Extending Continuous Cuts: Anisotropic Metrics and Expansion Moves

Carl Olsson, Martin Byröd, Niels Chr. Overgaard, Fredrik Kahl

Centre for Mathematical Sciences, Lund University, Sweden

P.O. Box 118, SE-221 00 Lund, Sweden

{calle,byrod,nco,fredrik}@maths.lth.se

Abstract

The concept of graph cuts is by now a standard method for all sorts of low level vision problems. Its popularity is largely due to the fact that globally or near globally optimal solutions can be computed using efficient max flow algorithms. On the other hand it has been observed that this method may suffer from metrication errors. Recent work has begun studying continuous versions of graph cuts, which give smaller metrication errors. Another advantage is that continuous cuts are straightforward to parallelize.

In this paper we extend the class of functionals that can be optimized in the continuous setting to include anisotropic TV-norms. We show that there is a so called coarea formula for these functionals making it possible to minimize them by solving a convex problem. We also show that the concept of α-expansion moves can be reformulated to fit the continuous formulation, and we derive approximation bounds in analogy with the discrete case. A continuous version of the Potts model for multi-class segmentation problems is presented, and it is shown how to obtain provably good solutions using continuous α-expansions. [1]

1. Introduction

Recent work has shown that continuous cuts can be used to successfully solve a number of application problems in computer vision [12, 5, 19, 14]. Applications include image segmentation, restoration, stereo and multi-label problems. While discrete (graph) cuts can be considered to be a mature technique, the full potential of continuous cuts is yet to be explored. In this paper we extend the underlying theory of the continuous framework in two directions. Our main contributions are:

- **A generalized coarea formula.** We show that the coarea formula holds for a wider class of functionals than for the usual Total Variational (TV) norm. More specifically, it is shown that there is a corresponding coarea formula for anisotropic metrics. This allows us to model and optimize energy functionals where cuts in certain directions are preferred. Equivalently we can think of this as a modified metric in which the shortest paths are not necessarily straight lines. Figure 1 illustraits differences between various metrics.

- **Continuous expansion moves.** In analogy with discrete cuts, we show that the concept of α-expansion can be recast in the continuous setting. We give a continuous counterpart to the discrete Potts model and derive similar optimality bounds as in the discrete case.

One motivation for studying continuous cuts is that the underlying problem is better suited for a continuous formulation. However, by the end of day, in order to perform computations on a finite precision computer, it is required to discretize the problem. In the continuous case, it turns out that the corresponding optimization problem can be recast as a convex variational problem and standard PDE-methods can be applied to efficiently compute the solution, e.g., [7]. In [10], it was shown experimentally that such continuous methods generally have smaller discretization (or metrication) errors than standard graph cuts. On the other hand, it is possible to approximate continuous cuts with discrete graph cuts to arbitrary precision using the results in [3]. The drawback is that this produces significantly larger graphs.

An alternative approach to continuous cuts is to consider the dual problem of continuous max flows [1, 18]. This could also be a fruitful and promising strategy, but not one we will follow in this paper. A possible disadvantage is that max flow algorithms (both discrete and continuous) are not as straightforward to parallelize as the primal problem.

After this work was submitted the papers [20, 13] have appeared. In [20] the authors presents an algorithm for solving the continuous maximal flows with anisotropic flow constraints (similar to [1]). This is related to our energy formulation through duality. In [13] a continuous counterpart to the Potts model for multi class segmentation is presented. The proposed model is very similar to ours, however, the coarea formula is not proven. Instead the authors demonstrate (experimentally) that the obtained solutions are often close to binary. An a posteriori error bound based on the
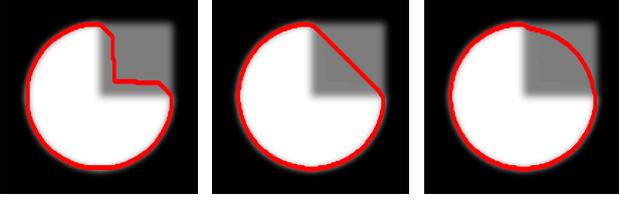
Figure 1. An illustration of geodesics obtained using various metrics. Left: Discrete graph cut. Middle: Isotropic continuous cuts. Right: Anisotropic continuous cuts (matrix field with a large component in the radial direction). See Section 7 for further details.

obtained solution to the relaxed convex program is given. In contrast we establish the coarea formula which enables us to give the same a priori bound as in the discrete case.

