Fast communication

Signal segmentation and denoising algorithm based on energy optimisation

Sasan Mahmoodi[a],*, Bayan S. Sharif[b]

[a]*Psychology Department, School of Biology, Henry Wellcome Building, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK*
[b]*School of Electrical, Electronic and Computer Engineering, Merz Court, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*

Abstract

A nonlinear functional is considered in this short communication for time interval segmentation and noise reduction of signals. An efficient algorithm that exploits the signal geometrical properties is proposed to optimise the nonlinear functional for signal smoothing. Discontinuities separating consecutive time intervals of the original signal are initially detected by measuring the curvature and arc length of the smoothed signal. The nonlinear functional is then optimised for each time interval to achieve noise reduction of the original noisy signal. This algorithm exhibits robustness for signals characterised by very low signal to noise ratios.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Nonlinear energy optimisation; Signal segmentation; Signal smoothing

1. Introduction

Energy optimisation for signal and image processing applications has received considerable attention in recent years. Image and signal restoration known as the 'inverse problem' was initially considered by using an energy optimisation approach in [1] and further developed by Rudin et al. [2] to introduce the total variation method. The notion of bounded variation (BV) was later employed in signal and image restoration problems based on nonlinear optimisation (e.g. in [3]). On the other hand, Kass et al. [4] initially introduced the 'snake algorithm' for object segmentation in images. This method was further developed mathematically by Mumford et al. [5–7] for image segmentation and smoothing and was subsequently approximated and implemented by using different approaches e.g. [8–15]. The solutions to the Mumford–Shah functional which are piecewise continuous functions from an appropriate Banach space and contours representing the discontinuities, minimise the functional consisting

*Corresponding author. Tel.: +44 191 222 6246; fax: +44 191 222 5622.

E-mail address: sasan.mahmoodi@ncl.ac.uk (S. Mahmoodi).

of three terms: (1) fidelity term indicating that the smoothed image should be as close as possible to the original image, (2) smoothing term requiring that the smoothed image should be as smooth as possible and (3) contour length removing the unnecessary contours and smoothing the contour representing discontinuities as much as required. Mumford and Shah conjectured that there exists a minimiser for their functional, although the proof for this conjecture is still an open problem. Existance of the solution is however proved for the special case where the smoothing term is not present in the functional and the solutions are reduced to piecewise constant functions surrounded by contours representing discontinuities (see e.g. [16,6,7]). The study of this functional is difficult because it involves two unknown variables of different nature. One of them is a piecewise continuous function on an N-dimensional space and the other representing the contour is associated with a $(N-1)$-dimensional set. Furthermore, the reason that Euler–Lagrange equations cannot be employed in this functional is the lack of differentiability. Therefore different approximation methods have been suggested in the literature (e.g. see [8–11,17]). However, the Mumford–Shah functional is in principle inappropriate for signal processing applications, since the notion of a contour is not defined in a signal processing context. Therefore, a functional based on the Mumford–Shah model and appropriate for signal processing applications is considered in this communication. For implementation purposes, regularisation is initially performed by smoothing the original signal and a newmethod based on geometrical properties of the smoothed signal, is further employed for segmentation and denoising. In comparison with signal restoration approaches used in the inverse problem, three major differences between our method and a restoration method such as BV regularisation in [3] can be observed. (1) Energy optimisation in [3] leads to restoration and therefore segmentation is implicitly achieved by using Lagrange multipliers employed in the active set strategy, while in our method, segmentation and denoising are explicitly employed in the formulation of the functional. (2) Original and denoised signals are considered piecewise constant in [3], whereas in this communication, signals are considered piecewise continuous whose samples are either acquired (for original signal) or computed (for smoothed signal). (3) Spline algorithm and tube method based on taut-string algorithm minimising signal length within a tube of radius α (e.g. see [18]), are employed for implementation in [3], while we consider geometrical properties of the smoothed signal in our simulation. The proposed method in this communication is computationally efficient and exhibits robustness even with SNRs less than 1. In Section 2, the mathematical background is discussed, whilst Section 3 deals with implementation issues. Results are presented in Section 4. Finally conclusions are drawn in Section 5.

