Shape Priors in Variational Image Segmentation: Convexity, Lipschitz Continuity and Globally Optimal Solutions

Daniel Cremers, Frank R. Schmidt and Frank Barthel
Department of Computer Science
University of Bonn, Germany

Abstract

In this work, we introduce a novel implicit representation of shape which is based on assigning to each pixel a probability that this pixel is inside the shape. This probabilistic representation of shape resolves two important drawbacks of alternative implicit shape representations such as the level set method: Firstly, the space of shapes is convex *in the sense that arbitrary convex combinations of a set of shapes again correspond to a valid shape. Secondly, we prove that the introduction of shape priors into variational image segmentation leads to functionals which are convex with respect to shape deformations.*

For a large class of commonly considered (spatially continuous) functionals, we prove that – under mild regularity assumptions – segmentation and tracking with statistical shape priors can be performed in a globally optimal manner. In experiments on tracking a walking person through a cluttered scene we demonstrate the advantage of global versus local optimality.

1. Introduction

Related Work. The computation of image segmentations by minimization of appropriate functionals on continuous space goes back to the late 80's [10, 13]. To cope with missing or misleading low-level information (due to background clutter, partial occlusions or noise), numerous researchers have advocated the introduction of statistical shape priors on a variational level (cf. [6, 15, 19]). Most recent work on continuous image segmentation has been focused on level set representation in which a shape is defined as the boundary given by the zero level set of an embedding function $\phi : \mathbb{R}^d \to \mathbb{R}$:

$$C = \{x \in \mathbb{R}^d \mid \phi(x) = 0\}. \quad (1)$$

Shape priors can be defined on the space of embedding functions [19, 16, 5]. To guarantee a unique correspondence between silhouettes and embedding functions, one typically constrains all analysis (statistical shape learning and segmentation) to the space of signed distance functions – see Figure 1, right side. The main advantages of the level set formulation (over parametric boundary representations) are that one is independent of a specific choice of parameterization and one can handle topological changes such as splitting and merging of boundaries during the optimization process.

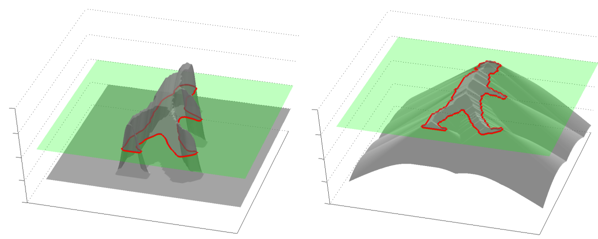

Figure 1. A relaxed notion of shape. In this paper, we introduce a novel definition of *shape* as a function $q : \mathbb{R}^d \to [0, 1]$ specifying the probability that a pixel $x \in \mathbb{R}^d$ is part of the shape (left). In contrast to the commonly used signed distance representation (right), the resulting image segmentation with statistical shape priors corresponds to the minimization of convex functionals over convex domains.

Although the above approaches allow to reliable segment and track familiar shapes in challenging image sequences, they all suffer from two well-known limitations:

- The space of signed distance functions is not a linear space. As a consequence, linear combinations of shapes (i.e. signed distance functions) no longer correspond to valid shapes. The construction of statistical distributions (means and covariance matrices) gives rise to challenging optimization problems [15].
- The resulting cost functionals are generally not convex. As a consequence, respective optimization by gradient descent merely leads to locally optimal solutions with very little insight as to how far computed solutions are from the globally optimal ones [5].

More recently, continuous image segmentation has been formulated on the basis of convex functional minimization [4, 3]. In a spatially discrete setting related approaches can be minimized using graph cuts [8, 1]. While shape priors have recently been introduced in globally optimal image segmentation [18], the present formulation is fundamentally different from [18] because it applies to continuous space, to arbitrary spatial dimension (including shape priors for volumetric segmentation), and to a larger class of cost functionals and statistical shape priors.

