Multistage Graph-based Segmentation of Thoracoscopic Images

Guillaume-Alexandre Bilodeau[1] , Yueyun Shu[1], Farida Cheriet[1,2]

[1]École Polytechnique de Montréal, C.P.6079, succ. Centre-ville, Montréal, QC, Canada, H3C 3A7

[2]Sainte-Justine Hospital, 3175 Côte Sainte-Catherine, Montréal, QC, Canada, H3T 1C5

guillaume-alexandre.bilodeau@polymtl.ca, shu.yue-yun@polymtl.ca, farida.cheriet@polymtl.ca

Abstract

This paper presents a novel method of graph-based segmentation using multiple criteria in successive stages to segment thoracoscopic images acquired during a diskectomy procedure commonly used for thoracoscopic anterior release and fusion for scoliosis treatment. Starting with image pre-processing, including Gaussian smoothing, brightness and contrast enhancement, and histogram thresholding, a standard graph-based method is applied to produce a coarse segmentation of thoracoscopic images. Next, regions are further merged in a multistage graph-based process based on features like grey-level similarity, region size and common edge length. Experimental results show that our approach achieves good spatial coherence, accurate edge location and appropriate segmentation of the regions of interest from a sequence of thoracoscopic images.

Keywords: Thoracic diskectomy, thoracoscopic images, Video-assisted thoracoscopic surgery (VATS), graph-based segmentation, region merging

1 Introduction

Thoracic diskectomy [1] is a video-assisted thoracoscopic surgery (VATS) which mainly consists of removing the intervertebral disc at some levels of the spine to treat symptomatic thoracic herniation causing spinal cord and nerve root compression. Thanks to many advances and improvements this minimally invasive surgery could also be used for instrumentation, correction

and fusion of thoracic and thoracolumber scoliosis. During this procedure, instruments and an endoscope are inserted through small incisions on the patient's body and the surgeon performs the operation by viewing the images acquired by the endoscope on a video monitor. The endoscope acquires image sequences of the instruments and the surgical site while the surgeon is removing the intervertebral disc. The advantages of this procedure are minimal inconvenience to patients by reducing the exposure of anatomical organs and the blood loss. Compared to the conventional surgeries, this method reduces patient risk and costly in-hospital recovery periods.

Some challenges still need to be overcome to facilitate the adoption of this procedure for the treatment of scoliosis, because the surgeon needs a lot of training as he loses depth perception since the surgical site is viewed indirectly via the monocular endoscope. Furthermore, context depth cues are scarce as the lens of the endoscope is typically very close to the organs being imaged. For these reasons, computer vision techniques can assist surgeons by enhancing thoracoscopic images to avoid any damages to the spinal cord. Thoracoscopic images obtained in the context of thoracic diskectomy are very challenging as intensity fall-off rapidly with radial distance from the image center, strong specular reflections are caused by the instruments and moist tissues, large changes in apparent size of objects due to motion and focusing of the endoscope, and exposed tissues or cuts which cause variations in textures and colors of the background. Furthermore, the whole image is made of shades of red. Fig. 2 shows sample thoracoscopic images.

To help surgeons, the goal of this collaborative work with Ste-Justine Children's Hospital in Montreal is to localize the boundary of the operated disc cavity, to allow its 3D reconstruction and positioning relative to operating instruments and a 3D model of the spine obtained from MRI images acquired before surgery. The relative distance between the instruments and the spinal cord could be estimated based on the 3D registration of the cavity with the preoperative 3D model of the spine.

In this paper, we focus on the segmentation of the cavity to localize it in images. It is the first challenging step toward a 3D navigation system for minimally invasive surgeries of the spine. The contribution of this paper is a novel multistage graph-based segmentation method that overcomes usual graph-based segmentation drawbacks by progressively integrating higher-level criteria. The proposed method is designed to handle thoracoscopic images acquired in the context of thoracic diskectomy procedure.

The paper is divided as follows. Previous works on endoscopic images and existing segmentation methods are reviewed in section 2. In section 3, we present the details of our segmentation method. In section 4, results are presented and analyzed. We conclude in section 5.

2 Related work

With medical imaging playing an increasingly prominent role in the diagnosis and treatment of diseases, segmentation techniques have been applied for extracting clinically useful information about anatomic structures through different modalities such as X-rays, CT, MRI, PET and ultrasound [2]. One common observation about medical applications of segmentation methods is that a single method can seldom fulfill the task. It is often necessary to combine several segmentation techniques from different categories to obtain a useful result. Although there are numerous segmentation techniques of medical images, few applications on thoracoscopic surgical images were reported in the literature.

