Abstract

The Perceptual Organization of Visual Flows

Ohad Ben-Shahar

2003

Locally parallel dense patterns - visual flows - define a perceptually coherent structure of particular significance to perceptual organization. Geometrically, it is argued that a proper way to investigate these structures requires the frame field approach from differential geometry, a study that leads to the notion of visual flow curvatures and to constraints on their mutual behavior. These curvatures are then used to develop a theory, and a rigorous model, of visual flow "good continuation" that extends common terminology from Gestalt psychology and from computational studies of curves.

The geometrical theory is then applied in three ways. Firstly, psychophysical exploration of the role of visual flow curvatures in human perception shows that sensitivity to these curvatures greatly affects orientation-based texture segmentation. Secondly, a contextual framework for the computation of coherent visual flows from images is developed and applied to texture, shading, and color analysis. Unlike existing approaches, the proposed framework is able to handle both sparse, dense, and multivalued data sets, while preserving line and point singularities and rejecting large scale non-flow structures. Lastly, the geometrical theory is linked to the functional organization of primary visual cortex to accurately predict the distribution of long range horizontal connections and to support their identification with those obtained mathematically.

The Perceptual Organization of Visual Flows

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the degree of
Doctor of Philosophy

By

Ohad Ben-Shahar

Dissertation Director: Steven W. Zucker

December 2003

To my wife, Galit,
my daughter, Shir,
and my mother, Carmela.

Without you, I would have never set on this journey and even if I had done so, would I have had the strength to carry it all the way through?

Acknowledgements

This has been a long journey and I am indebted to many people for helping me on my way. First and foremost, I would like to thank Steven Zucker. I have been extremely fortunate to have Steve as my advisor for the past six years, to benefit from his vast knowledge and experience, and to enjoy his unique views and ideas while having the freedom to develop my own, even when they sometimes diverged from his. More importantly, I thank him for showing me the path to true scholarship, leading me from asking just “How” to pondering first over the “What” and the “Why”, which too often and too lightly are ignored in our community. Steve’s singular commitment to basic questions, multidisciplinary research, and the “big picture”, has greatly inspired my thinking and undoubtedly will shape my future as a scholar.

I also wish to thank the other members of my committee: Drew McDermott, for his keen ability to penetrate directly to the core of ideas, James Duncan, for the special effort he has made to make himself available to me, and Brian Scholl, for the years of inspiring collaboration, numerous chats, lab meetings, and joint projects.

Many visitors and neighbors provided invaluable insight in their area of expertise. I am particularly thankful to Allen Tannenbaum for a close examination of the mathematical theory, Michael Stryker for the discussions we had held about the physiological implications, and Brian Scholl for expanding my phsychophysical horizons with object-based attention.

Without doubt, my years at Yale were made particularly enjoyable thanks to the constant interaction with my fellow students at the lab. I am particularly grateful to Jonas August and Patrick Huggins for uncountable hours of debates and blackboard discussions, and for reading, commenting, and correcting the English of almost everything I have written all these years. Many thanks also to Gang Li, Athos Georghiades, Todd Zickler, Melissa Koudelka, and all those who joined us for discussions and student seminars at the CVC Round Table. In addition, special thanks to Mauro Maggioni for many inspiring discussions, and to Steve Haker who patiently examined and helped with the math early on.

If an army travels on its stomach, computer researchers runs on technical support. I am grateful to Claudia Villano, and particularly to Mark Wogahn, for their indispensable service in keeping our systems up and running. I cannot recall a single time when Mark did not respond immediately to fix my problems, no matter how small or big, and giving me the feeling that I was virtually the only (or at least, the most important) user in the department. A similar level of appreciation is also due to the administrative staff of the Computer Science department, and especially to Judy Smith, Linda Dobb, and Amy Toensmeier, all of whom made consistent efforts to promptly address all my needs.

Certain images used in this dissertation were contributed by various people. For their contribution I wish to thank Pierre Kornbrobst (Odyssee lab, INRIA, France), Yaki Levi (Inspectech, Israel), David Fitzpatrick (Duke University), Michael Stryker (UCSF), Patrick Huggins (Yale), Gang Li (Yale), Pamela Davis (UCLA), and Paul Bourke (Swinburne University of Technology, Australia).

Finally, I owe an inexpressible debt of gratitude to my family members who made all this possible. Eden and Or, our two beautiful children, who were born in the U.S.A. during our years at Yale, reminded me what is truly important in life and helped me preserve my sanity (with the occasional opposite effect as well). My brothers in Israel, Eran and Harel, provided me with constant inspiration, encouragement, and a source of great pride. But most importantly, I am forever thankful to three women who jointly made a heavy sacrifice and carried on their shoulders the true burden of my higher education: to my first daughter Shir, who was brought to a foreign country and culture when she was just a baby, to my mother Carmela (Caroline), who painfully let her son go across the ocean to be seen only years apart, and to my wife Galit, who willingly left behind her own family and friends, and who put her own career on hold for the sake of her husband's. This dissertation is dedicated to the three of them, without whom none of it would have happened.

Contents

List of Figures

Preface

This dissertation is not a typical computer vision dissertation, neither in content nor in structure. While a typical thesis explores a particular vision related issue from a mathematical/computational point of view only, the one that follows does so from a multitude of perspectives, including psychophysical and physiological ones. This multidisciplinary approach is directly reflected in the arrangement of the chapters, each of which is devoted to one of the disciplines under consideration. Most chapters are self contained (so for example, the computational experimental results are part of the computational chapter, instead of a chapter of their own) and their order follows naturally from the way a contribution in one discipline stimulates investigation in another. This organization should help different readers focus more easily on their area of interest. In particular, readers interested only in computational aspects can skip Chapter 3 (psychophysics) and Chapter 5 (physiology). Readers interested solely in human perception and psychophysics can read through Chapter 3 only while readers interested solely in computational neuroscience should read through Chapter 5 excluding Chapter 3. The different chapters however, combine together to a certain Gestalt that, as this author believes, is greater than its parts. In the the spirit of the Gestalt psychology that greatly inspired this research, reading all chapters in their designated order is thus strongly recommended.

Chapter 1

Introduction

Computer vision is the science (and art) of inferring information about the world from its images (i.e., its 2D visual projections). Depending on the application, the nature of the desired information may vary significantly, and yet, a good understanding of the image features[1] involved in the inference process is paramount for the successful completion of the task.

Oftentimes the measurement of a particular image feature is already an inference process that acts upon fundamental and directly measurable units. A typical example is that of image curves, or contours, which earn their global identity due to a grouping process of individual edge measurements (edgels). But easier said than done: local edge measurements, both in biological and computer systems, are often noisy; even their existence or absence typically loom in uncertainty. Thus, which edgel goes with which? Which are spurious? And which should be filled in? The human visual system appears indifferent to these kind of difficulties as somehow we manage to effortlessly "see" the "right" contours almost always. Given the complexity of the problem, and the lack of constraints on which contours may be found in images, this flawless performance is rather puzzling.

The divide between locally measurable units (intensity points, edgels, motions, etc...) and global perceptual entities (curves, textures, surfaces, volumes, objects, scenes) was addressed by the Gestalt psychologists [Wertheimer, 1955a; Koffka, 1935; Helson, 1933; Ellis, 1955; Gordon,

---

[1] Here we use the notion of image feature in a broad sense to refer to any property of the two dimensional visual signal that can be measured and used in the inference process.

1989; Palmer, 1999a] almost a century ago. To Wertheimer, Koffka, Köhler, and their contemporaries the missing link was between the primitive representation of the signal based on its "parts" and the concrete and definite sensation of "wholes" which is very different than their sum. In Max Wertheimer's words [Wertheimer, 1955a]:

> *I stand at the window and see a house, trees, sky.*
>
> *Theoretically I might say there were 327 brightnesses and nuances of colour. Do I have ”327”? No. I have sky, house, and trees. It is impossible to achieve ”327” as such.*

The (partial) solution the Gestalt psychologists provided to the problem took the form of a set of rules by which sensory data are *organized* into coherent and stable perceptual "wholes" based on principles such as *proximity*, *similarity*, *good continuation*, *closure*, and *past experience*. These original principles are now known as Wertheimer's laws of *perceptual organization* [Wertheimer, 1955b] and research into possible new ones has never ceased (see [Pinna et al., 2003] for a recent example).

It is probably more accurate to characterize Wertheimer's laws of perceptual organization as phenomenological observations rather than "laws", and indeed the Gestalt movement suffered significantly because neither their laws, nor the interactions between them, were ever made sufficiently quantitative and concrete as to offer much predictive power. That being said, however, the importance of their work cannot be overstated. For computational vision, the revolutionary idea in perceptual organization lies in the implication that general purpose processes, i.e., processes that employ only minimal *a priori* knowledge about the world, can (1) impose structure on unfamiliar scenes, (2) recover non-accidental causal relationships, (3) organize sensory data into perceptual wholes that are likely to belong together, (4) overwhelmingly reduce the complexity of representation, and (5) facilitate other visual processes that are likely to fail without such an organization. These points were brought to the fore of computational vision research almost three decades ago [Zucker et al., 1975; Lowe, 1985], perhaps most forcefully in Witkin and Tenenbaum's monograph on the role of structure in vision [Witkin and Tenenbaum, 1983]:

> *People's ability to perceive structure in images exists apart from both the perception of tri-dimensionality and from the recognition of familiar objects. That is, we organize*

> *the data even when we have no idea what it is we are organizing. What is remarkable is the degree to which such naively perceived structure survives more or less intact once a semantic context is established ... It is almost as if the visual system has some basis for guessing what is important without knowing why.*
>
> *We argue that the aim of perceptual organization is the discovery and description of spatio-temporal coherence and regularity... A description that decomposes the image into constituents that capture regularity or coherence therefore provides descriptive chunks that act as "semantic precursors", in the sense that they deserve or demand explanation. ... By anticipating semantically important relationships, it provides a basis for interpretation in terms of stored models, and for learning new model.*

Based on such ideas, perceptual organization of sensory data and the recovery of coherent structure have progressively penetrated computer vision research. These efforts, however, focused almost exclusively on the study of curve-like structures (e.g., [Ullman, 1976; Fischler and Bolles, 1986; Kass et al., 1988; Sha'ashua and Ullman, 1988; Parent and Zucker, 1989; Donald and Riseman, 1992; Fischler, 1994; Alter and Basri, 1996; Williams and Jacobs, 1997; Saund, 1998; Jonas et al., 1999; Kimia et al., 1999; Sharon et al., 2000]). Perhaps this was because of the frequency at which these structures appear in visual stimuli, the importance associated with edge detection (both computationally and physiologically [Hubel and Wiesel, 1977]), and the fact that the Gestalt principle of good continuation was explicitly phrased and demonstrated in terms of curves [Wertheimer, 1955b].

But of course, curves are not the only generic structures to which the notion of good continuation, or more generally that of structural coherence, can be applied. In this thesis we set course to study a different class of structures which we call *visual flows*. Unlike curves, these structures played only a secondary role in computational perceptual organization, despite ample psychological and physiological evidence for their importance. The rest of this introductory chapter introduces the notion of visual flows and presents their many instances in visual stimuli. It also discusses their perceptual significance and it reviews related computational approaches that have been proposed to handle them. In the following chapters, these structures are investigated from a multitude of perspectives; we study them geometrically to better understand their behavior and to inject formal meaning to their structural coherence (just as it was done in various ways for curves); we develop both a computational framework and a system that compute coherent visual flows from images in a contextual, distributed, and biologically-motivated fashion; we examine the biological plausibility of the geometrical analysis through the psychophysics of texture segmentation and the plausibility of the computational model through the functional organization of the primary visual cortex. These and other contributions are first introduced in Section 1.4.

1.1 Flows in visual stimuli

Visual flows are visual structures defined by their orientation content - dense visual precepts characterized by local parallelism and slowly varying dominant local orientation (almost everywhere). Although we will mostly deal with structures whose local parallelism and varying dominant orientation are *perceived directly*, we require that these properties only *characterize* the object of interest, i.e., they should exist at the representational level but not necessarily beyond it.

1.1.1 Texture

Perhaps the most direct example of a visual flow occurs in the domain of textures, where it is called a *texture flow*. In natural visual stimuli such structures occur at different scales and as a result of a variety of processes. For example, hair growth typically organizes animal furs in a locally parallel manner, and similar structure results from the morphogenesis of other biological tissues such as muscle and skin. Pigmentation in the animal kingdom often results in locally parallel stripes, as is frequent in seashells, fish, certain kinds of antelopes, and, of course, zebras. Texture flows also occur as a result of physical processes: flows of fluids, accumulation of geological layers, and rock formations, to name but a few. A sample of patterns is illustrated in Figure 1.1.

Texture flows are also common in artifacts and man-made objects due to production processes that include linear or rotational motions, as well as repetitions. It is a common structure in technical visualization and for centuries it has been used by artists as a tool to convey both the shape and shading of smoothly varying surfaces and their discontinuities (Fig. 1.2).

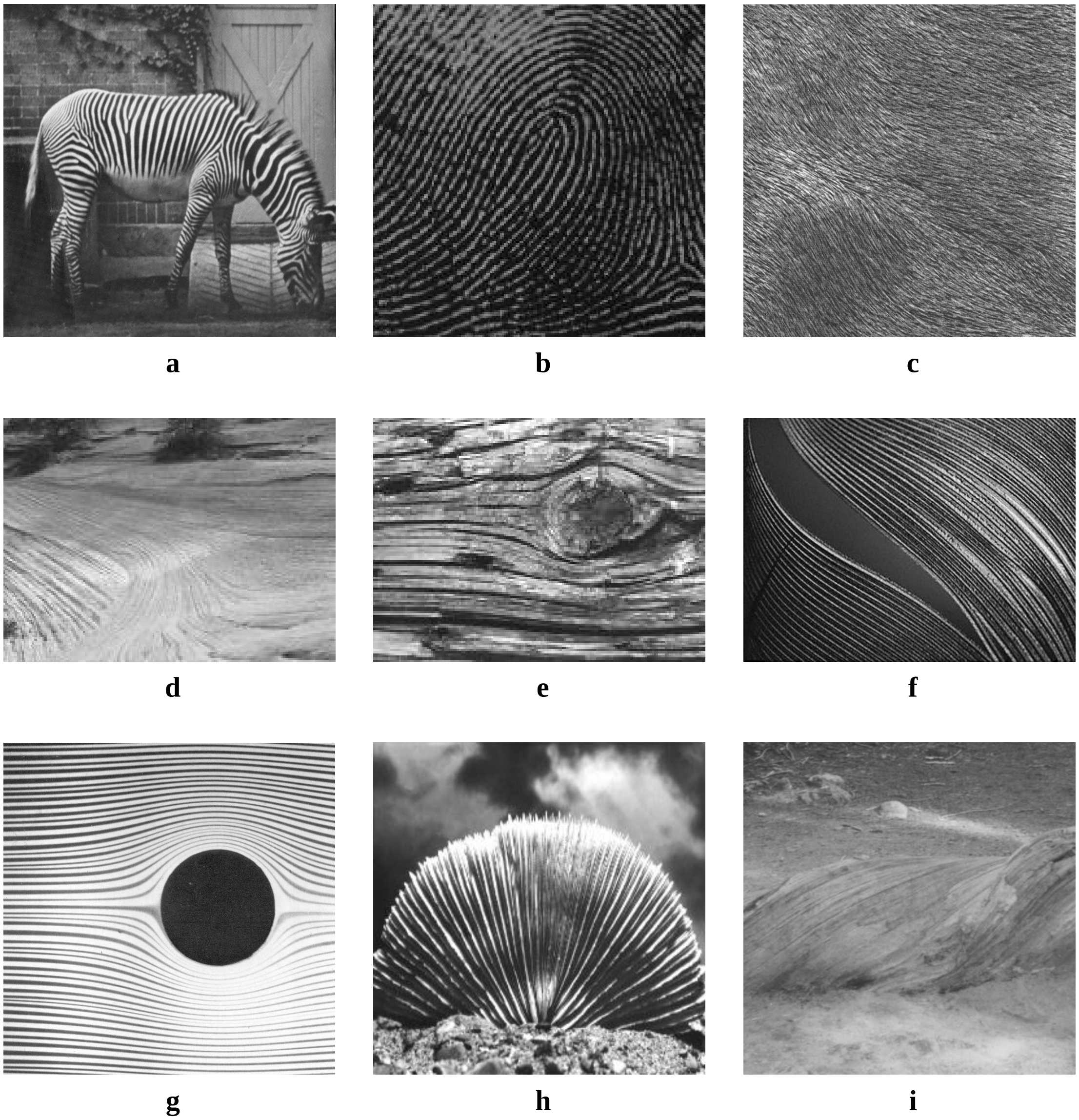


Figure 1.1: **Natural texture flows.** **(a)** Stripes on a zebra. **(b)** Fingerprint. **(c)** Animal fur [Brodatz, 1966]. **(d)** Rock formation (Zion national park). **(e)** Grain of a tree stump [Brodatz, 1966]. **(f)** Detail of a feather. **(g)** Fluid flow [Van Dyke, 1982] **(h)** Sea shell. **(i)** Exposed tree roots (Yosemite national park).

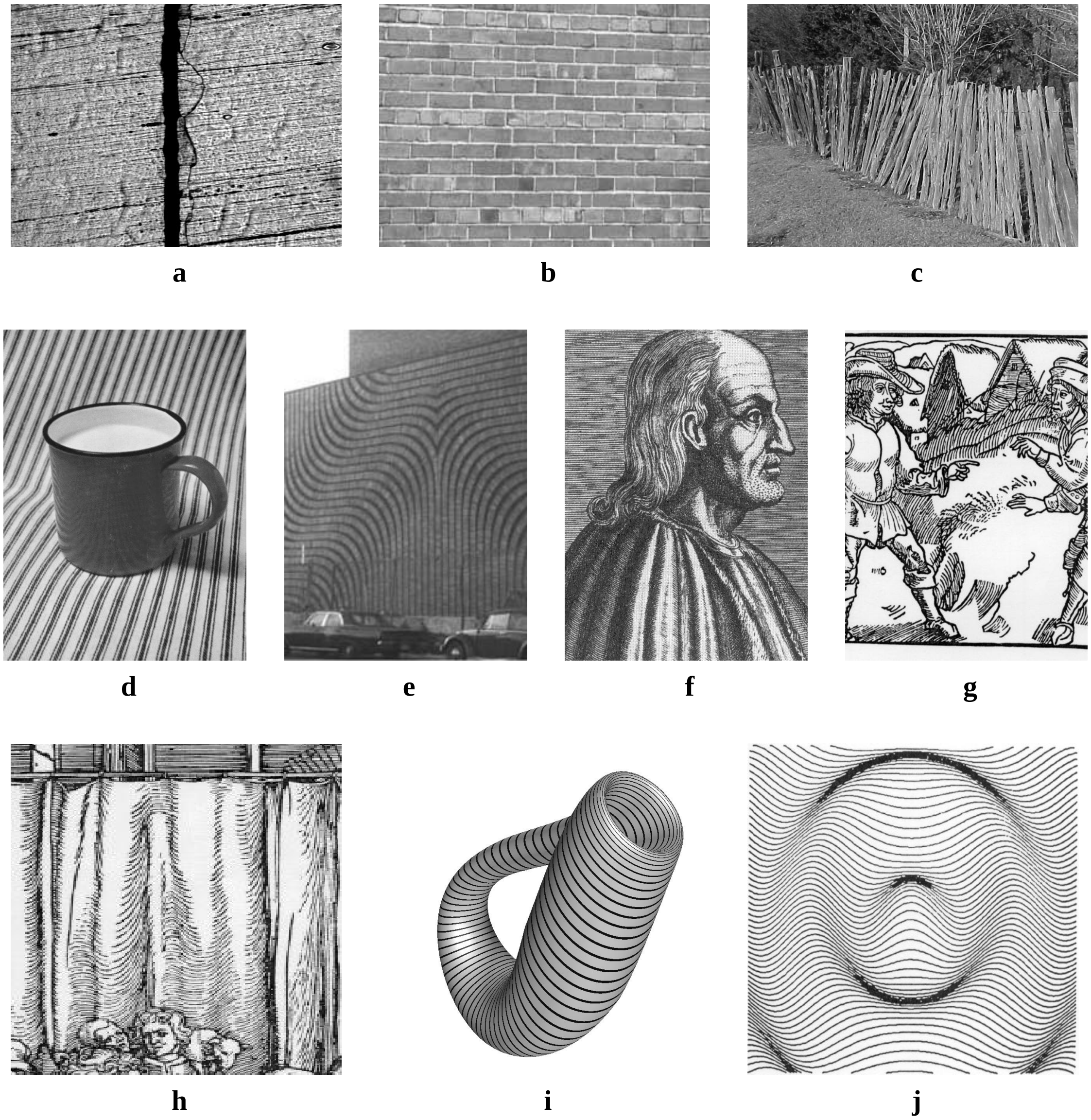


Figure 1.2: **Texture flows in man-made objects and the arts.** **(a)** Back side of a silicon wafer (courtesy of Yaki Levi, Inspectech Ltd., Israel). **(b)** A brick wall. **(c)** Wooden fence. **(d)** Table cloth. **(e)** Lincoln hospital, The Bronx, New York City. **(f)** A woodcut of the philosopher St. Anselm of Canterbury. **(g)** A detail from an early woodcut by Dürer [Kruth, 1963, panel 31]. **(h)** A detail from a later woodcut by Dürer [Kruth, 1963, panel 175]. **(i)** A textured Klein bottle (courtesy of Paul Bourke, Swinburne University of Technology, Australia). **(j)** A technical drawing of a 2D function.

Texture flows are visual flows whose locally parallel and oriented nature, thus their “flow”, is observed directly. The notion of flow in this case is not only intuitive but has a formal justification as well. In mathematics, the notion of flow arises mostly in the context of dynamical systems and first order differential equations. In the image plane, this amounts to systems of the form

$$\left(\frac{dx}{dt}, \frac{dy}{dt}\right) = \mathbf{w}(x, y) = (u(x, y), v(x, y)) \tag{1.1}$$

where $\mathbf{w}(x, y)$ is a differentiable vector field. Given $\mathbf{w}$, a *trajectory*[2] of $\mathbf{w}$ is defined as a differentiable parametrized curve $\alpha(t) = (x(t), y(t))$ such that $\alpha'(t) = \mathbf{w}(\alpha(t))$. The fundamental theorem of local existence and uniqueness of solutions of Equation 1.1 guarantees that a unique trajectory passes through each point $\mathbf{q} \in \mathbb{R}^2$ and that different trajectories do not cross. Furthermore, the trajectory passing through q “varies differentiably with q” [Do Carmo, 1976]. Formally, this is characterized as follows:

Theorem 1 (e.g., in [Do Carmo, 1976]) *Let $\mathbf{w}$ be a vector field in an open set $U \subseteq \mathbb{R}^2$. For each $\mathbf{p} \in U$ there exist a neighborhood $V \subset U$ of $\mathbf{p}$, an interval I, and a* differentiable *mapping $\alpha : V \times I \to U$ such that for each fixed $\mathbf{q} \in V$, the curve $\alpha(\mathbf{q}, t)$, $\quad t \in I$, is the trajectory of $\mathbf{w}$ passing through q.*

The map α is called the (local) *flow* of $\mathbf{w}$ at p and the fact that it is differentiable implies that there exists a neighborhood $P \subset V$ of $\mathbf{p}$ in which all trajectories are *roughly parallel to each other*. Coloring selected trajectories with distinct colors thus yields a visual pattern reminiscent of those in Figures 1.1 and 1.2, which concludes the analogy and the justification for the term texture (or more generally, visual) *flows*.

The association of texture flows to formal flows implies that every visual cue whose representation could involve spatially varying orientation, local parallelism, or differentiable vector fields, may be viewed as a visual flow. Next we briefly survey a few additional examples and their classification as visual flows.

---

[2] also called path, orbit, streamline, or integral line.

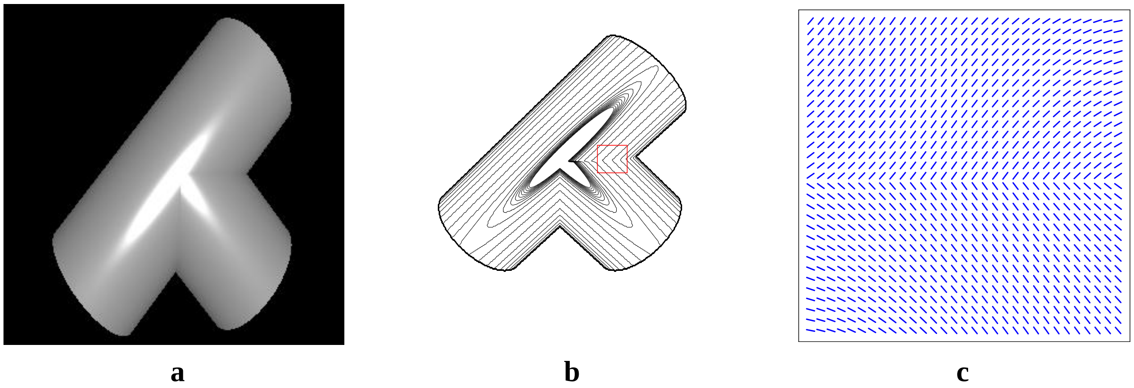


Figure 1.3: **Shading as a visual flow.** **(a)** An image of a piece of an industrial pipe. **(b)** The level set structure of the intensity and a region of interest (ROI). **(c)** The shading flow field in the ROI. Note that although the intensity is continuous, the field undergoes sharp discontinuities to signal fusion of distinct parts.

1.1.2 Shading

Shading, one of the primary low level cues in vision, has been studied computationally mostly in the context of shape inference [Horn and Brooks, 1989]. Represented directly, shading is a scalar function over the image plane. However, any scalar function can be represented, at least up to certain local contrast variations, via its level sets – the curves along which the intensity remains constant, also called iisophotes [Koenderink, 1990] or isoluminance contours. Underlying this "topographic map" is a vector field, sometimes called the *shading flow field* [Breton and Zucker, 1996], that makes explicit the direction of the isophotes at any given location. The geometry of this field is a precursor for shape and its interaction with edge geometry provides useful information for edge classification [Šára, 1994; Huggins and Zucker, 2001a].

The shading flow field of smooth surfaces is a visual flow. It possesses local parallelism and slowly varying dominant local orientation almost everywhere. Under these conditions, discontinuities in the shading flow field signal corresponding discontinuities in the three dimensional scene and thus their detection and localization is critical for shape description and analysis (Figure. 1.3).

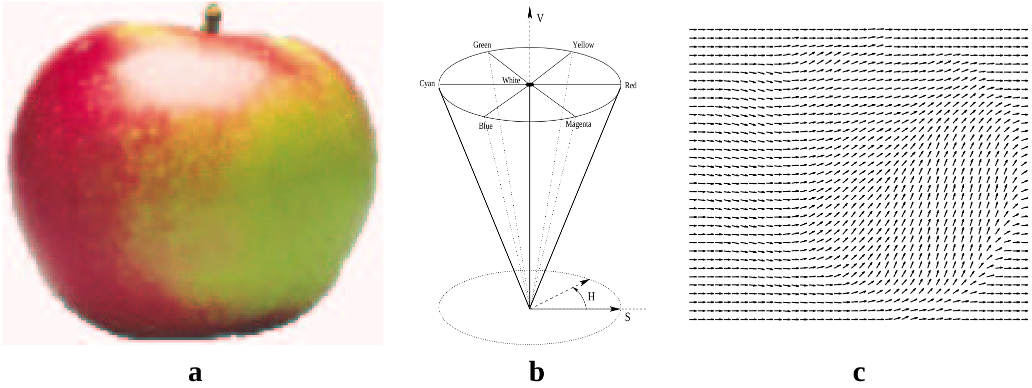


Figure 1.4: **Color (hue) as a visual flow.** **(a)** A typical color image. **(b)** The HSV color space. The hue component is periodic. **(c)** Portrayed as an oriented vector at each point, the hue field of the image is a visual flow (locally parallel with dominant local orientation everywhere).

1.1.3 Color

While much of computer vision and image processing research focuses exclusively on the luminance domain, many imaging devices, most notably the human eye, provide multispectral representations which in the brain are interpreted as color images.

The direct representation of color images follows either the measurement or the synthesis process to abstract each spectral channel as a distinct scalar function over the image plane (e.g., the RGB, CMY, or the YIQ color models [Foley et al., 1996]). In contrast, a perceptually-motivated representation of color images should separate the achromatic (lightness, or value) and chromatic components of the color image. Such a distinction is made explicit in the HSV color representation in which the chromatic component is further described with two independent variables for hue and saturation (Figure 1.4b).

Of the three perceptual dimensions of HSV, the hue is periodic and can be represented as an angle. Although it is not spatial or directly observable, this angle associates a *geometric* meaning to the hue component of the color (or alternatively, to the chromatic component as a whole) and thus suggests a geometrical treatment. Portrayed as an oriented vector at each point, the hue component of color images is in fact a vector field over the image plane. Based on the discussion above, we thus consider it as a visual flow (Figure 1.4).

a **b**

Figure 1.5: **Motion as a visual flow. (a)** A pair of images in a motion sequence. **(b)** The optical flow obtained from the pair (all images courtesy of Pierre Kornprobst ,Odyssee Lab, INRIA). Color of vectors encode magnitude. When the magnitude is positive, the field is locally parallel and has dominant local orientation almost everywhere.

1.1.4 Motion

The notion of motion is clearly associated with direction and thus it should be of no surprise that visual cues that emerge from motion can be represented as visual flows. Whether one refers to the true *motion field* or the *optical flow field* (which associates a motion vector to each pixel in an image sequence under the assumption that intensity is preserved [Nalwa, 1993]), the discrete nature of the visual world, as well as the motion of its finite size units (i.e., objects), dictate that motion related fields are piecewise smooth, thus locally parallel with dominant local orientation almost everywhere. Hence in our terms they too are visual flows[3]. One example is shown in Figure 1.5.

1.1.5 Stereo

The discussion above demonstrates that visual flow-like structures can be found in a variety of visual cues, either explicitly or implicitly; If the visual cue can be represented as a spatially varying scalar in $\mathcal{S}^1$, the visual flow is treatable directly; If, on the other hand, the measurement lives in $\mathbb{R}$, the visual flow is implicit in its level sets. For some cues among the latter, a visual flow representation may prove pointless. For others, it makes explicit critical properties of interest. For others yet, it may emphasize aspects of the input which become useful under certain conditions only. This may be the case, for example, with stereo vision.

---

[3] Note that like texture, but less so in shading or color, motion or optical flow fields may vanish in large portions of the image. Nevertheless, in places where they are strictly positive, they do behave as visual flows.

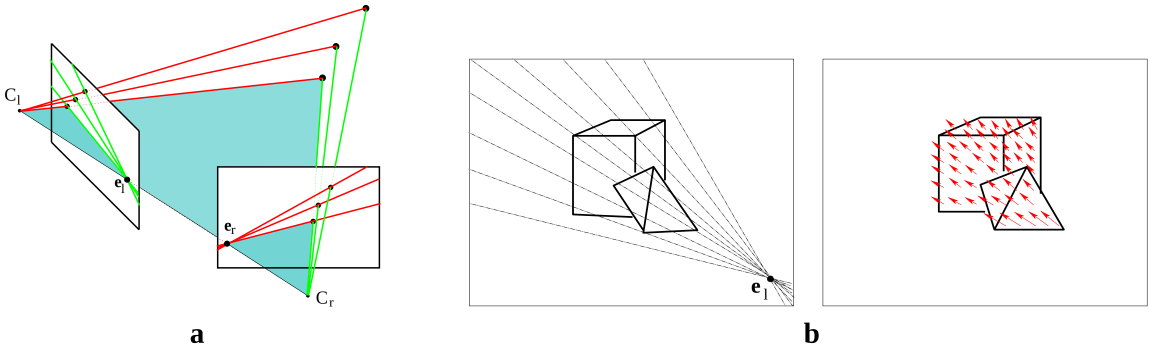


Figure 1.6: **Stereo as a visual flow.** **(a)** Epipolar geometry dictates that correspondence between points in stereo pairs is constrained to a family of image lines, all of which intersect at the epipoles (which may be at infinity) **(b)** The disparity flow field of each image is the field of disparity vectors relative to the other image. This field always corresponds to the geometry of the epipolar lines in the other image. It has a single point singularity (unless the epipolar lines are strictly parallel) and may have additional line singularities where the sign of disparity changes. Additionally, singularities in the *magnitude* of this field corresponds to surface discontinuities in the scene.

Not unlike motion pairs, stereo information, in particular correspondence and disparity, is oriented in nature, albeit in a restricted manner dictated by the epipolar geometry. Once the latter is determined, either through camera calibration or image rectification, representing the disparity as a vector field may have no added value. However, when the epipolar geometry is not known, such a representation can assist in recovering it. In particular, the epipolar constraint dictates that the trajectories of the *disparity flow field* are all straight lines. Thus viewing disparity as a flow field can assist in detecting the epipole (i.e., the image point where the epipolar lines intersect) since this is the only point where the field undergoes a generic point singularity. A depiction of disparity as a visual flow is illustrated in Figure 1.6.

1.1.6 Overlapping visual flows

Once a geometrical structure is defined, it is interesting to consider the superposition of several of them. Indeed, certain classes of visual flows are different from others in that they *can* spatially overlap. In other words, it is possible for more than one of them to occupy the same part of the image. For example, in the texture domain, textures with multiple dominant orientations at a point

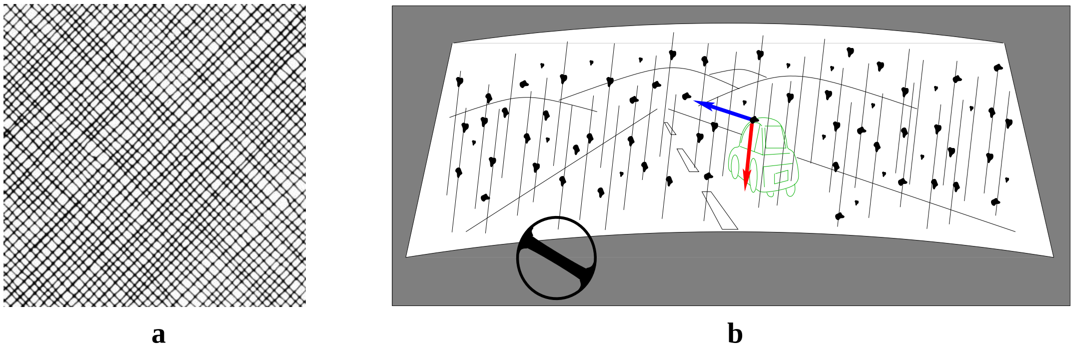


Figure 1.7: **Overlapping visual flows.** **(a)** A natural texture pattern with two dominant orientations at every point. **(b)** A schematic view of traffic in a snowy day. The motion field of a scene like this one contains regions of overlapping motions (e.g., blue motion of the car and red motion of the snow).

not only exist in their own right (see [Hornung, 1976; Brodatz, 1966] for many examples), but can arise from transparency, or what was termed *texture laciness* [Watanabe and Cavanagh, 1996]. In the motion domain, such structures can result from complex dynamic scenes such as persons walking in a snow storm (vertical and horizontal motions overlap) or a moving observer viewing a swarm of flying insects in a cluttered environment [Langer and Mann, 2001, 2002]. Examples of such overlapping flows are illustrated in Figure 1.7.

Unfortunately, the existence of overlapping flows was largely ignored by virtually all methods that analyze oriented patterns. Acknowledging their importance, the framework developed in this thesis is tailored to handle such structures, both at the representational and the computational levels.

1.2 Perceptual significance of Texture flows

Of all visual flows, texture flows are observed directly and their locally parallel nature, as well as their dominant local orientation, is visually explicit. Thus, texture flows provide a unique opportunity to examine the effect of the flow structure on the perception and interpretation of visual stimuli. Indeed, psychologists and psychophysicists have been investigating these patterns from at least three different points of view – preattentive vision, perceptual organization, and shape perception – to suggest that their role is critical.

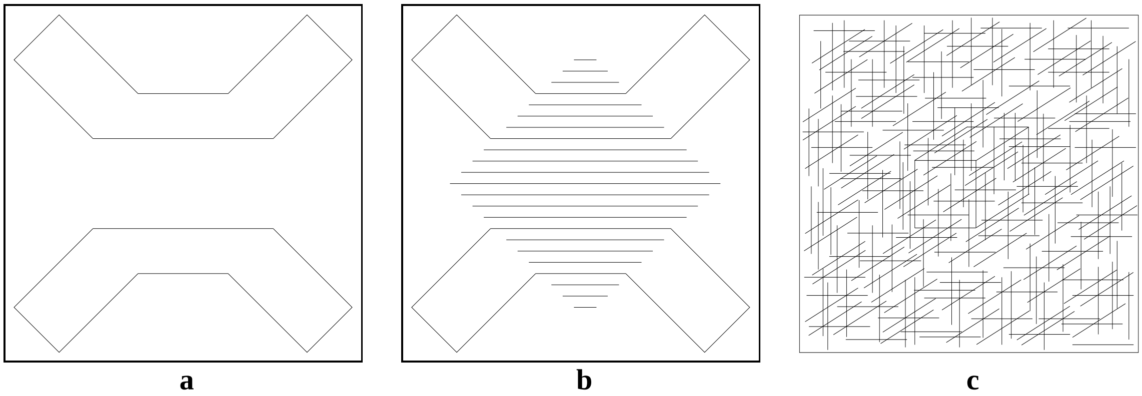


Figure 1.8: **Social conformity of a line. (a)** This image is unambiguously interpreted as two polygons. **(b)** Adding few parallel lines induce a grouping process that changes the global interpretation of the scene to a cross occluded by a diamond. **(c)** Social conformity of a line is a form of geometric camouflage. Can you tell which well known object exists in this image? (answer in Figure 1.9).

1.2.1 Texture flows and perceptual organization

From its inception by the Gestalt psychologists, perceptual organization became recognized as that part of the visual process that imposes organization of sensory input through grouping, figure-ground segregation, and modal and amodal completion. Psychological observations suggest that most of these processes apply to texture flows as well, albeit separately from the corresponding processes that act on curves.

Grouping of 2D locally parallel structure and the tendency of the visual system to organize it into coherent units has been demonstrated at least as early as 75 years ago [Galli and Zama, 1931]. This observation was made an explicit grouping principle by Kanizsa [Kanizsa, 1979], who dubbed it the “social conformity of a line”, and its essence is rather intuitive: locally parallel lines, or texture flow in general, tend to become coherent units to the degree that their elements completely lose their individual identity. As Figures 1.8 and 1.9 demonstrate, the effect of this grouping on the interpretation of visual stimuli is profound.

The tendency of the visual system to organize locally parallel structure also was demonstrated by the Moiré patterns of Glass [Glass, 1969, 1973; Prazdny, 1986; Wilson and Wilkinson, 1998]. In these structures, now called Glass patterns, two superimposed and slightly transformed copies

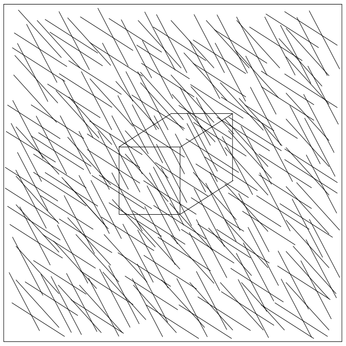


Figure 1.9: **Social conformity of a line.** By rotating the distractors in Figure 1.8c the “social organization” of the lines in the image changes and the cube is revealed.

of the same random dot pattern give rise to a globally coherent pattern of local parallelism almost everywhere (indeed, a texture flow). Surprisingly, however, it is neither similarity (as *all* dots are identical) nor proximity (see Figure 1.10) that act as a grouping principle in this case.

Texture flows also appear to posses the same “gap completion” property that has been investigated so extensively for curves (e.g., [Sha’ashua and Ullman, 1988; Kimia et al., 1999; Williams and Jacobs, 1997]). However, texture flow completion cannot be explained as multiple completions of individual curves. The problem is not only the combinatorics of correspondence between segment pairs along the occluder boundary, but also the fact that such a one-to-one correspondence does not necessarily exist. Similarly, it can be argued that perceptual filling-in of texture flow is insensitive to the exact locations at which the integral lines meet the occluder, an observation that stands in contrast to the qualitative changes (e.g., an emergence of inflection points) that such small perturbations may introduce to curve completions. A few illustrations that demonstrate all these points are shown in Figures 1.11 and 1.12).

1.2.2 Texture flows in low and high level vision

In addition to perceptual organization, texture flows have been shown to take a critical role in many visual processes, from preattentive texture segmentation to high level shape perception. For example, texture segregation has long been attributed to changes in the distribution of elementary features [Julesz, 1981, 1986], among which orientation is particularly conspicuous [Beck, 1966]

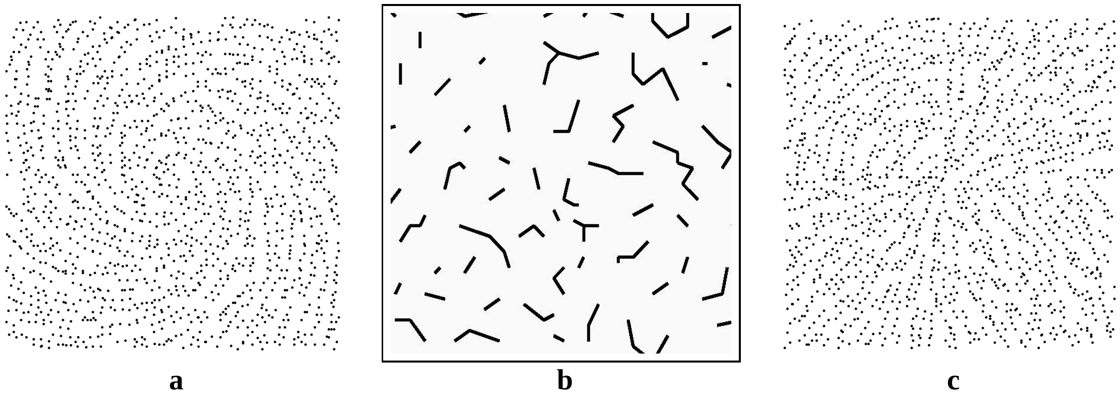


Figure 1.10: **Organization of parallel structure in Glass patterns.** **(a)** A swirly Glass pattern. **(b)** A zoom into the nearest neighbor graph of a typical Glass pattern implies that the emergent structure cannot be a direct result of proximity. In fact, as was pointed by Glass [Glass, 1973] and later again by Stevens [Stevens, 1978], if two points p and q in the Glass pattern are perceptually correlated, it is almost always the case that a few other uncorrelated points lie *closer* to each of p and q. **(c)** Another Glass pattern with easily segmented patches. Interestingly, the nearest neighbor graph in the neighborhood of the discontinuities looks similar to the one in panel b and provides no local indication for segmentation.

and deeply linked to the functional organization of the primary visual cortex [Hubel and Wiesel, 1977]. It has been shown that orientation-based texture segmentation depends primarily on one or two orientation gradients [Landy and Bergen, 1991; Nothdurft, 1991, 1993; Mussap and Levi, 1999] and most importantly, that orientation variations can occur *within* otherwise coherent regions, as is inherent in our informal definition of visual flows (Section 1.1). Unfortunately, however, most of the psychophysical research has concentrated on orientation-defined textures *without* such internal variations [Nothdurft, 1985a; Landy and Bergen, 1991; Wolfson and Landy, 1995; Regan et al., 1996; Caputo, 1997; Kwan and Regan, 1998; Wolfson and Landy, 1998; Caputo and Casco, 1999; Li, 1998a; Motoyoshi and Nishida, 2001] and undermined the observation that orientation-defined textures are rarely constant in natural images. We address this issue in great detail in Chapter 3.

On the other side of the spectrum, texture flows have been shown to play a role in 3D shape perception as well. While computational studies argue that under certain conditions shape information *can* be extracted from locally parallel surface contours [Stevens, 1981, 1983], psychological observations have shown that such structures greatly facilitate the perception of shape in the human visual system [Todd and Reichel, 1990]. One illustration is shown in Figure 1.13 (see also Figure 1.2i,j).

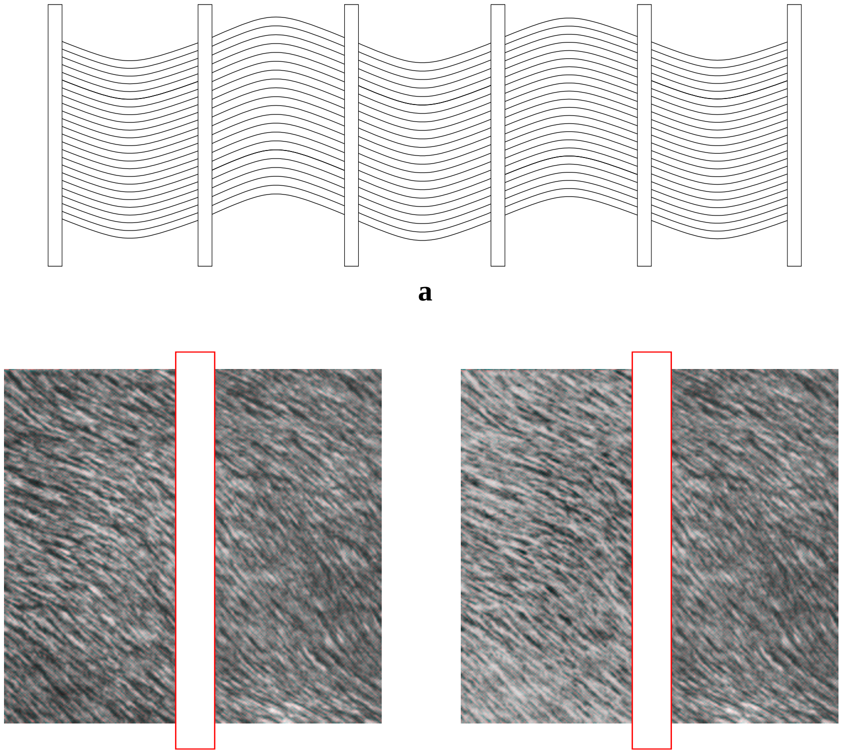


Figure 1.11: **Amodal completion of texture flow. (a)** This flow is amodally completed under the occluders and appears uniform and coherent. It requires a great deal of scrutiny to realize that each of its unoccluded segments contains a different number of lines, which precludes the completion of texture flow gaps by a line-by-line process. **(b)** The animal fur in this image [Brodatz, 1966] is easily completed behind the occluder despite the fact that no individual hair is seen at both of its sides **(c)** This fur is similarly completed behind the occluder despite the fact that now its left part is the complement of the original image and the position of all individual hairs is shifted.

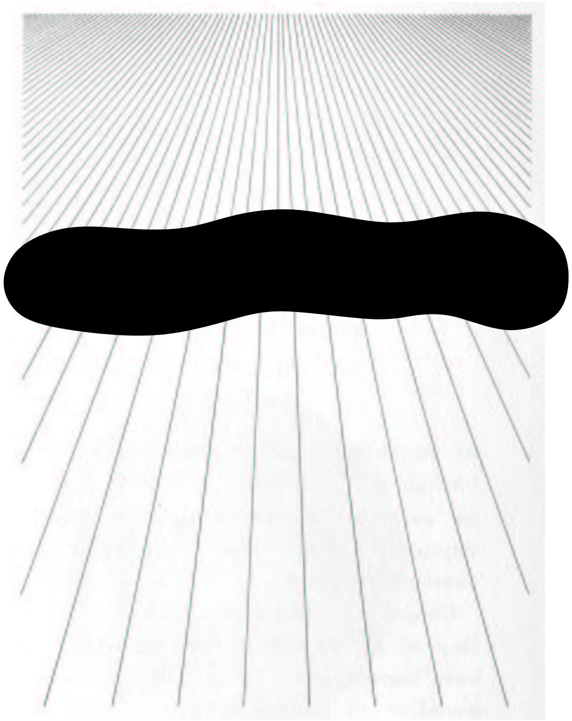


Figure 1.12: **Amodal completion of texture flow.** This flow is naturally completed behind the occluder to result in the perception of a single planar surface. However, the number of flow lines that meet the top part of the occluder is more than twice the one at the bottom. Thus, no one-to-one correspondence between lines is possible, which excludes a curve-based completion process. An original variation on a demonstration by Gibson [Gibson, 1950, page 93], this figure also illustrates how texture gradients alone are not sufficient to result in perceptual discontinuities. Note that without the occluder this image is normally perceived as a step in depth.

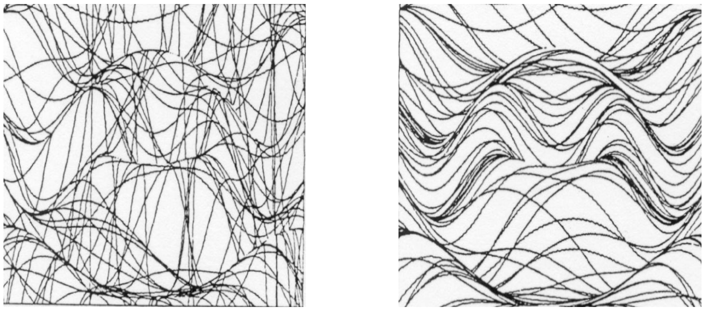


Figure 1.13: **Texture flow facilitates shape perception.** Although both images depict the same surface, its shape is apparent only when the set of surface contours possess some degree of parallelism (reproduced from [Todd and Reichel, 1990])

1.3 Computational studies of visual flows

Locally parallel structures have been of interest to computer vision researchers for more than two decades now and it is interesting to examine the evolution of the computational accounts in terms of methods and goals. Thus, this section is divided into early (up to late 90's) and recent (1998 and later) work, both of which are briefly summarized next.

1.3.1 Early work

Following Glass's work on random dots Moiré patterns [Glass, 1969, 1973], it was Stevens [1978] who first raised the issue of computing locally parallel structure by formalizing Glass's own lateral excitation idea into a computational, histogram based, approach. While Stevens focused on Glass patterns and advocated a discrete and sparse representation, Zucker [1983] argued for a dense, vector field-based representation and emphasized the role of locally parallel structure in the general context of perceptual organization. It was also the initial attempt to link the flow's local behavior to its *global structure*. In the same spirit, the importance of recovering the flow's global structure was argued by Kass and Witkin [1987], who described a filtering method for extracting the flow's local orientation, but provided no robust method for the integration of the inherently noisy measurements. Another filter, based on the gradient of a Gaussian, was proposed by Rao and Schunck [1991]. Not unlike Kass and Witkin [1987], they also obtained a more reliable measurement of dominant orientation by appropriately averaging the filter's response over a neighborhood. Other schemes were developed by Jain and colleagues [Rao and Jain, 1992; Shu and Jain, 1993] to fit the flow's local behavior to a simple first-order model (the phase portrait) via least squares. Their methods have a few important advantages: they yield a global description, they allow a classification of the flow pattern via its differential properties [Koenderink and van Doorn, 1976], and they are able to localize point singularities. On the other hand, they cannot handle multi valued data (overlapping flows or textures of multiple dominant orientations) and they behave poorly along line discontinuities.

1.3.2 Recent work

Following the original work on scale space [Witkin, 1983; Koenderink, 1984a] and a wealth of work on nonlinear diffusion of scalar [Perona and Malik, 1990; You et al., 1996; Weickert, 1997; Weickert et al., 1998; Yezzi, 1998; Sochen et al., 1998; Acton, 1998; Black et al., 1998] and vector valued images [Sapiro and Ringach, 1996; Sochen et al., 1998; Blomgren and Chan, 1998], the trend in the analysis of oriented patterns has recently shifted to formal models of orientation diffusion.

The aim of orientation diffusion is to denoise orientation fields and to obtain a multi scale representation of the data set. Not unlike the diffusion of raw gray levels, it is commonly acknowledged that this process should preserve "interesting" events - discontinuities and singularities in the structure - "as long as possible" in the scale space (e.g.,[Sochen et al., 1998, p. 310]). However, unlike the diffusion of gray levels, dealing with orientations introduces new complication since the signal lies in a non Euclidean space. Thus, the periodic behavior of directional data entails special care in the definition of energy functionals and in operations like averaging and differentiation. These issues were originally raised by Perona [Perona, 1998] who used an internal coordinate θ in $\mathcal{S}^1$ to represent the directional data, embedded it in $\mathbb{C}$ via exponentiation to carry the continuous diffusion, and used a physically-motivated energy functional to define a plausible discrete diffusion.

Energy minimization is also the basis of the orientation diffusion proposed by Tang *et al.* [Tang et al., 2000] who represented the directional data as unit length vector field in $\mathbb{R}^n$. Borrowing from the theory of harmonic maps in liquid crystals, they proposed to minimize the harmonic energy of a pattern by flowing toward its critical points via gradient descent. Using different norms one can change the sensitivity of this diffusion process to large scale discontinuities and hence the rate at which such structures in the orientation field are smoothed away. Not unlike the original Perona and Malik diffusion [Perona and Malik, 1990], this practical behavior manifests itself as an "edge stopping" function that is incorporated into the divergence form of the diffusion expression. It should be noted that both Tang *et al.* [Tang et al., 2000] and the related Total Variation (TV) formulation of Chen and Shan [Chan and Shen, 1999] incorporated a projection term to make sure the solution stays in $\mathcal{S}^1$. As noted in [Tang et al., 2000], this projection is difficult to control numerically and additional renormalization is usually required.

A different but related approach to orientation diffusion was suggested by Kimmel and Sochen [2000] who represented the directional data as a 2D surface embedded in a higher dimensional Riemannian manifold. Denoising is done by smoothing the signal surface via the Beltrami flow, where distances are measured *on* the surface itself, rather than along the domain that supports it. The outcome of this process is an edge stable diffusion process which is readily generalizable to higher dimensions.

Orientation diffusion methods are a powerful tool for progressively transforming a noisy set of measurements into a piecewise smooth representation. Their inhomogeneous nature allows them to adaptively damp their influence in areas "suspected" as structural edges and thus to preserve the prominent structure to deeper levels of the scale space. But, orientation diffusion methods have a few limitations that put their effectiveness in the analysis of oriented patterns in question. For example, existing methods are not designed to deal naturally with "holes" or missing measurements, which undermines both occlusion and the more general fact that measurements may be intrinsically sparse. At the same time, the same methods will diffuse even non-flow input to some meaningless average, thus creating structure from non-structure which should have been rejected altogether. In addition, diffusion schemes do not deal with data containing *multiple* dominant orientations at a point, which prevents the analysis of overlapping flows as described above. And finally, orientation diffusion may distort certain kinds of perceptual discontinuities which do not conform to the edge-like discontinuity structure assumed by their edge stopping function. All these limitations are discussed in detail in Chapter 4 (in particular, Section 4.1), where we develop a computational approach that can overcome them.

1.4 Overview and contributions

In this thesis we investigate visual flows in a multidisciplinary fashion. Our research “pendulum” will swing back and forth between four corners of a pyramid of disciplines – one geometrical, one computational, one psychophysical, and one physiological. Results in one corner will swing the pendulum to a different one, where they will trigger new questions, fresh investigations, and important consequences. When the pendulum comes to rest, we will have accumulated novel (and sometimes surprising) contributions both to computational vision, computational neuroscience, and the understanding of human perception.

The study of any visual structure must begin with the understanding of its formal properties, characteristics, and parameters that govern its behavior. Thus, we begin our journey in Chapter 2 by placing our pendulum in the geometry corner. We analyze formally the intuition behind visual flows to show that it is governed locally by *two* curvatures. We further show that the behavior of these curvature functions is not independent and that they necessarily covary in any visual flow.

One goal of our work is to develop a computational approach to the perceptual organization of visual flow. As with the Gestalt principle used for the perceptual organization of curves, a notion of “good continuation” is needed for our structures as well. Following the initial geometrical analysis, we introduce such a notion based on curvatures. We rigorously derive a unique local model that formalizes “good”, or coherent visual flows and study its properties. For example, we prove that our model, which in a certain representation takes the form of a right helicoid, is closely related to orientation diffusion processes, especially those which depend on extremizing the p-harmonic energy.

Naturally, the emergence of curvatures as intrinsic local descriptors has further computational implications, but it first raises the question about their possible manifestation at the perceptual level. Moving to the psychophysical corner, in Chapter 3 we thus explore this issue by conducting an exhaustive psychophysical study about the sensitivity to curvatures in orientation-based texture segmentation. We find that significant sensitivity indeed exists and show that previous models are at best incomplete.

With both geometrical and psychophysical support for the importance of flow curvatures, our pendulum now swings forward to the computational corner. The advantage of having a model for the local behavior of a coherent texture flow lies in the ability to assess the degree to which a particular measurement is compatible, or consistent, with the context in which it is embedded, and whether or not that context is part of a single coherent object. Motivated by the columnar architecture of the primary visual cortex [Hubel and Wiesel, 1977], in Chapter 4 we apply this contextual computation through a relaxation labeling network. We construct the compatibilities for our network from the right helicoidal model and show how it may be made stable along discontinuities. We implement all these ideas into one computational system that can handle sparse data sets and multi-valued inputs; It can fill in missing measurements over short distances and it can reject large scale non-flow structures; It is able to preserve both singularities and line discontinuities, including those which orientation diffusion is prone to distort. To the best of our knowledge, no computational approach to the analysis of visual flows possesses all these properties. We apply our system to numerous synthetic and natural images, both for texture flows and shading flows.

The progress in the psychophysical and the computational corners now shifts our attention the physiological one. Does the results presented thus far have any implications for computational neuroscience? On one hand, could our results entail new ideas with regard to the functional organization of the primary visual cortex (V1)? And on the other hand, is there evidence to suggest that V1 performs computations reminiscent to those in our computational system?

Questions such as these, which we address in Chapter 5, are typically difficult to answer, primarily because neurophysiological data accumulates slowly, partially, and with very little relation to, or guidance from, rigorous computational models. However, in this “detective game”, certain bits and pieces serve as clues. Since our computational system realizes a contextual inference approach which is based on the interaction between basic computational units, we focus our attention on the anatomical elements that are responsible for such interactions – the long range horizontal connections between cells in V1. To date, the most exhaustive data available on these connections is their distribution in the functional (orientation) domain, thus we similarly analyze our computational model to see whether or not its predicted patterns of interaction coincide with the measured data. The correspondence we find is striking, up to second statistical order. Our findings therefore support the possibility that long range horizontal connections in V1 indeed integrate data based on curvature-sensitive good continuation and suggest that further physiological inquiry in this direction is essential.

