IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2008, Anchorage, Alaska. ©IEEE.

Matching Non-rigidly Deformable Shapes Across Images: A Globally Optimal Solution

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

Abstract

While global methods for matching shapes to images have recently been proposed, so far research has focused on small deformations of a fixed template.

In this paper we present the first global method able to pixel-accurately match non-rigidly deformable shapes across images at amenable run-times. By finding cycles of optimal ratio in a four-dimensional graph – spanned by the image, the prior shape and a set of rotation angles – we simultaneously compute a segmentation of the image plane, a matching of points on the template to points on the segmenting boundary, and a decomposition of the template into a set of deformable parts.

In particular, the interpretation of the shape template as a collection of an a priori unknown number of deformable parts – an important aspect of higher-level shape representations – emerges as a byproduct of our matching algorithm. On real-world data of running people and walking animals, we demonstrate that the proposed method can match strongly deformed shapes, even in cases where simple shape measures and optic flow methods fail.

1. Introduction

Related work Following a series of seminal works in the late 80's [8, 1, 10, 1, 14], image segmentation by minimizing appropriate energies has become a central focus in Computer Vision research. More recently, it was shown that under certain conditions globally optimal solutions can be obtained [2].

However, for most real-world applications of image segmentation purely low-level information does not provide the desired segmentations. For segmenting and tracking familiar objects, researchers have therefore introduced prior knowledge about the shape of objects [13, 17, 5, 4]. The resulting segmentation processes were shown to reliably segment the objects of interest despite missing or misleading information, e.g. due to noise, background clutter or partial occlusions. Nevertheless, most existing approaches suffer from the following limitations:

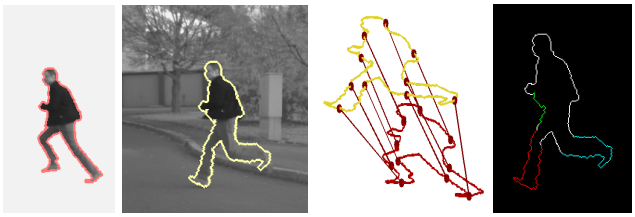


Figure 1. Given a silhouette in one image, we want to identify a corresponding silhouette in another image. The proposed method simultaneously determines this matching, a point alignment between the matched silhouettes and a decomposition into deformable parts (see color-coded image).

- Respective energies are often minimized *locally*. Not only does this require appropriate initialization, there is also no guarantee regarding the quality of computed solutions.
- Respective shape priors are typically based on rather simple geometric distances. As a consequence, they often require large amounts of training data to sufficiently cover the space of permissible shapes. Collecting the necessary training silhouettes as well as inferring alignments and statistical models is a tedious process. In addition, the resulting segmentation process only allows for very little generalization to novel views.
- The notion of *point correspondences* in the matching of shapes is often treated superficially. It is either handled by a reparameterization in the learning process [5] or completely ignored, reverting to an implicit representation optimized by the level set method [13].

Two combinatorial solutions to optimally find shapes in images were proposed in [3, 6]. Yet, both have practical limitations: The first one only applies to *open* curves. The second suffers from a quadratic memory complexity, prohibiting pixel-accurate globally optimal segmentations in

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

reasonable-sized images for many years to come.

In [16] we proposed a more efficient combinatorial solution which allows to generate pixel-accurate segmentations in a matter of seconds. Yet, it merely allows for rather simple elastic deformations of a single template. In particular, rotation is only handled as a global transformation. If parts of the shape rotate locally – like the fingers of a hand – performance deteriorates significantly. Such methods are ill suited for tracking deformable objects like the silhouettes of walking (or running) persons and animals.

A large body of literature is dedicated to the study of shape. Sophisticated shape distances typically take into account the correspondence of point pairs. Moreover, the partitioning of shapes into a collection of *deformable parts* was identified to play an important role in human notions of shape similarity. The identification of point or part correspondences is a challenging combinatorial problem [7, 11].

Contribution In this work we introduce the first pixel-accurate globally optimal algorithm to impose a shape similarity measure in image segmentation which considers large scale deformations. By minimizing a single functional, the proposed algorithm computes an optimal matching of template points to image pixels, while simultaneously partitioning the silhouette into a set of deformable parts.

