C. Rasmussen et al. (Eds.), *Pattern Recognition*,
Tübingen, Sept. 2004. LNCS Vol. 3175, pp. 36–44.

Kernel Density Estimation and Intrinsic Alignment for Knowledge-driven Segmentation: Teaching Level Sets to Walk

Daniel Cremers[1] and Stanley J. Osher[2] and Stefano Soatto[1]

[1] Department of Computer Science
University of California at Los Angeles, USA

[2] Department of Mathematics
University of California at Los Angeles, USA

Abstract. We address the problem of image segmentation with statistical shape priors in the context of the level set framework. Our paper makes two contributions: Firstly, we propose a novel multi-modal statistical shape prior which allows to encode multiple fairly distinct training shapes. This prior is based on an extension of classical kernel density estimators to the level set domain. Secondly, we propose an intrinsic registration of the evolving level set function which induces an invariance of the proposed shape energy with respect to translation. We demonstrate the advantages of this multi-modal shape prior applied to the segmentation and tracking of a partially occluded walking person.

1 Introduction

When interpreting a visual scene, human observers generally revert to higher-level knowledge about expected objects in order to disambiguate the low-level intensity or color information of the given input image. Much research effort has been devoted to imitating such an integration of prior knowledge into machine-vision problems, in particular in the context of image segmentation.

Among variational approaches, the level set method [16, 10] has become a popular framework for image segmentation. The level set framework has been applied to segment images based on numerous low-level criteria such as edge consistency [13, 2, 11], intensity homogeneity [3, 22], texture information [17, 1] and motion information [6].

More recently, it was proposed to integrate prior knowledge about the shape of expected objects into the level set framework [12, 21, 5, 20, 8, 9, 4]. Building up on these developments, we propose in this paper two contributions. Firstly, we introduce a statistical shape prior which is based on the classical kernel density estimator [19, 18] extended to the level set domain. In contrast to existing approaches of shape priors in level set segmentation, this prior allows to well approximate arbitrary distributions of shapes. Secondly, we propose a translation-invariant shape energy by an intrinsic registration of the evolving level set function. Such a closed-form solution removes the need to locally update explicit pose parameters. Moreover, we will argue that this approach is more

accurate because the resulting shape gradient contains an additional term which accounts for the effect of boundary variation on the location of the evolving shape. Numerical results demonstrate our method applied to the segmentation of a partially occluded walking person.

2 Level Set Segmentation

Originally introduced in the community of computational physics as a means of propagating interfaces [16][3], the level set method has become a popular framework for image segmentation [13, 2, 11]. The central idea is to implicitly represent a contour C in the image plane $\Omega \subset \mathbb{R}^2$ as the zero-level of an embedding function $\phi : \Omega \to \mathbb{R}$:

$$C = \{x \in \Omega \mid \phi(x) = 0\} \tag{1}$$

Rather than directly evolving the contour C, one evolves the level set function ϕ. The two main advantages are that firstly one does not need to deal with control or marker points (and respective regridding schemes to prevent overlapping). And secondly, the embedded contour is free to undergo topological changes such as splitting and merging which makes it well-suited for the segmentation of multiple or multiply-connected objects.

In the present paper, we use a level set formulation of the piecewise constant Mumford-Shah functional, c.f. [15, 22, 3]. In particular, a two-phase segmentation of an image $I : \Omega \to \mathbb{R}$ can be generated by minimizing the functional [3]:

$$E_{cv}(\phi) = \int_\Omega (I-u_+)^2 H\phi(x)dx + \int_\Omega (I-u_-)^2\big(1-H\phi(x)\big)dx + \nu \int_\Omega |\nabla H\phi|dx, \tag{2}$$

with respect to the embedding function ϕ. Here $H\phi \equiv H(\phi)$ denotes the Heaviside step function and u_+ and u_- represent the mean intensity in the two regions where ϕ is positive or negative, respectively. While the first two terms in (2) aim at minimizing the gray value variance in the separated phases, the last term enforces a minimal length of the separating boundary. Gradient descent with respect to ϕ amounts to the evolution equation:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E_{cv}}{\partial \phi} = \delta_\epsilon(\phi) \left[\nu \operatorname{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) - (I-u_+)^2 + (I-u_-)^2\right]. \tag{3}$$

Chan and Vese [3] propose a smooth approximation δ_ϵ of the delta function which allows the detection of interior boundaries.