2. Mathematical background

A noisy signal, $g(t)$ contaminated with a mean order stationary Gaussian noise [19], can be approximated by a piecewise continuous function $f(t)$ containing class C^2 functions $f_i(t)$ over time intervals (t_{i-1}, t_i) so that $f(t)$ is as close as possible to the original signal $g(t)$ and is as smooth as possible over each time interval. However, the smoothing process is only applied over intervals where the signal fluctuations are considered comparable to noise fluctuations. We therefore consider the following nonlinear functional whose optimisation leads to the desired smooth functions $f_i(t)$ over the time intervals (t_{i-1}, t_i).

$$E(f, S) = \frac{1}{2} \sum_i \int_0^t \left[(f_i(t) - g(t))^2 + \mu \left(\frac{\mathrm{d}f_i}{\mathrm{d}t} \right)^2 \right] \\ \times S_i(t)\, \mathrm{d}t, \tag{1}$$

where $E(f, S)$ is the energy term to be optimised, μ is a constant coefficient, and $S_i(t)$ is a rectangular time-domain window function representing the time interval:

$$S_i(t) = \begin{cases} 1 & : \quad t_{i-1} < t < t_i, \\ 0 & : \quad \text{otherwise.} \end{cases}$$

Let us initially consider the optimisation of functional (1) in the interval defined by $S_i(t)$. Further, let $S_i(t)$ be fixed and then vary $f_i(t)$.

Functional (1) is therefore rewritten as

$$E_i(f_i, S_i) = \int_{t_{i-1}}^{t_i} \left[(f_i(t) - g(t))^2 + \mu \left(\frac{\mathrm{d}f_i}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t. \tag{2}$$

As can be seen from (2), with fixed time intervals, the integrand of the functional is convex, hence there exists a minimiser for the functional with a set of fixed intervals. In what follows we calculate this minimiser. Let δf_i represent a class C^2 function, and then we calculate the variations of $E_i(f_i, S_i)$ by varying $f_i(t)$ with δf_i:

$$\delta E_i = E_i(f_i + m\delta f_i, S_i) - E_i(f_i, S_i)$$
$$= \frac{1}{2} \left(\int_{t_{i-1}}^{t_i} \left[(f_i + m\delta f_i - g)^2 \right. \right.$$
$$\left. + \mu \left(\frac{\mathrm{d}(f_i + m\delta f_i)}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t$$
$$\left. - \int_{t_{i-1}}^{t_i} \left[(f_i - g)^2 + \mu \left(\frac{\mathrm{d}f_i}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t \right).$$

By simple mathematical manipulations, it is concluded that:

$$\frac{\delta E_i}{\delta f_i} = \lim_{m \to 0} \frac{\delta E_i}{m}$$
$$= \int_{t_{i-1}}^{t_i} (f_i - g)\delta f_i \, \mathrm{d}t + \mu \int_{t_{i-1}}^{t_i} \left(\frac{\mathrm{d}f_i}{\mathrm{d}t} \right) \left(\frac{\mathrm{d}\delta f_i}{\mathrm{d}t} \right) \mathrm{d}t. \tag{3}$$

Integrating by part and treating this problem as free boundary condition [20,21], Eq. (3) is written as

$$\frac{\delta E_i}{\delta f_i} = \frac{\mathrm{d}f_i}{\mathrm{d}t} \delta f_i|_{t_i} - \frac{\mathrm{d}f_i}{\mathrm{d}t} \delta f_i|_{t_{i-1}}$$
$$+ \int_{t_{i-1}}^{t_i} \left[(f_i - g) - \mu \left(\frac{\mathrm{d}^2 f_i}{\mathrm{d}t^2} \right) \right] \delta f_i \, \mathrm{d}t = 0. \tag{4}$$

Since $\delta f_i \neq 0$ over the time interval (t_{i-1}, t_i), and in order to satisfy Eq. (4), we should have:

$$\mu \frac{\mathrm{d}^2 f_i(t)}{\mathrm{d}t^2} = f_i(t) - g(t), \quad t_{i-1} < t < t_i \tag{5}$$

with Neumann boundary conditions:

$$\frac{\mathrm{d}f_i}{\mathrm{d}t}|_{t_i} = 0, \quad \frac{\mathrm{d}f_i}{\mathrm{d}t}|_{t_{i-1}} = 0. \tag{6}$$

