

# NON-LOCAL UNSUPERVISED VARIATIONAL IMAGE SEGMENTATION MODELS

X. BRESSON AND T.F. CHAN\*

**ABSTRACT.** New image denoising models based on non-local image information have been recently introduced in the literature. These so-called "non-local" denoising models provide excellent results because these models can denoise smooth regions or/and textured regions simultaneously, unlike standard denoising models. Standard variational models s.a. Total Variation-based models are defined to work in a small local neighborhood, which is enough to denoise smooth regions. However, textures are not local in nature and requires semi-local/non-local information to be denoised efficiently. Several papers have introduced non-local filters and non-local variational models for image denoising. Yet, few studies have been done to develop unsupervised image segmentation models based on non-local information. This will be the goal of this paper. We define and study three unsupervised non-local segmentation models. These models will be based on the continuous global minimization approach for image segmentation recently introduced in [10, 6]. The energy of [10, 6] is a first order energy composed of the weighted Total Variation norm and a linear term. The first proposed non-local segmentation model will extend the Total Variation regularization term of [10, 6] to the non-local Total Variation energy. We will see that the non-local energy can segment fine and small structures better than the standard Total Variation energy. The second model will extend the data-based term of [10, 6] to a non-local term using the Chan-Vese model. The proposed non-local Chan-Vese model will overcome the main limitation of the original model, that does not work with local intensity inhomogeneities. Finally, the third model will also extend the data-based term of [10, 6] to a non-local term using the Mumford-Shah energy. The original Mumford-Shah energy is designed to work for piecewise smooth images only. We suggest to extend it to textures, defining a non-local Mumford-Shah model that works with real-world images. Numerical minimization schemes presented in this paper are based on continuous and discrete (graph cut) approaches. Experimental results will illustrate the improvements provided by the three proposed non-local unsupervised segmentation models.

**Keywords:** unsupervised image segmentation, non-local/graph approach, active contour, level set method, graph cut, Chan-Vese model, Mumford-Shah model, Total Variation energy, global minimization, continuous minimization, graph cut minimization.

## 1. INTRODUCTION

Many papers have been recently introduced on image denoising using spatially non-local (NL) information [7, 20, 21, 48]. These new denoising models produce excellent results compared with standard models such as [46, 44]. The main reason is based on the fact that most standard denoising models work with *local* image information, which is enough to denoise (piecewise) *smooth* images but not *textures*. It is indeed well-known that variational models and PDEs are efficient models to implement local computations but these models process textures like noise and textures are lost or partially lost during the denoising process. Recently, new denoising models have been developed on *non-local* image information. One of the recent and influential models is the *non-local means* algorithm of Buades, Call and Morel [7], which produces very good denoising results. This model is a filter-based model, whose filter is computed using distance between patches of image intensity. Gilboa and Osher defined in [20] the *variational formulation of NL-means*, with a non-local partial differential equation (PDE). We notice that the two denoising models [7, 20] are also related with Szlam, Maggioni and Coifman's model [48], defined in the context

---

\* Department of Mathematics, University of California, Los Angeles, CA 90095-1555, USA, xbrsson(at)math.ucla.edu, tonyc(at)college.ucla.edu. This research is supported by NSF #DMS-0610079 and ONR #N00014-06-1-0345.

of diffusion geometry. Although other researches have been done on non-local image denoising models, few studies have been developed on unsupervised image segmentation using non-local image information. This will be the objective of this paper. We believe and we will show that standard segmentation models can be enhanced in several ways using a NL extension of standard models. We propose in this paper three NL segmentation models based on the global continuous minimization approach for unsupervised image segmentation recently introduced in [10, 6].

The image segmentation approach proposed in [10, 6] makes the important step from *local* continuous minimization solution to *global* minimization solution. Global minimization solution is important to design robust image segmentation models (and other image processing models) because a global solution is *independent* of a good initial position of the contour unlike e.g. standard active contour models [29, 8, 30]. Although [10, 6] define a global minimization approach for the Chan-Vese's model [11], the geodesic active contour model [8, 30] and the Mumford-Shah's model [50, 39], this global minimization approach can naturally be applied to any active contour model as shown in [38]. Traditionally active contour segmentation models are solved using the *level set method* [43] which depends on slow minimization process and needs regular re-distancing the level set function. Some algorithms have been developed to improve the speed of active contours s.a. [32] but these approaches still compute a local minimizer. Unlike the level set method, the paradigm introduced in [10, 6] is able to compute a global minimizer for *any kind of active contour (boundary-, region-, shape-based) model*. We will present the image segmentation model of [10, 6] in the first section of this paper. This presentation will also be useful to introduce the proposed NL segmentation models since these models are based on [10, 6]. The first section will also emphasize that *continuous* segmentation models can determine global minimum, like recent discrete (*graph cut*) segmentation models introduced in [4]. [4] proposed a combinatorial approach for semi-supervised segmentation based on the parametric max flow/min cut developed in [5]. This discrete segmentation model is very fast and provides a global discrete minimum. However, the global discrete minimizer does not have a sub-pixel accuracy unlike continuous approaches. Besides, graph cut models are known to be anisotropic models and dependant of the grid. Finer grids are necessary to extract fine structures (metrification error) and memory allocation for 3D images need special schemes. Continuous models are isotropic, independent of the grid and are not limited to memory issues for 3D images. But, they need special implementations (parallel implementations or GPU implementation) to converge in a very fast way [49, 45]. Finally, one might say that the continuous segmentation model [10, 6] and the discrete segmentation model [4] are related to each other. The model [4] can be seen as the discrete analogue segmentation model to the continuous model [10, 6], or vice-versa.

The first NL segmentation model extends the TV regularization term of [10, 6] to the NL-TV energy. We will see that the non-local energy can segment *fine and small structures* better than the standard TV energy. Besides, this NL extension introduces a *graph representation* for images. Graph is an obvious representation to encode non-local information, such as neighboring pixels that are spatially far way but share close image features. This new image representation, based on a graph of intensity patches like in [7, 20], enhances the segmentation result. The reason is as follows. The regularization process of [10, 6] is defined by the TV norm, which is thus equivalent to *smooth iso level sets using their curvature* (the higher the curvature of the level set is, the stronger the regularization is). This can be an issue with small structures (with high curvature) to be segmented since they can be lost with the standard TV (especially in the case of medical images). In the situation where the regularization process is done on a graph defined from the image and its structures, then the regularization will better preserve fine and small structures which are repetitive in the image domain (which is the case in natural images and explains ) while smoothing out the noise. We will provide synthetic and medical results to show the improvements providing by this NL segmentation model. We will also present two minimization schemes, one based on a continuous approach and one based on a discrete (graph cut) approach based on the parametric max flow/min cut of [5], which is very fast<sup>1</sup>. The second NL segmentation model will extend the data-based term

<sup>1</sup>Our code can be found at <http://www.math.ucla.edu/~xbresson>

of [10, 6] to a NL term using the Chan-Vese model [11]. The proposed NL-CV model will be able to integrate simultaneously *semi-local and global* image information through a specific graph. The NL-CV model will improve the original model, that does not work with images with *local intensity inhomogeneities*. The implementation will be done with a continuous approach (a graph cut approach is also possible). The third NL segmentation model will extend the data-based part of [10, 6] to a NL term using the Mumford-Shah model [39]. This extension is interesting because the original MS energy was designed to work for piecewise smooth images only. We propose a MS energy for smooth or/and textured images by defining a NL-MS model that is *piecewise regular (regular can mean smooth or/and textured)*. We investigate two numerical schemes to implement the NL-MS model. The first one is based on the level set formulation of the MS energy as proposed in [50]. The second one is based on an elliptic approximation of the MS energy introduced in [2]. We will see that not only the NL-MS is able to segment images, this model also provides better denoising results than the state-of-the-art.

The outline of this paper is as follows. In section 2, we introduce the continuous global minimization scheme introduced in [10, 6]. Any (boundary-, region-, shape-based) active contour model, that are usually solved by the level set method, can be solved in an efficient way with this new segmentation approach. This minimization is global, i.e. no good initial condition are needed to get a global solution and it can be efficiently implemented using fast continuous minimization schemes. Then, the rest of the paper will introduce our contributions:

- Section 3 presents the first extension of [10, 6] to a NL framework. We change the regularization process of Section 2 by considering the NL-TV norm defined on a graph of image patches. This NL regularization, based on prior information about image structures, is able to better segment finer and smaller structures unlike the original approach. Continuous and discrete (graph cut) minimization schemes are introduced.

- Section 4 extends the data-based term of [10, 6] to a NL term using the CV model. The proposed NL-CV model improves the original CV model because the proposed model can segment images with local intensity inhomogeneities. A continuous minimization scheme is also introduced.

- Section 5 extends the data-based term of [10, 6] to a NL term using the Mumford-Shah model. The NL-MS overcomes the original MS model that is defined to work for piecewise smooth images only. The proposed NL-MS is defined for piecewise smooth or/and textured images. Two continuous algorithms are introduced to implement the NL-MS model. The first model is based on a level set formulation of the original MS energy. The second model is based on an elliptic approximation of the original energy. This model not only segments textured images but also produces better denoising results than state-of-the-art.

## 2. CONTINUOUS GLOBAL MINIMIZATION APPROACH FOR IMAGE SEGMENTATION

### 2.1. Proposed Energy.

We present the continuous global minimization approach for unsupervised image segmentation introduced in [10, 6] (see also [38]). The variational segmentation model presented in this section will be extended to a non-local framework in the rest of the paper. As we previously said in the introduction, the continuous global minimization approach introduced in [10, 6] is important because it makes the step between continuous local minimization approaches (mostly used with standard variational and PDE-based models) to global minimization approaches. Global minimization models are important because they are robust to bad initial condition, which is not the case with standard segmentation models such as active contour models [29, 8, 30]. Indeed, standard active contours have to be initially located close to the solution to successfully segment objects of interest. Authors in [10, 6] introduced a new paradigm to design continuous global minimization algorithm for image segmentation based on the active contour model and the level set method [43]. Although [10, 6] developed a global segmentation approach for the Chan-Vese model [11], the

geodesic active contour model [8, 30], and the two-phase piecewise smooth Mumford-Shah model [39], the objective of this section is to present the continuous global minimization approach to any (boundary-, region-, shape-based) active contour model. It is possible to re-formulate any active contour model to compute a global minimization solution. In the large literature on active contour models, the active contour model is implemented with the level set method. We will develop at this end of this section, the new advantages to use our implementation of active contour models. In a nutshell, our segmentation model is well-posed, determines a global minimum, is easy to code and much faster than the level set method. We notice that we have specified the word "continuous" in the proposed global minimization method. We emphasize the continuous nature of the minimization process because there are also discrete global minimization algorithm based on graph cut techniques. We will compare continuous and discrete approaches in Section 2.2.

In image segmentation, a successful variational model is the active contour model, initially proposed by Kass, Witkin and Terzopoulos in [29]. A lot of literature has been written on active contour models since this segmentation allows to easily combine different kind of information such as boundary, region and shape prior [8, 30, 11, 12]. The general energy of any active contour model is as follows (we focus on two-phase segmentation models but they can be extended to multi-phase):

$$(1) \quad E_{AC}(C) = \int_C g_b ds + \lambda \int_{C_{in}} g_r^{in} dx + \lambda \int_{C_{out}} g_r^{out} dx, \quad C : \gamma \rightarrow \mathbb{R}^N$$

where  $AC$  stands for active contour,  $C$  is a closed contour (curve/surface) parameterizes by the set  $\gamma$  ( $\gamma := [0, 1]$  for curve),  $g_b : \Omega \rightarrow \mathbb{R}_+$  is an arbitrary edge function (for example an edge detector function in the geodesic active contour model [8, 30] or a distance function of a set of points in the surface interpolation model [53]),  $ds$  is the contour element,  $C$  is the boundary between  $C_{in, out} \subset \Omega$  which are the inside and outside region of  $C$  in the image domain  $\Omega$ ,  $g_r^{in, out} : \Omega \rightarrow \mathbb{R}$  are arbitrary inside and outside region detector functions (use to force region intensity statistics such as mean, variance, histogram or/and shape prior),  $dx$  is the volume element and  $\lambda$  is a positive constant. Energy (1) encodes most of existing active contour models. The well-know geodesic/geometric active contour model proposed in [8, 30] is given when  $g_r^{in} = g_r^{out} = 0$  and  $g_b = g_b(|\nabla u_0(C(s))|)$  such that  $E_{GAC}(C) = \int_C g_b(|\nabla u_0(C(s))|) ds$ . This segmentation model has been successfully used in many applications because this model is well-posed and the level set method can be used to handle the segmentation flow. Another well-know active contour model given with (1) is the Chan-Vese (CV) model [11]. Indeed, if we choose  $g_b = 1$ ,  $g_r^{in} = (\mu_{in} - I)^2$  and  $g_r^{out} = (\mu_{out} - I)^2$  then we have  $E_{CV}(C) = \int_C ds + \lambda \int_{C_{in}} (\mu_{in} - I)^2 + \lambda \int_{C_{out}} (\mu_{out} - I)^2$ , which is the two-phase piecewise constant approximation of the Mumford-Shah model [39]. The two-phase piecewise smooth approximation of the Mumford-Shah model proposed in [50] can also be introduced considering  $g_b = 1$ ,  $g_r^{in} = |\nabla s_{in}|^2 + (s_{in} - I)^2$  and  $g_r^{out} = |\nabla s_{out}|^2 + (s_{out} - I)^2$  such as  $E_{VC}(C) = \int_C ds + \lambda \int_{C_{in}} |\nabla s_{in}|^2 + (s_{in} - I)^2 + \lambda \int_{C_{out}} |\nabla s_{out}|^2 + (s_{out} - I)^2$ . Many other active contour models fall into energy (1). How can we compute a (local) minimum of (1)?

Energy (1) is usually minimized with the well-known level set method developed by Osher and Sethian in [43], which represent an evolving contour in a higher dimensional space (co-dimension 1) in such way that changes of topology can be dealt with in an automatic way. The level set function not only represents the contour by its zero level set (non-parametric representation) but also the inside and outside regions of the closed zero level set. In [37], Merriman, Bence and Osher show that operators which identifies the zero level set is the Dirac operator  $\delta$  and the Heaviside operator  $H$  provides the inside and outside regions. Thus the level set formulation of e.g. the CV model is  $E_{CV}^2(\phi) = \int_{\Omega} |\nabla H(\phi)| dx + \lambda \int_{\Omega} (\mu_{in} - I)^2 H(\phi) dx + \lambda \int_{\Omega} (\mu_{out} - I)^2 (1 - H(\phi)) dx$ , where  $\phi : \Omega \rightarrow \mathbb{R}$  is the level set function. Then, the minimization is carried out assuming that operators  $\delta, H$  are regularized and using standard Euler-Lagrange equation technique. The level set minimizer of the Chan-Vese energy is local but experimental results in [11] suggest that the model has the tendency to compute a global minimizer since the model is able to deal with interior contours even with an initial condition contour position far away from the objects.