The key idea is to map every possible solution to a cycle in a four-dimensional graph spanned by the image, the prior shape and a set of rotation angles. The optimal cycle in the graph is then found in effectively linear time. Due to a penalty on changes in the rotation angle, the silhouette is simultaneously partitioned into the (automatically determined) optimal number of parts. On real-world sequences we demonstrate that our algorithm provides accurate matching of substantially deformed persons and animals.

2. Ratio Energies for Shape Knowledge in Image Segmentation

The central contribution of this work is a generalization of the method in [16] for matching shapes to images: Where previously an essentially rigid model was used, we are able to handle shape templates consisting of an (a priori unknown) number of deformable parts.

In [16] the countour of a rigid template, given as a curve $S : \mathbb{S}^1 \to I\!R^2$, is elastically matched to an image $I : \Omega \subset I\!R^2 \to I\!R$, i.e. an object silhouette $C : \mathbb{S}^1 \to \Omega$ is determined. We only consider uniform parameterizations of C, i.e. with constant derivative $||C_s(s)||$ everywhere. Simultaneously a monotone matching $m : \mathbb{S}^1 \to \mathbb{S}^1$ is determined which brings points on C in correspondence with points on S: A point $C(s)$ is matched to the point $S(m(s))$. The two functions are combined into a single function $\Gamma : \mathbb{S}^1 \to \Omega \times \mathbb{S}^1$, with $\Gamma(s) = (C(s), m(s))$. The optimal pair of image contour and matching Γ is found by minimizing an energy in the form of a line integral. This integral consists of a data term encouraging the contour to coincide with image edges, a shape similarity term and a regularization term for the matching:

$$E(\Gamma) = E_{data}(\Gamma) + \nu\, E_{shape}(\Gamma) + \lambda\, E_{reg,m}(\Gamma), \quad (1)$$

The data term is realized as a positive edge detector function assigning low values to high image gradients:

$$g(\mathbf{x}) = \frac{1}{1 + |\nabla I(\mathbf{x})|}, \quad (2)$$

The shape term corresponds to a comparison of the tangent angles $\alpha_C(s)$ on C with the corresponding tangent angle $\alpha_S(m(s))$ on S. The squared cyclic distance on $\mathbb{S}^1$ is taken. The regularity term for m penalizes length distortions: While a part of C may correspond to a part of S of different length, the amount of distortion is penalized. This amount is given by the derivative of a scaled version of the matching function $\tilde{m}(s) = \frac{l(S)}{l(C)} m(s)$. The used penalty function is

$$\Psi(\tilde{m}') = \begin{cases} \tilde{m}' - 1 & \text{if } K \geq \tilde{m}' \geq 1 \\ (\tilde{m}')^{-1} - 1 & \text{if } \frac{1}{K} \leq \tilde{m}' < 1 \\ \infty & \text{otherwise} \end{cases} \quad (3)$$

where K is a pre-defined constant limiting the maximal length distortion. Together, these terms result in the minimization problem

$$\min_{\Gamma=(C,m)} \int_{\mathbb{S}^1} g(C(s))\, ds + \lambda \int_{\mathbb{S}^1} \Psi(\tilde{m}'(s))\, ds + \nu \int_{\mathbb{S}^1} |\alpha_C(s) - \alpha_S(m(s))|^2\, ds \quad (4)$$

This functional can be written as a ratio energy: Instead of the uniform parameterization one can parameterize by arc length and divide by the length of C. The (globally) optimal Γ can be found by the Minimum Ratio Cycle algorithm [12] on a suitable regular graph.

For matching deformable shapes across images, the model described above has two severe limitations:

- The shape dissimilarity encoded in (4) merely allows for small elastic deformations of a rigid template. While invariance to global rotations can be introduced by rerunning the algorithm for various rotations of the template, local rotations of parts of the shape (the fingers of a hand or the legs of a running person) are not supported.
- When allowing more severe deformations (as done in this work), the edge indicator function $g(\mathbf{x})$ is too weak as a data term: If the image contains regions with high gradients like the twigs or leaves of trees, the optimal object silhouette is placed within these regions with no shape deformation at all.

Both drawbacks will be resolved in the following section.

