11th IEEE International Conference on Computer Vision, Rio de Janeiro, 2007.

Globally Optimal Image Segmentation with an Elastic Shape Prior*

Thomas Schoenemann and Daniel Cremers
Department of Computer Science
University of Bonn, Germany

Abstract

So far global optimization techniques have been developed independently for the tasks of shape matching and image segmentation. In this paper we show that both tasks can in fact be solved simultaneously using global optimization. By computing cycles of minimal ratio in a large graph spanned by the product of the input image and a shape template, we are able to compute globally optimal segmentations of the image which are similar to a familiar shape and located in places of strong gradient. The presented approach is translation-invariant and robust to local and global scaling and rotation of the given shape. We show how it can be extended to incorporate invariance to similarity transformations.

The particular structure of the graph allows for run-time and memory efficient implementations. Highly parallel implementations on graphics cards allow to produce globally optimal solutions in a few seconds only.

1. Introduction

Most state-of-the-art approaches to image segmentation are based on low-level cues such as edges [11] or region statistics [15]. Only in very few cases can the corresponding optimization problem be solved globally optimal. Most notably this is true for the approach of Jermyn and Ishikawa [10] to edge-based segmentation and the region-based approach of Greig et al. [7] for binary images. The latter can be extended to a two-phase piecewise constant Mumford-Shah functional [15, 2] with known mean intensities.

In recent years researchers have suggested to enhance purely low-level segmentation schemes by imposing prior knowledge, favoring segmentations that are in some sense similar to a given shape. Representative works are the pioneering contribution of Grenander et al. [8], the variational formulation of Cremers et al. [5] as well as level set formulations by Leventon et al. [13] for edge-based models and Tsai et al, Rousson and Paragios and Cremers et al. [17, 16, 4] for region-based approaches. These works demonstrate that shape priors allow to drastically improve the segmentation of familiar objects in the presence of prominent noise, partial occlusions and background clutter.

All these approaches use local optimization where usually each set of parameters is solved for alternatingly. The often huge number of parameters (accounting for mean intensities, rotation, translation, scale and deformation) to estimate suggests that the approaches are likely to get stuck in local optima. Moreover, stable gradient descent requires a delicate tuning of the respective time step sizes.

We emphasize two works not depending on initialization: the first is the work of Coughlan et al. [3] to find open contours in images. It makes intensive use of a training phase. The second is the approach of Felzenszwalb [6]. While this approach is able to find optimal closed contours in polynomial time, due to the cubic run-time in practice it has to rely on heuristics.

Sophisticated shape priors and shape similarity measures typically involve the computational challenge of determining appropriate matches between the points of the segmenting contour and points on a shape template. Such shape alignments are usually computed via dynamic time warping [14, 1]. Yet, to efficiently determine a segmentation which *globally* optimizes both the low-level edge consistency and the similarity to one or more shape templates has so far remained an open challenge.

In this paper we show that indeed segmentations and shape alignments can be determined simultaneously, resulting (with the exception of a few degenerate cases) in a globally optimal image segmentation with a translation-invariant elastic shape prior. To this end, we propose to compute cycles of minimal ratio in a large graph representing the product space spanned by the input image and all points of the shape template. The presented approach is robust against local and global scaling and rotation. We show extensions to incorporate rotational invariance.

The structure of the graph allows for a both memory and run-time efficient implementation of the globally optimal graph-theoretic algorithm. GPU-implementations allow to

*This work was supported by the German Research Foundation, grant #CR-250/1-1.

find globally optimal shape-consistent segmentations in a matter of seconds.

2. Global Optimization in Computer Vision

Our approach combines two fields where global optimization techniques have been thoroughly researched: the fields of image segmentation and shape matching. We begin with a short review of the techniques relevant for our work.

2.1. Elastic Shape Matching

In elastic shape matching one is looking for the optimal way to deform a given shape $C : [0, l(C)] \to \mathbb{R}^2$ into a shape $S : [0, l(S)] \to \mathbb{R}^2$, where both shapes are parameterized by arc length and $l(\cdot)$ denotes the length of a curve.

Formally one is looking for a *matching* or *alignment* of the shapes, a strictly increasing diffeomorphism $m : [0, l(C)] \to [0, l(S)]$ expressing that the points $C(s)$ and $S(m(s))$ correspond to each other. The optimal matching is determined by minimizing an appropriate measure of the total deformation.