Based on this important observation, Chan, Esedoglu and Nikolova (CEN) proposed in [10] to reformulate the original CV model in order to determine a global minimizing solution. The new functional of CEN consists in minimizing  $E_{CV}^3(u) = \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} ((\mu_{in} - I)^2 - (\mu_{out} - I)^2) u dx$ ,  $u : \Omega \rightarrow [0, 1]$ . It is possible to prove that the steady state of  $E_{CV}^3$  is the same as  $E_{CV}$ , which means that both energies have common minima. Besides, CEN stated a theorem that showed that global minima (not unique but global minimizers are similar) can be determined for  $E_{CV}^3$  (using any optimization algorithm including standard gradient descent method). The theorem states that for any minimizer  $u^*$  of  $E_{CV}^3$ , then any thresholded function of  $u^*$  gives also a global minimizer for  $E_{CV}^3$ . The paper of CEN made the important step from the continuous level set method, which gives local minima, to a new continuous optimization approach, which produces global minima. In [6], Bresson et al. extended the global minimization approach of the CV model to the geodesic active contour model [8, 30] using the weighted TV norm  $\int_{\Omega} g_b |\nabla u|$  and to the two-phase piecewise smooth Mumford-Shah energy as in [50]. [6] also proposed a fast numerical scheme to compute the global minimum, which is much faster than using the level set method.

Let us now extend the continuous global minimization approach introduced in [10, 6] to any active contour model, defined by (1). The level set formulation of (1) is:

$$(2) \quad E_{LSM}(\phi) = \int_{\Omega} g_b |\nabla H(\phi)| + \lambda \int_{\Omega} g_r^{in} H(\phi) + \lambda \int_{\Omega} g_r^{out} (1 - H(\phi)), \quad \phi : \Omega \rightarrow \mathbb{R}$$

where  $\phi$  is the level set function, defined positive inside in its zero level set representing the contour  $C$ . As we already noticed, energy (2) is equal to (1) since the first term  $\int_{\Omega} g_b |\nabla H(\phi)| = \int_{\Omega} g_b |\nabla \phi| \delta(\phi) = \int_C g_b ds$  is the weighted length of  $C$  and the second term is equal to  $\int_{C_{in}} g_r^{in} dx + \int_{C_{out}} g_r^{out} dx$  since  $C_{in}, C_{out}$  are respectively defined by  $H(\phi), 1 - H(\phi)$ . The Euler-Lagrange (EL) equation of (2) is given by:

$$(3) \quad \left( \nabla \cdot \left( g_b \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (g_r^{in} - g_r^{out}) \right) \delta(\phi) = 0.$$

We notice that the region term  $g_r^{in} - g_r^{out}$  used in this approach is the one defined in the context of region competition as introduced by Zhu and Yuille in [55], where two regions fights to maximize a region statistic criterium (such as the intensity mean in the CV model). However, we could have also considered a simplified region term  $g_r$  instead of  $g_r^{in} - g_r^{out}$  that can encode other feature such as a shape prior like in [12]. We do not change the steady state when we multiply an EL equation by a positive function, then (3) have the same steady state as  $\nabla \cdot \left( g_b \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (g_r^{in} - g_r^{out}) = 0$ .

If we regularize  $H, \delta$  such that they do not vanish over the whole image domain  $\Omega$  then the energy of  $\nabla \cdot \left( g_b \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (g_r^{in} - g_r^{out}) = 0$  is equal to  $\int_{\Omega} g_b |\nabla \phi| + \lambda (g_r^{in} - g_r^{out}) \phi$ . However, the previous energy is homogeneous of degree 1 in  $\phi$ . This means that this evolution equation does not have a stationary solution if the minimization of  $\phi$  is not restricted such as  $[0, 1]$ . Thus, the following variational model is proposed to globally minimized the general active contour energy (1):

$$(4) \quad E_{GMAC}(u) = \int_{\Omega} g_b |\nabla u| + \lambda \int_{\Omega} g_r^{in} u + \lambda \int_{\Omega} g_r^{out} (1 - u), \quad u : \Omega \rightarrow [0, 1],$$

where *GMAC* stands for global minimization of active contour models. Before introducing the global minimization theorem. We would like to emphasize the connection between the level set energy (2) and the standard level set method. Usually, there are two approaches with the level set method in the literature. The first one is based on a level set energy (2) and the second is based on an evolution flow such as [41]:

$$(5) \quad \phi_t + \left( \nabla \cdot \left( g_b \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (g_r^{in} - g_r^{out}) \right) |\nabla \phi| = 0$$

This evolution flow is well-posed since there exists a viscosity solution [14]. Besides, the previous flow also solves the following active contour evolution flow:

$$(6) \quad C_t + (g_b \kappa + \lambda (g_r^{in} - g_r^{out})) \mathcal{N} = 0,$$

when  $C$  is embedded in  $\phi$  as its zero level set. The steady state solution of (5) is the same as (3) and thus it is directly connected with the new energy (4).

The following theorem states the existence of global minima for any active contour model defined by (1) (no initial condition is needed).

**Theorem 2.1.1:** *Suppose that  $g_b : \Omega \rightarrow \mathbb{R}_+$ , for any given  $g_r^{in}, g_r^{out} : \Omega \rightarrow \mathbb{R}$  and  $\lambda \in \mathbb{R}_+$ , if  $u_\star$  is any minimizer of  $E_{GMAC}$ , then for almost every  $\nu \in \mathbb{R}$  we have that the characteristic/indicator function of sets  $\Omega_C(\nu) := \{x \in \Omega : u(x) > \nu\}$  (where  $C$  is the boundary of the set  $\Omega_C$ ) is a global minimizer of  $E_{GMAC}$  and  $E_{AC}$  defined in (4) and (1).*

**Proof.** See annex.

A global minimizer of (4) also provides a global minimizer for the active contour energy (1). The relation is given through a characteristic function of a set  $\Omega_C$ :  $E_{GMAC}(u = \chi_{\Omega_C}) = \int_C g_b ds + \lambda \int_{C_{in}} g_r^{in} dx + \lambda \int_{C_{out}} g_r^{out} dx = E_{AC}(C)$ . Hence, any characteristic function of sets  $\Omega_{\Omega(\nu)}$  which globally minimizes  $E_{GMAC}$  also globally minimizes the active contour energy  $E_{AC}$ . We sum up the previous ideas. Most of existing active contour models can be defined through this variational model:

$$\inf_{C:\gamma \rightarrow \mathbb{R}^N} E_{AC}(C) = \int_C g_b ds + \lambda \int_{C_{in}} g_r^{in} dx + \lambda \int_{C_{out}} g_r^{out} dx.$$

The minimizing solution  $C$  can be given using the LSM (where the contour is embedded into the zero level set  $C := \phi^{-1}(0)$ ) given by the PDE:

$$\phi_t = \left[ \nabla \cdot \left( g_b \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (g_r^{in} - g_r^{out}) \right] |\nabla \phi|$$

or the equivalent LSM is given by the variational model with its minimizing flow:

$$\begin{cases} \inf_{\phi:\Omega \rightarrow \mathbb{R}} E_{LSM}(\phi) &= \int_{\Omega} g_b |\nabla \phi| + \lambda H(\phi) g_r^{in} + \lambda (1 - H(\phi)) g_r^{out} dx \\ \phi_t &= \left[ \nabla \cdot \left( g_b \frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda (g_r^{in} - g_r^{out}) \right] \delta(\phi) \end{cases}$$

The minimization approach to globally minimize the active contour energy is given by:

$$\inf_{u:\Omega \rightarrow [0,1]} E_{GMAC}(u) = \int_{\Omega} g_b |\nabla u| + \lambda g_r^{in} u + \lambda g_r^{out} (1 - u) dx$$

where the contour globally minimizing (1) is given by thresholding any minimizer of  $E_{GMAC}$  between  $[0, 1]$ . We notice that the proposed segmentation energy (4) that provides a global minimizing solution to the segmentation problem is convex in  $u$ , lower-semicontinuous and composed of the weighted TV norm and a linear term in  $u$  (1<sup>st</sup> order energy). We also observe that the minimization problem over contours  $C$  in (1) is a non-convex problem, difficult to optimize. However, working with function  $u$  defined on  $\Omega$  makes the segmentation problem convex, which makes possible the computation of a global minimum.

## 2.2. Relations with Level Set Method and Graph Cuts.

In this section, we enumerate the main advantages of the proposed segmentation energy (4). We also compare this continuous global segmentation model with the level set method (LSM) and the discrete (graph cut) model introduced by Boykov in [4].

Let us start with the main advantages of the LSM. The LSM uses an Eulerian/ non-parametric representation of the contour, which can deal with natural changes of topology without extra processes. The LSM is very flexible to introduce different kind of information (boundary, region, shape) in the segmentation process, which makes it a very successful algorithm to solve the image segmentation problem. The LSM can also be applied to other various applications such as fluid dynamical flows, PDE on manifolds, etc. (see [47, 42]). Numerical schemes to implement the LSM are based on hyperbolic conservation laws and upwind implementations. These schemes are stable and can be highly accurate (see WENO schemes [28]). The main limitations of the LSM are as follows. The LSM computes a local minimum, which means that the initial condition is critical to get satisfactory

segmentation results. The LSM needs to regularly re-distance the level set function as a signed distance function (which means solving the Eikonal equation  $|\nabla\phi| = 1$ ) to guaranty correct and smooth evolution of the zero level set embedding the contour (the LS function cannot be too flat nor steep near the zero level set and cannot admit discontinuities in its derivatives). Standard evolution schemes for LSM are slow. [32] proposed to use an AOS scheme to speed up the PDE-based flow process but the proposed algorithm still computes a local minimum and needs to re-distance the LS function. The proposed global minimization approach introduced in Section 2.1 proposes to solve the limitations of the LSM in the context of image segmentation. Indeed, the proposed global minimization computes a global minimum, which makes the segmentation algorithm independent of the initial condition. Besides, the global minimum also corresponds to a minimum of the standard LSM. The proposed global minimization approach also uses an Eulerian/non-parametric representation of the contour as the standard LSM, which allows different topology. The proposed global minimization approach is also flexible to introduce different kind of information (boundary, region, shape) in the segmentation process. Besides computing a global minimum, other advantages of the proposed segmentation model is the simplicity of implementation (few lines of matlab) and the good computational speed for the minimization task. There is also no need to re-distance the level set function in our approach. Authors in [6] showed that the TV-based energy (4) can actually be quickly minimized using any recent fast TV minimization technique. For example, [6] used Chambolle’s projection algorithm introduced in [9] for its good speed and easy implementation. New fast algorithms can be also be used.

Recent discrete minimization models have been introduced to solve the image segmentation problem [4]. These minimization algorithms, called graph cut algorithms, are combinatorial models that minimize energies with discrete values. These algorithms are very fast such as the parametric maximum flow/minimum cut introduced by Boykov and Kolmogorov in [5] (linear time complexity). However, graph cut algorithms do not have sub-pixel accuracy unlike continuous algorithms. Besides, graph cuts can only use anisotropic operators such as the anisotropic gradient operator  $|\nabla u| = |\partial_x u| + |\partial_y u|$  instead of the isotropic operator  $|\nabla u| = \sqrt{|\partial_x u|^2 + |\partial_y u|^2}$ . In the case of graph cuts, the length of a contour needs successively finer grids to be correctly approximated. In fact, the quality of the approximation highly depends on the order of connectivity of the defined graph, and can lead to systematic metrification errors [45]. Continuous models do not have these limitations of anisotropic scheme and metrification error. Unlike continuous algorithms, graph cut algorithms do not have a stopping criteria for the minimization task. Graph cuts are efficient discrete energy minimization techniques even though they are more difficult to implement than continuous algorithms. However, recent continuous minimization algorithms have been developed and compete in terms of speed with graph cuts. For example, in the case of the ROF (TV-L2) minimization model [46], the Split-Bregman algorithm [23] produces very fast computational times close to the graph cut model [15]. Parallel implementations and GPU implementations of continuous PDEs are also possible and lead to very fast minimization algorithms [49, 45].

Finally, we would like to point out the continuous global minimization approach [10, 6] can also be used in other image processing problems. Other image segmentation models based on the proposed global optimization approach [10, 6] have been proposed in the literature, see [34, 38, 40, 26, 49]. Besides, Kolev, Klodt, Brox, Esedoglu and Cremers in [31] and Zach, Pock and Bischof in [52] used a global optimization framework to define global 3D surface reconstruction, independent of the initial condition. Finally, authors in [51] proposed a video tracking algorithm also based on a continuous global minimization framework.

### 3. NON-LOCAL IMAGE SEGMENTATION MODEL BASED ON NON-LOCAL TV

#### 3.1. Proposed Non-Local Variational Energy.

In this section, we extend the TV regularization part in energy (4) to the NL-TV energy. This extension improves segmentation results for fine and small structures. In the original model introduced in (4), the regularization process is based on minimizing the TV norm of  $u$ , which is equivalent to smooth iso level sets using their curvature (the higher the curvature of the level set is, the stronger the regularization is). Hence, the standard TV norm limits the segmentation of small structures, which hold high curvature. In the case where the regularization process is done using NL image information, then the regularization will better preserve small structures which are repetitive in the image domain (which is the case in natural images) while smoothing out the noise. Thus, the "scale space" generated by the NL-TV is different from the scale space of TV. For TV, the scale space based on curvature removes small structures with high curvature first, then larger structures with lower curvature. In the case of the NL-TV, the scale space is defined in a statistical way since structures that are not too repeated will be first removed, then structures which are more repetitive. The *cardinality* of similar structures is thus the *feature* of the NL scale space. Finally, we observe that the NL-TV energy uses prior image information, unlike the standard TV.

The extended NL energy for image segmentation is as follows:

$$(7) \quad E_{GMAC}^{NL}(u) = \int_{\Omega} g_b |\nabla_{NL} u| + \lambda g_r^{in} u + \lambda g_r^{out} (1 - u) dx, \quad u : \Omega \rightarrow [0, 1],$$

where  $\nabla_{NL}$  is the NL gradient operator. As in the approach introduced in Section 2, the segmentation result will be obtained by first computing the minimum of (7), then thresholding it between  $[0, 1]$ . This segmentation procedure will be justified as follows. Firstly, the minimization of (7) is global since the energy (7) is convex w.r.t.  $u$ . Secondly, it may be possible to prove the limit of the NL-TV norm to a continuous TV norm on manifold of image patches. Using the co-area formula on manifold, it is then possible to use Theorem 2.1.1 in Section 2.1. Thus, for any minimizer of (7), a thresholded function of this minimizer gives also a global minimum. Before studying the proposed NL segmentation task, we introduce some notations and definitions on NL quantities.

