The Elastic Ratio: Introducing Curvature into Ratio-based Globally Optimal Image Segmentation

Thomas Schoenemann, Simon Masnou, Daniel Cremers

The Elastic Ratio: Introducing Curvature into Ratio-based Globally Optimal Image Segmentation

Thomas Schoenemann, Simon Masnou and Daniel Cremers

***Abstract*— In this paper we present the first globally optimal ratio-based image segmentation method allowing to impose curvature regularity of the region boundary. The proposed method is fully unsupervised and compares favorably to other such approaches.**

To identify the optimal foreground region in the image, the algorithm minimizes the ratio of flux over a weighted sum of length and curvature regularity of the region boundary. The key concept is to find cycles in a product graph where each node corresponds to a pair of image locations.

Furthermore our results allow to draw conclusions about certain global optima of a reformulated snakes functional which is independent of parameterization: the proposed algorithm allows to find parameter sets where the modified snakes functional has a meaningful solution and simultaneously provides the corresponding global solution.

I. Introduction

Curvature regularity plays an important role in many fields of computer vision and image processing - among them image segmentation, perceptual organization and inpainting. Psychological studies have identified curvature as a key component for human scene interpretation.

In this paper we present a method that links curvature with image segmentation: the optimal foreground region is obtained by the minimization of an energy functional of the region boundary that we term the *elastic ratio*:

$$\frac{\displaystyle\int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^{\perp} \, ds}{\nu \, \mathcal{L}(C) + \displaystyle\int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q \, ds} \tag{1}$$

Here $I : \Omega \to \mathbb{R}$ denotes a gray-level image defined on a domain $\Omega \subset \mathbb{R}^2$, $C : [0, \mathcal{L}(C)] \to \Omega$ is a positively oriented parameterization by arc-length of the region boundary that we assume to be smooth, simple and closed with length $\mathcal{L}(C)$, $\big(C'(s)\big)^{\perp}$ coincides with $C'(s)$ rotated by $+\frac{\pi}{2}$, $\kappa_C(s)$ is the curvature vector at the point $C(s)$ – i.e., $\kappa_C(s) = C''(s)$ due to arc-length parameterization [22] – q and ν are positive real weighting factors and $\cdot$ denotes the scalar product in $\mathbb{R}^2$. Remark that changing the orientation of a curve changes the sign of the associated elastic ratio.

Figure 1 demonstrates the effect of curvature in this functional: it compares a version with length-regularity only to a version with curvature-regularity only. The results with curvature are much closer to human perception and we observed this consistently throughout our experiments.

Thomas Schoenemann and Daniel Cremers are with the Department of Computer Science, University of Bonn, Germany.

Simon Masnou is with the Laboratoire Jacques-Louis Lions, Université Paris 6, France.

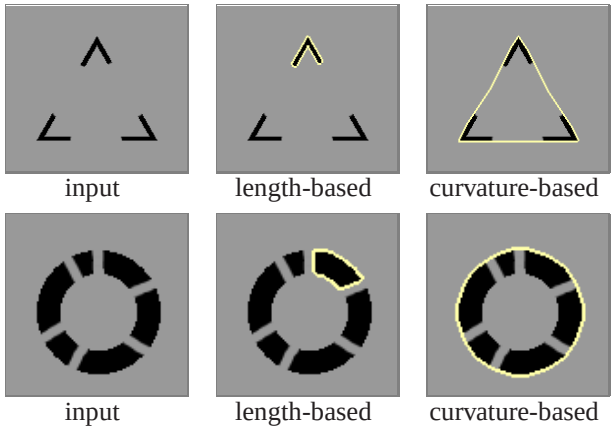


Fig. 1. Effects of length-based and curvature-based regularization in image segmentation on artificial images. Results were generated using functionals (4) and (1), the latter with absolute curvature ($q = 1, \nu = 0$). Note that curvature regularity gives rise to fundamentally different segmentations. In particular it allows gap closing and contour completion.

The proposed method is fully unsupervised and allows to compute global minimizers of a discrete version of the energy. It is of practical value since it deals with shading effects where region-based methods have problems.

It is also of theoretical value since the computation of global minimizers helps when analyzing the usefulness of a functional: it allows to identify the strengths and weaknesses of a model.

Moreover, the class of optimizable functionals is not limited to (1): we allow almost arbitrary dependencies on the position of a point $C(t)$ on the curve, the tangent vector $C'(t)$ and the curvature vector $\kappa_C(t)$ at $C(t)$. That is, we consider ratios of form

$$\frac{\displaystyle\int_0^{\mathcal{L}(C)} h(C(s), C'(s), \kappa_C(s)) \, ds}{\displaystyle\int_0^{\mathcal{L}(C)} g(C(s), C'(s), \kappa_C(s)) \, ds}$$

where $h : \Omega \times \mathbb{S}^1 \times \mathbb{R}^2 \to \mathbb{R}$ is arbitrary and $g : \Omega \times \mathbb{S}^1 \times \mathbb{R}^2 \to \mathbb{R}^+$ such that the denominator is strictly positive for all closed curves C with strictly positive length.

Among these functionals, we study what we call the *snakes ratio*:

$$\frac{-\displaystyle\int_0^{\mathcal{L}(C)} |\nabla I(C(s))|^p \, ds}{\nu \mathcal{L}(C) + \int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q \, ds} \tag{2}$$

As we shall see later, the snakes ratio is closely related to a parameterization-invariant formulation of the famous snakes model [40]: it allows to compute global minimizers of this reformulation and simultaneously provides a parameter set that leads to meaningful global optima.

Before entering the details of our method, we first give an introduction on related work on curvature and image segmentation.

II. Related Work and Contribution

In this work we propose a method for image segmentation which relies on results on the influence of curvature regularity in various fields. Both areas will now be reviewed, then the contribution of this paper is described.

A. Curvature in Vision and Image Processing

Following results in Psychophysics [39], the curvature has been introduced into various fields of computer vision. It was first seen in the context of shape completion through the celebrated Euler's elastica energy $\int_0^{\mathcal{L}(C)} |\kappa_C(s)|^2 \, ds$, see [57], [35], [48] and the subsequent developments in [25], [41], [32], [17]. Other applications of curvature are image segmentation [40], [3], [2], [51], [12], inpainting [4], [15], [45], [28], [30], image smoothing and denoising (see [56] and references in [11]), image analysis [52], [26] or surface interpolation and smoothing [36], [55], [24], [6], [29].

A large body of this literature is devoted to optimizing curvature-dependent functionals in a continuous and/or discrete setting through a minimizing flow yielding *local* optimizers. Very few methods are able to compute *global* optimizers, as for instance [46], [45] where total absolute curvature is optimized globally in the context of inpainting and [3], [2] where Amini, Weymouth and Jain showed how to globally optimize line integrals depending on curvature by means of dynamic programming. However, the run-time of their method is guaranteed to be quadratic in the number of image pixels. In practice this is too slow for reasonable image sizes. As a consequence they only consider curves in a band around an initial curve.

B. Image Segmentation as an Optimization problem

Over the past few decades numerous methods have been proposed for image segmentation. We focus on those that are based on minimizing a suitable energy functional. The corresponding functionals typically combine a data term with a regularity term (for a recent review see [19]). Both terms can incorporate either region or edge properties, yet the regularity term generally penalizes certain properties of the region boundaries.

a) Region-based Methods: In region-based image segmentation the intensity inside the region is assumed to be approximately constant [49], [16], to vary only slowly [49] or to be generated by a suitable probability model [61]. Recent methods allow to integrate flux into region-based formulations [38], [58], [42], [43], in particular for the segmentation of long elongated structures.

In the region-based framework, global optima are usually not available in polynomial time. A region-based functional with a very basic smoothness term could recently be optimized globally [23]. For length regularity and *given* region statistics global optima can be computed [7], [14], [50]. For curvature regularity we are only aware of local methods [27], [51].

b) Edge-based Methods: Ever since the pioneering work of Kass, Witkin and Terzopoulos [40], edge-based methods [13], [60], [37], [59] have formed one of the major approaches to image segmentation.

In their seminal work on the snakes model, Kass et al. [40] propose to minimize an energy computed along the region boundary described by a curve $C : [0, 1] \to \Omega$. The criterion to be minimized reads:

$$\alpha \int_0^1 |C'(t)|^2 dt + \beta \int_0^1 |C''(t)|^2 dt - \lambda \int_0^1 |\nabla I(C(t))|^2 \, dt$$

where α, β and λ are real positive constants. The first two terms of the criterion measure the smoothness of the curve (recall that the second derivative $C''(t)$ is directly related to the curvature – they actually coincide for arc-length parameterization of the curve) whereas the third term aims at attracting the curve to places with large image gradients. It is a well known issue of the model that for some particular values of the parameters, no meaningful global optima can be found: the only minimizers would be either points or curves with infinite length. In this paper we state a parameterization-invariant version of the functional and show that the proposed method allows to identify meaningful global optima of this version.

Amini et al. [3] optimize a modified snakes functional using dynamic programming. To avoid meaningless solutions, they enforce the curve to be polygonal with an a priori known number of points, such that no two neighboring points are less than a given distance apart. Although their method provides polynomial time solutions, due to its quadratic run-time complexity it is only applied in a small band around an initial curve.