In the corresponding Bayesian interpretation, the length constraint given by the last term in (2) corresponds to a prior probability which induces the segmentation scheme to favor contours of minimal length. But what if we have more informative prior knowledge about the shape of expected objects? Building up on recent advances [12, 21, 5, 20, 8, 9, 4] and on classical methods of non-parametric density estimation [19, 18], we will in the following construct a shape prior which statistically approximates an arbitrary distribution of training shapes (without making the restrictive assumption of a Gaussian distribution).

[3] See [10] for a precursor containing some of the key ideas of level sets.

Fig. 1. Sample training shapes (binarized and centered).

3 Kernel Density Estimation in the Level Set Domain

Given two shapes encoded by level set functions ϕ_1 and ϕ_2, one can define their distance by the set symmetric difference (cf. [4]):

$$d^2(H\phi_1, H\phi_2) = \int_\Omega \big(H\phi_1(x) - H\phi_2(x)\big)^2 \, dx. \tag{4}$$

In contrast to the shape dissimilarity measures discussed in [20, 8], the above measure corresponds to an L_2-distance, in particular it is non-negative, symmetric and fulfills the triangle inequality. Moreover it does not depend on the size of the image domain (as long as both shapes are entirely inside the image).

Given a set of training shapes $\{\phi_i\}_{i=1\ldots N}$ – see for example Figure 1 – one can estimate a statistical distribution by reverting to the classical Parzen-Rosenblatt density estimator [19, 18]:

$$\mathcal{P}(\phi) \propto \frac{1}{N} \sum_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}\, d^2(H\phi, H\phi_i)\right). \tag{5}$$

This is probably the theoretically most studied density estimation method. It was shown to converge to the true distribution in the limit of infinite training samples (under fairly mild assumptions). There exist extensive studies as to how to optimally choose the kernel width σ. For this work, we simply fix σ to be the mean nearest-neighbor distance:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \min_{j \neq i} d^2(H\phi_i, H\phi_j). \tag{6}$$

The intuition behind this choice is that the width of the Gaussians is chosen such that on the average the next training shape is within one standard deviation.

In contrast to existing shape priors which are commonly based on the assumption of a Gaussian distribution (cf. [12]), the distribution in (5) is a multi-modal one (thereby allowing more complex training shapes). We refer to [7] for an alternative multi-modal prior for spline-based shape representations.

4 Translation Invariance by Intrinsic Alignment

By construction the shape prior (5) is not invariant with respect to certain transformations of the shape ϕ such as translation, rotation and scaling. In the following, we will demonstrate how such an invariance can be integrated analytically by an intrinsic registration process. We will detail this for the case of translation. But extensions to rotation and scaling are straight-forward.

Assume that all training shapes $\{\phi_i\}$ are aligned with respect to their center of gravity. Then we define the distance between a shape ϕ and a given training shape as:

$$d^2(H\phi, H\phi_i) = \int_\Omega \big(H\phi(x - x_\phi) - H\phi_i(x)\big)^2 \, dx, \tag{7}$$

where the function ϕ is evaluated in coordinates relative to its center of gravity x_ϕ given by:

$$x_\phi = \frac{\int x \, H\phi \, dx}{\int H\phi \, dx}. \tag{8}$$

This intrinsic alignment guarantees that in contrast to (4), the distance (7) is invariant to the location of the shape ϕ. The corresponding shape prior (5) is by construction invariant to translation of the shape ϕ. Analogous intrinsic alignments with respect to scale and rotation are conceivable but will not be considered here.