### 3.2. Definitions and Notations.

The NL gradient operator is defined for the pair of points  $(x, y) \in \Omega \times \Omega$  as in [54, 21]:

$$\nabla_{NL} u(x, y) = (u(y) - u(x)) \sqrt{w(x, y)} : \Omega \times \Omega \rightarrow \mathbb{R}.$$

where  $w$  is the edge function of a graph. The NL gradient of  $u$  does not define a vector in the standard sense since it is a mapping from  $\Omega \times \Omega$  to  $\mathbb{R}$ . However, we will speak of NL vectors for mappings  $p : \Omega \times \Omega \rightarrow \mathbb{R}$  such as  $\nabla_{NL} u$ . The inner product between two NL vectors  $p_1, p_2 : \Omega \times \Omega \rightarrow \mathbb{R}$  at point  $x \in \Omega$  is defined as:

$$\langle p_1, p_2 \rangle(x) = \int_{\Omega} p_1(x, y) p_2(x, y) dy : \Omega \rightarrow \mathbb{R},$$

which gives the norm of a NL vector  $p : \Omega \times \Omega \rightarrow \mathbb{R}$  at point  $x \in \Omega$  as follows:

$$|p|(x) = \sqrt{\int_{\Omega} p(x, y)^2 dy} : \Omega \rightarrow \mathbb{R}_+.$$

Hence, the norm of the NL gradient of a function  $u : \Omega \rightarrow \mathbb{R}$  at  $x \in \Omega$  is defined as:

$$(8) \quad |\nabla_{NL} u|(x) = \sqrt{\int_{\Omega} (u(y) - u(x))^2 w(x, y) dy} : \Omega \rightarrow \mathbb{R}_+.$$

The NL divergence operator can be defined by the standard adjoint relation with the NL gradient (generalized divergence theorem):

$$\langle \nabla_{NL} u, p \rangle = - \langle u, \text{div}_{NL} p \rangle, \quad \forall u : \Omega \rightarrow \mathbb{R}, \quad \forall p : \Omega \times \Omega \rightarrow \mathbb{R},$$

which defines the NL divergence of the NL vector  $p : \Omega \times \Omega \rightarrow \mathbb{R}$  at  $x \in \Omega$ :

$$\text{div}_{NL} p(x) = \int_{\Omega} (p(x, y) - p(y, x)) \sqrt{w(x, y)} dy : \Omega \rightarrow \mathbb{R},$$

### 3.3. Non-Local BV Space.

Smoothness of a function  $u$  on a graph can be measured using the Sobolev norm as proposed in [48]:

$$(9) \quad \|u\|_{H^1}^2 = \|u\|_{L^2(\Omega)}^2 + \int_{\Omega} |\nabla_{NL}u|^2 dx,$$

where  $\int_{\Omega} |\nabla_{NL}u|^2 dx$  is the NL-Dirichlet energy, related with the graph Laplacian (through the derivation of (9)). The graph Laplacian plays a fundamental role in many approaches and applications. It allows to generalize the linear heat diffusion equation to graph representation. In (9), the smaller  $\|u\|_{H^1}$ , the smoother is  $u$  on the graph. It is also interesting to see how vary the function  $u$  on the graph. Let us introduce the space of graph-based TV/NL-TV.

**Definition 3.3.1:** *The graph-based TV/NL-TV norm of  $u \in L^1(\Omega)$  is defined by:*

$$\begin{aligned} TV_{NL}(u) &= \int_{\Omega} |\nabla_{NL}u| dx \\ &= \sup_{|p| \leq 1} \int_{\Omega} u \operatorname{div}_{NL} p \, dx, \end{aligned}$$

where  $p : \Omega \times \Omega \rightarrow \mathbb{R}$ . We define the graph-based BV space as the space of all functions  $u \in L^1(\Omega)$  satisfying  $TV_{NL}(u) < \infty$ . The graph/NL BV space is endowed with the norm:

$$\|u\|_{BV_{NL}} = \|u\|_{L^1(\Omega)} + TV_{NL}(u)$$

The smaller  $\|u\|_{BV_{NL}}$ , the less variations have  $u$  on the graph.

### 3.4. Well-Posedness of the Proposed NL Segmentation Model.

The non-local segmentation model proposed in Section 3.1 is similar to Section 2. It consists in minimizing a convex energy (7), then thresholding it between  $[0, 1]$  to get the segmentation result. The only difference is the regularization process. (4) uses the standard TV energy and (7) uses the NL-TV energy defined from the graph of image patches. The graph of image patches actually defines a *new image representation* and the NL segmentation task is done on this new representation. The objective of this section is to show (under some assumptions) that the proposed segmentation task is well-defined on the new image space representation.

Let us introduce this new space of image representation based on patches of image intensity. The patch of intensity around a pixel located at  $x$  in the image domain defines a point in a higher dimensional space, see Figure 1. For example, if the size of the patch is  $5 \times 5$  then the space of all patches has  $n = 25$  dimensions. We now make the assumption that the set of patches in an image belongs to a compact differentiable submanifold  $\mathcal{M}$  of  $\mathbb{R}^n$ , where the dimension of  $\mathcal{M}$  is much smaller than the ambient space  $d < n$ . The relevance of the existence of such a manifold  $\mathcal{M}$  is an open issue, see [3, 27]. However, we will assume in this work this manifold exists to show the well-posedness of our image segmentation. Based on [3, 27], Coifman and Lafon in [13, 33] studied the convergence of the graph Laplace operator defined on a graph to the Laplace-Beltrami operator defined on a manifold (vertices of the graph are actually samples from a smooth manifold). We use in this section the same arguments to show that the NL-TV norm defined on the graph converges to the TV norm defined on a manifold  $\mathcal{M}$ .

[ Figure 1. ]

**Proposition 3.4.1:** *Let us assume that the set of patches in an image belongs to a compact differentiable submanifold  $\mathcal{M} \subset \mathbb{R}^n$ , where the dimension of  $\mathcal{M}$  is much smaller than the*

ambient space  $d < n$ . If  $u \in L^1(\mathcal{M})$ , then, the NL-TV norm converges to the TV norm on  $\mathcal{M}$ :

$$\int_{\Omega} \frac{1}{\epsilon^{\frac{d+2}{4}}} |\nabla_{NL\epsilon} u| \xrightarrow{\epsilon \rightarrow 0} \int_{\mathcal{M}} |\nabla_{\mathcal{M}} u|.$$

**Proof.** See Annex.

Hence, using the previous proposition, the NL energy (7) is equal at the limit  $\epsilon \rightarrow 0$  to:

$$E_{GMAC}^{NL}(u) = \int_{\mathcal{M}} g_b |\nabla_{\mathcal{M}} u|(\hat{x}) + \lambda g_r^{in} u(\hat{x}) + \lambda g_r^{out} (1 - u(\hat{x})) d\hat{x}$$

Using the coarea formula on manifold  $\mathcal{M}$ :

$$\int_{\mathcal{M}} |\nabla_{\mathcal{M}} u|(\hat{x}) d\hat{x} = \int_{\nu} \int_{\mathcal{M}} |\nabla_{\mathcal{M}} \chi_{\Omega_u(\nu)}|(\hat{x}) d\hat{x} d\nu,$$

we can extend Theorem 2.1.1 to the new image representation defined by the manifold of intensity patches  $\mathcal{M}$ . Thus, if  $u_*$  is any minimizer of  $E_{GMAC}^{NL}$ , then for a.e.  $\nu \in [0, 1]$  we have that the characteristic function of sets  $\Omega_C(\nu) := \{x \in \mathcal{M} : u(x) > \nu\}$  is also a global minimizer of  $E_{GMAC}^{NL}$ . Thresholding  $u_*$  in  $\mathcal{M}$  leads to threshold  $u_*$  in  $\Omega$ . Thus, the segmentation algorithm is based on the same two steps of Section 2.1: minimizing a convex energy (7) and thresholding it to get the minimizer. Thus, the proposed non-local segmentation algorithm can be seen as carrying out the segmentation on the manifold of image patches. We remind that the previous arguments are based on the existences of a continuous manifold  $\mathcal{M}$ . This question is still open even though some works tend to justify the existence of manifolds from graphs [3, 27, 13, 33, 25].

### 3.5. Numerical Schemes.

#### 3.5.1. Minimization by Dual/Projection Approach.

Energy (7) can be minimized using the approach proposed in [6]. This approach consists in introducing a new function  $v : \Omega \rightarrow \mathbb{R}$  and splitting (7) such that:

$$(10) \quad E_{GMAC}^{NL}(u, v) = \int_{\Omega} g_b |\nabla_{NL} u| + \lambda (g_r^{in} - g_r^{out}) v + \frac{1}{2\theta} (u - v)^2 dx,$$

where  $\lambda, \theta > 0$  are two constants ( $\theta$  should be small). Since (10) is convex w.r.t.  $u, v$ , then its minimizer can be computed by minimizing (10) w.r.t.  $u$  and  $v$  separately, and iterating until convergence. Thus, the following minimization problems are considered:

- (1)  $u$  being fixed, we search for a solution to:

$$\inf_{v \in [0, 1]} \int_{\Omega} \lambda (g_r^{in} - g_r^{out}) v + \frac{1}{2\theta} (u - v)^2 dx,$$

whose solution is given by  $v = \min(\max(u - \lambda\theta(g_r^{in} - g_r^{out}), 0), 1)$ .

- (2)  $v$  being fixed, we search for a solution to:

$$(11) \quad \inf_u \int_{\Omega} g_b |\nabla_{NL} u| + \frac{1}{2\theta} (u - v)^2 dx,$$

whose solution is given by the minimum of the NL-ROF model.

An elegant way to minimize the NL-ROF model was proposed by Gilboa and Osher in [21]. They extend the projection algorithm of Chambolle [9] define for the ROF model to the NL-ROF (NL-TV+L2) model. Using the dual definition of NL-TV in Section 3.3, minimizing (11) is equivalent to this min-max problem:

$$\inf_u \sup_{p \leq g_b} \int_{\Omega} u \operatorname{div}_{NL} p + \frac{1}{2\theta} (u - v)^2 dx$$

The inf and sup can be swapped according to the minimax theorem [16] which gives the minimizing solution:

$$u(x) = v - \theta \operatorname{div}_{\text{NL}} p_*, \quad x \in \Omega$$

$p_*$  is the solution of the max problem:

$$\sup_{|p| \leq g_b} \int_{\Omega} \langle \nabla_{\text{NL}} v, p \rangle + \frac{\theta}{2} |\operatorname{div}_{\text{NL}} p|^2 dx$$

which is given by the steady state of the following semi-implicit iterative scheme:

$$(12) \quad p^{n+1}(x, y) = \frac{p^n + \tau \nabla_{\text{NL}}(\operatorname{div}_{\text{NL}} p^n - v/\theta)}{1/g_b + \tau |\nabla_{\text{NL}}(\operatorname{div}_{\text{NL}} p^n - v/\theta)|}, \quad (x, y) \in \Omega \times \Omega, \quad n > 0$$

where  $\tau < 1/(\|\nabla_{\text{NL}} \cdot\| \cdot \max_{\Omega} g_b)$  guaranties the convergence of the iterative scheme.

The iterative scheme (12) is easy to implement. The NL gradient, divergence and norm are discretized as follows ( $x$  is discretized by  $i$  and  $y$  by  $j$ ):

$$\begin{aligned} (\operatorname{div}_{\text{NL}} p)_i &= \sum_{j \in \mathcal{N}_i} (p_{ij} - p_{ji}) \sqrt{w_{ij}} \\ (\nabla_{\text{NL}} v)_{ij} &= (v_j - v_i) \sqrt{w_{ij}} \\ |p|_i &= \sqrt{\sum_{j \in \mathcal{N}_i} (p_{ij})^2}. \end{aligned}$$

which gives:

$$u_i = v_i - \theta \sum_{j \in \mathcal{N}_i} (p_{ij} - p_{ji}) \sqrt{w_{ij}},$$

and

$$p_{ij}^{n+1} = \frac{p_{ij}^n + \tau ((\operatorname{div}_{\text{NL}} p_j - \operatorname{div}_{\text{NL}} p_i) \sqrt{w_{ij}} + (v_j - v_i) \sqrt{w_{ij}}/\theta)}{1/g_b + \tau \sqrt{\sum_{j \in \mathcal{N}_i} ((\operatorname{div}_{\text{NL}} p_j - \operatorname{div}_{\text{NL}} p_i) \sqrt{w_{ij}} + (v_j - v_i) \sqrt{w_{ij}}/\theta)^2}}.$$

Weights  $w_{ij}$  can be determined in many ways. We consider the standard weights as in [20].

Experimental results are presented in Figures 2, and 3. We compare our NL segmentation model defined in (7) with the standard approach introduced in Section 2. For both experiments, we use the Chan-Vese model [11] for comparison. This means that we consider in (7) the edge detector  $g_b = 1$  and the region detector function as  $g_r^{in} - g_r^{out}(x) = (\mu_{in} - u_0(x))^2 - (\mu_{out} - u_0(x))^2$ , where  $\mu_{in}, \mu_{out}$  are the inside and outside intensity mean defined by  $\mu_{in} = \frac{\int_{\Omega} I u(x) dx}{\int_{\Omega} u(x) dx}$ ,  $\mu_{out} = \frac{\int_{\Omega} u_0(1-u(x)) dx}{\int_{\Omega} (1-u(x)) dx}$ . Figure 2 shows that the proposed NL segmentation model performs better to segment fine and small structures unlike the standard CV model. Figure 3 presents the segmentation of a real-world/medical image. We observe on Figures 3(d) and 3(e) that the NL segmentation model gives better segmentation results for finer structures than the CV model.

[ Figure 2. ]

[ Figure 3. ]

### 3.5.2. Minimization by Graph-Cut Approach.

In this section, we develop a numerical scheme based on graph cut techniques to minimize a discrete version of (7). As we already said in Section 2.2, graph cut techniques are limited by anisotropic scheme, metrification error and are not as easy to code as continuous minimization schemes. However, these minimization techniques are very fast.

Unlike Section 2, we adopt in this section a discrete point of view. In the discrete setting, segmentation (and other tasks s.a. denoising) can be formulated in a Bayesian framework using Markov Random Fields (MRF) [19]. The MRF framework aims at finding

the maximum a posteriori (MAP) probability of quantities (like intensity in image denoising or binary function in image segmentation) using the Bayes principle. Bayes principle states that the MAP is given by the product of the a priori probability of the unknowns (data-based term) and the likelihood probability (regularization term). The MRF is composed of several components. A set  $\mathcal{P} = \{1, \dots, N_p\}$  of pixels  $p$  (analogue to an open set  $\Omega \subset \mathbb{R}^N$  in the continuous framework), a neighborhood system  $\mathcal{N} = \{\mathcal{N}_p | p \in \mathcal{P}\}$ , where  $\mathcal{N}_p$  is a subset of pixels in  $\mathcal{P}$  describing the neighborhood of  $p$  and a field of random variables  $\mathcal{U} = \{\mathcal{U}_p | p \in \mathcal{P}\}$  (analogue to the continuous function  $u : \Omega \rightarrow \mathbb{R}$  in the continuous framework). Each random variable  $\mathcal{U}_p$  takes a value  $u_p$  in some discrete set  $\mathcal{F} = \{f_1, \dots, f_{N_f}\}$  of the possible labels (like intensity  $\mathcal{F} = \{0, \dots, 255\}$  in image denoising, corresponding to  $\mathcal{F} = \mathbb{R}$  in the continuous framework). We abbreviate  $\{\mathcal{U}_1 = u_1, \dots, \mathcal{U}_{N_p} = u_{N_p}\}$  with  $\mathcal{U} = u$  where  $u = \{u_p | p \in \mathcal{P}\}$  is an outcome/configuration of  $\mathcal{U}$ . An important and necessary condition for  $\mathcal{U}$  to be a MRF is that each random variable  $\mathcal{U}_p$  depends on the neighbor random variable  $Pr(\mathcal{U}_p | \mathcal{U}_{\mathcal{P}-\{p\}}) = Pr(\mathcal{U}_p | \mathcal{U}_{\mathcal{N}_p})$ ,  $\forall p \in \mathcal{P}$ , where  $\mathcal{U}_{\mathcal{N}_p} = \{\mathcal{U}_q | q \in \mathcal{N}_p\}$ . The fundamental result to make MRF practical is the Hammersley-Clifford theorem [19]. This theorem states that the probability of a particular configuration is the sum over all cliques in the neighborhood system  $\mathcal{N}$ , that is  $Pr(u) \propto \exp(-\sum_C V_C(u))$ .  $V_C$  is a clique potential, which gives the prior probability of a particular outcome of the elements of the clique  $C$ . Most of current models restrict the MRF with clique potentials involving pairs of neighboring pixels  $Pr(u) \propto \exp(-\sum_{p \in \mathcal{P}, q \in \mathcal{N}_p} V_{p,q}(u_p, u_q))$ . The field  $U$  cannot be directly observed in the experiment. It has to be estimated based on some observations (such as the noisy image in image denoising). The likelihood probability  $Pr(O|u)$  relates  $u$  with the observation  $O$ . It is usually assumed that noise is i.i.d. then  $Pr(O|u) = \prod_{p \in \mathcal{P}} G(u_p, O_p)$ . If  $G(u_p, O_p) = \exp(-D(u_p, O_p))$  then the likelihood is  $Pr(O|u) = \exp(-\sum_{p \in \mathcal{P}} D(u_p, O_p))$ . We want to obtain the configuration  $u \in \mathcal{F} \times \dots \times \mathcal{F} = \mathcal{F}^{N_p}$  that maximizes the MAP  $Pr(u|O)$ . Bayes principle states that  $Pr(u|O) \propto Pr(O|u)Pr(u)$ . It follows that the MAP estimate  $u$  should minimize the posterior energy function (using the *log* function):

$$(13) \quad E(u) = \sum_{p \in \mathcal{P}, q \in \mathcal{N}_p} V_{p,q}(u_p, u_q) + \sum_{p \in \mathcal{P}} D(u_p, O_p), \quad u : \mathcal{P} \rightarrow \mathcal{F}$$