Often one wants to exclude *global* deformations from this measure, i.e. a shift of the center of mass, a rotation of the complete shape and a scaling of it should not be penalized. What remains are *local* deformations such as the motion of the thumb of a hand.

Basri et al. [1] proposed a deformation measure based on two aspects: (1) points of similar curvature should be matched. If curvature is identical for all pairs of corresponding points the shapes are identical up to global rotation and translation. (2) a curve piece of C should be matched to a piece of S of equal length. This implies an ideal derivative[1] of $m'(s) = 1$. Deviations from both ideals are penalized, leading to the total deformation costs:

$$\int_0^{l(C)} |\kappa_C(s) - m'(s)\kappa_S(s)| \, ds \; + \; \lambda \int_0^{l(C)} \Psi(m'(s)) \, ds \quad (1)$$

Here $\kappa_C(s)$ denotes the curvature of C at $C(s)$ and likewise for $\kappa_S(s)$. As penalty function $\Psi(\cdot)$ in this work we choose

$$\Psi(m') = \begin{cases} m' - 1 & \text{if } m' \geq 1 \\ \frac{1}{m'} - 1 & \text{otherwise} \end{cases} \quad (2)$$

This function offers the advantage of symmetry, i.e. comparing C to S yields the same costs as the comparison of S to C. The optimal alignment can be found via dynamic time warping, see [14, 1].

2.2. Minimum Ratio Cycles

For the problem of image segmentation, Jermyn and Ishikawa [10] proposed to consider the ratio of two line integrals. Such problems are discretized and reduced to the problem of finding a cycle C in a graph where C minimizes

$$\min_C \frac{\sum_{e \in C} n(e)}{\sum_{e \in C} d(e)} \quad (3)$$

Here each edge is assigned a numerator weight $n(e)$ and a denominator weight $d(e)$ representing a small piece of the respective integral. Under fairly mild requirements on $d(e)$ (see [12]) this can be solved globally optimal using the Minimum Ratio Cycle algorithm. The key observation [12] is that for a ratio $\tau > \tau_{opt}$, with τ_{opt} the optimal ratio, a graph with same topology and edge weights $w(e) = n(e) - \tau d(e)$ must possess a negative cycle.

Negative cycles can be found via the Moore-Bellman-Ford algorithm for distance calculations. For graphs without negative cycles the algorithm computes the distance from a given root node r to all other nodes in the graph. It relies on a distance label $d(v)$ and a parent entry $p(v)$ for each node v in the graph. Initially $d(r)$ is set to 0 and $d(v) = \infty$ for all other nodes v. The root node is added to a queue. Then, as long as there are nodes in the queue, the front one is removed and expanded. Expanding a node v means to check for all outgoing edges $e = (v, v')$ whether $d(v) + w(e)$ is lower than $d(v')$. If so, a shorter path to v' was found, $d(v')$ is set to the new distance and $p(v')$ is set to v. If v' is not already in the queue it is added at the end.

If the graph has negative cycles, the algorithm will not terminate and from some point on the parent graph will permanently contain cycles. Regularly checking the parent graph for cycles allows to detect and extract negative cycles.

To determine the optimal ratio one starts with an upper bound τ on the optimal ratio and adjusts it until all negative cycles vanish. Every time a cycle is found the ratio τ is set to its ratio. The last extracted cycle is of optimal ratio.

For integral numerator and denominator weights in the worst case the run-time is proportional to the number of nodes times the number of edges and a factor given by the margin spanned by the edge weights (see [10]). In practice we observe a run-time linear in the number of nodes.

3. Elastic Shape Priors in Ratio Functionals

In this paper we simultaneously solve the problems of image segmentation and shape matching: given an image $I : \Omega \to \mathbb{R}$ in the image plane Ω, we find the optimal region boundary C (parameterized by arc length) that both accumulates strong gradients along the curve and is similar to a given shape S. We use an edge indicator function $g : \Omega \to \mathbb{R}$ which can be arbitrary. In practice we choose

$$g(x) = \frac{1}{1 + |\nabla I(x)|}$$

[1] If scale invariance is desired the ideal value is $l(S)/l(C)$.

This function is averaged along the curve C:

$$\frac{\int_0^{l(C)} g(C(s))\, ds}{l(C)} \tag{4}$$

Without the denominator long curves would be extremely disfavored. Thanks to the ratio the length of a curve does not enter as a bias. By itself minimizing (4) is not a sensible problem: a global minimum is given by an infinitesimally short curve located where $g(\cdot)$ is lowest. This is not a problem for our approach as we also impose similarity of C to the shape template S.

