On the Statistical Interpretation of the Piecewise Smooth Mumford-Shah Functional

Thomas Brox and Daniel Cremers

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

Abstract. In region-based image segmentation, two models dominate
the field: the Mumford-Shah functional and statistical approaches based
on Bayesian inference. Whereas the latter allow for numerous ways to
describe the statistics of intensities in regions, the first includes spa-
tially smooth approximations. In this paper, we show that the piece-
wise smooth Mumford-Shah functional is a first order approximation of
Bayesian a-posteriori maximization where region statistics are computed
in local windows. This equivalence not only allows for a statistical inter-
pretation of the full Mumford-Shah functional. Inspired by the Bayesian
model, it also offers to formulate an extended Mumford-Shah functional
that takes the variance of the data into account.

1 Introduction

Since the beginning of image analysis research, there has been enormous interest
in image segmentation. While the topic was handled in a quite heuristic manner
for a long time, a more systematic approach to the problem has been initiated
by three seminal works in the 1980s: the Bayesian formulation of Geman and
Geman [9], the energy functional of Mumford and Shah [18, 19], and the snakes
model by Kass, Witkin, and Terzopoulos [13]. In all these works, the formerly
purely algorithmic description of a segmentation method has been replaced by
its formulation as an optimization problem. This systematic description based on
sound mathematical concepts has considerably improved the understanding of
image segmentation and, hence, supported the development of new models and
better algorithms. The initially large gap between sound energy formulations
and efficient ways to find solutions of these energies, in particular in case of the
Mumford-Shah functional, was bridged by the works of Ambrosio and Tortorelli
[1], Morel and Solimini [16, 17], as well as the use of level set representations of
contours by Caselles et al. [6], Chan and Vese [7], and Paragios and Deriche [22].
A further type of optimization strategy has emerged in the spatially discrete
case with graph cut methods [10, 3].
Whereas all three approaches to image segmentation are based on energy min-
imization, their motivation is quite different. In [26], Zhu and Yuille outlined
many relations between the methods and algorithmic implementations such as

region merging or region growing. In particular, they established a link between a Bayesian approach to image segmentation and the piecewise constant case of the Mumford-Shah functional, sometimes called the *cartoon limit*. Zhu and Yuille also suggested a more general energy functional that replaces the constant approximation of image regions by arbitrary intensity distributions. This formulation was used particularly in level set based segmentation approaches where full Gaussian distributions [24], Laplace distributions [11], and nonparametric kernel densities [14] have been suggested.

Zhu and Yuille established relations between Bayesian methods and the cartoon limit of the Mumford-Shah functional, yet in their work, they ignored the part of the functional that allows also for piecewise *smooth* approximations. In the present paper, we complete their work by showing that the Mumford-Shah functional can be interpreted as a first-order approximation of a Bayesian model with probability densities estimated in local windows. Such types of densities have been used in [4] in the scope of contour-based pose estimation. Similar to the work of Zhu and Yuille [26], this equivalence allows to generalize the Mumford-Shah functional. We demonstrate this by proposing a functional which allows to approximate the input intensity by a piecewise smooth Gaussian distribution including mean *and* variance.

2   The Mumford-Shah Functional

The idea of Mumford and Shah was to find a piecewise smooth approximation $u : (\Omega \subset \mathbb{R}^2) \to \mathbb{R}$ of the image $I : (\Omega \subset \mathbb{R}^2) \to \mathbb{R}$ and an edge set K[1] separating the pieces of u, such that u is close to I and the total length of the edge set is minimal. This can be expressed as minimizing the functional

$$E(u, K) = \int_{\Omega - K} (u - I)^2 \mathbf{dx} + \lambda \int_{\Omega - K} |\nabla u|^2 \mathbf{dx} + \nu \, |K| \to \min, \qquad (1)$$

where $\lambda \geq 0$ and $\nu \geq 0$ are constant weighting parameters. An interesting special case arises for $\lambda \to \infty$, where u is required to be piecewise constant. This case, already discussed by Mumford and Shah in [19], is also known as the *cartoon limit* and can be written in short form

$$E(u, K) = \sum_i \int_{\Omega_i} (u_i - I)^2 \mathbf{dx} + \nu_0 \, |K| \to \min, \qquad (2)$$

where Ω_i denotes the piecewise constant regions separated by K and ν_0 is the rescaled version of the parameter ν in (1). Due to the quadratic error measure, given Ω_i, the solution of u_i is the mean of I within Ω_i. A related approach was independently developed by Blake and Zisserman [2]. In the spatially discrete case, (2) is related to the Potts model [23].

