

SALIENT LEVEL LINES SELECTION USING THE MUMFORD-SHAH FUNCTIONAL

Yongchao Xu^{1,2}, Thierry Géraud^{1,2}

Laurent Najman²

¹ EPITA Research and Development Laboratory
(LRDE)
14-16, rue Voltaire
FR-94276 Le Kremlin-Bicêtre, France

² Université Paris-Est,
Laboratoire d'Informatique Gaspard-Monge (LIGM),
A3SI, ESIEE Paris, Cité Descartes, BP 99
FR-93162 Noisy-le-Grand, France

ABSTRACT

Many methods relying on the morphological notion of *shapes*, (i.e., connected components of level sets) have been proved to be very useful for pattern analysis and recognition. Selecting meaningful level lines (boundaries of level sets) yields to simplify images while preserving salient structures. Many image simplification and/or segmentation methods are driven by the optimization of an energy functional, for instance the Mumford-Shah functional. In this article, we propose an efficient shape-based morphological filtering that very quickly compute to a locally (subordinated to the tree of shapes) optimal solution of the piecewise-constant Mumford-Shah functional. Experimental results demonstrate the efficiency, usefulness, and robustness of our method, when applied to image simplification, pre-segmentation, and detection of affine regions with viewpoint changes.

Index Terms— Level lines, Tree of shapes, Energy minimization, Pre-segmentation, Morphological shaping.

1. INTRODUCTION

In natural images, meaningful contours are usually smooth and well-contrasted. Recently, many authors claim that significant contours of objects in images coincide with segments of the image level lines [1]. Each connected level line is the contour of a *level set*, or *shape*, a connected set of pixels without holes. The inclusion relationship of level sets allows for representing an image by a tree, called a *tree of shapes* [2], which is invariant to contrast changes. Image simplification or segmentation can then be defined by selecting meaningful level lines in that tree. That subject has been investigated in the past ten years by [3, 4, 5, 6]. In [7] Lu et al. propose also a tree simplification method for image simplification purpose using the binary partition tree [8] and a knee function.

Following the seminal work of Mumford and Shah [9], finding relevant contours is often tackled thanks to an energy-based approach, as a compromise between some image-driven force (image contrast along contours, data fidelity, etc.) and the regularity of contours. Minimizing the Mumford-Shah functional tends to find a simplified or segmented image into regions. Curve evolution methods [10, 11] are usually

used to solve such an energy minimization problem. They have solid theoretical foundations, yet they are often computationally expensive.

In this paper we propose to formalize the piecewise-constant Mumford-shah functional on an image, subordinated to the tree of shapes of this image. The selection of the salient level lines corresponds to a meaningful locally optimal solution of the energy minimization problem. The main contribution is the proposition of an efficient greedy algorithm which takes into account the meaningfulness of the set of level lines. Simply put, a level line is easier to remove when it has a low degree of meaningfulness and when it favors a great decreasing of energy. Our algorithm drives very fast to a relevant local optimum in the sense that no more level lines can be removed while decreasing energy. The reason why we claim that we reach a relevant optimum is that meaningful level lines are hard to be removed during the proposed process. Note that our method actually belongs to the class of *morphological shapings* described in [12].

In [13], the authors proposed an efficient greedy algorithm to minimize the Mumford-Shah functional on a certain hierarchy, which leads to a global optimal segmentation on that hierarchy. In [14], the authors gave a detailed review of the tree (including the tree of shapes) filtering strategies. The works in [15] and [16] are closest to what we propose here. They both select meaningful level lines for image simplification and segmentation purpose using the piecewise-constant Mumford-Shah functional. In [15] the whole image domain is initially considered as a single region; level lines of the tree of shapes are browsed from root to leaves and are successively removed until the functional cannot decrease anymore. However, this top-down decision is based upon a non-significant energy variation since it is computed from the very few pixels lying between a shape and their immediate sub-shapes. Actually, our work is related to the one described in [16], where at each removal step, the level line which decreases the most the functional is selected. As a consequence, the iterative process of [16] requires not only to compute a lot of information to be able to update the functional value after each level line suppression, but also to find at each step, among all remaining level lines, the one candidate to the next removal. Hence [16] is computationally expensive, while what we propose here is

fast.