With all corners of the pyramid now connected under the same theme, our pendulum swings back to the computational corner for one last time. As we have mentioned, visual flows can be found in many and diverse forms. It is thus natural to examine the effectiveness of our computational system to domains where the visual flow is not as explicit as in, say, texture flows. In Chapter 6 we thus shift our attention to the domain of color (see Section 1.1.3) to apply our results to the denoising of color images. We introduce the notions of hue fields and color curvatures and propose, for the first time, a perceptual organization approach to handling this problem. We present results on a variety of synthetic and natural images to demonstrate how effective such an approach can be.

More remains to be done. Since the pendulum can (and should) swing many more times between the disciplines, Chapter 7 concludes this dissertation with selected future directions.

Chapter 2

The geometry of visual flows

2.1 Visual flow representation

Formalizing visual flows requires making explicit their basic characteristics – local parallelism and varying dominant orientation – through a choice of representation. While this can be done in different ways, the perceptual properties of visual flows (Sec. 1.2) strongly suggest that any formal representation must be continuous (as opposed to discrete) in the image plane.

Not surprisingly, the most direct way to represent visual flows is to use a scalar orientation function $\theta(x,y):\mathbb{R}^2 \rightarrow \mathcal{S}^1$ that defines a dominant orientation at each point (x,y) (e.g., in [Perona, 1998]). However, other representations are possible as well. For example, any visual flow can be represented as a 2D submanifold of $\mathbb{R}^2 \times \mathcal{S}^1$. Using local charts to cover this manifold, or a global parametrization and an adequate periodic function of orientation (e.g. [Sochen et al., 2001]), this representation can be embedded as a surface in $XY\theta$, i.e., the space $\mathbb{R}^3$ whose Z axis represents orientation. A *single valued* visual flow patch then takes the form $s(x,y) = (x,y,\theta(x,y))$, i.e., up to 2π-periodicity in the θ direction it is a graph of the scalar function $\theta(x,y)$ mentioned above. Note that although the scalar and surface representations appear very similar, it is important to realize that they are *not* identical, as each makes explicit different aspects of the data (e.g. [Marr, 1982, pp. 20-22]), as well as suggesting different analysis tools.

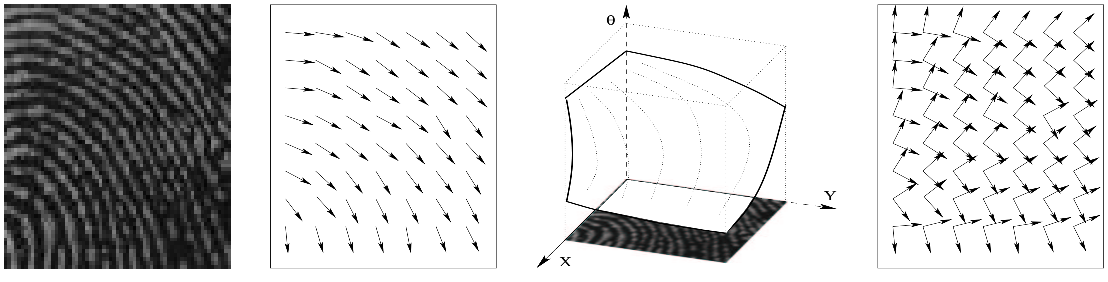
Alternatively, visual flows can be represented as unit length vector fields $\hat{\mathbf{E}}_T(x,y):\mathbb{R}^2 \rightarrow \mathbb{R}^2$,

$\hat{\mathbf{E}}_T(x,y) = (V_1(x,y), V_2(x,y))$ with $V_1^2 + V_2^2 = 1$, (e.g., in [Tang et al., 2000]). We insist on a *unit length* field because magnitude is neither part of our informal geometrical definition, nor is it available in all cases. Without magnitude, each such vector field $\hat{\mathbf{E}}_T(x,y)$ is an equivalence class of all vector fields $\mathbf{w}(x,y) = (u(x,y), v(x,y))$ whose trajectories completely overlap (c.f. Sec. 1.1.1), thus emphasizing their geometry rather than their velocities.

Finally, visual flows can be represented as a 2D *frame field* $E(\mathbf{q}) = \{\hat{\mathbf{E}}_T(\mathbf{q}), \hat{\mathbf{E}}_N(\mathbf{q})\}$ over the image plane [O'Neill, 1966]. Naturally, one would choose $\hat{\mathbf{E}}_T(\mathbf{q})$ as a basis vector tangent to the flow, and $\hat{\mathbf{E}}_N(\mathbf{q})$ as normal to it. Pushing the role of representation once again, we stress that frame fields are *not* identical to the unit length vector fields mentioned above since they make explicit certain aspects of visual flows that are not emphasized by the latter.

An important issue in the representation of visual flows is the domain of orientations. While certain flows, such as color and motion, can take orientations in the range $[0^\circ, 360^\circ)$, others, notably texture, are naturally constrained to $[0^\circ, 180^\circ)$. This difference can be handled explicitly, for example by using a *field of directions* [Do Carmo, 1976] instead of regular vector fields. However, since directions and π-periodic signals can be mapped to 2π-periodic orientations with a simple linear mapping [Perona, 1998; Granlund and Knutsson, 1995], we use both cases interchangeably.

In this chapter we rely primarily on the frame field representation, although we also use the others when appropriate. Selected representations are illustrated in Figure 2.1.



a **b** **c** **d**

Figure 2.1: **Representations of visual flows.** **(a)** An example of visual flow (cropped from Figure 1.1b). **(b)** Representation as a unit length vector field. **(c)** Representation as a surface patch in $XY\theta$. The height of this surface represents the local orientation of the texture at the corresponding spatial location. **(d)** Representation as a frame field. Note that although both vector and frame fields are drawn discretely, they are in fact defined continuously over $\mathbb{R}^2$.

2.2 The method of moving frames

The emergence of frame field representations is deeply rooted in differential geometry and dates back as far as the mid 19^{th} century to Frenet's and Serrent's work on curves in $\mathbb{R}^3$. Although the theory of surfaces was already richly developed at that time, the success of the celebrated Frenet equations, and his frame field approach, led Darboux to apply it to the study of surfaces (circa. 1880). Only a few years later, it was Cartan who brought this approach to full generality (early 1900) to realize its maximal power.

The idea behind the "method of moving frames" is strikingly simple and elegant: given a geometrical object of interest (a curve, a surface, a metric space, etc...), assign a natural frame to each of its points; then express the rate of change of the frame as it moves about the object only using the frame itself (which is, after all, a local basis). Doing so leads to intrinsic properties of the object under study and greatly simplifies its investigation and understanding. In the rest of this section we review the geometrical background necessary for our study of visual flows. Much of this material follows standard differential geometry texts [Do Carmo, 1976; O'Neill, 1966] albeit adapted and constrained to $\mathbb{R}^2$ on which visual flows are defined.

2.2.1 Preliminaries

Let $\mathbf{v_p}$ denote a *tangent vector* to $\mathbb{R}^2$, i.e., is a vector $\mathbf{v}$ whose tail (or application point) is set at $\mathbf{p}$ such that its head reaches $\mathbf{v} + \mathbf{p}$. The set of all tangent vectors to $\mathbb{R}^2$ at point $\mathbf{p}$ is called the *tangent space* to $\mathbb{R}^2$ at $\mathbf{p}$, denoted by $T_{\mathbf{p}}(\mathbb{R}^2)$. A *vector field* V on $\mathbb{R}^2$ is a function that assigns a tangent vector $V(\mathbf{p})$ to $\mathbb{R}^2$ at each point $\mathbf{p}$. Tangent vectors are naturally used to describe directions in vector spaces, and here it is first used in the definition of directional derivatives:

Definition 1 *let f be a differentiable real-valued function on $\mathbb{R}^2$, and let $\mathbf{v_p}$ be a tangent vector to $\mathbb{R}^2$. The derivative of f with respect to $\mathbf{v_p}$ is the number*

$$\mathbf{v_p}[f] = \frac{d}{dt}\left(f(\mathbf{p} + t\mathbf{v})\right)|_{t=0} \ .$$

If $\mathbf{v}$ is given in the usual Cartesian coordinate system as $\mathbf{v} = (v_1, v_2)$, it follows from the chain rule that $\mathbf{v_p}[f] = \nabla f(\mathbf{p}) \cdot \mathbf{v}$. Although this quantity clearly is scalable by the size of $\mathbf{v}$, we shall nevertheless refer to it as a *directional derivative*. In elementary calculus this problem is avoided by restricting $\mathbf{v}$ to unit vectors, a property our visual flows possess by definition.

Since vector fields are fields of tangent vectors, it is natural to extend the notion of directional derivative to vector fields. This yields the notion of the *operation of a vector field* V *on a function* f:

Definition 2 *Given a vector field* V *and a function* f*, both defined on* $\mathbb{R}^2$*, the operation of* V *on* f *is the real valued scalar function* $V[f] : \mathbb{R}^2 \to \mathbb{R}$ *defined as* $V[f](\mathbf{p}) = \mathbf{v}_{\mathbf{p}}[f]$*.*

As can be appreciated from its definition, the directional derivative at a point $\mathbf{p}$ is a function that converts tangent vectors at $\mathbf{p}$ to real numbers. It is also not too difficult to realize that $\mathbf{v}_{\mathbf{p}}[f]$ is linear. Functions that share these two proprieties are called *1-forms*:

Definition 3 *A 1-form* ϕ *on* $\mathbb{R}^2$ *is a real valued function on the set of all tangent vectors to* $\mathbb{R}^2$ *such that* ϕ *is linear at each point, that is*

$$\phi(a\mathbf{v} + b\mathbf{w}) = a\phi(\mathbf{v}) + b\phi(\mathbf{w}) \qquad \forall a, b \in \mathbb{R} \quad \forall \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}(\mathbb{R}^2) \ \ .$$

In the same spirit, there is a natural way to evaluate a 1-form ϕ on a vector field V to obtain a real valued function $\phi(V)$: at each point $\mathbf{p}$, define $\phi(V)$ as $\phi(V(\mathbf{p}))$. This way we can think of 1-forms as machines that convert vector fields to real-valued functions. One particularly important and common 1-form is the *differential of* f

Definition 4 *If* f *is a differentiable real-valued function on* $\mathbb{R}^2$*, the differential* df *of* f *is the 1-form that satisfies*

$$df(\mathbf{v}_{\mathbf{p}}) = \mathbf{v}_{\mathbf{p}}[f] \qquad \forall \mathbf{v}_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^2) \ \ .$$

We now turn to extend the basic notion from real-valued functions to vector fields on $\mathbb{R}^2$, which leads to the definition of the *covariant derivative* of a vector field with respect to tangent vectors and vector fields.

Definition 5 *Let* W *be a vector field on* $\mathbb{R}^2$*, and let* $\mathbf{v}_{\mathbf{p}}$ *be a tangent vector to* $\mathbb{R}^2$ *at* $\mathbf{p}$*. The covariant derivative of* W *with respect to* $\mathbf{v}_{\mathbf{p}}$ *is the following tangent vector at point* $\mathbf{p}$*:*

$$\nabla_{\mathbf{v}} W = \frac{d}{dt} \left(W(\mathbf{p} + t\mathbf{v}) \right) |_{t=0} \quad .$$

Clearly, $\nabla_{\mathbf{v}} W$ measures the initial rate of change of $W(\mathbf{p})$ as $\mathbf{p}$ moves in the $\mathbf{v}$ direction. Unlike directional derivatives, though, which convert tangent vectors to real numbers, covariant derivatives convert tangent vectors to tangent vectors. In practice, to apply $\nabla_{\mathbf{v}}$ to a vector field W, all we need to do is to compute the directional derivative of its Euclidean coordinates. Consequently, it is straightforward to realize that covariant derivatives are linear both in $\mathbf{v}$ and W.

Finally, by taking covariant derivatives pointwise, we can further extend this notion to *covariant derivative of a vector field W with respect to a vector field V*. The result is the vector field $\nabla_V W$ defined as $\nabla_V W(\mathbf{p}) = \nabla_{V(\mathbf{p})} W$.

2.2.2 Frame fields and the connection equation

The key construction that will accompany our study of visual flows is the following:

Definition 6 *A* frame field *on $\mathbb{R}^n$ is a set $\{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \ldots, \hat{\mathbf{E}}_n\}$ of n vector fields such that at all points*

$$\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \qquad 1 \leq i, j, \leq n \ .$$

As before, we restrict our discussion to $\mathbb{R}^2$, or the image plane, where visual flows are defined, thus a frame field is a pair of orthonormal vector fields on $\mathbb{R}^2$. For now, we also assume that the frame is differentiable. Note that $\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j$ denotes a point-wise extension of the dot product operator to two *vector fields*, thus yielding a real-valued function defined on $\mathbb{R}^n$.

By construction, each frame $\{\hat{\mathbf{E}}_1(\mathbf{p}), \hat{\mathbf{E}}_2(\mathbf{p})\}$ in a frame field $\{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2\}$ on $\mathbb{R}^2$ is a basis for $T_{\mathbf{p}}(\mathbb{R}^2)$ (i.e., the tangent space at that point). This means that every tangent vector $\mathbf{v}_{\mathbf{p}}$ at $\mathbf{p}$ can be expressed as a linear combination of $\hat{\mathbf{E}}_1(\mathbf{p})$ and $\hat{\mathbf{E}}_2(\mathbf{p})$

$$\forall\, \mathbf{v}_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{R}^2), \;\; \exists\, \alpha_1, \alpha_2 \quad \Rightarrow \quad \mathbf{v} = \alpha_1\, \hat{\mathbf{E}}_1(\mathbf{p}) + \alpha_2\, \hat{\mathbf{E}}_1(\mathbf{p})$$

where, by orthogonal expansion, the coefficients α_i are simply

$$\alpha_i = \mathbf{v} \cdot \hat{\mathbf{E}}_i(\mathbf{p}) \quad 1 \leq i \leq 2 \ .$$

The essential idea in frame field theory is to express in this way not just arbitrary vectors, but the *covariant derivatives of the vectors that constitute the frame itself.* In other words, since the frame is a basis, it follows that, at any given point $\mathbf{p}$, there exist coefficients that satisfy the following set of equations:

$$\begin{aligned}\nabla_{\mathbf{v}}\hat{\mathbf{E}}_1(\mathbf{p}) &= w_{11}(\mathbf{v})\hat{\mathbf{E}}_1(\mathbf{p}) + w_{12}(\mathbf{v})\hat{\mathbf{E}}_2(\mathbf{p}) \\ \nabla_{\mathbf{v}}\hat{\mathbf{E}}_2(\mathbf{p}) &= w_{21}(\mathbf{v})\hat{\mathbf{E}}_1(\mathbf{p}) + w_{22}(\mathbf{v})\hat{\mathbf{E}}_2(\mathbf{p})\end{aligned} \tag{2.1}$$

where orthogonal expansion dictates that

$$w_{ij}(\mathbf{v}) = (\nabla_{\mathbf{v}}\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) \quad 1 \leq i, j \leq 2 \ . \tag{2.2}$$

Since the coefficients $w_{ij}(\mathbf{v})$ are defined for arbitrary tangent vectors, they are functions from tangent vectors to real numbers. In fact, they are much more than that:

Theorem 2 (e.g., in [O'Neill, 1966]) *Let* $\{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2\}$ *be a frame on* $\mathbb{R}^2$ *and for each tangent vector* $\mathbf{v}$ *at point* $\mathbf{p}$ *define* $w_{ij}(\mathbf{v})$ *as in Eq. 2.2, i.e.,*

$$w_{ij}(\mathbf{v}) = (\nabla_{\mathbf{v}}\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) \quad 1 \leq i, j \leq 2 \ .$$

Then each $w_{ij}(v)$ *is a 1-form and* $w_{ij} = -w_{ji}$.

Proof: Since by construction w_{ij} are real-valued functions on the set of tangent vectors, to prove they are 1-forms requires showing their linearity. Since covariant derivatives are linear (see above) we get

$$\begin{aligned} w_{ij}(a\mathbf{v} + b\mathbf{w}) &= (\nabla_{a\mathbf{v}+b\mathbf{w}}\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) \\ &= ((a\nabla_{\mathbf{v}}\hat{\mathbf{E}}_i + b\nabla_{\mathbf{w}}\hat{\mathbf{E}}_i) \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) \\ &= (a\nabla_{\mathbf{v}}\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) + (b\nabla_{\mathbf{w}}\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) \\ &= a w_{ij}(\mathbf{v}) + b w_{ij}(\mathbf{w}). \end{aligned}$$

To prove that $w_{ij} = -w_{ji}$ we need to show that $w_{ij}(\mathbf{v}) = -w_{ji}(\mathbf{v})$ for an arbitrary tangent vector $\mathbf{v}$. In fact, this result follows directly from the orthonormality property $\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j = \delta_{ij}$ of frame fields. Since this dot product is constant (either 0 or 1), its directional derivative with respect to $\mathbf{v}$ is zero. Thus

$$0 = \mathbf{v}[\hat{\mathbf{E}}_i \cdot \hat{\mathbf{E}}_j] = (\nabla_{\mathbf{v}}\hat{\mathbf{E}}_i \bullet \hat{\mathbf{E}}_j)(\mathbf{p}) + (\hat{\mathbf{E}}_i \bullet \nabla_{\mathbf{v}}\hat{\mathbf{E}}_j)(\mathbf{p}) = w_{ij}(\mathbf{v}) + w_{ji}(\mathbf{v})$$

where the last step follows directly from the definition of $w_{ij}(\mathbf{v})$ in Eq. 2.2. **Q.E.D.**

The 1-forms w_{ij} are called the *connection forms* of the frame field and their skew symmetric property simplifies the expression of the covariant derivatives of the frame (Eq. 2.1) to the following equation, which is known as the *connection equation*:

$$\begin{pmatrix} \nabla_{\mathbf{v}}\hat{\mathbf{E}}_1 \\ \nabla_{\mathbf{v}}\hat{\mathbf{E}}_2 \end{pmatrix} = \begin{bmatrix} 0 & w_{12}(\mathbf{v}) \\ -w_{12}(\mathbf{v}) & 0 \end{bmatrix} \begin{pmatrix} \hat{\mathbf{E}}_1 \\ \hat{\mathbf{E}}_2 \end{pmatrix}$$

2.3 Visual flow curvatures

Of all possible ways of attaching a frame field to a visual flow, one is particularly natural and involves one basis vector *tangent* to the flow ($\hat{\mathbf{E}}_T$) and another one *normal* to the flow ($\hat{\mathbf{E}}_N$). Attaching an appropriate frame to each point of the visual flow results in a moving frame. Following the discussion above, the covariant derivatives of this frame field are direction dependent (Figure 2.2) and are described fully by the connection equation

$$\begin{pmatrix} \nabla_{\mathbf{v}}\hat{\mathbf{E}}_T \\ \nabla_{\mathbf{v}}\hat{\mathbf{E}}_N \end{pmatrix} = \begin{bmatrix} 0 & w_{12}(\mathbf{v}) \\ -w_{12}(\mathbf{v}) & 0 \end{bmatrix} \begin{pmatrix} \hat{\mathbf{E}}_T \\ \hat{\mathbf{E}}_N \end{pmatrix}. \tag{2.3}$$

It is clear that knowledge of the 1-form w_{12} is all that is needed to characterize the local behavior of the visual flow through its covariant derivatives. However, as a 1-form, w_{12} is a linear function of the tangent vector $\mathbf{v}$, and thus its values for all possible tangent vectors can be expressed as a linear combination of its values for two basis vectors only

$$w_{12}(\mathbf{v}) = w_{12}(a\ \mathbf{v}_1 + b\ \mathbf{v}_2) = a\ w_{12}(\mathbf{v}_1) + b\ w_{12}(\mathbf{v}_2).$$

In the frame field approach taken throughout this thesis, it is natural to use the frame once again, this time as the basis in which other vectors, $\mathbf{v}$ included, are described. This leads to the following description of w_{12}

$$w_{12}(\mathbf{v}) = w_{12}(a\ \hat{\mathbf{E}}_T + b\ \hat{\mathbf{E}}_N) = a\ w_{12}(\hat{\mathbf{E}}_T) + b\ w_{12}(\hat{\mathbf{E}}_N) \tag{2.4}$$

which immediately suggests that local changes in visual flow orientation are governed by the two

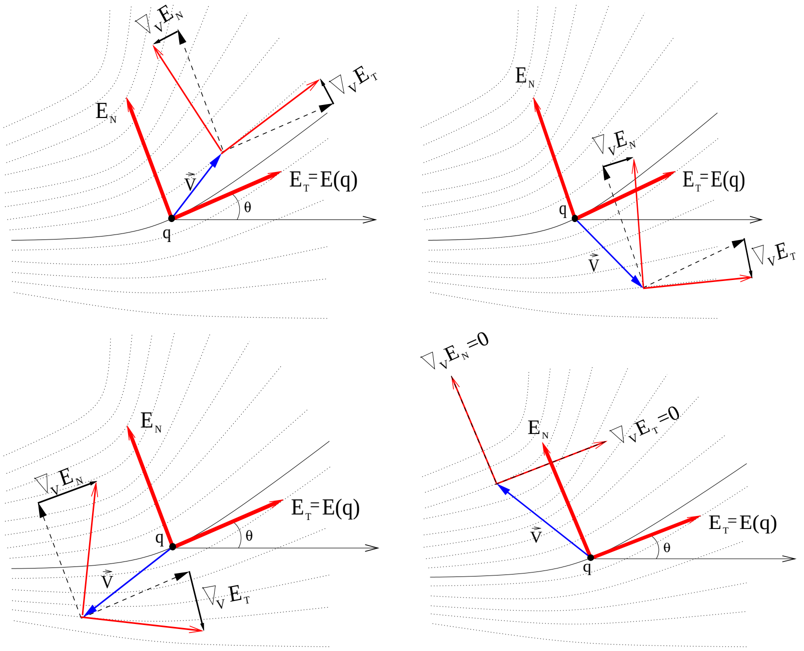


Figure 2.2: **The visual flow connection equation in action.** The changes in a tangential/normal frame that is attached to the visual flow depend on the direction of motion and their initial rate is determined by the connection equation of the underlying pattern. Here, for example, different motion directions (coded as blue vectors $\vec{\mathbf{V}}$) induce either a small counterclockwise rotation (top left), a small clockwise rotation (top right) a large clockwise rotation (bottom left), and no rotation at all (bottom right).

numbers $w_{12}(\hat{\mathbf{E}}_T)$ and $w_{12}(\hat{\mathbf{E}}_N)$, which we denote as

$$\begin{aligned} \kappa_T &\triangleq w_{12}(\hat{\mathbf{E}}_T) \\ \kappa_N &\triangleq w_{12}(\hat{\mathbf{E}}_N) \end{aligned} \tag{2.5}$$

and call the *tangential* and *normal* curvatures, respectively. Indeed, they represent the initial rate of change in the orientation of the flow in the two intrinsic directions – the tangential and normal – defined by the flow itself.

Of course, we can work out these curvatures in the other representations of visual flows as well (see Sec. 2.1). For example, by Definition 2, the tangential and normal curvatures are just the operation of the vector fields $\hat{\mathbf{E}}_T$ and $\hat{\mathbf{E}}_N$ on the scalar representation $\theta(x, y)$. Since these two vector fields are also expressed via $\theta(x,y)$

$$\begin{aligned}\hat{\mathbf{E}}_T &= (\cos\theta, \sin\theta)\\ \hat{\mathbf{E}}_N &= (-\sin\theta, \cos\theta)\ ,\end{aligned} \tag{2.6}$$

we thus get the expected projection relationship

$$\begin{aligned}\kappa_T &= \nabla\theta \cdot (\cos\theta, \sin\theta)\\ \kappa_N &= \nabla\theta \cdot (-\sin\theta, \cos\theta)\end{aligned} \tag{2.7}$$

which immediately implies that

$$||\nabla\theta|| = \sqrt{\kappa_T^2 + \kappa_N^2} \quad . \tag{2.8}$$

Another interesting insight is provided by evaluating the two visual flow curvatures in the unit length vector field representation $\hat{\mathbf{E}}_T(x,y)$, which gives:

$$\begin{aligned}\kappa_T &= ||\nabla \times \hat{\mathbf{E}}_T||\\ \kappa_N &= \ \nabla \cdot \hat{\mathbf{E}}_T \quad .\end{aligned} \tag{2.9}$$

The fact that the two curvatures are directly related to the curl and divergence of this field shows that, in general, it is neither irrotational nor incompressible (c.f. [Streeter and Wylie, 1981]).

It should be mentioned that while the connection equation 2.3 describes the local behavior of orientation for the general two dimensional case, it is equally useful for the one-dimensional case of curves. Now, only $\nabla_{\hat{\mathbf{E}}_T}$ is relevant (there is only one way to move along the one dimensional curve, and that is the tangential direction) so Eq. 2.3 simplifies to

$$\begin{pmatrix} \nabla_{\hat{\mathbf{E}}_T}\hat{\mathbf{E}}_T \\ \nabla_{\hat{\mathbf{E}}_T}\hat{\mathbf{E}}_N \end{pmatrix} = \begin{bmatrix} 0 & w_{12}(\hat{\mathbf{E}}_T) \\ -w_{12}(\hat{\mathbf{E}}_T) & 0 \end{bmatrix} \begin{pmatrix} \hat{\mathbf{E}}_T \\ \hat{\mathbf{E}}_N \end{pmatrix} . \tag{2.10}$$

In its more familiar form, where T,N, and κ replace $\hat{\mathbf{E}}_T$,$\hat{\mathbf{E}}_N$, and κ_T, respectively, this is the classical *Frenet equation* [O'Neill, 1966; Do Carmo, 1976] (where primes denote derivatives by arclength):

$$\begin{pmatrix} T' \\ N' \end{pmatrix} = \begin{bmatrix} 0 & \kappa \\ -\kappa & 0 \end{bmatrix} \begin{pmatrix} T \\ N \end{pmatrix} . \tag{2.11}$$

2.4 The curvature covariation constraint

Since the frame field point of view results in two curvatures at each point of the visual flow, across the structure we obtain two curvature *functions* – $\kappa_T(x,y)$ and $\kappa_N(x,y)$. Not unlike the Gaussian and mean curvatures of surfaces, these two curvature functions are not completely independent. In particular, curvatures of valid visual flows must satisfy the following constraint:

Theorem 3 (The curvature integrability constraint) *Given any texture flow* $\{\hat{\mathbf{E}}_T, \hat{\mathbf{E}}_N\}$*, its curvature functions* $\kappa_T(x,y)$ *and* $\kappa_N(x,y)$ *must satisfy*

$$\nabla\kappa_T \cdot \hat{\mathbf{E}}_N - \nabla\kappa_N \cdot \hat{\mathbf{E}}_T = \kappa_T^2 + \kappa_N^2 \quad .$$

Proof: Rewrite the system in Eq. 2.7 as follows

$$\begin{aligned} \theta_x &= \kappa_T \cos\theta - \kappa_N \sin\theta \\ \theta_y &= \kappa_T \sin\theta + \kappa_N \cos\theta \quad . \end{aligned} \tag{2.12}$$

Differentiate θ_x and θ_y by y and x, respectively, substitute Eq. 2.12 again to obtain

$$\begin{aligned} \theta_{xy} &= \kappa_{T_y} \cos\theta - \kappa_{N_y} \sin\theta - \theta_y^2 \\ \theta_{yx} &= \kappa_{T_x} \sin\theta + \kappa_{N_x} \cos\theta + \theta_x^2 \quad , \end{aligned}$$

and impose integrability, i.e. $\nabla \times \nabla\theta = \theta_{yx} - \theta_{xy} = 0$, to get

$$\nabla\kappa_T \cdot (-\sin\theta, \cos\theta) - \theta_y^2 = \nabla\kappa_N \cdot (\cos\theta, \sin\theta) + \theta_x^2 \quad .$$

Substituting both Eq. 2.8 and Eq. 2.6 completes the proof. **Q.E.D.**

The integrability constraint on visual flow curvatures suggests that unlike curvature (κ) and torsion (τ) for curves, not every combination of visual flow curvatures will integrate into a valid visual flow. One particular important consequence of this observation is the following:

Theorem 4 (Covariation of visual flow curvatures) *Unless* κ_T *and* κ_N *both equal zero, they cannot be constant simultaneously in a neighborhood around q, however small.*

Thus, in general, *at least one of the visual flow curvatures must vary, or the two curvatures need to covary, in any neighborhood of the texture flow*. Since κ_T is an "intra-streamline" property while κ_N is an "inter-streamline" property, Theorem 4 can be interpreted as a constraint, or perhaps a Gestalt quality [Wertheimer, 1955a], that the flow as a whole imposes on the collection of streamlines. We will encounter this constraint again later after formalizing visual flow good continuation.

2.5 Visual flow good continuation - The harmonic minimal models

Since a geometrical point of view reveals that the local behavior of visual flows is governed by two curvatures, it is natural to ask how we can use this insight to compute coherent visual flows from images. The amount of work in this area (see Sec. 1.3) suggests that this is a difficult problem of substantial interest to computer vision, and the depth of the gap between orientation local measurements and the desired (or, in the case of texture flows, the perceptual) outcome is illustrated in Figure 2.3.

Unlike existing approaches to this problem, ours is motivated by perceptual organization, and is based on asking "when are nearby visual flow measurements parts of a single coherent whole" (Figure 2.4). Following the discussion above, it becomes clear that visual flow measurements should involve position, orientation, and two curvatures, all of which must be part of the decision about geometrical coherence.

A more formal version of the problem is as follows: Given all of $\theta_0 = \theta(\mathbf{q})$, $K_T = \kappa_T(\mathbf{q})$, and $K_N = \kappa_N(\mathbf{q})$ at spatial position $\mathbf{q}$, find/design an orientation function $\theta_{\{\theta_0,K_T,K_N\}}(x,y)$ that reflects coherence, or good continuation, in the neighborhood of $\mathbf{q}$. Of course, while the local model $\theta_{\{\theta_0,K_T,K_N\}}$ must agree with the measurements *at* $\mathbf{q}$, the issue of how all these parameters should vary around $\mathbf{q}$ is in fact an underdetermined question.

Before studying this question, it is worth mentioning that the same approach, and similar questions, were already asked in the past in the domain of curves, not only in computational vision [Parent and Zucker, 1989; Yen and Finkel, 1998], but also in studies of visual integration in biological visual systems [Field et al., 1993b; Li, 1998b]. In this case, the Frenet equation [Do Carmo, 1976; O'Neill, 1966] suggests that local measurements of curves must involves the local tangent (i.e., ori-

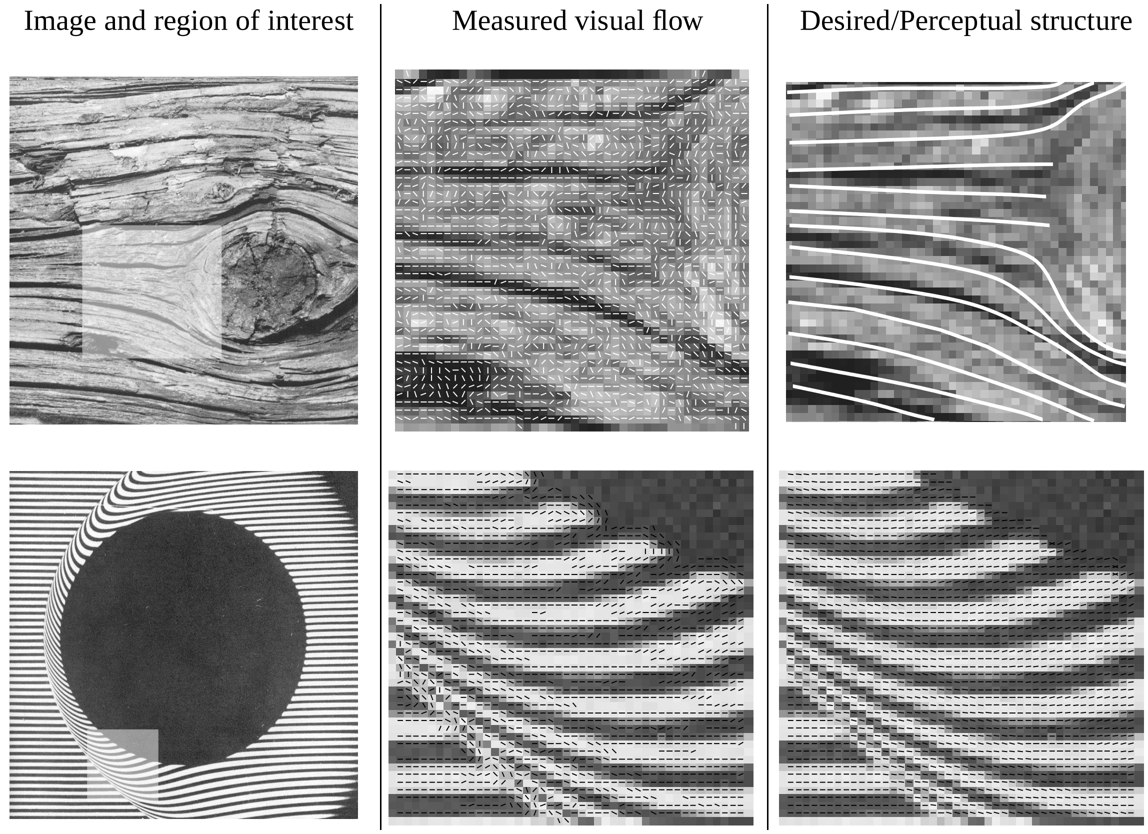


Figure 2.3: **The computational gap in visual flow measurements.** Demonstrated here for texture flow, the set of local orientation measurements is typically noise, incoherent, and often times partial. Similar problems occur in other visual flows such as shading, color, and motion. The desired outcome for the tree stump was drawn manually by a human observer (note the lack of structure in the knot area); the one for the fluid flow [Van Dyke, 1982] was produced by the computational system described in this thesis.

entation) and the curve's curvature. Thus, a local model for good continuation must depend upon both of these parameters. A common local model in this case is *cocircularity* [Parent and Zucker, 1989] and the *osculating circle* – the unique circle of a given tangent and curvature at a point (Figure 2.5).

The osculating circle is not the only possible local model that agrees with a given tangent and curvature at a point. In fact, the number of possible models is infinitely large[1], and each imposes a slightly different behavior around point $\mathbf{q}$ on the parameters involved. The osculating circle, in par-

---

[1] For example, a second order Taylor expansion of the parametric form of the curve in the Frenet frame gives a parabolic local model, rather than a circle.

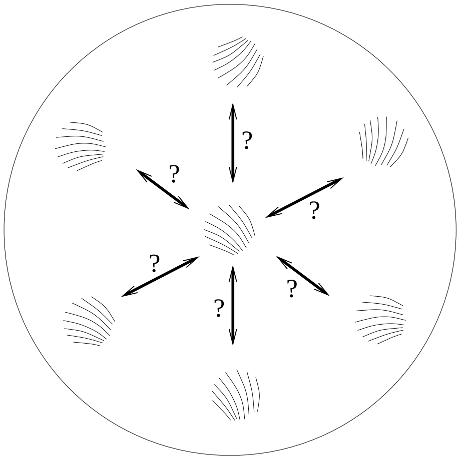


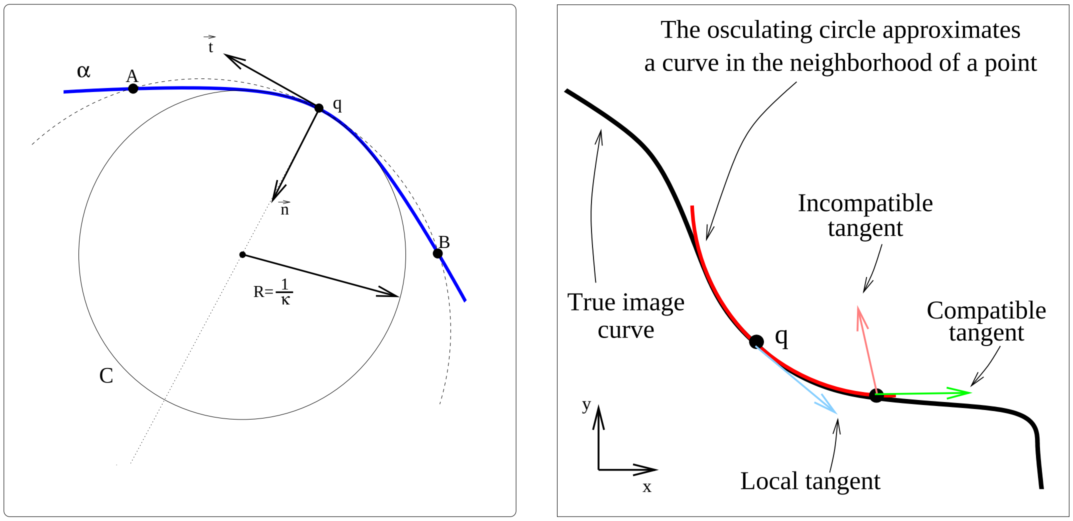
Figure 2.4: **Visual flow computation via contextual examination.** Given the orientation and curvatures of a flow patch, what is a desired orientation and curvatures of nearby coherent patches? What is an orientation local model that reflects good continuation? Answering such questions provides the ability to assess the degree to which a particular measurement is compatible, or consistent, with the context in which it is embedded, and whether or not that context is part of a single coherent structure.

ticular, imposes constancy of curvature; indeed, its curvature remains constant in the neighborhood of **q**, perhaps the simplest such constraint in our context. The osculating circle is also closely related to both old [Barrow and Tenenbaum, 1981, page 88] and recent [Kimia et al., 1999] computational notions of good continuation. What we would like to do is to construct an osculating circle-like object for visual flows.

We first observe that developing an osculating object for visual flows is slightly more challenging than one for curves. Not only that we have to deal with *two* curvatures, κ_T, and κ_N, but unlike the osculating circle, we cannot impose constancy on these curvatures (Theorem 4). Thus, visual flow good continuation must emerge as a different kind of invariance on the curvatures. A natural approach toward this goal is to formalize the intuition that "good" flows change slowly[2]. Doing so suggests employing some measure of variation in the neighborhood of the point **q**, but this, in turn, depends on the choice of representation. On one hand, using the scalar function representation

---

[2] Slow changes reflect the understanding that noise induces rapid variations. Note that a similar requirement was already part of our original informal categorization of these structures (Sec. 1.1).



a **b**

Figure 2.5: **The osculating circle for good continuation of curves. (a)** Given the tangent and curvature of a curve at a point **q**, there is a unique circle osculating point **q** with the same curvature and orientation. Locally, this circle behaves very similar to the curve, and thus it can be used as a model for how the curve extends beyond point **q**. **(b)** Decision about whether or not nearby curve measurements are part of the same whole can be established by comparing them to predictions made by the osculating circle. In this case the pink tangent is incoherent to the blue one because it is significantly different than predictions made by the latter's osculating circle. For opposite reasons, the green and blue tangents should be considered mutually coherent (although strictly speaking, the decision depends also on the curvature associated with the green tangent).

$\theta(x, y)$ suggests looking for the critical points of the harmonic energy

$$\int\int ||\nabla\theta||^2 \, dxdy \tag{2.13}$$

in the spirit of Tang *et al.* [Tang et al., 2000]. On the other hand, viewing the visual flow as a surface suggests minimizing the surface tension and looking for critical points of the area functional

$$\int\int \sqrt{1+\theta_x^2+\theta_y^2} \quad dxdy. \tag{2.14}$$

Unfortunately, either choice affects the minimization, and thus the possible solutions, while neither choice limits the number of solutions in any practical sense given the initial data we can use (i.e., orientation and curvature at a *single* point).

Avoiding this choice, however, and requiring that our solution satisfies *both* constraints, results in a surprising conclusion:

Theorem 5 (The harmonic minimal models) *Assume (without loss of generality) that* $\mathbf{q} = (0,0)$ *and* $\theta(0,0) = 0$*. A function* $\theta(x,y)$ *which satisfies* $\kappa_T(\mathbf{q}) = K_T$ *and* $\kappa_N(\mathbf{q}) = K_N$ *and is a critical point of both functionals (2.14) and (2.13) is either*

$$\begin{array}{lrcl} \textit{a plane} & \theta(x,y) & = & K_T x + K_N y, \\ \textit{a right Helicoid} & \theta(x,y) & = & \tan^{-1}(\frac{K_T x + K_N y}{1 + K_N x - K_T y}), \\ \textit{or a left Helicoid} & \theta(x,y) & = & \tan^{-1}(\frac{K_T x + K_N y}{1 - K_N x + K_T y}). \end{array}$$

Proof: The critical points of functional (2.14), functions whose Laplacian vanishes, are known as *harmonic functions* [Axler et al., 1992]. The critical points of functional (2.13), surfaces whose mean curvature vanishes, are known as *minimal surfaces* [Nitsche, 1989; Do Carmo, 1976]. It was in the early part of this century that Hamel [Hamel, 1923] (see also in Nitsche [1989]) proved that besides the plane, the only real harmonic minimal surfaces in Euclidean space of three dimensions are the helicoids, which can be represented in a suitable coordinate system by $\theta(x,y) = \tan^{-1}\left(\frac{y}{x}\right) + C$. This key result in the theory of minimal surfaces was later extended by Graustein [Graustein, 1940] to include imaginary solutions as well.

With the application of Hamel's theorem it remains to include the fact that the surface must be a graph of a 2D function over the image plane up to 2π-periodicity in the θ direction. In addition, the surface solution should induce a given pair of curvatures $(\kappa_T(\mathbf{q}), \kappa_N(\mathbf{q})) = (K_T, K_N)$ at $\mathbf{q}$. Without loss of generality, set $\mathbf{q}$ as the origin, i.e., $\mathbf{q} = (0,0)$, and let $\theta(0,0) = 0$. This aligns the frame $\{\hat{\mathbf{E}}_T, \hat{\mathbf{E}}_N\}$ at the origin in the direction of the global coordinate system and results, according to Eq. 2.7, in

$$\nabla\theta(0,0) = (K_T, K_N) \ . \tag{2.15}$$

Consider a helicoid in $XY\theta$ (the planar case is trivial). It is clear that in order for it to be be a graph of a 2D function (up to 2π-periodicity) its *directrix* must be perpendicular to the XY plane and its pitch must be exactly $\pm 2\pi$. All these helicoids are equal up to some translation $(\bar{x}, \bar{y}, \bar{\theta})$ in

$XY\theta$ and thus can be expressed in Cartesian coordinates by

$$s(x,y) = (x, y, \theta(x,y)) = (x, y, \tan^{-1}\left(\pm\frac{y-\bar{y}}{x-\bar{x}}\right) + \bar{\theta}). \qquad (2.16)$$

The given data $\theta(0,0) = 0$ and Eq. 2.15 are now used to determine the vector $(\bar{x}, \bar{y}, \bar{\theta})$. From Eqs. 2.16 and 2.15 we get

$$\begin{array}{llll} \theta_x = \pm\frac{\bar{y}-y}{(\bar{y}-y)^2+(\bar{x}-x)^2} & \Rightarrow & \theta_x(0,0) = \pm\frac{\bar{y}}{\bar{y}^2+\bar{x}^2} = K_T \\ \theta_y = \mp\frac{\bar{x}-x}{(\bar{y}-y)^2+(\bar{x}-x)^2} & \Rightarrow & \theta_y(0,0) = \mp\frac{\bar{x}}{\bar{y}^2+\bar{x}^2} = K_N \end{array}$$

which together with $\theta(0,0) = 0$ yields

$$(\bar{x}, \bar{y}, \bar{\theta}) = \left(\mp\frac{K_N}{K_T^2+K_N^2}, \pm\frac{K_T}{K_T^2+K_N^2}, \pm\tan^{-1}\left(\frac{K_T}{K_N}\right)\right).$$

Substituting these two translation vectors into expression 2.16 yields the following two solutions

$$\theta(x,y) = \pm\tan^{-1}\left(\frac{K_T}{K_N}\right) \pm \tan^{-1}\left(\frac{\mp K_T+(K_T^2+K_N^2)y}{\pm K_N+(K_T^2+K_N^2)x}\right).$$

Applying the following trigonometric identity

$$\tan^{-1}(A) + \tan^{-1}(B) = \tan^{-1}\left(\frac{A+B}{1-A\cdot B}\right)$$

results in the two helicoidal functions in the theorem and completes the proof. **Q.E.D.**

The three harmonic minimal models are depicted in Figure 2.6. Since they are all osculating objects, i.e., agree with a given orientation and curvatures at a point, in the small they are very similar. In fact, the two helicoids even have identical local shape at **q** (i.e., the principle curvatures and directions of their surface representation coincide at **q**). However, differences do exist between these models, especially in the behavior they impose on the curvature functions κ_T and κ_N in the neighborhood of **q**. Naturally, in neighborhoods larger than the most immediate, these curvature differences result in orientation differences as well. It is the nature of all these differences that eventually signals out one of the harmonic minimal models as a better one in representing good continuation of visual flows.

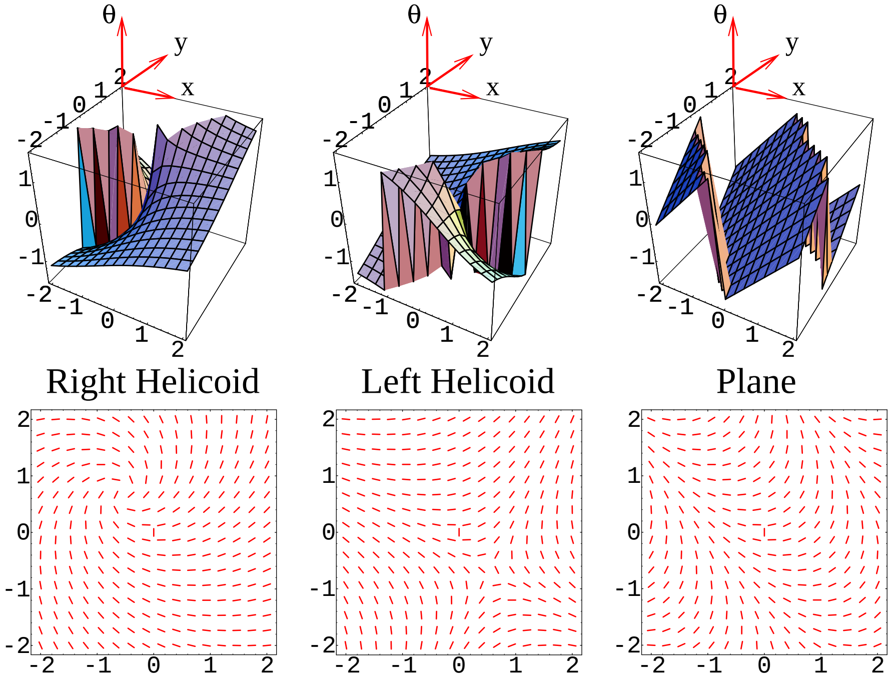


Figure 2.6: **The harmonic minimal models for coherence visual flows**. These are the only orientation functions that satisfy both the initial condition at $\mathbf{q}$ and are simultaneous critical points of energy functional (2.14) and (2.13). We illustrate these three solutions for $K_T = \kappa_T(\mathbf{q}) = 0.9$ and $K_N = \kappa_N(\mathbf{q}) = 0.7$ both as surfaces in $XY\theta$ and as flows in the image plane. The point $\mathbf{q}$ is marked with the short vertical marker. To set a visual context, we plot these solutions in a large neighborhood to show the singular point of the helicoids and the periodicity over the range $(-\frac{\pi}{2}, \frac{\pi}{2}]$.

2.6 Curvature covariation revisited and the right helicoidal model

The discussion so far suggests several main observations with regard to the modeling and computation of coherent texture flow: that (1) texture flow measurements should include two curvature measurements (in addition to position and orientation), that (2) nearby measurements should be considered coherent if each adequately agrees with the corresponding predictions made by the other's osculating object, and that (3) natural measures of variations lead to exactly three candidate osculating objects – the harmonic minimal models.

Additionally, the review of the osculating circle emphasized the fact that different local models for the good continuation of curves can be compared through the behavior they impose on the curvature function. Could such a comparison of the harmonic minimal models differentiate between the three to identify one of them as more appropriate? The (surprising) answer is given in the following theorem:

Theorem 6 (Identical curvature covariation of the right helicoidal model) *Of the three harmonic minimal models (Theorem 5), and of all orientation functions in general, the right helicoid is the only one with identical covariation of κ_T and κ_N in the neighborhood of $\mathbf{q}$:*

$$\frac{\kappa_T(x,y)}{\kappa_N(x,y)} = \mathit{const} = \frac{\kappa_T(\mathbf{q})}{\kappa_N(\mathbf{q})} = \frac{K_T}{K_N} \qquad \forall (x,y) \in N(\mathbf{q}). \tag{2.17}$$

Proof: It is immediate to verify that the right helicoid has a constant ratio of curvatures by substituting its orientation function into Eq. 2.7. The fact that this behavior is unique to the right helicoid can be derived by substituting Eq. 2.7 into Eq. 2.17 and reducing the system to a single quasi-linear differential equation [John, 1982]

$$\kappa_N(\mathbf{q})(\theta_x \cos\theta + \theta_y \sin\theta) - \kappa_T(\mathbf{q})(-\theta_x \sin\theta + \theta_y \cos\theta) = 0 \tag{2.18}$$

which can be rewritten as

$$\vec{\mathbf{n}} \cdot (\kappa_N(\mathbf{q})\cos\theta + \kappa_T(\mathbf{q})\sin\theta, \kappa_N(\mathbf{q})\sin\theta - \kappa_T(\mathbf{q})\cos\theta, 0) = 0 \tag{2.19}$$

where $\vec{\mathbf{n}}$ denotes a normal to the solution surface in $XY\theta$. The characteristic curves of this PDE in $XY\theta$ depend only on θ and are everywhere straight lines parallel to XY plane. Since the angle $\alpha(\theta)$ that these lines have with respect to the X direction grows linearly with θ

$$\alpha(\theta) = \theta - \tan^{-1}\left(\frac{\kappa_T(\mathbf{q})}{\kappa_N(\mathbf{q})}\right),$$

they constitute a ruling that defines a right helicoid.

For the sake of completion, it should be mentioned that Eq. 2.17 is feasible only when $K_N \neq 0$. If K_N vanishes, the appropriate ratio to consider is $\kappa_N(x,y)/\kappa_T(x,y)$, and the same proof carries through. **Q.E.D.**

Theorems 3 and 4 show that any (but constant) visual flow must have some covariation in its curvature functions. The right helicoid is the only local model that imposes identical covariation on κ_T and κ_N, thus putting them on equal ground and guarantees that neither dominates the other. In the next section we derive several other properties of the right helicoid, all of which reinforce its selection as the osculation object for the modeling of visual flow good continuation.

2.7 Properties and invariances of the right helicoidal model

Following the discussion in Sec. 2.1, a flow-centered right helicoid (i.e., using the frame $\{\hat{\mathbf{E}}_T, \hat{\mathbf{E}}_N\}$ at $\mathbf{q}$ as the coordinate system) can be represented either as a scalar orientation function

$$\theta(x,y) = \tan^{-1}\left(\frac{K_T x + K_N y}{1 + K_N x - K_T y}\right) \quad , \tag{2.20}$$

a surface in $XY\theta$

$$S(x,y) = \left(x, y, \tan^{-1}\left(\frac{K_T x + K_N y}{1 + K_N x - K_T y}\right)\right), \tag{2.21}$$

or a unit length vector field

$$\hat{\mathbf{E}}_T(x,y) = \frac{(1 + K_N x - K_T y, K_T x + K_N y)}{\sqrt{(1 + K_N x - K_T y)^2 + (K_T x + K_N y)^2}} \quad . \tag{2.22}$$

To obtain the expression of its curvature functions around the origin we simply substitute expression 2.20 into Eq. 2.7, or expression 2.22 into Eq. 2.9:

$$\begin{aligned} \kappa_T(x,y) &= \frac{K_T}{\sqrt{(1+K_N x - K_T y)^2 + (K_T x + K_N y)^2}} \\ \kappa_N(x,y) &= \frac{K_N}{\sqrt{(1+K_N x - K_T y)^2 + (K_T x + K_N y)^2}} \quad . \end{aligned} \tag{2.23}$$

Clearly, centering and aligning the model with the local coordinate frame does not affect its generality since we can adopt such an object-centered view for each point on the structure. However, it may be useful to extend the description to a fixed coordinate system in which the osculating object describes the consistent structure around an arbitrary point $\mathbf{q}_0 = (x_0, y_0)$, for arbitrary orientation $\theta(\mathbf{q}) = \theta_0$, and of course, for two given curvatures K_{T0} and K_{N0}. In general, if $\Theta_{K_{T0},K_{N0}}(\mathbf{q})$ is the flow-centered model, its general form should be an appropriate Euclidean transformation of this function, i.e.,

$$\Theta_{\mathbf{q}_0,\theta_0,K_{T0},K_{N0}}(\mathbf{p}) = \theta_0 + \Theta_{K_{T0},K_{N0}}(R_{\theta_0}(\mathbf{p} - \mathbf{q}_0)) \tag{2.24}$$

where $\mathbf{p}$ is the position vector around the point of interest $\mathbf{q}_0$ and R_{θ_0} is the rotation matrix

$$R_{\theta_0} = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{pmatrix} .$$

In our case, substituting expression 2.20 into Eq. 2.24 yields the following explicit formula

$$\theta_{x_0,y_0,\theta_0,K_{T0},K_{N0}}(x,y) = \tag{2.25}$$

$$\theta_0 + \tan^{-1}\left(\frac{(K_{T0}(x-x_0)+K_{N0}(y-y_0))\cos\theta_0 + (-K_{N0}(x-x_0)+K_{T0}(y-y_0))\sin\theta_0}{1+(K_{N0}(x-x_0)-K_{T0}(y-y_0))\cos\theta_0 + (K_{T0}(x-x_0)+K_{N0}(y-y_0))\sin\theta_0}\right)$$

which models a helicoidal flow around an arbitrary visual flow measurement in the the image plane.

As a surface, the helicoid has been studied by differential geometers for centuries. Together with the catenoid, the helicoid was the first of the (non trivial) minimal surfaces to be found (by Meusnier,1776), the following of which was discovered no less than 60 years later (by Scherk, 1835). The properties of minimal surfaces have been studied extensively since then (c.f. [Nitsche, 1989; Dierkes et al., 1992]), but to our best knowledge, this is the first time that helicoids have been studied as a representation of 2D oriented patterns. As we now discuss, this perspective results in a wealth of new properties, all of which naturally support the notion of good continuation for visual flows. To recover these properties, we will also exploit the different representations mentioned in Sec. 2.1. Note that the degenerate helicoid of $K_T = K_N = 0$ is a constant flow for which all the properties hereafter follow trivially. The proofs, however, ignore this case.

Theorem 7 *The right helicoid, and more generally, the three harmonic minimal models, are the only functions (that satisfy the initial data and) whose p-Laplacian $\Delta_p(\theta)$ vanishes simultaneously for all p.*

Proof: The p-Laplacian is defined as the following differential operator

$$\Delta_p\theta(x,y) \triangleq \nabla\cdot(||\nabla\theta||^{p-2}\nabla\theta). \tag{2.26}$$

It is immediate to show that the right helicoid, and the other two harmonic minimal models, have a vanishing p-Laplacian. This only requires substituting their orientation functions from Theorem 5 to the expression above. We next need to show that the three harmonic solutions are the *only ones* having this property.

First note that

$$\begin{aligned} p = 2 \quad &\Rightarrow \quad \Delta_p \theta(x,y) = \Delta \theta(x,y) = \theta_{xx} + \theta_{yy} \\ p = 1 \quad &\Rightarrow \quad \Delta_p \theta(x,y) = \nabla \cdot (||\nabla\theta||^{-1} \nabla\theta) = (\theta_x^2 + \theta_y^2)^{-\frac{3}{2}} (\theta_{xx}\theta_y^2 - 2\theta_x\theta_y\theta_{xy} + \theta_{yy}\theta_x^2). \end{aligned}$$

Assuming non vanishing gradient, both operators are well defined, thus a function $\theta(x,y)$ that satisfies both $\Delta_1\theta(x,y) = 0$ and $\Delta_2\theta(x,y) = 0$ must satisfy the following two constraints simultaneously

$$\begin{gathered} \theta_{xx} + \theta_{yy} = 0 \\ \theta_{xx}\theta_y^2 - 2\theta_x\theta_y\theta_{xy} + \theta_{yy}\theta_x^2 = 0. \end{gathered}$$

Summing up both sides we get

$$\theta_{xx}(\theta_y^2 + 1) - 2\theta_x\theta_y\theta_{xy} + \theta_{yy}(\theta_x^2 + 1) \overset{\triangle}{=} H = 0$$

which requires the mean curvature of the surface representation in $XY\theta$ (Eq. 2.21) to vanish everywhere. In other words, a function whose p-Laplacian vanishes for all p must be a minimal surface. Naturally, it is time to apply Hamel's theorem [Hamel, 1923] once again to conclude that the only candidates for $\Delta_p = 0 \;\; \forall p$ can be either a plane or a helicoid, namely, one of our harmonic minimal solutions from Theorem 5. **Q.E.D.**

Theorem 7 becomes more than an intellectual curiosity once we realize that functions with vanishing p-Laplacian $\Delta_p(\theta)$ are critical points of what is sometimes called the p-harmonic energy. As with anisotropic diffusion flows that emerged in the image processing literature [Perona and Malik, 1990], the p-Laplacian (Eq. 2.26) can be viewed as a weighted diffusion process in which the conductance function is inhomogeneous and changes spatially with the image content. Varying p changes the sensitivity of the process to large scale edges in the data and thus may achieve both isotropic and anisotropic diffusions. Ideas in this spirit were indeed used in the orientation diffusion literature as well (e.g., see [Tang et al., 2000]) and Theorem 7 shows that the right helicoid (or the three harmonic minimal models in general) represents a stationary point of all these diffusion processes, regardless of the chosen metric p.

