

L1 PATCH-BASED IMAGE PARTITIONING INTO HOMOGENEOUS TEXTURED REGIONS

M. Oliver, G. Haro, V. Fedorov, C. Ballester

DTIC – Universitat Pompeu Fabra

ABSTRACT

This paper proposes a novel patch-based variational segmentation method that considers adaptive patches to characterize, in an affine invariant way, the local structure of each homogeneous texture region of the image and thus being capable of grouping the same kind of texture regardless of differences in the point of view or suffered perspective distortion. The patches are computed using an affine covariant structure tensor defined at every pixel of the image domain, so that they can automatically adapt its shape and size. They are used in a segmentation model that uses an L^1 -norm fidelity term and fuzzy membership functions, which is solved by an alternating scheme. The output of the method is a partition of the image in regions with homogeneous texture together with a patch representative of the texture of each region.

Index Terms— Image segmentation, variational methods, L1-fidelity, patch-based methods, affine invariant patch similarity.

1. INTRODUCTION

Image simplification (or segmentation) is one of the central problems in image analysis and computer vision. The goal is to partition the image into regions which share common features – such as color, intensity, texture, or depth – while at the same time locate the most regular and accurate contours that define the sharp boundaries of these regions. Often, a representative feature of every region is also extracted; this information can be used, for example, for image cartooning, or image interpretation.

In the literature on image segmentation, variational approaches are among the most popular [26, 23, 28, 10, 24, 8]. From these references, it is now well known that a good segmentation can be obtained by minimizing an appropriate energy functional. The Mumford-Shah functional is one of the most popular with this underlying variational criterion (see, e.g., [18] and references therein). Let us briefly recall that Mumford and Shah [19] proposed a joint smoothing and edge detection energy formulation for the segmentation problem.

By minimizing that energy, one tries to approximate the input image u , defined on an image domain Ω , by a piecewise smooth function v , similar to u in each segment of $\Omega \setminus B$ and, at the same time, to reduce the complexity of its discontinuity set B . The L^2 norm of the error $v - u$ is used to measure the similarity between v and u . Due to the theoretical and numerical complexity of the Mumford-Shah functional, a simplification of the previous problem has been used where v is considered piecewise constant in $\Omega \setminus B$.

As the gray level or color value of a single pixel is neither discriminative nor robust enough to be used for computing the similarity between u and v , specially for natural textured images, the use of patches is a common practice for establishing image similarities and correspondences in different image processing and computer vision applications (such as e.g. denoising or stereo matching among many others [20, 1, 13]). Traditionally, these patches have been defined as squared or circular windows. One of the main problems with these windows is that if the center of the patch is close to an object boundary the patch contains mixed information from different objects. Simple improvements are bilateral weights [25] or adaptive patches that try to follow the local geometry of the image [6, 9, 30]. However, patches of fixed size have two main problems: (1) poor discrimination power when applied to textural structures not observable within the size of the neighborhood because of the wrong scale selected, and (2) lack of robustness to transformations of the local texture due to perspective transformations or changes in the point of view. In contrast, the patches proposed in [7] – following the idea that there is a relation between dominant orientation and an appropriate scale and adaptive neighborhoods [16, 2] – are ellipses on the image domain that automatically adapt their size and orientation to the local structure of the image. These kind of patches, in combination with the affine invariant similarity measure introduced in [7] allow to identify similar local image patterns that have suffered different affine transformations.

The focus of this work is to use the L^1 version of this patch-based affine invariant similarity measure in a Mumford-Shah-based segmentation functional to partition the image into regions that share the same local structure up to an affine distortion. Thanks to the use of the L^1 norm in our fidelity term, the proposed model also provides the representative sharp texture for each region, which consists of a patch con-

All authors acknowledge partial support by MINECO/FEDER UE project, reference TIN2015-70410-C2-1-R and by GRC reference 2014 SGR 1301, Generalitat de Catalunya.

taining contrast preserving texture as a result of a weighted median vector process. As a further consequence of the use of the L^1 norm it is robust to impulse noise and outliers. The L^1 norm was also used for segmentation purposes in [15, 12]. As in our proposal, [15] also uses fuzzy membership functions [29] and the TV norm to estimate the length of the segmentation boundaries. On the other hand, the classical structure tensor has been used in the literature for texture segmentation purposes [21, 22, 11]. but these methods are not robust to affine transformations in the texture. In this paper, we propose an L^1 -norm variational model that uses fuzzy membership functions and outputs a partition of the input image in regions with homogeneous texture regardless of affine distortions together with a representative patch containing the normalized texture of each region.

