Statistical Significance Based Graph Cut Regularization for Medical Image Segmentation

Abstract

Graph cut minimization formulates the image segmentation as a linear combination of problem constraints. The salient constraints of the computer vision problems are data and smoothness which are combined through a regularization parameter. The main task of the regularization parameter is to determine the weight of the smoothness constraint on the graph energy. However, the difference in functional forms of the constraints forces the regularization weight to balance the inharmonious relationship between the constraints. This paper proposes a new idea: bringing the data and smoothness terms on the common base decreases the difference between the constraint functions. Therefore the regularization weight regularizes the relationship between the constraints more effectively. Bringing the constraints on the common base is carried through the statistical significance measurement. We measure the statistical significance of each term by evaluating the terms according to the other graph terms. Evaluating each term on its own distribution and expressing the cost by the same measurement unit decrease the scale and distribution differences between the constraints and bring the constraint terms on similar base. Therefore, the tradeoff between the terms would be properly regularized. Naturally, the minimization algorithm produces better segmentation results. We demonstrated the effectiveness of the proposed approach on medical images. Experimental results revealed that the proposed idea regularizes the energy terms more effectively and improves the segmentation results significantly.

Keywords: Medical Image Segmentation; Statistical Significance; Graph Cut Minimization; Markov Random Fields; Regularization.

1 Introduction

Medical image segmentation is the process of labeling pixels or voxels in a medical image data set with a class label. One of the purposes of segmentation of medical data is obtaining the anatomical structure of organs such as kidney, liver or their parts of interest such as tumorous sections. The segmentation knowledge assists medical experts in recognizing abnormalities, making diagnosis for the body parts of interest. It is also useful to guide the surgeon. By all means, a precise segmentation is important for both diagnosis and surgery in terms of reliability. There are many different medical image segmentation techniques such as active contours [1], level sets [2], and model based approaches [3]. Markov Random Field (MRF) based segmentation methods are one of the most successful techniques for the medical applications [4, 5, 6]. MRF formulates the segmentation problem constraints as potential functions. The linear sum of the potential functions through a regularization parameter forms the energy formulation. The energy formulation can be minimized by the minimization algorithms such as graph cuts, belief propagation, etc. Researchers still work on developing MRF based algorithms for the medical image segmentation [7, 8].

The graph cut approach is one of the successful MRF applications for the vision problems. The approach forms a graph structure with an objective energy function, and then minimizes the function to solve the problem. The graph structure is constituted as nodes and edges. Each edge is assigned a cost according to the relationship between the connected nodes. All the costs on the graph are joined linearly through a regularization parameter to form the objective energy function. The costs of the edges on the graph are determined through potential functions. However, in the same energy formulation, the potential function of each constraint can have different functional forms. The difference in functional forms between the constraints may cause scale and distribution differences between the energy terms. The regularization parameter attempts to tune the difference between the potential functions, however its main function should be determining the weight of the smoothness constraint. We claim that, if the scale and distribution difference between the potential functions are decreased, balancing the trade-off between the energy terms would be more effective, which results in a better restricted solution space and produces a more meaningful solution. In this paper, we introduce a new empirical method based on statistical significance theory. The method measures the statistical significance of each potential function. Since different functional forms are expressed on the common measurement type, the scale and distribution differences between the data and smoothness terms are decreased. As a result, the regularization weight properly balances the relationship of the harmonic energy terms.

We test the effectiveness of the proposed approach on the medical image segmentation problem which is usually harder than the general segmentation, because many organs are in contact with each other and tissues have similar intensity variations. The interactive approach partially overcomes the hardness of medical image segmentation by initially marking some pixels as object and background. The marked pixels help the minimization algorithm by providing hints about the intention of the user. Joly and Boykov [5] proposed a successful graph cut application of the interactive graph cut. In that application, the potential function of data constraint is based on the marked pixels histogram, whereas, the potential function of the smoothness constraint is the intensity difference between the neighboring pixels. Note that the data and smoothness constraints have different functional forms. The different functional forms of the data and the smoothness terms of the interactive graph cut method make it a very good candidate for our method of decreasing the differences between functional forms. In addition, any improvement of the popular interactive graph cuts by our method would indicate the applicability of the proposed idea. Note however that the proposed method is not limited to the interactive graph cuts. It can be applied for any MRF based optimization systems that include functional forms with different scales and distributions.

The rest of the paper is organized as follows: In Section 2, we give a brief background about the markov random fields, regularization weight, graph cut minimization, and statistical significance theory. We introduce the method on the background/foreground graph cut segmentation in Section 3. Section 4 includes the experimental results on medical images and Section 5 provides concluding remarks.

2 Necessary Background

2.1 Ill posed problems and Prior Constraints

Computational vision includes a set of problems which attempt to find physical properties of the 3D world from 2D images. However, 2D images contain limited information about the 3D world. This information loss makes the vision problems ill-posed which do not satisfy one or more of the following well-posed requirements: existence, uniqueness, and stability of the solution [9]. Using only the data knowledge is not adequate to solve the ill-posed problems. There should be prior assumptions to decrease the complexity of the solution space. The most popular prior assumption in computational vision is the smoothness constraint which assumes that the physical properties of the neighborhood pixels generally do not change abruptly [10, 11, 12]. The regularization theory incorporates the smoothness constraint to the data constraint in order to find a stable desired solution.

2.1.1 Markov Random Fields

Most computer vision problems can easily be defined and solved in a discrete framework. However the classical regularization approach [11, 21] cannot be defined over discrete variables. The counterpart to the regularization is the MRF modeling in discrete framework. One of the discrete problems in computer vision is image labeling which is assigning a set of labels to image pixels or features. Image segmentation can be posed as a labeling procedure. Let $d = \{d_i\}_{i \in S}$ be the observed data from an input image. S is the set of sites i and represents an image pixel or feature. The corresponding labels at the image sites can be given by $f = \{f_i\}_{i \in S}$. Labeling procedure assigns the label f_i to the site i for the observed data d_i.

