Active Contours without Level Sets

Romain Yildizoglu, Jean-François Aujol, Nicolas Papadakis

ACTIVE CONTOURS WITHOUT LEVEL SETS

Romain Yildizoglu, Jean-François Aujol

Univ. Bordeaux, IMB, UMR 5251
F-33400 Talence, France.

Nicolas Papadakis

CNRS, LJK, UMR 5224, MOISE (INRIA/LJK)
38041 Grenoble, France

ABSTRACT

This paper deals with the problem of segmenting an image with active contours. We explain how recent convexification methods allow now to use active contours without level sets with simple and efficient first order schemes. We recall different algorithms proposed in the literature, and we propose a new variant. Numerical experiments in 2D and 3D confirm the interest of the approach.

Index Terms— Segmentation, active contours, convexification, total variation

1. INTRODUCTION

Segmentation remains one of the main problem in image processing . Numerous methods have been proposed to tackle this issue during the last 25 years. A very popular class of approaches are the one relying on active contour description with snakes [1, 2], level set representations [3, 4] or graph modeling [5, 6]. Our topic here is to sum up the different variational approaches that have raised so far with respect to active contour models. More precisely, we fill focus on the simple binary segmentation problem, known as the two-phase piecewise constant Mumford-Shah problem, defined as follows. Let us denote by $I : x \in \Omega \mapsto [0; 1]$ a gray-scale image of interest defined on the domain $\Omega \subset \mathbb{R}^2$. The binary segmentation problem consists in estimating a binary mask $u : x \in \Omega \mapsto \{0; 1\}$, that separates I into two different areas corresponding to 2 mean colors c_1 and c_2. The problem can be formalized as a minimization problem:

$$(u^*, c_1^*, c_2^*) = \underset{u\in\{0;1\}^{|\Omega|}, c_1\in[0;1], c_2\in[0;1]}{\operatorname{argmin}} J(u, c_1, c_2),$$

where the energy J is defined as:

$$\begin{aligned} J(u, c_1, c_2) = &\int_\Omega |Du| + \lambda \int_\Omega |I(x) - c_1|^2 u(x)dx \\ &+ \lambda \int_\Omega |I(x) - c_2|^2 (1 - u(x))dx, \end{aligned} \tag{1}$$

and $\lambda \geq 0$ weights the influence of the data term with respect to the penalization of the perimeter of the segmented area $S_u = \{x, u(x) = 1\}$. As u is binary, the contour length $|\partial S_u|$ is given by the total variation of u, $\int_\Omega |Du|$ [7]. Notice that in the case when u is smooth, then $\int_\Omega |Du| = \int_\Omega |\nabla u|\, dx$. The two last energy terms model the image I with two homogeneous regions characterized by c_1 in the segmented area S_u and c_2 in its complementary region $\Omega \backslash S_u = \{x, u(x) = 0\}$. In the case when c_1 and c_2 are fixed, it is a straightforward application of the direct method of calculus of variations to show that there exists a minimizer for problem (1) [8].

Considering the variables separately, the energy is convex in u and in (c_1, c_2). It is nevertheless not convex in (u, c_1, c_2). As a consequence, an alternative minimization scheme has to be considered. When u is fixed, an explicit expression of the optimal values of c_1 and c_2 can be computed by derivating (1). On the other hand, minimizing such energy with respect to u and within a variational framework is limited by two main technical issues. First of all, the binary unknown u does not live in a continuous space. Then, the analytic derivation of the regularization term gives the curvature $div\left(\frac{\nabla u}{|\nabla u|}\right)$, that is not defined everywhere and involves numerical approximations at points x such that $|\nabla u(x)| = 0$.

In the following, we will chronologically review some main approaches allowing to compute u within a variational framework, and propose a new algorithm (11). It is interesting to see how the two previously introduced issues have been successively solved, and how the resulting algorithms are each time not only more accurate but also faster, more stable and simpler to parametrize.