We specify functions  $V_{p,q}(u_p, u_q)$  and  $D(u_p, O_p)$  to formulate the equivalent discrete MRF energy of (7) as follows:

$$(14) \quad E_{GMAC}^{GC}(u) = \sum_{p \in \mathcal{P}, q \in \mathcal{N}_p} g_{b_{p,q}} \cdot |u_p - u_q| + \lambda \sum_{p \in \mathcal{P}} g_{r,p}^{in} u_p + g_{r,p}^{out} (1 - u_p), \quad u : \mathcal{P} \rightarrow \{0, 1\}$$

where  $g_{b_{p,q}}$  can be interpreted as a cost of discontinuity between  $p$  and  $q$ . It is natural to consider the clique potential in (14) as the length of discontinuity sets which represents the contours between the two regions  $C_{in}$  and  $C_{out}$ . This also represents the discretization of the TV norm in the case of graph cut approach such as [15]. The anisotropic property of this discretized TV clearly appears here. We can also define a discrete edge detector function as  $g_{b_{p,q}} = \exp(-(I_p - I_q)^2 / 2\sigma^2) \cdot 1/|p - q|$ . Besides, the MRF analogue of the CV model can be considered when we have  $D(u_p, O_p = I_p) = u_p(\mu_{in} - I_p)^2 + (1 - u_p)(\mu_{out} - I_p)^2$  (see also [17]). Other region detector functions can be defined such as histogram-based terms. A shape prior can also be encoded with  $D(u_p, O_p = S_p) = (u_p - S_p)^2$ . Actually all continuous active contour can have their MRF analogues.

As we said, one of the main interests to use the graph cut framework is the speed of minimization, most of time very fast. The way to minimize the MRF energy (14) is related with combinatorial optimization algorithms. In the case of multi-phases (Potts model) the problem is NP-hard. In the case of the Ising model ( $\mathcal{F} = \{0, 1\}$ ), fast algorithms have been developed to compute the exact global minimum of (14) in linear time. The first algorithm has been written by [24], who proposed a graph cut algorithm for binary image denoising. Many other algorithms have been proposed, see [5] for more references. All these algorithms are based on the technique called parametric maximum flow/minimum cut (max flow/min cut). The computation of the minimum cut is well-known as the graph cut problem. The min cut, also known as a s-t cut, is an optimal cut in a certain graph

defined for splitting the nodes belonging to the source from the nodes being in the sink. There are many different ways to compute a globally minimum s-t cut [18, 22] but a recent and efficient algorithm was proposed by [5], which outperform the standard combinatorial optimization techniques. We use their algorithm in this paper, which available on-line.

The extension of the discrete MRF energy (14) to a non-local framework is as follows:

$$(15) \quad E_{GMAC}^{NL,GC}(u) = \sum_{p \in \mathcal{P}, q \in \mathcal{N}_p^{NL}} w_{p,q} \cdot |u_p - u_q| + \lambda \sum_{p \in \mathcal{P}} g_{r,p}^{in} u_p + g_{r,p}^{out} (1 - u_p), \quad u : \mathcal{P} \rightarrow \{0, 1\},$$

where  $w_{p,q}$  is the weight function of the graph of image patches (like in Section 3.5.1). In this situation, the neighborhood  $\mathcal{N}_p^{NL}$  of  $p$  is a non-local neighborhood. The contour of segmented objects are naturally given by the discontinuity set of the binary function  $u$ .

Experimental results are presented in Figures 4 and 5. We also observe that the minimization speed with the graph cut approach is at least twice faster than the continuous approach.

[ Figure 4. ]

[ Figure 5. ]

#### 4. NON-LOCAL CHAN-VESE MODEL

##### 4.1. Motivations.

In the previous section, we extend the regularization part of energy (4) to the NL-TV. In this section, we will extend the data-based term of (4) to a NL term using the Chan-Vese (CV) model [11]. The CV model, also called active contours without edges, detects boundaries of objects lying in images based on the assumption that images are composed of two homogeneous regions (like two regions with Gaussian distributions). This model is more robust than standard snake models [29, 8, 30] which are based on the detection of large image gradients. Precisely, the CV model determines the optimal two-phase/binary piecewise constant approximation of an image, where the two regions are approximated by the mean intensity values. The main limitation of the CV model is with images with local intensity inhomogeneities or non-homogeneous images (Figure 7). The proposed NL-CV model will be able to overcome this limitation using semi-local and global image information. The integration of semi-local and global image information is done with a specific graph, different from Section 3. Semi-local information is important to segment objects that have non-homogeneous intensity. However, semi-local information can lead to bad results depending on the initial position of the contour. This is why the combination of both semi-local and global image information is important because global image information introduce global consistency and thus avoid bad local segmentation. We notice that the term "non-local" for the NL-CV model can be confusing. Indeed, the original CV model uses global image information (intensity means inside and outside the evolving active contour). The original CV model is thus already "non-local". We decide to keep the word non-local for the new CV model because we use a *graph* representation (which defines non-local information) to integrate semi-local an global information at the same time.

The NL-CV energy is defined from the following CV energy (4) in Section 2.1:

$$(16) \quad E_{CV}(u, \mu_{in}, \mu_{out}) = \int_{\Omega} |\nabla u| + \lambda(\mu_{in} - u_0(x))^2 u + \lambda(\mu_{out} - u_0(x))^2 (1 - u) dx, \quad u : \Omega \rightarrow [0, 1].$$

where  $\mu_{in}, \mu_{out} \in \mathbb{R}$ . The second term (data-based term) in (16) is then extended to a non-local/graph formulation as follows:

$$(17) \quad E_{CV}^{NL}(u, \mu_{in}, \mu_{out}) = \int_{\Omega} |\nabla u| + \lambda \int_{\Omega \times \Omega} (\mu_{in}(y) - u_0(x))^2 u(x) w(x, y) + (\mu_{out}(y) - u_0(x))^2 (1 - u(x)) w(x, y) dy dx, \quad u : \Omega \rightarrow [0, 1]$$

where  $u_0 : \Omega \rightarrow \mathbb{R}$  is the given image,  $\mu_{in}, \mu_{out} : \Omega \rightarrow \mathbb{R}$  are two functions and  $w : \Omega \times \Omega \rightarrow \mathbb{R}_+$  is the edge function of a graph specifically designed for the NL-CV model, presented in the next section. Finally, we notice that (17) is convex w.r.t. function  $u, \mu_{in}, \mu_{out}$  individually but not w.r.t. functions  $u, \mu_{in}, \mu_{out}$  simultaneously. The minimization of  $u$  is thus global for fixed  $\mu_{in}, \mu_{out}$ . However, we will need to update  $\mu_{in}, \mu_{out}$  to minimize (17) and determine the optimal segmentation result. Since (17) is not globally convex to all variables at the same time, the minimization result is dependent of the initial condition of  $u$ . Fortunately, experiments will show good robustness w.r.t. different initial conditions.

#### 4.2. Properties of the Non-Local Chan-Vese Model.

In energy (17), functions  $\mu_{in}, \mu_{out}$  represent respectively the inside and outside non-local means of Buades, Coll and Morel in [7]. It is easy to show it using the Euler-Lagrange equations of (17) w.r.t.  $\mu_{in}, \mu_{out}$ :

$$\begin{aligned} \mu_{in}(y) &= \frac{\int_{\Omega} u_0(x) u(x) w(x, y) dx}{\int_{\Omega} u(x) w(x, y) dx} \\ \mu_{out}(y) &= \frac{\int_{\Omega} u_0(x) (1 - u(x)) w(x, y) dx}{\int_{\Omega} (1 - u(x)) w(x, y) dx}. \end{aligned}$$

If one consider function  $u$  as a characteristic function of a set  $C_{in}$  (representing the inside contour), then functions  $\mu_{in}, \mu_{out}$  are specifically the inside and outside non-local means of a contour represented by the discontinuities of the characteristic function  $u$ :

$$(18) \quad \begin{aligned} \mu_{in}(y) &= \frac{\int_{C_{in}} u_0(x) w(x, y) dx}{\int_{C_{in}} w(x, y) dx} \\ \mu_{out}(y) &= \frac{\int_{C_{out}} u_0(x) w(x, y) dx}{\int_{C_{out}} w(x, y) dx}. \end{aligned}$$

Energy (17) is minimized w.r.t.  $u$  using the same technique as in [6]. We re-write (17) as follows:

$$(19) \quad \begin{aligned} E_{CV}^{NL}(u) &= \int_{\Omega} |\nabla u| + \lambda V(x) u(x) + C, \\ \text{where } V(x) &= \int_{\Omega} ((\mu_{in}(y) - u_0(x))^2 - (\mu_{out}(y) - u_0(x))^2) w(x, y) dy. \end{aligned}$$

where  $C$  is a constant independent of  $u$ , which is useless in the minimization process of  $u$ . Then, we split the original energy into two simpler minimization tasks by introducing a new function  $v : \Omega \rightarrow \mathbb{R}$  as follows:

$$(20) \quad \inf_{v \in [0, 1], u} E_{CV}^{NL, 2}(u, v) = \int_{\Omega} |\nabla u| + \lambda V v + \frac{1}{2\theta} (u - v)^2, \quad u : \Omega \rightarrow \mathbb{R}, v : \Omega \rightarrow [0, 1].$$

The energy (20) is convex w.r.t  $u, v$  then the minimization is done alternatively w.r.t.  $u, v$ . The minimization w.r.t.  $u$  is given by the minimizer of the ROF model and the minimization w.r.t.  $v$  is given by  $v = \min(\max(u - \lambda\theta V, 1), 0)$ .

### 4.3. Weight Function for Non-Local Chan-Vese Model.

The NL-CV model use different weights than the NL segmentation model introduced in Section 3. In the case of Section 3, the weights were computed like in [20] as  $w(x, y) = e^{-\|p_x - p_y\|_2^2/h}$ , where  $p_x, p_y$  are patches of image intensity centered at  $x, y$  and  $h$  is a positive parameter which acts as a scale parameter. The weight function of Section 3 is thus computed from the  $L^2$  distance between two image patches. We do not use these weights (more specific to image regularization) in the NL-CV model.

The weights of the NL-CV model are defined as follows. We want to overcome the problem of images with local intensity inhomogeneities. We thus need semi-local information to accurately segment boundaries of non-homogeneous objects. And we also need global image information to avoid a bad local minimum due to bad semi-local information. A combination of both semi-local and global information leads to a graph defined e.g. by these weights:

$$(21) \quad w_{\text{NL-CV}}(x, y) = \frac{1}{Z} \left( \alpha_{\text{SL}} e^{-\|x-y\|^2/2h_{\text{SL}}^2} + \alpha_{\text{G}} e^{-\|x-y\|^2/2h_{\text{G}}^2} \right),$$

where SL, G stand for semi-local and global, constants  $h_{\text{SL}} \ll h_{\text{G}}$  because the first term of  $w_{\text{NL-CV}}$  is designed to capture local boundaries of non-homogeneous objects and the second term captures global image information. Constants  $\alpha_{\text{SL}}, \alpha_{\text{G}}$  s.t.  $\alpha_{\text{SL}} + \alpha_{\text{G}} = 1$  weight the contribution of semi-local information and global information. Constant  $Z$  is the normalization factor. If we consider only the second part of (21), i.e.  $w_{\text{NL-CV}} = e^{-\|x-y\|^2/2h_{\text{G}}^2} / \sqrt{2h_{\text{G}}^2}$  and makes  $h_{\text{G}}$  large (the weight function tends to a constant function on the whole image domain) then the NL-CV model (17) converges to the original CV model (16). Indeed, for  $h_{\text{G}} \rightarrow \infty$  then  $\mu_{\text{in}}(y) = \int_{C_{\text{in}}} u_0(x)w(x, y)dx / \int_{C_{\text{in}}} w(x, y)dx \rightarrow \int_{C_{\text{in}}} u_0(x)dx / \int_{C_{\text{in}}} dx$  which is the mean value in  $C_{\text{in}}$  like in the CV model. We also have  $\int_{\Omega} \int_{\Omega} (\mu_{\text{in}}(y) - u_0(x))^2 u(x)w(x, y)dydx \rightarrow |\Omega| \int_{\Omega} (\mu_{\text{in}} - u_0(x))^2 u(x)dx$ , which is (up to a constant) the second term in the CV model (16). If we now consider only the first part of weights (21), i.e.  $w_{\text{NL-CV}}(x, y) = e^{-\|x-y\|^2/2h_{\text{SL}}^2} / \sqrt{2h_{\text{SL}}^2}$ , then we can interpret  $\mu_{\text{in}}, \mu_{\text{out}}$  as the semi-local means evaluated inside and outside the contour implicitly defined by the discontinuity set of  $u$ . A current point of the contour is then classified to one phase or the other by comparing the distance between the semi-local mean inside and outside the contour with the current intensity of the point. Thus, the weights (21) are used to compute a local mean value weighted by semi-local and global image information. Experiments show that semi-local means or global means (original CV model) alone are not enough to give a consistent segmentation result for the textured pictures. Global means are not able to deal with non-homogenous objects. Semi-local means can produce some inconsistencies in the segmentation task, that can be removed using global image information at the same time, see Figure 6.

[ Figure 6. ]

### 4.4. Results.

A direct implementation of the NL-CV model with weights (21) can be time consuming. However, since the choice of the graph is arbitrary, we discretize the weights by considering only a *small* number of pixels in the computation of the non-local means (18). More precisely, we consider two sets of points. The first set of points is for points close to  $x$  (semi-local part) and the second set is for points far from  $x$  (global part). This allows to decrease the computational time since we consider only a small portion of pixels in the whole support of the image.

