SIMULTANEOUS BOTTOM-UP/TOP-DOWN PROCESSING IN EARLY AND
MID LEVEL VISION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEHMET ERKUT ERDEM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

NOVEMBER 2008

Approval of the thesis:

SIMULTANEOUS BOTTOM-UP/TOP-DOWN PROCESSING IN EARLY AND MID LEVEL VISION

submitted by **MEHMET ERKUT ERDEM** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**

_____

Prof. Dr. Müslim Bozyiğit
Head of Department, **Computer Engineering**

_____

Assoc. Prof. Dr. Sibel Tarı
Supervisor, **Computer Engineering Dept., METU**

_____

Examining Committee Members:

Prof. Dr. Volkan Atalay
Computer Engineering Dept., METU

_____

Assoc. Prof. Dr. Sibel Tarı
Computer Engineering Dept., METU

_____

Prof. Dr. Kemal Leblebicioğlu
Electrical and Electronics Engineering Dept., METU

_____

Prof. Dr. Neşe Yalabık
Computer Engineering Dept., METU

_____

Asst. Prof. Dr. Pınar Duygulu-Şahin
Computer Engineering Dept., Bilkent University

_____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name   :   Mehmet Erkut ERDEM

Signature   :

ABSTRACT

SIMULTANEOUS BOTTOM-UP/TOP-DOWN PROCESSING IN EARLY AND MID
LEVEL VISION

ERDEM, Mehmet Erkut

Ph.D., Department of Computer Engineering

Supervisor: Assoc. Prof. Dr. Sibel Tarı

November 2008, 175 pages

The prevalent view in computer vision since Marr is that visual perception is a data-driven bottom-up process. In this view, image data is processed in a feed-forward fashion where a sequence of independent visual modules transforms simple low-level cues into more complex abstract perceptual units. Over the years, a variety of techniques has been developed using this paradigm. Yet an important realization is that low-level visual cues are generally so ambiguous that they could make purely bottom-up methods quite unsuccessful. These ambiguities cannot be resolved without taking account of high-level contextual information. In this thesis, we explore different ways of enriching early and mid-level computer vision modules with a capacity to extract and use contextual knowledge. Mainly, we integrate low-level image features with contextual information within unified formulations where bottom-up and top-down processing take place simultaneously.

Keywords: bottom up/top down paradigms in computer vision, image denoising, image segmentation, skeleton extraction, PDE methods

ÖZ

ERKEN VE ORTA DÜZEY GÖRMEDE ALTTAN ÜSTE/YUKARIDAN AŞAĞI
EŞZAMANLI İŞLEME

ERDEM, Mehmet Erkut

Doktora, Bilgisayar Muhendisliği Bölümü Bölümü

Tez Yöneticisi: Doç. Dr. Sibel Tarı

Kasım 2008, 175 sayfa

Marr'dan bu yana bilgisayarlı görmeye hakim olan genel kanı, görsel algının veri güdümlü, alttan üste doğru işleyen bir süreç olduğu yönündedir. Bu kapsamda imge verisi ileri beslemeli bir biçimde işlenirken ard arda gelen birbirlerinden bağımsız alt modüller basit, düsük düzey ipuçlarını daha karmaşık, soyut algı birimlerine dönüştürürler. Yıllar boyu bu model kullanılarak çok çeşitli sayıda yöntem geliştirilmiştir. Oysa önemli bir nokta, düşük düzey ipuçlarının genelde belirsiz bir yapıda olmasıdır ki bu tümüyle alttan üste doğru işleyen yöntemleri bir hayli başarısız yapmaktadır. Bu belirsizlikler yüksek düzey bağlamsal bilgi kaale alınmadan giderilemez. Bu tezde erken ve orta düzey bilgisayarlı görme modüllerinin farklı biçimlerde bağlamsal bilgi çıkarma ve kullanma yetisiyle kuvvetlendirilmesi incelenmektedir. Esas olarak, düşük düzey imge öznitelikleri bağlamsal bilgi ile alttan üste ve yukarıdan aşağıya doğru işlemenin eşzamanlı yeraldığı birleştirilmiş formülasyonların bünyesinde bütünleştirilmektedir.

Anahtar Kelimeler: bilgisayarlı görmede alttan üste/yukarıdan aşağıya yaklaşımlar, imge gürültü temizleme, imge bölütleme, iskelet çıkarma, kısmi türevleme denklem metodları

ACKNOWLEDGMENTS

I could not have made it this far without them.

I am also eternally grateful to two special people, my grandmothers Şükran Erdem and Nuran Halatçı, who have had significant influence on me at different periods of my life.

To my family and to the memory of my grandmother Şükran Erdem

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF SYMBOLS

f	grayscale input image
u	evolving image due to diffusion filtering or energy minimization
x, y or (x, y)	spatial coordinate in an image
t	artificial time variable in evolution equations, in the scope of diffusion: scale
$\frac{\partial}{\partial a}$	partial derivative with respect to a
$\nabla \cdot a$	divergence of a, $\nabla \cdot a = \frac{\partial a}{\partial x} + \frac{\partial a}{\partial y}$
∇a	gradient of a
$\nabla^2 a$	Laplacian of a
T	diffusion time, scale parameter
G_σ	Gaussian kernel with standard deviation σ
σ	scale parameter (standard deviation of a Gaussian kernel), in the scope of the Ambrosio-Tortorelli model: smoothing radius $\sigma = \sqrt{\frac{2\alpha}{\beta}}$
$a * b$	convolution of b with a
i, j or (i, j)	discrete pixel location
h_x , h_y	grid sizes in numerical implementation
$\frac{d}{dt}$	total time derivative
Δt	time step size in numerical implementation
$a_{i,j}^k$	discretized variable a at pixel (i, j) after iteration k
n	additive noise model
$|a|$	magnitude of vector a
α	weight of the smoothness term (scale parameter)
Ω	rectangular image domain, $\Omega \subset \mathbb{R}^2$
$\partial\Omega$	image boundary

$\frac{\partial a}{\partial n}\big|_{\partial\Omega}$	derivative of a in the direction normal to $\partial\Omega$
g	diffusivity function
λ	in the scope of diffusion models: contrast parameter, in the scope of linear elasticity mode: Lamé constant
a_b	abbreviation for $\frac{\partial a}{\partial b}$
a_{bc}	abbreviation for $\frac{\partial a}{\partial b \partial c}$
Φ	flux function
a_σ	convolution of a with G_σ
β	in the scope of Mumford-Shah model: weight of the data fidelity term (scale-space parameter), in the scope of shape matching energy: weight of the regularization term
Γ	edge set or shape boundary
ρ	blurring radius
v	Ambrosio-Tortorelli edge strength function for sufficiently small values of ρ, and TSP surface for sufficiently large values of ρ
χ_A	characteristic function that is defined to be identically one on A and is zero elsewhere
$\lfloor a \rfloor$	largest integer number b with $b \le a$
$\epsilon,\ \epsilon_1,\ \epsilon_2$	small positive constants
ϕ	in the scope of Chan Vese model: level set function, in the scope of context-influenced filtering: contextual feedback measure, in the scope of distance transform: distance surface or function
$c_1,\ c_2$	scalars denoting average gray values of object and background regions
$\lambda_1,\ \lambda_2$	weights of the fidelity terms of the object and background regions
μ	in the scope of Chan-Vese model: length parameter, in the scope of segmentation framework that employs a global shape model: weight of the prior shape term
$H(z)$	Heaviside function
$\delta(z)$	derivative of the Heaviside function with respect to z (Dirac delta function)
ϵ	scalar used in the regularized Heaviside function
Δ_-^a	backward difference in the direction of a
Δ_-^a	forward difference in the direction of a
$curv(a)$	curvature of a level curve of a, $curv(a) = \nabla \cdot \left(\frac{\nabla a}{|\nabla a|} \right)$

D	diffusion tensor
J	structure tensor
cv	context modulated diffusivity function
c	spatially varying multiplier within the modified diffusivity function cv
V	variable defining type of the feedback in the context-influenced filtering ($V = 0$ for *positive feedback*, $V = 1$ for *negative feedback*)
$\mathbf{f} = (f^1, \dots, f^m)$	multi-channel input image
$\mathbf{u} = (u^1, \dots, u^m)$	multi-channel evolving image
$\mathbf{v} = (v^1, \dots, v^m)$	multi-channel edge strength function
m	number of color or feature channels
ϕ^{dc}	contextual feedback measure based on directional consistency of edges
ϕ^{h}	contextual feedback measure based on edge continuity
ϕ^{te}	contextual feedback measure based on texture edges
ϕ^{ls}	contextual feedback measure based on local scale
η_s	neighborhood of a pixel i defined as $\pm s$ pixels along the orthogonal edge direction $\nabla u_i^{\perp}$
s	in the scope of directional consistency measure: number of neighboring pixels, in the scope of extraction of skeleton point: arclength in the direction of ∇v
ε^{dc}	a scalar that determines the decay rate of ϕ^{dc}
h	in the scope of edge continuity measure: an offset $\in [0, 1]$, in the scope of segmentation framework that employs a global shape model: scale parameter, in the scope of local deformation model: displacement vector (deformation function)
D_i^*	distribution of the contextual similarities between the patch centered at the pixel i and the patches which are Δx pixels to the direction $(*)$
n	height of a square image patch
ε^{te}	a scalar that determines the decay rate of ϕ^{te}
p_i^1, p_i^2	p-value's returned by the Wilcoxon Mann-Whitney test for the distributions D_i^{up}-D_i^{down} and D_i^{left}-D_i^{right}, respectively
σ_i	local scale of an image pixel i
Ω_i	$n \times n$ image patch around an image pixel i
ε^{ls}	a scalar that determines the decay rate of ϕ^{ls}
M	width of an image

N	height of an image
$\mathcal{O}$	big-Oh notation for computational complexity analysis
S, χ	a shape in the form of a silhouette (binary image)
∂S	shape boundary
U	surface estimated by solving a Poisson equation
$\mathcal{R}_\sigma$	shape representation based on integral kernels
$\bar{\Phi}$	mean TSP surface
Φ_i	i^{th} principle component
p	in the scope of segmentation framework that employs a global shape model: number of principle components, in the scope of shape skeletons: a skeleton point
$\mathbf{w} = (w_1, \cdots, w_p)$	shape variability parameters, weights for the principle components
$\Phi_{\mathbf{w}}$	shape prior defined by the global shape model
t_x, t_y	translation parameters
θ	rotation parameter
S_1	in the scope of shape matching: reference shape in the form of a binary image, in the scope of skeleton extraction: the closure of the set of zero-crossings of $\frac{d|\nabla v|}{ds}$ (skeleton points)
S_2, T	template shape in the form of a binary image
$\bar{\mu}$	Lamé constant
∂_a	abbreviation for $\frac{\partial}{\partial a}$
v_T	TSP surface representing the template shape T
μ_1	weight of the prior shape term
μ_2	weight of the matching energy
S_1^+	the points in S_1 where the curvature of the level curves is positive
S_1^-	the points in S_1 where the curvature of the level curves is negative
S_0	the set of points where $|\nabla v| = 0$
ψ	angle between the vectors ∇v and $\nabla \left(\frac{d|\nabla v|}{ds} \right)$
$\overline{\theta}, \underline{\theta}$	upper and lower threshold values for Shah's skeleton pruning algorithm
r	skeletal radius

LIST OF ABBREVIATIONS

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
AT	Ambrosio-Tortorelli
CV	Chan-Vese
MAP	Maximum a posteriori
MS	Mumford-Shah
PCA	Principle Component Analysis
PDE	Partial differential equation
RGB	Color model in terms of red, green, and blue
TSP	Tari-Shah-Pien
TUBITAK	The Scientific and Technological Research Council of Turkey (Türkiye Bilimsel ve Teknolojik Araştırma Kurumu)
TV	Total Variation

CHAPTER 1

INTRODUCTION

Everyday, our visual system interprets the external world it sees without facing any difficulty. From the plain man's point of view vision is an effortlessly easy process. At first glance, this could give rise to the illusion that constructing computers that make inferences related to the external world being seen through images is a simple task. However, for the time being, it has been proved that this is not the case. Computational vision has appealed and still appeals many researchers from many different fields.

Computational vision is very important not only from a practical point of view (e.g. perception-driven robots) but also from a scientific point of view, as it helps us to understand the underlying mechanisms of the visual systems of living organisms, including humans, by providing a systematic framework. It is noteworthy that in the 1960s the preliminary studies on computational vision emerged as a part of artificial intelligence. However, it was not until the 1980s that computational vision started to be defined as an area of science in modern sense. In this respect, David Marr is generally acknowledged as one of the most influential scientists who defined the area. According to Marr, visual perception can be interpreted as an information processing task, which requires a clear understanding of *what is to be computed*, *how the computation is to be done* and *how it is to be realized physically* [96]. This interpretation implies that there is a *single* visual perception problem that is solved everyday by living organisms that had experienced very different evolutionary paths; and computational vision aims to solve it by using computers.

Since Marr, in the computational study of vision, it has been commonly assumed that visual perception is a data-driven, bottom-up process that consists of several independent visual modules (Figure 1.1). In this view, image data is mainly processed in a feed-forward fashion, and a sequence of processes (modules) transforms simple low-level cues into more complex abstract perceptual units. Each process receives an input from its preceding process
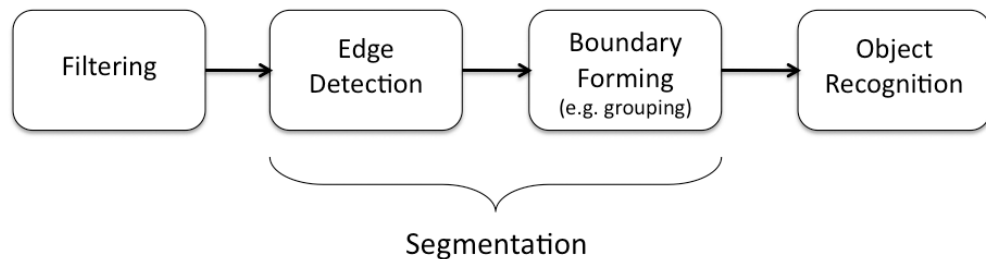
Figure 1.1: Visual modules and the information flow.

and produces an output to serve as an input for its succeeding process. During this activity, low-level cues are always assumed to be reliable and higher-level knowledge has no effect on the early stages.

Roughly speaking, visual modules can be described in three groups: *early visual processing* (or simply *filtering*), *segmentation* and *object recognition. Early visual processing* is responsible for smoothing the input image by filtering out irrelevant image information such as noise, and hence it has a critical role for the further processing modules. The task of *segmentation* is to partition the (filtered) image into coherent regions that are likely to correspond to objects exist in the image. After possible object regions are extracted, the third module *object recognition,* as its name implies, is in charge of recognizing known objects and analyzing the properties of them.

Since the early days of computational vision, the most common way of determining whether a region is coherent or not has relied on very simple image-based criteria such as uniformity of intensity or color. Subsequent to the filtering process that eliminates noise, segmentation involves identifying regions of uniform features. This is realized by carrying out two sequential steps: *edge detection* and *boundary forming.* In the *edge detection* step, edge points are identified by applying some local edge operators that are sensitive to the changes in local brightness. The general assumption is that different image regions have different intensities. Following this, in the *boundary forming* step, extracted local edge fragments are grouped together to obtain region boundaries. The grouping process is performed, for example, by forcing the smoothness and the continuity of bounding contours.

The main problem with this approach is that noisy pixels have similar characteristics to image edges. They correspond to the image points where an abrupt change in the intensity occurs. When a low-pass denoising filter is applied to the image, the filter not only removes the noise but also blurs intensity discontinuities, making the edges difficult to detect. If the

cut-off frequency of the low-pass filter is set to a large value to reduce the blurring, then one cannot eliminate noise. The noise which is passed to the second level in the hierarchy is enhanced more than the signal itself by the edge detection process [58].

In Figure 1.2, we present another drawback of this approach on a noise-free natural image. Here, we identify the image edges using the Canny edge detector [34], one of the widely used edge detection methods. As the details extracted from three separate image points (subimages labeled with a, b and c) clearly illustrate, the color is smoothly varying along the region that corresponds to the sky. It can be clearly seen that there are no extraneous edges inside that region, which makes partitioning the region as a whole possible. However, in the other parts of the image, the method results in cluttered edges that do not correspond to actual object or scene boundaries. The main reason is that the criteria based on color uniformity becomes inadequate to describe these image regions.
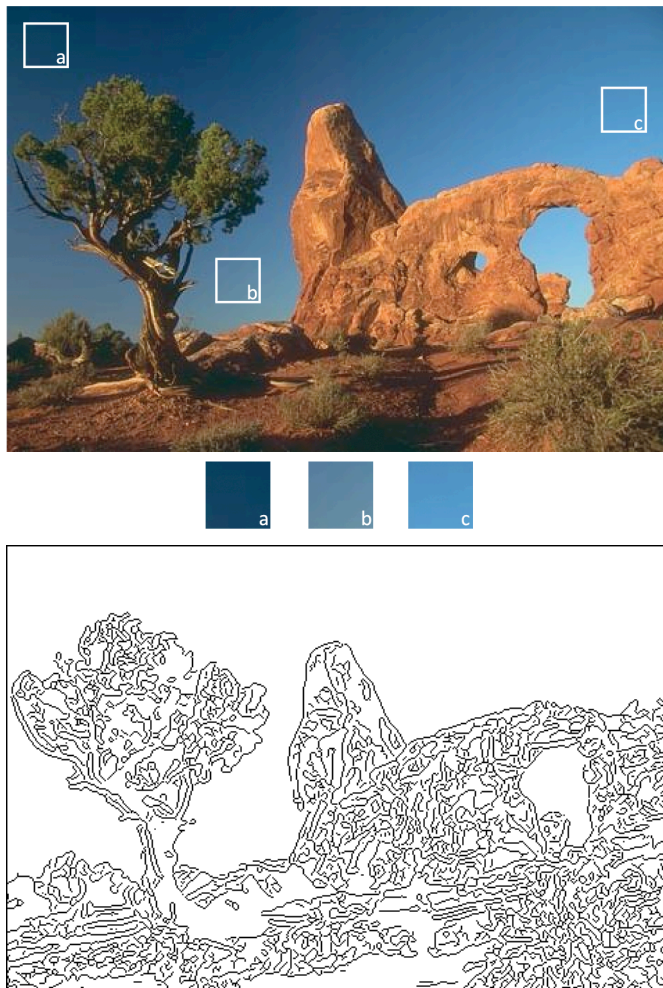
Figure 1.2: A natural image and a sample edge detection result.

One way to increase the robustness of the segmentation process is to consider additional image features other than intensity and color information. One such possible image feature is texture, which can simply be described as a visual cue that is composed of regularly arranged similar local patterns. Conveniently, it provides prominent information about image regions. For example, Figure 1.3 includes a texture mosaic image that consists of five regions. Since each region has its own local structure characteristics, it is possible to easily distinguish them by a segmentation method that relies on texture uniformity.
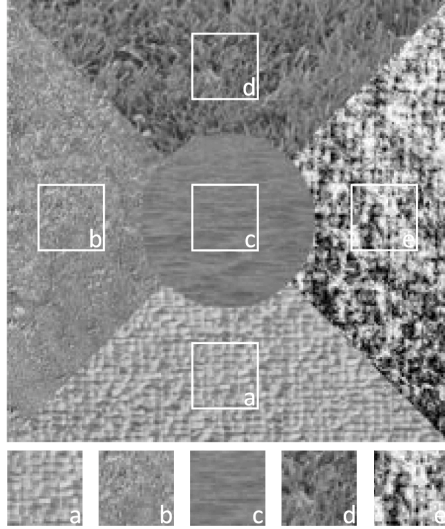
Figure 1.3: A mosaic image (taken from [72]) and texture samples extracted from it.

Although, in principle, employing texture information improves the segmentation results, texture features are in many cases insufficient as well, especially when the natural images are taken into account. For example, consider the cheetah image provided in Figure 1.4. Although our visual system can reliably extract the apparent boundary of the animal, most probably, a segmentation method that only relies on texture uniformity will fail in extracting them. The trouble is that even if the animal has the dominant texture composed of fur with round black spots (the subimage labeled with a), the ear of the animal does not contain any texture at all (the subimage labelled as b). This prevents the ear and the body regions of the animal to be grouped together, and thus makes obtaining a single region that corresponds to cheetah very difficult.

Another major challenge in image segmentation is to handle occlusions. In general, it is impossible to delineate the occluded object boundaries correctly by utilizing the cues depend
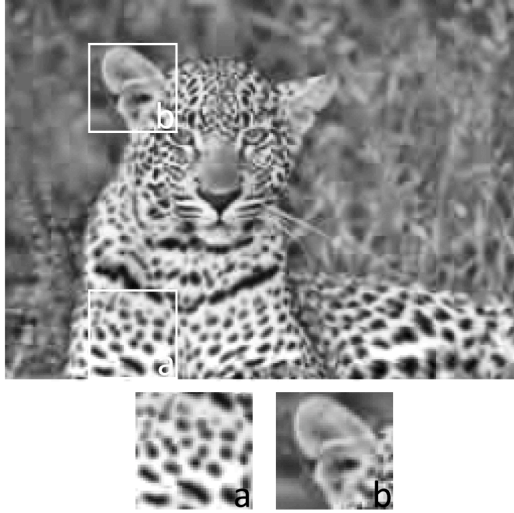
Figure 1.4: A cheetah image and the details extracted from two different image regions.

merely on raw image data. Figure 1.5 shows two such examples. In the images, the hands are partially occluded by the pencils that results in some parts of the regions describing the hands missing. What provides our visual system with the ability to reliably separate the hands from the background is the prior experience. What a hand could possibly look like is somewhat encoded in our brains, and guides the segmentation process. Hence, we can see the hands as single objects although some portions of them are not visible.

Figure 1.5: Partially occluded hand images.

As these examples illustrate, the corrupting influences such as noise, texture, occlusion, etc. greatly affect image segmentation. Most of the time, low-level visual cues are so ambiguous that they could make traditional, purely bottom-up segmentation methods quite unsuccessful. Due to the unidirectional information flow, any misinterpretation in the early

stages disturbs the later ones. The only way to resolve these ambiguities is to integrate low-level image features with high-level contextual information. Interestingly, there is an increasing number of studies in cognitive neuroscience that are consistent with this idea, e.g. [71, 85, 86, 102]. These studies all support existence of top-down information flow in the human visual system, and emphasize the importance of feedback connections between different visual areas that provide contextual influences.

Without doubt, this view requires a modified Marr paradigm in which different visual modules interact with each other through feed-forward/feedback pathways. The consistency of such interactions among the visual modules highly depends on making use of a shared language, which involves utilizing a common representational framework. Thus, it can be argued that representation is the key issue for such kind of computational models.

Generally speaking, the notion of context has a broad meaning. It typically includes *prior information* (i.e. application or aim specific knowledge), *global image features* or *statistics*, and *neighboring events* (e.g. nearby objects, nearby pixels, etc.)[139, 156]. Context mainly provides a top-down knowledge which influences the processing of visual cues so that some possible interpretations are excluded while the more consistent ones are enforced during the computations.

1.1 Major Contributions of This Thesis

Our main focus in this thesis is to enrich low and mid-level computer vision modules with a capacity to extract and use contextual knowledge. In particular, this thesis makes the following contributions:

1. A novel image smoothing method that considers local context has been introduced. The proposed approach employs high-level contextual measures extracted from local image regions to influence the ongoing filtering process.

2. Different local contextual measures have been suggested for modulating the filtering operation. Each performs several seemingly different visual tasks like elimination of unintuitive edges due to noise or texture, enhancing weak edges, extracting texture boundaries, and preserving texture during denoising.

3. Two new prior-shape based segmentation frameworks which utilize Tari-Shah-Pien (TSP) surfaces [143] as shape priors have been presented, yielding robustness against undesirable conditions such as noise, low contrast, weak edges and partial occlusions.

4. A new skeleton extraction approach that relies on a special function which simultaneously encodes the *boundaryness* and *medialness* has been suggested for highly-textured natural images. Using extracted shape skeletons, we can identify image patches of uniform characteristics that can be utilized within various image analysis tasks.

1.2 Organization of the Thesis

The organization of this thesis is as follows:

Chapter 2 and Chapter 3 survey the literature on filtering and segmentation methods. In particular, Chapter 2 is on a detailed review of important linear and nonlinear diffusion models, and variational segmentation methods. In Chapter 3, we examine important segmentation frameworks presented in the last decade that employ prior shape knowledge to guide the segmentation.

Chapter 4 presents a context-influenced image smoothing method with four different choices for local contextual measures. In Chapter 5, we suggest two new prior-shape guided segmentation methods utilizing TSP surfaces as shape priors. Chapter 6 is about a new skeleton based analysis of natural images that divides them into constituent components while simultaneously forming a representation for the components which ties shape and image features.

In Chapter 6, the basic machinery for constructing the function which simultaneously encodes the boundaryness and medialness is the context-guided image smoothing framework that we proposed in Chapter 4. Thus, these two chapters are closely related. Even though the simultaneous bottom-up/top-down processing plays a key role in this thesis, the work presented in Chapter 5 differs from the ones suggested in Chapter 4 and Chapter 6 in terms of how the higher-level information, providing the top-down flow, is incorporated and utilized in the computations. Within the prior-guided segmentation frameworks introduced in Chapter 5, the object specific knowledge is given a priori, i.e. it is explicitly provided by the user. In this respect, Chapter 5 can be read separately considering the background provided in Chapter 2 and Chapter 3.

Finally, Chapter 7 summarizes our major contributions and discusses future works.

CHAPTER 2

EARLY VISUAL PROCESSING

In the previous chapter, the traditional bottom-up paradigm that dominated the studies in computational vision was analyzed, and some interesting and challenging cases for the bottom-up segmentation approaches were presented that motivate the proposed works in this thesis.

These challenges generally stem from the unidirectional information flow. For instance, consider detecting edges that plays a key role in image segmentation. Since edges are described as abrupt intensity changes in the images, edge detection in fact requires a kind of derivative computation through local edge operators. If the image data is noisy, differentiation yields inaccurate results. Hence, the input image needs to be smoothed out before applying the operators. As discussed in the previous chapter, this leads to a *chicken and egg dilemma* since the smoothing process should eliminate the noise, and yet preserve the edges. Needless to say, this requires a clear understanding of possible edge points which is the actual goal of edge detection.

About twenty years ago, the dilemma in the one way information flow between smoothing and edge detection prompted eminent researchers to propose methods that combine smoothing and edge detection [20, 62, 100, 109]. Most of these unified frameworks are based on nonlinear diffusion models and variational regularization/segmentation methods. Although, partial differential equations (PDEs), which appear in many diffusion models, and variational approaches have been widely and successfully used in areas such as physics and engineering for many years, they have been an area of intensive research in image processing and computer vision only for the past two decades, yet they have provided effective solutions especially for image smoothing and restoration tasks.

We prefer to start our review with the linear diffusion equation to provide a motivation for these studies. It is also used as an illustrative example to analyze the relation between

variational regularization frameworks and diffusion equations. The subsequent parts of this chapter are devoted to important nonlinear diffusion models and variational regularization methods. The numerical implementation details of some of these works are also provided since they are closely related with the proposed works. The chapter concludes with a brief discussion on the reviewed works.

2.1 Linear Diffusion

The linear diffusion (heat) equation is the oldest and best investigated PDE method in image processing. Let $f(x)$ denote a grayscale (noisy) input image and $u(x, t)$ be initialized with $u(x, 0) = u^0(x) = f(x)$. Then, the linear diffusion process can be defined by the equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (\nabla u) = \nabla^2 u \tag{2.1}$$

where $\nabla \cdot$ denotes the divergence operator. Hence,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} . \tag{2.2}$$

The diffusion process can be seen as an evolution process with an artificial time variable t denoting the *diffusion time* where the input image is smoothed at a constant rate in all directions. Starting from the initial image $u^0(x)$, the evolving images $u(x, t)$ under the governed equation represent the successively smoothed versions of the initial input image $f(x)$, and thus create a *scale space* representation of the given image f, with $t > 0$ being the scale. As we move to coarser scales, the evolving images become more and more simplified since the diffusion process removes the image structures at finer scales. Figure 2.1 and Figure 2.2 show example scale space representations sampled at different diffusion times for two different images. In fact, the notion of scale is an essential part of early visual processing, where the main task is to separate the image into relevant and irrelevant parts.

It is shown that the solution of the linear diffusion equation with the given initial condition $u(x, 0) = f(x)$ for a specific diffusion time T is equivalent to the convolution of the input image $f(x)$ with the Gaussian kernel $G_\sigma(x)$ with standard deviation $\sigma = \sqrt{2T}$ [74, 79, 155]. Thus, linear diffusion can be regarded as a low-pass filter. The correspondence between the diffusion time variable t and the standard deviation σ clearly depicts the effect of t on the evolving images. The higher the value of t, the higher the value of σ, and the more smooth the image becomes. This relation also provides the following explicit solution to (2.1):

$$u(x, t) = \left(G_{\sqrt{2T}} * g \right)(x) \qquad \text{with} \qquad G_\sigma(x) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{|x|^2}{2\sigma^2} \right) . \tag{2.3}$$

$T = 0$ $T = 1.25$ $T = 2.5$

$T = 5$ $T = 10$ $T = 20$

Figure 2.1: Linear diffusion results for different diffusion times.

$T = 0$ $T = 5$ $T = 10$

$T = 20$ $T = 40$ $T = 80$
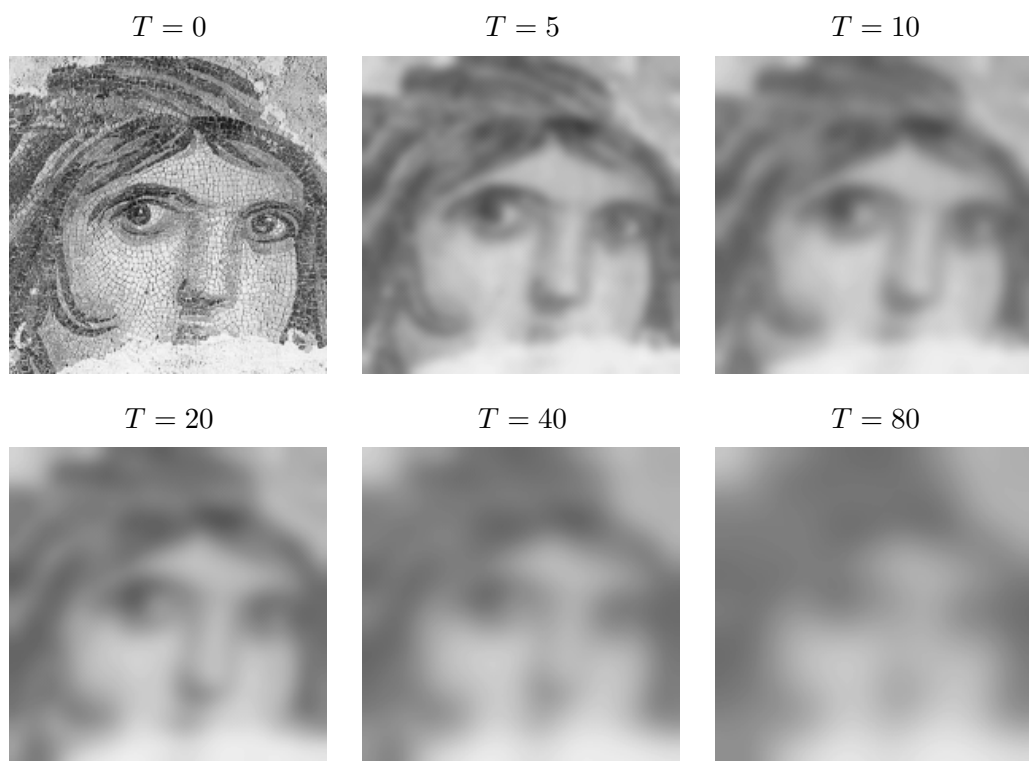
Figure 2.2: Linear diffusion results for different diffusion times.

Numerical Implementation

Since we deal with digital images, solving the linear diffusion equation requires discretization in both spatial and time coordinates. Central differences are the typical choices for the spatial derivatives:

$$\frac{d^2 u_{i,j}}{dx^2} \approx \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h_x^2}, \qquad \frac{d^2 u_{i,j}}{dy^2} \approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h_y^2} \tag{2.4}$$

where $u_{i,j}$ denotes the gray value or the brightness of the evolving image at pixel location (i,j).

The values of h_x and h_y are generally set to 1 as digital images are discretized on a regular pixel grid. For the remainder of this thesis, we take $h_x = h_y = 1$. This leads to the following space-discrete equation for (2.1):

$$\frac{du_{i,j}}{dt} = u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j} . \tag{2.5}$$

The straightforward approach to solve (2.5) is to consider an iterative scheme with an explicit time discretization, where homogeneous Neumann boundary condition is imposed along the image boundary

$$\frac{u_{i,j}^{k+1} - u_{i,j}^{k}}{\Delta t} = u_{i+1,j}^{k} + u_{i-1,j}^{k} + u_{i,j+1}^{k} + u_{i,j-1}^{k} - 4u_{i,j}^{k} \tag{2.6}$$

where Δt is the time step, and u^k represents the restored image u at iteration k. Numerical stability condition for the discrete scheme requires that $\Delta t \leq 0.25$.

Relation Between Variational Regularization and Diffusion Equations

Interestingly, there is a strong relation between variational regularization methods and diffusion equations [124]. The variational regularization methods formulate smoothing process as a functional minimization via which a noise-free approximation of a given image is to be estimated. Most of these formulations assume an additive noise model

$$f(x) = u(x) + n(x) \tag{2.7}$$

where $f(x)$ and $u(x)$ respectively denote the given noisy image and the desired denoised image, and $n(x)$ represents the additive noise.

Consider the Tikhonov energy functional [145] as an illustrative example:

$$E(u) = \int_\Omega \left((u - f)^2 + \alpha |\nabla u|^2 \right) dx \tag{2.8}$$

where

11

- $\Omega \subset \mathbb{R}^2$ is connected, bounded, open subset representing the image domain,

- f is an image defined on Ω,

- u is the smooth approximation of f,

- $\alpha > 0$ is the scale parameter.

The first term in $E(u)$ is the *data fidelity* term that penalizes the deviations between u and f, and thus forces the restored image to be close to the original image. The second term is called the *regularization* or *smoothness* term which penalizes the high gradients, and gives preference to smooth approximations. The relative importance of these two terms are defined by the scale parameter α.

The minimizing function u formally satisfies the Euler-Lagrange equation

$$(u - f) - \alpha \nabla^2 u = 0 \tag{2.9}$$

with the Neumann boundary condition $\frac{\partial u}{\partial n}\big|_{\partial \Omega} = 0$.

It is possible to rewrite (2.9) as

$$\frac{u - u^0}{\alpha} = \nabla^2 u \qquad \text{with} \qquad u^0 = f, \tag{2.10}$$

which may be regarded as an implicit time discretization of the linear diffusion equation (2.1) where a single time step $(T = \alpha)$ is used. Note that diffusion time (scale selection) problem is not really eliminated by the variational regularization, it is replaced with a new parameter α that determines the strength of the smoothness prior.

The main drawback of linear diffusion filtering is that the smoothing process does not consider information regarding important image features such as edges. It follows that same amount of smoothing to be applied at every image location. As a result, the diffusion process does smooth not only noise, but also image edges.

2.2 Perona-Malik Type Nonlinear Diffusion [109]

The main theory behind nonlinear diffusion models is to use nonlinear PDEs to create a scale space representation that consists of gradually simplified images where some image features such as edges are maintained or even enhanced. The earliest nonlinear diffusion model proposed in image processing is the so-called *anisotropic diffusion*[1] by Perona and

---

[1]In fact, Perona-Malik equation is an isotropic nonhomogeneous equation as it uses a scalar-valued diffusivity. A true example of anisotropic diffusion model, edge-enhancing diffusion [153], will be summarized in Section 2.6.

Malik [109].

In their formulation, they replaced the constant diffusion coefficient of linear equation (2.1) by a smooth nonincreasing diffusivity function g with $g(0) = 1$, $g(s) \geq 0$, and $\lim_{s \to \infty} g(s) = 0$. As a consequence, the diffusivities become variable in both space and time. The Perona-Malik equation is

$$\frac{\partial u}{\partial t} = \nabla \cdot (g(|\nabla u|)\nabla u) \tag{2.11}$$

with homogeneous Neumann boundary conditions and the initial condition $u^0(x) = f(x)$, f denoting the input image.

Perona and Malik suggested two different choices for the diffusivity function:

$$g(s) = \frac{1}{1 + s^2/\lambda^2}, \tag{2.12}$$

$$g(s) = e^{-\frac{s^2}{\lambda^2}} \tag{2.13}$$

where λ corresponds to a contrast parameter. These functions share similar characteristics, and result in similar effects on the diffusivities.

We review the 1D physical analysis of the Perona-Malik diffusion below since it clearly demonstrates the role of the contrast parameter λ and the main behavior of the equation [154]. For 1D case, the Perona-Malik equation is as follows:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \underbrace{(g(|u_x|)u_x)}_{\Phi(u_x)} = \Phi'(u_x)u_{xx} \tag{2.14}$$

with $g(|u_x|) = \frac{1}{1 + |u_x|^2/\lambda^2}$ or $g(|u_x|) = e^{-\frac{|u_x|^2}{\lambda^2}}$.

Figure 2.3 shows the diffusivity functions and the corresponding flux functions for linear diffusion and Perona-Malik type nonlinear diffusion. One can easily observe that for linear diffusion the diffusivity is constant ($g(s) = 1$), which results in a linearly increasing flux function. As a result, all points, including the discontinuities, are smoothed equally. For Perona-Malik, the diffusivity is variable and decreases as $|u_x|$ increases. It is evident that the decay in diffusivity is particularly rapid after the contrast parameter λ. This leads to two different behaviors in the diffusion process. Since $\frac{\partial u}{\partial t} = \Phi'(u_x)u_{xx}$, for the points where $|u_x| < \lambda$, $\Phi'(u_x) > 0$ which corresponds to lost in the material. For the points where $|u_x| > \lambda$, on the contrary, $\Phi'(u_x) < 0$ which generates an enhancement in the material. Hence, although the diffusivity is always nonnegative, one can observe both *forward* and *backward* diffusions during the smoothing process, and the contrast parameter λ separates the regions of forward diffusion from the regions of backward diffusion.

13
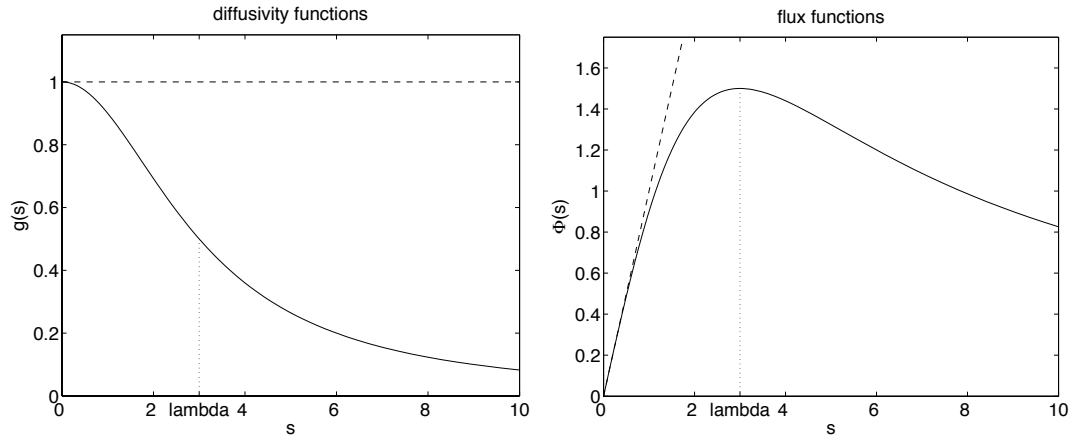
Figure 2.3: Diffusivities and the corresponding flux functions for linear diffusion (*plotted in dashed line*) and Perona-Malik type nonlinear diffusion (*plotted in solid line*). For Perona-Malik diffusivity $g(s) = \frac{1}{1+s^2/\lambda^2}$ is used with $\lambda = 3$.

If we consider the 2D case, the diffusivities are reduced at the image locations where $|\nabla u|^2$ is large. As $|\nabla u|^2$ can be interpreted as a measure of edge likelihood, this means that the amount of smoothing is low along image edges. In particular, the contrast parameter λ specifies a measure that determines which edge points are to be preserved or blurred during the diffusion process. Even edges can be sharpened due to the local backward diffusion behavior as discussed for the 1D case. Since the backward diffusion is a well-known ill-posed process, this may cause an instability, the so-called *staircasing effect*, where a piece-wise smooth region in the original image evolves into many unintuitive piecewise constant regions. Figure 2.4 shows an example where this instability occurs. The unintuitive regions such as the one at the woman's face and shoulder are clearly visible in Figure 2.4(b). A possible solution to this drawback is to use regularized gradients in diffusivity computations [36] (Figure 2.4(c)).

Replacing the diffusivities $g(|\nabla u|)$ with the regularized ones $g(|\nabla u_\sigma|)$ leads to the following equation:

$$\frac{\partial u}{\partial t} = \nabla \cdot (g(|\nabla u_\sigma|)\nabla u) \tag{2.15}$$

where $u_\sigma = G_\sigma * u$ represents a Gaussian-smoothed version of the image. Taking the equivalence of the Gaussian smoothing and the linear scale space into account, ∇u_σ can also be considered as the gradient computed at a specific scale $\sigma > 0$.

Some example results of regularized Perona-Malik filtering with different diffusion times are shown in Figure 2.5 and Figure 2.6. It is evident from these images that the corresponding smoothing process diminishes noise while retaining or even enhancing edges since it considers

14
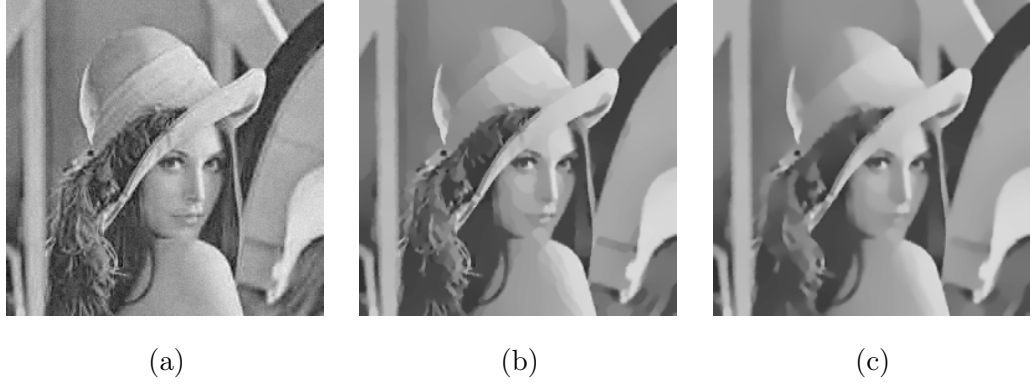
Figure 2.4: The staircasing effect. (a) Original noisy image. (b) Perona-Malik diffusion. (c) Regularized Perona-Malik diffusion.
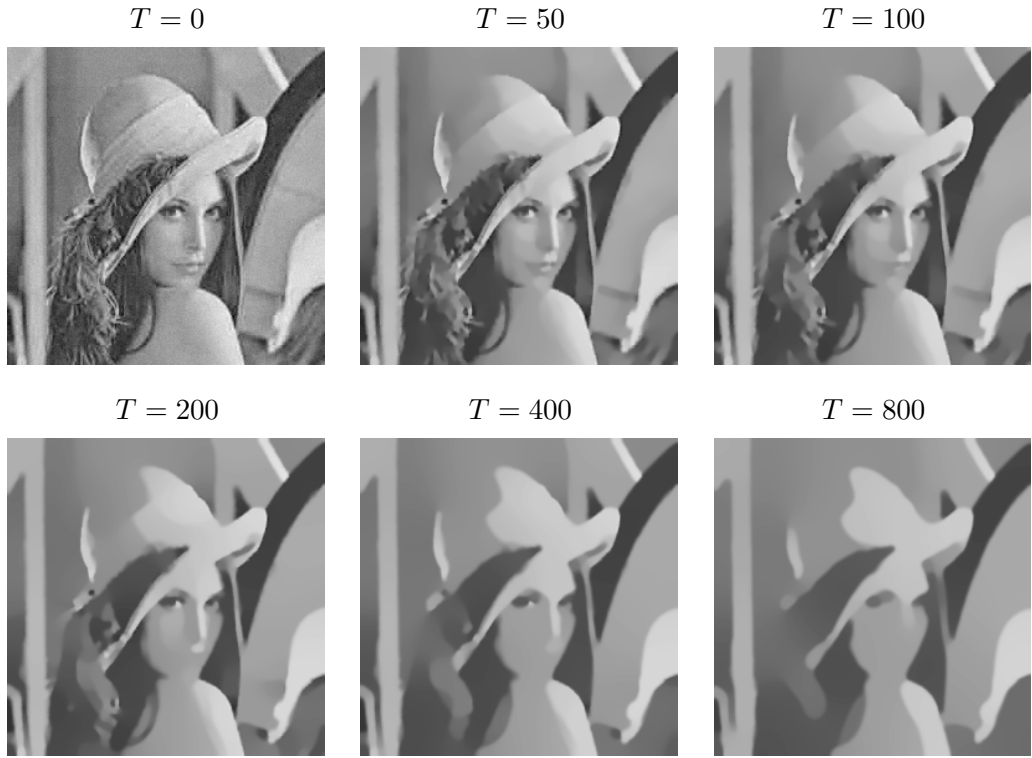
Figure 2.5: Regularized Perona-Malik results for different diffusion times ($\lambda = 1$, $\sigma = 1$).

a kind of a priori edge knowledge.

Numerical Implementation

For numerical implementation, we use central differences to approximate the gradient magnitude at a pixel (i, j) in the diffusivity estimation, $g_{i,j} = g(|\nabla u_{i,j}|)$:

$$|\nabla u_{i,j}| = \sqrt{\left(\frac{du_{i,j}}{dx}\right)^2 + \left(\frac{du_{i,j}}{dy}\right)^2} \approx \sqrt{\left(\frac{u_{i+1,j} - u_{i-1,j}}{2}\right)^2 + \left(\frac{u_{i,j+1} - u_{i,j-1}}{2}\right)^2}. \quad (2.16)$$
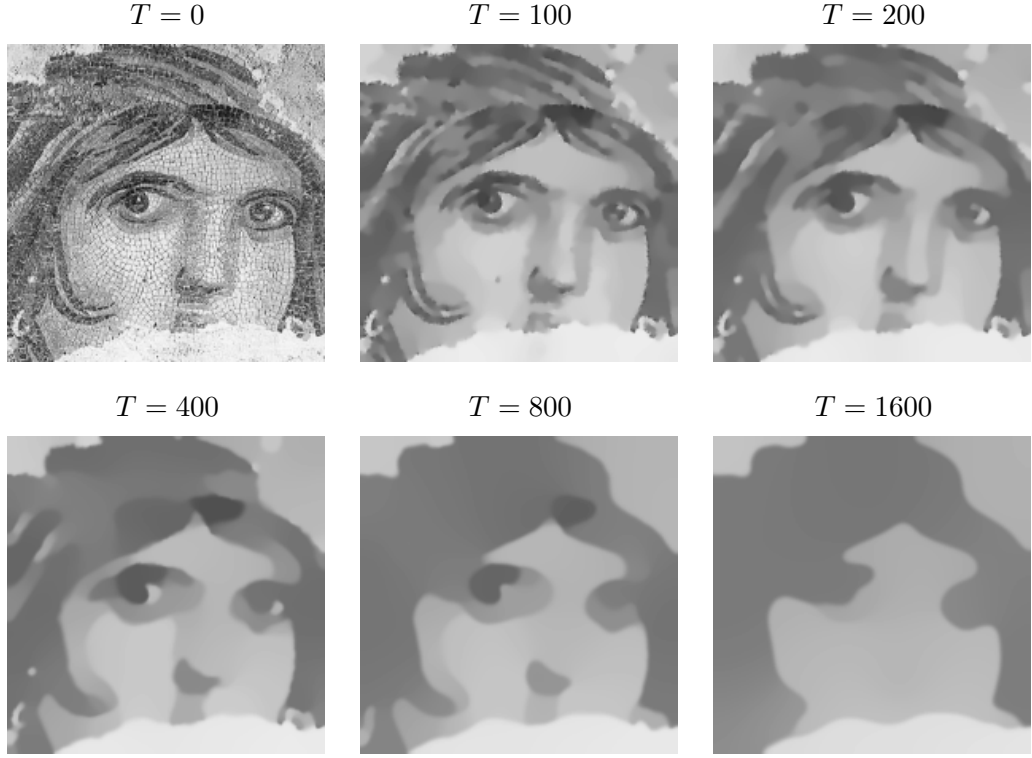
Figure 2.6: Regularized Perona-Malik results for different diffusion times ($\lambda = 1$, $\sigma = 1$).

The Perona-Malik equation (2.11) is first discretized w.r.t. spatial variables. This results in the following space-discrete equation:

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{\partial}{\partial x}\left(g(|\nabla u|)u_x\right) + \frac{\partial}{\partial y}\left(g(|\nabla u|)u_y\right), \\
\frac{du_{i,j}}{dt} &= g_{i+\frac{1}{2},j} \cdot (u_{i+1,j} - u_{i,j}) - g_{i-\frac{1}{2},j} \cdot (u_{i,j} - u_{i-1,j}) \\
&+ g_{i,j+\frac{1}{2}} \cdot (u_{i,j+1} - u_{i,j}) - g_{i,j-\frac{1}{2}} \cdot (u_{i,j} - u_{i,j-1}).
\end{aligned}
\tag{2.17}
$$

This discretization scheme requires the diffusivities to be estimated at mid-pixel points (Figure 2.7). They are simply computed by taking averages of the diffusivities over neighboring pixels:

$$
g_{i\pm\frac{1}{2},j} = \frac{g_{i\pm1,j} + g_{i,j}}{2}, \qquad g_{i,j\pm\frac{1}{2}} = \frac{g_{i,j\pm1} + g_{i,j}}{2} .
\tag{2.18}
$$

The time derivative in (2.17) can be discretized using forward difference. This yields an iterative scheme with an explicit time discretization, where homogeneous Neumann boundary condition is imposed along the image boundary

$$
\begin{aligned}
\frac{u_{i,j}^{k+1} - u_{i,j}^{k}}{\Delta t} &= g_{i+\frac{1}{2},j}^{k} \cdot u_{i+1,j}^{k} + g_{i-\frac{1}{2},j}^{k} \cdot u_{i-1,j}^{k} + g_{i,j+\frac{1}{2}}^{k} \cdot u_{i,j+1}^{k} + g_{i,j-\frac{1}{2}}^{k} \cdot u_{i,j-1}^{k} \\
&- \left(g_{i+\frac{1}{2},j}^{k} + g_{i-\frac{1}{2},j}^{k} + g_{i,j+\frac{1}{2}}^{k} + g_{i,j-\frac{1}{2}}^{k}\right) \cdot u_{i,j}^{k}
\end{aligned}
\tag{2.19}
$$

16

Figure 2.7: Discretization grid used in (2.17).

with Δt denoting the time step. For the Perona-Malik diffusion, the stability requirement is again $\Delta t \leq 0.25$.

2.3   Mumford-Shah (MS) Functional [100]

The formulation of Mumford and Shah [100] is based on a functional minimization via which a piecewise smooth approximation of a given image and an edge set are to be recovered simultaneously. In this unified formulation, smoothing and edge detection processes work jointly to partition an image into segments. The Mumford-Shah (MS) model is:

$$E_{MS}(u, \Gamma) = \beta \int_{\Omega} (u - f)^2 dx + \alpha \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + length(\Gamma) \qquad (2.20)$$

where

- $\Omega \subset \mathbb{R}^2$ is connected, bounded, open subset representing the image domain,

- f is an image defined on Ω,

- $\Gamma \subset \Omega$ is the edge set segmenting Ω,

- u is the piecewise smooth approximation of f,

- $\alpha, \beta > 0$ are the scale space parameters of the model.

The first term in E_{MS} is the *data fidelity* term which forces u to be close to the original image f. The next two terms are the generic priors that provide certain knowledge about the solution. Specifically, the second term, the so-called *regularization* or *smoothness* term, gives preference to piecewise smooth images by penalizing high gradients. Since the integral is over $\Omega \setminus \Gamma$, this prior is turned off at image boundaries, and thus it excludes image edges to be smoothed out. The third term is a penalty term on total edge length which prevents

the image to be split into many regions. Additionally, it implicitly imposes smoothness of the boundaries.

Generally, the unknown edge set Γ of a lower dimension makes the minimization of the MS model very difficult. Hence, in literature several approaches for approximating the MS model are suggested. In the next subsections, we review two such approaches which are proposed by Ambrosio and Tortorelli [3], and Chan and Vese [40], respectively.

2.3.1 Ambrosio-Tortorelli (AT) Approximation of the MS Functional [3]

Ambrosio and Tortorelli [3] suggested an approximation for the MS model by following the Γ convergence framework [24]. The basic idea is to introduce a smooth edge indicator function v which is more convenient than using the characteristic function χ_Γ as the edge indicator. The new function v depends on a parameter ρ, and as $\rho \to 0$, $v \to 1 - \chi_\Gamma$. That is, $v(x) \approx 0$ if $x \in \Gamma$ and $v(x) \approx 1$ otherwise. The result is the functional

$$E_{AT}(u,v) = \int_\Omega \left(\beta(u-f)^2 + \alpha(v^2|\nabla u|^2) + \frac{1}{2}\left(\rho|\nabla v|^2 + \frac{(1-v)^2}{\rho} \right) \right) dx \; . \qquad (2.21)$$

In the Ambrosio-Tortorelli (AT) model, the continuous function v encodes the boundary information. The value of v at a point can be interpreted as a measure of *boundaryness* where the low values indicate the edge points. That is, $v \approx 0$ along the boundaries and grows rapidly away from them. Thus, the function v may be thought as a blurred version of the edge set. The parameter ρ specifies the level of blurring (Figure 2.8), and as $\rho \to 0$, $\frac{1}{2}\int_\Omega \left(\rho|\nabla v|^2 + \frac{(1-v)^2}{\rho} \right) dx$ approximates the cardinality of the edge set Γ.

Piecewise smooth image u and the edge strength function v are simultaneously computed
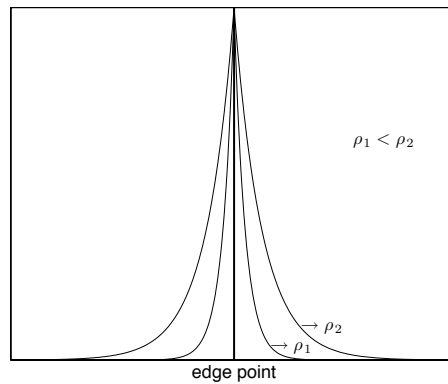
Figure 2.8: An example 1D edge strength function $(1-v)$ for two different values of ρ.

via the solution of the following system of coupled PDEs:

$$\frac{\partial u}{\partial t} = \nabla \cdot (v^2 \nabla u) - \frac{\beta}{\alpha}(u - f); \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = 0 , \tag{2.22}$$

$$\frac{\partial v}{\partial t} = \nabla^2 v - \frac{2\alpha|\nabla u|^2 v}{\rho} - \frac{(v-1)}{\rho^2}; \quad \left.\frac{\partial v}{\partial n}\right|_{\partial\Omega} = 0 \tag{2.23}$$

where $\partial\Omega$ denotes the boundary of Ω and n denotes the outer unit normal vector to $\partial\Omega$.

Notice that PDE for each variable can be interpreted as a biased diffusion equation that minimizes a convex quadratic functional in which the other variable is kept fixed:

Keeping v fixed, (2.22) minimizes a convex quadratic functional given by

$$\int_\Omega \left(\alpha v^2 |\nabla u|^2 + \beta(u - f)^2 \right) dx . \tag{2.24}$$

The data fidelity term in (2.22) provides a bias that forces u to be close to the original image f. In the regularization term, the edge strength function v specifies the boundary points and guides the smoothing accordingly. Since $v \approx 0$ along the boundaries, no smoothing is carried out at the boundary points, thus the edges are preserved.

Keeping u fixed, (2.23) minimizes a convex quadratic functional given by

$$\frac{\rho}{2} \int_\Omega \left(|\nabla v|^2 + \frac{1 + 2\alpha\rho|\nabla u|^2}{\rho^2} \left(v - \frac{1}{1 + 2\alpha\rho|\nabla u|^2} \right)^2 \right) dx . \tag{2.25}$$

The reciprocal relationship between v and $|\nabla u|^2$ can be clearly observed in the above functional. It asserts that the function v is nothing but a smoothing of $\frac{1}{1+2\alpha\rho|\nabla u|^2}$ with a blurring radius proportional to ρ and reciprocal to $|\nabla u|$. Ignoring the smoothness term $\rho|\nabla v|^2$, which mildly forces some spatial organization by requiring the edges to be smooth, and by letting $\rho \to 0$ [38, 131],

$$v \approx \frac{1}{1 + 2\alpha\rho|\nabla u|^2} . \tag{2.26}$$

Considering this approximation and the relation between variational regularization and diffusion equations (Section 2.1), we can interpret (2.22) as a Perona-Malik type nonlinear diffusion at a specific scale. Replacing v in (2.22) with $1/(1 + 2\alpha\rho|\nabla u|^2)$ yields the PDE

$$\frac{\partial u}{\partial t} = \nabla \cdot (g(|\nabla u|)\nabla u) - \frac{\beta}{\alpha}(u - f) \tag{2.27}$$

where $g(|\nabla u|) = \left(\frac{1}{1+|\nabla u|^2/\lambda^2} \right)^2$ with $\lambda^2 = 1/2\alpha\rho$. Thus, $\sqrt{1/2\alpha\rho}$ can be seen as a contrast parameter.