Contribution. In this work, we introduce a relaxed definition of the term *shape* which is based on assigning to each point in the spatial domain a probability that it is inside a given shape. Subsequently we propose a variational formulation to compute a segmentation of the image plane which is statistically consistent with a set of training shapes. The proposed formulation resolves the above limitations of existing approaches: Firstly, we prove that the space of shapes forms a convex set. Secondly, we prove that the optimization with respect to shape deformations amounts to the minimization of a convex functional. The class of optimizable functionals includes edge-based and region-based data terms as well as statistical and dynamical shape priors. Thirdly, reverting to the theory of Lipschitz functionals, we prove that both shape deformations and transformations can be optimized globally in continuous space. Experimental results on tracking walking people trough clutter and occlusion demonstrate the advantages of global over local optimality.

2. A Probabilistic Definition of Shape

Definition 1 (Shape)**.** In the following, we define as a *shape* in $\mathbb{R}^d$ a function

$$q : \mathbb{R}^d \to [0, 1],$$

which assigns to any pixel $x \in \mathbb{R}^d$ a probability $q(x)$ that x is part of the object – see Figure 1, left side. The space of all shapes will be denoted by $\mathcal{Q}$. While all statements in this paper directly apply to shapes and segmentation in arbitrary dimension d, we will for simplicity only consider the case of planar shapes ($d = 2$).

In contrast to explicit (parametric) representations of shape such as polygons or splines, the above implicit representation does not depend on a specific choice of parameterization. Shape alignment therefore does not require the estimation of point correspondences and propagation of shape does not require control point regridding. In contrast to alternative implicit representations of shape such as the signed distance function or alternative representations [7], the value of q has a clear probabilistic interpretation.

The above notion of *shape* fundamentally differs from classical definitions of shape in that we replace the hard decision of “a point is part of the shape” or “a point lies on the boundary of a shape” by a relaxed probability associated with each point. The key contribution of this paper is to show that this *relaxation* in the definition of shape gives rise to a number of advantages in the context of shape modeling and shape inference, most prominently it enables us to compute image segmentations with shape priors in a globally optimal manner.

3. Convex Shape Spaces

In this section, we show that the above probabilistic definition of shape has an important consequence for statistical shape modeling and shape inference as the space of shapes and any linear subspace form convex spaces.

Proposition 1. *The space $\mathcal{Q}$ of all shapes (as defined above) forms a convex set.*

Proof. Let $q_1 : \mathbb{R}^2 \to [0, 1]$ and $q_2 : \mathbb{R}^2 \to [0, 1]$ be two elements of $\mathcal{Q}$. Then for any convex combination $q_\gamma = \gamma q_1 + (1 - \gamma) q_2$ of these shapes, we have $q_\gamma \in \mathcal{Q}$:

$$0 \leq q_\gamma(x) \leq 1 \qquad \forall x \in \mathbb{R}^2,\ \gamma \in [0, 1].$$

□

This property of the shape space $\mathcal{Q}$ implies that any convex combination of a set

$$\chi = \{q_1, \dots, q_N\} \tag{2}$$

of training shapes will correspond to a valid shape. In particular, the mean

$$\mu = \frac{1}{N} \sum_{i=1}^{N} q_i(x)$$

is a function which assigns to each point $x \in \mathbb{R}^2$ the average of all probabilities. Similarly, statistical notions such as covariance matrices and eigenmodes can be easily defined.

Let us consider the subspace spanned by the first $n \leq N$ eigenmodes $\{\psi_1, \dots, \psi_n\}$ of the set χ:

$$\chi_n := \left\{ q_\alpha = \mu + \sum_{i=1}^{n} \alpha_i \psi_i \;\middle|\; q_\alpha(x) \in [0, 1] \right\}. \tag{3}$$

Lemma 1. *The set χ_n in* (3) *is convex.*

Proof. For any $n \leq N$, the set χ_n is the intersection of a linear space (spanned by the first n eigenmodes of χ) with the convex space $\mathcal{Q}$, therefore it is also a convex space. □

The set of training shapes χ can thus be approximated by nested low-dimensional spaces χ_n. Elements in these spaces are compactly represented by vectors $\alpha \in \mathbb{R}^n$ of eigencoefficients, modeling the shape $q_\alpha = q_0 + \alpha^\top \Psi$.

