

Image Segmentation by Piecewise Constant Mumford-Shah Model without Estimating the Constants

Xue-Cheng Tai Chang-Hui Yao

The paper is dedicated to the 70th birthday to Professor Lin Qun.

Abstract

In this work, we try to use the so-called Piecewise Constant Level Set Method (PCLSM) for the Mumford-Shah segmentation model. For image segmentation, the Mumford-Shah model needs to find the regions and the constant values inside the regions for the segmentation. In order to use PCLSM for this purpose, we need to solve a minimization problem using the level set function and the constant values as minimization variables. In this work, we test on a model such that we only need to minimize with respect to the level set function, i.e., we do not need to minimize with respect to the constant values. Gradient descent method and Newton method are used to solve the Euler-Lagrange equation for the minimization problem. Numerical experiments are given to show the efficiency and advantages of the new model and algorithms.

Keywords

PCLSM, Image Segmentation, Mumford-Shah model.

I. INTRODUCTION

Level set methods, originally introduced by Osher and Sethian [13], have been developed to be one of the most successful tools for the computation of evolving geometries, which appear in many practical applications. They use zero level set of some functions to trace interfaces separated a domain Ω into subdomains. For a recent survey on the level set methods see [15], [10], [14].

In [7], [6], [9] some variants of the level set methods of [13], so-called "piecewise constant level set method (PCLSM)", were proposed. The method can be used for different applications. In [12], [17], applications to inverse problems involving elliptic and reservoir equations are shown. In [7], [6], [9], [18], the ideas have been used for image segmentation problem.

Image segmentation is one of the foundational tasks of computer vision. Its goal is to partition a given image into regions which contain distinct objects. One of the most common forms of segmentation are based on assumption that distinct objects in an image have different approximately constant (or slowly varying) colors (or roughnesses in the case of monochrome imagery). In this paper, we use the Mumford-Shah model [11] for image segmentation with PCLSM. In [2], [8], [15], the Osher-Sethian level set idea was used to solve the Mumford-Shah model. No matter whether we use the Osher-Sethian method or PCLSM, we need to minimize with respect to the level set functions and the constant values. In this work, we shall propose a model such that only the level set function is the minimization variable. The advantage of this model is that we just have one variable to minimize. The resulted algorithms are more stable and have a slightly faster convergence rate. This model eliminates the question about when and how often to update the constant values.

The organization of our paper is as follows. In §II, firstly, we begin with a quick examination of what constitutes a piecewise constant level set method and the classical Mumford-Shah model. The

CIPR, Department of Mathematics, University of Bergen, Norway. Email: Tai@mi.uib.no. Web page <http://www.mi.uib.no/~tai>.

CIPR, Department of Mathematics, University of Bergen, Norway. E-mail: changhui.yao@cipr.uib.no

origin of the new idea is developed as well. In §III, gradient descent method and Newton method of [18] are used for the new model. At the end of this section, some remarks are given about how to obtain an initial guess for Newton method. In §IV, numerical experiments are shown to demonstrate the efficiency of the new model and the algorithms. For every example, the initial values of level set function and the values of some parameters are shown. The differences between one-variable model and two-variable model are compared in this section, too. In §V, we make some conclusions on this new model.

II. ONE-VARIABLE MUMFORD-SHAH MODEL WITH PCLSM

Firstly, We shall recall PCLSM of [7]. The essential idea of PCLSM of [7] is to use a piecewise constant level set function to identify the subdomains. Assume that we need to partition the domain Ω into subdomains Ω_i , $i = 1, 2, \dots, n$, and the number of subdomains is a priori known. In order to identify the subdomains, we try to identify a piecewise constant level set function ϕ such that

$$\phi = i \quad \text{in } \Omega_i, \quad i = 1, 2, \dots, n. \quad (1)$$

Thus, for any given partition $\{\Omega_i\}_{i=1}^n$ of the domain Ω , it corresponds to a unique PCLS function ϕ which takes the values $1, 2, \dots, n$. Associated with such a level set function ϕ , the characteristic functions of the subdomains are given as

$$\psi_i = \frac{1}{\alpha_i} \prod_{j=1, j \neq i}^n (\phi - j), \quad \alpha_i = \prod_{k=1, k \neq i}^n (i - k). \quad (2)$$