The rest of this paper is organized as follows. Some background information about the Mumford-Shah functional and the tree of shapes is provided in Section 2. Our proposed method is detailed in Section 3. In Section 4, we present some experimental results. We then conclude and give some perspectives in Section 5.

2. BACKGROUND

2.1. The Mumford-Shah Functional

According to the Mumford-Shah model [9], an image $f : \Omega \rightarrow \mathbb{R}$ or \mathbb{Z} is modeled as a piecewise-smooth function. A segmentation of f is defined as a pair (R, \tilde{f}) , where $R = R_1 \sqcup \dots \sqcup R_n$, each region R_i is a connected component in the image domain and $\forall i \neq j, R_i \cap R_j = \emptyset$, \tilde{f} is a regular function within each R_i . Let ∂R be the union of boundaries of the set of regions $\{R_i\}$, then $R \sqcup \partial R$ is the whole image domain Ω . The segmentation is given by the minimization of the Mumford-Shah functional $E(f, \partial R)$ defined by:

$$E(f, \partial R) = \iint_R (\tilde{f} - f)^2 dx dy + \mu \iint_R \|\nabla \tilde{f}\| dx dy + \nu |\partial R|, \quad (1)$$

where $|\partial R|$ represents the total length of the boundaries ∂R , and μ and ν are two positive parameters.

A special case of Eq. (1) is obtained by restricting the segmented image \tilde{f} to piecewise-constant functions, i.e., $\tilde{f} = \tilde{f}_i = \frac{1}{|R_i|} \iint f dx dy$ inside each region R_i , where $|R_i|$ denotes the area of the region R_i . Then, the simplified Mumford-Shah functional is given by:

$$E(f, \partial R) = \iint_R (\tilde{f}_i - f)^2 dx dy + \nu |\partial R|. \quad (2)$$

Since it is usually difficult to minimize the energy functional of Eq. (1), the simplified Mumford-Shah functional given by Eq. (2) is used instead.

2.2. The Tree of Shapes

For any $\lambda \in \mathbb{R}$ or \mathbb{Z} , the upper level sets \mathcal{X}_λ and lower level sets \mathcal{X}^λ of an image f are respectively defined by $\mathcal{X}_\lambda(f) = \{p \in \Omega \mid f(p) \geq \lambda\}$ and $\mathcal{X}^\lambda(f) = \{p \in \Omega \mid f(p) \leq \lambda\}$. Both upper and lower level sets have a natural inclusion structure: $\forall \lambda_1 \leq \lambda_2, \mathcal{X}_{\lambda_1} \supseteq \mathcal{X}_{\lambda_2}$ and $\mathcal{X}^{\lambda_1} \subseteq \mathcal{X}^{\lambda_2}$, which leads to two distinct and dual representations of an image, the max-tree and the min-tree [17].

Another tree has been introduced in [2]. A *shape* is defined as a connected component of an upper or lower level set where its holes have been filled in. Thanks to the inclusion relationship of both kinds of level sets, the set of shapes gives a unique tree, called *tree of shapes*. This tree features an interesting property: it is invariant to contrast changes. To put it differently, it is a self-dual, non-redundant, and complete representation of an image. Furthermore, such a tree inherently embeds a morphological scale-space (the parent of a

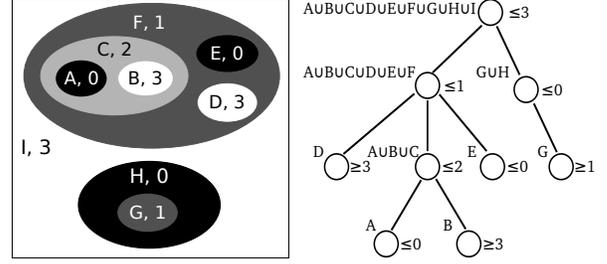


Fig. 1. An image (left) and its tree of shapes (right).

node/shape is a larger shape). An example on a simple image is depicted in Fig. 1.