Lemma 2. *The set $\mathcal{A}_n := \{\alpha \in \mathbb{R}^n | q_\alpha \in \mathcal{Q}\}$ of all feasible α is convex.*

Proof. Let α_1, α_2 be two elements of $\mathcal{A}_n$ and $\gamma \in [0; 1]$. Then we have to prove that the shape representing vector $\alpha := \gamma\alpha_1 + (1-\gamma)\alpha_2$ is also feasible, i.e. $\alpha \in \mathcal{A}_n$:

$$\begin{aligned} q_\alpha = q_0 + \alpha^\top \Psi &= \gamma \left(q_0 + \alpha_1^\top \Psi \right) + (1-\gamma) \left(q_0 + \alpha_2^\top \Psi \right) \\ &= \gamma q_{\alpha_1} + (1-\gamma) q_{\alpha_2} \in \mathcal{Q}. \end{aligned}$$

□

4. Convex Functionals on a Convex Domain

In the following, we propose to compute image segmentations by minimizing cost functionals of the following (very general) form:

$$E(\alpha) = E_i(q_\alpha) + \gamma E_s(\alpha) \tag{4}$$

In particular, we consider image energies of the form

$$\begin{aligned} E_i(q) \quad = \quad & \int f(x)\, q(x)\, dx + \int g(x)\, (1 - q(x))\, dx \\ & + \int h(x)\, |\nabla q(x)|\, dx, \end{aligned} \tag{5}$$

where f and g are arbitrary functions and $h \geq 0$. Note that for traditional (deterministic) representations of shape by binary functions $q : \mathbb{R}^2 \to \{0, 1\}$, the first two terms correspond to integrals of f and g over the inside and outside of the shape, while the last term corresponds to an integral of h along the boundary of the shape. The last term is often referred to as the weighted Total Variation norm [17]. In this sense, the above functional is an extension of classical segmentation schemes [13, 2, 11] from the traditional shape space to the space $\mathcal{Q}$. Meaningful choices of f, g and h for image segmentation are given by:

$$f = -\log p_{ob}(I), \;\; g = -\log p_{bg}(I), \;\; h = \frac{1}{1 + |\nabla I|}, \tag{6}$$

where p_{ob} and p_{bg} represent the color histograms (probabilities) of object and background [14], while h acts as an edge indicator.

For the shape energy E_s in (4), we consider the following statistical shape priors.

1. **Static uniform shape priors:** The distribution of training shapes is assumed to be uniform within the eigenmode space $\mathcal{A}_n$. Such a model was introduced for level set functions in [19], it corresponds to setting $\gamma = 0$ in (4).

2. **Static Gaussian shape priors:** The distribution of training shapes is assumed to be Gaussian, leading to a Mahalanobis type energy of the form:

$$E_s(\alpha) = \alpha^\top \Sigma^{-1} \alpha, \tag{7}$$

A related model was proposed for level set functions in [16].

3. **Dynamical shape priors:** The evolution of shape vectors α is modeled by a linear dynamical system, giving rise to the following shape energy for the segmentation of an image at time t:

$$E_s(\alpha) = (\alpha - v_t)^\top \Sigma^{-1} (\alpha - v_t), \tag{8}$$

where $v_t = \sum_i A_i \hat{\alpha}_{t-i}$ is the prediction by the Markov chain based on shape estimates $\hat{\alpha}_{t-1}, \ldots, \hat{\alpha}_{t-k}$ obtained for the last k images. A related model for level set functions was introduced in [5].

Proposition 2. *The image segmentation with statistical shape priors according to* (4) *amounts to the minimization of a convex functional over a convex set.*

Proof. Due to Lemma 2 we know that the optimization domain $\mathcal{A}_n$ of feasible α values is a convex set. Therefore, we need to prove that the functional (4) is convex in $\alpha \in \mathcal{A}_n$ for any choice of data term and shape prior discussed above.