2. Continuous Cuts

Consider a two class segmentation problem. Let Ω be a region of $\mathbf{R}^n$ representing the image domain. If $g_f : \Omega \mapsto \mathbf{R}$ and $g_b : \Omega \mapsto \mathbf{R}$ are two functions representing the costs of assigning pixels to foreground and background, respectively, then the segmentation problem can be posed as

$$\min_S \int_S g_f \, dx + \int_{\Omega \setminus S} g_b \, dx. \qquad (1)$$

If the functions g_f and g_b are not smooth this formulation may result in noisy segmentations, as this formulation do not take into account the classifications of neighboring pixels. Therefore a common approach (e.g. [12]) is to add a smoothing term which penalizes classifications where neighboring pixels does not belong to the same class. In the continuous setting the smoothing term is usually the curve length $\int dS$, that is, we penalize segmentations where the set S has a large boundary. Let u be an indicator function of the set S,

$$u(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

then in a distributional sense

$$\int_{\partial S} d\sigma = \int_\Omega |\nabla u| \, dx, \qquad (3)$$

where $|\cdot|$ is the euclidean distance $|(\xi_1, \xi_2)| = \sqrt{\xi_1^2 + \xi_2^2}$. If we let $g = g_f - g_b$ then we may write the segmentation problem (ignoring constants) as

$$\min_{u(x) \in \{0,1\}} E(u) \qquad (4)$$

where $E(u) = E^s(u) + \lambda E^d(u)$ and

$$E^s(u) = \int_\Omega |\nabla u| \, dx, \quad E^d(u) = \int_\Omega gu \, dx. \qquad (5)$$

The functional itself is convex. However we are not minimizing over a convex set since u is restricted to the set of indicator functions.

2.1. The Thresholding Theorem

In [18] it was observed that if u_0 is any minimizer of E on the convex set of functions u with values in the interval $[0, 1]$, then an optimal solution of (4) can be obtained by thresholding u_0 at any level s in $[0, 1]$. The reason is that E fulfills the following identity, for any admissible u,

$$\int_0^1 (E(u^s) - E(u)) \, ds = 0, \qquad (6)$$

where u^s is the threshold at s,

$$u^s(x) = \begin{cases} 1 & \text{if } u(x) > s \\ 0 & \text{otherwise.} \end{cases} \qquad (7)$$

Since any u_0^s is feasible in (4) we have that $E(u_0^s) - E(u_0) \leq 0$ for all s, so in view of (6) equality holds except on a set of dt-measure zero.

To show (6) we first consider the data term E^d of (5). In this case we may use the layer cake representation for u,

$$u = \int_0^\infty u^s \, ds. \qquad (8)$$

Inserting this into data term E^d, switching the order of integration, we see that (6) holds for E^d, because $u^s = 0$ for $s > 1$.

Ultimately showing (6) for the smoothness term E^s comes down to showing the coarea formula [9] for the TV-norm. Let

$$J(u) = \int_\Omega |\nabla u| dx. \qquad (9)$$

Then the coarea formula says that

$$J(u) = \int_{-\infty}^\infty J(u^s) \, ds. \qquad (10)$$

Again we use that $u^s = 0$ for $s > 1$ because $0 \leq u(x) \leq 1$. Therefore (10) is equivalent to (6).

2.2. The Relation to Discrete Graph Cuts

The formulation (5) reveals the connection to the discrete graph cuts model. If the euclidean distance is replaced by $|(\xi_1, \xi_2)| = |\xi_1| + |\xi_2|$ and we choose the gradient discretization as

$$\bar{\nabla} u(i,j) = \begin{pmatrix} u(i,j) - u(i-1,j) \\ u(i,j) - u(i,j-1) \end{pmatrix} \qquad (11)$$

then $\sum_{i,j} |\bar{\nabla} u(i,j)|$ is the cost in the discrete version on a four-connected grid.

3. Anisotropic Smoothness Terms

In this section we will show that the coarea formula for the total variation functional J in (9) can be extended to anisotropic smoothness terms as well. The proof is inspired by the proof of the isotropic version in [9].

An Anisotropic Coarea Formula. Let $A : \Omega \to \mathbf{R}^{N \times N}$ denote a matrix valued function. Assume that $A(x) = [a_{ij}(x)]_{1 \le i,j \le N}$ is strongly positive definite for every $x \in \Omega$, in the sense that there exist positive numbers κ, K with $0 < \kappa < K$ such that

$$\kappa |\xi|^2 \le \xi^T A(x) \xi \le K |\xi|^2 \quad (\xi \in \mathbf{R}^N, x \in \Omega). \quad (12)$$

Then we introduce the following *anisotropic total variation* functional,

$$J_A(u) = \int_\Omega \sqrt{\nabla u(x)^T A(x) \nabla u(x)} \, dx, \quad (13)$$

for any $u \in C_0^1(\Omega)$ (differentiable functions with compact support), preliminarily. This functional satisfies a coarea formula for functions of bounded variation (BV) similar to the classical one:

Theorem 3.1 (Anisotropic Coarea) *For any $u \in BV(\Omega)$ we have*

$$J_A(u) = \int_{-\infty}^{\infty} J_A(u^s) \, ds. \quad (14)$$

Since functions in $BV(\Omega)$ are not differentiable in the ordinary sense, we first have to extend the definition of J_A to cover the case when u is only weakly differentiable. Let $Q = Q(x)$ denote the the symmetric matrix $Q(x)$ such that $A(x) = Q(x)Q(x)$. For $u \in C_0^1(\Omega)$ we have,

$$J_A(u) = \int_\Omega |Q \nabla u| dx = \sup_{|\xi(x)| \le 1} \int_\Omega (Q \nabla u)^T \xi dx \quad (15)$$

$$= \sup_{|\xi(x)| \le 1} \int_\Omega -u \operatorname{div}(Q\xi) \, dx \quad (16)$$

$$= \sup_{\eta^T A^{-1} \eta \le 1} \int_\Omega -u \operatorname{div} \eta \, dx, \quad (17)$$

where ξ and η denote C^1-vector fields with compact support in Ω, and $\eta = Q\xi$. Let $\|\xi\|_A = \sqrt{\xi^T A \xi}$ and $\|\eta\|_A^* = \sqrt{\eta^T A^{-1} \eta}$. Then we may define the anisotropic total variation J_A for functions $u \in \mathrm{BV}(\Omega)$ by the formula

$$J_A(u) = \sup_{\xi \in C_0^1(\Omega; \mathbf{R}^N), \|\xi(x)\|_A^* \le 1} \int_\Omega -u \operatorname{div} \xi \, dx. \quad (18)$$

Notice that the requirement (12) on A implies that $\sqrt{\kappa} J(u) \le J_A(u) \le \sqrt{K} J(u)$, where $J(u) = \int_\Omega |\nabla u| \, dx$ is the usual, isotropic, total variation.

Differentiation. Before we prove the coarea formula we state a theorem from integration theory that will be needed. Let f be a non-decreasing real-valued function defined on an interval $I = [a, b]$, then a theorem of Lebesgue (see [16]

Thm. 5.3, p. 100) states that f is differentiable almost everywhere on I, that is, the limit

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} := f'(x) \quad (19)$$

exists for almost all $x \in I$. Moreover, if we set $f'(x) = 0$ for x where the above limits fail to exist, then f' is measurable and integrable with

$$\int_a^b f'(x) \, dx \le f(b) - f(a). \quad (20)$$

Proof of the Coarea formula. We will restrict ourselves to the case when $u \in C^1(\Omega) \cap BV(\Omega)$. The general case can be handled with standard approximation arguments.

STEP I. Assume first that $u \ge 0$, and insert the layer cake representation (8) in the formula (18) to get the inequality,

$$J_A(u) = \sup_{\|\xi\|_A^* \le 1} \int_\Omega - \left(\int_0^\infty u^s \, ds \right) \operatorname{div} \xi \, dx \quad (21)$$

$$\le \int_0^\infty \left(\sup_{\|\xi\|_A^* \le 1} \int_\Omega -u^s \operatorname{div} \xi \, dx \right) ds \quad (22)$$

$$= \int_0^\infty J_A(u^s) \, ds. \quad (23)$$

For $u \le 0$ we use that $J_A(-w) = J_A(w)$ and $J_A(\mathrm{const.} + w) = J_A(w)$, for any w, to derive the following inequality:

$$J_A(u) = J_A(-u) \le \int_0^\infty J_A((-u)^r) \, dr \quad (24)$$

$$= \int_0^\infty J_A(1 - u^{-r}) \, dr \quad (25)$$

$$= \int_0^\infty J_A(u^{-r}) \, dr = \int_{-\infty}^0 J_A(u^s) \, ds. \quad (26)$$

Finally, a general u may be written as the difference $u = u^+ - u^-$ between two positive functions. The trivial inequality $J_A(u) \le J_A(u^+) + J_A(u^-)$ then implies that the following inequality holds:

$$J_A(u) \le \int_{-\infty}^\infty J_A(u^s) \, ds. \quad (27)$$

STEP II. We are now going to prove the reverse inequality. Define the function,

$$m(t) = \int_{\{x \in \Omega; u(x) \le t\}} \|\nabla u\|_A \, dx. \quad (28)$$

Then m is nondecreasing on $\mathbf{R}$, hence it follows from Lebesgue's theorem that $m'(t)$ exists a.e. (wrt. Lebesgue measure), and that Lebesgue's inequality holds:

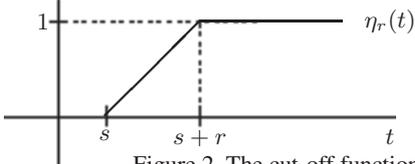$$\int_{-\infty}^\infty m'(t) \, dt \le m(\infty) - m(-\infty) = J_A(u). \quad (29)$$

Figure 2. The cut-off function η_r.

