On Local Region Models and the Statistical Interpretation of the Piecewise Smooth Mumford-Shah Functional

Supporting online material

Thomas Brox and Daniel Cremers

Computer Vision Group, University of Bonn,
Römerstr. 164, 53117 Bonn, Germany
{brox,dcremers}@cs.uni-bonn.de

September 5, 2007

Abstract

This report provides additional results on local region statistics and accompanies the work in [2]. In particular, we derive the Euler-Lagrange equations of the local Gaussian region model including variance as well as the local nonparametric model. Moreover, we show some experimental results on contour tracking with local region statistics.

1 Euler-Lagrange Equations for Local Region Statistics

Here we show that for energies based on local region statistics, in contrast to their first-order approximations, one can compute precise shape gradients. In order to minimize the Mumford-Shah functional, one usually alternates the optimization of the smooth region approximations given a preliminary partitioning and, vice-versa, the optimization of the partitioning given the smooth approximations. In fact, this is not a gradient descent but a coordinate descent. In case of local region statistics, the smooth region approximations are available in an analytic form by means of convolution expressions. Thus, their dependence on the partitioning can be considered when computing the Euler-Lagrange equations of the functional. This is *not* possible for the Mumford-Shah energy, since here the smooth approximation can only be computed numerically. Considering the dependence of the statistics on the partitioning, generally leads to additional secondary terms that respond to the changes of the distributions by moving points from one region to another.

Precise shape gradients including secondary terms are not unusual, though they are rarely implemented. A general work on this topic is [1]. In [6] secondary terms are computed for a homogeneous model with nonparametric densities, in [5] the secondary terms are given for the homogeneous Laplace distribution. In the latter work, it is shown empirically that these secondary terms have very little effect in case of the homogeneous Laplace distribution and can be neglected. In [9] it was shown that in case of the homogeneous Gaussian region model the secondary terms in fact turn out to be zero. In [7] the secondary terms for a local Gaussian distribution with fixed variance are derived, and they are brought in a form that allows for an efficient implementation. In case of local region models, implementation of these terms can have positive effects regarding the convergence of the energy [8].

In the following, we derive the Euler-Lagrange equations for the local Gaussian model including variance as well as the local nonparametric model. Contours are represented as zero-level lines of a number of level set functions $\Phi_i : \Omega \to \mathbb{R}$. The number of regions N is assumed to be fixed. The Euler-Lagrange equations we derive are subject to the side conditions: $\bigcup_i \Omega_i = \Omega$ and $\Omega_i \cap \Omega_j = \varnothing \; \forall i \neq j$, i.e., there must be no vacuum or region overlap. How to integrate such side conditions efficiently in case of

multi-region segmentation is described in [3]. In the frequent two-region case, the side conditions are automatically satisfied by using a single level set function that separates foreground and background.

1.1 Gaussian model

Let us first consider the local Gaussian model, i.e., the energy functional

$$E(\Phi_i) = \sum_{i=1}^{N} \int_\Omega H(\Phi_i(\mathbf{x})) \left(\frac{(I(\mathbf{x}) - \mu_i(\mathbf{x}, \Phi_i)^2}{2\sigma_i^2(\mathbf{x}, \Phi_i)} + \frac{1}{2} \log \sigma_i^2(\mathbf{x}, \Phi_i) \right) + \nu |\nabla H(\Phi_i(\mathbf{x}))| \, \mathbf{dx}, \tag{1}$$

where H denotes the step function. For its derivative H' to exist, we assume that H is regularized appropriately. Note that the means and standard deviations depend on the contour represented by Φ_i. The idea of the exact shape gradient is to take exactly this dependency into account. The contribution of the length constraint to the Euler-Lagrange equations is

$$\nu H'(\Phi_i(\mathbf{x})) \operatorname{div} \left(\frac{\nabla \Phi_i(\mathbf{x})}{|\nabla \Phi_i(\mathbf{x})|} \right) \tag{2}$$

and does not depend on the region model. We can thus concentrate on the data term. For better readability, we do not write the subindex i in the following, but simply write Φ for Φ_i, μ for μ_i, and σ for σ_i. We are interested in the Euler-Lagrange equation of

$$E(\Phi) = \int_\Omega H(\Phi(\mathbf{x})) \left(\frac{(I(\mathbf{x}) - \mu(\mathbf{x}, \Phi)^2}{2\sigma^2(\mathbf{x}, \Phi)} + \frac{1}{2} \log \sigma^2(\mathbf{x}, \Phi) \right) \mathbf{dx} \tag{3}$$