The statistical priors (7) and (8) introduced above for E_s are both quadratic in α with the positive definite Hessian Σ^{-1} and therefore convex. The functional (5) is convex in q, the first two terms being linear in q and the weighted TV norm being convex. Since q_α is linear in α – see equation (3) – the image energy (5) is a composition of a convex and a linear function and thus convex in α. □

Optimization by Iterated Projections

Since the energy E is convex, a gradient descent approach would always lead to the global optimum α_0. If this α_0 is within the domain $\mathcal{A}_n$ of feasible α, we have thus found a way to compute $\hat{E} := \min_{\alpha \in \mathcal{A}_n} E(\alpha)$. But problems may occur if α_0 is outside of this convex set $\mathcal{A}_n$. In this case, we will perform an iterative projection scheme that will define a sequence $(\alpha_k)_{k \in \mathbb{N}}$ of α values outside of $\mathcal{A}_n$ which will converge towards an α value $\alpha' \in \mathcal{A}_n$ (cf. Figure 2). At first, we project the given shape function q_{α_k} onto $\mathcal{Q}$ which amounts to setting all values to 1 or 0 which are above or below these levels respectively. By projecting this function onto $\mathbb{R}^n$, we receive the next α-value of our sequence. Combining this iterated projecting with a gradient descent approach in respect to α, we are able to calculate the minimum $\hat{E}$ of the energy functional within $\mathcal{A}_n$ itself.

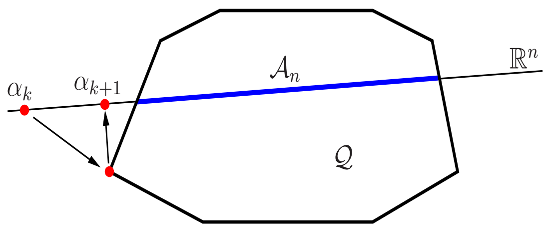

Figure 2. Iterated Projections. The intersection of the convex shape set $\mathcal{Q}$ with $\mathbb{R}^n$ results in the convex set $\mathcal{A}_n$. Starting with $\alpha_k \notin \mathcal{A}_n$, a solution in $\mathcal{A}_n$ is obtained by iterated projections.

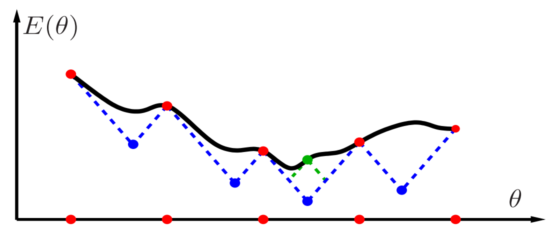

Figure 3. Lipschitz approach. If E (solid line) is Lipschitz continuous (cf. Definition 2), then one can globally minimize it in a continuous sense by iteratively finding lower bounds.

5. Lipschitz Optimization for Transformations

In addition to the deformations modeled by linear combinations of eigenmodes, meaningful shape priors should allow for transformations of shapes such as translation and rotation. To this end, we define an energy functional $E(\alpha, \theta)$ that depends on shape parameters $\alpha \in \mathcal{A}_n$ and transformation parameters $\theta \in \mathrm{SE}(2)$ modeling rigid body motion:

$$E(\alpha, \theta) = E_i(q_\alpha(\theta x)) + \gamma E_s(\alpha) \tag{9}$$

We assume that every shape of χ is registered with respect to rigid body transformation. Then every shape lies within a radius of ρ near the origin, and all shapes are bounded by a function q_{supp}:

$$\forall \alpha \in \mathcal{A}_n\colon\ q_\alpha(x) \le q_{\text{supp}}(x) = \begin{cases} 1, & \text{if } |x| \le \rho \\ 0, & \text{else} \end{cases} \tag{10}$$

While the functional $E(\alpha, \theta)$ is convex in the deformation parameters α, it is generally not convex in θ. Nevertheless, we will show that under mild regularity assumptions, the function

$$\hat{E}(\theta) := \min_{\alpha \in \mathcal{A}_n} E(\alpha, \theta), \tag{11}$$

can be globally optimized on $\mathrm{SE}(2)$ using the idea of Lipschitz continuity (cf. [9, 12]).