The outline of this paper is as follows. Sect. 2 details our variational proposal. In Sect. 3, the optimization method is presented while the experimental results are presented in Sect. 4. Finally, the paper is concluded in Sect. 5.

2. THE MODEL

Let $u : \Omega \rightarrow \mathbb{R}^M$ be a given image defined on $\Omega \subset \mathbb{R}^2$ with values in \mathbb{R}^M , where $M = 1$ for gray level images and $M = 3$ for standard color images. Our aim is to characterize the regions of the image domain Ω having homogeneous texture regardless of differences in the point of view or suffered perspective distortion.

To this goal, we consider a spatially varying Riemannian metric defined on the image domain Ω that captures these distortions. In particular, let $T_u(x)$ be the metric given, in matricial form, by the affine covariant tensor associated to u proposed in [7]. It guarantees affine covariant regions by associating an appropriate adaptive patch to each point $x \in \Omega$ which is defined as follows. Let $\mathbf{p}_u : \Omega \rightarrow L^q(\Delta_t)$ denote the function given by $\mathbf{p}_u(x) := \mathbf{p}_u(x, \cdot)$, where $\mathbf{p}_u(x, h) := u(x + T_u(x)^{-\frac{1}{2}}h)$, and $h \in \Delta_t$, a disc centered at the origin with radius proportional to t . Figure 1 displays an image and some of its patches defined by the affine covariant tensors $T_u(x)$, for some pixels x .

To compare two patches $\mathbf{p}_u(x), \mathbf{p}_u(y)$ in an appropriate manner, we propose to use the multiscale affine invariant patch-similarity measure of [7], that automatically transforms the patches in comparison:

$$\mathcal{D}_t^{a,q}(\mathbf{p}_u(x), \mathbf{p}_u(y)) = \int_{\Delta_t} g_t(h) \left\| u(x + T_u(x)^{-\frac{1}{2}}h) - u(y + T_u(y)^{-\frac{1}{2}}h) \right\|_{L^q}^q dh, \quad (1)$$

where $q > 0$, $\|\cdot\|_{L^q}$ denotes the norm in L^q , $t > 0$ represents the scale of the patch and allows to control the support in the patch comparison, g_t is a geodesic weighting function having effective support in Δ_t . In the following, we will consider $q = 1$.



Fig. 1: Left, in red, the adaptive patch associated to the affine invariant structure tensor for some pixels of the image. Right, some of the patches transformed into discs.

In order to define our segmentation model, let \mathcal{P}_u be the set of all patches obtained from image u and defined using the affine covariant tensor metric $T_u(x)$ associated to u . That is,

$$\mathcal{P}_u = \{\mathbf{p}_u(x), \quad x \in \Omega\}. \quad (2)$$

Let us notice that, thanks to the tensors, these elliptical patches can be considered defined on the *normalized disc* Δ_t . Figure 1 illustrates, on the right, four normalized discs corresponding to four patches of the image.

We propose to simplify the set of all patches \mathcal{P}_u by estimating an optimal finite set of representative patches $\{\mathbf{p}_{\Omega_1}, \dots, \mathbf{p}_{\Omega_N}\}$, where $\Omega = \cup_{i=1}^N \Omega_i$ is a partition of the image domain into N disjoint open regions Ω_i , for $N \in \mathbb{N}$, such that each region contains the pixels with similar patches and \mathbf{p}_{Ω_i} is the patch associated to the region Ω_i . In other words, Ω_i contains all the pixels with local homogeneous texture regardless of differences in the point of view or suffered perspective distortion. We propose to do it by minimizing the following energy:

$$E(\mathbf{p}, B) = \ell(B) + \lambda \int_{\Omega} \mathcal{D}_t^{a,1}(\mathbf{p}(x), \mathbf{p}_u(x)) dx, \quad (3)$$