First we make use of the analytic expressions for μ and σ in dependence of Φ:

$$\begin{aligned} \mu(\mathbf{x}) &= \frac{\int_\Omega K(\mathbf{x} - \zeta) H(\Phi(\zeta)) I(\zeta) \mathbf{d}\zeta}{\int_\Omega K(\mathbf{x} - \zeta) H(\Phi(\zeta)) \mathbf{d}\zeta} \\ \sigma^2(\mathbf{x}) &= \frac{\int_\Omega K(\mathbf{x} - \zeta) H(\Phi(\zeta)) I^2(\zeta) \mathbf{d}\zeta}{\int_\Omega K(\mathbf{x} - \zeta) H(\Phi(\zeta)) \mathbf{d}\zeta} - \mu^2(\mathbf{x}), \end{aligned} \tag{4}$$

where K can be any appropriate kernel function. In the models discussed in this paper $K = G_\rho$ is an isotropic Gaussian kernel with standard deviation ρ. For computing the Euler-Lagrange equation of (3) we employ the Gâteaux derivative, i.e., we seek

$$\left. \frac{dE(\Phi(\mathbf{x}) + \epsilon h(\mathbf{x}))}{d\epsilon} \right|_{\epsilon \to 0} = 0 \tag{5}$$

for any function $h(\mathbf{x})$. We obtain

$$\begin{aligned} \left. \frac{dE(\Phi(\mathbf{x}) + \epsilon h(\mathbf{x}))}{d\epsilon} \right|_{\epsilon \to 0} &= \int_\Omega H'(\Phi(\mathbf{x})) \left(\frac{(I(\mathbf{x}) - \mu(\mathbf{x}))^2}{2\sigma^2(\mathbf{x})} + \log \sigma(\mathbf{x}) \right) h(\mathbf{x}) \, \mathbf{dx} \\ -\frac{1}{2} \int_\Omega H(\Phi(\mathbf{x})) &\frac{2(I(\mathbf{x}) - \mu(\mathbf{x})) \mu_\Phi(\mathbf{x}) \sigma^2(\mathbf{x}) + ((I(\mathbf{x}) - \mu(\mathbf{x}))^2 - \sigma(\mathbf{x})^2) \sigma_\Phi^2(\mathbf{x})}{\sigma^4(\mathbf{x})} \, \mathbf{dx} \end{aligned} \tag{6}$$

with

$$\mu_\Phi(\mathbf{x}) = \frac{\int_\Omega H'(\Phi(\zeta)) K(\mathbf{x} - \zeta)(I(\zeta) - \mu(\mathbf{x})) h(\zeta) \, \mathbf{d}\zeta}{\int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{x} - \mathbf{z}) \, \mathbf{dz}} \tag{7}$$

and

$$\sigma_\Phi^2(\mathbf{x}) = \frac{\int_\Omega H'(\Phi(\zeta)) K(\mathbf{x} - \zeta) \left(I^2(\zeta) - \frac{\int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{x} - \mathbf{z}) I^2(\mathbf{z}) \mathbf{dz}}{\int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{x} - \mathbf{z}) \mathbf{dz}} \right) h(\zeta) \mathbf{d}\zeta}{\int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{x} - \mathbf{z}) \, \mathbf{dz}} - 2\mu(\mathbf{x}) \mu_\Phi(\mathbf{x}). \tag{8}$$