2. CHAN-VESE LEVEL SET FORMULATION

In their seminal paper [4], Chan and Vese proposed to solve this problem using a level set formulation. Introducing the surface ϕ whose 0 level represents the contour of interest, we can define the binary function $u(x) = H(\phi(x))$, where $H(.)$ is the Heaviside function that is 1 if its argument is positive and 0 otherwise. Reformulating problem (1) in terms of ϕ, that now lives in a continuous space, we get:

$$\begin{aligned} J(\phi) = &\int_\Omega |\nabla H(\phi(x))|dx + \lambda \int_\Omega |I(x) - c_1|^2 H(\phi(x))dx \\ &+ \lambda \int_\Omega |I(x) - c_2|^2 (1 - H(\phi(x)))dx. \end{aligned} \tag{2}$$

The contour length penalization term can be then rewritten as: $\int_\Omega |\nabla H(\phi(x))|dx = \int_\Omega \delta(\phi)|\nabla\phi(x)|dx$, where $\delta(y)$ (the derivative of the Heaviside function) is the Dirac distribution.

For numerical purposes due to the computation of the Dirac distribution, a regularization of the Heaviside function is needed. Given $\epsilon > 0$, the final modeling considers C^2 approximations $H_\epsilon(y)$ and $\delta_\epsilon(y)$ to extend the support of these functions and thus be able to perform discrete computations. In order to deal with the non differentiability of the Total Variation, a regularization of the term may also be relevant by defining $|\nabla\phi|_\eta = \sqrt{\phi_x^2 + \phi_y^2 + \eta^2}$, for $0 < \eta << 1$.

With all these approximations, one can compute the Euler-Lagrange derivatives to design an iterative minimizing algorithm. Introducing an artificial time step τ and given an initialization $\phi^0(x)$, the gradient descent approach leads to, for $k \geq 0$: $\phi^{k+1} = \phi^k - \tau\partial_{\phi^k} J(\phi^k)$. We then have

$$\phi^{k+1} = \phi^k + \tau\delta_\epsilon(\phi^k)\left[div\left(\frac{\nabla\phi^k}{|\nabla\phi^k|_\eta}\right) - \lambda(I - c_1)^2 + \lambda(I - c_2)^2\right]. \tag{3}$$

Notice that the level set function has to be updated into a signed distance function to get an accurate representation of the contour length. As the level set evolves in the neighborhood of its 0 level, such method estimates a local minimum of the energy function that depends on the initialization ϕ^0.

3. NIKOLOVA-ESEDOGLU-CHAN CONVEX FORMULATION

More recently, it was proposed in [8] to relax the binary problem and let $u(x)$ takes its values in the interval $[0; 1]$. The relaxed energy (1) is then convex in u, defined over the convex function set $\mathcal{A} := \{u \in BV(\Omega, [0; 1]^{|\Omega|})\}$. Hence a global optimal solution u^* may be computed using again a gradient descent scheme:

$$u^{k+1} = P_{\mathcal{A}}\left(u^k + \tau\left(div\left(\frac{\nabla u}{|\nabla u|_\eta}\right) - \lambda(I - c_1)^2 + \lambda(I - c_2)^2\right)\right), \tag{4}$$

the solution being projected onto the the convex set $\mathcal{A}$ after each iteration through $P_{\mathcal{A}}(u) = \min(\max(u, 0), 1)$[1].

To recover a binary map u from the global optimal solution u^* of the relaxed energy function, one can use a useful theorem (see [8, 9]), based on the co-area formula. It shows that for almost any threshold $\mu \in (0; 1)$, the characteristic function $u_\mu = H(u^* - \mu)$ is also a global minimum of the original binary energy function (1). Remark that the solution u^* is in practice binary almost everywhere, but there is no proof of such a conjecture. This remarkable result has led to numerous applications in computer vision (see e.g. [10]).

Notice that the problem has been reformulated in [11] in order to be convex with respect to u, c_1 and c_2. Using the strategy of [9], the problem complexity is increased by adding two new dimensions corresponding to the discrete possible values of c_1 and c_2. We will not detail this approach, as it involves a huge computational cost for accurate estimations of c_1 and c_2.

[1]The authors also propose a non constraint formulation of the problem that directly deals with values of u in $[0; 1]$.