where $\mathbf{p} = \sum_i \mathbf{p}_{\Omega_i} \chi_{\Omega_i}$ is a piecewise constant patch function, i. e., it associates a homogeneous texture to each point x of a connected component Ω_i of $\Omega \setminus B$, and $B = \cup_{i=1}^N \partial\Omega_i$, being $\partial\Omega_i$ and χ_{Ω_i} the boundary and the characteristic function of Ω_i , respectively. By analogy and by an abuse of notation we have denoted by $\mathcal{D}_t^{a,1}(\mathbf{p}(x), \mathbf{p}_u(x))$ the patch similarity $\int_{\Delta_t} g_t(h) \left\| \mathbf{p}_{\Omega_i}(h) - u(x + T_u(x)^{-\frac{1}{2}}h) \right\|_{L^1} dh$, for $x \in \Omega_i$. Observe that our choice of $q = 1$ allows to obtain pure representative patches for each region.

As $\Omega \setminus B = \cup_{i=1}^N \Omega_i$, with $\Omega_i \cap \Omega_j = \emptyset$ for all $i \neq j$, we can rewrite (3) taking advantage of the fact that the functions

χ_{Ω_i} are in $BV(\Omega; \{0, 1\})^1$ as follows

$$E(\mathbf{p}, \chi) = \sum_{i=1}^N \left[\int_{\Omega} |\nabla \chi_{\Omega_i}(x)| dx + \lambda \int_{\Omega} \int_{\Delta_t} g_t(h) \left\| \mathbf{p}_{\Omega_i}(h) - u(x + T_u^{-\frac{1}{2}}(x)h) \right\|_{L^1} \chi_{\Omega_i}(x) dh dx \right], \quad (4)$$

where $\chi = (\chi_{\Omega_1}, \dots, \chi_{\Omega_N})$ such that $\sum_i \chi_i(x) = 1, \forall x \in \Omega$. Again, $\mathbf{p} = \sum_i \mathbf{p}_{\Omega_i} \chi_{\Omega_i}$ is piecewise constant, i.e., a unique patch \mathbf{p}_{Ω_i} associated to each region Ω_i . Note that the Total Variation (TV) of χ_{Ω_i} in the first term is equal to the length of the boundary $\partial\Omega_i$. Energy (4) measures both the smoothness of the segmentation boundaries and the fidelity of the approximating piecewise patch function \mathbf{p} to the manifold of patches \mathcal{P}_u of the input image u . The output of the method is a partition of the image in regions with homogeneous texture together with a patch representative of the texture of each region (see Figure 2).

The variational model (4) is defined using characteristic functions $\chi_{\Omega_i} \in BV(\Omega; \{0, 1\})$, and the constraint $\sum_i \chi_i(x) = 1$ for each pixel $x \in \Omega$ implies that each pixel only belongs to a unique region Ω_i . But the set $BV(\Omega; \{0, 1\})$ is not convex and, moreover, the Euler-Lagrange equations for non-continuous non differentiable functions lead to difficulties in numerical implementations. Thus, following the idea proposed in [15] we relax the characteristic functions to be fuzzy membership functions [29] belonging to the set

$$\mathcal{C} = \left\{ (\omega_1, \dots, \omega_N) \mid \omega_i \in BV(\Omega; [0, 1]), 0 \leq \omega_i(x) \leq 1, \sum_{i=1}^N \omega_i(x) = 1, \forall x \in \Omega \right\}. \quad (5)$$

Now, $\omega_i(x)$ describes the fuzzy membership of a pixel x that may well belong simultaneously to more than one region; in other words, $\omega_i(x)$ can be understood as the probability that x belongs to the region Ω_i . Let ω be $\omega = (\omega_1, \dots, \omega_N)$ which is often denoted as an N -phase fuzzy membership function. In this framework, our model (4) writes

$$\min_{(\mathbf{p}, \omega) \in L^1(\Omega; L^1(\Delta_t)) \times \mathcal{C}} \bar{E}(\mathbf{p}, \omega) = \underbrace{\sum_{i=1}^N \int_{\Omega} |\nabla \omega_i(x)| dx}_{E_s(\omega)} + \lambda \underbrace{\sum_{i=1}^N \int_{\Omega} \mathcal{D}_t^{\mathbf{a},1}(\mathbf{p}(x), \mathbf{p}_u(x)) \omega_i(x) dx}_{E_d(\mathbf{p}, \omega)}. \quad (6)$$

The energy formulation (6) is convex with respect to \mathbf{p} and ω separately but not jointly. The proof of the existence of minimizers for $\bar{E}(\mathbf{p}, \omega)$ follows the lines of proof of Theorem 1 in [15], but they are not unique.

