A First Derivative Potts Model for Segmentation and Denoising Using MILP

Ruobing Shen\textsuperscript{1}, Gerhard Reinelt\textsuperscript{1}, and Stéphane Canu\textsuperscript{2}

\textsuperscript{1} Institute of Computer Science, Heidelberg University, 69120 Heidelberg, Germany 
ruobing.shen@informatik.uni-heidelberg.de
\textsuperscript{2} LITIS, Normandie University, INSA Rouen, 76800 Rouen, France

Abstract. Unsupervised image segmentation and denoising are two fundamental tasks in image processing. Usually, graph based models such as multicut are used for segmentation and variational models are employed for denoising. Our approach addresses both problems at the same time. We propose a novel MILP formulation of a first derivative Potts model, where binary variables are introduced to directly deal with the $\ell_0$ norm. As a by-product the image is denoised. To the best of our knowledge, it is the first global mathematical programming model for simultaneous segmentation and denoising. Numerical experiments on real-world images are compared with multicut approaches.

Keywords: image segmentation, denoising, Potts model, mixed integer programming, multicut.

1 Introduction

Segmentation is a fundamental task for extracting semantically meaningful regions from an image. In this paper we consider the problem of partitioning a given image into an unknown number of segments. Since we assume that no prototypical features about the segments are available, it is a so-called unsupervised image segmentation problem. In a general setting this problem is NP-hard. Exact optimization models such as the multicut problem \cite{1,2} are based on integer linear programming (ILP) and solved using branch-and-cut methods.

Another aspect of image processing is denoising. Main tools for denoising are variational methods like the approach with Potts priors which was designed to preserve sharp discontinuities (edges) in images while removing noises. Given $n$ signals, denote their intensities $y = (y_1, y_2, \ldots, y_n)$ (e.g. grey scale or color values) and define $w = (w_1, w_2, \ldots, w_n)$ as the vector of denoised values. The classical (discrete) Potts model (named after R. Potts \cite{3}) has the form

$$\min_w \|w - y\|_k + \lambda \|\nabla^1 w\|_0,$$

where the first part measures the $\ell_k$ norm difference between $w$ and $y$, and the second part measures the number of oscillations in $w$. Recall that the discrete first derivative $\nabla^1 x$ of a vector $x \in \mathbb{R}^n$ is defined as the $n - 1$ dimensional vector $(x_2 - x_1, x_3 - x_2, \ldots, x_n - x_{n-1})$ and the $\ell_0$ norm of a vector gives its number of nonzero entries.
The scalar $\lambda$ is a parameter for regularization. Recently, various modifications and improvements have been made for the Potts model, see [4] for an overview.

In general, solving the discrete Potts model (1) is also NP-hard. In [5] local greedy methods are used to solve it. Recently, [6] uses a mixed integer linear programming (MILP) formulation to deal with the $\ell_0$ norm for a similar problem in statistics called the best subset selection problem.

We combine segmentation and denoising in one framework that is motivated by the above mentioned two models. We assume that the input is an image given as grey scale values (RGB images can be easily transformed) for pixels located on an $m \times n$ grid. Let $V = \{p_{1,1}, \ldots, p_{m,n}\}$ denote this set of pixels. For representing relations between neighboring pixels, we define the corresponding grid graph $G = (V, E)$ where $E$ contains edges between pixels which are horizontally or vertically adjacent. A general segmentation is a partition of $V$ into sets $\{V_1, V_2, \ldots, V_k\}$ such that $\bigcup_{i=1}^k V_i = V$, and $V_i \cap V_j = \emptyset$, $i \neq j$. So in graph-theoretical terms the image segmentation problem corresponds to a graph partitioning problem.

The paper is organized as follows. In Section 2 we introduce our MILP formulation of problem (1) for the 1D signal case. We then review the multicut problem in Section 3 and introduce the multicut constraints for the grid graph. Section 4 presents our main MILP formulation for 2D images. Computational experiments for real application problems are presented in Section 5. We conclude and point to future work in Section 6.