3. Matching Non-rigidly Deformable Shapes Across Images

In this section we generalize the above framework from elastically deforming templates to flexible templates consisting of an a priori unknown number of deformable parts. To this end we increase the dimension of the optimization space by adding a rotation function $a(\cdot)$ modelling the local rotation of parts. The extended function

$$\Gamma : \mathbb{S}^1 \;\rightarrow\; \Omega \times \mathbb{S}^1 \times \mathbb{S}^1, \tag{5}$$

with $\Gamma(s) = \big(C(s), m(s), a(s)\big)$, now associates each point $C(s)$ with a point $S(m(s))$ on the template and a local rotation angle $a(s)$. The energy function is extended by a regularization of a:

$$\begin{aligned} E(\Gamma) \;\;=\;\; & E_{data}(\Gamma) \,+\, \nu\, E_{shape}(\Gamma) \\ & +\lambda\, E_{reg,m}(\Gamma) + \rho E_{reg,a}(\Gamma), \end{aligned} \tag{6}$$

While $E_{reg,m}$ is the same as in (4), the other terms are modified in order to solve the problems mentioned in the previous section.

3.1. Enforcing a Part Decomposition

One of the challenges is to decompose the prior contour S into the optimal number of locally rotating parts. Such a part decomposition corresponds to a piecewise constant angle function. We encourage such functions through a regularity term which penalizes the absolute derivative of a:

$$E_{reg,a}(\Gamma) = \int_{\mathbb{S}^1} |a'(s)|\, ds$$

3.2. Patch Comparisons for Matching Shapes Across Images

When matching shapes across images, one is given not only a prior contour S but also a reference image J which contains the prior contour. We now want to find a contour C in the image I that is similar to S *and* encloses an intensity pattern similar to the one enclosed by S.

To this end the edge indicator function is combined with a patch comparison: When assigning a point $\mathbf{y}$ on S to a point $\mathbf{x}$ on C, we compare a patch of J centered at $\mathbf{y}$ with a rotated patch of I, centered at $\mathbf{x}$. The rotation angle is given by the function $a(\cdot)$ introduced above. For the patch comparison we use a robust L_1-measure. With $\mathbf{R}_a$ denoting the rotation by angle a, this is expressed as

$$\begin{aligned} h(\mathbf{x}, \mathbf{y}, a) \;\;=\;\; & \beta \int_{B_r} w(\mathbf{y}, \mathbf{z}) \left| J(\mathbf{y}+\mathbf{z}) - I(\mathbf{x} + \mathbf{R}_a \mathbf{z}) \right| d\mathbf{z} \\ & + g(\mathbf{x}) \end{aligned}$$

with free parameter β, a disc B_r of radius r and a weighting function $w(\cdot,\cdot)$. The weights depend on the angle between $\mathbf{z}$ and the (inwards pointing) curve normal $\mathbf{n}_S$ of the prior shape S:

$$w(\mathbf{y}, \mathbf{z}) = \begin{cases} 1 & \text{if } \mathbf{z} = 0 \\ \max\left(0, \frac{1}{||\mathbf{z}||}\, \mathbf{z}^\top \mathbf{n}_S(\mathbf{y})\right) & \text{else} \end{cases}$$

3.3. Shape and Regularization Energies for Deformable Templates

When allowing local rotations of deformable parts the rotation angles should be taken into account for the shape similarity measure: Now the rotated angle $(\alpha_C(s) - a(s))$ is compared with the prior angle $\alpha_S(m(s))$. All together the following energy minimization problem arises:

$$\begin{aligned} \min_{\Gamma=(C,m,a)} \;\; & \int_{\mathbb{S}^1} h(C(s), S(m(s)), a(s))\, ds \\ & + \nu \int_{\mathbb{S}^1} \left| (\alpha_C(s) - a(s)) - \alpha_S(m(s)) \right|^2 ds \\ & + \lambda \int_{\mathbb{S}^1} \Psi(\tilde{m}'(s))\, ds \;+\; \rho \int_{\mathbb{S}^1} |a'(s)|\, ds \end{aligned} \tag{7}$$

4. Efficient Minimization by Finding Cycles in a Graph

The minimization problem (7) is mapped to finding the optimal cycle in a graph: By design of the $\Psi(\cdot)$ function the length of the object silhouette C can be no longer than $K \cdot |S|$. This allows to partition the function Γ into $K \cdot |S|$ parts, some of which may be empty (or missing). Then, each point on the prior contour S corresponds to at most K points on the contour C. When also discretizing the image locations and the rotation angles, the node set

$$\mathcal{V} = \mathcal{P} \times \mathcal{A} \times \big\{0, \ldots, K \cdot |S|\big\} \tag{8}$$

of a graph arises, where $\mathcal{A}$ is a set of rotation angles and $\mathcal{P}$ the set of image pixels. Each valid Γ corresponds to a cycle in the graph, containing at most one node of form $(\cdot, \cdot, t)$ for fixed t. Therefore the last component partitions the node set into *frames* of nodes, indexed by t.