In MRF framework, the labeling problem is formulated using bayesian statistics as in Eq.1. Maximum a posteriori (MAP) is the most common solution for MRF. In MAP criteria, the true labeling configuration f^* is estimated by

$$f^* = \arg\max_f P(f|d) = \arg\max_f p(d|f)P(f). \tag{1}$$

Maximum a posteriori (MAP) solution of the problem says that the more probable configuration f^* minimizes the posterior energy.

$$f^* = \arg\min_f E(f,d) = \arg\min_f (E_d(f,d) + E_s(f)), \tag{2}$$

where $E_d(f,d)$ is the data energy and denotes the data constraints of the segmentation problem; $E_s(f)$ is the prior energy and denotes the smoothness constraint. Formulated posterior energy can be minimized in different ways such as iterated conditional modes [13], stochastic gradient descent [14, 15], and simulated annealing [16, 17]. We used the graph cut approach [4, 18] (section 2.3) as MRF solver to introduce the proposed idea.

According to the Hammersley-Clifford Theorem [19], posterior energy can be written as a sum of unary and pairwise potentials.

$$E(f,d) = E_d(f,d) + \lambda E_s(f) = \sum_{i \in S} U(f_i, d_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} V(f_i, f_j), \tag{3}$$

where U is unary potential and measures the similarity between data d_i and label f_i; V is pairwise potential between adjacent sites and encourages the neighboring sites f_i, f_j to have similar labels; N_i denotes the neighbors of site i; λ is the regularization weight which balances the trade-off between data and smoothness energy terms. The unary potential function constitutes the data constraint of the energy formulations, similarly, pairwise potential function constitutes the smoothness constraint. Different potential functions U and V were employed by researchers such as absolute differences, squared differences, truncated absolute differences [20], probability histograms [5], and many other robust techniques. Different functional forms can be used for each potential function in the same energy function.

2.2 Regularization Weight in Computer Vision

The minimization of the energy formulation in Eq.3 gives the most probable labeling configuration f^*. It is desired to minimize both data and smoothness constraints; however there is often a trade-off between them. The regularization weight lambda balances the relationship between the closeness of the data to the solution and its smoothness. Choosing an optimal regularization parameter is important to construct a meaningful solution. If λ is small, the minimization solution will be noisy. On the other hand, if λ is large, the labeling configuration will not fit the observed data. Although the main function of λ is to tune the tradeoff between the smoothness and the data, it generally also serves as the main mechanism for adjusting the scale and distribution differences between the constraints. As a result, the optimal λ selection process becomes more complicated. The main motivation for our method is to decrease the scale and distribution differences between the constraints so that λ can only be used for tuning the tradeoff between the smoothness and the data.

In recent years, it is realized that regularization parameter depends on scene structure variations and image noise statistics [22, 23]. Therefore, different image sets need different regularization parameters for the optimal performance. Only a few papers focused on the optimal choice of regularization weight from the observed image. Zhang and Seitz [23] proposed a probabilistic mixture model for the regularization weight determination of a stereo pair. Peng and Veksler [24], on the other hand, work on the regularization weight for segmentation problem. The ultimate goal of these techniques is to estimate the optimal regularization parameter which produces optimal solution for the problem. We also aim to get the optimal solution but our method is not a parameter selection method. The main contribution of the proposed method lies in increasing the tuning performance of the regularization parameter by bringing the constraints on the common base.

2.3 Graph Cut Minimization for the Segmentation Problem

Graph cut algorithm [25] is one of the most popular energy minimization approaches which solves the object-background labeling problem successfully [4, 26, 27]. The algorithm begins by building a graph $G = (E, V)$. V consist of set of nodes that correspond to labeling sites and two extra nodes which denote object and background terminals. E are the edges that connect the nodes. Each graph edge is given a nonnegative cost using potential functions. The optimal labeling configuration is determined by finding the minimum cost cut on this graph by minimizing the graph energy functional.

$$E(f, d) = \sum_{i \in S} U(f_i, d_i) + \lambda \sum_{i \in S} \sum_{j \in N_i} V(f_i, f_j). \quad (4)$$

The first sum in the energy functional is called the data term which confines the labeling configuration to be close to observed data. The second sum is the smoothness term and confines the neighboring sites to have similar labels. For the segmentation problem, data costs can be modeled by potential functions including squared differences, absolute differences or truncated absolute differences between the pixel nodes and the terminal nodes. Smoothness cost on the other hand, can be modeled by using local intensity gradient, laplacian zero crossing or gradient direction [20].

The interactive graph cut [5] is one of the convenient and widely used graph cut approaches for the medical applications. The user at the beginning selects some pixels as object and background to direct the algorithm what is intended to segment.

The data terms of the energy are determined with the help of the marked pixels. The potential functions for the data terms are

$$U_{obj}(f_i, d_i) = -\ln Pr(I_i|\text{“}obj\text{”}) \tag{5}$$

and

$$U_{bcg}(f_i, d_i) = -\ln Pr(I_i|\text{“}bcg\text{”}), \tag{6}$$

where I_i is intensity at site i; “obj” denotes the intensity distribution of the pixels marked as object; “bcg” denotes the intensity distribution of the pixels marked as background. $Pr(I_i|\text{“}obj\text{”})$ and $Pr(I_i|\text{“}bcg\text{”})$ are the probabilities of intensity I_i which are the number of occurrences of intensity I_i in the object and the background intensity distributions respectively. The potential function for the smoothness cost is

$$V(f_i, f_j) = \exp(-\frac{(I_i - I_j)^2}{2\sigma^2}).\frac{1}{dist(i,j)}, \tag{7}$$

where I_i and I_j are intensities at sites i and j; $dist(i, j)$ is the Euclidean distance between two sites; σ is the system parameter. Note that data and smoothness terms have different functional forms: data energy terms are determined based on probability histograms derived from marked pixels, on the other hand, smoothness energy terms are the intensity difference between neighboring sites. We applied the proposed idea for the interactive graph cut approach to show the effects of decreasing the differences in functional forms of the energy constraints.

2.4 Statistical Significance

Statistical significance is a probability value (p-value) which is the measurement of randomness [28]. If the observed outcome of an experiment is statistically significant, this means that “probably” this outcome cannot happen randomly. To measure the statistical significance of the outcome of an experiment, probability and cumulative probability distribution functions of the experiment should be known. If the distribution of the outcome is not a known distribution such as exponential, a non-parametric technique can be used to form the distributions [29]. The idea behind the nonparametric techniques is that observed outcome represents the distribution function from which it was drawn. Therefore, we use the possible outputs of the experiment to form the probability distributions. The area under the probability distribution function forms the cumulative distribution function. After distribution functions are constituted, the p-value of the observed outcome is calculated. The location of the outcome on cumulative distribution determines its p-value. A p-value in the tail shows that the observed outcome would rarely occur by chance. A p-value in the middle however shows that the outcome probably occurred by chance.

