

# FAST LABEL: Easy and Efficient Solution of Joint Multi-Label and Estimation Problems

Ganesh Sundaramoorthi  
KAUST, Saudi Arabia

ganesh.sundaramoorthi@kaust.edu.sa

Byung-Woo Hong  
Chung-Ang University, Seoul, Korea

hong@cau.ac.kr

## Abstract

*We derive an easy-to-implement and efficient algorithm for solving multi-label image partitioning problems in the form of the problem addressed by Region Competition [19]. These problems jointly determine a parameter for each of the regions in the partition. Given an estimate of the parameters, a fast approximate solution to the multi-label sub-problem is derived by a global update that uses smoothing and thresholding. The method is empirically validated to be robust to fine details of the image that plague local solutions. Further, in comparison to global methods for the multi-label problem, the method is more efficient and it is easy for a non-specialist to implement. We give sample Matlab code for the multi-label Chan-Vese problem in this paper. Experimental comparison to the state-of-the-art in multi-label solutions to Region Competition shows that our method achieves equal or better accuracy, with the main advantage being speed and ease of implementation.*

## 1. Introduction

Low level processing of images into multiple regions, i.e., image segmentation, is a common pre-processing task in a number of applications in computer vision. In these problems, it is required to partition the image domain, i.e., label the image, into multiple regions of homogeneous features represented by parameters. In many cases, the parameters that determine the partitioning of the domain are not known and must be estimated with the labels in a *joint* multi-label and parameter estimation problem. Region Competition [19] and multi-phase piecewise constant Mumford-Shah [18] are such problems, which this paper addresses. Note the algorithm introduced here does not apply to piecewise *smooth* Mumford-Shah (e.g., [10, 17]). While these problems are classical, they remain difficult to optimize so as to achieve a *desired* optimizer, and therefore, they have been the subject of much research.

Existing methods to solve joint multi-label and parame-

ter estimation problems can be divided into two categories. The first category is local methods, which use local label updates based on steepest descent (e.g., [19, 18]). They are beneficial when good initialization is available. The second category is global methods, which are based on global solutions to the multi-label sub-problem using relaxations (e.g., [6, 14]). These are useful in the case that one does not want to bias the solution to the initialization as much. Local methods are sensitive to fine details of the image, while global methods are less sensitive to fine details of the image. In applications where there are a large number of labels and/or images (e.g., multi-label co-segmentation [12] - simultaneously segmenting a group of images into multiple regions), it is beneficial to derive faster to solutions than existing techniques. Further, existing methods are relatively sophisticated for a non-specialist, relying on advanced optimization techniques, thus making implementation difficult.

In this work, we derive a new method for joint multi-label and parameter estimation problems that apply to problems solved by Region Competition. We call the method *Fast-Label*. Fast-Label uses global iterative *updates* rather than local updates of local methods, and thus they are less sensitive than local methods to fine irrelevant details of the image. The advantage over existing global methods is both computational *speed* and *ease of implementation*. Fast-Label can be coded efficiently in a few lines of Matlab code, making it easy for a non-specialist to code and modify. Although Fast-Label uses global updates to approximate a solution, we do not have theoretical guarantees of a global optimum. However, extensive experiments show there is no compromise in the accuracy of the partitioning when compared to global methods. In some cases, we even obtain higher segmentation accuracy. The advantage in speed makes Fast-Label feasible in applications with many images and labels, e.g., [12] (see also Section 4.2).

Our work is based on the simple idea that without regularization in the labels, the globally optimal labeling can be computed given the parameters in a simple thresholding step. Our method is then designed to approximate the effects of regularization by rewriting the regularization in a

form that resembles the joint problem without regularization. A simple additional thresholding step then approximates the effects of regularization. Our work relates to the work of [15, 8], where the main idea is to be able to approximate a curve evolution arising from gradient descent of an energy with simple operations of thresholding and smoothing. While some of our ideas are based on those works, we are not interested in approximating the effects of gradient descent flows, as those are susceptible to undesirable local minima. We are interested in using some of those ideas to derive global iterative updates given the parameters. This leads to robustness to fine structure, and a fast method.

## 2. General Multi-label Problem

We consider the following energy:

$$E(\{R_i, p_i\}_{i=1}^N) = \sum_{i=1}^N \int_{R_i} F(I(x); p_i) dx + \alpha \int_{R_i} W_{R_i}(x) dx \quad (1)$$

where the first term of the energy represents the data fidelity term (or label cost) and the second term is regularization. The  $R_i \subset \Omega$  are the regions to be determined and they are such that their interiors are mutually disjoint and their union forms the domain of the image  $\Omega$ .  $I : \Omega \rightarrow \mathbb{R}^k$  is the image. The parameters  $p_i$  can either be finite (e.g.,  $\mathbb{R}^k$ ) or infinite dimensional (probability distributions in the range space of the image). The goal of the multi-label problem is to optimize the energy in  $R_i$  (given  $N$ ), and the parameters  $p_i$  are unknown.

An example of a commonly used  $F$  is  $F(I(x); p_i) = |I(x) - p_i|^2$  where  $p_i \in \mathbb{R}^k$ . This is considered in [18] and a special case of [19] when the noise model of the intensity inside the regions are assumed Gaussian and the noise variance is the same in each  $R_i$ . Another example is probability distributions,  $p_i : \mathbb{R}^k \rightarrow [0, 1]$ , of an image feature within region  $R_i$ , as in Region Competition. In this case,  $F(I(x); p_i) = -\log p_i(I(x))$  where  $p_i(I(x))$  is the probability that  $I(x)$  fits distribution  $p_i$ .