Since the parameters α and β define the relative importance of the regularization term, the scale is determined by the ratio α/β. Keeping the value of α fixed, decreasing the value of β results in more simplified results (Figure 2.9(b)-(c)). Moreover, the scale space parameters α and β also define the detail level in segmentation. With the ratio α/β fixed, the level of detail increases with the increasing α (Figure 2.9(b)-(d)).

19

(a)

(b)

(c)

(d)

Figure 2.9: Example segmentation results (u and $1 - v$). (a) Source image. (b)-(d) Corresponding segmentations obtained with the parameters (b) $\alpha = 1$, $\beta = 0.01$, $\rho = 0.01$, (c) $\alpha = 1$, $\beta = 0.001$, $\rho = 0.01$, and (d) $\alpha = 4$, $\beta = 0.04$, $\rho = 0.01$, respectively.

Numerical Implementation

Equations (2.22) and (2.23) can be simultaneously solved for u and v using standard numerical discretization techniques such as finite differences. We first discretize the coupled system with respect to spatial variables. This leads to the following space-discrete system of equations:

$$\begin{aligned}
\frac{du_{i,j}}{dt} &= v_{i+\frac{1}{2},j}^2 \cdot (u_{i+1,j} - u_{i,j}) - v_{i-\frac{1}{2},j}^2 \cdot (u_{i,j} - u_{i-1,j}) \\
&+ v_{i,j+\frac{1}{2}}^2 \cdot (u_{i,j+1} - u_{i,j}) - v_{i,j-\frac{1}{2}}^2 \cdot (u_{i,j} - u_{i,j-1}) \\
&- \frac{\beta}{\alpha} (u_{i,j} - f_{i,j}) \;, \tag{2.28} \\
\frac{dv_{i,j}}{dt} &= v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j} \\
&- \frac{2\alpha |\nabla u_{i,j}|^2 v_{i,j}}{\rho} - \frac{(v_{i,j} - 1)}{\rho^2} \;. \tag{2.29}
\end{aligned}$$

As in the discretization of the Perona-Malik equation, the diffusivities represented by the edge strength function v at mid-pixel points can be computed by taking averages over neighboring pixels:

$$v_{i\pm\frac{1}{2},j} = \frac{v_{i\pm1,j} + v_{i,j}}{2}, \qquad v_{i,j\pm\frac{1}{2}} = \frac{v_{i,j\pm1} + v_{i,j}}{2} \;. \tag{2.30}$$

The time derivatives in (2.28) and (2.29) can be discretized using forward differences, where regularization terms and the bias terms on the right hand side of each equation are evaluated at times k and $k + 1$, respectively.

$$\begin{aligned}
\frac{u_{i,j}^{k+1} - u_{i,j}^k}{\Delta t} &= \left(v_{i+\frac{1}{2},j}^k\right)^2 \cdot u_{i+1,j}^k + \left(v_{i-\frac{1}{2},j}^k\right)^2 \cdot u_{i-1,j}^k + \left(v_{i,j+\frac{1}{2}}^k\right)^2 \cdot u_{i,j+1}^k + \left(v_{i,j-\frac{1}{2}}^k\right)^2 \cdot u_{i,j-1}^k \\
&- \left(v_{i+\frac{1}{2},j}^k + v_{i-\frac{1}{2},j}^k + v_{i,j+\frac{1}{2}}^k + v_{i,j-\frac{1}{2}}^k\right) \cdot u_{i,j}^k - \frac{\beta}{\alpha}\left(u_{i,j}^{k+1} - f_{i,j}\right), \tag{2.31} \\
\frac{v_{i,j}^{k+1} - v_{i,j}^k}{\Delta t} &= v_{i+1,j}^k + v_{i-1,j}^k + v_{i,j+1}^k + v_{i,j-1}^k - 4v_{i,j}^k \\
&- \frac{\alpha\left(\left(u_{i+1,j}^k - u_{i-1,j}^k\right)^2 + \left(u_{i,j+1}^k - u_{i,j-1}^k\right)^2\right) v_{i,j}^{k+1}}{2\rho} - \frac{\left(v_{i,j}^{k+1} - 1\right)}{\rho^2} \tag{2.32}
\end{aligned}$$

where Δt denotes the time step.

Although the suggested scheme is neither fully explicit nor fully implicit, it still allows us to compute u^{k+1} and v^{k+1} by using forward recursion as in an explicit scheme. A numerical stopping criteria for the iterative scheme can be defined in the sense that the rate of change of u is less than a threshold. The corresponding minimization procedure is summarized in Algorithm 1.

Algorithm 1 Minimization of the Ambrosio-Tortorelli Model

1: Initialize the variables with $u^0 = f$, $v^0 = \frac{1}{1+2\alpha\rho|\nabla u^0|^2}$

2: **for** $k = 0$ to $kmax$ **do**

3:     Solve (2.31) for u^{k+1} using $\lfloor\sqrt{2\alpha/\beta}\rfloor$ steps

4:     **if** $|u^{k+1} - u^k| < \epsilon|u^k|$ **then**

5:         stop iterations

6:     **end if**

7:     Solve (2.32) for v^{k+1}

8: **end for**

2.3.2 Chan-Vese (CV) Approximation of the MS Functional [40]

Chan and Vese [40] proposed an approximation for the MS segmentation model by following the level-set based curve evolution formulation [106, 128]. Level sets provide an implicit contour representation where an evolving curve is represented with the zero-level line of a level set function (Figure 2.10). The basic aim of Chan and Vese (CV) model is to partition a given image into two regions that are likely to correspond object and background regions by embedding the object boundary by the zero-level curve of a 3D level set function.

Let ϕ be a level set function. Then, the Chan-Vese functional is

$$E_{CV}(c_1, c_2, \phi) = \lambda_1 \int_\Omega (f-c_1)^2 H(\phi)dx + \lambda_2 \int_\Omega (f-c_2)^2(1-H(\phi))dx + \mu \int_\Omega |\nabla H(\phi)|dx \quad (2.33)$$

where λ_1, $\lambda_2 > 0$ and $\mu \geq 0$ are fixed parameters. The length parameter μ can be interpreted as a scale parameter since it determines the relative importance of the length term. The possibility of detecting smaller objects/regions increases with decreasing μ.
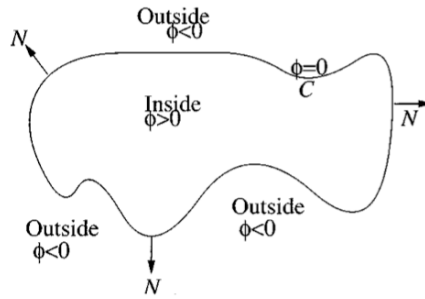
Figure 2.10: A curve can be represented as the zero-level line of a level set function (image taken from [40]).

The model represents the segmented image with the variables c_1, c_2 and $H(\phi)$, where $H(\phi)$ denotes the Heaviside function of the level set function ϕ defined by

$$H(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 . \end{cases} \tag{2.34}$$

The Heaviside function of the level set function, $H(\phi)$, specifies object and background regions in the observed image f, while the last term in (2.33), $\int_\Omega |\nabla H(\phi)|$, expresses the length of the object boundary. On top of that, the scalars c_1 and c_2 denote the average gray values of object and background regions indicated by $\phi \geq 0$ and $\phi < 0$, respectively. Thus, the CV model can be seen as a two-phase piecewise constant approximation of the MS model, which can theoretically be obtained by letting the weight α of the smoothness term in (2.20) tend to infinity, and forcing a two-region segmentation.

To segment a given image, the functional (2.33) needs to be minimized with respect to c_1, c_2, and ϕ. Keeping ϕ fixed, the average gray values c_1 and c_2 can be easily estimated by

$$c_1 = \frac{\int_\Omega f(x) H(\phi(x)) dx}{\int_\Omega H(\phi(x)) dx}, \tag{2.35}$$

$$c_2 = \frac{\int_\Omega f(x)(1 - H(\phi(x))) dx}{\int_\Omega (1 - H(\phi(x))) dx} . \tag{2.36}$$

Keeping c_1 and c_2 fixed and using the calculus of variations for the functional (2.33), the gradient descent equation for the evolution of ϕ is derived as

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\mu \, \nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda_1 (f - c_1)^2 + \lambda_2 (f - c_2)^2 \right] . \tag{2.37}$$

Figure 2.11 illustrates segmentation of a noisy image that contains objects of different shapes and sizes. We initialize the level set function ϕ with $\phi_0 = -\sqrt{(x - 100)^2 + (y - 100)^2} + 90$. As the zero-level line of the evolving level set function ϕ is attracted to object boundaries, a more accurate piecewise constant approximations of the original image f is recovered. Although some of the objects in the image have holes, they can be automatically detected by the CV model without considering additional curves since the level set formulation allows change of topology.

Numerical Implementation

In the numerical approximation of the CV model, generally, a regularized Heaviside function is used. For the remainder of this thesis, the following regularization is considered:

$$H_\varepsilon(z) = \frac{1}{2} \left(1 + \frac{2}{\pi} arctan \left(\frac{z}{\varepsilon} \right) \right), \tag{2.38}$$

$$\delta_\varepsilon(z) = \frac{dH_\varepsilon(z)}{dz} = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + z^2} . \tag{2.39}$$

23
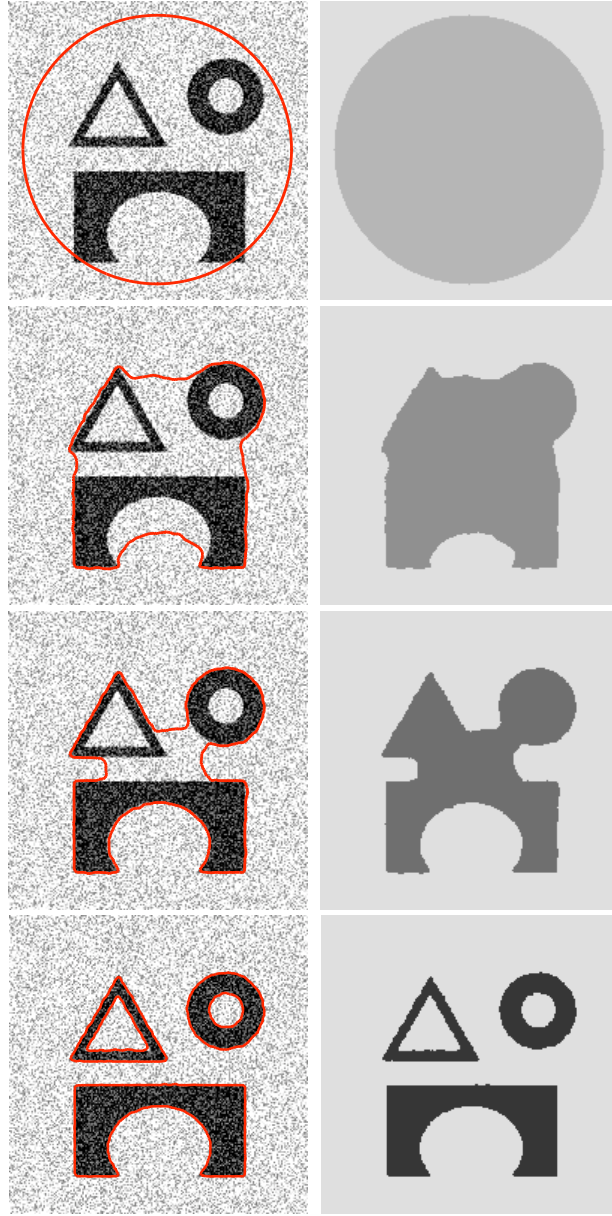
Figure 2.11: Example segmentation results (evolving contour ϕ superimposed on the original image f and the corresponding piecewise constant approximations of f). The parameters and the initial level set function are chosen as $\lambda_1 = \lambda_2 = 1$, $\mu = 0.5 \cdot 255^2$, $\varepsilon = 1$, and $\phi_0 = -\sqrt{(x-100)^2 + (y-100)^2} + 90$.

The evolution equation of ϕ (2.37) can be discretized by using standard finite differences as

$$
\begin{aligned}
\frac{\phi_{i,j}^{k+1} - \phi_{i,j}^{k}}{\Delta t} = \; \delta(\phi_{i,j}^{k}) \; \Bigg[&\mu \Delta_{-}^{x} \cdot \left(\frac{\Delta_{+}^{x} \phi_{i,j}^{k+1}}{\sqrt{(\Delta_{+}^{x} \phi_{i,j}^{k})^2 + (\phi_{i,j+1}^{k} - \phi_{i,j-1}^{k})^2/4}} \right) \\
&+ \mu \Delta_{-}^{y} \cdot \left(\frac{\Delta_{+}^{y} \phi_{i,j}^{k+1}}{\sqrt{(\phi_{i+1,j}^{k} - \phi_{i-1,j}^{k})^2/4 + (\Delta_{+}^{y} \phi_{i,j}^{k})^2}} \right) \\
&- \lambda_1 \left(f_{i,j} - c_1(\phi^k) \right)^2 + \lambda_2 \left(f_{i,j} - c_2(\phi^k) \right)^2 \Bigg] \qquad (2.40)
\end{aligned}
$$

where (i, j) denotes the pixel position, Δt is the time step, and forward and backward differences are defined as

$$
\Delta_{-}^{x} \phi_{i,j} = \phi_{i,j} - \phi_{i-1,j}, \quad \Delta_{+}^{x} \phi_{i,j} = \phi_{i+1,j} - \phi_{i,j},
$$
$$
\Delta_{-}^{y} \phi_{i,j} = \phi_{i,j} - \phi_{i,j-1}, \quad \Delta_{+}^{y} \phi_{i,j} = \phi_{i,j+1} - \phi_{i,j}.
$$

The minimization procedure is summarized in Algorithm 2. Keeping ϕ fixed, first the average gray values of object and background regions c_1 and c_2 are estimated. Next, the level set function ϕ is evolved according to (2.40). A numerical stopping criteria can be defined in the sense that the rate of change of ϕ or the overall energy (2.33) is less than a threshold.

Algorithm 2 Minimization of the Chan-Vese Model

1: Initialize the level set function with $\phi^0 = \phi_0$

2: **for** $k = 0$ to $kmax$ **do**

3:     Estimate $c_1(\phi^k)$ and $c_2(\phi^k)$ using (2.35) and (2.36), respectively

4:     Solve (2.40) for ϕ^{k+1}

5:     Check whether a numerical stopping criteria on ϕ is reached

6:     **if** it is reached **then**

7:         stop iterations

8:     **end if**

9: **end for**

2.4 Total Variation (TV) Regularization [122]

Rudin et al. [122] formulated image restoration as minimization of the total variation (TV) of a given image under certain assumptions on the noise. The Total Variation (TV) regularization model is generally defined as:

$$E_{TV}(u) = \int_\Omega \left(\frac{1}{2}(u - f)^2 + \alpha|\nabla u| \right) dx \tag{2.41}$$

where

- $\Omega \subset \mathbb{R}^2$ is connected, bounded, open subset representing the image domain,

- f is an image defined on Ω,

- u is the restored version of g,

- $\alpha > 0$ is a scalar.

The gradient descent equation for (2.41) is defined by

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) - \frac{1}{\alpha}(u - f); \quad \left. \frac{\partial u}{\partial n} \right|_{\partial \Omega} = 0 \ . \tag{2.42}$$

Since the value of α specifies the relative importance of the fidelity term, it can be interpreted as a scale parameter that determines the level of smoothing. In Figure 2.12, we illustrate the role of this parameter. The restored images are given in Figure 2.12(b)-(d). As it can be clearly seen, level of smoothing increases with increasing α.

In the original formulation of Rudin et al., the observed image f was assumed to be degraded by additive Gaussian noise with zero mean and known variance σ^2. Hence, in order to restore a given image, they propose to solve the constrained optimization problem

$$\min_u \int_\Omega |\nabla u| dx \tag{2.43}$$

subject to

$$\int_\Omega (u - f)^2 dx = \sigma^2 \ . \tag{2.44}$$

When TV regularization is defined as a constrained optimization problem, $\frac{1}{\alpha}$ can be considered as a Lagrange multiplier, which has to be determined by taking the given constraint into account.

In Section 2.1, we reviewed the relation between Tikhonov regularization and linear diffusion. In a similar manner, we can associate TV regularization with a nonlinear diffusion filter, the so-called *TV flow* [4, 52]. Ignoring the fidelity term in (2.42) leads to the PDE

$$\frac{\partial u}{\partial t} = \nabla \cdot (g(|\nabla u|)\nabla u) \tag{2.45}$$

26
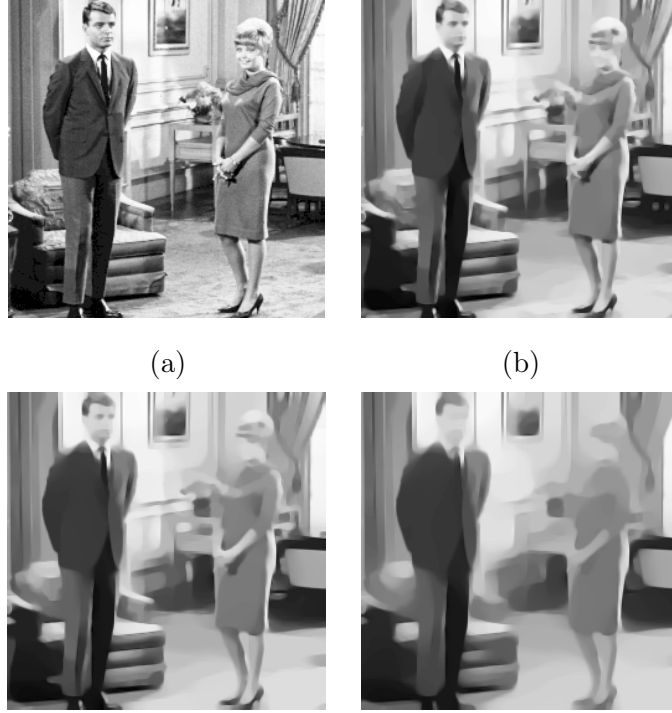
Figure 2.12: Example restoration results. (a) Source image. (b)-(d) Corresponding segmentations obtained with (b) $\alpha = 50$, (c) $\alpha = 100$, and (d) $\alpha = 200$.

with $u^0 = f$ and the diffusivity function $g(|\nabla u|) = \frac{1}{|\nabla u|}$. Notice that this diffusivity function has no additional contrast parameter as compared with the Perona-Malik diffusivities in (2.12) and (2.13). Figure 2.13 and Figure 2.14 depict TV scale space examples sampled at different diffusion times for two different images. It is evident from these images that the corresponding smoothing process yields segmentation-like, piecewise constant images.

Numerical Implementation

Like the models that we discussed in the previous sections of this chapter, the evolution equation of u (2.42) can be discretized by using standard finite differences. An important point is that the solution of TV regularization or equivalently TV flow leads to singular diffusivities as shown in (2.45). In numerical implementations based on standard discretization, this leads to stability problems as the image gradient tends to zero. A common solution to this problem is to add a small positive constant ϵ to image gradients.

After ϵ-regularization, the space-discrete version of (2.42) can be written as:
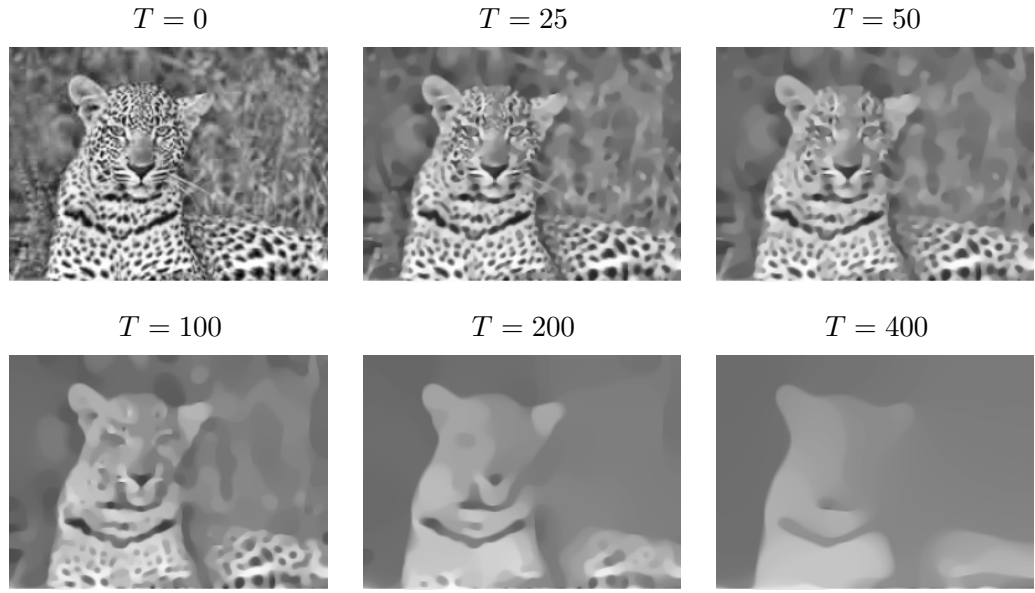
Figure 2.13: TV flow results for different diffusion times.
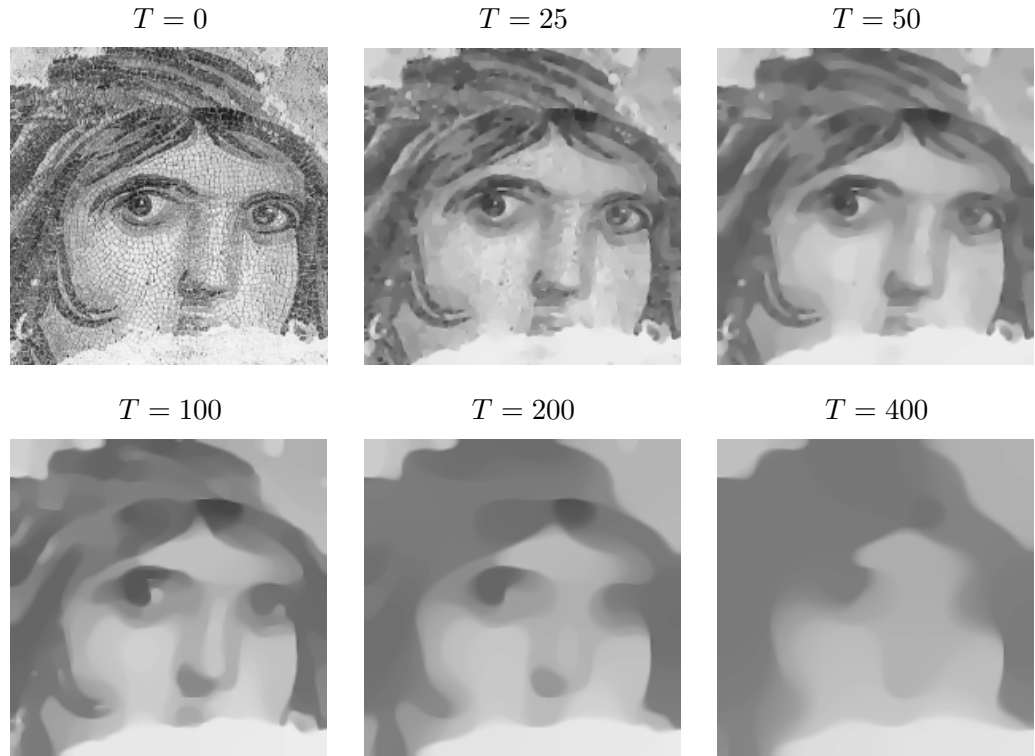
Figure 2.14: TV flow results for different diffusion times.

$$\frac{\partial u}{\partial t} \;=\; \frac{\partial}{\partial x}\left(\frac{u_x}{\sqrt{|\nabla u|^2 + \epsilon^2}}\right) + \frac{\partial}{\partial y}\left(\frac{u_y}{\sqrt{|\nabla u|^2 + \epsilon^2}}\right) - \frac{1}{\alpha}(u - f)$$

$$
= \frac{u_{xx}\left(u_y^2 + \epsilon^2\right) - 2u_x u_y u_{xy} + u_{yy}\left(u_x^2 + \epsilon^2\right)}{\left(u_x^2 + u_y^2 + \epsilon^2\right)^{\frac{3}{2}}} - \frac{1}{\alpha}(u - f) ,
$$

$$
\frac{du_{i,j}}{dt} = \frac{\frac{d^2 u_{i,j}}{dx^2}\left(\left(\frac{du_{i,j}}{dy}\right)^2 + \epsilon^2\right) - 2\left(\frac{du_{i,j}}{dx}\right)\left(\frac{du_{i,j}}{dy}\right)\left(\frac{d^2 u_{i,j}}{dxdy}\right) + \frac{d^2 u_{i,j}}{dy^2}\left(\left(\frac{du_{i,j}}{dx}\right)^2 + \epsilon^2\right)}{\left(\left(\frac{du_{i,j}}{dx}\right)^2 + \left(\frac{du_{i,j}}{dy}\right)^2 + \epsilon^2\right)^{\frac{3}{2}}}
$$

$$
- \frac{1}{\alpha}\left(u_{i,j} - f_{i,j}\right) \tag{2.46}
$$

where $\frac{d^2 u_{i,j}}{dx^2}$, $\frac{d^2 u_{i,j}}{dy^2}$ and $\frac{du_{i,j}}{dx}$, $\frac{du_{i,j}}{dy}$ can be approximated as in (2.4) and (2.16), respectively and

$$
\frac{d^2 u_{i,j}}{dxdy} \approx \frac{u_{i+1,j+1} - u_{i+1,j-1} - u_{i-1,j+1} + u_{i-1,j-1}}{4} . \tag{2.47}
$$

When explicit time discretization is employed, numerical solution to (2.46) can be computed by the following iterative scheme, where homogeneous Neumann boundary condition is imposed along the image boundary:

$$
\begin{aligned}
\frac{u_{i,j}^{k+1} - u_{i,j}^k}{\Delta t} = {} & \left(\left(\frac{u_{i+1,j}^k - u_{i-1,j}^k}{2}\right)^2 + \left(\frac{u_{i,j+1}^k - u_{i,j-1}^k}{2}\right)^2 + \epsilon^2\right)^{-\frac{3}{2}} \\
& \cdot \left[\left(u_{i+1,j}^k - 2u_{i,j}^k + u_{i-1,j}^k\right)\left(\left(\frac{u_{i,j+1}^k - u_{i,j-1}^k}{2}\right)^2 + \epsilon^2\right) \right. \\
& \quad - \frac{1}{8}\left(u_{i+1,j}^k - u_{i-1,j}^k\right)\left(u_{i,j+1}^k - u_{i,j-1}^k\right) \\
& \qquad \left(u_{i+1,j+1}^k - u_{i+1,j-1}^k - u_{i-1,j+1}^k + u_{i-1,j-1}^k\right) \\
& \quad + \left. \left(u_{i,j+1}^k - 2u_{i,j}^k + u_{i,j-1}^k\right)\left(\left(\frac{u_{i+1,j}^k - u_{i-1,j}^k}{2}\right)^2 + \epsilon^2\right)\right] \\
& - \frac{1}{\alpha}\left(u_{i,j}^k - f_{i,j}\right)
\end{aligned} \tag{2.48}
$$

where Δt denotes the time step. ϵ-regularization requires the stability condition $\Delta t \leq 0.25\epsilon$, and thus a sufficiently large number of iterations is needed for small values of ϵ.

2.5 A Common Framework for Curve Evolution, Segmentation and Anisotropic Diffusion [131]

In Section 2.3.1, we reviewed the Ambrosio-Tortorelli approximation of the Mumford-Shah functional. The key idea of this approximation is to utilize a continuous edge strength function v. The value of v approaches to 0 at the object boundaries and grows rapidly as image gradients become small. In [131], Shah suggested a modification to the Ambrosio-Tortorelli model (2.21), where the quadratic cost functions in both the data fidelity and the

smoothing terms are replaced with L^1-functions. The modified energy is:

$$E_S(u,v) = \int_\Omega \left(\beta\,|u-f| + \alpha\,v^2|\nabla u| + \frac{1}{2}\left(\rho|\nabla v|^2 + \frac{(1-v)^2}{\rho} \right) \right) dx \;. \qquad (2.49)$$

As $\rho \to 0$, this energy functional converges to the following functional:

$$E_{S2}(u,\Gamma) = \frac{\beta}{\alpha}\int_\Omega |u-f|dx + \int_{\Omega\backslash\Gamma} |\nabla u|dx + \int_\Gamma \frac{J_u}{1+\alpha J_u}ds \qquad (2.50)$$

with $J_u = |u^+ - u^-|$ indicating the jump in u across Γ. u^+ and u^- denote intensity values on two sides of Γ, respectively, and thus each boundary point is weighted according to its level of contrast.

Minimizing E_S corresponds to the gradient descent equations:

$$\frac{\partial u}{\partial t} = 2\nabla v \cdot \nabla u + v|\nabla u|\,curv(u) - \frac{\beta}{\alpha v}|\nabla u|\frac{(u-f)}{|u-f|}; \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = 0\,, \qquad (2.51)$$

$$\frac{\partial v}{\partial t} = \nabla^2 v - \frac{2\alpha|\nabla u|v}{\rho} - \frac{(v-1)}{\rho^2}; \quad \left.\frac{\partial v}{\partial n}\right|_{\partial\Omega} = 0 \qquad (2.52)$$

with $curv(u) = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right)$.

The second equation (2.52) is very similar to the evolution equation of v (2.23) in the AT model; only $|\nabla u|^2$ is replaced with $|\nabla u|$. The determining factor of the model is the new evolution equation of u (2.51). Replacing L^2-norms in both the data fidelity and the smoothness terms by their L^1-norms generates shocks in u and thus object boundaries are recovered as actual discontinuities. As it can be clearly seen from Figure 2.15, the suggested smoothing process of u gives rise to more cartoon-like, piecewise constant images (these results are obtained by using a half-quadratic approximation of Shah's modified energy proposed in [76]). However, the robust norms utilized attract the image towards the cartoon limit and catch unintuitive regions such as the one at the man's right shoulder and the ones on the floor. It is important to remark that the effect of the scale space parameters α and β on segmentation results is similar compared to the one in AT model (cf. Figure 2.9). The amount of smoothing is determined by the ratio α/β, and increasing the value of α while keeping α/β fixed leads to more detailed segmentations.

One of the underlying assumptions of the original MS model and AT approximation is that the filtered image varies from the observed image by Gaussian noise. Hence, when a source image is corrupted by impulse noise, the corresponding smoothing process produces inadequate results. However, replacing the L^2-norm with the L^1 in the modified model yields to a robust data fidelity term that can cope with impulse noise. For example, consider the noisy image given in Figure 2.16(a), which is degraded with 5% salt and pepper noise.
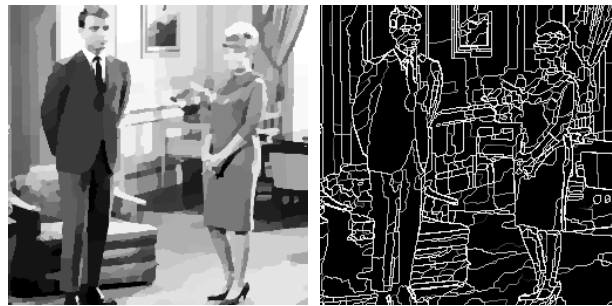
Figure 2.15: Example segmentation results (u and $1 - v$). (a) Source image. (b)-(d) Corresponding segmentations obtained with parameters (b) $\alpha = 1$, $\beta = 0.01$, $\rho = 0.01$, (c) $\alpha = 1$, $\beta = 0.001$, $\rho = 0.01$, and (d) $\alpha = 4$, $\beta = 0.04$, $\rho = 0.01$.

| (a) | (b) | (c) |
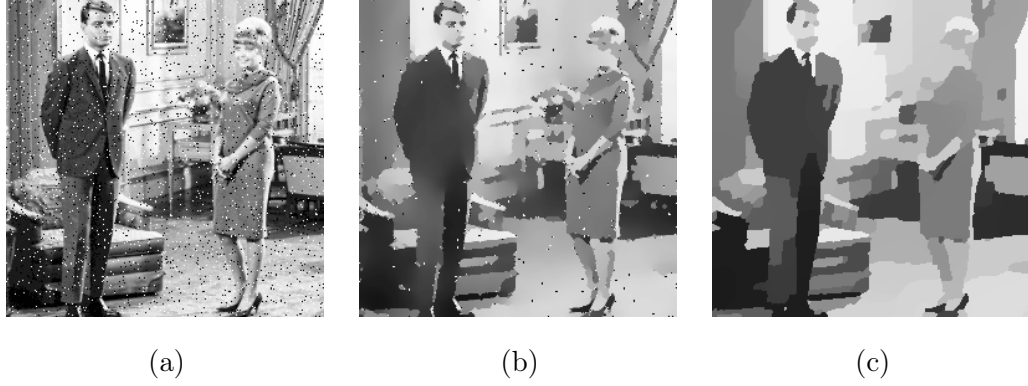
Figure 2.16: Segmentation of a noisy image degraded with 5% salt and pepper noise. (a) Source image. (b) Reconstruction using AT model. (c) Reconstruction using Shah's modified functional (both results are obtained with $\alpha = 1$, $\beta = 0.01$, $\rho = 0.01$).

Figure 2.16(b) and (c) depict the outcomes of the AT approximation and the modified model, respectively. As they demonstrate, the modified model eliminates the impulse noise during smoothing, however, noise still present in the AT result.

2.6 Edge Enhancing Diffusion [153]

The nonlinear diffusion model proposed by Perona and Malik employs a scalar-valued diffusivity function to guide the smoothing process as summarized in Section 2.2. The diffusivities are reduced at the image locations where the magnitude of image gradient $|\nabla u|^2$ is large, and as a result, the edges are preserved or even enhanced. In [153], Weickert suggested an alternative approach that additionally takes direction of the image gradients into account. The suggested model is an anisotropic nonlinear diffusion model with better edge enhancing capabilities.

In general, any anisotropic nonlinear diffusion can be described by the equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(\nabla u)\nabla u) \tag{2.53}$$

where u is the smoothed image that is initialized with the input image f (i.e. $u^0(x) = f(x)$), and D represents a matrix-valued diffusion tensor that describes the smoothing directions and the corresponding diffusivities. One can easily observe that for linear diffusion the diffusion tensor can be defined as $D(\nabla u) = I$, which results in a constant diffusion coefficient for all image points in all directions. Similarly, for Perona-Malik type nonlinear diffusion, $D(\nabla u) = g(|\nabla u_\sigma|)I$. Such a choice reduces the amount of smoothing at image edges, but

in an equal amount in all directions. In actual anisotropic setting, the diffusion tensor D is defined as a function of the structure tensor given by

$$J(\nabla u) = \nabla u \nabla u^T = \begin{bmatrix} u_x^2 & u_x u_y \\ u_x u_y & u_y^2 \end{bmatrix}. \tag{2.54}$$

The structure tensor $J(\nabla u)$ can be interpreted as an image feature describing the local orientation information. It has an orthonormal basis of eigenvectors v_1 and v_2 with $v_1 \parallel \nabla u$ and $v_2 \perp \nabla u$, and the corresponding eigenvalues $\lambda_1 = |\nabla u|^2$ and $\lambda_2 = 0$. It is important to note that noise significantly affects the tensor estimation. Thus the given image u is usually convolved with a Gaussian kernel G_σ with a relatively small standard deviation σ as a presmoothing step and the structure tensor is computed accordingly by using $\nabla u_\sigma = \nabla(G_\sigma * u)$ instead of ∇u.

The main idea behind edge enhancing diffusion is to use the structure tensor as an image/edge descriptor to construct a diffusion tensor that reduces the amount of smoothing across the edges while smoothing is still carried out along the edges. In order to perform this, Weickert proposed to utilize same orthonormal basis of eigenvectors $v_1 \parallel \nabla u_\sigma$ and $v_2 \perp \nabla u_\sigma$ estimated from the structure tensor $J(\nabla u_\sigma)$ with the following choice of eigenvalues satisfying $\frac{\lambda_1(|\nabla u_\sigma|)}{\lambda_2(|\nabla u_\sigma|)} \to 0$ for $|\nabla u_\sigma| \to \infty$

$$\lambda_1(|\nabla u_\sigma|) = \begin{cases} 1 & \text{if } |\nabla u_\sigma| = 0 \\ 1 - exp\left(-\frac{3.31488}{(|\nabla u_\sigma|/\lambda)^8}\right) & \text{otherwise,} \end{cases} \tag{2.55}$$

$$\lambda_2(|\nabla u_\sigma|) = 1 \tag{2.56}$$

where λ denotes the contrast parameter.

Such a choice preserves and enhances image edges by reducing the diffusivity λ_1 perpendicular to edges for sufficiently large values of $|\nabla u_\sigma|$. Specifically, the diffusion tensor is given by the formula

$$D = \begin{bmatrix} (u_\sigma)_x & -(u_\sigma)_y \\ (u_\sigma)_y & (u_\sigma)_x \end{bmatrix} \cdot \begin{bmatrix} \lambda_1(|\nabla u_\sigma|) & 0 \\ 0 & \lambda_2(|\nabla u_\sigma|) \end{bmatrix} \cdot \begin{bmatrix} (u_\sigma)_x & -(u_\sigma)_y \\ (u_\sigma)_y & (u_\sigma)_x \end{bmatrix}^{-1}. \tag{2.57}$$

Figure 2.17 and Figure 2.18 illustrate example results of edge enhancing diffusion filter for different diffusion times. As it can be clearly seen from these images, the corresponding smoothing process diminishes noise and fine image details while retaining and even enhancing edges as in the Perona-Malik type nonlinear diffusion. On the other hand, the corners become more rounded in the anisotropic model compared to the Perona-Malik filter (cf. Figure 2.5 and Figure 2.6) since edge enhancing diffusion allows smoothing along edges while preventing

33
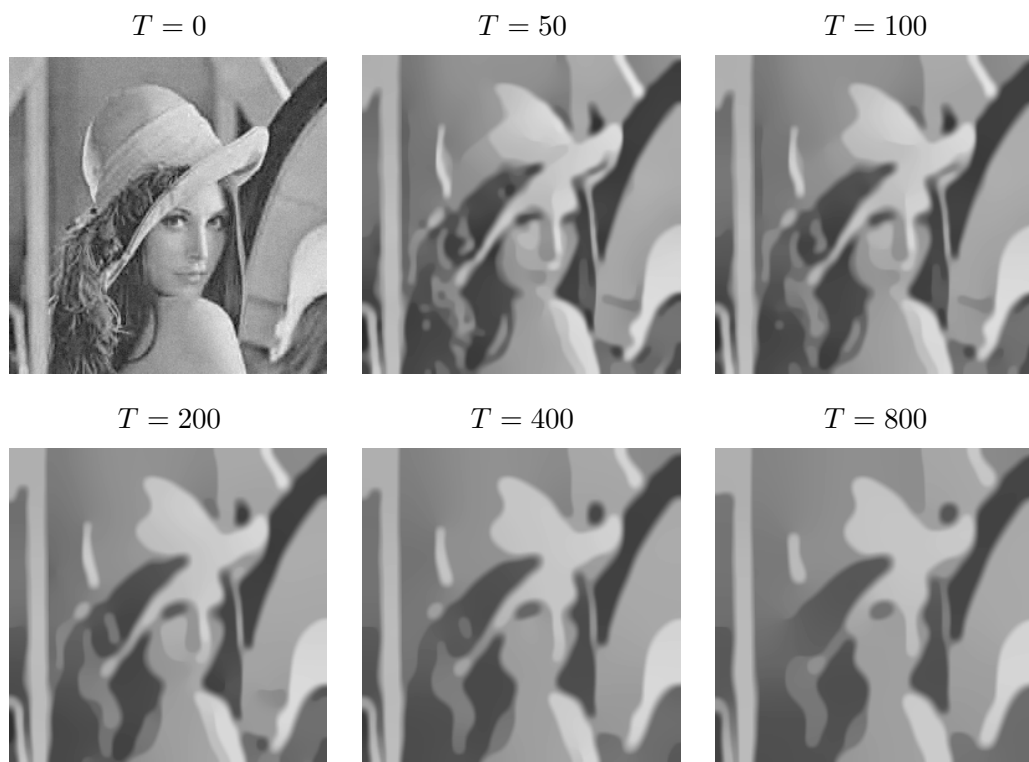
Figure 2.17: Edge enhancing diffusion results for different diffusion times ($\lambda = 2$, $\sigma = 1$).
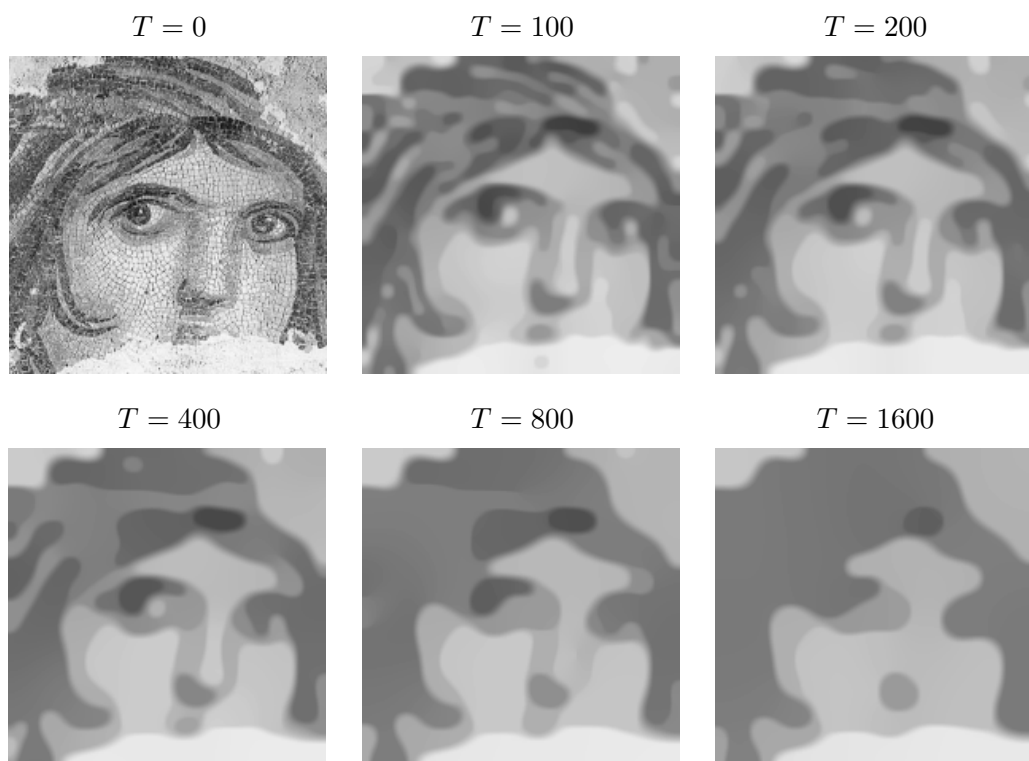
Figure 2.18: Edge enhancing diffusion results for different diffusion times ($\lambda = 1.8$, $\sigma = 1$).

smoothing across them. As discussed in [154], this causes a slight shrinking effect in the image structures, which eliminates fine or thin structures better than the Perona-Malik model. Thus, through this process one can capture semantically more correct image regions.

2.7   Discussion

All of the studies reviewed here offer elegant solutions for image segmentation and restoration problems as they utilize a unified approach that combines smoothing and edge detection. Variational methods formulate the problems in terms of energy minimization where the functional to be minimized depends on a global criterion. Typically, these functionals include regularization terms, such as terms enforcing smoothness of image regions or continuity of contours, that bias the solutions towards the expected results. One can minimize these functionals by solving the corresponding Euler-Lagrange equations. The nonlinear diffusion models, on the other hand, do not bother with devising such energies, and function directly on the PDEs.

As we discussed in Section 2.1, there is a close relationship between these two approaches as the Euler-Lagrange equations obtained from the variational formulations can be associated with nonlinear diffusion models. The local interactions between the image points, which are governed by the diffusion models, give rise to solutions that are consistent with the global criterion considered even though the computations are carried out in a local fashion. The critical point in these studies is that as the models get more and more complicated, implementation of the numerics becomes much more challenging.

It is important to note that the prior information included in the regularization terms of these models is very generic. This usually restricts the class of images to be segmented (e.g. the Mumford-Shah model enforces the segmented image to be piecewise smooth). One way to extend these models is to take additional image features other than intensity information into account. For instance, in [2, 123], the authors suggested to use Gabor filter responses to describe the input images, resulting in formulations that deal with textured images.

CHAPTER 3

PRIOR-GUIDED SEGMENTATION

In Chapter 2, important diffusion models and variational regularization frameworks were reviewed. Most of these works use very generic prior knowledge such as smoothness of image regions and continuity of contours to bias the segmentation. For natural images, these generic priors generally yield poor results due to the ambiguities in low-level visual cues. The only way to resolve these ambiguities is to consider high-level prior information. Interestingly, in the last decade, there is an increasing number of works that employ prior shape knowledge of objects within a given class to guide the segmentation process (e.g. [22, 44, 73, 87, 89, 105, 116, 120, 135, 148]). In this chapter, some of these important studies are reviewed in two parts.

Although the main focus of this thesis is on variational formulations, the first part of this chapter is dedicated to the studies that are formulated within a probabilistic framework. These studies are mainly proposed for detecting or recognizing objects within a given class, yet they also provide simultaneous segmentation results.

The studies that we survey in the remaining part of this chapter are influential examples of variational frameworks that integrate low-level image features with a priori information about the shape of an object of interest. Main goal of these frameworks is to delineate actual boundaries of the objects in the presence of corrupting influences such as noise and partial occlusion. The advantages and disadvantages of these formulations are compared and contrasted at the end of the chapter.

3.1 Probabilistic Formulations

Objects within a specific class have a large variability in their shape and appearance that makes segmenting out them in images a very difficult task. In the following sections, we will review some important examples of prior shape based segmentation methods that aim to solve this problem by utilizing probabilistic formulations. As stated before, these studies are mainly proposed for detecting objects from images. However, the detection results also provide a way to obtain simultaneous segmentation of objects. The key idea behind all these methods is to learn some kind of shape and appearance models from a set training images and to use these models in recognition and segmentation.

3.1.1 The Work of Borenstein and Ullman [22]

The work of Borenstein and Ullman is one of the earliest methods proposed in this domain. Borenstein and Ullman specifically aimed to obtain top-down segmentation of an object within a given class. The essential difference between their method and the studies that are summarized in the subsequent sections is that their approach does not include any object recognition procedure.

In their method, they use image patches (or fragments) together with their figure-ground labels to model the class-specific object knowledge. Segmenting a novel input image is obtained by following an approach similar to solving a jigsaw-puzzle. From a set of image patches representing the given class, the ones that constitute a consistent cover of the image data are tried to be grouped together. The method also defines an object vs. background separation since the figure-ground segmentations of the patches are provided in advance. Figure 3.1 presents an overview of this procedure.

The segmentation method can be divided into two main stages: a learning phase where image patches representing a specific object class are extracted, and an actual segmentation phase which uses these patches to partition a novel image into two as figure and ground. In the learning phase, the training images are divided into two as class and non-class images and candidate image patches of various sizes are extracted from the class images. Then, an optimal set of class-specific image patches, which are more common in class images compared to non-class images, are selected by considering a similarity measure based on the normalized correlation. These patches are the ones that best describe the given class when the variability of local shape and appearance is taken into account. During the learning phase, a figure-ground label is also stored for each image patch. While these labels are directly extracted
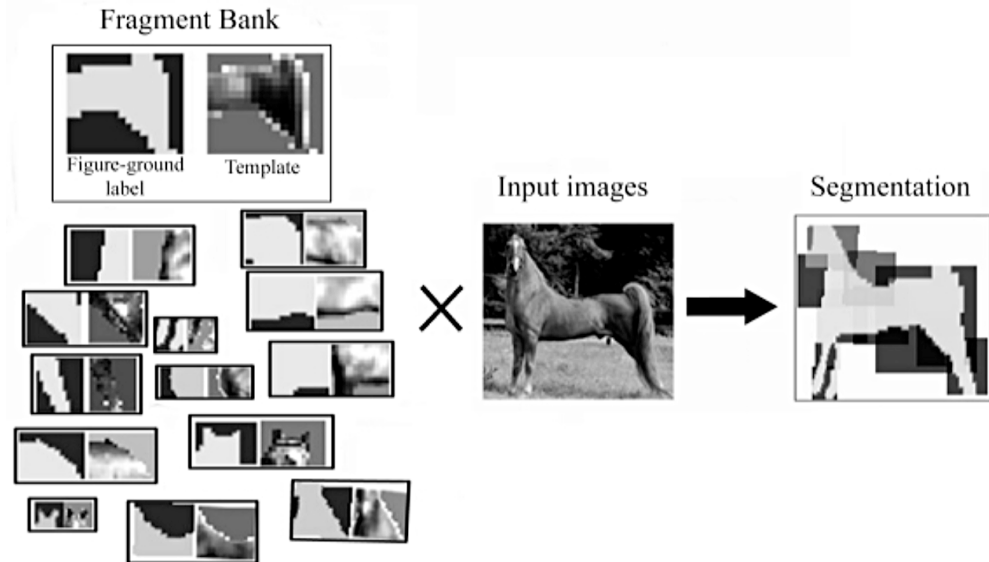
Figure 3.1: Overview of the class-specific segmentation (image taken from [22]).

from the segmented training examples in [22], Borenstein and Ullman suggested to learn them automatically in their former work [23].

In the segmentation phase, the class-specific segmentation is obtained by using a greedy algorithm that estimates an optimal cover of the novel image with the patches extracted in the learning phase. Among all possible covers, this algorithm searches for the one that maximizes a function that measures the individual match quality of the patches with the novel image, the local consistency of the cover, and the reliability of the assembled patches.

The main issue with this top-down approach is that the resulting segmentation is defined only by the figure-ground labels of the image patches that participate the estimated optimal cover. Thus, this process generally yields approximate (inaccurate) object boundaries that are not fully consistent with the image data. In a recent work [21], the version of this method that automatically learns the figure-ground labels of the patches [23] is combined with a bottom-up segmentation algorithm [134]. As reported in [21], while this combined approach increases the quality of the segmentations, it does not completely solve this problem.

3.1.2 The Work of Leibe et al. [87, 88]

The method proposed by Leibe et al. offers a combined approach to object recognition and segmentation. As in the work of Borenstein and Ullman, their approach uses a patch-based idea that is composed of two stages: learning a class-specific object model from a set of training images, and the following object recognition procedure.

The initial part of the learning phase consists of constructing a codebook of local appearances of a given object class. This is performed by first applying an interest point detector to training images and extracting the most informative regions from each image. Subsequent to this, all extracted patches are represented by using a local descriptor, and then the (visually) similar ones are grouped together. The grouping process is based on an unsupervised clustering method called the agglomerative clustering. Next, each cluster is described by its mean, yielding a dense set of local appearances. These resulting cluster centers correspond to the local structures appear on objects within the given class and define the codebook entries. In the final part of the learning phase, these codebook entries are used to built a class-specific object model called *the Implicit Shape Model*. This model includes both the extracted codebook entries (local appearances), and the information about where they may occur on the object by means of non-parametric spatial probability distributions (one for each entry). These distributions are estimated by performing an additional pass over the training images and comparing every codebook entry with the extracted local image patches around the detected interest points. If the similarity between an entry and an image patch is above a threshold, the position information is stored in the shape model relative to the object center. The whole learning procedure is illustrated in Figure 3.2. To obtain a simultaneous class-specific top-down segmentation, this shape model can be enriched by additionally keeping the figure-ground segmentations of the codebook entries.

After the object shape model is constructed, when a test image is provided, the recognition phase carries out the following steps. First, an interest point detector is applied to the test image and local image patches are extracted around the points detected. Then, these patches are compared to the entries in the codebook. For every match having a similarity value above a threshold, the corresponding codebook entry provides a probabilistic vote for potential location of the object center by taking the learned spatial probability distribution
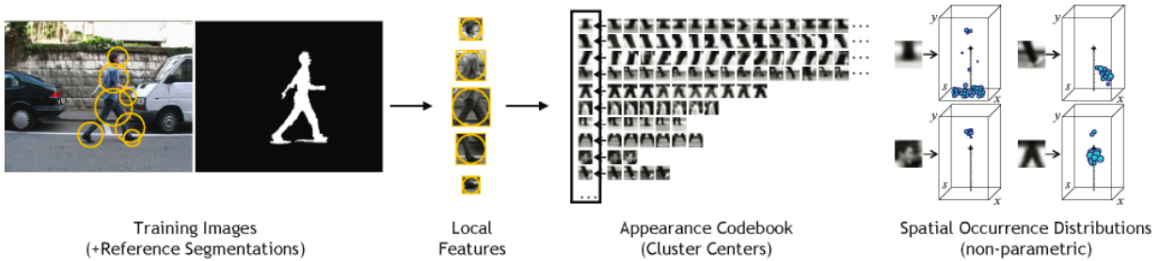
Figure 3.2: Overview of the learning procedure (image taken from [88]).

into account. Next, the votes are collected by using a Generalized Hough Transform and the local maxima in the vote space are estimated using the Mean-Shift mode estimation algorithm. Every local maximum point provides an object hypothesis, thus corresponds to an initial recognition result. The object hypotheses can be refined by backprojecting the participating votes to the image and considering the nearby image patches that are excluded by the interest point detector. More importantly, if the figure-ground segmentations of the codebook entries are available, a joint top-down segmentation result can be obtained. The key idea is to build pixel-wise probabilistic segmentation maps for both the object and the background, where the probability values are estimated by taking the weighted averages of the figure-ground labels coming from the activated codebook entries containing the corresponding pixel. In [87, 88], Leibe et al. also showed that it is possible to improve the initial recognition results by using the top-down segmentation result. For this, they used the obtained figure-ground knowledge to eliminate the ambiguous information due to the background pixels.

Figure 3.3 presents an overview of the summarized joint class-specific recognition and segmentation procedure. As it can be seen from the example provided in the figure, the recognition method proposed by Leibe et al. provides a significant advantage as it can cope with multiple objects. On the other hand, the joint segmentation procedure results in fuzzy object boundaries that are not completely consistent with the image data as in the work of Borenstein and Ullman.
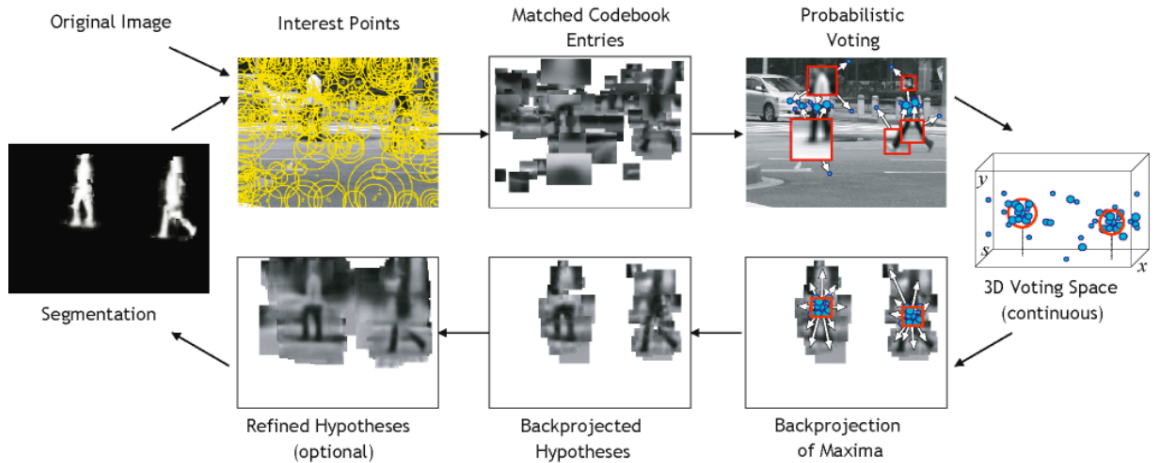
Figure 3.3: Overview of the joint class-specific recognition and segmentation procedure (image taken from [88]).

3.1.3 The Works of Shotton et al. [135] and Opelt et al. [105]

We prefer to review these two prior shape based methods within the same section since the motivations behind them are very closely related. While the works that we summarized in the previous sections use local image regions (patches) to represent class-specific object knowledge, the ones suggested by Shotton et al. and Opelt et al. employ local contour information for this purpose. The main intention of the authors in both works is to detect objects in images. Although this is the case, the detection results provide some kind of joint-segmentation results. Both methods include learning and testing stages. In the learning stage, a class-specific object model/detector that relies on local contour features is learned from a set of training images containing class and non-class images. The testing phase uses the learned model to recognize objects in novel images.

In [135], the learning process is performed in two stages in a semi-supervised manner. In the initial phase, a moderate set of class images containing reference figure-ground segmentations is used to construct a codebook of local contours. The local contours are simply extracted from the edge maps of randomly selected local regions on the segmentation masks. Each entry in the codebook corresponds to a local contour and its spatial position relative to the object center. Next, a location-specific classifier, which is modeled as weighted summation of the responses of a number of codebook entries, is learned from class and non-class images by using a boosting algorithm. The algorithm not only selects the most informative local contours within the codebook entries, but also estimates the optimal parameters for the classifier. Given a novel image, detecting an object corresponds to computing a classification map using the learned classifier. Simply, the recognition process consists of the following steps. First, the edge map of the image is extracted. Then, each local contour participating the learned classifier is matched with the extracted image edges using the oriented chamfer matching and the ones that are activated are used to estimate possible locations for the object center. This process can be performed by computing the total response of the learned classifier for every hypothesis which results in a classification map where each point in the map gives a measure of the object presence. In the final phase of the learning stage, the learned classifier is evaluated on a set of unlabeled class images, and the set of non-class images used in the initial learning phase. The false detections in the non-class images and the detection results in the unlabeled training data are then used to train a more accurate classifier. Some sample recognition results are presented in Figure 3.4. As seen in the figure, the local contour fragments that are involved in detection provide approximate boundaries

41
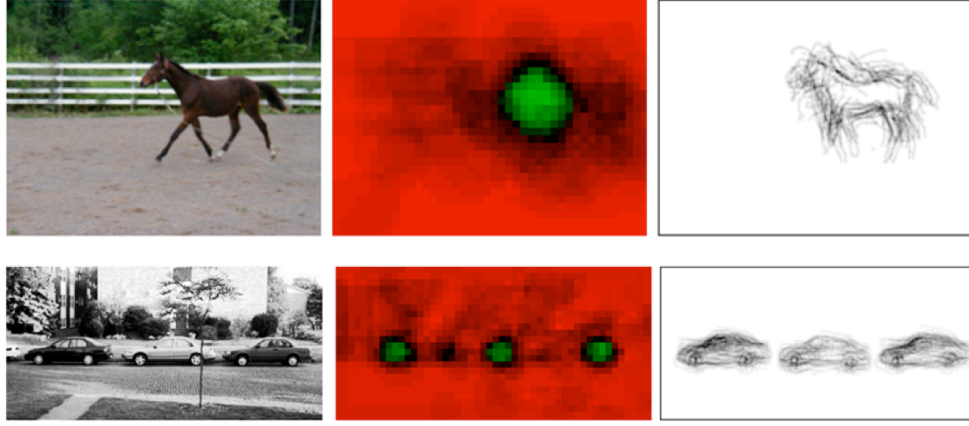
Figure 3.4: Sample results of the recognition method of Shotton et al. (image taken from [135]).

for the objects.