To measure this similarity the optimal alignment of C to S must be determined. This is done simultaneously with the determination of C. As similarity measure one might be tempted to use functional (1). However, the comparison of curvature is difficult to integrate when C itself is unknown. We therefore defer invariance to similarity transformations to section 5. For translation invariance we compare the tangent angles $\alpha_C(s)$ and $\alpha_S(m(s))$ of the curves. To ease notation in the following we simply write $|\alpha_C(s) - \alpha_S(m(s))|^2$. In fact, to accurately capture the cyclic nature of angles the minimum of this and $(2\pi - |\alpha_C(s) - \alpha_S(m(s))|)^2$ is taken.

Our cost function consists of the average edge indicator value along C plus the deformation and scaling costs for the optimal alignment, normalized by the length of C:

$$\min_{C,m} \quad \frac{\int_0^{l(C)} g(C(s))\, ds}{l(C)} + \lambda \frac{\int_0^{l(C)} \Psi(m'(s))\, ds}{l(C)} + \nu \frac{\int_0^{l(C)} |\alpha_C(s) - \alpha_S(m(s))|^2\, ds}{l(C)} \tag{5}$$

This is minimized simultaneously over the region boundary C and the alignment m. While this is surely not the only conceivable cost function it offers the immense advantage that it can be optimized globally via the Minimum Ratio Cycle algorithm. This algorithm is very efficient as it does not perform an explicit search over all possible start points. Moreover, as demonstrated in section 6 this functional leads to very accurate results.

Our algorithm does not exclude self-intersecting curves. However, while for image segmentation these are usually unwanted, this is different for the task of *locating* deformable objects in an image. This is demonstrated in Figure 1: in the prior shape (the black hand) all fingers are spread apart. To locate them in the image some line segments must occur repeatedly, which was accurately determined by the approach. Comparing to the outer hull of the hand – as done for example in Level Sets methods – would produce a weaker similarity.

prior shape

location of deformed version

Figure 1. For the task of locating deformable objects self-intersections and repeated line segments should not be excluded.

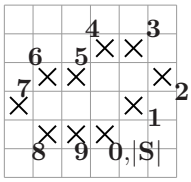


Figure 2. A shape is represented as a sequence of pixels on a discrete grid.

4. Product Graphs for Combined Matching and Segmentation

We solve problem (5) in discretized form. The shape S is given as a list of $|S|+1$ pixels where the first equals the last and each pixel is contained in an eight-neighborhood of its predecessor. This is illustrated in Figure 2. The region boundary C to be optimized is represented in the same way. The tangent angle for the shape pixel with number s is denoted as $\alpha_S(s)$.

In the discrete setting the matching $m(\cdot)$ aligns image pixels to shape pixels. A shape pixel may be aligned to several image pixels (implying $m' < 1$) or remain unaligned ($m' > 1$). We limit the maximal and minimal derivative of the matching m according to

$$\frac{1}{K} \leq m' \leq K \tag{6}$$

where K is some pre-defined integral constant. This implies that at most K image pixels are aligned to the same shape pixel and at most $K-1$ subsequent shape pixels are skipped. Both run-time and memory depend linearly on K and are therefore polynomial even for $K = |S|$, which would allow to reduce the entire shape to a single pixel.

The problem can now be expressed as the search for a cycle in a graph that minimizes a ratio of form (3).

4.1. Graph Structure

For the simultaneous search of region boundary and alignment we use the Minimum Ratio Cycle (MRC) algorithm on a product graph. Product graphs in combination with Minimum Ratio Cycles first appeared in [9] where the product of the pixel sets of several images was taken.

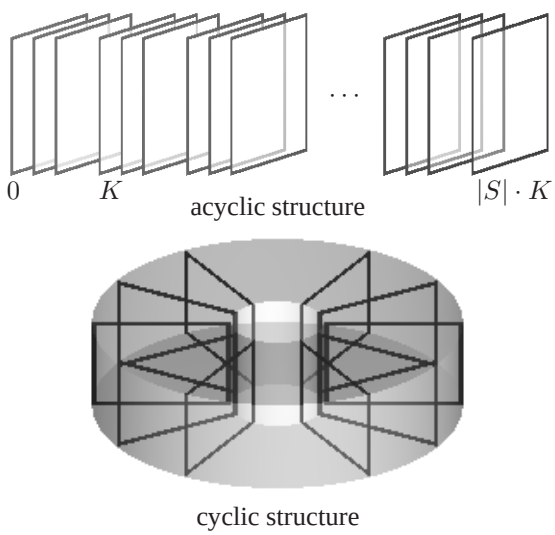


Figure 3. The topology of our graph: the graph is based on an acyclic structure (top). For each of the first $|S|$ shape nodes a group of K frames is created. These frames represent the image with one node for each pixel. As the frame $|S| \cdot K$ is identical to the frame 0 the graph actually has the cyclic structure shown in the bottom. For the construction of edges see text.