The first integral in (6) is clearly the usual part considered when applying coordinate descent. The second integral takes charge of the changes in the distribution by varying Φ. Changing the order of integration and substituting $\mathbf{x} \to \mathbf{y}$ and $\zeta \to \mathbf{x}$ in this second integral, we obtain the Euler-Lagrange equation

$$\begin{aligned} 0 = \frac{\partial E(\Phi)}{\partial \Phi} &= H'(\Phi(\mathbf{x})) \left(\frac{(I(\mathbf{x}) - \mu(\mathbf{x}))^2}{2\sigma^2(\mathbf{x})} + \log \sigma(\mathbf{x}) \right) \\ -\frac{1}{2} H'(\Phi(\mathbf{x})) &\int_\Omega \frac{K(\mathbf{y} - \mathbf{x}) H(\Phi(\mathbf{y}))}{\sigma^4(\mathbf{y}) \int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{y} - \mathbf{z}) \, \mathbf{dz}} \left(2(I(\mathbf{y}) - \mu(\mathbf{y}))(I(\mathbf{x}) - \mu(\mathbf{y})) \sigma^2(\mathbf{y}) \right) \mathbf{dy} \\ -\frac{1}{2} H'(\Phi(\mathbf{x})) &\int_\Omega \frac{K(\mathbf{y} - \mathbf{x}) H(\Phi(\mathbf{y})) ((I(\mathbf{y}) - \mu(\mathbf{y}))^2 - \sigma^2(\mathbf{y}))}{\sigma^4(\mathbf{y}) \int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{y} - \mathbf{z}) \, \mathbf{dz}} \left(I^2(\mathbf{x}) - \frac{\int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{y} - \mathbf{z}) I^2(\mathbf{z}) \, \mathbf{dz}}{\int_\Omega H(\Phi(\mathbf{z})) K(\mathbf{y} - \mathbf{z}) \, \mathbf{dz}} - 2\mu(\mathbf{y})(I(\mathbf{x}) - \mu(\mathbf{y})) \right) \mathbf{dy}. \end{aligned}$$

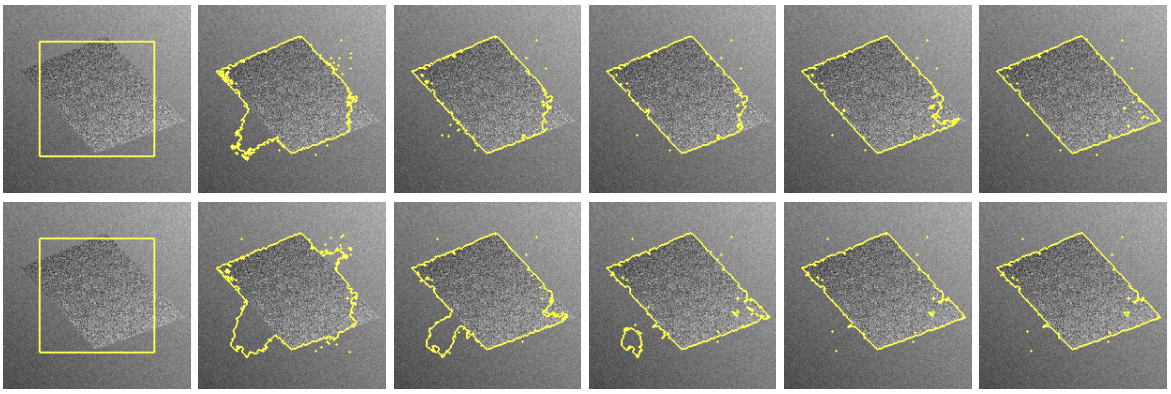

Figure 1: **Top row:** Contour evolution with coordinate descent. **Bottom row:** Contour evolution with gradient descent. **From left to right:** Result after 0, 50, 100, 200, 250, and 400 iterations, $\rho = 12, \nu = 2$.