Now, fix $s \in \mathbf{R}$ and let a "cut-off" function $\eta_r : \mathbf{R} \to [0, 1]$ be defined by the equation (see Fig. 2),

$$\eta_r(t) = \begin{cases} 0 & \text{if } t < s \\ (t - s)/r & \text{if } s \leq t < s + r, \text{ and} \\ 1 & \text{for } t > s + r. \end{cases} \quad (30)$$

Then we have,

$$\int_\Omega -\eta_r(u) \operatorname{div} \xi \, dx = \int_\Omega \eta_r'(u) \nabla u^T \xi \, dx \quad (31)$$

$$= \frac{1}{r} \int_{\{s \leq u \leq s+r\}} \nabla u^T \xi \, dx. \quad (32)$$

Now, consider difference quotients for the function m: Assume the vector field $\xi(x)$ satisfies $\|\xi\|_A^* \leq 1$ on Ω, then

$$\frac{m(s+r) - m(s)}{r} = \frac{1}{r} \int_{\{s \leq u \leq s+r\}} \|\nabla u\|_A \, dx \quad (33)$$

$$\geq \frac{1}{r} \int_{\{s \leq u \leq s+r\}} \nabla u^T \xi \, dx \quad (34)$$

$$= \int_\Omega -\eta_r(u) \operatorname{div} \xi \, dx \quad (35)$$

Suppose $m'(s)$ exists at $s \in \mathbf{R}$ and take limits on both sides of the above inequality, then we get

$$m'(s) \geq -\int u^s \operatorname{div} \xi \, dx,$$

because $\eta_r(u) \to u^s$ as $r \to 0$. This conclusion holds for any choice of ξ with $\|\xi\|_A^* \leq 1$, hence $m'(s) \geq J_A(u^s)$, whenever the derivative of m exists. Now, (29) implies,

$$J_A(u) \geq \int_{-\infty}^\infty m'(t) \, ds \geq \int_{-\infty}^\infty J_A(u^s) \, ds.$$

This is the desired inequality, so the proof is complete.

Remark. (a) In (12) we have assumed that $A(x)$ is positive definite. It is actually enough to have a (positive) semidefinite $A(x)$. Although (17) might not make sense, we could work with (16).

(b) It is possible to generalize the class of functionals even further. Let

$$J_\phi(u) = \int_\Omega \phi(x, \nabla u(x)) dx. \quad (36)$$

If ϕ is convex and positively one-homogeneous in its second argument, i.e. $\phi(x, t\xi) = t\phi(x, \xi)$, for any $t > 0$, then

essentially the same proof will hold. A similar duality as in (15) can be constructed via the conjugate function of ϕ (see [2]). Note that this framework includes the metrics considered in [11]. Related is the anisotropic ROF-model considered in [8], but they do not use the corresponding coarea formula.

4. A Continuous Formulation for Multi-Class Segmentation

Next we turn to the problem of multi-class segmentation, that is, we want to assign each pixel in the image to a predefined class of labels. Even though this problem is NP-hard, the discrete graph cut algorithm has proven to produce high-quality solutions via α-expansion moves [4]. It is even possible to derive bounds on the sub-optimality of the obtained solution. We will show that a continuous version of the α-expansion moves can be formulated and that the same bounds hold.

Let $\mathbb{N}$ be a subset of $\mathbb{Z}$ and Ω a subset of $\mathbb{R}^2$. We call a function $l : \Omega \mapsto \mathbb{N}$ a labeling of Ω. If $l(x) = \alpha$ we say that x is labeled α, and $S_\alpha = \{ x \mid l(x) = \alpha \}$ is the set off all points labeled α.

The multi-class segmentation problem can now be posed as minimizing the functional

$$E(l) = E^s(l) + \lambda E^d(l), \quad (37)$$

over all labeling functions $l(x) \in \mathbb{N}$, where

$$E^s(l) = \sum_\alpha \int_{\partial S_\alpha} d\sigma_A, \quad E^d(l) = \sum_\alpha \int_{S_\alpha} g_\alpha dx. \quad (38)$$

Here we use $d\sigma_A$ to indicate that the curve length is taken with respect to the (possibly anisotropic) metric induced by the matrix field $A(x)$.

The difficulties with this optimization problem are that the functional is non-convex and the labeling function may only attain values in $\mathbb{N}$.

4.1. Relation to the Discrete Potts Model

Next we will see that the functional presented is a continuous counterpart to the popular Potts model. If the pixels p and q belong to the same neighborhood the Potts model assigns the cost

$$E_{\text{potts}}(p, q) = \begin{cases} 0 & l(p) = l(q) \\ 1 & l(p) \neq l(q). \end{cases} \quad (39)$$

The Potts model is discontinuity preserving, that is, the penalty $|l(p) - l(q)|$ is independent of the distance between the class values. This is desirable since functionals that do not have this property tend to produce labelings where the transitions between classes are taken in several steps. Note that such a cost can never be convex in the label function.

To see the connection to the Potts model we define the indicator function of a label as

$$\chi_\alpha(x) = \begin{cases} 1 & \text{if } x \in S_\alpha \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

We let $E_\alpha^s(l)$ and $E_\alpha^d(l)$ denote the individual terms in the sums of the smoothness term and the data term. If we assume that $A(x) = I$ then we may write the smoothness terms as

$$E_\alpha^s(l) = \int_\Omega |\nabla \chi_\alpha| dx. \tag{41}$$

As in the two-class case the connection to the discrete formulation is that if instead of using the euclidean distance we use $|(x_1, x_2)| = |x_1| + |x_2|$ and choose the gradient discretization as (11) then $\sum_{i,j,\alpha} |\bar{\nabla} \chi_\alpha(i,j)|/2$ is the cost in the Potts model on a four-connected grid.