The solution to differential equation (5) with boundary conditions (6) is a minimiser of energy functional (1) over time interval (t_{i-1}, t_i). Let us now find the time interval (t_{i-1}, t_i) minimising energy functional (1). This can be specified by varying the time interval and considering the variations of f_i. Let us therefore vary t_i and calculate f_i's variations in a neighborhood of t_i. Fig. 1 shows how f_i and f_{i+1} change if t_i moves to t_i' where $t_i' = t_i \pm \delta t_i$. If we assume that t_i corresponds to a discontinuity and then consider the behavior of f by early and late movement of t_i by δt_i, i.e. variations of f are considered in a neighborhood (say I) of t_i'. If we further define f^+ and f^- corresponding to $t_i' = t_i + \delta t_i$ and $t_i' = t_i - \delta t_i$, respectively, as can be seen in Fig. 1.

$$f^+ = \begin{cases} f_i^+ & : \quad t \in I \quad \& \quad t < t_i + \delta t_i, \\ f_{i+1}^+ & : \quad t \in I \quad \& \quad t > t_i + \delta t_i, \\ \text{unchanged} & : \quad t \notin I, \end{cases}$$

$$f^- = \begin{cases} f_i^- & : \quad t \in I \quad \& \quad t < t_i - \delta t_i, \\ f_{i+1}^- & : \quad t \in I \quad \& \quad t > t_i - \delta t_i, \\ \text{unchanged} & : \quad t \notin I. \end{cases}$$

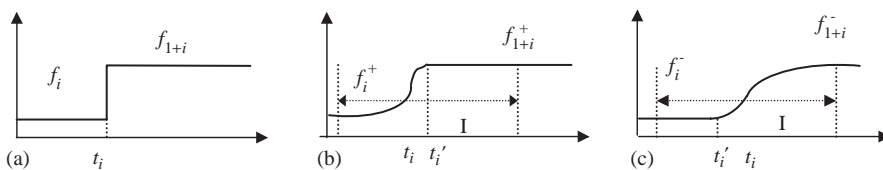And if S^+ and S^- are also defined in a similar manner, then the variation in the energy functional

Fig. 1. Variations of f by varying t_i.

is given as

$$\delta E = E(f^+, t_i + \delta t_i) - E(f^-, t_i - \delta t_i)$$
$$= \left(\int_0^\infty \left[(f^+ - g)^2 + \mu \left(\frac{\mathrm{d}f^+}{\mathrm{d}t} \right)^2 \right] S^+(t)\,\mathrm{d}t \right.$$
$$\left. - \int_0^\infty \left[(f^- - g)^2 + \mu \left(\frac{\mathrm{d}f^-}{\mathrm{d}t} \right)^2 \right] S^-(t)\,\mathrm{d}t \right),$$

$$\delta E = \int_I \left[(f^+ - g)^2 + \mu \left(\frac{\mathrm{d}f^+}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t$$
$$- \int_I \left[(f^- - g)^2 + \mu \left(\frac{\mathrm{d}f^-}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t. \tag{7}$$

By $\delta t_i \to 0$, Eq. (7) can be approximated as

$$\frac{\delta E}{\delta t_i} = \left[(f^+ - g)^2 + \mu \left(\frac{\mathrm{d}f^+}{\mathrm{d}t} \right)^2 \right]$$
$$- \left[(f^- - g)^2 + \mu \left(\frac{\mathrm{d}f^-}{\mathrm{d}t} \right)^2 \right]. \tag{8}$$

Minimisation of energy functional (1) with respect to t_i requires that Eq. (8) is set equal to zero, i.e.:

$$\left[(f^+ - g)^2 + \mu \left(\frac{\mathrm{d}f^+}{\mathrm{d}t} \right)^2 \right]$$
$$- \left[(f^- - g)^2 + \mu \left(\frac{\mathrm{d}f^-}{\mathrm{d}t} \right)^2 \right] = 0. \tag{9}$$