(9)

The additional terms are rather complex. However, one can verify that for a kernel of infinite width they cancel out. This confirms the result in [9] and indicates that these terms are only important if the kernel width is rather small. In order to allow for a more efficient implementation of these terms we rearrange them and write them by convolutions. With the following abbreviations

$$
\begin{aligned}
F_1(\mathbf{x}) &:= (K * H(\Phi))(\mathbf{x}) = \int_\Omega K(\mathbf{x}-\mathbf{y}) H(\Phi(\mathbf{y}))\mathbf{dy} \\
F_2(\mathbf{x}) &:= (K * (H(\Phi)I)(\mathbf{x}) = \int_\Omega K(\mathbf{x}-\mathbf{y}) H(\Phi(\mathbf{y}))I(\mathbf{y})\mathbf{dy} \quad \rightarrow \quad \mu(\mathbf{x}) = \frac{F_2(\mathbf{x})}{F_1(\mathbf{x})} \\
F_3(\mathbf{x}) &:= (K * (H(\Phi)I^2)(\mathbf{x}) = \int_\Omega K(\mathbf{x}-\mathbf{y}) H(\Phi(\mathbf{y}))I^2(\mathbf{y})\mathbf{dy} \quad \rightarrow \quad \sigma^2(\mathbf{x}) = \frac{F_3(\mathbf{x})}{F_1(\mathbf{x})} - \mu^2(\mathbf{x}) \\
F_4(\mathbf{x}) &:= \left(\bar{K} * \frac{H(\Phi)((I-\mu)^2 - \sigma^2)}{\sigma^4 F_1}\right)(\mathbf{x}) = \int_\Omega \frac{K(\mathbf{y}-\mathbf{x}) H(\Phi(\mathbf{y}))((I(\mathbf{y})-\mu(\mathbf{y}))^2 - \sigma^2(\mathbf{y}))}{\sigma^4(\mathbf{y}) F_1(\mathbf{y})}\mathbf{dy} \\
F_5(\mathbf{x}) &:= \left(\bar{K} * \frac{H(\Phi)\left(2I\sigma^2 - 2\mu(I-\mu)^2\right)}{\sigma^4 F_1}\right)(\mathbf{x}) \\
&= \int_\Omega \frac{K(\mathbf{y}-\mathbf{x}) H(\Phi(\mathbf{y}))\left(2I(\mathbf{y})\sigma^2(\mathbf{y}) - 2\mu(\mathbf{y})(I(\mathbf{y})-\mu(\mathbf{y}))^2\right)}{\sigma^4(\mathbf{y}) F_1(\mathbf{y})}\mathbf{dy} \\
F_6(\mathbf{x}) &:= \left(\bar{K} * \frac{H(\Phi)\left(\sigma^2\left(\frac{F_3}{F_1} - 2I\mu\right) - (I-\mu)^2\left(\sigma^2 - \mu^2\right)\right)}{\sigma^4 F_1}\right)(\mathbf{x}) \\
&= \int_\Omega \frac{K(\mathbf{y}-\mathbf{x}) H(\Phi(\mathbf{y}))\left(\sigma^2(\mathbf{y})\left(\frac{F_3(\mathbf{y})}{F_1(\mathbf{y})} - 2I(\mathbf{y})\mu(\mathbf{y})\right) - (I(\mathbf{y})-\mu(\mathbf{y}))^2\left(\sigma^2(\mathbf{y}) - \mu^2(\mathbf{y})\right)\right)}{\sigma^4(\mathbf{y}) F_1(\mathbf{y})}\mathbf{dy},
\end{aligned}
$$

where $\bar{K}$ denotes the mirrored kernel K, the Euler-Lagrange equation (1.1) becomes more concise

$$
H'(\Phi(\mathbf{x}))\left(\frac{(I(\mathbf{x})-\mu(\mathbf{x}))^2}{2\sigma^2(\mathbf{x})} + \log\sigma(\mathbf{x}) - \frac{1}{2}\left(I^2(\mathbf{x})F_4(\mathbf{x}) + I(\mathbf{x})F_5(\mathbf{x}) + F_6(\mathbf{x})\right)\right) = 0 \tag{10}
$$

and can be implemented rather efficiently using recursive filters. The convolutions $F_1, ..., F_3$ are already needed to implement the primary terms. Hence, the secondary terms require only three additional convolutions and one additional division. The supplementary additions and multiplications are comparatively cheap.

