Image Segmentation with Elastic Shape Priors via Global Geodesics in Product Spaces

Thomas Schoenemann, Frank R. Schmidt and Daniel Cremers
Department for Computer Science
University of Bonn, Germany

Abstract

We propose an efficient polynomial time algorithm to match an elastically deforming shape to an image. It is based on finding a globally optimal geodesic in the product space spanned by the image and the prior contour. To this end a branch-and-bound scheme is combined with shortest path techniques.

We compare this algorithm with a recently proposed ratio minimization approach. While we show that generally the ratio is the better model, for many instances the two perform similarly. We identify a class of problems where the proposed method is likely to be faster.

1 Introduction and Related Work

For decades researchers have striven to develop machine vision algorithms which can compete with or even outperform the human visual system. Despite many efforts this remains a challenging problem.

The human visual system makes heavily use of prior world knowledge. As a consequence researchers have endeavored to integrate such prior knowledge into computer vision tasks. Among these tasks, in this paper we address two very prominent ones: image segmentation and the tracking of deformable objects.

In both areas a critical issue is to make the methods robust against getting stuck in poor local minima. For tracking usually one reverts to stochastic optimization techniques such as particle filtering [1, 9] or Kalman filtering [8]. Yet, these techniques neither offer a guarantee to find the global minimum nor a means to verify whether a computed solution is optimal.

For the more general problem of image segmentation with prior knowledge such techniques are not common - the search space is simply too large to be sampled in a dense way. Instead, meanwhile there are a number of globally optimal algorithms at hand. These will be reviewed in greater detail in the following. For completeness we mention some local methods [12, 3, 7, 16, 19, 2, 5] based on contour evolution.

Among the global methods we differentiate between those based on a single template and those based on a set of templates. To our knowledge, the only global method in the latter category is the recent region-based work of Cremers et al. [6].

Concerning methods based on a single template, as far as we know the work of Felzenzwalb [10] provides the only global method which can include region-based terms. Its

BMVC 2008 doi:10.5244/C.22.28

practical relevance is however limited due to the quadratically increasing memory consumption.

More practicable global methods are based on matching contours to images. After the pioneering work of Coughlan et al. [4] on open contours, recently the authors of this paper achieved contour closing [18]. This latter method is based on minimizing ratio functionals.

In this paper we propose an efficient alternative algorithm to match closed, deformable curves to images. The underlying principle is to find a global geodesic in the product space spanned by the image and the prior curve. Its optimization algorithm is based on a combination of branch-and-bound and shortest path techniques. Experimentally we show that this novel method gives solutions that are comparable to those given by ratio optimization, but is faster for a certain class of problems.

2 Matching Shapes to Images via Global Geodesics

In this section, we will cast the problem of matching a deformable contour to an image as a problem of finding a globally minimal geodesic in the space spanned by the image and the contour. Let

$$S : \mathbb{S}^1 \to I\!R^2 \tag{1}$$

be a given closed template curve and $I : (\Omega \subset I\!R^2) \to I\!R$ be a given image containing an object whose boundary is similar to the shape S up to elastic deformations. The goal is to find the boundary curve $C : \mathbb{S}^1 \to \Omega$ of the object and an orientation-preserving correspondence function $m : \mathbb{S}^1 \to \mathbb{S}^1$ which puts into correspondence pairs of points on C and on S. The joint computation of C and m allows to impose measures of shape similarity which take into account the correspondence of parts that was shown to be of importance for reproducing human notions of shape similarity [11, 14].

Clearly the joint space of all segmentations C and all correspondence functions m contains exponentially many solutions. It is therefore of utmost importance to appropriately parameterize this space and – if possible – identify polynomial-time algorithms to determine the best among all possible solutions. In the following, we will propose such an algorithm.