Caselles et al. [13] propose to minimize a line integral of a positive edge indicator function $g(|\nabla I|)$, where g decreases monotonically with increasing gradient strength. Although this model is interesting for local optimization, it is easily seen that the global minimizers are meaningless: any degenerate curve reduced to a point in the image is a minimizer of this functional. Meaningful global optima can be found when seed nodes are given for foreground and background [8].

Methods that work without seed nodes are usually built on ratio functionals. Shi and Malik [54] propose normalized cuts, which leads to an NP-hard optimization problem. Using relaxation techniques one can find a solution which is independent of initialization.

The ratio regions of Cox et al. [18] provide meaningful global optima together with a polynomial time complexity. However, the complete search over all starting points takes prohibitively long in practice. Moreover the method is limited to planar graphs.

Finally, Jermyn and Ishikawa propose in [38] a class of ratio functionals for image segmentation, including the ratio of flux over length, whose global minimizers can be efficiently found by iterative negative cycle detection in a suitable graph. In this paper we extend their approach to include curvature regularity.

C. Contribution

In this paper we extend the class [38] of globally optimizable ratio functionals: we show how to integrate curvature regularity of the region boundary. The proposed method is fully unsupervised and allows a great variety of data and regularity terms: in the numerator any functional dependence on the (local) curvature of the curve is allowed. In the denominator we require functions yielding positive integrals for all closed curves.

Moreover the proposed algorithm allows to draw conclusions about certain global optima of a parameterization-invariant version of the snakes model: it allows to find a parameter set where the snakes functional has a meaningful global optimum for the given image and simultaneously provides the corresponding global solution.

A preliminary version of this work appeared in [53]. This extended version additionally addresses the issues of an efficient parallel implementation as well as theoretical considerations of the continuous optimization task: we address the existence of minimizers in the continuous domain and prove the convergence of discrete minimizers to "continuous" minimizers as the resolution increases.

III. The Jermyn-Ishikawa segmentation method

I.H. Jermyn and H. Ishikawa proposed in [38] to optimize a criterion with two nice properties:

1) its discrete counterpart can be *globally* optimized in polynomial time, which is somewhat unusual: for most segmentation models based on the optimization of an energy, the global optima are either
2) it can indifferently be written as an edge-based or a meaningless or not efficiently computable. region-based criterion. Most criteria in the literature fall either in one or the other category.

We will first state the contour-based formulation, then turn to the region-based one.

A. Contour-based Problem Statement

In Jermyn and Ishikawa's approach the image domain is segmented into two regions separated by an *oriented* curve C. In the most general setting, this curve minimizes a ratio of the form

$$\frac{\int_0^{\mathcal{L}(C)} \vec{v}(C(s)) \cdot \big(C'(s)\big)^\perp \, ds}{\int_0^{\mathcal{L}(C)} g(C(s)) \, ds} \tag{3}$$

where the curve C is parameterized with arc-length, $\vec{v} : \Omega \to \mathbb{R}^2$ is a vector field, $\big(C'(s)\big)^\perp$ is the unit normal to the curve that coincides with $C'(s)$ rotated by $+\frac{\pi}{2}$ and $g : \mathbb{R}^2 \to \mathbb{R}^+$ is a positive weight function.

In particular, Jermyn and Ishikawa propose to minimize the average outward flux that we call *length ratio* in this paper:

$$\frac{\int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^\perp \, ds}{\mathcal{L}(C)} \tag{4}$$

It is worth noticing that we are dealing with oriented curves C and do not fix the orientation: when changing the orientation, the curve normals $\big(C'(s)\big)^\perp$ switch sign and so does the entire numerator term. Hence minimizing (4) over oriented curves is equivalent to maximizing its absolute value.

It therefore amounts to finding curves perpendicular to strong image edges. The normalization by length avoids trivial optima (zero-length or infinite-length curves). Yet, in practice the found regions tend to be small and usually do not coincide with human perception. As a remedy Jermyn and Ishikawa proposed a balloon force. To understand how this works we must first look at the region-based interpretation.

B. Conversion to the Region-based Form

Criterion (3) is only apparently a pure edge criterion. As observed in [38], the Gauss-Green Theorem forms a bridge between curve energies and region energies. Denoting C a simple and smooth curve in $\mathbb{R}^2$ and C_{in} the connected region enclosed by C, the Gauss-Green Theorem states that for any smooth field $\vec{v}$ on $\mathbb{R}^2$

$$\int_{C_{in}} \operatorname{div} \vec{v} \, dx = -\int_0^{\mathcal{L}(C)} \vec{v}(C(s)) \cdot \vec{n}_C(s) \, ds$$

where $\vec{n}_C(s)$ is the *inner* unit normal to C at $C(s)$. It is easily observed that any continuous function f defined on $\mathbb{R}^2$ can be associated with a vector field $\vec{v}_f$ such that $f = \operatorname{div} \vec{v}_f$ by the simple formula $\vec{v}_f(x,y) = \frac{1}{2}\big(\int_0^x f(t,y)\,dt, \int_0^y f(x,t)\,dt\big)$. With this formula, the minimization of (4) can equivalently be written as

$$\max_C \frac{\left| \int_{C_{in}} \Delta I(x) \, dx \right|}{\mathcal{L}(C)},$$

where $\Delta I = \operatorname{div}(\nabla I)$ is the Laplace operator. This expression gives some indications on the optimal regions and confirms what can be observed experimentally: the optimal regions are likely to contain high values of ΔI with constant sign, which often occurs for small regions in the vicinity of edges, i.e. in zones where ∇I changes a lot.

To include the balloon force weighted with β one only has to add $\pm\beta$[1] to $\Delta I(x)$. The arising problem can be written as

$$\max_C \frac{\left| \int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^\perp \, ds \pm \int_{C_{in}} \beta \, dx \right|}{\mathcal{L}(C)}, \tag{5}$$

and will be termed *extended length ratio* in the following. It provides substantially better results in practice.

C. Problems with Region-based Terms

In the considered framework, the inclusion of region-based terms is actually a delicate issue. The reason is the use

[1]Here $\pm$ means taking the sign which yields the highest energy.

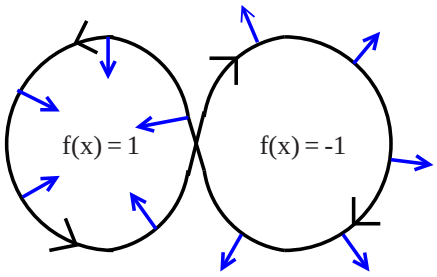


Fig. 2. The trouble with self-intersecting curves: the considered algorithm provides curve normals $\big(C'(s)\big)^\perp$ (shown in blue) that are partially inner normals and partially outer ones. With these normals, the region integral is evaluated as $\int_0^{\mathcal{L}(C)} \vec{v}_f(C(s)) \cdot \big(C'(s)\big)^\perp \, ds = -2$, where $\operatorname{div} \vec{v}_f = f$. A correct application of the Gauss-Green theorem using consistent normals gives the correct region integral $\int_{C_{in}} f \, dx = 0$.

of a contour-based algorithm which does not exclude self-intersecting curves. Alternatively one could use the region-based method in [43]. However, this method presently allows only for a very restricted class of functionals.

In principle the Gauss-Green theorem is valid for a large class of self-intersecting curves, e.g. the one in Figure 2. The problem lies on the side of the optimization algorithm: to calculate the region integral by processing along the curve, one either needs consistently the outer curve normal everywhere or consistently the inner normal. Yet, the optimization algorithm provides the outer normal for a part of the curve and the inner normal for the other part. To get consistent normals one would have to consider the entire curve at once. Yet, to get a polynomial time complexity the algorithm can only keep track of small fragments of the curve.

Therefore, given a vector field $\vec{v}_f$ such that $\operatorname{div} \vec{v}_f = f$, the region integral for the oriented curve in Figure 2 is calculated as

$$\int_0^{\mathcal{L}(C)} \vec{v}_f(C(s)) \cdot \big(C'(s)\big)^\perp \, ds = -2$$

because $\big(C'(s)\big)^\perp$ is defined as the unit tangent $C'(s)$ rotated by $+\frac{\pi}{2}$. However, a correct computation should give

$$\int_{C_{in}} \operatorname{div} \vec{v}_f \, dx = \int_{C_{in}} f \, dx = 0.$$

This implies that self-intersecting curves are not assigned the desired costs which makes the inclusion of region terms a non-trivial design task: one needs to make sure that the costs assigned to self-intersecting curves do not form the global minimum of the arising modified optimization problem. This seems to impose severe limits on the use of region terms. For this reason we will not consider region terms in the remainder of this paper.

IV. Introducing Curvature Regularity into Ratio Optimization

We have previously seen that the ratio of flux over length often leads to very small minimizing curves. One way to get longer curves is the addition of region integrals. In the last section we discussed the problems with such region terms.

As mentioned in the introduction, we propose a different remedy to the problem: in addition to the length of the curve, we will also penalize its curvature and thus consider the minimization of (1):

$$\min_C \frac{\int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^\perp \, ds}{\nu \, \mathcal{L}(C) + \int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q \, ds}$$

Why should this particular energy help to avoid small curves more than the length ratio? The reason is that the curvature term discourages direction changes: the curvature along a line segment is zero. Hence the curvature term will not grow in parts where the curve goes straight so that for long and not oscillating curves the denominator will be roughly similar to the length only and the ratio will not differ much from the length ratio. In contrast, for small curves, the curvature term will be much larger than the length so the ratio will be much smaller, in absolute value, than the length ratio.