Experimental results are presented on Figures 7 and 8. They show that the NL-CV model is able to improve the original CV model because it can deal with the local intensity inhomogeneities in images.

[ Figure 7. ]

[ Figure 8. ]

#### 4.5. Discussion and Related Works.

The NL-CV model overcomes the standard limitation of the CV model, which cannot deal with local intensity inhomogeneities. The weights defined for the NL-CV model in (21) are able to integrate semi-local and global information in the same framework. This allows to segment non-homogenous objects without getting stuck in a bad local minimum. The idea to improve the CV model was already proposed by Vese and Chan in [50], where authors minimized the two-phase piecewise-smooth original Mumford-Shah energy. This segmentation model is able to deal with non-homogeneous smooth objects. However, since the model does not use any global information, it is sensitive to bad local minima. Besides, the model is based on a level set formulation and thus it does not produce a global minimum for fixed functions  $\mu_{in}, \mu_{out}$  unlike our model. Another related work is [35, 36]. Authors produce very good segmentation results for non-homogeneous objects. However, their model did not introduce global image information like in our model. Besides the segmentation model is based on a level set formulation like in [50]. The proposed model is thus less robust to local minima than our model and the computational time to minimize their energy w.r.t. the level set function is higher than the minimization of our energy w.r.t. function  $u$ . Mory and Ardon in [38] also introduced a segmentation to deal with local inhomogeneities. They defined their algorithm based on the global minimization approach [10, 6], which allow them to determine a global minimum for a fixed region term. However, they did not use any global image information, which can be useful to avoid bad local minimum as we previously explained.

Another NL-CV model can be this one:

$$E_{CV}^{NL,2}(u, \mu_{in}, \mu_{out}) = \int_{\Omega} |\nabla u| + \lambda \int_{\Omega \times \Omega} |\mu_{in}(y) - u_0(x)| u(x) w(x, y) + |\mu_{out}(y) - u_0(x)| (1 - u(x)) w(x, y) dy dx, \quad u : \Omega \rightarrow [0, 1]$$

where  $\mu_{in}, \mu_{out}$  are in this case the weighted medians inside and outside the active contour. Medians are related with contrast invariant filters, and are more robust to a change of illumination. Finally, this model can also be combined with the segmentation model introduced in Section 3 by changing the TV energy in (17) with the NL-TV energy.

## 5. NON-LOCAL MUMFORD-SHAH MODEL

### 5.1. Motivations.

In Section 3, we extend the regularization term of energy (4) to the NL-TV. In Section 4, we extend the data-based term of energy (4) using the Chan-Vese Model. In this section, we consider another data-term extension of (4) based on the well-known Mumford-Shah (MS) energy [39]. The MS energy is defined as follows:

$$(22) \quad E_{MS}(u, C) = \int_{\Omega \setminus C} |\nabla u|^2 + \lambda \int_{\Omega} (u - u_0)^2 + \beta \int_C ds,$$

where  $u_0 : \Omega \rightarrow \mathbb{R}$  is the original image,  $u : \Omega \rightarrow \mathbb{R}$  is the regularized image,  $C : \gamma \rightarrow \mathbb{R}^2$  is the contour between smooth intensity regions and  $\int_C ds$  is the length of  $C$ . The MS energy (22) looks for the optimal pair  $(u, C)$  that represents the best piecewise smooth approximation of the given image where  $C$  is the boundary between smooth regions lying in the image. The existence of a minimizer for the MS energy is a difficult problem mainly

because of the presence of the length term. Nevertheless, it has been shown that minimizing solution lies in the space of functions with special bounded variation (SBV). See the book [1] for more details. The MS energy (22) is an optimal segmentation and denoising models for piecewise *smooth* images. However, this model fails to segment and denoise *textured* images, which limits the model to work optimally with real-world images. The main objective of this section is to extend the original MS model to define a piecewise *regular* approximation of the given image. Regular approximation means constant, smooth or/and textured approximation. The new MS energy (22) will use the non-local Dirichlet/ $H^1$  energy instead of the standard  $H^1$  energy  $\int |\nabla u|^2$ :

$$(23) \quad E_{MS}^{NL}(u, C) = \int_{\Omega \setminus C} |\nabla_{NL} u|^2 + \lambda \int_{\Omega} (u - u_0)^2 + \beta \int_C ds,$$

where the NL- $H^1$  is defined with (8) as:

$$\int |\nabla_{NL} u|^2 = \iint (u(y) - u(x))^2 w(x, y) dy dx,$$

where  $w$  are the weights defined as in [20] and Section 3, which correspond to the  $L^2$  distance between patches of image intensity. The next section is focused on the implementation of the NL-MS energy (23).

## 5.2. Numerical Schemes for the Non-Local Mumford-Shah Model.

### 5.2.1. Numerical Scheme based on Energy (4).

A direct computation of the Euler-Lagrange equations of (22) is not possible because the energy is not differentiable. There are two approaches to design a numerical scheme for the MS implementation:  $\Gamma$ -convergence approach and level set method. In this section, we investigate the approach based on the level set method as developed by Vese and Chan in [50]. In this approach, the contour  $C$  is embedded in a level set function. Then, using a regularized formulation of the length of  $C$ , authors in [50] differentiate the MS energy and get the Euler-Lagrange equations.

The MS energy (22) is defined in the context of energy (4) as follows:

$$(24) \quad E_{MS}^{VC}(u, r_{in}, r_{out}) = \int_{\Omega} |\nabla u| + \gamma (|\nabla r_{in}|^2 + \lambda (r_{in} - u_0)^2) u + \gamma (|\nabla r_{out}|^2 + \lambda (r_{out} - u_0)^2) (1 - u) dx, \quad u : \Omega \rightarrow [0, 1].$$

where  $u_0 : \Omega \rightarrow \mathbb{R}$  is the given image,  $r_{in}, r_{out} : \Omega \rightarrow \mathbb{R}$  are two functions. Then, the second term (data-based term) in (16) is then extended to a non-local/graph formulation as follows:

$$(25) \quad E_{MS}^{VC, NL}(u, r_{in}, r_{out}) = \int_{\Omega} |\nabla u| + \gamma (|\nabla_{NL} r_{in}|^2 + \lambda (r_{in} - u_0)^2) u + \gamma (|\nabla_{NL} r_{out}|^2 + \lambda (r_{out} - u_0)^2) (1 - u) dx, \quad u : \Omega \rightarrow [0, 1].$$

Minimizing (25) w.r.t. functions  $r_{in}$ :

$$\inf_{r_{in}} \int_{\Omega} (|\nabla_{NL} r_{in}|^2 + \lambda (r_{in} - u_0)^2) u dx, \\ \int_{\Omega \times \Omega} w(x, y) (r_{in}(y) - r_{in}(x))^2 u(x) dy dx + \int_{\Omega} \lambda (r_{in}(x) - u_0(x))^2 u(x) dx$$

provides the following Euler-Lagrange equations:

$$\int_{\Omega} w(x, y) (r_{in}(x) - r_{in}(y)) (u(x) + u(y)) dy + \lambda (r_{in}(x) - u_0(x)) u(x) = 0$$

which can be solved using a fixed point method:

$$r_{in}(x) = \frac{\lambda u(x)u_0(x) + \int_{\Omega} w(x,y)r_{in}(y)(u(x) + u(y))dy}{\lambda u(x) + \int_{\Omega} w(x,y)(u(x) + u(y))dy}.$$

Equivalently, the minimizing solution w.r.t. function  $r_{out}$  is:

$$r_{out}(x) = \frac{\lambda(1 - u(x))u_0(x) + \int_{\Omega} w(x,y)r_{out}(y)(2 - u(x) - u(y))dy}{\lambda(1 - u(x)) + \int_{\Omega} w(x,y)(2 - u(x) - u(y))dy}.$$

As in Section 4.2, energy (25) is minimized w.r.t.  $u$  using the same technique as in [6]. We re-write (25) as follows:

$$E_{MS}^{VC,NL}(u) = \int_{\Omega} |\nabla u| + \gamma V(x)u(x) + C,$$

where  $V(x) = (|\nabla_{NL}r_{in}|^2 + \lambda(r_{in} - u_0)^2) - (|\nabla_{NL}r_{out}|^2 + \lambda(r_{out} - u_0)^2),$

where  $C$  is a constant independent of  $u$ , useless in the minimization process w.r.t.  $u$ . Then, we split the original energy into two simpler minimization tasks by introducing a new function  $v : \Omega \rightarrow \mathbb{R}$  as follows:

$$(26) \quad \inf_{v \in [0,1], u} E_{MS}^{NL,2}(u, v) = \int_{\Omega} |\nabla u| + \gamma Vv + \frac{1}{2\theta}(u - v)^2, u : \Omega \rightarrow \mathbb{R}, v : \Omega \rightarrow [0, 1].$$

(26) is convex w.r.t  $u, v$  then the minimization is done alternatively w.r.t.  $u, v$ . The minimization w.r.t.  $u$  is given by the minimizer of the ROF model and the minimization w.r.t.  $v$  is given by  $v = \min(\max(u - \gamma\theta V, 1), 0)$ .

Figure 9 presents the segmentation of a textured square, which cannot be segmented with the original Mumford-Shah model [39].

[ Figure 9. ]

### 5.2.2. Numerical Scheme based on Elliptic Approximation of Length.

In the previous section, we develop a numerical scheme for the NL-MS model based on [50] and energy (4). As we previously said, there is another approach, based on the  $\Gamma$ -convergence tool, to deal with the implementation of the MS energy. This approach consists in approximating the MS energy by a sequence  $E_{MS}^{\varepsilon}$  of regular functionals defined on Sobolev spaces and prove the convergence of  $E_{MS}^{\varepsilon}$  to  $E_{MS}$  as  $\varepsilon \rightarrow 0$  in the  $\Gamma$ -convergence framework. Several approximations of the MS energy has thus been introduced in the literature. Our numerical scheme will be based on the elliptic approximation of the MS energy as introduced in [2]. We will see that not only the NL-MS is able to segment textured images unlike the original model, the NL-MS also provides better denoising results than the state-of-the-art. This is not surprising because a piecewise regular approximation of the MS energy is a more realistic model. The elliptic approximation of the NL-MS energy is defined as follows:

$$(27) \quad E_{MS,\varepsilon}^{NL}(u, v) = \int_{\Omega} v^2 |\nabla_{NL}u|^2 + \frac{\lambda}{2} \int_{\Omega} (u - u_0)^2 + \frac{\beta}{2} \left( \varepsilon |\nabla v|^2 + \frac{(v - 1)^2}{4\varepsilon} \right),$$

where  $u, v, u_0 : \Omega \rightarrow \mathbb{R}$ ,  $w(x, y) : \Omega \times \Omega \rightarrow \mathbb{R}_+$  and  $\lambda, \beta, \varepsilon > 0$ . The last term of (27) approximates the length of the discontinuities between piecewise regular regions. Euler-Lagrange equations of (27) lead to the minimizing flow for  $u$ :

$$(28) \quad \begin{aligned} u_t &= -(2 \operatorname{div}_{NL}(v^2 \nabla_{NL}u) + \lambda(u - u_0)), \\ &= -\left(2 \int_{\Omega} (v^2(x) + v^2(y))w(x, y)(u(y) - u(x))dy + \lambda(u(x) - u_0(x))\right), \end{aligned}$$

and  $v$ :

$$(29) \quad \begin{aligned} v_t &= -\left(2v|\nabla_{\text{NL}}u|^2 + \beta\left(\varepsilon|\nabla v|^2 + \frac{(v-1)^2}{4\varepsilon}\right)\right), \\ &= -\left(2v(x)\int_{\Omega}(u(y)-u(x))^2w(x,y)dy + \beta\left(\varepsilon\Delta v(x) + \frac{v(x)-1}{2\varepsilon}\right)\right). \end{aligned}$$

We observe that minimizing flows (28) and (29) are NL extensions of standard MS minimizing flows. The first term of (28) is like a non-local Perona-Malik term [44] which is useful to stop the diffusion flow along edges between regular (textured) regions.

Figures 10 and 11 present our experiments. We observe that the NL-MS model is not only able to segment textures but it simultaneously denoises images better than the standard model NL- $H^1$  [20].

[ Figure 10. ]

[ Figure 11. ]

## 6. CONCLUSION

In this paper, we have proposed three unsupervised image segmentation models based on non-local image information. These three non-local segmentation models are based on the continuous global minimization approach for image segmentation recently introduced in [10, 6]. Results show the improvements of standard segmentation algorithms in a "non-local" framework.

We remind in Section 2 the advantages of the continuous global minimization scheme for image segmentation introduced in [10, 6]. Any active contour model, that are usually solved by the level set method, can be solved in a efficient way with this new segmentation approach. Besides, the minimization is global, i.e. no good initial condition are needed to get a global solution and it can be efficiently implemented using fast continuous minimization schemes. Section 3 has extended the Total Variation regularization term of [10, 6] to the non-local Total Variation energy. We have seen that the regularization process defined on a graph of image patches can segment fine and small structures better than the standard Total Variation energy. We have presented a continuous minimization algorithm and a fast graph cut minimization algorithm. Section 4 has extended the data-based term of [10, 6] to a non-local term using the CV model. The proposed NL-CV model has overcome the main limitation of the original CV model, that does not work with local intensity inhomogeneities. Finally, Section 5 has extended the data-based term of [10, 6] to a non-local term using the Mumford-Shah model. The original Mumford-Shah energy is designed to work for piecewise smooth images only. Our proposed NL-MS is defined for piecewise smooth or/and textured images. Two continuous numerical algorithms are introduced. The first algorithm is based on a level set formulation of the original MS energy. The second model is based on an elliptic approximation of the original energy. The NL-MS model not only segments textured images but also produces better denoising results than state-of-the-art.

We notice that the NL-TV regularization term used in Section 3 can also be used in Sections 4 and 5 to improve the segmentation results. Future works will investigate the improvements of the weight function (21) used in the NL-CV model. These weights do not use any prior image information such as patches of image intensity. The computational speed of the NL-CV and the NL-MS can also be improved.