2 The First Derivative Potts Model: 1D

We are given $n$ signals $p = (p_1, \ldots, p_n)$ with coordinates $z = (z_1, \ldots, z_n)$ (in some interval $D \subseteq \mathbb{R}$) and intensities $y = (y_1, \ldots, y_n)$. We call a function $f$ piecewise constant over $D$ if there is a partition of $D$ into subintervals $D_1, \ldots, D_k$ such that $D = \bigcup_{i=1}^k D_i$, where $D_i \cap D_j = \emptyset$, and $f$ is constant when restricted to $D_i$. Throughout the paper, we assume the input images or signal intensities $y$ contain noise, and the task of segmentation and denoising is transformed into a piecewise constant fitting problem. The fitting value for signal $p_i$ is denoted $w_i = f(z_i)$.

In 1D the associated graph $G(V, E)$ is simply a chain, where $V = \{p_i \mid i \in [n]\}$ and $E = \{e_i = (p_i, p_{i+1}) \mid i \in [n-1]\}$ with $[n]$ denotes the set $1, 2, \ldots, n$. We propose to formulate problem (1) as an MILP by introducing $n - 1$ binary variables $x_{e_i}$. We define $x_{e_i} = 1$ if and only if the end nodes of $e_i$ are in different segments. If $x_{e_i} = 1$, the edge is called active, otherwise it is dormant. Since $w$ is constant within the same segment, it follows that $w_{i+1} - w_i \neq 0$ if and only if $x_{e_i} = 1$. Thus the active edges define the segments and the number of segments is $\sum_{i=1}^{n-1} x_{e_i} + 1$. See the left part of Figure 1 for an example.

A mixed integer programming (MIP) formulation for (1) is

$$\min \sum_{i=1}^{n} |w_i - y_i| + \lambda \sum_{i=1}^{n-1} x_{e_i}$$

(2a)

$$|w_{i+1} - w_i| \leq M x_{e_i}, \quad i \in [n-1],$$

(2b)

$$w_i \in \mathbb{R}, \quad i \in [n],$$

(2c)

$$x_{e_i} \in \{0, 1\}, \quad i \in [n-1].$$

(2d)
Note that we use the $\ell_1$ norm because it can be easily modeled with linear constraints. Namely, constraint (2a) is replaced by the two constraints $w_{i+1} - w_i \leq Mx_{e_i}$ and $-w_{i+1} + w_i \leq Mx_{e_i}$, and the term $|w_i - y_i|$ is replaced by $\varepsilon^+_i + \varepsilon^-_i$ where $w_i - y_i = \varepsilon^+_i - \varepsilon^-_i$ and $\varepsilon^+_i, \varepsilon^-_i \geq 0$. From now on, for simplicity, we will just specify models in form (2).

Constraint (2a) enforces that the pixels corresponding to the end nodes of a dormant edge have the same fitting value. We set the constant $M$ as $y^* = y_{\text{max}} - y_{\text{min}}$, where $y_{\text{max}}$ is the largest value of $y_i$ and $y_{\text{min}}$ the smallest. The regularization parameter is set to $\lambda = \frac{1}{2}\sigma_1 y^*$, where $\sigma_1$ is a user-defined parameter. When there exists extreme outliers, model (2) will not treat the outliers as separate segments, since doing so would incur the penalty $2\lambda$. The solution of (2) gives fitting value $w_i$ for the signal $p_i$ and the boundary of the segments are given by the edges $x_e = 1$.

### 3 The Multicut Problem

The multicut problem [1] formulates the graph segmentation problem as an edge labeling problem. For a partition $V = \{V_1, V_2, \ldots, V_k\}$ of $V$, the edge set $\delta(V_1, V_2, \ldots, V_k) = \{uv \in E \mid \exists i \neq j$ with $u \in V_i$ and $v \in V_j\}$ is called multicut induced by $\mathcal{V}$. Similar to Section 2, we introduce binary edge variables $x_e$ and represent the multicut as the set of active edges.