The assignment of curve points $C(s) \in \Omega$ and correspondences $m(s) \in \mathbb{S}^1$ to all points of the template S is equivalent to a mapping:

$$\Gamma : \mathbb{S}^1 \to \Omega \times \mathbb{S}^1, \tag{2}$$

which assigns to each point $s \in \mathbb{S}^1$ a point $\Gamma(s) = (C(s), m(s))$. Geometrically this mapping Γ can be seen as a cyclic path in the space $\Omega \times \mathbb{S}^1$ – see Figure 1.

Among all such mappings how should one define an optimal one? In the following we specify three requirements:

- **Edge detector:** We want the contour C in the image to pass through areas of high intensity gradient $|\nabla I|$. The computer vision literature offers a wealth of suitable data terms. Yet, for simplicity we only consider the function

  $$g(\mathbf{x}) = 1/(1 + |\nabla I(\mathbf{x})|),$$

  which assigns low values to high image gradients.

- **Shape similarity:** We want the shape C in the image to be similar to the template shape S. Again among a wealth of possible similarity measures we chose the simple one that encourages all pairs of corresponding edglets to have the same tangent angle $\alpha \in \mathbb{S}^1$. More specifically we want the squared cyclic difference (on the manifold $\mathbb{S}^1$) $|\alpha_C(s) - \alpha_S(m(s))|^2_{\mathbb{S}^1}$ to be small for all $s \in \mathbb{S}^1$. While this measure provides invariance to translation, it is clearly not invariant to global rotations of the template S. We will come back to this aspect later.

- **Regularity:** The correspondence function m assigns to each point on the template S its corresponding point on the curve C. In order to impose regularity of the assignment, we only consider orientation-preserving reparameterizations and disfavor local stretching or shrinking in the assignment process. Suppose we have a piece dC of the curve C which corresponds to a piece dS of the prior curve S. Then the length distortion is given by the ratio $|dC|/|dS|$. We use the penalty function

$$\Psi\left(\frac{|dC|}{|dS|}\right) = \begin{cases} \frac{|dC|}{|dS|} - 1 & \text{if } K \geq \frac{|dC|}{|dS|} \geq 1 \\ \left(\frac{|dC|}{|dS|}\right)^{-1} - 1 & \text{if } \frac{1}{K} \leq \frac{|dC|}{|dS|} < 1 \\ \infty & \text{otherwise} \end{cases}, \tag{3}$$

  where K is a predefined constant limiting the maximum length distortion.

With these notations the optimal assignment $\Gamma = (C, m)$ is defined as a global minimizer of the geodesic energy

$$E_{\text{geo}}(\Gamma) = \int_{\mathbb{S}^1} \left[g(C) + \lambda \Psi(m') + \nu |\alpha_C - \alpha_S \circ m|^2_{\mathbb{S}^1} \right] dC \tag{4}$$

This is a geodesic energy since minimization results in the global geodesic with winding number one on a manifold[1]

$$M \subset \mathbb{R}^2 \times \mathbb{S}^1 \times \mathbb{R}^2 \times \mathbb{R}$$

where the M only allows paths $\Gamma' : \mathbb{S}^1 \to M$ which correspond to a combination $C \times m \times C' \times m'$, i.e. the last two components are induced by the first two. Additionally one requires that C' be non-zero and m' be positive. Functional (4) can then be written as:

$$E_{\text{geo}}(\Gamma' \in M) = \int_{\Gamma'} \left[(C'\ m'\ C''m'') \mathbf{A} (C'\ m'\ C''m'')^\top \right]^{\frac{1}{2}} dt \tag{5}$$

where

$$\mathbf{A} = \left[g(C) + \lambda \Psi(m') + \nu |\alpha_C - \alpha_S \circ m|^2_{\mathbb{S}^1} \right]^2 \cdot \begin{pmatrix} \mathbf{I_{2\times 2}} & \mathbf{0_{2\times 4}} \\ \mathbf{0_{4\times 2}} & \mathbf{0_{4\times 4}} \end{pmatrix}$$

with $\mathbf{I}$ the identity matrix. Although the above matrix $\mathbf{A}$ is only positive *semi*-definite, the arising product is never 0 due to the requirement on C'. It therefore allows the interpretation as geodesic energy.

