An Adaptive Subdivision Scheme for Quadratic Programming in Multi-Label Image Segmentation

Marko Rak
rak@isg.cs.ovgu.de
Tim König
tim.koenig@isg.cs.ovgu.de
Klaus-Dietz Tönnies
toennies@isg.cs.ovgu.de

Department of Simulation and Graphics
University of Magdeburg, Germany

Convex quadratic optimization is one of the most widely used concepts in image segmentation. It facilitates a wide range of information sources, such as edge, intensity, texture and shape. The problem is especially challenging for the multi-label case, even being NP-hard in its most general setting. Therefore, fast "solutions", as the α-expansion of [1], are limited to local optimality. Addressing this problem, several approaches relax the labeling integrality condition, resulting in quadratic programs (QPs) like in [2] and in [4], which can be solved in polynomial time.

Although this is efficient in a theoretical sense, large-scale QPs that arise from typical multi-label tasks can rarely be used for image segmentation directly due to either time or space constraints, or both. We address this issue by an adaptive domain subdivision scheme, reducing the problem to a short sequence of spatially smoothed medium-scale QPs, which subsequently better approximate the large-scale program. Our scheme is globally optimal in terms of the approximated problem.

Putting our main focus on the subdivision, we restrict ourselves to minimization of the popular but rather simple piecewise constant Mumford-Shah functional. Therefore, we seek for a labeling that trades off the length of the labeling border and the approximation of image intensity u by known reference intensitites u_i for each label i. For discrete domains the associated energy can be written as

$$E = \sum_{j\in\mathcal{P}} \sum_i \alpha_i \, (u-u_i)^2 \, x_i^j + \gamma \sum_{(j,k)\in\mathcal{N}} \sum_i (x_i^j - x_i^k)^2, \tag{1}$$

with $\mathcal{P}$ being the set of domain elements and $\mathcal{N}$ being the set of neighborhood relations. The binary variable x_i^j indicates whether domain element j mutually exclusively belongs to label i or not.

Circumventing the hardness of general convex quadratic labelings, we remove the binarity constraint while keeping the rest of the problem fixed. For any quadratic labeling energy the problem then transforms into a $[0,1]$-relaxed binary QP (2 – 4) in $\mathbf{x} = (x_1^1 x_1^2 \dots x_l^{p-1} x_l^p)^\mathrm{T}$.

$$\mathbf{x}^\mathrm{T}\mathbf{H}\mathbf{x} + \mathbf{f}^\mathrm{T}\mathbf{x} \quad \to \min! \tag{2}$$

$$\mathbf{A}\mathbf{x} = \mathbf{1} \tag{3}$$

$$\mathbf{0} \preceq \mathbf{x} \preceq \mathbf{1} \tag{4}$$

Here, the element-wise mutual label exclusiveness degrades to affinity, i.e., $\forall j : \sum_i x_i^j = 1$, which is reflected by Eq. (3). For the Mumford-Shah example, matrix $\mathbf{H}$ and vector $\mathbf{f}$ subsume the neighborhood relations and the intensity approximation terms of Eq. (1), respectively.

Clearly, the above problem is of large scale, because $\mathbf{x}$ contains $p \cdot l$ variables, where p is the number of domain elements and l is the number of labels. Treating this issue, we seek for an adaptive domain subdivision that gradually becomes finer near borders of the (unknown) labeling. If such a domain subdivision would be known beforehand then we could set up l label variables for each of the r subdivision regions and arrange them altogether into a vector $\mathbf{z} = (z_1^1 z_1^2 \dots z_l^{r-1} z_l^r)^\mathrm{T}$. Furthermore, we could set up a projector $\mathbf{P}$ which by

$$\mathbf{x} = \mathbf{P}\mathbf{z} \tag{5}$$

superimposes the regional basis in $\mathbf{z}$ to the element basis in $\mathbf{x}$ according to the subdivision. Inserting the superimposition (5) into the QP (2 – 4), we would obtain a spatially smoothed approximation of the large-scale program. The approximate QP in $\mathbf{z}$ would be of medium scale, because $r \ll p$ unless the mild assumption of spatial labeling coherence is violated.

Certainly, we do not know the subdivision beforehand,which is why we have to (re-)construct it alongside the labeling. To attain this goal we draw on a hierarchical domain subdivision via quadtree, octree and their generalization to higher dimensions. The key idea is to create a short sequence $(0, 1, \dots, i, \dots, c)$ of superimpositions

$$\mathbf{x} = \mathbf{P}_i \mathbf{z}_i \tag{6}$$

Figure 1: Comparison of our adaptive subdivision scheme to the large-scale quadratic program on standard image data. First column: input images, second column: labeling of the large-scale quadratic program, third column: labeling of our adaptive subdivision scheme for $c = 5$

that stem from subsequent adaptive subdivision refinement steps. More precisely, we initially decompose the domain up to a preset tree level c above ground, i.e., above the 0-level where regions equal domain elements. We then alternate $c+1$ times between two steps. In the first step, we set up and solve the current approximate QP

$$\mathbf{z}_i^\mathrm{T}\mathbf{P}_i^\mathrm{T}\mathbf{H}\mathbf{P}_i\mathbf{z}_i + \mathbf{f}^\mathrm{T}\mathbf{P}_i\mathbf{z}_i \quad \to \min! \tag{7}$$

$$\mathbf{A}\mathbf{P}_i\mathbf{z}_i = \mathbf{1} \tag{8}$$

$$\mathbf{0} \preceq \mathbf{z}_i \preceq \mathbf{1}, \tag{9}$$

based on the tree leaves known so far. Its solution is then transformed from regional to element basis via Eq. (6) and "thresholded" to a hard labeling, i.e., taking the most certain label for each element. In the second step, leaves of the tree are subdivided if they touch a border of the hard labeling. The refined subdivision tree is then passed into the next alternation.

As the alternation elapses, the number of regions r_i increases, improving on the problem approximation. Due to the adaptivity of the subdivision, we keep the medium-scale property, i.e., $r_i \ll p$ for each QP (7 – 9) of the sequence. Keeping space requirements at its lowest, we can assemble $\mathbf{P}_i^\mathrm{T}\mathbf{H}\mathbf{P}_i$, $\mathbf{f}^\mathrm{T}\mathbf{P}_i$ and $\mathbf{A}\mathbf{P}_i$ directly without ever constructing the large-scale quantities $\mathbf{H}$, $\mathbf{f}$ and $\mathbf{A}$ explicitly.

We drew an experimental comparison of our scheme to the large-scale QP based on minimization of the Mumford-Shah functional (1). We used synthetic image data with strong noise as well as standard images from the Berkeley Segmentation Dataset [3], some examples of which are given in Figure 1. Experiments show high quality labelings widely independent of the initial tree level. Our scheme even introduces additional regularization, preventing the emergence of spurious segments. Depending on the complexity of the multi-label problem, our scheme outperformed the large-scale program by, at least, factors between two and seven. The gain in performance even allowed for near-interactive multi-label segmentation in our experiments on medium-sized problems.

[1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23: 1222–1239, 2001.

[2] O. Dalmau-Cedeno, M. Rivera, and P. P. Mayorga. Computing the alpha-channel with probabilistic segmentation for image colorization. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 1–7, 2007.

[3] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 416–423, 2001.

[4] M. Rivera, O. Dalmau-Cedeno, and J. Tago. Image segmentation by convex quadratic programming. In *Proc. Int. Conf. Pattern Recognit.*, pages 1–5, 2008.