Figure 1 compares coordinate descent and gradient descent (including the additional terms). The gradient descent converges slightly faster. On the other hand, computation takes 7.4ms per region and iteration in case of coordinate descent, whereas the additional terms slow down the gradient descent to 13ms per region and iteration.

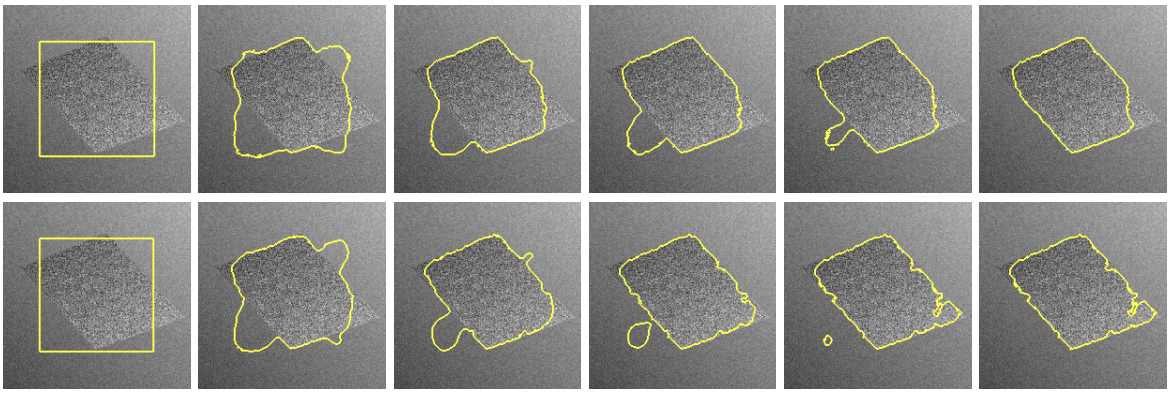

Figure 2: **Top row:** Contour evolution with coordinate descent. **Bottom row:** Contour evolution with gradient descent. **From left to right:** Result after 0, 50, 100, 200, 250, and 400 iterations, $\rho = 12, \nu = 4$.

1.2 Nonparametric model

Similar to the local Gaussian model, we can compute a shape gradient for the local nonparametric model by deriving the Euler-Lagrange equation of

$$E(\Phi) = -\int_\Omega H(\Phi(\mathbf{x})) \log p(I(\mathbf{x}), \mathbf{x}, \Phi)\, \mathbf{dx}. \tag{11}$$

Again we exploit the possibility to express $p(I(\mathbf{x}), \mathbf{x})$ in dependence of Φ analytically:

$$p(I(\mathbf{x}), \mathbf{x}) = \frac{\int_\Omega K(\mathbf{x}-\zeta) H(\Phi(\zeta)) K_h(I(\mathbf{x}) - I(\zeta)) \mathbf{d}\zeta}{\int_\Omega K(\mathbf{x}-\zeta) H(\Phi(\zeta)) \mathbf{d}\zeta}, \tag{12}$$

where $K : \mathbb{R}^2 \to \mathbb{R}_0^+$ denotes a spatial kernel as above, whereas $K_h : \mathbb{R} \to \mathbb{R}_0^+$ is a tonal kernel due to the Parzen estimator. Plugging (12) into (11) we obtain

$$\left.\frac{dE(\Phi(\mathbf{x}) + \epsilon h(\mathbf{x}))}{d\epsilon}\right|_{\epsilon \to 0} = -\int_\Omega H'(\Phi(\mathbf{x})) \log p(I(\mathbf{x}), \mathbf{x}) h(\mathbf{x})\, \mathbf{dx} - \int_\Omega H(\Phi(\mathbf{x})) \frac{p_\Phi(I(\mathbf{x}), \mathbf{x})}{p(I(\mathbf{x}), \mathbf{x})} \mathbf{dx} \tag{13}$$