The property that most distinguishes the right helicoid from the other two harmonic minimal models is the identical covariation of its curvatures in the image plane. While the proof of the next property follows immediately from the differentiation of Eq. 2.17, it is worth making it explicit to emphasize that the identical covariation applies to much deeper levels:

Theorem 8 : *Let $\kappa_T(x,y)$ and $\kappa_N(x,y)$ be the flow curvature functions induced by the right helicoid through Eq. 2.7. Let D^{ij} be the differential operator $D^{ij} \triangleq \frac{\partial^i}{dx^i}\frac{\partial^j}{dy^j}$. Then*

$$\frac{D^{ij}\kappa_T(x,y)}{D^{ij}\kappa_N(x,y)} = \mathit{const} = \frac{K_T}{K_N} \qquad \forall i,j \geq 0 \ .$$

The covariation of curvatures induced by the right helicoid implies additional geometrical and object-centered invariances on the corresponding visual flow in the image plane. For example, everywhere on the flow the direction along which it changes the most is constant relative to the direction of the flow itself. Furthermore, at each point of a right helicoidal visual flow both curvature functions changes the most along the same direction. This two invariances are formally phrased as follows:

Theorem 9 *If $\theta(x,y)$ is a texture flow induced by the right helicoid, then*

$$\cos\alpha(x,y) = \frac{\nabla\theta}{||\nabla\theta||} \cdot \hat{\mathbf{E}}_T = \mathit{const}$$

where $\alpha(x,y)$ is the angle that $\nabla\theta$ makes with the flow's frame. Moreover, the gradient vectors of its curvature functions $\nabla\kappa_T$ and $\nabla\kappa_N$ are linearly dependent (i.e., parallel).

Proof: From Eqs. 2.7 and 2.8 we obtain

$$\cos\alpha(x,y) = \frac{\nabla\theta}{||\nabla\theta||} \cdot \hat{\mathbf{E}}_T = \frac{\kappa_T}{\sqrt{\kappa_T^2 + \kappa_N^2}} = \frac{\kappa_T}{\kappa_T\sqrt{1+const}} = const \ \ .$$

Differentiating Eq. 2.17 by x and by y yields

$$\frac{\partial\kappa_T/\partial x}{\partial\kappa_N/\partial x} = \frac{K_T}{K_N} = \frac{\partial\kappa_T/\partial y}{\partial\kappa_N/\partial y} \ \ ,$$

which proves the linear dependency of the gradient vectors. **Q.E.D.**

The property that uniquely characterizes the right helicoidal flow is the identical covariation of its curvature function (Theorem 6). In itself, however, this property does not reveal the nature of this covariation. The next property suggests that even in this respect, the right helicoid is a most suitable object to reflect good continuation of visual flow:

Theorem 10 *Given a right helicoidal visual flow* $\theta(x,y) = \tan^{-1}(\frac{K_T x + K_N y}{1 + K_N x - K_T y})$ *around point* $\mathbf{q}$*, its two curvature functions* κ_T *and* κ_N *are inversely decaying along integral lines. More specifically, if* s *is the arc length along the streamline through* $\mathbf{q}$*, then*

$$\begin{aligned} \kappa_T(s) &= \frac{K_T}{1+s\cdot K_N} \\ \kappa_N(s) &= \frac{K_N}{1+s\cdot K_N} \ . \end{aligned} \tag{2.27}$$

Proof: An analytical expression for the behavior of the flow curvatures along streamlines can be derived by first computing the arc-length parametrization of the streamline $\alpha(s) = (x(s), y(s))$, and then substituting it into the curvature functions (using Eq. 2.7, or directly through Eq. 2.23). To do so, one must solve the helicoidal flow PDE that is induced by its vector field representation in Eq. 2.22:

$$\begin{aligned} \frac{dx}{ds} &= \frac{1+K_N x - K_T y}{\sqrt{(1+K_N x - K_T y)^2 + (K_T x + K_N y)^2}} \\ \frac{dy}{ds} &= \frac{K_T x + K_N y}{\sqrt{(1+K_N x - K_T y)^2 + (K_T x + K_N y)^2}} \ . \end{aligned} \tag{2.28}$$

Since this equation is highly non linear, solving it might prove difficult. Fortunately, it is easy to construct an equivalent *linear* equation whose trajectories overlap those of Eq. 2.28

$$\begin{aligned} \frac{dx}{dt} &= 1 + K_N x - K_T y \\ \frac{dy}{dt} &= K_T x + K_N y \ . \end{aligned} \tag{2.29}$$

The analytical solution to Eq. 2.29 is given by

$$\begin{aligned} x(t) &= e^{tK_N}(c_1 \cos tK_T - c_2 \sin tK_T) - \frac{K_N}{K_T^2+K_N^2} \\ y(t) &= e^{tK_N}(c_1 \sin tK_T + c_2 \cos tK_T) + \frac{K_T}{K_T^2+K_N^2} \end{aligned} \tag{2.30}$$

where the constants, obtained by applying the the initial conditions, are:

$$\begin{aligned} c_1 &= \frac{K_N}{K_T^2+K_N^2} \\ c_2 &= \frac{-K_T}{K_T^2+K_N^2} \end{aligned} \tag{2.31}$$

This yields the following parametrization of the streamline through the origin

$$\begin{aligned} x(t) &= \tfrac{-K_N+e^{tK_N}(K_N\cos tK_T+K_T\sin tK_T)}{K_T^2+K_N^2} \\ y(t) &= \tfrac{K_T+e^{tK_N}(-K_T\cos tK_T+K_N\sin tK_T)}{K_T^2+K_N^2} \quad . \end{aligned} \tag{2.32}$$

Since the solution in Eq. 2.35 was obtained from the linear version of our helicoidal visual flow, it is not arc-length parametrized. To obtain the correct curvature functions, we first need to reparametrize the streamline by arc-length. By integrating the length of the tangent to the streamline we first obtain its arc-length function

$$s(t) = \frac{\sqrt{e^{2tK_N}} - 1}{K_N} \tag{2.33}$$

to which an inverse function can be found analytically

$$t(s) = \frac{1}{K_N}\log\sqrt{1+2sK_N+s^2K_N^2} \quad . \tag{2.34}$$

Substituting this function back to Eq. 2.35 provides the final, arc-length parametrization, of the streamline through the origin:

$$\begin{aligned} x(s) &= \tfrac{1-K_N+sK_N}{K_T^2+K_N^2}\left[K_T\sin(\tfrac{K_T}{K_N}\log(1+sK_N)) + K_N\cos(\tfrac{K_T}{K_N}\log(1+sK_N))\right] \\ y(s) &= \tfrac{1+K_T+sK_N}{K_T^2+K_N^2}\left[K_N\sin(\tfrac{K_T}{K_N}\log(1+sK_N)) - K_T\cos(\tfrac{K_T}{K_N}\log(1+sK_N))\right] \end{aligned} \tag{2.35}$$

Computing the tangent vector of this curve

$$t'(s) = (x'(s), y'(s)) = \left(\cos(\frac{K_T}{K_N}\log(1+sK_N)), \sin(\frac{K_T}{K_N}\log(1+sK_N))\right)$$

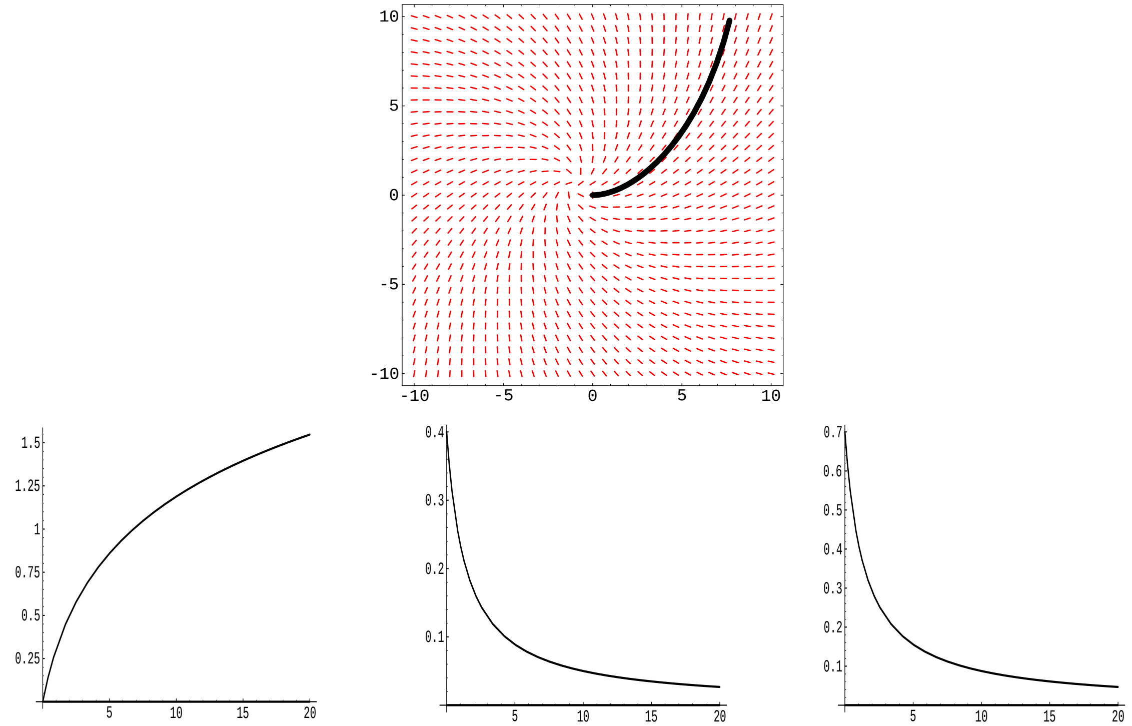
confirms that this is indeed an arc-length parametrization.

With the arc-length parametrization of the streamline now obtained, what is needed to complete the proof is its substitution in the orientation and curvature functions of the right helicoid. We get

$$\theta(s) = \theta(x(s), y(s)) = \frac{K_T}{K_N}\log(1+sK_N) \quad , \tag{2.36}$$

and

$$\begin{aligned} \kappa_T(s) &= \tfrac{K_T}{1+s\cdot K_N} \\ \kappa_N(s) &= \tfrac{K_N}{1+s\cdot K_N} \quad . \end{aligned} \tag{2.37}$$



Orientation along streamline

κ_T along streamline

κ_N along streamline

Figure 2.7: **Behavior of curvature along helicoidal flow streamline.** Illustrated here is helicoidal visual flow of $K_T = 0.4$ and $K_N = 0.7$ and it streamline through $\mathbf{q}$. The three graphs show the orientation (in radians), tangential curvature, and normal curvature functions along the streamline, all as a function of its archlength.

Example of the three functions for a streamline of one particular helicoidal visual flow (of $K_T = 0.4$ and $K_N = 0.7$) is illustrated in Fig. 2.7. **Q.E.D.**

The monotonically decaying curvatures along right helicoidal streamlines should not be taken lightly. In contrast, the two other harmonic minimal models not only have non-monotonically changing curvatures, they even change sign. Since both curvature extrema and inflection points have long been considered significant geometrical events for segmentation and part decomposition (e.g.,[Koenderink, 1984b; Richards et al., 1987; Hoffman, 1998]), it is inappropriate to include them in a model for coherence. Indeed, this does not happen with the right helicoid.

The advantage of having a model for coherent structure around an arbitrary visual flow measurement can be used to determine whether this measurement is consistent with its context, thus whether or not two nearby measurements are part of the same whole. While we develop this idea further in the chapters ahead, this contextual approach establishes an important necessary constraint on which local models can represent coherence. Informally, for two nearby measurements to be part of the same whole, they must be *mutually consistent* with each other. When consistency is based on a osculating object, this implies that the first measurement must agree with predictions made by the osculating object of the second, and vice versa (see Sec. 2.6, page 41). More formally, if $\theta_{x_0,y_0,\theta_0,K_{T0},K_{N0}}(x,y)$ predicts that a coherent visual flow at x_1, y_1 should have orientation θ_1 and curvatures K_{T1}, K_{N1}, then the osculating objects of this prediction, i.e., $\theta_{x_1,y_1,\theta_1,K_{T1},K_{N1}}(x,y)$, should in turn predict that a coherent visual flow at x_0, y_0 would have the original orientation and curvature values θ_0, K_{T0} and K_{N0}. The following definitions and theorem formalize this notion.

Definition 7 *Let* $\mathcal{F} = \mathbb{R}^2 \times \mathcal{S}^1 \times \mathbb{R} \times \mathbb{R}$ *be the space of all visual flow measurements, each of which is a 5-tuple* $F = \{x_0, y_0, \theta_0, K_{T0}, K_{N0}\}$. *Let* $\pi_\sigma(F)$ *be the projection operator that extracts component* σ *from* F, *with* $\sigma \in \{x, y, \theta, K_T, K_N\}$. *Let* $\Upsilon_F(x,y) : \mathbb{R}^2 \to \mathcal{F}$ *be an arbitrary osculating object for the measurement* F, *predicting coherent flow measurements of points* (x,y) *near* F. *We define the* binary relation $\overset{\Upsilon}{\approx}$ *between points in* $\mathcal{F}$ *as follows*

$$F_1 \overset{\Upsilon}{\approx} F_2 \qquad \Longleftrightarrow \qquad F_2 = \Upsilon_{F_1}(\pi_x(F_2), \pi_y(F_2))$$

i.e., $F_1 \overset{\Upsilon}{\approx} F_2$ *if* F_2 *is a prediction of the osculating object around* F_1.

Note that this definition uses Υ instead of Θ from Eq. 2.24 to emphasize that the model returns a 5-tuple as oppose to orientation only. In practice, however, the curvature are derived from the orientation function through Eq. 2.7.

The definition of the osculating object as a binary relation on all visual flow measurements is a technical step that aids in formalizing the constraint on valid osculating objects that we informally described above:

Definition 8 *Let* $\overset{\Upsilon}{\approx}$ *be the binary relation associated with the visual flow osculating object* $\Upsilon_F(x,y)$. $\Upsilon_F(x,y)$ *is said to be* admissible *if* $\overset{\Upsilon}{\approx}$ *is an equivalence relation [Simmons, 1963] on the space of all visual flow measurements, i.e., if it satisfies*

$$\begin{array}{lllll} \textit{reflexivity} & F_1 \overset{\Upsilon}{\approx} F_1 & & & \\ \textit{symmetry} & F_1 \overset{\Upsilon}{\approx} F_2 & \Longleftrightarrow & F_2 \overset{\Upsilon}{\approx} F_1 & \\ \textit{transitivity} & F_1 \overset{\Upsilon}{\approx} F_2 \ \textit{and} \ F_2 \overset{\Upsilon}{\approx} F_3 & \Longleftrightarrow & F_1 \overset{\Upsilon}{\approx} F_3 & . \end{array} \tag{2.38}$$

Admissibility of osculating objects captures the intuition that geometrical consistency of visual flow patches cannot be a "one way street". Osculating objects that do not have this property are ill-defined for our purposes and should be avoided. While the set of all admissible osculating objects contains other objects as well, it is nevertheless satisfying that we can claim the following:

Theorem 11 *The right helicoidal local model for visual flows is an admissible osculating object.*

Proof: Once a general expression for an osculating object is given in a fixed coordinate system (Eq. 2.25), its admissibility can be examined directly. Let $\mathbf{q} = (x_0, y_0)$ and $\mathbf{p} = (x_1, y_1)$ be two nearby points and assume (w.l.o.g) that $\mathbf{q} = (0,0)$, $\theta(\mathbf{q}) = 0$, $\kappa_T(\mathbf{q}) = K_{T0}$, and $\kappa_N(\mathbf{q}) = K_{N0}$, or put differently, that $F_1 = (0,0,0,K_{T0},K_{N0})$. Let $\theta_1(x,y)$ the right helicoid around $\mathbf{q}$, from which we can obtain the orientation and curvatures at $\mathbf{p}$ (Eqs. 2.20 and 2.23):

$$\begin{array}{lllll} \theta_1 & = & \theta(x_1,y_1) & = & \tan^{-1}\left(\frac{K_{T0}x + K_{N0}y}{1+K_{N0}x - K_{T0}y}\right) \\ K_{T1} & = & \kappa_T(x_1,y_1) & = & \frac{K_{T0}}{\sqrt{(1+K_{N0}x-K_{T0}y)^2+(K_{T0}x+K_{N0}y)^2}} \\ K_{N1} & = & \kappa_N(x_1,y_1) & = & \frac{K_{N0}}{\sqrt{(1+K_{N0}x-K_{T0}y)^2+(K_{T0}x+K_{N0}y)^2}} \ . \end{array}$$

Using $F_2 = (x_1, y_1, \theta_1, K_{T1}, K_{N1})$, we now construct the osculating object for this measurement and check for the three conditions in the definition (note that reflexivity is trivially true by construction). From Eq. 2.25, we get that the osculating object for F_2 is

$$\begin{array}{lll} \theta_2(x,y) & = & \tan^{-1}\left(\frac{K_{T1}x_1 + K_{N1}y_1}{1+K_{N1}x_1 - K_{T1}y_1}\right) - \\ & & \tan^{-1}\left(\frac{K_{T1}(x_1-x)+K_{T1}^2(xy_1-yx_1)+K_{N1}(y_1+xy_1K_{N1}-y(x_1K_{N1}+1)}{1+K_{N1}(x+x_1)-K_{T1}(y+y_1)+(K_{T1}^2+K_{N1}^2)(xx_1+yy_1)}\right) \end{array}$$

and we can substitute it into Eq 2.7 to extract its curvature functions as well, all of which are evaluated at $\mathbf{q}$ and compared to F_1. However, using the trigonometric identity

$$\tan^{-1}(A) - \tan^{-1}(B) = \tan^{-1}\left(\tfrac{A-B}{1+A\cdot B}\right)$$

we can simplify $\theta_2(x,y)$ to reveal that

$$\theta_2(x,y) = \tan^{-1}\left(\frac{K_T x + K_N y}{1 + K_N x - K_T y}\right) \equiv \theta_1(x,y)$$

i.e., that the osculating object for F_2 is *identical* to that for F_1. This immediately proves both symmetry and transitivity of the right helicoid and completes the proof. **Q.E.D.**

We complete our exploration for geometrical properties of the right helicoid with the observation that although different curvature estimates of visual flow at point **q** give rise to different helicoidal functions around **q** (Eq. 2.20), all these instances of the model are in fact the same object in $\mathbb{R}^2 \times \mathcal{S}^1$. The proof of this observation also demonstrates for the second time (see proof of Theorem 5) how the function in Eq. 2.20 is indeed a helicoid.

Theorem 12 *All values of K_T, K_N at point* **q** *induce the same osculating helicoid up to some* translation *in $XY\theta$.*

Proof: Apply the following substitution

$$x = v\cos(u) - \frac{K_N}{\xi^2} \qquad y = v\sin(u) + \frac{K_T}{\xi^2} \qquad \xi = \sqrt{K_T^2 + K_N^2}$$

into the surface representation of the right helicoid (Eq. 2.21). This yields the following alternative parametrization

$$s(u,v) = \left(v\ \cos(u) - \frac{K_N}{\xi^2}, v\ \sin(u) + \frac{K_T}{\xi^2}, \tan^{-1}\left(\frac{K_T\cos(u) + K_N\sin(u)}{-K_T\sin(u) + K_N\cos(u)}\right)\right) \quad .$$

We now seek to rephrase the third coordinate as a translation-based expression (the other two are already in the appropriate form), i.e., we seek to solve the equation

$$\tan^{-1}\left(\frac{K_T\cos(u) + K_N\sin(u)}{-K_T\sin(u) + K_N\cos(u)}\right) = u + \bar{\theta}(K_T, K_N) \quad . \tag{2.39}$$

In other words, we need to find $\bar{\theta}(K_T, K_N)$ that solves the equation

$$K_T\cos(u) + K_N\sin(u) - (K_N\cos(u) - K_T\sin(u))\tan(u + \bar{\theta}(K_T, K_N)) = 0 \quad . \tag{2.40}$$

This equation can be simplified to

$$\sec(u+\bar{\theta})(K_T\cos(\bar{\theta})-K_N\sin(\bar{\theta}))=0$$

whose solution is independent of u

$$\bar{\theta}(K_T,K_N)=\pm\cos^{-1}\left(\frac{K_N}{\sqrt{K_T^2+K_N^2}}\right) \quad . \tag{2.41}$$

Resolving when each of these two functions is an appropriate solution to Eqs. 2.39 and 2.40 yields the final familiar helicoidal form

$$s(u,v)=\left(v\ \cos(u)-\frac{K_N}{\xi^2},v\ \sin(u)+\frac{K_T}{\xi^2},u+\text{sign}(K_T)\ \cos^{-1}\left(\frac{K_N}{\xi}\right)\right)$$

and proves the theorem. **Q.E.D.**

Finally, it is worth mentioning that the osculating helicoid is closely related to the osculating circle - the local model for good continuation of curves that we mentioned in section 2.5 (see page 35). Two main reasons contribute to this relationship. Firstly, lifting the osculating circle to $XY\theta$ yields a helix, while helicoids, by construction, are collections of helices. Secondly, the helix is the only space curve whose ratio of “curvatures” (curvature and torsion) is constant [Forsyth, 1912]. Interestingly, in this chapter we have shown that a similar invariance exists for the helicoid through the curvatures of its induced visual flow. Implications of this intimate relation between the two osculating objects are further discussed in Chapter 7.

Chapter 3

Sensitivity to curvature in orientation-based texture segmentation

Texture segregation (by human observers) has long been attributed to changes in the distribution of elementary features across the visual field [Julesz, 1981, 1986]. The study of orientation, a conspicuous feature, has led to models of orientation-based texture segmentation that depend on the magnitude of one or two orientation gradients [Landy and Bergen, 1991; Nothdurft, 1991] with additional influence from the relative configuration between the orientation textons and the global texture boundary [Nothdurft, 1992; Wolfson and Landy, 1995]. However, the emergence of curvatures as intrinsic local descriptors of visual flows, and their role in determining coherent structure (Ch. 2), raises the question about their possible manifestation at the perceptual level, an avenue that has not been considered thus far. In this chapter, we pause from computational issues and conduct an exhaustive psychophysical study of the sensitivity to curvatures in orientation-based texture segmentation.

Before we begin, few words about terminology are in order. Texture flows, the visual structures whose "flow" is perceived directly, are commonly called in the psychophysical literature *orientation textures* or *orientation-defined textures* (ODTs). The partition of such stimuli to coherent perceptual parts is typically referred to as *orientation-based texture segmentation* (OBTS). Thus, these are the terms we will use in this chapter.

3.1 Orientation defined textures and the psychophysics of OBTS

The visual perception of texture plays a fundamental role in many aspects of vision from figure-ground segregation to 3D shape and depth perception. The ability to effortlessly segregate texture stimuli into coherent parts has long been attributed to the changes in the spatial distribution of elementary features, sometimes called textons [Julesz, 1981, 1986]. Of these features, one that has been studied extensively is *orientation*. While ODTs are frequent in natural and artificial visual stimuli (Figs. 1.1 and 1.2), textures are rarely characterized solely by orientation. Nevertheless, understanding the effect of orientation on texture segmentation is essential due to its neurophysiological basis [Hubel and Wiesel, 1977], its central role in perceptual organization and its close relationship to shape perception (see Sec. 1.2).

3.1.1 Basic observations on ODTs and OBTS

Orientation-defined textures in visual stimuli result from pattern formation processes that cover surfaces (and sometimes volumes) in the real world: fur might cover a bear, grass a field, or stripes a zebra. The visual appearance of these surfaces as ODTs is thus influenced by two basic factors: the behavior of the pattern formation processes and the interaction between surfaces during the imaging process.

It follows that discontinuities in ODTs, and thus perceptual edges and OBTS in general, can arise in two fundamental ways. The first is occlusion of one (textured) object by another; each textured surface projects to an ODT region that meets the other along the occlusion boundary. This rather common scenario raises the question of how the ODT varies in the neighborhood of the occlusion boundary. Computational studies show that there are two possible generic events, and that they can be classified as folds and cuts [Huggins and Zucker, 2001a], as is illustrated in Fig. 3.1.

The pattern formation processes that cover surfaces in the real world often enjoy smoothness properties. However, they need not be smooth everywhere. This observation leads us to the second cause of discontinuities – those that result from singularities in the pattern formation processes themselves. This issue is studied extensively in the formation of biological patterns and morphogenesis [Turing, 1952; Murray, 1989; Corbit and Garbary, 1995] and is well exemplified by the sin-

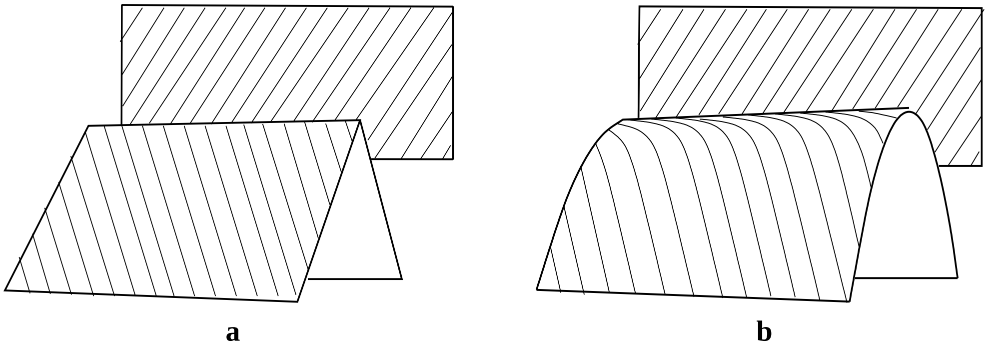


Figure 3.1: **Cuts and folds as generic events near occlusion boundary** (adapted from [Huggins and Zucker, 2001a]). **(a)** A generic 'cut' event implies that the ODT intersects the occlusion boundary transversely. **(b)** A generic 'fold' event implies that the ODT approaches a tangent configuration with the occlusion boundary. Note that this happens whether or not the texture on the surface has constant orientation.

gularities of the striped patterns that cover, for example, both zebras and many sea shells (Fig. 1.1). Importantly, it suggests that interesting (singularity) events can occur in ODTs that lead to a variety of configurations, none of which can be classified as occlusion boundaries.

Regardless of the process that creates ODT discontinuities, one universal property holds for all – natural ODTs are seldom constant in the neighborhood of discontinuities. In fact, every-day ODT stimuli are not likely to have constant orientation even *within* coherent regions; this requires an accidental match between the surface geometry, the texture formation process, and the observer's view-point. Furthermore, perspective projection dictates that even completely parallel lines in the world are likely to give rise to a non-constant retinal ODT.

Since ODTs are generically not constant, either within coherent regions or in the neighborhood of orientation edges, it is noteworthy that much of their psychophysical investigation in the last two decades has focused on stimuli of piecewise constant orientation [Nothdurft, 1985a; Landy and Bergen, 1991; Wolfson and Landy, 1995; Regan et al., 1996; Caputo, 1997; Kwan and Regan, 1998; Wolfson and Landy, 1998; Caputo and Casco, 1999; Li, 1998a; Motoyoshi and Nishida, 2001]. Part of our goal in this chapter is to bring this larger context to the fore of the psychophysical exploration of ODTs.

3.1.2 OBTS and orientation gradients

Although orientation and ODTs are geometrical objects, the study of OBTS has exploited very little of their intrinsic geometry. Filter-based approaches (e.g.,[Malik and Perona, 1990; Landy and Bergen, 1991]) consider the geometrical content only as a basis for computing scalar energies from which segmentation is derived through nonlinear transformation (typically, rectification) and detection of areas of high gradient. Feature-based models [Nothdurft, 1991, 1993; Mussap and Levi, 1999] suggest more generally that OBTS depends on the relationship between two orientation gradients (Fig. 3.2), namely – the change in orientation *between* coherent regions ($\Delta\theta_{between}$) and the change in orientation *within* regions ($\Delta\theta_{within}$). As expected, the former must be significantly larger than the latter for segmentation to occur.

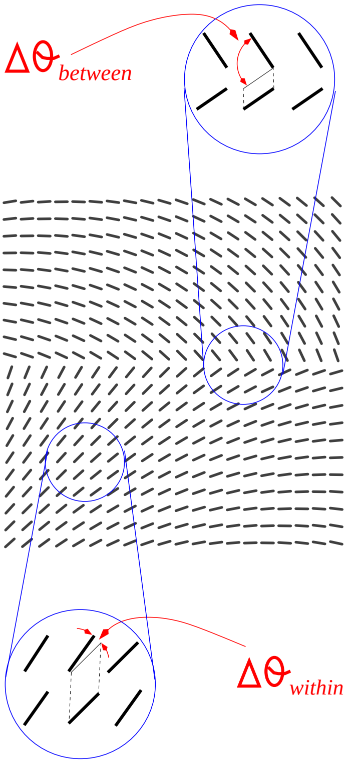


Figure 3.2: **OBTS and orientation gradients**. Existing models [Nothdurft, 1991; Landy and Bergen, 1991; Mussap and Levi, 1999] predict that OBTS depends on the relationship between two orientation gradients, one within and the other between perceptually coherent regions. These models predict that segmentation occurs reliably if and only if the ratio $\frac{\Delta\theta_{between}}{\Delta\theta_{within}}$ is significantly larger than 1.

Although the notion of orientation gradient (sometimes called orientation contrast) is invoked by most models for OBTS, the fact that gradients are *vector* quantities is typically overlooked. In particular, although orientation gradients, like any other gradient, have both magnitude and direction, their directional property is generally ignored. This may be justifiable in the case of $\Delta\theta_{between}$ since it describes the change of orientation across 1D perceptual edges. The same cannot be said about $\Delta\theta_{within}$, though, since it is supposed to characterize the behavior of orientation in 2D regions. Thus, whenever the vectorial nature of $\Delta\theta_{within}$ becomes important, we use the mathematical gradient symbol $\nabla\theta$ instead. In other cases we use the symbol $\Delta\theta_{within}$ but we also keep in mind that it is defined as the magnitude of $\nabla\theta$, i.e., as $\Delta\theta_{within} = ||\nabla\theta||$.

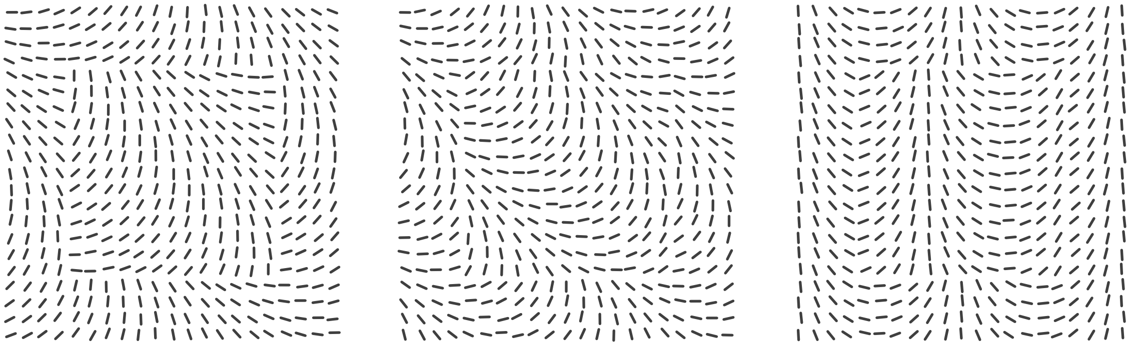


Figure 3.3: **Orientation gradients are not enough to predict OBTS. (a)** Inspired by Nothdurft [1991], this stimulus has a clear orientation-defined figure segmented away from its ground, both of which have the same orientation gradient $\nabla\theta$ within their interior. Here $\nabla\theta = \frac{(45°,45°)}{\sqrt{32}}$ and its magnitude is $\Delta\theta_{within} = 11.25°$. Consistent with existing models, the large orientation discontinuity of $\Delta\theta_{between} = 90°$ across the figure's boundary induces easy segmentation. **(b)** Still consistent with current models, the segmentation of the same figure becomes more difficult as the the ratio $\frac{\Delta\theta_{between}}{\Delta\theta_{within}}$ decreases. Here $\nabla\theta = \frac{(18°,18°)}{\sqrt{2}}$, $\Delta\theta_{within} = 18°$, and $\Delta\theta_{between} = 36°$. However, this case also reveals that the saliency of the square's boundary is not uniform (compare its lower left corner to the upper right or the upper left ones) and suggests segmentation may be influenced by more than the scalar values of $\Delta\theta_{within}$ and $\Delta\theta_{between}$. **(c)** The last hypothesis is further emphasized by rotating the direction of $\nabla\theta$. Now, the orientation gradient vector points horizontally, $\nabla\theta = (18°, 0°)$ but its magnitude is still $\Delta\theta_{within} = 18°$ and $\Delta\theta_{between}$ is still 36°. Nevertheless, the overall perceptual effect is very different. In particular, note how the saliency of the left and right boundaries has diminished almost completely despite the fact that the orientation discontinuity across them is identical to that across the top and bottom boundaries.

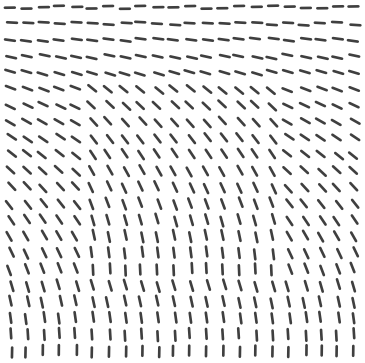


Figure 3.4: **Another example that orientation gradients are not enough to predict OBTS.** Here $\nabla\theta$ (and thus, $\Delta\theta_{within}$) is constant within the figure and within the ground, and $\Delta\theta_{between} = 18°$ is constant across the figure's edges (a square). Nevertheless, the saliency of the top edge is significantly higher than that of the bottom edge, which is hardly detectable without scrutiny. Note that since the orientation texels along one side of the bottom edge are exactly perpendicular to it, the fact that it is less salient relative to the top edge justifies a closer reexamination of previous predictions [Wolfson and Landy, 1995]. In particular, it raises the possibility that configural effects are themselves modulated by ODT variations.

This sharper semantics calls for investigating the relationship between $\nabla\theta$ (as a vector) and performance in OBTS. Indeed, contrary to predictions from existing models, changing only the *direction* of $\nabla\theta$ (without changing its magnitude, $\Delta\theta_{within}$), carries significant perceptual consequences (Fig. 3.3). In fact, we found that perceptual differences exist even for a fixed combination of $\nabla\theta$ and $\Delta\theta_{between}$ (Fig. 3.4). One must conclude that other factors also play a role in OBTS.

3.1.3 OBTS and ODT curvatures

To study of OBTS in the larger (and more realistic) context of spatially varying ODTs, this variability needs to be captured adequately. The geometrical analysis in Chapter 2 suggests that this should be done through ODT curvatures. Two main observations are important. First, examining OBTS from the point of view of ODT curvatures does not exclude existing findings or insight from classical models. This is because, as Eq. 2.8 indicates, the scalar measure of variability within coherent regions can be expressed directly in terms of curvatures

$$\Delta\theta_{within} = ||\nabla\theta|| = \sqrt{\kappa_T^2 + \kappa_n^2}\ . \tag{3.1}$$

Second, recall that Eq. 2.7 shows clearly that the *same* orientation gradient $\nabla\theta$ can give rise to *different* combinations of curvatures (e.g., by changing θ without changing $\nabla\theta$)

$$\begin{aligned} \kappa_T &= \nabla\theta \cdot (cos\theta, sin\theta) \\ \kappa_N &= \nabla\theta \cdot (-sin\theta, cos\theta) \ . \end{aligned} \tag{3.2}$$

The analogy of this observation to the perceptual evidence (Figs. 3.3 and 3.4) therefore raises the possibility that OBTS may relate to texture curvatures and their discontinuities. Our main goal in this chapter is to explore this possibility from a psychophysical point of view.

3.2 Experiment 1 – Sensitivity to curvature discontinuities in OBTS

To explore the role of curvature discontinuities in OBTS we conducted a two-alternative forced choice texture segmentation experiment with stimuli designed specifically to test segmentation performance as a function of $\nabla\theta$, $\Delta\theta_{between}$, and the discontinuities (or contrast) in the texture curvatures, $\Delta\kappa_T$ and $\Delta\kappa_N$. This section describes this experiment and its results.

3.2.1 Stimuli overview

All stimuli consisted of ODTs which were portrayed as arrays of 21x21 bright oriented segments on dark background[1]. Viewed by subjects from an approximate distance of 1 meter, all stimuli spanned 10 degrees of visual angle. Each ODT contained two lines of orientation discontinuity – one diagonal and one horizontal – which together defined a perceptual “figure”; either a left-pointing or right-pointing triangle (Fig. 3.5). The 25’ long oriented segments were initially positioned along a regular grid of 10x10 degrees and then randomly jittered up to 5’ in each direction to avoid grid artifacts.

Similar to previous experiments [Nothdurft, 1985b,a; Landy and Bergen, 1991; Nothdurft, 1991; Mussap and Levi, 1999; Wolfson and Landy, 1995], we designed the figure and ground to have a fixed orientation contrast (i.e., $\Delta\theta_{between}$ is constant along the figure’s boundaries in any given

---

[1] Although other methods to depict ODTs have also been used in the past [Landy and Bergen, 1991], the use of discrete oriented texels predominates the OBTS literature [Nothdurft, 1991, 1993, 1992; Mussap and Levi, 1999; Wolfson and Landy, 1995] and it allowed us a better comparison to previous studies with everywhere changing ODTs [Nothdurft, 1993, 1992; Mussap and Levi, 1999].

trial). Following past studies with textures of *varying* orientation [Nothdurft, 1991, 1992; Mussap and Levi, 1999], we also set the orientation gradient *within* the figure and the ground to be constant (i.e., $\nabla\theta_{figure} = \nabla\theta_{ground} = \nabla\theta =$*constant* in any given trial).

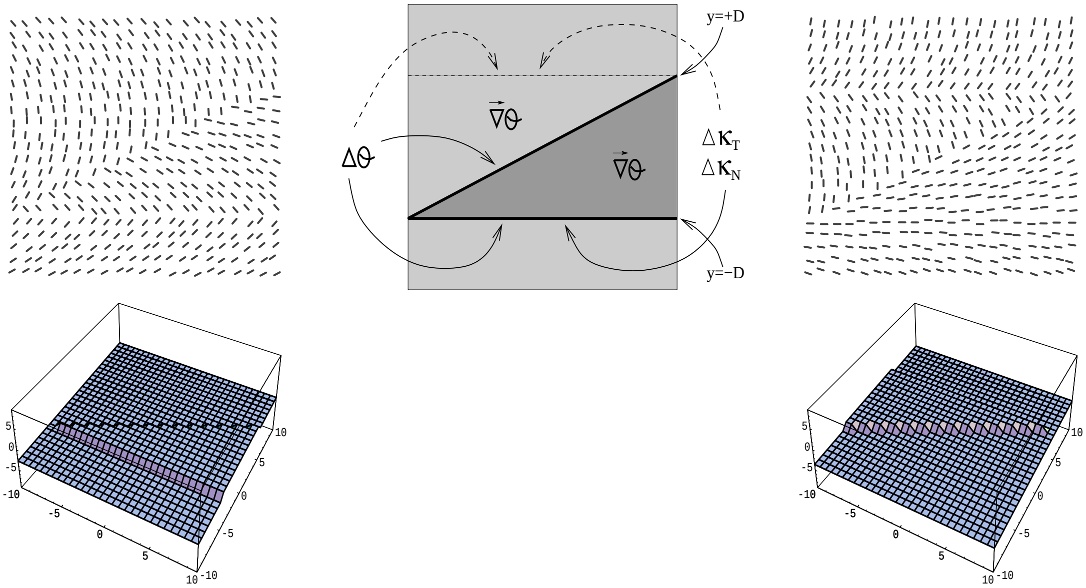


Figure 3.5: **Basic stimulus design for the OBTS experiment.** The subject's task was to identify one of two possible orientation-defined figures, either a left-pointing (left) or a right-pointing (right) triangle. The *same* diagonal discontinuity appeared in all stimuli and thus played no role in the subject's decision. Along the other edge (either at top or bottom), all of $\Delta\theta_{between}$, $\Delta\kappa_T$, and $\Delta\kappa_N$ were constant (within a trial). $\nabla\theta$, the orientation gradient, was also constant (within each trial) and identical for both the figure and the ground. Two examples are shown, both as ODTs and 3D surfaces in $XY\theta$. Both stimuli have $\nabla\theta = (0°, 5°)$ and $\Delta\theta_{between} = 75°$. To optimize printing, stimuli are shown in reversed contrast. Those used in the experiment had bright segments on dark background.

Unlike previous experiments, however, we designed our stimuli to also have constant *curvature* discontinuities. In other words, both $\Delta\kappa_T$ and $\Delta\kappa_N$ – the jump in tangential and normal curvatures, respectively – were constant (within a trial) between the figure and the ground along their defining horizontal edge (Fig. 3.5). Fortunately, the iso-curvature curves of ODTs with constant $\nabla\theta$ are straight lines, thus designing stimuli with constant curvature discontinuities is computationally feasible (see Sec. 3.2.2). To avoid compromising the explored psychophysical measures, and to allow for meaningful conclusions on configural effects (see Sec. 3.2.5, 3.3, and 3.4), all stimuli were designed to have an orientation gradient $\nabla\theta$ of a carefully selected (vertical) direction, which resulted in a one-to-one mapping between $\nabla\theta$ and the previously used $\Delta\theta_{within}$. Therefore, in the following we use them interchangeably.

For each combination of $\nabla\theta$ and $\Delta\theta_{between}$, three different combinations of $\Delta\kappa_T$ and $\Delta\kappa_N$ were tested. Note that once $\nabla\theta$, $\Delta\theta_{between}$, and *one* $\Delta\kappa$ is set, the other $\Delta\kappa$ is fully determined from Eq. 3.2 (see Sec 3.2.2). The chosen combinations of $\Delta\kappa_T$ and $\Delta\kappa_N$ covered three fundamental behaviors for the curvatures. One was defined by $\Delta\kappa_T = 0$, the second by $\Delta\kappa_N = 0$, and the third by $\Delta\kappa_T = \Delta\kappa_N$. In other words, the first class of orientation discontinuities had no discontinuity in tangential curvature, the second lacked discontinuity in normal curvature, and the last was characterized by (the same) discontinuity in both curvatures (Fig. 3.6). Note again that such triples of stimuli were created for *each* combination of $\nabla\theta$ and $\Delta\theta_{between}$, thus allowing us to investigate the effect of curvature discontinuities *independently* of the orientation gradients.

3.2.2 Stimuli details

Once the jittered position of the texels was set, their orientation was calculated from the underlying orientation function of the region they belonged to:

$$\begin{aligned} \theta_{ground}(x,y) &= \theta_0 + \Delta\theta_{within} \cdot y \\ \theta_{figure}(x,y) &= \theta_0 + \Delta\theta_{within} \cdot y + \Delta\theta_{between} \quad . \end{aligned} \tag{3.3}$$

Each stimulus had a fixed combination of $\Delta\theta_{within}$ and $\Delta\theta_{between}$ chosen from the following sets

$$\begin{aligned} \Delta\theta_{within} &\in \{5^\circ, 10^\circ, 15^\circ, 20^\circ, 25^\circ, 30^\circ\} \\ \Delta\theta_{between} &\in \{5^\circ, 10^\circ, 15^\circ, \ldots, 85^\circ, 90^\circ\} \quad . \end{aligned}$$

The free parameter θ_0 in Eq. 3.3 was set such that the corresponding curvature functions κ_T and κ_N achieved predefined discontinuities along a given horizontal line positioned at $y = \pm D$, where D corresponds to 2.5° (see Fig. 3.5). Since the two curvatures are coupled through Eq. 3.2, only one discontinuity is needed (say $\Delta\kappa_T$) to fully determine the other and thus fully determine θ_0. With the origin at the center of the stimulus, the solution to θ_0 takes the following form

$$\theta_0 = -\frac{1}{2}\left[\Delta\theta_{between} + 2\sec^{-1}\left(\frac{2\Delta\theta_{within}\sin\frac{\Delta\theta_{between}}{2}}{\Delta\kappa_T}\right) + 2D\Delta\theta_{within}\right] . \tag{3.4}$$

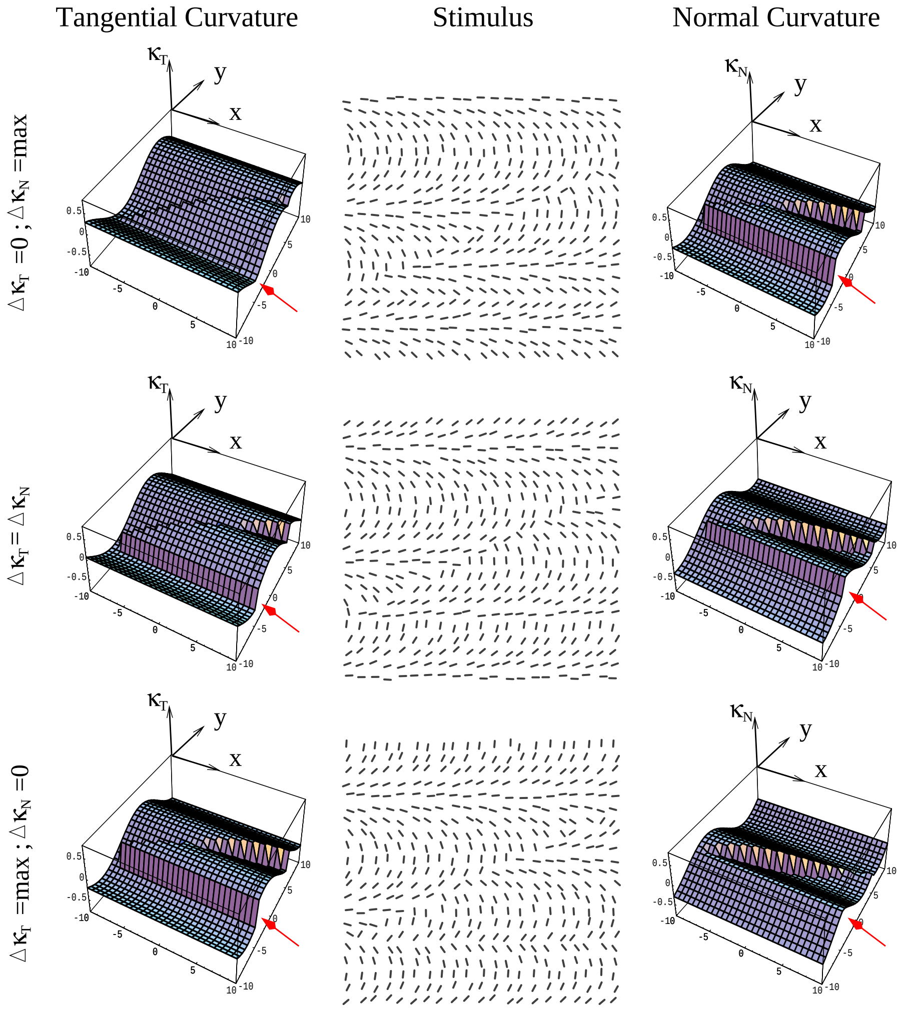


Figure 3.6: **The three cases of curvature discontinuities examined in Experiment 1.** For each combination of $\nabla\theta$ and $\Delta\theta_{between}$, three cases of curvature discontinuities were formed. Here we show the three cases for a left pointing triangle with $\Delta\theta_{within} = 20°$ and $\Delta\theta_{between} = 90°$. The 3D graphs show κ_T and κ_N as height functions over the image domain. The red arrows mark the discontinuities in curvatures, which are depicted, when present, as an abrupt change in height. Note how no discontinuity in one curvature (e.g., $\Delta\kappa_T = 0$) implies a large (in fact, the largest possible) discontinuity in the other curvature. Since the diagonal discontinuity remains fixed across the different triangles, the variation in $\Delta\kappa_T$ and $\Delta\kappa_N$ along it is irrelevant to the subject's decision.

Since we seek only real solutions, the trigonometric component of the last expression defines limits on the possible curvature discontinuities:

$$|\Delta\kappa| \leq \Delta\kappa_{max} \;\; \triangleq \;\; 2\Delta\theta_{within} \sin\left(\frac{\Delta\theta_{between}}{2}\right) . \tag{3.5}$$

This limit thus defines the three possible combinations for the curvature discontinuities which we used in the experiment

$$\begin{array}{rcl} \Delta\kappa_T = 0 & \Rightarrow & \Delta\kappa_N = \Delta\kappa_{max} \\ \Delta\kappa_T = \Delta\kappa_N & \Rightarrow & \Delta\kappa_T = \Delta\kappa_N = \frac{\Delta\kappa_{max}}{\sqrt{2}} \\ \Delta\kappa_N = 0 & \Rightarrow & \Delta\kappa_T = \Delta\kappa_{max} \qquad . \end{array} \tag{3.6}$$

With θ_0 set and plugged into Eq. 3.3, segments were plotted as antialiased lines centered at their jittered positions and having length of 25’. All stimuli were precomputed and stored before the beginning of the experiment and later displayed on a high-resolution flat color monitor (Dell UltraScan P991) using a 1 Ghz Pentium-III PC.

3.2.3 Subjects and Procedure

Six subjects (5 naive and the author) participated in this experiment. All subjects had normal or corrected-to-normal vision and each ran 6 sessions of 1080 trials (total of 6480 trials). Sessions were run on successive days and each consisted of 10 blocks of 108 trials (separated by short breaks). Each session lasted for an hour and included one or two practice blocks of 72 trials.

Each trial started with a 600 ms presentation of a mask (ODT of randomly oriented segments), followed by 200 ms of brief stimulus presentation and another 500 ms of post-stimulus mask[2]. A post marker prompted the subject to choose either a left- or right-pointing triangle. Each stimulus (i.e., a particular combination of $\Delta\theta_{within}$, $\Delta\theta_{between}$, and $\Delta\kappa_T$ and $\Delta\kappa_N$) was presented 20 times, 10 for each of the two possible figures, and their order was randomized and set before the experiment. To prevent observer strategies involving scrutinizing parts of the display, stimuli were randomly jittered up to $\pm 1.5°$ in each direction. During debriefing, all subjects reported using no special strategy other than global judgment in making their decisions.

---

[2] Stimulus presentation was set to 200 ms to approximately average presentation time from previous related studies; 160 ms in Nothdurft [1992] and 250ms in Wolfson and Landy [1995].

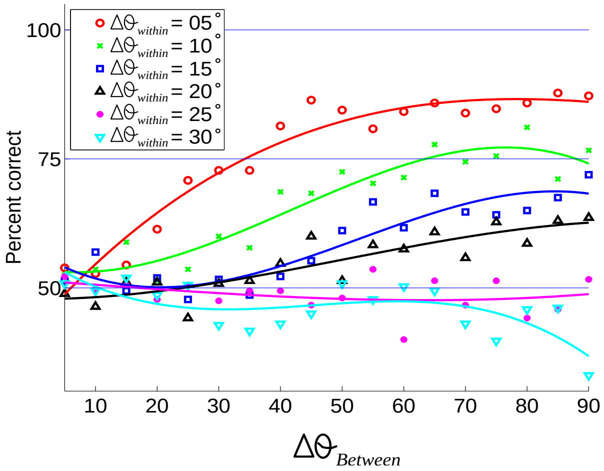


Figure 3.7: **Average accuracy of OBTS plotted against $\Delta\theta_{between}$.** Different colors represent the accuracy for different values of $\Delta\theta_{within}$. Note how performance decreases with larger $\Delta\theta_{within}$ and smaller $\Delta\theta_{between}$. Here, as well as in subsequent figures, graphs are 3^{rd} order polynomials fitted to the data.

3.2.4 Results

If current models of OBTS of non-constant ODTs [Nothdurft, 1991; Mussap and Levi, 1999] were valid, varying the behavior of curvature across the figure/ground boundary while holding $\nabla\theta$ and $\Delta\theta_{between}$ fixed should lead to no differences in segmentation performance. While Nothdurft [1992] demonstrated configural variations that affect segmentation performance in the presence of fixed $\Delta\theta_{within}$ and $\Delta\theta_{between}$ (see Sec. 3.2.5), here we show that similar conclusions can be drawn based on curvature discontinuities (i.e., based on intrinsic parameters only).

We first analyzed our data while pooling over the curvature dimension to control for, and replicate, existing findings. The results (Fig. 3.7) agree qualitatively with previously reported performance [Nothdurft, 1991; Mussap and Levi, 1999] and indicate that on average, segmentation accuracy decreases with larger $\Delta\theta_{within}$ and smaller $\Delta\theta_{between}$. Reliable segmentation occurs only when the ratio $\frac{\Delta\theta_{between}}{\Delta\theta within}$ is sufficiently larger than 1. The detection thresholds (75% detection accuracy) that we obtained ($\frac{\Delta\theta_{between}}{\Delta\theta within} \approx 7$) were higher than those reported in the past due to the different stimuli, procedure, and other experimental parameters that we used.

Taking curvature into account, however, reveals a different and intriguing pattern because the different curvature-based classes of orientation discontinuities produce *significant and consistent differences* in segmentation performance (Fig. 3.8). Three qualitative observations follow from such a curvature-based analysis. First, for small $\nabla\theta$ ($\Delta\theta_{within} \leq 10°$), segmentation of discontinuities with $\Delta\kappa_T = 0$ (red graphs) was substantially inferior[3]. In fact, subjects were able to reliably detect this class of discontinuities only for $\Delta\theta_{within} \leq 5°$ and $\Delta\theta_{between} \geq 75°$. $\Delta\theta_{within}$ greater than $5°$ prevented any segmentation of this class of discontinuities, regardless of $\Delta\theta_{between}$. Secondly, for intermediate $\nabla\theta$ values ($15° \leq \Delta\theta_{within} \leq 20°$), the only detectable class of orientation discontinuities was the one which was discontinuous in *both* curvatures (green graphs). This superior performance was found to be statistically significant both on average (Fig. 3.9) and individually for each $\Delta\theta_{between}$ beyond the detection threshold (not shown for space considerations). Finally, the largest $\nabla\theta$ values ($\Delta\theta_{within} \geq 25°$) collapse all cases to chance level, indicating that OBTS is uniformly impossible for this class of patterns.

It is particularly instructive to examine the differences in segmentation performance between discontinuity classes in light of the average performance for each $\Delta\theta_{within}$ (Fig. 3.7). For example, the blue graph in Fig. 3.7 that corresponds to average performance with $\Delta\theta_{within} = 15°$ shows no crossing of the 75% detection threshold. However, splitting this average into its components (Fig. 3.8, left panel on second row) reveals that this low average is dominated by the two discontinuity classes of $\Delta\kappa_T = 0$ and $\Delta\kappa_N = 0$, both of which are hardly raised above chance level. But to conclude that no reliable segmentation is possible at $\Delta\theta_{within} = 15°$ is clearly incorrect as the $\Delta\kappa_T = \Delta\kappa_N$ graph in this category shows excellent segmentation performance for $\Delta\theta_{between} \geq 50°$. Since the $\Delta\kappa_T = \Delta\kappa_N$ class of discontinuities was typically the most salient for the smaller values of $\nabla\theta$ too, we conclude that, in general, orientation discontinuities of equal orientation contrast are *not* all perceptually equivalent and that those with $\Delta\kappa_T = \Delta\kappa_N$, where the discontinuities in κ_T and κ_N are maximized simultaneously, are the most salient ones.

---

[3] Note that the combination of $\Delta\kappa_T = 0$ with small $\Delta\theta_{within}$ and small $\Delta\theta_{between}$ produced performance levels significantly less than chance. Inspection of the stimuli suggests that this is due to the "edge hallucination" phenomenon reported by Nothdurft [1992].

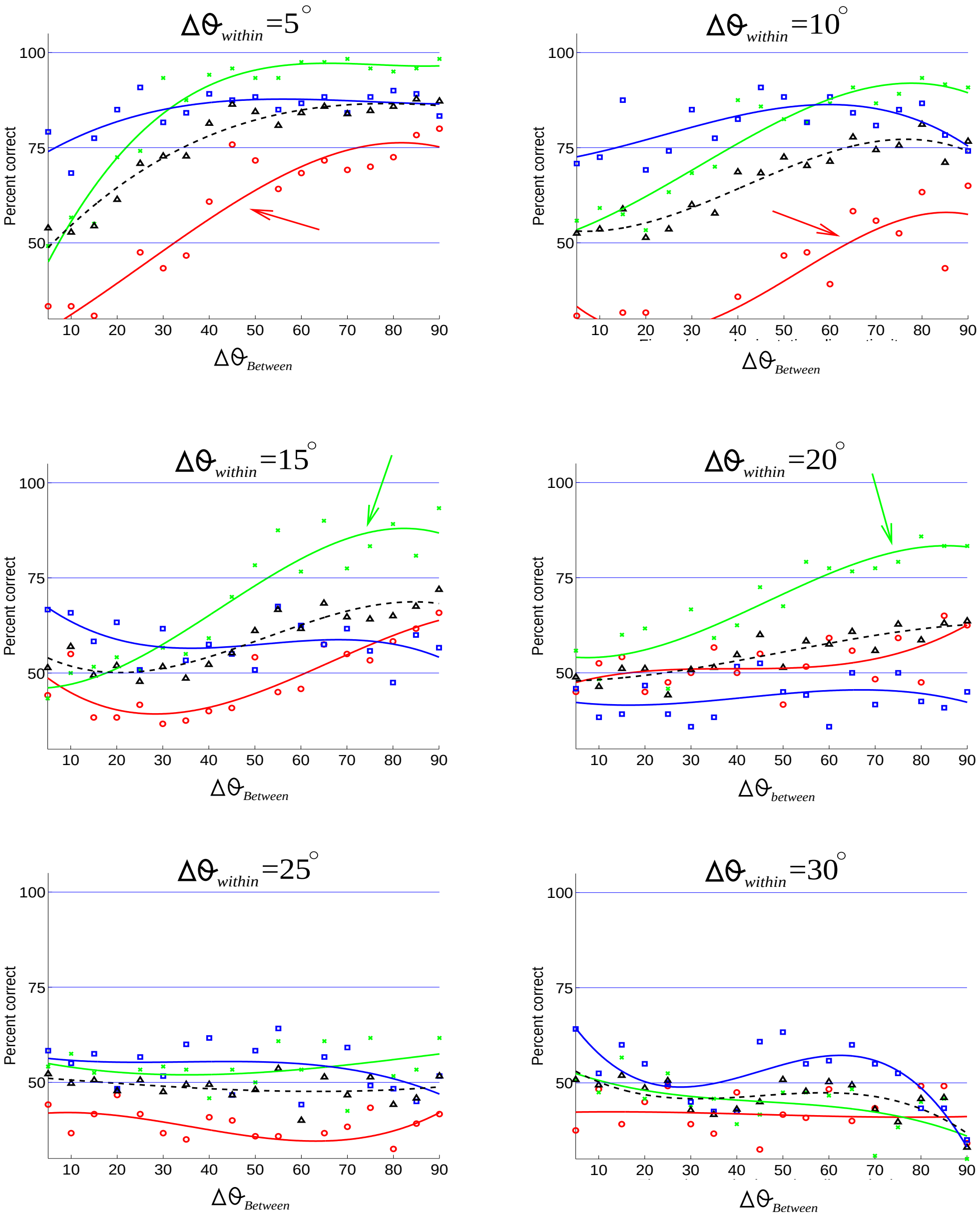


Figure 3.8: **Effect of curvature discontinuities on OBTS performance.** Each panel corresponds to a different value of $\Delta\theta_{within}$. X axes represent $\Delta\theta_{between}$. Red graphs correspond to $\Delta\kappa_T = 0$ discontinuities. Blue graphs correspond to $\Delta\kappa_N = 0$ discontinuities. Green graphs correspond to discontinuities with $\Delta\kappa_T = \Delta\kappa_N$. Dashed black graphs are the average for each case (replicated from Fig. 3.7). Note how the different curvature classes produce markedly different graphs. In particular, for small gradients ($\Delta\theta_{within} \leq 10°$), $\Delta\kappa_T = 0$ discontinuities (graphs marked with red arrows) are much less salient while for intermediate gradients ($15° \leq \Delta\theta_{within} \leq 20°$) only discontinuities of $\Delta\kappa_T = \Delta\kappa_N$ cross the detection threshold (graphs marked with green arrows). At higher gradients, all classes collapse to chance level.