The edges in the graph interlink these frames. By connecting neighboring image pixels they represent a part of the image contour C. From the last two components of the nodes, the local rotation angles and the correspondence to points on the prior contour are inferred.

For the edges we differentiate between two cases: When assigning the first image pixel to a shape point, one can choose the rotation angle freely. In contrast, when assigning a second (or third etc.) pixel to a shape point, we do not allow a change of rotation angle. Mathematically speaking there are the transitions (where $\mathbf{x}$ and $\mathbf{x}'$ are neighboring

pixels in an 8-neighborhood):

$$(\mathbf{x}', a', s' \cdot K + l) \rightarrow (\mathbf{x}, a, s \cdot K)$$
$$\text{for } s - K \leq s' < s \text{ and } 0 \leq l < K$$

$$(\mathbf{x}', a, s \cdot K + l') \rightarrow (\mathbf{x}, a, s \cdot K + l)$$
$$\text{for } 0 \leq l' < l < K$$

Each edge e is assigned a numerator weight $n(e)$ and a denominator weight $d(e)$. The former reflects the integral along the corresponding part of Γ, multiplied by the length of the covered image contour. The latter weight reflects this length. When following a (discrete) cycle Γ in the graph, the ratio energy

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

reflects the energy of the corresponding continuous Γ.

4.1. Efficiently Finding the Optimal Cycle

The optimal cycle in the described graph is found via the Minimum Ratio Cycle algorithm [12] introduced to Computer Vision by Jermyn and Ishikawa [9].

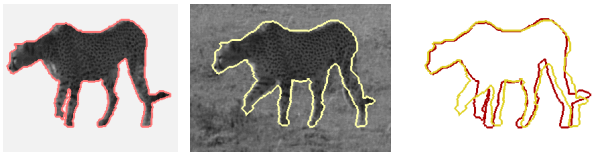
However, some cycles do not correspond to a valid Γ. These cycles contain several nodes of form $(\mathbf{x}, a, 0)$. If such a cycle is returned as global optimum, a recursive partitioning process is started: Frame 0 is partitioned into as many components as nodes of the above form were found, such that no two of these nodes are in the same component. For each component a recursive call is made, where frame 0 only contains the indicated nodes. If one of the recursive calls finds an invalid cycle, the corresponding component is partitioned further and more recursive calls are made. In practice the optimal valid cycle is found in linear time w.r.t. the number of image pixels.

4.2. Search-Space Pruning

While theoretically an arbitrarily fine set of rotation angles can be sampled, in practice the available memory limits the resolution. For a single rotation angle an image of size 352×288 requires roughly 750 MegaByte of memory. This increases linearly with the number of rotation angles.

To optimally exploit the available memory, we introduce mild pruning: We only consider rotation angles between $-75°$ and $+75°$ in steps of $15°$. Additionally we limit the maximal distance between a point on the image contour and its corresponding point on the prior shape contour to be no more than D_{max} in both x- and y- direction, where D_{max} is chosen between 50 and 130.

With these restrictions, the running time of our GPU-based implementation is in the order of 5 minutes, where the major source is the patch comparison.



reference image matched to next image contour deformation

Figure 2. All parts of the cheetah visible in the reference image are correctly matched to the next one.

5. Experiments

On several real-world sequences containing moving humans and animals we demonstrate that the proposed method is able to simultaneously match the prior contour to the correct image position, determine a matching between the two contours and partition the prior shape into a set of deformable parts. To allow significant deformations we set the maximal length distortion to $K = 5$. The patch weight β is set to $10/(\int_{B_r} w(\mathbf{0}, \mathbf{z})\, d\mathbf{z})$, with a disc radius of $r = 5$. Length distortion and deviation of tangent angles are penalized by $\lambda = \nu = 0.5$. The change of local rotation angles requires a higher weight – we set $\rho = 5$ – as it only enters once for each part of the limb.