As the minimization of (1) will be intensively studied in this paper, we now prove the existence of minimizers under some mild assumptions on the image I and on the length of curves.

A. Existence of minimizers

Assuming that Ω is bounded, the existence of minimizers of (1) is rather easy to prove among closed curves with length uniformly bounded by a constant and that admits a uniform parameterization in the Sobolev space $\mathrm{W}^{2,q}([0,1],\bar{\Omega})$, $q > 1$ [31]. Remark that the constant can be chosen arbitrarily large. For simplicity, we assume that the image I is continuously differentiable on $\bar{\Omega}$.

Remark that if $\int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^\perp \, ds < 0$ then changing the orientation of C also changes the sign of the integral. Thus, minimizing (1) is equivalent to maximizing

$$\frac{\left| \int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^\perp \, ds \right|}{\nu \, \mathcal{L}(C) + \int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q \, ds} \tag{6}$$

Using a uniform parameterization on $[0,1]$ this ratio can be rewritten as

$$\frac{\left| \int_0^1 \nabla I(C(t)) \cdot C'(t)^\perp \, dt \right|}{\nu \, \mathcal{L}(C) + [\mathcal{L}(C)]^{1-2q} \int_0^1 |C''(t)|^q \, dt} \tag{7}$$

where $C'(t)^\perp$ denotes the vector $C'(t)$ rotated by $\frac{\pi}{2}$. Let $A > 0$ and define

$$\mathrm{W}^{2,q}_A([0,1],\bar{\Omega}) = \{C \in \mathrm{W}^{2,q}([0,1],\bar{\Omega}),\ \mathcal{L}(C) \leq A\}.$$

We assume in the sequel that I and A are such that there exists at least a simple closed curve in $\mathrm{W}^{2,q}_A([0,1],\bar{\Omega})$ for which the numerator in (7) is non zero, otherwise the problem is trivial.

Take a maximizing sequence $(C_n)_{n\in\mathbb{N}}$ of simple closed curves in $\mathrm{W}^{2,q}_A([0,1],\bar{\Omega})$ with uniform parameterization on $[0,1]$. With no loss of generality, we can assume that there exists $a_1 > 0$ such that, for every $n \in \mathbb{N}$

$$\frac{\left|\int_0^1 \nabla I(C_n(t)) \cdot C_n'(t)^\perp \, dt\right|}{\nu\,\mathcal{L}(C_n) + [\mathcal{L}(C_n)]^{1-2q} \int_0^1 |C_n''(t)|^q \, dt} \geq a_1$$

Due to the regularity of the image I, there exists also a_2 such that

$$\left|\int_0^1 \nabla I(C_n(t)) \cdot C_n'(t)^\perp \, dt\right| \leq a_2 \mathcal{L}(C_n) \leq a_2 A$$

thus

$$\nu\,\mathcal{L}(C_n) + [\mathcal{L}(C_n)]^{1-2q} \int_0^1 |C_n''(t)|^q \, dt \leq \frac{a_2 A}{a_1} \tag{8}$$

and therefore there exists a constant a_3 such that for every $n \in \mathbb{N}$

$$\int_0^1 |C_n''(t)|^q \, dt \leq a_3$$

Observing that Ω is bounded and $|C_n'(t)| = \mathcal{L}(C_n) \leq A$ for every $t \in [0,1]$ and every $n \in \mathbb{N}$, due to the assumption of uniform parameterization, we conclude that the sequence $(C_n)_{n\in\mathbb{N}}$ is uniformly bounded in $\mathrm{W}^{2,q}([0,1],\bar{\Omega})$. Therefore (see for instance [31]), there exists a subsequence, still denoted as $(C_n)_{n\in\mathbb{N}}$, that converges weakly in $\mathrm{W}^{2,q}([0,1],\bar{\Omega})$ and strongly in $\mathrm{C}^1([0,1],\bar{\Omega})$ to a limit curve C. In addition, $\mathcal{L}(C_n) \to \mathcal{L}(C)$ – in particular $C \in \mathrm{W}^{2,q}_A([0,1],\bar{\Omega})$ – and

$$\int_0^1 |C''(t)|^q \, dt \leq \liminf_{n\to\infty} \int_0^1 |C_n''(t)|^q \, dt. \tag{9}$$

Remark that C is not necessarily simple since tangential auto-contacts may occur in the limit.

Let us now check that the limit curve C has strictly positive length. From (8), we deduce that

$$\int_0^{\mathcal{L}(C_n)} |\kappa_{C_n}(s)|^q \, ds = [\mathcal{L}(C_n)]^{1-2q} \int_0^1 |C_n''(t)|^q \, dt \leq \frac{a_2 A}{a_1}$$

Extending Fenchel's Theorem [22][Theorem 5.7.3] to $\mathrm{W}^{2,q}$ curves by approximation, we know that for every $n \in \mathbb{N}$, $\int_0^{\mathcal{L}(C_n)} |\kappa_{C_n}(s)| \, ds \geq 2\pi$. By the Hölder inequality, it follows that

$$\mathcal{L}(C_n)^{q-1} \int_0^{\mathcal{L}(C_n)} |\kappa_{C_n}(s)|^q \, ds \geq (2\pi)^q$$

thus $\mathcal{L}(C_n)^{q-1} \geq \frac{a_1 (2\pi)^q}{a_2 A}$. Passing to the limit, we conclude that $\mathcal{L}(C) > 0$. Therefore, we can deduce from (9) that

$$\nu\,\mathcal{L}(C) + \frac{\int_0^1 |C''(t)|^q \, dt}{[\mathcal{L}(C)]^{2q-1}} \leq \liminf_{n\to\infty} \left(\nu\,\mathcal{L}(C_n) + \frac{\int_0^1 |C_n''(t)|^q \, dt}{[\mathcal{L}(C_n)]^{2q-1}}\right).$$

Besides, the continuity of ∇I and the pointwise convergence of $C_n(t)$ to $C(t)$ and $C_n'(t)$ to $C'(t)$ for every $t \in [0,1]$ imply that

$$\int_0^1 \nabla I(C_n(t)) \cdot C_n'(t)^\perp \, dt \to \int_0^1 \nabla I(C(t)) \cdot C'(t)^\perp \, dt$$

and we finally get that

$$\frac{\left|\int_0^1 \nabla I(C(t)) \cdot C'(t)^\perp \, dt\right|}{\nu\,\mathcal{L}(C) + [\mathcal{L}(C)]^{1-2q} \int_0^1 |C''(t)|^q \, dt} \geq \limsup_{n\in\mathbb{N}} \frac{\left|\int_0^1 \nabla I(C_n(t)) \cdot C_n'(t)^\perp \, dt\right|}{\nu\,\mathcal{L}(C_n) + [\mathcal{L}(C_n)]^{1-2q} \int_0^1 |C_n''(t)|^q \, dt}$$

The sequence (C_n) being maximizing, we conclude that C is a curve – limit of simple curves – that maximizes (6) in $\mathrm{W}^{2,q}_A([0,1],\bar{\Omega})$. Remark that the same proof could be used to establish the existence of minimizers among all $\mathrm{W}^{2,q}$ curves, simple or non simple.

B. The Class of Optimizable Functionals

The discrete approach that we use can handle a much wider class of functionals with a great variety of data and regularity terms: the integrands in both numerator and denominator can depend on the point $C(t)$ of the curve, on the tangent vector $C'(t)$ and the curvature $\kappa_C(t)$ at $C(t)$, so that the class of solvable problems reads:

$$\frac{\int_0^{\mathcal{L}(C)} h(C(s), C'(s), \kappa_C(s)) \, ds}{\int_0^{\mathcal{L}(C)} g(C(s), C'(s), \kappa_C(s)) \, ds} \tag{10}$$

Here $h : \Omega \times \mathbb{S}^1 \times \mathbb{R}^2 \to \mathbb{R}$ is arbitrary and the integral of $g : \Omega \times \mathbb{S}^1 \times \mathbb{R}^2 \to \mathbb{R}^+$ over any closed curve with positive length must be strictly positive.

In the next section, we will describe how ratios of the form (10) can be globally optimized after a suitable discretization. It should be noted that self-intersecting curves can occur. Yet, in our experiments we observed them only for very small length weights ν.

V. The Optimal Curve as Cycle in a Graph

To globally optimize functionals of form (10) we discretize the space of all possible curves: a curve is now defined as a contiguous subset among finitely many line segments. Nevertheless we have the continuous optimization task in mind and the optimal discrete curve is viewed as a polygonal approximation of the optimal continuous one: its length and curvature estimates approximate the values of the continuous solution. Convergence is discussed in the next section.

To be able to use the path-based method described later on, we build a graph where each edge e is assigned a numerator edge weight $n(e)$ and a denominator edge weight $d(e)$. Each closed curve C in the discrete search space corresponds to some cycle Γ in the graph. In the end we will minimize the ratio problem

$$\min_{\Gamma} \frac{\sum_{e\in\Gamma} n(e)}{\sum_{e\in\Gamma} d(e)} \tag{11}$$

The numerator and denominator sums should hence reflect the respective integrals in (10).

For the length ratio (4) Jermyn and Ishikawa [38] build a graph with one node for each image pixel and edges connecting pixels in an 8-neighborhood. The edge weights correspond to the respective integrals along the arising line segments.

For functionals of form (10) a more elaborate graph structure is needed: the optimization algorithm supports only edge weights depending on *single* edges. However, if edges directly correspond to line segments, one cannot approximate the curvature of the desired continuous curve: line segments always have zero curvature.