In a neighborhood of the discontinuity, t_i satisfying Eq. (9), corresponds to the discontinuity point. Geometrical interpretation of Eq. (9) is that t_i is the intersection of the two functions $(f^+ - g)^2 + \mu(\mathrm{d}f^+/\mathrm{d}t)^2$ and $(f^- - g)^2 + \mu(\mathrm{d}f^-/\mathrm{d}t)^2$ in the neighborhood of t_i. To clarify Eq. (9), let us consider variations of functional (1) with respect to t_i in two different cases: (1) when t_i corresponds to a discontinuity such as the case shown in Fig. 1 and 2 when t_i is far from any discontinuity. If t_i is varied to $t_i + \delta t_i$ and $t_i - \delta t_i$, corresponding variations of functional (1) can be calculated

respectively as

$$\delta E^+ = \int_I \left[(f^+ - g)^2 + \mu \left(\frac{\mathrm{d}f^+}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t$$
$$- \int_I \left[(f - g)^2 + \mu \left(\frac{\mathrm{d}f}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t \tag{10}$$

and

$$\delta E^- = \int_I \left[(f - g)^2 + \mu \left(\frac{\mathrm{d}f}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t$$
$$- \int_I \left[(f^- - g)^2 + \mu \left(\frac{\mathrm{d}f^-}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t. \tag{11}$$

if t_i is a discontinuity point then it can be verified that $\delta E^+ > 0$ and $\delta E^- < 0$. This implies that between $t_i - \delta t_i$ to $t_i + \delta t_i$, there exist a point where E is minimised. By δt_i approaching zero, the point minimising E approaches a discontinuity. However, if t_i is far from any discontinuity, then it is obvious that $f^+ = f^- = f$. Hence $\delta E^+ = \delta E^- = 0$, thus implying that E remains constant and hence t_i is not a minimiser of the functional. It should be noted that although points far from any discontinuity satisfy Eq. (9), they are not minimisers of functional (1) and hence are not considered as solutions to our minimisation problem.

At this stage, it is interesting to examine the behavior of functional (1) and Eq. (5) by varying coefficient μ. If $\mu \to 0$, then the first term in functional (1) becomes dominant implying that f_i follows $g(t)$ and its fluctuations. However, if $\mu \to \infty$, the first term in functional (1) can be ignored and according to Eq. (5), f_i is heavily smoothed and can therefore be approximated as a line. Eqs. (5), (6) and (9) are the optimised solutions to the optimisation problem of functional (1). However, implementation of a nonlinear equation such as Eq. (9) along with a differential equation such as (5) especially in the presence of noise could be highly complicated and computationally expensive. An efficient algorithm is therefore proposed in this short communication to find the optimal solutions to functional (1) in

the presence of noise. We initially consider the linear functional:

$$E(f, S) = \frac{1}{2} \int_0^t \left[(f(t) - g(t))^2 + \mu \left(\frac{\mathrm{d}f}{\mathrm{d}t} \right)^2 \right] \mathrm{d}t. \quad (12)$$

Using Euler–Lagrange equations [20,21], the optimal solution for the entire time interval is given by

$$\mu \frac{\mathrm{d}^2 f(t)}{\mathrm{d}t^2} = f(t) - g(t). \quad (13)$$

The solution to differential equation (13) is a function $f(t)$ that smooths fluctuations as well as discontinuities of the original noisy signal $g(t)$. The objective in this communication is to detect discontinuities related to the original noisy signal using the geometrical properties of the smoothed function $f(t)$. There are two geometrical features related to the smoothed parts of $f(t)$ corresponding to discontinuities:

(1) Arc length variations that are maximised with respect to t i.e. [22,23],

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\sqrt{\mathrm{d}t^2 + \mathrm{d}y^2}}{\mathrm{d}t} = \sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}t} \right)^2}$$

or

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \sqrt{1 + (f')^2} \quad (14)$$

is maximised.

(2) Value of the curvature of the smoothed function defined as [22,23]:

$$\mathrm{Curv}(f) = \frac{f''}{\sqrt{(1 + (f')^2)^3}} \quad (15)$$

varies from a local maximum/minimum to another local minimum/maximum, crossing zero.