with

$$p_\Phi(I(\mathbf{x}), \mathbf{x}) = \frac{\int_\Omega K(\mathbf{x}-\zeta) H'(\Phi(\mathbf{x})) \left(K_h(I(\mathbf{x})) - I(\zeta)) - \frac{\int_\Omega K(\mathbf{x}-\mathbf{z}) H(\Phi(\mathbf{z})) K_h(I(\mathbf{x}) - I(\mathbf{z})) \mathbf{dz}}{\int_\Omega K(\mathbf{x}-\mathbf{z}) H(\Phi(\mathbf{z})) \mathbf{dz}} \right) h(\zeta) \mathbf{d}\zeta}{\int_\Omega K(\mathbf{x}-\mathbf{z}) H(\Phi(\mathbf{z})) \mathbf{dz}}. \tag{14}$$

Again, the second part of (13) is due to considering the dependence of $p(I(\mathbf{x}), \mathbf{x})$ on Φ. In this second part, we change the order of integration and substitute $\mathbf{x} \to \mathbf{y}$ and $\zeta \to \mathbf{x}$. Then we obtain the following Euler-Lagrange equation:

$$-H'(\Phi(\mathbf{x})) \left(\log p(I(\mathbf{x}), \mathbf{x}) + \int_\Omega \frac{K(\mathbf{y}-\mathbf{x}) H(\Phi(\mathbf{y})) \Big(K_h(I(\mathbf{y}) - I(\mathbf{x})) - p(I(\mathbf{y}), \mathbf{y}) \Big)}{p(I(\mathbf{y}), \mathbf{y}) \int_\Omega K(\mathbf{y}-\mathbf{z}) H(\Phi(\mathbf{z})) \mathbf{dz}} \mathbf{dy} \right) = 0 \tag{15}$$

By using binned histograms, the integrals can be implemented via convolutions like in the Gaussian model.
Figure 2 compares such an implementation including the additional term to the simple coordinate descent. One can observe that the gradient descent reaches a different, not necessarily better, local minimum. The computation time increases from 90ms to 152ms per region and iteration for an implementation with 32 bins. Note that in contrast to the Gaussian model, the additional term does not cancel out when $\rho \to \infty$.

2 Contour Tracking

The present section briefly demonstrates a reasonable practical application for local region statistics or the piecewise smooth Mumford-Shah functional, respectively: the tracking of silhouettes in image sequences. As we have seen, local minima are the major problem of local region statistics. Hence, in order to work well in practice, the application should provide sufficiently close contour initializations.

In contour tracking, this is the case, at least as long as the motion between successive frames is not too fast. Another such application is, for instance, medical image segmentation [7].

Figure 3 demonstrates tracking the silhouette of a car in a typical traffic scenario. The texture feature space proposed in [4] has been employed in order to enrich the available input data. Nonetheless, regarding the image statistics globally across the whole image, there is no cue that would distinguish the tracked car from the background. Locally, however, the car can be well separated from its surroundings in most of the frames. Hence, a local Gaussian region model including variance can track the car quite well, whereas a homogenous model, though nonparametric, clearly fails, as Figure 4 reveals. Note that in this sequence, the camera moves towards the tracked object. Consequently, the displacements of the object in the image become larger from frame to frame. The last frames in Figure 3 demonstrate that this can lead to initializations that are too far from the sought solution. Such problems could be avoided in a more sophisticated tracking framework that comprises a prediction of the displacement. Here, we purposely show the result of the pure segmentation approach that separates foreground and background in each frame.