¹The BV space is the space of real-valued functions whose total variation is bounded (finite).

3. OPTIMIZATION METHOD

To minimize the functional (6), we introduce an auxiliary variable $\mathbf{v} = (v_1, \dots, v_N) \in \mathcal{C}$ representing the fuzzy N -phase membership function ω and we penalize its deviation from ω by a quadratic term as follows

$$\min \tilde{E}(\mathbf{p}, \omega, \mathbf{v}) = E_s(\omega) + \lambda E_d(\mathbf{p}, \mathbf{v}) + \frac{1}{2\theta} \underbrace{\sum_{i=1}^N \int_{\Omega} (\omega_i(x) - v_i(x))^2 dx}_{E_c(\omega, \mathbf{v})}, \quad (7)$$

where $\theta > 0$ is small enough to enforce \mathbf{v} to be the closest possible to ω . This energy can be minimized by alternatively fixing two variables and minimizing with respect to the third one since the functional \tilde{E} is convex w.r.t each variable, and iterate until convergence. In the following we describe how we minimize each of them.

3.1. ω -subproblem; dual formulation algorithm. The subproblem for ω is

$$\min_{\omega} (E_s(\omega) + \lambda E_c(\omega, \mathbf{v})) \quad (8)$$

As the problem (8) is separable in the variables ω_i , we can solve each problem independently, that is:

$$\min_{\omega} \int_{\Omega} |\nabla \omega_i(x)| dx + \frac{1}{2\theta} \int_{\Omega} (\omega_i(x) - v_i(x))^2 dx \quad (9)$$

This minimization is done using a dual formulation and Chambolle's algorithm [5]:

Proposition 1. *The solution of Eq. (9) is given by*

$$\omega_i(x) = v_i(x) + \theta \operatorname{div}(\xi(x)) \quad (10)$$

where the vector function ξ is obtained by the following iterative fixed-point scheme:

$$\xi^{n+1}(x) = \frac{\xi^n(x) + \tau \nabla (\theta \operatorname{div}(\xi(x)) + v_i(x))}{1 + |\theta \operatorname{div}(\xi(x)) + v_i(x)|} \quad (11)$$

taking $\xi^0 = 0$ and $\tau \leq 1/8$.

3.2. \mathbf{v} -subproblem. The subproblem for $\mathbf{v} = (v_1, \dots, v_N)$ is

$$\min_{\mathbf{v} \in \mathcal{C}} \left(\lambda E_d(\mathbf{p}, \mathbf{v}) + \frac{1}{2\theta} E_c(\omega, \mathbf{v}) \right) \quad (12)$$

Again, the problem is separable in the variables v_i . As it is differentiable with respect to v_i , we can obtain a closed solution, plus a projection onto the convex set \mathcal{C} :

$$v_i(x) = \omega_i(x) - \lambda \theta \mathcal{D}_t^{\mathbf{a},1}(p(x), p_u(x)), \quad \forall i. \quad (13)$$

To include the projection onto \mathcal{C} , the set of fuzzy membership functions, the expression (13) is replaced by

$$v_i(x) = \min\{\max\{\omega_i(x) - \lambda \theta \mathcal{D}_t^{\mathbf{a},1}(p(x), p_u(x)), 0\}, 1\} \quad (14)$$

