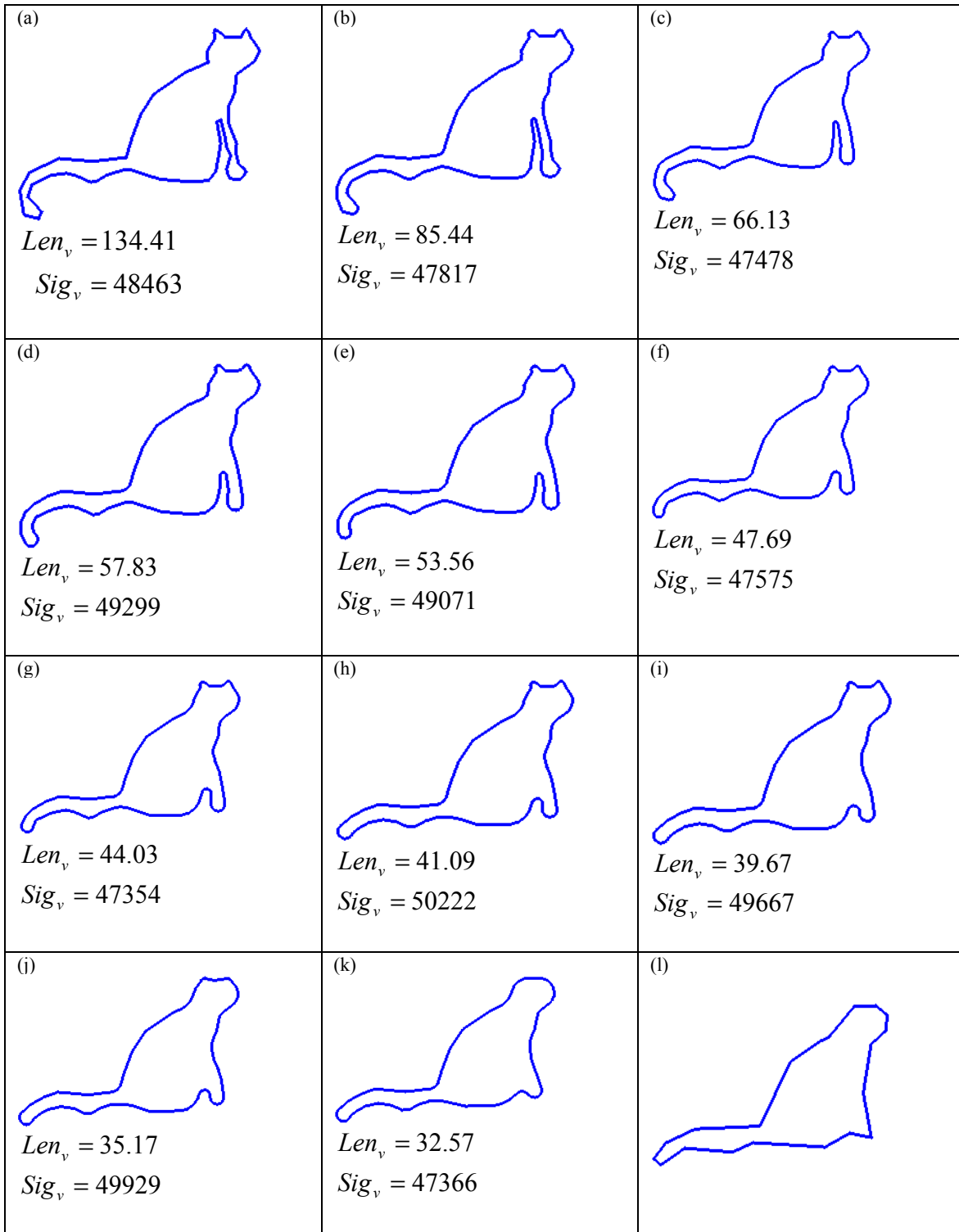


Fig. 5.15. Perceptual LODs of a Cat Object

(a) A cat object. (b - k) Selected LODs created with the perceptual simplification method. (l) A final representation of the simplest LOD in (k) created by applying the Douglas-Peucker algorithm with a tolerance of 3% of the object's maximum dimension.