---

[1] Since our focus lies on image segmentation, we will only consider edge sets which are sets of closed curves [17].

The model in (2) can be simplified further by assuming a fixed number of regions N. In particular, the case $N = 2$ and its level set formulation by Chan and Vese [7] has become very popular. A discrete version of the binary case has been introduced by Lenz and Ising for modeling ferromagnetism already in the 1920s [15, 12].

3   Bayesian Model and Local Region Statistics

An alternative approach to image segmentation can be derived using Bayes' rule

$$p(K|I) = \frac{p(I|K)p(K)}{p(I)}. \tag{3}$$

Here one seeks for a partitioning by the edge set K that maximizes the a-posteriori probability given the image I. The first factor in the nominator is in general approximated by an intensity distribution in the regions $i = 1, ..., N$ separated by K. The second factor is the a-priori probability of a certain partitioning K. Usually, the total length of the edge set K is assumed to be small,

$$p(K) = \exp(-\nu_B |K|), \tag{4}$$

but other more sophisticated shape priors can be integrated here, as well [8]. Assuming independence of intensities at different locations $\mathbf{x}$, one can write

$$p(I|K) = \prod_{\mathbf{x} \in \Omega} p(I(\mathbf{x})|K, \mathbf{x})^{\mathbf{dx}}, \tag{5}$$

a continuous product with $\mathbf{dx}$ being the infinitesimal bin size. With the partitioning of Ω by the edge set K into disjoint regions $\Omega = \bigcup_i \Omega_i$, $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$, the product over the whole domain Ω can be separated into products over the regions:

$$\prod_{\mathbf{x} \in \Omega} p(I(\mathbf{x})|K, \mathbf{x})^{\mathbf{dx}} = \prod_i \prod_{\mathbf{x} \in \Omega_i} p(I(\mathbf{x})|\mathbf{x}, \mathbf{x} \in \Omega_i)^{\mathbf{dx}}. \tag{6}$$

For convenience we define the conditional probability density to encounter an intensity s at position $\mathbf{x}$ given that $\mathbf{x} \in \Omega_i$ as

$$p_i(s, \mathbf{x}) := p(s|\mathbf{x}, \mathbf{x} \in \Omega_i). \tag{7}$$

Note that we have here a family of probability densities $p_i(s, \mathbf{x})$ for all $\mathbf{x} \in \Omega$, i.e.,

$$\begin{aligned} &p_i(s, \mathbf{x}) : \mathbb{R} \to \mathbb{R}_0^+ \\ &p_i(s, \mathbf{x}) \geq 0 \quad \forall s \in \mathbb{R}, \forall \mathbf{x} \in \Omega \int_{\mathbb{R}} p_i(s, \mathbf{x}) ds = 1 \forall \mathbf{x} \int \Omega. \end{aligned} \tag{8}$$

In general, it is preferable to express the maximization of (3) by the minimization of its negative logarithm. With the above assumptions, this leads to the energy functional

$$E(K) = \sum_i \int_{\Omega_i} -\log p_i(I(\mathbf{x}), \mathbf{x}) \mathbf{dx} + \nu_B |K|. \tag{9}$$

It obviously resembles the cartoon limit of the Mumford-Shah functional. We will come back to this issue in the next section.

There are several possibilities how to model the probability densities p_i. Typically, one assumes a homogeneous Gaussian distribution in each region Ω_i:

$$p_i(s) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(s-\mu_i)^2}{2\sigma_i^2}\right),$$ (10)

where μ_i and σ_i denote the mean and standard deviation of I in region Ω_i. Other choices like a Laplace distribution [11] or a nonparametric density [14] are possible, as well. All these models apply the same probability density to all points in a region. Hence, we will call them spatially *homogeneous* region models.