In [53] we used a graph where a node corresponds to a pair of an image pixel and an incoming direction. In this work we take a slightly different approach: each node in the graph corresponds to a pair of image pixels. More precisely the search space consists of all pairs of pixels that are spaced apart from one another not more than a certain distance R : if $\mathcal{P}$ is the pixel set of the image, the node set $\mathcal{V} \subset \mathcal{P}^2$ of the graph is expressed as (with $|\cdot|$ the L_2-norm)

$$\mathcal{V} = \{\, (\vec{p}_1, \vec{p}_2) \,|\, 0 < |\vec{p}_1 - \vec{p}_2| \leq R \,\}.$$

Edges in the graph connect nodes sharing an image pixel. More precisely the edge set is

$$\mathcal{E} = \{\, \big((\vec{p}_1, \vec{p}_2), (\vec{p}_2, \vec{p}_3)\big) \in \mathcal{V}^2 \,\}.$$

An edge now represents an oriented polygonal curve consisting of two line segments, both of length $\leq R$. We now turn to the question of how to define suitable edge weights to approximate the continuous functional (10).

A. Estimating Curvature, Normals and Tangent Angles

To define the edge weights one first needs to calculate the tangent vector – or, equivalently, the tangent angle with respect to the x-axis – and the curvature of the corresponding part of the curve. Recall that an edge represents a polygonal curve consisting of two adjacent line segments, say $\vec{p}_1\vec{p}_2$ and $\vec{p}_2\vec{p}_3$. For these two line segments the corresponding tangent angles with respect to the x-axis $\alpha_{1,2}$ and $\alpha_{2,3}$ are computed using the C++-function `atan2` on the difference vector of the respective end points. This vector also allows to calculate curve normals: to this end the vector is normalized and rotated by 90 degrees.

Estimating curvature is a more difficult issue. To allow optimal convergence properties we follow the results of Bruckstein, Netravali and Richardson [10]. Denoting the length of the two line segments $l_{1,2}$ and $l_{2,3}$, the absolute curvature at point $\vec{p}_2$ is estimated as

$$|\kappa|(\vec{p}_1, \vec{p}_2, \vec{p}_3) = \frac{|\alpha_{1,2} - \alpha_{2,3}|_{\mathbb{S}^1}}{\frac{1}{2}\min(l_{1,2}, l_{2,3})} \tag{12}$$

where the angle difference is taken on the manifold $\mathbb{S}^1$ to correctly account for the jump over 2π. The sign of the curvature (if needed) is determined by whether the angle between the line segments exceeds 180 degrees or not. We discuss below convergence properties of this process.

B. Computing the Edge Weights

We first give edge weights for the two ratios (1) and (2) considered in the experimental section. For the numerator we discretize each line segment via the method of Bresenham [9]. Then, we evaluate the data term for each pixel using the above mentioned segment normal and sum the obtained values.

For the denominator we evaluate length-based and curvature-based terms separately. The length of the curve is readily calculated as the sum of all line segment lengths. For the curvature term we evaluate the expression (12) and take the desired power of it. Finally, to get the integral of this term, the length of the segment needs to be included in the weights. Here again we follow the results of [10], summing over all edges $(\vec{p}_1, \vec{p}_2, \vec{p}_3)$ the quantity

$$\tfrac{1}{2}\min(l_{1,2}, l_{2,3}) \cdot [|\kappa|(\vec{p}_1, \vec{p}_2, \vec{p}_3)]^q \,.$$

The general case (10) is more difficult to handle: e.g. it could contain terms like $\int_0^{\mathcal{L}(C)} I(C(s))\,|\kappa_C(s)|\,ds$ where the integrand depends on position and curvature simultaneously. This makes it much harder to reflect the continuous functional in terms of discrete sums. Our solution is presently to calculate the Bresenham lines, then calculate the values of $h(\cdot,\cdot,\cdot)$ and $g(\cdot,\cdot,\cdot)$ for each pixel, using the above given estimates for normals and curvature. These values are then summed to form the edge weights.

C. Convergence of the Process

We defer the question of how to compute the optimal polygonal curve to the next section and assume for the moment that it is solved. We should mention, however, that the optimization algorithms requires the weights $n(e)$ and $d(e)$ to be subject to some regular quantization, i.e. they must be multiples of a certain $\epsilon > 0$. Termination is guaranteed for any such quantization, but the complexity depends on ϵ. Details are given in section VI-D.

We prove now that the limit of a converging sequence of discrete simple minimizers is a minimizer of (1) in the continuous domain. Let us first recall that the usual way to study relations between discrete and continuous minimizers involves a particular notion of convergence for functionals, the Γ-convergence [20]. It has the particularly useful property: if a sequence of energy functionals F_n Γ-converges to a functional F and a sequence (x_n) of minimizers of F_n converges to x then x is a minimizer of F. In this framework the results of Bruckstein et al. in [10] are directly related to our problem. Bruckstein et al. consider the space of rectifiable curves with finite total absolute curvature endowed with the metric d defined by

$$d(C_1, C_2) = \inf_{\Psi:[0,1]\to[0,1]} \sup_{t\in[0,1]} |C_1(t) - C_2(\Psi(t))|$$

with C_1, C_2 parameterized on $[0,1]$ and Ψ in the class of all homeomorphisms from $[0,1]$ to $[0,1]$. Then they prove, using the discrete definition of curvature (12) and using d as convergence metric for sequence of curves, that the discrete counterpart of $\int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q\,ds$ computed on polygons with n edges Γ-converges to $\int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q\,ds$ as n tends to ∞

and the maximal length of polygon edges tends to zero. Now remark that the existence (in the continuous domain) of a curve maximizing (6) is equivalent – if (6) is not trivially zero – to the existence of a curve minimizing

$$\frac{\nu\,\mathcal{L}(C) + \displaystyle\int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q\,ds}{\left|\displaystyle\int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^{\perp}\,ds\right|} \tag{13}$$

in the class of $\mathrm{W}^{2,q}$ curves with length uniformly greater than a suitable constant. If $\frac{1}{n}$ denotes the pixel size, let us define F_n as the functional that associates any polygon P_n defined on the grid with

$$F_n(P_n) = \sum_{e \in P_n} d(e)$$

where $d(e)$ is computed as in the previous section and P_n is assumed to have a maximal edge length smaller than $\frac{\delta}{n}$ with δ a constant independent of P_n and n. According to the result by Bruckstein et al., F_n Γ-converges to the functional

$$F(C) = \nu\,\mathcal{L}(C) + \int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q\,ds.$$

Besides, remark that the smoothness of I implies that its discrete gradient computed with finite differences uniformly converges to the continuous gradient ∇I. Take any sequence of simple polygons (P_n) with uniformly bounded length that converges for the metric d to a limit curve C. Let $\mathrm{int}(P_n)$ and $\mathrm{int}(C)$ denote the sets enclosed by P_n and C, respectively, and $\mathbb{1}_{\mathrm{int}(P_n)}$, $\mathbb{1}_{\mathrm{int}(C)}$ the associated characteristic functions. By the theory of functions of bounded variation [1] and possibly taking a subsequence, the derivatives $D\mathbb{1}_{\mathrm{int}(P_n)}$ weakly-$\star$ converge to $D\mathbb{1}_{\mathrm{int}(C)}$ as $n \to \infty$. It follows from the Gauss-Green Theorem for BV functions [1] that

$$\sum_{e \in P_n} n(e) \to \int_0^{\mathcal{L}(C)} \nabla I(C(s)) \cdot \big(C'(s)\big)^{\perp}\,ds$$

and we deduce that the ratio $\dfrac{\sum_{e\in P_n} d(e)}{\left|\sum_{e\in P_n} n(e)\right|}$ Γ-converges to (13) as n tends to ∞. Therefore, taking a sequence of simple discrete minimizers of this ratio, there exists a subsequence that converges to a minimizer of (13) in the continuous domain. Such minimizer being non degenerate according to our assumption that the length is uniformly bounded from below, we conclude that for any sequence of simple discrete minimizers of $\dfrac{\sum_{e\in\Gamma_n} n(e)}{\sum_{e\in\Gamma_n} d(e)}$, there exists a subsequence that converges to a minimizer of (1) as $n \to \infty$. This achieves the proof of convergence.

VI. Ratio Optimization over Cycles in a Graph

We now address the task of finding the cycle of optimal ratio in the described graph, i.e. how to solve the task

$$\min_{\Gamma} \frac{\displaystyle\sum_{e\in\Gamma} n(e)}{\displaystyle\sum_{e\in\Gamma} d(e)}$$

Minimum Ratio Cycle Algorithm

Input: A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with two edges weights $n(e)$ and $d(e)$ for each edge.
Output: A cycle Γ minimizing the ratio $\sum_{e\in\Gamma} n(e) / \sum_{e\in\Gamma} d(e)$.

1) Find an upper bound λ on the optimal ratio λ_{opt}
2) Compute edge weights $w(e) = n(e) - \lambda d(e)$ for each edge $e \in \mathcal{E}$.
3) Call the Moore-Bellman-Ford algorithm (Fig. 4) for the graph $\mathcal{G}$ and the edge weights w. If it returns a negative cycle, set λ to its ratio and go to 2). Otherwise output the last found cycle and **stop**.