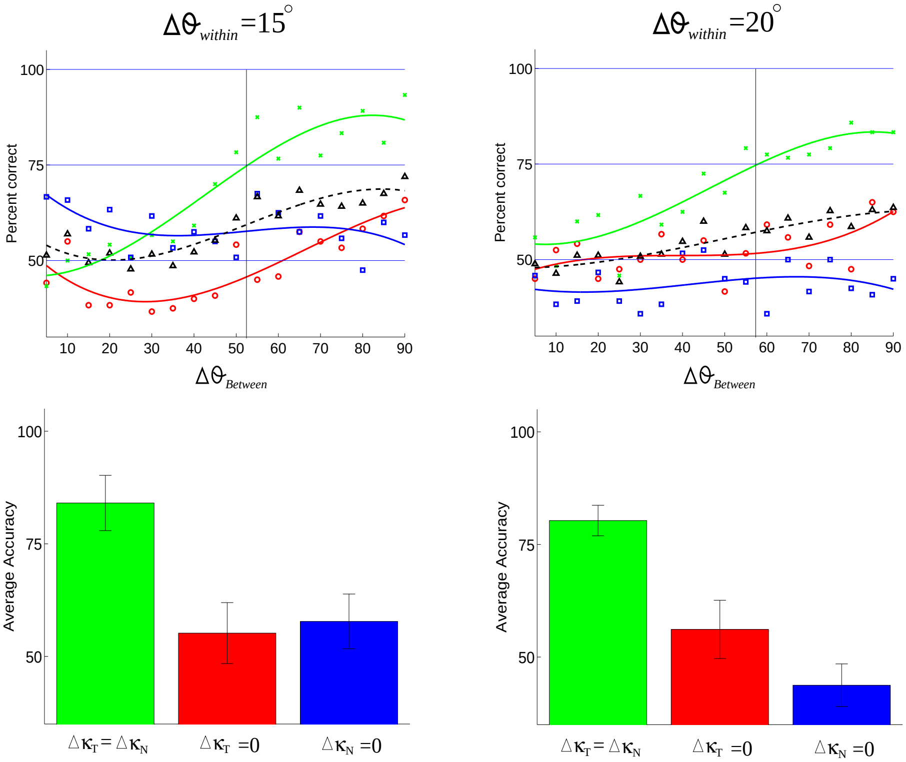


Figure 3.9: **Statistical significance of curvature-based OBTS performance.** When the $\Delta\kappa_T = \Delta\kappa_N$ class of discontinuities is superior to the others ($15° \leq \Delta\theta_{within} \leq 20°$), the differences in detection accuracies are statistically significant. Since an unbiased indicator for superiority in performance is signaled by the 75% accuracy threshold, we checked the statistical differences between the cases by identifying this threshold for the graphs that crossed them (marked by the black vertical lines) and collecting all data points beyond it. The two bar plots show the averages of these data sets for both $\Delta\theta_{within} = 15°$ and $\Delta\theta_{within} = 20°$. Error bars represent ±1 standard error. In both cases, the difference between the $\Delta\kappa_T = \Delta\kappa_N$ mean (green) and the other two is highly significant ($p < 0.001$).

3.2.5 Discussion - ODT curvatures and configural effects

The results of Experiment 1 show that OBTS clearly is affected by more than just the orientation gradients $\Delta\theta_{within}$ and $\Delta\theta_{between}$, and suggest that curvature discontinuities can play a major role in segmentation performance. Factors other than $\Delta\theta_{within}$ and $\Delta\theta_{between}$ that affect OBTS were previously observed psychologically [Olsen and Attneave, 1970], but have been barely explored systematically. The main related observation was made by Wolfson and Landy who found that OBTS is improved if the "orientation of texels [is] parallel, and to some extent perpendicular, to the [orientation defined] texture edge" [Wolfson and Landy, 1995, p. 2876]. This observation is significant because the configural factor it refers to relates directly to the generic cut and fold organizations we described earlier (Sec. 3.1.1; Figure 3.1). Equally important is what is missing from this observation, namely, $\nabla\theta$ and $\Delta\theta_{between}$. In other words, it suggests that the configuration around the orientation edge affects its saliency, but that this effect is independent of the orientation gradients.

While the lack of dependency on orientation gradients may be a direct outcome of the constrained piecewise-constant ODTs that Wolfson and Landy [1995] experimented with, their observation that "OBTS is improved if the orientation of texels [is] parallel, and to some extent perpendicular, to the texture edge" in itself makes predictions about intermediate configurations rather difficult. In Fig 3.4, for example, the top (salient) edge of the square is defined by an orientation jump from $27°$ to $45°$, while the bottom (hardly seen) edge is defined by an orientation jump from $90°$ to $108°$. Thus, the top edge is much further from a "parallel configuration" than the bottom one is from a "perpendicular configuration". Why, then, is the top edge significantly more salient than the bottom one? Could the spatially varying nature of this ODT contribute to the perceptual outcome?

To the best of our knowledge, the only systematic study of configural effects in spatially *varying* ODTs was carried out by Nothdurft [1992]. Employing methods from his earlier studies [Nothdurft, 1985a, 1991], Nothdurft systematically examined OBTS performance as a function of the average configuration of the figure's texels relative to the figure/ground boundary. Predating Wolfson and Landy [1995], he too observed a "systematic dependence of correct bar detection on the alignment of border and element orientations", albeit with no contribution from perpendicular configurations. Unfortunately, [Nothdurft, 1992] examined only ODTs of one particular $\Delta\theta_{within}$, leaving open the question about the interaction of this parameter with configural effects in OBTS. Furthermore, his choice of the direction of $\nabla\theta$ relative to the orientation edges in the stimuli introduced deviations of up to 30° from the desired configuration (texels either parallel or perpendicular to the edge), thus incorporating a significant amount of uncertainty into the results[4].

A critical aspect of the similar observations made by Nothdurft [1992] and Wolfson and Landy [1995] is that both are based on, and phrased in terms of, the orientation edge itself. Wolfson and Landy [Wolfson and Landy, 1995] go a step further to propose a computational "energy model that can account [for the detection of the edge and its saliency] by *giving extra weight to the oriented channel which is oriented similarly to the edge*" (emphasis added). Unfortunately, such an explanation introduces a chicken-and-egg problem; the outcome (i.e., the orientation edge) must be given as an input to the computational process from which it is supposed to emerge (and after all, the goal of OBTS is to find these edges). Wolfson and Landy recognized this as well, and commented that a model that takes into account the orientation of the texture edge (or any other aspect of it) is "not particularly compelling" [Wolfson and Landy, 1995, p.2872]. Indeed, an appropriate model should rely on intrinsic data only and should predict configural modulations as a side effect.

The above discussion raises a question about our stimuli: have all of the interacting factors in defining our ODTs - the positions, orientations, and curvatures - somehow conspired to work together in an unanticipated way, for example by producing edge configurations for which configural effects were already observed? Two principle outcomes are possible. First, we might find that our curvature-based findings are indeed simply a rephrasing of the previously observed configural effects. Such an outcome is not undesired; it will imply that all existing findings can be explained based on intrinsic measurements only, thus avoiding the chicken-and-egg problem.

Alternatively, we might find that our curvature-based findings are incompatible with the previously reported configural effects. Such an outcome will have significant consequences of a different flavor: it will suggest that previously observed configural effects are neither universally true, nor are they independent of orientation gradients and ODT curvatures.

---

[4] Deviations up to 13.3° from the studied configuration were also present in [Wolfson and Landy, 1995].

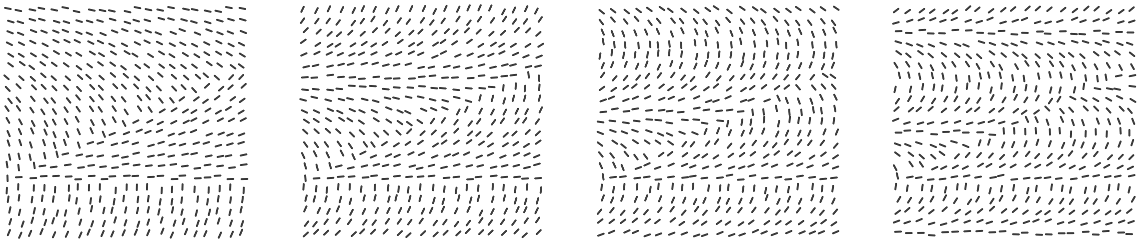


Figure 3.10: **A possible relationship between curvature discontinuities and edge configuration?** With $\Delta\theta_{between}$ set to 90°, the condition $\Delta\kappa_T = \Delta\kappa_N$ consistency produces a parallel/perpendicular configuration at the orientation edge. Shown here (left to right) are the left-triangle stimuli for $\Delta\theta_{within} \in \{5^\circ, 10^\circ, 15^\circ, 20^\circ\}$.

While Fig 3.4 implies that configural factors *may* be modulated by ODT variations (see page 70), Fig. 3.6 suggests that at least in some cases, the salient $\Delta\kappa_T = \Delta\kappa_N$ condition produces a parallel/perpendicular configuration at the orientation edge. Fig 3.10 further implies that this correspondence is independent of $\Delta\theta_{within}$ and thus not accidental, at least when $\Delta\theta_{between} = 90^\circ$. All this conflicting evidence suggests that a thorough examination of the relationship between configural effects and our curvature-based findings is required. The rest of this chapter is devoted to this issue, which we address both computationally and psychophysically. The conclusion that emerges is that previously observed configural effects are, in fact, incompatible with the results based on curvature, and that they are modulated by ODT variations in the proximity of orientation edges.

3.3 Computational reexamination of configural effects

To simplify forthcoming wording, let the *configuration hypothesis* be the observation that performance improves when texels are oriented parallel (and “to some extent perpendicular”) to the perceived texture edge [Nothdurft, 1992; Wolfson and Landy, 1995]. We now formalize it into a quantitative saliency measure to examine whether it is compatible with the results of Experiment 1. To do so, we assume that the orientation of the texture edge is available as a parameter, although to do so temporarily ignores the chicken-and-egg problem we mentioned above (Sec 3.2.5). Despite this shortcoming, the analysis is useful in several respects: (i) it provides an explicit interpretation of our curvature-based findings in terms of previous configural observations, thus facilitating the comparison between them, (ii) it allows us to evaluate the configuration hypothesis for a continuum of configurations and orientation gradients, and (iii) it provides one line of investigation that leads to the design of the psychophysical control experiment in Sec. 3.4.

3.3.1 Formal saliency based on parallel configuration

Let θ_t denote the orientation of texels near orientation edges and let θ_e be the orientation of the edge itself. Denote by $s_t(\theta_t, \theta_e)$ the configural saliency induced by texels near the orientation edge. Since there is no agreement in the literature about the contribution of perpendicular configurations, in the following we assume that OBTS performance improves only when the orientation of texels is *parallel* to the orientation edge, a configuration which we call "parallel". Allowing graceful degradation of this improvement with changes in θ_t, we define a normalized $s_t(\theta_t, \theta_e)$ as

$$s_t(\theta_t, \theta_e) \triangleq 1 - \frac{1}{90} \cdot D_{90}(\theta_t, \theta_e), \tag{3.7}$$

where $D_{90}(\alpha, \beta)$ is the angular distance between the two given orientations, i.e., a number in the range $[0, 90]$ (Fig 3.11a). In the rest of this paper we assume that $\theta_e = 0°$ (in accordance with the orientation edges in our stimuli), and simplify s_t to

$$s_t(\theta_t) \triangleq 1 - \frac{1}{90} \cdot D_{90}(\theta_t, 0), \tag{3.8}$$

as illustrated in Fig 3.11b.

By definition, any orientation edge is created by a one dimensional discontinuity in the 2D orientation field (or function). Since our edges are horizontal, we label the orientation of the field along the two sides of the discontinuity by θ_{down} and θ_{up}. Consequently, the configural saliency of the edge is some combination of the configural saliency of the texels on the two sides of the discontinuity. An intuitive and straight forward combination is the L_∞ norm (i.e., the maximum) of the saliencies of the two sides. Incorporating also the contribution of the orientation gradient $\Delta\theta_{between}$ across the edge, we therefore model the edge's perceptual saliency $S(\theta_{down}, \theta_{up})$ in terms of θ_{down} and $\Delta\theta_{between}$ as follows (note that $\theta_{up} = \theta_{down} + \Delta\theta_{between}$):

$$S(\theta_{down}, \Delta\theta_{between}) \triangleq \frac{\Delta\theta_{between}}{90} \cdot \max\left[s_t(\theta_{down}), s_t(\theta_{down} + \Delta\theta_{between})\right] \tag{3.9}$$

This function, whose parameters are (1) the texels orientation on one side of the edge, and (2) the orientation gradient across the edge, is illustrated in Fig 3.11c. To examine if it is a good predictor we computed different cross-sections that correspond to previous findings in the literature (Fig 3.11d). In particular, compare the cross section for $\Delta\theta_{between} = 90^\circ$ with Fig. 12C in Wolfson and Landy [1995], or the one for $\theta_{down} = 135^\circ$ to their Fig. 12D (or Figs 12H,I after appropriate rotation of the X axis).

Equipped with a quantitative model that formalizes the configuration hypothesis, we now apply it to the curvature based stimuli used in Experiment 1, make saliency predictions with regard to them, and compare these predictions to the psychophysical results. As we show, this comparison reveals important deficiencies in the configuration hypothesis.

To apply the saliency model to the three cases of curvature discontinuities, we first find the behavior of the orientation field around the discontinuities that they induce. In other words, for each curvature discontinuity and $\Delta\theta_{between}$, we find its corresponding θ_{down} by evaluating the orientation functions in Eq. 3.3 at the discontinuity position, with θ_0 solved through Eq. 3.4 for the three cases of curvature discontinuity (Eq. 3.6). This computation yields the following values, which turn out to be independent of $\Delta\theta_{within}$:

$$\begin{aligned} \theta_{down}(\Delta\kappa_T = 0) &= -\tfrac{\Delta\theta_{between}}{2} - 90^\circ \\ \theta_{down}(\Delta\kappa_T = \Delta\kappa_N) &= -\tfrac{\Delta\theta_{between}}{2} - 45^\circ \\ \theta_{down}(\Delta\kappa_N = 0) &= -\tfrac{\Delta\theta_{between}}{2} \end{aligned} \tag{3.10}$$

Plugged into the saliency function $S(\theta_{down}, \Delta\theta_{between})$, this evaluates the configural saliency of the edges in our stimuli, *as predicted from the configuration hypothesis*. In the same spirit, we also evaluate the predicted saliency of parallel configurations. The four resultant saliency graphs are plotted to scale in Fig 3.12.

Fig. 3.12 is the main result of this section as it summarizes the relationship between curvature-based edges and their expected configural saliency based on the configuration hypothesis. A few observations are immediate, and most clearly is the one that, for any given $\Delta\theta_{between}$, parallel configurations (black graph) are always more salient than any of the curvature-defined edges. This confirms that the saliency measure captures what the configuration hypothesis was designed to express intuitively.

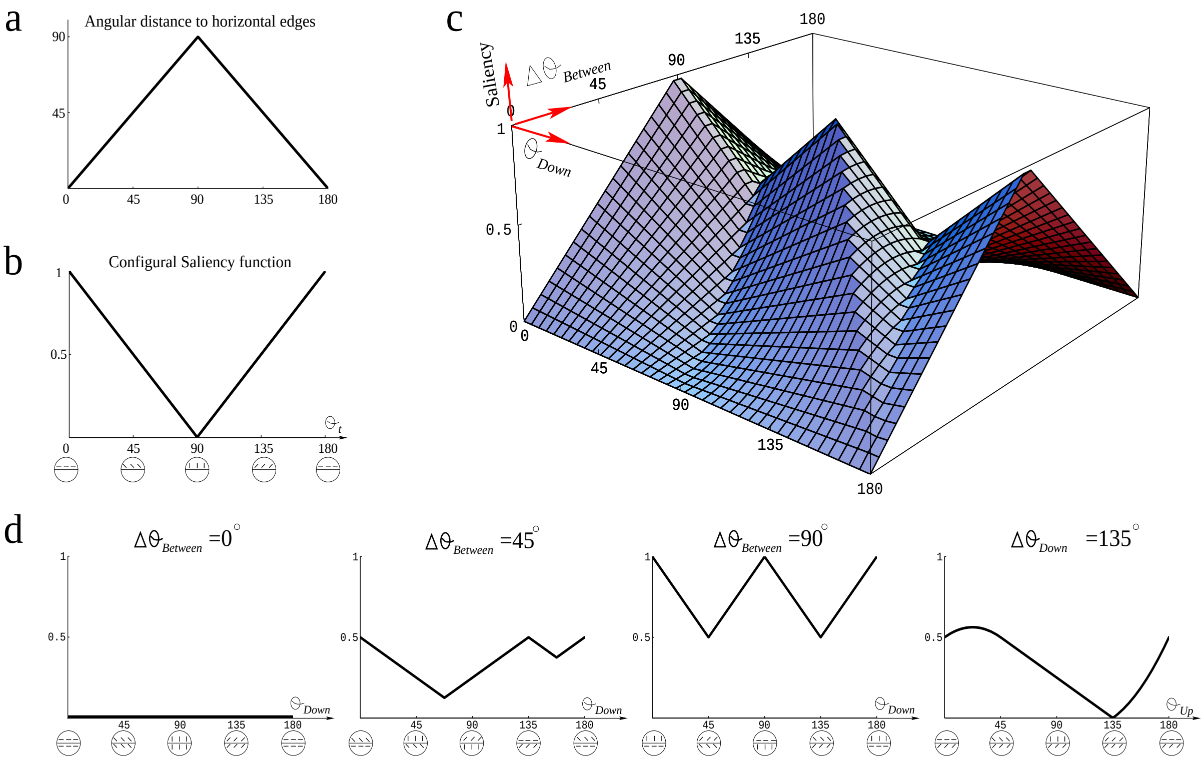


Figure 3.11: **Formalization of the configuration hypothesis.** **(a)** Since the configuration hypothesis is based on the relative angle between the orientation edge and the texture elements, some notion of “distance” between orientations is needed. We use the most natural angular distance which measures the shortest distance on the unit circle (modulo $90°$). Shown here is that distance from the horizontal orientation, i.e., from $0°$. **(b)** The saliency function (1 is most salient) based on the configuration hypothesis. Again, here we show saliency of horizontal edges. **(c)** The saliency function for all possible horizontal orientation edges (Eq. 3.9). It vanishes for edges of $\Delta\theta_{between} = 0°$ and it peaks for parallel configurations of $\Delta\theta_{between} = 90°$. **(d)** Few cross sections of $S(\theta_{down}, \Delta\theta_{between})$. The three cross sections on the left correspond to fixed values of $\Delta\theta_{between}$. The cross section on the right corresponds to a fixed $\theta_{down} = 135$ and varying θ_{up} (i.e., varying $\Delta\theta_{between}$ as well). See text for comparison to psychophysical measurements in the literature (note that zero saliency level corresponds to chance level in OBTS experiments).

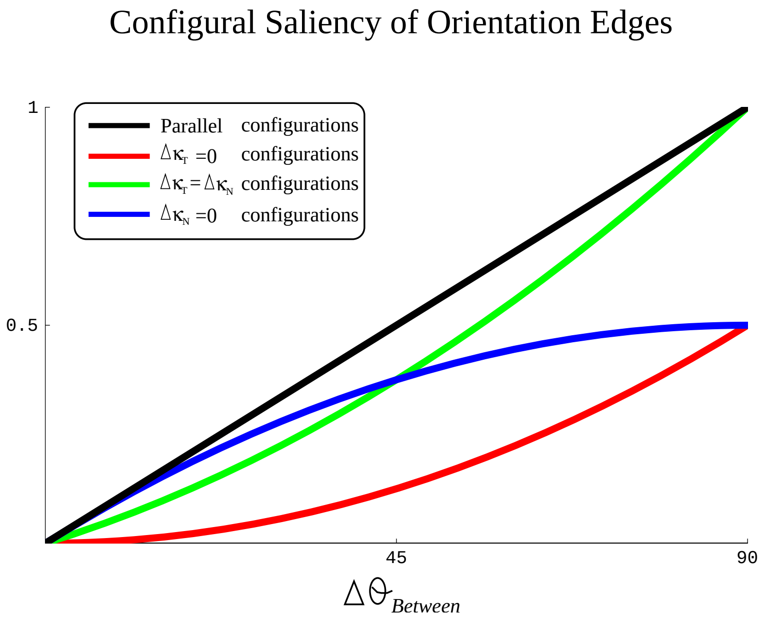


Figure 3.12: **Predicted saliency of orientation edges based on the configuration hypothesis.** Shown here is the configuration-modulated saliency of orientation edges as a function of their orientation gradient $\Delta\theta_{between}$. The black graph depicts the predicted saliency of parallel configurations. The red, green, and blue graphs depict the predicted saliency of the orientation edges induces by the $\Delta\kappa_{\ T} = 0$, $\Delta\kappa_T = \Delta\kappa_N$, and $\Delta\kappa_N = 0$ edges, respectively (the three cases are color-coded similar to previous figures).

Based on Fig. 3.12, one can make certain predictions with regard to the *relative saliency* of the three curvature-based orientation edges used in Experiment 1. In particular, we observe the following:

P1: $\Delta\kappa_T = 0$ discontinuities are the least salient for all $\Delta\theta_{between} < 90°$.

P2: $\Delta\kappa_T = 0$ and $\Delta\kappa_N = 0$ discontinuities are equally salient for $\Delta\theta_{between} = 90°$.

P3: $\Delta\kappa_T = \Delta\kappa_N$ discontinuities are the most salient for all $\Delta\theta_{between} > 45°$.

P4: $\Delta\kappa_N = 0$ discontinuities are the most salient for all $\Delta\theta_{between} < 45°$.

P5: $\Delta\kappa_N = 0$ and $\Delta\kappa_T = \Delta\kappa_N$ discontinuities are equally salient for $\Delta\theta_{between} = 45°$.

As can be observed in Fig. 3.8, predictions P1 and P3 are largely confirmed by Experiment 1. The rest, however, are either wrong or depend critically on $\Delta\theta_{within}$, in clear contradiction to the lack of such dependency in the configuration hypothesis and its derived saliency measure. In particular:

- P2 appears to be correct for $\Delta\theta_{within} \geq 15^\circ$ only, but not for smaller ones . For example, the relative performance at $\Delta\theta_{within} = 10^\circ$ and $\Delta\theta_{between} = 90^\circ$ shows clearly that the $\Delta\kappa_N = 0$ (blue) discontinuity is significantly more salient than the $\Delta\kappa_T = 0$ (red) one.

- P4 appears to be partially correct for smaller $\Delta\theta_{within}$ but not for larger ones. For example, it is clearly wrong for $\Delta\theta_{within} = 20^\circ$.

- P5 is wrong for $\Delta\theta_{within} > 10^\circ$.

The observations above strongly suggest that there is a significant gap between the configuration hypothesis and performance in OBTS. It may be argued that all it takes to fix the wrong predictions, and the configuration hypothesis itself, is the incorporation of $\Delta\theta_{within}$ into Eq. 3.9 in a way implied by signal detection theories and recent related findings in the OBTS literature [Nothdurft, 1991; Mussap and Levi, 1999]. However, doing so will reduce the saliency of *all* orientation edges by a factor proportional to $\Delta\theta_{within}$ and thus will do no more than to equally shift down (and rectify at zero) all the graphs in Fig. 3.12. Since such a transformation does not change the *relative* order of the graphs, the *relative* saliencies and the wrong predictions that follow from them will still hold. We conclude that even an expansion of the model to address basic signal detection concerns is insufficient to explain the results of Experiment 1.

3.3.2 Formal saliency based on other configuration hypotheses

The discussion in Sec. 3.3.1 revolves around the hypothesis that performance in OBTS improves when the texels' orientation is parallel to the perceived texture edge [Nothdurft, 1992; Wolfson and Landy, 1995]. However, it is possible that the configural content of this hypothesis is incomplete. In fact, Wolfson and Landy themselves relaxed the conclusion, saying that "performance [in OBTS] improves when the texels' orientation is parallel, *and to some extent perpendicular,* to the perceived texture edge" [Wolfson and Landy, 1995, p. 2876, emphasis added]. To examine this and other variations, we repeated the analysis of Sec. 3.3.1 for a variety of other saliency models, including models that assign some saliency to perpendicular configurations, and models that modulate saliency non-linearly as the configuration shifts from the optimal ones. All these models either failed to predict the previous findings of Wolfson and Landy [1995] even at the most basic levels (e.g., they failed to reproduce the correct cross sections similar to those in Fig. 3.11D), or they resulted in saliency predictions similar to those produced from the original model of Sec. 3.3.1. Examples of the predictions from a non linear model that assigns some saliency to perpendicular configurations are illustrated in Fig 3.13.

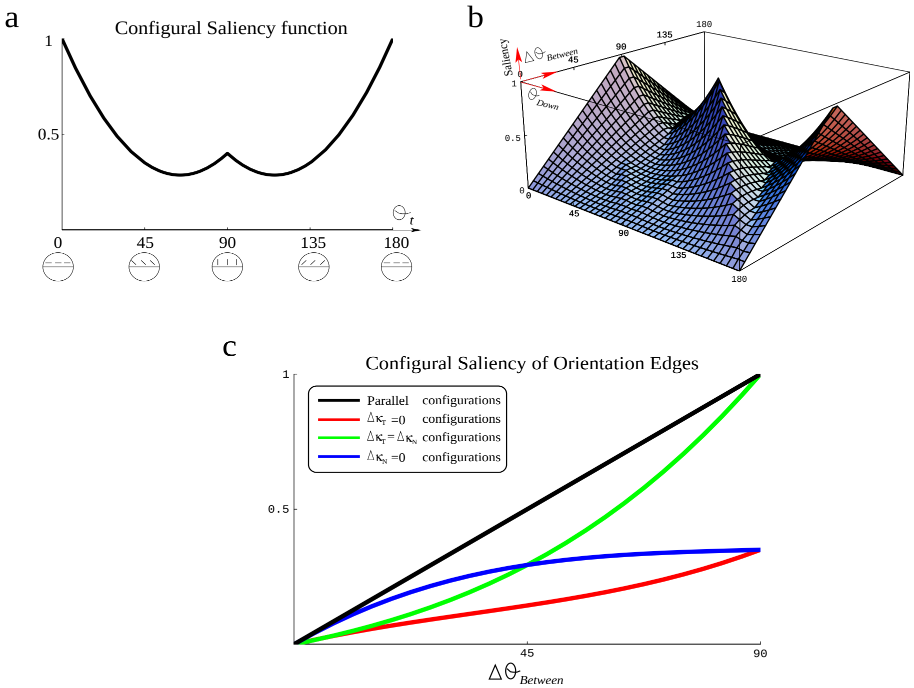


Figure 3.13: **The configuration-modulated saliency based on a different saliency model.** The model presented here has non-linear (parabolic) saliency degradation with assignment of some saliency to perpendicular configurations as well. **(a)** The saliency function of the variant model (compare to Fig. 3.11b). **(b)** The corresponding saliency function for all possible horizontal orientation edges (compare to Fig. 3.11c). **(c)** The configuration-modulated saliency of orientation edges as a function of their orientation gradient $\Delta\theta_{between}$. Different cases are color-coded similar to Fig. 3.12. Note that despite some perturbation in the graphs themselves, their mutual relationship and relative ordering is preserved. Thus, this saliency model, which "improves performance when texels are parallel, and to some extent perpendicular, to the perceived texture edge" [Wolfson and Landy, 1995], produces qualitatively similar predictions to those discussed in Sec. 3.3.1.

3.3.3 Discussion

The formalization of the configuration hypothesis into a saliency measure provides a basis to relate it to the curvature-based findings of Experiment 1 and leads to the conclusion that there exists a gap between the simple configural effect that this hypothesis articulates and actual performance in OBTS. Naturally, this raises the possibility that the configuration hypothesis is not explanatory even for those configurations it designates as optimal. Thus, we now wish to scrutinize the relationship, and differences in saliency, between parallel configurations and the curvature-based edges of $\Delta\kappa_T = \Delta\kappa_N$ that emerged as optimal in the conditions tested by Experiment 1. Although some data relevant to these questions may be hidden implicitly in the interaction between the saliency graphs (Fig. 3.12) and the curvature-based findings (Fig. 3.8), we next address these issues directly in Experiment 2,

To compare psychophysically two classes of stimuli whose differences are likely to be subtle, one first needs to identify those cases that maximize the differences between them. Fig. 3.12 suggests that the saliency of $\Delta\kappa_T = \Delta\kappa_N$ discontinuities (green graph) is somewhat less than that of parallel configurations (black graph). It shows that the differences collapse at very small and very large $\Delta\theta_{between}$, and that they are maximized at $\Delta\theta_{between} = 45°$. Fig. 3.12 thus implies that the best candidates for revealing differences between the two classes of stimuli are those with $\Delta\theta_{between} = 45°$. This conclusion, however, is based on the saliency measure we developed thus far, while in Experiment 2 we seek a far less constrained argument, i.e., one that is completely independent of any particular modeling step. As it turns out, even without any modeling of saliency, stimuli of $\Delta\theta_{between} = 45°$ are indeed the best candidates to probe OBTS for differences between the configuration hypothesis and curvature-based edges.

3.4 Experiment 2 – Curvature-based vs. parallel configurations

The goal of Experiment 2 is to examine the differences in saliency between parallel configurations and orientation edges with $\Delta\kappa_T = \Delta\kappa_N$, while all other factors ($\Delta\theta_{within}$ and $\Delta\theta_{between}$) are held constant. We first observe that it makes no sense to look for possible differences for *all* values of $\Delta\theta_{between}$. As Fig. 3.10 and Eq. 3.10 show, the two cases collapse when $\Delta\theta_{between} = 90°$

(i.e., edges of maximal orientation gradient and strict parallel/perpendicular configuration also have $\Delta\kappa_T = \Delta\kappa_N$) and thus no differences should be expected then. Eq. 3.10 shows, however, that in general, these two classes of edges will be different. To maximize the chance of detecting perceptual differences, we need to find the $\Delta\theta_{between}$ for which the differences between the two classes are also maximized. In other words, we need to find $\Delta\theta_{between}$ for which parallel configurations have curvature discontinuities which *differ the most* from the $\Delta\kappa_T = \Delta\kappa_N$ condition. For this reason we first need to evaluate the curvature discontinuities induced by parallel configurations of all possible $\Delta\theta_{between}$.

Given $\Delta\theta_{between}$, we first solve Eq. 3.3 for θ_0 that gives rise to a parallel configuration at an arbitrary edge position $y = D$ (recall that all our orientation edges are horizontal). Since parallel configurations can be characterized by parallel texels on either side, a proper analysis should consider both the case where $\theta_{ground}(x, D) = k \cdot 180°$ and the one where $\theta_{figure}(x, D) = k \cdot 180°$, for $k \in \{0, 1, 2, \ldots\}$. As it turns out, the two cases yield the same conclusion, which is also independent of k. Thus, in the following we describe the results for parallel configurations of $\theta_{ground}(x, D) = 0°$, which implies that $\theta_0 = -D\Delta\theta_{within}$ (see Eq. 3.3).

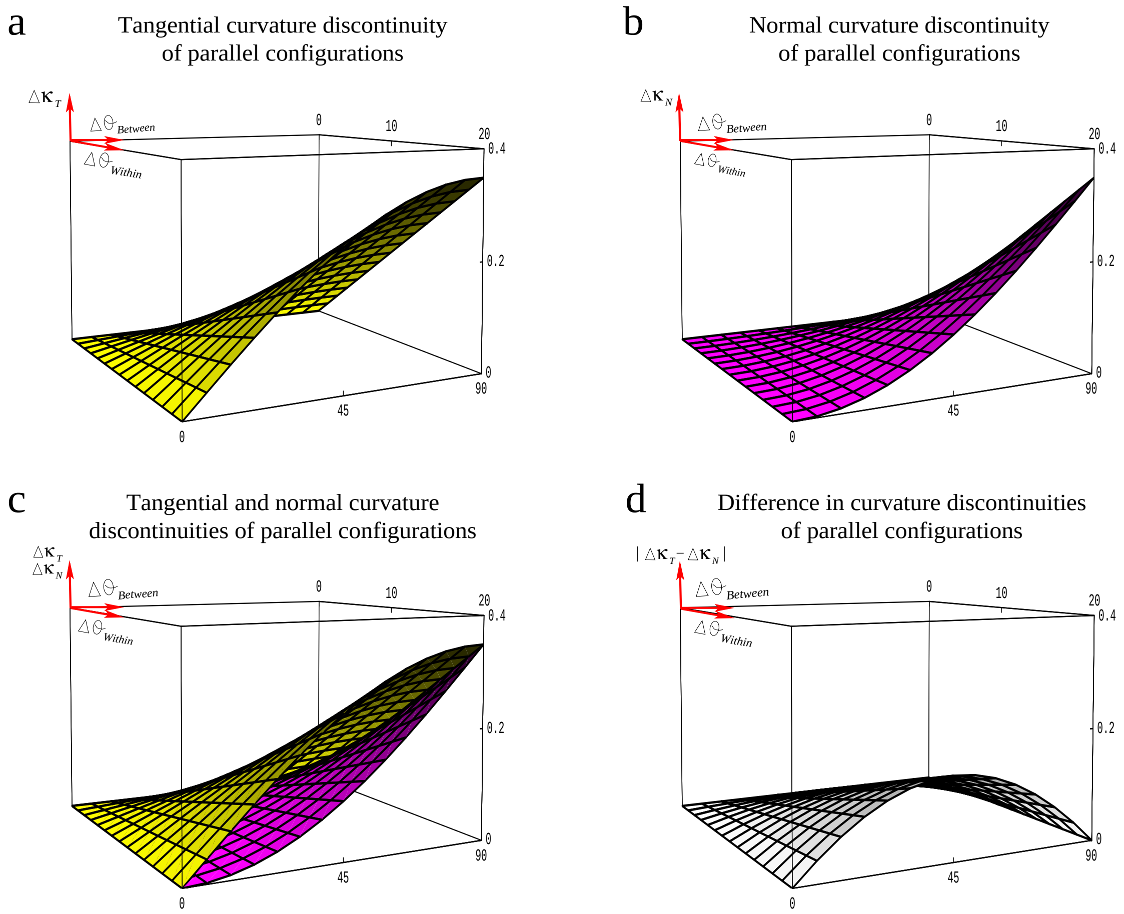
With θ_0 determined, so are θ_{ground} and θ_{figure} (Eq. 3.3). Now we compute their corresponding curvature functions $\kappa_T(x, y)$ and $\kappa_N(x, y)$ from Eq. 3.2, subtract, and evaluate at the edge position to yield both $\Delta\kappa_T$ and $\Delta\kappa_N$:

$$\theta_{ground}(x, D) = 0 \quad \Rightarrow \quad \begin{cases} \Delta\kappa_T(x, D) = \Delta\theta_{within} \cdot \sin(\Delta\theta_{between}) \\ \Delta\kappa_N(x, D) = \Delta\theta_{within} \cdot [\cos(\Delta\theta_{between}) - 1] \,. \end{cases} \tag{3.11}$$

These functions are plotted against $\Delta\theta_{within}$ and $\Delta\theta_{between}$ in Fig. 3.14a,b.

Recall that we are looking for the conditions under which parallel configurations differ the most from the $\Delta\kappa_T = \Delta\kappa_N$ condition. This can be solved analytically by maximizing the magnitude of the difference between $\Delta\kappa_T(x, D)$ and $\Delta\kappa_N(x, D)$ in Eq. 3.11. The solution turns out to be $\Delta\theta_{between} = 45°$, regardless of $\Delta\theta_{within}$. The same solution is illustrated graphically in Fig. 3.14c,d.

We conclude that if the parallel and curvature-based configurations exhibit differences in saliency, they will be most prominent in orientation edges of $\Delta\theta_{between} = 45°$. Fig. 3.14 also shows that these differences are rather subtle and that they grow linearly with $\Delta\theta_{within}$.



Tangential curvature discontinuity of parallel configurations

Normal curvature discontinuity of parallel configurations

Tangential and normal curvature discontinuities of parallel configurations

Difference in curvature discontinuities of parallel configurations

Figure 3.14: **Curvature discontinuities of parallel configurations. (a)** Tangential curvature discontinuity $\Delta\kappa_T$ as a function of $\Delta\theta_{within}$ and $\Delta\theta_{between}$. **(b)** Normal curvature discontinuity $\Delta\kappa_N$ as a function of $\Delta\theta_{within}$ and $\Delta\theta_{between}$. **(c)** Superimposing both tangential and normal curvature discontinuities on the same graph shows how the difference vanishes for small and large $\Delta\theta_{between}$. **(d)** A graph of $|\Delta\kappa_T - \Delta\kappa_N|$ shows clearly that the difference between the two curvature discontinuities is maximized for $\Delta\theta_{between} = 45°$ (for any given $\Delta\theta_{within}$) and that it grows linearly with $\Delta\theta_{within}$.

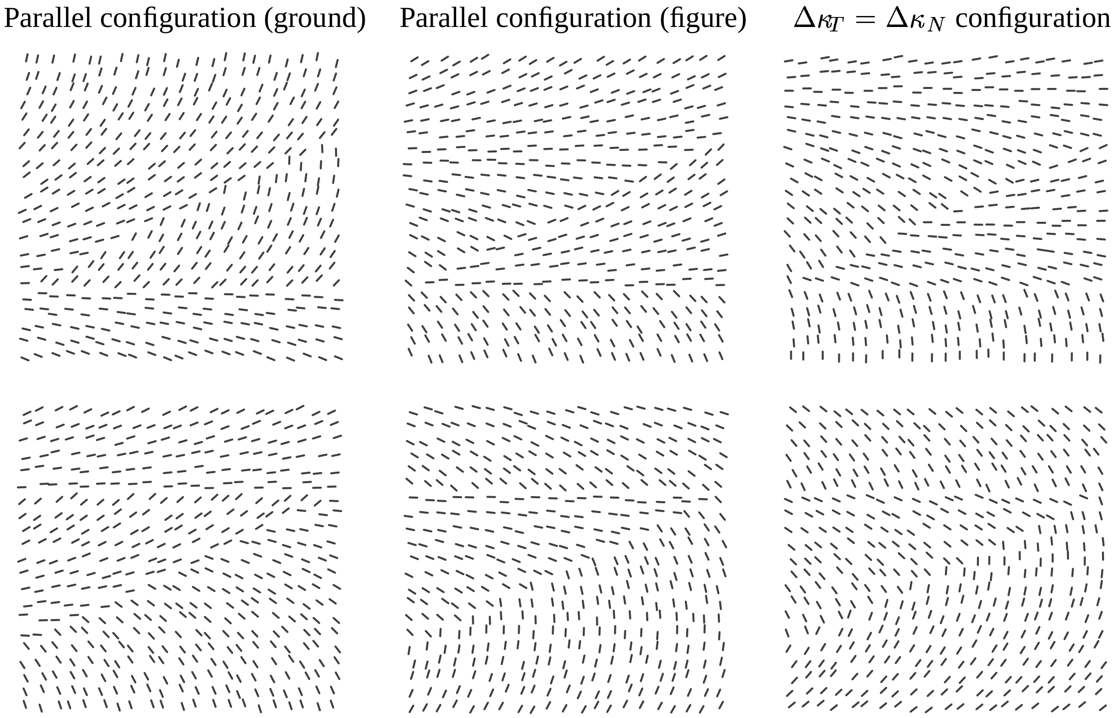


Figure 3.15: **Example of the stimuli used in Experiment 2.** Shown here for is the set of stimuli for $\Delta\theta_{within} = 5°$. Similar sets were constructed for $\Delta\theta_{within} \in \{10°, 15°, 20°\}$. In all cases $\Delta\theta_{between}$ was set to $45°$.

3.4.1 Stimuli

Stimuli used in Experiment 2 were constructed similarly to Experiment 1, although $\Delta\theta_{between}$ was fixed at $45°$ (as a result of the discussion above) and $\Delta\theta_{within}$ was limited to no more than 20° (beyond which performance dropped to chance level in Experiment 1). For each $\Delta\theta_{within} \in \{5°, 10°, 15°, 20°\}$ we computed three pairs of stimuli (left- and right-pointing triangles in each pair), to cover three basic configurations: (1) parallel configuration at the figure side of the edge, (2) parallel configuration at the ground side of the edge, and (3) a curvature-based edge of $\Delta\kappa_T = \Delta\kappa_N$. Fig. 3.15 illustrates the six stimuli that correspond to $\Delta\theta_{within} = 5°$.

3.4.2 Subjects and procedure

Five subjects from those who participated in Experiment 1 took part in this experiment and all (but one author) were naive to its purpose. All subjects had normal or corrected-to-normal vision and each ran one session of 360 trials, arranged in 10 blocks of 36 trials separated by short breaks. Each session lasted for less than 20 minutes and was preceded by one or two practice blocks of 72 trials. All other parameters in Experiment 2 were similar to those of Experiment 1 although now each stimulus was presented 30 times, 15 for each of the two possible figures. The presentation order was randomized and set before the experiment.

3.4.3 Results

To check for differences in performance between the detection of parallel configurations and those with $\Delta\kappa_T = \Delta\kappa_N$, we averaged the accuracy of detecting the two parallel configurations (cases 1 and 2 in Sec. 3.4.1) and subtracted from it the accuracy of detecting the $\Delta\kappa_T = \Delta\kappa_N$ configuration (case 3 in Sec. 3.4.1). Fig. 3.16 presents the average difference in performance as a function of $\Delta\theta_{within}$. It shows clearly that while parallel configurations are more salient for small $\Delta\theta_{within}$, they are less salient for larger $\Delta\theta_{within}$, where they are outperformed by the $\Delta\kappa_T = \Delta\kappa_N$ configuration.

This change in relative saliency is subtle and occurs gradually, but it was always statistically significant as soon as $\Delta\theta_{within}$ increased by more than 5° ($p < 0.01$ for $\Delta\theta_{within}$ increase of 10° and $p < 0.005$ for $\Delta\theta_{within}$ increase of 15°). The difference in the means for $\Delta\theta_{within} = 10^\circ$ and $\Delta\theta_{within} = 15^\circ$ was already statistically significant ($p < 0.05$) even though the increase in $\Delta\theta_{within}$ was only 5°.

The results of Experiment 2 clearly show that the configuration hypothesis is not generally true even for those parallel configurations which it strictly designates as optimal. It further shows clear dependency on $\Delta\theta_{within}$, in full agreement with Experiment 1 and the conclusions made in Sec. 3.3. These results reinforce Experiment 1 to suggest that OBTS is indeed sensitive to curvature discontinuities when these discontinuities become significant enough (as is possible in ODTs of large $\Delta\theta_{within}$). At this point, any advantage due to the parallel configuration is overwhelmed by

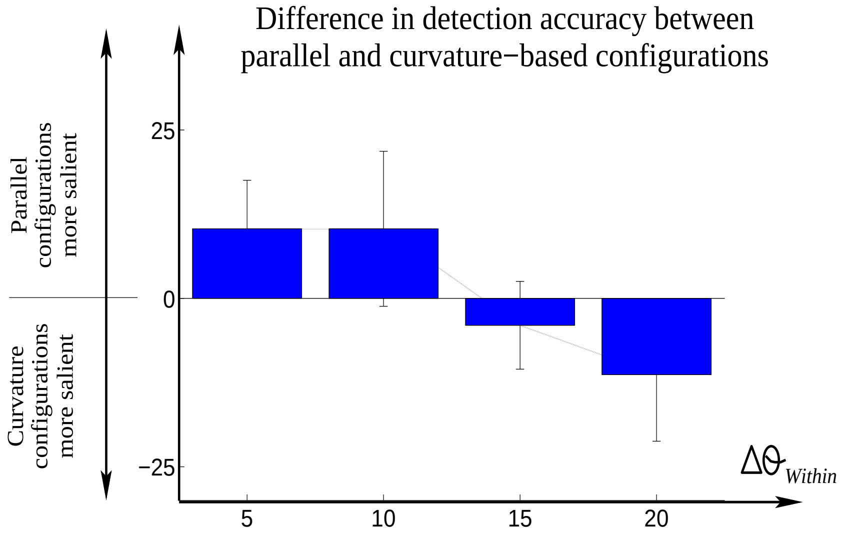


Figure 3.16: **Results of Experiment 2.** Shown here is the average difference in detection accuracy between parallel configurations and $\Delta\kappa_T = \Delta\kappa_N$ ones, plotted against $\Delta\theta_{within}$. Note that there are cases where parallel configurations are not the most salient.

the effect of the curvature discontinuities. Note that these findings are also consistent with Fig 3.14 which predicts that curvature discontinuities are likely to influence OBTS the most when $\Delta\theta_{within}$ is large.

3.5 Conclusions

It was shown in the past that OBTS is influenced by the two orientation gradients $\Delta\theta_{within}$ and $\Delta\theta_{between}$ [Nothdurft, 1991; Mussap and Levi, 1999]. It was further argued that OBTS is independently modulated by particular configural organizations in the neighborhood of the texture boundary [Nothdurft, 1992; Wolfson and Landy, 1995]. In this chapter we have shown that, in addition to the above processes, OBTS is also sensitive to discontinuities (or contrast) in the two texture curvatures $\Delta\kappa_T$ and $\Delta\kappa_N$, and it appears to improve the most when both discontinuities are maximized simultaneously, i.e., when $\Delta\kappa_T = \Delta\kappa_N$. Ch. 2 showed how that the notion of curvature emerges naturally from a geometrical model of ODTs and how it is fundamentally linked to the interaction between positions and orientations. From a mathematical perspective, then, it would have been surprising had an effect of curvature not been found.

The introduction of ODT curvatures into the psychophysical investigation of OBTS carries the advantage of emphasizing the spatially varying nature of ODTs and the observation that constant (or piecewise constant) ODTs are in fact accidental in natural, every-day stimuli. With curvatures, such variations in ODTs can be integrated into stimuli in a fully controlled way, thus allowing a direct link to human perception and performance.

As we discussed in the introduction to this chapter (Sec. 3.1), most classical studies of ODTs ignored the spatially varying nature of ODTs, focusing instead on piecewise constant structures only. Studying configural effects within this limited scope, Wolfson and Landy [Wolfson and Landy, 1995] concluded that OBTS is improved if the orientation of texels is parallel (and to some extent perpendicular) to the texture edge. Indeed, piecewise constant ODTs have $\kappa_T = \kappa_N = 0$, their boundaries emerge solely from orientation discontinuities, and thus saliency considerations involving only boundary configuration can be attractive. As we demonstrated, such arguments also are closely related to the fold and cut organization induced by the interactions of surfaces during the imaging process.

However, in the natural world, most ODTs are spatially varying, which suggests that configural factors should be examined within a larger scope, as we further did in the current study. Indeed, we conclude that there are significant gaps between the classical configuration hypothesis [Nothdurft, 1992; Wolfson and Landy, 1995] and actual performance in OBTS. We have shown that configural effects are neither independent of spatial variations (captured, e.g., by $\Delta\theta_{within}$), nor are they universal: while parallel configurations appear superior in ODTs of small internal variation of $\Delta\theta_{within} \leq 10°$, larger $\Delta\theta_{within}$ results in significant texture curvatures, at which point the advantage due to parallel configurations is superseded by the one due to the curvature discontinuities. Interestingly, the results by Nothdurft [Nothdurft, 1992] may already contain a flavor of this phenomenon since the peak performance he measured deviates up to $30°$ from the strictly parallel configuration (see his Figure 3A-C). Unfortunately, due to a similar deviation in the orientation of the texels themselves relative to the orientation edge, it is impossible to asses whether his data are indicative of something other than the configuration.

The advantage of the curvature-based explanation over the one based on configural modulations is fundamental. The former is based on intrinsic data only and thus can be developed into a predictive computational model. The latter, on the other hand, is undermined by a chicken-and-egg problem; it needs the outcome of the segmentation process (namely, the perceptual edge and its orientation) in order to make predictions about its own occurrence. In this spirit it is instrumental to reexamine Figure 3.4 which we discussed in Sec. 3.1.2 and 3.2.5. Figure 3.17 presents the same stimulus, now with its orientation and curvature functions also depicted as height functions. Clearly, while the orientation gradient across the square's perimeter is constant, the discontinuities in curvatures are not; they are simultaneously large only along the top edge. According to the results of Experiment 1, this edge should be the more salient, a prediction that agrees well with the perceptual evidence. Notably, this prediction does not require the orientation of the texture boundary as an input; it is made using intrinsic and local information only.

Since a main conclusion in this paper is the fact that orientation gradients are insufficient to explain OBTS, it remains to examine our results, both computational and psychophysical, in a larger context. For example, given the central role of *feature gradients* in early vision [Julesz, 1986; Nothdurft, 1993], our results justify the reexamination of feature gradients in preattentive vision more generally. Furthermore, since many other perceptual features (e.g., shading, motion, color) can be represented in terms of orientation, their intrinsic geometry, and segmentation, should be investigated analogously.

Lastly, combining our results with the observation that the complement of segmentation is visual *integration* suggests that limiting the discussion on visual integration to curves only [e.g. Field et al., 1993b] may miss an important functional aspect of biological vision systems. Current "association field" models, popular in psychophysics [Field et al., 1993b; Hess and Field, 1999; Kapadia et al., 2000, 1995], physiology [Schmidt et al., 1997; Bosking et al., 1997; Kapadia et al., 2000, 1995] and computational modeling [Parent and Zucker, 1989; Yen and Finkel, 1998], typically ignore curvature, and never consider *normal* curvature. Thus, the formal analysis and psychophysics outlined in this thesis may provide the first step toward explaining the many findings which are inconsistent with curve integration, both psychophysically and physiologically ([e.g. Matsubara et al., 1985; Ts'o et al., 1986; Polat and Sagi, 1993; Adini et al., 1997; Kisvárday et al., 1997; Sincich and Blasdel, 2001]). We address this issue in greater detail in Chapter 5.

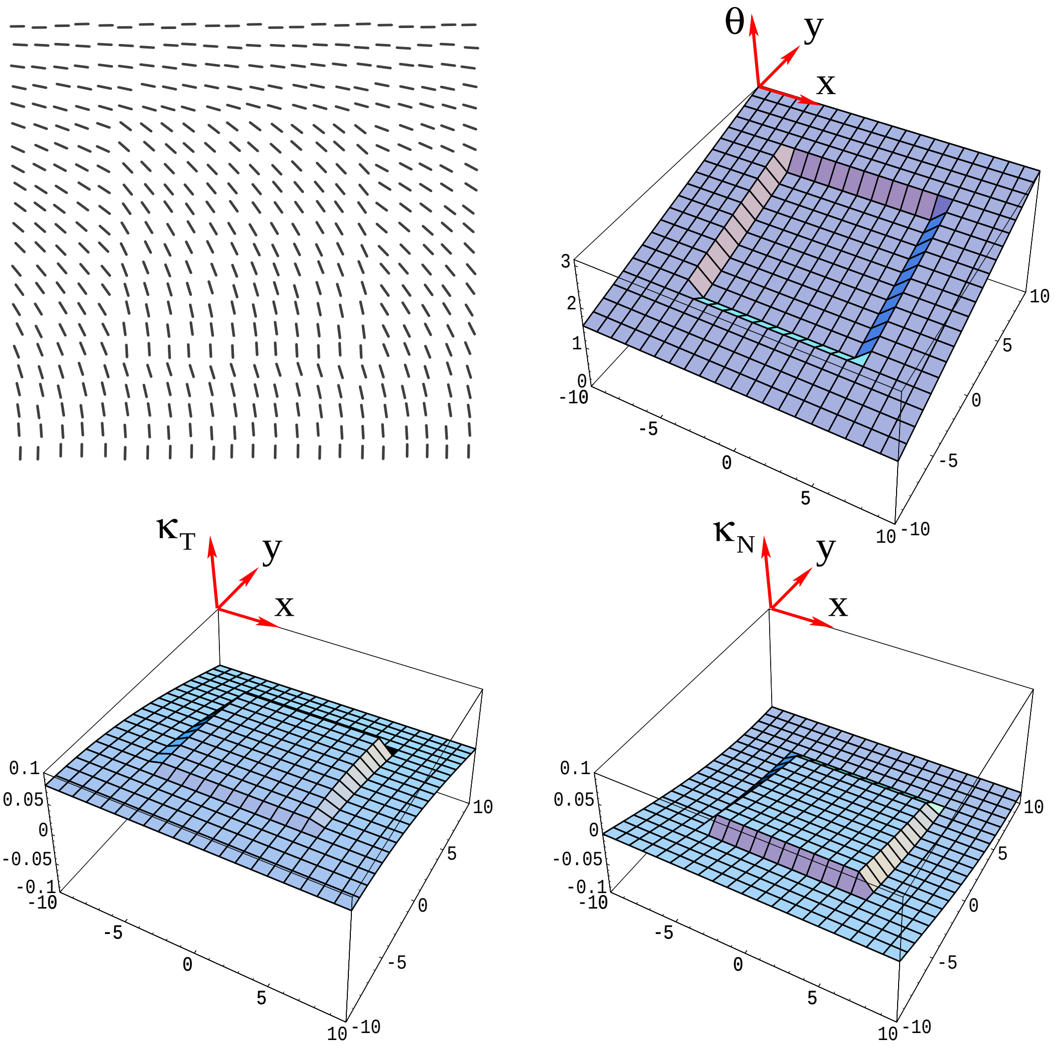


Figure 3.17: **Curvature discontinuities predict the saliency of orientation edges.** This figure shows the stimulus from Figure 3.17 with its orientation and curvature functions depicted as height functions. Although the orientation gradient along the square's perimeter is constant (as seen best in the 3D plot), the jump in the curvatures is not. In particular, while $\Delta\kappa_T$ is virtually zero at the bottom edge, it is large, and roughly equal to $\Delta\kappa_N$, at the top edge. Thus, intrinsic considerations alone can predict not only the location of edges, but their relative saliency as well.

Chapter 4

Computational perceptual organization of visual flows

With the importance of visual flow curvatures established both geometrically and psychophysically, and a model for flow coherence now in place, we now develop these ideas in a computational setting to extract piecewise coherent visual flows from images.

Although the analysis of oriented patterns has been part of computer vision research for almost two decades (see Sec. 1.3), we question the view that it is well understood, or solved (e.g., [Weickert, 1999, pp. 111,114]). In many cases, even the goals of this task have not been fully stated, let alone achieved. For example, it is commonly assumed that the initial data to be analyzed is dense and single valued. Many algorithms, especially those based on diffusion, cannot be applied otherwise. However, in many cases, the set of initial measurements is *not* dense and sometimes it is multi-valued, as is especially evident for textures (e.g., see [Watanabe and Cavanagh, 1996; Keeble et al., 1997], Sec. 1.1.6, or many examples in [Brodatz, 1966; Hornung, 1976]). Additionally, it is possible that the underlying true structure in the stimulus is *not* piecewise smooth across the entire domain, thus violating the common assumption taken, explicitly or implicitly, by all existing methods (Fig. 4.1). In such cases it is more reasonable to segment away the “non-structure” rather than to transform it to some fictitious coherent one. In other words, while it is reasonable to seek separation of noise from the locally oriented signal, it is *not* reasonable to create meaningless locally oriented

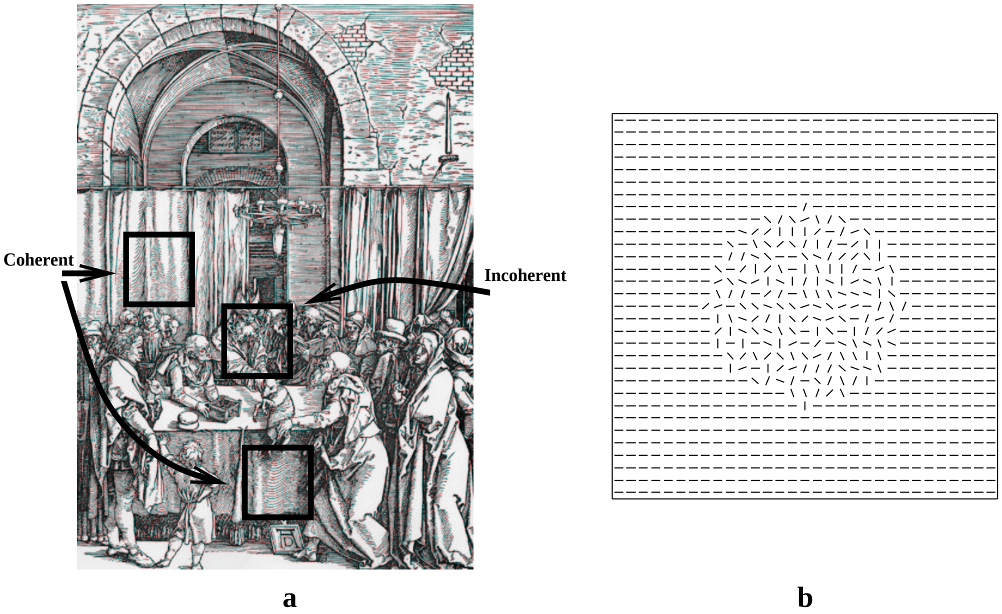


Figure 4.1: **Coherent and incoherent oriented structures in images.** It is common that coherent oriented structures are intermixed with incoherent ones. Thus, the assumption that the measured data assumes a piecewise smooth behavior is not always correct. **(a)** A Dürer woodcut "The high priest rejects Joachim's offering" [Kruth, 1963] with both kind of structures marked. **(b)** An illustrative synthetic example of a set of initial orientation measurements that reflects two regions, one coherent and another incoherent. Any algorithm that attempts to organize locally oriented data into coherent parts must be able to reject incoherent areas instead of creating a meaningless signal from the "noise".

signal from noise. This takes particular significance in the analysis of oriented patterns, since the coherent representation one seeks is not a goal in itself but typically an intermediate step toward other computations [Kass and Witkin, 1987].