Invariance to certain group transformations by intrinsic alignment of the evolving shape as proposed in this work is different from numerically optimizing a set of explicit pose parameters [5, 20, 8]. The shape energy is by construction invariant to translation. This removes the necessity to intermittently iterate gradient descent equations for the pose. Moreover, as we will see in Section 6, this approach is conceptually more accurate in that it induces an additional term in the shape gradient which accounts for the effect of shape variation on the center of gravity x_ϕ. Current effort is focused on extending this approach to a larger class of invariance. For explicit contour representations, an analogous intrinsic alignment with respect to similarity transformation was proposed in [7].

5 Knowledge-driven Segmentation

In the Bayesian framework, the level set segmentation can be seen as maximizing the conditional probability

$$\mathcal{P}(\phi \,|\, I) = \frac{\mathcal{P}(I \,|\, \phi)\, \mathcal{P}(\phi)}{\mathcal{P}(I)}, \tag{9}$$

with respect to the level set function ϕ, where $\mathcal{P}(I)$ is a constant. This is equivalent to minimizing the negative log-likelihood which is given by a sum of two energies:

$$E(\phi) = \frac{1}{\alpha} E_{cv}(\phi) \; + \; E_{shape}(\phi), \tag{10}$$

with a positive weighting factor α and the shape energy

$$E_{shape}(\phi) = -\log \mathcal{P}(\phi), \tag{11}$$

where $\mathcal{P}(\phi)$ is given in (5).

Minimizing the energy (10) generates a segmentation process which simultaneously aims at maximizing intensity homogeneity in the separated phases and a similarity of the evolving shape with respect to the training shapes encoded through the statistical estimator.

Gradient descent with respect to the embedding function amounts to the evolution:

$$\frac{\partial \phi}{\partial t} = -\frac{1}{\alpha}\frac{\partial E_{cv}}{\partial \phi} - \frac{\partial E_{shape}}{\partial \phi}, \tag{12}$$

with the image-driven component of the flow given in (3) and the knowledge-driven component is given by:

$$\frac{\partial E_{shape}}{\partial \phi} = \frac{\sum \alpha_i \frac{\partial}{\partial \phi} d^2(H\phi, H\phi_i)}{2\sigma^2 \sum \alpha_i}, \tag{13}$$

which simply induces a force in direction of each training shape ϕ weighted by the factor:

$$\alpha_i = \exp\left(-\frac{1}{2\sigma^2}\, d^2(H\phi, H\phi_i)\right), \tag{14}$$

which decays exponentially with the distance from shape ϕ_i.

6 Euler-Lagrange Equations for Nested Functions

The remaining shape gradient in equation (13) is particularly interesting since the translation-invariant distance in (7) exhibits a two-fold (nested) dependence on ϕ. The computation of the corresponding Euler-Lagrange equations is fairly involved. For space limitations, we will only state the final result:

$$\begin{aligned}\frac{\partial}{\partial \phi}\, d^2(H\phi, H\phi_i) = 2\,\delta\left(\phi(x)\right) \Bigg[&\Big(H\phi(x) - H\phi_i(x + x_\phi)\Big) \\ &- \frac{(x - x_\phi)^t}{\int H\phi dx} \int \Big(H\phi(x) - H\phi_i(x + x_\phi)\Big) \nabla H\phi(x)\, dx \Bigg].\end{aligned} \tag{15}$$

Note that as for the image-driven component of the flow in (3), the entire expression is weighted by the δ-function which stems from the fact that the function d only depends on $H\phi$. While the first term in (15) draws $H\phi$ to the template $H\phi_i$ in the local coordinate frame, the second term compensates for shape deformations which merely lead to a translation of the center of gravity x_ϕ. Not surprisingly, this second term contains an integral over the entire image domain because the change of the center of gravity through local deformation of ϕ depends on the entire function ϕ. In numerical experiments we found that this additional term increases the speed of convergence by a factor of 3 (in terms of the number of iterations necessary).