The second property might be true for the intervals of $f(t)$ corresponding to noise as well. However, the intervals that correspond to this change of curvature value for discontinuities are significantly greater than those for noise. This is due to the fact that the local mean of the original noisy signal changes in discontinuities, whereas it remains unchanged or slightly changed in intervals characterised by noise only. Based on the mathematical background discussed in this section, an

algorithm is proposed in Section 3 to detect discontinuities. Eq. (5) with boundary conditions (6) is then applied to each separate interval to obtain the smoothed function $f(t)$ corresponding to the chosen μ. Finally, an adaptive algorithm is proposed to choose the most optimal μ for a given noise level.

3. Implementation

Eq. (13) is initially solved to obtain $f(t)$ which smooths noise as well as discontinuities for an initial μ. The intervals of zero curvature are then detected and only one point corresponding to maximum arc length for each interval of zero curvature is chosen as a discontinuity point. Once discontinuity points along the signal are detected (segmentation step), Eq. (5) is solved for each interval using the boundary condition (6) to obtain the smoothed signal (noise reduction step). To obtain the most optimal μ for a given noise, $E = (1/t) \int_0^t [(f(t) - g(t))^2] \, \mathrm{d}t$ is calculated for different values of μ. This process continues by changing μ and hence $f(t)$ until E approaches the noise variance. In this scheme, with an initial μ, if E is less than noise variance, μ increases, otherwise it decreases.

4. Results

The algorithm proposed in Sections 2 and 3 is applied for the purpose of time interval segmentation and signal smoothing in each interval. Fig. 2(top row) shows an original noiseless signal. Gaussian noise is added to this signal to obtain noisy signals with SNR = 2.1, 1.0 and 0.42 as shown in Fig. 2(middle row). The smoothed signals using the proposed algorithm are obtained with $\mu = 100$ as depicted in Fig. 2(bottom row).

The adaptive algorithm to find the most optimised μ, proposed in Section 3, is applied to some noisy signals contaminated with Gaussian noise. The results are depicted in Fig. 3. Gaussian noise is added to three different original noiseless signals, shown in Fig. 3, to obtain noisy signals with SNR = 0.9. The smoothed signals are
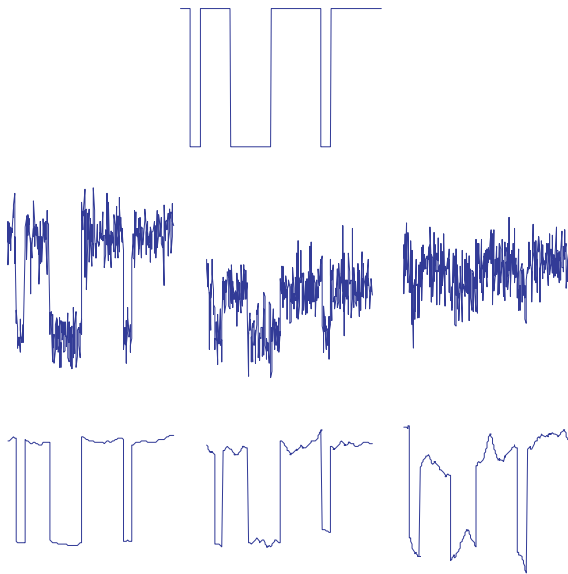
Fig. 2. Original noiseless signal (top row), noisy signals contaminated with Gaussian noise with SNR = 2.1, 1.0 and 0.42 (middle row) and their corresponding smoothed signals with $\mu = 100$ (bottom row).
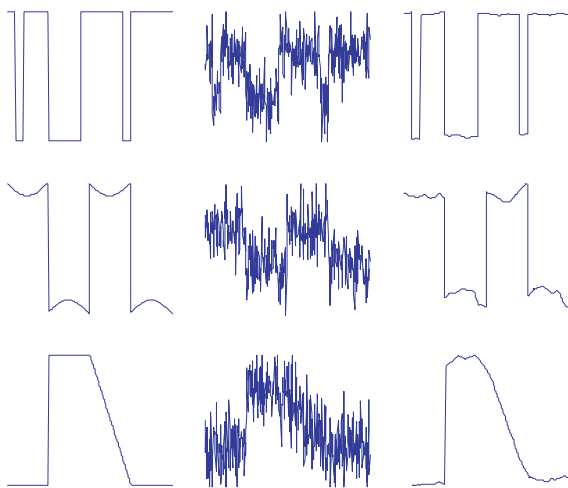

Fig. 3. Original noiseless signals (left column), noisy signals contaminated with Gaussian noise with SNR = 0.9 (middle column) and smoothed signals (right column) with optimised $\mu = 225$ (top), $\mu = 280$ (middle) and $\mu = 265$ (bottom).

obtained by applying the adaptive algorithm as depicted in the same figure, resulting in optimised values for μ (see Fig. 3).