Fig. 3. Ratio optimization after Lawler [44]. Shown is the linear search variant, also known as Dinkelbach's method [21].

over all cycles Γ, where $n(e)$ and $d(e)$ are subject to a regular ϵ-quantization with $d(e) \geq 0$ and such that the denominator sum is strictly positive for any cycle in the graph. To this end we use a variant of the Minimum Ratio Cycle algorithm proposed by Lawler [44]: instead of binary search we use linear search as proposed by Dinkelbach [21], which is much faster in practice.

The basic algorithm is shown in Figure 3. It is based on iterated negative cycle detection in a graph with single edge weights. Let λ be some ratio and define edge weights

$$w(e) = n(e) - \lambda d(e).$$

Now suppose the graph contains a negative cycle Γ w.r.t. the edge weights $w(e)$. By applying equivalence transformations one sees that any such cycle must be of better ratio than λ and vice versa:

$$\begin{aligned}
& \sum_{e\in\Gamma} w(e) < 0 \\
\Leftrightarrow \quad & \sum_{e\in\Gamma} [n(e) - \lambda d(e)] < 0 \\
\Leftrightarrow \quad & \sum_{e\in\Gamma} n(e) < \lambda \cdot \sum_{e\in\Gamma} d(e) \\
\Leftrightarrow \quad & \frac{\sum_{e\in\Gamma} n(e)}{\sum_{e\in\Gamma} d(e)} < \lambda
\end{aligned}$$

Notice that the third line is valid only because of the positivity of all conceivable denominator sums. This is the reason for the previously introduced restriction on the denominator.

The above equivalence transformation shows that the graph contains a negative cycle w.r.t. $w(e)$ if and only if the optimal ratio is lower than λ. If one is able to find negative cycles, this immediately gives rise to the algorithm in Figure 3: starting from some upper bound on the optimal ratio, negative cycle detection and ratio adjustments are alternated. Every time a negative cycle is found, λ is set to its ratio. The last found cycle must be of optimal ratio.

Negative cycle detection is performed efficiently by the Moore-Bellman-Ford algorithm [33], [47], [5] for distance calculations. The algorithm, depicted in Figure 4, is based on dynamic programming: starting from an initial distance labeling the distance label of any node is reduced whenever the labels of its predecessors allow such an improvement. If the graph does not have negative cycles, the algorithm terminates

Moore-Bellman-Ford Algorithm

Input: A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with (possibly negative) edge weights $w(e)$ for each edge. A root node $r \in \mathcal{V}$.
Output: A distance label $d(v)$ and a predecessor node $p(v)$ for every node $v \in \mathcal{V}$ in the graph. If the graph contains negative cycles such a cycle is returned.

1) Set $d(r) = 0$, $d(v) = \infty$ for $v \in \mathcal{V} \backslash \{r\}$. Mark $p(v)$ as invalid for all v.
2) Set `changes := false`
   For all $v \in \mathcal{V}$: check all incoming edges $e = (w, v)$. **If** $d(w) + w(e) < d(v)$
   $d(v) = d(w) + w(e)$, $p(v) = w$
   `changes := true`
3) **If** `changes = false` **stop.**
   Otherwise check the predecessor entries p for cycles. If a cycle is found, return the cycle. Else go to 2).

Fig. 4. Distance calculation and negative cycle detection via the Moore-Bellman-Ford algorithm [33], [47], [5].

with the correct distance labeling. Otherwise, after a few steps the parent entries will permanently contain cycles. Regularly checking for cycles then allows to extract a negative cycle, which is necessary to update the ratio.

While the basic algorithm in Figure 3 must be carried out sequentially, the negative cycle detection in Figure 4 allows a lot of freedom for the implementation. We now discuss how to efficiently implement negative cycle detection, both in a sequential and in a parallel way. The key for efficiency lies in how to implement step 2) in Figure 4. Concerning the numerical implementation we noticed that both double precision and integer optimization lead to the global optimum. Since double precision is not available on current GPUs, we use integer operations for both implementations.

A. Sequential Negative Cycle Detection

Efficient sequential implementations make use of a queue for implementing step 2 in Figure 4. Nodes whose distance label *cannot* change in the present iteration (because none of their neighbors changed their label in the last one) will then not be visited. Every time the distance label of a node is changed, the node is added to the end of a queue. As long as there are nodes in the queue, the front one is removed and its neighbors are checked for possible distance improvements. While the worst case complexity remains the same, in practice significant speed-ups are obtained.

To optimize the run-time an explicit representation of the entire graph is suitable. However, its memory consumption is very high: only images up to size 256×256 can be processed with 2 GigaByte of memory. We therefore implemented a version where edges (and their weights) are computed on-the-fly. This solves the memory issues, but increases the run-time significantly.

B. Parallel Negative Cycle Detection

State-of-the-art graphics hardware allows highly parallel implementations of a certain class of algorithms. This class does not contain the queue-based implementation just described. However, in the form described in Figure 4, step 2) can be implemented in parallel. To this end one uses two buffers of distance labels, where the second is updated based on the first. Distances and parent pointers are stored in matrices, i.e. there are no node structures at all. The cycle check is done on the CPU every 25 iterations, its computational costs (including memory transfer between GPU and CPU) are negligible in practice.

C. Choosing the Root Node

For the Moore-Bellman-Ford algorithm for distance calculation (Fig. 4) a root node must be fixed. While the choice of this root node does not affect the optimality property of the ratio optimization process, it can have significant influence on the performance.

For the parallel implementation it is useful to add an extra root node and connect it to every node by an edge weighted with 0. This amounts to initializing all distance labels with 0. After k iterations the distance label of any node contains the weight of the cheapest path of length k passing through it. While in theory one can still have $|\mathcal{V}|$ iterations until a negative cycle arises, in practice we expect a number of iterations in the order of the length of the most negative cycle in the graph.

This initialization could be used for the sequential implementation as well. However, we do not consider this sensible: first of all, the memory requirements are high since initially every node in the graph is added to the queue. Also one will have to visit every node in the graph at least once, which reduces the efficiency of the method in practice. For the first negative cycle detection we choose a root node in the center of the image. In subsequent calls the root node is selected as one of the nodes in the last found cycle.

D. Complexity of the Method

The described graph to estimate curvature contains $\mathcal{O}(|\mathcal{P}| \cdot R^2)$ nodes. Since each node is connected with $\mathcal{O}(R^2)$ neighbors, there are $\mathcal{O}(|\mathcal{P}| \cdot R^4)$ edges. The Moore-Bellman-Ford algorithm is known to terminate in time $\mathcal{O}(nm)$ on a graph with n nodes and m edges. This gives us a worst case complexity of $\mathcal{O}(|\mathcal{P}|^2 R^6)$ for one negative cycle detection.

Finally there is the issue of the number of distance calculation that need to be performed. Let $\epsilon > 0$ be the level of quantization, w_n be the maximum absolute numerator weight and w_d the maximal denominator weight, both before quantization. One can show [38] that the number of iterations is then $\mathcal{O}(m^3 w_d^2 w_n / \epsilon^3)$ in the worst case, with m the number of edges. In practice the number of iterations is less than 50 for $\epsilon = 10^{-3}$ and a radius $R = 3$.

In total this results in a run-time of $\mathcal{O}(|\mathcal{P}|^5 R^{18} w_d^2 w_n / \epsilon^3)$. While this seems very high, in practice we observe a linear dependence on the number of image pixels. On the GPU, even images of size 640×480 are processed in less than half an hour using a radius of 3.

VII. Minimum Ratio Cycles and Snakes

In this section we show that the presented class of optimizable ratio functionals allows to draw conclusions about a

parameterization-invariant version of the snakes model. In the original work of Kass et al. [40] the model was stated as

$$-\int_0^1 |\nabla I(C(s))|^2\,dt + \alpha\int_0^1 |C'(t)|^2\,dt + \beta\int_0^1 |C''(t)|^2\,dt.$$

This expression is not invariant to the chosen parameterization of the curve. Kass et al. probably chose this formulation as it allows to remove numerical instabilities when dealing with explicit parameterizations of the curve. From a today's perspective one would want a parameterization-invariant formulation which might read like this:

$$-\int_0^{\mathcal{L}(C)} |\nabla I(C(s))|^2\,ds + \lambda\nu\mathcal{L}(C) + \lambda\int_0^{\mathcal{L}(C)} |\kappa_C(s)|^2\,ds \tag{14}$$

Here we have modified the role of the weighting parameters: one is now given a relative weight $\nu > 0$ to balance the relation of length and curvature regularity as well as a weight $\lambda > 0$ to influence both regularity terms at once.

This functional is closely related to the *snakes ratio* we introduced in (2) and recall here for convenience:

$$\frac{-\int_0^{\mathcal{L}(C)} |\nabla I(C(s))|^p\,ds}{\nu\mathcal{L}(C) + \int_0^{\mathcal{L}(C)} |\kappa_C(s)|^q\,ds}$$

When applying the Minimum Ratio Cycle algorithm to this ratio with $p = q = 2$, one ends up computing a $\lambda_{\text{opt}} \leq 0$ and an optimal curve C_{opt} such that

$$-\int_0^{\mathcal{L}(C_{\text{opt}})} |\nabla I(C_{\text{opt}}(s))|^2\,ds + |\lambda_{\text{opt}}|\nu\mathcal{L}(C_{\text{opt}}) + |\lambda_{\text{opt}}|\int_0^{\mathcal{L}(C_{\text{opt}})} |\kappa_{C_{\text{opt}}}(s)|^2\,ds = 0$$

and any other curve has larger energy with respect to the same parameter $|\lambda_{\text{opt}}|$. Hence, the snakes ratio provides valuable insights into the modified snakes model (14): given a relative weight ν between length and curvature regularity, minimizing the snakes ratio provides an absolute regularity weight $|\lambda_{\text{opt}}|$ for which the parameterization-invariant snakes model (14) has a meaningful optimum and the algorithm also provides the associated optimal curve. This means that now a model can be optimized globally for which previously only local solutions were available.