The learning phase of the object detection method suggested by Opelt et al. [105] can be divided into two parts. In the first part, a codebook of local contour fragments of a given class is constructed from a set of training images where every object is specified by a bounding box, and a set of labeled validation images where the object centers are given for the class images. First, the Canny edge detector is applied to the images to extract the edge information. Next, starting from a random seed along the extracted boundaries within the bounding boxes, the candidate local contour fragments of various sizes are extracted from the training images and associated with the corresponding object centers. For each candidate, a cost function is used to measure how well the local contour fragment matches with the validation set (using the chamfer distance) and how consistent the relative object center information is with respect to the true object centers given for class images in the validation set. Subsequent to this, the optimal ones are selected and grouped together using agglomerative clustering. Each cluster is described by its medoid (the local contour fragment whose average distance to all the fragments in the cluster is minimal) and an optimized codebook is built. In the final part of the learning phase, Adaboost is utilized to learn a strong detector by combining several weak detectors that are formed using the constructed codebook. Each weak detector is a composition of 2-3 local contour fragments and is activated when (1) the participating fragments are matched with the image edges, and (2) the hypotheses about the object center are consistent and in agreement with the true object centers that are specified for the class images. Once the strong detector is learned, the recognition phase consists of the following steps. First, the image edges of the novel
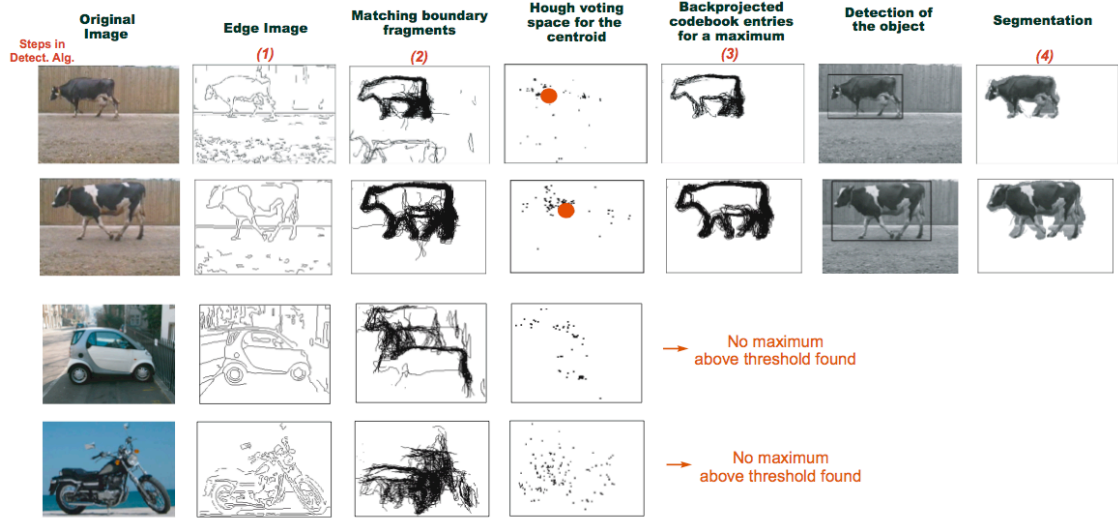
Figure 3.5: The overview of the contour-based recognition procedure proposed by Opelt et al. (image taken from [105]).

image are detected. Then, the local contours participating the weak detectors that form the strong detector are matched to this edge map. Each weak detector provides a probabilistic vote for the potential location of the object center. These votes are then collected by using a Generalized Hough Transform and the local maxima in the vote space are estimated using the Mean-Shift mode estimation algorithm as in the work of Leibe et al. [87, 88]. Every local maximum point above a certain threshold specifies an object hypothesis. Opelt et al. also showed that an approximate segmentation can be obtained by backprojecting the codebook entries that take part in these object hypotheses to the image plane. An overview of the joint recognition and segmentation process is presented in Figure 3.5 using some sample images.

3.2 Variational Formulations

In the following sections, we will briefly summarize important examples of the variational frameworks that integrates low-level image features with a priori information about the shape of an object of interest. As stated before, the main goal of these frameworks is to accurately delineate the actual boundaries of objects when corrupting influences due to missing regions, partial occlusion and noise appear in images. This is a very challenging task since the objects within a given class have a large variability in their shape as well. In general, the frameworks within this domain differ in terms of how they represent the prior shape knowledge, the boundary detection rule that they are based on, and how they extend

it to allow the influence of the shape prior.

3.2.1 The Work of Leventon et al. [89]

The work of Leventon et al. is one of the pioneering works that use prior shape knowledge within a variational formulation. It extends a level set implementation of the geodesic active contours model [35, 77] in such a way that the segmenting curve evolves according to two competing forces dictated by the image and the prior shape model, respectively. For this, Leventon et al. incorporated an additional prior shape term into the curve evolution equation. While the data-driven term locally attracts the curve towards the object boundary by taking the curvature and the image gradients into account, the additional term globally controls its shape.

In [89], the prior shape model is constructed by using a moderate set of training curves, which correspond to the various appearances of a shape of interest, as follows. First, each training curve is embedded as zero-level line of a higher dimensional surface by using signed distance function (Figure 3.6). Following that, a Principal Component Analysis (PCA) is performed on these distance surfaces. This process yields a linear low-dimensional space where a shape within the given class can be represented in terms of the mean distance surface and the coefficients of the estimated principal components (Figure 3.7). As reported in [89], in order to capture meaningful statistics (true variations in shape) within this space, it is important to align the training data before the analysis.

The level-set based segmentation approach uses an implicit contour representation where an evolving curve is represented with the zero-level line of a higher dimensional surface (Sec-
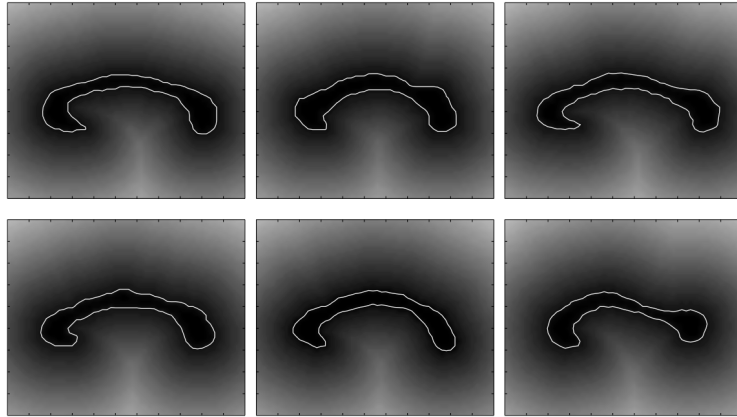
Figure 3.6: A prior shape model is constructed by utilizing distance surfaces with their zero-level lines representing the training curves (image taken from [89]).

44
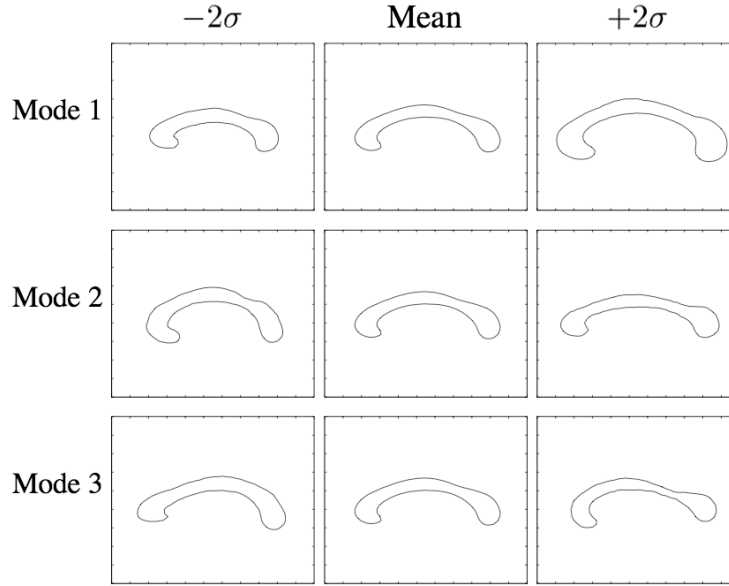
Figure 3.7: The mean shape and three principal components describing the shape variability (image taken from [89]).

tion 2.3.2). Partitioning an image into two regions (object vs. background) is performed by iterating between two processes. The first process is the estimation of the shape parameters (the coefficients of the principal components) and the rigid pose parameters of the most probable shape. The prior shape information and the image information are utilized together with the evolving shape (the current level set function) to determine a maximum a posteriori (MAP) estimate. The second process is the main curve evolution process, which is derived by adding a prior shape term to the evolution equation of geodesic active contours model. This additional term imposes prior knowledge about the shape of the object of interest and provides a global force that evolves the level set function towards the MAP estimate. As illustrated in [89], these alternating processes result in more accurate segmentation results since the shape of the segmenting curve is enforced by the prior shape model. The critical point is that this model describes only the global variations of a class of shapes.

3.2.2 The Work of Tsai et al. [148]

This study presents a segmentation method where the figure-ground separation is guided by a prior shape model. As in most of the works that employ prior shape knowledge, the segmentation method is composed of two main phases: a training phase where the prior shape model is constructed, and an actual segmentation phase where an image is partitioned into two as object vs. background by using the learned model. Figure 3.8 presents an
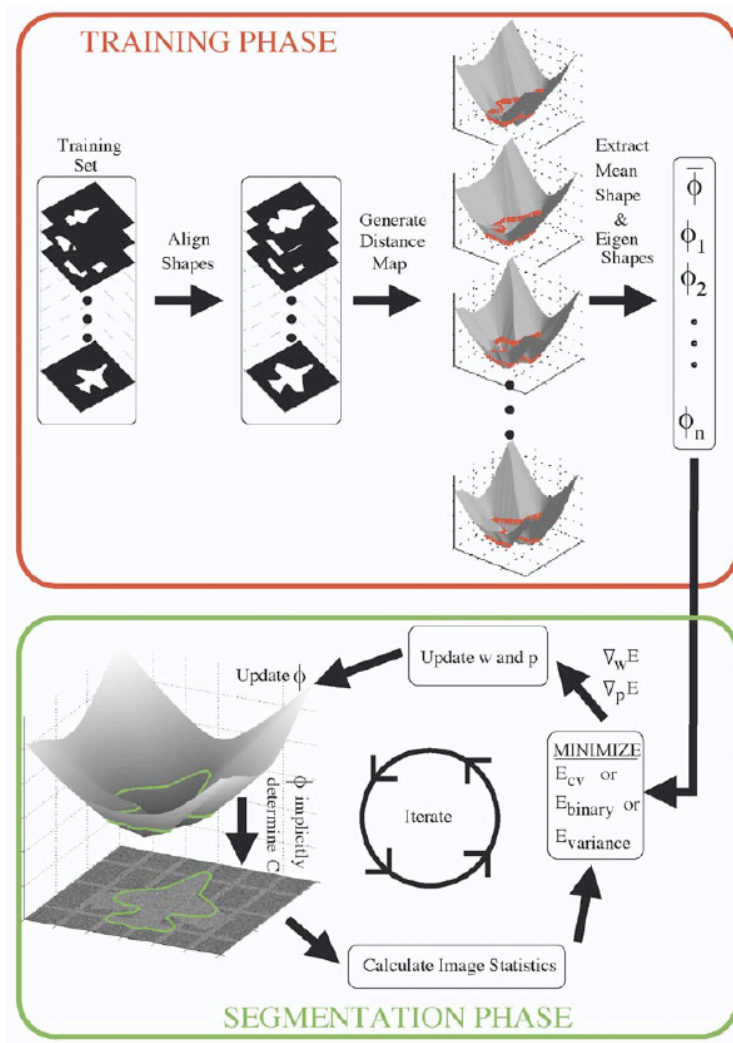
Figure 3.8: The overview of the segmentation method proposed by Tsai et al. (image taken from [148]).

overview of this method.

In their work, Tsai et al. represented the prior shape knowledge by adopting the linear shape model that we summarized in Section 3.2.1. Thus, the training phase of the method is nearly identical to the one proposed in [89]. The model is constructed by using a set of binary images that contains various appearances of a shape of interest. As discussed in Section 3.2.1, aligning the training data before the construction process is essential to obtain a more accurate shape model. For this, Tsai et al. suggested a shape alignment algorithm, formulated as an energy minimization problem, that removes the shape variabilities due to differences in pose. Once the training shapes are aligned using the corresponding gradient descent equations, a signed distance surface is estimated for each aligned shape. Then, the

mean distance surface and the principal components of variation modeling a linear shape space is constructed by applying PCA to the set of estimated distance surfaces.

The main difference between the segmentation method of Tsai et al. and the previously discussed method of Leventon et al. resides in how the constructed prior shape model is utilized in the segmentation phase. In the work of Leventon et al., segmentation is carried out by alternating between two processes. While first process employs the prior shape model to compute the most probable shape by taking the evolving segmenting curve and the image gradients into account, the second process uses this shape to impose an additional force to a level set implementation of the geodesic active contours model. On the other other hand, the method suggested by Tsai et al. follows a more unified approach. They embed the prior shape model within region-based active contours models [40, 158] by replacing the level set functions in the original functionals with the shape representation defined by the model. Thus, segmenting an image involves minimizing a single segmentation functional with respect to the shape and pose parameters by applying gradient descent. At each step of the gradient descent, these parameters are updated according to the current image statistics, yielding a new level set function. The zero-level curve of this level set function, in return, determines the segmenting curve, and thus the new image statistics for the next update. This cycle is repeated until convergence is achieved.

The experiments on both synthetic and medical data presented in [148] reveal promising segmentation results. The region-based active contours models, which are utilized as the boundary detection rule in the formulation, provide significant robustness against noise and missing regions. The linear shape model embedded within this segmentation model serves as a shape prior that controls the shape of the segmenting curve, resulting in accurate object boundaries. However, as in [89], the prior shape model describes only the global deformations within a shape class, and thus greatly affects the performance.

3.2.3 The Work of Rousson and Paragios [120]

Following an idea similar to the ones presented in [89, 148], Rousson and Paragios suggested a two-phase procedure for segmentation. While the first phase, which can be interpreted as a learning phase, constructs a prior shape model from a collection of sample shapes, the second phase carries out the actual segmentation process by using a unified segmentation model that combines the learned shape model with a data-driven segmentation model.

In [120], the shape prior is constructed as follows. First, the sample shapes within the training set are aligned by following a variational approach. The distance surface repre-

sentations of the training shapes are utilized within this alignment process. Subsequent to this, a pixel-wise probabilistic shape model, where each pixel is characterized by a Gaussian distribution, is estimated by taking aligned distance surfaces into consideration (Figure 3.9). This is performed by alternating between two processes until convergence is reached. While the first process computes the mean values and the variances in each pixel under some regularity conditions, the second process relaxes the estimated mean surface towards a signed distance surface. The main difference between the estimated probabilistic shape prior and the one utilized in [89, 148] lies in how the shape variations are described within these models. The linear shape formulation in [89, 148] models only the global variations in shapes by means of a Principal Component Analysis of the training set. On the other hand, the probabilistic shape model accounts for a pixel-wise confidence map for local variations. During segmentation, this confidence map locally determines in what extent the segmentation process respects the shape prior.

The actual segmentation process is formulated as an energy minimization problem that partitions an image into two regions as object vs. background. Considering the constructed shape prior along with a similarity transformation, Rousson and Paragios defined a unified energy which combines a level-set based segmentation model with a term imposing certain constraints on the shape of the object to be segmented. While the data-driven segmentation model supervises the segmentation according to the image data, the additional shape term
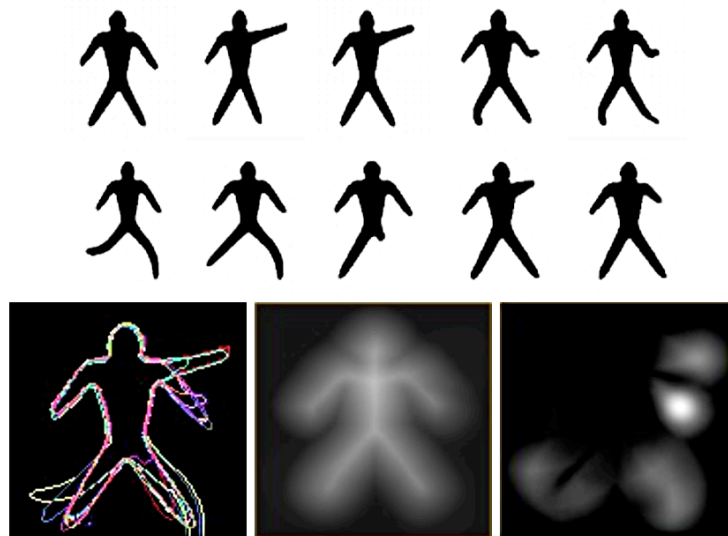
Figure 3.9: The aligned shapes are used to build a shape model that explains local variations in a class of shapes with a confidence map (image taken from [120]).

attracts the contour in a direction that results in a segmenting contour belonging to the predefined shape class, simultaneously registering the prior shape with the evolving level set. Recently, Rousson and Paragios suggested an alternative segmentation model [121] that uses the same probabilistic shape formulation, yet recovers the similarity transformation parameters in a more robust way by using the image data alone.

3.2.4   Diffusion Snakes - The Work of Cremers et al. [44]

The level-set based methods we discussed so far implicitly represent a curve as the zero-level line of a high-dimensional function. This method differs from these works in that it makes use of an explicit shape representation. The researchers utilized spline curves to learn a statistical shape prior and subsequently to partition an image with the prior shape knowledge. Spline-based formulations provide an explicit parameterization of curves as they define a curve by a collection of control points.

Learning a statistical shape prior involves extracting shape statistics from a set of training shapes. This is performed by fitting a spline curve to each training shape and aligning the collection of the resulting curves with respect to similarity transformations and cyclic permutation of the control points. Since the spline control points of the aligned curves lie in the same vector space, Cremers et al. used the distribution of these control points to model the prior shape knowledge. Assuming a Gaussian probability distribution, their statistical shape prior includes a mean shape (the mean control point vector) and a (regularized) covariance matrix.

Following that, Cremers et al. suggested a joint segmentation functional that integrates an implementation of the Mumford-Shah model, which represents the segmenting edge set as a closed spline curve, with a shape energy in the form of a Mahalanobis distance that penalizes the deviations from the derived shape prior. Segmenting an image involves minimizing the joint energy with respect to two variables: the control points of the segmenting curve and the smooth approximation of the original image. The steady-state solutions are obtained by alternating between the corresponding gradient descent equations. Mainly, the control points evolve under an image-driven force that attracts the curve towards the object boundaries and a shape-driven force that controls the shape of the segmenting curve. This process also provides a closed-form, parameter-free solution for the similarity transformation parameters. The name of the framework, *Diffusion Snakes*, comes from the simultaneously performed diffusion process that estimates a smooth approximation of the original image.

Figure 3.10 illustrates a sample segmentation result obtained by the reviewed spline-
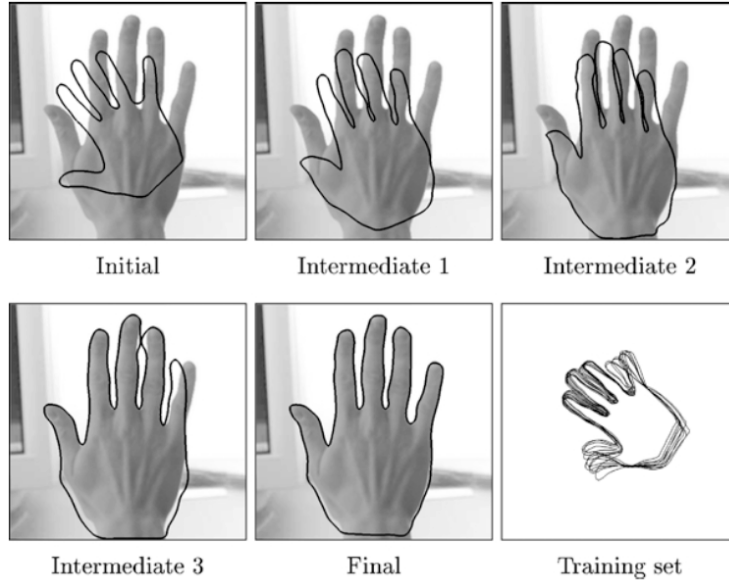
Figure 3.10: The shape statistics are extracted from a collection of sample shapes and integrated with a spline-based Mumford-Shah segmentation model (image taken from [44]).

based formulation. In general, the framework accurately delineates object boundaries even when corrupting influences due to occlusion, clutter and noise appear in images. However, as revealed in [43], when the training shapes representing the prior knowledge are visually different from each other, the resulting statistical shape prior fails to capture complex shape deformations. Taking this into account, in [43], Cremers et al. suggested to use nonlinear shape statistics within the Diffusion Snakes formulation, producing an improvement in segmentation results.

3.2.5 The Works of Riklin-Raviv et al. [115, 116]

The prior-based segmentation approaches proposed by Riklin-Raviv et al. differ from the studies we reviewed so far in that they ignore the shape variability. The researchers specifically aimed at segmenting out of an object of interest from a given image by taking a perspective [116] or projective [115] transformation model into account (Figure 3.11). The shape priors in these studies are composed of a single reference shape, and thus they do not need an explicit learning phase.

Both works formulate the prior-based segmentation process as a functional minimization via which a segmenting contour and a set of transformation parameters are to be recovered simultaneously. The general form of the functionals is nearly equivalent. Each functional is

(a)   (b)
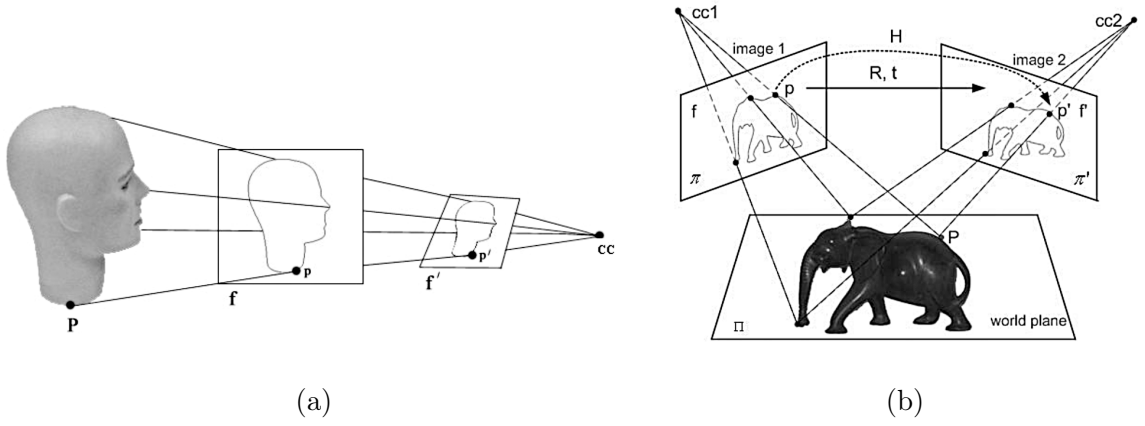
Figure 3.11: The methods proposed by Riklin-Raviv et al. use a single reference shape to segment an object of interest by considering transformation models based on (a) perspective transformation (image taken from [116]), (b) projective transformation (image taken from [115]).

an extended Chan-Vese functional (Section 2.3.2) with an additional prior shape term which mainly measure the difference in the areas of the evolving object region and the transformed prior. While the Chan-Vese functional leads to an image force that attracts the zero-level curve of the evolving level set function to the object boundaries, the prior shape term enforces the zero-level curve towards the given prior shape under transformation. The key difference between the proposed functionals lies in the transformation model utilized in the prior shape term.

The shape term proposed in [116] uses a single (segmented) image of an object of interest to partition the novel images of that object taken from the same camera center (Figure 3.11(a)). Riklin-Raviv et al. suggested a quite elegant formulation that utilizes a generalized cone representation, which is constructed from the reference image, as the shape prior. As illustrated in Figure 3.12, the cross sections of the generalized cone describe the appearances of a given object under the perspective transformation, and Riklin-Raviv et al. used the transformed versions of this cone model to constrain the evolving level set. In [115], Riklin-Raviv et al. extended this work by considering a projective transformation model (Figure 3.11(b)) so that it could partition a new image of a given object as seen through a different view point. For this, they use an explicit formulation of the homography matrix for the transformation model and a binary image (function) to represent the prior shape, resulting in more complex gradient descent equations.

Besides the methods discussed in this section, Riklin-Raviv et al. also suggested two
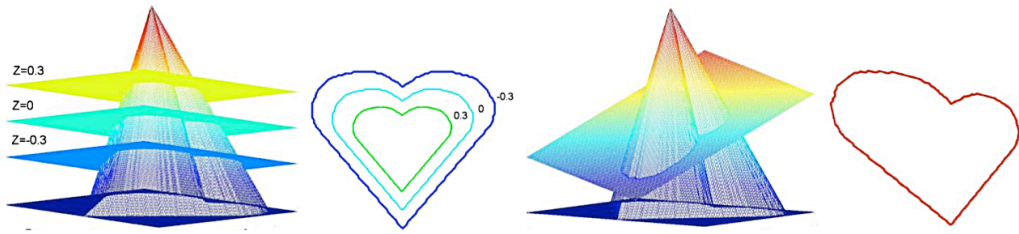
Figure 3.12: The framework that takes perspective transformation into consideration uses a generalized cone representation to describe possible deformations of the prior shape (image taken from [116]).

noteworthy segmentation methods in [114, 117]. The importance of these studies is that they do not utilize prior shape models that are explicitly constructed from a single reference shape or a set of sample shapes of an object. Instead, in these works, the prior shape knowledge is emerged from the input image(s) as the segmentation process carries on. The so-called *mutual segmentation* framework [117] simultaneously segments two ambiguous images of the same object taken from different viewpoints by taking a projective transform model into account. The evolving object contour in each image functions as a shape prior for the other image. The segmentation method proposed in [114], on the other hand, aims at extracting boundaries of symmetric objects in the presence of perspective distortion. In particular, it uses the symmetrical counterpart of the evolving object contour as a dynamic shape prior to supervise the segmentation.

3.2.6   The Work of Hong et al. [73]

Most of the studies we reviewed so far utilize representations that globally model the prior shape. In [44, 89, 148], the authors learned global variations of a shape of interest through a set of training shapes which corresponds to the various appearances of an object. Riklin-Raviv et al. ignored the shape variability at all and considered only a perspective [116] or projective [115] transformation model. The probabilistic framework of Rousson and Paragios [120], on the other hand, provides relaxation for the global representations as it uses a pixel-wise confidence map to locally determine in what extent the segmentation process respects the shape prior. In [73], Hong et al. suggested an alternative formulation which takes a single reference shape into account within a local deformation model that constrains the shape of the evolving contour. It is mainly obtained by extending the Chan-Vese segmentation functional with an additional shape energy in the same way proposed in the works
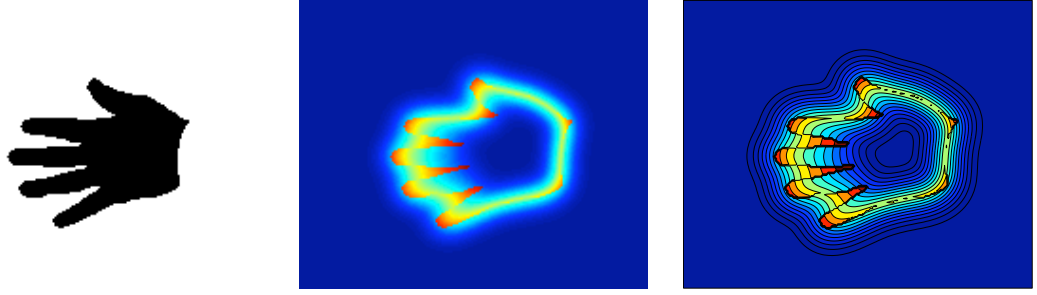
Figure 3.13: The shape representation using integral kernels encodes the local properties of shape geometry. This makes obtaining meaningful correspondences between two shapes possible.

of Riklin-Raviv et al. [116, 115]. The shape energy basically acts as a matching energy functional that determines the correspondences between the evolving object contour and the given prior shape. In their work, these correspondences are implicitly described by a local displacement map that transforms the shape prior into the evolving contour.

Particularly, the proposed shape energy consists of a *data* term and a *regularization* term. The data term measures the similarity between the evolving shape and the transformed prior, providing a driving force for the registration. On the other hand, the regularization term in the energy forces the local deformation map to be smooth. The key idea of the work of Hong et al. is to use a shape representation, the so-called *integral kernels*, in the data term (Figure 3.13). This representation assigns each point a distinct value based on the local geometry, and thus produces the estimated correspondences to be meaningful. In the regularization term, Hong et al. employed the *linear elasticity* model proposed in [28] to obtain a smooth deformation map. Since this model does not handle large deformations, they performed a global registration before applying the local deformation model.

During the segmentation process, the Chan-Vese functional provides an image force that pulls the zero-level curve of the evolving level set function towards the object boundaries. However, the shape of the evolving object contour is restricted by the additional prior shape term that simultaneously registers the shape prior to the image data. The main advantage of using a local deformation model over using a representation that models global shape variations is that it provides a more flexible way to manage the shape variations within an object class as it does not demand a large number of sample shapes. However, one of the drawbacks of this approach is that it does not always retain the specified shape. To overcome this weakness, in [73], Hong et al. suggested an extra regularization term that preserves the

local shape features. In Chapter 5, we utilize a similar local deformation model within our prior-based segmentation work that uses TSP surfaces [143] as shape priors. Chapter 5 also includes a detailed review of the representation based on integral kernels, as it shares some similar characteristics with the TSP surface.

3.3 Discussion

Segmenting out objects in natural images is a difficult task since low-level image cues are generally highly ambiguous. The diffusion models and variational regularization methods reviewed in Chapter 2 utilize very generic priors to resolve these ambiguities, but do usually produce poor segmentation results. On the other hand, relatively recent works make use of high-level a priori information about the shape of an object of interest. All of the studies investigated in this chapter are considered important examples in this domain. We mainly categorized these studies into two as probabilistic and variational formulations, and reviewed them in separate sections. The characteristics common to all these formulations is that they are all composed of two main phases: a training (learning) phase and an actual segmentation phase. While a prior shape model representing a priori knowledge about the target object is built in the training phase, the following segmentation phase partitions an image into two as object vs. background by taking the learned model into account.

The studies that are formulated within a probabilistic framework are mostly proposed for detecting objects from images, yet the detection process leads to simultaneous segmentation of objects. Due to this motive, these methods are not really concerned about whether they extract accurate object boundaries or not, and most of the time, the obtained segmentation results are not fully consistent with the image data. The training phase uses a moderate set of training images, which is composed of class and non-class images, to learn some kind of shape and appearance models by selecting the most informative local image patches or contours that are more common in class images compared to non-class images. As the objects within a specific class have a large variability in their shape and appearance, the success of these methods highly depends on the training set and the learning process. These works generally differ in terms of

1. how they represent the object-specific knowledge (utilizing local image regions with figure-ground labels in [22, 87] and local contour fragments in [105, 135]), and

2. how they use the probabilistic object models in joint recognition and segmentation

54

process.

The variational formulations reviewed in this chapter are global formulations in that they are obtained by integrating the prior shape information directly into generic variational segmentation frameworks. They were mainly originated from the applications in medical imaging where extracting specific parts of the biological structures such as brain has a critical importance. Thus, their main goal is to accurately delineate the actual boundaries of objects in the presence of corrupting influences due to missing regions, partial occlusion and noise. The segmentation results (extracted object boundaries) obtained by the variational formulations are generally compatible with the image data as the formulations unify high-level shape knowledge and the low-level image cues. In general, these formulations can be analyzed in terms of

1. how they represent the prior shape knowledge (PCA of signed distance representations in [89, 148], Gaussian probability distribution on control points in [44], a probability density function of signed distance representations in [120], and a single reference shape within a deformation model in [73, 115, 116]),

2. the boundary detection rule that they are based on (level-set based geodesic active contours model in [89], spline-based model in [44], and region-based active contours models in [73, 115, 116, 120, 148]), and

3. how they extend it to allow the influence of the shape prior.

To summarize, the probabilistic formulations offer solutions for segmentation that are more general than the ones suggested by the variational methods. They represent object-specific shape knowledge by models based on local visual cues such as image patches or contour fragments. However, during joint recognition and segmentation of objects, since they use these models in a top-down manner, they usually do not identify correct object boundaries. In this respect, the variational frameworks are more advantageous as they integrate low-level image features with high-level contextual information within a unified formulation.

CHAPTER 4

CONTEXT-INFLUENCED FILTERING

The prevalent view in computer vision since Marr [96] is that the early vision is a data-driven and bottom-up process. Over the years, a variety of techniques has been developed using this paradigm. Most of these techniques cannot handle the ambiguities in low-level visual cues due to the unidirectional information flow. In Chapter 2, it was argued that unified approaches have advantages over the bottom-up methods. The main reason for this is that they offer formulations which perform smoothing and edge detection in a simultaneous way. These formulations are mainly based on generic priors that bias the solutions towards the expected results. For example, the Mumford-Shah model contains regularization terms that force segmented images to be piecewise smooth while preventing them to be split into many regions.

The variational segmentation frameworks reviewed in the previous chapter improve outcomes of the unified methods by incorporating prior shape knowledge in computations. In this chapter, we go in a different direction and address the influence of context in image smoothing by developing a regularization framework based on the Ambrosio-Tortorelli approximation [3] of the Mumford-Shah segmentation functional [100]. More specifically, we embed contextual knowledge extracted from local image regions in the Euler-Lagrange equations of the AT model.

We propose various contextual feedback measures[1], each for a seemingly different visual task, and incorporate them in the regularization process. The smoothing examples that utilize these measures demonstrate the potential of the method under challenging cases such as difficult noise types, non-uniform contrast, existence of multi-scale patterns and textures.

---

[1]Most of the work presented in Section 4.2.1 and Section 4.2.2 has been in collaboration with Aysun Sancar Yılmaz and was previously presented in the First International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2007) [57] as an early version of this work. Full version is published in Journal of Mathematical Imaging and Vision (JMIV) [56].

4.1 Image Smoothing by Considering Local Context

Contextual influences have been explored within both computer vision and neuroscience community, e.g. [1, 13, 14, 15, 48, 49, 50, 68, 70, 75, 81, 84, 92, 101, 108, 112, 113, 125, 137, 138, 140, 142, 146, 147, 156, 159]. In the work proposed here, context is a collection of nearby local neighborhoods or nearby local image patches, and the contextual measures are aggregate quantities that are computed from these collections. Our work differs from the above efforts in the sense that we use the AT model (a PDE framework) as the computational platform.

As previously discussed in Section 2.3.1, the formulation of Ambrosio and Tortorelli suggests a numerical solution to the MS model by replacing the unknown segmentation contour in the model with a continuous function called the *edge strength function*. An important aspect of this formulation is that it can be interpreted as a biased nonlinear diffusion model with an additional constraint which enforces smoothness of the boundaries. Although the AT model gives fairly good segmentations for a limited set of images, as illustrated in Figure 4.1, the reconstruction results are affected by a heavy noise or a texture. We argue that some of the challenging cases for the AT model can be resolved by considering the context.

In the following, we give a general review of methods that modified the AT model. After that, in Section 4.1.2, we introduce our regularization framework which considers the context. Section 4.1.2 also presents the details of our numerical implementation, and discusses how the proposed regularization can be extended to multi-channel images.

4.1.1 Related Work

Over the years, a variety of modifications to the AT model were proposed. As reviewed in Section 2.5, Shah [131] considered replacing the quadratic cost functions in both the data fidelity and the regularizer with L^1-functions. Consequently, shocks form in the smoothed image and the object boundaries are recovered as actual discontinuities without being affected by the noise or the texture much.

An extreme modification to the AT approximation has been proposed by Tari et al. [143] by using sufficiently large values of ρ and interpreting v as a smoothed distance function. The resulting distance function, which is called the *TSP surface*, allows to perform skeletal analysis in grayscale images. The properties of this surface will be examined in detail in the following chapters.
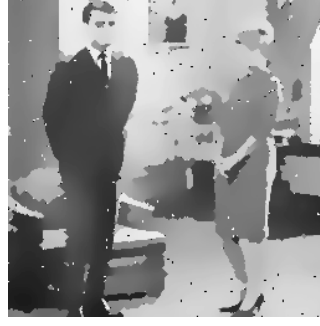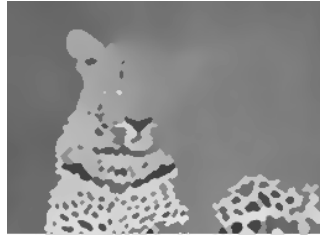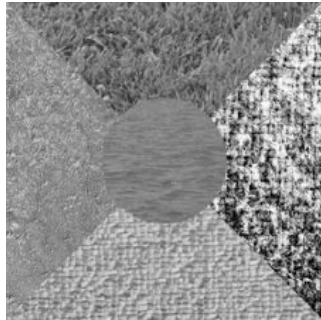
Figure 4.1: Images which cannot be handled by the AT model. (a)-(b) *couple* image corrupted with 5% salt and pepper noise and its reconstruction using AT. (c)-(d) A noisy test image –70% of the pixels are degraded with uniform noise– and its reconstruction using AT. (e)-(f) An animal image and its reconstruction using AT. (g)-(h) A texture mosaic image and its reconstruction using AT.

In [11, 12], Bar et al. presented a promising approach for image deblurring in the presence of impulse noise. The success of their method stems mostly from the use of robust data fidelity measure obtained by replacing the L^2-norm with the L^1. The use of non-smooth cost functions such as the L^1-norm in the data fidelity term in order to deal with outliers is well motivated both theoretically and experimentally (e.g. [30, 45, 46, 104, 131]).

Teboul et al. [144] presented a modification to the AT model by replacing the quadratic cost $|\nabla v|^2$ with an L^1-cost which leads to singular diffusivity. The major drawback of singular diffusivities is the numerical difficulty [33]. The cost function choice in [144] also leads to directional smoothing. As explored by Weickert [153] (Section 2.6), directional smoothing can offer significant feature preserving capabilities. However, the models get complicated and the numerics is not as simple as in the case of isotropic diffusion. Our regularization approach is similar to that of Weickert [153] in the sense that we consider gray values in a local image region to steer diffusion. However, our contextual feedback measures are more general and higher level. Moreover, we remain in the isotropic setting.

In the following, we propose a simple yet effective modification to the AT model that converts it to a regularizer with much better feature preserving capabilities without resorting to additional energies, prior models and difficult to minimize cost functions.

4.1.2   New Regularization Method

Key to our approach is the link between regularization and diffusion filters [19, 99, 103, 124, 130]. After casting the AT approximation as a biased diffusion filter, we use local image measures to steer the diffusion. The local measures are computed from a collection of local neighborhoods that are not necessarily centered on the pixel to be regularized. These measures are referred to as *contextual feedback measures* or simply *feedback measures* throughout the thesis.

We utilize spatially adaptive approach in [47] which has also been investigated under the computational frameworks of image algebra [118] and mathematical morphology [127]. Our approach can also be related to the one in Gilboa et al. [63] in the sense that both of the methods consider contextual knowledge apart from the other variational denoising frameworks which are based on the local features of the image such as derivatives. While [63] is based on a non-local convex regularization term that depends on the contextual similarity suggested in [31], we remain in the local and simple framework of the AT approximation.

In the proposed method, as in the AT approximation, there are two coupled processes: the image process u and the edge process v. The edge process v evolves according to the

same dynamics defined by the AT model:

$$\frac{\partial v}{\partial t} = \nabla^2 v - \frac{2\alpha|\nabla u|^2 v}{\rho} - \frac{(v-1)}{\rho^2}; \quad \left.\frac{\partial v}{\partial n}\right|_{\partial\Omega} = 0 \ . \tag{4.1}$$

On the other hand, the image process u evolves according to

$$\frac{\partial u}{\partial t} = \nabla \cdot ((cv)^2 \nabla u) - \frac{\beta}{\alpha}(u - f); \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = 0 \ , \tag{4.2}$$

$$cv = \phi v + (1 - \phi)V \ .$$

In (4.2), V is either 0 or 1 decided by the specified feedback type, and ϕ denotes the corresponding contextual feedback measure which takes values in $[0, 1]$ depending on u and/or v in a collection of neighborhoods. Thus, ϕ is at a higher place in the visual hierarchy than u and v. Note that with $\phi = 1$ for all image points, the coupled system reduces to the gradient descent for the AT model. We discuss the role played by the multiplier c by considering an abstract view (Figure 4.2) of the discrete approximations of the pair of coupled PDEs which are given by (4.7) and (4.8) in the subsequent numerical implementation section.

For any image point i, the value of u_i^{k+1} is computed using u_i^k, the neighborhood of u_i^k, f_i, and the feedback which depends on ϕ_i. Notice that the direct connection from v to u in the AT model is replaced by an indirect path; first rising higher up in the visual hierarchy and then coming back to steer the diffusion of u.
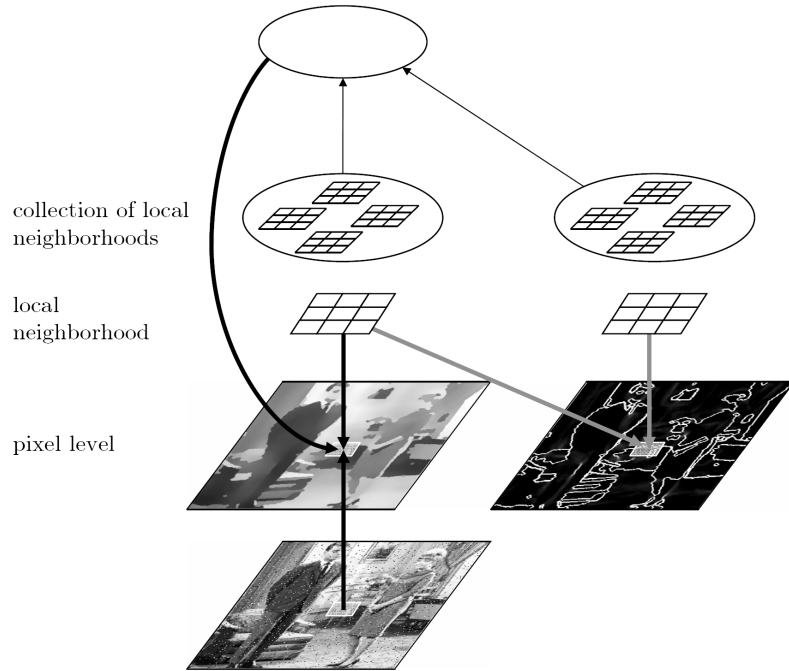
Figure 4.2: Proposed architecture. Aggregate quantities computed from a collection of neighborhoods are utilized to steer the evolution of the processes u and v. See text.

Suppose that we wish to eliminate an accidentally occurring feature, such as a high gradient due to noise. One can define ϕ_i such that it has low values for the accidental occurrences, and let it modulate the diffusivity in (4.2) as

$$(cv)_i^2 = (\phi_i v_i + (1 - \phi_i)\, 1)^2 \ . \tag{4.3}$$

Note that the lower the value of ϕ_i, the higher the diffusivity shifts towards the maximum value 1 to smooth out the feature. We call this type of modulation *negative feedback*.

Suppose that we wish to prevent an accidental elimination of a feature of interest, e.g. preserving a fine texture or encouraging edge formation. One can define ϕ_i such that it is low for meaningful occurrences, and let it modulate the diffusivity in (4.2) as

$$(cv)_i^2 = (\phi_i v_i + (1 - \phi_i)\, 0)^2 \ . \tag{4.4}$$

Thus, the lower the value of ϕ_i, the higher the diffusivity shifts towards the minimum value 0 to preserve the meaningful feature. We call this type of modulation *positive feedback*.

A computational advantage of the proposed approach is that both negative and positive feedbacks are implemented as a regularization. This provides robustness. Another source of robustness is the separate evolution of the edge indicator v_i from its modulated form $(cv)_i$, even though they eventually seem to converge to each other. The separate evolution of these variables prevents oscillations in the edge indicator, and allows us to effectively use both positive and negative feedback. Finally, notice that as indicated by (4.3) and (4.4), c_i is not explicitly computed. We may say that c_i as an indirect measure of feedback.

Numerical Implementation

The partial differential equations (4.2) and (4.1) can be simultaneously solved for u and v using standard numerical discretization techniques such as finite differences. The coupled system is first discretized with respect to spatial variables, yielding the following space-discrete system of equations:

$$
\begin{aligned}
\frac{du_{i,j}}{dt} &= (cv)_{i+\frac{1}{2},j}^2 \cdot (u_{i+1,j} - u_{i,j}) - (cv)_{i-\frac{1}{2},j}^2 \cdot (u_{i,j} - u_{i-1,j}) \\
&+ (cv)_{i,j+\frac{1}{2}}^2 \cdot (u_{i,j+1} - u_{i,j}) - (cv)_{i,j-\frac{1}{2}}^2 \cdot (u_{i,j} - u_{i,j-1}) \\
&- \frac{\beta}{\alpha}\,(u_{i,j} - f_{i,j}) \ , \\[2pt]
\frac{dv_{i,j}}{dt} &= v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j} \\
&- \frac{2\alpha|\nabla u_{i,j}|^2 v_{i,j}}{\rho} - \frac{(v_{i,j} - 1)}{\rho^2} \ .
\end{aligned}
$$

(4.5)

(4.6)

This discretization scheme requires the modified diffusivities represented by cv to be estimated at mid-pixel points. They are simply computed by taking averages over neighboring pixels as in the discretization of the AT model.

The time derivatives in (4.5) and (4.6) can be discretized using forward differences, where regularization terms and the bias terms on the right hand side of each equation are evaluated at times k and $k+1$, respectively. This yields an iterative scheme where homogeneous Neumann boundary condition is imposed along the image boundary

$$
\begin{aligned}
\frac{u_{i,j}^{k+1} - u_{i,j}^{k}}{\Delta t} &= \left((cv)_{i+\frac{1}{2},j}^{k}\right)^2 \cdot u_{i+1,j}^{k} + \left((cv)_{i-\frac{1}{2},j}^{k}\right)^2 \cdot u_{i-1,j}^{k} \\
&+ \left((cv)_{i,j+\frac{1}{2}}^{k}\right)^2 \cdot u_{i,j+1}^{k} + \left((cv)_{i,j-\frac{1}{2}}^{k}\right)^2 \cdot u_{i,j-1}^{k} \\
&- \left((cv)_{i+\frac{1}{2},j}^{k} + (cv)_{i-\frac{1}{2},j}^{k} + (cv)_{i,j+\frac{1}{2}}^{k} + (cv)_{i,j-\frac{1}{2}}^{k}\right) \cdot u_{i,j}^{k} \\
&- \frac{\beta}{\alpha}\left(u_{i,j}^{k+1} - f_{i,j}\right), \\
\frac{v_{i,j}^{k+1} - v_{i,j}^{k}}{\Delta t} &= v_{i+1,j}^{k} + v_{i-1,j}^{k} + v_{i,j+1}^{k} + v_{i,j-1}^{k} - 4v_{i,j}^{k} \\
&- \frac{\alpha\left(\left(u_{i+1,j}^{k} - u_{i-1,j}^{k}\right)^2 + \left(u_{i,j+1}^{k} - u_{i,j-1}^{k}\right)^2\right)v_{i,j}^{k+1}}{2\rho} \\
&- \frac{\left(v_{i,j}^{k+1} - 1\right)}{\rho^2}.
\end{aligned}
$$

(4.7)

(4.8)

with Δt denoting the time step.

Although the suggested scheme is neither fully explicit nor fully implicit, it still allows us to compute u^{k+1} and v^{k+1} by using forward recursion as in an explicit scheme. A numerical stopping criteria for the iterative scheme can be defined in the sense that the rate of change of u is less than a threshold. The corresponding procedure is summarized in Algorithm 3.

Extension to Multi-Channel Images

A straightforward way to smooth a multi-channel image is to diffuse each channel independently. For an m-channel image $\mathbf{f} = (f^1, \ldots, f^m)$, this yields the following system of coupled equations for $\mathbf{u} = (u^1, \ldots, u^m)$ and $\mathbf{v} = (v^1, \ldots, v^m)$:

$$
\frac{\partial u^i}{\partial t} = \nabla \cdot ((c^i v^i)^2 \nabla u^i) - \frac{\beta}{\alpha}(u^i - f^i); \quad \left.\frac{\partial u^i}{\partial n}\right|_{\partial\Omega} = 0 \quad (i = 1, \ldots, m), \quad (4.9)
$$

$$
\frac{\partial v^i}{\partial t} = \nabla^2 v^i - \frac{2\alpha|\nabla u^i|^2 v^i}{\rho} - \frac{(v^i - 1)}{\rho^2}; \quad \left.\frac{\partial v^i}{\partial n}\right|_{\partial\Omega} = 0 \quad (i = 1, \ldots, m). \quad (4.10)
$$

Algorithm 3 The iterative scheme for the context-influenced smoothing method

---

1: Initialize the variables with $u^0 = f$, $v^0 = \frac{1}{1+2\alpha\rho|\nabla u^0|^2}$

2: **for** $k = 0$ to $kmax$ **do**

3:    Estimate the contextual feedback measure ϕ

4:    Compute the modified diffusivity $(cv)^k$ by using (4.3) or (4.4)

5:    Solve (4.7) for u^{k+1} using $\lfloor \sqrt{2\alpha/\beta} \rfloor$ steps

6:    **if** $|u^{k+1} - u^k| < \epsilon |u^k|$ **then**

7:        stop iterations

8:    **end if**

9:    Solve (4.8) for v^{k+1}

10: **end for**

We remark that different edge locations will be created in different channels. To overcome this inconsistency, we can diffuse each channel with a common edge indicator function as proposed in [110] using the following set of coupled PDEs [2]:

$$\frac{\partial u^i}{\partial t} = \nabla \cdot ((c^i v)^2 \nabla u^i) - \frac{\beta}{\alpha}(u^i - f^i); \quad \frac{\partial u^i}{\partial n}\Big|_{\partial\Omega} = 0 \qquad (i = 1, \ldots, m) , \qquad (4.11)$$

$$\frac{\partial v}{\partial t} = \nabla^2 v - \sum_{i=1}^{m} \left(\frac{2\alpha|\nabla u^i|^2 v}{\rho} \right) - \frac{(v-1)}{\rho^2}; \quad \frac{\partial v}{\partial n}\Big|_{\partial\Omega} = 0 . \qquad (4.12)$$

For the multi-channel case, the contextual feedback measures may be estimated in various ways. First, each ϕ^i and consequentially $(c^i v)$ may be computed directly from the ith channel of the image. Second, a common feedback measure may be estimated from a weighted average of the multi-channel data. For color images, the average readily corresponds to an intensity image. A common feedback measure may also be estimated via a statistical measure extracted from channels. When the latter strategy is adopted in the experiments, the median of the individually computed measures is used for the contextual feedback measures that will be discussed in Sections 4.2.1 and 4.2.4. On the other hand, for the feedback measure introduced in Section 4.2.3, the measure is estimated by considering the distribution of contextual similarities computed from vector-valued image patches.

---

[2]An alternative color image segmentation method within the AT regularization framework is proposed in [29] by modeling images as manifolds.

4.2 Contextual Feedback Measures

In the following subsections, we present different choices for the contextual feedback measure which are respectively based on directional consistency of the edges (ϕ^{dc}), edge continuity (ϕ^h), texture edges (ϕ^{te}), and local scale (ϕ^{ls}). The measures presented in Section 4.2.1 and Section 4.2.3 are to be interpreted as negative feedback measures whereas the ones presented in Section 4.2.2 and Section 4.2.4 are to be interpreted as positive feedback measures.

Since the feedback measure induces a multiplier for the diffusivity function, we can integrate different contextual feedback measures by simply taking the multiplier c as the product of individual c values, and consider a coalition during the ongoing filtering process.

4.2.1 Directional Consistency: ϕ^{dc}

Segmentation, when defined as a sequential bottom-up process, is composed of the following three steps: smoothing, edge detection and edge linking. The purpose of the edge linking step is to force global consistency to locally detected edges in order to form a coherent edge set. In this step, the edge pixels that are detected based on the magnitude of image gradients are linked to give a connected edge set if their gradient directions are in agreement. The unlinked edge pixels are discarded. Interestingly, this last step is what the MS model or its AT approximation lacks. We induce such an effect in our diffusion model by increasing the relative persistence of the edge pixels which are consistent with their neighbors by increasing the diffusivity at the inconsistent ones. We consider a feedback measure ϕ_i^{dc} such that $\phi_i^{dc} \to 1$ on the preferred configurations and $\phi_i^{dc} \to 0$ on the incoherent ones, and let c_i has the following form:

$$c_i = 1 + (1 - \phi_i^{dc})\frac{1 - v_i}{v_i} \ . \tag{4.13}$$

First, notice that c_i increases in proportion to the image gradient $|\nabla u_i|$ which is proportional to $\frac{1 - v_i}{v_i}$ (refer to (2.26) in Section 2.3.1). Second, notice that the overall diffusivity coefficient $(cv)_i^2$ can be estimated as follows, without explicitly computing the variable c_i:

$$(cv)_i^2 = \left(\phi_i^{dc} v_i + \left(1 - \phi_i^{dc}\right) 1\right)^2 \ . \tag{4.14}$$

The value of the diffusivity is bounded by 1. It attains its maximum as $\phi_i^{dc} \to 0$ and decays to a value determined by the edge indicator v_i as $\phi_i^{dc} \to 1$. In the experiments, we compute ϕ_i^{dc} as

$$\phi_i^{dc} = exp\left(\varepsilon^{dc}\left[\frac{1}{|\eta_s|}\sum_{j \in \eta_s} \nabla u_i \cdot \nabla u_j - 1\right]\right) \ . \tag{4.15}$$

64

In (4.15), η_s represents the neighborhood of pixel i defined as $\pm s$ pixels along the orthogonal edge direction $\nabla u_i^\perp$. The parameter ε^{dc} is a scalar which determines the decay rate of the ϕ_i^{dc} function. Particularly, the definition of neighborhood utilized here is adaptive in the sense defined in [47]. If the neighboring pixels are coherent (having similar edge directions), then the average angle between ∇u_i and ∇u_j's is close to 0, yielding $\phi_i^{dc} \to 1$. Therefore, when ϕ_i^{dc} is high, occurrence of a low-level feature (high gradient) is not accidental and we can rely on the edgeness measure. On the other hand, as ϕ_i^{dc} approaches to 0, an occurrence of the same low level feature is very likely to be accidental and we warp the diffusivity value towards 1.

The importance of directional consistency is best observed if the input image contains impulse noise. Our method can be related to the ones in [37, 41] in the sense that some regularization is performed on the noisy image points. While [37, 41] use hard decisions on noisy image points and smooth out only these points, our method uses a soft decision strategy by means of the local region measure ϕ^{dc} to define the new diffusivity $(cv)^2$ for each image point.

4.2.2   Edge Continuity: ϕ^h

In boundary detection methods, the principle of edge continuity is used to eliminate streaking or breaking up of an edge contour due to noise or non-uniform contrast. This procedure is commonly referred as *hysteresis* due to successful application of threshold retardation in the Canny edge detector [34]. In our diffusion model, we lower the diffusivity at pixels that correspond to the broken parts of boundary segments to favor edge formation. There may be various choices for the selection of the feedback measure ϕ^h. The important point is to decrease the modified diffusivity $(cv)_i^2$ if the neighboring site encourages formation of an edge, i.e. having a low v value. As shown in Section 2.3.1, there is a reciprocal relationship between v_i and $|\nabla u_i|$ given by $v_i \approx \frac{1}{1+2\alpha\rho|\nabla u_i|^2}$. A decreasing diffusivity can be achieved by increasing the estimate of the image gradient which is used in estimating the diffusivity. Therefore, a natural choice is to add an offset $h_i \in [0,1]$ indicating a support in favor of edge formation to the gradient term in the diffusivity estimate:

$$
\begin{aligned}
(cv)_i^2 &= \left(\frac{1}{1 + h_i + 2\alpha\rho|\nabla u_i|^2} \right)^2 \\
&= \left(\frac{1}{1 + 2\alpha\rho(|\nabla u_i|^2 + h_i^*)} \right)^2 .
\end{aligned}
\tag{4.16}
$$

We can rewrite

$$(cv)_i^2 = \left(\phi_i^h v_i + \left(1 - \phi_i^h\right) 0\right)^2 \tag{4.17}$$

by letting

$$\phi_i^h = \frac{1}{1 + h_i v_i} \ . \tag{4.18}$$

There is a subtle difference between ϕ^{dc} discussed in the previous subsection and ϕ^h. Whereas ϕ^{dc} is high for the non-accidental occurrences of certain low-level features, e.g. image gradients, ϕ^h is high for the accidental occurrences of the same feature. As ϕ^h approaches to 0, we can rely on a non-accidental occurrence of the feature of interest and warp our diffusivity towards 0. Therefore, the modified edge indicator cv is a linear combination of v and the maximum edgeness value which is 0.