We consider the following class of regularizers:

$$W_{R_i}(x) = \frac{1}{\sigma} \phi(x) (G_\sigma * \mathbf{1}_{R_i^c})(x), \quad (2)$$

where  $G_\sigma$  is the Gaussian kernel of standard deviation  $\sigma$ ,  $\mathbf{1}_{R_i^c}$  is the indicator function on  $R_i^c = \Omega \setminus R_i$ , and  $\phi : \Omega \rightarrow \mathbb{R}^+$ . To gain intuition for this regularization, consider when  $\phi = 1$ . For a point  $x \in R_i$ ,  $W_{R_i}(x)$  will be close to zero when all points in a disc, centered at  $x$ , of radius about  $\sigma$  are also inside  $R_i$ . It is close to 1 when most points around  $x$  are outside  $R_i$ . Therefore, the term enforces spatial regularity. The greater  $\sigma$ , the greater neighborhood size for enforcing

spatial regularity. In the limit when  $\sigma \rightarrow 0$ , one can show that

$$\int_{R_i} W_{R_i}(x) dx \rightarrow \frac{1}{2} \int_{\partial R_i} \phi(x) ds(x), \quad (3)$$

where  $s$  denotes the arc-length parameter. Details are in Appendix A. Hence in the case that  $\phi = 1$ ,

$$\sum_{i=1}^N \int_{R_i} W_{R_i}(x) dx \rightarrow \sum_{i \sim j} \text{Len}(\partial R_i \cap \partial R_j) + \frac{1}{2} \text{Len}(\partial \Omega), \quad (4)$$

where the summation on the right is over all adjacent pairs of regions, and  $\text{Len}(\cdot)$  denotes length. Thus, when  $\phi = 1$ , the regularization reduces to the standard length regularization, commonly used in segmentation algorithms. The case when  $\phi$  varies spatially implies penalizing length less at locations where  $\phi$  is small.

The use of the regularizer (2) is for two reasons. First, it leads to an efficient approximate optimization, as we will see in the next section. Second, it imposes regularity in larger neighborhoods that is commonly used in literature, which is beneficial for segmentation. Further, the approximation does not compromise on accuracy of the segmentation as we show. The length penalty is an arbitrary choice of regularization whose main purpose is to achieve spatial regularity, and spatial regularity can be imposed in a number of ways, including our choice of regularization.

**Remark 1** An example of  $\phi$  that is not uniform is  $\phi(x) = \exp(-\beta |\nabla I(x)|)$  where  $\nabla I(x)$  denotes the gradient of the image. This choice of  $\phi$  implies that regularization is applied less in regions where the gradient is large, i.e., edges.

Another way of inducing spatial regularity without the use of the regularization term  $W_{R_i}$  is to choose  $\alpha = 0$ , and then choose  $F$  to be spatially regular. For example, suppose that  $F_o$  is a not spatially regular, then smoothing  $F(I; p_i) = G_\sigma * F_o(I; p_i)$  provides spatially regularity. This implies a spatially regular solution for  $R_i$ , without explicit regularization term in the energy.

## 3. The Fast-Label Algorithm

### 3.1. Approximate Optimization

We now consider an approximate scheme to optimize the energy given in (1) that converges rapidly in practice, and gives a good segmentation. The algorithm follows an EM algorithm for optimization where we optimize in  $p_i$  given  $R_i$ , and then in  $R_i$  given  $p_i$ . Given the  $R_i$ , the parameters  $p_i$  can be usually solved in closed form by solving the equation

$$\frac{\partial}{\partial p_i} \int_{R_i} F(I(x); p_i) dx = 0. \quad (5)$$

For example, in the case that  $F(I(x); p_i) = |I(x) - p_i|^2$ , we find that  $p_i = 1/|R_i| \int_{R_i} I(x) dx$ . To optimize in  $R_i$ ,

we first gain intuition about our method from the case when  $\alpha = 0$ . In this case, given an estimate for  $p_i$ , it is easy to see that a global optimal for  $R_i$  (this is an exact solution not an approximation) is

$$R'_i = \left\{ x \in \Omega : i = \arg \min_j F(I(x); p_j) \right\}. \quad (6)$$

To give an approximation when  $\alpha \neq 0$ , we assume that we have an initial estimate  $R'_i$ ,  $i = 1, \dots, N$  of the regions that is close to the optimum of  $E$ , holding  $p_i$  fixed, e.g., (6). We then replace  $W_{R_i}$  with  $W_{R'_i}$  in  $E$ . Then a global optimizer of  $E$  with this choice of  $W$  is

$$R_i = \left\{ x \in \Omega : i = \arg \min_j F(I(x); p_j) + \alpha W_{R'_j}(x) \right\}. \quad (7)$$

Our optimization is then summarized as follows:

1. Choose an initialization for  $\{R_i\}_{i=1}^N$  (see Section 3.2 for practical choices).
2. Compute the optimal parameters  $p_i$  for the current regions  $R_i$  using (5).
3. Compute

$$R'_i = \left\{ x \in \Omega : i = \arg \min_j F(I(x); p_j) \right\}. \quad (8)$$

4. Compute  $W_{R'_i}$ :

$$W_{R'_i}(x) = \frac{1}{\sigma} \phi(x) (G_\sigma * \mathbf{1}_{R_i^c})(x). \quad (9)$$

5. Update  $R_i$ :

$$R_i = \left\{ x \in \Omega : i = \arg \min_j F(I(x); p_j) + \alpha W_{R'_j}(x) \right\}. \quad (10)$$

6. Repeat steps 2-5 until  $R_i$  does not change.

The computational complexity of our method is  $O(NM)$  where  $N$  is the number of labels and  $M$  is the number of pixels in the image. Gaussian convolution can be done efficiently (in linear time) using an approximation with a low order IIR filter [7], which we use in the experiments.