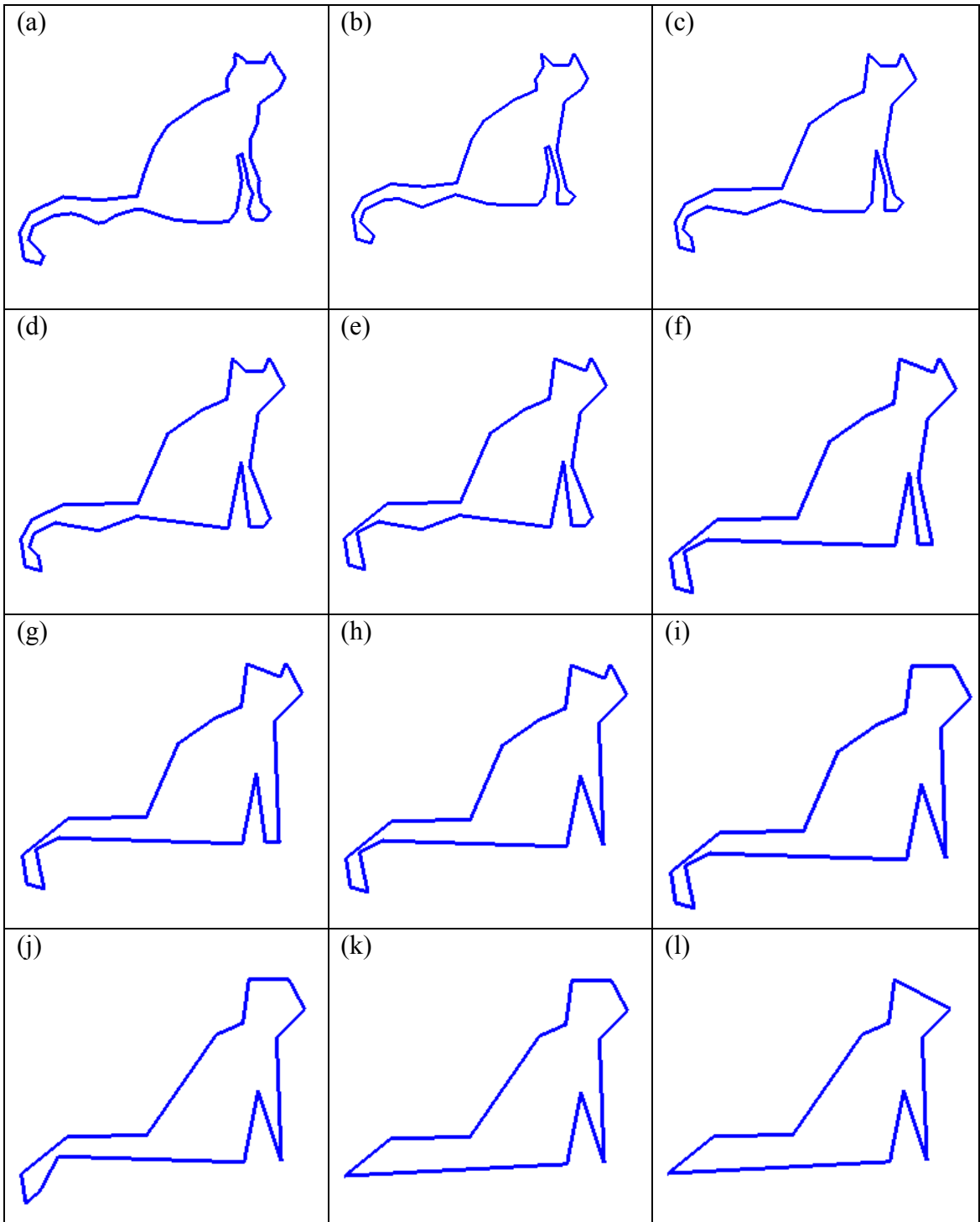


Fig. 5.16. Geometric LODs of a Cat Object

The cat object with LODs created by applying the Douglas-Peucker algorithm.

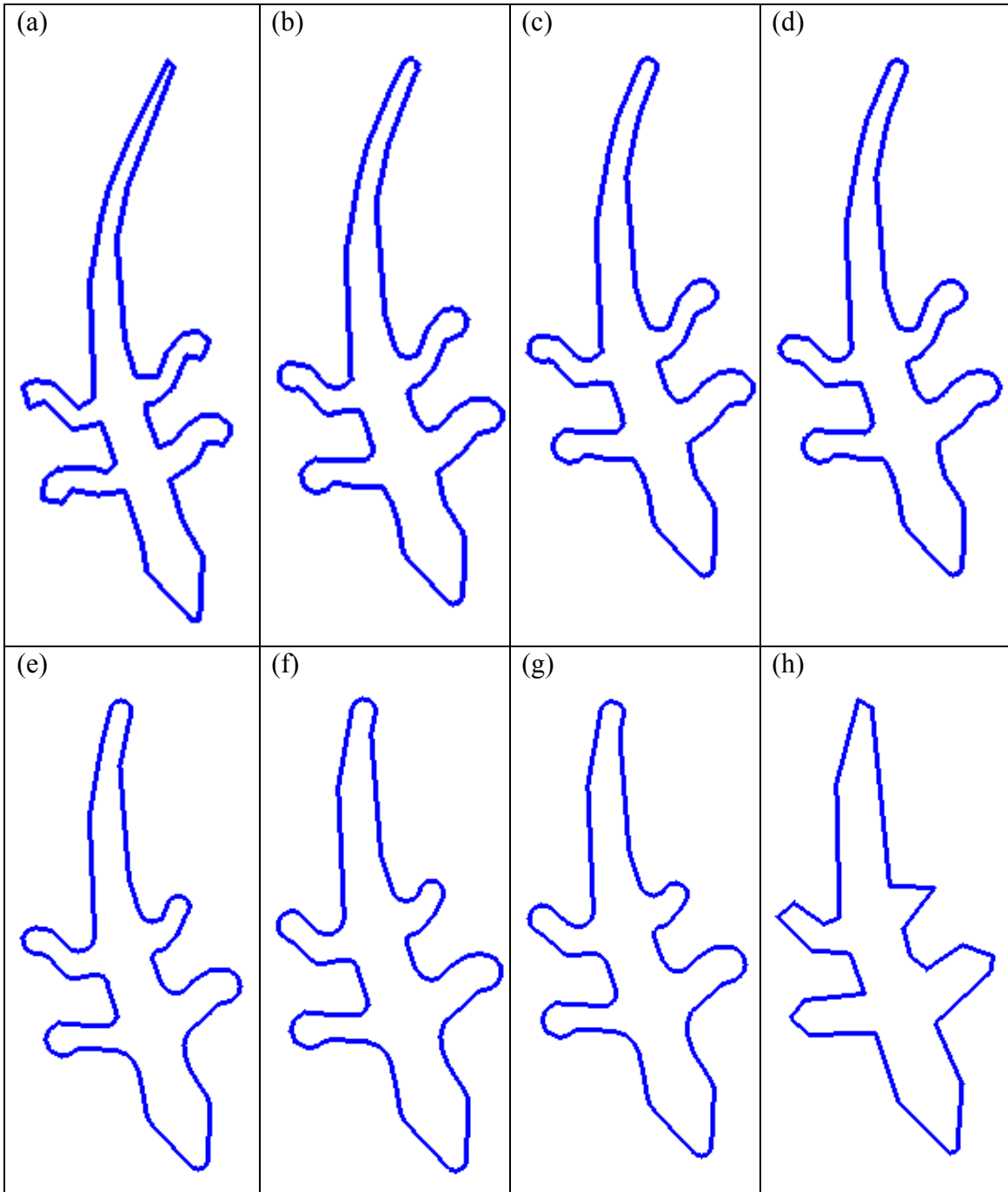


Fig. 5.17. Perceptual LODs of a Lizard Object.