VIII. Experiments

On several images from different domains we demonstrate the performance of the proposed method. We focus on the elastic ratio, but also discuss the snakes ratio (2). In particular we show:

- The elastic ratio with squared curvature allows object segmentation for a large variety of domains. The length weight ν was adjusted experimentally on a variety of images. We found 0.15 to give reliable results.
- Our fully unsupervised method is able to outperform region-based methods: it is less sensitive to shading effects which allows to find more precise boundaries.

Image		Run-time		
Name	Resolution	CPU-expl.	CPU-impl.	GPU
Seal	200×133	364s	812s	35s
Bunny	260×180	1567s	1593s	101s
Berkeley #3	321×481	N/A	8673s	1046s
Baseball #2	450×314	N/A	14810s	151s

TABLE I

Comparison of run-times for the different implementations. For the CPU run-time with explicit storage of edges and with on-the-fly computation are shown. Experiments were run on the same machine and using compiler optimization.

- Our method is robust to noise, i.e. even for very noisy pictures it produces results comparable to those on noise-free pictures. We stress that there is no need to adjust any parameters.

A. Length Ratio vs. Elastic Ratio

In Figure 5 we show a comparison of ratio functionals on images containing objects in front of cluttered background. In two cases the length ratio finds a meaningful object: here the entire object boundary has a high contrast. On these images the elastic ratio produces comparable results.

In the majority of cases, however, the length ratio tends to find small homogeneous regions. The figure shows that in many of these cases the elastic ratio is able to locate meaningful objects. This trend is confirmed by Figure 6, where we show some results on the Berkeley database.

For this reason Jermyn and Ishikawa [38] proposed to integrate a suitably weighted balloon force. Figure 7[2] demonstrates that there are fairly large parameter ranges giving rise to almost identical segmentations. The functional therefore seems robust to the choice of the area weight. However, there is no parameter which works well for all the shown images.

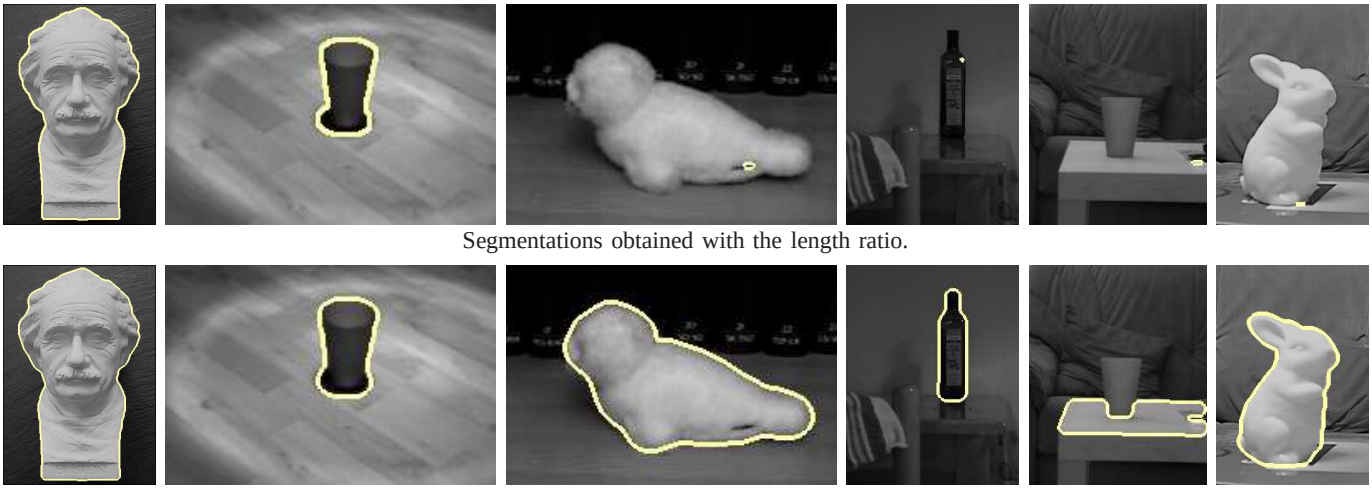
Nonetheless the extended length ratio produces meaningful objects in several cases and we consider it somewhat complementary to the elastic ratio: each gives rise to segmentations that cannot be produced with the respective other one. In general, whether one wants to favor objects with large area or with low curvature of the region boundary will depend on the application.

B. Efficiency on CPU and GPU

Due to the large search space an efficient optimization of the elastic ratio is desirable. We implemented the algorithm both on the CPU and on the GPU as described in Section VI.

The run-times for several images are given in Table I. For the smallest image the explicit graph uses roughly half the system memory. Here the explicit storage of edges is about twice as fast as the implicit one – both on the CPU. For the second image the two perform almost identical: here the entire system memory of 4 GB was needed for explicit storage.

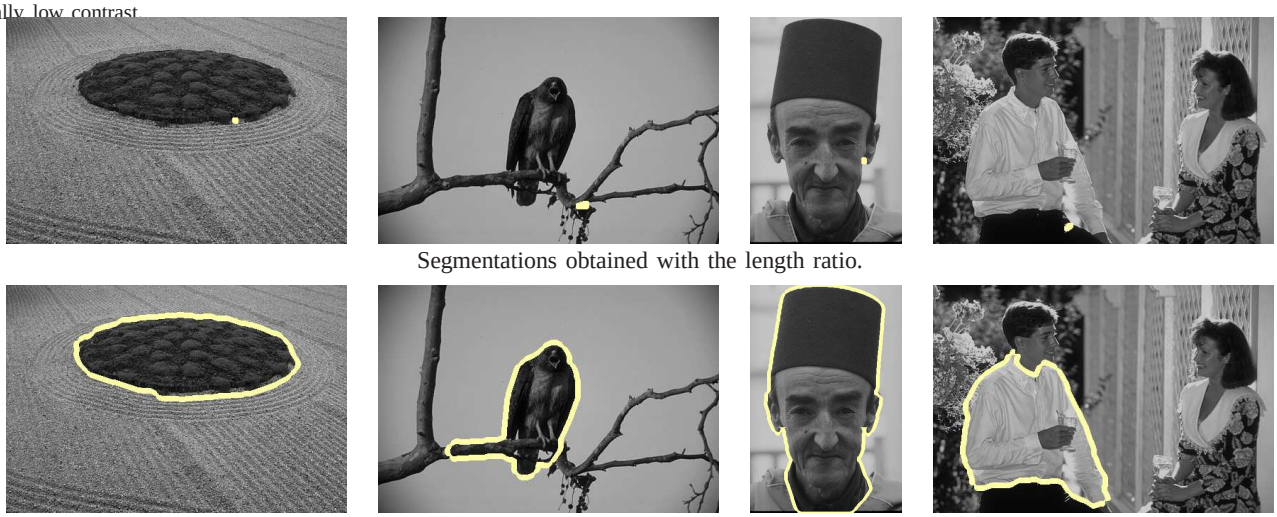
[2]We thank Greg Mori for sharing his data with us.



Segmentations obtained with the length ratio.

Segmentations obtained with the elastic ratio.

Fig. 5. The elastic ratio gives rise to more meaningful segmentations than the length ratio. In particular, it is able to find objects in the presence of partially low contrast.



Segmentations obtained with the length ratio.

Segmentations obtained with the elastic ratio.

Fig. 6. The results on the Berkeley database confirm: the elastic ratio is better suited for object segmentation than the length ratio.

The speed-up of the GPU version over the CPU one with implicit storage is between a factor of 8 and a factor of 100. The huge deviations are due to the different natures of the algorithms (queue-based vs. full parallel). In particular, these differences result in a different sequence of intermediate ratios.

All given run-times are quite high, lying above half a minute. However, we once again emphasize that the proposed method separates objects from the background in a *fully* unsupervised manner (i.e. does not rely on any user input).

C. Robustness and Comparison to Region-based Approaches

For a comparison to region-based approaches we implemented variants of Mumford-Shah-like functionals [49][3]:

$$E(u_1,u_2,\Omega_1)=\sum_{i=1,2}\int_\Omega \frac{(I-u_i)^2}{\sigma_I^2}\,1_{\Omega_i}+\lambda|\nabla u_i|^2dx+\nu|\partial\Omega_i|. \tag{15}$$

where a piecewise smooth approximation by two functions $u_1,u_2:\Omega\rightarrow\mathbb{R}$ and a partition of the image plane Ω into two disjoint regions Ω_1 and Ω_2 is computed by alternating globally optimal updates for u_1,u_2 and Ω_1. The data fidelity terms – normalized with respect to the intensity variance σ_I^2 – are only imposed in the regions indicated by the characteristic functions 1_{Ω_i} associated with region Ω_i, and $|\partial\Omega_i|$ denotes the Euclidean boundary length of Ω_i. The update with respect to u_i (for fixed Ω_i) is obtained by solving the Euler-Lagrange equations

$$(I-u_i)\,1_{\Omega_i}\;+\;\lambda\,\sigma_I^2\,\Delta u_i=0,\quad i=1,2 \tag{16}$$

using Successive Over-Relaxation (SOR). The update of Ω_1 for fixed u_i can be computed in globally optimal manner for a discrete approximation on a regular grid using graph cuts [7].[4] For $\lambda\rightarrow\infty$ we obtain piecewise constant approximations with scalar constants u_i given by the mean gray value in region Ω_i.