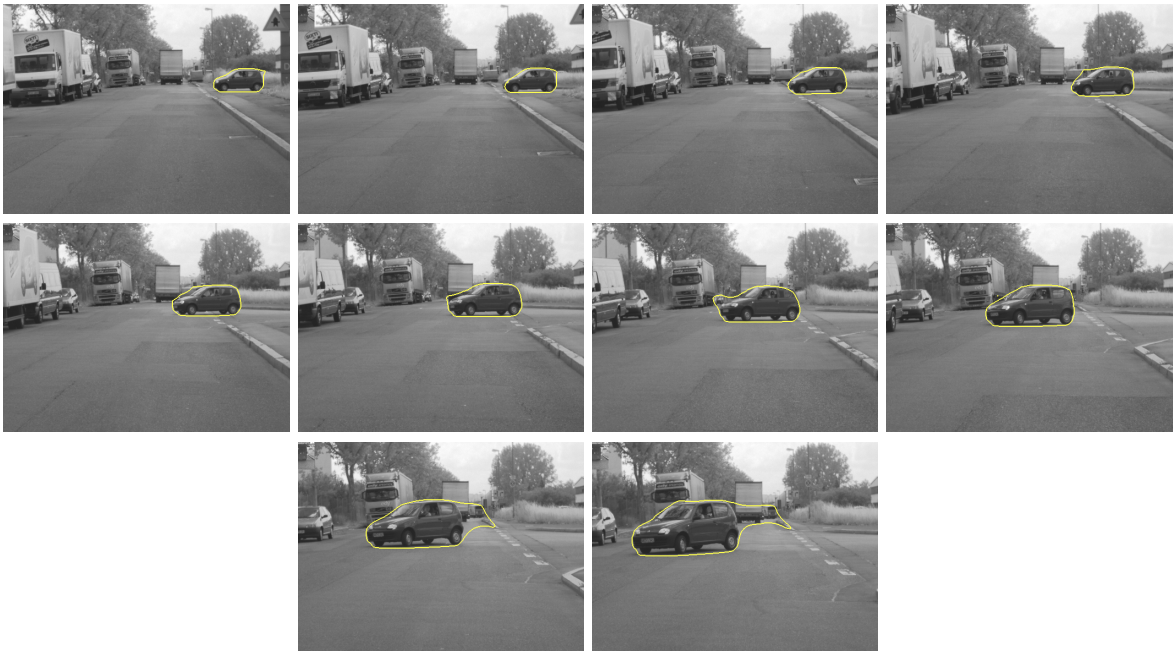

Figure 3: Tracking the silhouette of a car with the Mumford-Shah functional including variance. $\nu = 45, \lambda = 72$. Input channels are, apart from the gray value, the texture features from [4]. Every fifth frame is shown. Note that the camera is moving towards the tracked object. Locally, the car can be well separated from the background (compare Figure 4). Only when the silhouette changes too fast, the contour from the previous frame is not a sufficiently close initialization anymore.

Figure 4: Same setting as in Figure 3 but modeling foreground and background with a global Parzen distribution. Only the strong smoothness constraint ($\nu = 45$) prevents the contour from capturing even larger parts of the background. Clearly, object and background cannot be distinguished based on spatially uniform feature distributions.

References

[1] G. Aubert, M. Marlaud, O. Faugeras, and S. Jehan-Besson. Image segmentation using active contours: calculus of variations or shape gradients? *SIAM Journal on Applied Mathematics*, 63(6):2128–2154, 2003.

[2] T. Brox and D. Cremers. On local region models and the statistical interpretation of the piecewise smooth Mumford-Shah functional. *International Journal of Computer Vision*. Submitted.

[3] T. Brox and J. Weickert. Level set segmentation with multiple regions. *IEEE Transactions on Image Processing*, 15(10):3213–3218, 2006.

[4] T. Brox and J. Weickert. A TV flow based local scale estimate and its application to texture discrimination. *Journal of Visual Communication and Image Representation*, 17(5):1053–1073, Oct. 2006.

[5] M. Heiler and C. Schnörr. Natural image statistics for natural image segmentation. *International Journal of Computer Vision*, 63(1):5–19, 2005.

[6] J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky. A nonparametric statistical method for image segmentation using information theory and curve evolution. *IEEE Transactions on Image Processing*, 14(10):1486–1502, 2005.

[7] J. Piovano, M. Rousson, and T. Papadopoulo. Efficient segmentation of piecewise smooth images. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *LNCS*, pages 709–720. Springer, 2007.

[8] M. Rousson. Personal communication, May 2007.

[9] M. Rousson and R. Deriche. A variational framework for active and adaptive segmentation of vector-valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 56–62, Orlando, Florida, Dec. 2002.