In the light of these observations, the computational task in hand can be described as follows:

> **given** a noisy, not necessarily dense, and possibly multi valued field of orientations in the image plane, **compute** a piecewise coherent (smooth), dense, and possibly multi valued representation, while preserving discontinuities, singularities, and the boundaries between coherent and incoherent parts, the latter of which should be rejected.

To our knowledge, no work to date posses all these properties[1].

---

[1] The reader should observe that contrary to contemporary approaches to the analysis of oriented patterns (sec. 1.3.2),

4.1 Visual flows and orientation diffusion

It is instructive to examine the performance of state-of-the-art approaches to the analysis of oriented patterns in the context of our computational task. These methods, collectively known as *orientation diffusion* (see Sec. 1.3.2), are a powerful tool in attempting a particular task: progressively transforming the (originally noisy) set of orientation measurements into a piecewise smooth representation. Their inhomogeneous nature allows them to adaptively damp their influence in areas “suspected” as structural edges and thus to preserve the prominent structure to deeper levels of the scale space. Despite all of their advantages, orientation diffusion methods share a number of limitations in respect to our task. We discuss them next, albeit in no particular order.

By design, diffusion schemes are not designed to deal with stimuli or data sets of *multiple* dominant orientations at a point, nor they are suitable to deal with *missing*, or non-dense sets of measurements, something which is inherent to the analysis of many oriented patterns. For example, consider the textured pattern in Fig. 4.2a, duplicated from Fig. 1.7. While clearly, this should be considered as a multiply oriented pattern, no local measurement scheme will produce a *dense* field of directions from this image, as significant part of it contains no contrast at all. Indeed, while edge detectors, by definition, can capture only a sparse oriented structure for this image (Fig. 4.2c), even the commonly used gradient-based measurement (e.g., as in [Perona, 1998]) fails to yield a dense field, as is depicted by the “holes” in the orientation map in Fig. 4.2d. Of course, these holes correspond to ill-behaved areas of zero gradient, a situation that can be “fixed” with a certain amount of blurring. Unfortunately, by the time the blurring is deep enough to yield a dense field, the underlying structure may undergo qualitative changes, and “the baby is thrown out with the bath water”[2].

---

a scale space representation is *not* part of our task. While a multi scale representation could be an important goal in its own right, we claim that it is secondary in importance in the context of the motivation discussed thus far and do not consider it further.

[2] It should be noted that inpainting methods (e.g., [Bertalmio et al., 2000]) that use diffusion to fill in holes in images are ill suited to handle general non-dense sets of orientation measurements which we are concerned with here. Three reasons contribute to this observation. First, inpainting methods need a dense margin of minimal width around the hole from which information is propagated inward. In general, the distribution of missing measurements may be such that none of them has such a dense margin around it (One extreme example is when the missing measurements are distributed like a chess board). Second, by the very nature of inpainting algorithms, the information around ”holes” is assumed to be correct. In our context, the distribution of noise renders such an assumption void. Third, inpainting techniques require manual initialization of areas defined as ”holes”. This clearly is unfeasible for images like the one in Fig. 4.2.

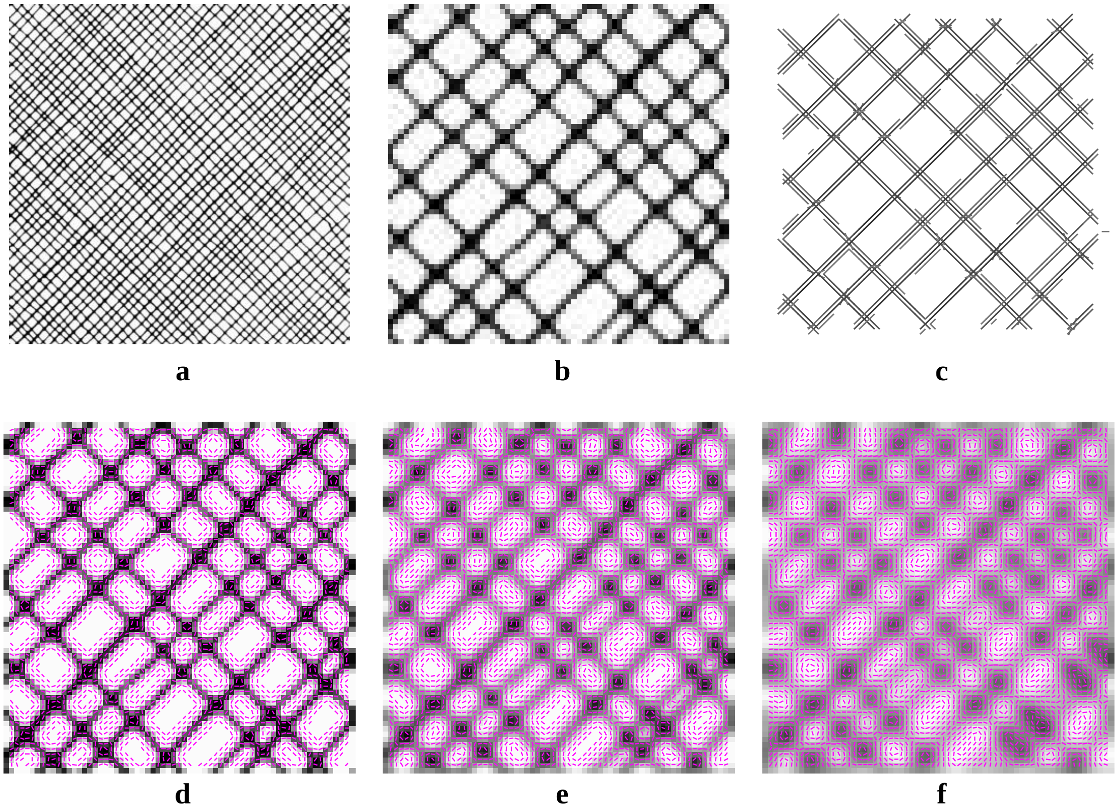


Figure 4.2: **Measurements taken from oriented patterns are not necessarily dense, nor do they always correspond to a single dominant orientation at a point.** Both of these observations stand in contrast to most diffusion formulations to date. **(a)** A multi oriented texture pattern from [Hornung, 1976]. **(b)** A detail. **(c)** A structurally correct, but highly sparse output of an edge detector [Iverson and Zucker, 1995]. Such an orientation map *cannot* serve as initial condition for a diffusion process. **(d)** Even measurements based on the intensity gradient can result in “holes” at areas of zero contrast. **(e)** A better behaved image can be obtain by blurring. Indeed, after blurring the image with a Gaussian ($\sigma = 1.0$, 3x3 kernel), many (but not all) holes now contain directional information. **(f)** Only when the blurring is deep enough ($\sigma = 1.5$, 5x5 kernel) the field of measurements becomes dense. However, by this time, the underlying oriented structure, as captured by the gradient measurements, is completely distorted and ceases to represent the original image.

Another limitation of orientation diffusion approaches is their incapacity to deal "appropriately" with "non-structure", i.e., with orientation fields that present no coherence at all (perceptual or otherwise), such as the one in the central disk of Fig. 4.1b. Applying a diffusion process to such pattern can lead to one of two outcomes. Typically, the pattern will be denoised quickly towards an average value determined by the surrounding pattern (or towards an unpredictable average, if no surrounding data exist). Alternatively, in orientation diffusion schemes that employ an "edge stopping" diffusion (e.g., [Perona, 1998; Tang et al., 1999]), the incoherent pattern may be diffused towards an everywhere high gradient pattern. We argue that neither of these outcomes is desired. Rather than creating a signal that does not represent the data, it is more reasonable to mark and reject those incoherent regions while preserving the boundaries between them and the coherent ones.

A final problem in orientation diffusion relates to its dynamics on a certain class of generic structures. We introduce this issue by first discussing the different configurations of discontinuities one might encounter in orientation fields.

4.1.1 A classification of line discontinuities

Consider a line discontinuity (as opposed to a point singularity) in a piecewise smooth 2D orientation field[3], as illustrated in Fig 4.3a. What are the generic configurations that this kind of discontinuity can take? A more intuitive representation of this question may be done by "lifting" the orientation field representation to $\mathbb{R}^2 \times \mathcal{S}^1$, or $XY\theta$. In this space, each smooth part of the field is represented as a smooth surface patch and the line discontinuity is represented by the way these two surface patches interact. Typically, they will give rise to a step-like configuration. This, however, is not the only possibility.

Assume that the line discontinuity has a regular parameterization $C(s) : [0,1] \mapsto \mathbb{R}^2$ and let $\vec{n}(s)$ be its field of normals. Assuming the curvature $\kappa(s)$ along the curve is small enough, the following two curves

$$C^{+}(s) \triangleq C(s) + \epsilon\vec{n}(s)$$

$$C^{-}(s) \triangleq C(s) - \epsilon\vec{n}(s).$$

---

[3] While here we focus on 2D fields, this analysis can be generalized to higher dimensions where discontinuities of this sort take the form of submanifolds of codimension 1.

are regular and approach $C(s)$ asymptotically as ϵ decreases. We are interested in the profile of the oriented pattern along $C^+(s)$ and $C^-(s)$, i.e., in the two functions

$$\theta^+(s) = \theta(C^+(s))$$

$$\theta^-(s) = \theta(C^-(s)).$$

Since the orientation field is smooth on both sides of the discontinuity, $\theta^+(s)$ and $\theta^-(s)$ are smooth also. Therefore, both functions trace rectifiable curves in $\mathbb{R}^1 \times \mathcal{S}^1$

$$\alpha^+(s) \triangleq (s, \theta^+(s)) \quad \text{and} \quad \alpha^-(s) \triangleq (s, \theta^-(s)).$$

Assume that $\theta^+(0) \neq \theta^-(0)$ and $\theta^+(1) \neq \theta^-(1)$, fix $\theta^+(s)$, and examine the ways with which $\theta^-(s)$ can get from the point $(0, \theta^-(0))$ to $(1, \theta^-(1)$:

1. $\theta^-(s)$ can intersect $\theta^+(s)$ *zero* times (Fig. 4.3b).

2. $\theta^-(s)$ can intersect $\theta^+(s)$ a *countable* number of times, either transversely[Guillemin and Pollack, 1974] or nontransversely (Fig. 4.3c,d).

3. $\theta^-(s)$ can intersect $\theta^+(s)$ *noncountable* number of times (Fig. 4.3e). The smoothness of both curves implies that in this case they intersect nontransversely.

Fig. 4.4 visualizes these three cases both as direction fields in the image plane and as surface patches in $XY\theta$. Inspired by their visual appearance in Figs. 4.3 and 4.4, we name these configurations *step*, *cross*, and *bifurcation*, respectively.

It remains to verify which of the three configurations is interesting from the point of view of vision. While the step configuration is clearly a frequently encountered one, we argue that both the cross and the bifurcation are of special interest to vision also. One example, from the domain of shading (Sec. 1.1.2) is shown in Fig. 4.5. The Caterpillar image in Fig. 4.5a is representative of the class of objects that have curved parts that meet at a surface discontinuity. A synthetic version of such a Lambertian worm-like object is shown in Fig. 4.5b. This particular object is illuminated and viewed from above and its shading flow field (as a level set contour map) is shown in Fig. 4.5c. It can be easily shown that the vertical discontinuity is an orientation cross that corresponds to

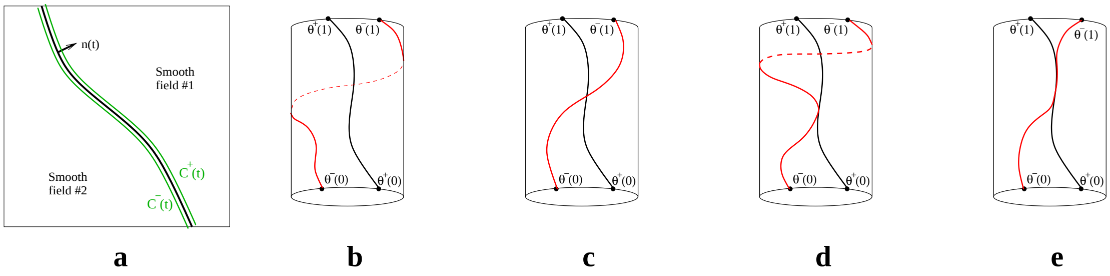


Figure 4.3: **Topological interaction of orientation profiles. (a)** A line discontinuities in 2D orientation fields can be thought of as a curve (submanifold of codimension 1) which separates two regions of smooth fields, each of which has a particular and independent behavior in the proximity of the edge. The oriented data along the two sides of the discontinuity can be depicted as two curves in $\mathbb{R}^1 \times \mathcal{S}^1$ whose mutual interaction can be one of three: **(b)** Zero number of intersections, **(c)** Countable number of (transverse) intersections. **(c)** Countable number of (nontransverse) intersections. **(e)** Noncountable number of (nontransverse) intersection points.

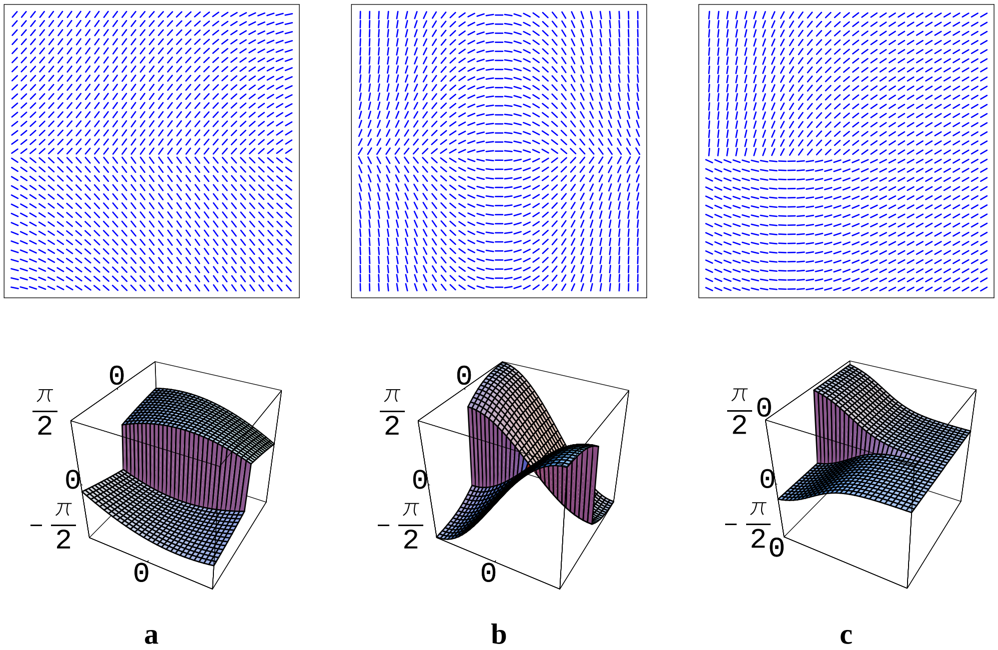


Figure 4.4: **The three possible configurations of line discontinuities. (a)** Step edge. **(b)** Cross edge. **(c)** Bifurcation edge. All configurations are shown both as visual flows and surfaces in $XY\theta$.

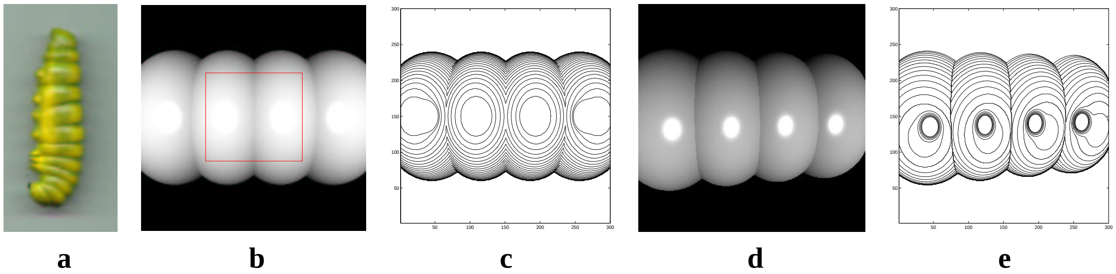


Figure 4.5: **Example of an orientation cross.** **(a)** An prototypical natural object having curved parts meeting at surface discontinuities. **(b)** A synthetic version of the same worm like object illuminated and viewed from above. **(c)** The shading flow field in the marked region depicted as a contours map. Note the orientation cross. **(d)** The same object under different viewing and illumination conditions. **(e)** The shading flow field under the new conditions. Note how the cross configuration is preserved (albeit shifted spatially).

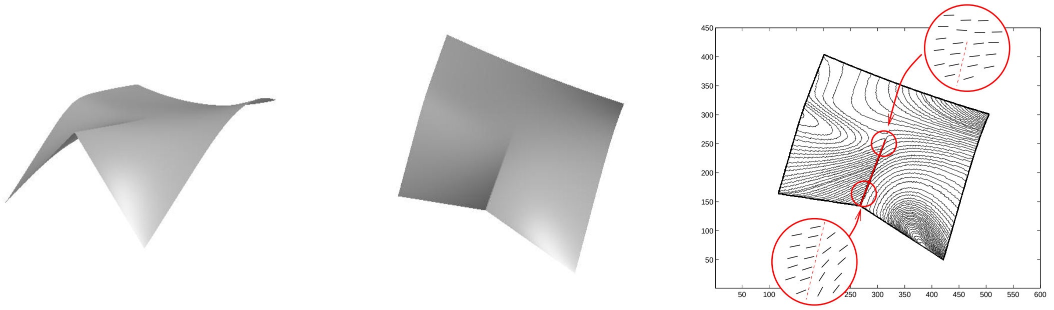
Figs. 4.3c and 4.4b. Furthermore, since the edge is a result of a *surface* discontinuity, rather than two abutting folds (which is accidental), there is an open set V of viewing angles from which this relationship is preserved. Changing the view point smoothly in V gives rise to a smooth change in $\theta^{+}(s)$ and $\theta^{-}(s)$, hence there exists an open set of view points from which the cross configuration is also preserved. The same argument can be applied to perturbations in the light source position and small changes to the shape (as long as the surface discontinuity is kept fixed). In all these cases, $\theta^{+}(s)$ and $\theta^{-}(s)$ undergo a smooth transition, which cannot eliminate the cross[4]. We conclude that orientation crosses do not only happen in visual stimuli, but in fact they are generic structures. Similar arguments apply to shading bifurcations, one example of which is illustrated in Fig. 4.6.

4.1.2 Structure distortion in orientation diffusion

With a better understanding of the line discontinuities that may arise in visual flows, we now argue that orientation diffusion is in fact not equipped to deal with two of the structures that we identified above, namely orientation crosses and bifurcations. As will be demonstrated, this may lead to severe structural distortions in the orientation field. While we illustrate this claim using the orientation diffusion based on harmonic map theory [Tang et al., 2000], similar arguments hold for related formulations also.

---

[4] Note that if the cross corresponds to a nontransverse intersection of $\theta^{+}(s)$ and $\theta^{-}(s)$, a small perturbation *can* result in a qualitative change. Indeed, these cases *are* accidental and only rarely encountered.



a **b** **c**

Figure 4.6: **Example of an orientation bifurcation. (a)** A prototypical object that contains a shading bifurcation due to the interaction between its smooth and discontinuous parts. Shapes like this one are common, for example, in the automobile industry. **(b)** Another view of the same object, this time without occlusion. **(c)** The intensity level sets of the object with two details along the bifurcation.

Following [Perona and Malik, 1990] and the main stream in diffusion literature, a key aspect in anisotropic orientation diffusion is that it can be interpreted as a variation of isotropic diffusion "weighted" by an "edge stopping", or adaptive diffusivity function [Tang et al., 2000]. The process is consequently characterized in general by the following PDE in divergence form

$$I_t = \nabla \cdot (w(||\nabla I||)\nabla I)$$

where $w(x)$ is the "edge stopping" function and I is the diffused quantity. A desired robust behavior is achieved when $w(x)$ is mostly monotonically decreasing as $|x|$ increases (see also [Black et al., 1998]). This induces a fast diffusion over noisy but "mostly smooth" regions, and suppresses the diffusion across large scale discontinuities where $||\nabla I||$ is large, i.e., across edges.

Indeed, such a behavior is imposed by the diffusion process in Tang et al. [2000]

$$\frac{\partial I_i}{\partial t} = \nabla \cdot (||\nabla I||^{p-2} \nabla I_i) + I_i ||\nabla I||^p \quad ,$$

where I_i is the i's component of the unit vector representation of the oriented data and p is the metric of the harmonic energy to be extremized. Note that in this case

$$w(||\nabla I||) \triangleq ||\nabla I||^{p-2}$$

which for $p < 2$ posses exactly the desired monotonicity properties. Indeed, this works well for step edges. The problem arises in recognizing that global structure discontinuities may not necessarily exhibit the local "step" behavior all along. In fact, there are cases where an orientation discontinuity ceases to behave as a step (regardless of the spatial scale considered) and appears more "noise-like" in terms of the deviation from a smooth surface. As was discussed above, this is exactly what happens in the proximity of orientation crosses and orientation bifurcations, where the deviation from a smooth orientation field vanishes gradually. To an anisotropic diffusion process, this area should be smoothed out as fast as the "edge stopping" function permits it.

Two problems emerge. First, it is likely that by the time the diffusion achieves its primary goal and eliminates noise from the data set, it will also induce a "closing" of the cross or bifurcation. Figure 4.7 illustrates this behavior with the harmonic map diffusion of Tang et al. [2000]. For illustration purposes we show the data both as fields of directions in the image plane, and as a surface in $XY\theta$. Note how the bifurcation quickly collapses up to a total elimination of the discontinuity. This stands in clear contrast to the stable behavior in the proximity of the orientation step. Figure 4.8 illustrates the same undesired behavior with the diffusions proposed in [Chan and Shen, 1999] and Kimmel and Sochen [2000].

Second, if two or more such processes are developed in the same region, or when other singular structures are present, they all might interact in an unpredicted way to induce significant global structural distortions - either line discontinuities or new singularities. For example, consider again the Lambertian worm-like object in Fig. 4.5b. Its shading flow field around the surface discontinuity is illustrated in Fig. 4.9, although now it is depicted as a field of directions (as opposed to contour map). This field has one cross, two (asymptotic) bifurcations, and two singularities, all in close proximity. As each of the cross and bifurcations collapses slowly, they interact and give rise to a new set of discontinuities that are not present in the data. Not only does this misrepresent the raw data, but it also violates the desired scale space principle of *causality* which has been a driving force behind the use of scale space analysis in general [Witkin, 1983], and behind orientation analysis in particular [Perona, 1998, p. 466].

To summarize, although orientation diffusion is very effective when the noisy signal is dense, piecewise smooth, single valued, and all its discontinuities are orientation *steps*, it suffers from a few important limitations. As we have argued, orientation diffusion is not designed to deal with non dense or multi valued orientation fields; it is likely to create fictitious structure from incoherent

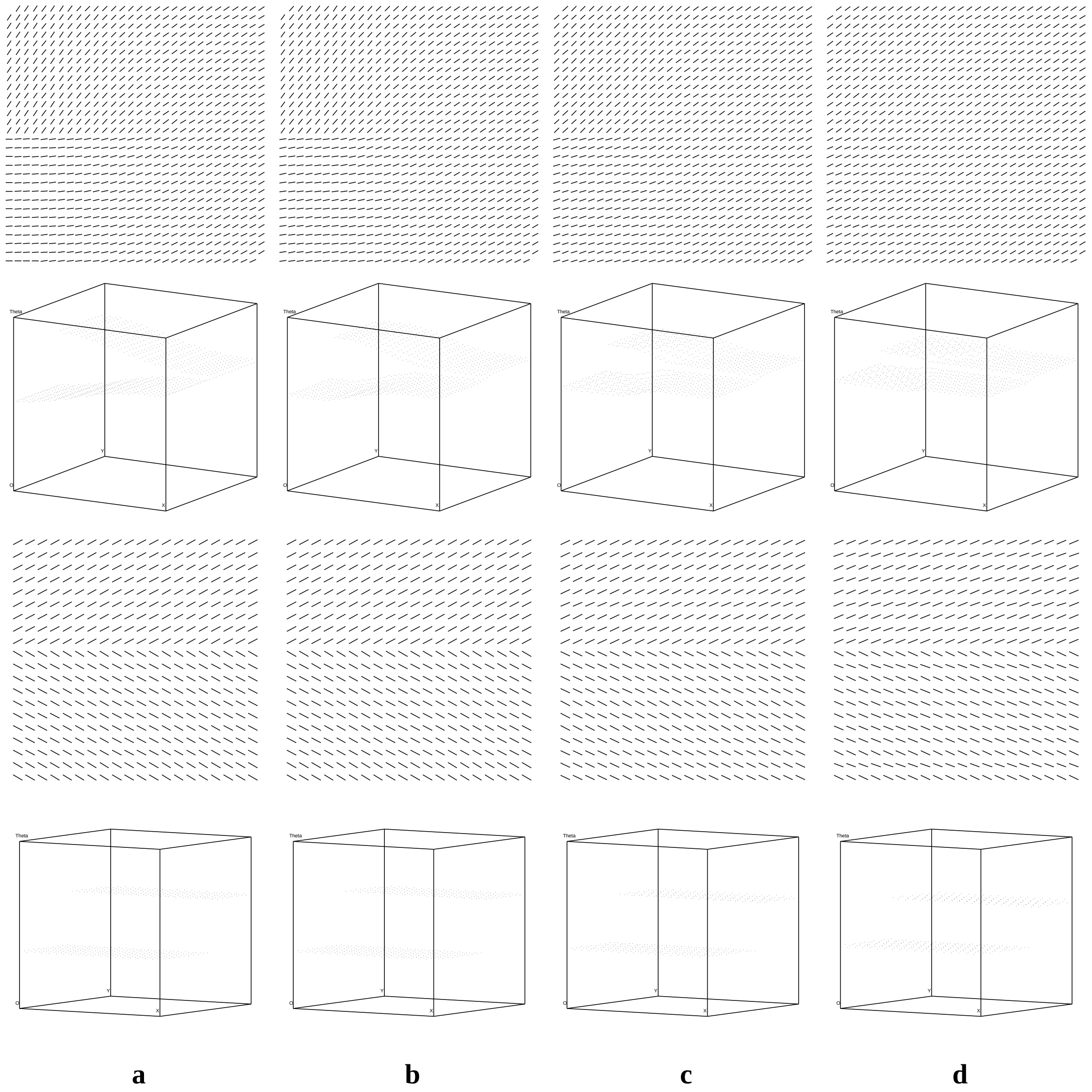


Figure 4.7: **Collapse of an orientation bifurcation under orientation diffusion.** Demonstrated here is the orientation diffusion due to Tang *et al.* [Tang et al., 2000], with $P = 1$ and step size= 0.002. For comparison, we illustrate the same number of iterations on a step discontinuity of the same magnitude. It is evident that unlike a step, a bifurcation is not able to maintain its structure for a long time. **(a)** Original oriented pattern. **(b)** After 500 iterations. **(c)** After 1500 iterations. **(d)** After 2500 iterations.

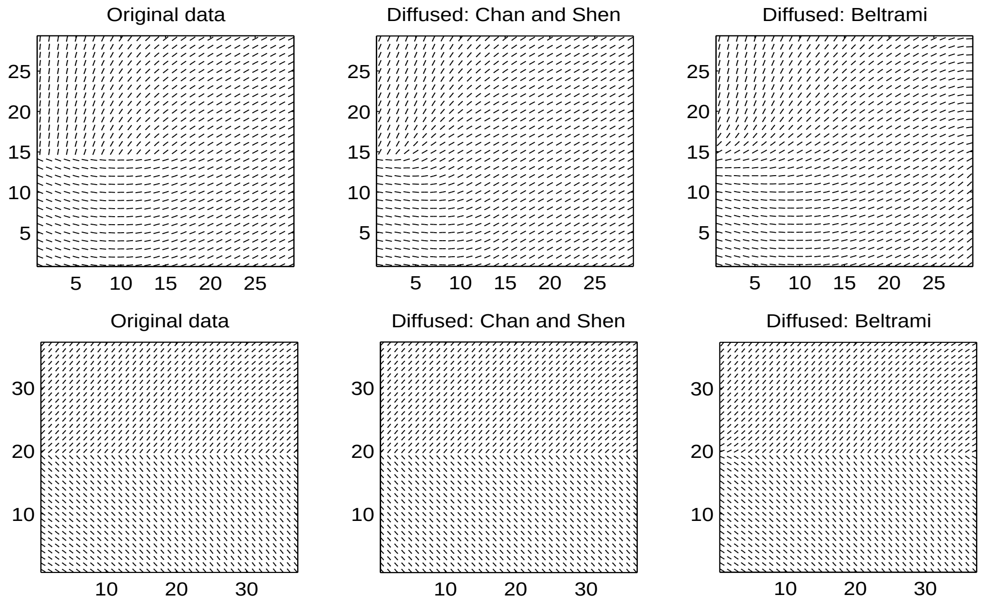
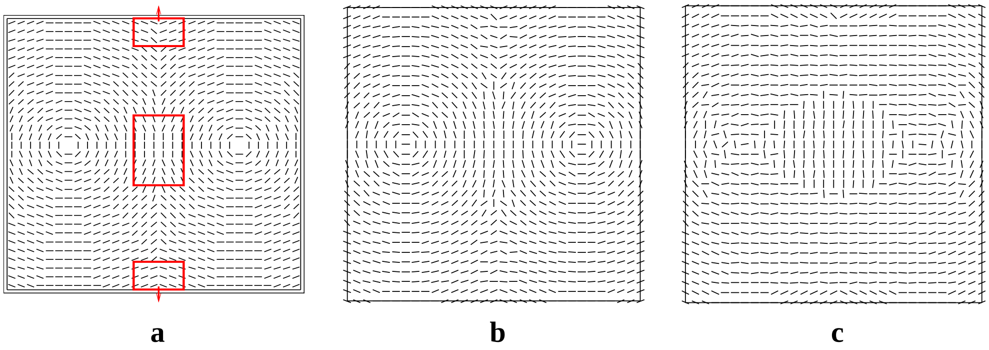


Figure 4.8: **Collapse of an orientation bifurcation under other orientation diffusion methods.** Demonstrated here are the methods due to Chan and Shen [Chan and Shen, 1999] and Kimmel and Sochen Kimmel and Sochen [2000]). For comparison, the stable behavior on a step discontinuity (under the same number of iterations and step size) is illustrated at the bottom row.



a **b** **c**

Figure 4.9: **Example of a qualitative structure distortion in orientation diffusion. (a)** The shading flow field around the central surface discontinuity of the worm like object in Fig. 4.5b (it is the field of tangents to the isoluminance contour map in Fig. 4.5c). Three areas are highlighted. The central one is an orientation cross, and the other two are orientation bifurcations (albeit they occur at infinity). Note also the two point singularities. **(b)** The result of 50 iterations of anisotropic diffusion (Tang *et al.* [Tang et al., 2000], $P = 1$, step size=0.1). Note how each of the three areas start to collapse and a new set of line discontinuities begin to emerge. **(c)** The result of 100 iterations, at which the structure distortion dominates the pattern.

inputs, and it is prone to qualitatively distort the underlying structure around orientation crosses and bifurcations. In the rest of this chapter we present an alternative computational framework that addresses all these concerns.

4.2 Perceptual organization via contextual inference

Unlike existing approaches to the analysis of visual flows, ours is motivated by perceptual organization, and is based on asking “when are nearby visual flow measurements parts of a single whole” (Figure 2.4). Chapter 2 addressed two aspects of this question by deriving the local measurements (two curvatures, in addition to position and orientation) and a model of local coherence (the right helicoid). The advantage of having such a model lies in the ability to assess the degree to which a particular measurement is compatible, or consistent, with the context in which it is embedded. A natural framework with which one can pursue this task while maximizing the average consistency over a domain of interest is *relaxation labeling*. We developed such a relaxation network for the organization of coherent visual flows and derived the compatibility function which governs its behavior from the right helicoidal model. This chapter describes this system in detail.

4.2.1 Relaxation labeling

Following early studies on contextual constraints in the interpretation of line drawings [Waltz, 1975], it has become widely acknowledged that interpretation of sensory data is highly unreliable unless carried out in some non-local context. Relaxation labeling [Hummel and Zucker, 1983; Kittler and Illingworh, 1985] is a formal computational framework for doing exactly that. Closely related to popular models of neural networks [Hopfield and Tank, 1985] and polymatrix games Miller and Zucker [1999] , it involves representing the interpretation problem as the assignment of labels to nodes in a graph whose edges represent the contextual structure.

Let $I = \{i \mid i = 1..n\}$ be a set of nodes, each of which may take any label λ from the set Λ. Let $p_i(\lambda)$ denote the probability, or confidence, in the assignment of label λ to node i. In general, this measure needs to satisfy two constraints:

$$p_i(\lambda) \geq 0 \quad \forall i, \lambda \qquad \text{and} \qquad \sum_{\lambda \in \Lambda} p_i(\lambda) = 1 \quad \forall i. \tag{4.1}$$

Of all possible assignments (henceforth denoted as the space $\mathbb{K}$), those which satisfy $p_i(\lambda) \in \{0, 1\} \quad \forall i \in I, \forall \lambda \in \Lambda$ are called *unambiguous* since they assign a unique label to each node. The role of relaxation labeling is to start from a given, typically ambiguous labeling assignment, and iteratively change it toward a “better” one, where better refers to the degrees of ambiguity and consistency of the assignment of labels at different nodes.

The fundamental mechanism by which contextual information is propagated in a relaxation labeling network is a *compatibility function* $r_{ij}(\lambda, \lambda')$ which quantifies the contextual information conveyed by label λ' at node j about label λ at node i. In conjunction with the confidence measure p_i, the *contextual support* that a label λ at node i receives from its neighborhood is defined (almost exclusively) to be:

$$s_i(\lambda) = s(i, \lambda; \bar{r}, \bar{p}) = \sum_{j=1}^{n} \sum_{\lambda'=1}^{m} r_{ij}(\lambda, \lambda') p_j(\lambda'), \tag{4.2}$$

which can be viewed as a sum of all compatibilities weighted by the confidences. In the same spirit, the supports can be averaged to yield a scalar measure of consistency across the entire network, or what is traditionally called the *average local consistency*

$$A(\bar{p}) = \sum_i \sum_\lambda p_i(\lambda) s_i(\lambda) = \sum_i \sum_\lambda \sum_j \sum_{\lambda'} p_i(\lambda) r_{ij}(\lambda, \lambda') p_j(\lambda'). \tag{4.3}$$

Finally, it remains to clarify when an assignment $\bar{p}$ is *consistent*. Intuitively, this happens when the labels assigned at each node maximally agree with their context. Had we dealt with unambiguous assignments only, this could have been expressed as a maximization of support at all nodes simultaneously, i.e.,

$$s_i(\lambda) \geq s_i(\lambda') \quad \forall i \in I.$$

For ambiguous assignments, the same idea is expressed via the following set of inequalities:

$$\sum_\lambda p_i(\lambda) s_i(\lambda; \bar{p}) \geq \sum_\lambda \tilde{p}_i(\lambda) s_i(\lambda; \bar{p}) \quad \forall \tilde{p}_i \in \mathbb{K}, \quad \forall i \in I.$$

Using all this machinery, relaxation labeling maps inconsistent labeling to consistent ones via an iterative process. Here we follow the algorithm by Hummel and Zucker [Hummel and Zucker, 1983] which is described by the following confidence update rule:

$$p_i^{t+1}(\lambda) \leftarrow \Pi_{\mathbb{K}} \left[p_i^t(\lambda) + \delta s_i^t(\lambda) \right] , \tag{4.4}$$

where $\Pi_{\mathbb{K}}$ is a projection operator that projects its argument onto $\mathbb{K}$, and δ is a constant step size. A fundamental result from the theory of relaxation labeling relates it to polymatrix game theory, and states that this algorithm converges to consistent labeling [Hummel and Zucker, 1983] related to the Nash equilibria of the polymatrix game [Miller and Zucker, 1999]. Furthermore, if the compatibility function is symmetric $r_{ij}(\lambda, \lambda') = r_{ji}(\lambda', \lambda)$, the update rule constitutes a gradient ascent on $A(\bar{p})$ that maximizes it locally [Hummel and Zucker, 1983]. Similar result for the original ad hoc rule of Rosenfeld et al. [1976]

$$p_i^{t+1}(\lambda) \leftarrow \frac{p_i^t(\lambda) s_i^t(\lambda)}{\sum_{\lambda'} p_i^t(\lambda') s_i^t(\lambda')}$$

were obtained later by Pelillo [Pelillo, 1997].

4.2.2 A relaxation network for visual flow

A direct abstraction of the relaxation process for visual flow should involve a 2D image-based network of nodes $i = (x, y)$ (i.e., pixels) whose labels are drawn from the set

$$\Lambda = \{(\theta, \kappa_T, \kappa_N) \mid \theta \in (-\frac{\pi}{2}, \frac{\pi}{2}] \ , \ \kappa_T, \kappa_N \in [K_{min}, K_{max}]\} \ \cup \ \{\text{no-flow}\}$$

after it has been quantized appropriately. Since labels at each node compete with each other, this network structure is appropriate if no multiple orientations are allowed at any given node, e.g., as is the case with shading or hue flows. To allow multiple orientations at a point, and motivated by the columnar architecture of the visual cortex [Hubel and Wiesel, 1977], we replace this 2D abstraction with a 5D network of nodes $i = (x, y, \theta, \kappa_T, \kappa_N)$ whose labels are either T ($TRUE$) or F ($FALSE$). This two-label paradigm was first suggested by Parent and Zucker [Parent and Zucker, 1989] for their curve inference relaxation labeling algorithm. For each node $i, p_i(T)$ denotes the confidence that a visual flow of orientation θ and curvatures κ_T, κ_N passes through pixel (x, y).

Since by definition of the confidence measure $p_i(F) = 1 - p_i(T)$, we need to maintain and update the confidence of only one label at a node.

Under the two-label relaxation labeling, and providing we make one more intuitive design decision, most of the machinery reviewed in Section 4.2.1 above takes a particularly convenient form. In particular, we impose the following behavior on the compatibility structure between any two nodes:

$$\begin{aligned} r_{ij}(F,T) &= - \; r_{ij}(T,T) \\ r_{ij}(F,F) &= +\beta \, r_{ij}(T,T) \\ r_{ij}(T,F) &= -\beta \, r_{ij}(T,T), \end{aligned} \tag{4.5}$$

where $0 \leq \beta \leq 1$ is a constant. These constraints formalize the intuition that (1) if an evidence at node j is *compatible* with a flow at node i, it is equally *incompatible* with the *lack* of flow at node i, and that (2) the lack of evidence is β times less informative than its existence. With these general guidelines the support function at each node now becomes

$$s_i(T) = \sum_{j=1}^{n} r_{ij}(T,T) \left[(1+\beta)p_j(T) - \beta\right] = -s_i(F).$$

The relationship $s_i(T) = -s_i(F)$ is particularly useful from the point of view of the gradient ascent rule (Eq. 4.4). It implies that except for the boundary of $\mathbb{K}$, the gradient ascent procedure stays always inside the space and thus requires no projection. In other words, the projection operation $\Pi_{\mathbb{K}}$, which otherwise may become much more complicated [Mohammed et al., 1983], reduces to

$$\Pi_{\mathbb{K}}(x) = \Pi_0^1(x) \stackrel{\triangle}{=} \min(1, \max(0, x)). \tag{4.6}$$

As a result of the two-label paradigm and the compatibility constraints (4.5), only quantities that depend solely on the T label need to be represented explicitly. Thus we use p_i,s_i, and r_{ij} to refer to $p_i(T)$,$s_i(T)$, and $r_{ij}(T,T)$, respectively, and rewrite the relaxation rule (Eq. 4.4) as

$$p_i^{t+1} \leftarrow \Pi_0^1 \left[p_i^t + \delta s_i^t\right] = \Pi_0^1 \left[p_i^t + \delta \sum_j r_{ij}((1+\beta)p_j^t - \beta)\right] \tag{4.7}$$

with t being the iteration number. Note that choosing $\beta = 0$ to reflect the assignment of zero information to lack of evidence further simplifies the relaxation rule to

$$p_i^{t+1} \leftarrow \Pi_0^1 \left[p_i^t + \delta \sum_j r_{ij} p_j^t\right]. \tag{4.8}$$

4.2.3 Discrete geometric compatibilities for visual flow

It remains to specify a discrete compatibility function r_{ij} in order to complete the design of the relaxation network. Following Chapter 2, we derive these compatibilities from the right helicoidal model, with an emphasis on two issues; the effects of quantization and boundary stability.

Measurement quantization dictates that every possible node i represents an *equivalence class* C_i of measurements, each of which induces a helicoidal field of compatible labels in the neighborhood of i. In the continuum, the union of all these helicoidal fields that correspond to points in C_i forms a consistent 5D "volume" V_i. After quantization, this volume results in a set of consistent, and thus excitatory labels. It is important to mention that since each node i corresponds to a different set of position, orientation, and curvatures, the shape of the 5D consistency volume V_i around i will vary with i.

While the consistency volume V_i determines *which* of the nodes are compatible with a given node i, it does not determine the *values* of r_{ij}, which represent the likelihood that node j shares the same local model (i.e., a helicoid) with node i. Thus, we set r_{ij} to be the probability that C_j intersects the helicoid of a randomly selected point in C_i. In other words, if P_i is a random variable uniformly distributed in C_i, and $H(P_i)$ is the helicoid associated with it, we set

$$r_{ij} = \text{Prob}\left[H(P_i) \cap C_j \neq \emptyset\right].$$

In practice, we compute these values directly by sampling. We also surround the excitatory region with an inhibitory surround to compensate for the lack of intrinsic normalization of labels between all the nodes in the 5D network that share the same spatial position, and thus help localizing its response in the orientation domain. Example of one such compatibility field, for a particular value of K_T and K_N, is illustrated in Fig. 4.10.

4.2.4 Support normalization for boundary stability

Visual flow *line discontinuities* are curves in the image plane along which the flow undergoes a discontinuity or termination (Sec. 4.1.1). As we argued, an important part of the computational task is to preserve these structures. Fortunately, due to the high dimensionality of the space in which

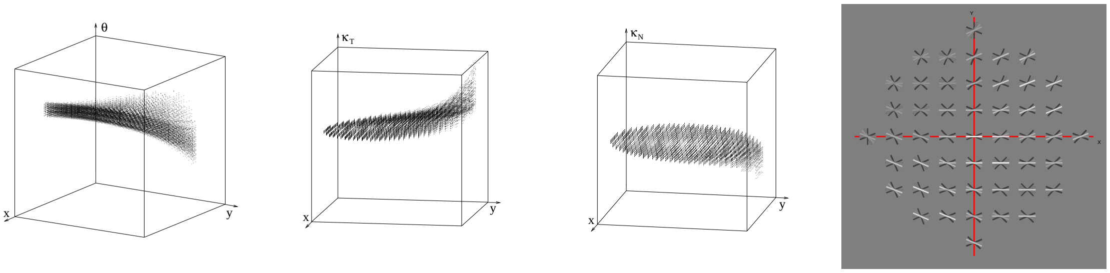


Figure 4.10: **Example of a 5D helicoidal consistency volume.** Shown are projections of the volume on the 3D subspaces $XY\theta$, $XY\kappa_T$, and $XY\kappa_N$ for one particular measurement of $\theta = 0$, $\kappa_T = 0.3$ and $\kappa_N = -0.1$. The 2D image depicts the same object but now as oriented segments on the image plane. Brightness represents degree of compatibility and the inhibitory surround is shown in black (see text).

the labeling process operates, most line discontinuities, either in orientation or curvature, can be viewed as visual flow boundaries. This is simply because in the 5D assignment space, each of the structures that meet along a discontinuity lies outside the context of its counterpart, and thus both are mutually uninfluential. Consequently, what the relaxation process is required to exhibit is stability along *boundaries* in the assignment space.

One way to achieve such stability is to introduce non-linearities to the currently linear support function $s_i(\lambda)$. This may allow for great flexibility in the type of discontinuities that can be handled, but it may significantly complicate the design and it may also affect the applicability of the fundamental results of relaxation labeling theory. Another approach, which is attractive mostly for curved discontinuities, is to endow the assignment space with labels that correspond to the discontinuity itself, e.g., its local orientation and curvature. This, however, has the disadvantage of further increasing the dimensionality of the assignment space and raising the computation time significantly. To stay within the simpler but better understood framework of linear support (Eq. 4.2) and to avoid an assignment space of dimensionality higher than 5, we equip our network with some degree of boundary stability by considering the minimum and maximum support that is expected from a coherent context and then normalizing the support function accordingly.

Given the compatibility volume V_i which corresponds to a particular visual flow node i, one can compute the integral of the compatibility coefficients assuming that a single, full confidence flow traverses through V_i. We denote it by s_{max} (Fig. 4.11). It is clear that the closer i is to a discontinuity

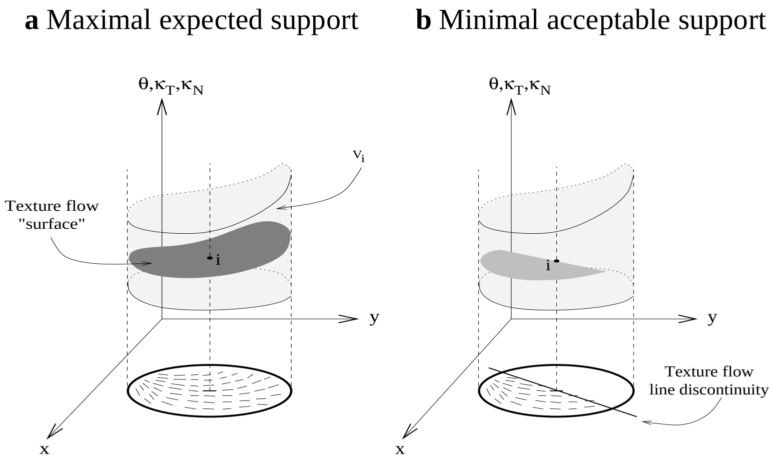

a Maximal expected support **b** Minimal acceptable support

Figure 4.11: **Stability to boundaries in the relaxation process.** Practical stability of the relaxation labeling process to line discontinuities in the flow can be achieved though the normalization of the support function from the interval $[s_{min}, s_{max}]$ to $[0, 1]$. **(a)** At each node i, s_{max} is determined by integrating the support gathered from a full confidence, compatible flow that transverses the entire consistency volume V_i. **(b)** The minimal accepted support s_{min} reflects a flow of some minimally accepted confidence $\rho_{min} < 1$ (depicted here by the brighter surface intensity) that terminates along a line that intersects i.

or a boundary in the data, the less context supports it. At the boundary, the process should remain stable, i.e., neither grow nor shrink. Thus, one can define the level of support for which no change in confidence is desired. We denote this level by s_{min} and observe that it depends on both the geometry of the discontinuity and the minimally accepted confidence of the existing structure. Here we assume the simplest model of discontinuity, namely one that *locally* occurs along a straight line. The support one can gather from such a structure of minimally accepted average confidence ρ_{min} (Fig. 4.11) can then be approximated by

$$s_{min} = \Delta_{min} \cdot \rho_{min} \cdot s_{max} \tag{4.9}$$

where normally, Δ_{min} would be set to 0.5 to indicate that only half of the structure is expected around the boundary, and ρ_{min} would be set to 0.5, which is the minimal confidence that cannot be disambiguated as the $TRUE$ label.

In the context of the two-label relaxation labeling paradigm and the gradient ascent update rule (Eq. 4.8), decrease in the confidence of a label occurs only if $s_i < 0$. Since we seek stability for all supports in the range $[s_{min}, s_{max}]$, it remains to normalize the support values by mapping this interval to the unit interval $[0, 1]$ before submitting it to the update rule. This can be done via the transformation

$$s_i \leftarrow \frac{s_i - s_{min}}{s_{max} - s_{min}} \quad . \tag{4.10}$$

4.3 Experimental results

We tested the proposed geometric model and relaxation network on a variety of inputs, both synthetic and natural. In all cases we defined the network and helicoidal compatibility fields to be π-periodic with 8 equivalence classes, and to support curvatures in the range $[-0.2, 0.2]$ with 5 equivalence classes. Ambiguity of the $TRUE$ and $FALSE$ labels at each node was resolved by confidence comparison at the end of the relaxation process. If the final distribution of labels contained multiple *adjacent* orientation responses (e.g., both $0°$ and $22.5°$) within the same orientation column, these were averaged vectorially according to their confidence measure to yield an unambiguous assignment. Multiple responses separated by at least one orientation class were always kept separate in the final representation.

The intrinsically distributed nature of the network computation suggest executing it on a massively parallel hardware. For example, a 128x128 image, 8 orientation classes, and 5 curvature classes, requires a network of 3,276,800 nodes, all running in parallel. The experimental results presented here were obtained on a serial computer, though, which affected the 5D relaxation runtimes considerably. Our system was implemented in C and executed under the Linux operating system (Redhat 7.X) on a 860MHz Pentium III desktop PC equipped with 128 MBytes of RAM. While the typical (serial) computation time consumed by each node was less than 1 millisecond per iteration, the total iteration time on a typical image size of 128x128 pixels was approximately one hour. The computation on the much smaller synthetic flows (35x35 pixels) consumed approximately 5 minutes per iteration. The results of these computations are discussed next.

4.3.1 Synthetic Visual flows

Fig. 4.12 shows an example of a synthetic visual flow composed of two parts. For such flows we computed exact initial measurements, corrupted them along the orientation and curvature dimensions with additive and/or salt-and-pepper noise, and then quantized the noisy signal to fit the quantization level of the network[5] and to reflect the possible flow labels that the relaxation labeling process can manipulate. Although the pattern in this figure is mostly continuous in terms of orientation, there is a vertical line of discontinuity in the tangential curvature. Such discontinuities can occur, for example, in shading flow fields where two parts of different surface curvatures meet smoothly. Thus, the preservation of these discontinuities is crucial for accurate shape from shading analysis and part decomposition. As can be seen in the figure, the relaxation labeling process is able to eliminate the noise, to fill in holes, and to converge to a globally coherent structure while preserving the discontinuity. Fig. 4.12 also provides a visualization of the change in the distribution of labels in the 5D assignment space (shown by its 3D projections). For lack of space we include this kind of visualization for one case only.

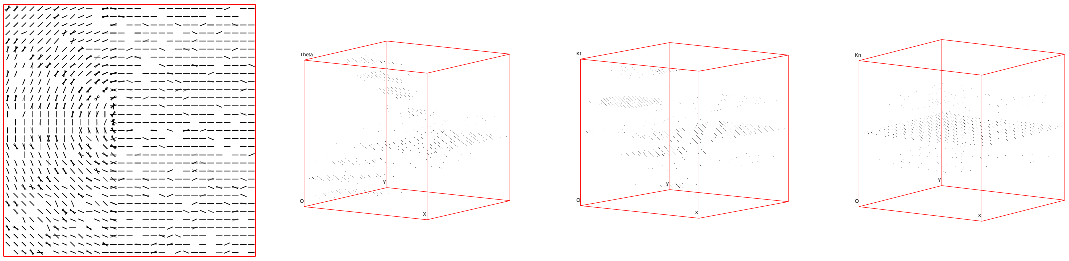
Fig. 4.13 illustrates the result of the relaxation process (20 iterations, $\delta = 0.25$) on other synthetic visual flows. These particular data sets were selected to demonstrate the different properties of our system, namely (1) the ability to handle missing measurements and “holes”, (2) the preservation of discontinuities and stability at boundaries, (3) the rejection of non-flow structures, and (4) the capacity to process overlapping or multi-oriented flows. Most of these counter the inherent limitations of orientation diffusion, as we discussed in Section 4.1.

4.3.2 Robustness in the presence of noise

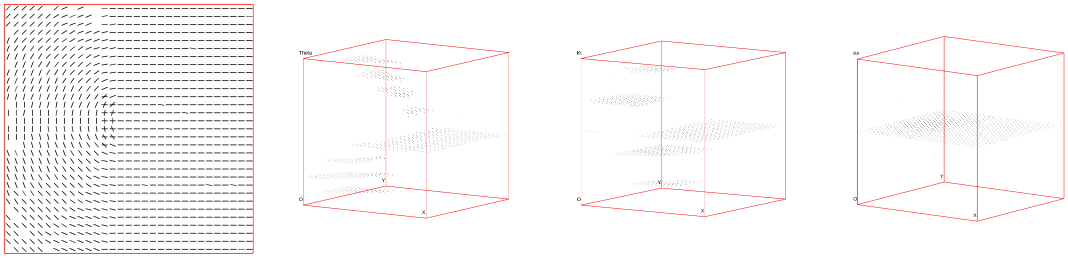
Any computational method that is designed to extract coherent signal from incoherent measurements must be robust in the presence of noise. To empirically examine this aspect of our approach we conducted an exhaustive test by progressively adding noise to a known (constant) flow and checking the behavior of the relaxation network for a fixed number of iterations. The results of this experiment are presented in Fig. 4.14.

---

[5] All of the examples illustrate the noisy signals *after* they were quantized.



(a) Initial $\theta(x, y)$ **(b)** Initial $(x, y, \theta(x, y))$ **(c)** Initial $(x, y, \kappa_T(x, y))$ **(d)** Initial $(x, y, \kappa_N(x, y))$



(e) Relaxed $\theta(x, y)$ **(f)** Relaxed $(x, y, \theta(x, y))$ **(g)** Relaxed $(x, y, \kappa_T(x, y))$ **(h)** Relaxed $(x, y, \kappa_N(x, y))$

Figure 4.12: **Organization of synthetic visual flow.** **(a)** A noisy synthetic visual flow composed of two parts. This structure reflects the shading flow on a bullet-shaped object with a parabolic and elliptic (spherical) parts. **(b-d)** Initial distribution of orientation and curvatures measurements. Note that the Z axes represents orientation, tangential curvature, and normal curvature, respectively. **(e-h)** Relaxed distributions ($\delta = 0.25$, 10 iterations). Missing measurements were filled in while noisy measurements were filtered out, making the entire structure globally coherent while keeping its discontinuity intact.

Panel **a** of this figure illustrates the results of 20 relaxation iterations ($\delta = 0.5$, $\Delta_{min} = 0.3$) on progressively noisier versions of the same signal. The system is able to reconstruct the original flow even for extremely high levels of noise, beyond which it converges to a new final state that represents the rejection of the structure altogether. This behavior, and the dynamics of the process, is best illustrated through the changes in the average local consistency (panel **b**). Note how for most noise levels, $A(\bar{p})$ quickly converges to a certain value that represents the original signal. Beyond a certain point (MSE≥ 0.569), it converges to a different value which represents the lack of flow everywhere. These two noise regions surround a narrow range of noise levels for which the behavior is unstable, i.e., the results are unpredictable and very sensitive to the initial conditions. This indeed was verified experimentally by repeating the test for a large number of trials (75 runs of 20 iterations each), and examining both the average and variance of $A(\bar{p})$. This examination reveals

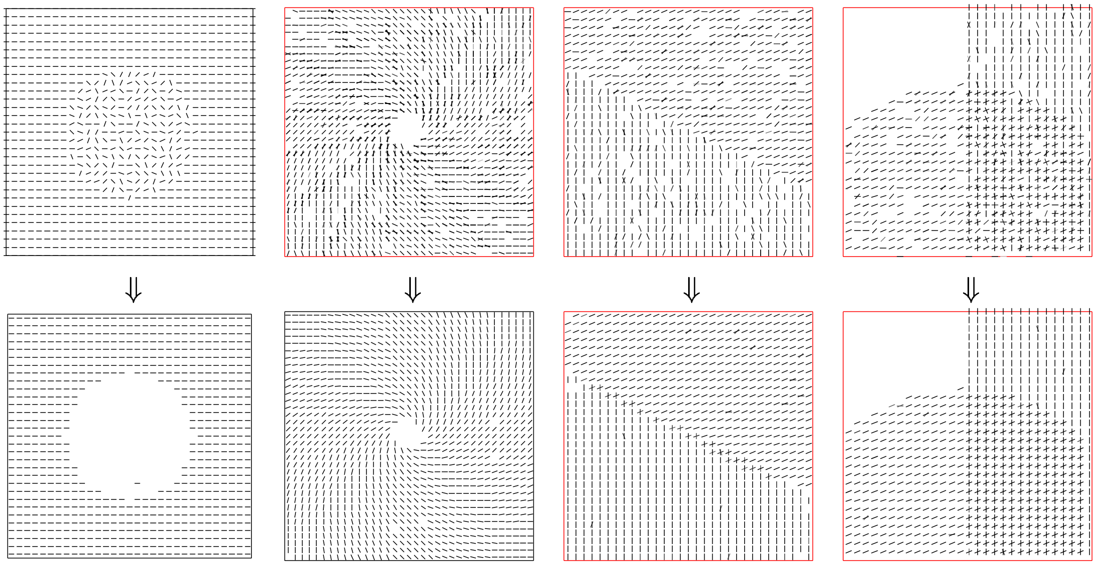


Figure 4.13: **Organization of synthetic flow structures.**

three distinct zones; two stable ones, in which significant variations in initial data are monotonically suppressed till convergence to one predictable outcome; and one unstable, in which small variations are amplified and the result is unpredictable.

4.3.3 Natural texture flows

Figs. 4.15 to 4.17 show the results of applying our system to natural texture flows. Initial orientation measurements for these flows were computed either from the intensity gradient or using a dark/bright line detector [Iverson and Zucker, 1995]. Initial curvatures estimates were approximated by the application of Eq. 2.7 or, in the lack of dense orientation measurements, curvature confidences were assigned equally to all possible curvature pairs. In the latter case, it was the network that in fact selected the most appropriate curvature based on the most salient structure around each node. The stability at boundaries in all these examples is wholly due to the support normalization process described in Sec 4.2.4. The limitation of this approach is apparent in the Cup image of Fig. 4.15 (note the "leakage" of the flow into the cup) and we address this issue further in Sec. 4.4.