[1] Formally one would take the phase space $T(\mathbb{R}^2) \times T(\mathbb{S}^1)$ where $T(\cdot)$ denotes the tangent bundle.

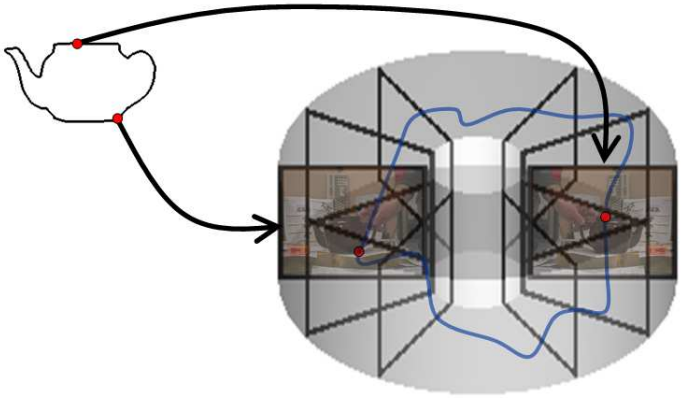


Figure 1: The structure of the graph: for any point on the prior contour there are K copies of the image in the graph. If a cycle in the graph passes through such a frame, this defines an assignment of a pixel in the image to the respective point on the prior contour.

3 Geodesic Energy versus Ratio Functionals

Geometrically our approach can be interpreted as follows: we consider the product space formed by the image plane Ω and the correspondence function m – see Figure 1. The intensity I of the input image and the tangent angles α_S of the given template induce a deformation of the space providing a norm as discussed above. The computation of an optimal matching hence boils down to the computation of geodesic (shortest) paths w.r.t. the given norm.

We compare this approach with our recent work [17] where we consider a ratio functional. This can be interpreted as an approach which aims at finding cycles Γ with minimal *average* cost. It can be written in the following manner:

$$E_{\text{ratio}}(C,m) = \frac{E_{\text{geo}}(C,m)}{\|C\|_2} = \int_{\mathbb{S}^1} \left[g(C) + \lambda \Psi(m') + \nu |\alpha_C - \alpha_S \circ m|^2_{\mathbb{S}^1} \right] ds \qquad (6)$$

The major difference is that here ds is used where dC appears in (4). This (partly) removes the bias towards smaller curves.

The global optimum of the ratio energy (6) is found by combining the Minimum Ratio Cycle algorithm of Lawler [15, 13] with a recursive splitting strategy [18]. The underlying algorithmic principle is iterated negative cycle detection. In the next section we show how to efficiently minimize the geodesic energy.

4 Efficiently Finding the Global Geodesic

In this paper we show how to find the global optimum of a discretized version of the geodesic energy (4), where the contour C consists of a discrete number of image pixels. Likewise, the prior contour S is discretized into an ordered set of points. The order is obtained by picking an arbitrary point and assigning it the number 0, then enumerating the other points in a clock-wise sense.

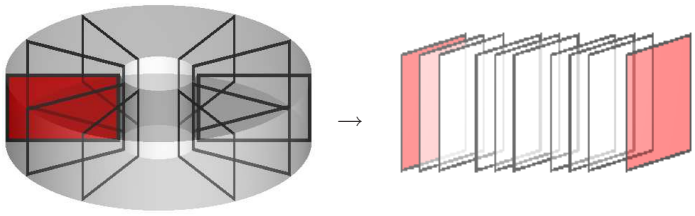


Figure 2: By cutting the graph at one place, an acyclic graph is obtained. Distance calculations in this graph are straightforward and highly parallelizable.

4.1 Optimization via Finding Cycles in Graphs

In the discrete setting, the simultaneous optimization over image contour C and matching m is mapped to the problem of finding the optimal cycle in the torus-like graph shown in Figure 1. In the following we give a brief overview of how this graph is constructed. For further details we refer to [18].