If ϕ is given as in (1), then we have $\psi_i(x) = 1$ for $x \in \Omega_i$, and $\psi_i(x) = 0$ elsewhere. We can use the characteristic functions to extract geometrical information for the subdomains and the interfaces between the subdomains. For example,

$$\text{Length}(\partial\Omega_i) = \int_{\Omega} |\nabla \psi_i| dx, \quad \text{Area}(\Omega_i) = \int_{\Omega} \psi_i dx. \quad (3)$$

Define

$$K(\phi) = (\phi - 1)(\phi - 2) \cdots (\phi - n) = \prod_{i=1}^n (\phi - i). \quad (4)$$

At every point in Ω , the level set function ϕ satisfies

$$K(\phi) = 0. \quad (5)$$

This level set idea has been used for Mumford-Shah image segmentation in [7]. For a given digital image $u_0 : \Omega \mapsto R$, which may be corrupted by noise and blurred, the piecewise constant Mumford-Shah segmentation model is to find subdomains Ω_i and constant values c_i by minimizing:

$$\sum_i \int_{\Omega_i} |c_i - u_0|^2 dx + \beta \text{Length}(\Gamma). \quad (6)$$

The curve Γ is the one that separates the domain Ω into subdomains Ω_i such that $\Omega = \bigcup_i \Omega_i \cup \Gamma$ and $\Gamma = \bigcup_i \partial\Omega_i$. In [2], the traditional level set idea of [13] was used to represent the curve Γ and to solve the problem (6). In [7], [16], [18], the PCLSM were used for identifying the regions and the constant values. The minimization problem there is:

$$\min_{\substack{\mathbf{c}, \phi \\ K(\phi)=0}} \left\{ F(\mathbf{c}, \phi) = \frac{1}{2} \int_{\Omega} |u(\mathbf{c}, \phi) - u_0|^2 dx + \beta \int_{\Omega} |\nabla \phi| dx \right\},$$

where

$$\mathbf{c} = (c_1, c_2, \dots, c_n), \quad u(\mathbf{c}, \phi) = \sum_{i=1}^n c_i \psi_i(\phi).$$

From our simulations, we have observed that two set of different values of (\mathbf{c}, ϕ) can produce the same u and thus may confuse the algorithm and slow down the convergence. The remedy we have used is to update the \mathbf{c} values less often than ϕ and do not update the \mathbf{c} values too earlier in the iterations. In this work, we shall propose another technique to tackle this problem, i.e. we shall eliminate the \mathbf{c} as a minimization variable. Instead, we express it as a function of the level set function ϕ . More precisely, we shall solve the following minimization problem to segment an image u_0 :

$$\min_{\substack{\phi \\ K(\phi)=0}} \left\{ F(\phi) = \frac{1}{2} \int_{\Omega} |u(\phi) - u_0|^2 dx + \beta \int_{\Omega} |\nabla \phi| dx \right\}, \quad (7)$$

where

$$u(\phi) = \sum_{i=1}^n c_i(\phi) \psi_i(\phi), \quad \text{and} \quad c_i(\phi) = \frac{\int_{\Omega} \psi_i(\phi) u_0 dx}{\int_{\Omega} \psi_i^2(\phi) dx}, \quad i = 1, 2, \dots, n. \quad (8)$$

At convergence, we have that

$$c_i(\phi) = \frac{\int_{\Omega_i} u_0 dx}{\text{Area}(\Omega_i)}, \quad i = 1, 2, \dots, n,$$

i.e. c_i is the mean value of u_0 in Ω_i . This value for c_i is consistent with the values obtained by the model of [2] and [7].

III. ALGORITHMS

The model (7)-(8) is rather similar to the models of [18]. Then we shall use the same algorithms used in [18] to solve (7)-(8). The augmented Lagrangian functional for the minimization problem (7)-(8) is

$$L(\phi, \lambda) = F(\phi) + \int_{\Omega} \lambda K(\phi) dx + \frac{r}{2} \int_{\Omega} |K(\phi)|^2 dx, \quad (9)$$

where $\lambda \in L^2(\Omega)$ is the multiplier and $r > 0$ is a penalty parameter. The first algorithm we shall use to solve the minimization problem is the following classical gradient descent method. It shall be emphasized that it is also possible to use some other types of gradient methods, for example, the Barzilai and Borwein (BB) gradient methods in [1], alternate step (AS) gradient methods in [4], [3], and monotone gradient methods in [5].

