A Multi-resolution Stochastic Level Set Method for Mumford-Shah Image Segmentation

Yan Nei Law, Hwee Kuan Lee, Andy M. Yip

Abstract

The Mumford-Shah model is one of the most successful image segmentation models. But existing algorithms for the model require a good initial guess to obtain good results and are therefore impractical. To make the model practical, it is essential to develop an algorithm which can compute the global or near global optimal solution efficiently. While gradient descent based methods are well-known to find a local minimum only, even many stochastic methods do not provide a practical solution to this problem either. We propose a hybrid approach which combines gradient based and stochastic optimization methods to resolve the problem of sensitivity to the initial guess. At the heart of our algorithm is a well-designed basin hopping scheme which uses global updates to escape from local traps in a way that is much more effective than standard stochastic methods. In our experiments, a very high quality solution is obtained within ten stochastic hops whereas the solutions obtained with other standard stochastic methods are incomparable even after thousands of steps. We also propose a multi-resolution approach to reduce the computational cost and enhance the search for the global minimum. Furthermore, we derived a simple but useful theoretical result relating solutions at different resolutions.[1]

I. Introduction

Image segmentation is indispensable in many applications since it facilitates the extraction of information and interpretation of image contents. For instance, in high throughput imaging, storage and organization of a large number of images according to the extracted information are required. Therefore, it is crucial to segment each image into meaningful partitions in an accurate, fast, automated and robust way. The segmentation problem is fundamental to image processing. But it is a difficult one since it is highly ill-posed.

While standard segmentation methods may work reasonably well when the images have simple contents and are of high quality, they perform poorly on practical images which are often non-ideal. For example, the nuclei of the breast cancer cells in Fig. 2 have a diverse intensity levels. Thus, simple thresholding or even adaptive thresholding [1] may miss many nuclei. In the brain MRI image shown in Fig. 8, the various matters and tumor have no sharp defining boundary. Most edge-based segmentation methods do not work. In the zebrafish image shown in Fig. 9, the proliferating-cell nuclear antigen (PCNA) clusters (black dots around the intervillus pockets) consist of many

Y. N. Law and H. K. Lee are with Bioinformatics Institute, Singapore. A. M. Yip is with the Department of Mathematics, National University of Singapore, Singapore. All correspondence should be addressed to Andy M. Yip. Email: andyyip@nus.edu.sg.

[1]Image data and computer codes written in Matlab are available upon request from the corresponding author.

clusters of cells. Most algorithms will treat spatially separated clusters as individuals. But sometimes it may be more useful to perceive them collectively as a region. Similarly, the villi also appear to be formed by many small white patches. This can make the recognition of the PCNA and villi much more difficult. The high noise level also poses additional difficulty which would fail methods like thresholding and watershed [1]. When these complexities can not be reduced by repeating or redesigning the experimental setup, sophisticated computational methods provide valuable alternatives.

There are many advanced approaches for image segmentation: clustering, histogram-based, region-growing, graph partitioning and optimization model-based. A recent survey in segmentation can be found in [2]. Among the various advanced segmentation approaches, optimization model-based methods often give very promising results, provided that the underlying optimization problem is solved accurately numerically. An advantage of this approach is that one can incorporate domain knowledge into the model explicitly through rigorous mathematical modeling.

The Mumford-Shah Segmentation Model

In this paper we consider the well-known Mumford-Shah segmentation model [3] which is one of the best models in segmentation. It can handle gracefully the aforementioned complex situations and is very robust to noise. A distinctive feature is that it is region-based which allows it to segment objects without edges. For the same reason, it can group a cluster of smaller objects into a larger object.

For a given image u_0, the piecewise constant Mumford-Shah model seeks for a set of curves C and a set of constants $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ which minimize an energy functional given by:

$$F^{\mathrm{MS}}(C, \mathbf{c}) = \sum_{i=1}^{n} \int_{\Omega_i} |u_0(x, y) - c_i|^2 dxdy + \mu \times \text{Length}(C). \tag{1}$$

The curves in C partition the image into n mutually exclusive segments Ω_i for $i = 1, 2, \ldots, n$. The idea is to partition the image so that the intensity of u_0 in each segment Ω_i is well-approximated by a constant c_i. The goodness-of-fit is measured by the fitting term $\int_{\Omega_i} |u_0(x, y) - c_i|^2 dxdy$. On the other hand, a minimum description length principle is employed which requires the curves C to be as short as possible. This increases the robustness to noise and avoids spurious segments. The parameter $\mu > 0$ controls the trade-off between the goodness-of-fit and the length of the curves C.

Optimization Methods

Solving the optimization problem associated with most sophisticated models, including the Mumford-Shah model, is far from trivial. Updating of the curves C in Eq. (1) is complicated in situations in which the curves need to be split or merged. A breakthrough in solving the Mumford-Shah optimization problem is achieved by Chan and Vese [4], and a more general version in [5]. They propose to use the level set method introduced by Osher and Sethian [6] to represent the curves C by the zeroth level of a set of functions Φ defined on the image domain. Simply put, it is a change of variable from C to Φ. As a result, the energy functional in (1) becomes $F^{\mathrm{MS}}(\Phi, \mathbf{c})$, see §II-C

for details. The problem then becomes a standard variational problem which is much easier to handle. Methods for solving such a variational problem can be largely divided into two categories: deterministic and stochastic.

In gradient descent based deterministic methods, a new solution is obtained by moving the current solution along a descent direction in which the energy is decreasing [7]. In greedy based deterministic methods, membership in pixel or region level is updated in a greedy way in order to optimize the energy locally, see the section Related Work. These deterministic methods work well when the energy functional is convex. When it is non-convex, which is the case of the Mumford-Shah model, the solution obtained can be far from the globally optimal solution. Moreover, tremendously different solutions may be obtained when different initial solutions are used.

Global optimization techniques such as interval arithmetics and subdivision schemes [8] are available. But they are often suitable only for problems of very small size due to their high complexity.

For non-convex energy functionals, Monte Carlo (MC) methods, a major class of stochastic methods, provide viable alternatives. They can overcome the problem of being trapped in a local minimum. In particular, simulated annealing implemented with the Metropolis algorithm [9] and a local updating scheme has been popular in solving many optimization problems. The annealing theorem says that such a method can converge to the global solution in the long run [10]. However, its convergence to the global solution is logarithmically slow. Moreover, it is often difficult to specify a good annealing schedule to make the method work, see Fig. 7. Thus, it still does not provide a practical solution to our problem.

To remedy the weaknesses of simulated annealing with local updates, many advanced MC methods such as generalized ensembles [11], [12], reweighting [13], [14], cluster algorithms [15] and basin hopping [16] have been developed over the past couple of decades to handle a variety of problems in condensed matter physics that cannot be solved efficiently using traditional simulated annealing. Some of these methods adopt global updating strategies. To the best of the authors' knowledge, they have not been utilized to solving the Mumford-Shah image segmentation problem. We show in this paper that basin hopping with global updates provides a very effective means to solving the problem.