By design of the $\Psi()$-function, a point on the prior contour S may correspond to at most K pixels on the image contour C. These K choices are represented by K frames of nodes, where each frame contains a node for each pixel in the image I.

The arising directed graph consists of $K \cdot |S|$ frames. Directed edges connect nodes only across frames and are directed in a clock-wise sense. Moreover, when cutting the graph at frame 0, an acyclic graph is obtained. For minimizing the geodesic energy it suffices to have one edge weight per edge (as opposed to two for minimizing ratio functionals), representing a certain part of the integral.

By construction each pair of image contour and alignment corresponds to a cycle in the graph. The reverse is not true: some cycles do not represent a valid pair of alignment and image contour. These *invalid* cycles wrap around multiple times in the torus. The task is therefore to find the optimal *valid* cycle in the graph, i.e. to optimize over all cycles that wrap around exactly once.

4.2 Optimal Paths

Before describing how to find the optimal closed (cyclic) contour and matching function, we consider the slightly easier problem to optimize open curves. These can be optimized by means of dynamic programming [4].

Let $\mathscr{P}$ be the set of image pixels, and suppose we know some set $X \subseteq \mathscr{P}$ of pixels which are likely correspondences for point 0 on the prior contour. For each pixel $x \in X$ we can determine the optimal *open* curve ending in x and starting in some pixel in X. To this end, the torus-like graph is cut open at frame 0. Additionally, another copy of frame 0 is added and the cut edges are connected to the respective nodes in this new frame. This process is visualized in Figure 2.

Shortest paths in the arising acyclic graph are then computed. To this end, the initial distance labels in frame 0 are set to 0 for all pixels $x \in X$ and to ∞ otherwise. Dynamic programming now allows to compute the shortest paths. The determined cost in the last frame are termed *end distances* in the following.

Since this distance calculation for open curves is the basis for solving the problem with closed curves, we point out a few important properties of the resulting end distances and the corresponding shortest paths. These will be needed in the next section.

1. A shortest path represents a valid cycle in the torus-like graph if and only if its start and end node represent the same image location (by construction no shortest path will ever correspond to an invalid cycle). The cost of such a path gives an upper bound on the minimal geodesic for closed curves, i.e. on the cost of the optimal valid cycle in the torus-like graph.
2. For all pixels $x \in X$, the end distance at x represents a lower bound for all *cycles* where x is the first image pixel aligned to the prior point 0.
3. If no $x \in X$ has a distance label smaller or equal to some known upper bound on the optimal cycle, then this optimal cycle cannot align a pixel in X to the prior point 0 (more precisely, no such pixel will be the *first* pixel aligned to 0).

4.3 A Branch-and-bound Algorithm for Contour Closing

In this section we solve the problem of matching *closed* contours to images. A simple approach would be an exhaustive search over the initial correspondence, where for each correspondence the optimal curve is determined via the above described dynamic programming approach.

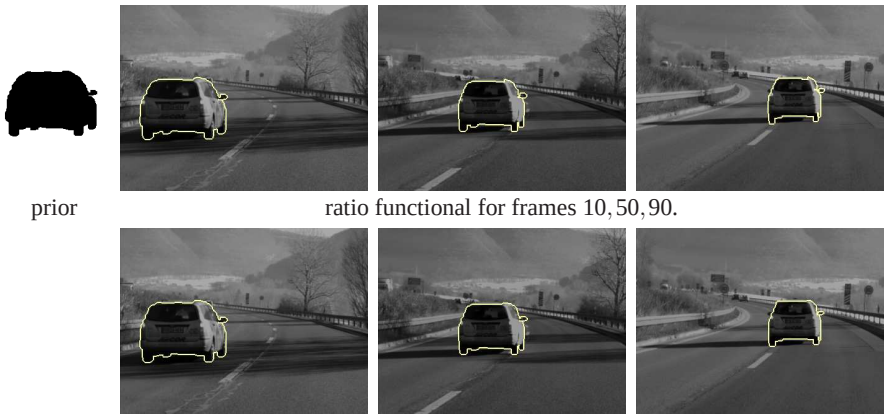
To overcome the quadratic run-time of this approach we resort to a branch-and-bound scheme, where the distance calculation for open curves serves to determine lower bounds on the cost of the optimal cycle. In general branch-and-bound schemes do not give polynomial time algorithms. We therefore emphasize that the proposed method *is* a polynomial time algorithm: its worst case complexity is quadratic in the number of image pixels. In practice we observe a linear run-time.