(a) A lizard object. (b - g) Selected LODs created with the perceptual simplification method. (h) A final representation of the simplest LOD in (h) created by applying the Douglas-Peucker algorithm with a tolerance of 6% of the object's maximum dimension.

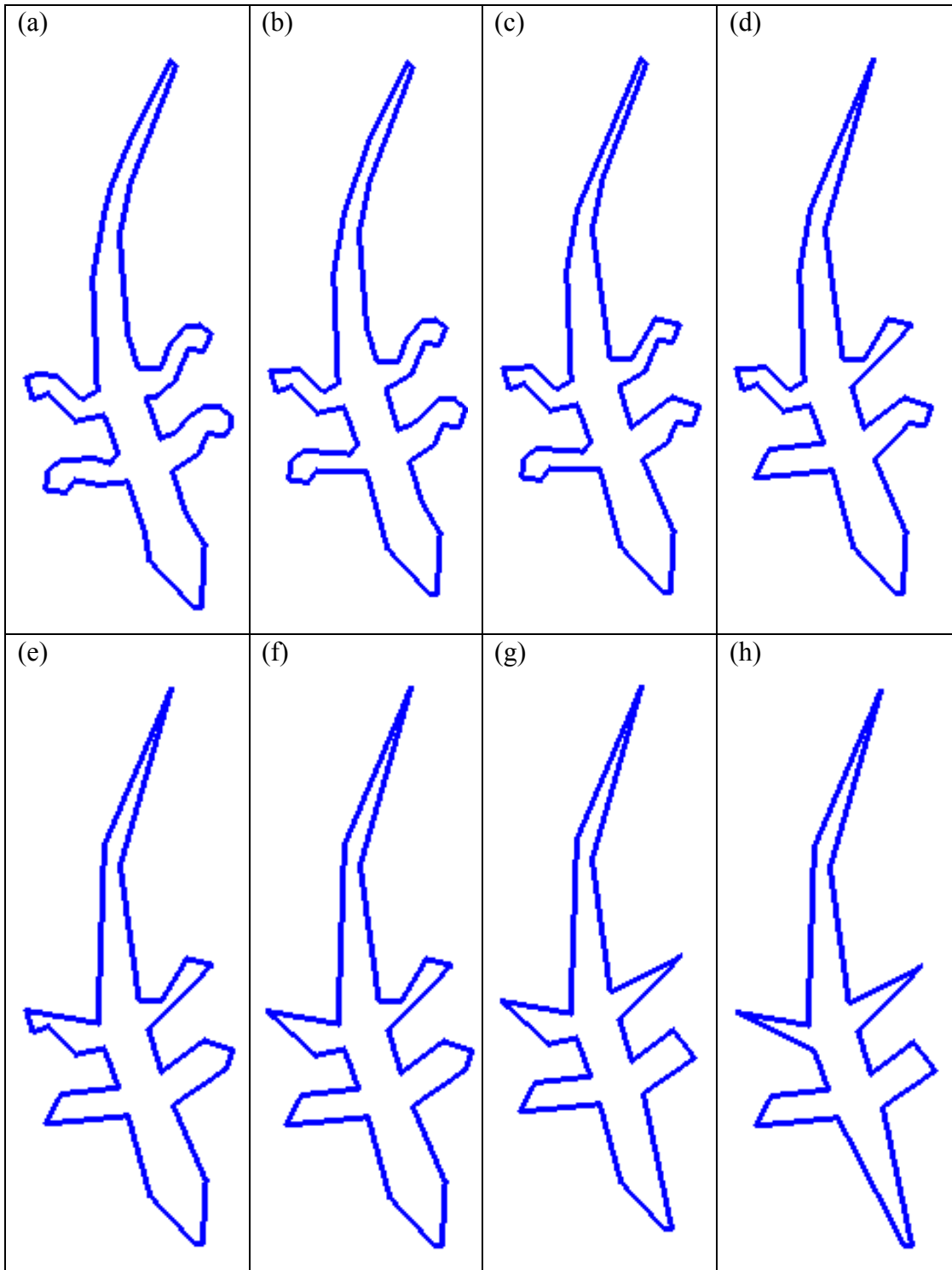


Fig. 5.18 Geometric LODs of a Lizard Object

The lizard object with LODs created by applying the Douglas-Peucker algorithm.

CHAPTER 6 CONCLUSIONS

6.1. Research Summary

The first section of this chapter reviews the research in this dissertation, and the next section discusses future work that could further expand and deepen this research. The last section completes this dissertation by exploring several new areas of research suggested by my experience with this work.

Chapter 3 presented the idea that objects can be represented as a collection of solid parts plus the information that is required to connect those parts. Applying this idea to the Blum Medial Axis Transform, a measure was created to grade the amount of substance information and the amount of connection information at every point on the medial axes. A new and significant idea is that this measure is allowed to be continuously variable, allowing points to have both substance and connection components rather than be forced into a binary classification of one or the other.

Chapter 3 also demonstrated how this measure can be used to decompose objects into perceptually defined parts. Because the measure is graded, the resulting parts classification is fuzzy as well. This corresponds well with our perception of objects, where there is often ambiguity in determining the parts of an object.

The idea of considering objects to be substance and connections is a new contribution to the field of medial form analysis. Typically, medial representations either maintain connecting information indirectly, requiring a great deal of work to extract it, or they do not maintain the information at all. This work demonstrates the value in creating distinct substance and connection components in an object representation.

Furthermore, the substance and connection scheme and the resulting fuzzy parts-classification address the longstanding instability problem with the Blum MAT. Using a substance-weighted Blum MAT, form calculations can remain stable with boundary perturbations that cause drastic changes to the underlying MAT. In addition to providing a solution to the Blum instability problem, this approach is intended for the general class of skeleton and medial form representations.

Chapter 4 describes a list of key components of preattentive form saliency. Based on the current body of perceptual theory and physical and psychophysical studies, aperture-based scale and relationship distance scale were discussed as fundamental components of saliency. Adding to these the importance of bilateral relationships produced the notion of figures as the basic building blocks of form.