Statistical significance theory has found minimal usage in computer vision. We previously used the statistical significance to combine the similarity measures for the stereo correspondence problem [30]. Each measure is expressed as p-values and compared with each other. Peng and Veksler [24], on the other hand, used the statistical significance method to normalize the segmentation features and to expressed them on the same base. We employ the statistical significance as the potential function measurement units for Eq.4, which has several advantages. First, the statistical significance based potential functions would measure the randomness of the label assignments of sites. If the measured statistical significance p-value is small, then the label assignment of a site would rarely occur by chance. On the other hand, if the the measured statistical significance p-value is large, then the label assignment of the site probably occurred by chance. As a result, using the statistical p-values as the potential function measurement units is more intuitive and meaningful. Second, since each unary and pairwise potential function of MRF framework can be expressed in statistical significance measurement units, linearly summing them, as in Eq.4, would be more meaningful. In addition, the regularization weight lambda would not get influenced from the scale and distribution differences of the potential functions. The main novelty of this paper is the new perspective of tuning the scale and distribution difference between potential functions of data and smoothness constraints by evaluating the constraints in terms of statistical significance.

3 The Proposed Method - Graph Cut Minimization Modified by Statistical Significance

Graph cut minimization is also a MRF based approach, and its data and smoothness constraints have different functional forms in the same energy formula. We introduce a new empirical method that aims to decrease the measurement difference between the constraints by expressing them using the same measurement type based on statistical significance.

3.1 Measuring p-values of Energy Terms

In order to measure the statistical significance of the data and smoothness constraints, cumulative probability distribution (cdf) of the terms should be known. The outputs of the potential functions of the observed image form the probability distributions of the terms. Since the distributions depend on the observed image, the distribution structures are not known beforehand. Therefore, we follow a non-parametric approach to form the cdf's. Data costs of a given graph form the probability distribution of data constraint. Similarly, the smoothness costs on the graph form the smoothness probability distribution. The data probability distribution function can be formalized as

$$Pr(U(f,d)) = \{U_{obj}(f_i, d_i), U_{bcg}(f_i, d_i) : \forall i \in S\} \tag{8}$$

where $U_{obj}(f_i, d_i)$ and $U_{bcg}(f_i, d_i)$ denote the data terms obtained from the unary potential function(Eq.5-6). The smoothness probability distribution function is

$$Pr(V(f)) = \{V(f_i, f_j) : \forall i \in S, \forall j \in N_i\} \tag{9}$$

where $V(f_i, f_j)$ denotes the smoothness terms obtained by the pairwise potential function(Eq.7). The area under the probability distributions produces the cdf's of data and smoothness constraints.

$$F(U(f,d)) = \sum_{-\infty}^{U(f,d)} Pr(U(f,d)). \tag{10}$$

$$F(V(f)) = \sum_{-\infty}^{V(f)} Pr(V(f)). \tag{11}$$

Each term is converted to p-value using the cdf's. The location of the terms on cumulative distributions determines their p-values.

$$pU(f_i, d_i) = F(U(f,d) < U(f_i, d_i)) \tag{12}$$

and

$$pV(f_i, f_j) = F(V(f) < V(f_i, f_j)). \tag{13}$$

are the p-values of data and smoothness terms respectively. The modified energy function can be formalized as

$$E(f,d) = \sum_{i \in S} F(U(f,d) < U(f_i, d_i)) + \lambda \sum_{i \in S} \sum_{j \in N_i} F(V(f) < V(f_i, f_j)). \tag{14}$$

The proposed method forms the cdf of the terms using all costs on the graph such that data costs form the data cdf, smoothness costs form the smoothness cdf. The smaller costs of the graph would be on the tail of the distributions which is considered as statistically significant part. Therefore, smaller costs are more statistically significant and take smaller p-values. Note that, Eq.4 and Eq.14 have similarity such that both equations aim to minimize the graph energy. However, the proposed method minimize the energy through the p-values.

The p-values for the data costs are determined according to other data costs on the graph. Similarly, the p-values for the smoothness costs are determined according to other smoothness costs on the graph. It can be interpreted as each cost is normalized relatively to other costs. The normalized energy terms makes the energy formulation more harmonious so the λ regularizes the terms more effectively.

Other advantages of the introduced idea can be expressed using a concrete example. Assume, for example, that data and smoothness terms have equal costs assigned by equations 5, 6, and 7 for pixel i. Assume also that the data cost is less observable, or less frequently seen, than the smoothness cost. Because the data cost is less observable, we can say that data term is statistically more significant than the smoothness term, albeit both terms have equal costs. In order to influence the final result of the energy minimization towards a statistically more significant solution, our approach assigns a lower value (Eq. 12) to the data cost than the smoothness value (Eq. 13) for pixel i. The direct connection between the assignment of costs and the statistical significance of the terms results in an effective dynamic regularization weight adjustment strategy between the data and smoothness terms. Furthermore, since the same approach assigns lower energy cost values for the statistically important pixels (pixels with both low data and smoothness p values), these pixels might push the final result of the energy minimization process towards a statistically significant solution.

4 Experiments

We set up experiments on the medical image segmentation problem to validate the proposed idea. We compared the graph cut energy minimization algorithm in Eq.4

with the modified energy minimization in Eq.14. For a reliable comparison, we construct the same graph structure for both approaches. The only difference is using Eq.5-6-7 for the graph cut and using Eq.12-13 for the modified graph cut. The aim of the experiment is to show that expressing the outputs of the potential functions as p-values make the constraints more harmonious. Therefore regularization weight regularizes the graph energy properly which produces better segmentation results. We run the graph cut and the modified graph cut approaches repeatedly with different regularization weights. To be able to test the whole regularization weight range, we re-formalized the energy minimizations in Eq.4 and Eq.14 as

$$E(f,d) = (1-\lambda)\sum_{i\in S} U(f_i,d_i) + \lambda \sum_{i\in S}\sum_{j\in N_i} V(f_i,f_j), \tag{15}$$

and

$$E(f,d) = (1-\lambda)\sum_{i\in S} F(U(f,d) < U(f_i,d_i)) + \lambda \sum_{i\in S}\sum_{j\in N_i} F(V(f) < V(f_i,f_j)), \tag{16}$$

respectively. Then, we changed the λ parameter between 0 and 1 with small increments. In order to quantitatively evaluate the performance of the approaches, we compared segmentations with the ground truth data which are delineated by an expert manually throughout the experiments.