Our Contribution

In this paper, we propose a hybrid approach, which we called *Stochastic Level Set Method*, to combine the strengths of advanced deterministic and stochastic optimization approaches. It is a kind of basin hopping method [16] which uses the gradient information of the energy function to make the best local move and uses a stochastic split-and-merge method to make a global move towards the global solution. The split-and-merge step is the crucial element in our algorithm. It is tailor-made for the Mumford-Shah model and can overcome the slowness of many standard stochastic methods by allowing a large change in the partition of the image. *Our work has solved a practical problem existed in current Mumford-Shah algorithms: The need to specify a "good" initial guess which is usually unknown in many imaging applications.*

To further reduce the computational cost, we use a multi-resolution method. A key issue is to ensure that the solution at a lower resolution indeed gives a good approximation to the original problem. *For this, we derived an*

important theoretical result relating solutions at different resolutions.

Related Work

Our method combines the best of deterministic and stochastic methods for the Mumford-Shah energy. A similar hybrid approach has been proposed for solving the shape-from-shading problem [17]. But we go beyond it by using the more sophisticated basin hopping method.

A greedy region merging method is proposed in [18] which starts with a very fine segmentation and progressively merges two regions at a time according to a predefined schedule. Other greedy schemes are considered in [19], [20]. In these methods, the membership of a small group of pixels is updated at a time. A different approach is proposed in [21] which reduces the optimization problem to a series of simple thresholding and filtering steps. These methods are very efficient to compute a local optimal solution of the Mumford-Shah energy. But they make no attempt to compute the global solution.

A very interesting result is obtained in [22]. They show that for $n = 2$ in Eq. (1) and for fixed constants $\mathbf{c} = (c_1, c_2)$, the non-convex Mumford-Shah energy can be reformulated as a convex energy. Thus, a global solution (for a fixed $\mathbf{c}$) can be easily computed. However, the result does not hold for $n > 2$. Therefore, when $n > 2$, we are dealing with a non-convex energy which makes existing algorithms impractical due to the need of a good initial solution — a luxury which is seldom available.

As with many variational models, the Mumford-Shah model can be cast in a probabilistic framework in which the pointwise maximum *a posteriori* estimate is equivalent to the solution of the original model [23]. In [24], a soft Mumford-Shah segmentation model in the probabilistic framework is also proposed. However, the algorithms used to compute the optimal distributions are deterministic. Thus the aim in these work is fundamentally different from us — we apply stochastic optimization methods to optimize a deterministic model whereas the above methods apply deterministic optimization methods to optimize stochastic models. These algorithms make no significant attempt to compute a global solution and may get trapped in a local minimum. Indeed, our framework may be applied to improve their solution quality.

A work which appears to be the closest to the present one is given by Juan *et al*. [25]. They consider stochastic methods to compute solutions to various active contour models, including the Mumford-Shah. The major difference, also our main contribution, is that our updates are global whereas theirs are local. In their method, each point on the active contour (or level set function) moves in its normal direction at a speed which is a random perturbation of the one suggested by the gradient descent. Thus the speed depends only on local quantities. In contrast, our hopping step is region based and is thus global which can escape from local minima effectively (see Fig. 4). Another important difference is that, being based on local updates, their method relies on simulated annealing that is constrained by the annealing theorem. Our method does not require any annealing schedule but can still overcome local traps and stabilize in a finite number of steps.

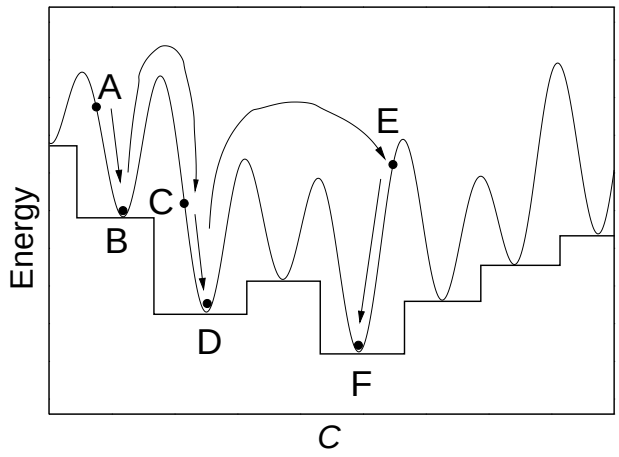

Fig. 1. A schematic plot of the basin hopping method with x-axis representing the multi-dimensional solution space C. The method essentially flattens the energy landscape to reduce energy barriers.

II. Algorithm

In this section, we present our proposed algorithm. In §II-A, we present the basic ideas behind basin hopping and explain why it is superior to sole deterministic or sole stochastic methods. Next, in §II-B, we present the outline of our proposed algorithm. A detailed description of each major component of our method is presented in §II-C–§II-F.

A. The Basin Hopping Method

The basin hopping method [16] combines deterministic and stochastic methods to remove energy barriers and escape from local minima. We illustrate the main ideas of the method through a schematic depicted in Fig. 1. Starting from an initial guess (point A), a gradient descent is performed to find the local minimum (point B). A stochastic hopping step is invoked to jump from point B to a new iterate, say point C, which is a global update move selected randomly from a small set of candidate moves. A gradient descent is performed again from point C to obtain a lower energy minimum at point D. Several hopping steps can be performed to obtain a better iterate, for example, an additional hop brings the iterate from point D to E and then to F.

In general, each point is attracted to a local minimum through a gradient descent process. In effect, basin hopping simplifies the energy landscape. This effective energy is depicted as the staircases in Fig. 1. Notice that the simplified energy has much less energy barriers. We may view the method as a series of hopping between local minima (basins). If the new local minimum gives a lower energy, then the new solution will be taken without reservation. Otherwise, a coin may be flipped to determine if the new iterate is accepted. For methods using local updates, it is necessary to allow temporary increases in energy. But for methods using global updates, they can still escape from local minima even if we accept only a new solution leading to a decreased energy.

In contrast, the gradient descent method of [4] computes a local minimum only. Good results can be obtained efficiently, provided that a good starting point is available. Greedy algorithms of [18]–[20] can be viewed as a series

of deterministic hops. A problem is that if the hopping leads to a bad solution, then there is no way out. But basin hopping explores many possibilities. Thus, it is clear that stochastic methods have an edge over deterministic ones for non-convex problems.

Basin hopping is clearly much better than simulated annealing with local updates, this is verified in Fig. 7. With basin hopping, a local minimum is quickly found using the gradient descent while in simulated annealing, a minimum is not found until the final temperature. The energy landscape is also not simplified in simulated annealing. The exploration of the search space using local updates is much slower than that using the global updates we proposed. Finally, once the temperature is lower than the energy barrier ΔE, the probability of escaping from this local minimum goes as $\sim \exp(-\Delta E/T)$, which is exponentially small.

The success of the basin hopping method highly depends on the design of the hopping strategy. Our hopping strategy is a stochastic region splitting and merging method which is specially designed for the Mumford-Shah energy and is effective for the diverse images we tested. On the other hand, repeated hopping and local optimization increase the computational cost. We alleviate this problem using a multi-resolution approach.