The basic principle is to successively split the set of image pixels $\mathscr{P}$ into disjoint subsets while maintaining a lower bound for each subset. Additionally a single upper bound is maintained which is updated every time a cycle is found. The algorithm terminates as soon as no lower bound is below the upper bound.

The algorithm starts with an infinite upper bound and a single partition of the pixel set, $\mathscr{P}$ itself. Its lower bound is set to 0. The algorithm then iteratively selects a component with lower bound below the upper bound. For this component the optimal open curves are determined (see previous section). Pixels with an end distance above the upper bound are removed from further consideration. For the remaining pixels the shortest paths are extracted. If cycles have formed, the respective pixels are also removed and the upper bound is updated.

If afterwards there are pixels remaining in the component, they are split into two parts. Numerous possibilities exist for the splitting rule, and in the end they are all heuristic. To avoid unnecessary further splitting, we make sure that the start and end node of the lowest cost path are separated (in practice the start node was usually removed in the above described process). Additionally we strive to get equally sized parts. The lower bounds for each part are set to the minimum of the respective distance labels.

This process is continued until no components are left to process. In the end the upper bound reflects the optimal energy. To get the optimal solution one additionally stores the corresponding cycle each time the upper bound is updated.



prior ratio functional for frames $10, 50, 90$.

geodesic energy for frames $10, 50, 90$.

Figure 3: Tracking a passing car. Both functionals are able to deal with contour deformation and changes in scale and lighting.

5 Extensions

So far we have focused on a translation-invariant approach for image segmentation with prior knowledge. In practice the desired amount of invariance depends on the application: for tracking one may not want full translational invariance as small motions are more likely. For other applications one might want to include rotational invariance.

It turns out that the results for the ratio functionals carry over to the geodesic energy: when tracking objects one can reduce the search space to a small window around the previous contour – see [18]. Rotational invariance is included by a sufficiently fine sampling of the rotation angle. When moving to the next rotation angle, one can use the previously determined energy as initial upper bound. This boosts performance significantly, since a tight upper bound allows to exclude many paths quite early in the optimization process. In the same manner the choice among multiple prior contours can be handled.

6 Global Geodesics vs. Ratio Minimizers in Tracking and Image Segmentation

In this section we evaluate the proposed method for shape-based tracking and image segmentation. We give a comparison to the ratio energy, including the quality of the computed global solutions as well as on the respective running times. The memory consumption is equal for both approaches.

Both methods were implemented on a GPU to exploit the parallelization properties. We use a Geforce 8800 GTX and the CUDA 1.1 programming framework.

6.1 Tracking

For tracking one usually deals with small deformations and can exploit spatial coherence. We therefore set $K = 2$ and $\lambda = \nu = 0.5$ and allow each point to move a distance of 15

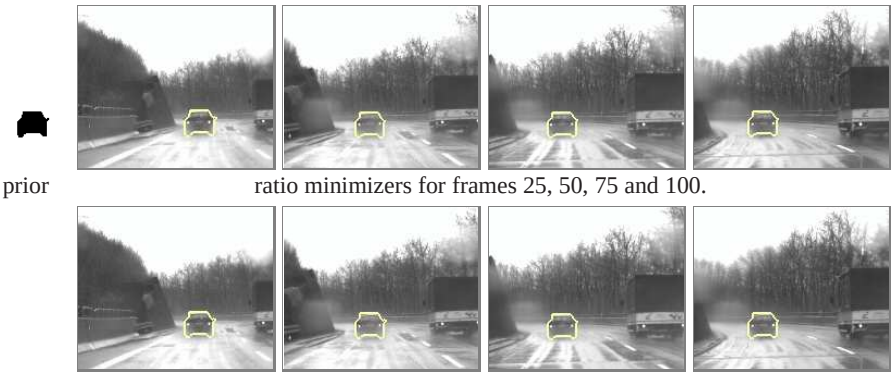


Figure 4: For tracking in bad weather, both functionals are competitive, with slightly better results for the line energy.