The key idea is that we can efficiently compute $\hat{E}(\theta)$ for a given θ by performing the proposed method of *iterated projections* introduced above. To find the global optimum of $\hat{E}(\theta)$, we iteratively subdivide the θ-domain into multiple smaller domains – see Figure 3. For every sub-domain $D \subset \mathrm{SE}(2)$, we calculate the energy at one chosen sample θ_0 which provides an *upper bound* for the global minimum. Provided that the gradient of $\hat{E}(\theta)$ is bounded, a *lower bound* for each sub-domain D can be determined. By performing a branch-and-bound method, we subdivide the sub-domains with the most promising lower bounds. In doing so, we iteratively find tighter lower bounds and terminate once sufficient accuracy is obtained. To determine the lower bound for a sub-domain, we assume that the functional does not oscillate too rapidly. In other words, we need to assume that the following Lipschitz condition holds:

Definition 2 (Lipschitz)**.** A functional $E : \mathrm{SE}(2) \to \mathbb{R}$ is called *Lipschitz continuous* if there exists a uniform $L \in \mathbb{R}$ such that for all $\theta_1, \theta_2 \in \mathrm{SE}(2)$ the following inequality is fulfilled:

$$|E(\theta_1) - E(\theta_2)| \le L \mathrm{dist}(\theta_1, \theta_2)$$

For differentiable functionals E this definition is equivalent to the property that the derivative $\frac{\mathrm{d}E}{\mathrm{d}\theta}$ is bounded by L.

In order to prove that $\hat{E}(\theta)$ is Lipschitz continuous, we proceed as follows: First, we will estimate a Lipschitz constant L for the functional $E(\alpha, \theta)$. Afterwards, we will show in Proposition 3 that the same constant L is also a Lipschitz constant for $\hat{E}$ itself. For simplicity, we assume that the $h \equiv 0$ in (6). Yet, the Lipschitz approach is easily extended to functionals that include an edge indicator h.

Lemma 3. *If all functions of the training set χ are Lipschitz continuous with constant L_χ, then $E(\alpha, \theta)$ is Lipschitz continuous.*

Proof. Any transformation $\theta \in \mathrm{SE}(2)$ can be written as $\theta x = R(x+t)$, with a rotation $R \in \mathrm{SO}(2)$ and a translation $t \in \mathbb{R}^2$.

$$\begin{aligned}
\left|\frac{\mathrm{d}E_i}{\mathrm{d}R}\right| &= \left|\int (f-g)(x) \frac{\mathrm{d}}{\mathrm{d}R} q(\theta x) \mathrm{dx}\right| \\
&= \left|\int (f-g)(x) \det\left(x+t, \nabla q(\theta x)\right) \mathrm{dx}\right| \\
&\le L_\chi \rho \int |(f-g)(x)|\, dx \\
\left|\frac{\mathrm{d}E_i}{\mathrm{d}t}\right| &= \left|\int \left(\nabla g(x-t) - \nabla f(x-t)\right) q(Rx) \mathrm{dx}\right| \\
&\le \int \left|\left(\nabla g(x-t) - \nabla f(x-t)\right)\right| q_{\text{supp}}(x) \mathrm{dx} \\
&\le \left[\int |\nabla f(x) - \nabla g(x)|^2 \, \mathrm{dx}\right]^{1/2} \sqrt{\pi} \rho
\end{aligned}$$

with q_{supp} defined in (10). Since $\left|\frac{\mathrm{d}E}{\mathrm{d}\theta}\right|^2 = \left|\frac{\mathrm{d}E_i}{\mathrm{d}R}\right|^2 + \left|\frac{\mathrm{d}E_i}{\mathrm{d}t}\right|^2$, we have found a uniform upper bound for ∇E that does not depend on $(\theta, \alpha) \in \mathrm{SE}(2) \times \mathcal{A}_n$. Thus, E is Lipschitz continuous. $\square$

Proposition 3. *Under the above regularity assumptions on the training shapes, the segmentation* $\operatorname{argmin}_{(\alpha,\theta)\in\mathcal{A}_n\times\mathrm{SE}(2)} E(\alpha,\theta)$ *can be determined in a globally optimal manner.*