In contrast, *local* region models take the spatial position into account, i.e., there is in general a different probability density at each point $\mathbf{x}$ in the region. For a Gaussian distribution this yields [4]:

$$p_i(s, \mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_i(\mathbf{x})} \exp\left(\frac{(s-\mu_i(\mathbf{x}))^2}{2\sigma_i(\mathbf{x})^2}\right).$$ (11)

Estimation of the parameters $\mu_i(\mathbf{x})$ and $\sigma_i(\mathbf{x})$ can be achieved using a window function, e.g. a Gaussian G_ρ with standard deviation ρ, and restricting the estimation only to points within this window:

$$\mu_i(\mathbf{x}) = \frac{\int_{\Omega_i} G_\rho(\zeta - \mathbf{x}) I(\zeta)\, d\zeta}{\int_{\Omega_i} G_\rho(\zeta - \mathbf{x})\, d\zeta} \qquad \sigma_i(\mathbf{x}) = \frac{\int_{\Omega_i} G_\rho(\zeta - \mathbf{x})(I(\zeta) - \mu_i(\mathbf{x}))^2\, d\zeta}{\int_{\Omega_i} G_\rho(\zeta - \mathbf{x})\, d\zeta}.$$ (12)

Obviously, the local region model converges to the corresponding homogeneous model for $\rho \to \infty$.

4   Bayesian Interpretation of the Mumford-Shah Functional

The Bayesian model from the last section is quite flexible in the choice of the probability density function. It further yields a nice statistical interpretation of the model assumptions and allows for the sound integration of a-priori information. On the other hand, the Mumford-Shah functional combines segmentation and image restoration by a piecewise smooth function. The reader may have already noticed similarities between the models in Section 2 and Section 3. In this section, we will investigate the relation between both segmentation approaches aiming at a statistical interpretation of the *full* Mumford-Shah functional.

We start with the Bayesian model in (9). A comparison to the cartoon model in (2) reveals a large similarity. As shown in [26], for a specific choice of the probability densities, both formulations turn out to be equivalent. Indeed, equivalence

of (2) and (9) is established by modeling the probability densities as Gaussian functions with fixed standard deviation

$$p_i(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(s-\mu_i)^2}{2\sigma^2}\right). \tag{13}$$

Applying the logarithm

$$\log p_i(s) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(s-\mu_i)^2}{2\sigma^2} \tag{14}$$

and plugging this into (9) yields

$$\begin{aligned} E(K) &= \sum_i \int_{\Omega_i} \frac{1}{2}\log(2\pi\sigma^2) + \frac{(I(\mathbf{x})-\mu_i)^2}{2\sigma^2}\,\mathbf{dx} + \nu_B\,|K| \\ &= \sum_i \int_{\Omega_i} \frac{(I(\mathbf{x})-\mu_i)^2}{2\sigma^2}\,\mathbf{dx} + \nu_B\,|K| + \text{const.} \end{aligned} \tag{15}$$

Due to the same fixed standard deviation in all regions, the logarithm term containing σ does not depend on K and, hence, is a negligible constant in the energy functional. Also the denominator $2\sigma^2$ is a constant and merely leads to a rescaling of the parameter ν_B. Thus, with $\mu_i \equiv u_i$, $\sigma = \sqrt{0.5}$, and $\nu_B = \nu_0$, (15) states exactly the same energy minimization problem as the cartoon model in (2).

With this equivalence in mind, the question arises, whether there exists a choice of the probability density function that relates the Bayesian model to the full, piecewise smooth Mumford-Shah functional stated in (1). Since (1) explicitly allows the approximation u to vary within a region, a homogeneous region model is obviously not sufficient. Local region statistics, on the other hand, include varying parameters in the region. Hence, having in mind that the equivalence of the Bayesian model and the cartoon model was established for a homogeneous Gaussian region model with fixed standard deviation, we take a closer look at the *local* Gaussian model, again with fixed standard deviation.