B. The Stochastic Level Set Method

The proposed algorithm is outlined as follows:

1) Reduce the resolution of the original image by L levels.
2) *Local optimization*: Apply the level set based gradient descent method with initial curves C^0 to obtain a new iterate $C^{\rm curr}$ and energy $E^{\rm curr}$.
3) Repeat the followings up to a predetermined number of steps or until all possible hops have been rejected:
   a) *Hopping*: Use the stochastic split-and-merge to hop to a configuration C'.
   b) *Local optimization*: Apply the level set based gradient descent method with initial curves C' to obtain a proposed new iterate $C^{\rm prop}$ and energy $E^{\rm prop}$.
   c) *Accept/reject*: If $\Delta E = E^{\rm prop} - E^{\rm curr} < 0$, then set $C^{\rm curr} \leftarrow C^{\rm prop}$ and $E^{\rm curr} \leftarrow E^{\rm prop}$.
4) Repeat the followings until the original resolution is reached:
   a) Increase the image resolution by 1 level.
   b) *Local optimization*: Apply the level set based gradient descent method with initial curves $C^{\rm curr}$ to obtain a new iterate $C^{\rm prop}$.
   c) *Accept*: Set the current iterate, $C^{\rm curr} \leftarrow C^{\rm prop}$.

We remark that the dependence of the Mumford-Shah energy on the constant $\mathbf{c}$ is ignored since it can be determined from the curves, see Eq. (3) below. This method can be formulated with simulated annealing [10] by setting the acceptance probability to $\min\{1, \exp(-\Delta E/T)\}$ and decreasing the value of T gradually. But our hopping scheme is global so that simulated annealing is not needed.

C. Level Set Methods for the Mumford-Shah Model

Splitting and merging of evolving curves are non-trivial, Vese and Chan [5] propose to use a set of functions $\Phi = (\phi_1, \ldots, \phi_m)$ defined on the image domain to encode the segments into the intersections of the positive and negative regions of Φ and to encode the dividing curves C into the locations at which one of component functions is zero. In this paper, we use two level set functions ϕ_1, ϕ_2 to represent up to 4 *phases*. Each phase may consist of more than one connected component. Each connected component is a segment. For each location (x, y), the phase that it belongs is determined by the following encoding:

$$\begin{aligned}
&\text{Phase 1:} \quad \phi_1(x,y) > 0 \quad \text{and} \quad \phi_2(x,y) > 0,\\
&\text{Phase 2:} \quad \phi_1(x,y) > 0 \quad \text{and} \quad \phi_2(x,y) < 0,\\
&\text{Phase 3:} \quad \phi_1(x,y) < 0 \quad \text{and} \quad \phi_2(x,y) > 0,\\
&\text{Phase 4:} \quad \phi_1(x,y) < 0 \quad \text{and} \quad \phi_2(x,y) < 0.
\end{aligned}$$

Then, the Mumford-Shah energy (1) can be expressed as

$$\begin{aligned}
F^{\mathrm{MS}}(\Phi, \mathbf{c}) &= \int_\Omega (u_0 - c_1)^2 H(\phi_1) H(\phi_2) dxdy + \int_\Omega (u_0 - c_2)^2 H(\phi_1)[1 - H(\phi_2)] dxdy \\
&+ \int_\Omega (u_0 - c_3)^2 [1 - H(\phi_1)] H(\phi_2) dxdy + \int_\Omega (u_0 - c_4)^2 [1 - H(\phi_1)][1 - H(\phi_2)] dxdy \\
&+ \mu \int_\Omega \left[|\nabla H(\phi_1)| + |\nabla H(\phi_2)| \right] dxdy. \qquad (2)
\end{aligned}$$

Here, $H(z)$ is the Heaviside function which is 1 if $z > 0$, $1/2$ if $z = 0$, and 0 otherwise. The crucial advantage of such a change of variable from C to Φ is that the minimization with respective to Φ is much easier to carry out. Moreover, splitting and merging of curves and subregions are done by simply "moving the level set functions up and down". In general, we can use m level set functions to represent up to $n = 2^m$ phases. See [5] for details.

We remark that it is possible to obtain many segments using only a small number of phases. But the segments within each phase are fitted with the same constant c_i by the model.

D. The Level Set Based Gradient Descent Method

The gradient descent method brings an initial solution to its respective local minimum. This method is used to optimize $F^{\mathrm{MS}}(\Phi, \mathbf{c})$ with respect to Φ. To optimize $F^{\mathrm{MS}}(\Phi, \mathbf{c})$ with respect to $\mathbf{c}$ as well, we use the alternating approach in [4]. Starting from a given initial guess Φ^0, we update Φ and $\mathbf{c}$ alternatively as follows:

1) For a given fixed Φ^{k-1}, compute $\mathbf{c}^k$ by solving the problem $\min_{\mathbf{c}} F^{\mathrm{MS}}(\Phi^{k-1}, \mathbf{c})$.
2) For a given fixed $\mathbf{c}^k$, compute Φ^k by applying one step of gradient descent starting at Φ^{k-1}.

For the ease of presentation, we shall call this alternating method as level set based gradient descent.

It can be easily shown that the problem in the first step has the following explicit solution:

$$c_i = \text{average of } u_0 \text{ over phase } i, \qquad \text{for } i = 1, 2, \ldots, n. \qquad (3)$$

Thus the constants can easily be determined from Φ.

For the second step, we denote by $\frac{\partial F^{\mathrm{MS}}}{\partial \Phi}$ the gradient of F^{MS} with respective to Φ. Then, the gradient descent method computes the next iterate by the following update:

$$\Phi^k = \Phi^{k-1} - \Delta t \frac{\partial F^{\mathrm{MS}}}{\partial \Phi}(\Phi^{k-1}, \mathbf{c}^k), \tag{4}$$

where $\Delta t > 0$ is the step length parameter. The detailed numerical implementation of Eq. (4) can be found in [5].

When the alternating minimization converges at the K-th step, we obtain a configuration $(\Phi^K, \mathbf{c}^K)$ which is the local minimum of F^{MS} starting from Φ^0.

E. Stochastic Region Split-and-Merge

The success of our proposed method depends largely on the type of hop performed. We design a very effective stochastic split-and-merge in which a large area of the image can be splitted and merged in a single step. This greatly reduces the number of hops to search for a global minimum. In our examples, it is sufficient to get a much improved segmentation within 10 split-and-merge steps whereas the results of standard simulated annealing with local updates are incomparable even after thousands of steps.

In the splitting step, we randomly choose a segment Ω_i (a connected component in a phase) according to the probability distribution $\mathrm{Prob}(\text{choosing } \Omega_i) = w_i / \sum_i w_i$, where

$$w_i = \int_{\Omega_i} (u_0(x,y) - \bar{c}_i)^2 dxdy.$$

Here, $\bar{c}_i$ is the average of u_0 over Ω_i. This weighting scheme is designed for the Mumford-Shah model since it prefers to split a segment having a high sum-of-differences-squared, which resembles the fitting term of the Mumford-Shah energy. We have also considered other possible weighting schemes. Since a sum-of-differences-squared of a connected component can be written as the product of region size and intensity variance:

$$w_i = |\Omega_i| \times \frac{1}{|\Omega_i|} \int_{\Omega_i} (u_0(x,y) - \bar{c}_i)^2 dxdy,$$

it is intuitive to try on a weighting scheme using either one of them. However, we found that these two schemes are unable to escape from local minima (results not shown in the paper). Using region size, the largest but homogeneous regions (e.g. background) are often picked. But they are often not good candidates to split. Using variance, very small regions (e.g. 2-pixel regions) are often picked. However, the move due to splitting these small regions is too small to escape from local minima. In contrast, using the sum-of-differences-squared can balance these two factors.