5. The α-Expansion Algorithm

As we have seen, minimizing our objective functional may seem troublesome since it is non convex and the labeling may only take certain discrete values. However, in this section we will show that it is possible to apply a continuous version of the popular α-expansion algorithm [4] , to obtain provably good solutions.

We first consider the α-expansion move. In this move we fix a class α and try to reduce the current labeling cost by allowing all pixels not labeled α to change their label to α. We let u be an indicator function of the pixels *not* in the class α after the move. As pixels in the class α are not allowed to change class we have the constraint $1 - u \geq \chi_\alpha$. The move can be characterized in terms of the indicator functions by the update rules

$$1 - u \mapsto \chi_\alpha \tag{42}$$
$$\chi_\beta u \mapsto \chi_\beta, \tag{43}$$

where $\beta \neq \alpha$, and $u : \Omega \mapsto \{0,1\}$. Inserting into (37) we see that our energy functional can be written

$$E(u) = E_\alpha^s(u) + \lambda E_\alpha^d(u) + \sum_{\beta \neq \alpha} (E_\beta^s(u) + \lambda E_\beta^d(u)) \tag{44}$$

where

$$E_\alpha^s(u) = \int_\Omega |Q\nabla(1-u)| dx, \ E_\alpha^d(u) = \int_\Omega g_\alpha(1-u) dx \tag{45}$$

$$E_\beta^s(u) = \int_\Omega |Q\nabla(\chi_\beta u)| dx, \ E_\beta^d(u) = \int_\Omega g_\beta \chi_\beta u \, dx \tag{46}$$

and $A(x) = Q(x)Q(x)$ as before. (Here we have written $E(u)$, and not as previously $E(l)$, as the labeling function will only depend on u). Hence the α-expansion move can be determined by optimizing the functional with respect to the binary function u.

5.1. The Thresholding Theorem II

In order to be able to compute the minimum of $E(u)$ efficiently we need to establish that the thresholding property (6) holds. We deduce it from the coarea formula proved in Section 3.

First we consider the smoothness terms E_α^s and E_β^s. Since $|Q\nabla(1-u)| = |Q\nabla u|$ we see that $E_\alpha^s(u)$ reduces to $J_A(u)$ for which the coarea formula (14) holds. Note that exchanging the terms $1 - u$ and u means that we are considering the boundary of the exterior of the set instead of the interior. Clearly these boundaries are the same. For $E_\beta^s(u)$ we note that

$$(\chi_\beta u)^t = \chi_\beta u^t \tag{47}$$

since χ_β is binary. Inserting into (14) we get

$$E_\beta^s(u) = J_A(\chi_\beta u) = \int_{-\infty}^{\infty} J_A(\chi_\beta u^t) dt = \int_{-\infty}^{\infty} E_\beta^s(u^t) dt. \tag{48}$$

Since $u(x) \in [0,1]$ and therefore u^t constant, both $E_\alpha^s(u^t)$ and $E_\beta^s(u^t)$ vanish, if $t \notin [0,1]$, and we conclude that the thresholding property holds.

The proof for the data term is exactly the same as for the two-class case (see Section 2.1).

5.2. Optimality Bound for α-expansion

In this section we give an optimality-bound similar to what is given for the discrete version in [4]. The proof is more or less the same. The only thing that needs to be done is to adjust the arguments to a continuous setting.

Let l^* be the globally optimal solution of $\min_l E(l)$, $\hat{l}$ a local minimum with respect to α-expansion moves, and S_α^* be the set $\{x \in \Omega; l^*(x) = \alpha\}$. We define the labeling l^α by

$$l^\alpha(x) = \begin{cases} \alpha & \text{if } x \in S_\alpha^* \\ \hat{l}(x) & \text{otherwise} \end{cases} \tag{49}$$

Since l^α can be reached from $\hat{l}$ with one α-expansion we have $E(\hat{l}) \leq E(l^\alpha)$, since $\hat{l}$ is a local minimum. Let $E|_S(l)$ be the restriction of the energy to a subset $S \subset \Omega$. Then we have

$$E\big|_{\text{int}(S_\alpha^*)}(l^\alpha) = E\big|_{\text{int}(S_\alpha^*)}(l^*), \tag{50}$$

$$E\big|_{\text{ext}(S_\alpha^*)}(l^\alpha) = E\big|_{\text{ext}(S_\alpha^*)}(\hat{l}) \tag{51}$$

by the definition of l^α. Furthermore on the boundary ∂S_α^* of S_α^* we have

$$E\big|_{\partial S_\alpha^*}(l^\alpha) \leq E\big|_{\partial S_\alpha^*}(l^*), \tag{52}$$

since both the left hand side and the right hand reduces to a measurement of the length of the boundary ∂S_α^*. (The

data term vanishes since ∂S_α^* has measure zero and the functions involved in the data term are bounded.) Since $\Omega = \text{int}(S_\alpha^*) \cup \text{ext}(S_\alpha^*) \cup \partial S_\alpha^*$ we have

$$E(\hat{l}) = E\big|_{\text{int}(S_\alpha^*)}(\hat{l}) + E\big|_{\text{ext}(S_\alpha^*)}(\hat{l}) + E\big|_{\partial S_\alpha^*}(\hat{l}) \quad (53)$$

$$E(l^\alpha) \le E\big|_{\text{int}(S_\alpha^*)}(l^*) + E\big|_{\text{ext}(S_\alpha^*)}(\hat{l}) + E\big|_{\partial S_\alpha^*}(l^*) \quad (54)$$