Algorithm 1: (Gradient descent method) Choose initial values for ϕ^0 and λ^0 . For $k = 1, 2, \dots$, do:

1. Use (8) to update $u = \sum_{i=1}^n c_i(\phi^{k-1}) \psi_i(\phi^{k-1})$.

2. Find ϕ^k from

$$L(\phi^k, \lambda^{k-1}) = \min_{\phi} L(\phi, \lambda^{k-1}). \quad (10)$$

3. Update the Lagrange-multiplier by

$$\lambda^k = \lambda^{k-1} + rK(\phi^k). \quad (11)$$

In order to get the optimal solution of (10), we can construct an artificial time variable and solve the following equation to steady state

$$\frac{\partial \phi}{\partial t} + \frac{\partial L}{\partial \phi} = 0.$$

It is easy to see that

$$\frac{\partial L}{\partial \phi} = (u - u_0)u'(\phi) - \beta \nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda K'(\phi) + rK(\phi)K'(\phi)$$

and the curvature term $\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right)$ is approximated by

$$\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \approx \nabla \cdot \left(\frac{\nabla \phi}{\sqrt{|\nabla \phi|^2 + \varepsilon}} \right).$$

In our numerical examples, we take $\varepsilon = 10^{-8}$. Normally, we choose a fixed step size Δt and do a fixed number of the following iterations to solve (10) approximately

$$\phi^{new} = \phi^{old} - \Delta t \frac{\partial L}{\partial \phi}(\phi^{old}, \lambda^{k-1}). \quad (12)$$

In the simulations, the above iteration is terminated when the inner iteration number is more than n_λ or

$$\left\| \frac{\partial L}{\partial \phi}(\phi^{new}, \lambda^{k-1}) \right\|_{L^2} \leq \frac{1}{10} \left\| \frac{\partial L}{\partial \phi}(\phi^{k-1}, \lambda^{k-1}) \right\|_{L^2}.$$

We shall compare this algorithm with a corresponding algorithm of [7]. The λ for the algorithm for [7] is updated using the same criteria.

Theories and numerical tests on image segmentation in [7], [6], [9], [18] and on inverse problem in [12], [17] have shown that gradient descent method is a stable but slow convergent algorithm with PCLSM. Therefore, different approaches have been used to accelerate the convergence. Newton method has been used in [18] to get faster convergence. We shall follow the same idea also for the minimization problem here.

Similarly as in [18], we define

$$R(\phi, \lambda) = \frac{1}{2} \int_{\Omega} |u(\phi) - u_0|^2 dx + \int_{\Omega} \lambda K(\phi) dx. \quad (13)$$

and

$$L_0(\phi, \lambda) = F(\phi) + \int_{\Omega} \lambda K(\phi) dx. \quad (14)$$

It is easy to see that L_0 is equal to L if we take $r = 0$ in (9). In addition, the functional L_0 given in (14) reduces to R if we take $\beta = 0$. Based on a small value of β , we can replace Hessian matrix

$$\begin{pmatrix} \frac{\partial^2 L}{\partial \phi^2} & \frac{\partial^2 L}{\partial \phi \partial \lambda} \\ \frac{\partial^2 L}{\partial \phi \partial \lambda} & 0 \end{pmatrix} \quad (15)$$

by a simplified Hessian matrix

$$\begin{pmatrix} \frac{\partial^2 R}{\partial \phi^2} & \frac{\partial^2 R}{\partial \phi \partial \lambda} \\ \frac{\partial^2 R}{\partial \phi \partial \lambda} & 0 \end{pmatrix}. \quad (16)$$

By the chain rule, we have

$$\begin{aligned}\frac{\partial^2 R}{\partial \phi^2} &= \left(u'(\phi)\right)^2 + (u - u_0)u''(\phi) + \lambda K''(\phi), \\ \frac{\partial^2 R}{\partial \phi \partial \lambda} &= \frac{\partial^2 R}{\partial \lambda \partial \phi} = K'(\phi), \\ u' &= \sum_{i=1}^n c'(\phi)\psi_i(\phi) + c(\phi)\psi_i'(\phi), \\ u''(\phi) &= \sum_{i=1}^n (c''(\phi)\psi_i + 2c'(\phi)\psi_i'(\phi) + c_i\psi_i''(\phi)).\end{aligned}$$