We know of no global optimization algorithms for either the piecewise smooth or piecewise constant version of the functional (the latter is obtained for $\alpha\rightarrow\infty$). In fact, solving

[3]Note that functional (15) is not identical with the original Mumford-Shah approach since the smoothness terms in expression (15) are extended into the entire domain Ω.

[4]A related efficient algorithm for minimizing the piecewise smooth Mumford Shah functional by alternating graph cuts and smooth approximations was independently developed and evaluated in greater detail in [34].

Fig. 7. Where the elastic ratio identifies body parts, the length ratio only finds small homogeneous regions. With a suitable area weight the extended length ratio can find meaningful regions.

the piecewise constant case reduces to the k-means problem for $\nu = 0$. The problem is known to be NP-hard.

The performances of the elastic ratio and (15) are compared in Figure 8. Already for the noise-free image we could not find a length parameter where the Mumford-Shah-type implementation separates the object from the background. For the highly noisy images, despite the adaptive smoothness terms numerous small regions arise. In contrast, the elastic ratio identifies the object almost perfectly *without* needing to adjust any parameters.

Lastly we address the robustness of the elastic ratio with respect to the length weight ν. Figure 9 demonstrates that for a fairly large range of ν the object is found. Up to a certain point the contour becomes more complex with increasing ν. Then, from this point on the length term becomes dominant and the functional approaches the length ratio.

D. Results for the Snakes Ratio

Figure 10 presents results for a slightly modified snakes ratio: for robustness we use the gradient absolute instead of its squared absolute. We recall from Section VII that all these results are global solutions of a suitably weighted parameterization-invariant reformulation of the snakes functional. Since our algorithm always finds parameter sets where the global solution has energy 0, we cannot draw conclusions about the entire functional. However, we believe that other meaningful parameter sets do not lead to significantly better results.

When using the balancing weight $\nu = 0.15$ for length against curvature – which works well for the elastic ratio –

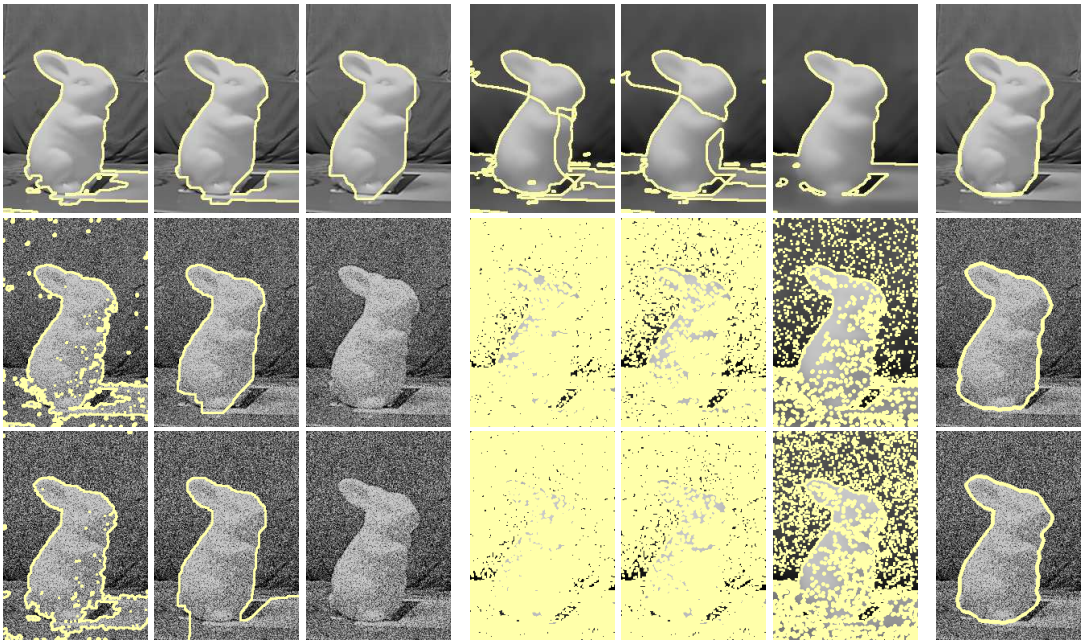


Fig. 8. Noise Robustness and Comparison to Region-based Approaches. The elastic ratio extracts the object almost perfectly. Moreover it is robust to noise, without the need to change any parameters. In contrast, both the piecewise constant Mumford-Shah [49] and the piecewise smooth approximations (see text) fail to differentiate the object from the background. Here we show local minimizers (for details see text) for three different length weights.

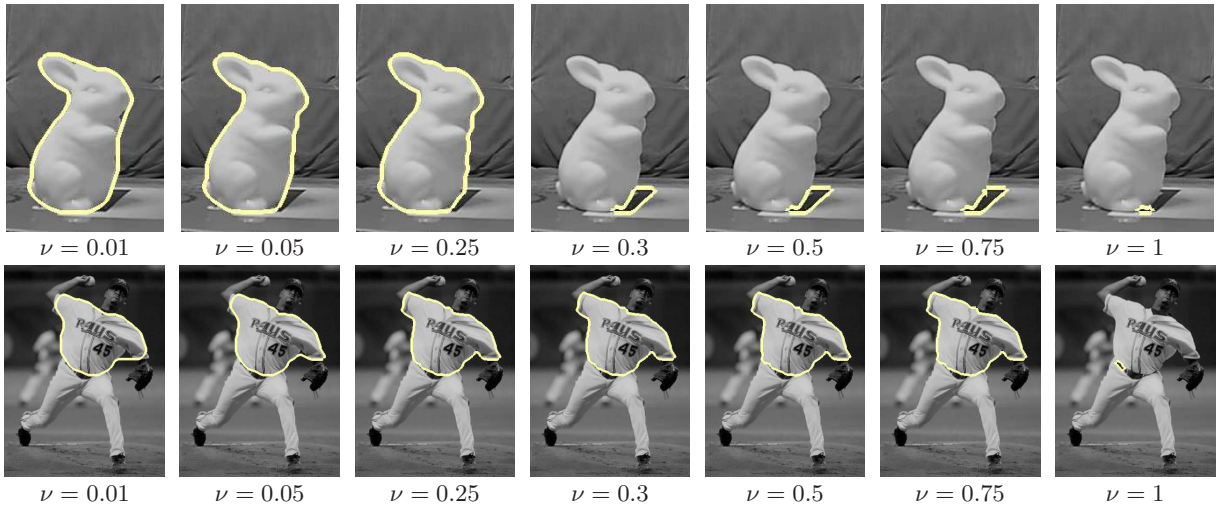


Fig. 9. Effect of the length weight on the elastic ratio: for a fairly large parameter range a meaningful part of the image is found.

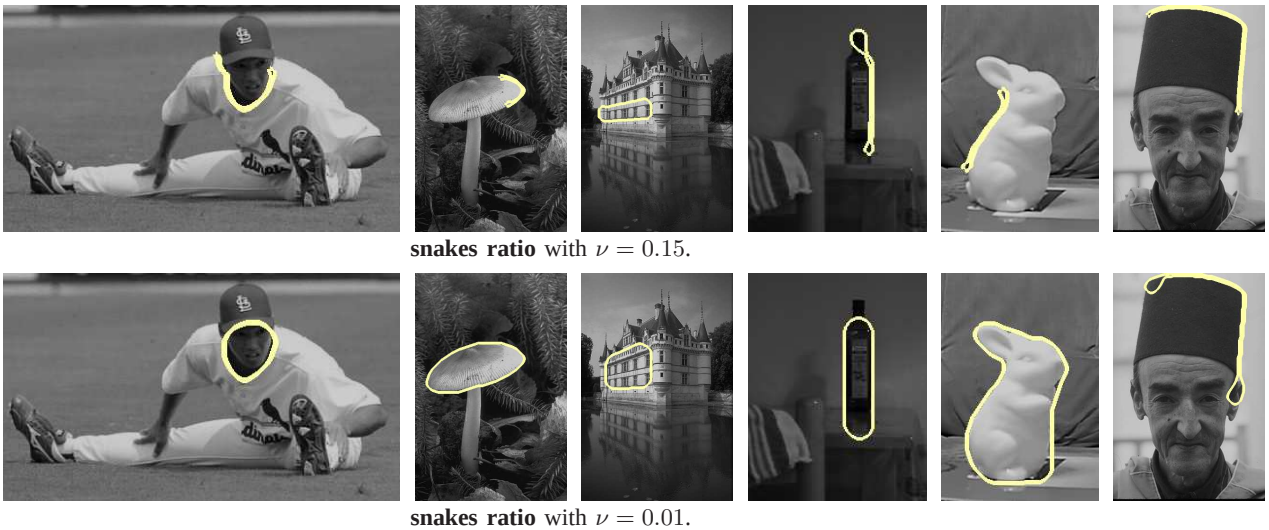


Fig. 10. The snakes ratio reveals: the gradient absolute is a bit weak as a data term.

the results are discouraging: in most cases the curve goes one way, turns around and goes almost exactly the same way back. We consider these solutions as valid in the original sense [40] since they do not self-intersect. While some line segments occur repeatedly with opposing directions of traversal, in the continuous solution space there will be a curve without repetitions and almost the same costs – at least if the image gradient is continuous.