Since $\hat{l}$ is a local minimum we get

$$E\big|_{\text{int}(S_\alpha^*)}(\hat{l}) + E\big|_{\partial S_\alpha^*}(\hat{l}) \le E\big|_{\text{int}(S_\alpha^*)}(l^*) + E\big|_{\partial S_\alpha^*}(l^*) \quad (55)$$

This gives us an estimate of the energy on $\text{int}(S_\alpha^*) \cup \partial S_\alpha^*$. Since $\hat{l}$ is a local minimum with respect to expansion moves for all the labels the same bound is valid for the other labels. Summing over α we get we get a bound on the entire domain Ω Note that as in the discrete case the boundary terms will occur twice on both sides. Letting $B = \cup \partial S_\alpha^*$ we obtain

$$E(\hat{l}) + E\big|_B(\hat{l}) \le E(l^*) + E\big|_B(l^*) \quad (56)$$

Finally since the $E\big|_B$ is nonnegative and $E\big|_B(l^*) \le E(l^*)$ we obtain the desired bound

$$E(\hat{l}) \le 2E(l^*) \quad (57)$$

which shows that any local minimum with respect to expansion moves is at worst 2 times the optimal energy.

6. Implementation

As we have seen the continuous cut problem can be posed as a convex optimization problem. It might still be quite difficult to solve since the objective function is not differentiable everywhere and we have the constraints $u(x) \in [0,1]$. However it was shown in [6] that the solution to such a problem can be obtained by solving a related unconstrained problem. Let J be a convex, positively one-homogeneous ($J(tu) = tJ(u)$ if $t \ge 0$) functional fulfilling the coarea formula (10). Then a solution to the problem

$$\min_{u(x) \in [0,1]} J(u) + \lambda \int_\Omega f u \, dx \quad (58)$$

can be obtained by thresholding the solution of

$$\min_w J(w) + \frac{\lambda}{2} \int_\Omega (w + f)^2 \, dx \quad (59)$$

at zero. The program (59) have a number of advantages. First of all it is an unconstrained minimization problem, second the the $\mathbb{L}_2$ term is very well conditioned. If the J functional is just the TV-norm then we recognize this as the Rudin-Osher-Fatemi (ROF) denoising model [17]. This problem has been well studied and there exist freely available implementations for solving it.
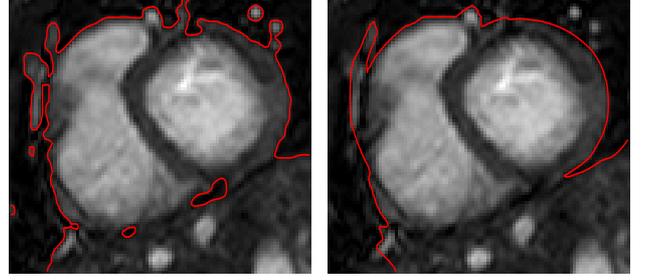

Figure 3. MR-slice of a heart. Left: Segmentation using the isotropic formulation. Right: Anisotropic segmentation where approximately circular cuts are preferred.

In the α-expansion case the smoothness term is somewhat more complicated. However, it turns out that with a simple approximation we may put it into this form. The problem is the term $|\nabla(\chi_\beta u)|$. Using the the triangle inequality we see that

$$|Q\nabla(\chi_\beta u)| \le |Q\nabla\chi_\beta|u + \chi_\beta|Q\nabla u|. \quad (60)$$

Looking at (58) we see that the first term on the right hand side of (60) will contribute to the data term while the second will contribute to a weighted TV-norm term. Although this is an approximation, in many cases (60) is actually fulfilled with equality. We also have to incorporate the constraint $(1 - u) > \chi_\alpha$. However this is easily done by modifying the data term f. In our implementation we use the approximate energy to create a starting guess for a gradient descent search. This allows us to use freely available methods for solving (59) such as the one presented in [7]. We used this method since it is a second order method and therefore very accurate. On the other hand this renders it slow for large problems. Solving the ROF model for an image of size 200×200 pixels (as in Figure 4) takes around 10 seconds and for the 600×600 image (in Figure 5) around 50 seconds. Although efficient implementations are outside the scope of this paper we mention that there are much more efficient methods, see [15] and the references therein.