A Matlab implementation of the scheme above for the multi-label Chan-Vese model, i.e., Region competition [19] assuming Gaussian intensity statistics with means varying with regions, and standard deviation constant among regions, is in Figure 1. It shows the ease of Fast-Label. Fast-Label converges rapidly and achieves an accurate segmentation.

In the case that  $\alpha = 0$ , and spatial regularity is obtained by smoothing  $F$  (as discussed in the previous section), steps

```
function L=multilabel_cv(I,N,alpha,sigma,itters)
% Multi-label Chan-Vese Matlab code

p=min(I(:):(max(I(:))-min(I(:)))/(N-1):max(I(:)));
f=zeros([size(I),N]);
H=zeros([size(I),N]);

h=fspecial('gaussian',ceil(sigma*[5,5]),sigma);

for i=1:itters,
    for j=1:N,
        if i>1, p(j)=mean( I(find(L==j)) ); end
        f(:,j)=(I-p(j)).^2;
    end
    [mF,L]=min(f,[],3);

    for j=1:N,
        H(:,j)=imfilter(1-double(L==j),h);
    end
    [mF,L]=min(f+alpha*H,[],3);
end
```

Figure 1. **MatLab Code for Fast-Label** in the case of the multi-label problem solved by Region Competition [19] (Chan-Vese). Not only is the code easy to implement, but the method also converges quickly, and attains as good or better solution than any state-of-the-art method for Region Competition. Although at first glance, it appears that a memory requirement of  $2N$  times the size of the image is required, this can be avoided and only 4 times the size of the image is required (independent of  $N$ ). This can be achieved with a few additional lines of code.

4 and 5 are not needed,  $R'_i = R_i$ , and eq. (8) is replaced by the smooth  $F$ . This scheme is advantageous in that the parameter  $\alpha$  is eliminated. We tested this scheme, and the results are similar to the approach above, although the two approaches are not equivalent mathematically.

It should be noted that in the case that  $F = 0$ ,  $\phi = 1$ ,  $N = 2$ , and  $\sigma \rightarrow 0$ , the algorithm defaults to an algorithm for approximating mean curvature flow considered in [15]. Indeed, in [15], the characteristic function of an initial region  $R$ , i.e.,  $\mathbf{1}_R - \mathbf{1}_{R^c}$  is smoothed and thresholded, and it is shown that such a scheme applied successively approximates mean curvature flow. In [8], the idea is generalized to data driven energies so as to approximate gradient descent flows. The main difference in our scheme is to use a good initialization to the problem, i.e., eq. (8), and then add in the effects of the regularization in the second step eq. (10) to achieve an approximate global solution given the parameters, *not to approximate a local gradient descent*. Further, we are not interested in the case  $\sigma \rightarrow 0$ , rather we use that as our regularity parameter.

**Remark 2** *There are two practical issues that we discuss. First, as the image has finite size, boundary conditions must be specified for the Gaussian smoothing. In practice, the best choice compromising accuracy and speed is symmetric boundary conditions. Also, if  $\sigma$  is chosen extremely small, then it would be hard to represent the kernel discretely, as*

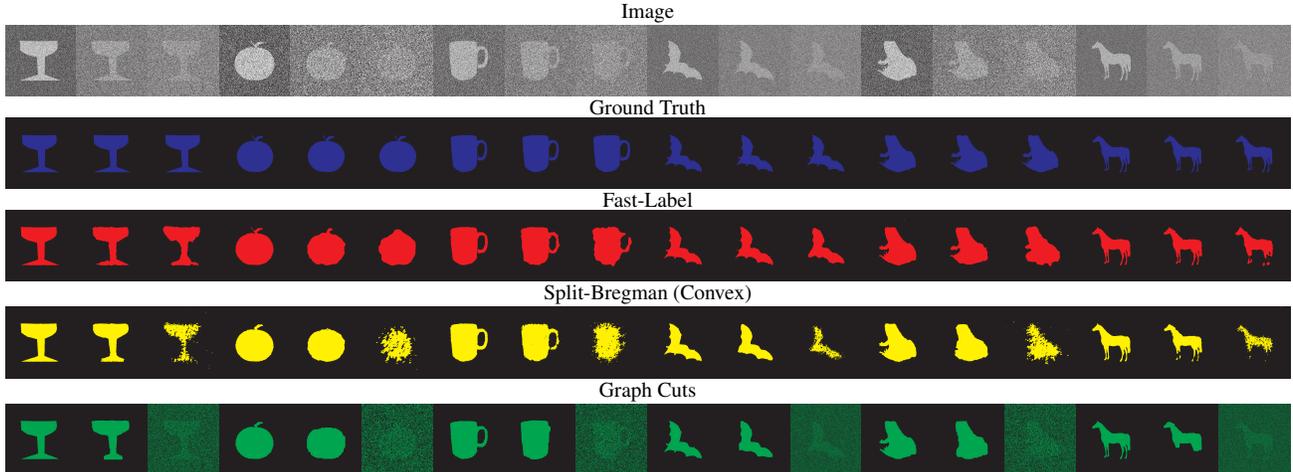


Figure 2. **Segmentation results on MPEG-7 shape image dataset.** In each block of  $5 \times 3$  images, [first row] shows the shape images with different Gaussian noise variance, in an increasing order, [second row] shows the ground truth of the shape, [third row] shows the segmentation result by our fast-label, [fourth row] shows the segmentation result by split-bregman, and [fifth row] shows the segmentation result by Graph Cuts.

numerical precision would limit the kernel to a delta function. This can be corrected with an up-sampling of the image; however, in experiments we seldom run into such a problem - accurate results are obtained at the native resolution of the image.