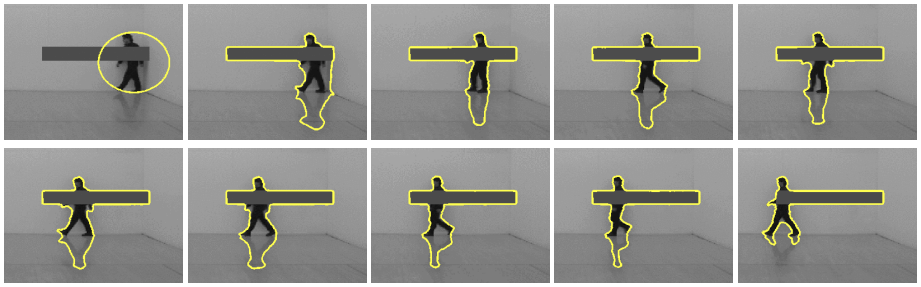

Fig. 2. Various frames showing the segmentation of a partially occluded walking person generated with the Chan-Vese model (2). Based on a pure intensity criterion, the walking person cannot be separated from the occlusion and darker areas of the background such as the person's shadow.

7 Tracking a Walking Person

In the following we apply the proposed shape prior to the segmentation of a partially occluded walking person. To this end, a sequence of a dark figure walking in a (fairly bright) squash court was recorded.[4] We subsequently introduced a partial occlusion into the sequence and ran an intensity segmentation by iterating the evolution (3) 100 times for each frame (using the previous result as initialization). For a similar application of the Chan-Vese functional (without statistical shape priors), we refer to [14]. The set of sample frames in Figure 2 clearly demonstrates that this purely image-driven segmentation scheme is not capable of separating the object of interest from the occluding bar and similarly shaded background regions such as the object's shadow on the floor.

In a second experiment, we manually binarized the images corresponding to the first half of the original sequence (frames 1 through 42) and aligned them to their respective center of gravity to obtain a set of training shape – see Figure 1. Then we ran the segmentation process (12) with the shape prior (5). Apart from adding the shape prior we kept the other parameters constant for comparability.

Figure 3 shows several frames from this knowledge-driven segmentation. A comparison to the corresponding frames in Figure 2 demonstrates several properties of our contribution:

- The shape prior permits to accurately reconstruct an entire set of fairly different shapes. Since the shape prior is defined on the level set function ϕ – rather than on the boundary C (cf. [5]) – it can easily reproduce the topological changes present in the training set.
- The shape prior is invariant to translation such that the object silhouette can be reconstructed in arbitrary locations of the image. All training shapes are centered at the origin, and the shape energy depends merely on an intrinsically aligned version of the evolving level set function.
- The statistical nature of the prior allows to also reconstruct silhouettes which were not part of the training set (beyond frame 42).

[4] We thank Alessandro Bissacco and Payam Saisan for providing the image data.

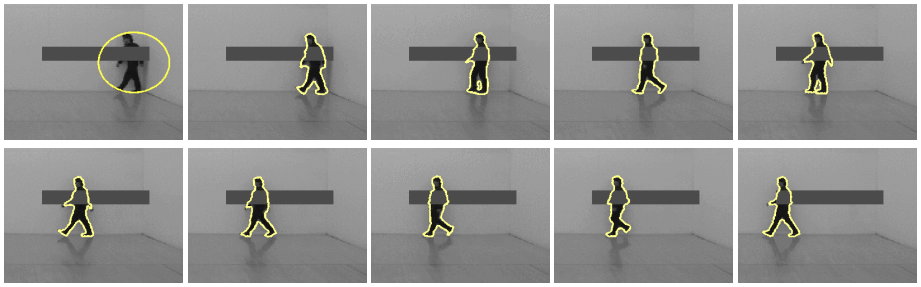

Fig. 3. Segmentation generated by minimizing energy (10) combining intensity information with the statistical shape prior (5). Comparison with the respective frames in Figure 2 shows that the multi-modal shape prior permits to separate the walking person from the occlusion and darker areas of the background such as the shadow. The shapes in the bottom row were not part of the training set.