7. Experiments

Next we will present a few experiments that demonstrates the possible benefits of being able to use an anisotropic smoothing term. Figure 1 shows a simple synthetic experiment with a circle. For the pixels outside the circle the cost is $g_b = 0$ (the cost of assigning pixels to background) and $g_b = 1$ (foreground cost). For the white pixels inside the circle $g_b = 1$ and $g_f = 0$. In addition there is a gray area where $g_b = g_f = 0.5$. The cost functions was also smoothed using a Gaussian kernel in order not to have perfect transitions between the classes. To the left is the discrete isotropic version. In this case curve length is

measured using the $\mathbb{L}_1$-distance ($|(x_1, x_2)| = |x_1| + |x_2|$). Because of the metric the solution is actually not unique. To the right is the isotropic segmentation. In this case the segmentation of the gray area is determined by the curve length of the boundary. On the second row is the anisotropic version. The matrix field is constructed such that the largest eigenvalues of $A(x)$ are $\sqrt{0.9}$ and $\sqrt{0.1}$. The eigenvector corresponding to the largest eigenvalue is directed in the normal direction of the circle.

Next we present a simple experiment that illustrates that it is possible to incorporate prior knowledge to affect the shape of the cut. Figure 3 shows an MR-slice of a heart. T o the left is the isotropic solution, and to the right is the anisotropic version. Here the matrix field $A(x)$ is constructed as to prefer boundaries which are roughly circles around the center of the left ventricle (similar as before).

Figure 4 shows an image of the amazon river delta. In this experiment the objective is to segment the river. To obtain probability distributions for the background and for the foreground classes we manually marked some areas from each class in the image. From these distributions the data term was constructed. The cost of assigning individual pixels to background is shown in the middle image, first row. To the right is the anisotropic segmentation. In this case the matrix field was constructed using the structure tensor. If $S(x)$ is the structure tensor then $A(x) = (I + S(x))^{-1}$. The resulting metric favors a cut if there is a clear edge (or corner) in the image. Furthermore it favors cuts in the same direction as the edge. To reduce sensibility to noise we also put a threshold on the eigenvalues of the structure tensor. If the eigenvalues are smaller than a certain threshold we set the eigenvalue to zero. In this way only strong edges will affect the segmentation. On the second row of Figure 4 is two isotropic segmentations. In the first one (left) we use the same weighting between the data term and smoothness term as for the anisotropic case. Since long thin structures have a proportionally large boundary the isotropic smoothness term removes some parts of the river. In the second one (middle) we select a weight such that the level of details of the river is approximately the same as for the anisotropic case. In this case we see that a few new areas of pixels outside on land is classified as river. This is also to be expected since in this case the smoothness term is weaker. By contrast, the anisotropic version manages to acheive a high level of detail without producing any spurious unwanted regions. For comparison we also plotted the discrete graph cut solution with the weight selected to give roughly the same level of detail as the anisotropic solution. Finally we present a four class segmentation experiment. Figure 5 shows the image that is to be segmented. We use the classes grass, bricks and asphalt, and the fourth class correspond to pixels not in any or the previous classes. To obtain probability distributions for the three first classes we manually marked
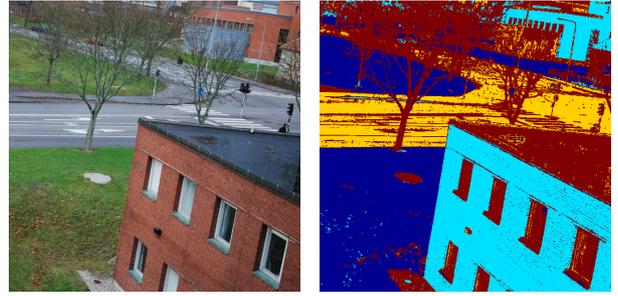

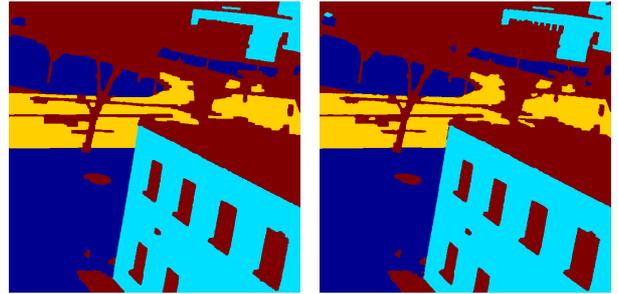Figure 5. Four class segmentation. Left: An image of an intersection. Right:Result with no smoothness term.