In the discrete implementation of (4.2), diffusivities are estimated at mid-grid points. Hence, h_i should be computed as a support from a suitably chosen neighbor. For example, a modified diffusivity $(cv)_{i+0.5,j}^2$ at a mid point between (i, j) and $(i + 1, j)$ may receive a support in the form of either $(1 - v_{i+0.5,j-1})$ or $(1 - v_{i+0.5,j+1})$. Notice that the lower the value of edge indicator at a neighboring site, the higher the support it provides.

Adding spatial organization to energies defining regularization with line process has been previously proposed by Black and Rangarajan [17]. In [19], Black et al. derived the necessary update equations. If we let $\frac{v_k^2 + v_l^2}{2}$ define a line process between site k and site l, then our development becomes equivalent to that of Black et al. Thus, solving new coupled equations are qualitatively equivalent to modifying the MS model with an additive term favoring unbroken contours as in Black and Rangarajan [17].

4.2.3   Texture Edges: ϕ^{te}

In the MS model, edgeness is measured by the image gradient. However in textured images, a large image gradient may be due to a texture within a region rather than from an edge. Consequently, the MS model or its AT approximation fail to capture object boundaries (Figure 4.1(e)-(h)). Our motivation in defining this feedback measure is to eliminate the texture and to preserve the structure. In this respect, our texture edges feedback strategy is related to the work of Aleman-Flores et al. [2]. Whereas they modify Perona-Malik to make the diffusivity a function of the texture gradient computed from the Gabor feature space, we let a measure computed from local patches modulate the diffusivity.

In order to capture the true object boundaries in a textured image, we need features that are at a higher level than the local derivatives. In the feed-forward step, we estimate the

contextual feedback measure ϕ^{te} using the probability map of the texture edges of the source image. Due to computational concerns, this probability map is extracted only once using a simple patch-based technique proposed by Wolf et al. [157]. The probability of an image point being near a texture edge is computed using a non-parametric test of distributions called Wilcoxon Mann-Whitney test.

For each point i, four distributions are sampled: D_i^{up}, D_i^{down}, D_i^{left} and D_i^{right}. These distributions correspond to the contextual similarities between the patch centered at the point i and the patches which are Δx pixels to the related direction. The similarities are simply computed by using the Euclidean distance between the $n \times n$ patches. If the point i lies on near a texture edge, then sampled distributions D_i^{up} and D_i^{down}, or D_i^{left} and D_i^{right} should be different from each other. This hypothesis is tested using Wilcoxon Mann-Whitney test. As in [157], the final probability is taken as the minimum of the p-value's p_i^1 and p_i^2 returned by the test for the distributions D_i^{up}-D_i^{down} and D_i^{left}-D_i^{right}. We use this probability value to define ϕ_i^{te} as

$$\phi_i^{te} = exp\left(-\varepsilon^{te}\left(min(p_i^1, p_i^2)\right)\right), \tag{4.19}$$

where ε^{te} is again a scalar parameter that determines the decay rate. For points that do not correspond to a texture edge, the estimated p-values are high making $\phi_i^{te} \to 0$ and $(cv)_i \to 1$.

This case is equivalent to the case discussed under directional consistency. Just like ϕ^{dc}, a low value of ϕ^{te} indicates an accidental occurrence of a feature of interest (i.e. high gradient). Therefore, we warp our diffusivity estimate towards the maximum value of 1 to diffuse more at the locations where the existence of a local feature is not supported by the context. The modulated diffusivity is given by

$$(cv)_i^2 = \left(\phi_i^{te} v_i + \left(1 - \phi_i^{te}\right) 1\right)^2. \tag{4.20}$$

4.2.4   Local Scale: ϕ^{ls}

The MS model decomposes an image into cartoon and noise components. During denoising, some important features of the image such as textures or fine details are also smoothed out as they are treated as noise by the model. This behavior is more apparent in the models with more robust norms, e.g. [122, 131]. In [64], a spatially varying fidelity term is proposed for the TV regularization (Section 2.4) to partly preserve local image structures. The term is based on local variance measures and determines the level of denoising.

Our framework can be also used to devise a texture preserving denoising model by lowering the diffusivity at the pixels inside textured regions. The contextual feedback measure ϕ^{ls}

determines whether a point lies on a textured region or not, and controls the level of smoothing accordingly.

We use robust statistics following [18] to determine the textured regions. For an image point i, the local scale is specified by

$$\sigma_i = median_{j \in \Omega_i}(|\nabla u_j - median_{j \in \Omega_i}(|\nabla u_j|)) \tag{4.21}$$

where higher values of σ_i means that the point i lies on a textured image region. In deriving σ_i, we use an $n \times n$ image patch around the image point i which is represented by Ω_i in (4.21). Considering σ_i, we define the new contextual feedback measure ϕ_i^{ls} as follows:

$$\phi_i^{ls} = exp\left(-\varepsilon^{ls} \ \sigma_i\right) \ . \tag{4.22}$$

For a pixel inside a textured region, the value of σ_i is high, making $\phi_i^{ls} \rightarrow 0$. Similar to ϕ_i^h, ϕ_i^{ls} is low for non-accidental occurrences of a high gradient. Therefore, we warp our diffusivity estimate towards the minimum value of 0 when ϕ_i^{ls} is low. That is, we compute the modulated edge indicator $(cv)_i$ as the convex combination of 0 and v_i

$$(cv)_i^2 = \left(\phi_i^{ls} v_i + \left(1 - \phi_i^{ls}\right) 0\right)^2 \ . \tag{4.23}$$

4.3   Experimental Results

In this section, we present experimental results on various synthetic and natural images. In order to demonstrate the performance of our regularization framework and to illustrate what it is gained by taking the context into account, we compare the results of our method with that of the AT model [3], Shah's modification [131] and the TV regularization [122]. In all experiments, we use a convergence tolerance of $\epsilon = 5 \times 10^{-4}$ unless stated otherwise. Due to computational concerns, the texture edges measure ϕ^{te} is extracted only once from the original image, and the local scale measure ϕ^{ls} is estimated at the 0th, 10th and 20th iterations.

The importance of the directional consistency of the edges is best observed if the image contains impulse noise. Thus, we first apply our regularization method to the noisy image given in Figure 4.3(a). Figure 4.3(b) and (c) illustrate smoothing results of the AT model for different values of smoothing radius, $\frac{\alpha}{\beta}$. The result in Figure 4.3(b) is obtained with $\alpha = 0.75$, $\beta = 0.005$, $\rho = 0.01$. When we increase the smoothing radius by choosing $\beta = 0.001$, diffusion is so strong that we even lose the upper body part of the woman. Yet, the noise is still present as seen in Figure 4.3(c). If we use a regularization term which

uses a contextual feedback by means of directional consistency of the edges, as discussed in Section 4.2.1, the image is denoised without blurring (Figure 4.3(d) and (e)). The perceptual difference between Figure 4.3(d) and (e) is in the sharpness level. The result in Figure 4.3(d) is obtained with the segmentation parameters specified as $\alpha = 1, \beta = 0.01, \rho = 0.01$ and the contextual feedback parameters $s = 2$ and $\varepsilon^{dc} = 0.25$ (these are the default values for the directional consistency parameters). For the result given in Figure 4.3(e), we use the same parameters except $\varepsilon^{dc} = 0.1$. The variable ε^{dc} determines the decay rate of the coherency function used in the segmentation process and specifies the level of sharpness. For large ε^{dc} value, the decay rate is high, and as a result the edges are smoothed out more depending on the contextual feedback. As observed, the resulting image is smoother. Conversely, for small ε^{dc} values, we get sharper results. The result given in Figure 4.3(e) is also comparable to that is obtained by Shah's modification presented in Figure 4.3(f) which uses the L^1-norm for both the data fidelity and the regularizer (a half-quadratic approximation of Shah's modified energy proposed in [76] is utilized). While both results are satisfactory, the reconstruction using the model due to Shah is blockier. The robust norms used attract the image towards the cartoon limit and catch unintuitive regions such as the one at the man's right shoulder and the ones on the floor. Better result can be obtained by *merely* replacing the data fidelity term in the AT model with the L^1-norm as in [30, 104].

In the AT model, increasing the value of α while keeping $\frac{\alpha}{\beta}$ fixed means decreasing the penalty of the length term, yielding more detailed reconstructions. Consequently, in Figure 4.4, the proposed modification is tested with again the image given in Figure 4.3(a), however, forcing the reconstruction to be more detailed by the proper choice of parameters. Figure 4.4(a) demonstrates the outcome of the proposed modification presented in the first experiment. The result displayed in Figure 4.4(c) is obtained by using the same parameters except $\alpha = 4$, $\beta = 0.04$. As the corresponding edge indicator functions respectively shown in Figure 4.4(b) and (d), even the detailed reconstruction with $\alpha = 4$ is noise free.

In the next example, Figure 4.5, we illustrate the effect of edge continuity (Section 4.2.2). For this, we use the sample image given in Figure 4.5(a), which is taken from [19]. The smoothing results are presented in Figure 4.5(b) and (c) together with the corresponding edge indicator functions. Figure 4.5(b) illustrates the outcome of the AT whereas Figure 4.5(c) illustrates the result obtained by considering edge continuity. Both results are obtained by using the same scale space parameters, $\alpha = 1$, $\beta = 0.01$, $\rho = 0.01$. As it can be clearly seen from the enlarged indicator functions given in Figure 4.5(d), the modified scheme eliminates broken contours.
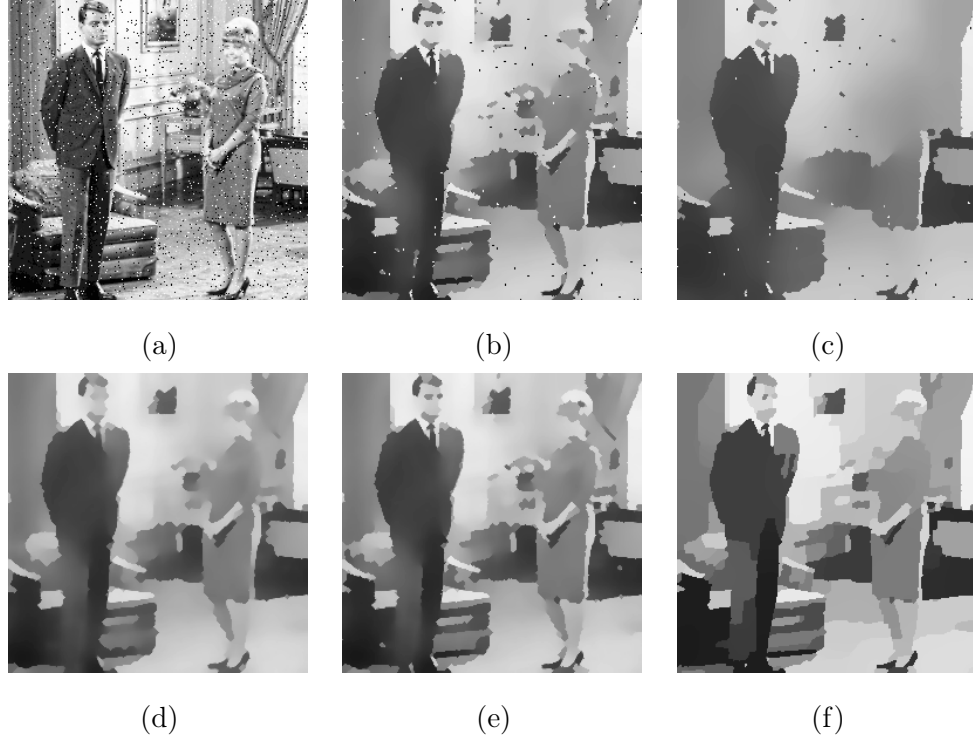
Figure 4.3: Considering directional consistency of the edges ϕ^{dc} as the contextual feedback measure eliminates impulse noise. (a) Source image corrupted with 5% salt and pepper noise. (b)-(c) Reconstructions using AT with two different smoothing radii. Notice that the noise is still present even when we lose the upper body part of the woman. (d)-(e) Reconstructions with directional consistency for two different sharpness levels. Notice that at comparable scales noise is completely eliminated. (f) L^1-reconstruction catches unintuitive regions.

In Figure 4.6, we demonstrate the results obtained with a regularization considering the coalition of edge continuity and directional consistency as the contextual feedback measure via the product of individual c functions. Recall that a combination can be achieved by multiplying individual c values. In the case of directional consistency, the multiplier c^{dc} is given by (4.13). In the case of edge continuity, the multiplier c^h is equal to ϕ^h as indicated by (4.17). The reconstruction results of the source image corrupted with 10% salt and pepper noise (Figure 4.6(a)) are given in Figure 4.6(b)-(d). They are obtained using the scale space parameters specified in the previous experiment. As it can be clearly seen, the result obtained with edge continuity, Figure 4.6(b), does contain noise. Figure 4.6(c) illustrates the result obtained with the modification that considers directional consistency. Finally, Figure 4.6(d) is the outcome of the combined framework that takes the coalition of two into account, which is not only noise free but also has stronger edges.

(a)   (b)

(c)   (d)
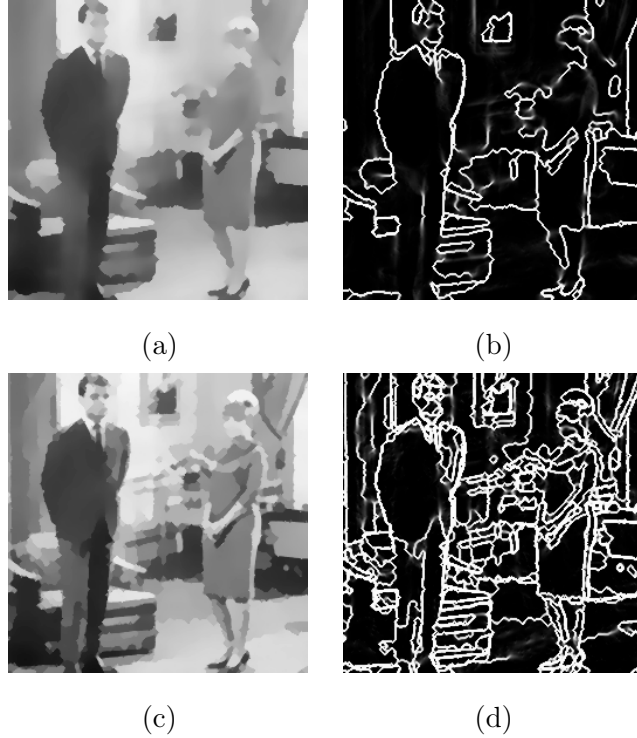
Figure 4.4: u and $1 - v$ functions computed with $\alpha = 1$, $\beta = 0.01$ and $\alpha = 4$, $\beta = 0.04$ respectively. Even in detailed reconstruction, directional consistency feedback is sufficient to remove noise completely.

In Figure 4.7, the framework considering the coalition of directional consistency of the edges and edge continuity is tested on a noisier image (Figure 4.7(a)). Figure 4.7(b)-(c) are the outcomes of the AT approximation with $\alpha = 2.5$, $\beta = 0.0001$, $\rho = 0.01$. Figure 4.7(b) is obtained with a convergence tolerance of $\epsilon = 4.8 \times 10^{-4}$ whereas Figure 4.7(c) is obtained with $\epsilon = 1.6 \times 10^{-4}$. The effect of decreasing β dramatically is to shut off the data fidelity term and to increase the relative importance of the length term. Notice that the relative increase (α/β) in the weight of the second term of the AT model is less than the relative increase $(1/\beta)$ in the weight of the third term. Our goal is to compare the effect of the length term, which is a part of the original model, to the effect of ϕ^{dc}. As it can be seen in Figure 4.7(c), eliminating the noise entirely results in smoothing out the whole rectangular region. This is mainly due to the perceived difference in the contrast. Figure 4.7(d)-(f) are the outcomes of the proposed framework that are produced by using the scale space parameters $\alpha = 1$, $\beta = 0.01$, $\rho = 0.01$. Figure 4.7(d) is obtained by using directional consistency with $\epsilon = 2.5 \times 10^{-4}$. Figure 4.7(e) is the outcome of the coalition of ϕ^{dc} and ϕ^h with $\epsilon = 1.25 \times 10^{-4}$. Notice that the results given in Figure 4.7(d) and (e) are visually

71
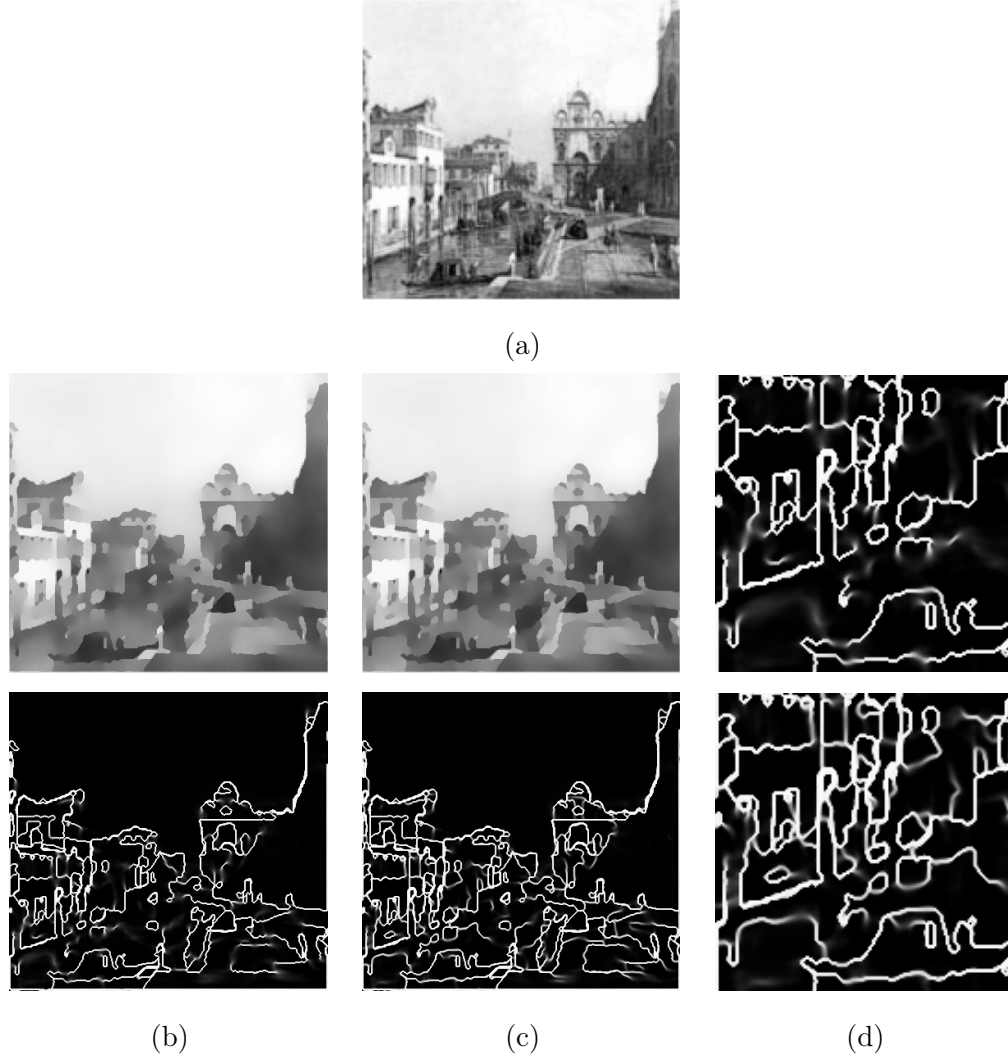
Figure 4.5: Considering edge continuity ϕ^h as the contextual feedback measure eliminates broken contours. (a) Source image. (b) Reconstruction using AT (u and $1-v$). (c) Reconstruction with the new method forcing edge continuity (u and $1-v$). (d) Details from the edge indicator functions given in (b) and (c), respectively.

similar. This is due to the fact that the contrast is almost constant in the image. Hence, broken lines do not occur. Figure 4.7(f) is obtained by using directional consistency with a larger neighborhood size, $s = 10$, and a convergence tolerance of $\epsilon = 1.6 \times 10^{-4}$. The perceptual difference between Figure 4.7(d) and (f) is clearly evident. Imposing coherency in a large neighborhood produces smoother object boundaries, however, at the expense of losing sharpness. Both reconstructions are qualitatively comparable to the one obtained by the TV regularization (Figure 4.7(g)). Note that the control parameters of the TV regularization were tuned to eliminate the noise and to capture the shape boundaries.

(a)   (b)

(c)   (d)
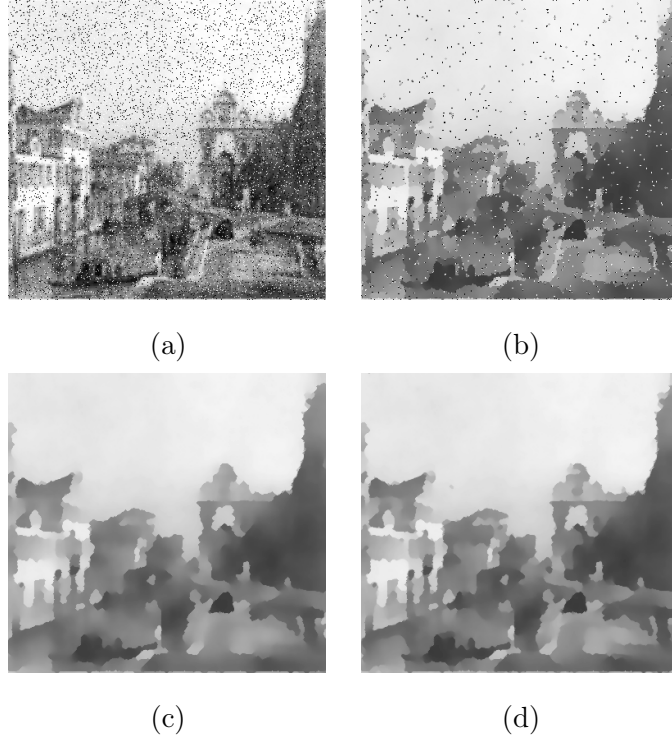
Figure 4.6: Considering a framework that uses the coalition of ϕ^{dc} and ϕ^h as the contextual feedback measure eliminates both noise and the broken contours. (a) Source image corrupted with 10% salt and pepper noise. (b) Reconstruction with edge continuity. (c) Reconstruction with directional consistency of the edges. (d) Reconstruction using both edge continuity and directional consistency of the edges.

Figure 4.8 - Figure 4.10 demonstrate the use of texture edges measure ϕ^{te} described in Section 4.2.3 for smoothing textured images. We observe that this measure, if used together with the other feedback measures, captures the actual object boundaries. Note that in our formulation, we do not decompose the original image into structure, texture and noise components as in [9, 150], we only retain the structure. We do so without using complicated norms.

Figure 4.8 and Figure 4.9 include two texture mosaic images [72] as the source images. Figure 4.8(b) and Figure 4.9(b) are the outcomes of our proposed method. Figure 4.8(b) is obtained by taking account of ϕ^{te} as the only contextual feedback measure with the segmentation parameters $\alpha = 100$, $\beta = 0.0075$, $\rho = 0.01$ and feedback parameters $n = 5$, $\Delta x = 15$ and $\varepsilon^{te} = 1000$ (these are the default values for the texture edges). On the other hand, Figure 4.9(b) is obtained by considering ϕ^{te} together with ϕ^{dc} and ϕ^h measures using the same parameters except $\alpha = 300$, $\beta = 0.005$, $\rho = 0.001$ and $\varepsilon^{te} = 100$. For each
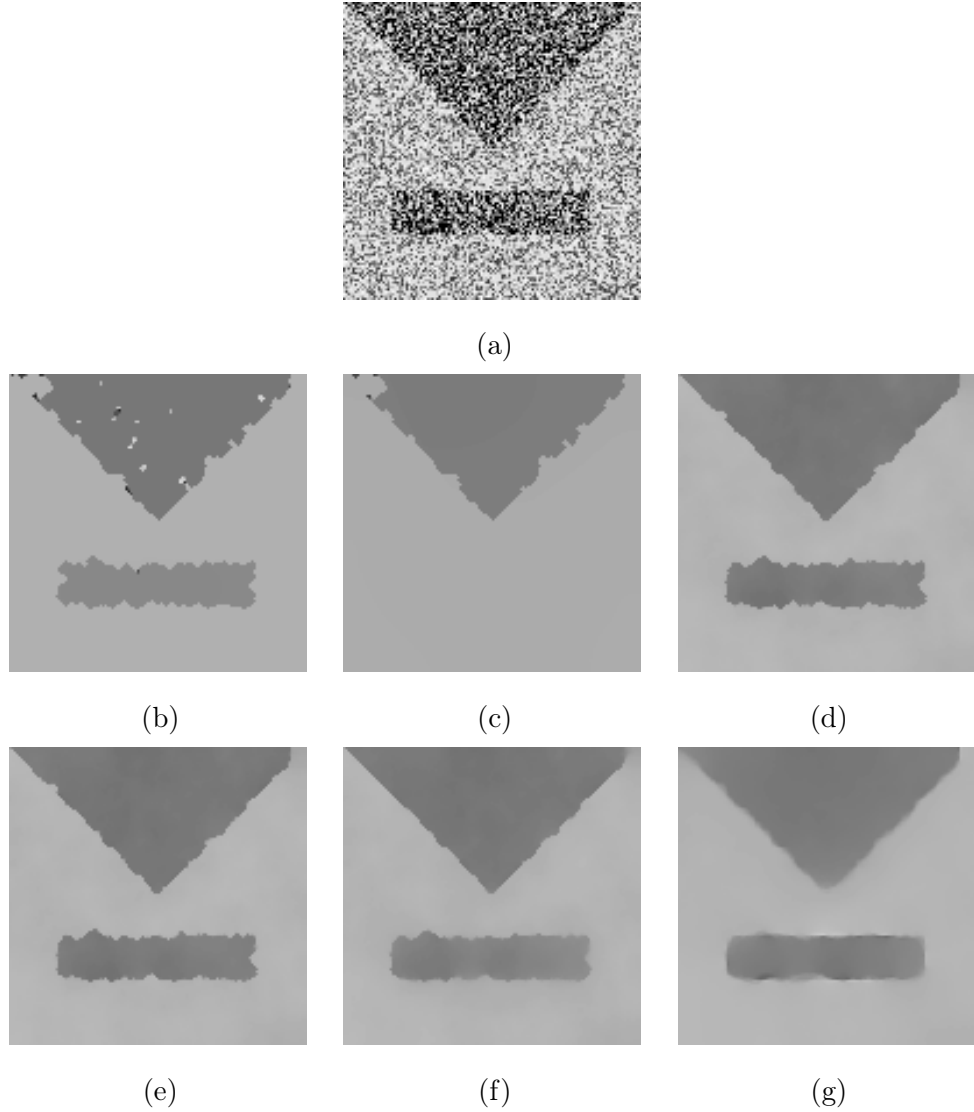
Figure 4.7: A difficult denoising case. (a) Source image, 70% of the pixels are degraded with uniform noise. (b)-(c) Reconstructions using AT with different choices of convergence tolerance. (d) Reconstruction considering directional consistency. (e) Reconstruction using the coalition of ϕ^{dc} and ϕ^h. (f) Reconstruction considering directional consistency with a larger neighborhood size. (g) Reconstruction using the TV regularization.

texture mosaic image, the texture edges measure and the resulting edge indicator function are also given. As it can be clearly seen in Figure 4.8(c)-(d) and Figure 4.9(c)-(d), while the estimated texture edges measures computed using the method proposed in [157] are noisy, the resulting edge indicator functions succeed to capture the actual object boundaries. The results of Shah's modification and the TV regularization for different scale parameters are also provided in Figure 4.8 and Figure 4.9 for comparison.
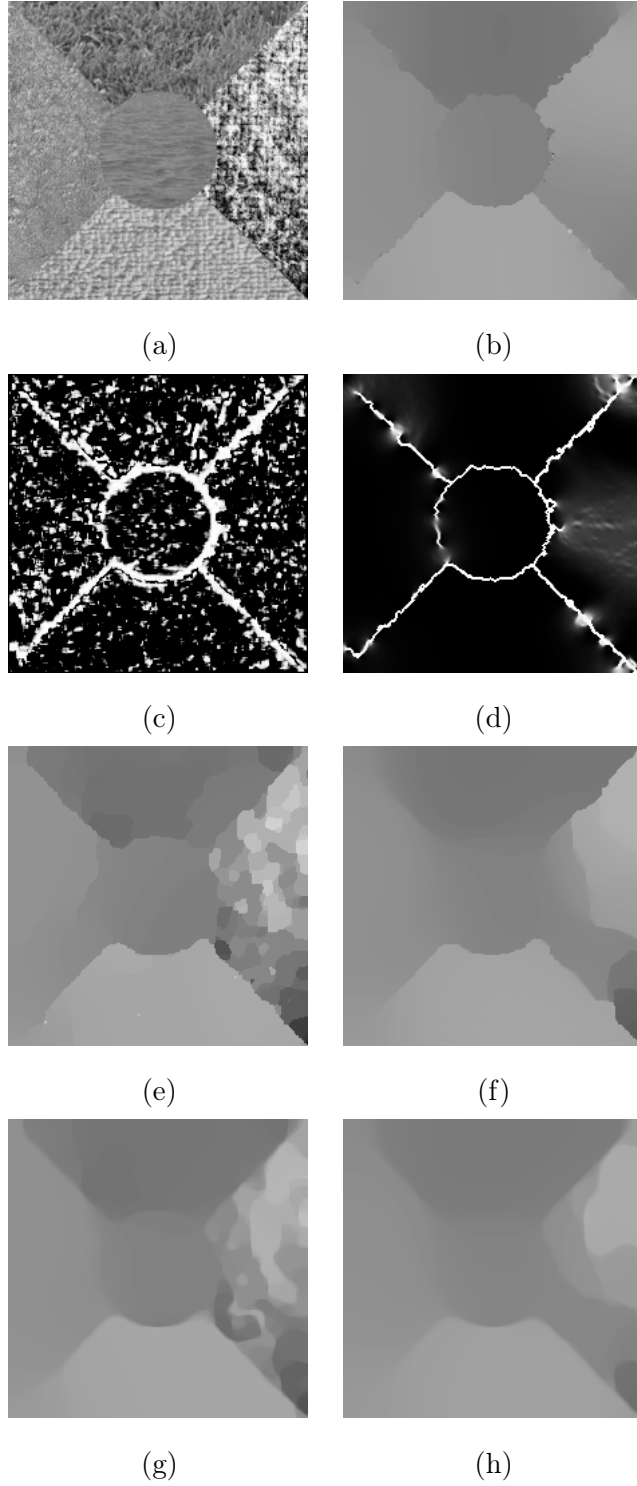
Figure 4.8: Actual region boundaries are captured by considering ϕ^{te} as the feedback measure. (a) Source image. (b) Reconstruction result with contextual feedback. (c) Texture edges measure ϕ^{te}. (d) Final edge indicator function. (e)-(f) Reconstructions using Shah's modification with different scale parameters. (g)-(h) Reconstructions using the TV regularization with different scale parameters.
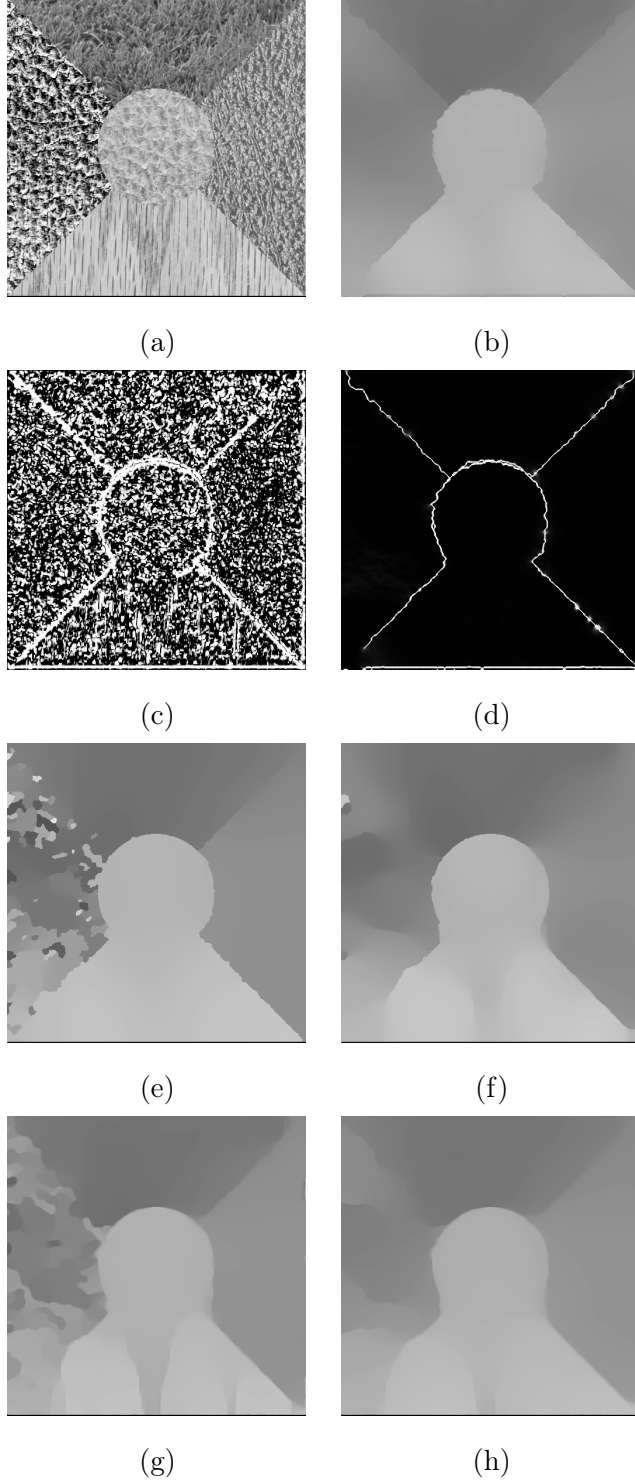
Figure 4.9: Coalition of three feedback measures. (a) Source image. (b) Reconstruction result with contextual feedback. (c) Textures edges measure ϕ^{te}. (d) Final edge indicator function. (e)-(f) Reconstructions using Shah's modification with different scale parameters. (g)-(h) Reconstructions using the TV regularization with different scale parameters.

In Figure 4.10, the framework taking the coalition of directional consistency, edge continuity and texture edges into account is tested with a natural animal image. The source image Figure 4.10(a) is taken from [60]. Figure 4.10(b) is the outcome of the combined framework with the segmentation parameters $\alpha = 200$, $\beta = 0.05$, $\rho = 0.001$ and the default contextual feedback parameters except $\varepsilon^{te} = 20$.

(a)

(b)

(c)

(d)

(e)

(f)
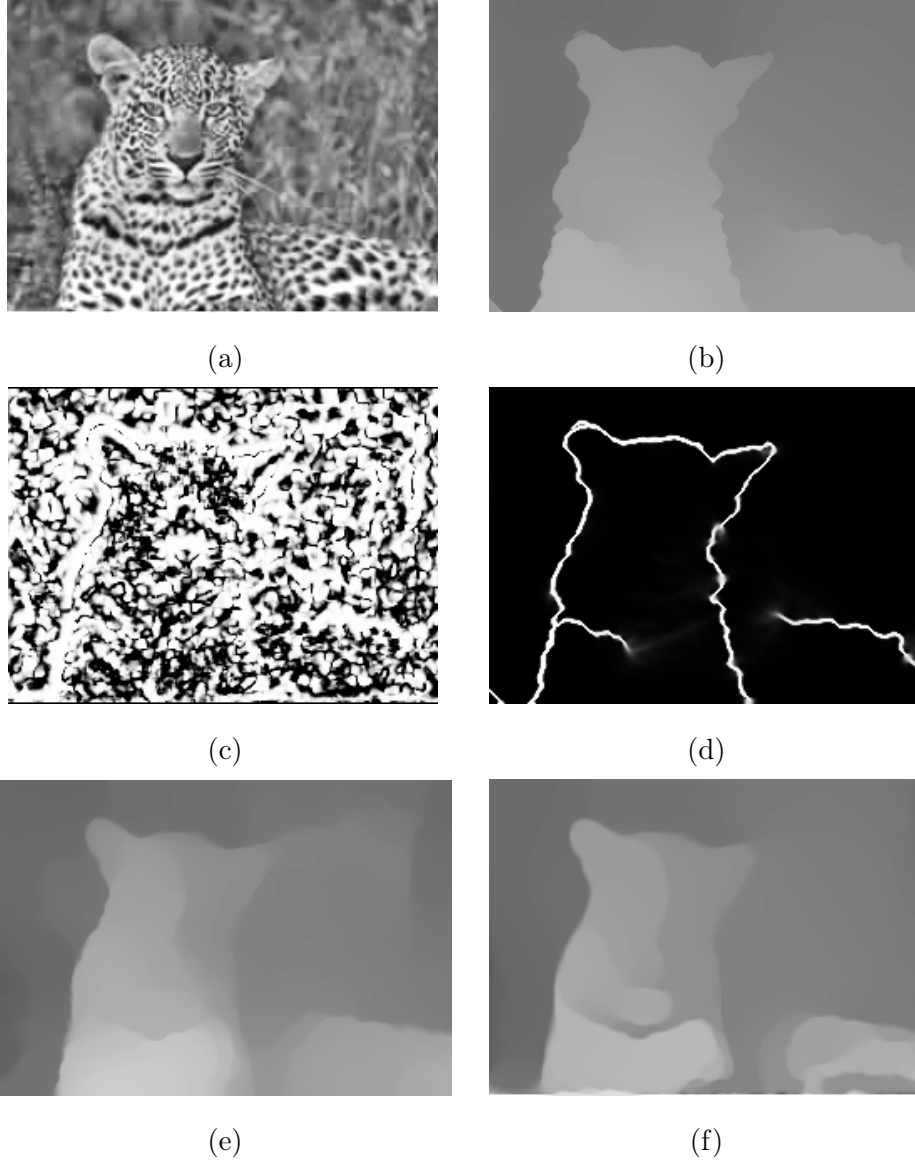
Figure 4.10: Coalition of three feedback measures. (a) Source image. (b) Reconstruction result with contextual feedback (c) Texture edges measure ϕ^{te}. (d) Final edge indicator function. (e) Reconstruction using Shah's modification. (f) Reconstruction using the TV regularization.

In Figure 4.7, we addressed the effect of decreasing β dramatically on the outcomes of the AT model. It mainly increases the relative importance of the length term and turns off the data fidelity term, leading to cartoon-like images. Next, by taking textured images into account, we compare the cartoon limit of the AT model with the regularization framework that utilizes coalition of directional consistency, edge continuity and texture edges.

Figure 4.11(a) illustrates a natural animal image. In Figure 4.11(b), we present the reconstruction of this image by using the cartoon limit of the AT model, with $\alpha = 5$, $\beta = 0.0001$, $\rho = 0.01$. Figure 4.11(c) is the outcome of the proposed regularization method obtained using $\alpha = 1$, $\beta = 0.01$, $\rho = 0.01$ and the default contextual feedback parameters except $\varepsilon^{te} = 2$. It is clearly seen from these images that both the cartoon model and our proposed framework give satisfactory results by eliminating fine texture details such as the black spots in the animal's skin and the bush in the background. It is essential to underline that our method does remove texture without forcing the result to be piecewise constant as the cartoon limit does.

In the next example, Figure 4.12, we consider three synthetic images, each having a similar pattern. While generating these images, we mainly swap the intensity values of the image regions. For the source images given in Figure 4.12(a) and (c), the cartoon limit smoothes out the circular patterns and gives rise to a dark object on a white background (Figure 4.12(b) and (d)). On the other hand, for the image having a gray background (Figure 4.12(e)), the same set of parameters which forces the AT model to cartoon limit, $\alpha = 0.1$, $\beta = 0.0005$, $\rho = 0.01$, yields a poor result as presented in Figure 4.12(f). Due to the contrast difference, elimination of the black circles leads to losing the object boundary. This is more visible in Figure 4.12(f) where the smoothing radius is decreased by choosing

| (a) | (b) | (c) |
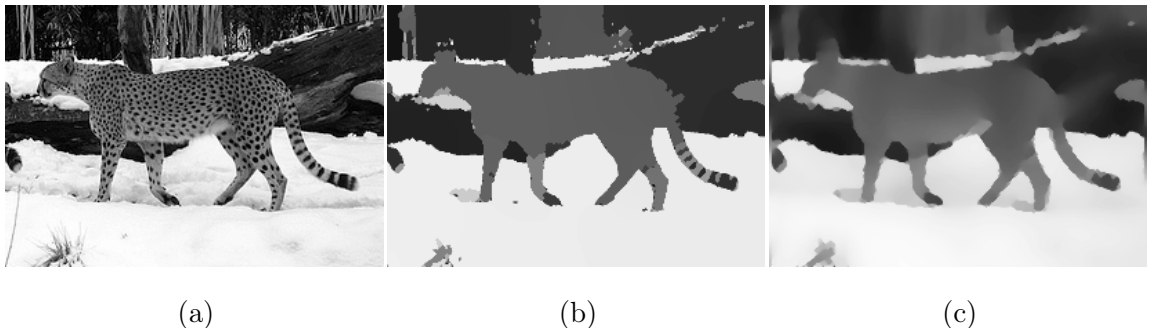
Figure 4.11: Comparison of the effect of cartoon limit of the AT model and that of our regularization framework on textured images. (a) Source image. (b) Reconstruction obtained by forcing the AT model to cartoon limit. (c) Reconstruction result with contextual feedback.

(a)   (b)   (c)   (d)

(e)   (f)   (g)   (h)
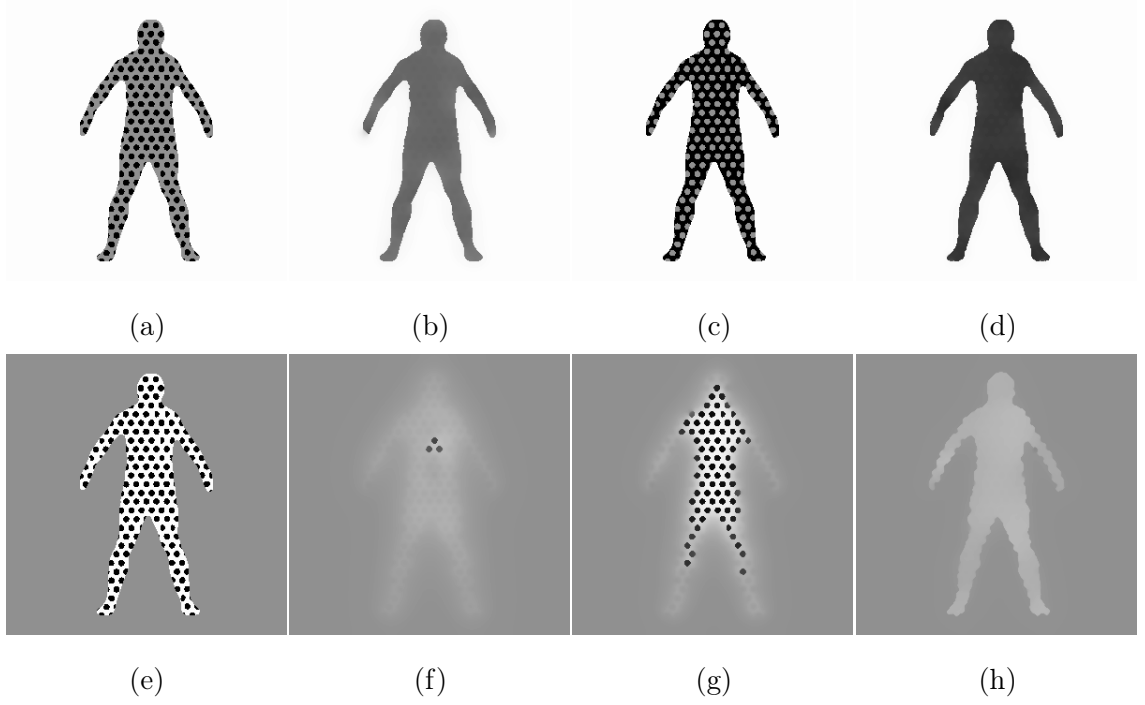
Figure 4.12: Comparison of the effect of cartoon limit of the AT model and that of our regularization framework on textured images. (a)-(d) Sample synthetic images containing textured regions and their reconstructions obtained by forcing the AT model to cartoon limit. Notice that the circular patterns are smoothed out. (e) Another synthetic textured image. (f)-(g) Reconstructions of (e) by forcing the AT model to cartoon limit. (h) Reconstruction of (e) with contextual feedback. Notice that while the cartoon limit of the AT model does not give any satisfactory result, our modification eliminates texture.

$\beta = 0.001$. Our regularization framework deals with this problem by taking the contextual feedback into account. As it can be clearly seen in Figure 4.12(f), the reconstruction result obtained by using $\alpha = 4$, $\beta = 0.005$, $\rho = 0.01$, $\Delta x = 10$, $\varepsilon^{te} = 10$ captures the actual object boundaries, and at the same time removes the circular patterns.

As discussed in Section 4.2.4, the MS model does not distinguish texture and fine details from noise. Hence, they are smoothed out during denoising. Figure 4.13 depicts how local scale measure affects the reconstruction results. For the image given in Figure 4.13(a), the outcomes of the AT model are presented in Figure 4.13(b)-(c). Figure 4.13(b) is obtained by using $\alpha = 10$, $\beta = 2.5$, $\rho = 0.001$ with a convergence tolerance of $\epsilon = 1 \times 10^{-4}$ while Figure 4.13(c) is obtained using $\alpha = 15$, $\beta = 1$, $\rho = 0.001$. Figure 4.13(d) is the outcome of the proposed method estimated using $\alpha = 10$, $\beta = 0.1$, $\rho = 0.001$, $\varepsilon^{ls} = 0.25$, $n = 25$ and the local scale measure depicted in Figure 4.13(e). The L^2-norms of the differences between

the original image and the reconstructions presented in Figure 4.13(b)-(d) are 6.0954×10^3, 7.6363×10^3 and 6.0735×10^3, respectively. Although the amount of signal eliminated from the source image in Figure 4.13(b) and (d) is approximately the same and the reconstructions are in a visually similar scale, Figure 4.13(b) is not entirely smooth and contains noisy pixels whereas the source image is denoised while preserving textures in Figure 4.13(d). Increasing the level of smoothing in the AT model results in noise-free results as presented in Figure 4.13(c), however the textured regions are also smoothed out during the process. Figure 4.14 illustrates the results of two more texture preserving denoising experiments. Figure 4.14(c) and (d) are obtained using the parameters $\alpha = 20$, $\beta = 0.1$, $\rho = 0.001$, $\varepsilon^{ls} = 0.125$, $n = 15$ and $\alpha = 4$, $\beta = 0.1$, $\rho = 0.001$, $\varepsilon^{ls} = 0.125$, $n = 25$ respectively. Figure 4.14(b) is taken from *Berkeley Segmentation Dataset* [97]. Note that the basic idea is just to lower the diffusion at textured locations. The measure ϕ^{ls} does not make a distinction between noise and texture.

(a)   (b)   (c)

(d)   (e)
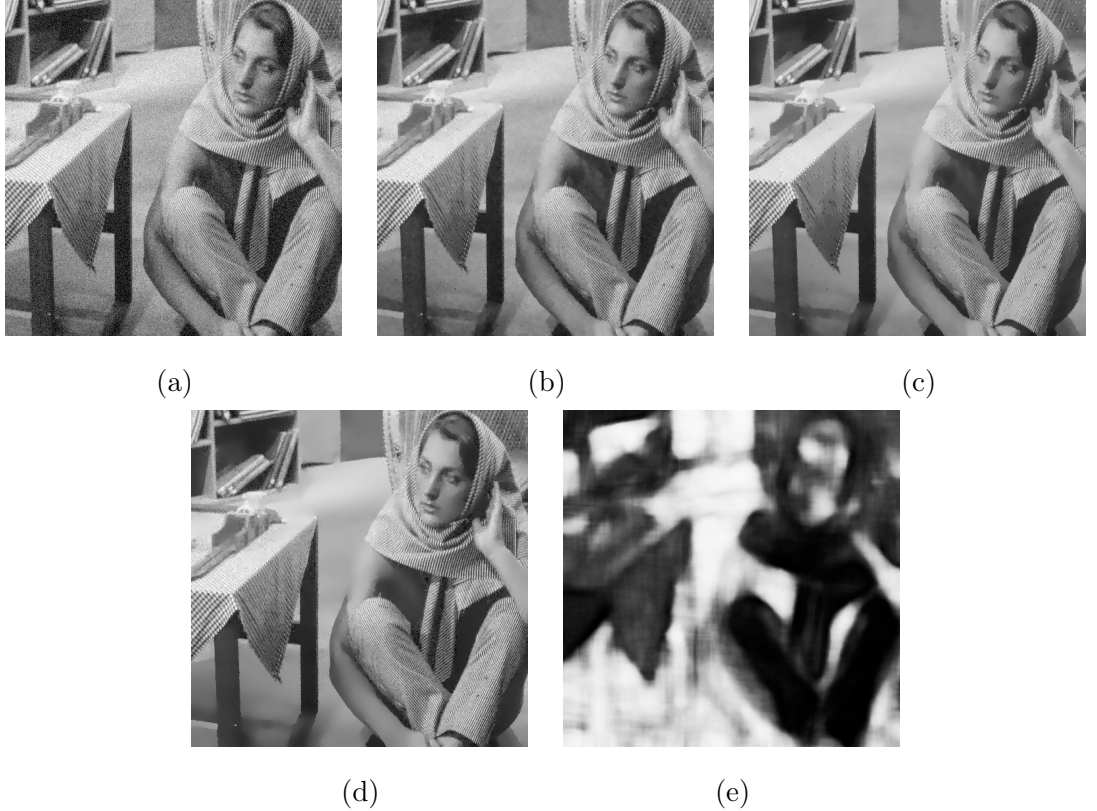
Figure 4.13: Texture preserving denoising with local scale measure ϕ^{ls}. (a) Source image. (b)-(c) Reconstructions using AT with different choices of smoothing levels. (d) Reconstruction with contextual feedback. Notice that texture in the fabric is preserved. (e) Local scale measure ϕ^{ls}.

(a)   (b)

(c)   (d)
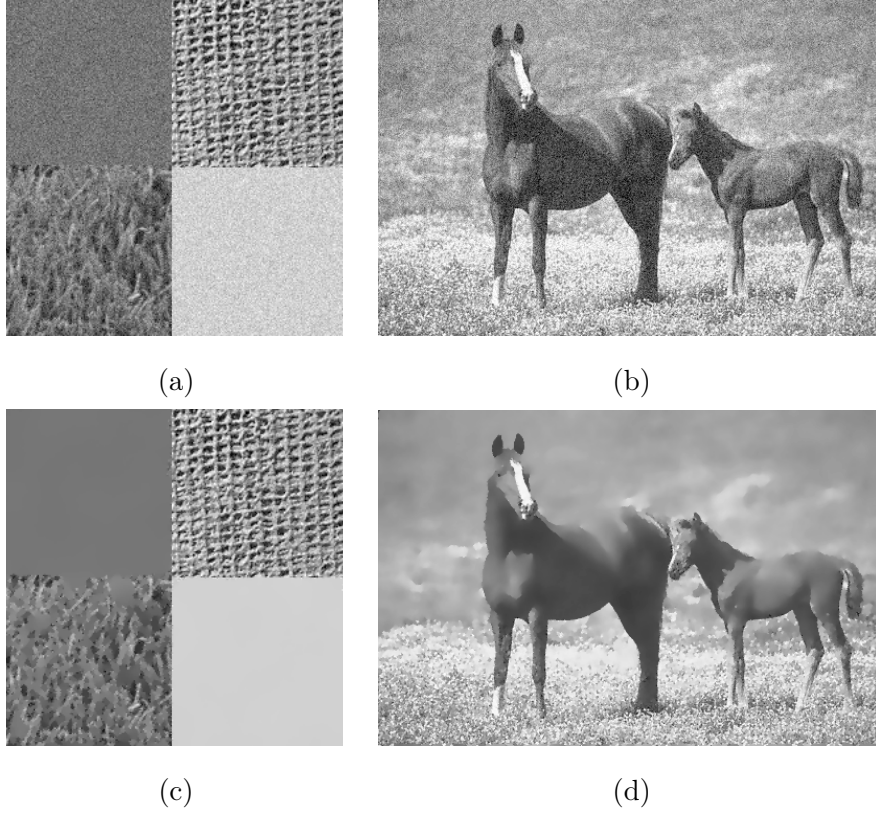
Figure 4.14: Some more experiments on texture preserving denoising. Notice the difference between non-textured and textured regions. (a)-(b) Source images. (c)-(d) Smoothed images.

Finally, we apply our method to color images as described at the end of Section 4.1.2. We use the RGB channels as our multi-channel data. For Figure 4.15, we have repeated the texture preserving denoising experiment using now colored versions of Figure 4.13(a) and Figure 4.14(b). The denoising results presented in Figure 4.15(b) are obtained by diffusing each channel separately with a common feedback measure estimated from the intensity image.

Figure 4.16 illustrates the results of using different strategies for color image smoothing of the image presented in Figure 4.16(a) (image taken from [97]). All smoothing results given in Figure 4.16(b)-(d) are obtained by considering the coalition of directional consistency, edge continuity and texture edges with the segmentation parameters $\alpha = 100$, $\beta = 0.1$, $\rho = 0.001$ and the default contextual feedback parameters except $\varepsilon_{dc} = 0.05$. Figure 4.16(b) is the result obtained by diffusing each channel separately by using the feedback measures estimated from corresponding multi-channel data. Figure 4.16(c) is obtained by again diffusing each channel separately but with a common feedback measure estimated from the intensity image. Finally, Figure 4.16(d) is the outcome of the color image smoothing with a common edge strength function and a feedback measure estimated from intensity image.
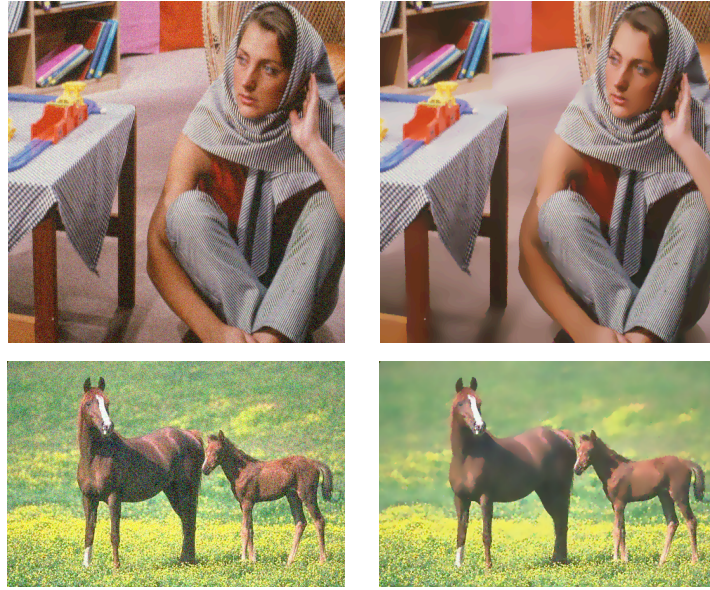
(a)   (b)

Figure 4.15: Texture preserving denoising on a color image. Notice the difference between non-textured and textured regions. (a) Source images. (b) Smoothed images.
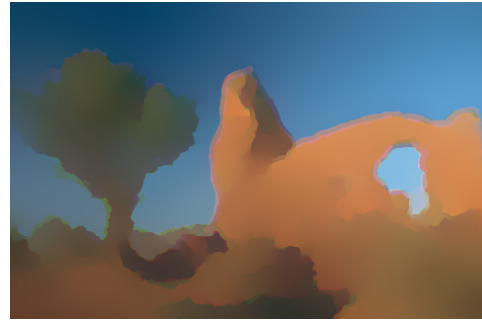
(a)   (b)
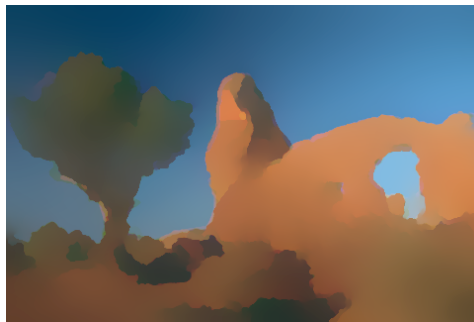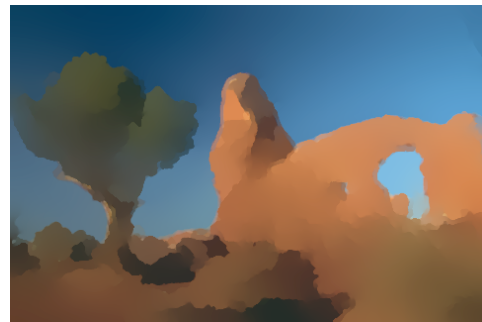
(c)   (d)

Figure 4.16: Color image smoothing. (a) Source Image. (b) Result of channel-by-channel smoothing. (c) Result of channel-by-channel smoothing with a common feedback measure estimated from intensity image. (d) Color image smoothing with a common edge strength function and a feedback measure estimated from intensity image.

In Figure 4.17, we demonstrate a case where using contextual feedback measures estimated from the intensity image yields an inaccurate smoothing result. The reconstructions of the color image shown in Figure 4.17(a), which is reproduced from [39], are given in Figure 4.17(c) and (e) together with the corresponding edge indicator functions provided in Figure 4.17(d) and (f). All of the smoothing results are obtained by considering the coalition of directional consistency, edge continuity and texture edges with the segmentation parameters $\alpha = 100$, $\beta = 0.1$, $\rho = 0.001$ and the default contextual feedback parameters except $\varepsilon^{dc} = 0.05$. Figure 4.17(c) is the outcome of the color image smoothing with a common edge strength function and a feedback measure estimated from intensity image. Since transforming the color image into intensity image makes the upper and the lower objects disappear (Figure 4.17(b)), these objects are smoothed out during processing and the reconstruction fails to capture the actual object boundaries. On the other hand, when the feedback measures are estimated by considering each channel simultaneously, we get a fairly good result (Figure 4.17(e)).