Another possible weighting scheme we considered is based on the energy change ΔE. This approach is quite natural since it leads to a greedy way to minimize the energy. In this scheme, we randomly choose a segment with probability proportional to a decreasing function of ΔE, e.g. $\exp(-\Delta E)$ and $\max\{-\Delta E, 0\}$. This scheme is inconsistent with the essence of basin hopping, which is to allow a temporary increase in energy so that large global changes can be achieved. This scheme often picks small local updates to avoid a large increase of energy. We found that ΔE often depends heavily on the length term. Thus, the inclusion of the length term may hinder the splitting of large regions. This is the reason why we exclude the length term in our splitting scheme.

To split the selected segment into two subsegments, we apply the ISODATA thresholding method [26] which computes an optimal threshold so as to make the intensity levels within each subsegment as close as possible. This can reduce the fitting term of the Mumford-Shah energy optimally.

In the merging step, we try to merge either one of the resulting subsegments obtained in the splitting step with one of the remaining $n-1$ phases. Thus, there is $2(n-1)$ potential mergers. The merger which causes the greatest reduction in the fitting term is used. In case the chosen merger has been rejected previously since the last hop is accepted, the merger which causes the next greatest reduction is used instead. Our merging step is deterministic. We found experimentally that stochastic merging based on the reduction in the fitting term does not give a better performance than the deterministic scheme, see Fig. 5 (Middle). Thus we use the deterministic scheme for simplicity.

F. A Multi-resolution Hybrid Approach

Down-sampling of the image reduces the search space and provides a less noisy image. This can further reduce the number of local minima. Computational cost for low resolution images is also relatively cheap. For these reasons, we apply the stochastic level set method at a low resolution image to search for a global minimum at the reduced resolution. This allows us to find a close-to-optimal solution quickly. Since detailed features are lost in the low resolution image, it is necessary to refine the solution at the original resolution. We do this by progressively increasing the resolution one level (a factor of two in each dimension) at a time until the original resolution is reached. Each time the resolution is increased, we apply the (deterministic) level set based gradient descent method to refine the solution. We found that only a small number of iterations are needed because a good solution is already obtained after applying the stochastic level set method at the lowest resolution. Our results show that this approach allows us to obtain a solution of better quality in much less time compared to the pure deterministic method in [5] and its multi-resolution version, see Fig. 2 and 5 (Left).

In order for the multi-resolution method to work, there are two main issues. Firstly, an outline of the target objects should be preserved at the lowest resolution. This restricts the number of resolution levels that can be reduced. The resolution of the images used in the experiments is reduced by $2-3$ levels with no problem in preserving the object outlines. Secondly, the optimization problems at different levels must have similar solutions. We found that if the image at one level lower is constructed by a simple local averaging, then the segmentation obtained from the lower resolution image is also optimal for the original image among all segmentations at the lower resolution. This result is stated precisely in the following theorem. The proof of the theorem can be found in the Appendix.

Theorem 1: Let $u=(u_{i,j})$ be a given $M\times M$ image. Let $v=(v_{i,j})$ be the $M/2\times M/2$ lower resolution image constructed by a simple averaging, that is,

$$v_{i,j}=\frac{1}{4}(u_{2i-1,2j-1}+u_{2i-1,2j}+u_{2i,2j-1}+u_{2i,2j})$$

for $1\le i,j\le M/2$. Then the solution of the Mumford-Shah problem with data u is identical to that with data v if the segmentation of u is restricted to the resolution one level lower, i.e. $u_{2i-1,2j-1}, u_{2i-1,2j}, u_{2i,2j-1}, u_{2i,2j}$ must belong to the same segment for each i,j.

III. METHODS

For the images used in our experiments, the breast cancer cell image is obtained within the first and second authors' institute. The brain MRI and zebrafish intestine images are obtained in [27] and [28] respectively.

To validate our method, we compare it with the pure gradient descent method proposed in [4], [5]. While other existing methods such as the multi-scale scheme in [20] and the thresholding method in [21] can be computational more efficient, they are deterministic and not designed for global optimization. Thus, the solutions obtained are qualitative similar. For this reason, it is sufficient to compare our method with the standard gradient descent method. To demonstrate the efficiency of our proposed hopping scheme, we also compare our method with the multi-scale version of the gradient descent method and the simulated annealing method.

In our results, we use 4 phases which are sufficient to obtain a good segmentation for each of the test images. The parameter μ is chosen manually. To test the robustness of the methods to the initial guess, we use various initial guesses when testing on the breast cancer cell image. These initial guesses have been used in [4], [5]. For the other two images, we choose the initial condition so that the pure gradient descent method gives the most visually appealing results.

We count the number of gradient descent steps and use it as the number of iterations. The computational cost spent on the stochastic split-and-merge steps is negligible. When applying the hopping steps, the image resolution is reduced by three levels for the cell and the MRI images. 10 hops are applied. 100 gradient descent steps are taken both initially and after each hop. Thus a total of 1100 gradient descent steps are taken at the lowest resolution (3rd level). At the 2nd, 1st and original levels, 100, 80 and 20 gradient descent steps are taken respectively. To standardize the number of iterations at different levels for comparison, we treat one iteration at L levels lower than the original level as $1/4^L$ iterations at the original level. This is because the number of pixels is reduced by a factor of $1/4^L$ and the cost per iteration is proportional to the number of pixels. As a result, our method takes 64 equivalent iterations at the original level. For the zebrafish image, the image resolution is reduced by two levels and the remaining settings are the same. Hence, our method takes 109 equivalent iterations at the original level.

For the pure gradient descent method, the number of iterations varies and is reported in each figure individually. We tried to use the smallest possible number of iterations while maintaining the quality of the resulting segmentations.

In our comparison to the simulated annealing, we start the simulated annealing scheme with assigning each pixel to a random phase membership. Each update is done by randomly choosing a pixel, proposing a new randomly chosen phase membership, and then accepting the proposed change with the probability $\min(1, \exp(-\Delta E/T))$. Here, ΔE is the change of energy due to the proposed change and T is the annealing temperature which is initially set to 6×10^7 and reduced by a factor of 0.98 in every $40 \times N$ (number of pixels) updates. We call N times of such local updates as one iteration. The computation cost of one iteration is comparable to that of one iteration of our stochastic level set method. The annealing scheme is chosen to obtain the largest decreased of energy with the least number of iterations.

All algorithms are implemented in MATLAB 7 with a Pentium D 3GHz machine. The total CPU times are recorded. The energy of the resulting segmentations are reported in Table I as a measure of the goodness of the

results.

IV. RESULTS

A. Robustness to the Initial Condition

In this test, the image we used consists of breast cancer cells. The results of the pure gradient descent and our hybrid methods utilizing different initial conditions are shown in Fig. 2. The gradient descent method is very sensitive to the initial condition. A reasonable result is obtained using the Initial Condition 1. But the other two results are poor. Further increasing the number of iterations does not improve the solution (results not shown here) since the iterations are trapped in a local minimum. In contrast, the stochastic level set method is very robust to the initial condition. For all the three distinct initial conditions, our method leads to essentially the same solution (pairwise difference less than 0.3%). Although we are not able to verify that the solution is globally optimal, the quality of the segmentation is much better and the energy is lower, cf. Table I.