Measures were then developed to quantify perceptual properties of figures. Visual conductance measures the likelihood that, in perceiving an object, we will visually connect two adjacent figures. Figure endness captures the special perceptual effects found at figural ends, and, while incomplete in its development, it measures at every point the degree to which we would perceive that point as the end of a figure. Visual mass reflects the overall size of an object by quantifying the retinal area that will be stimulated when viewing the object, and visual length is a measure of the elongation of a figure or of an entire object.

Based on these measures, form complexity and form saliency metrics were created. Visual length serves as a direct measure of form complexity. Visual potential reflects all of the figural properties and components of saliency discussed above; when it is integrated within a local aperture, it produces a local saliency measure, and when it is integrated over many contiguous medial points it produces a global saliency measure.

These complexity and saliency metrics have been grounded in figural models of perception. The measures produced sensible results on varied objects, and the results qualitatively matched those predicted by certain studies in the psychophysical literature. While there is a great deal of work that can be done to refine these measures, this research demonstrates the feasibility of generally applicable form metrics. The chapter also applied the

measures to object *parts* – protrusions and indentations. This, too, is new. By decomposing objects into their perceptual parts and providing a means to determine the saliency of each part as well as the object as a whole, this work provides a flexible approach for applying perceptual metrics to objects.

Chapter 5 presents a perception-based approach to object simplification. A perception-based measure was presented to schedule simplification operations; for each LOD, the medial point that carried the least visual saliency is selected at the region for simplification. This limits the simplifying operations to affecting the region of the object where modification is least noticeable.

A perceptual measure was presented to track the degree of simplification at each iteration. In this way, the method actually simplifies the underlying form of an object rather than its geometric representation. A perceptual measure was also used to compare the visual similarity of a potential LOD with the original object. This allows for some large geometric variations between LODs if their visual consequences are small.

The problem of object simplification was then reduced to a numerical optimization scheme for each LOD. Operating around the point with the lowest measure of saliency, the boundary in that local region was manipulated to produce an LOD with the least visual complexity that exhibited the global saliency of the original object.

Chapter 5 contributes a demonstration of qualitatively different simplification results from a perceptual approach as compared to a geometric-based method. In short, the perceptual method operates on the parts of an object rather than on geometrically local regions. Moreover, applying simplification operations to an object part can have global effects as the neighboring parts are adjusted to maintain saliency.

These results suggest that the traditional simplification algorithms may be able to retain greater visual fidelity in each LOD by accounting for some of the perceptual effects revealed by this work's perceptual approach. The contribution of this research includes not only the perceptual effects uncovered here that may be applied to simplification algorithms, but also the idea of viewing object simplification as a perceptual problem rather

than a geometric one. One product of that viewpoint is seen with the object simplification problem cast as a simple constrained optimization. This approach allows each component of the problem – scheduling, measuring visual complexity and visual similarity, and the simplification operations themselves – to be addressed independently within the well-understood framework of an optimization problem.

The results of Chapter 5 also highlight some limitations imposed by using line primitives for representing objects. The perceptual LODs required many *more* primitives than the original objects even though they have a much lower level of visual complexity. The rounding of boundaries caused by the perceptual simplification could be represented by a small number of curved primitives, but it requires a very large number of line primitives. This suggests new possibilities for matching rendering primitives with the needs of our visual system, a matter discussed in Section 6.3.

Tying all this work together brings us back to my original thesis statement:

Medial form representations can provide a means to quantify perceptual form cues with measures that qualitatively satisfy perceptual effects over a variety of objects.

Using such perceptual measures for interactive rendering reveals desirable properties and provides significantly different results than non-perceptually based approaches.

In this research, I have demonstrated how figural properties of form can be quantified with perceptual metrics and how using the Blum MAT as a medial representation allows these metrics to be computed. These figural properties embody important form cues that have been reported in the perception literature, and they include visual length, visual mass, visual conductance and figural endness. They are key components of form saliency metrics, and visual length is even used alone to measure elongation and visual complexity.

Even more significantly, I have demonstrated preattentive saliency measures that reflect perceptual phenomena. The measures use a parts-classification scheme that fits with our intuition about the parts of objects, and the method allows for the ambiguity of object parts that we encounter in our perception of the world. The approach even accounts for a

hierarchical parent/child influence among object parts. The results produced by the metrics qualitatively match results from psychophysical studies.

Applying these measures to the computer graphics problem of object simplification has produced a perceptually based simplification method. This approach produces qualitatively different results than a geometric-based method, and these results display important perceptual properties that are not present in geometric methods.

6.2. Future Work

There are several opportunities to improve upon the perceptual-based simplification method described in this work. The measure to detect ends of figures in Chapter 3 is ad-hoc and differs from the measure presented in Chapter 4 to measure the strength of endness response. Chapter 4's method, using an annulus to detect the amount of boundary within one medial radius, is useful for quantifying endness response for computing saliency. However, it does not adequately detect the endness condition in the midst of a complex branching of Blum MAT axes, a requirement for use in Chapter 3 for determining visual conductance. Because of this, conductance among axes is based on the end condition as computed in Chapter 3, but the endness component of saliency is computed using Chapter 4's calculation, and this produces unexpected results for certain configurations.

The solution to this problem is likely to be a different medial representation other than the Blum MAT. With a representation that does not see an explosion of axes for curved boundaries, the ad-hoc endness detection of Chapter 3 could be discarded and the annulus method could be refined. Possible candidates for such a representation include m-reps [Pizer99] or Hamilton-Jacobi Skeletons [Siddiqi02].

In my opinion, a medial representation other than the Blum MAT would be a valuable modification of this work. The MAT varies greatly with every boundary perturbation, and while the work of Chapter 3 isolates the calculations from this instability, the additional work required to maintain this insulation over the very large number of Blum MAT segments is considerable. Calculations such as the annular endness measure, discussed

above, are derived from attributes such as the tangent of the medial axis. However, because the axes vary wildly with boundary changes, attributes such as this must be averaged within some region. Performing such a calculation then requires all of the techniques used in Chapter 3, such as using visual conductance and shadowing. A representation that was less responsive to perceptually unimportant boundary details would aid in such calculations.