(a)   (b)

(c)   (d)   (e)   (f)
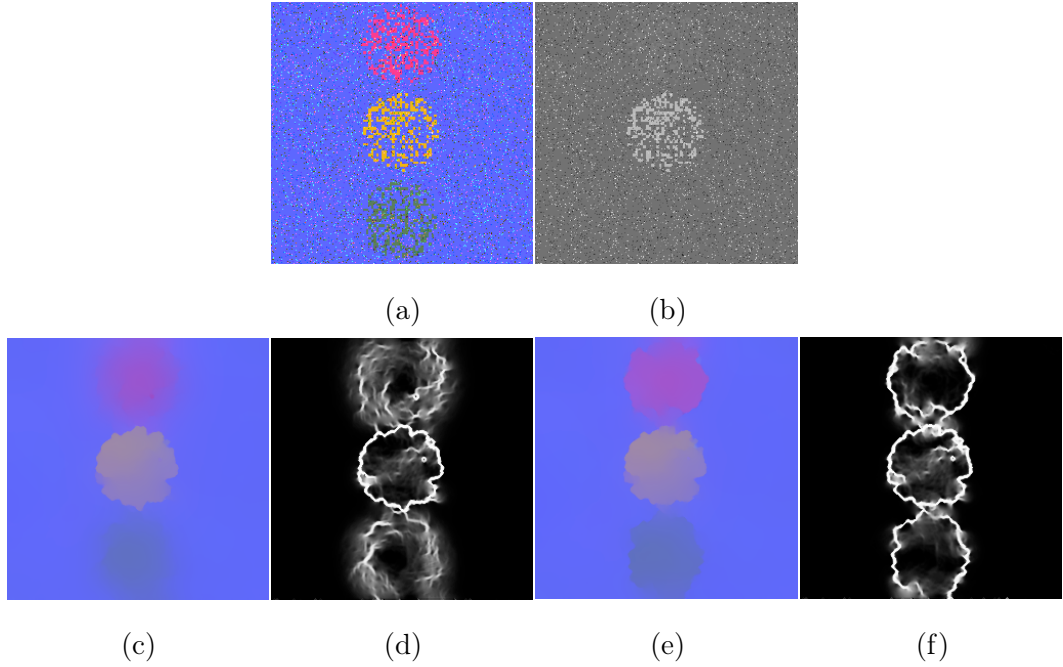
Figure 4.17: Coalition of measures in color image smoothing. (a) Source Image. (b) Corresponding intensity image. (c)-(d) Reconstruction result of color image smoothing with a common edge strength function and a feedback measure estimated from the intensity image (u and 1-v). (e)-(f) Reconstruction result of color image smoothing with a common edge strength function and a feedback measure estimated from all channels (u and $1 - v$).

Figure 4.18 shows some typical smoothing results of natural color images obtained with the proposed method. In all the cases, textured regions are smoothed out, and at the same time structures are preserved.

<div align="center">(a)   (b)</div>
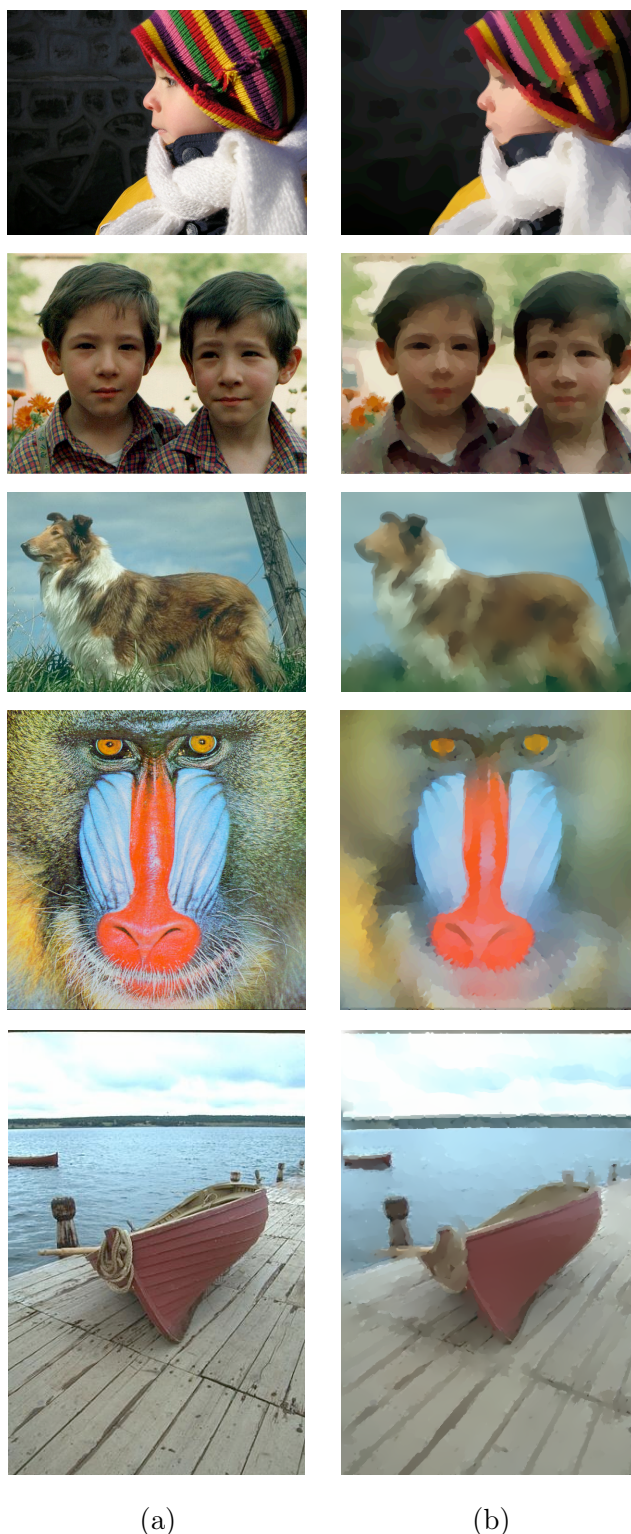
Figure 4.18: Smoothing of natural color images. (a) Source images. (b) Smoothing results.

Computational Complexity and Performance

In this section, we provide information regarding the computational complexity and the performance of our procedure. As we summarized in Algorithm 3, the proposed method consists of the steps that involve estimating the contextual feedback measure and the modified diffusivity, and evolving the smooth image and the edge strength function. Assuming the resolution of the input image is $M \times N$, the complexity of updating the piecewise smooth image and the edge strength function is $\mathcal{O}(MN)$ per iteration. Similarly, the complexity of determining the modified diffusivity at each iteration is $\mathcal{O}(MN)$ since the diffusivity is computed at each pixel using a convex combination where the coefficient corresponds to the feedback measure estimated from local context. Nonetheless, the overall complexity of the proposed method depends mainly on how the contextual feedback measure is estimated and incorporated into the general framework. As this measure is mostly determined according to the features extracted from local image regions, the computational complexity of this step is generally higher than that of others.

For instance, the complexity of computing the feedback measures proposed by us typically dominates the overall process: $\mathcal{O}(MNs)$ for directional consistency, $\mathcal{O}(MN\Delta x(n^2 + \log \Delta x))$ for texture edges, and $\mathcal{O}(MNn^2 \log n)$ for local scale. However, in the case of edge continuity, the complexity is $\mathcal{O}(MN)$ as it is determined taking only the neighboring pixels into account. In the current implementation, we compute the measures respectively based on the directional consistency of edges and edge continuity at each iteration. However, as we stated before, we do estimate the texture edge measure from the source image once at the beginning, and the local scale measure at specific time points due to computational concerns.

We have carried out the experiments presented here on a standard PC hardware (2.4GHz Intel Core 2 Duo system with 2GB of RAM) using nonoptimized MATLAB code. As the complexity analysis reveals, the time required to smooth an image mainly relies on the size of that image, and the contextual feedback measure or measures utilized in the smoothing process. As an example, consider the experiment presented in Figure 4.6. The size of the source image is 365×388. When the edge continuity is taken into account as the feedback measure, it takes 170 iterations and roughly 343 seconds to smooth the image. When the measure based on the directional consistency is used, the computation time is approximately 134 seconds, and the number of iterations for convergence is 99. On the other hand, as one can expect, considering the coalition of these two measures in the smoothing increases the computation time per iteration. The smoothing takes 94 iterations and about 265 seconds.

4.4 Summary and Discussion

Diffusivity modulated by the context is the key mechanism behind the proposed method. The modulated diffusivity $(cv)^2$ is a square of a convex combination of the edge indicator v and a fixed value which is either 1 (negative feedback) or 0 (positive feedback). Four different possibilities for contextual feedback measure are presented to perform several seemingly different tasks: elimination of unintuitive edges due to noise or texture, preserving texture, catching texture boundaries, and enhancing weak edges. Experimental results demonstrate that the proposed regularization is able to remove difficult noise types, capture texture boundaries and produce almost segmentation like results without using singular diffusivities that arise from cost functions such as the L^1-norm.

In the same computational frame, the contextual feedback measures can be combined and the list can be extended. As an example, continuous measures by Kokkinos et al. [80] that allow probabilistic discrimination between edges, textured and smooth areas can be directly utilized in our framework. Moreover, these type of measures can be used as weights in combining the effects of various feedback measures.

In the proposed approach, both negative and positive feedback are implemented as regularization. This may have the disadvantage that if a feature does not exist in the data (e.g. illusory contour), it will not emerge. There may be two solutions. Firstly, an occasional random noise may be added in order to generate spurious edges. If these edge hypotheses do not get sufficient support from the context, they will be eliminated during the regularization process. Secondly, the information rising up in the hierarchy may get connected to a database object whose shape is known and impose top-down constraints to the regularization process. In fact, assuming such an association is present, the next chapter focuses on two new prior-shape guided segmentation methods. Moreover, in Chapter 6, we will return back to this association problem and suggest a new skeleton based analysis of highly-textured natural images. We will extend the TSP method [143], employing the context-influenced smoothing framework proposed here, such that it becomes applicable to natural images. This approach will be simply utilized for extracting visual cues from an image, and consequently for giving insights into the objects and their parts in the images.

CHAPTER 5

PRIOR-GUIDED SEGMENTATION USING TSP SURFACES AS SHAPE PRIORS

In the previous chapter, an image smoothing method that considers the local context has been introduced. The proposed approach employs high-level contextual measures extracted from local image regions to influence the filtering process. We suggested different local contextual measures for modulation, and it was shown that these local contextual measures can resolve certain ambiguities in low-level visual cues. In this chapter, we propose new segmentation frameworks that aim to solve the ambiguities that the context-guided filtering framework cannot cope with.

Segmentation is generally defined as partitioning an image into coherent regions that are likely to correspond to objects in the image. However, as discussed in the introduction chapter, delineating the object boundaries accurately becomes particularly challenging when corrupting influences due to missing regions, partial occlusions and noise appear in images. The only way to resolve these ambiguities is to integrate low-level image features with high-level a priori shape knowledge. In this respect, how to represent a shape becomes the most important issue in devising a prior-shape guided segmentation framework. Hence, first, we address the representation issue by going through several shape representations that mainly encode the local geometry of shapes, yet our main focus is on the Tari-Shah-Pien (TSP) surface [143].

In the subsequent sections, we suggest two new prior-guided segmentation methods that make use of TSP surfaces as shape priors. These studies mainly differ from each other in terms of how they model the deformation of the prior shape. The first method, in particular,

utilizes a representational framework to model global shape variations[1]. On the other hand, the second method integrates a shape matching model that considers local deformation between two shapes with the Mumford-Shah segmentation model[2]. In each section, the performances of these methods are demonstrated on several illustrative images.

5.1 Representing the Prior Shape

There is an increasing number of studies in computer vision that employ prior shape knowledge to guide segmentation. We previously review some of them in Chapter 3. Demand to use prior knowledge about objects within a given class particularly leads to the difficulty of how to represent a shape. The major problem is that for the time being no one has achieved to develop a general theory of shape yet, and thus there is not a clear answer to the question "What is the best way of representing (prior) shapes". As a matter of fact, formulation of appropriate shape spaces and metrics is still an active research topic.

Representation also plays an important role in Marr's view of computational vision [96]. He emphasized that visual perception should be interpreted as an information processing task, which should be understood at three different levels, where representation is one of them. In particular, these levels are

- *Computational theory level (what is to be computed and why?)*

- *Representation and Algorithm level (how the computation is to be done?)*

- *Hardware implementation level (how the process is to be realized physically?)*

It is evident that the computational theory level is more important than the other levels as it describes the underlying computations in an abstract way. There may be several ways to perform the same computational process with a wide choice of representations and algorithms.

A binary image, also known as a silhouette, can be seen as a shape representation itself. However, it does not provide sufficient information about properties of a shape. Hence, in shape analysis literature, silhouettes are generally used as inputs to several other shape representation frameworks. In the proposed work, we prefer to use the TSP surfaces as

---

[1]This is a joint work with Aykut Erdem and was previously published in the Fifth International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2005) [55].

[2]This work has been developed in collaboration with Luminita Vese during my research visit at University of California, Los Angeles (UCLA).

(a)   (b)
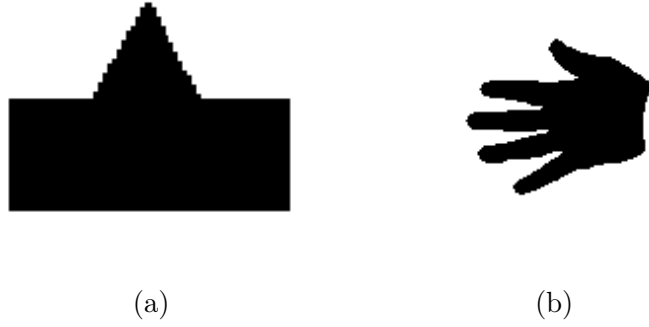
Figure 5.1: (a) A convex shape example. (b) A concave shape example.

shape priors. Before giving the details of this representation and discussing what does the TSP surface offer for this thesis work, let us first review the representations similar to the TSP surfaces. In the subsequent sections, the silhouettes given in Figure 5.1 are used as illustrative examples.

5.1.1 Shape Representation Based on the Distance Transform

The distance transform has been a widely used shape representation method. Given a shape in terms of silhouette, the key idea is to estimate a surface whose value at each internal point is the minimum distance of the point to the shape boundary. The distance transform is mainly used in skeleton estimation frameworks such as [126, 136], where the skeleton points are described as the ones which are equidistant from at least two boundary points. They are also successfully employed in level-set based segmentation frameworks that employ prior shape knowledge (e.g. [89, 120, 148]).

Let Γ denote the shape boundary, then the distance surface (or function) ϕ is defined as:

$$\phi(x) = \min_{y \in \Gamma} \ dist(x, y) \tag{5.1}$$

with x representing internal points of the silhouette. The quick way of computing distance function ϕ is to solve the Eikonal equation

$$|\nabla \phi| = 1 \ . \tag{5.2}$$

The solution of this equation can be easily determined by using the fast marching algorithm of Sethian [129]. Figure 5.2 illustrates estimated distance functions of the shapes given in Figure 5.1 and the corresponding level curves. One of the significant properties of the distance transform is that the zero-level curve of the distance functions corresponds to the
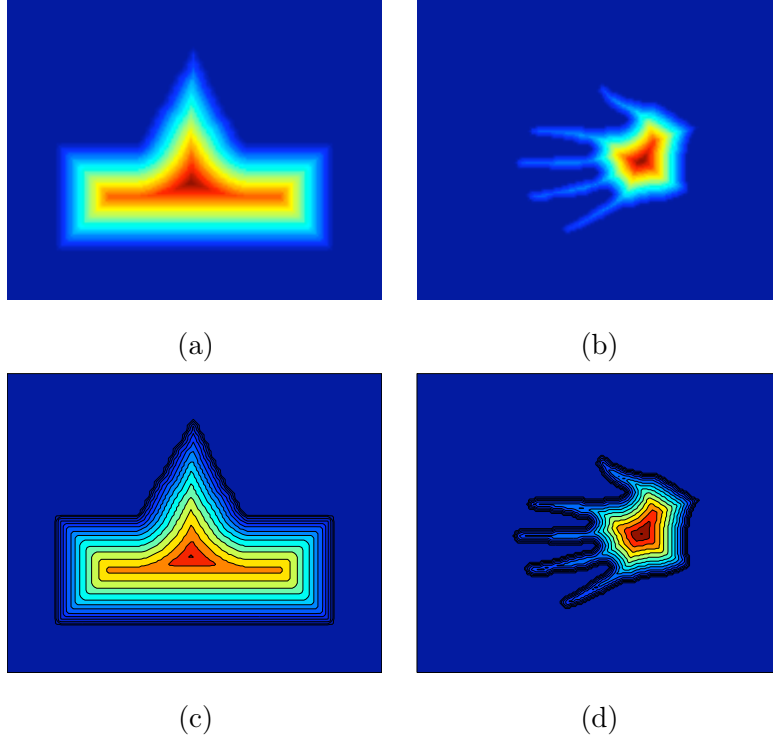
Figure 5.2: (a)-(b) Estimated distance functions. (c)-(d) Corresponding level curves.

shape boundary and the other level curves propagate toward the center of the shape at uniform speed, that is clearly visible in the provided examples.

5.1.2 Shape Representation Using a Poisson Equation [65]

In [65], Gorelick et al. proposed a novel shape representation which can also be seen as a surface estimated from a silhouette. They extracted some shape features by using the differential properties of this surface and utilized them for shape classification and retrieval.

Let S represent a silhouette and ∂S denote its boundary. Then, a surface U can be estimated from the given silhouette by solving the following Poisson equation

$$\nabla^2 U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = -1 \tag{5.3}$$

with $(x, y) \in S$ and the Dirichlet boundary condition $U(x, y) = 0$ along the shape boundary ∂S.

As in the distance transform, the zero-level curve of the surface U denotes the shape boundary. Unlike the distance transform, the proposed Poisson equation leads to a regularization on the corresponding level curves. Thus, the level curves of U represent smoothed analogs of the shape boundary. Moreover, as pointed out in [65], due to this regularization
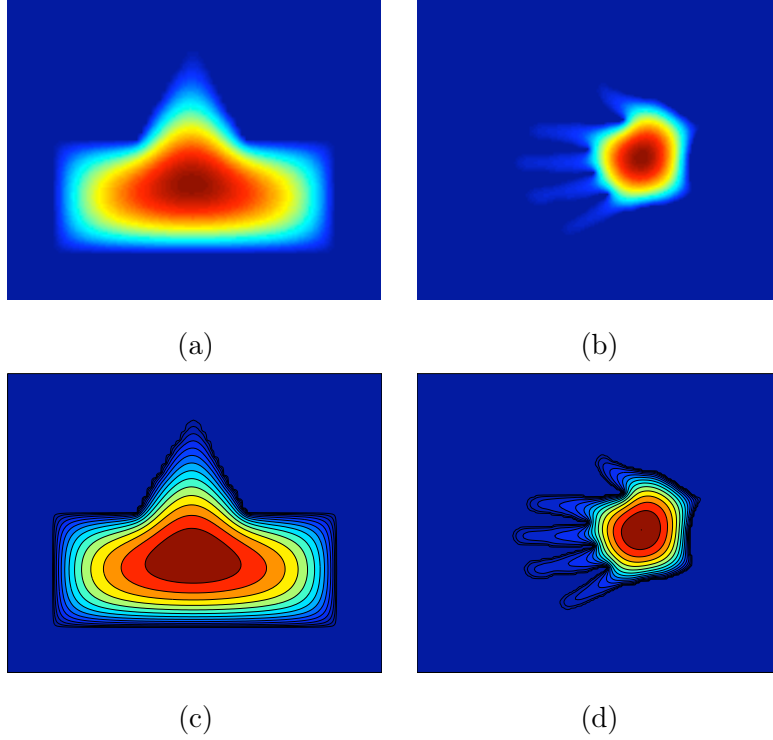
Figure 5.3: (a)-(b) Solutions to the Poisson equation. (c)-(d) Corresponding level curves.

the nearby points effect the values assigned to them, reflecting the local shape structure. Note that, in the distance transform, the value of each point inside the silhouette depends only on the nearest boundary point. Figure 5.3 depicts the solution to the Poisson equation obtained for the silhouettes given in Figure 5.1, and the corresponding level curves (cf. Figure 5.2).

5.1.3 Shape Representation Using Integral Kernels [73]

The shape representation using *integral kernels* was first introduced in [73]. As previously summarized in Section 3.2.6, Hong et al. used this representation as a shape prior within a local deformation model. The key idea is to describe the local structure of a shape in the form of a binary image χ by using a kernel representation.

Particularly, the proposed representation $\mathcal{R}_\sigma$ is determined by using a Gaussian kernel

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{|x|^2}{2\sigma^2}\right) \tag{5.4}$$

as

$$\mathcal{R}_\sigma(\chi; x) = \chi(x)(G_\sigma(x) * (1 - \chi(x))) + (1 - \chi(x))(G_\sigma(x) * \chi(x)) . \tag{5.5}$$

Figure 5.4 provides an illustrative example depicting how the representation given in
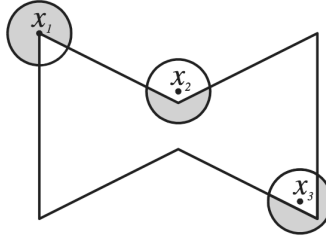
Figure 5.4: The representation using integral kernels is specified by the intersection area between the kernel and the inside or outside of the silhouette (image taken from [73]).

(5.5) is estimated. For the points at the bounding contour or inside the silhouette, the values assigned are specified by the intersection area between the kernel and the outside of the silhouette (see the points denoted by x_1 and x_3). On the other hand, for the points outside the silhouette, the values assigned are determined by the intersection area between the kernel and the inside of the silhouette (see the point denoted by x_2).

Like the representation based on the Poisson equation reviewed in the previous section, the shape representation using integral kernels encodes the local properties of shape geometry. The value assigned to each point is determined according to its position on the silhouette image (inside or outside) and the scale parameter σ. In particular, the parameter σ defines the radius of the disk kernel, and thus the local interactions between the points. Figure 5.5 shows the shape representation based on integral kernels and the corresponding level curves for two different choices of σ for the silhouettes given in Figure 5.1. It is important to note that for the concave shapes, different shape regions may affect the assigned values depending on the choice of the value of σ. The main advantage of this representational framework is that the representation can be directly estimated by convolving the binary images by integral kernels, not by solving a PDE.

5.1.4 Shape Representation Based on the TSP Surface [143]

As previously discussed in Section 2.3.1, the formulation of Ambrosio and Tortorelli uses sufficiently small values of ρ to obtain an approximate edge set represented by the edge strength function v. As $\rho \to 0$, the value of the v function becomes approximately 0 along the boundaries and grows rapidly away from them. In [143], Tari, Shah and Pien employed the AT model with significantly large values of ρ, yielding an alternative role to the AT edge strength function. As the value of ρ increases, the differential properties of the edge strength function v start to encode the skeleton information. The resulting representation is called
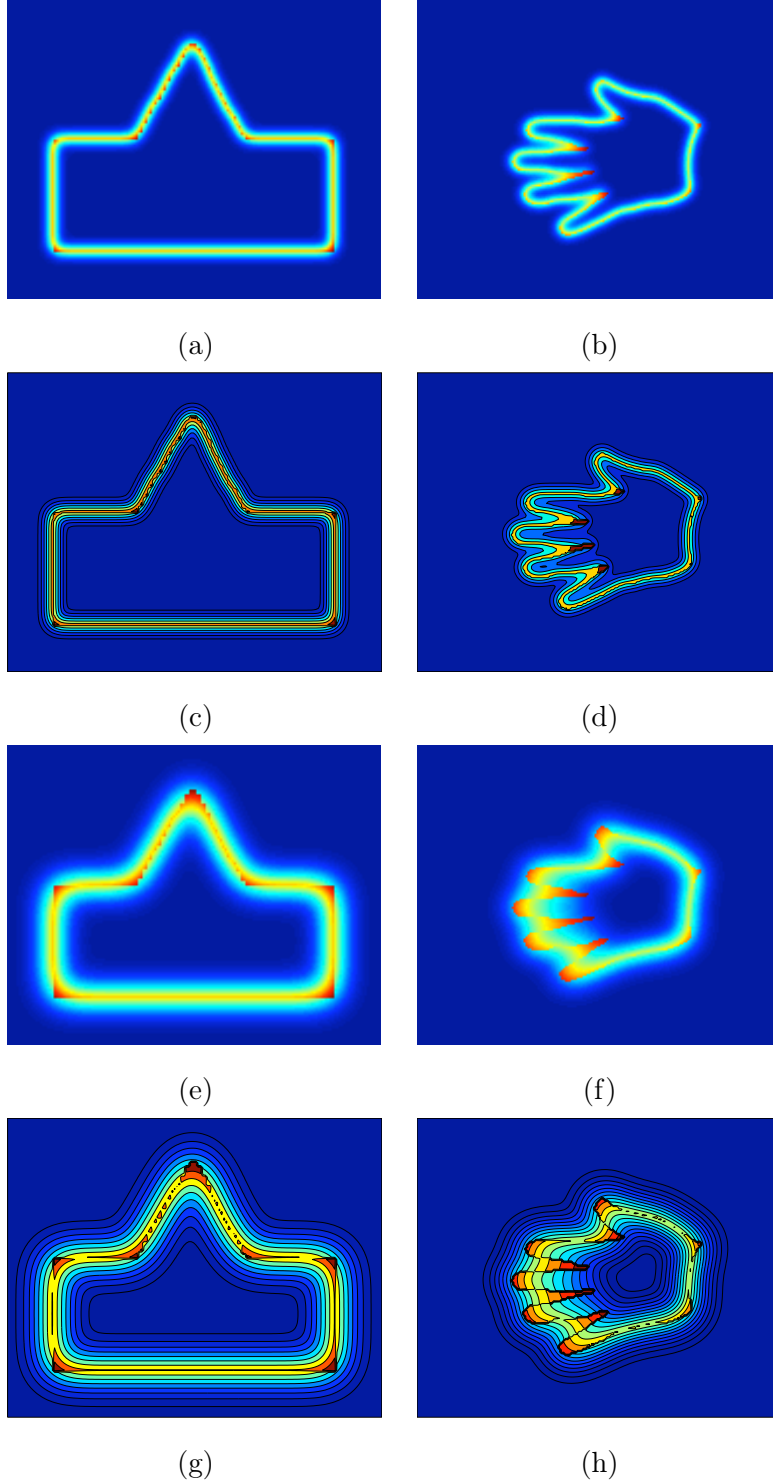
Figure 5.5: (a)-(b) Integral kernel representations with $\sigma = 3$ and (c)-(d) the corresponding level curves. (e)-(f) Integral kernel representations with $\sigma = 7$ and (g)-(h) the corresponding level curves.

the TSP surface throughout this thesis.

The most important advantage of the TSP method is that it can be directly applied to gray scale images. However, we first review the TSP surface on binary images. As pointed out in [143], the analysis on binary images is important since it clearly demonstrates the properties of the TSP surface.

For binary images the TSP surface v is the minimizer of the following functional:

$$E_{b\_TSP}(v) = \frac{1}{2} \int_{\Omega} \left(\rho |\nabla v|^2 + \frac{v^2}{\rho} \right) dx \tag{5.6}$$

subject to the boundary condition $v = 1$ along the shape boundary Γ. Note the change in the value of boundary likelihood for the binary formulation.

The v function is estimated by numerically computing the steady state solution of the PDE

$$\frac{\partial v}{\partial t} = \nabla^2 v - \frac{v}{\rho^2}, \quad v|_{\Gamma} = 1 . \tag{5.7}$$

For sufficiently large values of ρ, the v function becomes an approximate monotonic function of distance from the boundary. More specifically, the edge strength function v acts like a level set function whose level curves are smoothed analogs of the shape boundary that evolve in the inward normal direction with a curvature dependent velocity as in the curve evolution of Kimia, Tannenbaum and Zucker [78].

In the TSP formulation, the parameter ρ determines the level of smoothing of the shape boundary, thus it can be interpreted as a scale parameter. Figure 5.6 illustrates sample level curves of TSP surfaces computed for a cat shape for two different choices of ρ values. As it can be clearly seen, the amount of smoothing increases with increasing ρ. While the TSP surface computed using $\rho = 1$ looks like the solution of the distance transform, the level curves for $\rho = 128$ encode smoother versions of the shape boundary. In this respect, shape representation based on the TSP surface is more similar to the representation based on the Poisson equation than that based on the distance transform.

The TSP method can be easily extended to grayscale images since the edge strength function corresponding to shape boundary can be directly estimated from grayscale images using the AT model reviewed in Section 2.3.1. In the resulting framework, there are two separate scales represented by the parameters α, β and ρ. The ratio between α and β determines the level of smoothing of the grayscale image. With this ratio fixed, the higher the value of α, the more detailed is the image smoothing. Figure 5.7 illustrates the TSP surfaces and the corresponding level curves for the silhouettes given in Figure 5.1 for two
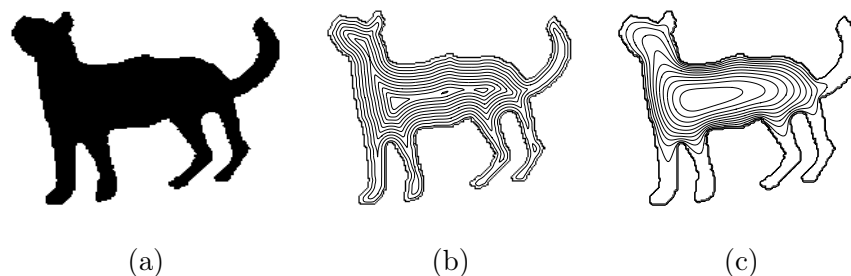
Figure 5.6: Analysis of the TSP method for two different values of ρ. (a) A cat shape. (b)-(c) The level curves of the v function for $\rho = 1$ and $\rho = 128$, respectively.

different choices of ρ. For these surfaces, the binary silhouettes are interpreted as grayscale images, and only the edge strength functions are allowed to evolve with $\alpha = 500$.

The shape skeletons can be easily estimated from the TSP surfaces by tracking the curvature extrama along the level curves of the v function. As Tari, Shah and Pien demonstrated in [143], there is a connection between the differential properties of the edge strength function and the curvature extrama of the evolving shape boundary (Figure 5.8). The details of the skeleton extraction process will be given in the next chapter.

5.1.5 What does the TSP surface offer?

All of the shape representations reviewed in the previous sections reflect local shape structure in their own ways. Distance transform assigns every silhouette point a value based on the nearest boundary point. The other representations use certain regularizations, and hence take more global properties into account. The introduced regularizations in the TSP surface and the surface based on the Poisson equation lead to their level curves to represent smoother versions of the shape boundary. However, from a computational perspective, the representation using integral kernels is more favorable than the others since one does not need to solve any PDE to obtain them.

In the beginning of this chapter, we emphasized that representation is an important issue for an information processing task. Depending on how the underlying process to be carried out, the chosen representation makes certain information more explicit while ignoring irrelevant ones. As we pointed out, the main focus of this thesis work is to enrich low and mid-level computer vision modules with a capacity to extract and use contextual knowledge. This view demands a modified Marr paradigm in which close interactions exist among different visual modules. The consistency of such interactions can be guaranteed only when a common
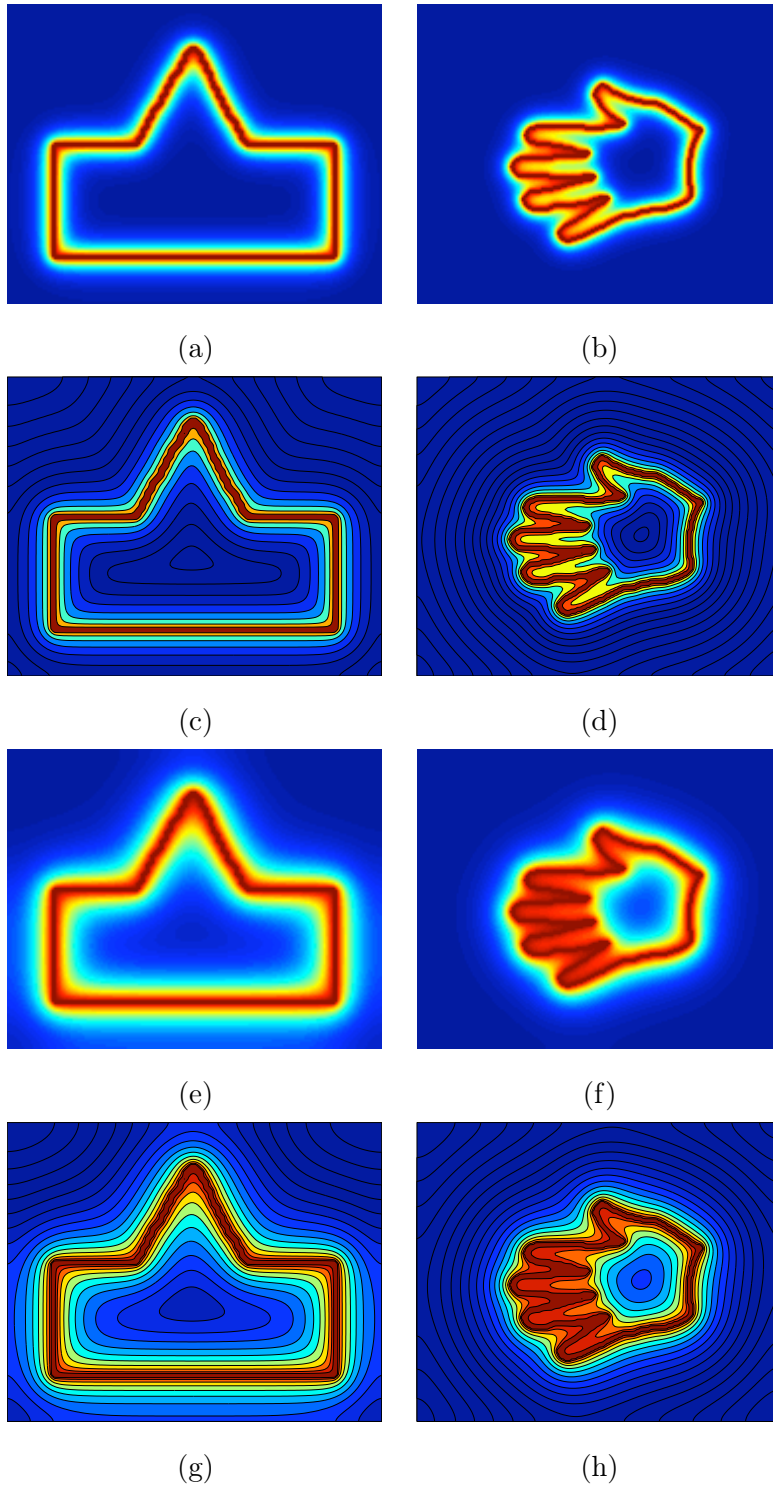
Figure 5.7: (a)-(b) TSP surfaces with $\rho = 4$ and (c)-(d) the corresponding level curves. (e)-(f) TSP surfaces with $\rho = 8$ and (g)-(h) the corresponding level curves.
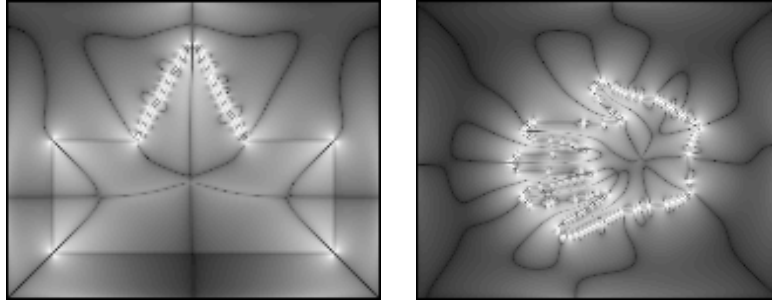
Figure 5.8: Differential properties of the TSP surfaces encode skeleton information.

representational framework is utilized.

Within the scope of this thesis, the TSP surface plays a key role. The segmentation frameworks presented in this chapter integrate the Mumford-Shah segmentation model with prior shape models which use the TSP surfaces as shape priors. In the next chapter, we will extend the TSP method such that it becomes applicable to highly-textured natural images with all their complexity. We divide images into constituent components while simultaneously forming a representation for those components which ties shapes and image features. The patches of uniform characteristics can be robustly and easily used in various image analysis tasks.

5.2 Prior-guided Segmentation Using a Global Shape Model

We previously reviewed some diffusion and variational regularization models in Chapter 2. Most of these models use very generic priors such as smoothness of image regions and continuity of contours to bias the segmentation. However, these priors are generally not sufficient enough to accurately extract object boundaries. In the last decade, there is an increasing number of works that employ prior shape knowledge of objects within a given class to guide the segmentation process. We previously summarized some of these variational frameworks in Chapter 3. In the following, inspired by the unlevel-sets formulation of Raviv et al. [116] (Section 3.2.5) and the statistical formulation of Leventon et al. [89] (Section 3.2.1), we present a new image segmentation method which utilizes a model describing the global variations of the prior shape to supervise the segmentation process.

5.2.1 Modeling Global Shape Variations

In our formulation, we prefer to represent the prior shapes in terms of TSP surfaces. As summarized in the previous section, TSP surfaces provide distance-transform like representations for the shape boundary, and reflect local shape characteristics.

Let Γ denote the boundary of a given shape, we can compute the corresponding TSP surface, v, by minimizing the functional given in (5.6). If one searches only the pose of an object of interest, ignoring the shape variability, then utilizing a single TSP surface representing the shape is sufficient enough to capture the prior information. On the other hand, for more general cases, variations in the shape are also required to be modeled. Here, we use a set of training curves, which correspond to the various appearances of a shape of interest, to devise a global model for shape variations.

Suppose we are given an ensemble of aligned shapes[3] whose boundaries are given by the curves $\Gamma_1, \cdots, \Gamma_p$, respectively. We first estimate the corresponding TSP surfaces, $v_1(x), \cdots, v_p(x)$. Then, following an idea presented by Leventon et al. [89] and later adopted in [148], we perform a Principal Component Analysis (PCA) on these TSP surfaces. This process yields a linear low-dimensional shape space where a shape within the given class can be represented in terms of the mean TSP surface and the coefficients of the estimated principal components. Specifically, the mean TSP surface $\bar{\Phi}(x)$ is the ordinary average of $v_1(x), \cdots, v_p(x)$. Let $\Phi_1, \cdots, \Phi_p$ be the principal components computed by PCA, then a possible shape from this ensemble is represented with

$$\Phi_{\mathbf{w}}(x) = \bar{\Phi}(x) + \sum_{i=1}^{p} w_i \Phi_i(x) \ . \tag{5.8}$$

Notice that this process does not guarantee that the resultant representation $\Phi_{\mathbf{w}}(x)$ will be a TSP surface, yet the experiments reveal that the global model yields sufficient segmentation results.

5.2.2 Segmentation Framework

In the previous section, a representation for the shape prior in terms of a mean TSP surface and principle components is developed. Now, we will discuss how this representation (5.8) can be utilized as a shape prior to guide the segmentation process.

We formulate image segmentation as an energy minimization problem. Our boundary detection rule is based on the AT functional. In this formulation, we let the edge strength

---

[3]The alignment algorithm proposed in [148] is used in the experiments.

function v be approximately 1 along the boundaries and decay away from the boundaries (cf. 2.21), leading to the following version of the AT model:

$$E_{AT'}(u,v) = \int_\Omega \left(\beta(u-f)^2 + \alpha((1-v)^2|\nabla u|^2) + \frac{1}{2}\left(\rho|\nabla v|^2 + \frac{v^2}{\rho}\right)\right) dx \ . \tag{5.9}$$

The simplest way of integrating prior shape information with the AT functional is to add two energies to arrive at a combined energy:

$$E = E_{AT'} + \mu E_{shape} \tag{5.10}$$

where μ is the parameter which enforces the shape similarity of the embedding surface in the overall segmentation process.

A straightforward choice for shape energy is simply

$$E_{shape}(v, t_x, t_y, h, \theta, \mathbf{w}) = \int_\Omega (v - T(\Phi_\mathbf{w}))^2 dx \tag{5.11}$$

where the pose transformation function T applied to the shape prior $\Phi_\mathbf{w}$ is defined as follows:

$$\begin{bmatrix} x' \\ y' \\ T(\Phi_\mathbf{w}) \end{bmatrix} = \underbrace{\begin{bmatrix} h & 0 & 0 \\ 0 & h & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{scale} \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{rotation} \begin{bmatrix} x \\ y \\ \Phi_\mathbf{w} \end{bmatrix} + \underbrace{\begin{bmatrix} t_x \\ t_y \\ 0 \end{bmatrix}}_{translation} \ . \tag{5.12}$$

However, we observed that such a straightforward choice may cause segmentation process to trap into local minima. Hence, we considered the normalized difference in our formulation:

$$E_{shape}(v, t_x, t_y, h, \theta, \mathbf{w}) = \frac{\int_\Omega (v - T(\Phi_\mathbf{w}))^2 dx}{\int_\Omega (v + T(\Phi_\mathbf{w}))^2 dx} \ . \tag{5.13}$$

Our segmentation process recovers both pose transformation parameters t_x, t_y, θ, and h and the shape variability parameters $\mathbf{w}$ simultaneously along with a piecewise smooth image u and the corresponding TSP surface v. For numerical implementation, we discretize the proposed functional with respect to spatial variables instead of discretizing the Euler-Lagrange equations obtained from the functional as we usually do. Upon casting the problem into a discrete setting, we arrive to the following optimization problem:

$$\begin{aligned} \min E(u,v,t_x,t_y,h,\theta,\mathbf{w}) \ = \ & \sum_{i=1}^{width}\sum_{j=1}^{height} \beta(u_{i,j}-f_{i,j})^2 + \alpha(1-v_{i,j})^2(u_x^2+u_y^2) \\ + \ & \frac{\rho}{2}(v_x^2+v_y^2) + \frac{v_{i,j}^2}{2\rho} \\ + \ & \mu\frac{\sum_{i=1}^{width}\sum_{j=1}^{height}(v_{i,j}-T(\Phi_\mathbf{w})_{i,j})^2}{\sum_{i=1}^{width}\sum_{j=1}^{height}(v_{i,j}+T(\Phi_\mathbf{w})_{i,j})^2} \end{aligned} \tag{5.14}$$

99

subject to the constraints

$$0 \leq u_{i,j} \leq 255 \,,$$

$$0 \leq v_{i,j} \leq 1 \,.$$

In (5.14), u_x, u_y, v_x, v_y denote the central difference approximations for x and y derivatives of $u_{i,j}$ and $v_{i,j}$, respectively:

$$u_x = \frac{u_{i+1,j} - u_{i-1,j}}{2} \,, \quad u_y = \frac{u_{i,j+1} - u_{i,j-1}}{2} \,, v_x = \frac{v_{i+1,j} - v_{i-1,j}}{2} \,, \quad v_y = \frac{v_{i,j+1} - v_{i,j-1}}{2} \,.$$

We now apply gradient descent to (5.14) to obtain segmentation. Minimizing the functional with respect to each unknown variable leads to the following evolution equations:

$$
\begin{aligned}
\frac{\partial u_{i,j}}{\partial t} =\ & \frac{\alpha}{2} \left[(1 - v_{i-1,j})^2 (u_{i,j} - u_{i-2,j}) - (1 - v_{i+1,j})^2 (u_{i+2,j} - u_{i,j}) \right. \\
& \left. +\ (1 - v_{i,j-1})^2 (u_{i,j} - u_{i,j-2}) - (1 - v_{i,j+1})^2 (u_{i,j+2} - u_{i,j}) \right] \\
& +\ 2\beta (u_{i,j} - f_{i,j}) \,,
\end{aligned}
\tag{5.15}
$$

$$
\begin{aligned}
\frac{\partial v_{i,j}}{\partial t} =\ & -2\alpha (1 - v_{i,j})(u_x{}^2 + u_y{}^2) + \frac{v_{i,j}}{\rho} \\
& +\ \frac{\rho}{4} \left[(v_{i,j} - v_{i-2,j}) - (v_{i+2,j} - v_{i,j}) + (v_{i,j} - v_{i,j-2}) - (v_{i,j+2} - v_{i,j}) \right] \\
& +\ 2\mu \left[\frac{(v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j})}{S} - \frac{D(v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j})}{S^2} \right] \,,
\end{aligned}
\tag{5.16}
$$

$$
\begin{aligned}
\frac{\partial t_x}{\partial t} =\ & 2\mu \left[\frac{\sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j}) \left(-\frac{\partial T(\Phi_{\mathbf{w}})}{\partial x'} \right)_{i,j}}{S} \right. \\
& \left. -\ \frac{D \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j}) \left(\frac{\partial T(\Phi_{\mathbf{w}})}{\partial x'} \right)_{i,j}}{S^2} \right] \,,
\end{aligned}
\tag{5.17}
$$

$$
\begin{aligned}
\frac{\partial t_y}{\partial t} =\ & 2\mu \left[\frac{\sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j}) \left(-\frac{\partial T(\Phi_{\mathbf{w}})}{\partial y'} \right)_{i,j}}{S} \right. \\
& \left. -\ \frac{D \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j}) \left(\frac{\partial T(\Phi_{\mathbf{w}})}{\partial y'} \right)_{i,j}}{S^2} \right] \,,
\end{aligned}
\tag{5.18}
$$

$$
\begin{aligned}
\frac{\partial h}{\partial t} =\ & 2\mu \left[\frac{\sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j}) \left(-\frac{\partial T(\Phi_{\mathbf{w}})}{\partial x'} \frac{\partial x'}{\partial h} - \frac{\partial T(\Phi_{\mathbf{w}})}{\partial y'} \frac{\partial y'}{\partial h} \right)_{i,j}}{S} \right. \\
& \left. -\ \frac{D \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j}) \left(\frac{\partial T(\Phi_{\mathbf{w}})}{\partial x'} \frac{\partial x'}{\partial h} + \frac{\partial T(\Phi_{\mathbf{w}})}{\partial y'} \frac{\partial y'}{\partial h} \right)_{i,j}}{S^2} \right] \,,
\end{aligned}
\tag{5.19}
$$

$$\frac{\partial \theta}{\partial t} \;=\; 2\mu \left[\frac{\sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j}) \left(-\frac{\partial T(\Phi_{\mathbf{w}})}{\partial x'}\frac{\partial x'}{\partial \theta} - \frac{\partial T(\Phi_{\mathbf{w}})}{\partial y'}\frac{\partial y'}{\partial \theta} \right)_{i,j}}{S} \right.$$
$$\left. - \frac{D \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j}) \left(\frac{\partial T(\Phi_{\mathbf{w}})}{\partial x'}\frac{\partial x'}{\partial \theta} + \frac{\partial T(\Phi_{\mathbf{w}})}{\partial y'}\frac{\partial y'}{\partial \theta} \right)_{i,j}}{S^2} \right] \;,\;\; (5.20)$$

$$\frac{\partial w_l}{\partial t} \;=\; 2\mu \left[\frac{\sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j}) \left(-T(\Phi_l)_{i,j} \right)}{S} \right.$$
$$\left. - \frac{D \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j}) \left(T(\Phi_l)_{i,j} \right)}{S^2} \right] \qquad (5.21)$$

where
$$D = \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} - T(\Phi_{\mathbf{w}})_{i,j})^2 \,, \quad S = \sum_{i=1}^{width} \sum_{j=1}^{height} (v_{i,j} + T(\Phi_{\mathbf{w}})_{i,j})^2 \,.$$

Our strategy is to alternate between these gradient descent equations during minimization. When we try to update a variable, we keep the other unknown variables fixed and apply the corresponding gradient descent equation. A numerical stopping criteria for the iterative scheme can be defined in the sense the rate of change of the evolving variables is less than a threshold. The overall minimization procedure is summarized in Algorithm 4.

Algorithm 4 Minimization of the proposed segmentation functional

---

1: Compute TSP surfaces, $v_1, \cdots, v_p$, representing the given training curves $\Gamma_1, \cdots, \Gamma_p$

2: Determine the mean TSP surface $\bar{\Phi}$ and the principal components $\Phi_1, \cdots, \Phi_p$ by performing PCA on $v_1, \cdots, v_p$

3: Initialize the variables with $u^0 = f$, $v^0 = \frac{2\alpha\rho|\nabla u^0|^2}{1 + 2\alpha\rho|\nabla u^0|^2}$

4: Set initial values for pose transformation parameters t_x, t_y, θ, h

5: Set initial values of shape variability parameters $w_1, \cdots, w_p$

6: **while** stopping criteria is not reached **do**

7:    Update u and v according to (5.15) and (5.16), respectively

8:    Update t_x and t_y using (5.17) and (5.18), respectively

9:    Update h and θ according to (5.19) and (5.20), respectively

10:    Compute the new shape variability parameters $w_1, \cdots, w_p$ using (5.21)

11: **end while**

5.2.3 Experimental Results

In this section, we demonstrate the segmentation results of our algorithm on various images. In all experiments, we set $\mu = 1$ unless stated otherwise and use sufficiently small values of smoothing radius, (i.e. $\sqrt{\alpha/\beta} = 0.1$ or $\sqrt{\alpha/\beta} = 1$). Typical ρ values are $8, 16, 32$.

As pointed out before, if the shape of the object of interest does not vary, we can use a single TSP surface representing the object as our shape prior. Following this, we start with some simple examples. We apply our method to the hand image given in Figure 5.9(a). The segmentation process takes the prior shape given in Figure 5.9(b) into account. We simply estimate the TSP surface representing the prior shape (Figure 5.9(c)) and integrate it with our segmentation framework. The shape prior is initialized as demonstrated in Figure 5.9(d), and then allowed to evolve according to the proposed framework. As seen in Figure 5.9(e), the segmentation given by our method can handle plane Euclidean transformations and accurately delineate the object boundaries.

Delineating the occluded object boundaries is in general a very difficult task. The segmentation methods that utilize cues depend merely on image data yields poor results. Figure 5.10(a) shows such an example that is generated from the hand image shown in Figure 5.9(a) by artificially introducing a ring type shape and cutting off some part of the pointer finger. In our formulation, the corrupting influences due to occlusion and missing

(a)   (b)   (c)

(d)   (e)
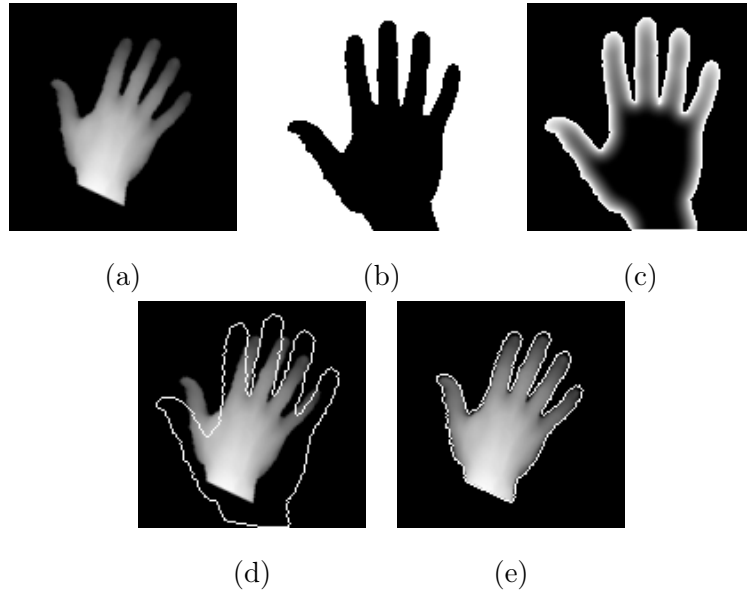
Figure 5.9: Segmentation result of a hand image. (a) Input image. (b) Prior hand shape. (c) Prior TSP surface computed with $\rho = 8$. (d) Initial one-level curve of the prior TSP superimposed on the input image. (e) Final segmentation result.

regions is compensated by the prior shape knowledge, and as a result our method gives accurate object boundaries (Figure 5.10(c)). Instead of taking μ constant and equal to 1, if we increase its value throughout the iterations, we can speed up the recovery process of the transformation parameters. Thus, the missing and occluded parts become apparent in the evolving TSP surface in less number of iterations as demonstrated in Figure 5.11.

An advantage of TSP surfaces is that they allow us to represent shape boundaries which are not simple closed curves within our formulation. In the following, we consider such an

(a)   (b)   (c)
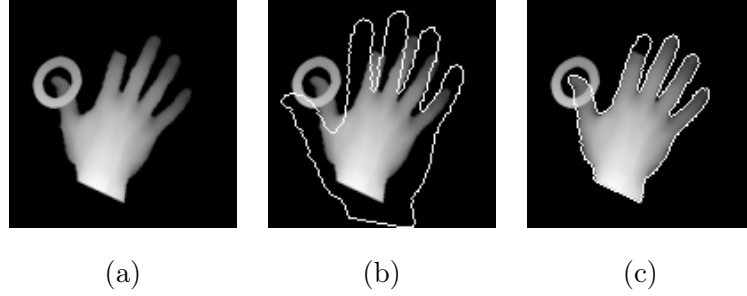
Figure 5.10: Segmentation of partially occluded hand with missing parts. (a) Input image. (b) Initial one-level curve of the prior TSP superimposed on the input image. (c) Final segmentation result.
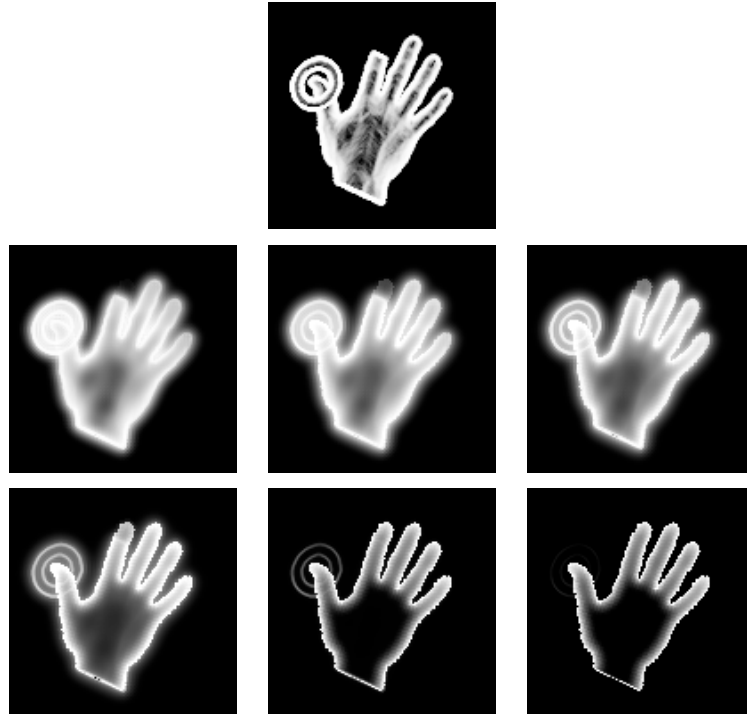
Figure 5.11: Evolution of the TSP surface of the hand image given in Figure 5.10(a) throughout iterations.

example where we consider a prior shape in the form of a line drawing with junctions as illustrated in Figure 5.12(a)). We estimate the TSP surface representing the prior shape (Figure 5.12(b)), and use it to segment out the object given in Figure 5.12(c). Although the object is composed of two blobs, as Figure 5.12(e) illustrates, the proposed work is able to extract both of them simultaneously. Notice that a formulation based on level sets cannot be applied here since the segmenting curve is no longer a simple closed curve.

For the experiments illustrated in Figure 5.9 - Figure 5.12, we have ignored the shape variability and used a single TSP surface to represent prior shape knowledge. In the next example, Figure 5.14, we demonstrate how the shape variations within an object class are dealt with by the proposed framework. For this, we use a set of airplane shapes given in Figure 5.13, each having a different appearance. We extract the TSP surface of each airplane in the training set, and then perform PCA on the surfaces estimated. This process yields a mean TSP surface and main principal components that we use to define the shape prior as discussed in Section 5.2.1. Once the prior is constructed, next, we test the performance of our segmentation framework on a set of novel examples, not belonging to the training set. Figure 5.14(a) shows these example airplanes that all have different poses. In Figure 5.14(b), we present the training curves superimposed on the input images considering the initial pose

(a)   (b)

(a)   (b)   (c)
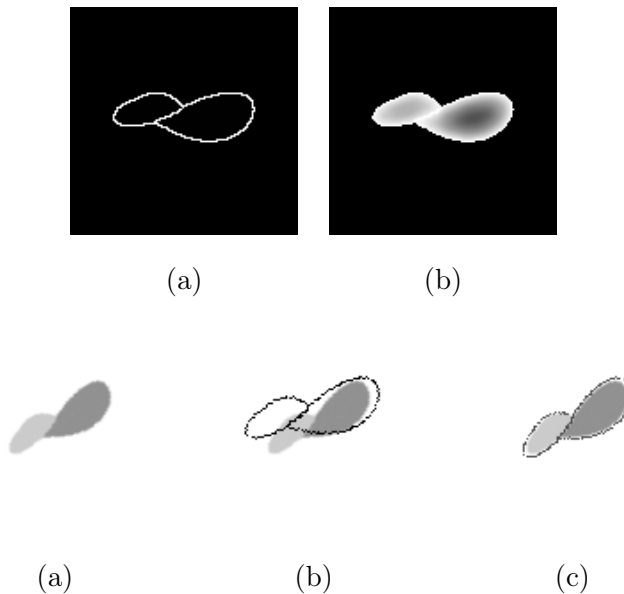
Figure 5.12: Segmenting an object considering a shape prior given in the form of a line drawing with junctions. (a) Prior shape. (b) Prior TSP surface computed with $\rho = 8$. (c) Input image. (d) Initial one-level curve of the prior TSP superimposed on the input image. (e) Final segmentation result.
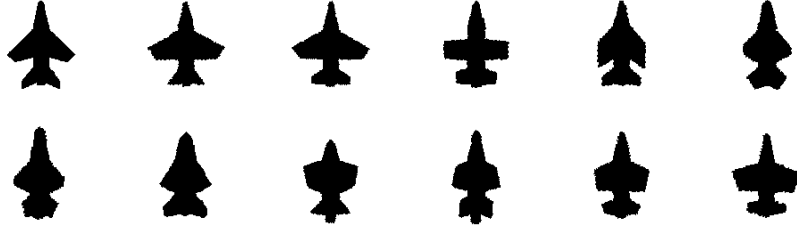
Figure 5.13: Set of airplane shapes taken from [148].