### 3.2. Initialization Method

One of the main issues facing all current methods of joint multi-label and parameter estimation problems, whether they be based on local updates or global updates, is initialization. While this issue is typically resolved differently depending on the particular application, we have explored two generic initializations, which seem to give promising results in the experiments that we have performed. One is a simple quantization of the image(s) into the number of pre-specified labels. The other is a random initialization: for each pixel, choose a random number between one and the number of desired labels to indicate the initial labeling. Both give accurate results. This is a highly desirable property as full automation may be possible.

## 4. Applications

### 4.1. Multi-label Segmentation

For a given image  $I : \Omega \rightarrow \mathbb{R}^k$  we assume that the image consists of disjoint regions  $R_i$ ,  $i = 1, 2, \dots, N$  and  $p_i$  is a constant, e.g., a piecewise constant model. A multi-label segmentation is obtained by minimizing the energy (1) with  $F(I(x); p_i) = |I(x) - p_i|^2$ ,  $p_i \in \mathbb{R}^k$  and  $\phi = 1$ .

### 4.2. Multi-label Co-segmentation

One of the benefits of our method is computational speed. We illustrate an application where speed is ben-

eficial, that is, multi-image, multi-label co-segmentation. This application may benefit from our efficient method. One of the latest algorithms for multi-label co-segmentation is presented in [12]. Suppose that we have  $K$  images  $I_j$ ,  $j = 1, \dots, K$ , the objective of co-segmentation is to identify regions that have similar properties among different images. These regions are also required to partition each image into regions of homogeneous features. We assume that there are at most  $N$  distinct regions in each image. We denote by  $R_i^j$  the  $i^{\text{th}}$  region in image  $j$ . It is assumed that  $R_i^j$ ,  $j = 1, \dots, K$  have a common probability distribution  $p_i$  of features. Thus,  $R_i^j$ ,  $j = 1, \dots, K$  are corresponding regions across the images in the collection. The goal is to partition each of the images based on the (unknown) distributions  $p_i$ . The MAP estimate leads to the following energy (to be optimized in the regions and distributions):

$$\sum_{j=1}^K \sum_{i=1}^N \int_{R_i^j} -\log p_i(I_j(x)) + \alpha W_{R_i^j}(x) dx. \quad (11)$$

For fixed  $R_i^j$ , the optimal estimate of the distributions is

$$p_i(r) = \frac{w_i(r)}{\int_{\mathbb{R}} w_i(r') dr'}, \quad w_i(r) = \sum_{j=1}^K \int_{\{\tilde{I}_j=r\} \cap R_i^j} \frac{ds(x)}{|\nabla I_j(x)|}, \quad (12)$$

where  $ds$  denotes the arc-length parameter. The estimate of the regions are then

$$\tilde{R}_i^j = \left\{ x \in \Omega : i = \arg \max_k p_k(\tilde{I}_j(x)) \right\} \quad (13)$$

$$R_i^j = \left\{ x \in \Omega : i = \arg \max_k p_k(\tilde{I}_j(x)) e^{-\alpha W_{R_k^j}(x)} \right\}. \quad (14)$$

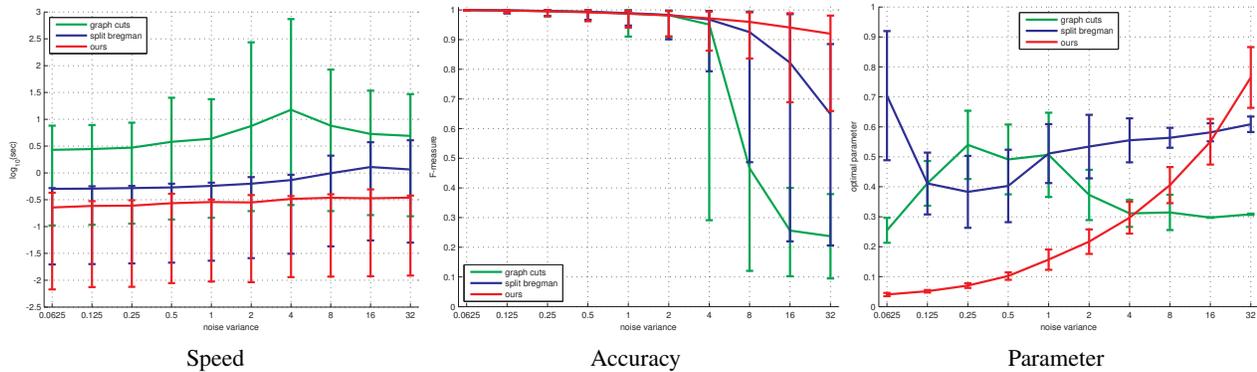


Figure 3. **Comparison of State-of-the-Art Methods for Bi-Label Segmentation.** Testing is on shape images with varying noise from the MPEG-7 dataset. [Left] shows the time (in log scale) versus the noise variance indicating the proposed fast-label is about 2-3 times faster than the Split-Bregman (SB) method [9], and about 10-30 times faster than Graph Cuts (GC) [6]. [Middle] shows the accuracy in terms of F-measure of all methods versus the noise variance. [Right] shows the optimal weight on regularization for each of the methods normalized between 0 and 1 versus the noise variance. It can be seen that proposed method has a clear relationship between regularization and noise level, while the others do not.