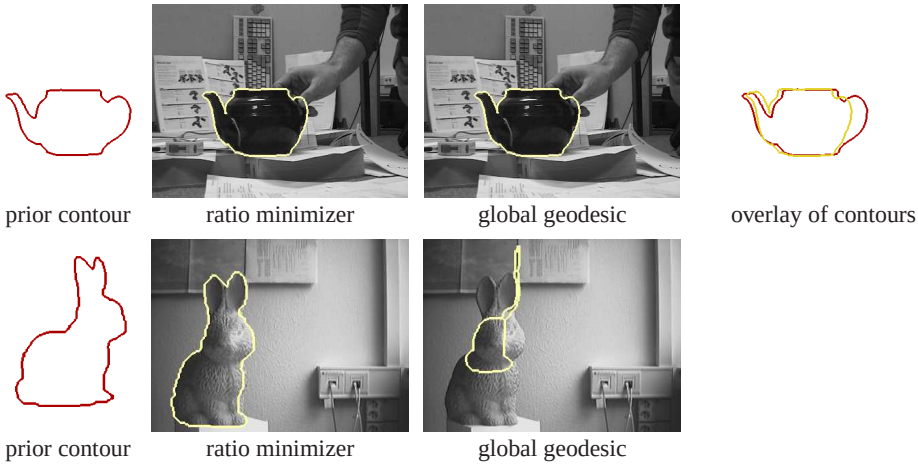
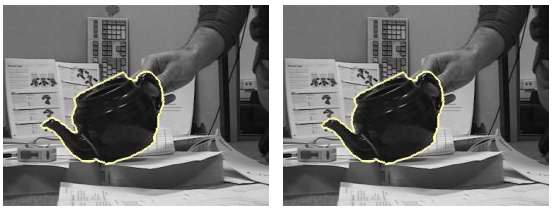


Figure 5: If there are few low-contrast places, both energies find the object reliably. However, for difficult tasks the geodesic energy reveals a (stronger) bias towards short curves.

pixels in each direction. Real-time performance is highly desirable and both approaches are made real-time capable by using the tight initialization from [18].

Figure 4 demonstrates that both methods are able to track a car in bad weather – over a hundred frames and more. Moreover, they are both essentially real-time capable: ratio minimization yields 25 fps, minimizing the geodesic energy 24.5 fps. A close inspection of the resulting segmentations showed that the geodesic energy gives slightly better results: it gives a better location of the mirror of the car and often has the better overall displacement.

A quite different tracking task is shown in Figure 3: here the contour undergoes significant deformation and scale changes. The results of both energies are so similar that it is impossible to weight one functional over the other. With respect to the run-time the geodesic energy is 30% faster.



ratio energy with rotational invariance.

geodesic energy with rotational invariance.

Figure 6: Both energies can handle rotational invariance.

6.2 Image Segmentation

For image segmentation one must deal with translation-invariance and stronger deformations (we set $K = 3$ and $\lambda = \nu = 0.25$). Here it becomes apparent that the ratio energy (6) is the better model since it is not so strongly biased towards shorter curves: Figure 5 reveals that the ratio excels in cases where the global geodesic (4) fails. Yet, such differences are only observed for a combination of a significant deformation and many places with low contrast. For easier tasks (such as the first row in Figure 5 the two approaches yield virtually the same results. Here the geodesic even outperforms the ratio: it is minimized in 4.2 seconds where the ratio needs 12 seconds with standard initialization and 9.7 s with the initialization from [18]. We found these run-times to be quite stable when testing other images in the sequence. Apparently minimizing the geodesic energy is faster when many parts of the contour correspond to strong edges.