(a)　　　　　　(b)　　　　　　(c)
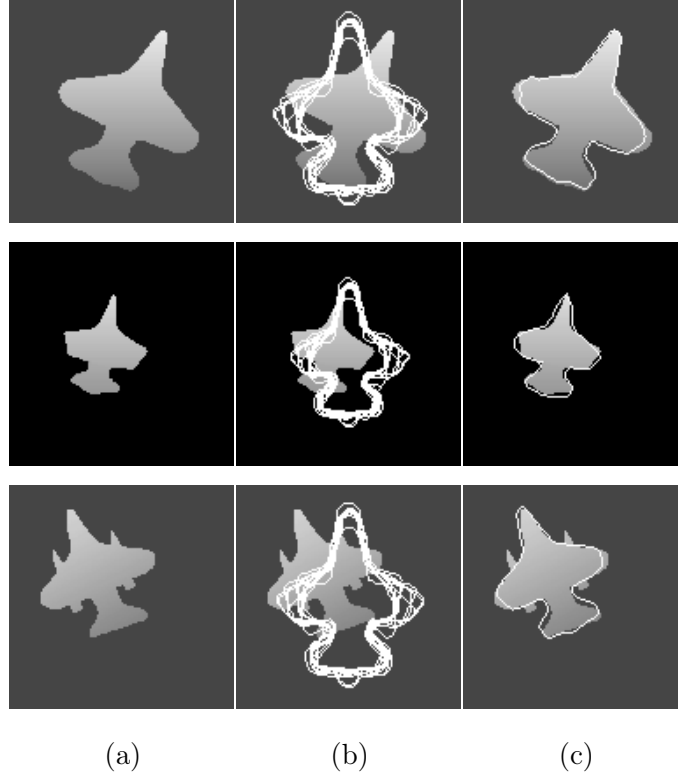
Figure 5.14: Segmentation of airplane images using a model describing the global variations of the prior shape. (a) Input images. (b) Training curves superimposed on the input image (with the initial pose parameters). (c) Final segmentation results.

parameters. The contours in Figure 5.14(c) are the outcomes of the proposed formulation. Since TSP surfaces have values that are approximately 1 along an edge and decay to 0 away from an edge, we obtain these contours by performing some morphological operations on the surfaces estimated. The results obtained show that the shape prior satisfactorily describes the global variations of the prior shape. For the airplane image in the bottom row, the additional parts in the wings are not parts of the segmentation as they are not previously observed in the training set. This can be interpreted as a drawback of the global shape model in that it does not describe the local deformations in shape.

5.3 Prior-guided Segmentation Utilizing a Local Deformation Model

In the previous section, we proposed a novel image segmentation method which utilizes TSP surfaces to model the global variations of the prior shape and to supervise the segmentation process. In this section, we present a new segmentation framework that make use of the TSP surfaces to determine meaningful correspondences between a given template and the evolving object region by taking account of a local deformation model similar to the formulation of Hong et al. [73] (Section 3.2.6).

5.3.1 Shape Matching Using a Local Deformation Model

The aim of shape matching is to determine correspondences between two shapes. One way to find these correspondences is to extract some features or structures from the given silhouettes and to reduce the problem to finding a matching between them (e.g. [7, 16, 78, 160]). We are interested in integrating a matching model into the Mumford-Shah segmentation model, and hence we avoid such kinds of frameworks and focus on a formulation that establishes the correspondences without performing any explicit shape analysis.

We approach shape matching as a registration problem where the task is to find a transformation between corresponding data [98]. Given two silhouettes, a reference S_1 and a template S_2, the trivial approach is to interpret the silhouettes S_1 and S_2 themselves as shape representations and to determine a deformation such that the distance between the reference shape and the transformed template shape is minimized. However, very likely that in such a formulation, the estimated correspondences between the silhouettes will be semantically wrong since each silhouette point is indistinguishable from others. Obtaining more meaningful matching results demands more descriptive representations. Hence, we first represent S_1 and S_2 by the corresponding TSP surfaces. As we review in Section 5.1.4, the TSP surfaces assign each point a distinct value based on the local geometry of the shapes.

Let v_1 and v_2 denote the TSP surfaces estimated from S_1 and S_2, respectively. Matching between the given shapes is estimated by minimizing the functional

$$E_{match}(h) = E_{fid}(h) + \beta E_{reg}(h) \qquad (5.22)$$

with respect to the unknown deformation function $h : \Omega \to \mathbb{R}^2$. The first term in the energy $E_{fid}(h)$ measures the similarity between the reference TSP surface and the template TSP surface which is transformed under the displacement vector field h.

In particular, the similarity term is defined as

$$E_{fid}(h) = \frac{1}{2} \int_{\Omega} (v_2(x + h(x)) - v_1(x))^2 \, dx \qquad (5.23)$$

where Ω denotes the domain in which v_1 and v_2 are defined. While this term provides a driving force for the registration, the regularization term $E_{reg}(h)$ forces the deformation field h to be smooth. The relative importance between smoothing and similarity terms is determined by the parameter β.

In literature, several approaches are proposed to regularize the deformation field h (see Part II of [98] for a review). In our formulation, we use the linear elasticity model proposed in [28], which is a physically motivated model:

$$E_{reg}(h) = \int_{\Omega} \left(\frac{\bar{\mu}}{4} \sum_{i,j=1}^{2} \left(\partial_{x_i} h_j + \partial_{x_j} h_i \right)^2 + \frac{\lambda}{2} (\nabla \cdot h)^2 \right) dx \qquad (5.24)$$

where ∂_{x_j} denotes the partial derivative with respect to x_j, and $\lambda, \bar{\mu} > 0$ are the Lamé constants that reflect material properties of an elastic body (a TSP surface in our framework). This model cannot handle large deformations, and hence we assume that a global registration is initially performed.

The deformation field h minimizing (5.22) formally satisfies the Euler-Lagrange equation

$$\frac{\partial h}{\partial t} = -\frac{E_{match}}{\partial h} = - \left(\frac{\partial E_{fid}}{\partial h} + \beta \frac{\partial E_{reg}}{\partial h} \right) \qquad (5.25)$$

where

$$\frac{\partial E_{fid}}{\partial h} = (v_2(x + h(x)) - v_1(x)) \, \nabla v_2(x + h(x)) \,, \qquad (5.26)$$

$$\frac{\partial E_{reg}}{\partial h} = - \left(\bar{\mu} \nabla^2 h + (\lambda + \bar{\mu}) \nabla (\nabla \cdot h) \right) \,. \qquad (5.27)$$

In the following, Figure 5.15 - Figure 5.18, we present example matching results obtained by the proposed method. From the silhouettes in Figure 5.15(a)-(b) - Figure 5.18(a)-(b), we first estimate the corresponding TSP surfaces on which we perform the actual matching process. These are given in Figure 5.15(c)-(d) - Figure 5.18(c)-(d). As discussed, finding correspondences between two shapes corresponds to determining a deformation field h that assigns a displacement vector $h(x)$ for every point x, minimizing (5.22). The deformation fields representing the matching between these shapes are illustrated in Figure 5.15(e) - Figure 5.18(e). As it can be clearly seen from Figure 5.15(f) and Figure 5.16(f), the resulting registration process determines meaningful correspondences between the given shapes for the first two examples. However, for the last two cases, although the template shapes

are accurately registered to the reference shapes (Figure 5.17(f) and Figure 5.18(f)), the matching results are not so good. This clearly depicts that the registration procedure yields reliable matching results only when the shapes to be registered are locally similar.
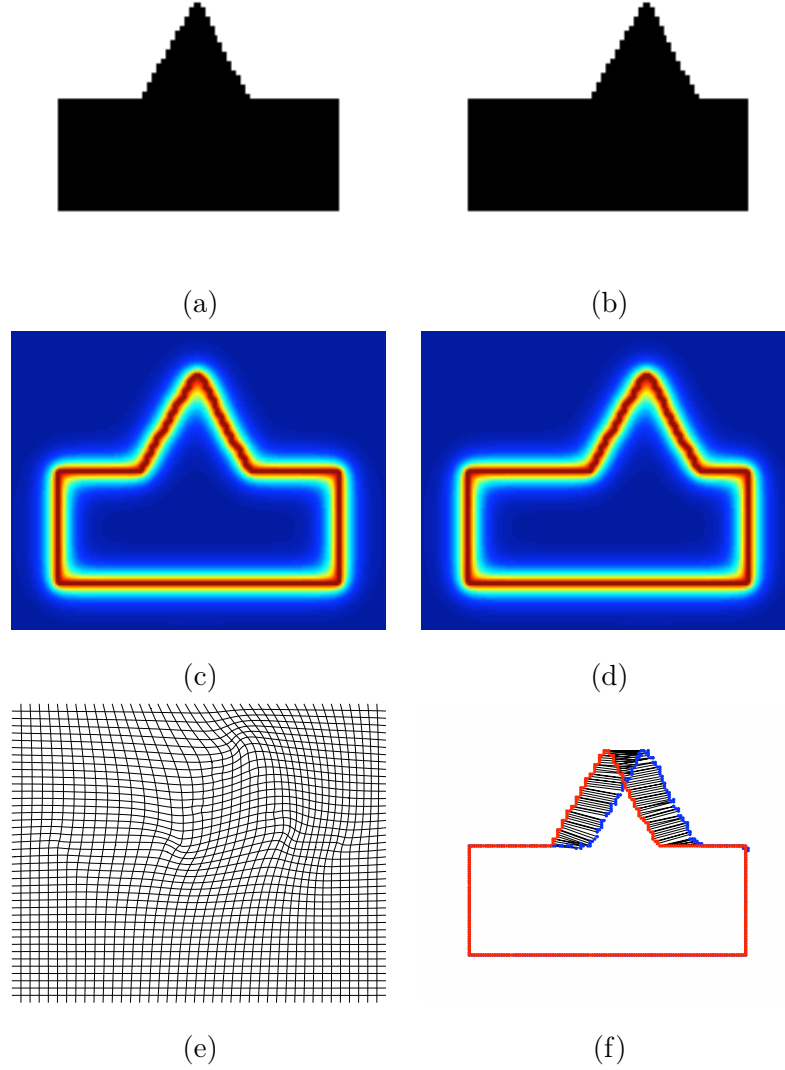
Figure 5.15: (a)-(b) Silhouettes of two artificial shapes. (c)-(d) Corresponding TSP surfaces estimated with $\rho = 4$. (e) Estimated deformation vector field. (f) Matching result.
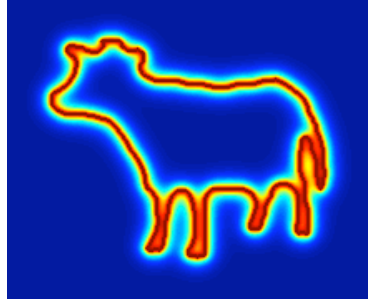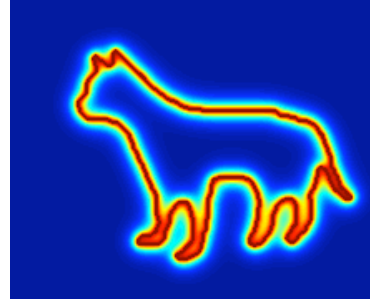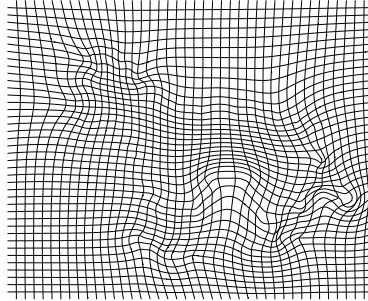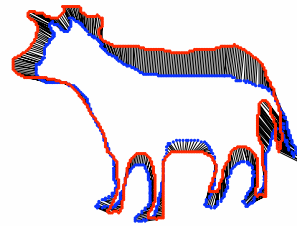
Figure 5.16: (a)-(b) *cow* and *cat* silhouettes (adapted from [26]). (c)-(d) Corresponding TSP surfaces estimated with $\rho = 4$. (e) Estimated deformation vector field. (f) Matching result.

(a)　　　　　　　　　　(b)

(c)　　　　　　　　　　(d)

(e)　　　　　　　　　　(f)
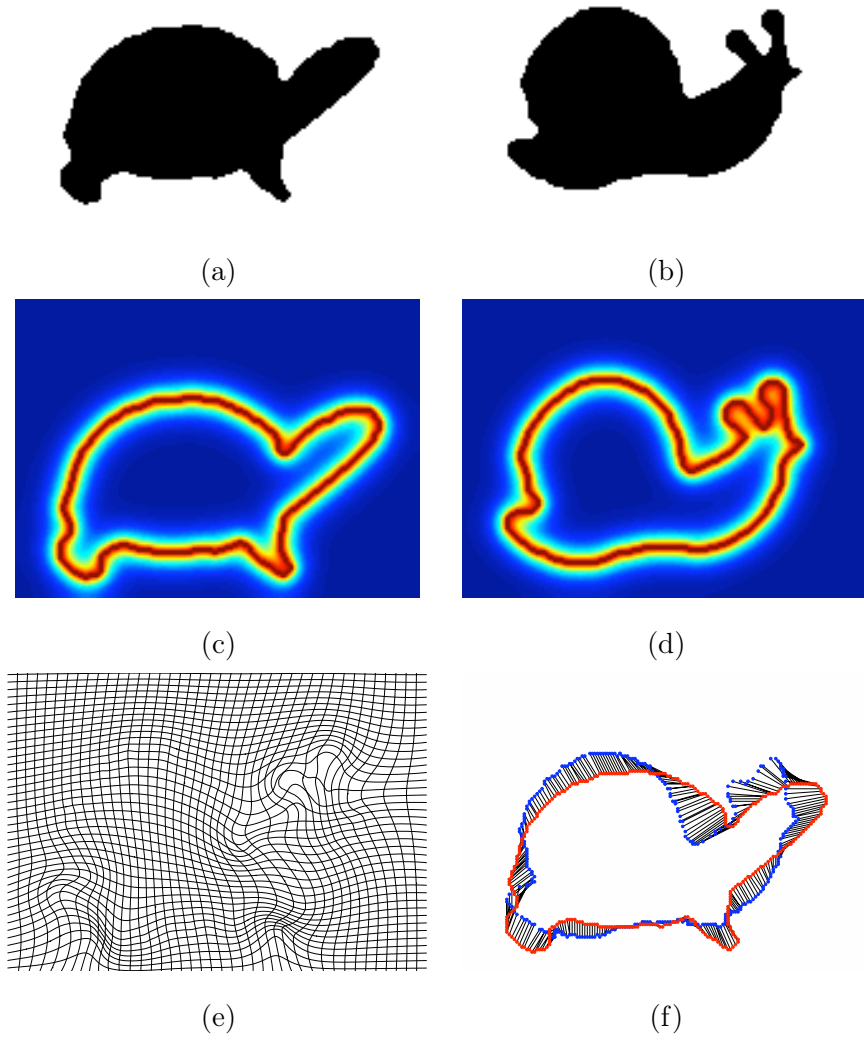
Figure 5.17: (a)-(b) Silhouettes of two artificial shapes. (c)-(d) Corresponding TSP surfaces estimated with $\rho = 4$. (e) Estimated deformation vector field. (f) Matching result.

(a)

(b)

(c)

(d)

(e)

(f)
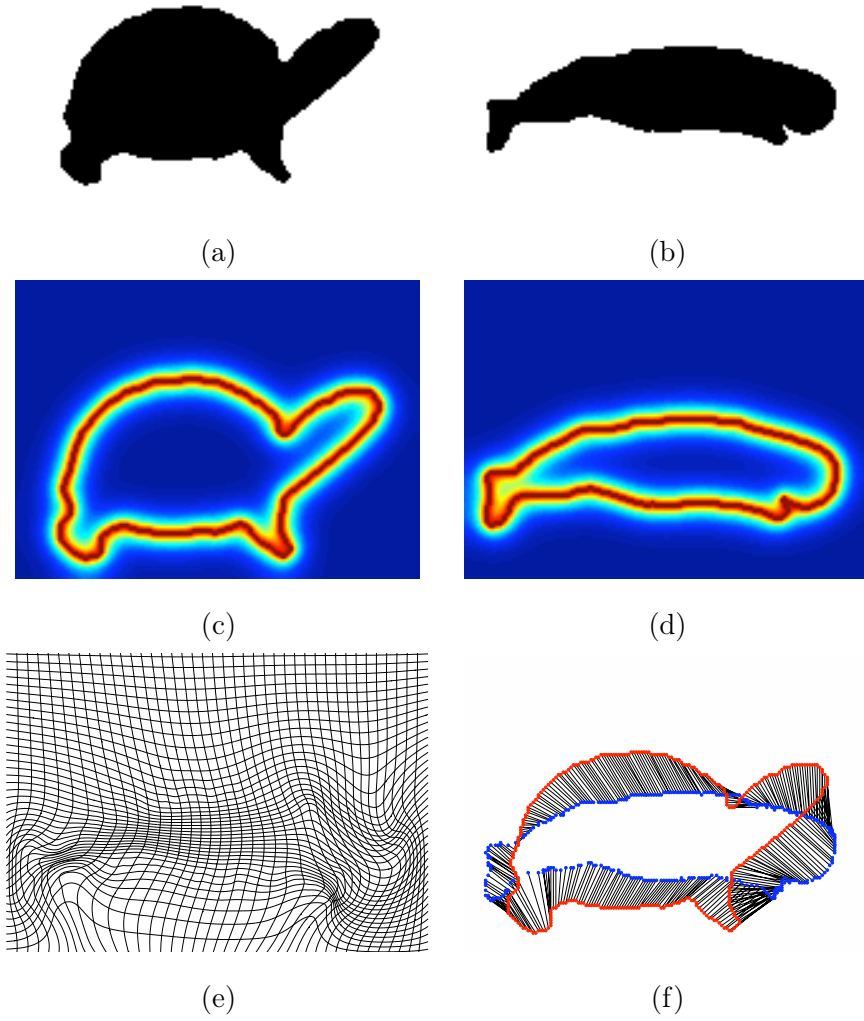
Figure 5.18: (a)-(b) Silhouettes of two artificial shapes. (c)-(d) Corresponding TSP surfaces estimated with $\rho = 4$. (e) Estimated deformation vector field. (f) Matching result.

111

5.3.2 Segmentation Framework

In this section, motivated by the formulation of Hong et al. [73], we present a new segmentation framework that make use of the TSP surfaces and the matching functional discussed in the previous section. Let T denote the given binary template and v_T represent the corresponding TSP surface. We formulate image segmentation by minimization of a pair of coupled functionals, E_1 and E_2 defined as follows:

$$E_1(\phi|h) = E_{CV}(\phi) + \mu_1 E_{shape}(\phi|h), \quad \mu_1 \geq 0 , \tag{5.28}$$

$$E_2(v, h|\phi) = E_{TSP}(v|\phi) + \mu_2 E_{match}(h), \quad \mu_2 \geq 0 . \tag{5.29}$$

The first functional E_1 is an extended Chan-Vese (CV) functional (2.33) with an additional prior shape term given by

$$E_{shape}(\phi|h) = \int_\Omega \left(H(\phi(x)) - T(x + h(x))\right)^2 dx . \tag{5.30}$$

This prior term is similar to the one used in [116] and measures the difference in the areas of the evolving object region represented by $H(\phi)$ and the transformed prior under the deformation h. As a result, while CV term leads to an image force that attracts the zero-level curve of the evolving level set function ϕ to object boundaries, the prior term enforces the zero-level curve towards the given template T under transformation.

The second functional E_2 consists of two terms. The term E_{TSP} is utilized to represent the evolving object region in terms of a TSP surface denoted by v, and defined by

$$E_{TSP}(v|\phi) = \alpha \int_\Omega v^2 |\nabla H(\phi)|^2 dx + \frac{1}{2} \int_\Omega \left(\rho |\nabla v|^2 + \frac{(1-v)^2}{\rho}\right) dx . \tag{5.31}$$

In fact, it is a modified AT functional (2.21) where the variable u is replaced with $H(\phi)$ and the fidelity term is excluded. The second term E_{match} in (5.29) corresponds to the matching energy defined in (5.22). Minimizing this functional with respect to the unknown variables v and h establishes a correspondence between the evolving object region and the prior shape that are represented as the TSP surfaces v and v_T, respectively.

The functionals E_1 and E_2 are coupled in the sense that the deformation field h found by (5.29) determines the level set ϕ; and the level set ϕ estimated by (5.28) specifies the TSP surface v and the deformation field h. These two functionals work jointly to partition an image into object vs. background regions.

Our strategy is to alternate between these functionals when we apply the gradient descent. We fix the deformation field h when we try to minimize the first functional E_1, and determine

the level set ϕ. Similarly, we fix the level set function ϕ when we try to minimize the second functional E_2, and estimate the evolving shape represented by the TSP surface v and the deformation field h. Note that we could have combine these two functionals and have started with it, interpreting it as a general framework. The main disadvantage of such a formulation is that when we apply the gradient descent, the deformation field h depends on not only the TSP surfaces, but also the binary silhouettes, which may yield inaccurate correspondences as we stated before.

We now apply gradient descent to the functionals E_1 and E_2 to obtain segmentation. Minimizing the first functional (5.28) with respect to ϕ leads to the following evolution equation:

$$\frac{\partial \phi}{\partial t} = -\frac{E_1}{\partial \phi} = -\left(\frac{E_{CV}}{\partial \phi} + \mu_1 \frac{E_{shape}}{\partial \phi}\right) \qquad (5.32)$$

with

$$\frac{E_{shape}}{\partial \phi} = 2\,\delta(\phi(x))\,(H(\phi(x)) - T(x + h(x)))\ . \qquad (5.33)$$

With fixed ϕ, minimizing the second functional (5.29) with respect to v and h leads to two evolution equations. While the evolution equation for h is similar to (5.25) where v_2 is replaced with v_T, and v_1 with v, the evolution equation for v is given by:

$$\frac{\partial v}{\partial t} = -\frac{E_2}{\partial v} = -\left(\frac{E_{TSP}}{\partial v} + \mu_2 \frac{E_{match}}{\partial v}\right) \qquad (5.34)$$

with

$$\frac{E_{TSP}}{\partial v} = \nabla^2 v - \frac{2\alpha |\nabla H(\phi)|^2 v}{\rho} - \frac{(v-1)}{\rho^2}\ , \qquad (5.35)$$

$$\frac{E_{match}}{\partial v} = (v_T(x + h(x)) - v(x))\ . \qquad (5.36)$$

The minimization procedure of the proposed framework is summarized in Algorithm 5. In the experiments, the iterations continue until the square of the overall energy $(E_1 + E_2)^2$ is no more decreasing.

5.3.3 Experimental Results

In this section, we present experimental results on various natural images. In order to demonstrate the performance of our framework and to illustrate what it is gained by introducing prior shape knowledge, we compare the results of our method with those of the CV model using the same values for the common parameters and the same initial conditions. Note that setting μ_1 to zero and excluding E_2 from computations reduce our formulation to the CV model, yielding a model that does not take prior shape into account. In all experiments, we

Algorithm 5 The proposed segmentation algorithm

1: Compute the TSP surface v representing the given binary template T

2: Initialize the variables with $\phi^0 = \phi_0$, $v^0 = v_T$, $h^0 = 0$

3: **while** stopping criteria is not reached **do**

4:     Estimate c_1 and c_2 using (2.35) and (2.36), respectively

5:     Transform the template T under the currently estimated deformation field h.

6:     Update the level set function ϕ according to (5.32) by taking the transformed template into account

7:     Update the TSP surface v by iterating (5.34) 500 times

8:     Estimate the deformation field h by iterating (5.25) 200 times

9: **end while**

set $\varepsilon = 1$, $\lambda_1 = \lambda_2 = 1$, $\rho = 4$, $\alpha = 500$, $\mu_2 = 10^{-4}$, $\lambda = 0$, $\bar{\mu} = 1$ unless stated otherwise, and assume that a global registration is initially performed.

Let us start with some simple examples. We apply our method to the hand image given in Figure 5.19(a). The segmentation process takes the prior shapes shown in Figure 5.19(b) into account. The level sets are initialized by considering these shapes and then are allowed to evolve according to the proposed framework. Since there are no corrupting influences in the images and hands correspond to single homogeneous regions, as seen in Figure 5.19(c)-(d), the segmentations given by our method and that by the CV model accurately delineate the object boundaries. The similar results obtained show that top-down knowledge provided by a shape prior is not always required for segmentation. It can be argued that it should be consulted only when an image is ambiguous. However, an advantage of the proposed work is that it segments an image while jointly registering the prior shape on the image data[4] (Figure 5.19(e)). This important issue will be discussed in detail through some illustrative examples later on.

A major challenge in image segmentation is to handle occlusions. In general, it is impossible to delineate the occluded object boundaries accurately by utilizing the cues depend merely on image data. Figure 5.20(a) shows two such examples that are generated from Figure 5.19(a) by artificially introducing occlusions. In the images, the hands are partially

---

[4]In [67], the authors introduced a model that also performs registration and segmentation simultaneously. The model depends on the CV model and a nonlinear elasticity regularization on the deformation vector field and does not employ any prior shape knowledge. Hence, it cannot fully cope with the ambiguities in images as compared to our formulation.

| (a) | (b) | (c) | (d) | (e) |
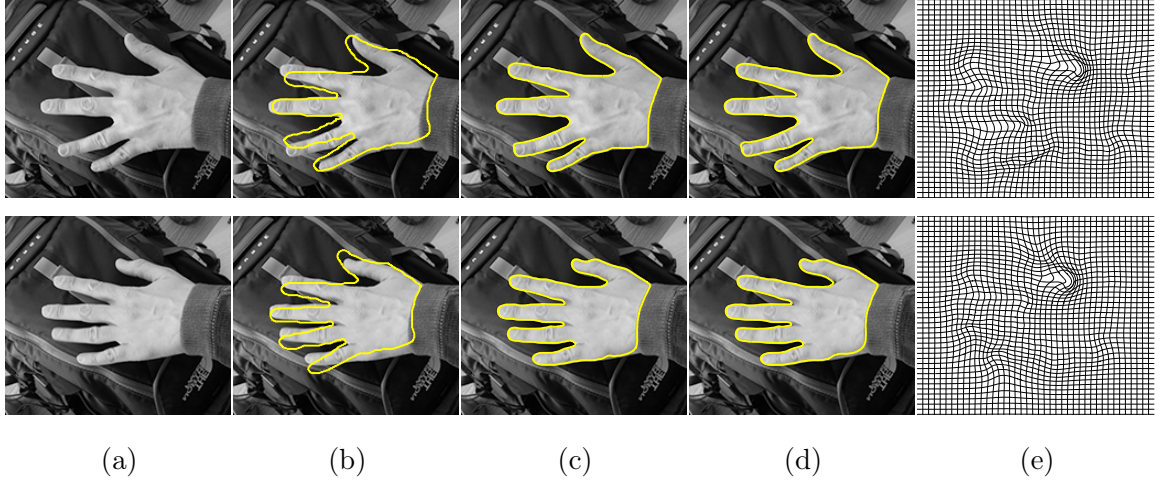
Figure 5.19: Segmentation results of two hand images. (a) Input images. (b) Unregistered prior shapes superimposed on the input images (initial zero-level curves). (c) Segmentation results without prior shape information. (d) Final segmentation results. (e) Deformed grids under the estimated displacement vector fields. (the parameters are $\mu = 15000$, $\mu_1 = 10000$ and $\beta = 0.01$).

occluded by the pencils, resulting in some parts of the regions describing the hands missing. Thus, without the shape priors (Figure 5.20(b)), the missing information cannot be recovered and the segmenting contours fail to capture the objects of interest as in Figure 5.20(c). When the prior shape knowledge is taken into account, the shapes of the evolving object regions are not allowed to deviate much from the prior shapes, and as a result ambiguous information due to partial occlusions are compensated and the final contours yield more acceptable approximations of object boundaries (Figure 5.20(d)).

We next consider images corrupted by noise. The hand images given in Figure 5.21(a) are noisy versions of the images presented in the bottom rows of Figure 5.19(a) and Figure 5.20(a), respectively. They were generated by corrupting 75% of the image pixels with uniform noise. Although there is a significant amount of noise in the images, the proposed framework is able to produce relatively good segmentation results (Figure 5.21(d)). Particularly, in segmenting the noisy hand image in the bottom row, our method copes with not only missing information due to noise, but also that due to partial occlusions. This experiment also illustrates the robustness of the CV model against noise. For the noisy hand image in the top row, the contour obtained by the CV model delineates the object boundary without taking any prior knowledge into account. Although this is the case, the one obtained by the proposed method is more smoother due to the regularization introduced on the deformation

(a)   (b)   (c)   (d)   (e)
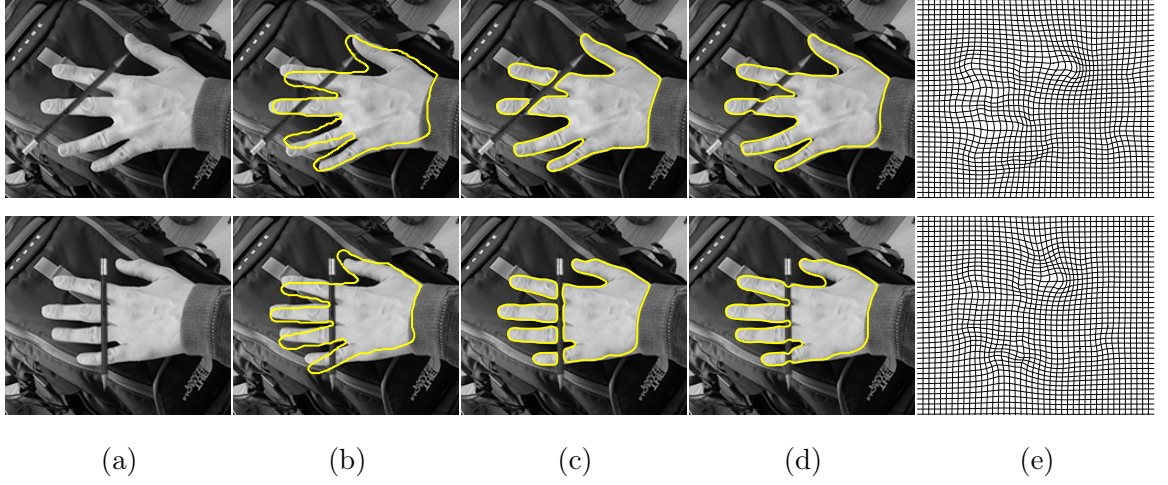
Figure 5.20: Segmentation of partially occluded hands. (a) Input images. (b) Unregistered prior shapes superimposed on the input images (initial zero-level curves). (c) Segmentation results without prior shape information. (d) Final segmentation results. (e) Deformed grids under the estimated displacement vector fields. (the parameters are $\mu = 25000$, $\mu_1 = 20000$ and $\beta = 0.025$).

(a)   (b)   (c)   (d)   (e)
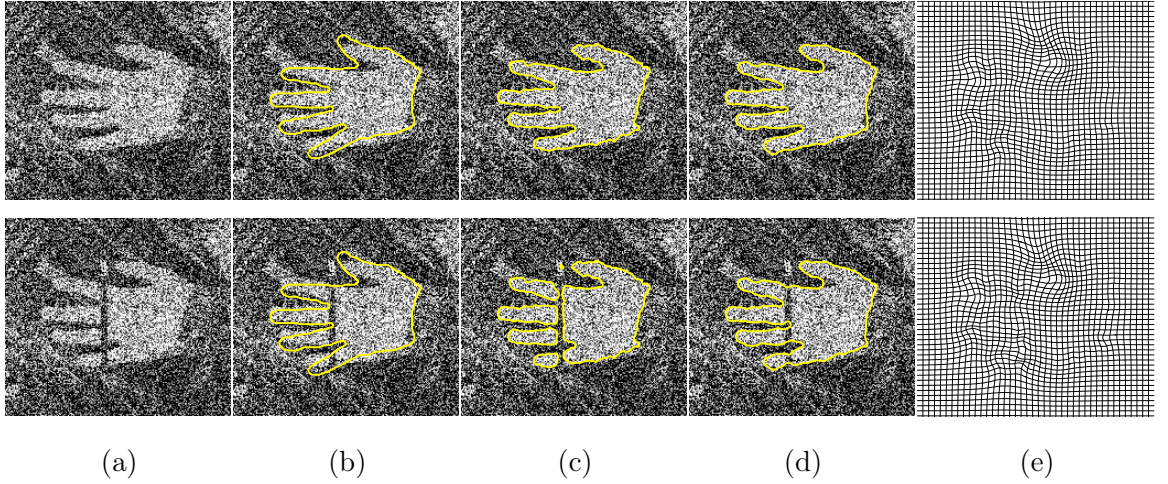
Figure 5.21: Segmenting hand images corrupted by significant amount of noise. (a) Input images. (b) Unregistered prior shapes superimposed on the input images (initial zero-level curves). (c) Segmentation results without prior shape information. (d) Final segmentation results. (e) Deformed grids under the estimated displacement vector fields. (the parameters are $\mu = 20000$, $\mu_1 = 15000$ (top row), $\mu_1 = 12500$ (bottom row) and $\beta = 0.01$).

field (Figure 5.21(e)) that is utilized to transform the prior shape, which guides the evolving contour.

In the next example, Figure 5.22, we demonstrate a major drawback of our formulation. For the images we illustrated in Figure 5.20 and Figure 5.21, the proposed framework handles missing or misleading information by establishing a correspondence between the evolving object region and the prior shape and guiding segmentation process accordingly. However, once the amount of missing information increases, this alters. Figure 5.22(a) shows example hand images under heavy occlusions. Note that a certain part of the thumb is behind the bar in the image in the top row, and fingers are nearly missing in the image in the bottom row. As seen in Figure 5.22(d), with increasing occlusion, there is not sufficient information for the employed local deformation model to compensate, thus the proposed formulation results in degraded segmentation results.

An advantage of our formulation is that it performs image segmentation with simultaneously registering the prior shape on the image data as we discussed in the beginning of this section. In the following, we show joint results of segmentation and registration processes on a set of horse images. Figure 5.23(a) contains the horse images that are used in this experiment (the horse image in the third row is taken from the *Weizmann Horse Database* [22] and that in the last row is generated by artificially modifying it). In Figure 5.23(b), the initial zero-level curves, which correspond to the unregistered shape priors, are superimposed on the input images. The contours in Figure 5.23(c) and (d) are the outcomes of the CV

(a)   (b)   (c)   (d)   (e)
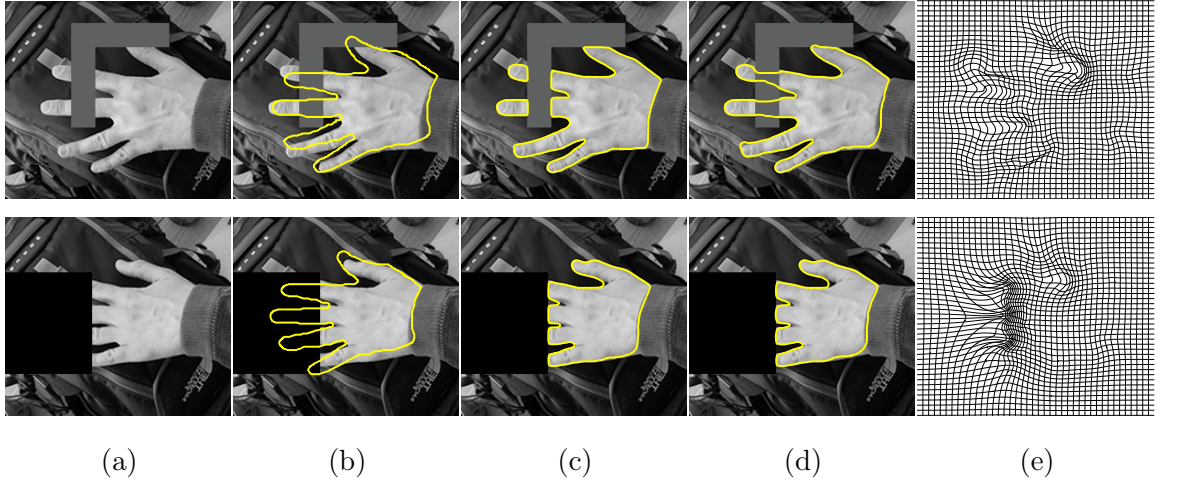
Figure 5.22: Segmenting hand images under heavy occlusions. (a) Input images. (b) Unregistered prior shapes superimposed on the input images (initial zero-level curve). (c) Segmentation results without prior shape information. (d) Final segmentation results. (e) Deformed grids under the estimated displacement vector fields. (the parameters are $\mu = 15000$, $\mu_1 = 20000$, $\beta = 0.005$ (top row) and $\mu = 30000$, $\mu_1 = 15000$, $\beta = 0.01$ (bottom row)).

<center>(a)   (b)   (c)   (d)   (e)</center>
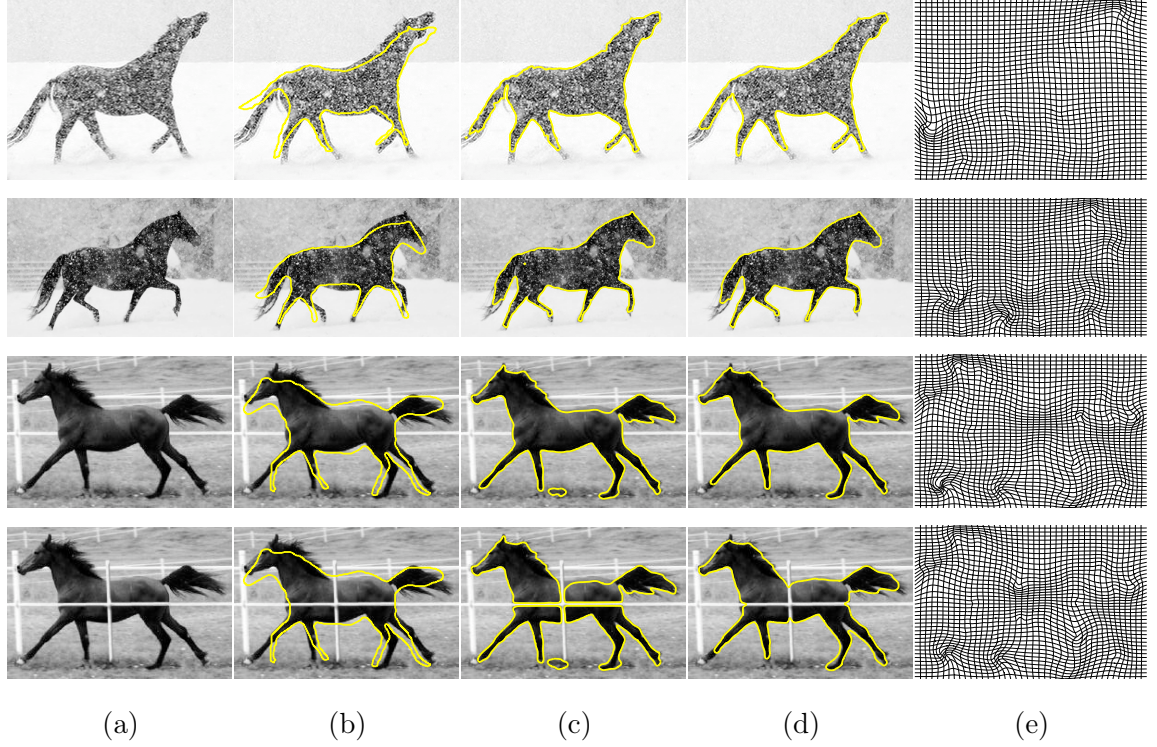
Figure 5.23: (a) Input images. (b) Unregistered prior shapes superimposed on the input images (initial zero-level curves). (c) Segmentation results without prior shape information. (d) Final segmentation results. (e) Deformed grids under the estimated displacement vector fields.

model and the proposed formulation, respectively. Table 5.1 lists the parameters utilized in segmentation.

Figure 5.24 presents the corresponding registration results, where the initial zero-level curves are broken into meaningful parts (head, tails, legs, etc.) and transformed under the estimated displacement vector fields shown in Figure 5.23(e) accordingly. We have used different colors for different parts of the curves in order to show how meaningful the obtained correspondences are. As these results confirm, even the CV model and our framework give the same segmentation results (e.g. for the horse images in the topmost two rows), our model is superior to the Chan-Vese model due to jointly performed registration process that allows to perform additional semantic analysis if an explicit shape analysis is carried out beforehand.

As discussed in Section 2.3.2, the CV model can be interpreted as a two-phase piecewise constant approximation of the MS model, where the given image is assumed to be composed of two regions as a dark object on a bright background, or a bright object on a dark

<center>118</center>

Table 5.1: The parameters for the images given in Figure 5.23

	μ	μ_1	β
first row	9000	10000	0.05
second row	15000	5000	0.01
third row	18000	5000	0.01
fourth row	18000	20000	0.01
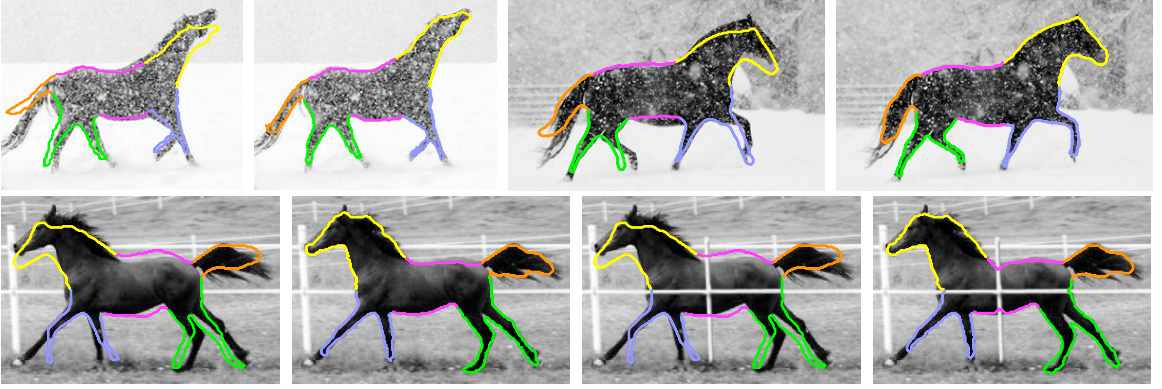
Figure 5.24: Analysis of the registration process. Correspondences between the prior shapes and the deformed versions of them are shown in same color.

background. Although the model can cope with noise (Figure 5.21) as it does not employ gradient information to extract object boundaries, two-phase piecewise constant assumption generally yields unrealistic results for textured images. A similar argument can be made for our formulation since it is based on the Chan-Vese model.

Consider the highly textured images presented in the top row of Figure 5.25. While the image in the left is a synthetic image containing a human pattern, that in the right is a natural starfish image. To eliminate the corrupting effects due to texture, we suggest to perform a priori smoothing process guided by the context-influenced filtering model proposed in Chapter 4. The outcomes of this smoothing process are shown in the bottom row of Figure 5.25 [5]. Figure 5.26 illustrates the segmentation results obtained by using the original and the smoothed images with the parameters listed in Table 5.2.

The contours in Figure 5.26(b) and (c) are the outcomes of the Chan-Vese model and our framework, respectively. The parameters λ_1 and λ_2 are chosen so that different weights

---

[5]These results are estimated by taking coalition of directional consistency, edge continuity and texture edges into consideration with the parameters $\alpha = 5$ (left column), $\alpha = 4$ (right column), $\beta = 0.01$, $\rho = 0.01$, $\Delta x = 10$ and $\varepsilon_{te} = 20$.

Table 5.2: The parameters used in segmenting the images presented in Figure 5.25

	μ	λ_1	λ_2	μ_1	β
first column	7000	0.7	1	5000	0.01
second column	100	0.25	1	100	0.05
third column	3000	1	0.25	1000	0.01
fourth column	1000	1	0.1	500	0.01
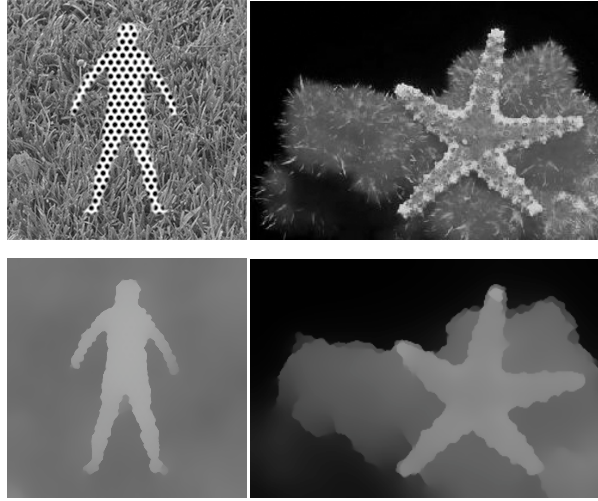
Figure 5.25: Input images and the corresponding smoothing results.

are assigned to the fidelity terms of the object and background regions during the minimization process. We extract more accurate object boundaries with prior smoothing, especially for the textured image containing the human pattern. Additionally, performing smoothing beforehand improves the speed of segmentation since smoothed images require less number of iterations.

In Figure 5.27, we present the registration results that are simultaneously estimated with the segmentation results given in Figure 5.26. While Figure 5.27(a) illustrates the unregistered curves, Figure 5.27(b) and (c) show the registered ones that are estimated by using the input and the smoothed images, respectively. Again, the colors depict the correspondences between the initial curves and the final ones. Note that all these curves are superimposed on the original unprocessed images whether they are computed from them or not.

(a)

(b)

(c)

(d)
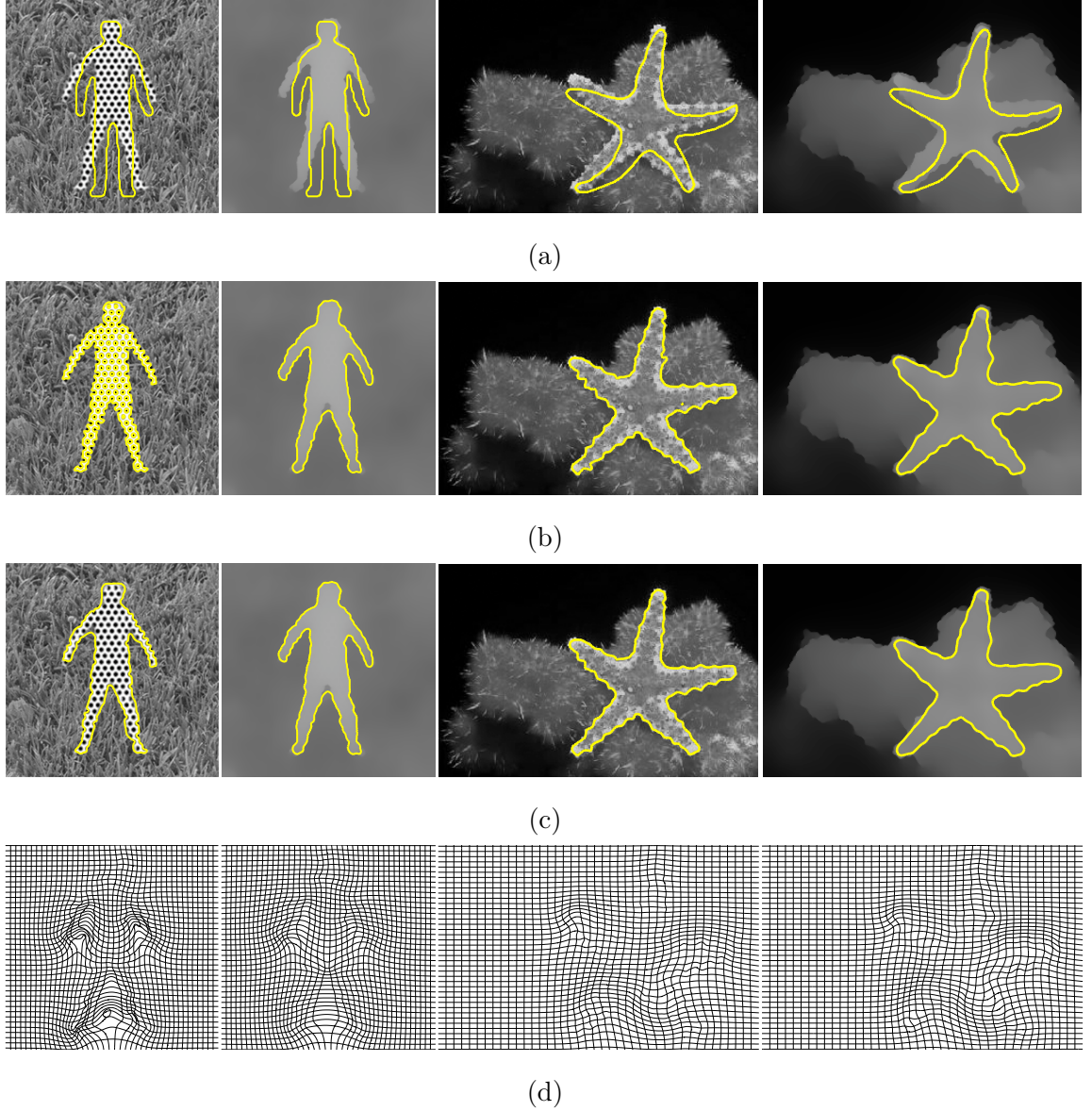
Figure 5.26: Segmenting input and processed images given in Figure 5.25. (a) Unregistered prior shapes superimposed on the images (initial zero-level curves). (b) Segmentation results without prior shape information. (c) Final segmentation results. (d) Deformed grids under the estimated displacement vector fields.

Comparison with the method of Hong et al.

Finally, we compare our prior-guided segmentation framework with the one proposed by Hong et al. [73], which also takes a local deformation model into account as we reviewed in Section 3.2.6. The shape priors that they employed are based on integral kernels (Section 5.1.3), and encode local geometry of shapes in a different way compared to the TSP surfaces. Consider the images shown in Figure 5.28(a) and Figure 5.29(a) that contain a
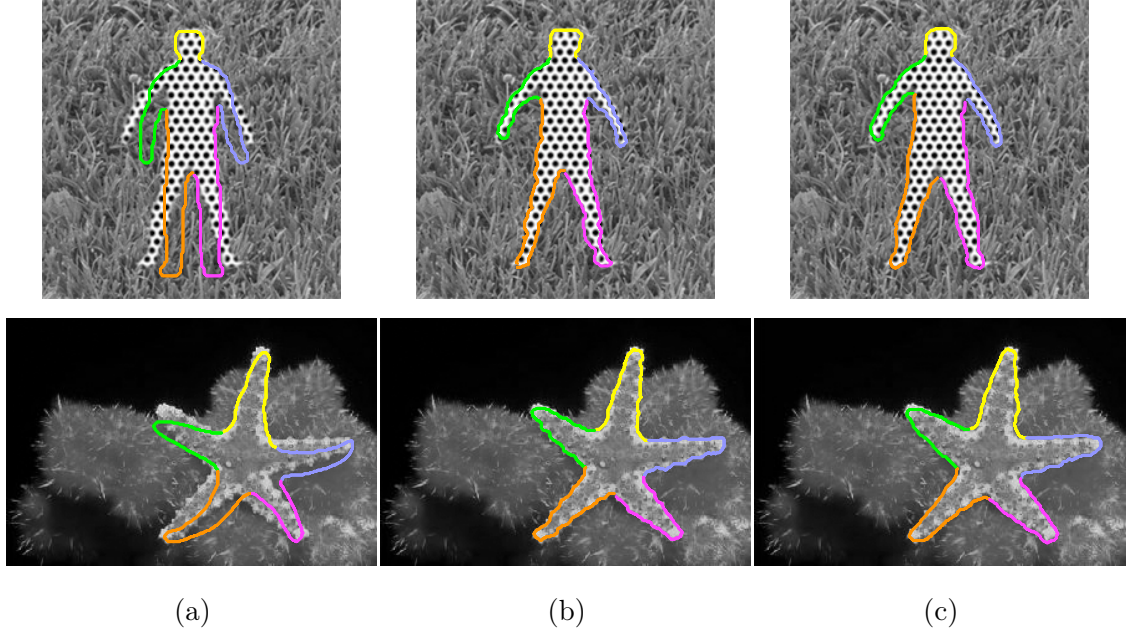
Figure 5.27: Analysis of the registration results (correspondences between the prior shapes and the deformed versions of them are shown in same color). (a) Initial curves. (b) Final curves obtained by taking input images into account. (c) Final curves obtained by taking processed images into account (contours superimposed on input images).

brain MRI and a hand, respectively. While the corpus callosum in the brain image has weak edges due to low contrast, the hand is partially occluded. As it can be clearly seen in Figure 5.28(c)-(d) and Figure 5.29(c)-(d), both methods can cope with these influencing factors and give very similar and good segmentation results. These examples support our previous viewpoint that there is not a unique way to represent the shape priors. Each representation has its own strengths and weaknesses, and choice of representations are highly dependent on applications and computational constraints. Particularly, although integral kernels are more computationally favorable than the TSP surfaces since they do not require solving a PDE, the TSP surfaces provide a much more rich representation in the sense that they additionally allow to perform shape analysis as they can be used to extract skeletons even though we did not explicitly utilize this property in our prior-guided segmentation frameworks.
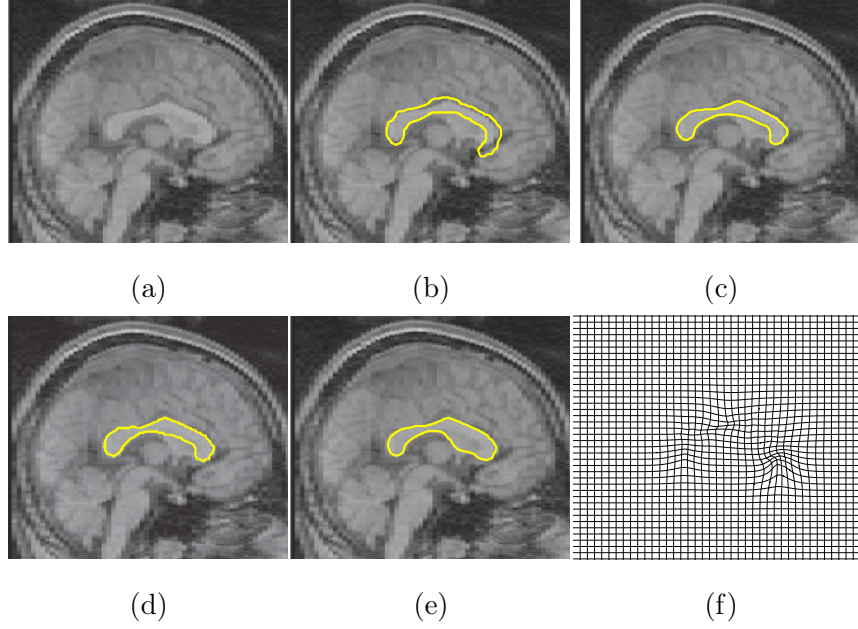
Figure 5.28: Segmenting a brain MRI image having weak edges. (a) Input image taken from [73].(b) Unregistered prior shape superimposed on the input image (initial zero-level curve). (c) Final segmentation result (the parameters are $\mu = 2000$, $\lambda_1 = 1$, $\lambda_2 = 0.25$, $\mu_1 = 1250$ and $\beta = 0.01$). (d) Segmentation result reported in [73]. (e) Segmentation result without prior shape information. (f) Deformed grid under the estimated displacement vector field.

Computational Complexity and Performance

In this section, we provide information about the computational complexity and the performance of the suggested segmentation framework. As we summarized in Algorithm 5, at each iteration, we first update the average gray values of object and background regions, and the evolving the level set function. Next, we compute the TSP surface representing the object region, and estimate the displacement vector field of the local deformation model accordingly. Assuming the resolution of the input image is $M \times N$, the complexity of all these processes is $\mathcal{O}(MN)$ per iteration as we go through every pixel during the computations. Hence, in general, the time required to segment an image depends on the size of that image.

We have perfomed the experiments presented here on a standard PC hardware (2.4GHz Intel Core 2 Duo system with 2GB of RAM) using nonoptimized MATLAB code. As an example, consider the experiment presented in Figure 5.29 where we partition an 365×388 hand image. With the current implementation, it takes 39 iterations and roughly 14 minutes to segment the image. It is important to remark that more efficient implementation schemes
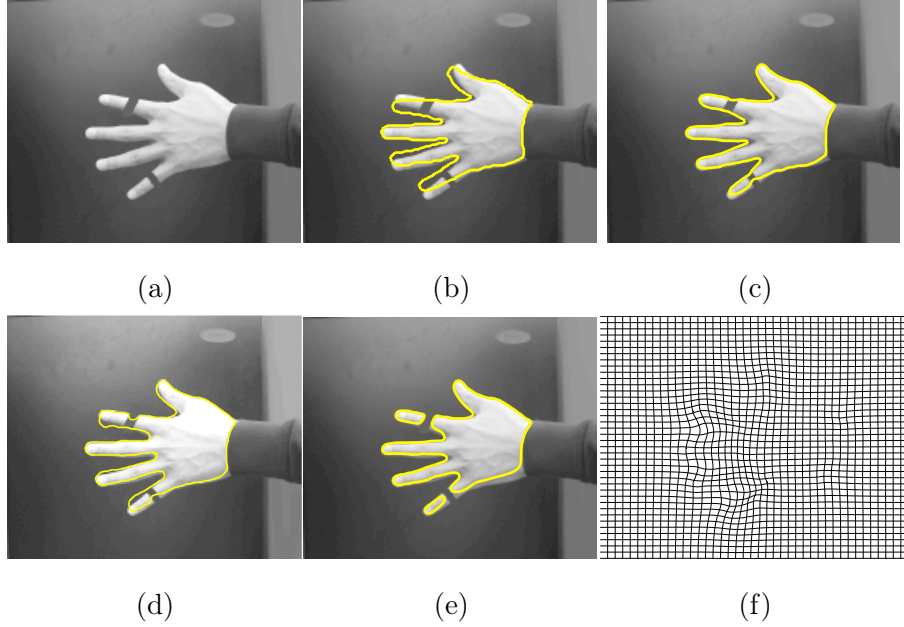
Figure 5.29: Segmenting a partially occluded hand. (a) Input image taken from [73].(b) Unregistered prior shape superimposed on the input image (initial zero-level curve). (c) Final segmentation result (the parameters are $\mu = 20000$, $\mu_1 = 17500$ and $\beta = 0.025$). (d) Segmentation result reported in [73]. (e) Segmentation result without prior shape information. (f) Deformed grid under the estimated displacement vector field.

could be considered to alleviate the computation time. For instance, since the Dirac delta function arises as a factor in the evolution equation of the level set function, the computations can be carried out within a narrow band around the zero level curve of the level set function as in [120], which significantly reduces the computational complexity per iteration. Moreover, one can utilize multigrid methods [25, 27] to more effectively solve the partial differential equations.