## 7. ANNEX

### 7.1. Proof of Theorem 2.1.

**Theorem 2.1.1:** *Suppose that  $g_b : \Omega \rightarrow \mathbb{R}_+$ , for any given  $g_r^{in}, g_r^{out} : \Omega \rightarrow \mathbb{R}$  and  $\lambda \in \mathbb{R}_+$ , if  $u_*$  is any minimizer of  $E_{GMAC}$ , then for almost every  $\nu \in \mathbb{R}$  we have that the characteristic/indicator function of sets  $\Omega_C(\nu) := \{x \in \Omega : u(x) > \nu\}$  (where  $C$  is the boundary of the set  $\Omega_C$ ) is a global minimizer of  $E_{GMAC}$  and  $E_{AC}$  defined in (4) and (1).*

**Proof.** The proof is based on [10, 6]. First, it consists in using the co-area formula to decompose the weighted TV norm (using  $g_r := g_r^{in} - g_r^{out}$ ):

$$\begin{aligned} \inf_u E_{GMAC}(u) &= \int_{\Omega} g_b |\nabla u| + \lambda g_r u dx \\ &= \int_{\nu} \left( \int_{\Omega} g_b |\chi_{\Omega_C(\nu)}| + \lambda g_r \chi_{\Omega_C(\nu)} dx \right) d\nu \\ &= \int_{\nu} \left( Per_{g_b}(\Omega_C(\nu)) + \lambda Area_{g_r}(\Omega_C(\nu)) \right) d\nu. \end{aligned}$$

Thus, the minimization consists in solving the geometric problem:

$$(30) \quad \min_{\Omega} Per_{g_b}(\Omega) + \lambda Area_{g_r}(\Omega)$$

for a.e.  $\nu$  and  $Per_{g_b}$  and  $Area_{g_r}$  are the weighted length and area of  $\hat{\Omega}$ . Theory of functions with bounded variation always states the existence of a minimizer  $\hat{\Omega}$ , which implies:

$$(31) \quad Per_{g_b}(\Omega_{u_*}(\nu)) + \lambda Area_{g_r}(\Omega_{u_*}(\nu)) \geq Per_{g_b}(\hat{\Omega}) + \lambda Area_{g_r}(\hat{\Omega}), \quad \forall \nu,$$

where  $\Omega_{u_*}(\nu) := \{x \in \Omega : u_*(x) > \nu\}$ . If we integrate w.r.t.  $\nu$  then  $E_{GMAC}(u_*) \geq E_{GMAC}(\chi_{\hat{\Omega}})$ . Thus, if  $\hat{\Omega}$  is a minimizer of (30), then the characteristic function  $\chi_{\hat{\Omega}}$  is a minimizer of  $E_{GMAC}$ . However, we also have  $E_{GMAC}(u_*) \leq E_{GMAC}(\chi_{\hat{\Omega}})$  since  $u_*$  is also a minimizer of  $E_{GMAC}$ . Thus,  $E_{GMAC}(u_*) = E_{GMAC}(\chi_{\hat{\Omega}})$  and the inequality in (31) is actually an equality which implies  $\hat{\Omega} = \Omega_{u_*}(\nu)$  for a.e.  $\nu$ . Hence,  $\Omega_{u_*}(\nu)$  is a minimizer of (30) for a.e.  $\nu$  and the characteristic function  $\chi_{\Omega_C(\nu)}$  is also a minimizer of  $E_{GMAC}$ . Finally, characteristic functions of sets  $\Omega_C(\nu)$  globally minimized the AC energy (1) since  $E_{GMAC}(u = \chi_{\Omega_C}) = E_{AC}(C)$ .  $\square$

## 7.2. Proof of Proposition 3.4.1.

**Proposition 3.4.1:** *Let us assume that the set of patches in an image belongs to a compact differentiable submanifold  $\mathcal{M} \subset \mathbb{R}^n$ , where the dimension of  $\mathcal{M}$  is much smaller than the ambient space  $d < n$ . If  $u \in L^1(\mathcal{M})$ , then, the NL-TV norm converges to the TV norm on  $\mathcal{M}$ :*

$$\int_{\Omega} \frac{1}{\epsilon^{\frac{d+2}{4}}} |\nabla_{NL\epsilon} u| \xrightarrow{\epsilon \rightarrow 0} \int_{\mathcal{M}} |\nabla_{\mathcal{M}} u|.$$

**Proof.** As we said, we made the assumption that the set of all patches  $\{p_x | x \in \Omega\}$  are samples from a  $d$ -dimensional submanifold  $\mathcal{M} \subset \mathbb{R}^n$ . We use the work of Coifman and Lafon [13, 33, 3, 27] to show that the graph-based TV norm/NL-TV norm converges to the TV norm on manifold. In [13, 33], authors prove that the graph Laplacian converges to the Laplace-Beltrami operator on the manifold  $\mathcal{M}$ . The square norm of the graph-based gradient of function  $u$  at the point  $x \in \Omega$  is by definition (8):

$$|\nabla_{NL\epsilon} u|^2(x) = \int_{\Omega} w_{\epsilon}(x, y) (u(y) - u(x))^2 dy$$

The weight function  $w_{\epsilon}(x, y)$  is defined by:

$$(32) \quad w_{\epsilon}(x, y) = \exp\left(-\frac{\|p_x - p_y\|_2^2}{\epsilon}\right) = w_{\epsilon}(p_x, p_y),$$

where  $p_x, p_y$  are patches of intensity centered at  $x, y$ . (32) means that the weight function  $w_{\epsilon}$  for  $(x, y)$  has the same value for patches  $(p_x, p_y)$  (See Figure 1). Besides, function  $u$  at

$x$  has the same value for patch  $p_x$ . Thus, we can write:

$$\begin{aligned} |\nabla_{\text{NL}\epsilon} u|^2(x) &= \int_{\Omega} w_{\epsilon}(x, y)(u(y) - u(x))^2 dy \\ (33) \qquad \qquad \qquad &= \int_{\mathcal{M}} w_{\epsilon}(p_x, p_y)(u(p_y) - u(p_x))^2 dp_y = |\nabla_{\text{NL}\epsilon} u|^2(p_x), \end{aligned}$$

since it is assumed that the set of all patches  $\{p_x | x \in \Omega\}$  are samples from  $\mathcal{M}$ . We consider the new notation  $\hat{x} = p_x, \hat{y} = p_y$  such that:

$$|\nabla_{\text{NL}\epsilon} u|^2(\hat{x}) = \int_{\mathcal{M}} w_{\epsilon}(\hat{x}, \hat{y})(u(\hat{y}) - u(\hat{x}))^2 d\hat{y}$$

with  $w_{\epsilon}(\hat{x}, \hat{y}) = h(\frac{\|\hat{x} - \hat{y}\|^2}{\epsilon})$  exponentially decays. For  $\hat{y} \in \mathcal{M}$ , we consider the orthogonal projection  $v$  of  $\hat{y}$  on  $T_{\hat{x}}$ , the tangent plane to  $\mathcal{M}$  at  $\hat{x}$ . We can show that:

$$(34) \qquad \qquad \qquad |\nabla_{\epsilon} u|^2(\hat{x}) = \epsilon^{d/2} (\epsilon m_0 |\nabla_{\mathcal{M}} u|^2(\hat{x}) + O(\epsilon^{3/2})).$$

The Taylor expansion of  $u$  around  $\hat{x}$  is given by:

$$u(\hat{y}) = u(\hat{x}) + \sum_{i=1}^d s_i \frac{\partial u}{\partial s_i}(\hat{x}) + \text{Higher order terms.}$$

where  $s_i$  are the coordinate system along orthogonal geodesics on  $\mathcal{M}$ . [13, 33] show that if  $\hat{y} \in \mathcal{M}$  is in an Euclidean ball of radius  $C\sqrt{\epsilon}$ , then  $s_i = v_i + O(\epsilon^{3/2})$ , which implies:

$$(35) \qquad \qquad \qquad u(\hat{y}) = u(\hat{x}) + \sum_{i=1}^d v_i \frac{\partial u}{\partial s_i}(\hat{x}) + O(\epsilon).$$

[13, 33] also deduce that

$$(36) \qquad \qquad \qquad \|\hat{y}(v)\|^2 = \|v\|^2 + \left(\sum_{i=1}^d a_i v_i^2\right)^2 + O(\epsilon^{5/2})$$

$$(37) \qquad \qquad \qquad d\hat{y} = \left(1 + 2 \sum_{i=1}^d a_i^2 v_i^2\right) dv + O(\epsilon^{3/2})$$

where  $a_i$  is the curvature of the geodesic. Collecting (35), (36) and (37), we have

$$\begin{aligned} |\nabla_{\text{NL}\epsilon} u|^2(\hat{x}) &= \int_{\mathcal{M}} \left[ h\left(\frac{\|v\|^2}{\epsilon}\right) + \frac{1}{\epsilon} \left(\sum_{i=1}^d a_i v_i^2\right)^2 h'\left(\frac{\|v\|^2}{\epsilon}\right) \right] \\ &\quad \times \left[ u(\hat{x}) + \sum_{i=1}^d v_i \frac{\partial u}{\partial s_i}(\hat{x}) + O(\epsilon) - u(\hat{x}) \right]^2 \\ &\quad \times \left[ 1 + 2 \sum_{i=1}^d a_i^2 v_i^2 \right] dv + O(\epsilon^{3/2}) \\ &= \sum_{i=1}^d \left(\frac{\partial u}{\partial s_i}\right)^2(\hat{x}) \int_{\mathcal{M}} v_i^2 h\left(\frac{\|v\|^2}{\epsilon}\right) dv + \text{Higher order terms} \\ &= |\nabla_{\mathcal{M}} u|^2(\hat{x}) \int_{\mathbb{R}^d} v^2 h\left(\frac{\|v\|^2}{\epsilon}\right) dv + O(\epsilon^{(d+3)/2}), \end{aligned}$$

since  $h$  quickly decays. This concludes the proof of (34) considering  $\int_{\mathbb{R}^d} v^2 h(\frac{\|v\|^2}{\epsilon}) dv = \epsilon^{\frac{d+2}{2}} \int_{\mathbb{R}^d} v^2 h(\|v\|^2) dv = \epsilon^{\frac{d+2}{2}} m_0$ . Thus, we define the graph-based norm of the gradient operator as follows:

$$|\nabla_{\mathcal{M}\epsilon} u|(\hat{x}) := \frac{1}{\epsilon^{\frac{d+2}{4}}} |\nabla_{\text{NL}\epsilon} u|(\hat{x})$$

which gives the norm of the gradient operator on the manifold  $\mathcal{M}$  at the limit:

$$|\nabla_{\mathcal{M}\epsilon} u|(\hat{x}) \xrightarrow{\epsilon \rightarrow 0} |\nabla_{\mathcal{M}} u|(\hat{x}).$$

Finally, we have

$$\int_{\Omega} \frac{1}{\epsilon^{\frac{d+2}{4}}} |\nabla_{\text{NL}\epsilon} u|(x) dx \xrightarrow{\epsilon \rightarrow 0} \int_{\mathcal{M}} |\nabla_{\mathcal{M}} u|(\hat{x}) d\hat{x}. \quad \square$$

#### REFERENCES

- [1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, 2000.
- [2] L. Ambrosio and V.M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via  $\Gamma$ -convergence. *Communications on Pure and Applied Mathematics*, XLIII:999–1036, 1990.
- [3] M. Belkin. Problems of Learning on Manifolds, Ph.D. Dissertation, 2003.
- [4] Y. Boykov and G. Funka-Lea. Graph Cuts and Efficient N-D Image Segmentation. *International Journal of Computer Vision (IJCV)*, 70(2):109–131, 2006.
- [5] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, 2004.
- [6] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Fast Global Minimization of the Active Contour/Snake Models. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.
- [7] A. Buades, B. Coll, and J.M Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modeling and Simulation (MMS)*, 4(2):490–530, 2005.
- [8] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.
- [9] A. Chambolle. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.
- [10] T.F. Chan, S. Esedoglu, and M. Nikolova. Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.
- [11] T.F. Chan and L.A. Vese. Active Contours Without Edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- [12] Y. Chen, H.D. Tagare, S. Thiruvankadam, F. Huang, D. Wilson, K.S. Gopinath, R.W. Briggsand, and E.A. Geiser. Using Prior Shapes in Geometric Active Contours in a Variational Framework. *International Journal of Computer Vision*, 50(3):315–328, 2002.
- [13] R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis (ACHA)*, 21(1):5–30, 2006.
- [14] M.G. Crandall, H. Ishii, and P.L. Lions. Users’ guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematic Society*, 27(1):1–69, 1992.
- [15] J. Darbon and M. Sigelle. Image Restoration with Discrete Constrained Total Variation Part I: Fast and Exact Optimization. *Journal of Mathematical Imaging and Vision*, 26(3):277–291, 2006.
- [16] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. NorthHolland, Amsterdam, 1976.
- [17] N. El-Zehiry, S. Xu, P. Sahoo, and A. Elmaghraby. Graph Cut Optimization for the Mumford-Shah Model. In *Visualization, Imaging, and Image Processing*, 2007.
- [18] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton Univ. Press, 1962.
- [19] D. Geman and S. Geman. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [20] G. Gilboa and S. Osher. Nonlocal Linear Image Regularization and Supervised Segmentation. *SIAM Multiscale Modeling and Simulation (MMS)*, 6(2):595630, 2007.
- [21] G. Gilboa and S. Osher. Nonlocal Operators with Applications to Image Processing. *CAM Report 07-23*, 2007.
- [22] A.V. Goldberg and R.E. Tarjan. A New Approach to the Maximum-Flow Problem. *J. ACM*, 35(4):921–940, 1988.
- [23] T. Goldstein and S. Osher. The Split Bregman Method for L1 Regularized Problems, CAM Report 08-29, 2008.
- [24] D. Greig, B. Porteous, and A. Seheult. Exact Maximum A Posteriori Estimation for Binary Images. *J. Royal Statistical Soc. Series B*, 51(2):271–279, 1989.
- [25] M. Hein, J. Audibert, and U. Von Luxburg. From Graphs to Manifolds - Weak and Strong Pointwise Consistency of Graph Laplacians. In *Conference on Learning Theory*, pages 470–485, 2005.
- [26] N. Houhou, J.-P. Thiran, and X. Bresson. Fast Texture Segmentation Model based on the Shape Operator and Active Contour. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [27] P.S. Huggins and S.W. Zucker. Representing Edge Models via Local Principal Component Analysis. In *European Conference on Computer Vision (ECCV)*, pages 384–398, 2002.
- [28] G.-S. Jiang and D. Peng. Weighted ENO Schemes for Hamilton-Jacobi Equations. *Journal of Scientific Computing*, 21(6):2126–2143, 1999.
- [29] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, pages 321–331, 1987.
- [30] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A.J. Yezzi. Conformal Curvature Flows: From Phase Transitions to Active Vision. In *Archive for Rational Mechanics and Analysis*, volume 134, pages 275–301, 1996.