Then Newton method is given in the following:

Algorithm 2: (Newton method) Choose initial values ϕ^0, λ^0 . For $k = 1, 2, \dots$, do:

1. Use (8) to update $u = \sum_{i=1}^n c_i(\phi^{k-1})\psi_i(\phi^{k-1})$.
2. Update ϕ^k, λ^k from

$$\begin{pmatrix} \phi^k \\ \lambda^k \end{pmatrix} = \begin{pmatrix} \phi^{k-1} \\ \lambda^{k-1} \end{pmatrix} - \begin{pmatrix} \frac{\partial^2 R}{\partial \phi^2} & \frac{\partial^2 R}{\partial \phi \partial \lambda} \\ \frac{\partial^2 R}{\partial \phi \partial \lambda} & 0 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial L_0}{\partial \phi} \\ \frac{\partial L_0}{\partial \lambda} \end{pmatrix}. \quad (17)$$

For the algebraic system (17), we can solve a 2×2 system at each grid point, which means that the cost for Algorithm 2 is nearly the same as for Algorithm 1 at each iteration. Some remarks about the above algorithm are given in the following.

Remark 1: We have to find a relative good initial values for Newton method, which guarantees this algorithm convergent. Different ways can be taken to get such initial values. In our examples, we can take the data from Algorithm 1 by updating fixed iterations, which will be carried out under condition of large noise added image u_0 in (7). But for many of the test examples, the simple scaling procedure outlined in §IV is good enough to have Algorithm 2 converge.

Remark 2: In general, we will take a small value of β . If the interfaces are oscillatory, we increase the value of β and decrease Δt to keep the algorithms stable.

IV. NUMERICAL EXAMPLES

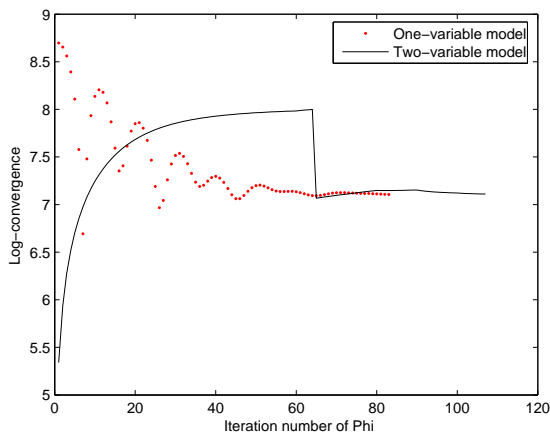
In this section, we will demonstrate the efficiency of the new model (7) and compare it with the algorithms given in [7]. We shall concentrate on two-phase segmentation only.

As in [18], we can use the following scaling procedure to get initial values for ϕ . First, we set the phase number $n = 2$. Then we scale u_0 to a function between 1 and 2 as in the following and take this as the initial values for ϕ , i.e.

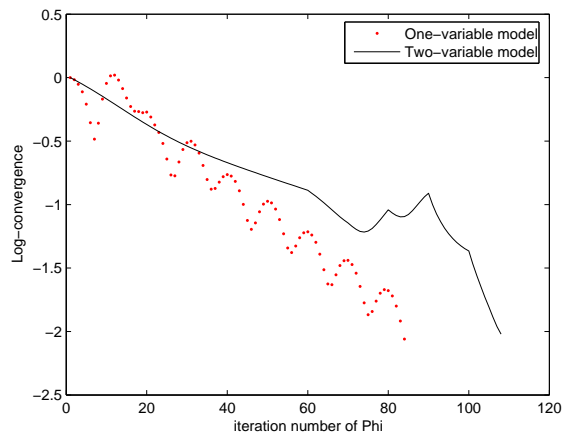
$$\phi^0(x) = 1 + \frac{u_0(x) - \min_{x \in \Omega} u_0}{\max_{x \in \Omega} u_0 - \min_{x \in \Omega} u_0}. \quad (18)$$

For some case, we do a fixed number of iterations of Algorithm 1 using the above ϕ^0 as the starting value and then use the obtained image as the initial values for Algorithm 2. We consider only two-dimensional grey scale images. To complicate the segmentation process we typically expose the original image with Gaussian distributed noise and use the polluted image as observation data u_0 . We report the signal-to-noise-ratio calculated by $\text{SNR} = \frac{V_1}{V_2}$, where V_1 and V_2 are the variance of the data and the variance of the noise, respectively. In the following, we shall use one-variable model to (7)-(8) and use two-variable model to refer to Algorithm 2 of [7].