B. The Hopping Process

In Fig. 3, we show some intermediate results during the hopping process. These results use the Initial Condition 3 in Fig. 2. After the first 100 iterations at the lowest level, the iterate is trapped in a local minimum. The result there is qualitatively similar to the one by the pure gradient descent at the original level, cf. Fig. 2, 3rd row 3rd column. After five rounds of hopping, a much improved result is obtained. As the resolution increases (after 10 hops), the refinements are very little, showing that the results in the lowest level are indeed very good. This indicates the potential to further lower the resolution when applying the hopping steps.

Next, we demonstrate the effect of our global hopping. In Fig. 4, we show the result before and after the fifth hop. Before the hop, the energy is 3.22×10^{12} and a large segment in the dark gray phase is chosen. Then it is splitted into two subsegments depicted in vertical and horizontal strips in the second picture. They roughly correspond to the background and cytoplasm respectively, cf. the original image in Fig. 2. Our splitting step is able to pick up this segment (with a high probability of 0.31) and try to correct it. The subsegment in vertical strips is then merged with the black background where the energy increases to 3.35×10^{12}. After the gradient descent is applied, the energy drops significantly to 3.05×10^{12} which is much lower than the energy before the hop. Therefore, the hop is accepted. This scenario is in analogy with the illustration in Fig. 1 where a hop brings the iterate from point D to E and then F.

C. Multi-resolution with hopping vs. multi-resolution without hopping

The result of multi-resolution gradient descent method without hopping is shown in Fig. 5 (Left). The Initial Condition 3 is used and the number of iterations for each level is the same as those for stochastic level set method. We can see from the figure that the result is trapped in a local minimum and looks similar to the result of the pure gradient descent method, cf. Fig. 2, 3rd row 3rd column. This indicates that a multi-resolution approach alone without hopping is insufficient to escape from a local minimum. The convergence profile is shown in Fig. 7 (Left).

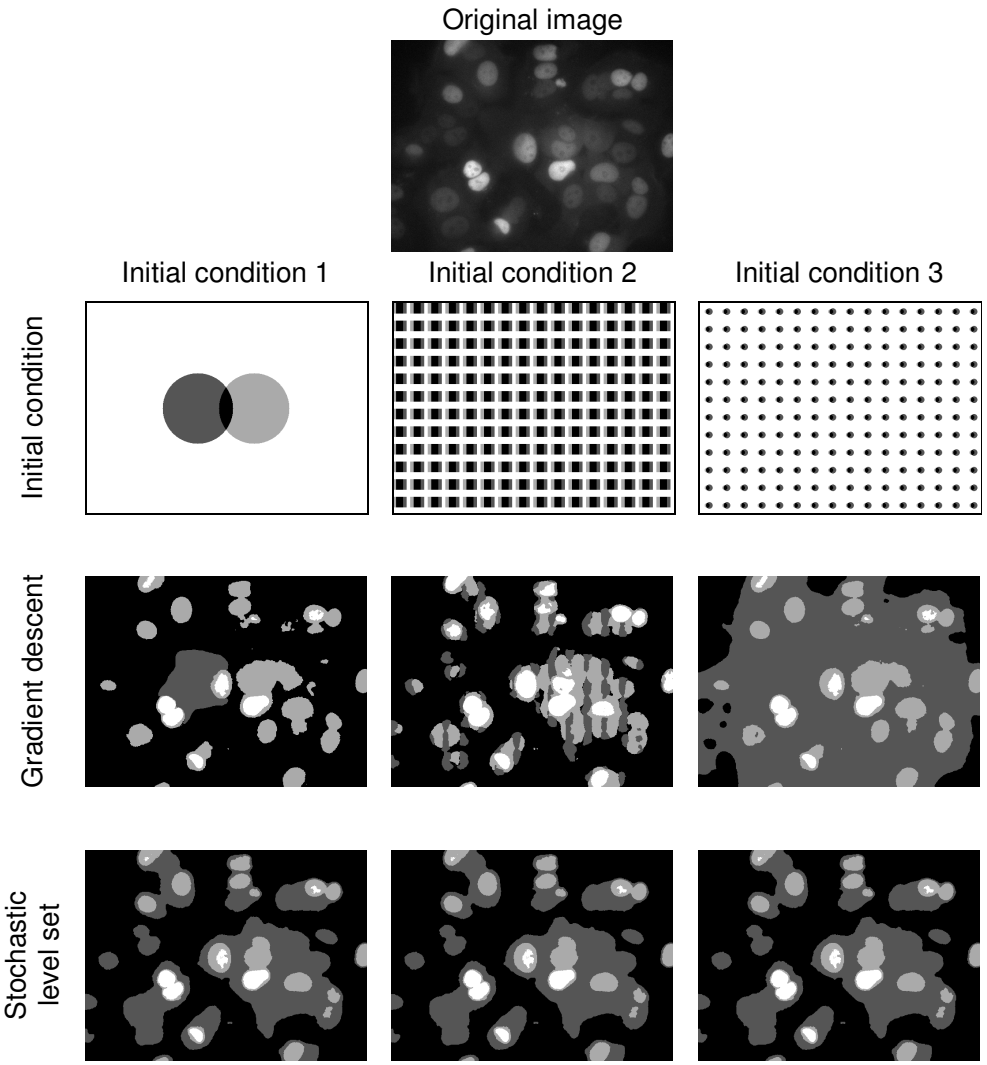

Fig. 2. Segmentation of the breast cancer cell image. 1st row: Original image with size 480×640. 2nd row: Three initial conditions. 3rd row: Gradient descent method with $\mu = 3 \times 10^7$ and 300 iterations. The CPU time is 690 sec. 4th row: Stochastic level set method with $\mu = 3 \times 10^7$. Number of equivalent iterations at the original level is 64. The CPU time is 135 sec. The difference between each pair of phase assignments is less than 0.3%.

D. Deterministic merging vs. stochastic merging

In the merging step, instead of choosing the merger which causes the greatest reduction in the fitting term, we randomly choose a combination (a subsegment and a phase) with probability proportional to the reciprocal of the fitting term to merge. The result is shown in Fig. 5 (Right). The Initial Condition 3 is used. The result is very similar to the result obtained by deterministic merging scheme, cf. Fig. 2, 4th row 3rd column. Since such a stochastic merging scheme does not significantly improve the result, we choose the simpler deterministic one.

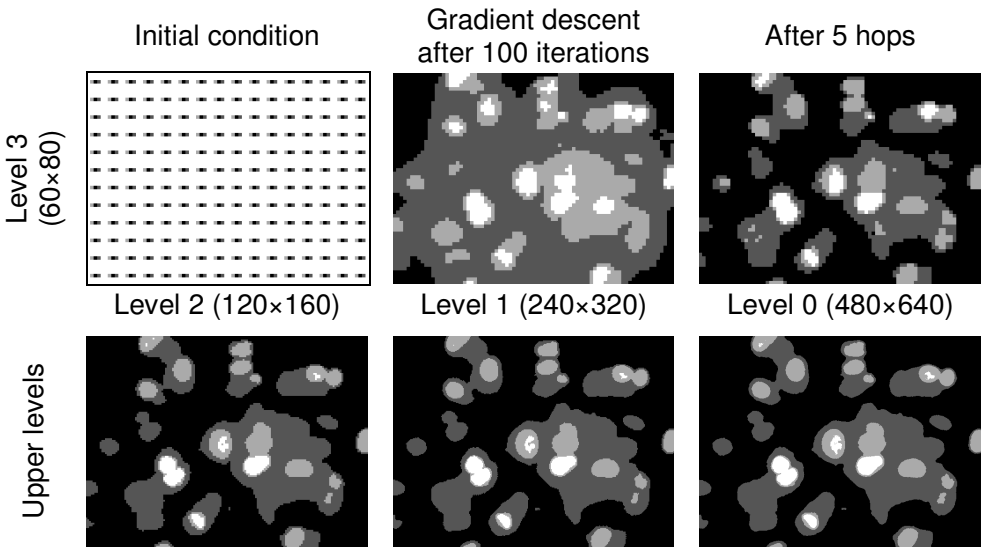

Fig. 3. Intermediate results of the stochastic level set method. Five rounds of hopping give a much improved result (cf. Fig. 2, 3rd row 3rd column). As the resolution increases (after 10 hops), the refinements are very little, showing that the results in the lowest level are indeed very good.