- [31] K. Kolev, M. Klodt, T. Brox, S. Esedoglu, and D. Cremers. Continuous Global Optimization in Multiview 3D Reconstruction. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 4679, pages 441–452, 2007.
- [32] G. Kuhne, J. Weickert, M. Beier, and W. Effelsberg. Fast Implicit Active Contour Models. In *DAGM Symposium on Pattern Recognition*, pages 133–140, 2002.
- [33] S. Lafon. Diffusion Maps and Geometric Harmonics, 2004.
- [34] S. Leung and S. Osher. Global Minimization of the Active Contour Model with TV-Inpainting and Two-Phase Denoising. In *Variational, Geometric, and Level Set Methods in Computer Vision (VLSM)*, Springer-Verlag, vol. 3752, pages 149–160, 2005.
- [35] C. Li, C.-Y. Kao, J.C. Gore, and Z. Ding. Implicit Active Contours Driven by Local Binary Fitting Energy. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2007.
- [36] C. Li, C.-Y. Kao, J.C. Gore, and Z. Ding. Minimization of Region-Scalable Fitting Energy for Image Segmentation. *IEEE Transactions on Image Processing*, 17(10):1940–1949, 2008.
- [37] B. Merriman, J.K. Bence, and S.J. Osher. Motion of multiple junctions: a level set approach. *Journal of Computational Physics*, 112(2):334–363, 1994.
- [38] B. Mory and R. Ardon. Fuzzy Region Competition: A Convex Two-Phase Segmentation Framework. In *Scale Space Variational Methods (SSVM)*, pages 214–226, 2007.
- [39] D. Mumford and J. Shah. Optimal Approximations of Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, 42:577–685, 1989.
- [40] K. Ni, X. Bresson, T. Chan, and S. Esedoglu. Local Histogram-Based Segmentation Using the Wasserstein Distance. (Submitted), 2008.
- [41] S. Osher. "Level Set Methods", in *Geometric Level Set Methods in Imaging, Vision and Graphics*. eds. S. Osher and N. Paragios, Springer-Verlag, NY, pages 3–20, 2003.
- [42] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer-Verlag, NY, 2003.
- [43] S. Osher and J.A. Sethian. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics*, 79(1)(12-49), 1988.
- [44] P. Perona and J. Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1252(629-639), 1990.
- [45] T. Pock, M. Unger, D. Cremers, and H. Bischof. Fast and Exact Solution of Total Variation Models on the GPU. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [46] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60(1-4):259 – 268, 1992.
- [47] J.A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Material Sciences*. Cambridge University Press, 1999.
- [48] A.D. Szlam, M. Maggioni, and R.R. Coifman. Regularization on Graphs with Function Adapted Diffusion Processes. *Journal of Machine Learning Research*, 2008.
- [49] M. Unger, T. Pock, and H. Bischof. Continuous Globally Optimal Image Segmentation with Local Constraints. In *Computer Vision Workshop*, 2008.
- [50] L.A. Vese and T.F. Chan. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *International Journal of Computer Vision*, 50(3):271–293, 2002.
- [51] C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. In *Lecture Notes in Computer Science, Pattern Recognition*, volume 4713, pages 214–223, 2007.
- [52] C. Zach, T. Pock, and H. Bischof. A Globally Optimal Algorithm for Robust TV-L1 Range Image Integration. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–8, 2007.
- [53] H. Zhao, S. Osher, B. Merriman, and M. Kang. Implicit and Nonparametric Shape Reconstruction from Unorganized Data using a Variational Level Set Method. *Computer Vision and Image Understanding*, 80(3):295314, 2001.
- [54] D. Zhou and B. Scholkopf. A Regularization Framework for Learning from Graph Data. In *Workshop on Statistical Relational Learning and Its Connections to Other Fields*, 2004.
- [55] S. C. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

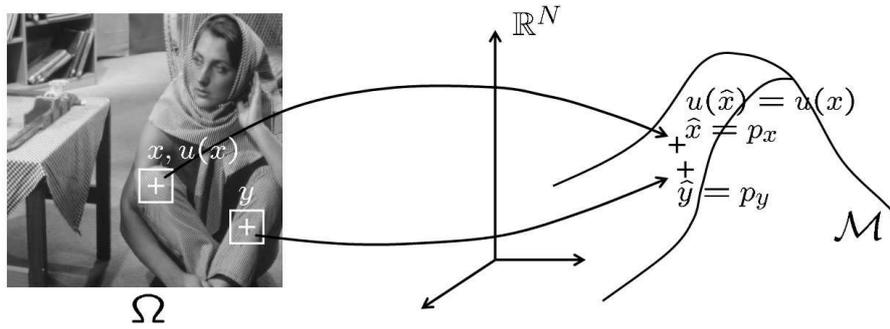
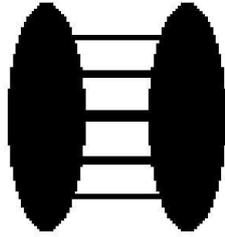
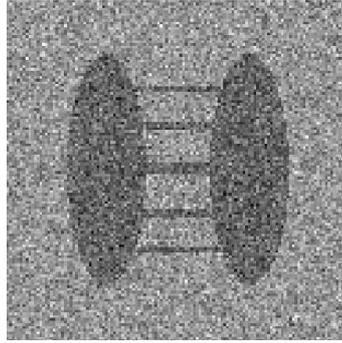


FIGURE 1. Illustration of the new image space representation  $\mathcal{M} \subset \mathbb{R}^N$  defined from image  $u$  in the spatial image domain  $\Omega$ . Two points  $x, y$  far away on  $\Omega$  can be close on the space of image patches  $\mathcal{M}$  if the intensity patches  $p_x, p_y = \hat{x}, \hat{y}$  are similar. The set of all image patches are samples of  $\mathcal{M}$ .



(a) Synthetic Image.



(b) Synthetic Image corrupted by Gaussian Noise.

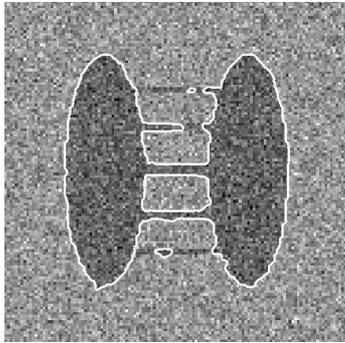
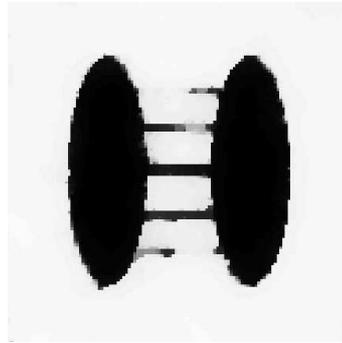
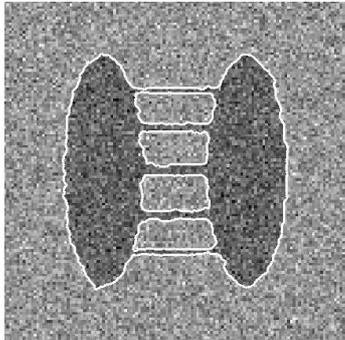
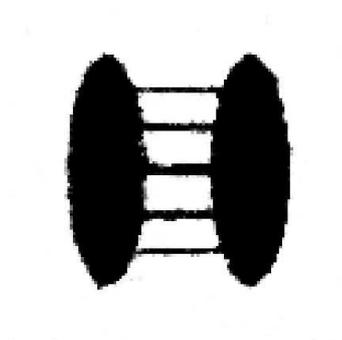
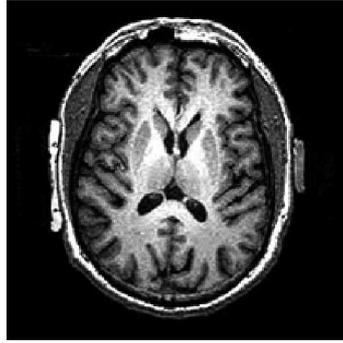
(c) *Standard TV + CV's Data Term.* Final Contour.(d) *Standard TV + CV's Data Term.* Final function  $u$ .(e) *NL-TV + CV's Data Term.* Final Contour.(f) *NL-TV + CV's Data Term.* Final function  $u$ .

FIGURE 2. Non-local Segmentation model (7) with NL-TV + Chan-Vese's data term. We observe that the NL segmentation model can segment fine structures (Figures (e-f)) unlike the standard model (Figures (c-d)). The minimization process is done by a *continuous* projection algorithm, see Section 3.5.1. The image size is  $128 \times 128$ . The iterative minimization process for continuous NL-TV+CV takes 120 iterations and the computational time is 3.7 sec.



(a) IRM Brain Image.

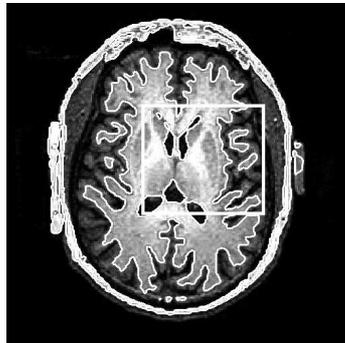
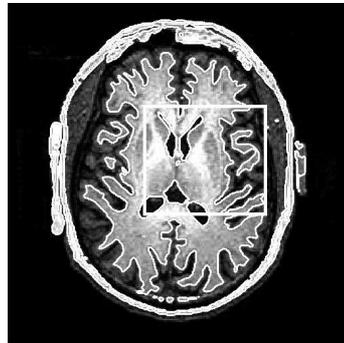
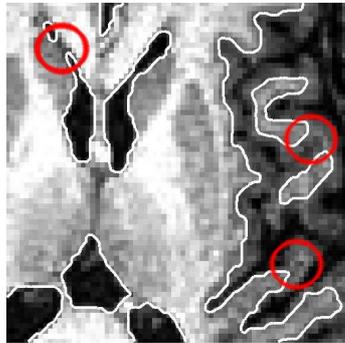
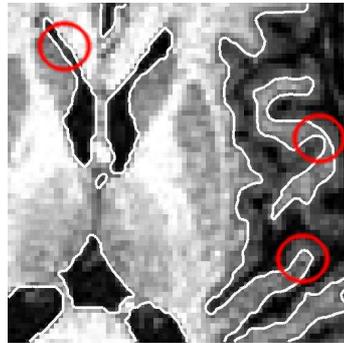
(b) *Standard TV + CV's Data Term*. Final Contour.(c) *NL-TV + CV's Data Term*. Final Contour.(d) *Standard TV + CV's Data Term* (Zoom In). Final Contour.(e) *NL-TV + CV's Data Term* (Zoom In). Final Contour.

FIGURE 3. Non-local Segmentation model (7) with NL-TV + Chan-Vese's data term. We observe that the NL segmentation model can segment fine structures (Figures (c) and (e)) unlike the standard model (Figures (b) and (d)). The minimization process is done by a *continuous* projection algorithm, see Section 3.5.1. The image size is  $210 \times 210$ . The iterative minimization process for continuous NL-TV+CV takes 30 iterations and the computational time is 1.8 sec.

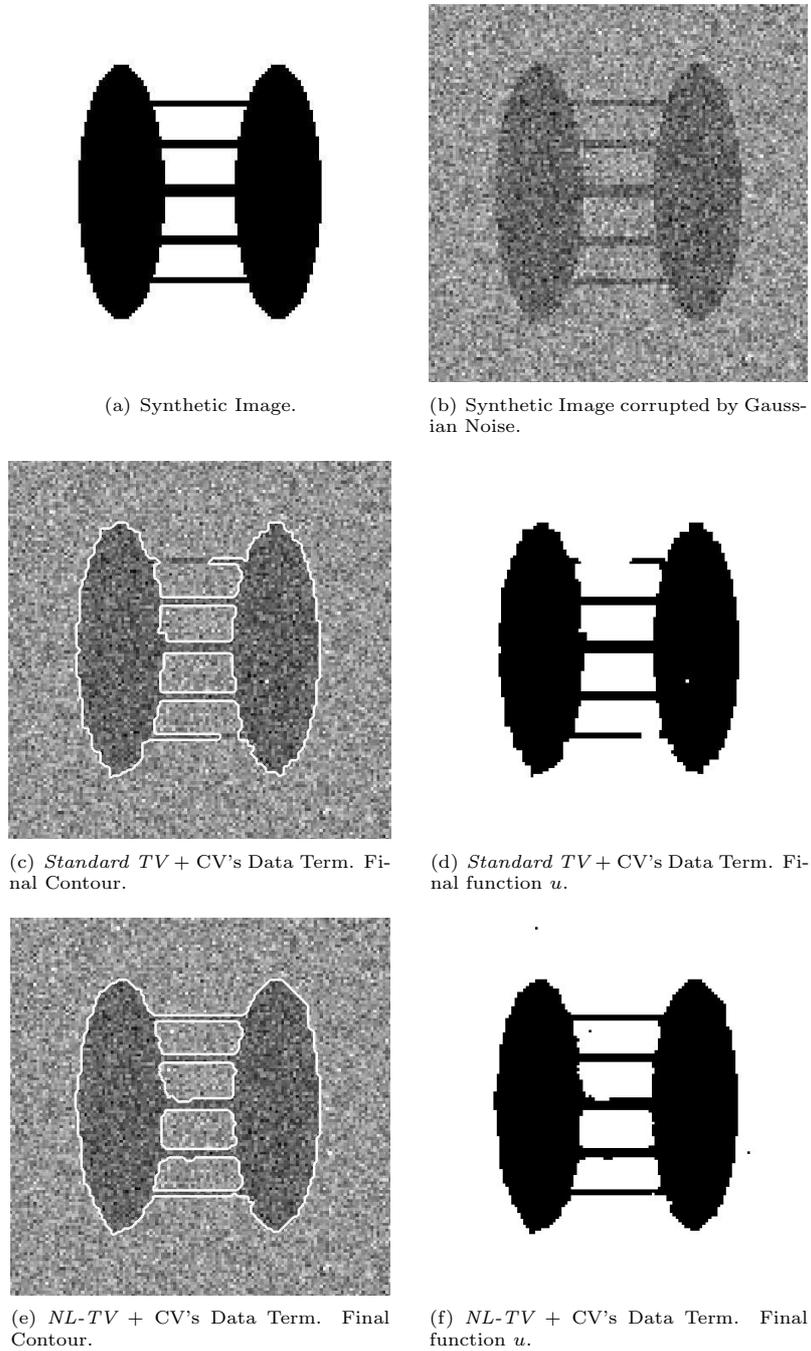
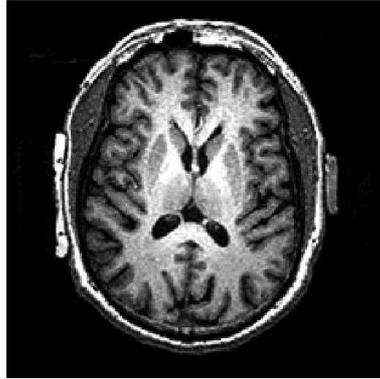


FIGURE 4. Non-local Segmentation model (15) with NL-TV + Chan-Vese's data term. Unlike Figure 2, the minimization process is done with a *discrete/graph cut* algorithm, see Section 3.5.2. The image size is  $128 \times 128$ . The iterative minimization process for discrete NL-TV+CV takes 10 iterations (update of inside and outside intensity means + energy minimization by graph cut) and the computational time is 0.5 sec. It is faster than the continuous approach but the result does not have a sub-pixel accuracy.



(a) IRM Brain Image.

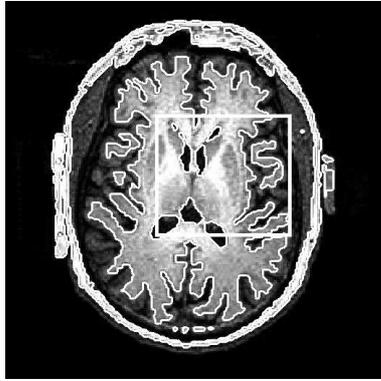
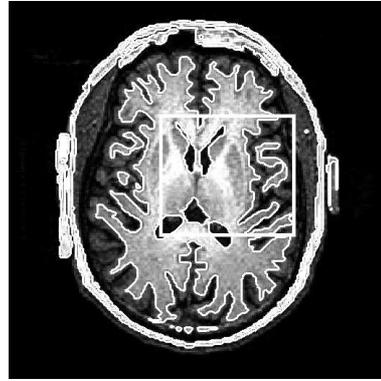
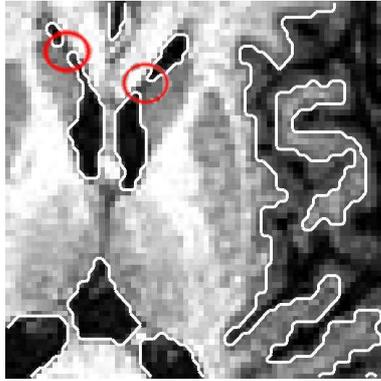
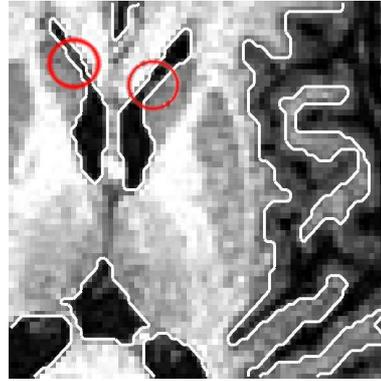
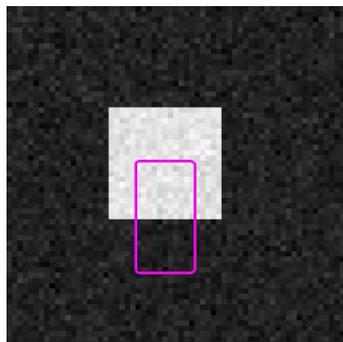
(b) *Standard TV + CV's Data Term*. Final Contour.(c) *NL-TV + CV's Data Term*. Final Contour.(d) *Standard TV + CV's Data Term* (Zoom In). Final Contour.(e) *NL-TV + CV's Data Term*. (Zoom In). Final Contour.