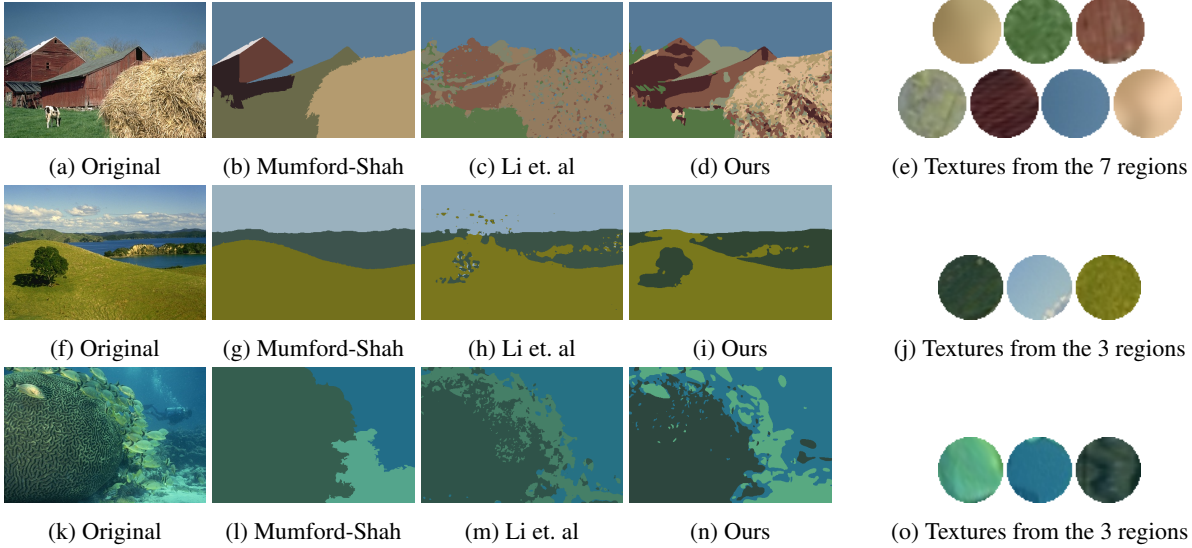


Fig. 2: Results on a selection of images from [17].

together with the constraint $\sum_{i=1}^N v_i(x) = 1, \forall x \in \Omega$.

3.3. p-subproblem. The subproblem for \mathbf{p} is

$$\min_{\mathbf{p}} E_d(\mathbf{p}, \mathbf{v}) = \sum_{i=1}^N \int_{\Omega} \mathcal{D}_t^{\mathbf{a},1}(\mathbf{p}(x), \mathbf{p}_u(x)) v_i(x) dx. \quad (15)$$

with $\mathbf{p} = \sum_i \mathbf{p}_{\Omega_i} \chi_{\Omega_i}$. For each region Ω_i , the unknown patch \mathbf{p}_{Ω_i} is given by a vector of size the number of pixels contained in the disc Δ_t . Thus, the solution for subproblem (15) is given by a weighted median vector. For computing the weighted median vector we use the algorithm proposed in [27, 3].

3.4. Initialization. As the functional is not jointly convex the final result has a high dependence on the initialization. Following the idea of [15] we initialize the algorithm using fuzzy c-means [4] over the set of patches of the input image. Afterwards, by applying (15), the median patch that corresponds to each initial region will be computed.

4. EXPERIMENTAL RESULTS

In order to display the associated segmented image, we first select, for each pixel, the maximum value of the membership functions at that pixel. This is a translation of the assumption that each pixel belongs only to the region with highest membership value (instead of belonging to more than one region with different probability). Then, the piecewise constant image is recovered straightforward.

To compute the affine covariant tensors and associated patches, we use the algorithm proposed in [7] and, for all the experiments, we select a radius of 90 and normalized discs inscribed in a square of side 51 which results in discs with area 2053 pixels (see [7] for details). The used convergence criterium in the minimization algorithm is the following: we

stop the iterative scheme if the relative error between two consecutive membership functions is less than a very small threshold.

In Fig. 2 we present some experimental results on images of the Berkeley Segmentation Dataset [17]. We compare our segmentation results with the method presented by Li et. al in [15] and the classical Mumford-Shah method, using the algorithm proposed by Koepfler et al. [14]. For our results we also present the output texture disc associated to each of the regions. For noisy images, such as the ones in [15], we obtain similar results together with an associated patch.

5. CONCLUSIONS AND FUTURE WORK

In this work we propose a new variational formulation for image segmentation that uses similarity among shape and size adaptive patches in an L^1 fidelity term and the total variation of fuzzy membership functions as relaxed length of the boundaries of the segmentation regions. The result is a partition of the image in regions of local homogeneous texture regardless of differences in the point of view or suffered local perspective or affine distortion, together with a patch, associated to each region, which contains the representative texture of its corresponding region.

The proposed method, built on Riemannian metrics intrinsic to the input image and on an L^1 data term, will allow to synthesize a textured output by using the representative tensor metric of each region to fill-in the region with the texture of the representative patch. This is an interesting direction for future research.