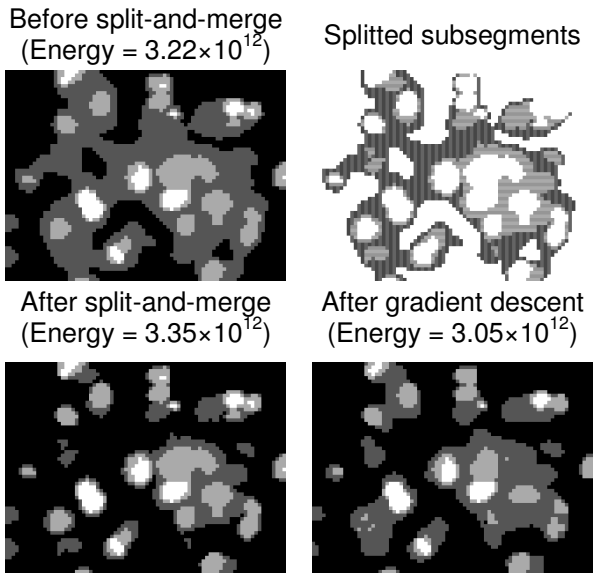

Fig. 4. Result before and after the fifth hop. In a single hop, a large area in the image is splitted and merged. Energy increases after the split-and-merge but decreases significantly after the gradient descent iterations. This scenario is in analogy with the illustration in Fig. 1, where a hop brings the iterate from D to E and then F.

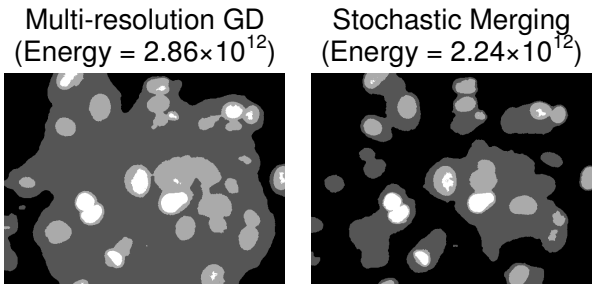

Fig. 5. Comparison with competitive approaches. (Left) Multi-resolution gradient descent method without hopping. The solution is trapped by a local minimum. (Right) Stochastic level set method but with a stochastic merging scheme. No significant improvement over the deterministic scheme is observed.

E. Stochastic level set method with global updates vs. simulated annealing with local updates

We performed a systematic investigation of different annealing schemes on the breast cancer cell image. Forty-one different annealing schemes are performed and we pick the best scheme for comparison with the stochastic level-set method.

The results of simulated annealing after 1000, 3000 and 8000 iterations are shown in Fig. 6. It takes about 20 hours for every 1000 iterations. Note that one iteration consists of 480×640 local updates, see the Method section. After 1000 iterations, the result is still incomparable to our result although it has already taken 20 hours. In terms of quality, the result after 3000 iterations is similar to ours. The disagreement between the two phase assignments is about 2%. Our energy is about 0.2% lower. To gain our confidence that our method can obtain solutions of good quality, we further run the simulated annealing until 8000 iterations. The result is only slightly (2%) different from the result after 3000 iterations. Compared to our result, the energy of simulated annealing after 8000 iterations is 2% lower than ours and the phase assignments differ only by 3%. Thus even after 8000 iterations, simulated annealing does not perform significantly better than our method.

Next, we compare the convergence profile of four different methods. The plots of the energy vs. number of iterations (up to 100 iterations) for the stochastic level set method (solid line), the gradient descent method (dashed line), the multi-resolution gradient descent method (dashed-dot line) and the simulated annealing method (dotted line) are shown in Fig. 7 (Left). For the stochastic level set and gradient descent methods, we start with the Initial Condition 3. For the simulated annealing method, we start with a random configuration. We observe that the stochastic level set method achieves a much lower energy level in less iterations than the other two methods. We also show in Fig. 7 (Right) the result of simulated annealing up to 8000 iterations. Our energy level is reached by simulated annealing only after about 3000 iterations after which the energy stabilizes.

F. More Examples

1) Brain MRI Image: The results on a brain MRI image are shown in Fig. 8. The objective is to segment out the tumor near the middle and to separate the grey matters from the white matters. In this image, the segments

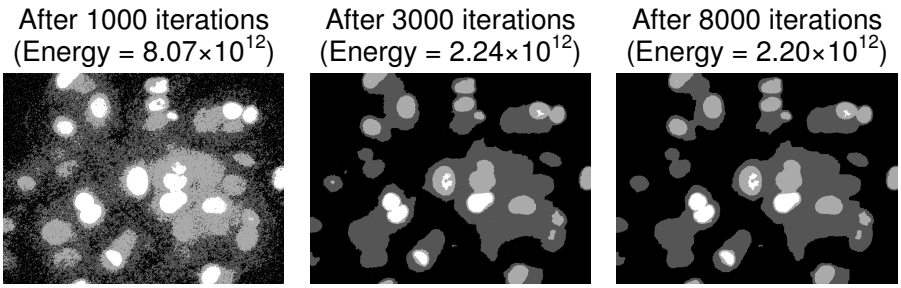

Fig. 6. Simulated annealing after 1000, 3000 and 8000 iterations. The CPU time for every 1000 iterations is about 20 hours.

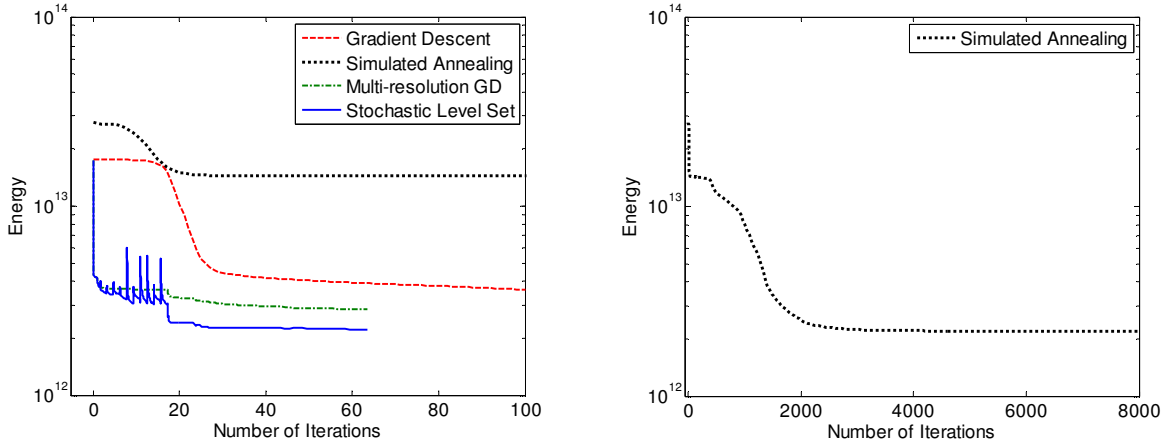

Fig. 7. Convergence profile of energy for various methods. (Left) 100 iterations for the gradient descent and simulated annealing methods. 64 iterations for the stochastic level set and multi-resolution gradient descent (GD) methods. (Right) 8000 iterations for the simulated annealing. Stochastic level set method obtains a much lower energy in smaller number of iterations. The spikes in the solid curve are due to a temporary increase in energy right after the merging step (but before the gradient descent iterations followed). Both hops that are accepted and rejected are shown.