We begin with an image of an old newspaper where only two phases are needed. One phase represents the characters and the other phase represents the background of the newspaper. The Log_{10} convergence for $\|F(\phi^k)\|_{L^2}$ and $\frac{\|K(\phi^k)\|_{L^2}}{\|K(\phi^0)\|_{L^2}}$ with gradient descent method for the one-variable model and two-variable model are plotted in Fig.1. We use the same values for β and Δt for both models. For this plot for the two-variable model, the \mathbf{c} values are updated as stated in the remarks of [7]. The plot shows that the one-variable model converges slightly faster. The cost functional converges to the same value for the two models. In Fig.2, we try to compare gradient descent method with Newton method. Fig.2(c) shows the level set function ϕ by Newton method at 10 iterations. This segmented is the same as that of gradient descent method at 108 iterations in Fig.2(d). In the tests for Fig.1 and Fig.2, we have used $\beta = 0.1, r = 10^7, \Delta t = 10^{-8}, n_\lambda = 10$. Fig.2(b) shows a part of convergent ϕ clearly.



(a) The Log_{10} -convergence of $\|F(\phi^k)\|_{L^2}$ for one-variable model and of $\|F(c, \phi^k)\|_{L^2}$ for two-variable model.



(b) The Log_{10} -convergence for $\frac{\|K(\phi^k)\|_{L^2}}{\|K(\phi^0)\|_{L^2}}$

Fig. 1. A comparison of the Log_{10} -convergence the one-variable model with the two-variable model by using gradient descent method.



Fig. 2. A comparison Newton method with gradient descent method for the one-variable model. (a) An old real newspaper. (b) A small partition of the convergent ϕ , it is a piecewise constant function $\phi = 1 \vee 2$. (c) Segmented image using Newton method at 10 iterations, $\beta = 0.1$. (d) Segmented image using gradient descent method at 126 iterations.

The next example is a numerical test on a car plate image. We want to challenge the segmentation by adding a large amount Gaussian distributed noise to the real image and to take the polluted image in Fig.3(b) as the observation data. For the noise, we have $\text{SNR} \approx 1.7$, which means the observation image is very noisy. We want to compare our proposed algorithms with the Chan-Vese method(CVM) [8]. The results are displayed in Fig.3(d)(e)(f). For Newton method, we have to get the initial values by Algorithm 1 because the noise is so large that the simple scaling procedure (18) is not good enough to make it convergent. The image in Fig.3(c), which is produced by gradient descent method at 288 iterations with $\beta = 0.65, r = 2 \times 10^5, \Delta t = 8 \times 10^{-9}, n_\lambda = 1$, is obtained. The convergence rate of Newton method is much faster than that of gradient descent method, though gradient descent method and Newton method need nearly the same computational costs. The image produced by gradient descent method and Newton method is also nearly the same. The CVM needs more iterations and computing cost, which is also the case for tests with other images.