The problems with indentations that were discussed in Chapter 4 would also be addressed with a different medial representation. Recall that small indentations change the visual significance of a parent figure very little, while large indentations can carry a higher value of visual significance than the entire parent figure. A multi-scale medial representation would allow this to be addressed. A large scale representation of a parent figure would not be affected by a small scale indentation, while a larger scale indentation could easily be detected and accounted for.

Finally, the slow speed of the simplification method described leaves it unsuitable for practical use. For example, on a Sun Ultra 10 workstation, computing the metrics for the original cat object takes 22 seconds, while the original lizard takes 1 minute 46 seconds. As the simplification progresses and objects become more rounded, running time is even longer. Considering the fact that a practical algorithm would call for an automatic optimization scheme that would require dozens if not hundreds of trials for each LOD, the current implementation would be unusable for all but the most simple objects. While there are many optimizations that can be applied to this method, the one with the greatest impact would be to limit the perceptual computations during the optimization loop. When modifying an object's boundary, only the medial axes in that local area are affected, so only those need to be recomputed. For example, simplifying around the left ear of the cat object affects, at most, 9 of the 93 medial axes in the object. This subset contains 78 units of arclength along the axes out of a total of 1560 over the entire object. In this typical case, recomputing perceptual metrics on only the medial axes that change would result in a 20 times increase in speed.

6.3. New Research Areas to Investigate

In computer graphics, the overwhelming need for simplification comes from 3D rendering, so an obvious and important research question to ask on the heels of this work is how to apply such perceptual simplification to 3D objects. The straightforward approach of extending these methods to 3D is, in my opinion, a lost cause. While the state-of-the-art in computing 3D medial surfaces is improving, it is a difficult problem. Furthermore, even with a perfect medial representation, computing the perceptual metrics on medial sheets instead of medial curves adds a vexing level of complexity. Medial branch points in 2D become branch curves in 3D, and while there are two choices of direction along a 2D axis, there are an infinite number of directions on a 3D surface. Given this, there is no direct extension of these metrics from 2D to 3D.

An area of research that is perhaps more promising than refining the actual simplification method presented here is to investigate ways to apply the lessons of this dissertation to more traditional simplification methods. Current methods could be augmented to detect figure-like parts on objects, such as protruding regions and indentations. Then, allowing interactions among neighboring parts during simplification, such as melting a protrusion into a larger scale parent part, would mirror some of the fundamental results of this work. I believe that these effects could be implemented using geometric-based measures. Also, this work points to the need to investigate alternatives to line and polygon primitives. Space-filling primitives like circles and “splats” show promising correspondence to the perceptual effects shown here.

In general, I believe that for adding perceptual effects to object simplification, the most accessible step is to develop a perception based similarity metric or one that incorporates perceptual principles using geometric measures. As shown in [Lindstrom01], such a metric could also be harnessed to determine scheduling of simplifications. This leaves the task of incorporating perceptual principles into the simplification operations themselves. I recommend an approach that considers operations affecting a more global region than the usual vertex-based operations. This would open the possibility of implementing effects during simplification operations such as modifying the size of surrounding regions

or changing the topology of objects based on perceptual triggers rather than geometric ones.

Another fruitful approach would be to use perceptual measures in a view-dependent simplification scheme. Our perception is based on projections of the 3D world into 2D retinal images. A view-dependent simplification algorithm would take 3D objects or an entire 3D scene and create a 2D projection based on the desired viewpoint. [Lindstrom01] presents a framework for this, creating a number of projections from viewpoints spaced on a spherical shell around an object. Using 2D image-based metrics, they create LODs that show maximum similarity across all viewpoints. A 2D perception-based metric could easily be included within their framework.

[Cohen98], [Lindstrom00] and others demonstrate the importance of including shading attributes such as color and texture in any similarity measure for simplification. These methods measure how much difference there is in a rendering of an LOD compared to a rendering of the original object. This can be viewed, in a sense, as a “distance” metric in image space. [Rushmeier95] and [Watson01] discuss perceptually based image metrics that can be used to measure visual similarity between LODs. Any of these image-based metrics can be combined with form-based perceptual measures. 2D objects can be segmented into parts using image-based analysis, and then form measures similar to the ones shown in this work could be applied to determine which parts have a low visual significance. Then, in these image regions, a higher image-based deviation could be tolerated than is indicated by the image-based metric alone without losing visual fidelity.

Finally, this work suggests a number of psychophysical experiments. To start, user studies could test the effectiveness of simplifications derived from the perceptual approach described here. A useful study would compare LODs created with this approach to LODs created with a geometric approach. Experiments are also warranted to study the perceptual measures of Chapter 4. To what degree do the local image saliency and overall object saliency measures match our actual perception? Is there a difference in our sensitivity to internal and external figures? To what degree should the saliency at ends of figures be emphasized? How large an aperture should be used for calculating the meas-

ures? What are the perceptual triggers for splitting an object into two or more objects, thereby changing topology? Studies into these questions would provide important direction for the further development of perceptual shape metrics for computer graphics.