Our first experiment is on a Liver CT image (Fig.1a). We summarized the procedures of the graph cut and modified graph cut minimizations on Liver CT image in Algorithm 1 and Algorithm 2.

The minimization step of the approach is repeated for different regularization weights between 0 and 1. The labeling f for each regularization weight is compared with the ground truth. The recovered object and background labels are regarded as errors if they are background and object pixels respectively in the ground truth. The minimum error labeling f gives the optimal regularization weight for the objective function (Fig.1b). We repeat the same procedure for the modified graph cut

Algorithm 1 Graph cut minimization

Construct the graph structure :
The image pixels are defined as sites S.
The spatial neighbors (4-connected) of pixels are defined as N.
Manually mark the certain pixels as object and background.
Define the potential functions $U(f)$ and $V(f)$:
Calculate the data terms using Eq.5-6.
Calculate the smoothness terms using Eq.7.
Construct the objective energy function (Eq.15) :
Combine all data and smoothness terms through a regularization weight.
Minimize the objective energy function using graph cut minimization.
Return the assigned labels as the segmentation result. (f of Eq.15)

minimization.

Algorithm 2 Modified graph cut minimization

Construct the graph structure :
The image pixels are defined as sites S.
The spatial neighbors (4-connected) of pixels are defined as N.
Manually mark the certain pixels as object and background.
(For a reliable comparison, we used the same marked pixels in Algorithm 1.)
Define the potential functions $U(f)$ and $V(f)$:
Calculate the data terms using Eq.5-6.
Calculate the smoothness terms using Eq.5-6.
Calculate the statistical significance of potential functions :
Form the probability distributions of each potential function using Eq.8-9.
Form the cumulative distribution functions using Eq.10-11.
Convert the each term to p-values using Eq.12-13.
Construct the objective energy function (Eq.16) :
Combine all p-values of terms through a regularization weight.
Minimize the objective energy function using graph cut minimization.
Return the assigned labels as the segmentation result. (f of Eq.16)

The minimization step of the modified approach is also repeated for different regularization weights between 0 and 1. The labeling f for each weight is compared with the ground truth. The minimum error labeling gives the optimal regularization weight for the objective function. Some of the segmentation results of the approaches are shown in Fig.1. In this paper, we used red for object and blue for background

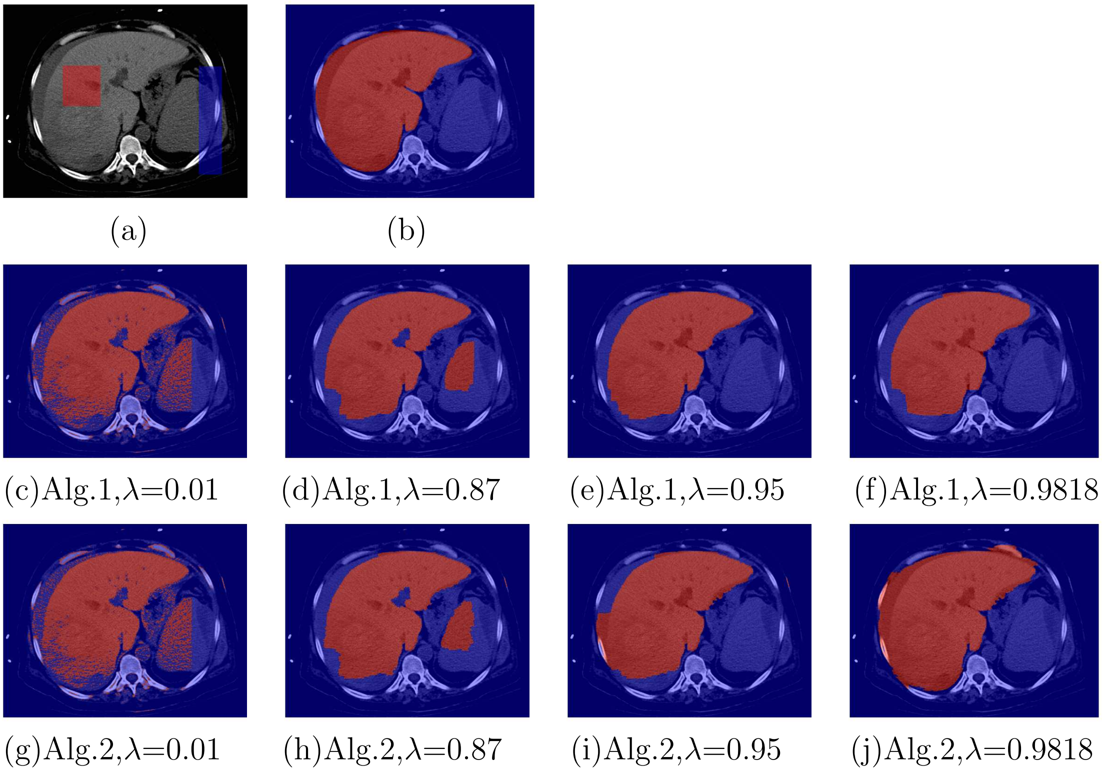


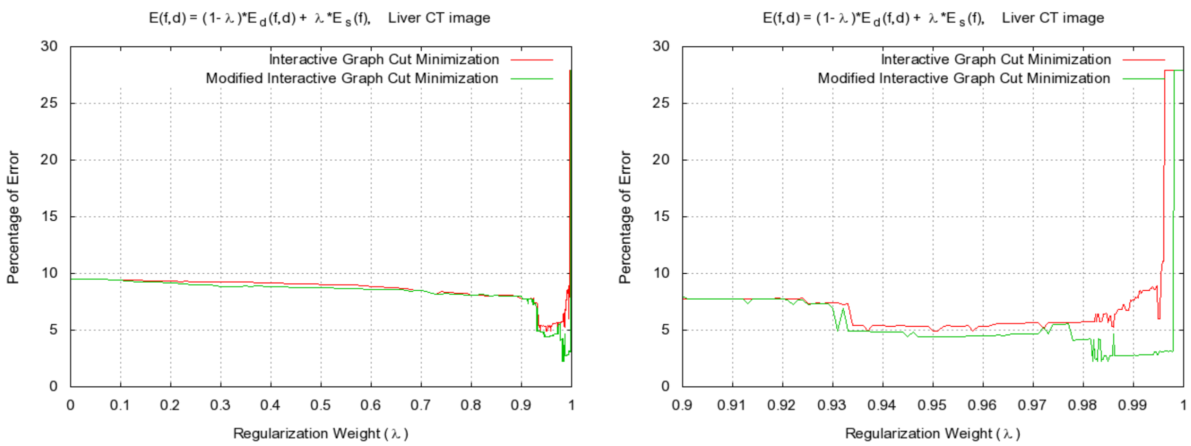
Figure 1: (a) Liver CT image marked as object (red) and background (blue). (b) Ground truth segmentation of Liver image. (c-f) Segmentation results obtained by graph cut minimization for different λ weights. (g-j) Segmentation results obtained by the modified graph cut minimization for different λ weights. Alg.1 and Alg.2 denote graph cut minimization and modified graph cut minimization respectively. The scores of the segmentations are in Table 1.