Since the standard deviation is fixed, we can focus on the local mean in (12):

$$\mu_i(\mathbf{x}) = \frac{\int_{\Omega_i} G_\rho(\zeta - \mathbf{x}) I(\zeta)\,d\zeta}{\int_{\Omega_i} G_\rho(\zeta - \mathbf{x})\,d\zeta}. \tag{16}$$

The numerator is a convolution of the image I with the Gaussian function G_ρ. The denominator is only for normalization in case the window hits the boundary of Ω_i. It ensures the preservation of the average gray value of μ_i in the domain Ω_i independent of ρ.

In order to bend the bow to the Mumford-Shah functional, we will relate this filtering operation to a regularization framework. Yuille and Grzywacz [25] as well as Nielsen et al. [21] showed that the outcomes of some linear filters are

exact minimizers of certain energy functionals with an infinite sum of penalizer terms of arbitrarily high order. More precisely, it was shown in [21] that filtering an image I with the filter

$$\hat{h}(\omega) = \frac{1}{1 + \sum_{k=1}^{\infty} \alpha_k \omega^{2k}} \tag{17}$$

given in the frequency domain, yields the minimizer of the following energy functional:

$$E(u) = \int_{\mathbb{R}} \left((u - I)^2 + \sum_{k=1}^{\infty} \alpha_k \left(\frac{d^k u}{dx^k} \right)^2 \right) dx. \tag{18}$$

In particular, this includes for $\alpha_k = \frac{\lambda^k}{k!}$, the Gaussian filter

$$\hat{h}(\omega, \lambda) = \frac{1}{1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \omega^{2k}} = \exp(-\lambda \omega^2). \tag{19}$$

This filter corresponds to the Gaussian G_ρ with standard deviation $\rho = \sqrt{2\lambda}$ in the spatial domain. Nielsen et al. further showed in [20] that for Cartesian invariants, such as the Gaussian, this correspondence can be generalized to higher dimensions. Therefore, the convolution result in (16) is the exact minimizer of

$$E(\mu_i) = \int_{\Omega_i} \left((\mu_i - I)^2 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{j_1 + j_2 = k} \left(\frac{d^k \mu_i}{dx^{j_1} dy^{j_2}} \right)^2 \right) \mathbf{dx} \tag{20}$$

with natural boundary conditions.

Based on these findings, we can proceed to generalize the piecewise constant case in (15). We plug the local Gaussian probability density from (11) with fixed standard deviation $\sigma = \sqrt{0.5}$ into the Bayesian model in (9):

$$\begin{aligned} E(\mu, K) &= \sum_i \int_{\Omega_i} \frac{1}{2} \log(2\pi\sigma^2) + \frac{(I(\mathbf{x}) - \mu_i(\mathbf{x}))^2}{2\sigma^2} \mathbf{dx} + \nu_B |K| \\ &= \sum_i \int_{\Omega_i} (I(\mathbf{x}) - \mu_i(\mathbf{x}))^2 \mathbf{dx} + \nu_B |K| + \text{const.} \end{aligned} \tag{21}$$

The means μ_i have in (16) been defined as the results of local convolutions. As we have just found, this convolution result is the minimizer of (20). Hence, we can write the Bayesian energy as:

$$E_B(\mu, K) = \sum_i \int_{\Omega_i} \left((\mu_i - I)^2 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{j_1 + j_2 = k} \left(\frac{d^k \mu_i}{dx^{j_1} dy^{j_2}} \right)^2 \right) \mathbf{dx} + \nu_B |K|. \tag{22}$$

Neglecting all penalizer terms of order $k > 1$ yields

$$E_{MS}(\mu, K) = \sum_i \int_{\Omega_i} \left((\mu_i - I)^2 + \lambda |\nabla \mu_i|^2 \right) \mathbf{dx} + \nu_B |K| + \text{const.} \tag{23}$$

which states exactly the Mumford-Shah functional in (1). Consequently, *minimizing the full piecewise smooth Mumford-Shah functional is equivalent to a first-order approximation of a Bayesian a-posteriori maximization based on local region statistics.* In particular, it is the approximation of the Bayesian setting with a Gaussian distribution, fixed standard deviation $\sigma = \sqrt{0.5}$, and a Gaussian windowing function where $\rho = \sqrt{2\lambda}$ and $\nu_B = \nu$.