BIBLIOGRAPHY

- [August99] August, J., Siddiqi, K and Zucker, S. Ligature Instabilities in the Perceptual Organization of Shape. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, pp. 42 – 48. (1999)
- [Biederman87] Biederman, Irving, Recognition-by-Components: A Theory of Human Image Understanding, *Psychological Review*, 94(2):115-147 (1987)
- [Blum67] Blum, Harry, A Transformation for Extracting New Descriptors of Shape, In W. Whalen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pp. 362-380. MIT Press, Cambridge, MA, 1967.
- [Blum78] Blum, Harry, and Roger N. Nagel, Shape Description Using Weighted Symmetric Axis Features, *Pattern Recognition*, 10:167-180 (1978)
- [Bolin98] Bolin, M. and G. Meyer. A Perceptually Based Adaptive Sampling Algorithm. *Proceedings SIGGRAPH 98*, pp. 299-309.
- [Brady84] Brady, Michael and Haruo Asada, Smoothed Local Symmetries and Their Implementation, *International Journal of Robotics Research*, 3(3):36-61 (1984).
- [Braunstein89] Braunstein, M. L., D. D. Hoffman, and A. Saidpour, Parts of Visual Objects: An Experimental Test of the Minima Rule, *Perception*, 18:817-826. (1989)
- [Bruce85] Bruce, J. W., P. J. Gibling and C. G. Gibson, Symmetry Sets, *Proc. Royal Society Edinburgh*, 101A:163-186 (1985)

- [Bruce96] Bruce, Vicki, Green, Patrick R., and Georgeson, Mark A., *Visual Perception: Physiology, Psychology, and Ecology*, Psychology Press, East Sussex, UK, 1996.
- [Burbeck95] Burbeck, CA, SM Pizer, Object representation by cores - identifying and representing primitive spatial regions, *Vision Research*, 35 (13):1917-1930. (July 1995)
- [Burbeck96] Burbeck, CA, SM Pizer, BS Morse, D Ariely, GS Zauberman, JP Rolland, Linking object boundaries at scale - a common mechanism for size and shape judgments, *Vision Research*, 36 (3):361-372. (February 1996)
- [Certain96] Certain, Andrew, Jovan Popovic, Tony DeRose, Tom Duchamp, David Salesin, and Werner Stuetzle, Interactive Multiresolution Surface Viewing, *Proceedings of SIGGRAPH 96* (New Orleans, LA, August 4-9, 1996). In *Computer Graphics Proceedings*, Annual Conference Series, 1996. ACM SIGGRAPH, pp. 91-98.
- [Clark76] Clark, James H., Hierarchical Geometric Models for Visible Surface Algorithms, *Communications of the ACM*, 19(10):547-554.
- [Cohen96] Cohen, Jonathan, Amitabh Varshney, Dinesh Manocha, Greg Turk, Hans Weber, Pankaj Agarwal, Frederick Brooks, William Wright, Simplification Envelopes, *Proceedings SIGGRAPH 96*,
- [Cohen98] Cohen, Jonathan, Marc Olano and Dinesh Manocha, Appearance-Preserving Simplification, *Proceedings of SIGGRAPH 97*
- [DeHaemer91] DeHaemer, Michael J. Jr. and Michael J. Zyda, Simplification of Objects Rendered by Polygonal Approximations, *Computers and Graphics*, 15(2):175-184.

- [DeRose94] DeRose, Tony, Michael Lounsbery, and Joe Warren. Multiresolution Analysis for Surfaces of Arbitrary Topological Type. Technical Report 93-10-05, Department of Computer Science and Engineering, University of Washington, January 1994.
- [Douglas73] Douglas, D. H. and T. K. Peucker, Algorithms for the reduction of the number of points required to represent a line or its caricature, *The Canadian Cartographer*, 10(2):112 – 122, 1973.
- [Eberly93] Eberly, D., R. Gardner, B. Morse, S. Pizer, C. Scharlach, Ridges for Image Analysis, Ridges for image analysis. University of North Carolina Computer Science Department technical report TR93-055, *Journal of Mathematical Imaging and Vision*, 4: 351-371 (1994).
- [Eberly94] Eberly, David, A Differential Geometric Approach to Anisotropic Diffusion, *Geometry-Driven Diffusion in Computer Vision*, ter Haar Romeny, Bart M. (Ed.), Kluwer Academic Publisher, Dordrecht, The Netherlands, pp. 380-382, 1994.
- [Erikson00] Erikson, Carl, Dinesh Manocha, and William Baxter, HLODs for Fast Display of Large Static and Dynamic Environments, *Proceedings ACM Symposium on Interactive 3D Graphics, 2001*.
- [Ferwerda96] Ferwerda, J.A., S. N. Pattanaik, P. Shirley, and D. Greenberg. A Model of Visual Adaptation for Realistic Image Synthesis. *Proceedings SIGGRAPH 96*, pp. 249-258.
- [Frome72] Frome, Francine S. (1972). A Psychophysical Study of Shape Alignment. Technical Report TR-198, University of Maryland Computer Science Center, College Park, Maryland.

- [Funkhouser92] Funkhouser, Thomas, Carlo H. Sequin, Seth J. Teller, Management of Large Amounts of Data in Interactive Building Walkthroughs, *Proceedings of SIGGRAPH 97*.
- [Garland97] Garland, Michael, Paul S. Heckbert, Surface Simplification Using Quadric Error Metrics. *Proceedings SIGGRAPH 97*, pp. 209-216.
- [Grossberg94] Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55:48-121.
- [He95] He, T., L. Hong, A. Kaufman, A. Varshney, and S. Wang. Voxel-based Object Simplification. *IEEE Visualization '95 Proceedings*, pp. 296-303 (1995).
- [Hershberger92] Hershberger, J. and J. Snoeyink, Speeding up the Douglas-Peucker line simplification algorithm, *Proc. 5th Intl. Symp. Spatial Data Handling*, IGU Commission on GIS, pp. 134 – 143, 1992.
- [Hoffman84] Hoffman, D. D., and W. A. Richards, Parts of recognition, *Cognition*, 18:65-96. (1984)
- [Hoppe93] Hoppe, Hugues, Tony DeRose, Tom Duchamp, John MmcDonald, and Werner Stuetzle, Mesh Optimization, *Computer Graphics (SIGGRAPH 1993Proceedings)*, 27:19-26.
- [Hoppe96] Hoppe, Hugues, Progressive Meshes, *Proceedings of SIGGRAPH 96* (New Orleans, LA, August 4-9, 1996). In *Computer Graphics Proceedings, Annual Conference Series*, 1996. ACM SIGGRAPH, pp. 99-108.
- [Hoppe97] Hoppe, Hugues, View-Dependent Refinement of Progressive Meshes, *Proceedings of SIGGRAPH 97*