segments. The Fig.1c-f show segmentation results for various λ parameters by graph cut minimization using Algorithm 1. The optimal λ is 0.95 with the minimum percentage error 4.91% (Fig.1e). The segmentation results in Fig.1g-j are obtained by modified graph cut minimization using Algorithm 2. The optimal λ is 0.9818 with the minimum percentage error 2.23 % (Fig.1j). All the other scores of segmentations are listed on Table 1.

We show the percentage errors for the whole range of λ between 0 and 1 on λ-Error graph (Fig.2a). The errors are calculated by comparing segmentations of each λ with the ground truth segmentation. The minimum points of the curves give the optimal regularization parameter for the approaches. Note that proposed approach produced smaller error rates with the regularization weights. For the optimal weight,

	Graph Cut Minimization			Modified Graph Cut Minimization			
Figure	λ	Error [pixels]	Error [%]	Figure	λ	Error [pixels]	Error [%]
Fig.1c	0.01	21085	9.51	Fig.1g	0.01	21018	9.48
Fig.1d	0.87	17878	8.07	Fig.1h	0.87	17662	7.94
Fig.1e	**0.95**	**10890**	**4.91**	Fig.1i	0.95	9785	4.42
Fig.1f	0.9818	12765	5.76	Fig.1j	**0.9818**	**4935**	**2.23**

Table 1: Detailed scores of Fig.1c-j. Bold scores are obtained by the optimal regularization weights.



(a) (b)

Figure 2: (a) λ-Error Graph of Liver CT in Fig.1. (b) The same λ-Error Graph of Liver CT in the range between 0.9 and 1.

it produced better segmentation than the original graph cut minimization. Suitable regularization parameters for the observed image are located between 0.9 and 1 for both approaches. Therefore, we redraw a detailed λ-Error graph for this range to show the effectiveness of the modified graph cut minimization (Fig.2b).

We carried out the same experiment following the steps of Algorithm1 and Algorithm2 for different medical images. Fig.3 shows another liver CT, a knee MRI and a lung CT. The first column of the image set shows the initial marked pixels by the user. The ground truth of images are shown on the second column. The segmentations in the third column are obtained by graph cut minimization with optimal regularization weight. The segmentations in the fourth column are obtained by modified graph cut approach with optimal regularization weight. Regularization weights and percentage errors of optimal segmentations are shown in Table 2. λ-

Error graphs of the experiments are shown at Fig.4. The interesting regularization parameter ranges of images are between 0.9 and 1. Therefore, we show errors in detailed graphs just for these ranges.

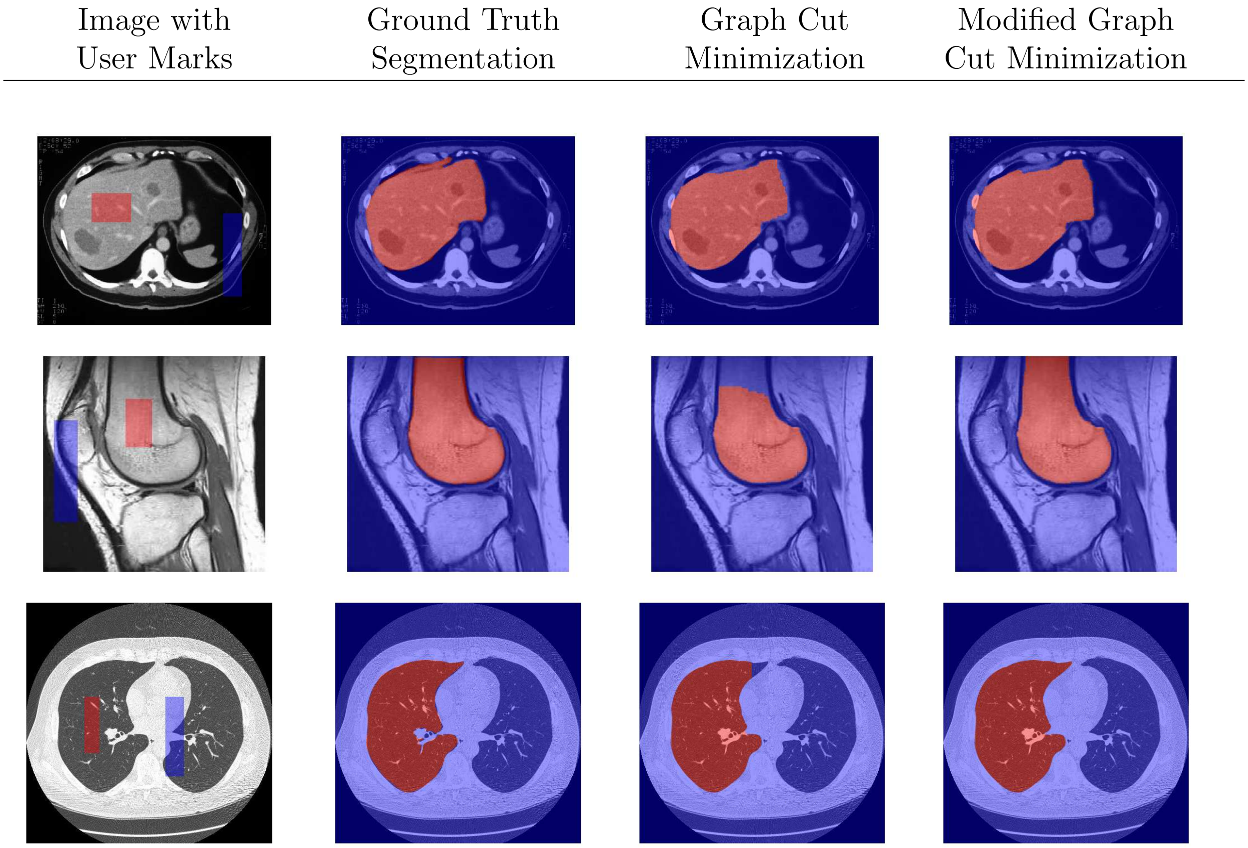


Figure 3: The first column: A Liver CT, a Knee MRI and a Lung CT [31] in vertical order. Images are marked initially with a user. The second column: Ground truth of each image set. The third column: Optimal segmentations by graph cut minimization. The fourth column: Optimal segmentations by modified graph cut minimization. Segmentation scores of both approaches are listed in Table 2.