What is the effect of neglecting the higher order terms, as done by the Mumford-Shah functional? The main effect is that the minimizers μ_i of the functional in (23) are less smooth than those of the functional in (22). Figure 1 depicts a comparison in case of the whole image domain being a single region. Obviously, the visual difference is almost negligible, and it can be further reduced by choosing λ in the first-order approximation slightly larger than in the regularizer containing the infinite sum of penalizers.
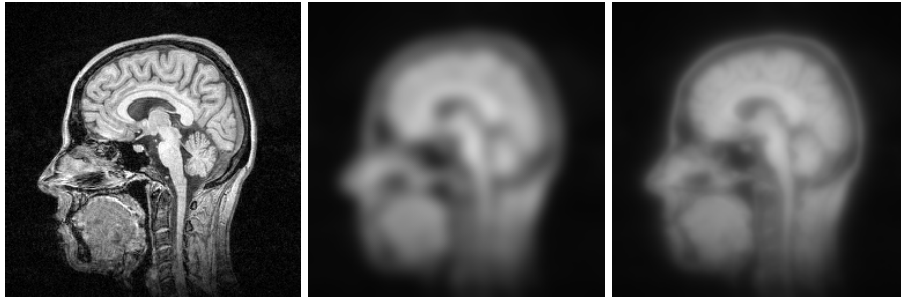

Fig. 1. Comparison of regularization with and without higher order penalizers. **Left:** Original image. **Center:** Smoothing result with the regularizer in (22) (Gaussian smoothing) for $\lambda = 20$. **Right:** Smoothing results with the regularizer in (23) for $\lambda = 20$.

5   Extending the Mumford-Shah Functional

In the previous section, we have shown that the full, piecewise smooth version of the Mumford-Shah functional is a first-order approximation of a Bayesian segmentation approach assuming local Gaussian distributions with a fixed standard deviation. In this section, we will make use of this relation in order to extend the Mumford-Shah functional in a way that it also takes the variance of the data into account. In the Bayesian formulation, this is easy to achieve, as shown in Section 3. Hence, we can take the Bayesian model and express the convolutions by regularization formulations.

With the full Gaussian model, the probability densities

$$p_i(s, \mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_i(\mathbf{x})} \exp\left(\frac{(s - \mu_i(\mathbf{x}))^2}{2\sigma_i(\mathbf{x})^2} \right). \tag{24}$$

depend on two functions $\mu_i(\mathbf{x})$ and $\sigma_i(\mathbf{x})$ given by (12). For $\rho \to \infty$ they are the mean and standard deviation of I in Ω_i, i.e., the minimizers of

$$\int_{\Omega_i} \left(\frac{(\mu_i - I)^2}{2\sigma_i^2} + \frac{1}{2}\log(2\pi\sigma_i^2) + \lambda\left(|\nabla\mu_i|^2 + |\nabla\sigma_i|^2\right) \right) \mathbf{dx} \qquad (25)$$

for $\lambda \to \infty$. This yields a generalized cartoon model. For $\rho \ll \infty$ we make use of the relation between Gaussian convolution and regularization stated in the previous section and obtain $\mu_i(\mathbf{x})$ and $\sigma_i(\mathbf{x})$ as the minimizers of

$$
\begin{aligned}
E(\mu_i, \sigma_i) = {} & \int_{\Omega_i} \left(\frac{(\mu_i - I)^2}{2\sigma_i^2} + \frac{1}{2}\log(2\pi\sigma_i^2) \right) \mathbf{dx} \\
& + \int_{\Omega_i} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{j_1 + j_2 = k} \left(\frac{d^k \mu_i}{dx^{j_1} dy^{j_2}} \right)^2 \mathbf{dx} \\
& + \int_{\Omega_i} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{j_1 + j_2 = k} \left(\frac{d^k \sigma_i}{dx^{j_1} dy^{j_2}} \right)^2 \mathbf{dx}
\end{aligned}
\qquad (26)
$$