have a complex topology and shape. The image is also noisy. The Mumford-Shah model can exhibit these complex segments as a solution. But it is non-trivial to realize. The pure gradient descent method groups the tumor and the white matters together for many different initial conditions. But our method can obtain the desirable segmentation with all the different initial conditions that we tried (only one is shown in the figure).

2) Proliferating-cell Nuclear Antigen: The results are shown in Fig. 9. The objects to segment consist of clusters of small molecules — proliferating-cell nuclear antigen (PCNA) in this case. They are depicted in the original image as black dots around the intervillus pockets. The stochastic level set method has no difficulty in segmenting them out. They are depicted in black color in Fig. 9. The black phase obtained by the gradient descent method contains many features which are not PCNA and are actually of lighter colors in the original image.

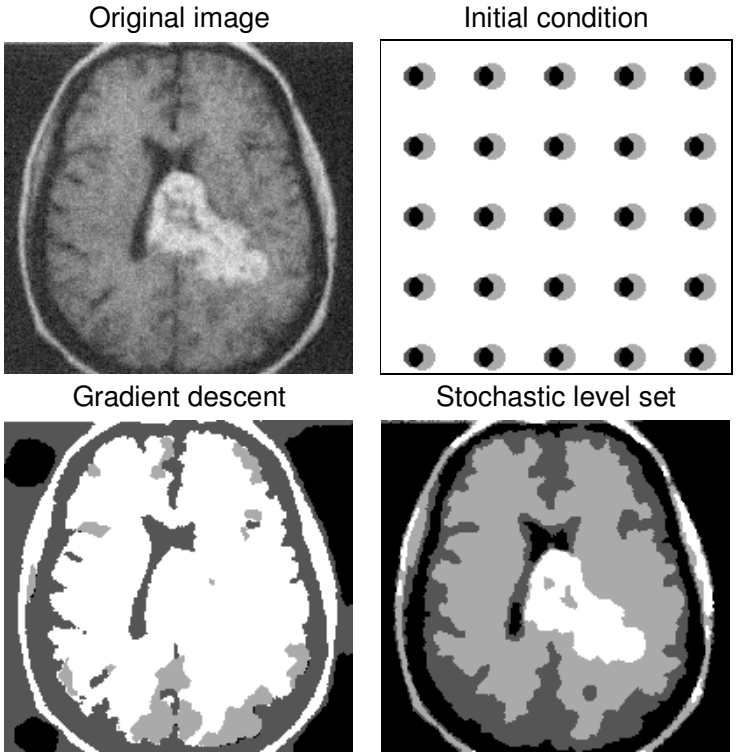

Fig. 8. Segmentation of the brain MRI image. (Top left) Original image with size 190×200. (Top right) Initial condition. (Bottom left) Gradient descent method with $\mu = 1000$ and 1000 iterations. The CPU time is 217 sec. (Bottom right) Stochastic level set method with $\mu = 1000$. The number of equivalent iterations at the original level is 64. The CPU time is 17 sec.

TABLE I

ENERGY OF THE SEGMENTATION RESULTS OBTAINED BY THE TWO METHODS. (IC STANDS FOR INITIAL CONDITION.)

Date set	Energy (Gradient descent)	Energy (Stochastic LS)	% improvement
Breast (IC 1)	3.83×10^{12}	2.24×10^{12}	42
Breast (IC 2)	3.81×10^{12}	2.24×10^{12}	41
Breast (IC 3)	2.90×10^{12}	2.24×10^{12}	22
Brain MRI	1.96×10^{7}	1.29×10^{7}	34
Zebrafish	2.06×10^{7}	1.46×10^{7}	29

Energy reduced about $20 \sim 40\%$.

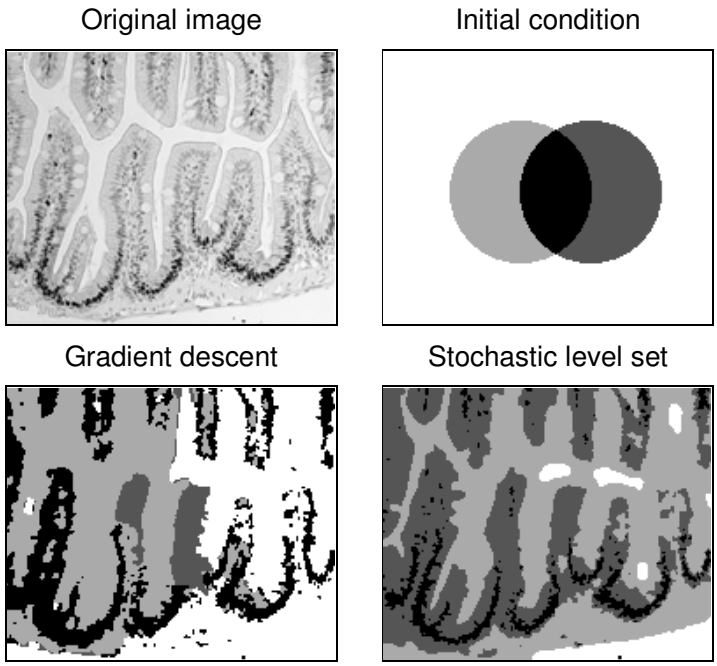

Fig. 9. Segmentation of the zebrafish image. (Top left) Original image with size 156×188. (Top right) Initial condition. (Bottom left) Gradient descent method with $\mu = 1000$ and 1000 iterations. The CPU time is 170 sec. (Bottom right) Stochastic level set method with $\mu = 1000$. The number of equivalent iterations at the original level is 109. The CPU time is 21 sec.

V. Discussion

In this paper, we propose a hybrid optimization method to solve the Mumford-Shah segmentation problem. Using numerous images with distinct features, we show that our method is very robust to the initial guess and outperforms standard deterministic and stochastic methods. It also obtains solutions of lower energy in much less computational resources. Our results suggest that coupling global and local updates is a promising approach to solving non-convex variational problems in image segmentation. Our intuition is that although the objective function is non-convex, its landscape is usually "not that complicated" either (at least for our examples). This is because of the regularization terms. However, designing an effective global updating scheme for a given problem can be non-trivial since the global landscape of the objective function is difficult to characterize. But for the Mumford-Shah model, splitting and merging based on the sum-of-differences-squared appear to be a good candidate. Finally, we remark that while we use gradient descent to carry out local optimization, it can be replaced by any other fast local optimization procedures, e.g. [21].