3. PROPOSED METHOD

3.1. About Energy Minimization

The basic idea in our proposed method is to minimize the simplified Mumford-Shah functional in Eq. (2) subordinated to the tree of shapes. The segmentation is restricted to the regions whose boundaries are level lines. Although this is not the general context for a segmentation, since the meaningful boundaries of objects may not coincide with full level lines, but segments of them [1]. However, the main edges of the image are included in them, and the level lines are contrast invariant and robust. So still this is very useful as an image simplification or pre-segmentation tool.

More specifically, for a given tree of shapes \mathcal{T} composed of a set of shapes $\{\tau_i\}$, any two successive shapes of \mathcal{T} are related by an edge reflecting the inclusion relationship, also known as the parenthesis between nodes of the tree. Let $\partial\mathcal{T} = \{\partial\tau \mid \tau \in \mathcal{T}\}$ be the union of boundaries of all the shapes of \mathcal{T} . The minimization of the energy functional of Eq. (2) restricted to \mathcal{T} is given by:

$$\min_{\mathcal{T}'} E(f, \partial\mathcal{T}'), \quad (3)$$

where \mathcal{T}' is a simplified version of \mathcal{T} by removing some shapes from \mathcal{T} and by updating the parenthesis relationship between the shapes of \mathcal{T}' .

The basic operation of the energy minimization problem of Eq. (3) is the merging of the element “regions” formed by the level lines $\partial\mathcal{T}$. For each given shape τ , the induced element “region” R_τ is defined as $R_\tau = \{p \mid p \in \tau, p \notin C(\tau)\}$, where $C(\tau)$ denotes all the children of the shape τ . Note that R_τ is not always a spatial connected component due to the fact that two successive level lines may have some part in common, this may cause a spatial separation of those pixels within R_τ . An element “region” R_τ is merged with the element “region” R_{τ_p} induced by its parent τ_p . This merging process can also be seen as a level line suppression, which will cause the update of the “region” $R'_{\tau_p} = R_{\tau_p} \cup R_\tau$. The parenthesis relationship for its children $\tau_{c1}, \dots, \tau_{ck}$ should

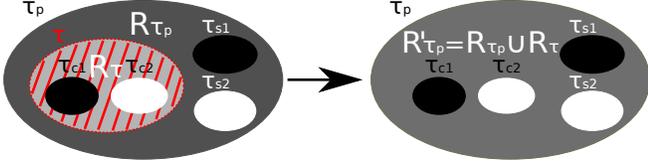


Fig. 2. Suppressing the node τ makes the “region” R_τ (covered with red oblique lines) merge with R_{τ_p} ; the result (depicted in the right image) is a simplified image.

also be updated to τ_p . Fig. 2 shows an example of a such merging operation.

Observe that the minimization problem of Eq. (3) is a combinatorial optimization. The computation of the optimum has an exponential complexity. Hence we propose to use the greedy algorithm to compute a local optimum. The removability of a level line $\partial\tau$ is decided by the sign of the simplified functional of Eq. (2) variation ΔE_τ while τ is suppressed. Let $S(f, R)$ be the sum of value of all the pixels inside region R . Then the functional variation ΔE_τ is given by

$$\Delta E_\tau = \frac{S^2(f, R_\tau)}{|R_\tau|} + \frac{S^2(f, R_{\tau_p})}{|R_{\tau_p}|} - \frac{S^2(f, R'_{\tau_p})}{|R'_{\tau_p}|} - \nu |\partial\tau|. \quad (4)$$

If ΔE_τ is negative, which means the suppression of τ decreases the functional, then remove τ . According to Eq. (4), the removability of a shape τ depends only on R_τ and R_{τ_p} . As a shape τ suppression triggers the update of R_{τ_p} , the removal of τ impacts also the removability of its relatives, i.e., its parent, its children and siblings. So the order of level line removal is critical.

The work in [16] bears a heavy computational update of ΔE for the nodes that are relatives of the suppressed level line. In addition, after each update, it also requires to find the new level line with the highest negative functional variation ΔE . We propose to fix that issue thanks to a reasonable ordering of level lines based on their meaningfulness. Our proposed algorithm is described in the following Section.

3.2. Proposed Algorithm

Initialization: The output tree, \mathcal{T}' , will be incrementally simplified. First, it is set to $\mathcal{T}' = \mathcal{T}$.