In our case the node set is the product of the pixel set $\mathcal{P}$ of the image I and the set of numbers between 0 and $K \cdot |S|$, representing K instances of each point of the shape template:

$$\mathcal{V} = \mathcal{P} \times \{0, \ldots, K \cdot |S|\} \tag{7}$$

The node $(\vec{p}, s \cdot K + l)$ with $l < K$ expresses that the image pixel $\vec{p}$ is aligned to the shape pixel with number s and that previously already l image pixels were aligned to the same shape pixel. From a sequence of such nodes one can therefore reconstruct both the alignment m and the region boundary C.

As the last shape pixel equals the first one the nodes $(\vec{p}, 0)$ and $(\vec{p}, K \cdot |S|)$ represent the same thing. They can therefore be thought of as identical or as linked by edges with weight 0. This introduces cycles into the otherwise acyclic directed graph we will now describe. The general topology is illustrated in Figure 3.

There are two kinds of edges in the graph. In both cases the contour C is developed by traversing from an image pixel $\vec{p}$ to a pixel $\vec{q} \in \mathcal{N}(\vec{p})$ where $\mathcal{N}(\vec{p})$ denotes the 8-neighborhood of $\vec{p}$. In both groups the denominator weight $d(e) = |\vec{q} - \vec{p}|$ of an edge e represents the piece of the denominator integral along C between $\vec{p}$ and $\vec{q}$.

In the first group the shape pixel stays constant and another image pixel is aligned:

$$(\vec{p}, s \cdot K + l) \rightarrow (\vec{q}, s \cdot K + l + 1) \quad \text{for } l < K - 1 \text{ and } \vec{q} \in \mathcal{N}(\vec{p})$$

The numerator weight for such an edge e covers the integral from $\vec{p}$ to $\vec{q}$ for the data and deformation term. It also covers the scaling costs of λ caused by the alignment of another image pixel to the same shape pixel. This corresponds to the non-linear part of Ψ in (2). With $\varphi(\cdot)$ denoting the angle of a vector, the costs are

$$\begin{aligned} n(e) \quad = \quad & \tfrac{1}{2} |\vec{q} - \vec{p}| \left(g(\vec{p}) + g(\vec{q}) \right) \\ & + \nu \, |\vec{q} - \vec{p}| \, |\alpha_S(s) - \varphi(\vec{q} - \vec{p})|^2 \\ & + \lambda \end{aligned}$$

In the second group of edges one traverses to a new shape pixel, possibly skipping some shape pixels in the process. As so far no image pixels were aligned to the new shape pixel, the second component of the target node is a multiple of K:

$$(\vec{p}, s \cdot K + l) \rightarrow (\vec{q}, (s + i) \cdot K) \quad \text{for } l < K, \;\; 1 \leq i \leq K \text{ and } \vec{q} \in \mathcal{N}(\vec{p})$$

The data term for the edge is as in the first case. For the deformation costs we consider the best shape direction in the skipped interval. Scaling costs (corresponding to the linear part of Ψ) are only added if shape pixels were skipped:

$$\begin{aligned} n(e) \quad = \quad & \tfrac{1}{2} |\vec{q} - \vec{p}| \left(g(\vec{p}) + g(\vec{q}) \right) \\ & + \nu \, |\vec{q} - \vec{p}| \min_{s \leq s' < s+i} |\alpha_S(s') - \varphi(\vec{q} - \vec{p})|^2 \\ & + (i - 1) \, \lambda \end{aligned}$$