With edge weights $c : E \to \mathbb{R}$ representing differences between pixels intensities, the multicut problem for unsupervised partitioning [2] can be formulated as the ILP

$$\begin{align*}
\min & \quad \sum_{e \in E} -c_e x_e + \sum_{e \in E} \lambda x_e \\
\text{s.t.} & \quad \sum_{e \in C \setminus \{e'\}} x_e \geq x_{e'}, & \forall \text{ cycles } C \subseteq E, e' \in C, \\
& \quad x_e \in \{0, 1\}, & \forall e \in E,
\end{align*}$$

where $\lambda$ again serves as the penalty term for the number of active edges. Constraints (3a) are called multicut constraints and they enforce the consecutiveness of the active edges, and in turn the connectedness of each segment. Thus each maximal set of vertices induced only by dormant edges corresponds to a segment. Problem (3) is NP-hard in general and while the number of inequalities can be exponentially large, violated constraints can be found efficiently using shortest path algorithms, and added iteratively until the solution is feasible [12].
It is well known that if the cycle $C$ is chordless, then constraints (3a) are facet-defining for the multicut polytope \[2\]. In a grid graph, although the number of such constraints is still exponential, it may be advantageous to first restrict to the 4-edge cycles of the grid (see in Sec. 5). The right part of Figure 1 shows a partition of a $4 \times 4$-grid graph into 3 segments where the dashed active edges form the multicut.

The multicut problem \[3\] has two main weaknesses. First, it is sensitive to noise, and usually requires some denoising methods as pre-processing. Second, the result heavily depends on the setting of the single parameter $\lambda$ and it is hard to control the desired number of segments.

4 The First Derivative Potts Model in 2D

We are given pixels on an $m \times n$-grid image $V = (p_{i,j})$ with intensity values $Y = (y_{i,j})$. The $i$-th row of $Y$ is denoted by $Y^r_i$ and the $j$-th column by $Y^c_j$. We divide $E = E^r \cup E^c$ into its horizontal (row) edges $E^r$ and its vertical (column) edges $E^c$. The problem is to find the fitting value $W = (w_{i,j}) \in \mathbb{R}^{m \times n}$, and the boundaries (in terms of active edges) of each segment.

The main formulation is modeled as a first derivative Potts model and is obtained by formulating (2) per row and column. In addition to the 1D signal case, we need the multicut constraints (3a) to enforce the consecutiveness of the active edges in 2D. This formulation overcomes the first weakness of the multicut problem. To tackle the second, we introduce edge weights for each row and column. We further add constraints, called the regularity constraints, to bound the number of active edges, per row and column.

4.1 Main formulation

Denote $e^r_{i,j} \in E^r$ as horizontal (row) edge $(p_{i,j}, p_{i,j+1})$ and $e^c_{i,j} \in E^c$ as vertical (column) edge $(p_{i,j}, p_{i+1,j})$. Our main formulation is

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} |w_{i,j} - y_{i,j}| + \sum_{i=1}^{m} \lambda^r_i \sum_{j=1}^{n-1} x_{e^r_{i,j}} + \sum_{j=1}^{n} \lambda^c_j \sum_{i=1}^{m-1} x_{e^c_{i,j}}$$

(4)

$$|w_{i+1,j} - w_{i,j}| \leq M^r_i x_{e^r_{i,j}}, \quad i \in [m], \quad j \in [n - 1],$$

(4a)

$$|w_{i,j+1} - w_{i,j}| \leq M^c_j x_{e^c_{i,j}}, \quad j \in [n], \quad i \in [m - 1],$$

(4b)

$$\sum_{e \in C \setminus \{e'\}} x_e \geq x_{e'}, \quad \forall \text{ cycles } C \subseteq E, e' \in C,$$

(4c)

$$\sum x_{e^r_{i,j}} \leq k^r_i, \quad i \in [m],$$

(4d)

$$\sum x_{e^c_{i,j}} \leq k^c_j, \quad j \in [n],$$

(4e)

$$w_{i,j} \in \mathbb{R}, \quad i \in [m], \quad j \in [n],$$

(4f)

$$x_e \in \{0, 1\}, \quad e \in E.$$

(4g)

The constants $M$ and $\lambda$ are computed similarly as in (2), now individually for row $i$ and column $j$. In addition, we have the regularity constraints (4d)–(4g). Constant $k$ is
computed as follows. We first compute the average intensity of each $4 \times 4$ pixel block of the image, and then calculate the absolute difference of their maximum and minimum value, denoted $Y^*$. So $Y^*$ somehow represents the global contrast of the image. Since the input image contains noise, we take only edges with weight greater than $\sigma_2 Y^*$ as potential active edges into account, where $0 < \sigma_2 < 1$ is some suitably chosen parameter. Then $k_i^r$ is computed as the number of elements in $\nabla_1 Y^r_i$ that are greater than $\sigma_2 Y^*$. Constant $k_j^c$ is computed similarly for each column $j$.