Step 1: Sort the set of shapes $\{\tau | \tau \in \mathcal{T}\}$ in the increasing order \mathcal{O} of shape meaningfulness.

Step 2: Propagate the shapes in the order \mathcal{O} , for each shape $\tau \in \mathcal{T}'$, compute ΔE_τ . If it is negative, remove τ from \mathcal{T}' , update R'_{τ_p} and update the parenthood relationship for its children on \mathcal{T}' .

Step 2 may need to be iterated until no shape can be removed. In practice, 2 or 3 propagations are enough, and most of the level lines are removed in the first propagation. In consequence, the complexity of the minimization problem of



(a) Input image. (b) Ballester, $\nu = 1k$. (c) Our, $\nu = 1k$.

Fig. 3. Comparison of our approach with the one of Ballester et al.. (a): $E = 1.2433e + 08$; (b): $E = 1.28113e + 07$, PSNR = 27.3; (c): $E = 1.15285e + 07$, PSNR = 28.7.

Eq. (3) is linear w.r.t. the number of shapes of \mathcal{T} . The image reconstructed from the last tree \mathcal{T}' is a locally optimal solution of Eq. (3) in the sense that any more level line removal will increase the simplified Mumford-Shah functional.

Instead of fixing the parameter ν in Eq. (4), a variant of the algorithm is to compute a $\nu_{\min}(\tau)$ for each level line $\partial\tau$ when $\Delta E_\tau = 0$. Then, for any $\nu > \nu_{\min}(\tau)$, the level line $\partial\tau$ will be removed by solving the minimization problem of Eq. (3). Consequently, by using this ν_{\min} as an attribute function characterizing each shape τ , the proposed method can be seen as one of the many variant of morphological shapings [12].

4. EXPERIMENTAL RESULTS

For all experiments, the shape meaningfulness we use is the average of gradient magnitude along the level lines. We have experienced with some other measurements, including a random order between the shapes. All the experiments show the usefulness of the sorting step based on level lines meaningfulness.

In Fig. 3, we compare our proposed method with the one of Ballester et al. on a classic image “house” (256×256) having originally 23578 level lines. Qualitatively, the image (b) and (c) obtained by the two methods using the same parameter $\nu = 1000$ are very close. The result (Fig. 3 (b)) given by the method of Ballester retains 30 level lines, and 27 level lines are selected by our approach. In spite of their strong simplifications, the salient structures are preserved in the simplified images. The distinct final functional minimized by the two approaches confirms that they fall into different local minimum. The CPU time for the Ballester method that we implemented and our approach are respectively 4s and 0.2s on a regular PC station.

Fig. 4 shows a quantitative evolution of our method w.r.t. increasing noise. The image “house” is corrupted with an additive Gaussian noise of variance $\delta^2 = 5, 10, 15, 20, 25$. All the results are obtained using the same parameter $\nu = 400$. Qualitatively, the simplification results from (a) to (f) are very close. They indeed select almost the same amount of level lines (around 32), and most of them coincide with the same salient structures in the image. Furthermore, the similar PSNR of each result also confirms the robustness of

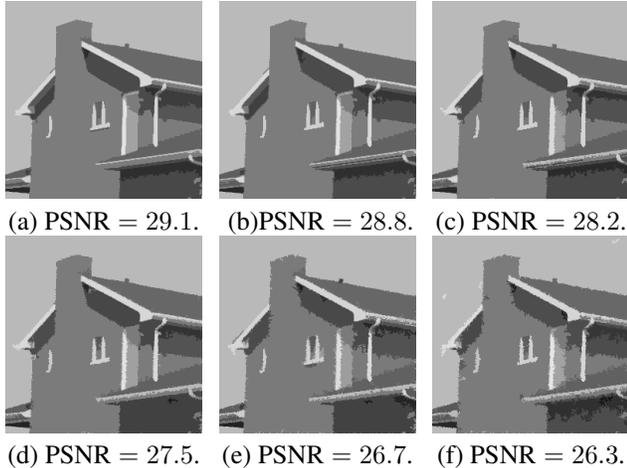


Fig. 4. Qualitative behavior of our method w.r.t. increasing noise. (a): Result applied to original image; (b-f): Results applied to images with additive Gaussian noise of variance going from 5 (b) to 25 (f) with a step of 5.