- [Hubel77] Hubel, D. H., T. N. Wiesel, Functional architecture of macaque monkey visual cortex *Proceedings of the Royal Society of London, Ser. B.*, 198, pp. 1-59.
- [Kimia96] Kimia, BB, Kaleem Siddiqi, Geometric Heat Equation and Non-Linear Diffusion of Shapes and Images, *CVIU*, 64(3):305-322. (November 1996)
- [Koenderink84] Koenderink, J. J., The Structure of Images, *Biological Cybernetics* 50:363-370 (1984)
- [Kovacs93] Kovacs, Ilona and Bela Julesz, A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation, *Proceedings of the National Academy of Sciences, USA*, 90:7495-7479. (1993)
- [Kovacs94] Kovacs, Ilona and Bela Julesz, Perceptual Sensitivity Maps within Globally Defined Visual Shapes, *Nature*, 370:644-646. (1994)
- [Kovacs98] Kovacs, Ilona and Bela Julesz, Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, 38:2323-2333 (1998).
- [Koenderink90] Koenderink, J. J., Solid Shape , The MIT Press, Cambridge, MA, 1990.
- [Lee82] Lee, D. T., Medial Axis Transformation of a Planar Shape, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4:363-369, 1982.
- [Lee95] Lee, T.S., D. Mumford, and P.H. Schiller, Neural Correlates of Boundary and Medial Axis Representation in Primate Striate Cortex, *Investigative Ophthalmology & Visual Science*, 36(4):S477. (1995)

- [Leyton87] Leyton, M, Symmetry-Curvature Duality, *Computer Vision, Graphics and Image Processing*, 38:327-341. (1987)
- [Leyton88] Leyton, M, A Process Grammar for Shape, *Artificial Intelligence*, 34:213-247. (1988)
- [Leyton89] Leyton, M, Inferring Causal History from Shape, *Cognitive Science*, 13:357-387. (1989)
- [Leyton92] Leyton, M, *Symmetry, Causality, Mind*, MIT Press, April 1992.
- [Lindstrom00] Lindstrom, Peter, Greg Turk, Image-Driven Simplification, *ACM Transactions on Graphics*, 19(3):204-241. (2002)
- [Luebke97] Luebke, David and Carl Erikson, View-Dependent Simplification of Arbitrary Polygonal Environments, *Proceedings of SIGGRAPH 97*
- [Marr78] Marr, D., and K. H. Nishihara, Representation and Recognition of the Spatial Organization of Three Dimensional Structure, *Proceedings of the Royal Society of London*, 200:269-294. (1978).
- [McGreevy91] McGreevy, MW, Virtual Reality and Planetary Exploration, *Online*, 13(8): 3-8, Computer Systems and Research Division, NASA Ames Research Center, Moffett Field, CA. (August 1991)
- [Michels65] Michels, Kenneth, Leonard Zusne, Metrics of Visual Form, *Psychological Bulletin*, 65(2):74-86 (1965).

- [Nelson85] Nelson, J.I., B. J. Frost, Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex, *Experimental Brain Research*, 61:54-61 (1985).
- [Ogniewicz92] Ogniewicz, R. L. and M. Ilg, Voronoi Skeletons: Theory and Applications, *Proc. CVPR '92* pp. 63-69, 1992
- [Orban79] Orban, G.A., H. Kato, P.O. Bishop, End-zone region in receptive fields of hypercomplex and other striate neurons in the cat. *Journal of Neurophysiology*, 42:818-832.
- [Orban79b] Orban, G.A., H. Kato, P.O. Bishop, Dimensions and properties of end-zone inhibitory areas in receptive fields of hypercomplex cells in cat striate cortex. *Journal of Neurophysiology*, 42:833-849.
- [Pattanaik98] Pattanaik, S. N. , J. A. Ferwerda, D. A. Greenberg, and M. D. Fairchild. A Multiscale Model of Adaptation and Spatial Vision for Realistic Imaging. *Proceedings SIGGRAPH 98*, pp. 287-298.
- [Pettet99] Pettet, Mark W., Shape and contour detection, *Vision Research*, 39:551-557 (1999).
- [Pizer87] Pizer, Stephen M., William B. Oliver, and Sandra H. Bloomberg, Hierarchical Shape Description via the Multiresolution Symmetric Axis Transform, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):505-511, July 1987.
- [Pizer98] Pizer, S M., D Eberly, Daniel S. Fritsch, and Bryan S. Morse, Zoom-Invariant Vision of Figural Shape: The Mathematics of Cores, *Computer Vision and Image Understanding*, 69(1):53-71, January 1998.

- [Pizer99] Pizer, S.M., D. Fritsch, V. Johnson, E Chaney, Segmentation, Registration, and Measurement of Shape Variation via Image Object Shape. *IEEE Transactions on Medical Imaging*, October 1999, 18(10): 851-865.
- [Pizer02] Pizer, Stephen M., Kaleem Siddiqi, Gabor Székely, James N. Damon, Steven W. Zucker, Multiscale Medial Loci and Their Properties, *To appear in IJCV*, 2002.
- [Polat93] Polat, Uri, Dov Sagi, Lateral Interactions Between Spatial Channels: Suppression and Facilitation Revealed by Lateral Masking Experiments, *Spatial Vision*, 33(7):993-999 (1993).
- [Polat94] Polat, Uri, Dov Sagi, The Architecture of Perceptual Spatial Interactions, *Vision Research*, 34(1):73-78 (1994).
- [Popovic97] Popovic, Jovan and Hugues Hoppe, Progressive Simplicial Complexes, *Proceedings of SIGGRAPH 97*.
- [Pspotka78] Pspotka, J, (1978). Perceptual processes that may create stick figures and balance. *Journal of Psychology: Human Perception and Performance*, 4:101-111.
- [Reddy96a] Reddy, Martin, A Measure for Perceived Detail in Computer Generated Images, *Technical Report ECS-CSG-19-96*, Department of Computer Science, University of Edinburgh, Edinburgh (1996)
- [Ramasubramanian99] Ramasubramanian, M., S. Pattanaik, and D. Greenberg. A Perceptually Based Physical Error Metric for Realistic Image Synthesis. *Proceedings of SIGGRAPH 99*, pp. 73-82 (1999).
- [Reddy96b] Reddy, Martin, SCROOGE: Perceptually-Driven Polygon Reduction, *Computer Graphics forum*, 15(4):191-203 (1996)