The λ-Error graphs of each image show the performance of both approaches for all ranges of the regularization weight. The minimum points of the curves indicate that proposed approach produces more effective segmentations with the optimal regularization weights. Moreover, the modified approach also produces better results for most of the regularization parameter ranges. The optimal segmentation results of both approaches and their percentage errors verify the effectiveness of the proposed approach.

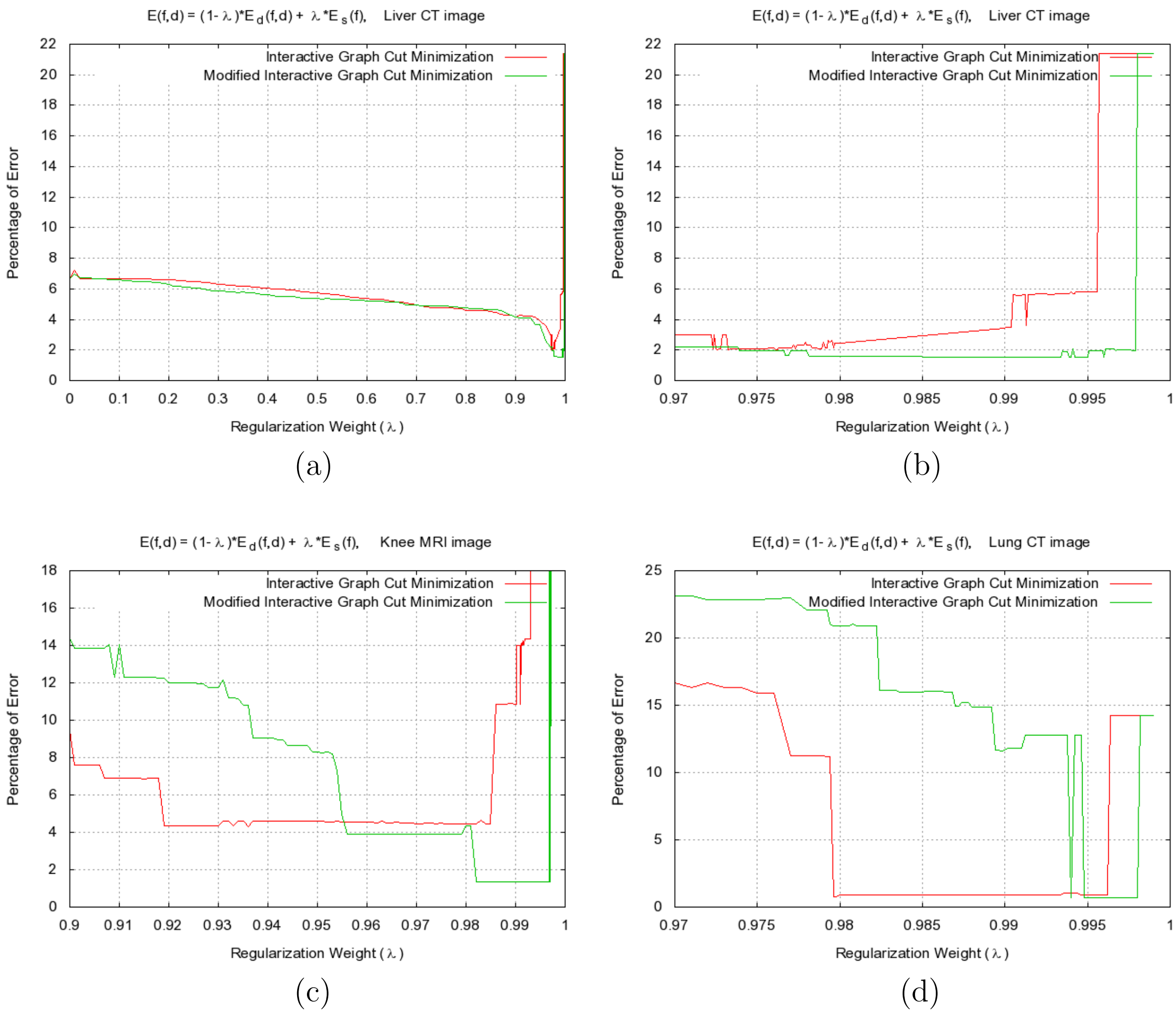


Figure 4: (a) λ-Error graph of Liver CT. (b) Detailed graph of Liver CT in range between 0.97 and 1. (c) λ-Error graph of Knee MRI. (d) λ-Error graph of Lung CT. Optimal regularization weight and errors for that weights are shown in Table 2.

In order to show that the proposed method is not limited for the interactive graph cut minimization, we implemented a small experiment on basic graph cut approach [18] which does not utilize any marked pixels from the user. Because the basic graph cut does not have the priori segmentation knowledge, it produces poor results for the medical image segmentation. However, proposed idea is still applicable for the basic graph cut approach, because it suffers the same problem as interactive graph cut minimization, which is minimizing the inharmonious energy terms. The data energy terms of the basic graph cut minimization is the squared difference between the pixels and their assigned labels. Similarly, the smoothness

	Graph Cut Minimization			Modified Graph Cut Minimization		
Figure	λ	Error [pixels]	Error [%]	λ	Error [pixels]	Error [%]
Liver CT	0.9751	2989	2.05	0.9921	2159	1.52
Knee MRI	0.936	5236	4.32	0.993	1613	1.33
Lung CT	0.98	2269	0.9	0.997	1731	0.69

Table 2: The detailed scores of the medical images in Fig.3.

terms are the intensity differences of neighboring pixels. We used the same liver CT images in Figure 1 and Figure 3, and modified the energy terms using the same steps of the Algorithm-2. (except the steps of the potential function definitions). Then we run the basic graph cut and the modified basic graph cut algorithms repeatedly for the whole λ range. Figure 5 shows the percentage errors of all approaches for the optimal λ parameters. The red bar belongs to the graph structure constructed by the potential functions in [18]. The green bar corresponds to the results after the proposed idea is applied to the potential functions on [18]. The blue bars on the other hand belong to the graph structures constructed by the potential functions in [5] (interactive approach). Similarly, the magenta bar corresponds to the results after the proposed idea is applied to the potential functions on [5]. Note that, interactive graph cut minimization produced lower errors than the basic graph cut minimization. As the figures show the proposed method is applicable for any MRF based minimization system that uses inharmonious constraints.