Equations (12), (13), and (14) are iterated until convergence to determine the co-segmentation. Coarse binning of  $w_i$  provides robustness to illumination change in images.

## 5. Experiments

Code for Fast Label is publicly available at the website [http://vision.ucla.edu/~ganeshs/fast\\_label/](http://vision.ucla.edu/~ganeshs/fast_label/).

In the first experiment, we test Fast-Label in the case of bi-label segmentation and estimation where  $p_1, p_2 \in \mathbb{R}$  and  $F(I(x); p_i) = (I(x) - p_i)^2$ ,  $i = 1, 2$ , which is the model considered in [4]. The model assumes that the image consists of two regions with approximate intensity  $p_1, p_2$  and a Gaussian noise model. In this experiment, we use the MPEG-7 shape image dataset and select 60 different shapes. Ten different levels of Gaussian noise are added to each of the images forming 600 images. We compare Fast-Label in speed and accuracy to Split-Bregman (SB) [9], which is a fast method for implementing the convex relaxation of [4] considered in [3], and Graph Cuts (GC) [2] without a penalty term for the label cost. Since all methods require a parameter for regularization, the optimal parameter is selected with respect to the accuracy. We report the average run time, accuracy in terms of F-measure to ground truth, and optimal parameter versus the noise levels. All methods are given the same initialization: one bit quantization of the image. Selected segmentation results are shown in Figure 2. Fast-Label is more robust with respect to noise, preserving details of the shapes, while the numerics of the other methods break down with larger noises. Figure 3 shows the comparison of the methods in terms of speed (left) and accuracy (middle). Fast-Label is 2-3 times faster than SB and 10-30 times faster than GC. Our method is as accurate as SB and GC up to intermediate noise level (2). At larger noise, the numerics of SB and GC break down, resulting in highly

inaccurate results, while our method remains accurate. Figure 3 (right) shows the optimal parameter versus noise level for all the methods (we have normalized the parameters between 0 and 1 for easy comparison). There is a clear trend between noise level and our parameter; this is not the case for SB and GC.

In the second experiment, we test Fast-Label on Brain-Web [5], a common dataset for evaluation of segmentation of brain MRI. The task is to label each pixel in an image with one of 12 labels. In the evaluation, we consider the most prominent regions, gray and white matter, as regions of interest. From the dataset, we select four different cases of 3D brain MRI volumes with various amounts of additive noise (0.1% to 12%), and segment the images slice-wise. We compare Fast-Label in speed and accuracy to multi-label segmentation using  $\alpha$ -expansion Graph Cuts (GC) [6], convex multi-label relaxation (CMR) [16], and multiphase level sets (MPLS) [18] (implemented using a fast level set narrowband method). For fair evaluation, the optimal values for the regularity parameter and the number of labels are selected based on the average F-measure w.r.t ground truth of gray and white matter. For all methods, we assume that the image consists of  $N$  regions with approximate intensity  $p_i$  in region  $R_i$ , and a Gaussian noise model. The initialization for all methods is a quantization of the image into  $N$ -labels. Selected segmentation results are shown in Fig. 4. Fast-Label is more robust to noise than the other methods, which break down at large noise levels. Quantitative comparison of Fast-Label to the others is presented in Fig. 5. Fast-Label is about 10-100 times faster than GC, 10-1000 times faster than CMR and 100-1000 faster than MPLS. In terms of the accuracy, our method is as good as GC up to 0.4% noise variance, but our method becomes more accurate than GC at larger noises, and is more accurate than

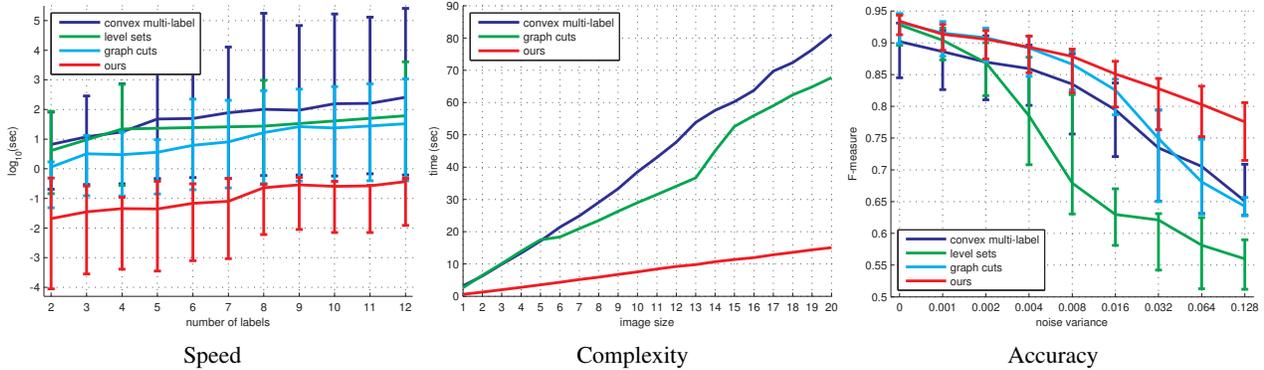


Figure 5. **Comparison of multi-label segmentation methods on the BrainWeb dataset.** [Left] shows the time (in log scale) versus the number of labels. The proposed fast-label method is about 10-100 times faster than graph cuts [6], 10-1000 times faster than convex multi-label [17] and 100-1000 times faster than multiphase level sets [18]. [Middle] shows the time versus the image size ( $1\times$ ,  $2\times$ , ...,  $20\times 256\times 256$ ), while the proposed fast-label method is linear in the image size, and graph cuts and convex multi-label are quadratic to cubic. [Right] shows the optimal F-measure versus noise level.