and the Bayesian energy can be written as

$$E_B(\mu, \sigma, K) = \sum_i E(\mu_i, \sigma_i) + \nu|K|. \qquad (27)$$

Based on the observation in Section 4, a qualitatively similar approach is obtained by neglecting the penalizer terms with $k > 1$

$$
\begin{aligned}
E_{MS}(\mu, \sigma, K) = {} & \int_{\Omega - K} \left(\frac{(\mu - I)^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2) \right) \mathbf{dx} \\
& + \lambda \int_{\Omega - K} \left(|\nabla\mu|^2 + |\nabla\sigma|^2 \right) \mathbf{dx} + \nu|K|,
\end{aligned}
\qquad (28)
$$

which we may call an extended version of the Mumford-Shah functional. Main advantage of this extension over the original Mumford-Shah functional is that the parameter ν gets invariant with respect to the image contrast. This contrast invariance becomes even more interesting when dealing with vector-valued input images and estimating a separate variance for each vector channel. The influence of each channel on the segmentation then only depends on its discriminative properties and not on the magnitude of the channel values. This allows for the sound integration of different input channels with different contrast and noise levels. For a proof in case of a global Gaussian model we refer to [5]. This proof can be adopted for the local Gaussian model in a straightforward manner.
Another advantage of taking the variance into account is the possibility to distinguish regions that are equal in their mean value, but differ in the variance.

Figure 2 illustrates a result obtained with the extended Mumford-Shah functional. For the experiment we used a level set implementation and expect two regions in the image. Our implementation is based on gradient descent and,

hence, can only ensure a local optimum that need not necessarily be the global one. The initial contour is shown in Figure 2a. The background region of the input image has been generated by a constant function at 127 and Gaussian noise with standard deviation 20. The circular foreground region contains a gradient ranging from 0 to 255. Gaussian noise with standard deviation 70 has been added to this region. The resulting contour and the local mean approximation for $\lambda = 32$ and $\nu = 32$ are shown in Figure 2b and Figure 2c, respectively. For comparison, we depict in Figure 2d the contour found with the same implementation but the standard deviation set fixed, i.e., the original Mumford-Shah functional. For this case, the parameter ν had to be increased to $\nu = 1000$ to obtain reasonable results. Since the two regions have different variances, which can only be exploited by the extended Mumford-Shah functional, the extension finds a more attractive solution than the original version. Larger ν in the original Mumford-Shah functional cannot improve the quality of the result as they lead to an over-smoothed contour not capturing the full circle anymore.

6 Summary

We have provided a statistical interpretation of the Mumford-Shah functional with piecewise smooth regions by showing its relations to Bayesian image segmentation with local region statistics. The link has been established by means of a theorem that relates Gaussian convolution to a regularization problem with an infinite sum of penalizers of arbitrarily high order. Based on this relation, we showed that the Mumford-Shah functional is equivalent to a first-order approximation of a Bayesian approach with Gaussian probability densities estimated with a Gaussian windowing function and the standard deviation set fixed. By means of this relation, we derived an extended version of the Mumford-Shah functional from the Bayesian model which includes the standard deviation as a spatially varying, dynamic function.

References

1. L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via γ-convergence. *Communications on Pure and Applied Mathematics*, XLIII:999–1036, 1990.
2. A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.
3. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
4. T. Brox, B. Rosenhahn, and J. Weickert. Three-dimensional shape knowledge for joint image segmentation and pose estimation. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *LNCS*, pages 109–116. Springer, Aug. 2005.
5. T. Brox, M. Rousson, R. Deriche, and J. Weickert. Colour, texture, and motion in level set based segmentation and tracking. Technical report 147, Dept. of Mathematics, Saarland University, Saarbrücken, Germany, Aug. 2005.

6. V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.

7. T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb. 2001.

8. D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: introducing statistical shape knowledge into the mumford-shah functional. *International Journal of Computer Vision*, 50(3):295–313, Dec. 2002.

9. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

10. D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society B*, 51(2):271–279, 1989.

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

12. E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31:253–258, 1925.

13. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

14. 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.

15. W. Lenz. Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern. *Physikalische Zeitschrift*, 21:613–615, 1920.

16. J.-M. Morel and S. Solimini. Segmentation of images by variational methods: a constructive approach. *Revista Matematica de la Universidad Complutense de Madrid*, 1:169–182, 1988.