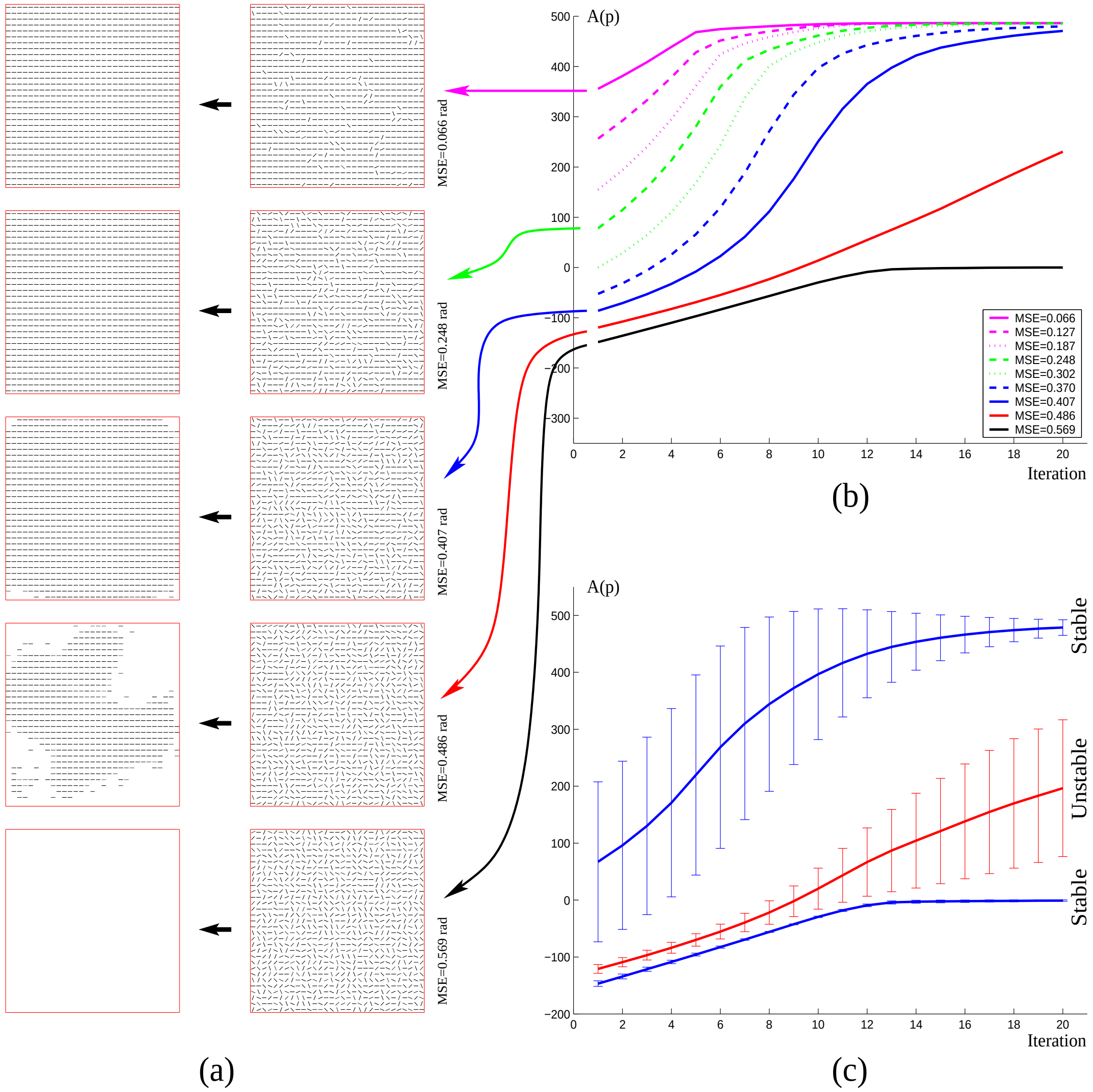


Figure 4.14: **Robustness to noise in texture flow computation. (a)** Each row illustrates a progressively more noisy (MSE values, in radians, are recorded in the figure) texture flow along with the result of 20 iterations of the relaxation process. **(b)** The dynamics of $A(\bar{p})$ as a function of initial noise level and number of iterations. The corresponding texture patches in panel a are indicated by the arrows. **(c)** Regions of stability and instability in the relaxation process (see text). Graphs represent mean $A(\bar{p})$ from 75 runs. Bars are 1 s.d.

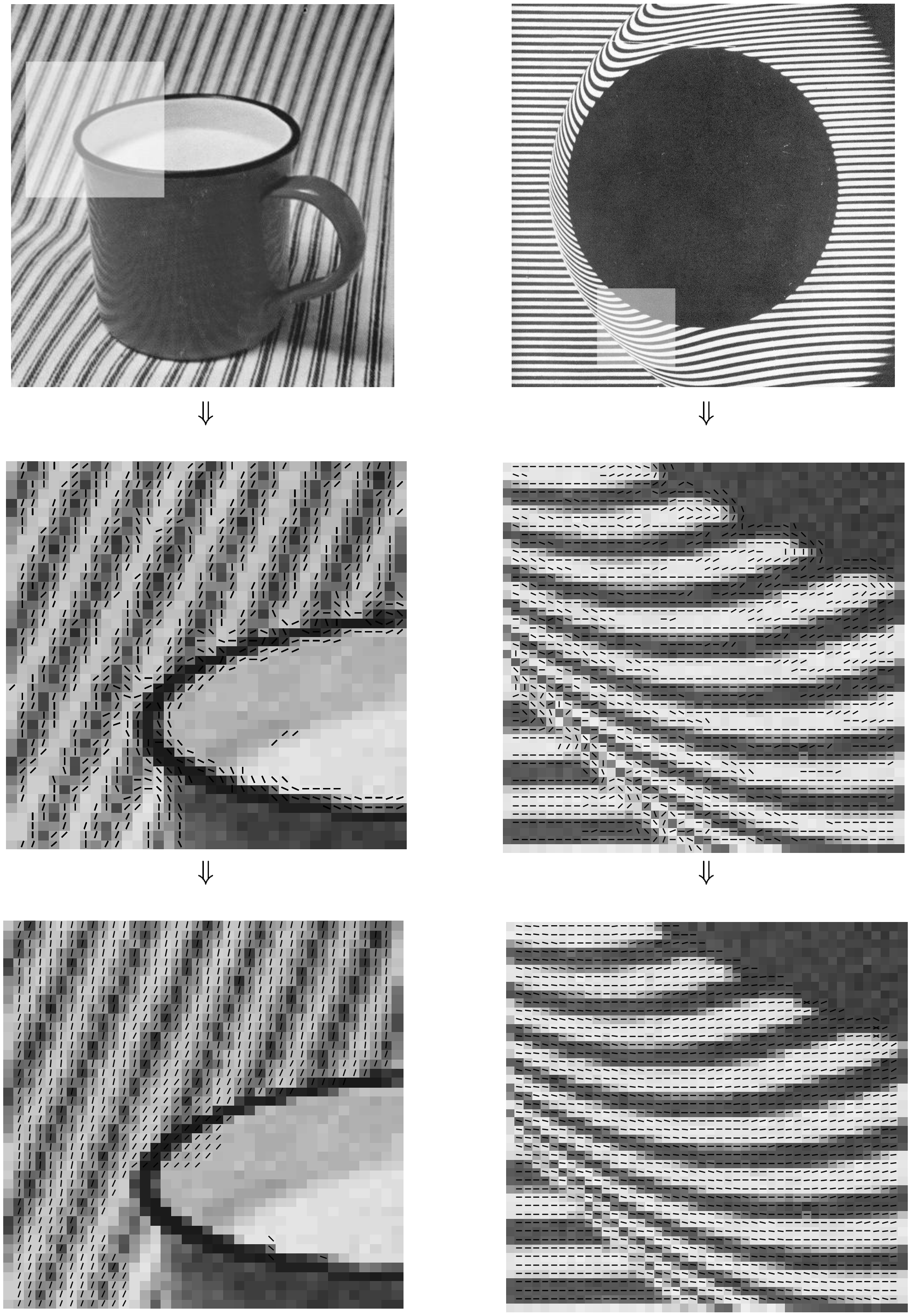


Figure 4.15: **Organization of natural texture flows**. Top row shows the images and a region of interest to which we applied the computation. Middle row shows the initial measurements superimposed on the intensity ROI. Bottom row shows the resultant flows.

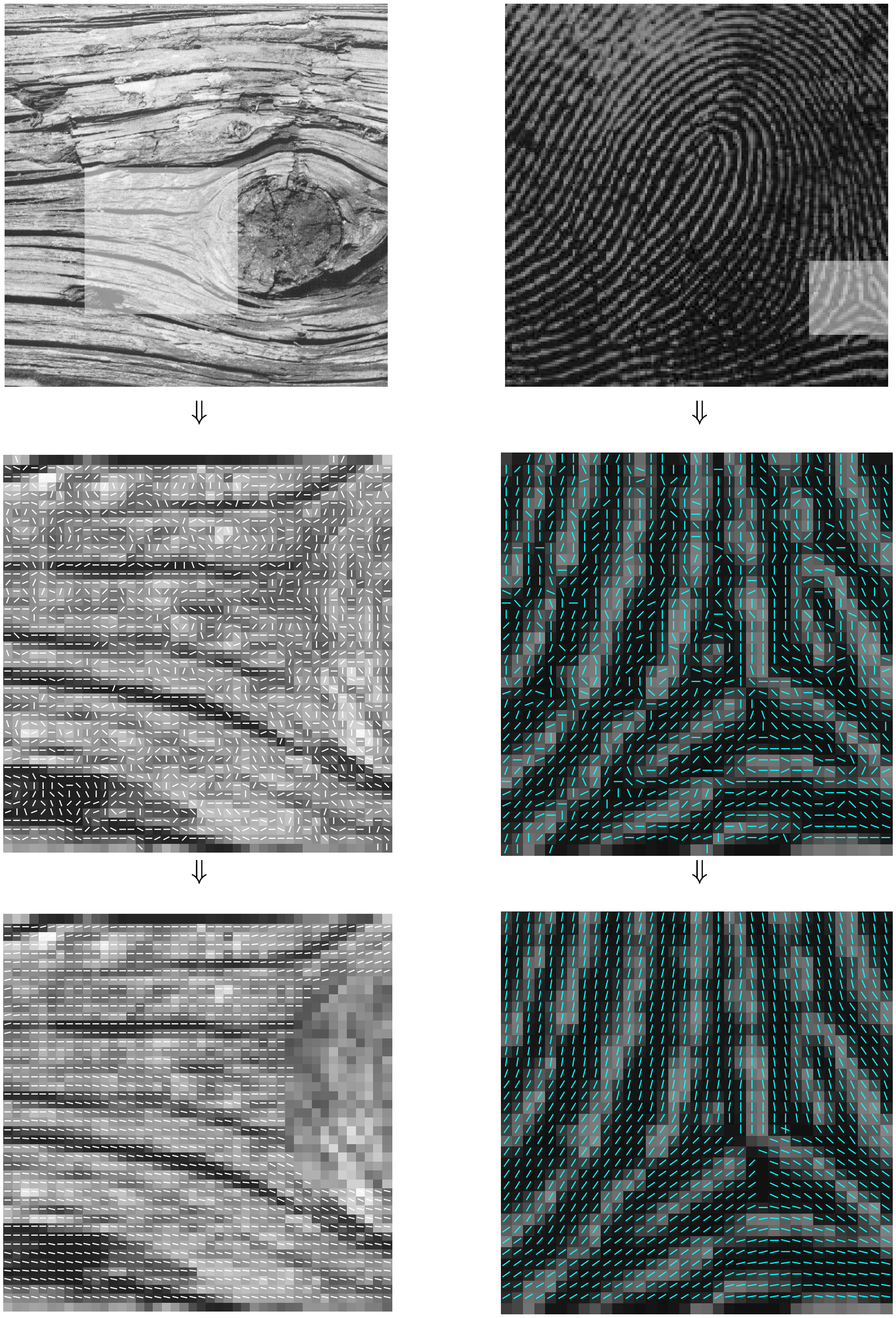


Figure 4.16: **Organization of natural texture flows**. Organization is similar to Fig 4.15. Compare, in particular, the recovery of the flow in the tree stump image to the desired perceptual structure in Fig. 2.3 and the rejection of measurements in the region of the knot.

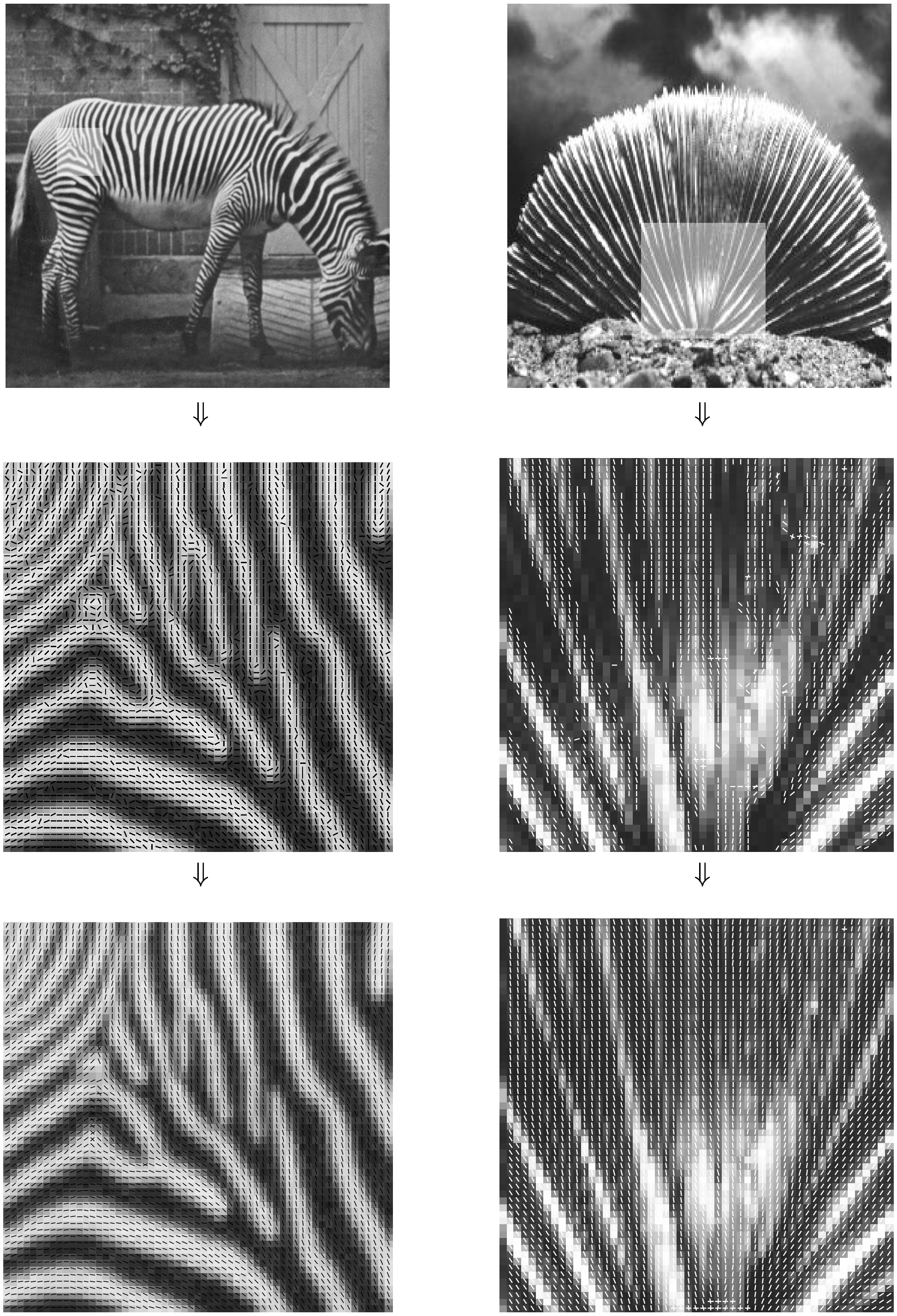


Figure 4.17: **Organization of natural texture flows**. Note the recovery of the singularity (and the short discontinuity below it) of the zebra’s texture and coherent dense representation achieved for the Shell.

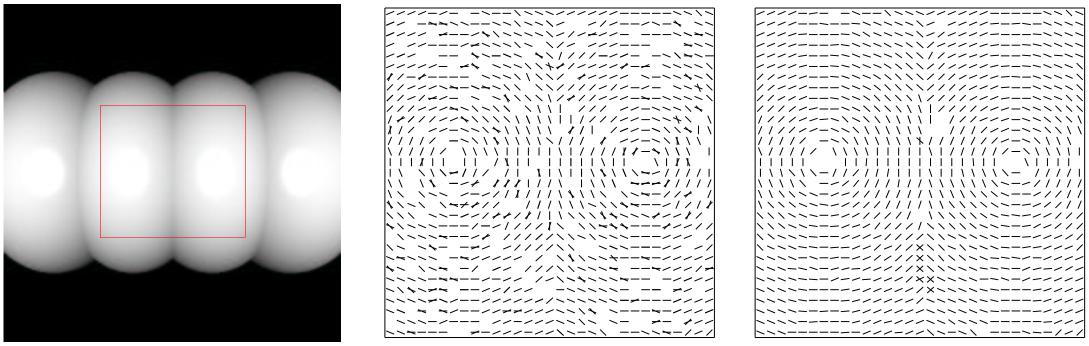


Figure 4.18: **The organization of coherent shading cross.** The worm-like object from Fig. 4.5 and noisy initial measurement of its shading flow. Unlike orientation diffusion (Fig. 4.9) the relaxation process extracts the coherent structure from the noisy data without collapsing the orientation cross and two bifurcation, and without distorting the global structure.

4.3.4 Natural shading flows

Shading represents an important class of visual flows (Secs. 1.1.2), especially for smooth untextured surfaces. A failure to preserve the geometrical information in the shading may result in a wrong interpretation of the visual data, as was discussed in Secs. 4.1.2 and 4.3.1. The examples presented in this section (Figs. 4.18 to 4.20) examine the performance of the relaxation network on a variety of shading flows. Fig. 4.18, for example, presents the analysis of the shading flow of the worm-like object from Fig. 4.5 and shows that unlike orientation diffusion (Fig. 4.9), all the essential structural features are preserved. Fig. 4.19 and 4.20 demonstrate the recovery of coherent shading on smooth untextured surfaces and the rejection of non-coherent shading structure from highly textured or non smooth portions of a scene. Such results can be particularly useful to prevent confusion in edge classification [Huggins and Zucker, 2001b] and shape interpretation tasks.

4.3.5 Overlapping and multi-oriented natural visual flows

The ability to handle overlapping and multi-oriented structures is intrinsic to the computational approach developed in this thesis, and one synthetic example was already demonstrated in Fig. 4.13 (Sec 4.3.1). Clearly, such structures are not confined to synthetic anomalies (Sec. 1.1.6). One natural example of multi-oriented texture is illustrated in Fig. 4.21. Given such an image, an initial estimate

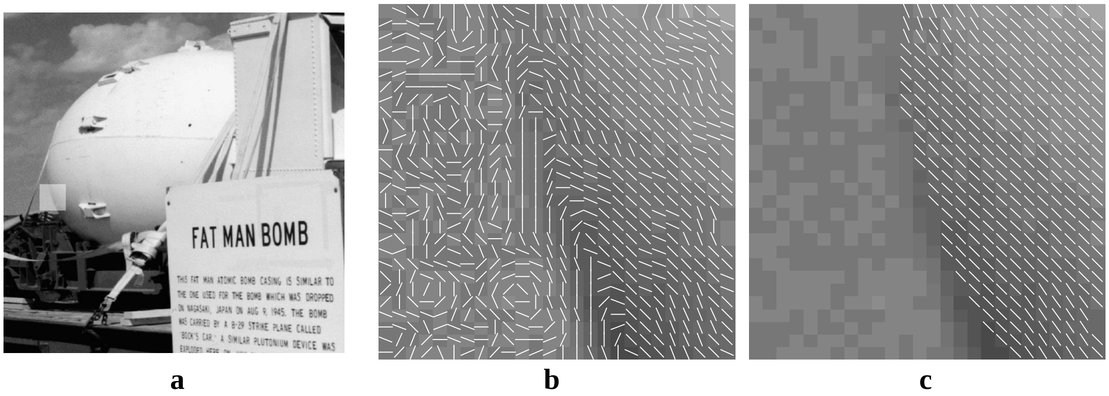


Figure 4.19: **The organization of coherent shading flow.** The smooth shape of the untextured fat-man bomb is in sharp contrast to the clouded background it occludes (left). The noisy initial measurement of its shading flow (center) are relaxed to a coherent structure on the bomb and rejection of the shading flow projected from the background (right).

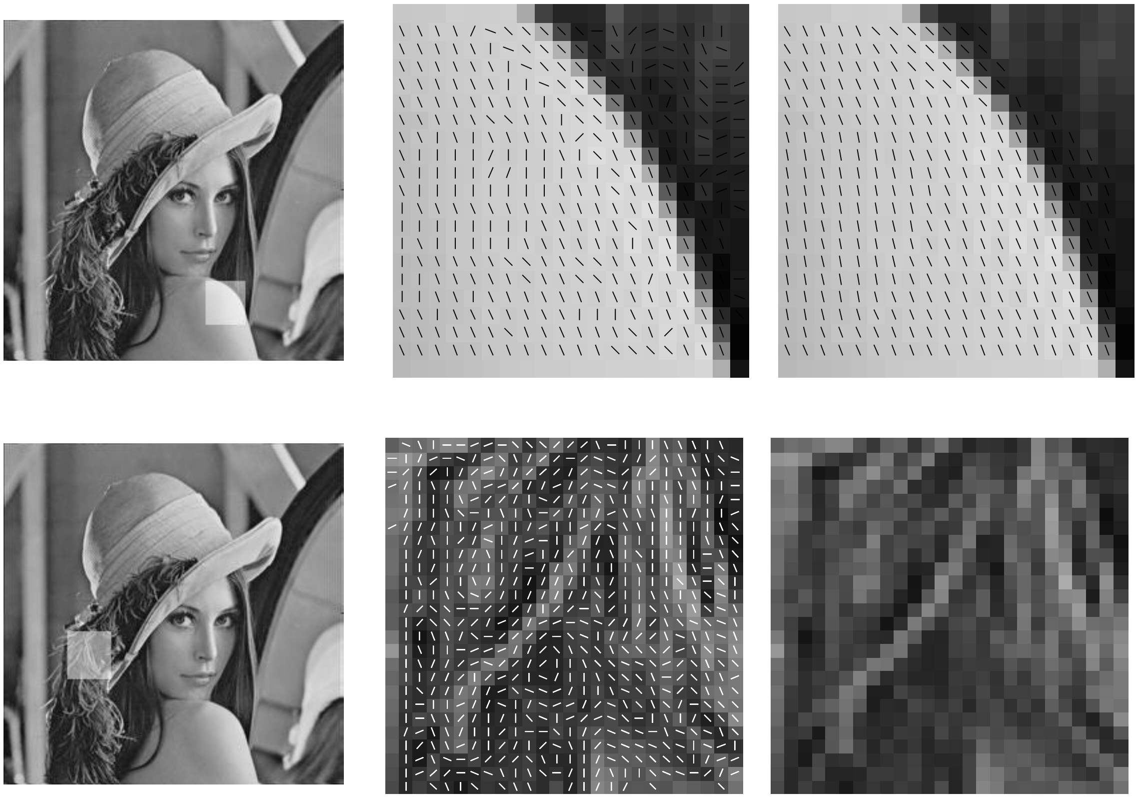


Figure 4.20: **The organization of coherent shading flow.** Although the shading on Lena's shoulder is very smooth, some noise is nevertheless present. The relaxation process extract the coherent structure, eliminate the noise, and at the same time rejects the incoherent shading flow from the feather area.

of a multi-valued texture flows can be extracted from it via a population of oriented filters. In our experiments, we used the Logical/Linear edge/line detector [Iverson and Zucker, 1995], which by design can return multiple edgels at a point.

Once the relaxation labeling network has been initialized by a process that can deal with multi-orientation patterns, the computational process cannot, and need not, distinguish between a single texture of multiple orientations and an overlapping arrangement of *distinct* visual flows. In fact, with only the geometrical (i.e., orientation) information in hand, such a distinction can be decided only by a more global process similar to the way transparency is decided for a collection of intensity patches [Kanizsa, 1979]. From the point of view of the relaxation network, every multi-directional texture is perceived, and thus analyzed, as a collection of *separate* visual flows that happen to share the same spatial region. This separation is achieved at no extra cost as a result of the high dimensionality of the space in which the relaxation operates. In particular, overlapping visual flows are likely to have different orientations wherever they overlap, thus the representations of these flows in the 5D assignment space will be mutually uninfluential.

4.4 Visual flows with boundaries

As is shown in Sec. 4.3, the result of the normalized relaxation process is usually excellent. Nevertheless, the fact that both the support function (Eq. 4.2) and the normalization (Eq. 4.10) are linear creates a delicate balance: while better structure inference and noise resistance suggest smaller s_{min}, it also implies that at discontinuities the flow will eventually grow uncontrollably.

In certain types of visual flows, however, discontinuities can be detected explicitly and embedded into the network to decouple the handling of discontinuities from the support normalization. The most obvious example of this case probably lies in the domain of shading, where discontinuities in shading flows are signaled by intensity edges and thus can be explicitly identified by edge detection. Similar observations can be made in relation to boundaries in hue flows, motion flows, and even texture flow.

Explicit boundaries can be thought of as dividing the visual flow field domain into distinct regions, implying that the computation of coherent structure on either side of the boundary can and

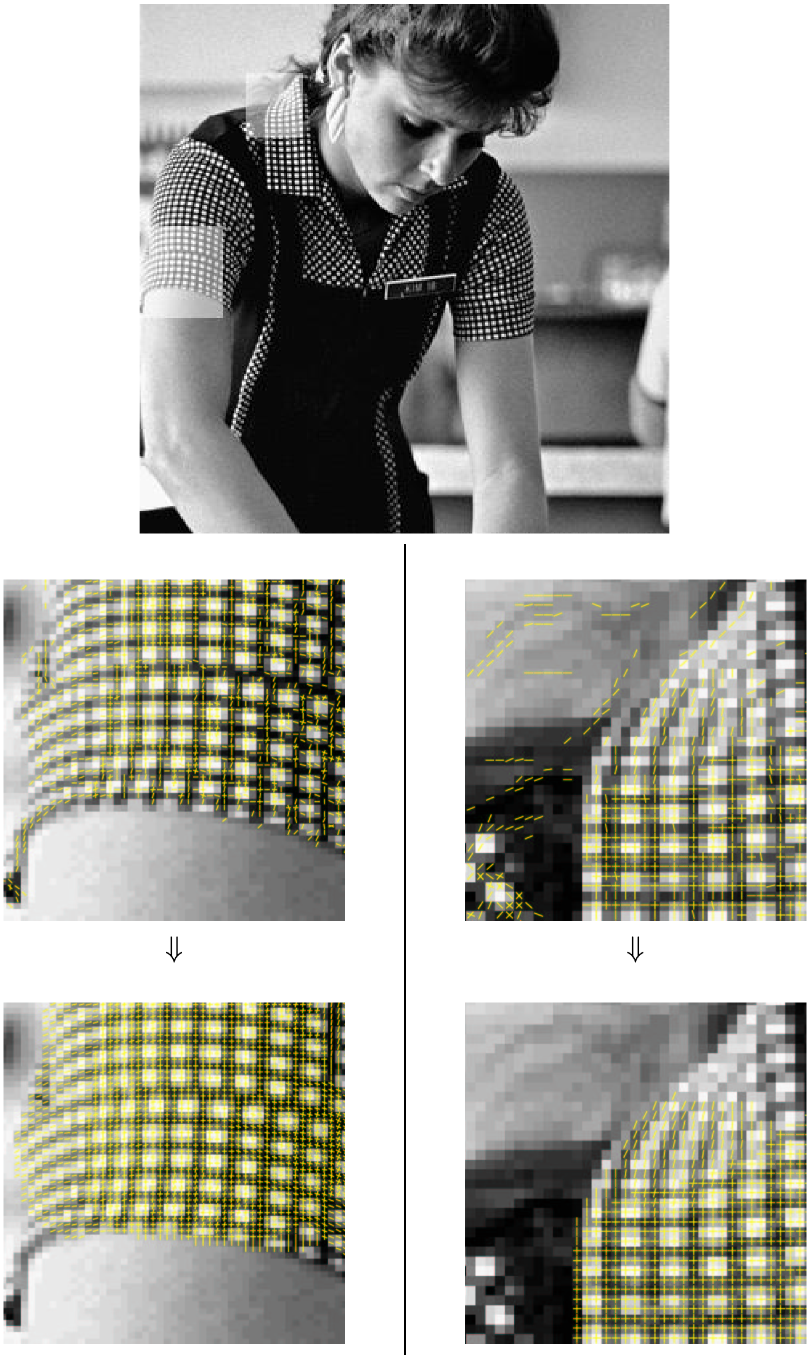


Figure 4.21: **The organization of natural multi-oriented texture flow.**

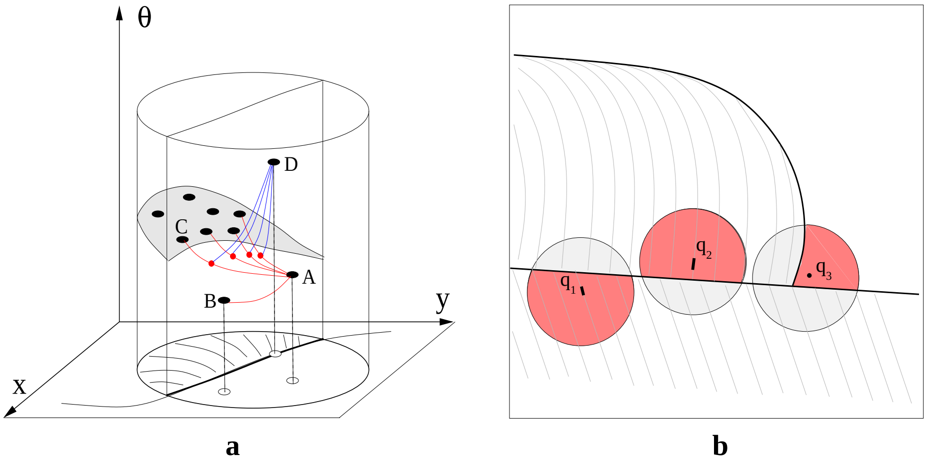


Figure 4.22: **Edge-flow non-linear interactions for boundary stability.** **(a)** Interaction between edge and flow nodes blocks the effective extent of compatibility fields to gather support only from one side of the boundary (see text). Thus, flow structure can never extend beyond explicit boundaries. **(b)** In the image plane, this mechanism can be seen as a data-driven local inhibition that dynamically alters the shape of compatibility fields based on boundary information. Here the image contains three distinct regions, two with flow structure and one without. The circles represent the spatial extent of the original compatibility fields around selected points. The highlighted regions are the effective regions in the presence of boundaries. Although points like q_3 have much structure in their original context, none of it is part in their effective neighborhood.

should be done separately. This is an intuitive but powerful argument. Once such separation is achieved in practice, it obviates the trade-off between high resistance to noise and strict stability along discontinuities.

To implement this idea in the framework of relaxation labeling, what is needed is a specialized set of interactions between edge nodes and nearby shading flow nodes in order to block the flow input if it comes from across the edge. Consider the configuration in Fig. 4.22a, where a flow structure in the image (XY) plane is bounded by the indicated edge and the compatibility structure is depicted as interaction between points in $XY\theta$. *Flow* cell A is connected to a set of other flow cells (B and C) which are a part of the its compatibility field. Although A is not active (there is no flow in its corresponding retinotopic position), its facilitory interaction with the cells on the other side of the edge may eventually raise its activity level. To prevent cell C from affecting A, an active *edge* cell D blocks the facilitory inputs from C, thus effectively limiting A's context to cell B only (or more generally, to those cells on *his* side of the edge). Unless enough of these cells are also active, A will not reach its activation potential, and thus will not signal any flow.

More generally, with the input from beyond boundaries blocked, and so long as s_{min} is positive, the flow on one side of the boundary will not extend across it, because the total support contributed to the other side will never exceed zero. This frees the selection of s_{min} from stability considerations and allows us to determine it solely on the basis of noise resistance and structural criteria. This mechanism can be viewed as a data-driven inhibition that dynamically alters the effective extent and shape of compatibility fields according to explicit boundary information, as illustrated in Fig. 4.22b. In this sense, not only that now we can drop the assumption that flow boundaries are locally straight (Sec. 4.2.4), but we can allow non smooth boundaries as well. Interestingly, a nonlinear veto mechanism that is reminiscent of the one proposed here also exists in biological systems in the form of shunting inhibition [Borg-Graham et al., 1998].

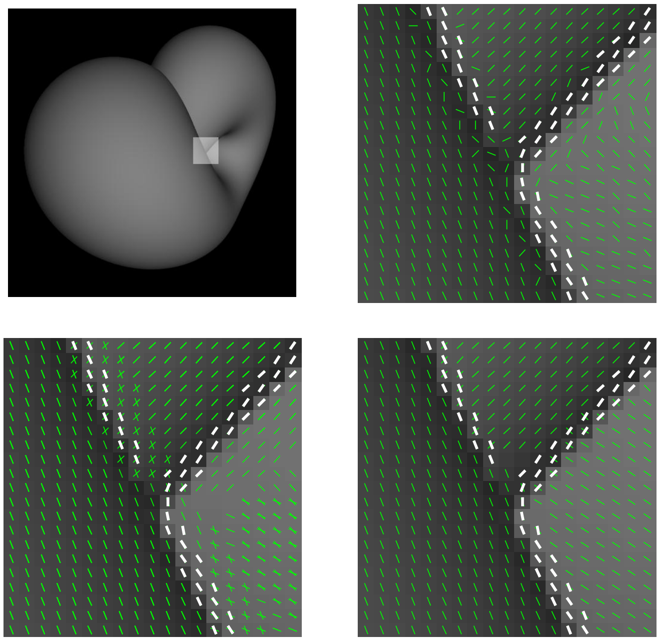


Figure 4.23: **Shading flow relaxation with edges as boundary conditions.** Shown are (left to right, top to bottom) image and ROI, initial shading flow (thin segments) and edges (thick segments), relaxation without boundaries (50 iterations), and relaxation with boundaries (150 iterations). Without considering boundaries, the coherent flow in one region (left half side) "leaks" across the edge before computation in another region (bottom right quarter) is finished. This is avoided when boundaries are taken into account, even if the number of iterations is much larger. Both relaxations use $\delta = 0.25$ and $s_{min} = s_{max}/8$.

We have tested this adaptive relaxation labeling network on a variety of synthetic and natural images, and compared the results to those obtained without explicit boundary information. Two examples involving shading flows and intensity edges are shown in Figs. 4.23 and 4.24. For these stimuli we used the Logical/Linear [Iverson and Zucker, 1995] and the Canny Canny [1986] detectors to detect edges, while the shading flow fields were measured using standard gradient operators.

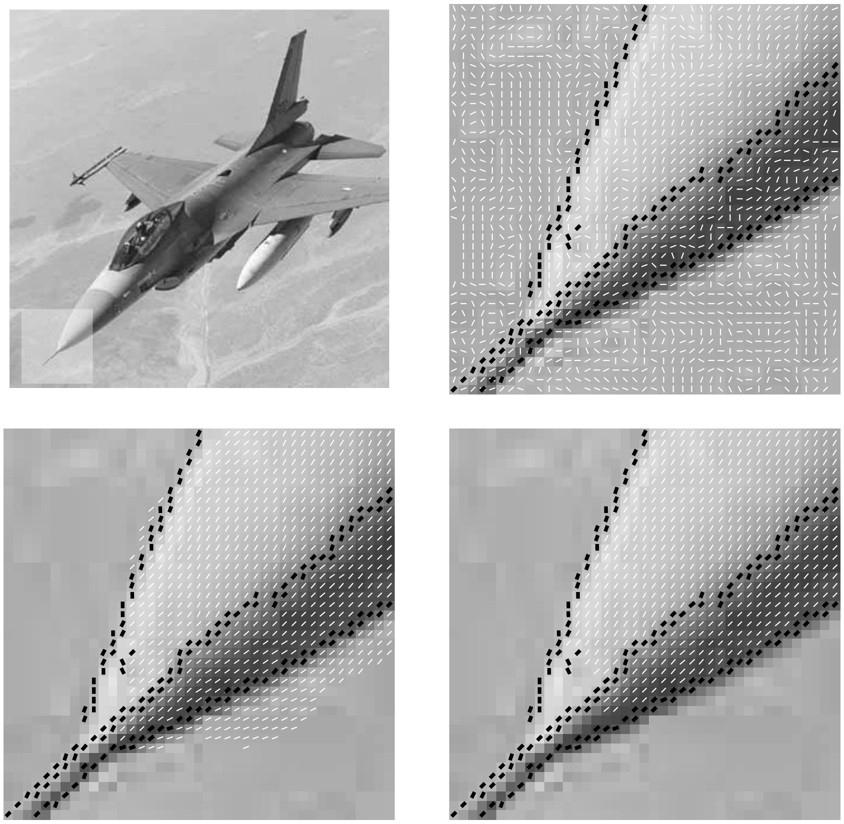


Figure 4.24: **Shading flow relaxation with edges as boundary conditions.** Ignoring boundaries causes the flow to extend beyond the F16 nose boundary (bottom left, 50 iterations), an outcome that is prevented when boundary information is explicitly integrated into the network (bottom right, 150 iterations). Both relaxations use $\delta = 0.25$ and $s_{min} = s_{max}/8$.

Chapter 5

Visual flow, curvature, and the functional organization of the primary visual cortex

The framework developed in chapter 4 is effective computationally. Since it is based on a distributed contextual computation between interconnected simple elements that represent oriented visual atoms, it is also biologically plausible. Indeed, neurons in primary visual cortex respond selectively to oriented stimuli such as edges and lines and the long-range horizontal connections between them are thought to facilitate some kind of contextual computation. This functional similarity suggests an in-depth examination of the relationship between the biological machinery and the computational one, as we do in this chapter.

The predominant position among neurophysiologists is that the contextual computation served by the long-range horizontal connections in V1 relates to curve integration. More specifically, many physiological and psychophysical findings suggest that the distribution of these connections is dictated by either colinear or association field models for curve integration. However, significant evidence of interactions inconsistent with these hypotheses is accumulating also. Thus, in this chapter we first scrutinize this issue and show that neither the colinear and association field models, nor natural random variations around them, can account for the available biological data on the distribution of long-range horizontal connection.

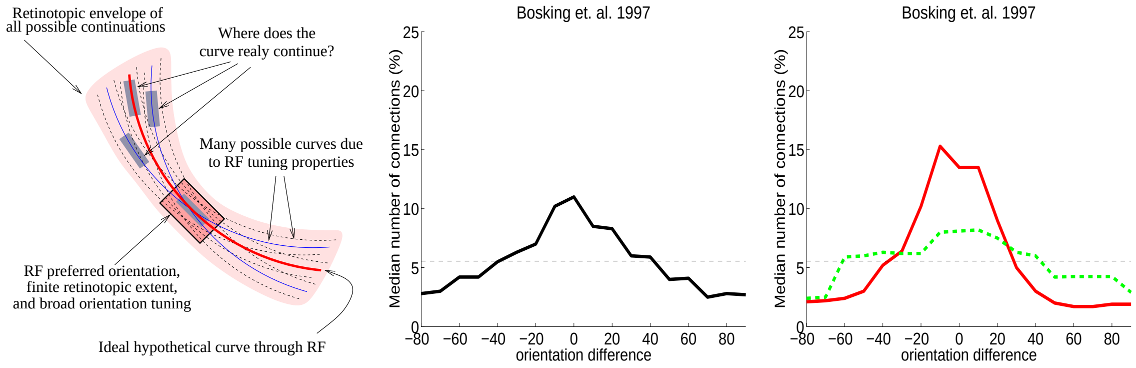
These findings motivate the search for more principled explanations, one of which emerges naturally from the geometrical model and the computational, curvature-based compatibility structures presented earlier in this thesis. Based on these developments, as well as on earlier studies of cocircular compatibilities [Parent and Zucker, 1989; Iverson, 1994], we generate predictions about long-range horizontal connections and compare them to anatomical data. The match proves to be striking; the curvature-based models predict quantitatively the (typically ignored) spread in projection distribution, its non-monotonic variance, and the differences found between individual neurons. Surprisingly, and for the first time, these results also indicate that texture (and shading) continuation can serve as alternative and complimentary functional explanations to contour integration by long-range horizontal connections. Because current anatomical data support both (curve and texture) integration models equally, and because both are important computationally, new testable predictions are derived to allow their differentiation and identification.

5.1 Long-range horizontal connections in V1

The receptive fields (RFs) of neurons in visual cortex characterize their response to patterns of light in the visual field. In primary visual cortex this response is often selective for stimulus orientation in a small region [Hubel and Wiesel, 1977]. The clustered long-range horizontal connections between such cells [Rockland and Lund, 1982] link those with non-overlapping RFs and are thought to facilitate contour integration [Field et al., 1993a]. However, there is no direct physiological evidence that these connections *only* support curve integration while, on the other hand, there also remains much ambiguity about the precise connections required to support the integration of curves. In this section we examine these issues in greater depth.

5.1.1 Biological data and integration models

The argument that associates long-range horizontal connections with curve integration begins with the realization that the finite spatial extent of RFs and their broad orientation tuning lead to significant uncertainties in the position and the local orientation measured from visual stimuli. This



a **b** **c**

Figure 5.1: **Visual integration and the distribution of long-range projections. (a)** Broad tuning in orientation and position introduce uncertainty in curve integration even if a single curve model (thick, red curve) is assumed through the RF. Determining which nearby RF the curve continues through can be facilitated by interaction between neurons with mutually aligned, retinotopically close RFs. **(b)** A fundamental measurable property of long-range connection is their distribution in the orientation domain, i.e., the percentage of connections between interconnected neurons as a function of preferred orientation (angular) difference. This graph shows the median distribution of 7 cells in primary visual cortex of tree shrew (from Bosking et al. [1997]). Qualitatively similar (though coarser) measurements are available on primates as well ([Malach et al., 1993]). **(c)** Connectivity distribution of individual cells reveals significant variability and qualitative differences between them. Shown here are two cells from Bosking et al. [1997].

causes a further uncertainty in determining which of the many nearby RFs signal the next section of a curve (Fig. 5.1a). All these uncertainties underlying curve integration can be reduced by interactions between neurons whose RFs are close in retinotopic coordinates. Starting with Mitchison and Crick and their hypothesis about interactions between iso-oriented RFs [Mitchison and Crick, 1982], physiological and anatomical findings have been accumulating to suggest a roughly *colinear* interaction. The main evidence supporting this conclusion is based on the distribution of angular differences between preferred orientations of connected cells. These distributions are computed by taking the orientation difference between a target cell and every other cell it is connected to with a long-range horizontal connection. Indeed, as is exemplified in Fig. 5.1b, these distributions have been shown to be unimodal, with maximal interaction between iso-oriented RFs [Ts'o et al., 1986; Gilbert and Wiesel, 1989; Weliky et al., 1995; Schmidt et al., 1997; Buzás et al., 1998; Bosking et al., 1997; Malach et al., 1993; Sincich and Blasdel, 2001; Schmidt and Löwel, 2002]. Furthermore, direct anatomical studies reveal long-range interactions between co-axial cells [Bosking et al., 1997; Schmidt et al., 1997] and indirect psychophysical experiments report a general *association field* [Field et al., 1993a; Kapadia et al., 1995, 2000] which emphasizes straight or slowly varying continuations while allowing some support for more rapidly varying continuations as well (Fig. 5.2a).

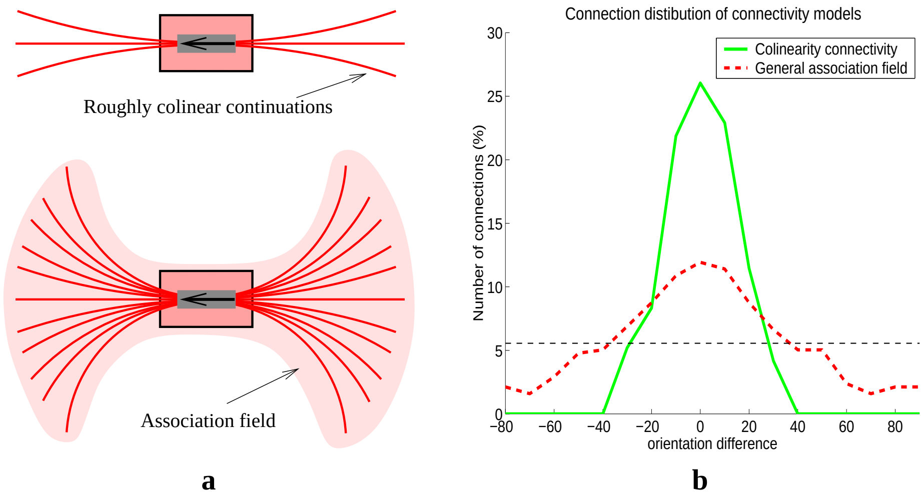


Figure 5.2: **Colinear facilitation, association fields, and their predicted distribution of connections.** **(a)** Informally, two visual integration, or continuation models are typically considered in the physiological and psychophysical literatures. Colinearity, the predominant model, predicts only few possible curve continuations (top). On the other hand, many possible continuations reveal an association field (bottom), similar to those observed psychophysically [Field et al., 1993a]. **(b)** The corresponding distribution derived from the colinearity and association field models. Observe that colinearity predicts a very narrow distribution which is clearly at odds with the significant spread frequently measured anatomically or electrophysiologically (compare to Fig. 5.1b). The association field leads to a wider spread, but like colinearity, it predicts a *fixed* distribution for all cells, a hypothesis refuted in recent studies (see text). The colinearity distribution (solid green) was calculated from the field depicted in Fig 5.7a while the association field distribution (dashed red) was calculated from the field in Fig 5.7e. The dashed horizontal line depicts the uniform distribution.

With the accumulation of these data, however, are a growing number of observations that are difficult to reconcile with the intuition that neural spatial integration is based on colinearity, or that it only serves curve integration. Facilitory interaction between cells of significant orientation difference [Kapadia et al., 1995], short range co-axial *inhibition* [Polat and Sagi, 1993], iso-orientation *side* facilitation [Adini et al., 1997], and strong correlations between iso-oriented, non-overlapping, and *parallel* receptive fields [Ts'o et al., 1986, see bottom cell on page 1163 or top cell on page 1164] are functionally inconsistent. Evidence of cross-orientation [Matsubara et al., 1985; Kisvárday et al., 1997] and non-axial [Gilbert and Wiesel, 1989] connections, plus roughly isotropic retinotopic extent [Malach et al., 1993; Sincich and Blasdel, 2001], suggest anatomical inconsistencies.

These inconsistencies prompt a closer examination of the interactions within visual cortex and their population statistics. As the evidence suggests, individual cells, or small collections of adjacent cells captured in tracer injections, may have qualitatively different connectivity distributions [Bosking et al., 1997]: some are narrow and high while others are very wide, as is illustrated in Fig 5.1c. When averaged, the pooled distribution is

- unimodal,
- peaks at zero orientation offset,
- indicates a non-negligible fraction of connections linking cells of significantly different orientation preferences [Malach et al., 1993; Kisvárday et al., 1994; Kisvárday et al., 1997; Bosking et al., 1997],
- crosses the uniform distribution at approximately ±40 degrees[1], and
- has a *non-monotonically* changing variance as the orientation difference increases [Malach et al., 1993; Bosking et al., 1997].

Neither colinearity nor association field models predict all of these features. While both models imply unimodal pooled distributions over orientation differences (Fig. 5.2b), they also suggest a fixed projection field and thus neither predicts *any* differences between the connection distributions of individual cells. Furthermore, colinearity is clearly at odds with the significant spread in the distribution of connections over orientation differences, whether it is measured via extracellular injections (e.g., [Bosking et al., 1997]) or the more elaborate intracellular protocol [Buzás et al., 1998].

---

[1] This crossing point provides a reference for the bias of projection patterns toward particular orientations; considering the offsets where the connection distribution crosses the uniform line quantifies this bias in way independent of scale or quantization level.

5.1.2 Integration models and random physiological variations

It is tempting to explain the apparent anomalies and inconsistencies between the predicted and measured distributions of long-range horizontal connections as random physiological variations, for example by asserting that anatomy only *approximates* the correct connections. We tested this explanation by applying different noise models to the colinearity and association field connectivity distributions from Fig. 5.2, and checked whether the resultant pooled distributions posses the properties listed above. The results of the most natural noise model are illustrated in Fig. 5.3b. Under this model, each long-range horizontal connection, ideally designated to connect cells of orientation difference $\Delta\theta$, is shifted to connect cells of orientation difference $\Delta\theta + \epsilon_\sigma$, where ϵ_σ is a normally distributed random variable with variance σ (and zero mean). As the figure shows, it takes an overwhelming amount of noise (s.d.$\geq 35°$) to transform the colinear distribution to one that resembles the measured data in terms of spread and peak height, but the non-monotonic behavior of the variance is *never* reproduced (for space considerations we omit the results of other connection-based noise models, or the noisy distributions based on the association field model, all of which were even less reminiscent of the measured physiological data).

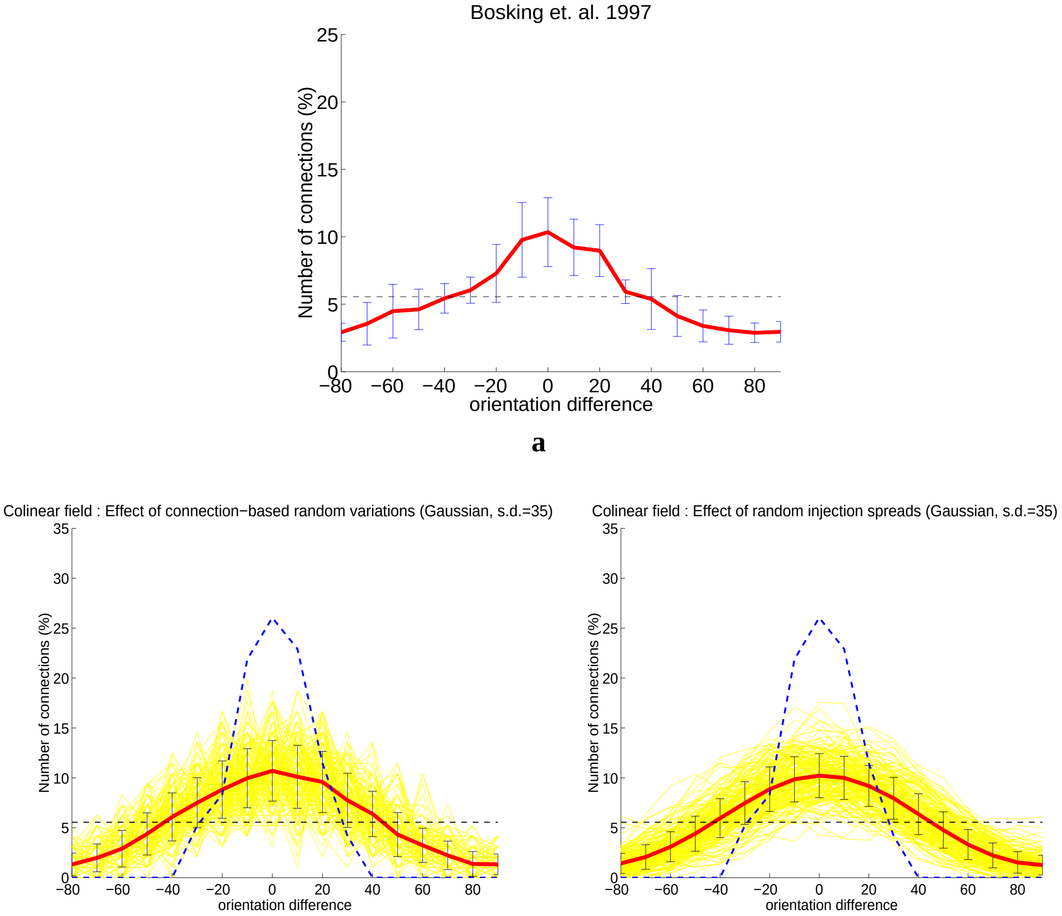
A second possible source for the inconsistencies between the predicted and measured distributions may be the extracellular injection protocol commonly in use by physiologists to trace long range horizontal connections (as in [Bosking et al., 1997]). Although cells stained by these injections are likely to have similar orientation preferences, their orientation tuning may nevertheless be different, sometimes significantly. Consequently, the distribution of presynaptic terminals (boutons) traced from the injection site may incorporate an artificial, random spread relative to the single orientation typically assumed at the injection site. Preliminary evidence from a recently developed single-cell protocol [Buzás et al., 1998] suggests that leakage in the injection site cannot bridge the gap between the predicted colinear distribution and those measured anatomically. However, we thoroughly examined this possibility by modeling the leakage in the injection site computationally as a Gaussian random variable of predefined variance. The base distribution (colinear or association field) of the computational cells selected by this process were then summed up and normalized, and the resultant (random) distribution was attributed to the original cell representing the injection site. Repeating this process many times yielded a collection of (different) distributions, for which we calculated an average and variance. The results are illustrated in Fig. 5.3c. Similar to random variations at the level of individual connections, here too it takes an overwhelming amount of noise (s.d.$\geq 35°$) to transform the colinear distribution to one that resembles the measured data in terms of spread and peak height, but the non-monotonic behavior of the variance is *never* reproduced.

The thinking around long-range horizontal connections has been dominated by their first-order statistics and its peak at zero orientation offset. However, the non-monotonicity of the variance was first reported almost a decade ago [Fig. 3d in Malach et al., 1993] and we have further confirmed it from the more detailed measurements in Bosking et al. [1997], as was illustrated in Fig. 5.3a. Since neither colinearity nor association field models can explain this aspect of the physiological data, even if much noise is allowed, it is necessary to consider whether this and the other subtle properties of the pooled data reflect genuine functional properties of long-range horizontal connections. Such novel properties are suggested by the geometrical and computational exploration discussed earlier in thesis, as well as in previous work on contextual curve inference [Parent and Zucker, 1989; Iverson, 1994]. In particular, two new functional aspects emerge; one that involves curvature as a natural consequence of orientation good continuation; and a second that involves 2D visual flow (texture or shading), rather than 1D curve integration. Our main goal in this chapter is to examine whether or not either of these new aspects can explain better the available biological data.

5.2 The computational columnar machine

Curve integration, the hypothesized functional role ascribed to long-range horizontal connections, is naturally based in differential geometry and thus it is first important to understand the fundamental relationship between the biological machinery and the mathematical formulation.

The tangent, or the local linear approximation to a curve, abstracts orientation preference, and the collection of all possible tangents at each (retinotopic) position can be identified with the orientation hypercolumn [Hubel and Wiesel, 1977]. Formally, since position takes values in the plane $\mathbb{R}^2$ (think of image coordinates x, y) and orientation in the circle S^1 (think of an angle θ varying between 0 and 2π), the primary visual cortex can be abstracted as the product space $\mathbb{R}^2 \times S^1$



a

(b) **(c)**

Figure 5.3: **Results of a statistical perturbation of colinearity connectivity distribution. (a)** Mean connection distribution computed from the data in Bosking *et al.* [Bosking et al., 1997], shown here for reference. Error bars are ±1 standard deviation. Note the unimodal distribution that peaks at approx 11%, the wide spread, the crossing of the uniform distribution (dashed horizontal line) around ±40 °, and the non-monotonic variance. Can all these features be replicated by applying noise to the base distribution induced by the standard colinearity model? **(b)** Result of simulating physiological deviation at the individual connection level Dashed blue line is the base colinear distribution. Yellow region is the superposition of individual applications of the noise model to the base distribution. Solid red graph is the expected distribution and error bars are ±1 standard deviation. Permitting large enough developmental variations (shown here is the result of Gaussian i.i.d. noise of s.d.$= 35°$) in the connections to model the first-order statistics significantly violates the underlying connectivity principle of good continuation but still cannot model the second-order statistics. **(c)** Results of simulating measurement errors due to leakage in the injection site. All parts are color-coded as in panel b. Again, permitting large enough injection spread to model the first-order statistics (shown here is the result of Gaussian noise of s.d.$= 35°$ and assuming 20 cells per injection site [Bosking et al., 1997]) cannot model the second-order statistics.

(Fig. 5.4). Points in this space represent both position and orientation to abstract visual edges of given orientation at a particular spatial (i.e., retinotopic) position.

Since any single tangent is the limit of any smooth curve passing through a given (retinotopic) point in a given direction, the question of curve integration becomes one of determining how two tangents at nearby positions are related (Colinearity, for example, asserts that the tangent orientation hardly changes for small displacements along the curve). In general terms, the angular difference between RFs captures only one part of the relationship between nearby tangents; their relative spatial offset also must be considered. Thus, in the mathematical abstraction, relationships between tangents correspond to relationships between points in $\mathbb{R}^2 \times S^1$, or the space we labeled $XY\theta$. Physiologically, these relationships are carried by the long-range horizontal connections, with variation in retinotopic position corresponding to $\mathbb{R}^2$, and variation along orientation hypercolumns corresponding to S^1 (Fig. 5.5).

In Chapter 2 we discussed how determining an appropriate relationships that reflects good continuation amounts, in mathematical terms, to determining a connection structure. As we show in this chapter, the relationship between these two types of “connections”, the mathematical and the physiological, appears to be more than linguistic.

A main consequence of geometrical analysis of Chapter 2 is that the relationship between nearby tangents depends on the covariant derivative: for curves the connection is dictated by one curvature; for visual flows two curvatures are required. By estimating these quantities at a given retinal point $\mathbf{q}$ it is possible to approximate the underlying geometrical object, and thus a coherent distribution of tangents, around $\mathbf{q}$. Such approximation, or osculating objects, can then be used to model the set of horizontal connections that are required to facilitate the response of a cell if its RF is embedded in a visual context that reflects good continuation.