4. DUAL FORMULATION OF THE CONVEX PROBLEM

The previous formulation still involves the regularization of the Total Variation $|\nabla u|_\eta$, defined for functions u taking their values in the continuous interval $[0; 1]$. Assuming periodic or homogeneous Dirichlet boundary conditions over $\partial\Omega$, an idea is then to consider the dual formulation of the Total Variation given by [12]:

$$\int_\Omega |Du| = \sup_{\mathbf{z}\in\mathcal{B}} \int_\Omega u\, div(\mathbf{z})dx \tag{5}$$

where the dual variable $\mathbf{z}$ is a vector defined in the unit circle with $\mathcal{B} = \{\mathbf{z} = (z_1, z_2), |\mathbf{z}| = \sqrt{z_1^2 + z_2^2} \leq 1\}$. If u is regular, then we have $\int_\Omega |Du| = \sup_{\mathbf{z}\in\mathcal{B}} \int_\Omega \nabla u \cdot \mathbf{z}dx$. This dual formulation allows to represent with $\mathbf{z}$ the unit vector direction of ∇u and will be the key point for dealing with the non differentiability that happens when $|\nabla u| = 0$. Fixing c_1 and c_2, we get the following minimization problem:

$$(u^*, \mathbf{z}^*) = \underset{u\in\mathcal{A}}{\operatorname{argmin}} \ \underset{\mathbf{z}\in\mathcal{B}}{\operatorname{argmax}} J(u, z),$$

where the energy J is defined as:

$$\begin{aligned} J(u, \mathbf{z}) = &\int_\Omega u\, div(\mathbf{z})dx + \lambda\int_\Omega |I(x) - c_1|^2 u(x)dx \\ &+ \lambda\int_\Omega |I(x) - c_2|^2(1 - u(x))dx. \end{aligned} \tag{6}$$

This energy is now continuously differentiable in $(u, \mathbf{z})$.

4.1. Minimization and convergence

Following the works of Pock and al. [10, 13], the energy (6) can be minimized with the parallelizable Arrow-Hurwicz algorithm [14], also known as the first-order Primal-Dual proximal point method [12, 15, 16], that consists of alternate maximizations over $\mathbf{z} \in B$ and minimizations over $u \in \mathcal{A}$. Introducing two time steps τ_u and $\tau_{\mathbf{z}}$ and starting from an initialization (u^0, z^0), the process reads, for $k \geq 0$:

$$\begin{cases} \mathbf{z}^{k+1} = P_{\mathcal{B}}(\mathbf{z}^k + \tau_{\mathbf{z}}\nabla u^k) \\ u^{k+1} = P_{\mathcal{A}}(u^k + \tau_u(div(\mathbf{z}^{k+1}) - \lambda(I - c_1)^2 + \lambda(I - c_2)^2)), \end{cases} \tag{7}$$

where the projections over the convex set $\mathcal{B}$ is given by:

$$P_{\mathcal{B}}(\mathbf{z}) = \begin{cases} \mathbf{z} & \text{if } |\mathbf{z}| \leq 1 \\ \frac{\mathbf{z}}{|\mathbf{z}|} & \text{otherwise.} \end{cases} \tag{8}$$

It was recently demonstrated in [13] that the previous Primal-Dual algorithm converges to the exact optimal solution u^* at the rate $O(1/\sqrt{N})$, for iteration N. It was shown that the Primal-Dual (7) corresponds to the specific case $\theta = 0$ of the more general algorithm defined for any $\theta \in [0; 1]$ that converges at the rate $O(1/N)$ for $\theta > 0$:

$$\begin{cases} \mathbf{z}^{k+1} = P_{\mathcal{B}}(\mathbf{z}^k + \tau_{\mathbf{z}}\nabla u^k) \\ \tilde{u}^{k+1} = P_{\mathcal{A}}(\tilde{u}^k + \tau_u(div(\mathbf{z}^{k+1}) - \lambda(I - c_1)^2 + \lambda(I - c_2)^2)) \\ u^{k+1} = \tilde{u}^{k+1} + \theta(\tilde{u}^{k+1} - \tilde{u}^k), \end{cases} \tag{9}$$

with the initialization $\tilde{u}^0 = u^0$ and some conditions over τ_u and $\tau_{\mathbf{z}}$ that are detailed below.