6. REFERENCES

- [1] P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *Int. J. Comput. Vis.*, 93(3):319–347, 2011.
- [2] C. Ballester and M. Gonzalez. Affine invariant texture segmentation and shape from texture by variational methods. *J. Math. Imaging Vis.*, 9(2):141–171, 1998.
- [3] M. Barni. A fast algorithm for 1-norm vector median filtering. *IEEE Trans. on Image Processing*, 6(10), 1997.
- [4] J. C. Bezdek, L. Hall, and L. Clarke. Review of mr image segmentation techniques using pattern recognition. *Med. Phys.*, 20(4):1033–1048, 1993.
- [5] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imag. Vis.*, 20, 2004.
- [6] C. A. Deledalle, V. Duval, and J. Salmon. Non-local methods with shape-adaptive patches (nlm-sap). *J. Math. Imaging Vis.*, 43(2):103–120, 2012.
- [7] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester. Linear multiscale analysis of similarities between images on riemannian manifolds: Practical formula and affine covariant metrics. *SIAM J. Imaging Sci.*, 8(3):2021–2069, 2015.
- [8] J. F. Garamendi and E. Schiavi. A multiclass anisotropic mumford-shah functional for segmentation of d-dimensional vectorial images. In *VISAPP*, 2017.
- [9] S. Grewenig, S. Zimmer, and J. Weickert. Rotationally invariant similarity measures for nonlocal image denoising. *J Vis Commun Im. Represent*, 22(2):117–130, 2011.
- [10] Y. Gu, Xiong, W., L. L. Wang, and J. Cheng. Generalizing mumford-shah model for multiphase piecewise smooth image segmentation. *IEEE T Im. Process*, 2016.
- [11] N. Houhou, J.-P. Thiran, and X. Bresson. Fast texture segmentation based on semi-local region descriptor and active contour. *Numer Math-Theory Me.*, 2009.
- [12] M. Jung. Piecewise-smooth image segmentation models with l1 data-fidelity terms. *J Sci Comput*, 70(3), 2017.
- [13] A. Kheradmand and P. Milanfar. A general framework for regularized, similarity-based image restoration. *IEEE T Image Process*, 23(12):5136–5151, 2014.
- [14] G. Koepfler, C. Lopez, and J.-M. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM J. Numer. Anal.*, 31(1):282–299, 1994.
- [15] F. Li, S. Osher, J. Qin, and M. Yan. A multiphase image segmentation based on fuzzy membership functions and l1-norm fidelity. *J Sci Comput*, 69(1):82–106, 2016.
- [16] T. Lindeberg and J. Garding. Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Im. Vis Comput*, 15, 1997.
- [17] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE I Conf Comp Vis*, volume 2, pages 416–423, July 2001.
- [18] J. M. Morel and S. Solimini. *Variational methods in image processing*. Birkhäuser, 1994.
- [19] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun Pure Appl Math*, 42(5), 1989.
- [20] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the non-local-means to super-resolution reconstruction. *IEEE T Image Process*, 18(1):36–51, 2009.
- [21] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *Computer vision and pattern recog.*, 2003.
- [22] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi. Integrated active contours for texture segmentation. *IEEE T Image Process*, 1(1):1–19, 2004.
- [23] E. Strelakovski and D. Cremers. Real-time minimization of the piecewise smooth Mumford-Shah functional. In *European Conference on Computer Vision*, pages 127–141, 2014.
- [24] J. H. Syu, S. J. Wang, and L. C. Wang. Hierarchical image segmentation based on iterative contraction and merging. *IEEE T Image Process*, 2017.
- [25] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.
- [26] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int. J. Comput. Vis.*, (50(3)), 2002.
- [27] T. Viero, K. Oistamo, and Y. Neuvo. Three-dimensional median-related filters for color image sequence filtering. *IEEE T Circ Sys Vid*, 4(2):129–142, 1994.
- [28] Y. Xu, T. Geraud, and L. Najman. Hierarchical image simplification and segmentation based on Mumford Shah salient level line selection. *Patt. Recog. Lett*, 2016.
- [29] L. A. Zadeh. Fuzzy sets. *Inf Control*, 8(3), 1965.
- [30] C. Zuo, L. Jovanov, H. Luong, B. Goossens, W. Philips, Y. Liu, and M. Zhang. Rotation invariant similarity measure for non-local self-similarity based image denoising. In *IEEE Image Proc*, pages 1618–1622, 2015.