5.1. Matching Across Images

In this section we consider the matching of a shape in a *fixed* reference image to several frames of an image sequence. As long as the order of limb parts is kept and the parts are not occluded, they are correctly matched. In Figure 2 we show the matching of a cheetah to the next image, where the rotation of the front leg was correctly recovered. The tail is not included in the segmentation as large parts of it are occluded in the reference frame.

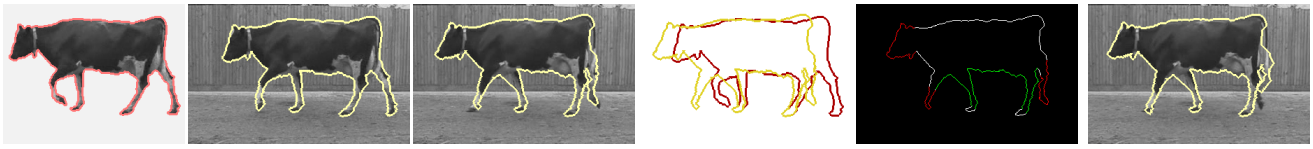
Figure 3 shows a sequence of a running human filmed from a moving car. Despite scale changes and even if the order of limbs changes in the image, the proposed algorithm provides a clear partitioning of the torso and the two legs (see color-coded image in the middle of Figure 3) in an unsupervised manner.

In Figure 4 we show the matching of a cow, where the legs and the soil have very similar intensities[1]. As long as the respective parts are visible, they are reliably found.

5.2. Comparison to State-of-the-art Methods

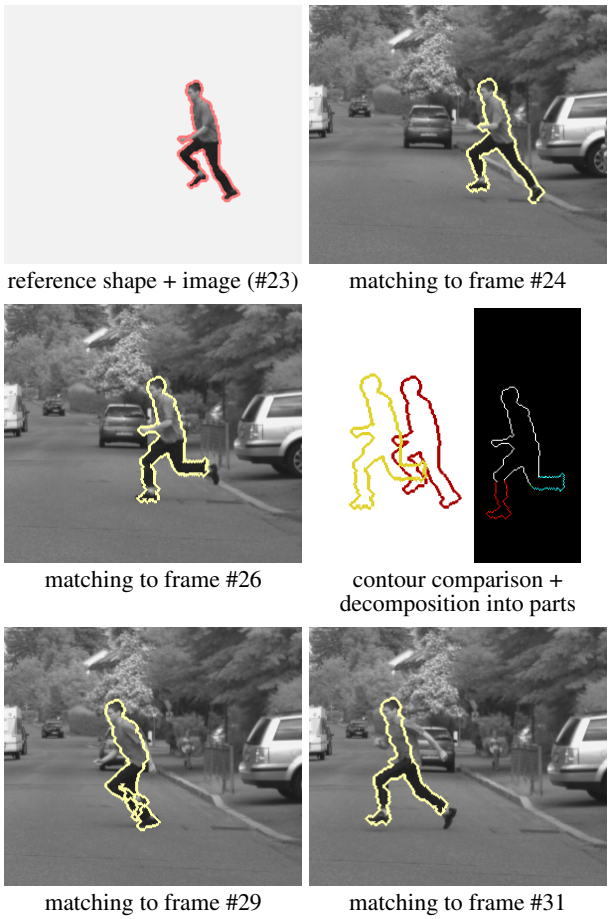
Figure 5 shows a comparison to several state-of-the-art approaches on a sequence where the whole scene is in motion as it was filmed from a moving car. As can be seen, the displacements are far too large for optic flow methods [15] to work well. In addition, such methods cannot include patch comparisons.

[1]Image data provided by D. Magee, University of Leeds. `http://www.robots.ox.ac.uk/ vgg/data/data-various.html`



reference shape + image (frame #8) | direct matching to frame #10 | direct matching to frame #14 | contour deformation for frame #14 | decomposition into parts (frame #14) | #15: the occluded hind leg is matched to the tail

Figure 4. Matching of a walking cow: Despite low contrast between legs and soil, as long as all parts are visible they are found reliably.