8 Conclusion

We combined concepts of non-parametric density estimation with level set based shape representations in order to create a statistical shape prior for level set segmentation which can accurately represent arbitrary shape distributions. In contrast to existing approaches, we do not rely on the restrictive assumptions of a Gaussian distribution and can therefore encode fairly distinct shapes.

Moreover, we proposed an analytic solution to generate invariance of the shape prior to translation of the object of interest. By computing the shape prior in coordinates relative to the object's center of gravity, we remove the need to numerically update a pose estimate. Moreover, we argue that this intrinsic registration induces a more accurate shape gradient which comprises the effect of shape or boundary deformation on the pose of the evolving shape.

Finally, we demonstrate the effect of the proposed shape prior on the segmentation and tracking of a partially occluded human figure. In particular, these results demonstrate that the proposed shape prior permits to accurately reconstruct occluded silhouettes according to the prior in arbitrary locations (even silhouettes which were not in the training set).

Acknowledgments

DC and SS were supported by ONR N00014-02-1-0720/N00014-03-1-0850 and AFOSR F49620-03-1-0095/E-16-V91-G2. SO was supported by an NSF IIS-0326388-01, "ITR: Intelligent Deformable Models", Agreement # F5552-01.

References

1. T. Brox and J. Weickert. A TV flow based local scale measure for texture discrimination. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3022 of *LNCS*, pages 578–590, Prague, 2004. Springer.
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. T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
4. T. Chan and W. Zhu. Level set based shape prior segmentation. Technical Report 03-66, Computational Applied Mathematics, UCLA, Los Angeles, 2003.
5. Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. Geiser. Using shape priors in geometric active contours in a variational framework. *Int. J. of Computer Vision*, 50(3):315–328, 2002.
6. D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In C. Dyer and P. Perona, editors, *IEEE Conf. on Comp. Vis. and Patt. Recog.*, volume 1, pages 53–58, June 2003.
7. D. Cremers, T. Kohlberger, and C. Schnörr. Shape statistics in kernel space for variational image segmentation. *Pattern Recognition*, 36(9):1929–1943, 2003.
8. D. Cremers and S. Soatto. A pseudo-distance for shape priors in level set segmentation. In N. Paragios, editor, *IEEE 2nd Int. Workshop on Variational, Geometric and Level Set Methods*, pages 169–176, Nice, 2003.
9. D. Cremers, N. Sochen, and C. Schnörr. Multiphase dynamic labeling for variational recognition-driven image segmentation. In T. Pajdla and V. Hlavac, editors, *European Conf. on Computer Vision*, volume 3024 of *LNCS*, pages 74–86, Prague, 2004. Springer.
10. A. Dervieux and F. Thomasset. A finite element method for the simulation of Raleigh-Taylor instability. *Springer Lecture Notes in Math.*, 771:145–158, 1979.
11. S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *Proc. IEEE Intl. Conf. on Comp. Vis.*, pages 810–815, Boston, USA, 1995.
12. M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. Conf. Computer Vis. and Pattern Recog.*, volume 1, pages 316–323, Hilton Head Island, SC, June 13–15, 2000.
13. R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE PAMI*, 17(2):158–175, 1995.
14. M. Moelich and T. Chan. Tracking objects with the chan-vese algorithm. Technical Report 03-14, Computational Applied Mathematics, UCLA, Los Angeles, 2003.
15. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
16. S. J. Osher and J. A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. of Comp. Phys.*, 79:12–49, 1988.
17. N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. of Computer Vision*, 46(3):223–247, 2002.
18. E. Parzen. On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
19. F. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.
20. M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden et al., editors, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 78–92, Copenhagen, May 2002. Springer, Berlin.
21. 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.
22. A. Tsai, A. J. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Trans. on Image Processing*, 10(8):1169–1186, 2001.