We compared the modified graph cut method with the other favorite segmentation methods in medical image segmentation. Since the active contours are also based on energy minimization, we selected two main approaches of active contours from the recent literature. One of them is the region based active contour model of Chan and Vese [32], which uses the statistical information of image intensity within each region. Another comparison method is the variational level set [33], which produces better results than the classical level set approach. For a fair comparison, we set the best parameter values and initialized the contours on appropriate locations. The segmentation results of the approaches for the Liver CT image are shown in

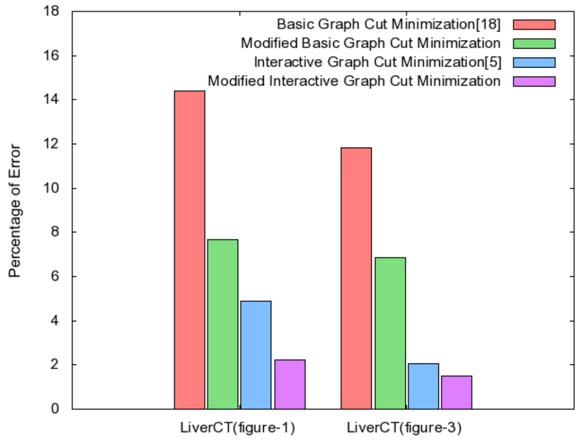


Figure 5: The percentage errors of the original graph cut[18], the modified original graph cut[18], the interactive graph cut[5], and the modified interactive graph cut approaches.

Fig. 6. The percentage errors of these segmentations are listed in Table 3. Because we initialize the contours very close the liver boundary, active contour algorithms almost know where to segment. However, the interactive graph cut approach has limited knowledge about the segmented object (only marked pixels). Therefore, active contour models produced better segmentation results than the interactive graph cut. On the other hand, the modified interactive segmentation (proposed method) regularize the energy terms more effectively, therefore it produces better results than both interactive graph cut and active contour models.

Segmentation Methods	[%] Percentage Error
Interactive Graph Cut [5][6]	4.91
Region Based Active Contour [32]	3.38
Variational Level Set [33]	3.39
Modified Interactive Graph Cut (Proposed Method)	2.23

Table 3: The percentage errors of the liver CT segmentations in Fig.6

5 Conclusions

The graph cut approach formulates the objective energy function as a linear combination of potential functions of data and smoothness constraints. Since the terms can

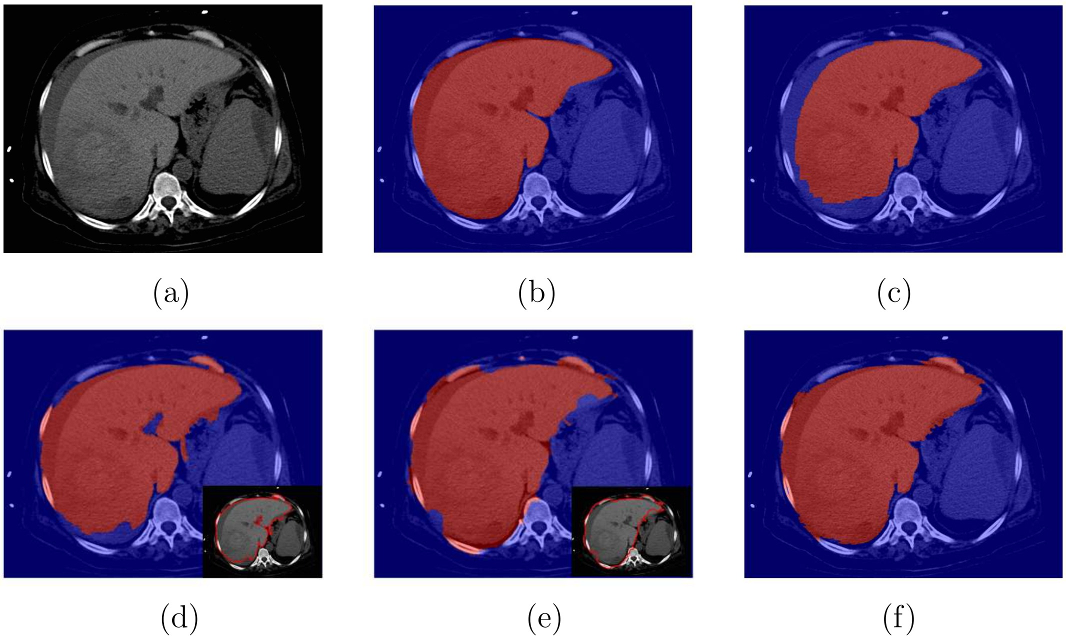


Figure 6: a) Liver CT image b) Ground Truth Segmentation c) Interactive Graph Cut [6] d) Region Based Active Contour Method [32] e) Variational Level Set Method [33] f) Modified Interactive Graph Cut (proposed method)

have different functional forms, forming the energy formulation can be interpreted as linear combination of inharmonious constraints. The regularization parameter regularizes the difference between the functional forms of the terms and determines the smoothness weights. We proposed an empirical method which increases the tuning performance of the regularization parameter. The method modifies the graph structure by measuring the statistical significance of each term and expresses them as p-values. Expressing the terms on the graph by the same measurement unit decreases the scale and distribution differences between the terms. The more harmonious terms are properly balanced by the regularization weight, so the minimization algorithm produces a better solution. Experimental results showed that the modified graph cut approach improves the graph cut segmentation significantly because of the new regularization.

Determining the regularization parameters is not a new area for the computational vision problems. The literature has different parameter estimation methods. However, none of them is concerned in increasing the effectiveness of the regularization weight. The novelty of the proposed method is increasing the tuning performance of the regularization weight. Another novelty of the method is using the statistical significance as a measurement unit for the terms with different scales and distributions. The literature generally uses the statistical significance as a decision measurement for an hypothesis. However, our approach uses it to bring the inharmonious terms on a common base.