CMR at all noise levels. MPLS is shown to be the least accurate since it is easily trapped in local minima. The middle plot in Fig. 5 shows the grow rate of the proposed method, GC and CMR versus varying image sizes with five labels. Note that the complexity of GC and CMR in the number of pixels is between  $O(N^2)$  and  $O(N^3)$  whereas for Fast-Label, it is  $O(N)$ , and this is verified experimentally.

In the third experiment, we apply Fast-Label to multi-label co-segmentation on the iCoseg dataset [1]. In iCoseg, each object class consists of different numbers of images and each image may have multiple objects of interest. Our method uses a color histogram as a feature like other methods. Since there is no prior information regarding the number of distinct objects within each image, it is necessary to find the optimal number of labels for each class. It is also necessary to choose an appropriate regularization, which is related to the distinctiveness of object feature which may vary across images within the same class. Thus, the optimal parameters for each object class need to be determined. The performance of the algorithm is evaluated based on the score given by  $\frac{1}{K} \sum_{j=1}^K \max_i \frac{|T^j \cap R_i^j|}{|T^j \cup R_i^j|}$  where  $T^j$  denotes the ground truth in image  $j$  and  $R_i^j$  denotes region with label  $i$  in image  $j$ . Illustrative results of our co-segmentation algorithm are presented in Figure 6. Within each  $3\times 3$  block of images, the top row shows three examples of the same class images, the middle row shows the ground truth for the class object, and the bottom row shows our partitioning labels. Table 1 shows quantitative comparison to other state-of-the-art methods. Results indicate that Fast-Label is, on average, 9.14 times better than second-best method. In addition to the improved accuracy, Fast-Label is computationally more efficient compared to the co-segmentation algorithm presented in [12]. The computational time reported in [12] varies from 30 mins to one hour for 30 images. Our

method takes 143 secs on average per parameter for 33 images in MatLab.

## 6. Conclusion

We have constructed an algorithm, called *Fast Label*, for the problem solved by Region Competition, i.e., a class of joint multi-label and parameter estimation problems. It is based on global updates of labels, and thus robust to local image structure that pose difficult for local methods. The advantage of Fast-Label over other global methods is both speed and simplicity of our implementation, while not compromising accuracy. Extensive experiments on multiple applications showed these advantages of speed and accuracy compared to state-of-the-art algorithms.

## Acknowledgements

GS was supported by KAUST baseline, CRF, and Visual Computing Center funding. BH was partly supported by NRF-2010-220-D00078 and NRF-2011-0007898 in Korea.

## A. Analysis of Regularization

We now examine the regularization term with the choice of (2) as  $\sigma \rightarrow 0$ . We assume that  $\partial R_i$  is a rectifiable set (that is piecewise differentiable). Note that as  $\sigma \rightarrow 0$ , for points  $x \in \partial R_i$  (where  $\partial R_i$  is assumed to be a differentiable curve) it is true that  $G_\sigma * \mathbf{1}_{R_i^c}(x) \rightarrow 1/2$  almost everywhere (in the measure theoretic sense). This can be seen as almost everywhere,  $\partial R_i$  can be approximated with its tangent line around the point  $x \in \partial R_i^c$ , and thus a small disc around the point  $x$  is bisected by the tangent line of  $\partial R_i$  at  $x$ . Therefore, half the weight of  $G_\sigma$  centered at  $x$  is on one side of the disc (where  $\mathbf{1}_{R_i^c} = 1$ ), and the other half of the weight of  $G_\sigma$  is on the other side of the disc (where  $\mathbf{1}_{R_i^c} = 0$ ). Hence, it is clear that  $G_\sigma * \mathbf{1}_{R_i^c}(x) \rightarrow 1/2$ .