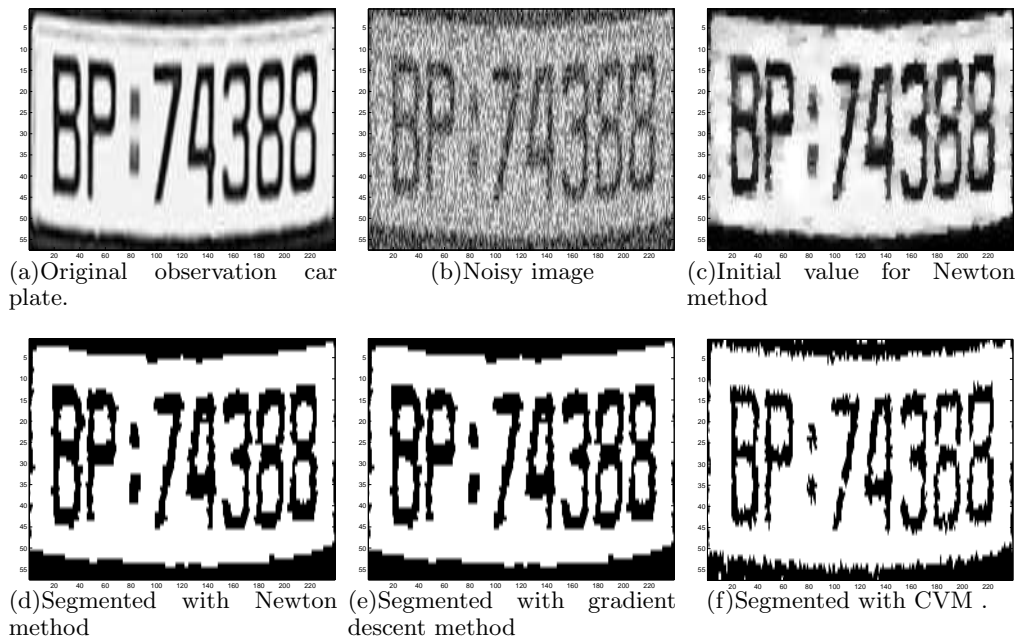


Fig. 3. A comparison of three algorithms for the segmentation of a car plate. (a) The original observation for a car plate. (b) The noisy car plate with $\text{SNR} \approx 1.7$. (c) The initial value of ϕ for Newton method (d) Segmented image using Newton method. (e) Segmented image by gradient descent method. (f) Segmented image with CVM.

In the last two examples, we apply our algorithms to some real images. One is a helicopter (c.f. Fig.4(a)) taken in a cloudy day. The results produced by the two algorithms are shown in Fig.4(b)(c). We have used $\beta = 0.1, r = 10^7, \Delta t = 1 \times 10^{-8}, n_\lambda = 10$ for the algorithms. The other image is a photo for a field, see Fig.5(a). From the segmented image shown in Fig.5(b)(c), we can easily identify the boundary of the different parts of the field. The result is obtained by using gradient descent method at 169 iterations using $\beta = 0.1, r = 10^7, \Delta t = 1 \times 10^{-8}, n_\lambda = 10$ and using Newton method at 9 iterations. This image is rather complicated. We can see that some objects only have a few pixels, even the density values of the image are not nearly constant at each subregion. But it seems there is no problem for our model and algorithms to get a good segmentation for this image.

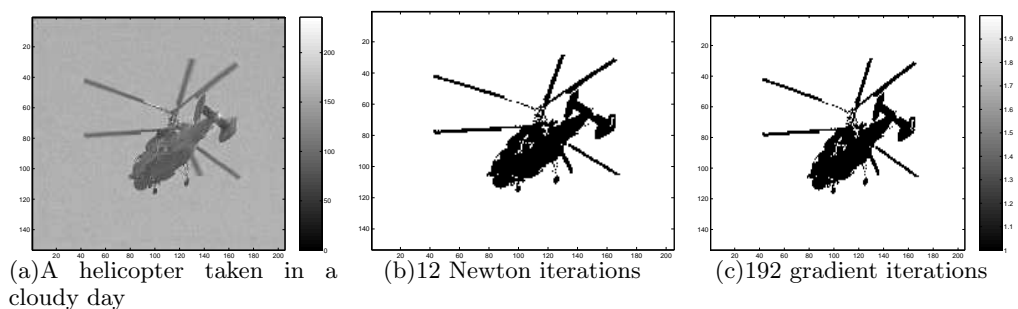


Fig. 4. (a) A helicopter taken in a cloudy day. (b) Segmented image using Newton method at 12 iterations with $\beta = 0.1$. (c) Segmented image using gradient descent method at 192 iterations.