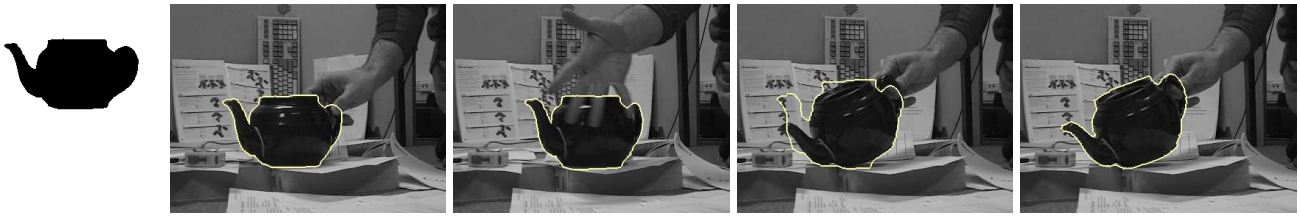
4.2. Efficient Optimization

To find the optimal cycle in the graph we use the Minimum Ratio Cycle algorithm[2] (see section 2.2). As shape templates are often of length 500 or more implementations with common graph classes are not feasible: already for very small images an immense amount of memory is needed. Fortunately the specific structure of the graph allows to implement the MRC algorithm without actually building a graph. One simply calculates a distance tensor and a tensor of parent pointers corresponding to the nodes. Edges are computed on the fly.

For the parent pointers it suffices to store the incoming direction and the change in the second component. We use two bytes per node, for $K \leq 5$ one byte suffices.

Distance calculations are performed by proceeding along t, where t denotes the second component of the nodes. Notice that except for $t = 0$ the distance matrix for t can be discarded as soon as $t + K^2 + 1$ is reached. The distance matrix for t can be computed in parallel as only distances for $t' < t$ are relevant. In a GPU implementation we observed speed-ups of a factor 250 and more.

[2] For the root node an extra node is added and connected to all nodes with second component 0. All edge weights are set to 0.



prior shape | after 3D-rotation | 3D-rotation + occlusion | no rotational invariance | with rotational invariance

Figure 4. The proposed method determines accurate segmentations in the presence of high background clutter and 3D-rigid body motion.

When reaching $t = K \cdot |S|$ the cycle detection step is performed. By the way the graph is constructed a cycle must contain a node with second component $t = K \cdot |S|$[3]. It is possible that it contains several such nodes, implying that the shape was aligned multiple times. While in intermediate solutions this occurs frequently, we only observed it once in the final solution. These are the degenerate cases where no global optimum is determined[4]. They can be interpreted as the scene not containing a similar enough shape.

If no cycle is found one has to check if the distance of any node with $t = K \cdot |S|$ is lower than the distance of the respective node for $t = 0$. If so the distance calculation is continued, otherwise the optimal ratio was found. In practice we observed at most 3 distance calculation rounds for the same ratio.

5. Comparison to Multiple Templates

Often it is desirable to have rotationally invariant shape priors. The described approach is only invariant to translation. To incorporate invariance against global rotations the rotation angle is sampled in regular intervals and a new shape created for each angle.

In a more general setting one is given several shape templates (e.g. different silhouettes of a 3D-object) and wants to determine the optimal one. This is achieved by sequentially comparing to each shape, resulting in a run-time linear in the number of shapes. The good news is that the subsequent comparisons are generally much faster than the first one: one can initialize with the ratio determined for the first template.

6. Experiments

In several experiments we demonstrate that our approach is able to produce accurate segmentations in the presence of high background clutter and low contrast. Moreover the segmentations are superior to the ones generated by the Mumford-Shah functional.

[3]This implies that the shape pixel 0 cannot be skipped. It is straightforward to remove this

[4]In such a case the global minimum can still be found in polynomial time by reverting to an explicit search over all start points.

In all cases *elastic* deformations are needed to get the desired contours. We fix $K = 3$ and $\lambda = \nu = 0.25$. For all images (resolutions around 384×288 pixels) 750 MegaBytes suffice to produce globally optimal segmentations.

In Figure 4 we segment a pot in front of cluttered background. The prior template was obtained from a frame early in the sequence. Segmentations after 3D-rotation on the table, occlusion and lifting are shown. The segmentation of the first image took 74.5 *minutes* on the CPU, on the GPU only 15 *seconds* – a speed-up factor of 290. The figure also shows that for large rotations rotationally invariance (see Sec. 5) is needed.

Figure 5 shows that to get meaningful segmentations indeed both scaling and deformation costs are needed. Again the prior shape was obtained several frames ago and the object rotated afterwards. The Figure also demonstrates that the Mumford-Shah functional – implemented by iterated graph min-cuts – is not able to extract the desired object.