When reducing the influence of the length term ($\nu \to 0$) larger regions are found. For most images these regions are very close to convex. Usually they do not correspond to meaningful objects.

IX. Conclusion

In this paper we introduced curvature into ratio optimization. We present a contour-based, fully unsupervised method which allows the global optimization of a discrete version of the energy.

From a practical viewpoint, we showed that the proposed method can handle shading effects where region-based approaches perform poorly. Our results also demonstrate that in contour-based approaches curvature is an important issue to mimic human perception.

From a theoretical viewpoint the computation of global optima allows to identify the strength and weaknesses of the cost functional. In this light we proved the existence of minimizers in the continuous domain and we proved that, as the resolution increases, sequences of minimizers of the discrete energy converge, possibly taking a subsequence, to a minimizer in the continuous domain.

Lastly we gave a parameterization-invariant reformulation of the snakes model and showed that the proposed method allows to identify meaningful global solutions of this reformulation.

Acknowledgments

This work was supported by the German Research Foundation, grant #CR-250/1-1. We thank Mathias Hauptmann and Frank R. Schmidt for helpful discussions. We also thank Greg Mori for sharing data with us.

References

[1] L. Ambrosio, N. Fusco, and D. Pallara. Partial regularity of free discontinuity sets,ii. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, 24:39–62, 1997.

[2] A. A. Amini. *Using Dynamic Programming for Solving Variational Problems in Vision: Applications Involving Deformable Models for Contours and Surfaces*. PhD thesis, The University of Michigan, Ann Arbor, 1990.

[3] A. A. Amini, T. E. Weymouth, and R. C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 12(9):855 – 867, September 1990.

[4] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. on Image Processing*, 10(8):1200–1211, 2001.

[5] R. E. Bellman. On a routing problem. *Quaterly of Applied Mathematics*, 16:87–90, 1958.

[6] A.I. Bobenko and P. Schröder. Discrete Willmore Flow. In *Eurographics Symposium on Geometry Processing*, 2005.

[7] Y. Boykov and M.-P. Jolly. Interactive organ segmentation using graph cuts. In *Intl. Conf. on Medical Image Computing and Comp. Ass. Intervention*, pages 276–286, 2000.

[8] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *IEEE Int. Conf. on Comp. Vision*, volume 1, pages 26–33, Nice, France, 2003.

[9] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.

[10] A.M. Bruckstein, A.N. Netravali, and T.J. Richardson. Epi-convergence of discrete elastica. In *Applicable Analysis, Bob Carroll Special Issue*, volume 79, pages 137–171, 2001.

[11] F. Cao. *Geometric Curve Evolution and Image Processing*, volume 1805 of *Lecture Notes in Mathematics*. Springer, 2003.

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

[13] V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert. Minimal surfaces based object segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19(4):394–398, 1997.

[14] A. Chambolle. Total variation minimization and a class of binary MRF models. In *Energy Min. Methods in Comp. Vision and Patt. Recog.*, pages 136–152, St. Augustine, Florida, 2005.

[15] T. F. Chan, S. H. Kang, and J. Shen. Euler's elastica and curvature based inpaintings. *J. Appl. Math.*, 2:564–592, 2002.

[16] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[17] G. Citti and A. Sarti. A cortical based model of perceptual completion in the roto-translation space. *Submitted*, 2008.

[18] I. J. Cox, S. B. Rao, and Y. Zhong. Ratio regions: a technique for image segmentation. In *IEEE Int. Conf. on Comp. Vision*, volume 2, pages 557–564, 1996.

[19] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *Int. J. of Comp. Vision*, 72(2):195–215, April 2007.

[20] G. Dal Maso. *An Introduction to Γ-Convergence*. Birkhäuser, Boston, 1993.

[21] W. Dinkelbach. On nonlinear fractional programming. *Manage. Sci.*, 13:492–498, 1967.

[22] M.P. do Carmo. *Differential Geometry of curves and surfaces*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.

[23] X. Dou, W. Xiaodong, A. Wahle, and M. Sonka. Globally optimal surface segmentation using regional properties of segmented objects. In *IEEE Int. Conf. on Comp. Vision and Patt. Recog.*, Anchorage, Alaska, June 2008.

[24] M. Droske and M. Rumpf. A level set formulation for Willmore flow. *Interfaces and Free Boundaries*, 6(3), 2004.

[25] Sharon E., Brandt A., and Basri R. Completion energies and scale. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 22(10):1117–1131, 2000.

[26] J. Elder and S.W. Zucker. Computing contour closure. In *Europ. Conf. on Comp. Vision*, volume 1 of *Lect. Not. Comp. Sci.*, pages 399–412, Cambridge, U.K., April 1996. Springer Verlag.

[27] S. Esedoglu and R. March. Segmentation with depth but without detecting junctions. *J. Math. Imaging and Vision*, 18:7–15, 2003.

[28] S. Esedoglu, S. Ruuth, and R. Tsai. Threshold dynamics for shape reconstruction and disocclusion. In *Proc. ICIP05, Int. Conf. on Image Proc., Genova, Italy*, volume II, pages 502–505, 2005.

[29] S. Esedoglu, S. Ruuth, and R. Tsai. Threshold dynamics for high order geometric motions. Technical report, UCLA CAM report, 2006.

[30] S. Esedoglu and J. Shen. Digital image inpainting by the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13:353–370, 2002.

[31] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Math. C.R.C. Press, 1992.

[32] C. Fantoni and W. Gerbino. Contour interpolation by vector field combination. *Journal of Vision*, 3:281–303, 2003.

[33] L. R. Ford. Network flow theory. Paper P-923, The Rand Corporation, Santa Monica, 1956.

[34] L Grady and C. Alvino. Reformulating and optimizing the Mumford-Shah functional on a graph - a faster, lower energy solution. In *Europ. Conf. on Comp. Vision, Marseille, France*, Oct. 2008.

[35] B.K.P. Horn. The curve of least energy. *ACM Trans. on Math. Software*, 9:441–460, 1983.

[36] L. Hsu, R. Kusner, and J. Sullivan. Minimizing the squared mean curvature integral for surfaces in space forms. *Experimental Math*, 1(3):191–207, 1992.

[37] A. Jalba, M. Wilkinson, and J. Roerdink. CPM: A deformable model for shape recovery and segmentation based on charged particles. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 26(10):1320–1335, 2004.

[38] I. H. Jermyn and H. Ishikawa. Globally optimal regions and boundaries as minimum ratio weight cycles. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 23(10):1075–1088, 2001.

[39] G. Kanizsa. Contours without gradients or cognitive contours. *Italian Jour. Psych.*, 1:93–112, 1971.

[40] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Comp. Vision*, 1(4):321–331, 1988.

[41] B.B. Kimia, I. Frankel, and A.-M. Popescu. Euler spiral for shape completion. *Int. J. of Comp. Vision*, 54:159–182, 2003.

[42] V. Kolmogorov and Y. Boykov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *IEEE Int. Conf. on Comp. Vision*, Beijing, China, October 2005.

[43] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in vision. In *IEEE Int. Conf. on Comp. Vision*, Rio de Janeiro, Brasil, October 2007.

[44] E. L. Lawler. Optimal cycles in doubly weighted linear graphs. In *Theory of Graphs: International Symposium*, pages 209–213, New York, USA, 1966. Gordon and Break.

[45] S. Masnou. Disocclusion : a variational approach using level lines. *IEEE Trans. on Image Processing*, 11:68–76, 2002.

[46] S. Masnou and J. M. Morel. Level-lines based disocclusion. In *Int. Conf. on Image Processing*, volume 3, pages 259–263, Chicago, USA, 1998.

[47] E.F. Moore. The shortest path through a maze. In *International Symposium on the Theory of Switching*, pages 285–292. Harvard University Press, 1959.

[48] D. Mumford. Elastica and computer vision. *Algebraic Geometry and Its Applications*, pages 491–506, 1994.

[49] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

[50] M. Nikolova, S. Esedoglu, and T. F. Chan. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[51] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, Segmentation and Depth*, volume 662 of *Lect. Not. Comp. Sci.* Springer, Berlin, 1993.

[52] P. Parent and S. W. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 11(8):823–839, 1989.

[53] T. Schoenemann and D. Cremers. Introducing curvature into globally optimimal image segmentation: Minimum ratio cycles on product graphs. In *IEEE Int. Conf. on Comp. Vision*, Rio de Janeiro, Brazil, October 2007.

[54] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 22(8):888–905, 2000.

[55] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher. Geometric Surface Processing via Normal Maps. In *ACM Transactions on Graphics*, 2003.

[56] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE's : a common framework for different applications. *IEEE Trans. on Patt. Anal. and Machine Int.*, 27(4), 2005.

[57] S. Ullman. Filling-in the gaps: the shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25:1–6, 1976.

[58] A. Vasilevskiy and K. Siddiqi. Flux-maximizing geometric flows. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 24(12):1565–1578, 2002.

[59] X. Xie and M. Mirmehdi. MAC: Magnetostatic active contour model. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 30(4):632 – 646, 2008.

[60] C. Xu and J. Prince. Generalized gradient vector flow external forces for active contours. *Signal Processing*, 71(2):131–139, 1998.

[61] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.