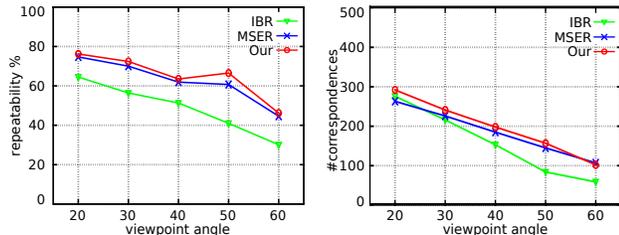


Fig. 5. Repeatability score (left) and number of correspondences (right) for the “Graffiti” sequence with increasing viewpoint angles (x-axis); Our method outperforms IBR and MSER.

the proposed method w.r.t. noise.

We also test the robustness of our proposed method w.r.t. the viewpoint changes, which is the most difficult type of transformation to cope with for affine region detectors [18]. In this experiment, all the connected components of the upper and lower level sets of the simplified image obtained with our method, are detected as affine regions. Fig. 5 shows the quantitative evolution of those regions applied to the “Graffiti” sequence available at [19]. In general, a high repeatability and a large number of correspondences are desired. Although the proposed method of salient level lines selection is not dedicated to general affine regions detection, yet the result in Fig. 5 (red) compared to two approaches (maximally stable extremal region (MSER) [20] and intensity extrema-based region detector (IBR) [21]) bearing some similarity, reflects that the proposed salient level lines selection method is quite robust w.r.t. viewpoint changes. In fact, the better result is obtained thanks to the high detection accuracy on those regions given by the perfectly preserved salient level lines.

In Fig. 6, we test our proposed method to color images from the Berkeley Segmentation Dataset [22]. The strategy is

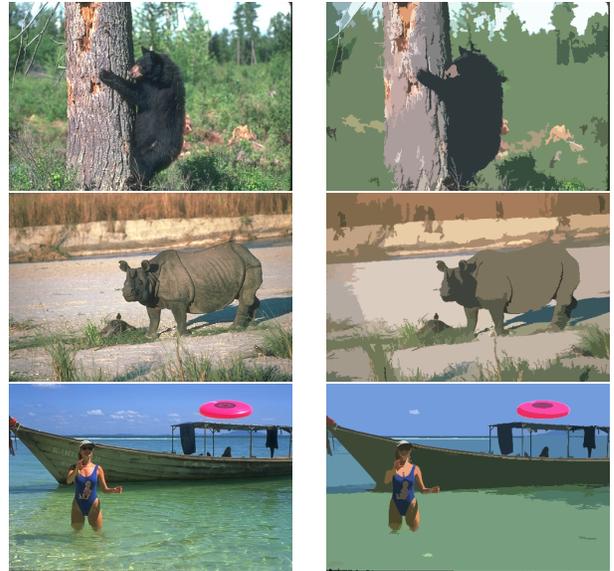


Fig. 6. Some pre-segmentation results obtained with our proposed method on the Berkeley Segmentation Dataset [22].

to apply individually the proposed method to the red, green, and blue channels. In order to obtain a pre-segmentation, a high parameter value $\nu = 2500$ is used, and a grain filter [2] is applied to get rid of too tiny shapes. Less than 50 level lines are selected for each channel of each image, which results in a ratio of level lines selection around 2300. The regions formed by the union of those selected salient level lines from the 3 channels are considered as the pre-segmented regions. The images shown in Fig. 6 are obtained by taking the average color inside each region, where the boundaries between salient regions remain intact. Finding an actual segmentation becomes a lot easier with such a pre-segmentation.