5.4   Summary

In many vision applications, extracting accurate object boundaries becomes particularly difficult when the images contain missing or misleading information due to partial occlusions, weak edges and noise. We have presented two new prior-shape based segmentation frameworks that aim to solve this problem. The key to our formulations is the TSP surface representation. In general, in the segmentation framework proposed in Section 5.2, we make use of this surface representation while modeling the global variations of the prior shape and subsequently guiding the segmentation process. On the other hand, our work suggested in

Section 5.3 utilizes TSP surfaces to determine meaningful correspondences between a given template and the evolving object region by taking account of a local deformation model. The deformation between similar local shapes, in return, supervises segmentation by enforcing the zero-level curve of the evolving level set function towards the transformed template.

In our preliminary work which integrates a global shape model with the Mumford-Shah functional, we have simply demonstrated the potential use of TSP surfaces as an aid for incorporating shape priors into image segmentation. Although it can be considered as a straightforward model as compared to our latter work, it deals with the shape variations within an object class and gives adequate results for a set of fairly simple synthetic images. However, one critical issue that must be addressed is that the shape prior obtained by adding linear combinations of the principal components to the mean TSP surface could not preserve the properties of a TSP surface. It may have either negative values or values that exceed 1 at some points even if the empirical results show that such situations are not observed a lot. In this regard, to obtain a more precise global shape model, one can use a nonnegative matrix factorization (e.g. [69, 83]) instead of principal component analysis.

Our latest method discussed in Section 5.3 is more robust than our former work against undesirable conditions such as noise, weak edges and partial occlusions as the experimental results demonstrate. The Chan-Vese model utilized in this work gives a two-phase piece-wise constant approximation of the input image, yielding a clear separation of object and background regions. Moreover, the integrated matching functional that uses a local deformation model handles shape variability better than the global shape model and does only need a single template to represent the shape prior. To illustrate this, in Figure 5.30, we repeat the experiment presented in the bottom row of Figure 5.14 by applying our former work. Although we use a single airplane template (Figure 5.30(b)) in the experiment, the segmentation result obtained is fairly good. The local deformation model transforms the prior shape towards the object boundaries by introducing additional parts in the wings that are not initially a part of the prior shape (Figure 5.30(e)). On the other hand, as it can be clearly seen in Figure 5.30(d), the global shape model in our former work is inadequate in describing these local variations in shape.

Another advantage of our new method is that it partitions an image into regions while simultaneously registering the prior shape on the image data that allows further semantic analysis to be performed on the extracted object. However, since the proposed approach utilizes a local deformation model that relies on linear elasticity, the shapes to be registered should be locally similar. For the objects having highly variable shapes, one need to use reg-
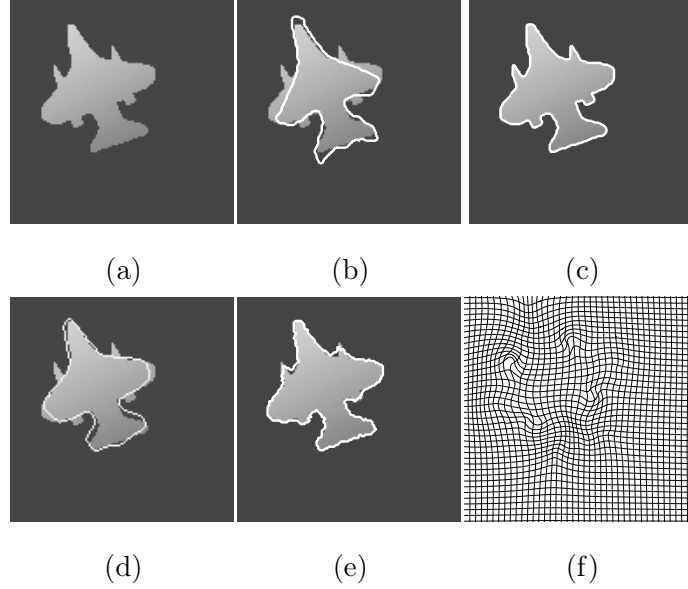
Figure 5.30: Comparison of the proposed segmentation methods. (a) Input image.(b) Unregistered prior shape superimposed on the input image (initial zero-level curve). (c) Final segmentation result (the parameters are $\mu = 3000$, $\mu_1 = 1500$ and $\beta = 0.01$). (d) Segmentation result of our previous work (presented in Figure 5.14). (e) Transformed prior. (f) Deformed grid under the estimated displacement vector field.

istration models that can deal with large deformations. Another drawback of the proposed work is that once the amount of missing or misleading information increases, the local deformation model cannot compensate the insufficient data and results in degraded segmentation results. This suggests that the local deformation model should work jointly with a global shape model to restrict implausible transformations.

As we emphasized before, representation has a critical role in any information processing task. From this perspective, computational frameworks that employ different prior shape representations have been suggested for prior-shape guided segmentation. Generally, it is not very clear how to completely compare the results of these studies as one cannot fully measure what is attained is whether due to the computational scheme or due to the underlying prior shape representation. As a futurework, we want to compare the shape representations previously discussed in Section 5.1 within a segmentation framework similar to our former model to explore their influences on the ongoing process. In this respect, we plan to decompose our former scheme into three coupled processes that respectively correspond to prior-guided segmentation, feature extraction, and local shape matching steps, which allows us to play with the local shape feature utilized in segmentation.

CHAPTER 6

SKELETONS FROM NATURAL IMAGES

In the previous chapter, we suggested two new prior-guided segmentation methods that make use of TSP surfaces as shape priors. In general, the prior shape knowledge helps us to attain more accurate segmentation results by resolving the missing or misleading information in the images. One key issue in the proposed works, as in most of the prior-shape based segmentation frameworks, is that an initial guess about the object of interest is required to initiate the segmentation process.

Finding out such an initial guess is often as difficult as the segmentation itself since it demands an analysis of the input image beforehand. In this chapter, we propose a skeleton based analysis of natural images which operates directly on the raw image with all its complexity. We divide the image into constituent components while simultaneously forming a representation for the components which ties shape and image features.

Our analysis is based on the function v which simultaneously codes the *boundaryness* and *medialness*. The localization of the one deteriorates as the localization of the other improves as a function of the choice of scale and contrast parameters. We show that, with the proper choice of parameters, v function, at a skeleton point, becomes an invertible function of the distance from putative shape boundaries. We use this connection to extract patches of uniform characteristics that can be robustly and easily used in various image analysis tasks. We provide illustrative results using textured images taken from various datasets.

6.1 Related Work

Skeleton extraction has been of interest to computer vision community for almost 40 years. Skeleton based representations have been used in numerous shape recognition methods quite successfully, e.g. [7, 10, 14, 126, 136, 160]. However, the vast majority of the skeleton extraction methods assume that shape is given (see [7, 111, 141] for an extensive list of references). Thus, these methods ignore the difficult problem of image segmentation.

The literature on skeleton-like information extraction from real images are rather limited (even though the earliest attempts go back to seventies [53, 90, 152]) due to ambiguities which are inherent to the processing of real images with all their complexity. One of the earliest attempts is the gray-value thinning [53, 149] by applying thinning to the results of binarization (thresholding) at various gray-values. Another early attempt is the Gradmat [152] method which defines the axis as the ridges of a medialness function. A medialness score is computed for every image point P based on the gradient magnitudes at all pairs of points that have P as their midpoint. The Gradmat method is highly sensitive to noise. However, when the basic idea is combined with regularization as in Cores [32], a robust medialness score can be computed.

There are some recent papers dealing with skeleton extraction from grayscale images [5, 42, 51]. However, they consider strong assumptions on the nature of input images such as piecewise constant images in [5] or existence of a bright object in a dark background in [42, 51]. None of these approaches can handle complicated natural images.

The limited literature on skeleton-like structure extraction from real images can be extended by including ridge extraction methods that operate on grayscale images to extract special image loci [54, 61, 91, 93, 94], watersheds [151], and morphological representations [66, 95, 127].

A theoretically interesting approach, which has not been used in practice, has been presented by Tari, Shah and Pien in [143]. They propose *the edge strength function* which is a regularized form of the regularized image gradient. The edge strength function, which has been constructed via a pair of coupled diffusion equations, is approximately an exponential function of the minimum distance to the putative edges. It has been shown that the medialness of a point is inversely related to the gradient of the edge strength function at that point, and this relation has been exploited to extract skeletons. The method is only applicable to piecewise smooth images with moderate noise, and the extracted skeletons are too complicated to be used in image analysis. One reason that the skeletons are too complicated is the

saddle point instability which has been reported in [6] and solved by using a large enough regularization parameter. However, the suggested solution works only for simply connected domains. Another limitation of the edge strength function is that its value does not give the skeletal radius (minimum distance from a skeleton point to the putative shape boundary). This is an important problem because the recovery of the putative shape boundary from the edge strength function is not trivial [130, 131].

In this chapter, we extend the work of Tari, Shah and Pien (TSP method) such that it becomes applicable to natural images. We first change the pair of coupled equations which are used to compute the edge strength function such that it better reflects the structure of the image regions. Second, we decouple the skeletal radius and the image gradient, and relate the skeletal radii directly to the minimum distance to the putative shape boundary. Using skeletal radii, we extract shape sections of uniform characteristic, e.g. texture, easily. These shape sections can be used in image analysis. We demonstrate our results using images from various datasets [82, 97, 107].

6.2   Computing Medialness From an Image

Our construction of medialness builds on our work on shape preserving filtering discussed in Chapter 4. The following pair of coupled PDEs defines an evolution both for a smooth image u and a function v, given an input image f and the parameters α, β, ρ:

$$\frac{\partial u}{\partial t} \;=\; \nabla \cdot ((cv)^2 \nabla u) - \frac{\beta}{\alpha}(u - f); \quad \left.\frac{\partial u}{\partial n}\right|_{\partial\Omega} = 0 \;, \tag{6.1}$$

$$\frac{\partial v}{\partial t} \;=\; \nabla^2 v - \frac{2\alpha|\nabla u|^2 v}{\rho} - \frac{(v-1)}{\rho^2}; \quad \left.\frac{\partial v}{\partial n}\right|_{\partial\Omega} = 0 \;. \tag{6.2}$$

If we fix $c = 1$, the system of coupled PDEs is reduced to the gradient descent for the Ambrosio-Tortorelli approximation [3] of the Mumford-Shah [100] segmentation functional which has been previously used by Tari, Shah and Pien [143]. Key point is that

$$cv = \phi v + (1 - \phi)V \tag{6.3}$$

where ϕ measures the strength of a feature of interest and is explicitly estimated at each image point. ϕ induces a spatially varying multiplier c for the diffusivity function, providing a higher-level contextual influence. The type of the influence is determined by V, which is either 0 for boosting the feature of interest or 1 for eliminating it. Note that with $\phi = 1$, at all image points, $c = 1$. Thus, the higher-level influence is turned off and the system is reduced to the one in [143].

The parameters α, β, ρ determine the image scale σ (i.e. the smoothing radius of u) and the contrast threshold λ. The image scale σ is given by $\sqrt{\frac{2\alpha}{\beta}}$. The contrast threshold λ is roughly given by $\sqrt{1/2\alpha\rho}$. Notice that the parameter α affects both the image scale and the contrast threshold. Increasing α while keeping both $\sqrt{\frac{2\alpha}{\beta}}$ and ρ constant decreases the contrast threshold, thus leads to an increase in the number of edges.

Notice that if we fix u, (6.2) is a gradient descent for

$$\int_\Omega \left(\frac{\rho^2}{1 + 2\alpha\rho|\nabla u|^2}|\nabla v|^2 + \left(v - \frac{1}{1 + 2\alpha\rho|\nabla u|^2} \right)^2 \right) dx . \tag{6.4}$$

Clearly, v is a blurred form of $\frac{1}{1+2\alpha\rho|\nabla u|^2}$, and the amount of blurring is an increasing function of ρ. Thus, as ρ decreases, the localization of edges improves. Consequently, for small values of the parameter ρ, v function encodes edgeness information.

Note that as in the edge strength function of Tari, Shah and Pien [143], the new v function computed from (6.1)-(6.3) is a nonlinearly smoothed form of a nonlinearly smoothed image gradient. The difference between the former and the later is the way the image is smoothed. We select ϕ such that the smoothing of u will preserve structures rather than noise or texture.

The construction of the medialness function is summarized in Algorithm 6. We start with a small ρ value ($\rho = 0.001$ is used in the experiments). Three different contextual feedback measures ϕ^{dc}, ϕ^h, ϕ^{te} denoting, respectively, the directional consistency, the edge continuity, and the texture edges are used to compute cv in (6.3). The numerical implementation follows the iterative scheme we previously introduced for the context-influenced filtering in Section 4.1.2. A numerical stopping criteria $|u^{k+1} - u^k| < \epsilon_1|u^k|$ is defined on the evolving image u for a small positive constant ϵ_1. This first phase yields a smooth image u^* and a function v which attains its minimum on the putative edges. After the numerical convergence with the first set of parameters, we set $f = u^*$ and rerun the coupled system using a different parameter set. First, we significantly lower both the contrast threshold and the image scale to shutdown the smoothing of u. Consequently, the influence of contextual feedback parameters ε^{dc} in (4.15), h in (4.18), and ε^{te} in (4.19) becomes negligible. Thus, we set all of them to zero. Second, we dramatically increase the value of ρ (we use $\rho = 128$ throughout the experiments). A second numerical stopping criteria $|v^{k+1} - v^k| < \epsilon_2|v^k|$ is defined on the evolving medialness function v for a small positive constant ϵ_2.

In the following, we demonstrate the role played by the parameter ρ which mainly defines the separate phases of our medialness computation procedure. We consider a cube image that is degraded by Gaussian noise with zero mean and non-constant (local) variance along with 2% impulse noise (Figure 6.1). Recall that as $\rho \to 0$, $v \to \frac{1}{1+2\alpha\rho|\nabla u|^2}$, and as ρ increases

Algorithm 6 The suggested iterative scheme for constructing a medialness function from a natural image

---

1: Initialize the variables with $u^0 = f$, $v^0 = \frac{1}{1+2\alpha\rho|\nabla u^0|^2}$

2: Set the contrast and scale parameters, λ and σ

3: Set ρ to a small value (e.g. $\rho = 0.001$)

4: Determine the contextual feedback measure ϕ^{te} [56]

5: **for** $k = 1$ to $kmax$ **do**

6:    Estimate the contextual feedback measures ϕ^{dc} and ϕ^h [57]

7:    Compute the modified diffusivity $(cv)^{k-1}$ by taking account of complementary feedback measures

8:    Solve (4.7) for u^k using $\lfloor\sigma\rfloor$ steps

9:    **if** $|u^k - u^{k-1}| < \epsilon_1|u^{k-1}|$ **then**

10:       stop iterations

11:    **end if**

12:    Solve (4.8) for v^k

13: **end for**

14: Set $f = u^k = u^*$, decrease the level of detail and smoothing by changing σ and λ, and increase the value of ρ (e.g. $\rho = 128$)

15: Turn off the influence of the contextual feedback measures by selecting $\varepsilon^{dc} = h = \varepsilon^{te} = 0$ at all image points, making $c = 1$

16: Solve (4.7) for u^{k+1} using $\lfloor\sigma\rfloor$ steps

17: Solve (4.8) for v^{k+1}

18: **while** $|v^{k+1} - v^k| \geq \epsilon_2|v^k|$ **do**

19:    $k := k + 1$

20:    Solve (4.7) for u^{k+1} using $\lfloor\sigma\rfloor$ steps

21:    Solve (4.8) for v^{k+1}

22: **end while**

---

v becomes blurred. As Figure 6.2 illustrates, the localization of edges deteriorates as ρ increases from 1 to 32. Notice that for $\rho = 32$, the edge localization is very poor, and the v function resembles a distance function. Figure 6.3 illustrates, in color, the v function computed from the noisy cube image with $\rho = 32$ and $\rho = 128$. As the details extracted from v for each surface of the cube depict, for large values of ρ, v function encodes some kind of medialness information, i.e. v is approximately a function of the minimum distance
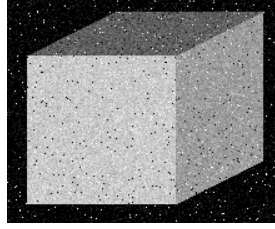
Figure 6.1: A cube image degraded by Gaussian noise with zero mean and non-constant variance along with 2% impulse noise.

| $\rho = 1$ | $\rho = 4$ | $\rho = 8$ | $\rho = 32$ |
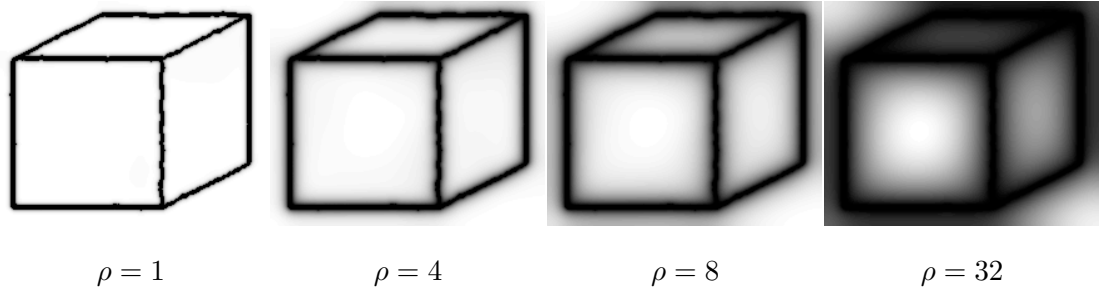
Figure 6.2: Deterioration in the edge localization.

| $\rho = 32$ | $\rho = 128$ |
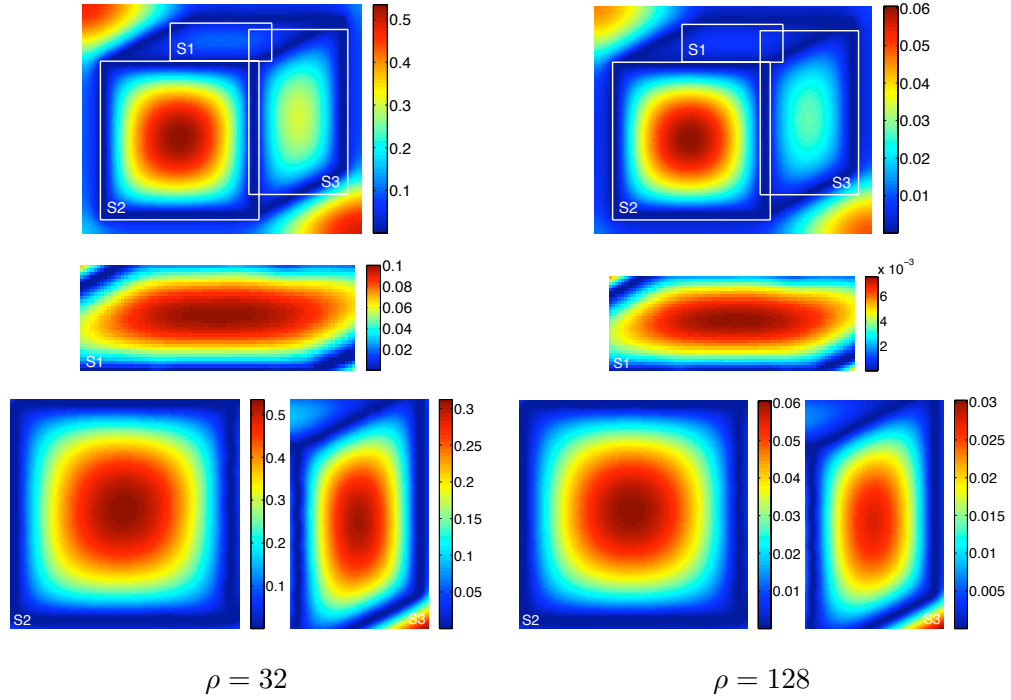
Figure 6.3: v function computed using $\rho = 32$ and $\rho = 128$, and the details extracted from v for each surface of the cube.

to the putative edges. The value of the parameter ρ mainly determines the growth rate of the approximate distance function. The rate of growth increases with the decreasing ρ. Consequently, v function encodes the medialness information more reliably as ρ increases.

Note that just like the function in [143], the new v function is an implicit function of the image gradient rather than the distance to shape boundary. In Section 6.3, we propose an approximation such that v function, at a skeleton point, becomes an invertible function of the minimum distance to the shape boundary.

6.3 From Medialness to Skeletons and Patches

In the following, we first review the skeleton extraction method of Tari, Shah and Pien. Then, we show that with our choice of parameters, i.e very large ρ, the medialness function v at a skeleton point becomes independent of the image gradient. We use this observation to compute the skeletal radius, i.e the distance from a skeleton point to the nearest boundary. The envelope of skeletal circles defines a patch of interest. We demonstrate how these patches can be used in image analysis.

6.3.1 Review of the Skeleton Extraction Method [143]

Skeleton points are determined from the surface v by the points where the surface gradient $|\nabla v|$ is extremum along the level curves of v.

Let S_1 denote the closure of the set of zero-crossings of $\frac{d|\nabla v|}{ds}$

$$\frac{d|\nabla v|}{ds} = \frac{((v_y^2 - v_x^2)v_{xy} - v_x v_y (v_{yy} - v_{xx})}{|\nabla v|^2} \tag{6.5}$$

where s denotes the arclength along the level curves of v. This is approximately equivalent to the curvature extrama. However, as computation of curvature involves second derivatives of v, it is more sensitive to noise than the computation of $|\nabla v|$.

Let $S_1^+ \subset S_1$ and $S_1^- \subset S_1$ be two disjoint set of points such that $S_1 = S_1^+ \cup S_1^-$. This separation can be achieved either using the sign of $\frac{d^2|\nabla v|}{ds^2}$ [143] or a sign test [132]. S_1^+ is approximately equivalent to the points where the curvature of the level curves is positive. Let S_0 denote the set of points where $|\nabla v| = 0$. Points in S_0 correspond to centers (maxima of v) or necks (saddle points of v).

The constructions are illustrated in Figure 6.4 on a binary cat shape. The v functions are computed using the binary formulation given in [143] with two different values of ρ. As it can be seen from the sample level curves of v given in Figure 6.4(b), the level of smoothing of the shape boundary increases with the increasing ρ. Moreover, as ρ increases, the cardinality of points of S_0 decreases. Increasing the value of ρ also results in more coarser skeleton as the less significant branches shrink (Figure 6.4(c)).

(a)

$\rho = 1$ $\rho = 128$

(b)

(c)
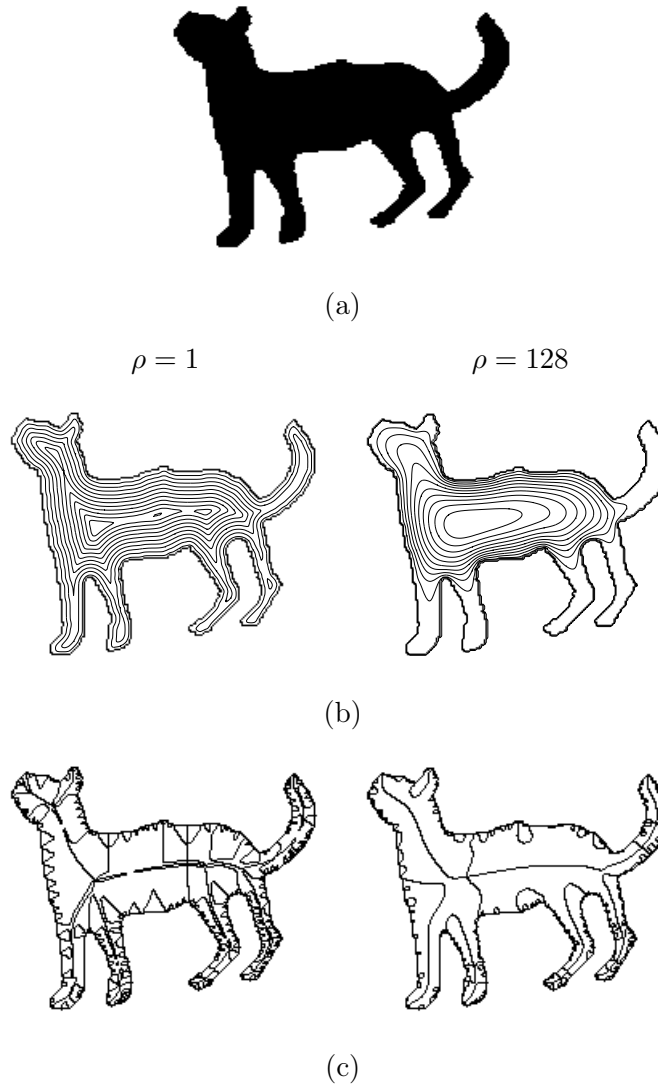
Figure 6.4: Illustrative constructions. (a) A cat shape. (b) The level curves of v for two different ρ values. As ρ increases, $|S_0|$ decreases. (c) S_1 for two different ρ values.

Roughly speaking, S_1^+ tracks the protrusion of the level curves of v, and S_1^- tracks the indentation of the level curves of v. When they join ($\frac{d^2|\nabla v|}{ds^2} = 0$), both branches terminate at a *disconnection point* [8]. The branches of S_1 that do not terminate at a disconnection point terminate at a point in S_0.

6.3.2   Distance From a Skeleton Point to the Nearest Boundary Point

Recall that v function is a nonlinearly smoothed form of $\frac{1}{1+2\alpha\rho|\nabla u|^2}$. Thus, the value of v at a point depends on the gradient $|\nabla u|$ at the point and its neighbors. Consequently, one can not estimate the radius of the skeletal circle (i.e. the distance from a skeleton point to the

134

nearest boundary) using the value of v. Difficulties associated using v function for extracting boundaries has been previously reported in [131].

We propose to replace the direct dependency of v to $|\nabla u|$, just at a skeleton point, with an indirect dependency through an absolute distance $|d|$ such that the function v attains its minimum at a distance $|d|$ from the skeleton point.

We consider a ribbon-like section of a shape, i.e. a section having slowly varying width (this will be justified later in Section 6.3.3), and the skeleton points in that shape section (the dotted line) in Figure 6.5. Note that the boundaries at $|d|$ distance from the skeleton point are the putative edges. Ignoring the curvature effect, the evolution equation is reduced to

$$\left(\frac{d^2}{dx^2} - \frac{2\alpha|u_x|}{\rho} - \frac{1}{\rho^2}\right)v = -\frac{1}{\rho^2}; \quad -d \leq x \leq d \tag{6.6}$$

with the conditions $v(-d) = v(d) \approx 0$.

In (6.6), $|u_x|$ is large at the putative edges. Thus, $\frac{1}{1+2\alpha\rho|\nabla u|^2}$ is practically zero at $x = \pm d$. On the other hand, $|u_x|$ is a small quantity unless $v \approx 0$. The second and third term in the left hand side are, therefore, negligible. The term in the right hand side affects the scaling of the solution. Multiplying the right hand side by ρ^2 does not make a qualitative difference. Consequently, the behavior is roughly governed by

$$\frac{d^2v}{dx^2} = -\frac{1}{\rho^2}; \quad -d \leq x \leq d \tag{6.7}$$

Therefore,

$$v(0) = \frac{d^2}{2\rho^2} . \tag{6.8}$$

Consequently, the radius r of the maximal circle is given by

$$r = d = \rho\sqrt{2v} . \tag{6.9}$$

In Figure 6.6, we illustrate the sample skeletal circles with the computed radii according
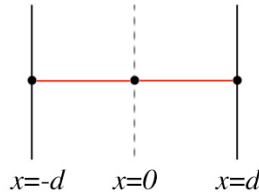
Figure 6.5: An illustration of a ribbon-like section of a shape and its skeleton (the dotted line).

135
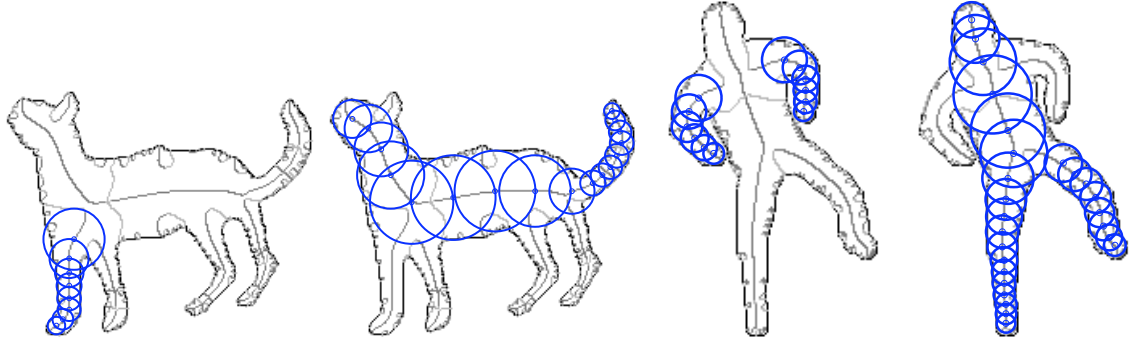
Figure 6.6: Sample skeletal circles with radii computed using (6.9).

to (6.9). Notice that the envelopes of the circles are quite in agreement with the corresponding ribbon-like shape sections.

6.3.3 Optional Pruning Strategies

As ρ increases, the skeleton points code less and less the noise and the secondary details (i.e. such branches get shorter and shorter), and the skeleton branches terminate at the end of ribbon-like sections of a shape. In [6], Aslan proposed a straightforward solution to eliminate the irrelevant branches based on a length criterion.

Ideally, in a symmetric ribbon-like portion of a shape, v is symmetric about the geometric axis. The positive skeleton points (S_1^+) coincide with the geometric axis, and are orthogonal to the level curves of v. As we approach to a disconnection point (i.e. the join of S_1^+ and S_1^-), the behavior of S_1^+ deviates from the ideal. Eventually the level curves of $\frac{d|\nabla v|}{ds}$ becomes tangent to a level curve of v at the disconnection point (Figure 6.7(e)-(f)). Shah [133] proposed the angle, ψ, between the vectors ∇v and $\nabla \left(\frac{d|\nabla v|}{ds} \right)$ as a measure of the deviation from the ideal that could be used to prune inaccurate branches.

As the skeleton points have ψ values close to $90°$ along the ribbon-like shape sections, considering a fairly high threshold does preserve only the branches of the skeleton identifying the main protrusions. On the other hand, setting the threshold too low fails to eliminate the branches related to noise or insignificant shape details. Shah suggested to use a two-step pruning procedure to obtain more accurate skeleton representations. This procedure requires two threshold values $\overline{\theta}$ and $\underline{\theta}$ for the angle ψ with $\overline{\theta} > \underline{\theta}$. In the first step, the skeleton points are pruned by making use of $\overline{\theta}$, which yields the significant protrusion branches representing the ribbon-like shape sections. Next, in the second step, the branches of the thresholded skeleton are traced in the direction of increasing v and extended by the skeleton points

136

ignored by the initial step as long as ψ remains greater than $\underline{\theta}$.

In Figure 6.7, we present sample pruning results obtained by the suggested algorithm. Figure 6.7(a) illustrates the unpruned skeleton branches. In Figure 6.7(b) and (c), we respectively use $\overline{\theta} = 60°$ and $75°$, and $\underline{\theta} = 10°$ to prune the skeleton branches extracted. As it can be clearly seen from these pruning results, the level of pruning increases with the increasing value of the thresholding angle $\overline{\theta}$. The enlarged pruning result given in Figure 6.7(d) illustrates the deviation in the ideal behavior of the protrusion branches (S_1^+) as we approach to a disconnection point. The pruned segments of the protrusion branches which do not represent the ribbon-like shape sections are shown in thick red lines.

The pruning strategy discussed above as a simpler alternative to Shah's former strategy [132] can be utilized to further justify the approximation we proposed in (6.8).

6.4 Experimental Results and Discussion

The first set of experiments is performed using the images presented in Figure 6.8 with the parameters listed in Table 6.1. The common feature of these images is that they all contain a single object which is distinguishable from the surroundings by means of intensity or texture difference.

Figure 6.9-Figure 6.11 illustrate the outcome of the proposed skeleton extraction procedure. First two rows display u and $1 - v$ after the first phase with small ρ. Notice that $1 - v$ is an edge map. The third row displays the level curves of v after the second phase. Final v accurately encodes the successive level curves of a smoothed distance function. Finally, the last row depicts the extracted skeleton points. Notice that even the function v after the

Table 6.1: Parameter sets used during the first and the second phases of the medialness computation for the images given in Figure 6.8

Images	$\rho = 0.001 \quad (\epsilon_1 = 5 \times 10^{-4})$		$\rho = 128 \quad (\epsilon_2 = 1 \times 10^{-5})$	
cheetah	$\lambda = 1.58$, $\sigma = 89.44$, $\varepsilon^{dc} = 0.25$, $\varepsilon^{te} = 20$	$\lambda = 0.04$		
cheetah2	$\lambda = 0.33$, $\sigma = 244.95$, $\varepsilon^{dc} = 0.25$, $\varepsilon^{te} = 100$	$\lambda = 0.04$		
bear	$\lambda = 3.95$, $\sigma = 46.19$, $\varepsilon^{dc} = 0.1$, $\varepsilon^{te} = 20$	$\lambda = 0.14$		
bear2	$\lambda = 1.58$, $\sigma = 141.42$, $\varepsilon^{dc} = 0.25$, $\varepsilon^{te} = 200$	$\lambda = 0.04$	$\sigma = 20$, $\varepsilon^{dc} = h = \varepsilon^{te} = 0$	
elephant	$\lambda = 3.53$, $\sigma = 20$, $\varepsilon^{dc} = 0.1$, $\varepsilon^{te} = 10$	$\lambda = 0.14$		
rhino	$\lambda = 2.24$, $\sigma = 44.72$, $\varepsilon^{dc} = 0.25$, $\varepsilon^{te} = 2$	$\lambda = 0.14$		
dog	$\lambda = 5.00$, $\sigma = 28.28$, $\varepsilon^{dc} = 0.25$, $\varepsilon^{te} = 10$	$\lambda = 0.14$		
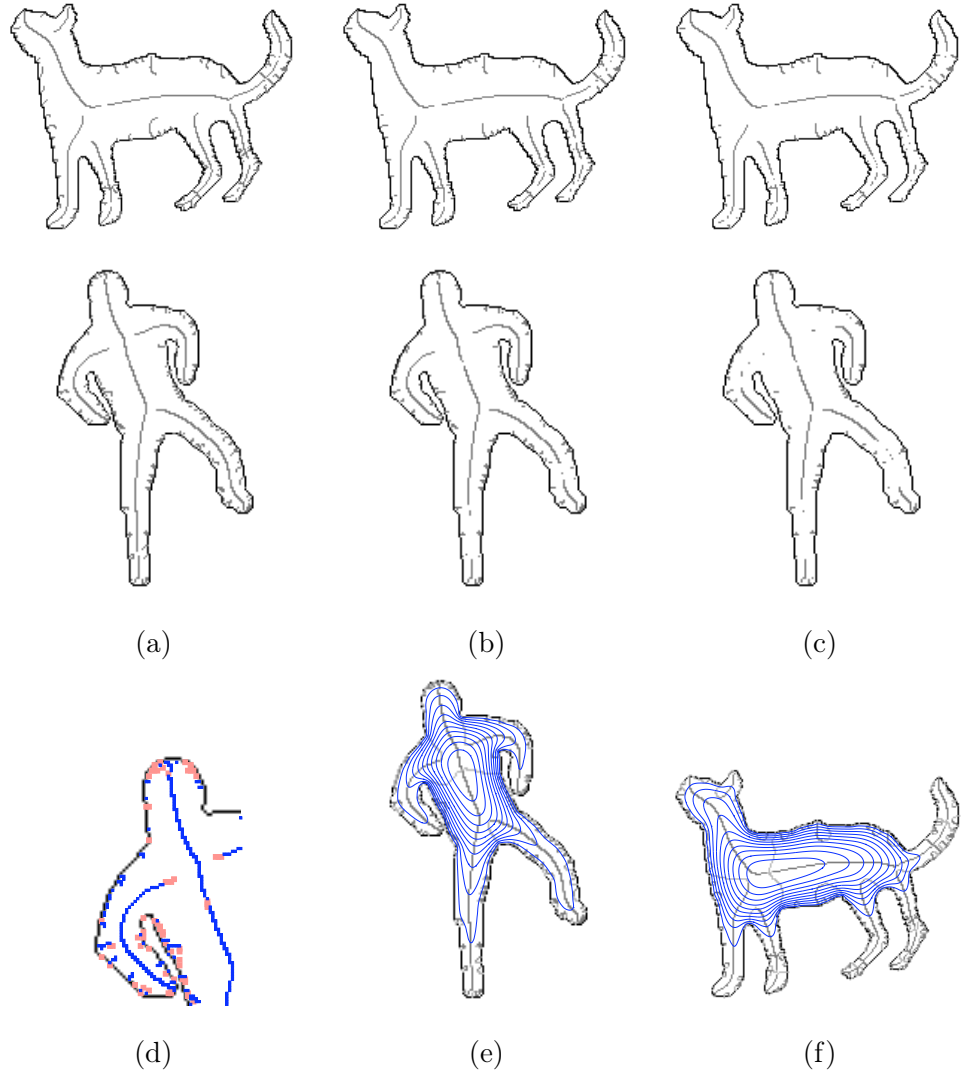
Figure 6.7: Examples of the pruning strategy suggested in [133]. (a) Original unpruned skeletons. (b)-(c) Sample pruning results with $\overline{\theta} = 60°$ and $\overline{\theta} = 75°$, respectively. (d) An enlarged pruning result from (b) where the pruned segments of the protrusion branches are shown in thick red lines. (e)-(f) Level curves of v superimposed on the skeleton branches. Notice that as we approach to a disconnection point, the level curves of $\frac{d|\nabla v|}{ds}$ becomes tangent to a level curve of v.

first phase contains unintuitive or double edges (Figure 6.9-*cheetah2*), or missing contours (Figure 6.10-*bear2*), skeletons can still be computed.

Each protrusion branch (S_1^+) emanates from a boundary protrusion and ends at an interior point. The v-value at a point on a protrusion branch (to an extent discussed in Section 6.3.2) is an invertible function of the distance to the nearest boundary point (6.8). In Figure 6.12, we concentrate on isolated protrusion branches and depict the recovery of shape

cheetah   cheetah2   bear   bear2
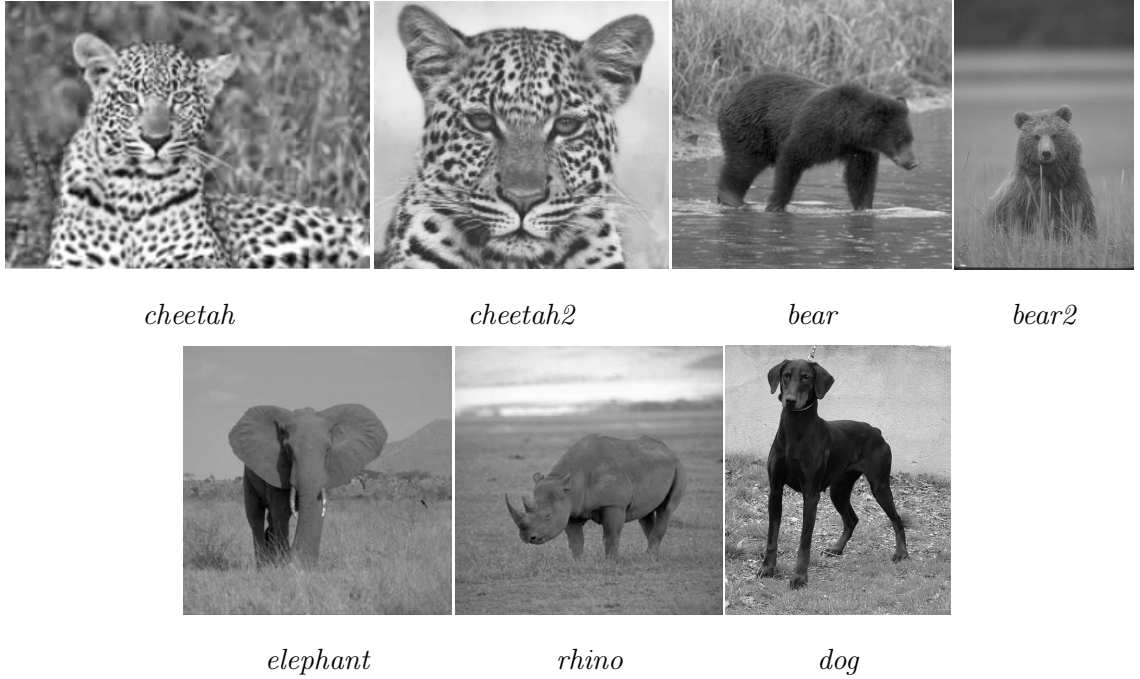
elephant   rhino   dog

Figure 6.8: Illustrative natural images used in the experiments (*bear2* image is taken from *Berkeley Segmentation Dataset* [97] and *bear, elephant, rhino* images are from [107]).

sections from the selected skeleton branches. Specifically, Figure 6.12(a) shows the plots of the skeletal radii for the selected protrusion branches extracted from the natural images presented in Figure 6.8. It can be seen from the corresponding skeletal circles presented in Figure 6.12(b) that the protrusion branches identify particular object sections. While for the *cheetah, cheetah2, bear2* and *elephant* images, the identified sections correspond to ears, the protrusion branch encodes one of the legs for the *bear* image. On the other hand, for the *rhino* and *dog* images, the selected branches respectively specify the whole main body, and the main body along with the rightmost leg of the animals. As we obtain the skeletal radii by considering a ribbon-like portion of a shape, the estimated radii becomes unreliable for the locations where a protrusion branch loses its association to a ribbon-like structure. Especially, the v values and the values of the associated skeletal radii along such locations becomes larger than expected due to the nearby indentations. This is clearly visible in the protrusion branches specified for the *cheetah, cheetah2,* and *bear* images.

Figure 6.13 illustrates the structured image regions identified by the protrusion branches using three different strategies. In Figure 6.13(a), we show the regions determined by employing the skeletal radius formula in (6.9) derived considering the ribbon-like portions of a shape. For the *cheetah, cheetah2,* and *bear* images, this process results in skeletal circles
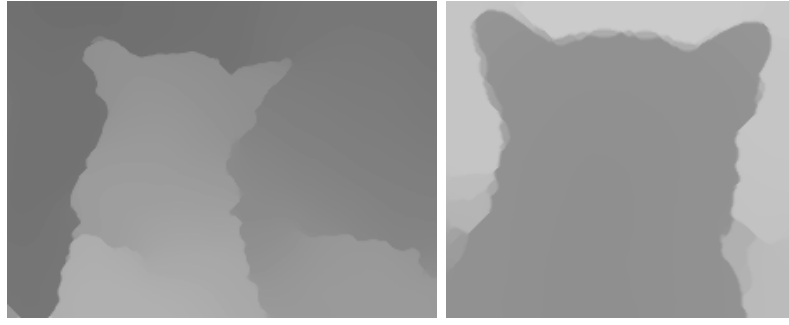
(a)
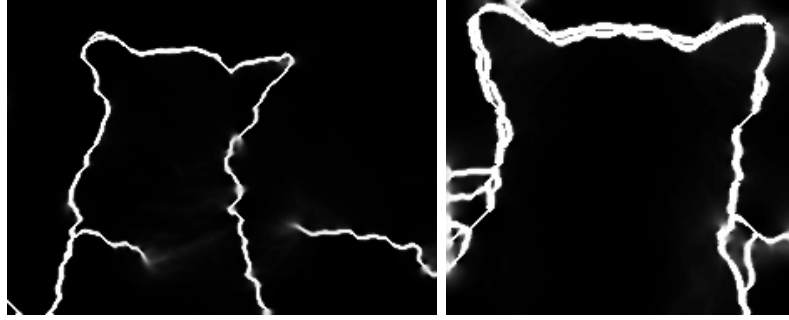
(b)
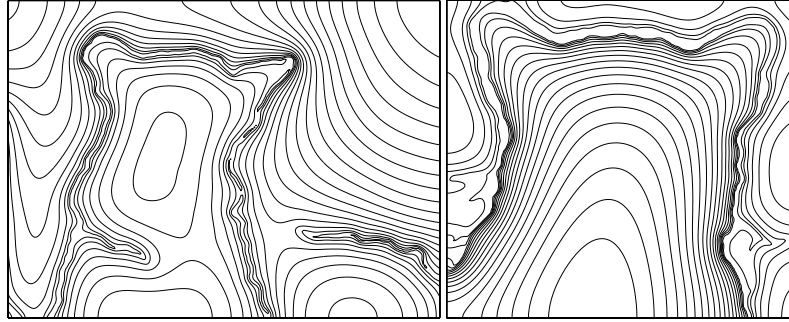
(c)
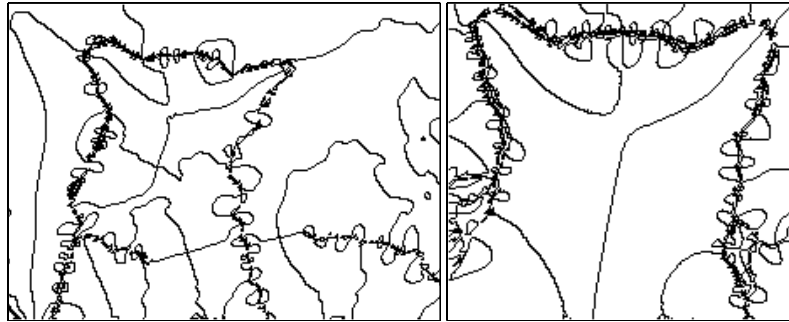
(d)

Figure 6.9: *cheetah* and *cheetah2* images. (a)-(b) u and $1 - v$ after the first phase with $\rho = 0.001$. (c) Level curves of v after the second phase with $\rho = 128$. (d) Skeletons extracted.
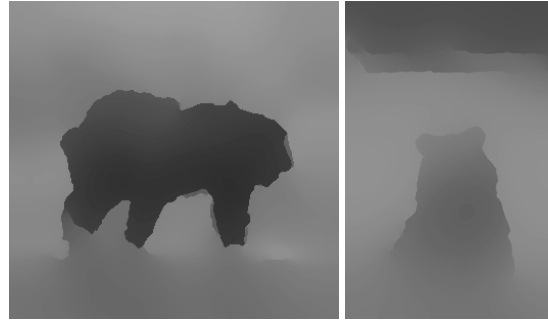
(a)

(b)
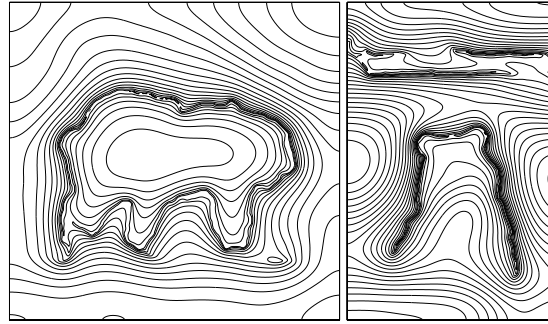
(c)
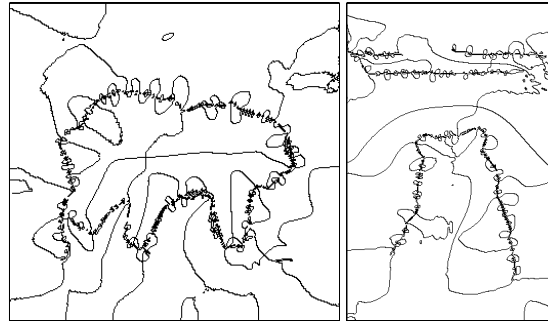
(d)

Figure 6.10: *bear* and *bear2* images. (a)-(b) u and $1 - v$ after the first phase with $\rho = 0.001$. (c) Level curves of v after the second phase with $\rho = 128$. (d) Skeletons extracted.
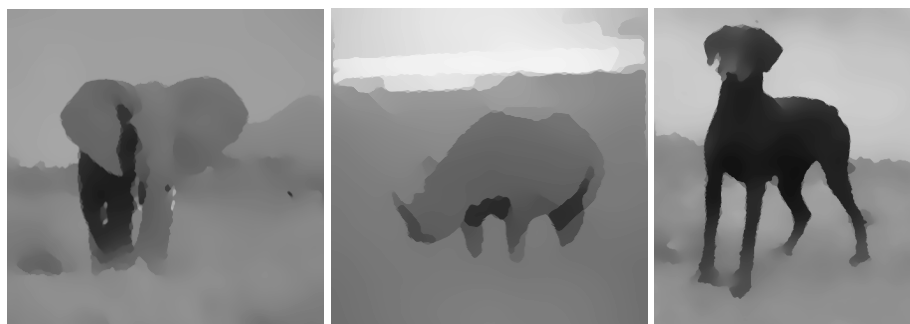
(a)
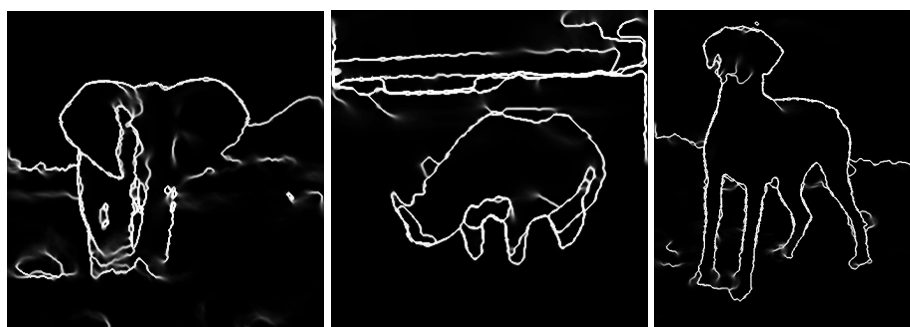
(b)
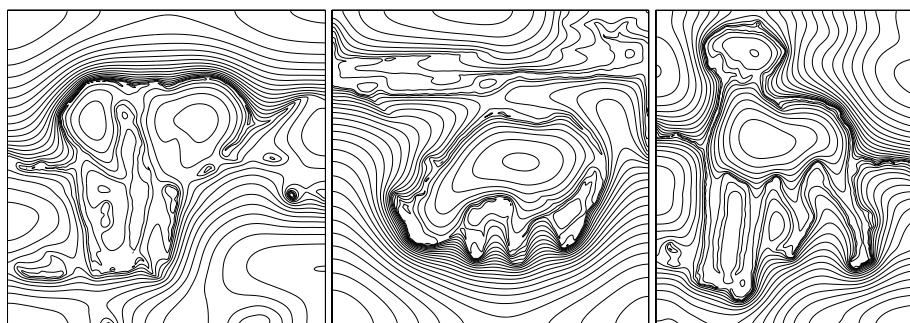
(c)
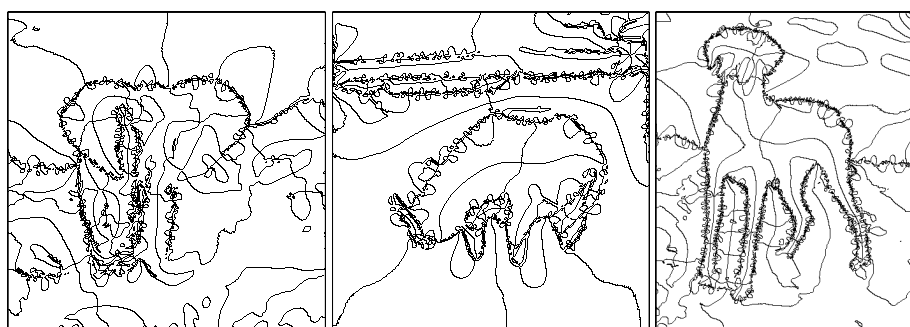
(d)

Figure 6.11: *elephant*, *rhino*, and *dog* images. (a)-(b) u and $1 - v$ after the first phase with $\rho = 0.001$. (c) Level curves of v after the second phase with $\rho = 128$. (d) Skeletons extracted.

(a)

(b)
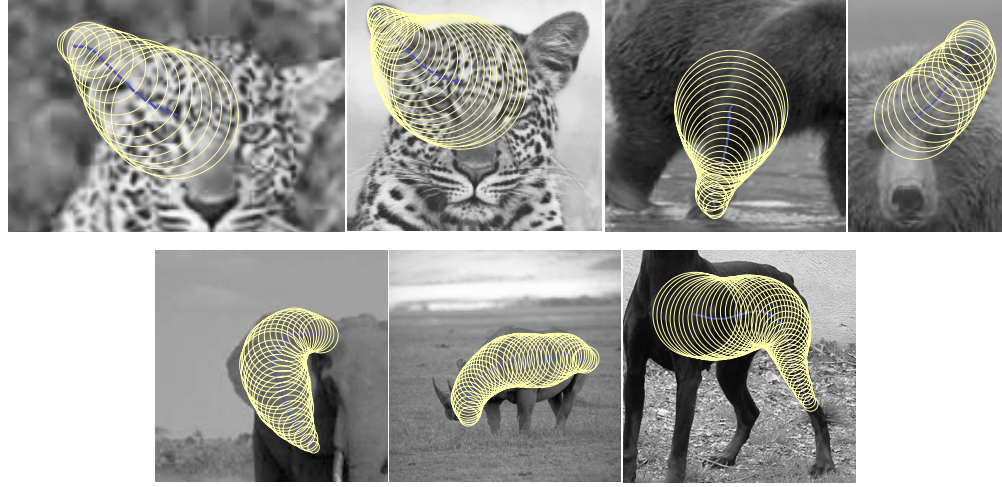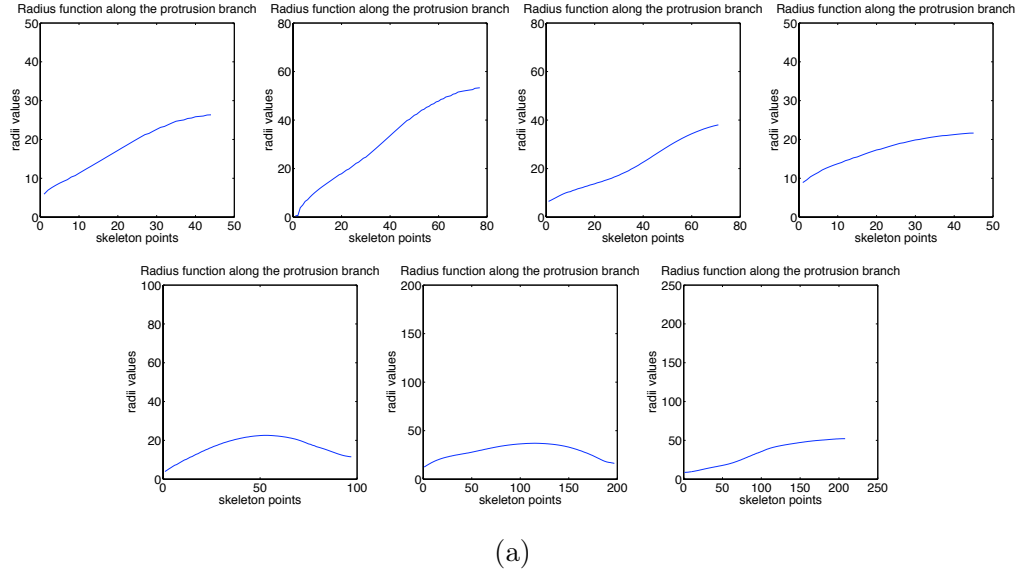
Figure 6.12: Analysis of isolated protrusion branches. (a) Plots of the skeletal radii starting from the boundary protrusions. (b) Circles associated with the skeleton points along the protrusion branches.

larger than expected, especially as the corresponding skeleton points approach to a discon-nection point (i.e. the join of S_1^+ and S_1^-). When we use the pruning algorithm proposed by Shah in [133] with $\overline{\theta} = 80°$ and $\underline{\theta} = 10°$, we can obtain more reliable results since it preserves the major ribbon-like shape sections. However, the extracted regions still contain pixels be-longing to the background as illustrated in Figure 6.13(b). Neglecting the multiplier $\sqrt{2}$ in (6.9), on the other hand, gives more conservative radii and yields smaller circles to be asso-ciated with skeleton points (Figure 6.13(c)). This allows us to extract more accurate feature statistics from the estimated image regions. Notice that, even the scale of the *cheetah* and

(a)

(b)

(c)
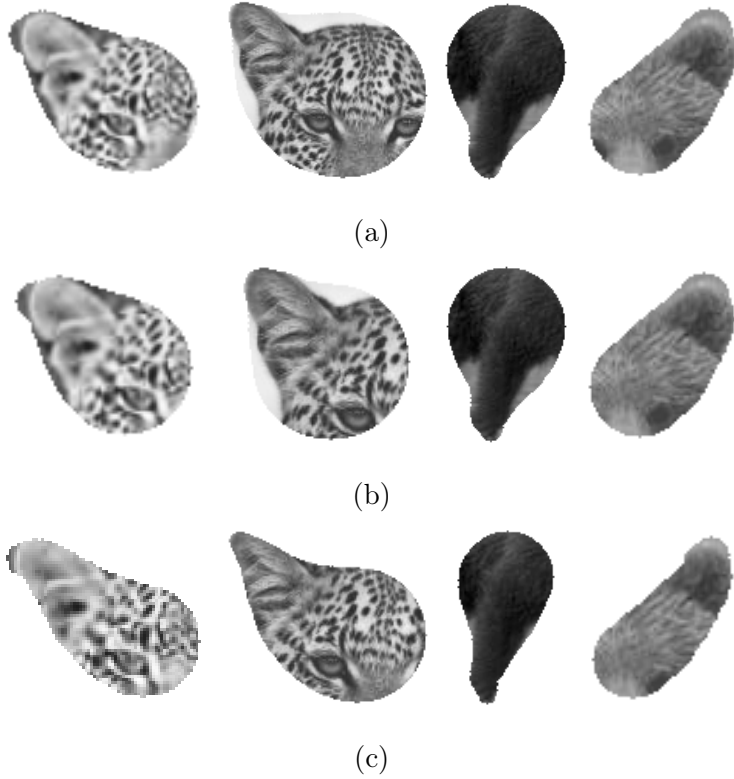
Figure 6.13: Sample structured image regions extracted with (a) the skeletal radii computed using (6.9), (b) the same radii in (a) but considering the pruned skeleton branches, and (c) the more conservative radii computed neglecting $\sqrt{2}$ in (6.9).

cheetah2 images are different, the extracted regions of interest show similar characteristics within each strategy by means of shape and appearance.