5. Conclusion

An efficient method is presented in this short communication for signal segmentation and noise reduction based on an energy optimisation method. Geometrical properties such as curvature and maximum arc length of the smoothed signal are employed to detect discontinuity points. Once discontinuities are detected, signal is smoothed for each time interval by solving the appropriate differential equations. This method can be generalised to 2D images by employing geometrical features of surfaces to detect discontinuities as object edges. However, contour length minimisation should also be taken into consideration to satisfy the Mumford–Shah functional requirements.

References

[1] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Winston & Sons, Washington, DC, 1977.

[2] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.

[3] W. Hinterberger, M. Hintermuller, K. Kunisch, M. von Oehsen, O. Scherzer, Tube methods for BV regularization, J. Math. Imaging Vision 19 (2003) 219–235.

[4] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, J. Comput. Vision 1 (1987) 321–331.

[5] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations, Springer-Verlage, New York, 2002.

[6] J.M. Morel, S. Solimini, Variational Methods in Image Segmentation, Birkhauser, Boston, 1995.

[7] D. Mumford, J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Commun. Pure Appl. Math. 42 (4) (1989) 577–688.

[8] L. Ambrosio, V.M. Tortorelli, Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence, Commun. Pure Appl. Math. XLIII (1990) 999–1036.

[9] A. Braides, Approximation of Free-Discontinuity Problems, Lecture Notes in Mathematics, vol. 1694, 1998.

[10] A. Braides, G. Dal Maso, Non-local approximation of the Mumford–Shah functional, Calc. Var. Partial Differential Equations 5 (4) (1997) 293–322.

[11] A. Chambolle, Image segmentation by variational methods: Mumford and Shah Functional and the discrete approximation, SIAM J. Appl. Math. 55 (3) (1995) 827–863.

[12] T.F. Chan, L.A. Vese, Active contours without edges, IEEE Trans. Image Process. 10 (2) (2001) 266–277.

[13] M. Hintermuller, W. Ring, An inexact-CG type active contour approach for the minimization of the Mumford–Shah functional, J. Math. Imaging Vision 20 (2004) 19–42.

[14] T.J. Richardson, Scale independent piecewise smooth segmentation of images via variational methods, Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, Mass Institute of Technology, 1989.

[15] A. Tsai, A. Yezzi, A.S. Willsky, Curve evolution implementation of the Mumford–Shah functional for image segmentation, denoising interpolation and magnification, IEEE Trans. Image Process. 10 (8) (2001) 1169–1186.

[16] G. Congedo, I. Tamanini, On the existance of solutions to a problem in multidimensional segmentation, Ann. Inst. Henri Poincare 8 (2) (1991) 175–195.

[17] M. Gobbino, Finite difference approximation of the Mumford–Shah functional, Commun. Pure Appl. Math. 51 (2) (1998) 197–228.

[18] E. Mammen, S. van de Geer, Locally adaptive regression splines, Ann. Statist. 25 (1997) 387–413.

[19] J.L. Melsa, A.P. Sage, An Introduction to Probability and Stochastic Processes, Prentice-Hall Inc., Englewood Cliffs, NJ, 1973.

[20] I.M. Gelfand, S.V. Fomin, Calculus of Variations, Prentice-Hall Inc., Englewood Cliffs, NJ, 1963.

[21] U.B. Manderscheid, Introduction to the Calculus of Variations, Chapman & Hall, London, 1991.

[22] M.P. Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall Inc., Englewood Cliffs, NJ, 1976.

[23] M.M. Lipschutz, Theory and Problems of Differential Geometry, Schaum's Outline Series, McGraw-Hill, New York, 1969.