4.2. Numerical implementation

In order to implement all the above mentioned algorithms, we need to describe the discretization of the involved gradient and divergence operators. To have an exact representation of the relation (5), the discretization of the following scalar product must be checked: $\langle u, div(\mathbf{z})\rangle = -\langle \nabla u, \mathbf{z}\rangle$, $\forall u \in \mathcal{A}, \mathbf{z} \in \mathcal{B}$. We now consider the discrete regular grid (i, j), $1 \leq i \leq L$, $1 \leq j \leq M$ representing the domain Ω. Looking at the discrete gradient operator as a vector of matrices $\nabla = [D_x; D_y]^T$, the chosen discretizations should check: $\langle D_x u, z_1\rangle + \langle D_y u, z_2\rangle = \langle u, D_x^T z_1 + D_y^T z_2)\rangle$. To that end, one can consider finite differences and take $\nabla u(i, j) = [u_x^+(i, j), u_y^+(i, j)]^T$ where the gradient with respect to the first dimension reads

$$u_x^+(i,j) = \begin{cases} u(i+1,j) - u(i,j) & \text{if } 1 \leq i < M, \\ 0 & \text{if } i = M. \end{cases}$$

The corresponding divergence operator is given by $div(\mathbf{z}) = (z_1)_x^- + (z_2)_y^-$, the gradient over the first dimension taken as:

$$z_x^-(i,j) = \begin{cases} z(i,j) & \text{if } i = 1, \\ z(i,j) - z(i-1,j) & \text{if } 1 < i < M, \\ -z(i-1,j) & \text{if } i = M. \end{cases}$$

4.3. Narrow band: local minima vs dimension reduction

When dealing with large images or even 3D images, as can be provided by MRI, local convergence can be an advantage to achieve segmentation in real time. With an adequate initialization, the local Chan and Vese method can segment a specific object even if there exists other regions in the image with similar color properties. On the other hand, the convex formulations permit to compute a global optimal solution but can not segment one single object. In this convex case, as the obtained solution should be binary almost everywhere, the algorithm will mainly acts at the boundaries of the segmented regions after a few iterations,. For these reasons, Baeza et al. proposed in [17] to define a narrow band method in order to speed up the process in case of high dimensional problems. However, such approach does not permit to compute a global minimum anymore. The corresponding intuitive algorithm is as follows. Lets $\mathbf{u}_0$ be an initial binary segmentation, a narrow band $B_0 \subset \Omega$ is defined in the spatial domain, around the boundaries $\partial S_{\mathbf{u}_0}$ of the corresponding region $S_{\mathbf{u}_0}$. The size of the narrow band is parametrized with $\beta > 0$ as:

$$B_0 = \{x \in \Omega, \min_{y \in S_{\mathbf{u}_0}} |x - y| < \beta\}.$$

The recursive algorithm then consists in computing a first solution u^* evolving only in B_0 and with fixed values in $\Omega \backslash B_0$. At convergence, the solution is binarized to obtained $\mathbf{u}_1$ that is used to define a new narrow band B_1 and so on... An additional term to the energy is necessary to make the process converge to a binary solution and prevent from narrow band cycles. At narrow iteration l performing into the band B_l defined at the neighborhood of the binary previous solution $\mathbf{u}_l$, and introducing $\gamma > 0$, the energy to minimize for $l \geq 0$ is:

$$\begin{aligned} J(u, \mathbf{z}) = &\int_{B_l} u\, div(\mathbf{z}) dx + \lambda \int_{B_l} |I(x) - c_2|^2 (1 - u(x)) dx \\ &+ \lambda \int_{B_l} |I(x) - c_1|^2 u(x) dx + \frac{\gamma}{2} \int_{B_l} |u - \mathbf{u}_l|^2 dx, \end{aligned} \tag{10}$$

with $u(x) = \mathbf{u}_l(x)$, for $x \in \Omega \backslash B_l$. The three last terms of this energy that correspond to the data now forms a uniformly convex operator with respect to u, with convexity parameter γ. As a consequence, the accelerated version of the Chambolle-Pock algorithm [13] can be considered to reach the rate of convergence $O(1/N^2)$ for the minimization inside the narrow band. Considering $\tau_u^0 \tau_{\mathbf{z}}^0 < 1/8$ and initializing $u^0 = \tilde{u}^0 = \mathbf{u}_l$, the new algorithm to find the optimal solution inside the band B_l reads:

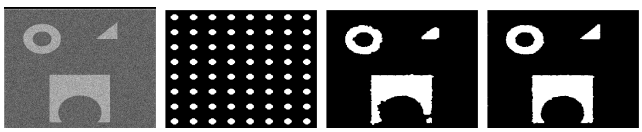
$$\begin{cases} \mathbf{z}^{k+1} = P_{\mathcal{B}}(\mathbf{z}^k + \tau_{\mathbf{z}}^k \nabla u^k) \\ \tilde{u}^{k+1} = P_{\mathcal{A}}\left(\frac{\tilde{u}^k + \tau_u^k (div(\mathbf{z}^{k+1}) - \lambda(I-c_1)^2 + \lambda(I-c_2)^2 + \gamma \mathbf{u}_l)}{1 + \tau_u^k \gamma}\right) \\ \theta^k = 1/\sqrt{1 + 2\gamma\tau_u^k}, \tau_u^{k+1} = \theta^k \tau_u^k, \tau_{\mathbf{z}}^{k+1} = \tau_{\mathbf{z}}^k / \theta^k \\ u^{k+1} = \tilde{u}^{k+1} + \theta^k (\tilde{u}^{k+1} - \tilde{u}^k). \end{cases} \tag{11}$$

At convergence, the obtained solution u^* is binarized to define the next estimate $\mathbf{u}_{l+1}$. To ensure the convergence of such an algorithm, the authors of [17] also showed that the binarization should be done carefully. More precisely, one should take $\nu << 1$ and choose the binarization threshold μ from among the two possible values $\mu = \nu$ and $\mu = 1 - \nu$, the one corresponding to the smaller energy (10). Notice that in practice, taking a fixed threshold does not perturb the convergence process.

5. EXPERIMENTS AND COMPARISONS

We compare the following algorithms: CV stands for the Chan-Vese approach (equation (3)); NEC stands for the Nikolova-Esedoglu-Chan approach (equation (4)); AH stands for the Arrow Hurwicz primal dual algorithm (7); CP stands for the Chambolle-Pock algorithm (9).

Fig. 1 presents a first segmentation result on a noisy synthetic image ($c_1 = 50$, $c_2 = 100$ and $\sigma = 35$). The non convex CV algorithm is limited by the initialization and cannot segment entirely the objects, the level set evolution being stopped by local noisy fronts (the initialization choice can also have a dramatic influence on the computation time). Table 1 (left) displays comparison results between CV, CEN, AH and CP algorithms for noisy data obtained with $\sigma = 15$. It gives the computation time with respect to the size of a synthetic image. It shows that the primal dual approaches (AH and CP) are clearly much faster than the other algorithms. Table 1 (right) shows the evolution of the number of missegmented pixels with respect to the SNR of the image. The primal dual approaches (AH and CP) seem to be more robust.



Noisy image Initialization CV CP

Fig. 1. Examples of the 2D segmentations obtained with the CV and the convex approach (CP) on a synthetic image.

Of course, c_1 and c_2 can be updated within the iterations as in [4] (but problem (6) is not convex with respect to u, $\mathbf{z}$, c_1 and c_2). Nevertheless, this has not been a problem from

size /method	CV	NEC	AH	CP
128×128	0.5	0.12	0.07	0.03
256×256	3.8	0.24	0.1	0.09
512×512	12.3	1.53	0.82	0.76
1024×1024	63	8.42	3.6	3.23
2048×2048	278	35.72	14.11	13.64

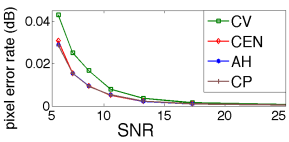
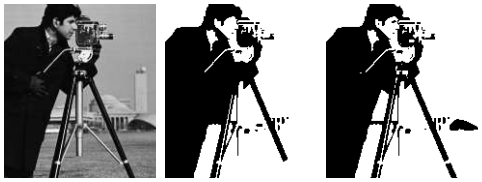