5 Computational Experiments

Computational experiments were performed using Cplex 12.6.1, on a Intel i5-4570 quad-core desktop, with 16GB RAM. We compare three different models on two images (see left part of Figure 2) taken from [2], all without preprocessing. The first one is model (3), the second is (4) without the regularity constraints (4f)–(4g), and the third one is model (4). We apply the cutting plane approach like in [1,2], i.e., first ignore the multicut constraints (3h), and check the feasibility of the solution, then add violated constraints on-the-fly. We implemented Cplex lazy constraints callback, and set the parameters to default. The two images are resized to $40 \times 40$ and $41 \times 58$, with Gaussian noise (first row of Figure 2 on the right) and Salt and Pepper (last two rows) added.

The right part of Figure 2 shows the segmentation results, the first two columns are model 1 with different parameters $\lambda$. Third and fourth column represent model 2 and 3. The number $S$ in Figure 2 stands for the number of segments. For example, $\lambda = 0.22$ on the top left corner, and its right image shows the result of changing $\lambda$ to 0.195 ($S$...
changes from 70 to 50). Model 1 takes less than one sec in all three instances. Although both found optimal solution of the first instance in less than 6 sec, model 2 and 3 did not converge in the second and third instance, given the time limit 50 and 100 sec. The Cplex optimality gaps are 1.91%, 1.75%, 27.1% and 21.9%, respectively.

Although requiring more computational time, model 2 and 3 seem to be more robust to noise, less sensitive to parameters, and give better segmentation results. In addition, we found it beneficial to add the 4-cycle constraints as hard constraints. This is demonstrated and applied in model 2 and 3, for the third instance. Without them, both models failed to find meaningful results in the 100 sec time limit. Furthermore, model 3 requires the fewest cut callbacks amongst all, on average 2 cuts, compared to 60 for model 2 and 707 for model 1. It closes optimality gaps faster than model 2 in all three instances.

6 Conclusions and Future Work

We present a combined unsupervised image segmentation and denoising framework, which is modeled as MILP that solves the problems to optimality. This is a global model, in the sense that firstly, it can use any heuristic results, like [5], as an initial solution. Secondly, it could improve the initial solution by finding a better upper bound within the branch and bound framework in Cplex. The main limitations of the paper are that problem (4) works on grid graph only, and it does not scale to larger images.

Because (4) is NP-hard, in the future, decomposition algorithms such as superpixel lattices [8] can be used as preprocessing, in order to work on larger images. We will also explore the possibilities of applying (4) to 3D images. Since the underlying problem is the piecewise constant fitting problem, applications beyond the scope of computer vision are also of interest.

References

1. B. Andres, J. H. Kappes, T. Beier, U. Köthe, and F. A. Hamprecht. Probabilistic image segmentation with closedness constraints. In International Conference on Computer Vision, pages 2611–2618, 2011.
2. J. H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schnörr. Globally optimal image partitioning by multicuts. In International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, pages 31–44, 2011.
3. R. B. Potts and C. Domb. Some generalized order-disorder transformations. Proceedings of the Cambridge Philosophical Society, 48:106–109, 1952.
4. T. Chan, S. Esedoglu, F. Park, and A. Yip. Total variation image restoration: Overview and recent developments. In Nikos Paragios, Yunmei Chen, and Olivier Faugeras, editors, Handbook of Mathematical Models in Computer Vision, pages 17–31. "Springer”, 2006.
5. R. M. H. Nguyen and M. S. Brown. Fast and effective l0 gradient minimization by region fusion. In International Conference on Computer Vision, pages 208–216, 2015.
6. D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. Ann. Statist., 44(2):813–852, 2016.
7. M. Storath and A. Weinmann. Fast partitioning of vector-valued images. SIAM J. Imaging Sci., 7:1826–1852, 2014.
8. A. P. Moore, S. J. D. Prince, and J. Warrell. Constructing superpixels using layer constraints. In International Conference on Computer Vision, pages 2117–2124, 2010.