Based on first principles, much of Chapter 2 was devoted to the derivation of the right helicoid as an osculating object for visual flows (Secs. 2.5 to 2.7). The same chapter also mentions cocircularity [Parent and Zucker, 1989; Zucker et al., 1989; Sigman et al., 2001; Geisler et al., 2001] and the osculating circle as an osculating object for curves. In both cases, different estimates of curvature give rise to objects in $XY\theta$ – helicoids for visual flows and helices for curve – whose points define both the spatial position and the local orientation of nearby RFs that are compatible

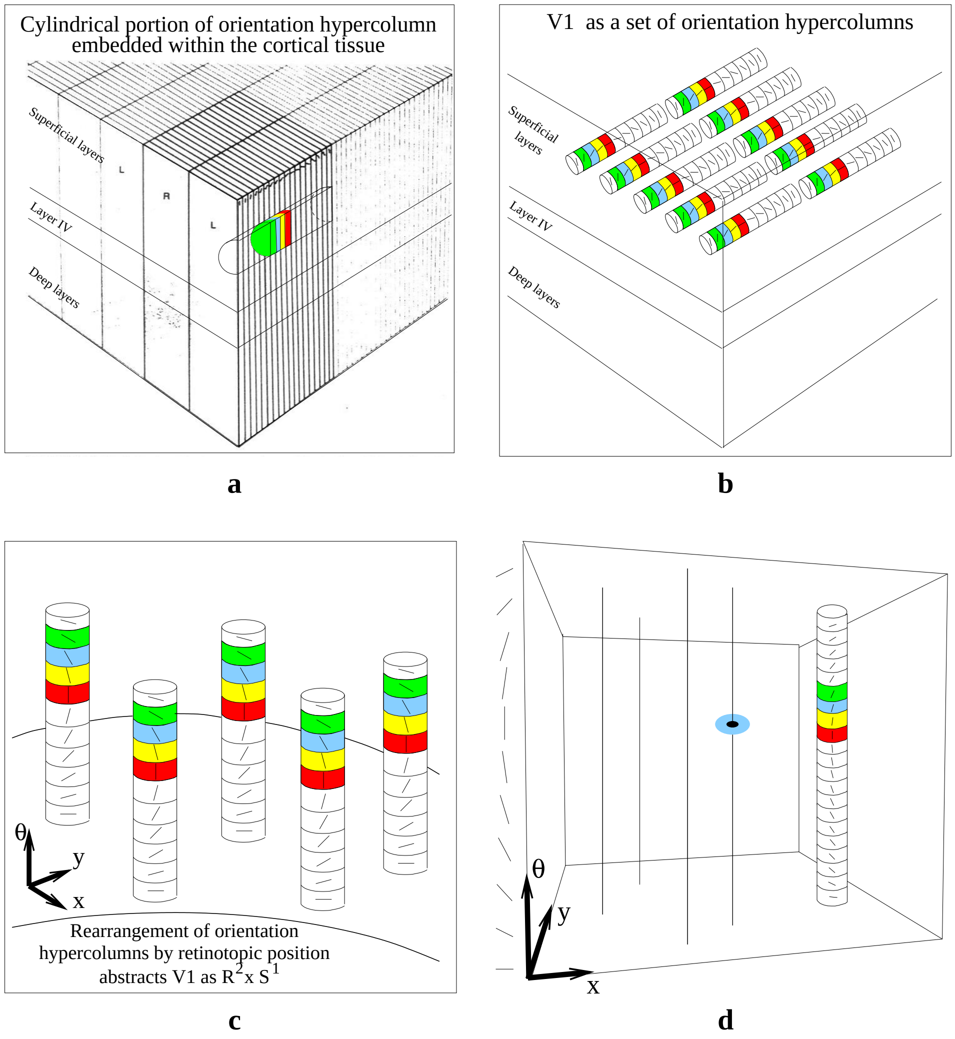


Figure 5.4: **Abstracting the primary visual cortex as $\mathbb{R}^2 \times \mathcal{S}^1$, or *position*×*orientation* space. (a)** The "ice cube" cartoon of visual cortex [Hubel and Wiesel, 1977] (cytochrome-oxidase blobs and distortions due to cortical magnification factor are not shown). A tangential penetration in the superficial layers reveals an orientation hypercolumn of cells whose RFs have similar spatial (retinotopic) coordinates. With cells of similar orientation tuning grouped by color, the hypercolumn is cartooned as a horizontal cylinder. **(b)** With ocular dominance columns omitted, the superficial layers of the primary visual cortex can be viewed as a collection of (horizontally arranged) orientation hypercolumns. **(c)** Drawing the cylinders vertically emphasizes that RFs of cells within a column overlap in retinotopic coordinates (x, y) and makes explicit this aspect of their organization. **(d)** Since different hypercolumns correspond to different retinotopic positions, the set of *all* hypercolumns abstracts the visible subspace of $\mathbb{R}^2 \times S^1$, with each column corresponding to a different vertical fiber in that space. The θ axis in this space corresponds to a tangential penetration within V1 hypercolumns (colors within the column represent different orientation tunings), and the XY plane corresponds to retinotopic coordinates.

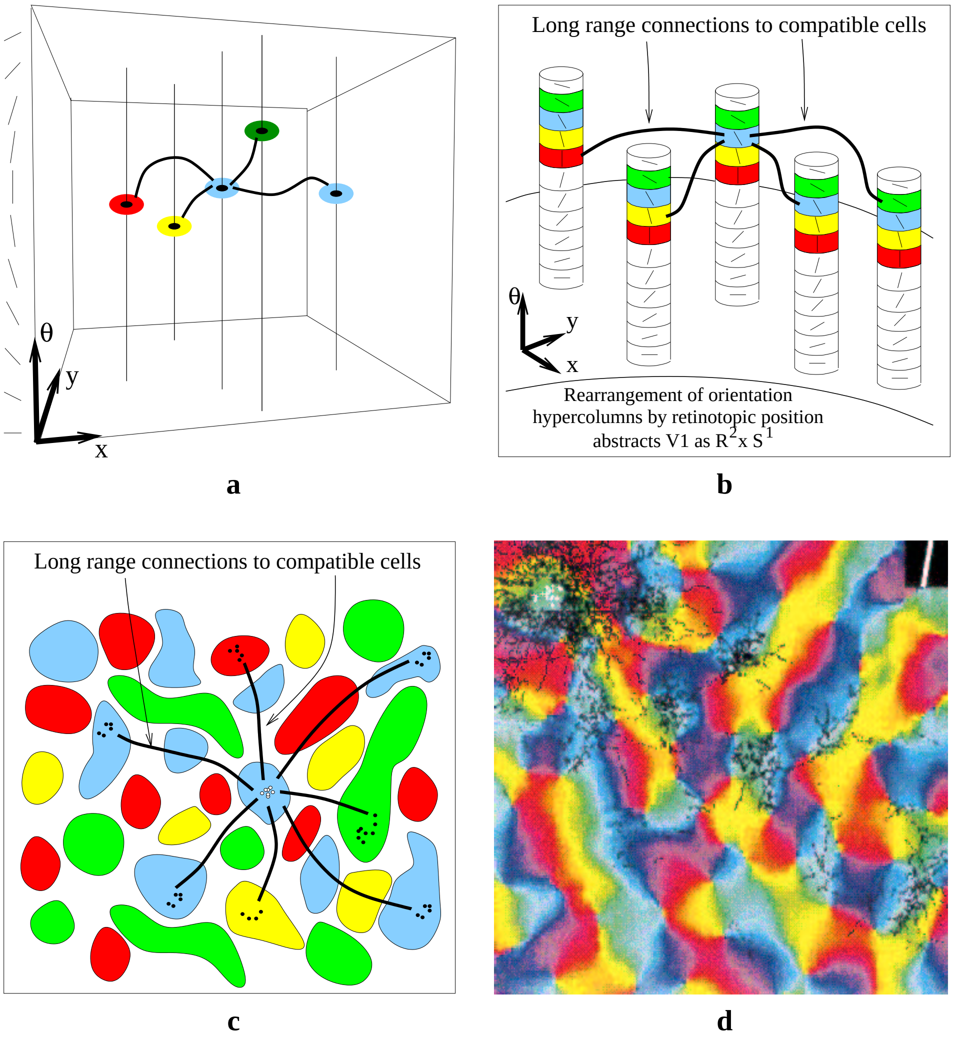


Figure 5.5: **Long-range horizontal connections as relationships between points in $\mathbb{R}^2 \times \mathcal{S}^1$. (a)** Since visual integration must involve both relative orientation and spatial offset of nearby RFs, it is more fully abstracted by relationships between points in $\mathbb{R}^2 \times \mathcal{S}^1$. The exact nature of these relationships is determined by the underlying integration model. **(b)** Redrawing $\mathbb{R}^2 \times \mathcal{S}^1$ fibers as orientation hypercolumns in V1 relates the integration model in $\mathbb{R}^2 \times \mathcal{S}^1$ to the distribution of long-range connections between the hypercolumns. **(c)** Collapsing the $\mathbb{R}^2 \times \mathcal{S}^1$ abstraction to a cortical orientation map, the integration model implies a particular set of long-range horizontal connections between orientation domains (colors represent orientation tuning similar to panels a,b and Fig. 5.4). Such links have been identified and measured through optical imaging and anatomical tracing (e.g., [Malach et al., 1993; Bosking et al., 1997; Buzás et al., 1998]) and thus can be compared to the model's predictions. **(d)** A real counterpart to the schematic in panel c. Reproduced from Bosking *et al.* [Bosking et al., 1997], this image shows an optical image of intrinsic signals combined with long-range horizontal connections traced through extracellular injection of biocytin. The white dots at the upper left corner represent the injection site while the black dots represent labeled boutons.

with the estimate at $\mathbf{q}$ (Fig. 5.6). Thus, these objects can be used to derive a set of interactions between compatible cells, as was discussed in Chapter 4 for visual flows, and in previous work for curves [Parent and Zucker, 1989; Iverson, 1994].

As with the effect of quantization on the computational compatibility fields (c.f. Sec. 4.2.3), broad RF tuning means that both the helix and the helicoid must be dilated appropriately, thus resulting in compatible "volumes" in $\mathbb{R}^2 \times S^1$ (Fig. 4.10) and possibly multiple compatible orientations at given spatial positions. This dilation should be reflected in the set of compatible RFs and the horizontal links to them, and although it takes part in all our calculations, we omit it from Fig. 5.6 to avoid clutter.

With the compatibility volumes in $XY\theta$ determined, their discrete projection to the image plane produces a more intuitive representation of their induced connectivity patterns between orientation hypercolumns. Examples of such compatibility structures are illustrated in Fig. 5.7. Each position in these compatibility fields represents one orientation hypercolumn while individual bars represent the orientation preference of single neurons, all of which are connected to the central cell in each field. Multiple bars at any given point represent multiple neurons in the same hypercolumn that are connected to the central cell, a result of the dilation of the compatible structure due to broad RF tuning. All fields assume that orientation tuning is quantized to 10 degrees (as in Bosking *et al.* [Bosking et al., 1997]) and their radius of influence is set to 4-5 non-overlapping hypercolumns to reflect a 6-8mm cortical range of horizontal connections [Gilbert and Wiesel, 1989] and hypercolumn diameter of 1.5mm (which also accounts for ocular dominance domains).

Fig. 5.7 indicates few important observations and questions with regard to curvature-based compatibility (and thus, connectivity) structures. First, it shows that while the majority of connections link roughly co-linear cells, some connect cells of large orientation differences. Could that be the source of spread observed in biological tissue?

Secondly, it shows clearly how different curvature values induce different projection fields, both for visual flows and curves. These differences are likely to translate to different connection distributions in the orientation domain, and thus they raise the possibility that the variability measured biologically relates to curvature sensitivity rather than noise.

And finally, the figure shows how the "sum" over curvatures of the curve compatibility fields

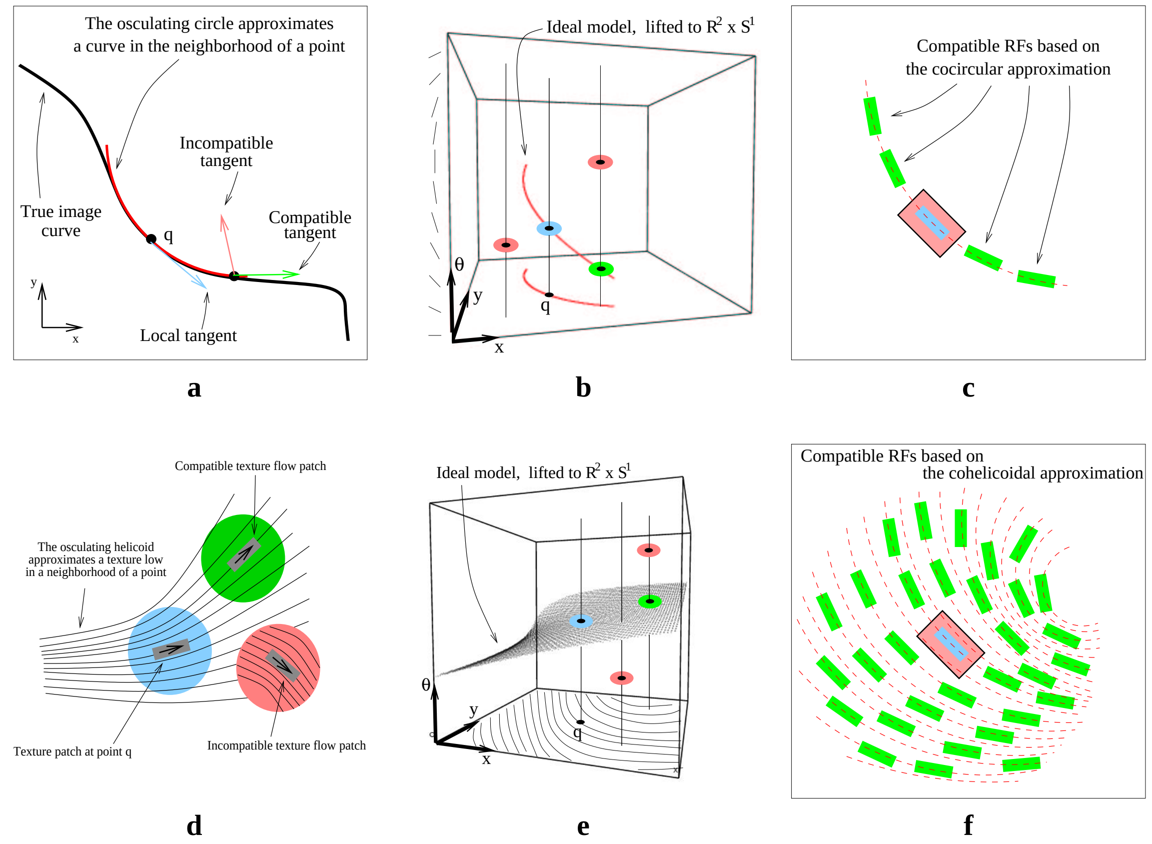


Figure 5.6: **Integration models and horizontal connections between compatible RFs.** **(a)** Estimate of tangent (light blue vector) and curvature at a point **q** permits modeling a curve with the osculating circle as a good-continuation approximation in its neighborhood. Given the approximation, compatible (green) and incompatible (pink) tangents at nearby locations can be explicitly derived. **(b)** With height representing orientation (see scale along the θ axis), the osculating circle lifts to a helix in $\mathbb{R}^2 \times S^1$ whose points define both the spatial location and orientation of compatible nearby tangents. Color-coded as in panel a, the green point is compatible with the blue one while the pink points are incompatible with it. **(c)** The consistent structure in panels a,b induces discrete set of RFs which can facilitate the response of the central cell. Shown here is an example for one particular curvature tuning at the central cell. **(d)** For visual flows, determination of good continuation requires two curvatures at a point, from which a local model of good continuation can be formulated. Given such a model, compatible (green) and incompatible (pink) flow patches at nearby locations can be explicitly derived. **(e)** In $\mathbb{R}^2 \times S^1$, our model for 2D orientation good continuation lifts to a right helicoidal patch, whose points define both the spatial location and orientation of compatible (green) nearby flow tangents. **(f)** As an abstraction for visual integration, the ideal geometrical model – the right helicoid – induces a discrete set of RFs which can facilitate the response of the central cell. Shown here is an example for one particular curvature tuning at the central cell.

gives an association field (Fig. 5.7e) reminiscent of recent psychophysical findings [Field et al., 1993a]. As a psychophysical entity, however, the association field is not necessarily a one-to-one reflection of connectivity patterns in the visual cortex. In fact, representing a “cognitive union” across displays of different continuations, the association field is unlikely to characterize any single cell.

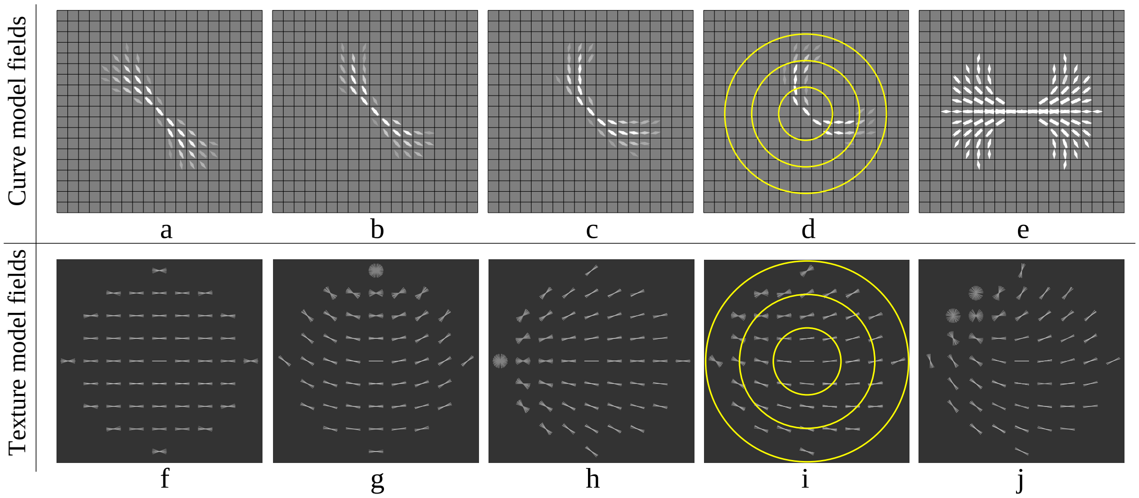


Figure 5.7: **Curve and visual flows projection patterns.** **(a-d)** Examples of co-circularity projection fields for cells with orientation preference of 150 degrees (center bars) and different values of curvature tuning (based on the implementation by Iverson [Zucker et al., 1989]). **(a)** $\kappa = 0.0$ (curvature in units of pixels^{-1}). **(b)** $\kappa = 0.08$. **(c)** $\kappa = 0.16$. **(d)** $\kappa = 0.24$. **(e)** The union of all projection fields of all cells with same orientation preference (0 degrees in this case) but different curvature tuning. Note the similarity to the schematic association field in Fig. 5.2a. **(f-j)** Examples of the helicoidal projection fields for cells with horizontal orientation preference (center bars) and different curvature tuning. **(f)** $(\kappa_T, \kappa_N) = (0.0, 0.0)$. **(g)** $(\kappa_T, \kappa_N) = (0.2, 0.0)$. **(h)** $(\kappa_T, \kappa_N) = (0.0, 0.2)$. **(i)** $(\kappa_T, \kappa_N) = (0.1, 0.1)$. **(j)** $(\kappa_T, \kappa_N) = (0.2, 0.2)$. The fields shown are just a few examples sampled from the models, both of which contain similar (rotated) projection fields for each of the possible orientation preferences in the central hypercolumn. The circles superimposed on panels d,i are used to characterize retinotopic distance zones for the predictions made in Fig. 5.12.

5.3 Biological and computational anatomy

The computational projection fields generated above contain all the geometrical information needed for predictions about long-range horizontal connections of individual cells (or after some averaging, that of tracer injection sites) in visual cortex. Thus we now turn to the central question: how well do these connectivity maps match the available data about projection fields in visual cortex? In particular, do they make better predictions than those arising from colinearity or association field models (Section 5.1)?

To answer these questions, we focused on anatomical studies that report population statistics [Malach et al., 1993; Bosking et al., 1997] and compared their data to predictions produced by performing "computational anatomy" on our model[2]. We randomly sampled the population of model-generated fields analogously to the way anatomists sample cells, or injection sites, in neural tissue and computed both individual connection distributions and their population statistics. To generate robust predictions we repeated these sampling procedures many times and calculated the expected values and standard errors of the frequency distribution. Since RF responses do not indicate whether or not the stimulus pattern is a curve (1D) or a texture (2D), it was necessary at this point to consider continuations for both. Thus we repeated all computations for both the visual flow and curve compatibilities.

5.3.1 Computational anatomy predicts biological measurements

Fig. 5.8 illustrates the main results computed from our models, and compares them to the corresponding anatomical data reported in the literature [Malach et al., 1993; Bosking et al., 1997]. The agreement of the computational process to the biological data is striking, both qualitatively and quantitatively. As with the association field, our model correctly predicts the spread of the pooled distribution with similar peak height (approximately 11% for orientation resolution of 10 degrees) and a similar orientation offset at which it crosses the uniform distribution (approximately ± 40 degrees). Unlike colinearity and association field models, however, ours predict qualitative differ-

---

[2] Anatomical studies such as Bosking *et al.* [Bosking et al., 1997] and Malach *et al.* [Malach et al., 1993] were preferred to psychophysical or electrophysiological studies because the latter two typically contribute no population statistics and are generally more difficult to interpret directly in terms of the structure of horizontal connections.

ences between distributions of individual neurons, or injection sites, similar to findings in the literature (c.f. Fig 5.1c). Most importantly, our model predicts a consistently non-monotonic standard deviation. At orientation resolution of 10 degrees, both the anatomical data and the computational models exhibit variance local minima at approximately ±30 degrees. This property holds both for a random sample of cells (Fig. 5.9), and for the computational population as a whole (omitted for space consideration).

5.3.2 Curvature quantization and population statistics

As we mentioned, the geometrical model must be quantized both in orientation and curvature before projection patterns can be computed and computational predictions can be made. We fixed the orientation quantization to the same level used in Bosking *et al.* [Bosking et al., 1997]. Curvature quantization, however, is not addressed in the physiological literature and thus it is necessary to examine its effect on the resultant connectivity distributions. We note that even with orientation represented to hyper-acuity levels there are sufficient number of cells to represent such quantization [Miller and Zucker, 1999].

Broad orientation tuning implies discrete orientation quantization and suggests even coarser curvature quantization. The results presented in Figs 5.8 and 5.9 are based on quantizing curvature into 5 classes[3]. This is a likely upper bound, given the broad band-pass tuning of cortical neurons that have been observed [Dobbins et al., 1987; Versavel et al., 1990] and modeled [Dobbins et al., 1989]. However, to study the effect of curvature quantization, we repeated the computations with both a smaller (3) and a larger (7) number of curvature classes. 3 is clearly the lower limit, which may correspond to the tree shrew [Bosking et al., 1997] or other simple mammals, and 7 is more than required computationally (see results of Sec. 4). We found that all of the properties predicted initially remain invariant under these changes. In particular, regardless of quantization level, the pooled distribution remains unimodal, it peaks at zero orientation difference with approximately 11%, it crosses the uniform distribution at ±40°, and it has non-monotonic variance with local minima at ±30° (with somewhat increased variance around zero orientation for higher quantization

---

[3] In the context of curves these 5 classes may be labeled as straight, slowly curving to the left, slowly curving to the right, rapidly curving to the left, and rapidly curving to the right.

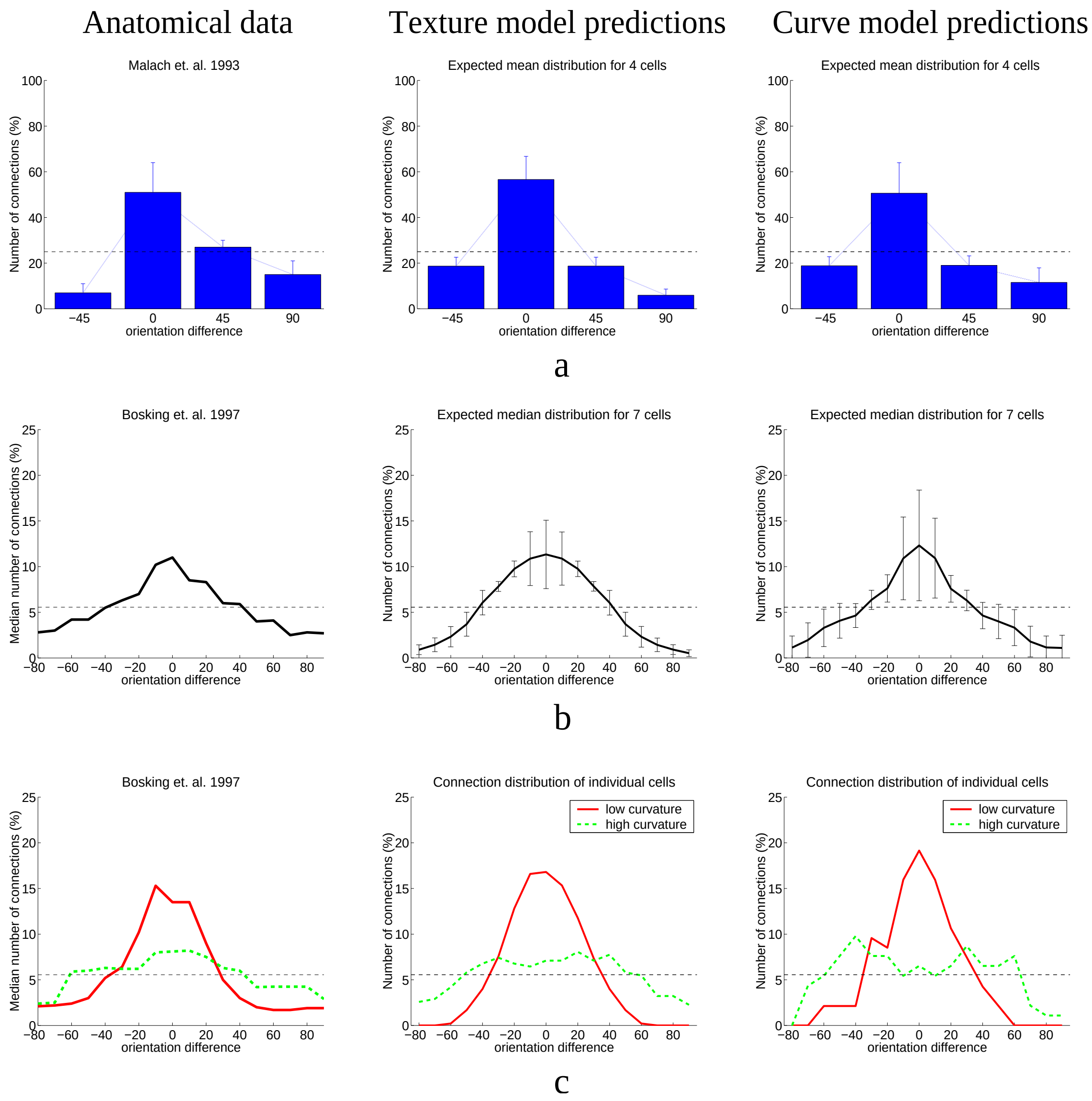


Figure 5.8: **Orientation-domain comparison of anatomical data and model predictions.** In all graphs, dashed horizontal lines represent the uniform distribution and error bars represent ±1 standard error. **(a)** Mean connection distribution of 4 injection sites from Malach *et al.* [Malach et al., 1993] vs. the computational prediction from our models (expected mean, N=4, 100 repetitions). The asymmetry in the pooled distribution measured by Malach *et al.* [Malach et al., 1993] likely derives from a bias at the injection site (see their Fig. 4D) rather than being intrinsic. **(b)** Median distribution of 7 injection sites from Bosking *et al.* [Bosking et al., 1997] against the computational prediction from our models (expected median, N=7, 100 repetitions). Note in particular the similarity in peaks' height, the orientation offset at which the graphs cross the uniform distribution, and the strongly non-monotonic behavior of the variance. **(c)** Two individual injection sites with qualitatively different connection distributions reproduced from Bosking *et al.* [Bosking et al., 1997]. The counterpart computational instances are sampled from our models.

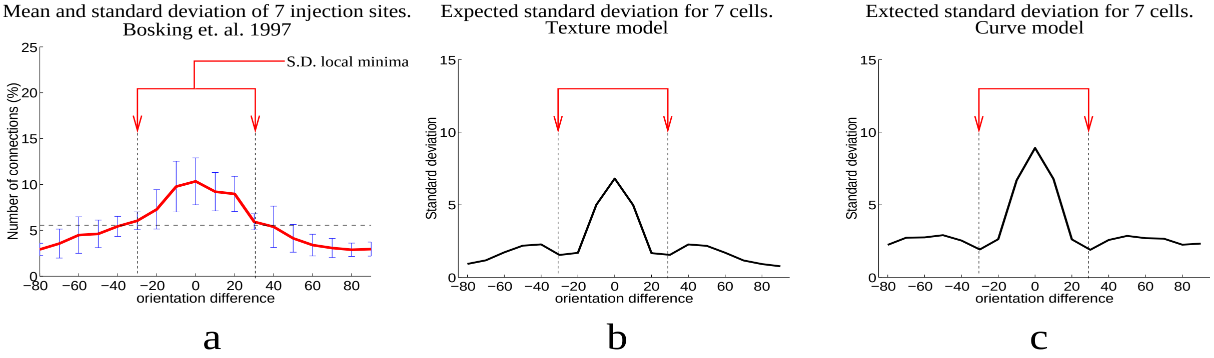


Figure 5.9: **Biological and computational population variances.** Although both the computational and the physiologically measured distributions of the mean are monotonically decreasing, their standard deviation is consistently non-monotonic. **(a)** While Bosking *et al.* [Bosking et al., 1997] used the population *median*, we further analyzed their published data (from 7 injection sites) to find its *mean* and *standard deviation*. It is evident that the s.d. is non-monotonic, with two local minima at ±30 degrees (marked with red arrows). **(b)** Expected standard deviation for the texture model. **(c)** Expected standard deviation for the curve model. Both graphs depict the expected s.d. for 7 randomly selected cells (N=7, 100 repetitions). and both show a similar non-monotonic behavior with pronounced s.d. local minima at approximately ±30 degrees. Note how the computational local minima coincide with the anatomical ones (arrows are copied from panel **a** and overlaid on the computational graphs). Compare also to the s.d. on the median graphs in Fig. 5.8b. Note that as with the distributions themselves, both computational models produce quantitatively similar s.d. results.

levels). Qualitative differences between individual neurons are predicted regardless of the number of curvature classes. All these results are illustrated in Fig. 5.10.

5.3.3 Relationship between cell distribution and connection distribution

Since both anatomical and computational studies must sample the population of (biological or computational) cells to measure the distribution of their horizontal connections, an important consideration is whether the underlying distribution of cells (based on their curvature tuning) can affect the *pooled* distribution of connections[4]. For example, if most cells in the population are tuned for zero or very small curvature, the pooled connection distribution may differ from that of a population dominated by high curvature cells.

---

[4] Obviously, no effect on the connectivity distribution of *individual* cells is expected.

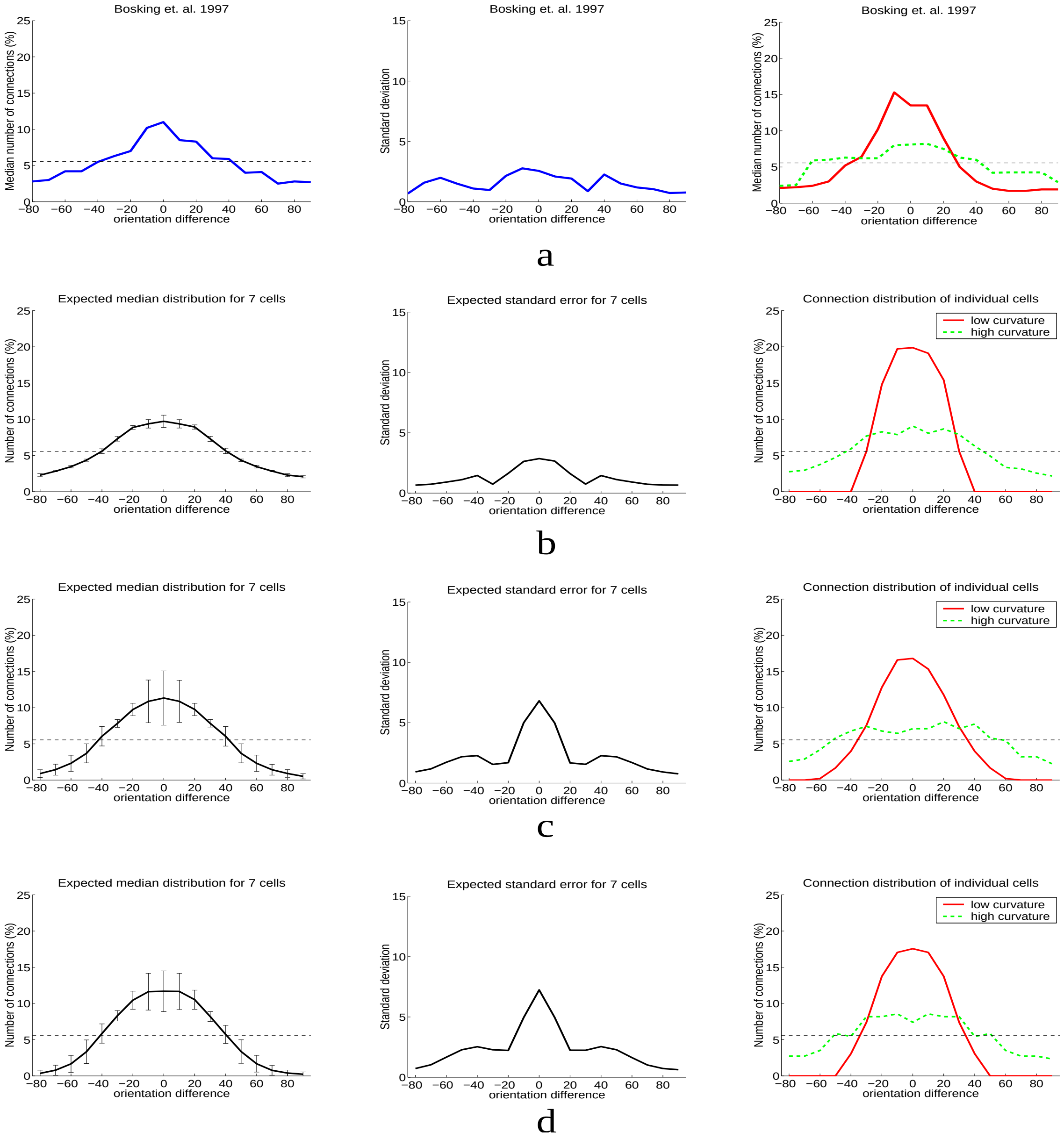


Figure 5.10: **Insensitivity of connectivity distribution to curvature quantization. (a)** Anatomical data from Bosking *et al.* [Bosking et al., 1997] shown for comparison with the computational predictions. **(b)** Computational predictions with 3 curvature classes. **(c)** Computational predictions with 5 curvature classes. **(d)** Computational predictions with 7 curvature classes. In all cases, left column depicts the expected median for 7 cells (bars are 1 s.d.), middle column depicts the expected standard deviation for 7 cells, and the right column shows two qualitatively different distributions from two different cells. For space considerations we show the results from the texture model only.

Unfortunately, not much data about such distribution of cells is yet available, partially because anatomists need not assume any particular cells' distribution for their measurements of projection fields. Some statistical studies of edge correlations in images have been attempted [Sigman et al., 2001; Geisler et al., 2001; August and Zucker, 2000], but these co-occurrence measurements neither depend on curvature nor do they necessarily imply any particular distribution of cells at the computational level. Nevertheless, our model raises the possibility of a curvature bias effect, which we discuss next.

The results presented in Figs 5.8 to 5.10 are based on the assumption that cells of different curvature tuning (or put differently, of different connectivity patterns) are distributed uniformly. Such an assumption follows naturally from the mathematical abstraction that allocates the same number of computational units to each portion of $\mathbb{R}^2 \times S^1$. However, the repeated findings of colinear-like interactions in primary visual cortex (see Sec. 5.1) suggests that cells tuned to zero or very small curvature(s) may be more frequent than others. Would such a bias in the distribution of cells affect the predictions made from our models?

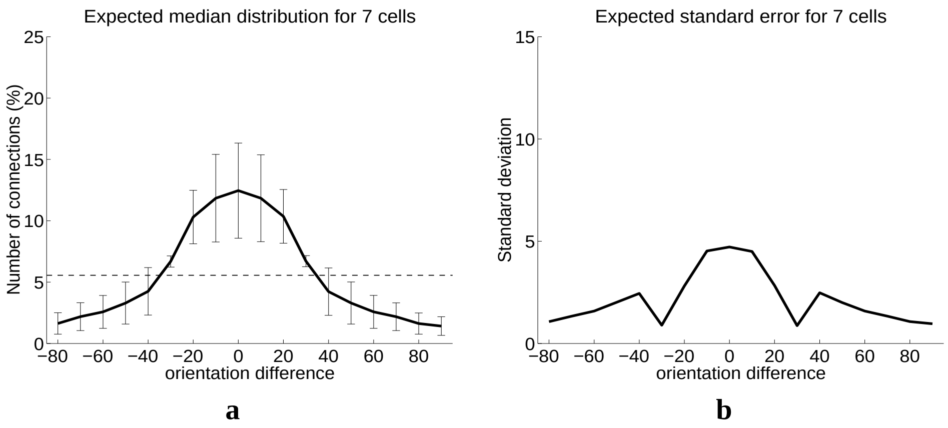


Figure 5.11: **Insensitivity of connectivity distribution to cells' distribution.** An empirical evaluation provides a statistical confirmation that the properties of our models persist even when a curvature-based bias is introduced on the population of cells. Illustrated here are the results from a population of cells distributed normally with $\sigma_T = \sigma_N = 0.15$ in which cells of zero curvature tuning are 8 times more frequent than those of maximal curvature tuning. The graphs in this figure correspond to the texture model with curvatures quantized to 3 classes each. **(a)** Expected median of 7 cells. Error bars are ±1 standard deviation. **(b)** Expected standard error for 7 cells. Compare both graphs to Fig. 5.8b, 5.9, and 5.10b.

We examined this question by repeating the computational anatomy process described above on a non-uniformly distributed population of cells. All computations were done on the more general 2D (texture) model, with population of cells distributed non-uniformly using 2D normal distribution of zero mean and variances σ_T and σ_N in the κ_T and κ_N dimensions, respectively. This distribution reflects a frequency of cells based on their curvature tuning and it emphasizes cells of zero or low curvatures with respect to all others. Fig 5.11 illustrates one example of the resultant statistical measures for normal cell distribution of $\sigma_T = \sigma_N = 0.15$, which induces an isotropic normal distribution, with zero curvature cells 8 times more frequent than cells tuned to the maximum value of curvatures. As expected, this non-uniform distribution slightly increased the peak height of the population statistics, but otherwise, all other features that were predicted from the uniform cell distribution, in particular the non-monotonic variance, were fully preserved. Similar results were obtained with other values of σ_T and σ_N, and for all quantization levels of the curvatures (Section 5.3.2). Thus, our conclusions are not biased by an implicit assumption about curvature-dependent distribution of cells.

5.4 Discussion

The findings presented from our computational anatomy support the functional identification of the long-range horizontal connections with those obtained mathematically. However, the question of why a visual flow integration is necessary becomes unavoidable, and we believe this issue is more than just formal mathematics. Clearly, certain physiological and psychophysical findings suggest the perceptual integration of 2D flow structures rather than 1D curves. These include iso-orientation side facilitation [Adini et al., 1997], functional and anatomical connections between retinotopically *parallel* receptive fields [Ts'o et al., 1986; Gilbert and Wiesel, 1989], and roughly isotropic retinotopic extent of projection fields [Malach et al., 1993; Sincich and Blasdel, 2001]. Although this class of patterns may seem less important than curves as a factor in perceptual organization, their perceptual significance has been established, as we discussed in length in Sec. 1.2. Furthermore, recent computational vision research, and this thesis, implicate them in the analysis of visual shading [Lehky and Sejnowski, 1988; Huggins et al., 2001], as well as in color (Ch. 6).

Whether or not projection patterns of cells in primary visual cortex come in different flavors (i.e., curve vs. flow integration) is an open question. To answer it one is likely to exploit the many physiologically measurable differences between these classes of projection patterns as suggested by Fig. 5.7. Unfortunately, the statistical data obtained so far do not distinguish between the two (curve and flow) integration models; without a spatial dimension, the statistical differences between the two models in the orientation domain are too fine to measure relative to the accuracy of current laboratory techniques.

Until full spatio-angular data are obtained, however, the inclusion of even *weak* spatial information is sufficient to generate further testable predictions. In particular, incorporating the retinotopic *distance* between linked cells into the statistics (or estimating it from their cortical distance) can produce predictions regarding the dependency of the distribution's spread and shape on the integration distance. This can be done by repeating the computational anatomy described before, only focusing on connections within a given retinal annulus, as is illustrated in Fig. 5.12. Interestingly, some verification for these predictions can be seen in the measurements of Kisvárday et al. [1997], whose top row in their Fig. 9 shows developing peaks resembling the ones in Fig 5.12b,c. Similar annular analyzes, which focus on sectors of the annuli in directions other than parallel to the RF's preferred orientation, provide measurable differences between curve and flow projection fields.

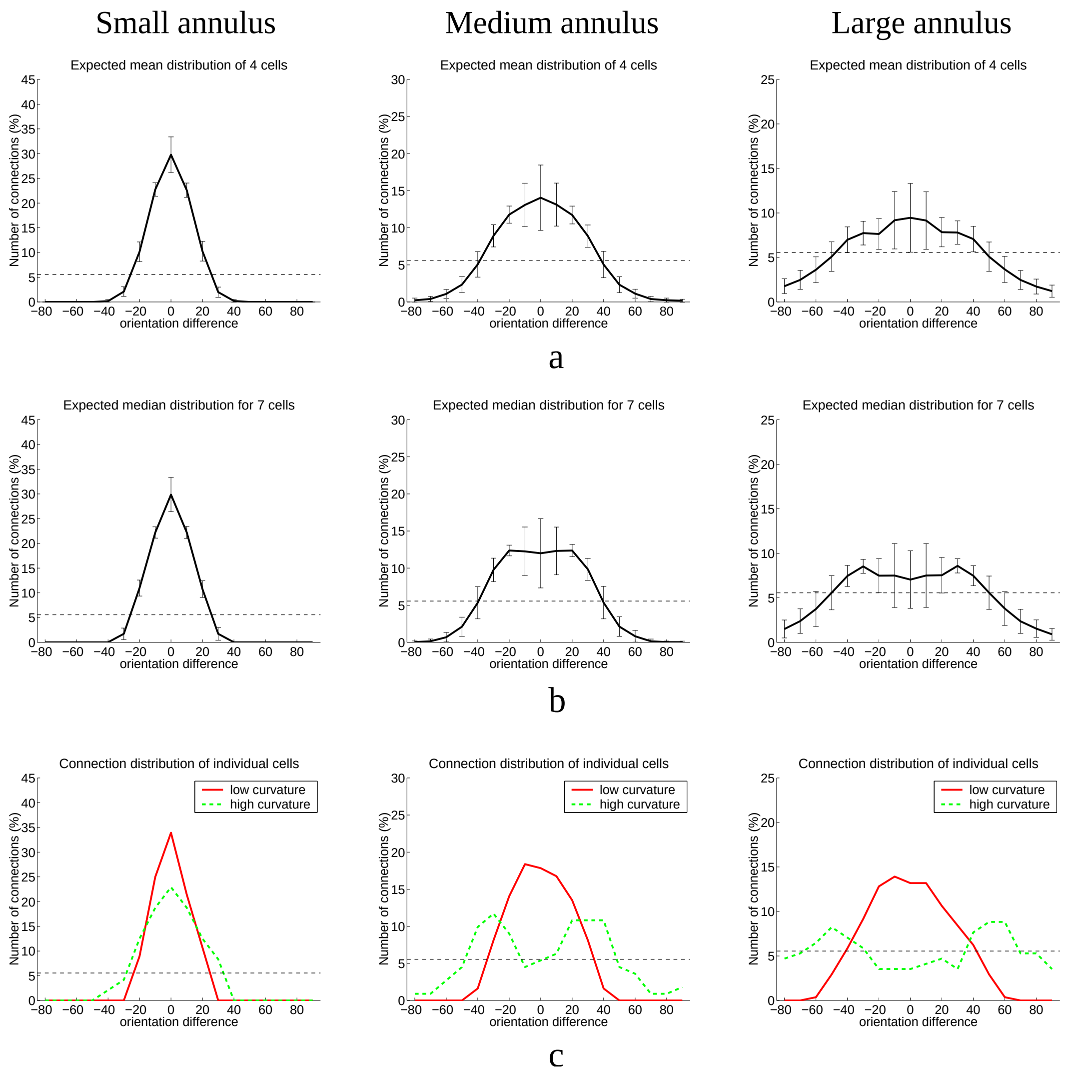


Figure 5.12: **Predictions of connectivity distributions by retinotopic distance.** Left, middle, and right columns correspond to predictions based on small, medium, and large annuli, respectively (circles in Fig. 5.7d,i). All annuli refer to distances *beyond* one orientation hypercolumn thus the small annulus should *not* be confused with distances less than the diameter of one hypercolumn. In a,b the same sampling procedure and the same sample set sizes described in Fig. 5.8 were repeated. **(a)** Expected mean distribution and standard error (N=4, 100 repetitions). Note the spread with increased retinotopic distance. **(b)** Expected median distribution and standard error (N=7, 100 repetitions). Note the developing symmetric peaks that further depart from the iso-orientation domain as the spatial distance increases. The correspondence of these peaks to the minima of the standard error is remarkable, thus designating them as statistical “anchors” suitable for empirical verification. **(c)** Individually tuned cells show the qualitative difference between distributions of high and low curvature cells. Evidently, the distributions of high curvature cells (dashed green graphs) are the ones that lead to the peaks mentioned in panel **b** above.

5.5 Summary

We have shown how the mathematical analysis and computational models developed earlier in this thesis predict and better explain both the pooled distribution of long-range horizontal connections in primary visual cortex and the connectivity distributions of individual cells/injection sites. These findings suggest a functional identification of the long-range horizontal connections with those obtained mathematically and imply, first and foremost, that curvature is at the heart of the contextual computation carried out by these horizontal connections. Furthermore, these findings indicate, for the first time, that visual flow continuation, in particular that of texture and shading, can serve as alternative and complimentary explanations to the functional role of long-range horizontal connections. While most researchers have limited this to curve integration via colinearity, we thus conclude in an enlarged context: differential geometry provides a foundation for connections in visual cortex that predicts both dependency on curvature(s) and an expanded functional capability. Since both curve and texture/shading flow continuation follow from the geometry, and both are important for vision, more targeted experiments are required to articulate their neural realization.

Chapter 6

A perceptual organization approach to color image denoising

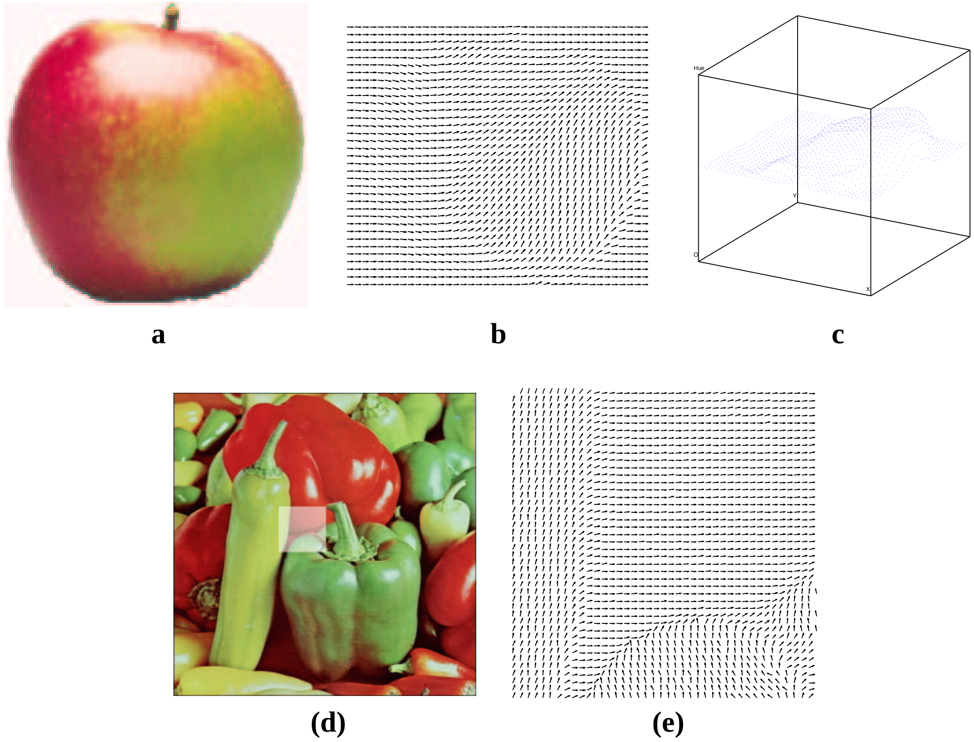
Chapter 5 above examines the relationship of our geometrical and computational work to computational neuroscience, and thus completes a multidisciplinary exploration of visual flows from four interrelated perspectives: geometrical, psychophysical, computational, and physiological. We now turn back one more time to the computational corner of our pyramid of disciplines to discuss yet another application of our work to (an apparently) different domain – the analysis of color images. The connection of our work to the domain of color already was made in the Introduction (Sec. 1.1.3) and in the present chapter we expand it and look closer at the particular application of color image denoising.

Following early attempts to denoise color images through independent smoothing of the RGB channels, practically all contemporary approaches focus on a variety of filtering processes applied appropriately to the color *vectorial* data. While some studies explore vector median and directional filters [Astola et al., 1990; Trahanias and Venetsanopoulos, 1993], most color image enhancement approaches are based on some form of anisotropic diffusion [Perona and Malik, 1990; Weickert, 1997; ter Haar Romeny, 1994], either on an explicit vectorial representation of the color, or based on differential geometrical properties of a manifold representation in a higher dimensional space [Tang et al., 2001; Sapiro and Ringach, 1996; Sochen et al., 1998; Kimmel et al., 2000].

Many of these efforts, as well as other related approaches [Boccignone et al., 2002; Blomgren and Chan, 1998; Chan and Shen, 1999], use the perceptually-motivated HSV color representation in which the achromatic (lightness, or value) and chromatic components of the color image are explicitly separated, and the latter is further described with two independent variables for hue and saturation. This representation has deep roots in the psychology of perception and the subjective experience of color, from Hering's opponent hues theory [Hering, 1878/1964] and Munsell's book of colors [Munsell, 1905], to accumulating psychophysical evidence on color sensitivity [De Valois and De Valois, 1990]. Combined with neurophysiological findings for color opponent cells, color specificity in V4, and perceptual impairments such as achromatopsia [Spillmann and Werner, 1990], it is now commonly speculated that neural structures that explicitly encode the hue, saturation, and lightness do exist and will eventually be found [Palmer, 1999b, p. 120].

Of the three perceptual dimensions of HSV, the hue is periodic and can be represented as an angle (Fig. 1.4). Although it is not spatial in nature, this angle associates a *geometric* meaning to the hue component of the color (or alternatively, to the chromatic component as a whole) and thus suggests a geometrical treatment. Although this idea is not completely new and has been exploited in the color diffusion literature through the diffusion of orientations (e.g., [Tang et al., 2001; Chan and Shen, 1999]), the computational study of visual flows proposed in this thesis suggests a completely different approach - one which is based on perceptual organization and a principle of *color good continuation*.

The basic idea is the following: if hue behaves like orientation, two nearby color patches should be mutually coherent if their hue orientations posses mutual *geometrical* good continuation. This amounts to assessing the degree to which each hue measurement is *geometrically* compatible with the context in which it is embedded, and whether or not that context is part of a single whole. Naturally, this suggests using the geometrical and computational infrastructure developed earlier to define two hue curvatures and a local model for the behavior of the color. This, in turn, specifies consistency constraints between nearby color measurements and defines a distributed contextual computation that replaces noisy pixels with coherent ones. Just like for texture and shading (Ch. 4), this process provides robust noise removal that preserves line discontinuities and singularities. These properties are specifically important given the observations that (1) the hue is typically piecewise smooth over the image, and (2) that it may vary greatly, albeit smoothly, even *within* perceptually coherent objects (Fig 6.1). In noisy color images, the noisy pixels do not posses the required good continuation with their context, thus can be detected, suppressed, and replaced.



a **b** **c**

(d) **(e)**

Figure 6.1: **Color images and hue fields.** These fields, which typically are piecewise smooth, can vary even within perceptually coherent objects. **(a)** A natural image of an apple with varying hue. **(b)** The corresponding hue field changes smoothly across the apple's surface. **(c)** A 3D representation of the hue field, where hue is represented as height. Identifying the top face with its bottom one leads to the space $XY\mathcal{H} \triangleq \mathbb{R}^2 \times \mathcal{S}^1$ in which the image's hue is a submanifold. **(d)** A natural image of peppers with a region of interest (ROI). **(e)** The hue field of the peppers image in the ROI is piecewise smooth. In general, occlusion boundaries between objects in the world induce hue singularities (up to some blurring from the imaging process) that must be preserved while noisy images are denoised.

6.1 Color geometry, hue curvatures, and hue coherence

Consider the HSV color space, in which a color image is a mapping $\mathcal{C} : \mathbb{R}^2 \to \mathcal{S}^1 \times [0,1]^2$, where $\mathcal{S}^1$ is the unit circle. The hue component across the image is a mapping $\mathcal{H} : \mathbb{R}^2 \to \mathcal{S}^1$ and

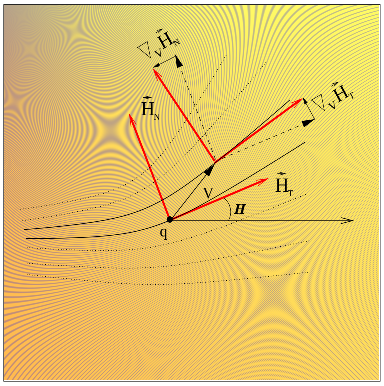


Figure 6.2: **The hue frame field and covariant derivatives.** Color represents the underlying, smoothly changing, hue values of an image patch. Parallel lines represent the corresponding hue field, and the differential geometry follows the general case of visual flows. See Sec. 2.3 and Fig 2.2 for additional explanations.

thus can be represented as a unit length vector field over the image plane, henceforth called the *hue field*. In many natural images, as well as visual artifacts, this hue field is typically piecewise smooth (Fig. 6.1) with orientation singularities that correspond to significant events in the scene (e.g., occlusion). Denoising color images requires suppressing spurious measurements *within* the coherent segments while preserving the boundaries *between* them. To achieve this goal, a notion of color coherence is first needed and since the hue field is a visual flow, the same geometrical analysis of Ch. 2 applies. In this section we briefly review its main points.

By attaching a frame field $\{\mathcal{H}_T, \mathcal{H}_N\}$ to each point in the image domain, we not only represent the hue vector itself, but also provide a local coordinate system in which all other vectors can be represented in a natural, object centered view (Fig. 6.2). Perhaps the most important vectors (other than the frame vectors themselves) are the covariant derivatives of $\mathcal{H}_T$ and $\mathcal{H}_N$. These covariant derivatives represent the initial rate of change of the frame in any given direction $\mathbf{v}$, a quantity which in the $\{\mathcal{H}_T, \mathcal{H}_N\}$ coordinates is captured by Cartan's *connection equation* (c.f. Eq. 2.3):

$$\begin{pmatrix} \nabla_{\mathbf{v}}\mathcal{H}_T \\ \nabla_{\mathbf{v}}\mathcal{H}_N \end{pmatrix} = \begin{bmatrix} 0 & w_{12}(\mathbf{v}) \\ -w_{12}(\mathbf{v}) & 0 \end{bmatrix} \begin{pmatrix} \mathcal{H}_T \\ \mathcal{H}_N \end{pmatrix} \tag{6.1}$$

The coefficient $w_{12}(\mathbf{v})$ is a *linear* function of the tangent vector $\mathbf{v}$, and thus is fully representable with two scalars – the *hue curvatures* – at each point:

$$\begin{aligned}\kappa_T &\overset{\triangle}{=} w_{12}(\mathcal{H}_T)\\ \kappa_N &\overset{\triangle}{=} w_{12}(\mathcal{H}_N)\ \ .\end{aligned} \tag{6.2}$$

These two curvatures can be expressed in terms of the hue itself and its gradient (c.f. Eq. 2.7)

$$\begin{aligned}\kappa_T &= \nabla\mathcal{H}\cdot(\cos\mathcal{H},\sin\mathcal{H})\\ \kappa_N &= \nabla\mathcal{H}\cdot(-\sin\mathcal{H},\cos\mathcal{H}),\end{aligned} \tag{6.3}$$

where $\mathcal{H}$ can be computed from the RGB channels via the standard transformation

$$\mathcal{H} = \tan^{-1}\left(\frac{2R-G-B}{(G-B)\sqrt{3}}\right)\ \ . \tag{6.4}$$

Since local changes in the hue are characterized (up to Euclidean transformation) by a pair of curvatures, it is natural to conclude that nearby coherent hue values should relate to each other based on these curvatures. Put differently, measuring a particular curvature pair $(\kappa_T(\mathbf{q}),\kappa_N(\mathbf{q}))$ at a point $\mathbf{q}$ should induce a field of coherent hue values, i.e., a hue function$\tilde{\mathcal{H}}(x,y)$, in the neighborhood of $\mathbf{q}$. Coherence of $\mathcal{H}(\mathbf{q})$ to its spatial context can then be determined by examining how well $\mathcal{H}(x,y)$ fits $\tilde{\mathcal{H}}(x,y)$ around $\mathbf{q}$.