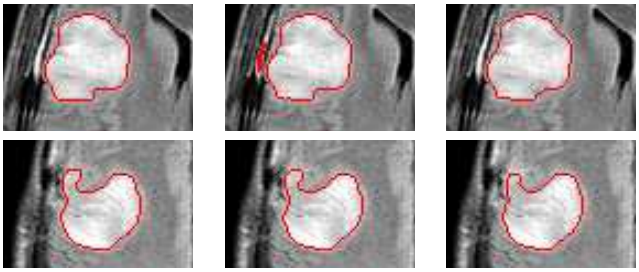
Table 1. Left: Computation time (in seconds) with respect to the image size on a synthetic image. The primal dual approaches (AH and CP) are much faster than the other ones. Right: Rate of mis-segmented pixels with respect to the SNR of the image. The primal dual approaches (AH and CP) are more robust.

a numerical point of view. Figure 2 shows such an example. Fig. 3 presents a segmentation result on a real 3D image (only 2 layers are shown here). The narrow band method is useful to deal with the high data dimension. Moreover, when using a too large regularization term with CP algorithm, one gets an over-smoothed segmentation result, whereas artifacts can be kept in the segmentation results with a too small regularization term. Thanks to the use of the narrow band method, we can avoid both problems.



Cameraman CV CP

Fig. 2. Examples of the 2D segmentations obtained with the CV and the convex approach (CP) on the cameraman image. Notice details such as the eye which can be recovered with CP.



CP + narrow band ($\lambda = 0.1, \gamma = 0.1$) CP ($\lambda = 0.1$) CP ($\lambda = 0.12$)

Fig. 3. Comparison of the 3D segmentations obtained with the CP algorithm with or without the narrow band method. The segmentation result is over-smoothed on the last image of the second row and an artifact is kept in the segmentation result in the second image of the first row. With the narrow band method both drawbacks are avoided.

6. CONCLUSIONS

In this paper, we have presented robust and fast algorithms based on convex functionals that deal with the binary segmentation problem. As multiphase level sets, such approaches can be easily extended to consider advanced image modeling. From the experimental tests, the convex formulation is a very reliable alternative to level sets methods.

Acknowledgments:

The authors would like to thank Professor Vicent Caselles for his useful comments about a preliminary version of this work. The first and second authors acknowledge the support of the French "Agence Nationale de la Recherche" (ANR), under grant NatImages (ANR-08-EMER-009), "Adaptivity for natural images and texture representations". The authors would like to thank Dr B. Henriques de Figueiredo , Pr H. Loiseau and Pr G. Kantor at Institut Bergonié for providing them with the scan data.

7. REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, vol. 1, pp. 321–331, 1988.

[2] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *IJCV*, vol. 22, pp. 61–79, 1997.

[3] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *JCP*, vol. 79, no. 1, pp. 12 – 49, 1988.

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

[5] Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *ICCV'01*, 2001, vol. 1, pp. 105–112.

[6] E. Bae, J. Yuan, and X.-C. Tai, "Global minimization for continuous multiphase partitioning problems using a dual approach," *IJCV*, vol. 92, pp. 112–129, 2011.

[7] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, vol. 147, Springer-Verlag, 2002.

[8] M. Nikolova, S. Esedoglu, and T. F. Chan, "Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models," *SIAM JAM*, 66(5):1632–1648, 2006.

[9] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, "A convex formulation of continuous multi-label problems," in *ECCV*, 2008, vol. 5304, pp. 792–805.

[10] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "Global solutions of variational models with convex regularization," *SIAM JIS*, vol. 3, pp. 1122–1145, 2010.

[11] E. Brown, T. F. Chan, and X. Bresson, "Completely Convex Formulation of the Chan-Vese Image Segmentation Model," *IJCV*, pp. 1–19, 2011.

[12] T. F. Chan, G. H. Golub, and Pep Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM JSC*, vol. 20, pp. 1964–1977, 1999.

[13] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *JMIV*, vol. 40, pp. 120–145, 2011.

[14] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*, vol. 2, Stanford Un. Press, 1964.

[15] R. T. Rockafellar, "Augmented lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 97–116, 1976.

[16] M. Zhu and T Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restauration," Tech. Rep., UCLA CAM Report 08-34, 2008.

[17] A. Baeza, V. Caselles, P. Gargallo, and N. Papadakis, "A narrow band method for the convex formulation of discrete multilabel problems," *SIAM JMMS*, 8(5):2048–2078, 2010.