Another noteworthy point is that the method adjusts the effect of the regularization weight by adjusting the weights of each term individually. Note that, the method produces a p-value for each term by evaluating the terms in their distributions. Because of the relative evaluation, the importance of each term is adjusted individually in the energy minimization. In recent years, estimating the adaptive regularization weights from the observed images have become important to obtain optimal performance. Individual adjustment of each weight makes the proposed method similar to these methods which try to adjust the regularization weight adaptively.

The proposed method improves the graph cut segmentation clearly, whereas it requires minimal extra computational load for the significance calculation. Cdf's are calculated from the costs on the graph which are already calculated for the original graph cuts. The only computational load is converting the cost values to p-values which is a linear time operation.

The general structure of the method makes it applicable to other energy minimization applications that involve regularization such as snakes, level sets or stereo correspondence. Although we did not propose the method to estimate the optimal regularization parameter, it can be employed in increasing the performance of the parameter estimation techniques in the literature for finding optimal regularization parameter.

References

[1] L. Lorigo, O. Faugeras, E. Grimson, R. Keriven, R. Kikinis, C. Westin, "Co-dimension 2 geodesic active contours for MRA segmentation", Inf. Process Med Imaging, vol.16, pp.126-139, 1999.

[2] S. Osher, N. Paragios, "Geometric level set methods in imaging, vision, and graphics", Springer, 2003.

[3] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, A. Willsky, "A shape based approach to curve evolution for segmentation of medical imagery", IEEE Trans. Med. Imaging, vol.22, pp.137-154, 2003.

[4] Y. Boykov, M.P. Jolly, "Interactive organ segmentation using graph cuts", Lecture Notes In Computer Science, Proc. of the Third International Conference on MICCAI, pp. 276-286, 2000.

[5] Y. Boykov, M.P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images", IEEE Conf. on Computer Vision, vol.1, pp.105-112, 2001.

[6] Y. Boykov, G. Funka-Lea, "Graph cuts and efficient N-D image segmentation", Int. J. Computer Vision, vol.70, pp.109-131, 2006.

[7] D. Freedman, T. Zhang, "Interactive graph cut based segmentation with shape priors", IEEE Conf. on Computer Vision and Pattern Recognition, vol.1, pp.755-762, 2005.

[8] J. Zhanga, Y. Wangb, X. Shia, "An improved graph cut segmentation method for cervical lymph nodes on sonograms and its relationship with nodes shape assessment", Computerized Medical Imaging and Graphics, vol.33, pp.602-607, 2009.

[9] J. Hadamard, "Sur les problmes aux drives partielles et leur signification physique", Princeton University Bulletin, 1902.

[10] D. Marr, T. Poggio, "Cooperative computation of stereo disparity", Science, pp.209-236, 1976.

[11] M. Bertero, T.A. Poggio, V. Torre, "Ill-posed problems in early vision", Proceedings of the IEEE, vol.76, pp.869-889, 1988.

[12] J. Marroquin, S. Mitter, T.A. Poggio, "Probabilistic solution of ill-posed problems in computational vision", J. Amer. Stat. Assoc., vol.82, pp.76-89, 1987.

[13] J. Besag, "On the statistical analysis of dirty pictures", J. of the Royal Statistical Society B., vol.48, pp.259-302, 1986.

[14] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, "Equations of state calculations by fast computing machines", J. of Chemical Physics, vol.21, pp.1087-1091, 1953.

[15] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images", IEEE Trans. Pattern. Anal. Mach. Intell., vol.6, pp.721-741, 1984.

[16] S. Kirkpatrick, C.D.J. Gelatt, M.P. Vecchi, "Optimization by simulated annealing", Science, vol.220, pp.671-680, 1983.

[17] P. Carnevali, L. Coletti, S. Patarnello, "Image processing by simulated annealing", IBM Journal of Research and Development, vol.29, pp.569-579, 1985.

[18] D. M. Greig, B. T. Porteous, A.H. Seheult, "Exact Maximum A Posteriori Estimation for Binary Images", J. Royal Statistical Soc., Series B., vol.51, pp.271-279, 1989.

[19] J.M. Hammersley, P. Clifford, "Markov field on finite graphs and lattices", unpublished, 1971.

[20] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", IEEE conf. on Computer Vision, vol.47, pp.7-42, 2002.

[21] A.N. Tikhonov, V.A. Arsenin, "Solutions of ill-posed problems", Washington, DC: Winston, 1977.

[22] K. Krajsek, R. Mester, "Maximum Likelihood Estimator for Choosing the Regularization Parameters in Global Optical Flow Methods", IEEE Conf. on Image Processing, pp.1081-1084, 2006.

[23] L. Zhang, S.M. Seitz, "Estimating optimal parameters for MRF stereo from a single image pair", IEEE Trans. Pattern. Anal. Mach. Intell., vol.29, pp.331-342, 2007.

[24] B. Peng, O. Veksler, "Parameter selection for graph cut based image segmentation", Proc. of the British Machine Vision Conference, 2008.

[25] Y. Boykov, O. Veksler, R. Zabih, "Fast approximate energy minimization via graph cuts", IEEE Trans. Pattern. Anal. Mach. Intell., vol.23, pp.1222-1239, 2001.

[26] Y. Boykov, V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts", IEEE conf. on Computer Vision, vol.1, pp.26-33, 2003.

[27] C. Rother, V. Kolmogorov, A. Blake, "GrabCut -interactive foreground extraction using iterated graph cuts", SIGGRAPH, August, 2004.

[28] J.E. Freund, "Mathematical statistics", Prentice Hall. 2007.

[29] R.A. Fisher, "The design of experiments", New York: Hafner. 1935.

[30] S. Candemir, Y.S. Akgul, "A nonparametric statistical approach for stereo correspondence", IEEE conf. on Computer and Information Sciences, pp.1-6, 2007.

[31] http://www.via.cornell.edu/databases/lungdb.html

[32] T.F.Chan, L.A.Vese, ”Active Contours Without Edges”, IEEE Trans. on Image Processing, vol.10, No.2., pp.266-277, 2001.

[33] C.Lee, C.Xu, C.Gui, M.D.Fox, ”Level Set Evaluation Without RE-initialization: A New Variational Formulation”, IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 430-436, 2005.