Proof. It suffices to prove that $\hat{E}(\theta)$ is Lipschitz continuous. Let L the Lipschitz constant of $E(\alpha,\theta)$ and θ_1, θ_2 two different transformations of SE(2). Then, there are two elements $\alpha_1, \alpha_2 \in \mathcal{A}_n$ fulfilling $\hat{E}(\theta_1) = E(\alpha_1, \theta_1)$ and $\hat{E}(\theta_2) = E(\alpha_2, \theta_2)$ resp., i.e.:

$$E(\alpha_1,\theta_1) \leq E(\alpha_2,\theta_1) \quad \wedge \quad E(\alpha_2,\theta_2) \leq E(\alpha_1,\theta_2)$$

Using the first inequality, we obtain

$$\hat{E}(\theta_2) - \hat{E}(\theta_1) \geq E(\alpha_2,\theta_2) - E(\alpha_2,\theta_1) \geq -L\mathrm{dist}(\theta_1,\theta_2),$$

while the second one gives:

$$\hat{E}(\theta_2) - \hat{E}(\theta_1) \leq E(\alpha_1,\theta_2) - E(\alpha_1,\theta_1) \leq L\mathrm{dist}(\theta_1,\theta_2)$$

Thus, $\hat{E}$ is Lipschitz continuous with Lipschitz constant L. □

6. Experimental Results

We introduced an algorithm which allows to compute globally optimal image segmentations with statistical shape priors. This algorithm is based on convex minimization of deformation parameters interlaced with Lipschitz optimization of transformation variables.

To clarify the effect of the Lipschitz approach, we will show a comparison of the algorithm run without and with the Lipschitz optimization for a sequence showing a person walking in a cluttered scene.[1]

To this end, we construct a dynamical shape prior by hand-segmenting a different sequence (showing a different person walking at a different pace). By box-filtering these binary functions, we receive probabilistic shape functions that are Lipschitz continuous according to definitions 1 and 2. After dimension reduction via PCA, we train an autoregressive model on the parameter space $\mathcal{A}_6$.

As image energy, we use the approach (6) where $f(x)$ and $g(x)$ are the negative log probability for the observed intensity given that the pixel x is part of the foreground or the background respectively. Neglecting the edge indicator term h in (5), we receive the following energy functional:

$$E(\alpha,\theta) := \int_{\mathbb{R}^d} \log\left(\frac{p_{bg}(I)}{p_{ob}(I)}\right) q_\alpha(\theta x)\mathrm{dx} + \gamma\,|\alpha - v_t|^2_{\Sigma^{-1}}\,,$$

where $|\alpha - v_t|^2_{\Sigma^{-1}} := (\alpha - v_t)^\top \Sigma^{-1}(\alpha - v_t)$ describes the energy of the dynamic shape prior – see equation (8).

During our experiments, we compare a pure gradient descent on α and θ with the proposed Lipschitz approach. Figure 4, top row, shows that the pure gradient descent works well in the presence of partial occlusions such as the table. Yet, it fails to cope with larger occlusions where the local optimization gets stuck in a local minimum with respect to θ. In addition, the gradient descent approach obviously requires an appropriate initialization. Both of these drawbacks are resolved by the proposed global optimization based on the combination of convexity and Lipschitz optimization – see Figure 4, second row.

In Figure 5, we show that our algorithm provides a reliable criterion to determine whether a computed result is consistent with data or not: Reliable segmentations correspond to low (negative) energy, while unreliable ones (full occlusion) correspond to high (positive) energy.

7. Conclusion

In this paper, we proposed an algorithm which computes globally optimal image segmentations with statistical shape priors in a continuous setting. To this end, we proposed a relaxed definition of shape where each point is associated with a probability that it is part of the shape. We proved that this definition leads to convex segmentation functionals optimized on convex shape spaces. Using Lipschitz theory, we proved that a large class of segmentation functionals with geometric, static or dynamical statistical shape priors can be globally minimized. Experimental comparison on tracking a walking person demonstrates that a person can be reliably tracked through clutter and occlusions without the need to (re)initialize.

Acknowledgments

This work was supported by the German Research Foundation (DFG), grants number CR250/1-2 and CR250/4-1.

References

[1] Y. Boykov and M.-P. Jolly. Interactive organ segmentation using graph cuts. In *MICCAI*, volume 1935 of *LNCS*, pages 276–286. Springer, 2000.