In the following, we concentrate on the structured image regions identified by the conservative skeletal radii (Figure 6.14(a)), and show what can be further inferred from these regions of interest. As shown in Figure 6.14(b) and (c), these regions can be used to obtain approximate object boundary fragments. The curves presented in Figure 6.14(b) are estimated using a simple thresholding mechanism which mainly inspects the minimum v-values around the boundary of the skeletal circles and discards the circles that are far from the object boundaries. These circles are shown in dashed lines in the figure. The resulting boundary fragments extracted in this way provide some hypotheses regarding the object segmentation.

In Figure 6.12-Figure 6.14, our focus is on the analysis of local regions encoded by isolated protrusion branches. However, sometimes, the local regions identified by isolated branches do not express meaningful shape sections. In Figure 6.15 and Figure 6.16, we illustrate

(a)

(b)

(c)
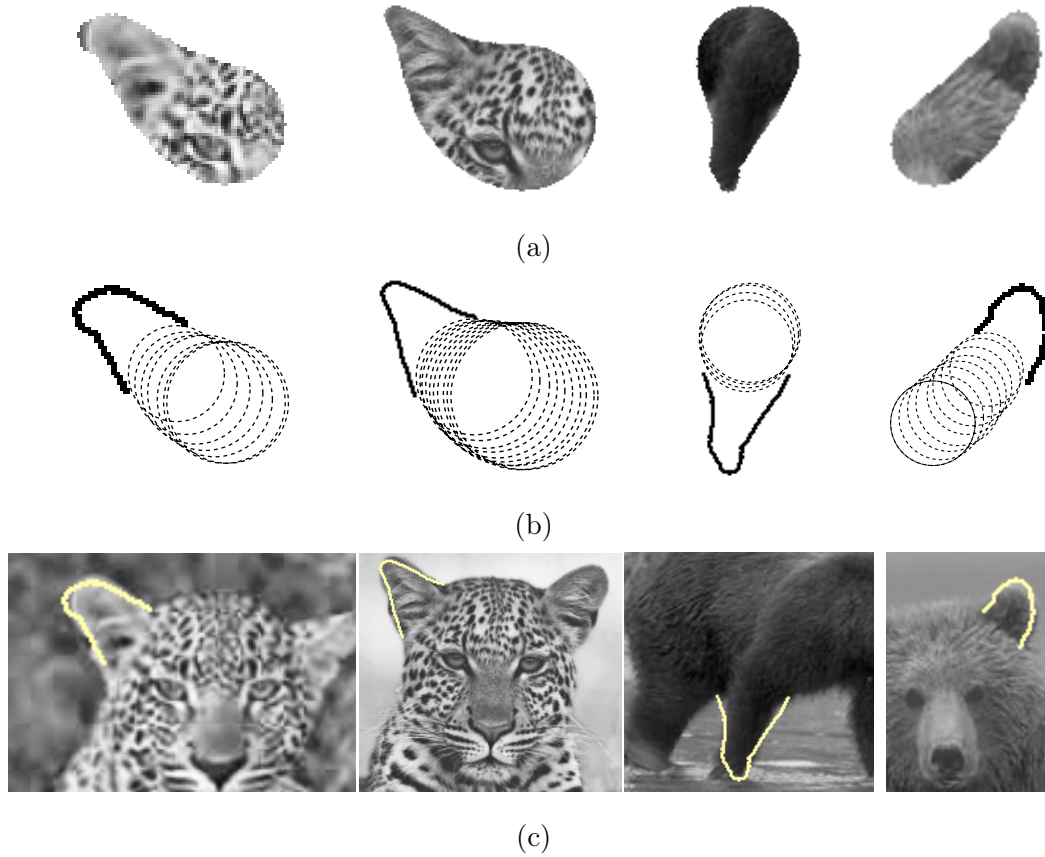
Figure 6.14: Analysis on the structured image regions extracted using the conservative skeletal radii. (a) Regions of interest. (b) Boundary extraction process (see text). (c) Approximate object boundary fragments inferred from (b).

two such cases. Mainly, in these images, we successively group the protrusion branches to obtain more accurate regions of interest. For each figure, we give the results of the region extraction process considering both the radii computed using (6.9), and the more conservative radii that we get by neglecting the multiplier $\sqrt{2}$ in (6.9). While Figure 6.15 is about the analysis of the region corresponding to the left ear of *elephant*, Figure 6.16 concentrates on the main body of *rhino*. We estimate the initial approximations of these regions using the branches that terminate at elliptic points in S_0 which can be regarded as region centers. As shown in the topmost left images of Figure 6.15 and Figure 6.16, these branches represent the most prominent features of the regions and give general ideas about the overall shape. The approximations can be further improved by successively adding the nearby protrusion branches to the computations. A similar procedure can be also employed for the cases where we could not attain a segregation of foreground object and background, e.g. the legs of *rhino* in Figure 6.11.

145

(a)   (b)
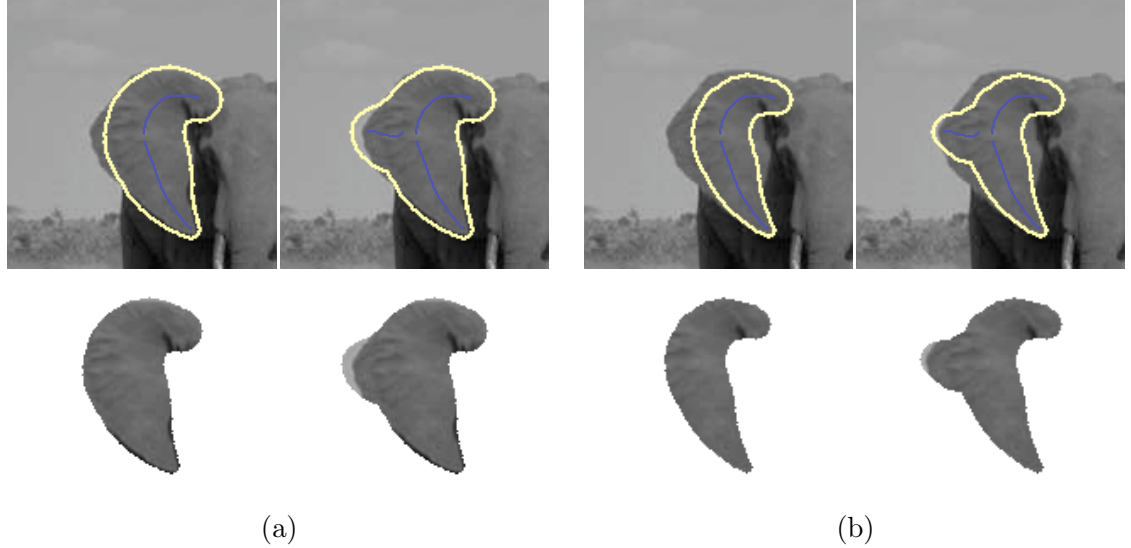
Figure 6.15: Successive approximation of the structured image region corresponding to the ear of *elephant* by considering (a) the skeletal radii computed using (6.9), and (b) the more conservative radii computed neglecting $\sqrt{2}$ in (6.9).

Figure 6.17 illustrates an example where the extracted skeleton has the TSP instabilities reported in [6, 7]. The details from the skeleton presented in Figure 6.17(a) clearly depicts the instability: One of the prominent branches expressing the main body and the rightmost leg is broken into several parts at the saddle points, introducing unintuitive branches. Hence, the initial approximation of the region encoded by the longest part does not fully span the rightmost leg. It possible to solve this problem by tracing the nearby skeleton points and by appending the ones belonging to broken protrusion branches to the initial part of the branch. When such a strategy is employed, a more accurate approximation can be obtained. This successive approximation process is illustrated in Figure 6.17(b) and (c), respectively considering the skeletal radius in (6.9) and the conservative one.

In the following, we show region extraction results on a set of bird images given in Figure 6.18 that are obtained by using the parameters specified in Table 6.2. The local image regions are again defined by the skeletons obtained by the proposed procedure considering the conservative radii. Here, our analysis is more local than the previous examples as we let the images partition into many fairly small regions. For this, we employ only the edge continuity and the directional consistency measures in the first phase which does not yield a complete segregation of the foreground objects. In the second phase where we increase the value of ρ to a sufficiently large value, we impose homogeneous Dirichlet boundary conditions ($v = 0$) along $\partial\Omega$, making the skeleton branches emanate from the image boundaries. We
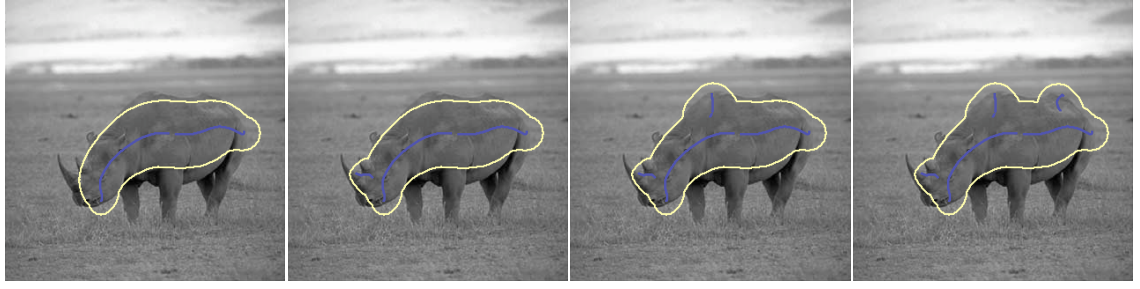
(a)

(b)

Figure 6.16: Successive approximation of the structured image region corresponding to the main body of *rhino* by considering (a) the skeletal radii computed using (6.9), and (b) the more conservative radii computed neglecting $\sqrt{2}$ in (6.9).

apply the pruning method suggested in [6] to eliminate short skeleton branches.

In Figure 6.19, the first row displays u after the first phase with small ρ. The second row displays v after the first phase and the extracted skeletons that are superimposed on it after pruning. In Figure 6.19(c), we display local image patches defined by the envelopes of the

(a)

(b)   (c)
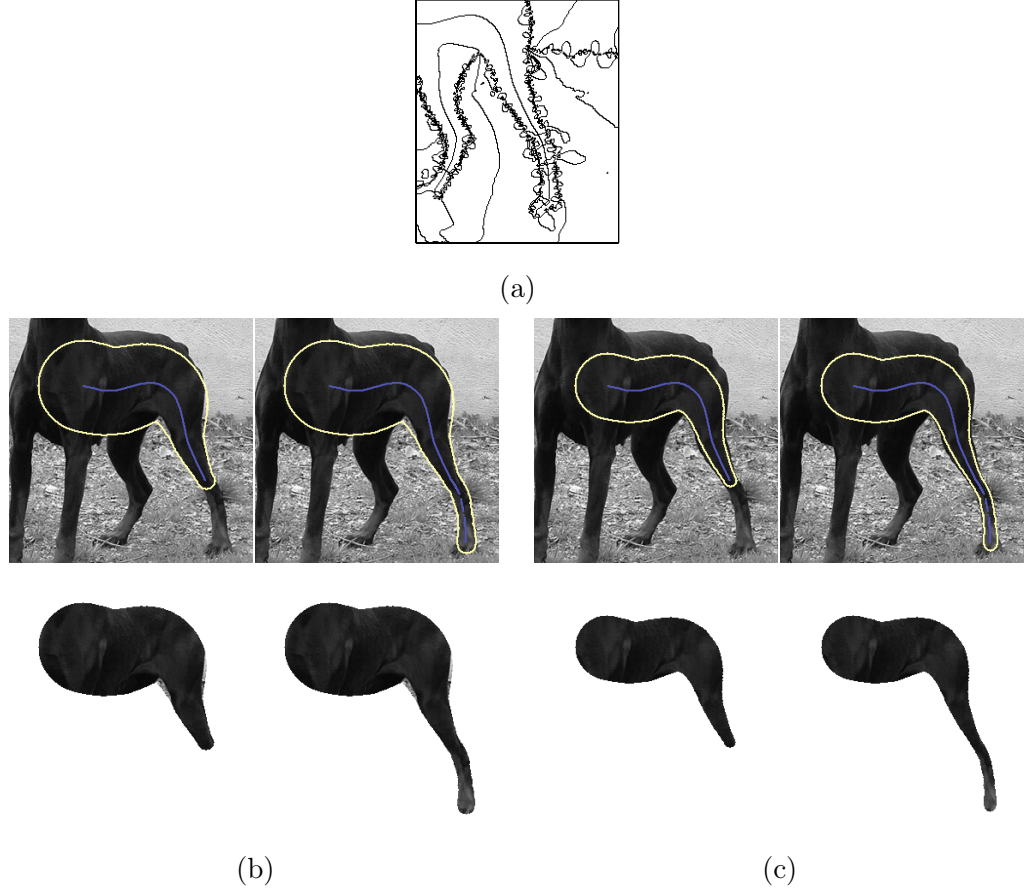
Figure 6.17: TSP instability problem [6, 7]. (a) Details from the extracted skeleton where the TSP instability occurs. (b)-(c) The region of interest identified by the protrusion branch of subject of interest, and its successive approximation obtained by appending the nearby skeleton points to the initial part of the branch (for two choices of the skeletal radii – the one derived considering ribbon-like shape portions and the more conservative one).

Table 6.2: Parameter sets used during the first and the second phases of the medialness computation for the images given in Figure 6.18

Images	$\rho = 0.001$ $(\epsilon_1 = 5 \times 10^{-4})$		$\rho = 128$ $(\epsilon_2 = 1 \times 10^{-6})$
duck1	$\lambda = 3.53$	$\sigma = 28.28,\ \varepsilon^{dc} = 0.25$	$\lambda = 0.20,\ \sigma = 1.41,\ \varepsilon^{dc} = h = 0$
duck2			
duck3	$\lambda = 5.00$		

skeletal circles associated with the extracted skeleton branches. As it can be clearly seen, these regions cover a significant portion of the images even when we use a pruning step.

Finally, we demonstrate an illustrative image analysis procedure where we perform simple foreground/background separation. We start with the local image regions presented in
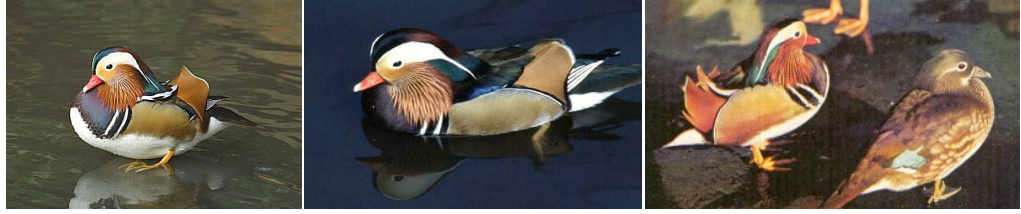
duck1   duck2   duck3

Figure 6.18: Bird images taken from [82].

(a)
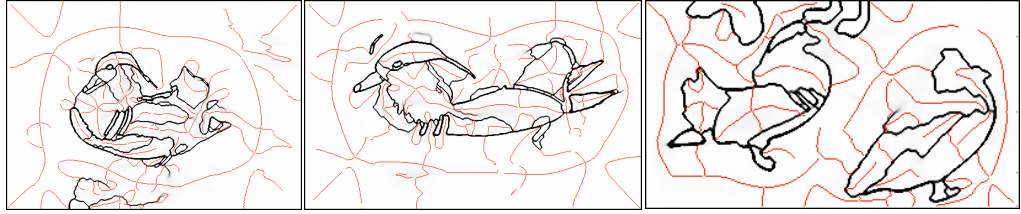
(b)

(c)

Figure 6.19: Region extraction results.(a) u after the first phase with small ρ. (b) v after the first phase with small ρ and the skeletons extracted (superimposed after pruning). (c) Local image regions defined by the skeletons extracted.

Figure 6.19(c). The extracted patches are of uniform characteristic that code specific sections of the image. We first describe each structured patch within each image by means of simple texture features (mean intensity, standard deviation, entropy, homogeneity of the co-occurrence matrix). Next, we perform k-means clustering on these features with $k = 4$ for each image. As the images all contain birds which are distinguishable from the surroundings by means of intensity or texture difference, one of the estimated clusters includes the regions that correspond to background. In Figure 6.20(a)-(b), we display the outcome of this

procedure. Even though we describe the structured patches using simple texture features, they provide a rough, but fairly good separation of the foreground from the background. Now that a rough foreground/background separation is available, we can perform a thorough analysis and refine the separation results. The main idea of this step is to restrict the attention to the initially estimated foreground portion of the image during re-estimation of the separation. We first compute the convex hull enclosing the initial foreground region, and use it in the second clustering phase. We describe each pixel in the convex hull by means of the entropy value calculated within a squared window centering the pixel (ignoring the pixels outside the convex hull). Next, we perform k-means clustering with $k = 2$ to classify the pixels into two as foreground and background. In Figure 6.20(c)-(d), we present the refined separation for the rightmost bird in the *duck3* image (11×11 square image patches are used in computing the entropy, and the holes inside the region defined by the foreground pixels are filled). Clearly, the final result is much more accurate than the initial separation result. Lastly, in Figure 6.20(e), we illustrate the outcome of an image inpainting method [119] where the estimated foreground region is used as a mask to remove the bird from the image (the image is resized to 50% of its original size before applying inpainting).

6.5   Conclusion

Skeleton based representations have been used in shape recognition methods quite successfully. However, their applicability to real images has been neglected with the exception of few works. The main reason is the ambiguities which are inherent to the processing of real images. In this paper, we have presented a skeleton based analysis which operates directly on a raw natural image with all its complexity.

Our analysis is based on a function v which simultaneously codes medialness and boundaryness. The accuracy in the localization of the former improves as the accuracy in the localization of the later deteriorates as a function of the parameter ρ. The v function is a nonlinearly smoothed form of the image gradient which is itself selectively smoothed so as to preserve shapes rather than texture and noise.

An important practical contribution is that, with the proper choice of parameters, v function becomes independent of the image gradient. We propose an approximation to v function on the skeleton points such that it becomes a function of the minimum distance to putative shape boundaries. The envelope of skeletal circles associated with the skeleton points defines a patch of interest of uniform characteristic. Consequently, we can divide the

(a)

(b)

(c)   (d)   (e)
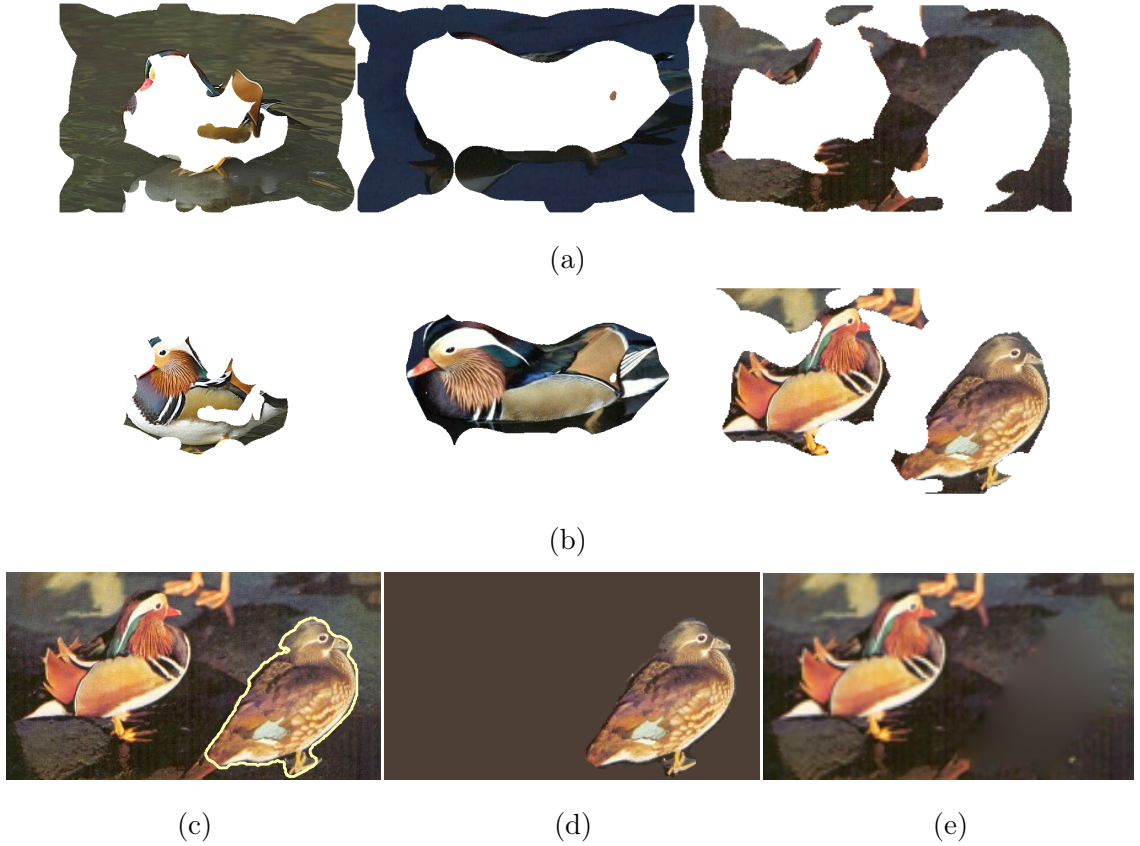
Figure 6.20: Foreground/background separation. (a)-(b) Background/foreground regions estimated by the first clustering phase. (c)-(d) The refined separation result for the rightmost bird in the *duck3* image. (e) An inpainting result.

image into constituent components while simultaneously forming a representation for those components which ties shapes and image features.

We have demonstrated the potential of the proposed skeleton based analysis on several illustrative examples. We have suggested a trivial foreground/background separation technique building on the patches of interest defined by the skeletal branches. Even though we use simple texture features to describe the structured image patches, we obtain fairly good separation results.

The current trend in recent object recognition studies is to employ shape features, e.g. silhouette or boundary based features [59, 105, 135], which are more reliable than the commonly used appearance based features extracted from square or circular local image patches. In this respect, local symmetry information carried by the skeletons is an important cue, which provides information regarding the shape of objects and their parts in images. The structured image regions identified by the proposed radius function describe local appear-

ance knowledge more accurately than the circular or square image patches as our region extraction procedure depends on local shape information as well. One can use the visual features extracted by this way within an object recognition framework to construct semantically more correct codebook of local appearances of a given object class. Once such a codebook is built, the knowledge about objects and their parts could be simply inferred by associating the visual cues estimated from input images with the ones in the codebook. An outcome of such a scheme would be an emergence of an object-specific knowledge from the images which can advance the prior-shape based filtering and segmentation frameworks by letting this knowledge initiate and supervise the segmentation process.

CHAPTER 7

SUMMARY AND CONCLUSION

Context has a broad meaning, and implies knowledge that affects the interpretations in a top-down manner. From computational vision point of view, the bottom-up processing strategies ignore context. This mainly results in an inaccurate perception, and consequently false inferences to be raised related to the external world being seen through images. In this respect, taking account of contextual information is very important to deal with missing or misleading image features. In general, contextual knowledge influences the processing of visual cues, which leads to resolution of the ambiguities or filling in the essential features that are not present in the data.

In this thesis, we explored various ways of improving low and mid-level computer vision modules such as image smoothing and segmentation with a capacity to use contextual data by integrating low-level image features with high-level information within unified formulations where bottom-up and top-down processing take place simultaneously. The proposed works are based on the AT approximation of the MS functional or the TSP formulation that depends on the AT model.

The AT model in general offers a global method that combines smoothing and edge detection within a unified formulation. The local interactions between the image points governed by the associated Euler-Lagrange equations give rise to solutions according to a global criterion which enforces smoothness of image regions and continuity of segmenting contours even though all the computations are carried out in a local fashion. However, this global criterion, or that is to say the global context, generally yields poor results as the computations are greatly affected by the ambiguities in the low-level visual cues.

Based on these findings, in Chapter 4, we have presented a novel regularization method which is derived from the AT model by letting the contextual knowledge extracted from local

image regions additionally modulate the diffusivity[1]. Specifically, the key idea behind this formulation is to embed some contextual measures in the Euler-Lagrange equations of the AT model, so that the diffusion process at each image point is more aware of the surrounding events and acts accordingly. To illustrate the influence of context in image smoothing, we suggested four different contextual feedback measures each apparently for a different visual task, and incorporated them in the regularization process to steer the ongoing diffusion. In the same computation framework, these measures can be integrated and the list can be extended with new feedback measures. The experiments performed on both gray and color natural images indicate the potential of the method under difficult noise types, non-uniform contrast, existence of multi-scale patterns and textures. More importantly, they show that it is possible to resolve some of the challenging cases for the AT model by considering local context.

In Chapter 5, we have looked into the task of segmenting out an object of interest in an image while the knowledge about the shape of the object is given a priori. Delineating accurate object boundaries is a very difficult task, especially when the images contain missing or misleading features due to partial occlusions, weak edges and noise. As previously illustrated in many prior-shape guided segmentation studies, introducing context in the form of high-level shape knowledge helps resolving this problem. One of the main novelties of this thesis is utilizing the TSP surfaces as prior shape representations. TSP surfaces provide distance-transform like representations for the shape boundary, and encode the local shape geometry accordingly.

We have presented two new segmentation frameworks that make use of TSP surfaces as shape priors which yield robustness against undesirable conditions such as noise, low contrast, weak edges and partial occlusions as our experimental results reveal. These frameworks mainly differ from each other in terms of how they model the deformation of the prior shape. In the segmentation framework proposed in Section 5.2, a model describing the global variations of the prior shape has been used to guide the segmentation process[2]. This framework has a fairly straightforward formulation, and can only deal with simple synthetic images. Yet, it should be noted that it demonstrates the potential use of TSP surfaces as prior shape representations. Our work proposed in Section 5.3, on the other hand, has employed

---

[1]An early version of this work was presented in the First International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2007) [57]. Full version is published in Journal of Mathematical Imaging and Vision (JMIV) [56].

[2]This is a joint work with Aykut Erdem and was previously published in the Fifth International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2005) [55].

TSP surfaces to determine meaningful correspondences between a given template and the evolving object region by considering a local deformation model similar to that of Hong et al. [73][3]. The deformation map estimated, in return, supervises segmentation by enforcing the zero-level curve of the evolving level set function towards the transformed template. This framework is more robust compared to our former method as the CV model utilized in this work results in a clear separation of object and background regions. The shape variability has been handled in a better way due to the integrated matching functional which considers a local deformation model. As a consequence, the method does only employ a single template to represent the shape prior and to capture the variations in the shape class, which eliminates the need for a set of training curves that correspond to the various appearances of the shape of interest. It is also important to remark that this method partitions an image into two as object versus background segments while jointly registering the prior shape on the image data, allowing further semantic analysis to be carried out on the object region extracted.

One critical issue in the prior-shape based segmentation frameworks is that an initial guess about the object of interest, which imposes top-down constraints, is required to initiate the segmentation process. Generally, finding out such a guess is as difficult as the segmentation process itself. As a step in that direction, in Chapter 6, we have proposed a skeleton based analysis of natural images which operates directly on the raw image with all its complexity. We divide the image into constituent components while simultaneously forming a representation for the components which ties shape and image features. In particular, considering our context-influenced regularization framework, we extend the TSP method such that it becomes applicable to natural images. We decouple the skeletal radius and the image gradient, and relate the skeletal radii directly to the minimum distance to the putative shape boundary. Using skeletal radii, we can identify local image regions of uniform characteristic. As our experimental results have clearly demonstrated, these structured image regions can be robustly and easily used in various image analysis tasks.

7.1 Future Work

A promising extension of this thesis could involve an interleaved segmentation and object recognition framework where the object-specific knowledge emerges from the images autonomously as in the emergence of the local contextual data within our context-influenced

---

[3]This work has been developed in collaboration with Luminita Vese during my research visit at University of California, Los Angeles (UCLA).

smoothing method, and supervises the recognition and segmentation processes simultaneously. Our image analysis procedure based on skeleton extraction gives us some hints about how such a framework can be attained, and its practical advantages over the similar joint recognition and segmentation studies. First of all, local symmetry information carried by the shape skeletons is an important cue which provides information regarding the shape of objects and their parts in images. Previously, in the initial part of Chapter 3, we reviewed some probabilistic formulations which employ models representing object knowledge based on local visual cues mostly as image patches or contour fragments to recognize and segment out objects in a scene. As we stated before, the appearance based features commonly used in these studies are extracted from square or circular image patches, and usually are inadequate to identify correct object boundaries. On the other hand, image features that can be computed from the local image regions described by the extracted skeleton branches (with the help of the suggested radius function) are more reliable since the local shape is taken into consideration while computing these structured image regions. This suggests that one could construct semantically more correct codebook of local appearances of a given object class by just using almost identical strategies utilized in these probabilistic formulations. Once such a codebook is contructed, the information regarding the objects and their parts could be simply inferred by associating the visual cues estimated from input images with the ones in the codebook.

Apart from the challenging joint recognition and segmentation task, there is a large number of possibilities to advance the proposed prior-shape guided segmentation frameworks. Integrating the local deformation model with a shape model describing the global variations of a given object class may result in a formulation that can cope with local shape variations, as well as with large amounts of occlusions and missing parts. In addition to this, as the TSP formulation allows estimating TSP surfaces directly from the grayscale images, one can employ this capability within a segmentation framework in an intelligent way, avoiding the constraints imposed by the CV model such as a dark object on a bright background, or a bright object on a dark background. Furthermore, TSP surfaces provide a rich shape representation in the sense that they can be used to perform explicit shape analysis by means of the skeletons extracted from them. In our formulations, we did not utilize this property. An interesting futurework could be devising a segmentation framework which is heavily built on skeletal information.

As another futurework, we plan to compare the shape representations that all encode local shape geometry in their own ways (Section 5.1) within a segmentation framework

similar to our former model by exploring their influences on the ongoing process. In this respect, we can decompose our former scheme into three coupled processes that respectively correspond to prior-guided segmentation, feature extraction, and local shape matching steps, which allows us to play with the local shape feature utilized in segmentation.

REFERENCES

[1] European Conference on Visual Perception (ECVP 2005), Symposium on the Role of Context in Recognition., August 2005.

[2] M. Aleman-Flores, L. Alvarez, and V. Caselles. Texture-oriented anisotropic filtering and geodesic active contours in breast tumor ultrasound segmentation. *J. Math. Imaging Vis.*, 28(1):81–97, 2007.

[3] L. Ambrosio and V. Tortorelli. On the approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Commun. Pure Appl. Math.*, 43(8):999–1036, 1990.

[4] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, 2001.

[5] C. Arcelli and L. Serino. Skeletonization of labeled gray-tone images. *Image and Vision Computing*, 23(2):159–166, 2005.

[6] C. Aslan. Disconnected Skeletons for Shape Recognition. Master's thesis, Department of Computer Engineering, Middle East Technical University, May 2005.

[7] C. Aslan, A. Erdem, E. Erdem, and S. Tari. Disconnected skeleton: Shape at its absolute scale. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(12):2188–2203, 2008.

[8] C. Aslan and S. Tari. An axis-based representation for recognition. In *ICCV*, volume 2, pages 1339–1346, 2005.

[9] J.-F. Aujol, G. Aubert, L. B.-F., and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *J. Math. Imaging Vis.*, 22(1):71–88, 2005.

[10] X. Bai and L. J. Latecki. Path similarity skeleton graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(7):1282–1292, 2008.

[11] L. Bar, N. Kiryati, and N. Sochen. Image deblurring in the presence of impulsive noise. *Int. J. Comput. Vision*, 70(3):279–298, 2006.

[12] L. Bar, N. Sochen, and N. Kiryati. Image deblurring in the presence of salt-and-pepper noise. In *Scale-Space*, pages 107–118, 2005.

[13] M. Bar. Visual objects in context. *Nature Reviews: Neuroscience*, 5:617–629, August 2004.

[14] E. Baseski, A. Erdem, and S. Tari. Dissimilarity between two skeletal trees in a context. *Pattern Recognition*. published online: 30 May 2008, doi:10.1016/j.patcog.2008.05.022.

[15] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16:2041–2066, 2004.

[16] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.

[17] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int. J. Comput. Vision*, 19(1):57–91, 1996.

[18] M. J. Black and G. Sapiro. Edges as outliers: Anisotropic smoothing using local image statistics. In *Scale-Space*, pages 259 – 270, 1999.

[19] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. Image Processing*, 7(3):421–432, March 1998.

[20] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, Cambridge, MA, 1987.

[21] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. In *CVPR Workshop on Perceptual Organization in Computer Vision*, 2004.

[22] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *ECCV*, pages 109–122, 2002.

[23] E. Borenstein and S. Ullman. Learning to segment. In *ECCV*, pages 315–328, 2004.

[24] A. Braides. *Approximation of Free-Discontinuity Problems.* Lecture Notes in Mathematics, Vol. 1694. Springer-Verlag, 1998.

[25] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Math. of Computation*, 31(138):333–390, 1977.

[26] X. Bresson, P. Vandergheynst, and J.-P. Thiran. A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional. *Int. J. Comput. Vision*, 68(2):145–162, 2006.

[27] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial (2nd ed.).* SIAM, Philadelphia, PA, USA, 2000.

[28] C. Broit. *Optimal Registration of Deformed Images.* PhD thesis, Computer and Information Science, University of Pennsylvania, 1981.

[29] A. Brook, R. Kimmel, and N. A. Sochen. Variational restoration and edge detection for color images. *J. Math. Imaging Vis.*, 18(3):247–268, 2003.

[30] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, volume 4, pages 25–36, 2004.

[31] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR*, volume 2, pages 60–65, 2005.

[32] C. A. Burbeck and S. M. Pizer. Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research*, 35(13):1917–1930, 1995.

[33] B. Burgeth, J. Weickert, and S. Tari. Minimally stochastic schemes for singular diffusion equations. In *Image Processing Based on Partial Differential Equations*, Mathematics and Visualization, pages 325–339. Springer Berlin Heidelberg, 2006.

[34] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

[35] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Int. J. Comput. Vision*, 22(1):61–79, 1997.

[36] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.

[37] R. H. Chan, C.-W. Ho, and M. Nikolova. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *IEEE Trans. Image Processing*, 14(10):1479–1485, 2005.

[38] T. Chan and L. Vese. Variational image restoration and segmentation models and approximations. *UCLA CAM-report 97-47*, September 1997.

[39] T. F. Chan, B.Y. Sandberg, and L.A. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11(2):130–141, 2000.

[40] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[41] S. Chen and X. Yang. A variational method with a noise detector for impulse noise removal. In *SSVM*, pages 442–450, 2007.

[42] D. H. Chung and G. Sapiro. Segmentation-free skeletonization of gray-scale images via PDE's. In *ICIP*, volume 2, pages 927–930, 2000.

[43] D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, 2003.

[44] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional. *Int. J. Comput. Vision*, 50(3):295–313, 2002.

[45] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part I: Fast and exact optimization. *J. Math. Imaging Vis.*, 26(3):261–276, 2006.

[46] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part II: Levelable functions, convex priors and non-convex cases. *J. Math. Imaging Vis.*, 26(3):277–291, 2006.

[47] J. Debayle and J.C. Pinoli. General adaptive neighborhood image processing: Part I. *J. Math. Imaging Vis.*, 25:245–266, 2006.

[48] G. Deng. Symbol mapping and context filtering for lossless image compression. In *ICIP*, volume 1, pages 526–529, 1998.

[49] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *Int. J. Comput. Vision*, 40(1):7–23, 2000.

[50] A. Desolneux, L. Moisan, and J.-M. Morel. *From Gestalt Theory to Image Analysis: A Probabilistic Approach.*, volume 35 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, NY, 2007.

[51] G. S. di Baja and I. Nystrom. 2D grey-level skeleton computation: a discrete 3D approach. In *ICPR*, volume 2, pages 455–458, 2004.

[52] F. Dibos and G. Koepfler. Global total variation minimization. *SIAM J. Numer. Anal.*, 37(2):646–664, 2000.

[53] R. P. W. Duin and P. W. Verbeek. Gray value skeleton. *Delft (TN/FL-PH) Internal Report*, 1979.

[54] D. Eberly, R. Gardner, B. Morse, S. Pizer, and C. Scharlach. Ridges for image analysis. *J. Math. Imaging Vis.*, 4(4):353–373, 1994.

[55] E. Erdem, A. Erdem, and S. Tari. Edge strength functions as shape priors in image segmentation. In *EMMCVPR*, pages 490–502, 2005.

[56] E. Erdem and S. Tari. Mumford-Shah regularizer with contextual feedback. *J. Math. Imaging Vis.* published online: 24 July 2008, doi:10.1007/s10851-008-0109-y.

[57] E. Erdem, A. Sancar Yilmaz, and S. Tari. Mumford-Shah regularizer with spatial coherence. In *SSVM*, pages 545–555, 2007.

[58] O. Faugeras. *Three-Dimensional Computer Vision.*, chapter 4.2. The MIT Press, 1993.

[59] R. Fergus, P. Perona, and A. Zisserman. A visual category filter for google images. In *ECCV*, volume 1, pages 242–256, 2004.

[60] M. Galun, E. Sharon, R. Basri, and A. Brandt. Texture segmentation by multiscale aggregation of filter responses and shape elements. In *ICCV*, volume 1, pages 716–723, 2003.

[61] J. M. Gauch and S. M. Pizer. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(6):635–646, 1993.

[62] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–639, 1984.

[63] G. Gilboa, J. Darbon, S. Osher, and T. Chan. Nonlocal convex functionals for image regularization. *UCLA CAM-report 06-57*, October 2006.

[64] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Variational denoising of partly textured images by spatially varying constraints. *IEEE Trans. Image Processing*, 15(8):2281–2289, August 2006.

[65] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt. Shape representation and classification using the poisson equation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):1991–2005, December 2006.

[66] J. Goutsias and D. Schonfeld. Morphological representation of discrete and binary images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1369–1379, 1991.

[67] C. Le Guyader and L. Vese. A combined segmentation and registration framework with a nonlinear elasticity smoother. *UCLA CAM-report 08-16*, March 2008.

[68] R. M. Haralick and J. S. J. Lee. Context dependent edge detection. In *CVPR*, pages 223–228, 1988.

[69] M. Heiler and C. Schnörr. Learning non-negative sparse image codes by convex programming. In *ICCV*, volume 2, pages 1667–1674, 2005.

[70] G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. In *ECCV*, volume 1, pages 30–43, 2008.

[71] S. Hochstein and M. Ahissar. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804, 2002.

[72] T. Hofmann, J. Puzicha, and J. M. Buhmann. An optimization approach to unsupervised hierarchical texture segmentation. In *ICIP*, volume 3, pages 213–216, 1997.

[73] B.-W. Hong, E. Prados, S. Soatto, and L. Vese. Shape representation based on integral kernels: Application to image matching and segmentation. In *CVPR*, volume 1, pages 833– 840, 2006.

[74] T. Iijima. Basic theory on normalization of a pattern (for the case of a typical one-dimensional pattern). *Bulletin of the Electrotechnical Laboratory*, 26:368–388, 1962 (in Japanese).

[75] Y. Jin and S. Geman. Context and hierarchy in a probabilistic image model. In *CVPR*, volume 2, pages 2145–2152, 2006.

[76] J. P. Kaufhold. *Energy Formulations of Medical Image Segmentations.* PhD thesis, Boston University College of Engineering, Department of Biomedical Engineering, Boston, MA, 2000.

[77] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *ICCV*, pages 810–815, 1995.

[78] B. B. Kimia, A. R. Tannenbaum, and S. W. Zucker. Shapes, shocks, and deformations I: The components of two-dimensional shape and the reaction-diffusion space. *Int. J. Comput. Vision*, 15(3):189–224, 1995.

[79] J. J. Koenderink. The structure of images. *Biol. Cybernetics*, 50:363–370, 1984.

[80] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Texture analysis and segmentation using modulation features, generative models and weighted curve evolution. *IEEE Trans. Pattern Anal. Mach. Intell.* published online: 30 January 2008, doi:10.1109/TPAMI.2008.33.

[81] G. C. Lai and R. J. P. de Figueiredo. Image interpretation using contextual feedback. In *ICIP*, volume 2, page 2623, 1995.

[82] S. Lazebnik, C. Schmid, and J. Ponce. A maximum entropy framework for part-based texture and object recognition. In *ICCV*, volume 1, pages 832–838, 2005.

[83] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(21):788â791, 1999.

[84] T. S. Lee. Contextual influences in visual processing. In M. D. Binder, N. Hirokawa, and U. Windhorst, editors, *Encyclopedia of Neuroscience*. Springer-Verlag. in Press.

[85] T. S. Lee, D. Mumford, R. Romero, and V. A.F. Lamme. The role of the primary visual cortex in higher level vision. *Vision Research*, 38(15-16):2429–2454, 1998.

[86] T. S. Lee and A. L. Yuille. Efficient coding of visual scenes by grouping and segmentation. In K. Doya, S. Ishii, A. Pouget, and R. P. N. Rao, editors, *Bayesian Brain: Probabilistic Approaches to Neural Coding*, chapter 8, pages 141–185. MIT Press, 2007.

[87] B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In *ECCV'04 Workshop on Statistical Learning in Computer Vision*, pages 17–32, 2004.

[88] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *Int. J. Comput. Vision*, 77(1-3):259–289, 2008.

[89] M. E. Leventon, E. L. Grimson, and O. D. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR*, pages 1316–1323, 2000.

[90] G. Levi and U. Montanari. A grey-weighted skeleton. *Information and Control*, 17(1):62–91, August 1970.

[91] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *Int. J. Comput. Vision*, 30(2):117–154, 1998.

[92] P. Lombardi, V. Cantoni, and B. Zavidovique. Context in robotic vision: Control for real-time adaptation. In *ICINCO*, volume 3, pages 135–142, 2004.

[93] A. M. López, F. Lumbreras, and J. Serrat. Creaseness from level set extrinsic curvature. In *ECCV*, volume 2, pages 156–169, 1998.

[94] J. B. A. Maintz, P. A. van den Elsen, and M. A. Viergever. Evaluation of ridge seeking operators for multimodality medical image matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(4):353–365, 1996.

[95] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):701–716, 1989.

[96] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman, San Francisco, 1982.

[97] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, pages 416–423, 2001.

[98] J. Modersitzki. *Numerical Methods for Image Registration.* Oxford University Press, 2004.

[99] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation.* Birkhauser, 1995.

[100] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.*, 42(5):577–685, 1989.

[101] K. Murphy, A. Torralba, and W. Freeman. Using the forest to see the trees: A graphical model relating features, objects and scenes. In *NIPS*, 2003.

[102] S. O. Murray, P. Schrater, and D. Kersten. Perceptual grouping and the interactions between visual cortical areas. *Neural Networks*, 17(5-6):695–705, 2004.

[103] M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space and edge detection filters. *J. Math. Imaging Vis.*, 7(4):291–307, 1997.

[104] M. Nikolova. A variational approach to remove outliers and impulse noise. *J. Math. Imaging Vis.*, 20(1-2):99–120, 2004.

[105] A. Opelt, A. Pinz, and Andrew Zisserman. A boundary-fragment-model for object detection. In *ECCV*, volume 2, pages 575–588, 2006.

[106] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulation. *J. Comput. Phys.*, 79:12–49, 1988.

[107] G. Papari, P. Campisi, N. Petkov, and A. Neri. A biologically motivated multiresolution approach to contour detection. *EURASIP J. Appl. Signal Process*, 2007(Article ID 71828):28 pages, 2007.

[108] D. Parikh, C. L. Zitnick, and T. Chen. Determining patch saliency using low-level context. In *ECCV*, volume 2, pages 446–459, 2008.

[109] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12:629–639, 1990.

[110] H. H. Pien and J. M. Gauch. Variational segmentation of multi-channel MRI images. In *ICIP*, volume 3, pages 508–512, 1994.

[111] S. M. Pizer, K. Siddiqi, G. Székely, J. N. Damon, and S. W. Zucker. Multiscale medial loci and their properties. *Int. J. Comput. Vision*, 55(2-3):155–179, 2003.

[112] J. Porway, K. Wang, B. Yao, and S. C. Zhu. A hierarchical and contextual model for aerial image understanding. In *CVPR*, pages 1–8, 2008.

[113] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *ICCV*, pages 1–8, 2007.

[114] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Segmentation by level sets and symmetry. In *CVPR*, volume 1, pages 1015–1022, 2006.

[115] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Prior-based segmentation and shape registration in the presence of perspective distortion. *Int. J. Comput. Vision*, 72(3):309–328, 2007.

[116] T. Riklin-Raviv, N. Kiryati, and N. A. Sochen. Unlevel-sets: Geometry and prior-based segmentation. In *ECCV*, volume 4, pages 50–61, 2004.

[117] T. Riklin-Raviv, N. Sochen, and N. Kiryati. Shape-based mutual segmentation. *Int. J. Comput. Vision*, 79(3):231–245, 2008.

[118] G.X. Ritter and J.N. Wilson. *Handbook of Computer Vision Algorithms in Image Algebra*. CRC Press, 1996.

[119] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, volume 2, pages 860– 867, 2005.

[120] M. Rousson and N. Paragios. Shape priors for level set representations. In *ECCV*, volume 2, pages 78–92, 2002.

[121] M. Rousson and N. Paragios. Prior knowledge, level set representations visual grouping. *Int. J. Comput. Vision*, 76(3):231–243, 2008.

[122] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D.*, 60:259–268, 1992.

[123] B. Sandberg, T. Chan, and L. Vese. A level-set and Gabor-based active contour algorithm for segmenting textured images. *UCLA CAM-report 02-39*, July 2002.

[124] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *J. Math. Imaging Vis.*, 12(1):43–63, 2000.

[125] O. Schwartz, A. Hsu, and P. Dayan. Space and time in visual context. *Nature Reviews: Neuroscience*, 8:522–535, July 2007.

[126] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):550–571, 2004.

[127] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.

[128] J. Sethian. *Level Set Methods.* Cambridge University Press, 1996.

[129] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proc. of the National Academy of Sciences of the USA*, 93(4):1591â1595, 1996.

[130] J. Shah. Segmentation by nonlinear diffusion. In *CVPR*, pages 202–207, 1991.

[131] J. Shah. A common framework for curve evolution, segmentation and anisotropic diffusion. In *CVPR*, pages 136–142, 1996.

[132] J. Shah. Grayscale skeletons and segmentation of shapes. *CVIU*, 99(1):196â109, 2005.

[133] J. Shah. Skeletons and segmentation of shapes. *Department of Mathematics, Northeastern University, Internal Report*, 2005.

[134] E. Sharon, A. Brandt, and R. Basri. Segmentation and boundary detection using multiscale intensity measurements. In *CVPR*, volume 1, pages 469–476, 2001.

[135] J. Shotton, A. Blake, and R. Cipolla. Contour-based learning for object detection. In *ICCV*, volume 1, pages 503– 510, 2005.

[136] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *Int. J. Comput. Vision*, 35(1):13–32, 1999.

[137] A. Singhal, J. Luo, and W. Zhu. Probabilistic spatial context models for scene content understanding. In *CVPR*, volume 1, pages 235–241, 2003.

[138] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. Conditional models for contextual human motion recognition. In *ICCV*, volume 2, pages 1808–1815, 2005.

[139] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman and Hall, 1993.

[140] A. Srivastava, A. B. Lee, E. P Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *J. Math. Imaging Vis.*, 18(1):17–33, 2003.

[141] C. Y. Suen and P. S. P. Wang, editors. *Thinning Methodologies for Pattern Recognition.*, volume 8 of *Series in Machine Perception and Artificial Intelligence*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1994.

[142] Q. Tang, N. Sang, and T. Zhang. Contour detection based on contextual influences. *Image and Vision Computing*, 25:1282–1290, 2007.

[143] S. Tari, J. Shah, and H. Pien. Extraction of shape skeletons from grayscale images. *CVIU*, 66(2):133–146, 1997.

[144] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Variational approach for edge preserving regularization using coupled PDE's. *IEEE Trans. Image Processing*, 7(3):387–397, March 1998.

[145] A. N. Tikhonov. *Solutions of Ill-posed Problems.* Winston and Sons, Washington D.C., 1977.

[146] A. Torralba and A. Oliva. Statistics of natural image categories. *Network-Computation in Neural Systems*, 14(3):391–412, 2003.

[147] G. T. Toussaint. The use of context in pattern recognition. *Pattern Recognition*, 10(3):189–204, 1978.

[148] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imaging*, 22(2):137–154, 2003.

[149] B.J.H. Verwer, L.J. van Vliet, and P.W. Verbeek. Binary and grey-value skeletons: Metrics and algorithms. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 7(5):1287–1308, 1993.

[150] L. A. Vese and S. J. Osher. Image denoising and decomposition with total variation minimization and oscillatory functions. *J. Math. Imaging Vis.*, 20(1-2):7–18, 2004.

[151] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(6):583–598, 1991.

[152] S. Wang, A. Rosenfeld, and A. Y. Wu. A medial axis transformation for grayscale pictures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 4(4):419–421, 1982.

[153] J. Weickert. Anisotropic diffusion filters for image processing based quality control. In *Proc. Seventh European Conference on Mathematics in Industry*, pages 355–362, 1994.

[154] J. Weickert. *Anisotropic Diffusion in Image Processing.* Teubner, Stuttgart, 1998.

[155] A. P. Witkin. Scale space filtering. *Proc. Eighth Int. Joint Conf. on Artificial Intelligence*, 2:1019–1022, 1983.

[156] L. Wolf and S. Bileschi. A critical view of context. *Int. J. Comput. Vision*, 69(2):251–261, 2006.

[157] L. Wolf, X. Huang, I. Martin, and D. Metaxas. Patch-based texture edges and segmentation. In *ECCV*, volume 2, pages 481–493, 2006.

[158] A. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *ICCV*, volume 2, pages 898–903, 1999.

[159] S. Zheng, Z. Tu, and A. Yuille. Detecting object boundaries using low-, mid-, and high-level information. In *CVPR*, pages 1–8, 2007.

[160] S. C. Zhu and A. L. Yuille. FORMS: A flexible object recognition and modeling system. *Int. J. Comput. Vision*, 20(3):187–212, 1996.

VITA

Personal Information

Surname, Name: Erdem, Mehmet Erkut

Date and Place of Birth: 21 March 1980, Ankara

Phone: (+90) 312 266 4577

Fax: (+90) 312 210 5544

E-mail: erkut@ceng.metu.edu.tr

Web: www.ceng.metu.edu.tr/~erkut

Research Interests

Computer vision (cognitive and computational aspects)

nonlinear diffusion equations, level sets, calculus of variation, symmetry representation

Education

M.Sc., Dept. of Computer Engineering, METU, Turkey, July 2003

- Thesis Title: Image-based Extraction of Material Reflectance Properties of a 3D Rigid Object
- Advisor: Volkan Atalay
- CGPA: 3.86 / 4.00 (High Honor)

B.Sc., Dept. of Computer Engineering, METU, Turkey, June 2001

- CGPA: 3.68 / 4.00 (High Honor)

Awards

- IEEE SIU 2004 Alper Atalay Best Student Paper Award (Third Prize), 2004
- Interpro Information and Communication Technologies Awards, Encouragement Award, 2002

Refereeing

- Reviewer, Pattern Recognition
- Additional Reviewer, ECCV 2006
- Additional Reviewer, ICCV 2005

Visiting Positions

Research Visit   **Oct 2007 - Dec 2007**

Department of Mathematics (worked with Luminita A. Vese)

University of California, Los Angeles (UCLA), Los Angeles, USA

Research Visit   **Jul 2004 - Aug 2004**

Virginia Bioinformatics Institute (worked with Volkan Atalay and Rengul Cetin-Atalay)

Virginia Polytechnic Institute and State University, Virginia, USA

Teaching Experience

Teaching Assistant   **Dec 2001 - Sep 2008**

Department of Computer Engineering,

Middle East Technical University, Ankara, TURKEY

courses include Image Processing, Dynamic Systems, Scientific Computing, Algorithms and Data Structures.

International Journal Publications

Published

E. Erdem and S. Tari, Mumford-Shah Regularizer with Contextual Feedback, *Journal of Mathematical Imaging and Vision*, published online: 24 Jul 2008, doi:10.1007/s10851-008-0109-y

C. Aslan, A. Erdem, E. Erdem, S. Tari, Disconnected Skeleton: Shape at its Absolute Scale, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, published online: 18 December 2007, doi:10.1109/TPAMI.2007.70842

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin, Vision-based continuous Graffiti-like text entry system, *Optical Engineering*, Vol. 43, Issue 3, pp. 553-558, March 2004

Submitted

E. Erdem and S. Tari, Skeletons from Natural Images, submitted

INTERNATIONAL CONFERENCE PUBLICATIONS

Book Chapters

E. Erdem, A. Sancar Yilmaz, S. Tari, Mumford-Shah Regularizer with Spatial Coherence, *Proc. of International Conference on Scale Space Methods and Variational Methods in Computer Vision 2007 (SSVM'07)*, Ischia, Italy, Springer LNCS, Vol. 4485, pp. 545-555, May-June 2007

A. Erdem, E. Erdem, S. Tari, Articulation Prior in an Axial Representation, *Workshop on the Representation and Use of Prior Knowledge in Computer Vision - in conjuction with ECCV 2006*, Graz, Austria, to appear in Springer LNCS, May 2006

E. Erdem, A. Erdem, S. Tari, Edge Strength Functions as Shape Priors in Image Segmentation, *Proc. of EMMCVPR 2005*, Florida, USA, Springer LNCS, Vol. 3757, pp. 490-502, November 2005

Other Conference Publications

E. Erdem, A. Erdem, V. Atalay, A. E. Çetin, Computer Vision Based Unistroke Keyboard System and Mouse for the Handicapped, *ICME 2003*, Baltimore, USA, Volume 2, pp. 765-768, July 2003

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin, Computer Vision Based Unistroke Keyboards, *7th International Symposium On Computer and Information Sciences*, Florida, USA, pp. 210-214, October 2002

National Conference Publications

E. Erdem, A. Erdem, U. Yılmaz, V. Atalay, Üç Boyutlu Katı Nesnelerin Yansıtma Özelliklerinin Görüntülerden Çıkarılması, *IEEE Sinyal İşleme ve Uygulamaları Kurultayı (SIU 2004)*, Kuşadası, Nisan 2004

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin, Engelliler İçin Bilgisayarlı Görmeye Dayalı Klavye ve Fare Sistemi, *IEEE Sinyal İşleme ve Uygulamaları Kurultayı (SIU 2003)*, İstanbul, Haziran 2003

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin, Kağıt Klavye: Bilgisayarlı Görmeye Dayalı Klavye, *IEEE Sinyal İşleme ve Uygulamaları Kurultayı (SIU 2002)*, Pamukkale, Haziran 2002

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin., Bilgisayarlı Görmeye Dayalı Fare Sistemi, *IEEE Sinyal İşleme ve Uygulamaları Kurultayı (SIU 2002)*, Pamukkale, Haziran 2002

International Conferences (Abstract Only)

E. Erdem and S. Tari, Mumford-Shah Regularizer with Contextual Feedback, *SIAM Conference on Imaging Science 2008*, San Diego, USA, July 2008

M. Iskar, E. Erdem, A. Erdem, S. Akman, A. Dickerman, V. Atalay, R. Cetin-Atalay, CAPRIS: A Database for Cancer Gene Promoter Related Motif Search, *International Symposium on Health Informatics and Bioinformatics*, Antalya, TURKEY, November 2005

E. Erdem, A. Erdem, V. Atalay, Image-based Extraction of Material Reflectance Properties of a 3D Rigid Object, *Eurographics 2003*, Poster Session, Granada, Spain, September 2003

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin, Computer based text entry for Wearable Computing, *6th IEEE International Symposium on Wearable Computers, Demonstrations*, Seattle, USA, October 2002

A. Erdem, E. Erdem, V. Atalay, A. E. Çetin, Computer Vision Based Mouse, *ICASSP 2002, Student Forum Session*, Florida, USA, May 2002

Seminars

Incorporating Contextual Data into Low and Mid Level Vision, *UCLA Dept. of Mathematics Image Processing Seminar*, Los Angeles, USA, November 2007

Incorporating Contextual Data into Low and Mid Level Vision, *UCLA Vision Lab Group Meetings*, Los Angeles, USA, November 2007

Mumford-Shah Regularizer with Contextual Feedback, *METU Dept. of Computer Engineering*, Ankara, TURKEY, May 2007

Mumford-Shah Regularizer with Contextual Feedback, *METU Institute of Applied Mathematics*, Ankara, TURKEY, May 2007

Personal Interests

An active member of Seruven (a group of people who think, write and discuss about comics and graphic novels). For more information, you can visit http://www.seruven.org.