FIGURE 5. Non-local Segmentation model (15) with NL-TV + Chan-Vese's data term. Unlike Figure 3, the minimization process is done with a *discrete/graph cut* algorithm, see Section 3.5.2. The image size is  $210 \times 210$ . The iterative minimization process for discrete NL-TV+CV takes 12 iterations (update of inside and outside intensity means + energy minimization by graph cut) and the computational time is 0.96 sec. It is faster than the continuous approach but the result does not have a sub-pixel accuracy.



(a) Synthetic Image. Initial Contour.

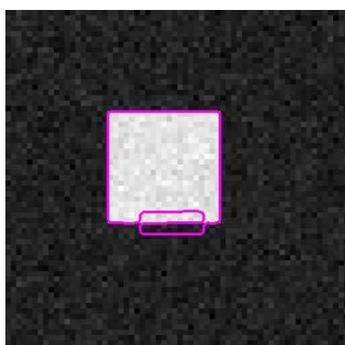
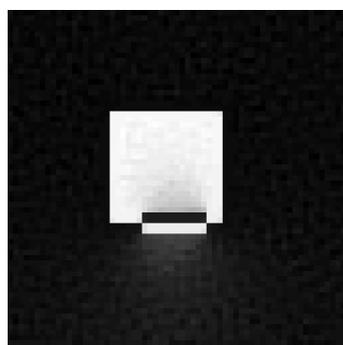
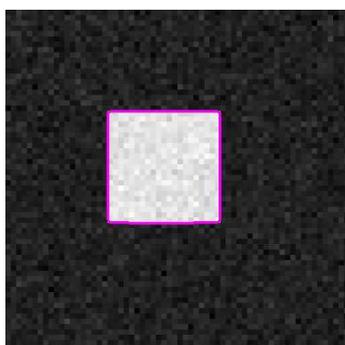
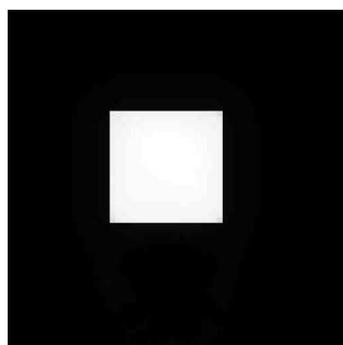
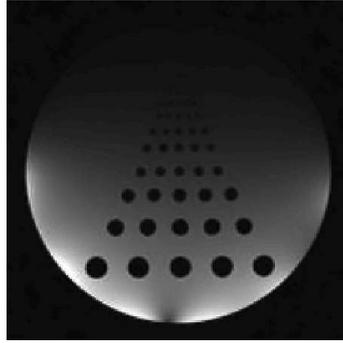
(b) NL-CV *without* Global Information. Final Contour.(c) NL-CV *without* Global Information. Final function  $u$ .(d) NL-CV *with* Global Information. Final Contour.(e) NL-CV *with* Global Information. Final function  $u$ .

FIGURE 6. Non-local Chan-Vese model (17). The NL-CV using local information *only* can produce some inconsistent segmentation results (Figures (b-c)). These inconsistencies can be removed using global image information at the same time (Figures (d-e)), avoiding bad local minimum.



(a) Synthetic Image.

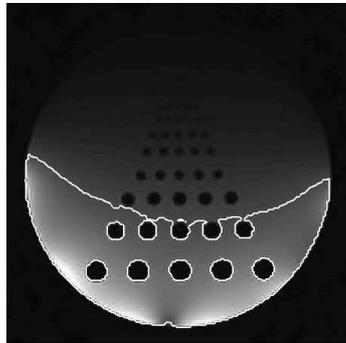
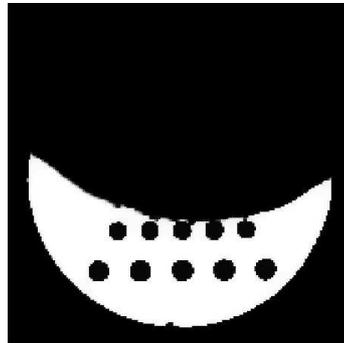
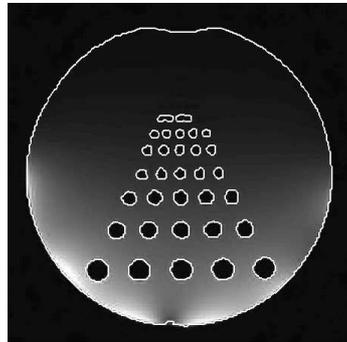
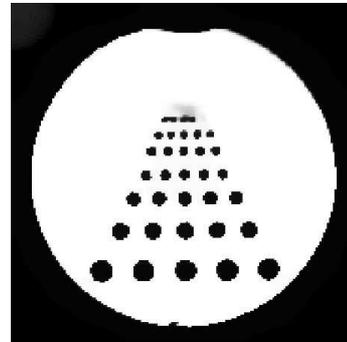
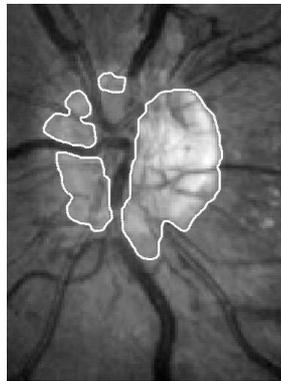
(b) *Standard CV Model*. Final Contour.(c) *Standard CV Model*. Final function  $u$ .(d) *Non-local CV Model*. Final Contour.(e) *Non-local CV Model*. Final function  $u$ .

FIGURE 7. Non-local Chan-Vese model (17). The NL-CV can segment objects with non-homogeneous intensities (Figures (d-e)), unlike the original CV model (Figures (b-c)). The image size is  $183 \times 183$ . The iterative minimization process for the NL-CV takes 280 iterations and the computational time is 71 sec.



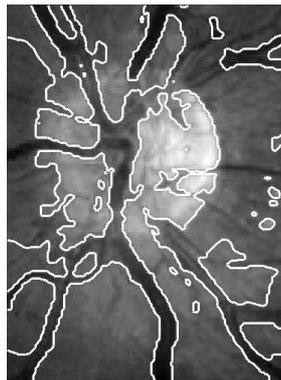
(a) Picture of Eye Blood Vessels.



(b) *Standard CV Model*. Final Contour.



(c) *Standard CV Model*. Final function  $u$ .



(d) *Non-local CV Model*. Final Contour.



(e) *Non-local CV Model*. Final function  $u$ .

FIGURE 8. Non-local Chan-Vese model (17). Unlike the original CV model (Figures (b-c)), the NL-CV can segment the eye blood vessels that have intensity inhomogeneities (Figures (d-e)). The image size is  $106 \times 145$ . The iterative minimization process for the NL-CV takes 130 iterations and the computational time is 28 sec.

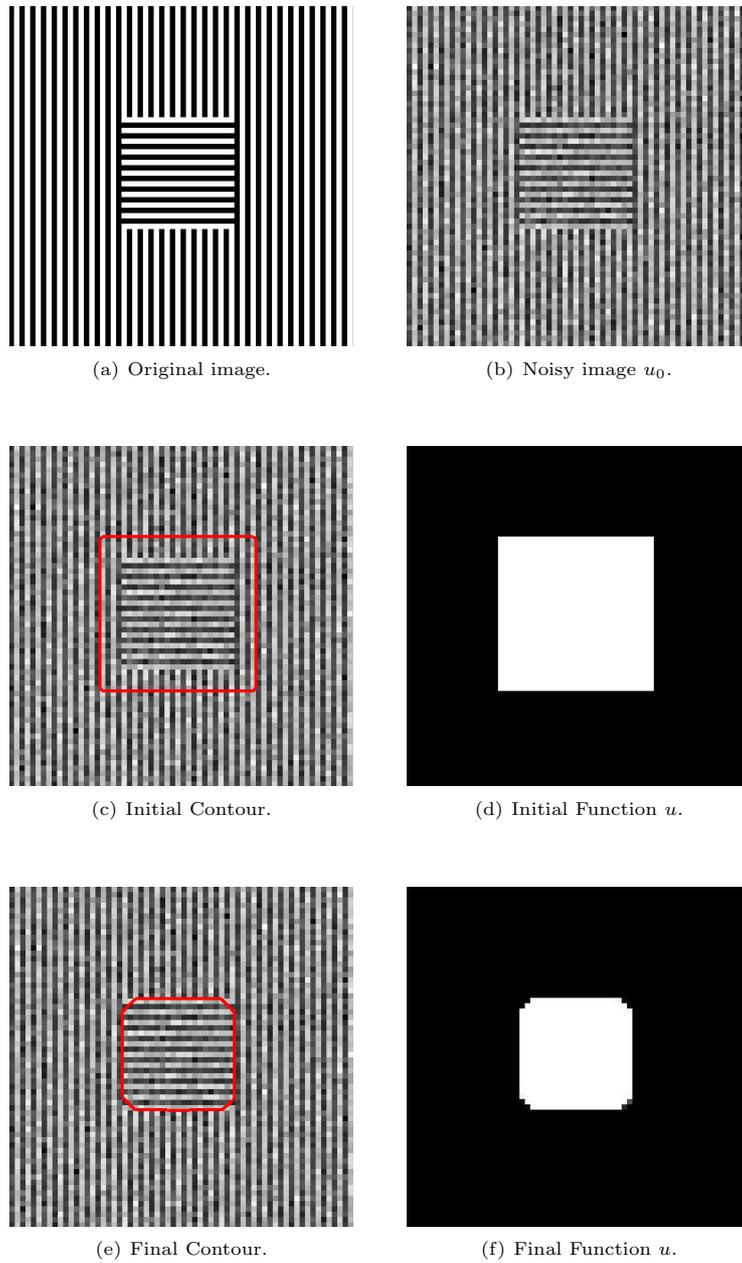


FIGURE 9. Non-local Mumford-Shah model (25). The NL-MS model can segment textures. The original MS model [39] is not able to segment the textured square because the mean value inside and outside the boundary are the same. The image size is  $64 \times 64$ . The iterative minimization process for NL-MS takes 90 iterations and the computational time is 0.2 sec.

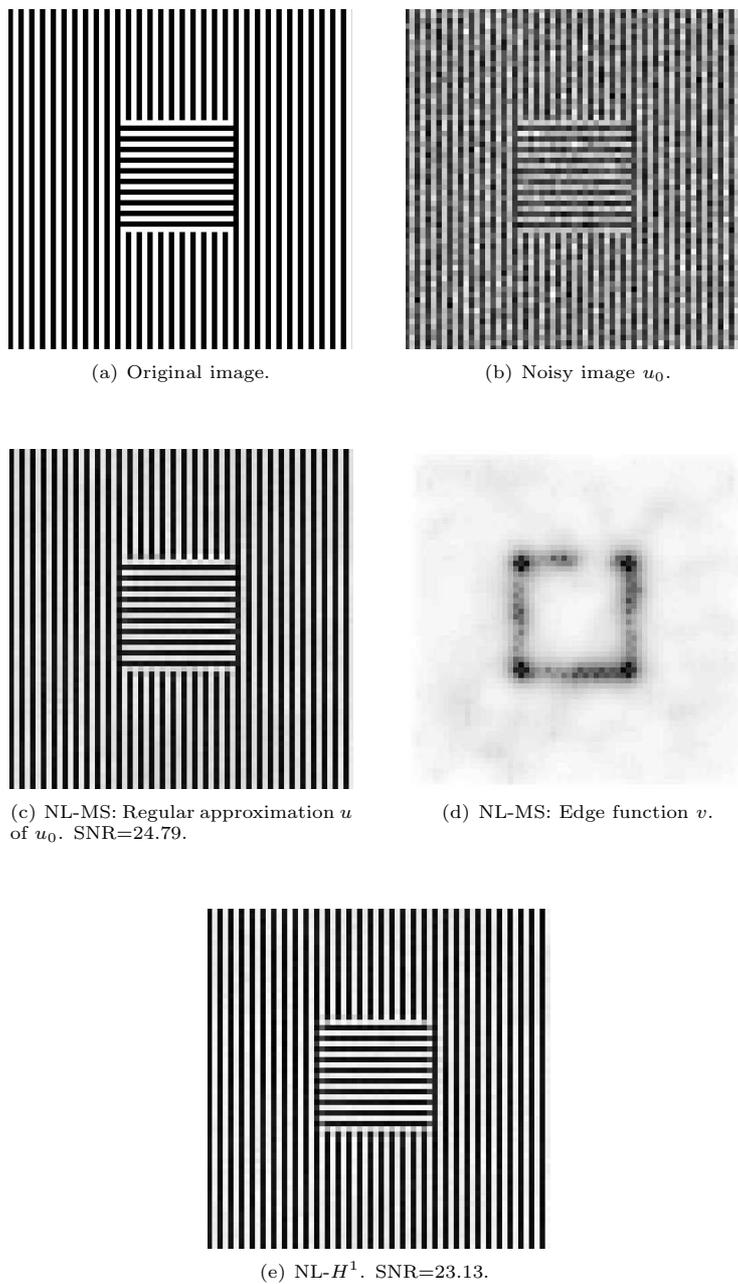


FIGURE 10. Non-local Mumford-Shah model (27). The NL-MS model can segment textures unlike the original model [39]. Boundary between textures are represented by function  $v$  with low values (dark part in Figure (d)). Besides, the NL-MS model can denoise images (Figure (c)) better than the standard model NL- $H^1$  [20] (Figure (e)). The image size is  $64 \times 64$ . The iterative minimization process for NL-MS takes 1500 iterations and the computational time is 5.8 sec.



(a) Original Image.

(b) Noisy image  $u_0$ .(c) NL-MS: Regular approximation  $u$  of  $u_0$ .  
SNR=17.67.(d) NL-MS: Edge function  $v$ .(e)  $NL-H^1$ . SNR=17.29.

FIGURE 11. Non-local Mumford-Shah model (27). Boundary between textures are represented by function  $v$  with low values (dark part in Figure (d)). The NL-MS better denoise images (Figure (c)) than the standard model  $NL-H^1$  [20] (Figure (e)). The image size is  $266 \times 247$ . The iterative minimization process for NL-MS takes 2650 iterations and the computational time is 245 sec.