[2] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 694–699, Boston, USA, 1995.

[3] A. Chambolle. Total variation minimization and a class of binary mrf models. In *Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition*, number 3757 in LNCS, pages 136–152. Springer, 2005.

[4] T. Chan, S. Esedoḡlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and de-

[1] While more accurate results may be obtained with a user-specified stick-figure model, one should keep in mind that the proposed method does not require any user interaction in the model building. In can directly be applied to arbitrary (binary) shapes in arbitrary dimension.

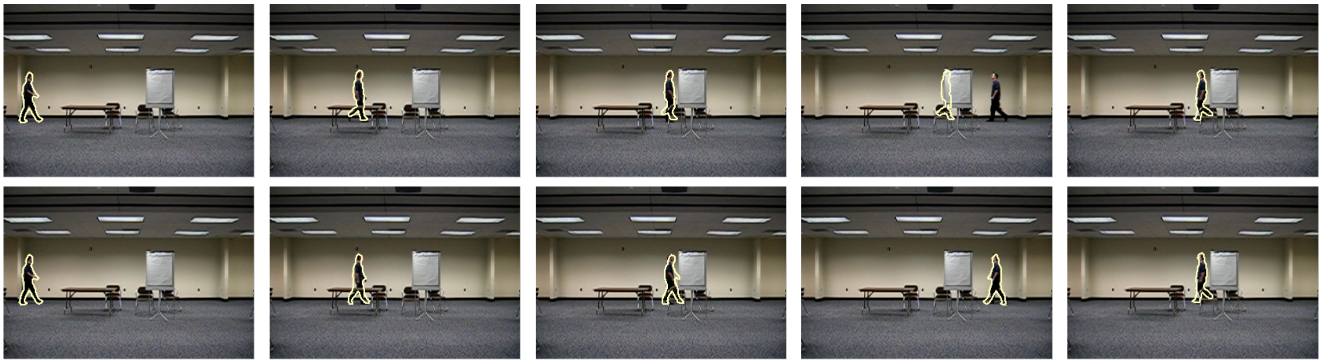

Figure 4. Local versus global optimality. Image segmentation with a dynamical shape prior, implemented by **gradient descent (top row)** and by **Lipschitz optimization (bottom row)**. While gradient descent can handle partial occlusion by the table, it fails to handle total occlusion (4th image). The proposed Lipschitz optimization, on the other hand, guarantees the globally optimal solution and therefore reliably tracks the person upon reappearing from behind the white board.

Figure 5. Detection of inconsistent solutions. The plot on the left shows the image energy E_i as a function of the frame number. Red crosses indicate the four frames shown on the right. Incorrect segmentation results due to a total occlusion of the object of interest can be automatically identified and suppressed (2nd and 3rd frame).

noising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[5] D. Cremers. Dynamical statistical shape priors for level set based tracking. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 28(8):1262–1273, August 2006.

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

[7] A. Duci and A. C. Mennucci. Banach-like metrics and metrics of compact sets, 2007.

[8] 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.

[9] P. Hansen and B. Jaumard. Lipschitz optimization. In P. Horst and P. Pardalos, editors, *Handbook of Global Optimization*. Kluwer Academic Publishers, 1995.

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

[11] S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *IEEE Int. Conf. on Computer Vision*, pages 810–815, 1995.

[12] H. Li and R. Hartley. The 3d-3d registration problem revisited. In *IEEE Int. Conf. on Computer Vision*, Rio de Janeiro, 2007.

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

[14] M. Rousson and D. Cremers. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In *MICCAI*, volume 1, pages 757–764, 2005.

[15] M. Rousson and N. Paragios. Prior knowledge, level set representations & visual grouping. *Int. J. of Computer Vision*, 76(3):231–243, March 2008.

[16] M. Rousson, N. Paragios, and R. Deriche. Implicit active shape models for 3d segmentation in MRI imaging. In *MICCAI*, volume 2217 of *LNCS*, pages 209–216. Springer, 2004.

[17] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

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

[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 *Comp. Vision Patt. Recog.*, pages 463–468, Kauai, Hawaii, 2001.