5. CONCLUSION AND PERSPECTIVES

In this paper, we presented an efficient morphological shaping to salient level lines selection, based on the minimization of the piecewise-constant Mumford-Shah functional. Our major contribution is to rely on a *meaningful* ordering of level lines in order to minimize this energy functional on the tree of shapes. As a consequence, the proposed greedy algorithm converges to a relevant local optimum very quickly compared with the similar work of Ballester et al.. We have shown that the proposed method allows for strongly simplifying images while preserving their salient structures. We have seen that a strong property of our proposal is its robustness to noise and to viewpoint changes. Furthermore simplification results can be used as pre-segmentations that are suitable for object recognition, scene analysis, or practical shape matching [23]. The authors are currently investigating some applications of the proposed simplification method. In addition, a major perspective of this work is to rely on shape-based morphology [12] to make this method hierarchical.

6. REFERENCES

- [1] V. Caselles, B. Coll, and J. Morel, “Topographic maps and local contrast changes in natural images,” *Intl. Journal of Computer Vision*, vol. 33, no. 1, pp. 5–27, 1999.
- [2] P. Monasse and F. Guichard, “Fast computation of a contrast-invariant image representation,” *IEEE Trans. on Image Processing*, vol. 9, no. 5, pp. 860–872, 2000.
- [3] A. Pardo, “Semantic image segmentation using morphological tools,” in *Proc. of ICIP*, 2002, pp. 745–748.
- [4] F. Cao, P. Musé, and F. Sur, “Extracting meaningful curves from images,” *JMIV*, vol. 22, pp. 159–181, 2005.
- [5] J. Cardelino, G. Randall, M. Bertalmio, and V. Caselles, “Region based segmentation using the tree of shapes,” in *Proc. of ICIP*, 2006, pp. 2421–2424.
- [6] Y. Xu, T. Géraud, and L. Najman, “Context-based energy estimator: Application to object segmentation on the tree of shapes,” in *ICIP*, 2012, pp. 1577–1580.
- [7] H. Lu, J.C. Woods, and M. Ghanbari, “Binary partition tree analysis based on region evolution and its application to tree simplification,” *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1131–1138, 2007.
- [8] P. Salembier and L. Garrido, “Binary partition tree as an efficient representation for image processing, segmentation and information retrieval,” *IEEE Trans. on Image Processing*, vol. 9, no. 4, pp. 561–576, 2000.
- [9] D. Mumford and J. Shah, “Optimal approximations by piecewise smooth functions and associated variational problems,” *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
- [10] J.A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, 1999.
- [11] T.F. Chan and L.A. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [12] Y. Xu, T. Géraud, and L. Najman, “Morphological Filtering in Shape Spaces: Applications using Tree-Based Image Representations,” in *ICPR*, 2012, pp. 485–488.
- [13] L. Guigues, J. P. Cocquerez, and H. L. Men, “Scale-sets image analysis,” *International Journal of Computer Vision*, vol. 68, no. 3, pp. 289–317, 2006.
- [14] P. Salembier and M. H. F. Wilkinson, “Connected operators,” *IEEE Signal Processing Mag.*, vol. 26, no. 6, pp. 136–157, 2009.
- [15] Y. Pan, “Top-down image segmentation using the mumford-shah functional and level set image representation,” in *Proc. of ICASSP*, 2009, pp. 1241–1244.
- [16] C. Ballester, V. Caselles, L. Igual, and L. Garrido, “Level lines selection with variational models for segmentation and encoding,” *Journal of Mathematical Imaging and Vision*, vol. 27, pp. 5–27, 2007.
- [17] P. Salembier, A. Oliveras, and L. Garrido, “Antiextensive connected operators for image and sequence processing,” *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570, 1998.
- [18] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, “A comparison of affine region detectors,” *Intl. J. Comput. Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.
- [19] “Affine covariant region detectors,” <http://www.robots.ox.ac.uk/~vgg/research/affine/>.
- [20] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” *BMVC*, vol. 22, no. 10, pp. 384–393, 2002.
- [21] T. Tuytelaars and L.V. Gool, “Matching widely separated views based on affine invariant regions,” *Intl. J. Comput. Vision*, vol. 59, no. 1, pp. 61–85, 2004.
- [22] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE PAMI*, vol. 33, no. 5, pp. 898–916, 2011.
- [23] J. L. Lisani, L. Moisan, P. Monasse, and J. M. Morel, “On the theory of planar shape,” *SIAM Multiscale Modeling and Simulation*, vol. 1, no. 1, pp. 1–24, 2003.