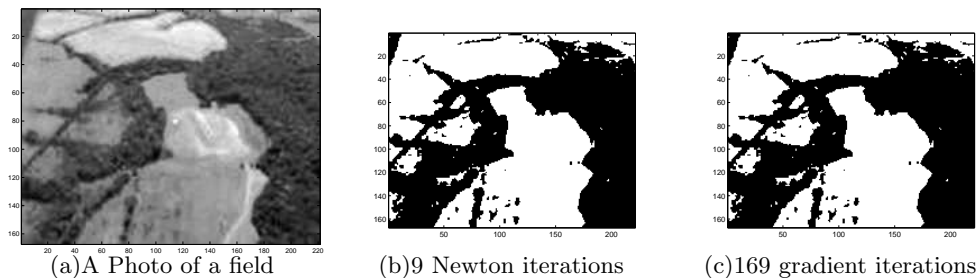


Fig. 5. (a) A Photo for a field. (b) Segmented image using Newton method at 9 iterations with $\beta = 0.1$. (c) Segmented image using gradient descent method at 169 iterations.

V. Conclusion.

We have also done many other tests for the proposed model with gradient descent method and Newton method. Even the image is very complicated and the density values of the image are not nearly constant inside each phase, or very noise, the proposed algorithms are able to find the phases properly. Newton method is still very fast like that in [18]. We can use gradient descent method to produce the initial segmentation for Newton method when the image u_0 has a large amount noise added.

REFERENCES

- [1] J. Barzilai and J.M. Borwein. Two point step size gradient methods. *IMA J.Numer. Anal.*, 8(2):141–148, 1988.
- [2] T. Chan and L. A. Vese. Active contours without edges. *IEEE Tran. image Proc.*, 10:266–277, 2001.
- [3] Yu-Hong Dai. Alternate step gradient method. *Optimization*, 52(4-5):395–415, 2003.
- [4] Yu-Hong Dai and Ya-Xiang Yuan. Alternate minimization gradient method. *IMA J. Numer. Anal.*, 23(3):377–393, 2003.
- [5] Yu-Hong Dai and Ya-Xiang Yuan. Analysis of monotone gradient methods. *Journal of Industrial and Management Optimization*, 1(2):181–192, 2005.
- [6] J.Lie, M. Lysaker, and X.-C. Tai. A binary level set model and some applications for Mumford-Shah image segmentation. *IEEE Trans. Image Proc.*, 15(4), 2006.
- [7] J.Lie, M. Lysaker, and X.-C. Tai. A variant of the level set method and applications to image segmentation. *Math. Comp.*, 2006.
- [8] L.A.Vese and T.F.Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.
- [9] J. Lie, M. Lysaker, and X.-C. Tai. Piecewise constant level set methods and image segmentation. In Ron Kimmel, Nir Sochen, and Joachim Weickert, editors, *Scale Space and PDE Methods in Computer Vision: 5th International Conference, Scale-Space 2005*, volume 3459, pages 573–584. Springer-Verlag, Heidelberg, April 2005.
- [10] M.Burger and S.Osher. A survey on level set methods for inverse problems and optimal design. CAM-report-04-02, UCLA, Applied Mathematics, 2004.
- [11] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl.Math*, 42:577–685, 1989.
- [12] Lars K. Nielsen, Xue-Cheng Tai, Sigurd Aanosen, and Magne Espedal. A binary level set model for elliptic inverse problems with discontinuous coefficients. CAM-report-05-51, UCLA, Applied Mathematics, 2005.
- [13] S.Osher and J.A.Sethian. Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *J.Comput.Phys.*, 79:12–49, 1988.
- [14] S.Osher and R.Fedkiw. An overview and some recent results. *J.Comput.Phys*, 169 No. 2:463–502, 2001.
- [15] Xue-Cheng Tai and Tony F. Chan. A survey on multiple level set methods with applications for identifying piecewise constant functions. *Int. J. Numer. Anal. Model.*, 1(1):25–47, 2004.
- [16] Xue-Cheng Tai, Oddvar Christiansen, Ping Lin, and Inge Skjaelaen. A remark on the MBO scheme and some piecewise constant level set methods. CAM-report-05-24, UCLA, Applied Mathematics, 2005.
- [17] Xue-Cheng Tai and Hongwei Li. Piecewise constant level set methods (PCLSM) for elliptic inverse problems. CAM-report-05-59, UCLA, Applied Mathematics, 2005.
- [18] Xue-Cheng Tai and Changhui Yao. Fast piecewise constant level set methods (PCLSM) with newton updating. CAM-report-05-52, UCLA, Applied Mathematics, 2005.