Clearly, the local coherence model $\tilde{\mathcal{H}}(x,y)$ should be a function of the local hue curvatures $(\kappa_T(\mathbf{q}),\kappa_N(\mathbf{q}))$, it should agree with these curvatures at $\mathbf{q}$, and it should extend around $\mathbf{q}$ according to some variation in both curvatures (as a consequence of the propositions above). Naturally, we adopt the same general model developed earlier in Ch. 2, which in $XY\mathcal{H}$ takes the form of a right helicoid

$$\tilde{\mathcal{H}}(x,y) = tan^{-1}\left(\frac{\kappa_T(\mathbf{q})x+\kappa_N(\mathbf{q})y}{1+\kappa_N(\mathbf{q})x-\kappa_T(\mathbf{q})y}\right)\ \ . \tag{6.5}$$

Recall that this local model possesses many properties that suit good continuation, and in particular it is both a minimal surface in the $(x,y,\tilde{\mathcal{H}}(x,y))$ surface representation, and a critical point of the p-harmonic energy for *all* p. Most importantly, it is the unique model that does not bias the changes in one hue curvature relative to the other, i.e., it satisfies

$$\frac{\kappa_T(x,y)}{\kappa_N(x,y)} = \text{const} = \frac{\kappa_T(\mathbf{q})}{\kappa_N(\mathbf{q})}\ \ .$$

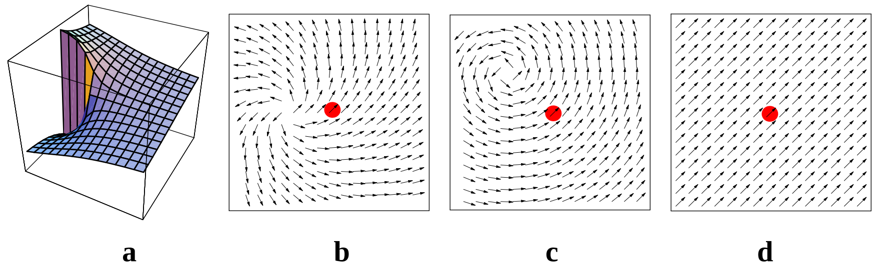


Figure 6.3: **A local model for coherent hue.** This model can be depicted, for example, as a surface (right helicoid) in $XY\mathcal{H}$ or a hue field in the image plane. In both cases, different orientation and curvature tunings at the central pixel (marked with red in the hue fields) yield different local behaviors around it. **(a)** A representation as a surface in $XY\mathcal{H}$ yields a right helicoid that wraps itself around $[-\pi, \pi)$ (hence the apparent discontinuity). **(b)** A hue field representation for $\mathcal{H} = 45^\circ$ at the central (red) pixel and $\kappa_T > 0$ and $\kappa_N > 0$. **(c)** Same with $\kappa_T > 0$ and $\kappa_N = 0$. **(d)** Same with $\kappa_T = 0$ and $\kappa_N = 0$.

Examples of the model for few different curvature tuning are illustrated in Fig 6.3. Note that unlike for texture flows, the local model for the hue function is not a *double* helicoid since the hue function is 2π-periodic while texture flows are π-periodic. The basic properties (Sec. 2.7) and proofs carry through, however.

6.2 Contextual color denoising

Having a model that represents coherence of hue values, a relaxation labeling network that denoises hue fields can be developed in a way described in Ch 4. Few differences do apply, however, and we discuss them briefly in this section.

The domain of color images dictates that every pixel has exactly one value. In other words, neither zero nor multiple hue values are permitted at each pixel. This observation has two important consequences on the architecture of the relaxation network. Firstly, the network need not be five dimensional but rather 2D only with nodes $i = (x, y)$ that represent image pixels. Secondly, unlike for texture, or even shading, we need *not* maintain a “no flow” label. The set of labels at each node can thus be drawn from the set $\Lambda = \{(\mathcal{H}, \kappa_T, \kappa_N) \mid \mathcal{H} \in [-\pi, \pi)\ ,\ \kappa_T, \kappa_N \in [-K, K]\}$ after it has been quantized appropriately. The fact that all these labels compete with each other at each node via the normalization constraint $\sum_{\lambda\in\Lambda} p_i(\lambda) = 1$ (see Eq. 4.1) eliminates the need for inhibitory interactions (p. 105) and suggests that ambiguous assignments at convergence can and should be resolved by a straight-forward confidence comparison (i.e. winner-takes-all).

As before, what governs the dynamics of the relaxation process, and ultimately its convergence state, is the compatibility relationships $r_{ij}(\lambda, \lambda')$ between different labels at different nodes. In our case, these compatibilities represent the degree to which two nearby pixels have consistent hue values. Derived from the right helicoid, these compatibility structures are similar to those of Sec. 4.2.3. The main difference is the fact that now they are based on 2π periodicity. Examples of a variety of compatibility fields for a particular hue value and different combinations of hue curvatures are illustrated as hue fields in Fig. 6.4 and as hue values in Fig. 6.5. An interesting property of these fields is their relationship to their curvature tuning. Note how higher curvature values introduce greater variations and additional colors into the fields, and how changing the curvature tuning while keeping constant the "total variation" $\kappa_T^2 + \kappa_N^2$ amounts to a rotational transformation of the field.

The 2D network architecture and the absence of a "no flow" label provide an additional advantage with regard to the behavior of the process around line discontinuities in the hue field. In fact, with the simplified architecture, stability along such discontinuities does not require support normalization (see Sec. 4.2.4) or any other ah-hoc mechanisms. Rather, such stability is achieved as a by-product of the competition of labels at each node. Indeed, the distribution of labels on the two sides of a (locally straight) discontinuity ensures that after each side becomes coherent, an equilibrium is achieved by means of symmetry.

Although support normalization is no longer needed for boundary stability, we nevertheless keep it in the system for a different reason, namely, to prevent the network from transforming large scale incoherent structures into coherent ones. Indeed, without a label to represent "no flow", such incoherent structures cannot be rejected. Unless no minimal acceptable support is defined, and without inhibitory interactions, even minute accidental evidence for coherent structure in these regions can slowly gain confidence. Although the pace of these changes is negligible relative to the elimination of noise in otherwise coherent regions, normalizing the support can prevent it altogether. The advantage to color denoising is that highly textured regions remain fixed at the same time where nearby coherent, but noisy, regions can be denoised.

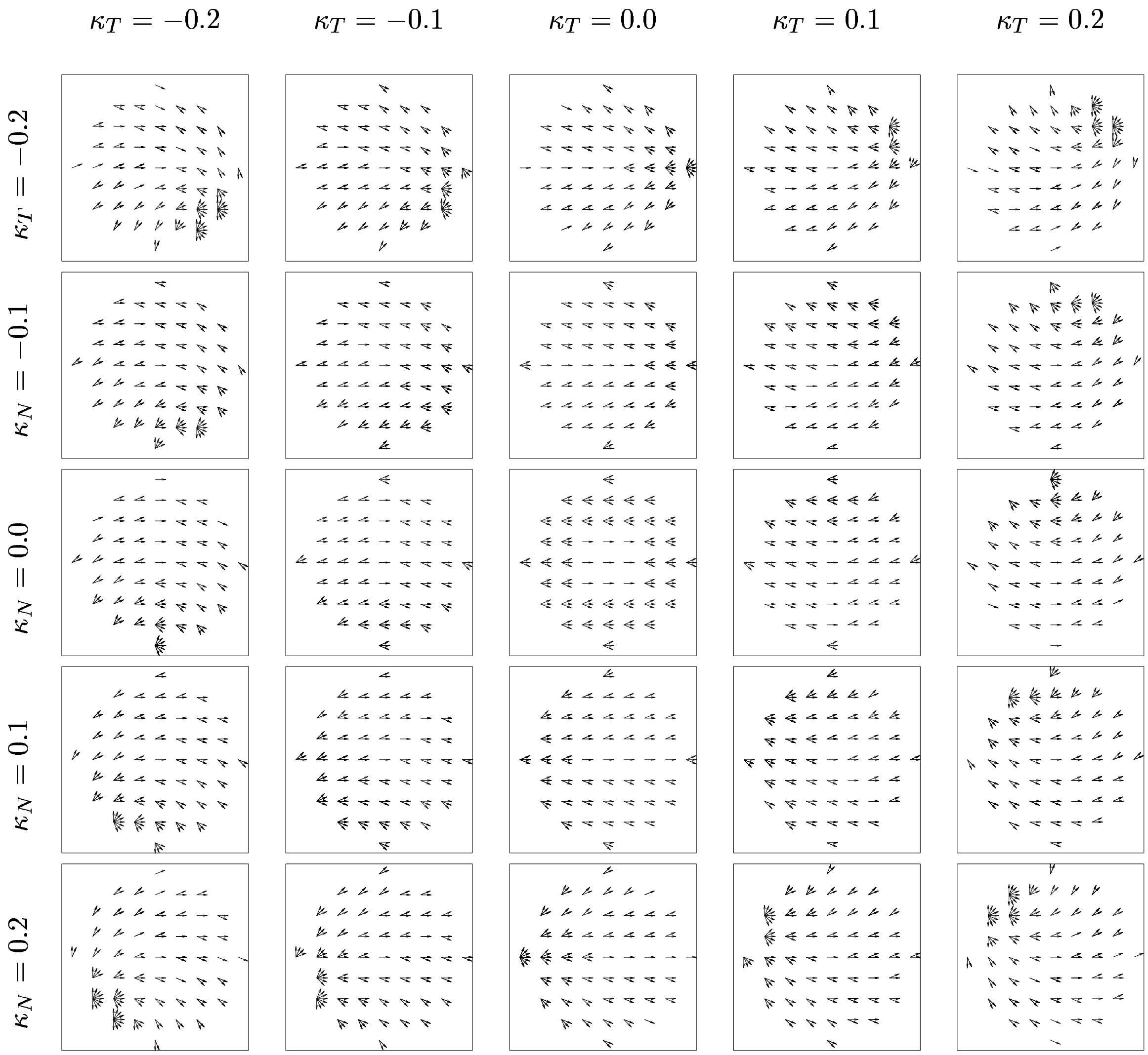


Figure 6.4: **Discrete hue compatibility fields.** Shown here is a collection of 9×9 hue compatibility fields for $\mathcal{H} = 0^\circ$ (the red color), where both κ_T and κ_N are selected from the set $\{-0.2, -0.1, 0.0, 0.1, 0.2\}$. Each field represents all compatible hue values in the neighborhood of the central label. Due to quantization, especially that of curvatures, a given label at the center may be compatible with more than one label at the same nearby location in its neighborhood, an outcome depicted by multiplicity of vectors at certain positions in the fields. The interpretation as hue values in shown in the next figure.

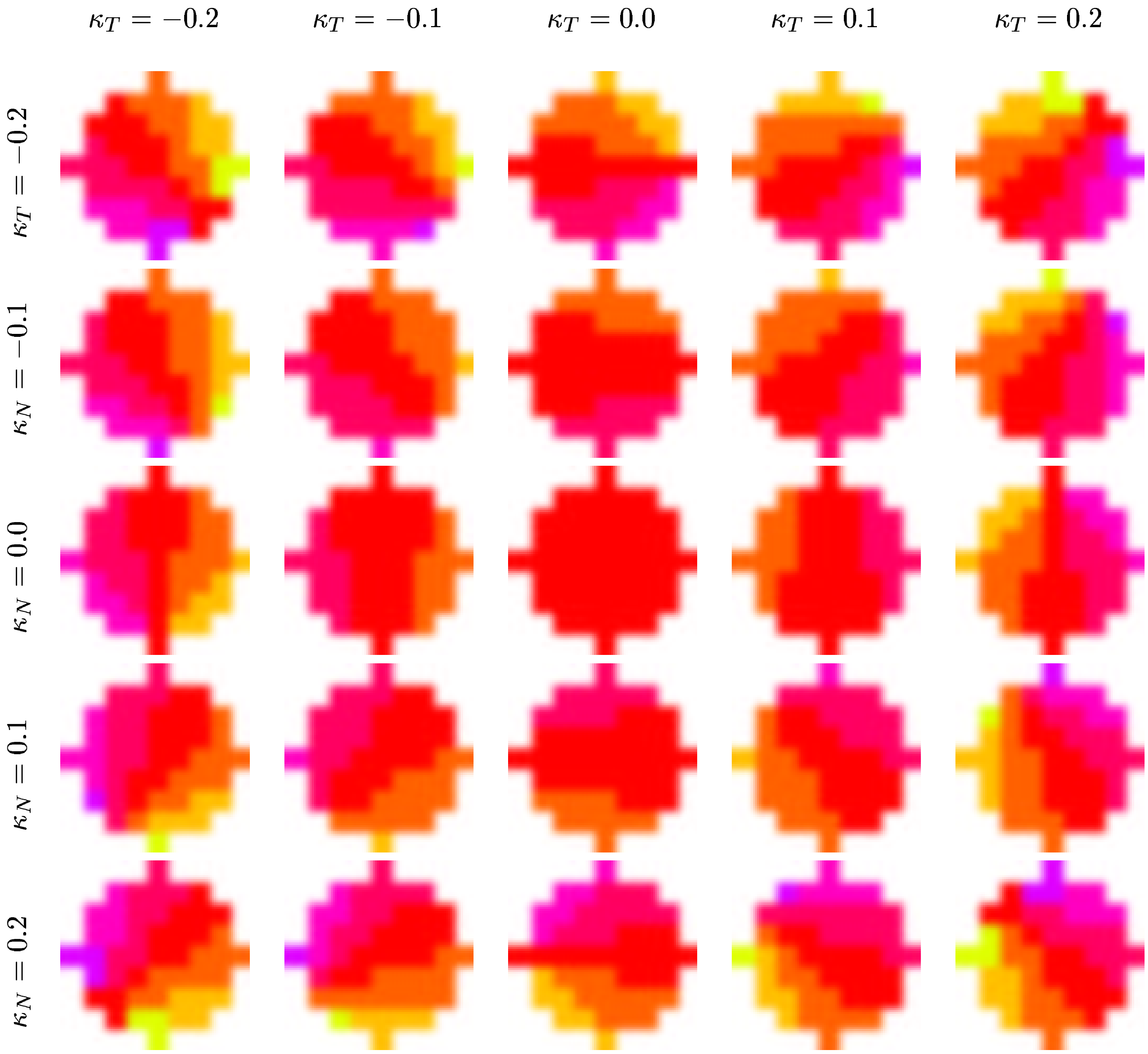
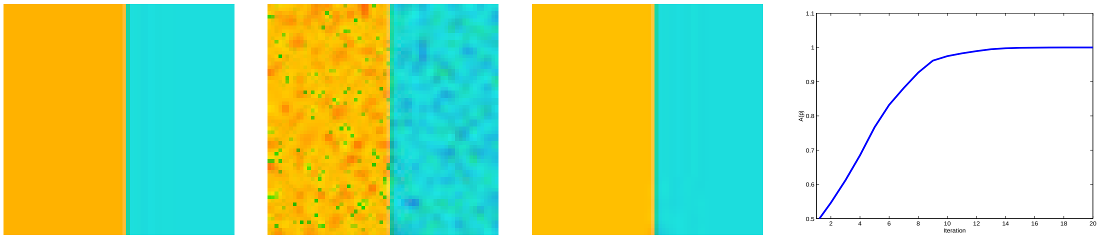


Figure 6.5: **Discrete hue compatibility patches.** The same compatibility fields in the previous figure are now depicted as color pixels and hue patches. Since the multiplicity of compatible hue values cannot be depicted with the color representation, the patches in this figure show only the most compatible hue value at each position.



a **b** **c** **d**

Figure 6.6: **Color denoising of synthetic step edge.** **(a)** A color step edge. **(b)** A noisy step edge. **(c)** The result of 20 iterations of relaxation labeling. The noise is completely eliminated while the edge structure is preserved. **(d)** is the graph of the average local consistency as a function of iteration. it illustrates that convergence on this input was effectively achieved at the 12^{th} iteration, when the assignments $p_i(\lambda)$ reached a steady state.

6.3 Experimental results

We applied the relaxation labeling network for denoising color images on a variety of synthetic and natural inputs. In all cases we quantized the hue uniformly to 32 equivalence classes and curvatures to 5 (as in Fig. 6.4). Step size was set to $\delta = 0.5$ and ambiguous hue assignments at steady state were resolved with the maximum confidence selection mentioned above. Since this selection necessarily resulted in one of the quantized hue values, a much finer orientation quantization was required here than in the texture case in order to avoid undesired visual artifacts.

Fig. 6.6 illustrates the relaxation behavior around a step color edge. A noiseless steady state was achieved with few iterations only, after which no additional changes occurred. A more interesting result is presented in Fig. 6.7 whose input represents an edge configuration that a typical nonlinear diffusion is likely to distort. Here the hue profiles along the two sides of the perceptual edge conforms to the cross configuration discussed in Sec. 4.1.1. Just as the distortion that occurs with shading flow (Fig. 4.9), the increased diffusion that occurs in the cross region leads to the collapse of the edge from the inside out and to a distortion of the underlying structure. As is illustrated in the figure, this does not happen with our approach.

Fig. 6.8 illustrates the results of our denoising approach on the Apple image from Fig. 6.1. To emphasize the contribution of our method, which acts on the oriented hue channel, here we show the result for a corruption along the hue dimension. Unlike typical diffusion processes, which contain

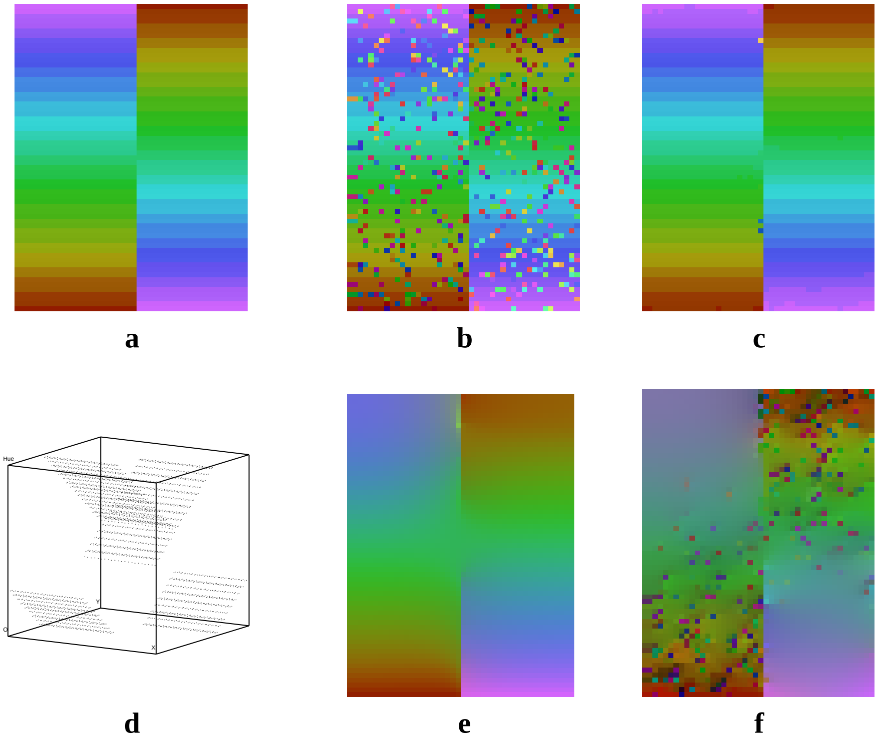


Figure 6.7: **Color denoising of synthetic cross edge.** This unique configuration is most apparent in panel d, where the hue values are represented as height (and the top face of the cube should be identified with its bottom). **(a)** The original cross edge. **(b)** A noisy cross. **(c)** Results of 30 iterations. **(d)** The original hue field in $XY\mathcal{H}$. **(e)** Results of 500 iterations of the Beltrami flow on the original image. **(f)** Results of 500 iterations of the Beltrami flow on the noisy image. While here we used the Beltrami flow for color images [Sochen et al., 1998; Kimmel et al., 2000], similar results were obtained with other orientation diffusion schemes (e.g., [Tang et al., 2000; Sapiro and Ringach, 1996]).

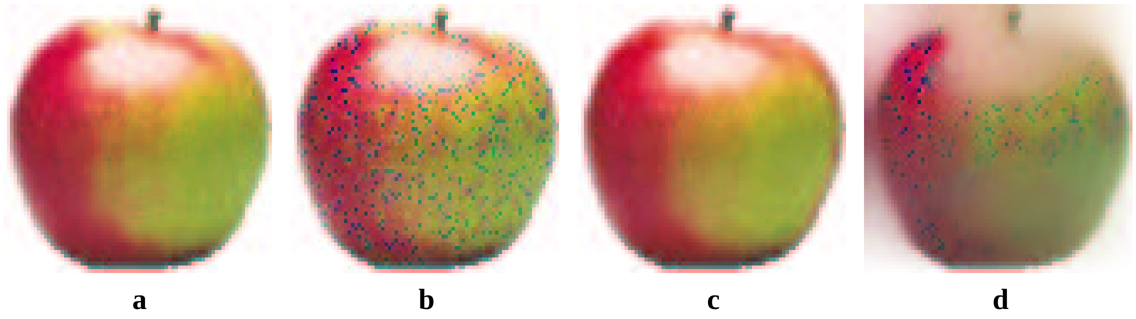


Figure 6.8: **Color denoising of a natural image with varying hue.** **(a)** Original image (from Fig. 6.1). **(b)** Noisy image. **(c)** Result of 25 denoising iterations. **(d)** Result of color diffusion [Sochen et al., 1998; Kimmel et al., 2000]. Note that although the noise is not yet removed, the color structure across the apple is completely blurred and distorted, as it is the case across the apple's boundaries. Compare to panel c.

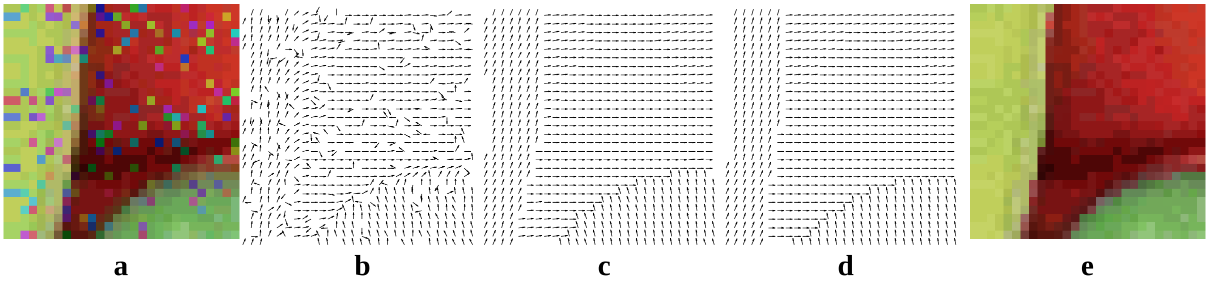


Figure 6.9: **Stability of hue field after convergence of the local average consistency.** **(a)** A detail from noisy Peppers image from the same ROI marked in Fig. 6.1 (but further zoomed in to allow better depiction of the hue field). **(b)** The corresponding noisy hue field. **(c)** The result of 20 iterations. **(d)** The result of 50 iterations is virtually identical. **(e)** The denoised color image.

no natural indicators for stopping the progress along the scale space, in our approach convergence is readily signaled by convergence of the average local consistency (c.f. Fig. 6.6D). The stability of the process once convergence has been achieved is demonstrated in Fig. 6.9.

Finally, Fig. 6.10 demonstrates the result of using our approach on a variety of natural images. In all cases the noise was completely removed, and convergence achieved, after 30 iterations of relaxation labeling or less.

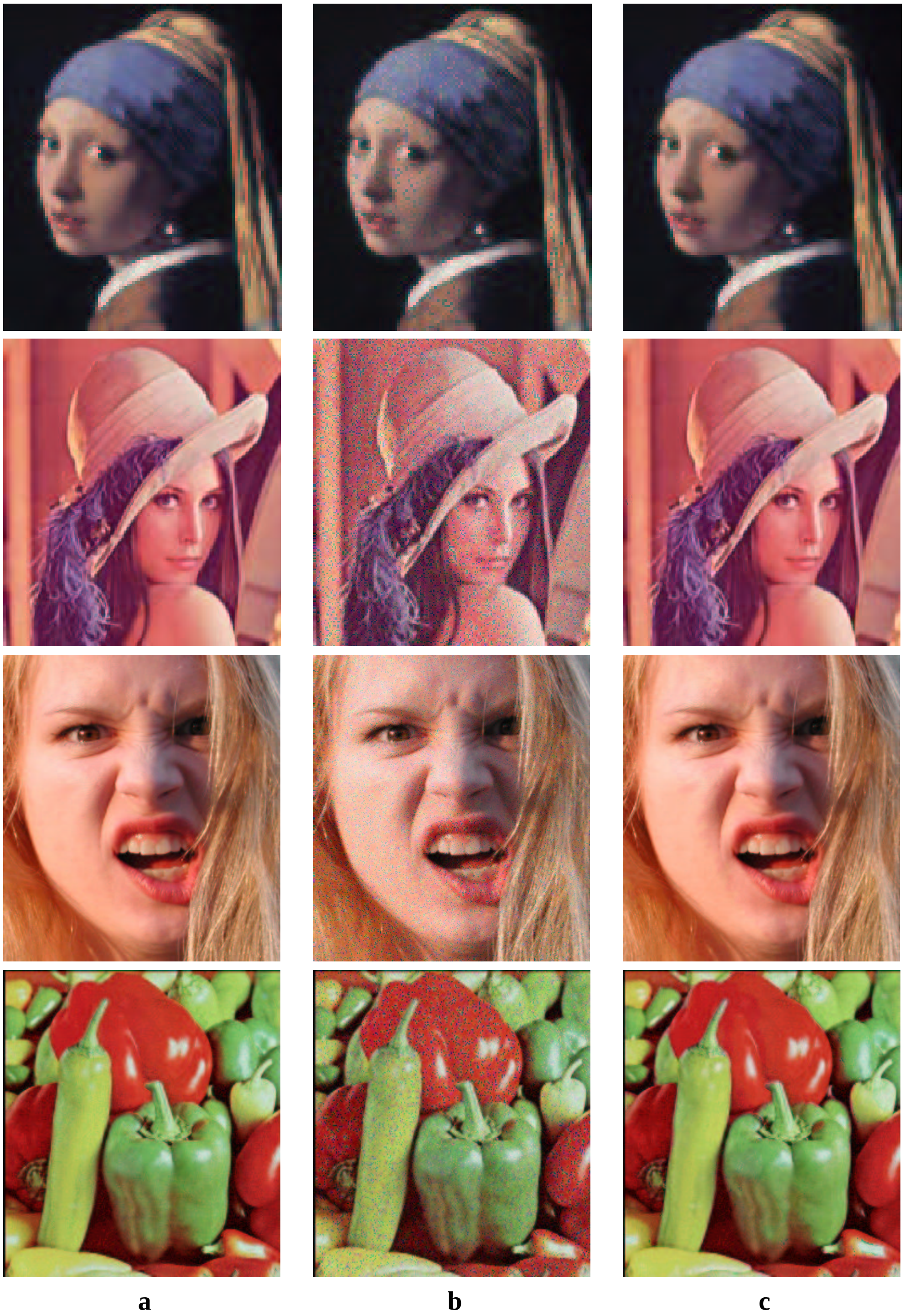


Figure 6.10: **Color denoising of the Vermeer, Lena, Sidney, and Peppers images**. **(a)** Original image. **(b)** Noisy image. **(c)** Convergence state of the relaxation process. We are grateful to Pamela Davis for the permission to use the Sidney image.

6.4 Conclusions

This section presented an application of the geometrical and computational study discussed earlier in this thesis to the domain of color. We have shown how the hue component of the color can be treated geometrically, how a notion of hue curvature arises naturally, and how contextual computation can extract coherent chromatic structures and eliminate noise while preserving discontinuities.

While thus far we have treated the saturation and lightness channels with standard denoising techniques (i.e., scalar anisotropic diffusion), an integrated good-continuation approach is in place. In particular, the combination of hue and saturation (sometimes called the chroma) yield yet another orientation-like representation and while this representation resides on $\mathcal{S}^2$ (as opposed to $\mathcal{S}^1$), the same curvature-based principle of good continuations is applicable there as well. Although the lightness channel does not encode direction directly, the behavior of its level sets does so indirectly and considering this information in color denoising may prove critical in preserving important visual structures (e.g., roof edges). Thus, color as a whole can be treated geometrically with a good continuation approach, a possibility which is a natural avenue for future computational research.

Chapter 7

Summary

In this thesis we have introduced visual flows and investigated them from four different but interrelated perspectives. Geometrically, we argued that a proper way to study these structures requires the frame field approach from differential geometry, a study that led to the notion of visual flow curvatures and to constraints on their mutual behavior. We introduced the notion of visual flow "good continuation" to extend common terminology from the psychological and computational studies of curves and we rigorously derived a unique local model – the right helicoid – that formalizes coherent visual flows. We proved numerous properties of this model, including the identical covariation of its flow curvatures, as well as other important invariances.

The geometrical insight developed in this thesis was then carried on in three ways. Firstly, we explored the role of visual flow curvatures in human perception and showed that sensitivity to these curvature greatly affect orientation-based texture segmentation. Secondly, we developed a contextual computation framework for the extraction and organization of coherent visual flows from images and demonstrated its application to texture and shading analysis. Unlike existing approaches, ours is able to handle both sparse, dense, and multivalued data sets, while preserving line and point singularities and rejecting large scale non-flow structures. We have further discussed a variation of the basic system to the domain of color and demonstrated its effectively in denoising color images. Lastly, we introduced a direct link between the developed computational structure and the columnar organization of the primary visual cortex. We proposed a novel theory for the functional role of the long range horizontal connection between orientation hypercolumns and used this theory to derive predictions about population statistics of these connections. The validity of the predictions, up to second statistical order, was confirmed from existing biological measurements and supported the functional identification of long-range horizontal connections with those obtained mathematically.

7.1 Future directions

The work in this thesis spans several dimensions of the visual sciences and thus supplies a wealth of future research directions. Geometrically, an extension of the analysis to higher dimensions can be extremely useful, e.g., for medical imaging applications. The most general (Euclidean) extension should consider $\mathcal{S}^m$ visuals flows defined over $\mathbb{R}^n$, with m and n not necessarily related. This requires a powerful generalization of the notion of curvatures and raises interesting questions with regard to the existence of helicoidal-like models for coherence in these spaces.

Computationally, our work raises many questions, most with both immediate practical consequences and long term implications. For example, recall that without explicit boundary information the degree to which the system is able extract coherent structure depends on boundary stability considerations through the value of s_{min}. Decoupling these aspects of the computation can be highly valuable. As mentioned, a solution that endows the assignment space with labels that model local boundaries explicitly may be of particular interest. Another approach may involve a simulated annealing type process in which the value of s_{min} gradually increases first to accommodate deep noise resistance and later to provide boundary stability. Naturally, such a process implies non-constant compatibilities (i.e., ones that depend upon the iteration number) and thus its repercussions might penetrate deep into the theory of relaxation labeling.

Even more important is the extension of the computational approach to implicit visual flows, i.e., visual cues whose oriented structure is implicit in their level sets. Currently, our process can extract the oriented structure and reconstruct the geometry of the levels sets, as we have demonstrated for shading. What is not dealt with are the *values* of these level sets. For shading, this missing aspect is not crucial as reconstruction of the level set geometry already carries enormous benefits with it. For other domains or applications, for example the denoising of color images, a computational process that handles both the level sets and their values may be necessary.

One property of the helicoidal model developed in this thesis is its intimate relationship to the osculating circle - a commonly used model for local coherence of curves. This relationship suggests a way to handle both curves and textures within a *unified* contextual computation. This issue should not be underestimated since it relates to a fundamental computational paradox. On one hand, the contextual organization of structure depends on the structure it seeks to organize. Indeed, as we have shown, the compatibilities for curves are different from those for visual flows. On the other hand, spatially localized detectors, in particular linear filters or biological receptive fields, typically do not provide information about the type of stimulus that crosses their receptive field, i.e., whether it is 1D (curve) or 2D (visual flow). Thus, which set of contextual interactions should be applied? It may be possible to address this chicken-and-egg problem by running two independent networks concurrently, but that implies an inefficient use of computational resources and an additional non-trivial decision process that collapses their results. The intimate relationship between the right helicoidal model and the osculating circle suggests that a unified approach is possible.

Another important issue of both theoretical and practical implications is the interaction of visual flows with boundary information. In Sec. 4.4 we developed such an interaction that uses explicit boundary information to improve flow computation and showed how it is naturally applicable to shading and edges. The opposite interaction is no less important. Could shading flow be used as a ”boundary condition” for a better extraction of edges and boundaries?

Geometrical and computational issues aside, the work in this thesis suggests additional directions in other vision sciences as well. Psychophysically, our results about sensitivity to curvature in OBTS justify the reexamination of feature gradients in preattentive vision more generally. For example, do shading flow curvatures affect intensity based segmentation? Are color segmentation, color consistency, and filling-in phenomena (Fig. 7.1) dependent upon hue curvatures? Related speculative implications exist at the physiological level as well. For example, the contextual and distributed approach to color image denoising suggests to search for long range interactions between color sensitive cells, either in V1 cytochrome oxidase blobs or perhaps in V4, which facilitates coherent color perception.

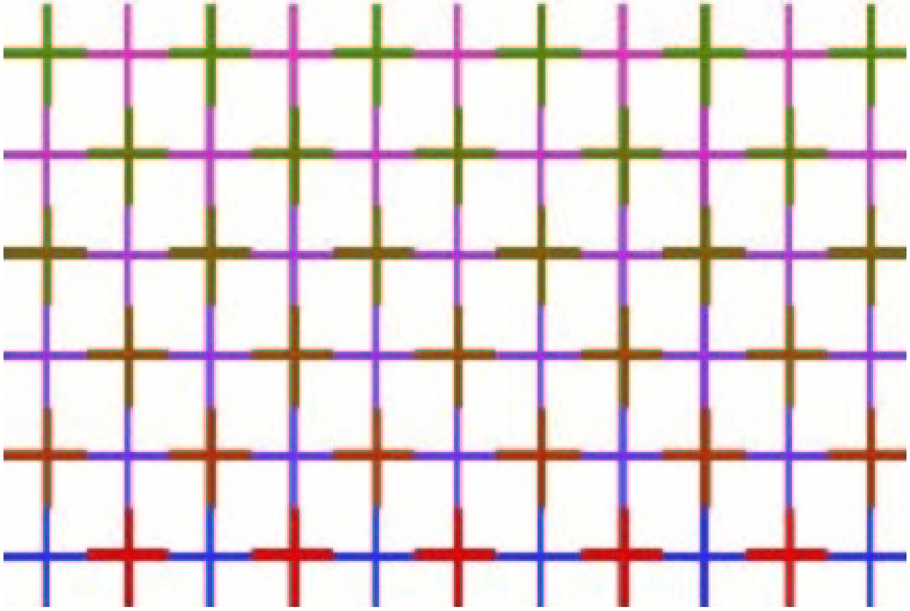


Figure 7.1: **A modified Ehrenstein illusion.** This stimulus shows how color filling in can occur between boundaries of different colors, thus suggesting a link between the approach proposed in this thesis and the psychology of color perception. Modified from Redies and Spillmann [1981], the diagonal neon bands in this illusion are best viewed on a fluorescent (e.g., CRT) display.

7.2 A final take home message

There is something unique about vision sciences in the way they combine different disciplines – from pure mathematics to ”wet” neurophysiology – and their scope of influence – from immediate industrial applications to long term understanding of the human mind. It is hoped that the benefit of interaction between different fields as demonstrated in this thesis will encourage both computer vision researchers, psychophysicists, and physiologists, to work in greater collaboration and mutual understanding. We believe that such a path will ultimately lead to better artificial vision systems alongside greater understanding of biological and human vision as well.

Bibliography

Acton, S., 1998. Multigrid anisotropic diffusion. IEEE Transactions on Image Processing 7 (3), 280–291.

Adini, Y., Sagi, D., Tsodyks, M., 1997. Excitatory-inhibitory network in the visual cortex: Psychophysical evidence. Proceedings of the National Academy of Sciences of the USA 94, 10426–10431.

Alter, T., Basri, R., 1996. Extracting salient curves from images: An analysis of the saliency network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 12–20.

Astola, J., Haavisto, P., Nuevo, Y., 1990. Vector median filters. Proceedings of the IEEE 78 (4), 678–689.

August, J., Zucker, S., 2000. The curve indicator random field: Curve organization via edge correlation. In: Boyer, K., Sarkar, S. (Eds.), Perceptual Organization for Artificial Vision Systems. Kluwer Academic Publishers.

Axler, S., Bourdon, P., Ramey, W., 1992. Harmonic Function Theory. Springer-Verlag.

Barrow, H., Tenenbaum, J., 1981. Interpreting line drawings as three-dimensional surfaces. Artificial Intelligence 17, 75–116.

Beck, J., 1966. Effect of orientation and the shape similarity on perceptual grouping. Perception & Psychophisics 1, 300–302.

Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C., 2000. Image inpainting. Proceedings of SIGGRAPH .

Black, M., Sapiro, G., Marimont, D., Heeger, D., 1998. Robust anisotropic diffusion. IEEE Transactions on Image Processing 7 (3), 421–432.

Blomgren, P., Chan, T., 1998. Color TV: Total variation methods for restoration of vector-valued images. IEEE Transactions on Image Processing 7 (3), 304–309.

Boccignone, G., Ferraro, M., Caelli, T., 2002. Generalized spatio-chromatic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (10), 1298–1309.

Borg-Graham, L., Monier, C., Fregnac, Y., 1998. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. Nature 292, 369–373.

Bosking, W., Zhang, Y., B., S., Fitzpatrick, D., 1997. Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex. The Journal of Neuroscience 17 (6), 2112–2127.

Breton, P., Zucker, S., 1996. Shadows and shading flow fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Brodatz, P., 1966. Textures: A Photographic Album for Artists and Designers. Dover Publications.

Buzás, P., Eysel, U., Kisvárday, Z., 1998. Functional topography of single cortical cells: an intracellular approach combined with optical imaging. Brain Research Protocols 3, 199–208.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6), 679–698.

Caputo, G., 1997. Object grouping contingent upon background. Vision Research 37 (10), 1313–1324.

Caputo, G., Casco, C., 1999. A visual evoked potential correlate of global figure-ground segmentation. Vision Research 39, 1597–1610.

Chan, T., Shen, J., 1999. Variational restoration of non-flat image features: Models and algorithms. CAM-TR 99-20, UCLA.

Corbit, J., Garbary, D., 1995. Fractal dimension as a quantitative measure of complexity in plant development. Vol. 262. pp. 1–6.

De Valois, R., De Valois, K., 1990. Spatial Vision. Oxford University Press.

Dierkes, U., Hildebrandt, S., Küster, A., Wohrlab, O., 1992. Minimal Surfaces. Springer-Verlag.

Do Carmo, M., 1976. Differential Geometry of Curves and Surfaces. Prentice-Hall, Inc.

Dobbins, A., Zucker, S., Cynader, M., 1987. Endstopped neurons in the visual cortex as a substrate for calculating curvature. Nature 329 (6138), 438–441.

Dobbins, A., Zucker, S., Cynader, M., 1989. Endstopping and curvature. Vision Research 29 (10), 1371–1387.

Donald, J., Riseman, E., 1992. Computing curvilinear structure by token-based grouping. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 264–270.

Ellis, W., 1955. A source book of Gestalt Psychology. Routledge & Kegan Paul Ltd.

Field, D., Hayes, A., Hess, R., 1993a. Contour integration by the human visual system: Evidence for a local “association field”. Vision Research 33 (2), 173–193.

Field, D., Hayes, A., Hess, R., 1993b. Contour integration in the human visual system: Evidence for a local ’association’ field,. Vision Research 33, 173–193.

Fischler, M., 1994. The perception of linear structure: A generic linker. In: Proceedings of the Image Understanding Workshop. Vol. 2. pp. 1565–1579.

Fischler, M., Bolles, R., 1986. Perceptual organization and curve partitioning. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1), 100–105.

Foley, J., van Dam, A., Feiner, S., Hughes, J., 1996. Computer Graphics. Principles and practice. Addison Wesley Publishing Company.

Forsyth, A., 1912. Lectures on the differential geometry of curves and surfaces. Cambridge, University press.

Galli, A., Zama, A., 1931. Untersuchungen über die wahrnehmung ebener geometrischen figuren die ganz oder teilweise von anderen geometrischen figuren verdecket sind. Zeitschrift für Psychologie 123, 308–348.

Geisler, W., Perry, J., Super, B., Gallogly, D., 2001. Edge co-occurrence in natural images predicts contour grouping performance. Vision Research 41 (6), 711–724.

Gibson, J., 1950. The Perception of the Visual World. The Riverside Press.

Gilbert, C., Wiesel, T., 1989. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. The Journal of Neuroscience 9 (7), 2432–2442.

Glass, L., 1969. Moiré effect from random dots. Nature 223 (5206), 578–580.

Glass, L., 1973. Perception of random dots inference patterns. Nature 246 (5432), 360–362.

Gordon, I., 1989. The gestalt theory. In: Theories of Visual Perception. John Wiley & Sons Ltd., Ch. 3, pp. 46–75.

Granlund, G., Knutsson, H., 1995. Signal Processing for Computer Vision. Kluwer Academic Publishers.

Graustein, W., 1940. Harmonic minimal surfaces. In: Trans. of the American Mathematical Society. Vol. 47. pp. 173–206.

Guillemin, V., Pollack, A., 1974. Differential Topology. Prentice-Hall, Inc.

Hamel, G., 1923. Zum gedachtnis an hermann amandus schwarz. Jahresber, D.M.V. 32, 6–13.

Helson, H., 1933. The fundamental propositions of gestalt psychology. Psychological Review 40, 13–32.

Hering, E., 1878/1964. Outline of the theory of the light sense. Harvard University Press, (translated by L. M. Hurvich and D. Jameson).

Hess, R., Field, D., 1999. Integration of contours: New insights. Trends in Cognitive Sciences 3, 480–486.

Hoffman, D., 1998. Visual Intelligence. W.W. Norton & Company.

Hopfield, J., Tank, D., 1985. Neural computation of decisions in optimization problems. Biological Cybernetics 52, 141–152.

Horn, B., Brooks, M. (Eds.), 1989. Shape from Shading. MIT Press, Cambridge, MA.

Hornung, C., 1976. Background Patterns, Textures, and Tints. Dover Publications.

Hubel, D., Wiesel, T., 1977. Functional architecture of macaque monkey visual cortex. Vol. 198. pp. 1–59.

Huggins, P., Chen, H., Belhumeur, P., Zucker, S., 2001. Finding folds: On the appearance and identification of occlusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 718–725.

Huggins, P., Zucker, S., 2001a. Folds and cuts: how shading flows into edges. In: Proceedings of the IEEE International Conference on Computer Vision.

Huggins, P., Zucker, S., 2001b. How folds cut a scene. In: Proc. of the 4^{th} Int. Workshop on Visual Form.

Hummel, R., Zucker, S., 1983. On the foundations of the relaxation labeling proceeses. IEEE Transactions on Pattern Analysis and Machine Intelligence 5, 267–287.

Iverson, L., 1994. Toward discrete geometric models for early vision. Ph.D. thesis, McGill University.

Iverson, L., Zucker, S., 1995. Logical/linear operators for image curves. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (10), 982–996.

John, F., 1982. Partial Differential Equations. Springer-Verlag New York Inc.

Jonas, A., Siddiqi, K., Zucker, S., 1999. Contour fragment grouping and shared, simple occluders. Computer Vision and Image Understanding 76 (2), 146–162.

Julesz, B., 1981. Textons, the elements of texture perception, and their interactions. Nature 290 (12), 91–97.

Julesz, B., 1986. Texton gradients: The texton theory revisited. Biological Cybernetics 54, 254–251.

Kanizsa, G., 1979. Organization in Vision: Essays on Gestalt Perception. Praeger Publishers.

Kapadia, M., Ito, M., Gilbert, C., Westheimer, G., 1995. Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. Neuron 15, 843–856.

Kapadia, M., Westheimer, G., , Gilbert, C., 2000. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. Journal of Neurophysiology 84, 2048–2062.

Kass, M., Witkin, A., 1987. Analyzing oriented patterns. Computer Vision, Graphics and Image Processing 37, 362–385.

Kass, M., Witkin, A., Trezopoulos, 1988. Snakes: Active contour models. International Journal of Computer Vision 1 (4), 321–331.

Keeble, D., Kingdom, F., Morgan, M., 1997. The orientational resolution of human texture perception. Vision Research 37 (21), 2993–3007.

Kimia, B., Frankel, L., Popescu, A., 1999. Euler spiral for shape completion. In: IEEE Computer Society Workshop on Perceptual Organization in Computer Vision.

Kimmel, R., Malladi, R., Sochen, N., 2000. Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images. International Journal of Computer Vision 39 (2), 111–129.

Kimmel, R., Sochen, N., 2000. Orientation diffusion or how to comb a porcupine? CIS 2000-02, Technion - Israel Istitute of Technology.

Kisvárday, Z., Kim, D., Eysel, U., Bonhoeffer, T., 1994. Relationship between lateral inhibition connections and the topography of the orientation map in cat visual cortex. The Journal of Neuroscience 6, 1619–1632.

Kisvárday, Z., Tóth, É., Rausch, M., Eysel, U., 1997. Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. Cerebral Cortex 7, 605–618.

Kittler, J., Illingworh, J., 1985. Relaxation labeling algorithms - a review. Image and Vision Computing , 206–216.

Koenderink, J., 1984a. The structure of images. Biological Cybernetics 50, 363–370.

Koenderink, J., 1984b. What does the occulding contour tell us about solid shape 13, 321–330.

Koenderink, J., 1990. Solid Shape. MIT Press, Cambridge, MA.

Koenderink, J., van Doorn, A., 1976. Local structure of movement parallax of the plane. Journal of the Optical Society of America 66 (7), 717–723.

Koffka, K., 1935. Principles of Gestalt Psychology. Routledge & Kegan Paul Ltd.

Kruth, W. (Ed.), 1963. The Complete Woodcuts of Albrecht Dürer. Dover Publiucations.

Kwan, L., Regan, D., 1998. Orientation-tuned spatial filters for texture-defined form. Vision Research 38, 3849–3855.

Landy, M., Bergen, J., 1991. Texture segregation and orientation gradient. Vision Research 31 (4), 679–691.

Langer, M., Mann, R., 2001. Dimensional analysis of image motion. Proceedings of the IEEE International Conference on Computer Vision , 155–162.

Langer, M., Mann, R., 2002. Tracking through optical snow. Workshop on Biologically Motivated Computer Vision , 181–188.

Lehky, S., Sejnowski, T., 1988. Network model of shape-from-shading: neural function arises from both receptive and projective fields. Nature 333, 452–454.

Li, Z., 1998a. A neural model of contour integration in the primary visual cortex,. Neural Computation 10, 903–940.

Li, Z., 1998b. A neural model of contour integration in the primary visual cortex. Neural Computation 10 (4), 903–940.

Lowe, D., 1985. Perceptual Organization and Visual Recognition. Kluwer Academic Publishers.

Malach, R., Amir, Y., Harel, M., Grinvald, A., 1993. Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. Proceedings of the National Academy of Sciences of the USA 90, 10469–10473.

Malik, J., Perona, P., 1990. Preattentive texture discrimination with early vision mechanisms. Journal of the Optical Society of America 7 (5), 923–932.

Marr, D., 1982. Vision. W.H.Freeman and Company.

Matsubara, J., Cynader, M., Swindale, N., Stryker, M., 1985. Intrinsic projections within visual cortex: Evidence for orientation specific local connections. Proceedings of the National Academy of Sciences of the USA 82, 935–939.

Miller, D., Zucker, S., 1999. Computing with self-excitatory cliques: A model and and application to hyperacuity-sclae computation in visual cortex. Neural Computation 11, 21–66.

Mitchison, G., Crick, F., 1982. Long axons within the striate cortex: Their distribution, orientation, and patterns of connections. Proceedings of the National Academy of Sciences of the USA 79, 3661–3665.

Mohammed, J., Hummel, R., Zucker, S., 1983. A gradient projection algorithm for relaxation methods 5 (3), 330–332.

Motoyoshi, I., Nishida, S., 2001. Visual response saturation to orientation constrast in the perception of texture boubdary. Journal of the Optical Society of America 18 (9), 2209–2219.

Munsell, A., 1905. A Color Notation. G.H.Ellis, Boston.

Murray, J., 1989. Mathematical Biology. Springer-Verlag.

Mussap, A., Levi, D., 1999. Orientation-based texture segmentation in strabismic amblyopia. Vision Research 39, 411–418.

Nalwa, V., 1993. A guided Tour of Computer Vision. Addison-Wesley Publishing.

Nitsche, J., 1989. Lectures on Minimal Surfaces. Cambridge Univ. Press.

Nothdurft, H., 1985a. Orientation sensitivity and texture segmentation in patterns with different line orientation. Vision Research 25 (4), 551–560.

Nothdurft, H., 1985b. Sensitivity for structure gradient in texture discrimination tasks. Vision Research 25 (12), 1957–1968.

Nothdurft, H., 1991. Texture segmentation and pop-out from orientation contrast. Vision Research 31 (6), 1073–1078.

Nothdurft, H., 1992. Feature analysis and the role of similarity in preattentive vision. Perception & Psychophisics 52 (4), 255–275.

Nothdurft, H., 1993. The role of features in preattentive vision: Comparison of orientation, motion, and color cues. Vision Research 33 (14), 1937–1958.

Olsen, R., Attneave, F., 1970. What variables produce similarity grouping? Perception & Psychophisics 83, 1–21.

O'Neill, B., 1966. Elementary Differential Geometry. Academic Press.

Palmer, S., 1999a. Gestalt psychology. In: Wilson, R., Keil, F. (Eds.), The MIT Encyclopedia of Cognitive Sciences. MIT Press.

Palmer, S., 1999b. Vision Science: Photons to Phenomenology. The MIT Press.

Parent, P., Zucker, S., 1989. Trace inference, curvature consistency, and curve detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (8), 823–839.

Pelillo, M., 1997. The dynamics of nonlinear relaxation labeling processes. Journal of Mathematical Imaging and Vision 7, 309–323.

Perona, P., 1998. Orientation diffusion. IEEE Transactions on Image Processing 7 (3), 457–467.

Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7), 629–639.

Pinna, B., Werner, J., Spillmann, L., 2003. The watercolor effect: a new principle of grouping and figure-ground organization. Vision Research 43, 43–52.

Polat, U., Sagi, D., 1993. Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking exteriments. Vision Research 33 (7), 993–999.

Prazdny, K., 1986. Some new phenomena in the perception of glass paterns. Biological Cybernetics 53, 153–158.

Rao, A., Jain, R., 1992. Computerized flow field analysis: Oriented texture fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (7), 693–709.

Rao, A., Schunck, B., 1991. Computing oriented texture fields. CVGIP: Graphical Models and Image Processing 53 (2), 157–185.

Redies, C., Spillmann, L., 1981. The neon color effect in the ehrenstein illusion. Perception 10, 667–681.

Regan, D., Hajdur, L., Hong, X., 1996. Two-dimensional aspect ratio idscrimination for shape defined by orientation texture. Vision Research 36 (22), 3695–3702.

Richards, W., Keonderink, J., Hoffman, D., 1987. Inferring 3d shapes from 2d silhouettes. Journal of the Optical Society of America 4, 1168–1175.

Rockland, K., Lund, J., 1982. Widespread periodic intrinsic connections in the tree shrew visual corte. Science 215 (19), 1532–1534.

Rosenfeld, A., Hummel, R., Zucker, S., 1976. Scene labeling by relaxation operations. IEEE Transactions on Systems, Man and Cybernetics 6 (6), 420–433.

Sapiro, G., Ringach, D., 1996. Anisotropic diffusion of multivalued images with applications to color filtering. IEEE Transactions on Image Processing 5 (11), 1582–1586.

Šára, R., 1994. Local shading analysis via isophotes properties. Ph.D. thesis, Johannes Kepler University in Linz.

Saund, E., 1998. Perceptual organization of occluding contours of opaque surfaces. In: IEEE Computer Society Workshop on Perceptual Organization in Computer Vision.

Schmidt, K., Goebel, R., Löwel, S., Singer, W., 1997. The perceptual grouping criterion of colinearity is reflected by anisotropies in the primary visual cortex. The European Journal of Neuroscience 9, 1083–1089.

Schmidt, K., Löwel, S., 2002. Long-range intrinsic connections in cat primary visual cortex. In: Payne, B., Peters, A. (Eds.), The Cat Primary Visual Cortex. Academic Press, pp. 387–426.

Sha'ashua, A., Ullman, S., 1988. Structural saliency: The detection of globally salient structures using a locally connected network. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 321–327.

Sharon, E., Brandt, A., Basri, R., 2000. Completion energies and scale. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10), 1117–1131.

Shu, C., Jain, R., 1993. Direct estimation and error analysis for oriented patterns. Computer Vision, Graphics and Image Processing 58 (3), 383–398.

Sigman, M., Cecchi, G., Gilbert, C., Magnasco, M., 2001. On a common circle: Natural scenes and gestalt rules. Proceedings of the National Academy of Sciences of the USA 98 (4), 1935–1940.

Simmons, G., 1963. Introduction to Topology and Modern Analysis. McGraw-Hill.

Sincich, L., Blasdel, G., 2001. Oriented axon projections in primary visual cortex of the monkey. The Journal of Neuroscience 21 (12), 4416–4426.

Sochen, N., Kimmel, R., Bruckstein, A., 2001. Diffusion and confusions in signal and image processing. Journal of Mathematical Imaging and Vision 14 (3), 195–209.

Sochen, N., Kimmel, R., Malladi, R., 1998. A geometrical framework for low level vision. IEEE Transactions on Image Processing 7 (3), 310–318.

Spillmann, L., Werner, J. (Eds.), 1990. Visual Perception: The Neurophysiological Foundations. Academic Press, Inc.

Stevens, K., 1978. Computation of locally parallel structure. Biological Cybernetics 29, 19–28.

Stevens, K., 1981. The visual interpretation of surface contours. Artificial Intelligence 17, 47–73.

Stevens, K., 1983. The line of curvature constraint and the interpretation of 3d shape from parallel surface contours. In: Proceedings of the International Joint Conference on Artificial Intelligence. pp. 1057–1061.

Streeter, V., Wylie, E., 1981. Fluid Mechanics. McGraw-Hill Ryerson Limited.

Tang, B., Sapiro, G., Caselles, V., 1999. Direction diffusion. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1245–1252.

Tang, B., Sapiro, G., Caselles, V., 2000. Diffusion of general data on non-flat manifolds via harmonic maps theory: The direction diffusion case. International Journal of Computer Vision 36 (2), 149–161.

Tang, B., Sapiro, G., Caselles, V., 2001. Color image enhancement via chromaticity diffusion. IEEE Transactions on Image Processing 10 (5), 701–707.

ter Haar Romeny, B. (Ed.), 1994. Geometry-Driven Diffusion in Computer Vision. Kluwer Academic Publishers.

Todd, J., Reichel, F., 1990. Visual perception of smoothly curved surfaces from double-projected contour patterns. Journal of Experimental Psychology: Human Perception and Performance 16 (3), 665–674.

Trahanias, P., Venetsanopoulos, V., 1993. Vector directional filters - a new class of multichannel image processing filters. IEEE Transactions on Image Processing 2 (4), 528–534.

Ts'o, D., Gilbert, C., Wiesel, T., 1986. Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. The Journal of Neuroscience 6 (4), 1160–1170.

Turing, A., 1952. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society B (London) 237, 37–72.

Ullman, S., 1976. Filling in the gaps: The shape of subjective contours and a model for their creation. Biological Cybernetics 25, 1–6.

Van Dyke, M., 1982. An Album of Fluid Motion. The Parabolic Press.

Versavel, M., Orban, G., Lagae, L., 1990. Responses of visual cortical neurons to curved stimuli and chevrons. Vision Research 30 (2), 235–248.

Waltz, D., 1975. Understanding line drawings of scenes with shadows. In: Winston, P. (Ed.), The Psychology of Computer Vision. McGraw-Hill, New York, pp. 19–91.

Watanabe, T., Cavanagh, P., 1996. Texture laciness: The texture equivalent of transparancy? Perceeption 25, 293–303.

Weickert, J., 1997. A review of nonlinear diffusion filtering. In: Romeny, B. t. H., Florack, L., Koenderink, J., Viergever, M. (Eds.), Scale-Space Theory in Computer Vision. Vol. 1252 of Lecture Notes in Computer Science. Springer-Verlag, pp. 3–28.

Weickert, J., 1999. Coherence-enhancing diffusion filtering. International Journal of Computer Vision 31 (2/3), 111–127.

Weickert, J., ter Haar Romeny, B., Viergever, M., 1998. Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing 7 (3), 398–410.

Weliky, M., Kandler, K., Fitzpatrick, D., Katz, L., 1995. Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to oriented columns. Neuron 15, 541–552.

Wertheimer, M., 1955a. Gestalt theory. In: Ellis, W. (Ed.), A source book of Gestalt Psych. Routledge & Kegan Paul Ltd., pp. 1–11.

Wertheimer, M., 1955b. Laws of organization in perceptual forms. In: Ellis, W. (Ed.), A source book of Gestalt Psych. Routledge & Kegan Paul Ltd., pp. 71–88.

Williams, L., Jacobs, D., 1997. Stochastic completion fields: A neural model of illusory contour shape and salience. Neural Computation 9 (4), 837–858.

Wilson, H., Wilkinson, F., 1998. Detection of global structure in galss patterns: implications for form vision. Vision Research 38, 2933–2947.

Witkin, A., 1983. Scale-space filtering. In: Proceedings of the International Joint Conference on Artificial Intelligence. pp. 1019–1022.

Witkin, A., Tenenbaum, J., 1983. On the role of structure in vision. In: Beck, J., Hope, B., Rosenfeld, A. (Eds.), Human and Machine Vision. Academic Press, pp. 481–542.

Wolfson, S., Landy, M., 1995. Discrimination of oriention-defined texture edges. Vision Research 35 (20), 2863–2877.

Wolfson, S., Landy, M., 1998. Examining edge- and region-based texture analysis mechanisms. Vision Research 38 (3), 439–446.

Yen, S., Finkel, L., 1998. Extraction of perceptually salient contours by striate cortical networks. Vision Research 38 (5), 719–741.

Yezzi, A., 1998. Modified curvature motion for image smoothing and enhancement. IEEE Transactions on Image Processing 7 (3), 345–352.

You, Y., Tannenbaum, A., Kaveh, M., 1996. Behavioral analysis of anisotropic diffusion in image processing. IEEE Transactions on Image Processing 5 (11), 1539–1553.

Zucker, S., 1983. Computational and psychophysical experiments in grouping: Early orientation selection. In: Beck, J., Hope, B., Rosenfeld, A. (Eds.), Human and Machine Vision. Academic Press, pp. 545–567.

Zucker, S., Dobbins, A., Iverson, L., 1989. Two stages of curve detection suggest two styles of visual computation. Neural Computation 1, 68–81, see also, Iverson, L.A., PhD thesis, McGill University, 1994.

Zucker, S., Rosenfeld, A., Davis, L., 1975. General purpose models: Expectations about the unexpected. In: Proceedings of the International Joint Conference on Artificial Intelligence. pp. 716–721.

Index