17. J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser, Basel, 1994.

18. D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 22–26, San Francisco, CA, June 1985. IEEE Computer Society Press.

19. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–685, 1989.

20. M. Nielsen, L. Florack, and R. Deriche. Regularization and scale space. Technical Report 2352, INRIA Sophia-Antipolis, France, Sept. 1994.

21. M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space and edge detection filters. *Journal of Mathematical Imaging and Vision*, 7:291–307, 1997.

22. N. Paragios and R. Deriche. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1/2):249–268, 2002.

23. R. Potts. Some generalized order-disorder transformation. *Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.

24. 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.

25. A. Yuille and N. M. Grzywacz. A computational theory for the perception of coherent visual motion. *Nature*, 333:71–74, 1988.

26. S.-C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, Sept. 1996.
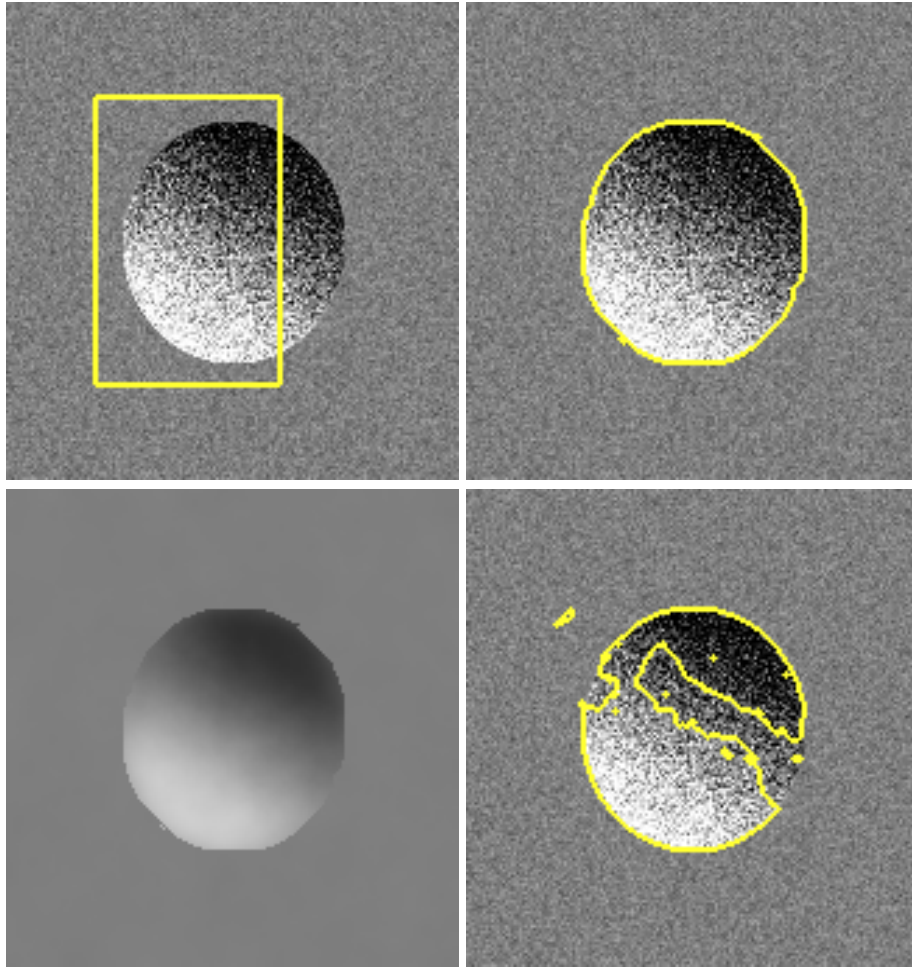
Fig. 2. Example for two regions. **Top left:** (a) Original image of size 162×171 pixels with the initial contour. **Top right:** (b) Contour obtained with the extended Mumford-Shah functional in (28). **Bottom left:** (c) Approximated mean μ. **Bottom right:** (d) Contour obtained with the original Mumford-Shah functional.