- [Rock93] Rock, I., CM Linnett, Is a perceived shape based on its retinal image?, *Perception*, 22 (1):61-76. (1993)
- [Rossignac93] Rossignac, Jarek. R. and Paul Borrel, Multi-Resolution 3D Approximations for Rendering Complex Scenes, *Geometric Modeling in Computer Graphics*, Falcidieno, B. and T. L. Kunii, editors, Springer-Verlag, Genova, Italy, pp. 455-465. Also published as Research Report RC 17697 (#77951), IBM Research Division, T. J. Watson Research Center, 1992.
- [Rushmeier95] Rushmeier, H., G. Ward, C. Piatko, P. Sanders, and B. Rust, Comparing Real and Synthetic Images: Some Ideas About Metrics. *Proceedings of the 6th Eurographics Workshop on Rendering*, pp. 82-91 (1995).
- [Schachter81] Schachter, Bruce, Computer Image Generation for Flight Simulation. *IEEE Computer Graphics and Applications*, 1(4):29-68 (1981).
- [Schroeder92] Schroeder, William J., Jonathan A. Zarge, and William E. Lorensen, Decimation of Triangle Meshes, *Computer Graphics (SIGGRAPH 1992 Proceedings)*, 26(2):65-70.
- [Seel94] Seel, M., Eine Implementierung Abstrakter Voronoidiagramme. Diplomarbeit, Universität des Saarlandes (1994).
- [Shaked98] Shaked, Doron and Alfred M. Bruckstein, Pruning Medial Axes, *Computer Vision and Image Understanding*, 69(2):156-169, 1998.
- [Shipley92] Shipley, T.F. and Kellman, P.J. (1992). Strength of visual interpolation depends on the ratio of physically-specified to total edge length. *Perception and Psychophysics*, 52:97-106.

- [Siddiqi96] Siddiqi, K., KJ Tresness, BB Kimia, Parts of visual form: psychophysical aspects, *Perception*, 25(4):399-424. (1996)
- [Siddiqi99] Siddiqi, K., A. Tannenbaum, S. Zucker, The Hamilton-Jacobi Skeleton, *ICCV 1999*, pp. 828-834.
- [Siddiqi02] Siddiqi, K. S. Bouix, A. R. Tannenbaum, S. W. Zucker. Hamilton-Jacobi Skeletons. *International Journal of Computer Vision*, to appear 2002.
- [Stanney02] Stanney, Kay M., Ed. *Handbook of Virtual Environments: Design, Implementation, and Applications*, Lawrence Erlbaum Associates, Mahway, New Jersey, 2002.
- [Székely92] Székely, G., Ch. Brechbuhler, O. Kubler, R. Ogniewicz, and T. Budinger, Mapping the human cerebral cortex using 3D medial manifolds, *Proceedings of Visualization in Biomedical Computing*, SPIE, 1808:124-132 (1992)
- [Székely96] Székely, G. *Shape Characterization by Local Symmetries*. Habilitationsschrift: Institut für Kommunikationstechnik, Fachgruppe Bildwissenschaft, ETH, Zurich. (1996)
- [Taylor93] Taylor, Russell M., Warren Robinett, Vernon L. Chi, Frederick P. Brooks, Jr., William V. Wright, R. Stanley Williams, and Erik J. Snyder, The Nanomanipulator: A Virtual-Reality Interface for a Scanning Tunneling Microscope, *Computer Graphics: Proceedings of SIGGRAPH '93*, August 1993.
- [Treisman80] Treisman, Anne M., Garry Gelade, A Feature-Integration Theory of Attention, *Cognitive Psychology*, 12(1):97-136. (1980)

- [Turk92] Turk, Greg, Re-tiling Polygonal Surfaces, *Computer Graphics (SIGGRAPH '92 Proceedings)*, 26:55-64.
- [Varshney94] Varshney, Amitabh, *Hierarchical Geometric Approximations*, Dissertation, UNC Chapel Hill, Chapel Hill, NC. 1994
- [Vos93] Vos, P.G., N. Bocheva, N. Yakinoff, and E. Heloper, Perceived location of two-dimensional patterns, *Vision Research*, 33(15):2157-69.
- [Walk92] Walkthrough Project: Final Technical Report to National Science Foundation Computer and Information Science and Engineering, UNC-CH Computer Science Technical Report TR92-026, Chapel Hill, NC. (1992)
- [Watson01] Watson, Benjamin, Alinda Friedman, and Aaron McGaffey. Measuring and Predicting Visual Fidelity. *Proceedings of SIGGRAPH 01*, pp. 213-220 (2001).
- [Westover90] Westover, Lee, Footprint Evaluation for Volume Rendering. *Proceedings of SIGGRAPH 90*, pp. 367-376.
- [Wolfe94] Wolfe, Jeremy M., Stacia R. Friedman-Hill, Alexander B. Bilsky, Parallel processing of part/whole information in visual search tasks, *Perception and Psychophysics*, 55:(5):537-550. (1994)
- [Wolfe97] Wolfe, J.M., S.C. Bennett, Preattentive object files: shapeless bundles of basic features, *Vision Research*, 37(1):25-43. (Jan 1997)
- [Wright93] Wright, RD, AN Katz, EA Hughes, Inattention and the perception of visual features, *Acta Psychol.*, 83(3):225-35, (August 1993)

[Young86] Young, Richard A., Simulation of Human Retinal Function with the Gaussian Derivative Model, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[Young87] Young, Richard A., The Gaussian Derivative Model for Spatial Vision: I. Retinal Mechanisms, *Spatial Vision*, 2(4):273-293, 1987.

[Zyda88] Zyda, Michael, Robert McGhee, Ron Ross, Douglas Smith and Dale Stregle, Flight Simulators for under \$1000, *IEEE CG&A*, 8(1):19-27. (Jan 1988)