The situation changes when including rotational invariance as in Figure 6: minimizing the ratio energy is now roughly 50% faster. This value was determined by averaging the run-times for several frames. None of the two approaches guarantees that the algorithm terminates after a single distance calculation if there is no solution with cost below the initial upper bound. Yet, it seems that the ratio minimization is less likely to make multiple calls.

7 Discussion and Conclusion

We introduced a geodesic formulation to match deformable shapes to images. The arising optimization task can be solved globally in polynomial – effectively linear – time using a combination of branch-and-bound and shortest path techniques.

The proposed method is inferior to the previously proposed ratio energies. It hence justifies the ratio normalization. Yet, in many cases the two approaches give similar results. If the optimal contour passes through many strong edges the geodesic is usually faster to minimize, particularly if there are strong deformations.

In cases where the solution is less obvious the ratio minimization is usually faster. This also holds when multiple prior contours are given, e.g. when including rotational invariance.

Acknowledgements This research was supported by the German Research Foundation (DFG), grants #CR-250/1-1 and #CR-250/2-1. We thank Bodo Rosenhahn and Daimler Research for sharing their data with us.

References

[1] A. Blake and M. Isard. *Active Contours*. Springer, London, 1998.

[2] 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. of Comp. Vision*, 28:145–162, 2006.

[3] T. F. Cootes, C. J. Taylor, D. M. Cooper, and J. Graham. Active shape models – their training and application. *Comp. Vis. Image Underst.*, 61(1):38–59, 1995.

[4] J. Coughlan, A. Yuille, C. English, and D. Snow. Efficient deformable template detection and localization without user initialization. *Comp. Vis. Image Underst.*, 78(3):303–319, 2000.

[5] D. Cremers and M. Rousson. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In J. S. Suri and A. Farag, editors, *Parametric and Geometric Deformable Models: An application in Biomaterials and Medical Imagery*. Springer, May 2007.

[6] D. Cremers, F. R. Schmidt, and F. Barthel. Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In *IEEE Int. Conf. on Comp. Vision and Patt. Recog.*, Anchorage, Alaska, June 2008.

[7] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Int. J. of Comp. Vision*, 50(3):295–313, 2002.

[8] F. Dellaert and C. Thorpe. Robust car tracking using kalman filtering and bayesian templates. In *Conference on Intelligent Transportation Systems*, 1997.

[9] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice (Statistics for Engineering and Information Science)*. Springer, New York, 2001.

[10] P. F. Felzenszwalb. *Representation and Detection of Shapes in Images*. PhD thesis, Massachusetts Institute of Technology, Sept. 2003.

[11] Y. Gdalyahu and D. Weinshall. Flexible syntactic matching of curves and its application to automatic hierarchical classication of silhouettes. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 21(12):1312–1328, 1999.

[12] U. Grenander, Y. Chow, and D. M. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes*. Springer, New York, 1991.

[13] 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.

[14] L. J. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 22(10):1185–1190, 2000.

[15] 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.

[16] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Int. Conf. on Comp. Vision and Patt. Recog.*, volume 1, pages 316–323, Hilton Head Island, SC, 2000.

[17] T. Schoenemann and D. Cremers. Globally optimal image segmentation with an elastic shape prior. In *IEEE Int. Conf. on Comp. Vision*, Rio de Janeiro, Brazil, October 2007.

[18] T. Schoenemann and D. Cremers. Globally optimal shape-based tracking in real-time. In *IEEE Int. Conf. on Comp. Vision and Patt. Recog.*, Anchorage, Alaska, June 2008.

[19] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. Model–based curve evolution technique for image segmentation. In *IEEE Int. Conf. on Comp. Vision and Patt. Recog.*, pages 463–468, Kauai, Hawaii, 2001.