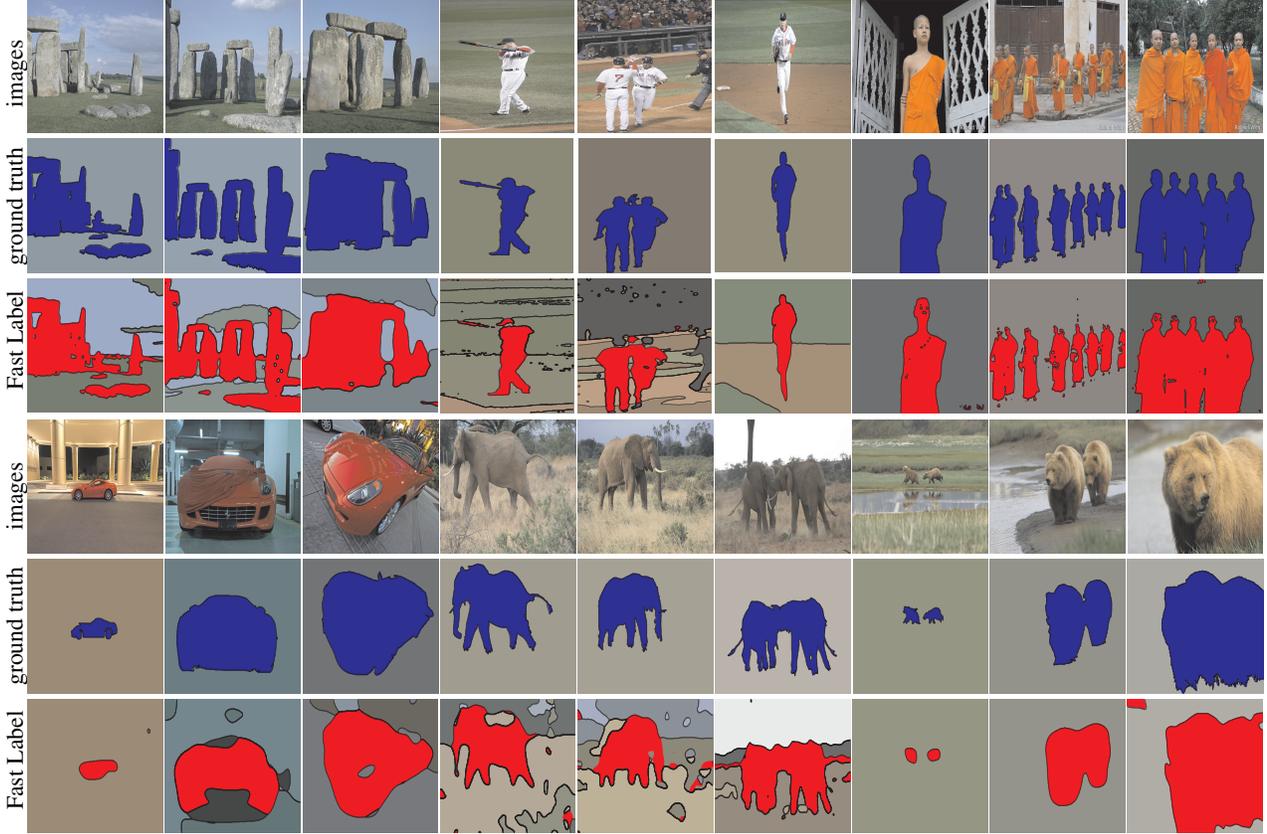


Figure 6. **Example results of Multi-Label Co-segmentation on iCoseg Dataset** using a color histogram feature. In each block of  $3 \times 3$  images, [top row] shows three examples within class images, [middle row] shows the ground truth, and [bottom row] shows our fast-label co-segmentation result. The quantitative evaluation of the results is presented in Table 1.

class	images	labels	ours	Joulin et al. [12]	<sup>†</sup> Joulin et al. [11]	Kim et al. [13]	<sup>‡</sup> Joulin et al. [11]
baseball player	25	5	<b>74.8</b>	62.2	53.5	51.1	24.9
brown bear	5	2	<b>82.1</b>	75.6	<b>78.5</b>	40.4	28.8
elephant	15	6	<b>65.0</b>	<b>65.5</b>	51.2	43.5	23.8
ferrari	11	4	<b>73.0</b>	<b>65.2</b>	63.2	60.5	48.8
football player	33	4	<b>64.3</b>	<b>51.1</b>	38.8	38.3	20.8
kite panda	7	3	<b>81.4</b>	57.8	58.0	<b>66.2</b>	58.0
monk	17	2	<b>85.0</b>	<b>77.6</b>	76.9	71.3	76.9
panda	21	2	<b>64.9</b>	<b>55.9</b>	49.1	39.4	43.5
skating	11	3	<b>82.7</b>	<b>64.0</b>	47.2	51.1	47.2
stonehenge	18	3	<b>90.7</b>	<b>86.3</b>	85.4	64.4	62.3

Table 1. **Co-segmentation results of Fast-Label in Comparison to State-of-the-Art** on iCoseg dataset. For each object class, the score is computed by the average intersection over union with respect to the ground truth with the optimal number of labels obtained. The evaluation results of all other methods that are compared to our method are reported in [12]. <sup>†</sup> multi-label algorithm. <sup>‡</sup> two-label algorithm.

Next, we see that for  $x \in \text{int}(R_i)$  (interior of  $R_i$ ), that  $G_\sigma * \mathbf{1}_{R_i^c}(x) \rightarrow 0$ , and more precisely, we have that

$$G_\sigma * \mathbf{1}_{R_i^c}(x) \approx \begin{cases} 1/2 & \text{dist}_{\partial R_i}(x) \leq \sigma, x \in R_i \\ 0 & \text{dist}_{\partial R_i}(x) > \sigma, x \in R_i \end{cases}, \quad (15)$$

where  $\text{dist}_{\partial R_i}(x)$  indicates the minimum Euclidean distance from  $x$  to a point on  $\partial R_i$ . Therefore the regularization

term becomes

$$\begin{aligned} \int_{R_i} W_{R_i}(x) dx &\rightarrow \frac{1}{\sigma} \int_{R_i \cap \{\text{dist}_{\partial R_i}(x) \leq \sigma\}} \frac{1}{2} \phi(x) dx = \\ &\frac{1}{2\sigma} \int_0^\sigma \int_{R_i \cap \{\text{dist}_{\partial R_i}(x) = r\}} \phi(x) ds(x) dr \rightarrow \\ &\frac{1}{2} \int_{\partial R_i} \phi(x) ds(x), \quad (16) \end{aligned}$$