Figure 6. Left: The discrete α-expansion solution is shown. Right: The continuous solution.
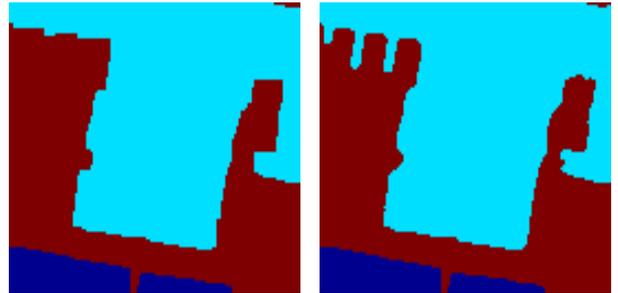

Figure 7. Close-ups of the results above: discrete (left) and continuous (right).

some areas from each class in the image. The pixels in these areas have then used to estimate the distributions. The distribution for the fourth class was taken to be uniform over the RGB space. The data term E_α^d was then selected as $g_\alpha(x) = 1 - p_\alpha(x)$, where $p_\alpha(x)$ is the probability of each pixel belonging to class α. Figure 7 shows the results when only using the data term. Figures 6 and 7 show the discrete and continuous solutions. In this case the weight λ was set to 1, and for the discrete solution a four connected grid was used. On a larger scale they look quite similar however upon closer inspection it is possible to see that the discrete version does prefer straight edges as opposed to smooth boundaries. Hence, if small metrication errors are important, one should use continuous α-expansion.
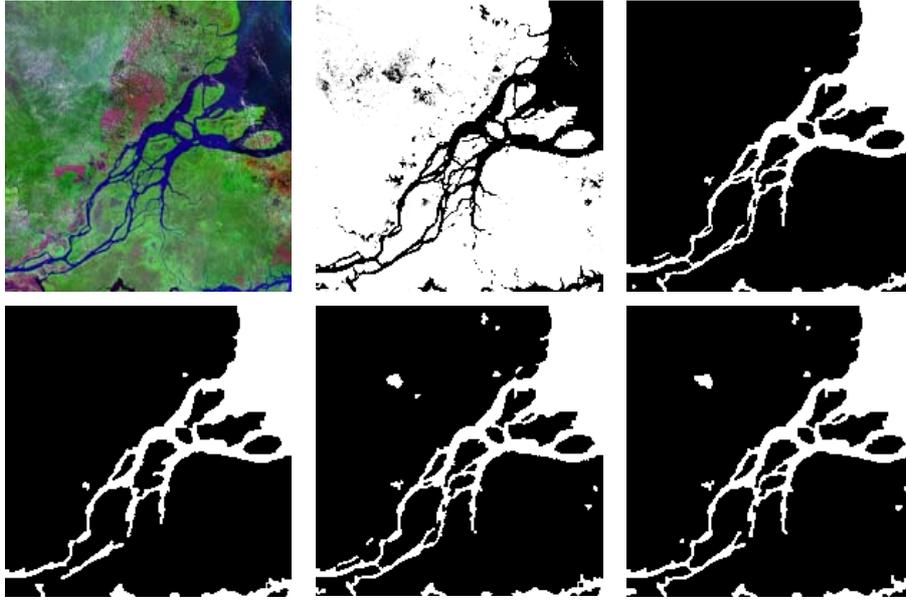
Figure 4. Segmentation of the amazon river delta. First row, left: Original image. Middle: Data term. Right: Anisotropic segmentation. Second row, left: Smooth isotropic segmentation. Middle: High detail isotropic segmentation. Right: High detail discrete isotropic graph cut.

References

[1] B. Appleton and H. Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(1), 2006.

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

[3] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *European Conference on Computer Vision (ECCV)*, Nice, France, 2003.

[4] Y. Boykov, O. Veksler, and Z. Ramin. Fast approximate energy minimization via graph cuts. *Transactions on Pattern Analysis and Machine Intelligence*.

[5] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging Vision*, 28:151–167, 2007.

[6] A. Chambolle. Total variation minimization and a class of binary mrf models. *Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition*, 2005.

[7] T. Chan, G. Golub, and P. Mulet. A nonlinear primal-dual method for total variation based image restorations. *SIAM Journal of Scientific Computing*, 20:1964–1977, 1999.

[8] S. Esedoglu and S. Osher. Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. *Comm. Pure Appl. Math.*, 57:1609–1626, 2004.

[9] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions Studies in Advanced Mathematics*. CRC-Press, 1st edition, 1992.

[10] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. *European Conf. on Computer Vision*, pages 332–345, 2008.

[11] V. Kolmogorov and Y. Boykov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *International Conference on Computer Vision (ICCV)*, Nice, France, 2005.

[12] M. Nikolova, S. Esedoglu, and T. F. Chan. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.

[13] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A convex relaxation approach for computing minimal partitions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, 2009.

[14] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, Marseille, France, 2008.

[15] T. Pock, M. Unger, D. Cremers, and H. Bischof. Fast and exact solution of total variation models on the GPU. In *CVPR Workshop on Visual Computer Vision on GPU's*, 2008.

[16] H. L. Royden. *Real Analysis*. Prentice Hall, 3rd edition, 1988.

[17] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[18] G. Strang. Maximal flow through a domain. *Mathematical Programming*, 26(2):123–143, 1983.

[19] W. Trobin, T. Pock, D. Cremers, and H. Bischof. Continuous energy minimization via repeated binary fusion.

[20] C. Zach, M. Niethammer, and J.-M. Frahm. Continuous maximal flows and Wulff shapes: Application to MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, 2009.