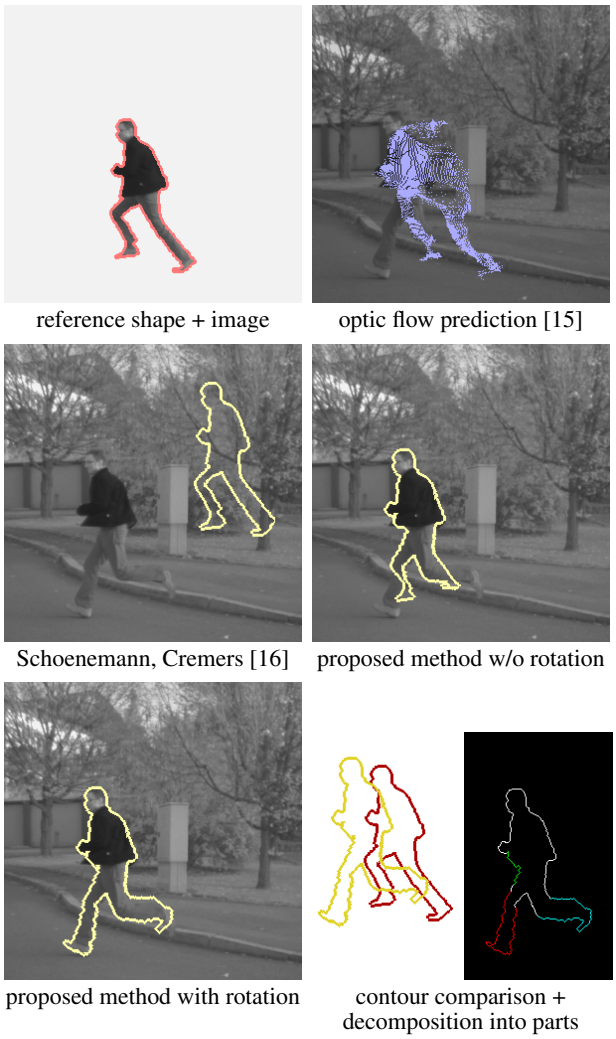


reference shape + image (#23) | matching to frame #24

matching to frame #26 | contour comparison + decomposition into parts

matching to frame #29 | matching to frame #31

Figure 3. The proposed algorithm provides a reliable segmentation and matching across up to seven frames. The arms are not matched as in the prior image only a small part of one arm forms a part of the silhouette. In some cases the right foot is not matched due to a white sock becoming visible.



reference shape + image | optic flow prediction [15]

Schoenemann, Cremers [16] | proposed method w/o rotation

proposed method with rotation | contour comparison + decomposition into parts

Figure 5. Where optic flow and elastic shape measures fail, the proposed method provides substantially better results.

Likewise, our simple elastic method [16] fails: Being based on a simple edge detector and not taking into account the local rotation of parts, the contour is placed in the trees.

In contrast, the contributions of this paper lead to substantial improvements: Without local rotations, the novel patch-based data term identifies those parts correctly that did *not* rotate, while misplacing the rotating legs. When also including local rotation even a combination of large displacements and significant rotation can be accounted for.

5.3. Tracking

So far we have matched a fixed template to several frames of an image sequence. It is natural to extend this to a tracking approach, where the template determined for the last frame is matched to the next. A result is shown in Figure 6. Until the point where the legs cross, head, torso and legs are tracked reliably. We emphasize that our shape measure is *not* designed for general human body motions.

Figure 6. By matching the silhouette determined for the preceeding frame to the next one, the running man is tracked over multiple frames.

We merely demonstrate how it performs on such data.

Conclusion

We present the first global method to impose shape priors which take into account the local rotation of parts in image segmentation. The method simultaneously matches shapes across images, computes an alignment of points on the matched contours and decomposes the shape into an a-priori unknown number of deformable parts.

The solution is determined in a globally optimal manner by computing optimal cycles in a 4D graph, which can be solved in effectively linear time.

On real-world image sequences we demonstrate matching of rotating parts in the presence of large displacements. In particular, our approach works reliably where simpler shape-based methods and optic flow fail.

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

References

[1] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

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

[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.

[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.

[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.

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

[7] 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.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 6(6):721–741, 1984.

[9] 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.

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

[11] 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.

[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.

[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, pp. 316–323, Hilton Head Island, SC, 2000.

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

[15] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, April 2006.

[16] 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.

[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.*, pp. 463–468, Kauai, Hawaii, 2001.