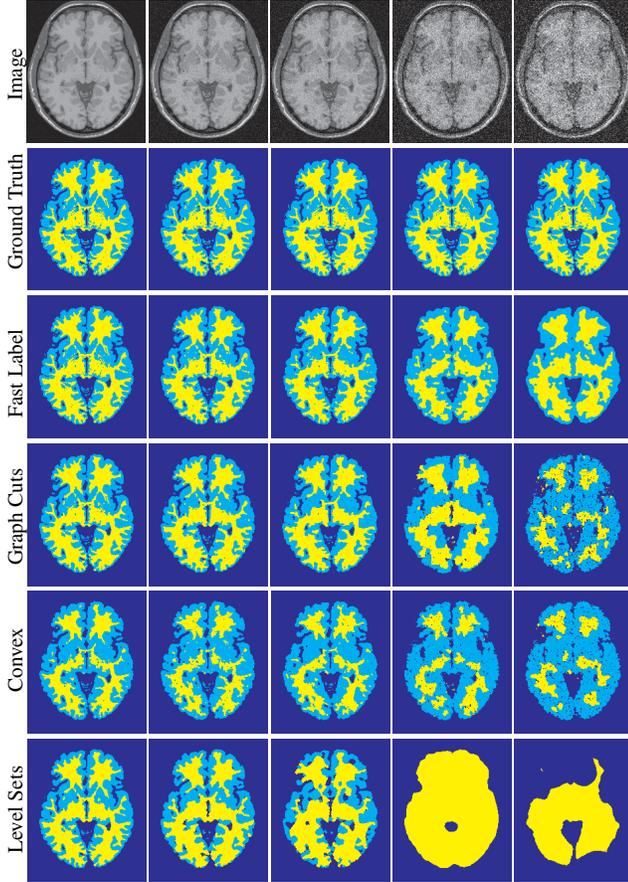


Figure 4. **Multi-Label Segmentation Results on Brain MRI.** Comparison of Fast-Label to state-of-the-art methods. The results are optimal in terms of F-measure with respect to the gray and white matter. [First row] brain MRI image with increasing noise in order. [Second row] ground truth. [Third row] our fast-label method. [Fourth row] graph cuts. [Fifth row] convex multi-label relaxation. [Sixth row] multiphase level sets.

where  $ds$  indicates the arc-length element. The equality from line 1 to line 2 is due to the Co-area formula, and the last line is due to the Lebesgue differentiation formula. Therefore, we see that as  $\sigma \rightarrow 0$  and  $\phi = 1$  that

$$\sum_{i=1}^N \int_{R_i} W_{R_i}(x) dx \rightarrow \sum_{i \sim j} \text{Len}(\partial R_i \cap \partial R_j) + \frac{1}{2} \text{Len}(\partial \Omega) \quad (17)$$

where  $\text{Len}(\cdot)$  denotes length, and the notation  $i \sim j$  denotes that  $R_i$  is adjacent to  $R_j$ .

## References

- [1] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *IEEE Conf. on CVPR*, 2010. 6
- [2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI, IEEE Transactions on*, 26(9), 2004. 5
- [3] X. Bresson, S. Esedoglu, P. Vanderheynt, J. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2), 2007. 5
- [4] T. Chan and L. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2), 2001. 5
- [5] C. Cocosco, V. Kollokian, R. Kwan, G. Pike, and A. Evans. Brainweb: Online interface to a 3d mri simulated brain database. In *NeuroImage*, volume 5, 1997. 5
- [6] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *Proc. IEEE Conf. on CVPR*, 2010. 1, 5, 6
- [7] R. Deriche. Recursively implementing the gaussian and its derivatives. Technical Report RR-1893, INRIA, 1993. 3
- [8] S. Esedoglu and Y.-H. R. Tsai. Threshold dynamics for the piecewise constant mumfordshah functional. *Journal of Computational Physics*, 211(1):367–384, 2006. 2, 3
- [9] T. Goldstein, X. Bresson, and S. Osher. Geometric applications of the split bregman method: Segmentation and surface reconstruction. *J. of Scientific Computing*, 45(1), 2010. 5
- [10] B.-W. Hong, Z. Lu, and G. Sundaramoorthi. A new model and simple algorithms for multi-label mumford-shah problems. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1219–1226. IEEE, 2013. 1
- [11] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *CVPR*, 2010. 7
- [12] A. Joulin, F. Bach, and J. Ponce. Multi-class cosegmentation. In *IEEE Conf. on CVPR*, 2012. 1, 4, 6, 7
- [13] G. Kim, E. Xing, L. Fei-Fei, and T. Kanade. Distributed cosegmentation vis submodular optimization on anisotropic diffusion. In *Conf. on ICCV*, 2011. 7
- [14] J. Lellmann and C. Schnorr. Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.*, 4(4):1049–1096, 2011. 1
- [15] B. Merriman, J. Bence, and S. Osher. Diffusion generated motion by mean curvature. *UCLA CAM Report*, 1992. 2, 3
- [16] T. Pock, A. Chambolle, D. Cremers, and H. Bischof. A convex relaxation approach for computing minimal partitions. In *IEEE Conference on CVPR*, 2009. 5
- [17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the mumford-shah functional. In *IEEE Conference on ICCV*, 2009. 1, 6
- [18] L. Vese and T. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *IJCV*, 50(3), 2002. 1, 2, 5, 6
- [19] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *PAMT, IEEE Trans. on*, 18(9), 1996. 1, 2, 3