So far all segmentations were obtained for a fixed prior shape. When given a video sequence containing articulated motion as in Figure 6 it is much more sensible to use a tracking approach: for each frame the shape determined for the last frame is sought. This way larger deformations are decomposed to a sequence of smaller ones.

Conclusion

We presented a graph-theoretic approach allowing for the first time to compute globally optimal image segmentations which are consistent both with the low-level edge information and with a higher-level elastic shape prior induced by a single shape template. The key idea is to find cycles of minimal ratio in a large product graph spanned by the input image and the shape template. As each node represents an image pixel as well as a shape pixel any cycle simultaneously encodes a segmentation as well as an alignment (or elastic registration) with the shape template.

The specific structure of the graph allows for runtime and memory efficient implementations without an actual graph. Thanks to the high parallelizability GPU-implementations are able to reduce the computation time by more than a factor of 250. This way, globally optimal shape-consistent segmentations are obtained in a matter of seconds.

Figure 6. Tracking of a deformable shape in the presence of low contrast.

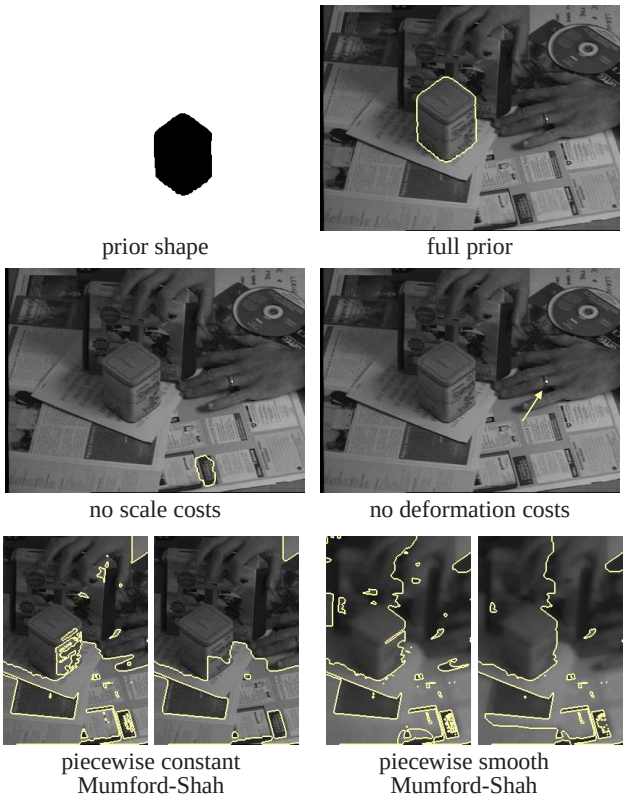


Figure 5. Effectiveness of the shape prior: the full prior is needed to find the Teabox. Moreover the Mumford-Shah functionals (shown for two phases with different length weights) are not helpful for extracting the Teabox.

References

[1] R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the similarity of deformable shapes. *Vision Research*, 38:2365–2385, 1998. 1, 2

[2] 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. 1

[3] 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. 1

[4] D. Cremers, S. J. Osher, and S. Soatto. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. of Comp. Vision*, 69(3):335–351, September 2006. 1

[5] 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. 1

[6] P. F. Felzenszwalb. *Representation and Detection of Shapes in Images*. PhD thesis, Massachusetts Institute of Technology, Sept. 2003. 1

[7] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum *a posteriori* estimation for binary images. *J. Roy. Statist. Soc., Ser. B.*, 51(2):271–279, 1989. 1

[8] U. Grenander, Y. Chow, and D. M. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes*. Springer, New York, 1991. 1

[9] H. Ishikawa and I. H. Jermyn. Region extraction from multiple images. In *IEEE Int. Conf. on Comp. Vision*, Vancouver, Canada, 2001. 3

[10] 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. 1, 2

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

[12] E. L. Lawler. Optimal cycles in doubly weighted linear graphs. In *Theory of Graphs: International Symposium*, pp. 209–213, New York, USA, 1966. Gordon and Break. 2

[13] 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. 1

[14] R. McConnell, R. Kwok, J. Curlander, W. Kober, and S. S. Pang. $\psi - s$ correlation and dynamic time warping: two methods for tracking ice floes in sar images. *IEEE Trans. on Geosc. and Rem. Sens.*, 29:1004–1012, 1991. 1, 2

[15] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989. 1

[16] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Europ. Conf. on Comp. Vision*, pages 78–92. Springer, 2002. 1

[17] 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. 1