APPENDIX I
PROOF OF THEOREM 1

Let Ω_0 and Ω_1 be the image domain at the original level (level 0) and at one level higher (level 1) respectively. Let Ω_l^k for $k = 1, \ldots, n$ be the kth phase in Ω_l for $l = 0, 1$ defined by a segmentation Φ. Let

$$F_0(\mathbf{c}, \Phi) = h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\Omega_0^k} (u_{i,j} - c_k)^2 + \mu \text{Length}(\Phi)$$

be the objective function at level 0 and let

$$F_1(\mathbf{c}, \Phi) = (2h)^2 \sum_{k=1}^{n} \sum_{(i,j)\in\Omega_1^k} (v_{i,j} - c_k)^2 + \mu \text{Length}(\Phi)$$

be the objective function at level 1. Here h^2 and $(2h)^2$ are the area of a pixel in level 0 and level 1 respectively.

Let $\tilde{\Phi}$ be an arbitrary segmentation at level 1 and let $\tilde{\mathbf{c}}(\tilde{\Phi})$ be the optimal constants determined from $\tilde{\Phi}$. Let $\tilde{\Omega}_l^k$ be the kth phase in Ω_l defined by $\tilde{\Phi}$. Let $\bar{v} = (\bar{v}_{i,j})$ a constant extension of v from Ω_1 to Ω_0, i.e. it is an $M \times M$ image such that $\bar{v}_{i,j} = v_{\lceil i/2 \rceil, \lceil j/2 \rceil}$. Then we have

$$\begin{aligned}
F_0(\tilde{\mathbf{c}}(\tilde{\Phi}), \tilde{\Phi}) &= h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_0^k} (u_{i,j} - \tilde{c}_k)^2 + \mu \text{Length}(\tilde{\Phi}) \\
&= h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_0^k} (\bar{v}_{i,j} - \tilde{c}_k + u_{i,j} - \bar{v}_{i,j})^2 + \mu \text{Length}(\tilde{\Phi}) \\
&= h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_0^k} (\bar{v}_{i,j} - \tilde{c}_k)^2 + \mu \text{Length}(\tilde{\Phi}) + h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_0^k} (u_{i,j} - \bar{v}_{i,j})^2 \\
&\quad + 2h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_0^k} (\bar{v}_{i,j} - \tilde{c}_k)(u_{i,j} - \bar{v}_{i,j}) \\
&= (2h)^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_1^k} (v_{i,j} - \tilde{c}_k)^2 + \mu \text{Length}(\tilde{\Phi}) + \text{constant} \\
&\quad + 2h^2 \sum_{k=1}^{n} \sum_{(i,j)\in\tilde{\Omega}_1^k} (v_{i,j} - \tilde{c}_k) \sum_{p,q=0}^{1} (u_{2i-p,2j-q} - v_{i,j}) \\
&= F_1(\tilde{\mathbf{c}}(\tilde{\Phi}), \tilde{\Phi}) + \text{constant}.
\end{aligned}$$

Let $(\mathbf{c}_1, \Phi_1)$ be the optimal solution of F_1 at level 1. Then, we have

$$F_0(\tilde{\mathbf{c}}(\tilde{\Phi}), \tilde{\Phi}) = F_1(\tilde{\mathbf{c}}(\tilde{\Phi}), \tilde{\Phi}) + \text{constant} \geq F_1(\mathbf{c}(\Phi_1), \Phi_1) + \text{constant} = F_0(\mathbf{c}(\Phi_1), \Phi_1)$$

for any $\tilde{\Phi}$ defined on Ω_1. Thus, Φ_1 is also a minimizer of F_0 on Ω_1.

REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Prentice Hall, 2002.

[2] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet another survey on image segmentation: Region and boundary information integration," *Lecture Notes in Comp. Sci., Proc. of ECCV '02(III)*, vol. 2352, pp. 408–422, 2002.

[3] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, pp. 577–685, 1989.

[4] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Tran. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.

[5] 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*, vol. 50, no. 3, pp. 271–293, 2002.

[6] S. Osher and J. A. Sethian, "Fronts propogating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulation," *J. Comput. Phys.*, vol. 79, pp. 12–49, 1988.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[8] R. B. Kearfott, *Rigorous Global Search: Continuous Problems*, ser. Nonconvex Optimization and Its Applications. Dordrecht: Kluwer Academic Publishers, 1996, vol. 13.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[10] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[11] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letters*, vol. 81, pp. 2050–2053, 2001.

[12] H. K. Lee, Y. Okabe, and D. P. Landau, "Convergence and refinement of the Wang-Landau algorithm," *Computer Physics Communications*, vol. 175, pp. 36–40, 2006.

[13] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.

[14] H. K. Lee and Y. Okabe, "Reweighting for nonequilibrium Markov processes using sequential importance sampling methods," *Physics Review E*, vol. 71, p. 015102(R), 2005.

[15] J.-S. Wang and R. H. Swendsen, "Cluster Monte Carlo algorithms," *Physica A*, vol. 167, pp. 565–579, 1990.

[16] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structure of Lennard-Jones clusters containing up to 110 atoms," *Journal of Physical Chemistry A*, vol. 101, pp. 5111–5116, 1997.

[17] A. Crouzil, X. Descombes, and J.-D. Durou, "A multiresolution approach for shape from shading coupling deterministic and stochastic optimization," *IEEE Tran. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1416–1421, 2003.

[18] G. Koepfler, C. Lopez, and J. M. Morel, "A multiscale algorithm for image segmentation by variational methods," *SIAM J. Numer. Anal.*, vol. 31, no. 1, pp. 282–299, 1994.

[19] B. Song and T. F. Chan, "A fast algorithm for level set based optimization," University of California, Los Angeles, CAM Report 02-68, 2002.

[20] T. F. Chan and S. Esedoglu, "A multiscale algorithm for Mumford-Shah image segmentation," University of California, Los Angeles, CAM Report 03-57, 2003.

[21] S. Esedoglu and Y. H. Tsai, "Threshold dynamics for the piecewise constant Mumford-Shah functional," *J. Comput. Phys.*, vol. 211, no. 1, pp. 367–384, 2006.

[22] T. F. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of denoising and segmentation models," *SIAM J. Appl. Math.*, vol. 66, pp. 1632–1648, 2006.

[23] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape," *International Journal of Computer Vision*, vol. 72, no. 2, pp. 195–215, 2007.

[24] J. Shen, "A stochastic-variational model for soft Mumford-Shah segmentation," *International Journal of Biomedical Imaging*, vol. 2006, Article ID 92329, 2006.

[25] O. Juan, R. Keriven, and G. Postelnicu, "Stochastic motion and the level set method in computer vision: Stochastic active contours," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 7–25, 2006.

[26] T. W. Ridler and E. S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Tran. Syst. Man Cybern.*, vol. SMC-8, pp. 630–632, 1978.

[27] *http://www.umassmed.edu/bmp/faculty/ross.cfm*.

[28] A.-P. G. Haramis, A. Hurlstone, Y. van der Velden, H. Begthel, M. van den Born, G. J. A. Offerhaus, and H. C. Clevers, "Adenomatous polyposis coli-deficient zebrafish are suspectible to digestive tract neoplasia," *EMBO Reports*, vol. 7, no. 4, pp. 444–449, 2006.