Some researches [3-8] have been conducted in recent years to process laparoscopic images for the development of 3D navigation system to assist surgeons. Most researchers focused on segmenting instruments in images and tracking their movements. Ueckery et al. [3] used a Bayesian classifier to segment images into two classes, which are organ and instrument. Lo et al. [4] used a similar method in their tissue/instrument segmentation. Recently, Boisvert et al. [5] used a support vector machine for instrument segmentation using a region growing technique combined with a classification approach. For the thoracoscopic images we are concerned with,

statistical or leaning methods are difficult to apply. Our data which represent a cavity operated during a thoracic discectomy procedure changes drastically during the operation and from patient to patient. Hence, specific features are hard to select and therefore features from our data cannot be reliably learned. A set of powerful discriminating features must be chosen for classification, or otherwise the classifier would not work well or even fail.

Other methods for processing specific endoscopic images are reported. Asari [6] designed a fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images. A differential region growing technique is used on the basis of a similarity criterion, and a dynamic hill-clustering method ensures the effectiveness of the terminating condition during the growth process. Bockholt et al. [7] used statistical texture analysis for bladder tumor detection. Karkanis et al. [8] presented an approach on the detection of tumors in colonoscopic video. It benefits from a color feature called color wavelet covariance (CWC), which is based on the covariance of second-order textural measures. The images processed in these applications do not share the same characteristics as ours.

For a multi-class segmentation of thoracoscopic images more general bottom-up approaches should be more suitable. Roubert [9] used a bottom-up approach specifically aimed at segmenting instruments. Focusing on distinguishing instruments from the background, thresholding and window filtering are used to filter out the regions that were considered non-instruments in the first step, and then edge detection and line fitting are used to label the instruments in the remaining image. This method cannot fit our goal since we cannot reject any region to facilitate further processing. In addition, line fitting does not apply to segmentation of unstructured tissues.

Analyzing the wealth of segmentation methods available in computer vision, we have reached the following conclusions. Methods like thresholding and clustering are not desirable options as they do not take into account position of similar pixels. Edge-based methods are not suited either, because we have to focus more on regions than boundaries because their appearance is changing with the point of view of the endoscope.

From our analysis, we believe the best option is region-based method where regions are constructed by grouping pixels. Among these methods, graph-based methods are of particular interest. Graph-based methods map an image onto a graph where nodes are composed of pixels/regions and edges are composed of links between neighboring nodes. Each node has a weight based on some feature and each edge has a weight generally defined by the weight difference of the nodes it connects. The algorithm will group nodes [10] or cut the graph into connected regions [11] by edge weight (reflecting similarity of pairs of nodes). It can be used without any supervision, and do not require a learning phase. Graph-based segmentation takes into account global image properties as well as local spatial relationships, and results in a regions map that is ready for further processing, e.g. region merging or labeling. Among many variations of graph-based segmentation methods, recursive shortest spanning tree (RSST) based algorithms [12] are of particular interest. The algorithm recursively finds the shortest link weight edge and eliminates it. Thus the two nodes connected by that edge are grouped into one region. The RSST algorithm groups similar pixels into homogeneous regions in a desired fashion for our application. It has been successfully applied in many segmentation tasks [10, 13, 14]. Most recent applications include object-oriented segmentation for next generation video standard [15, 16].

Although graph-based segmentation is a good bottom-up approach, it does not generate acceptable results all the time, because it has some intrinsic drawbacks. As mentioned above, graph-based segmentation relies solely on weight differences as merging criterion, and it will merge low weight difference regions iteratively. Normally, the algorithm will not stop merging until a certain number of regions are left [10] or will be using a threshold for the edge weight. In some other graph-based methods such as [17], an initial weight is defined and it will ultimately determine when to stop merging. The parameter that defines the stop condition is usually called K. We observe that there are three main drawbacks to graph-based segmentation:

- The optimal K is hard to determine, and inappropriate K can cause under-segmentation or over-segmentation. The value of K depends on the image characteristics and the specific application.
- The average intensity among regions tends to be closer when the regions grow bigger. Using a single criterion as average intensity difference may be reasonable at the beginning, but it may not be the most appropriate when regions grow to some extend.
- Prior knowledge or constraints are not used in the merging procedure, so the results are hardly controllable. For example, small regions sometime remains inside bigger ones up to the end, because weight differences among the small regions and their neighbors are too large. Those small regions are possibly spurious areas that we want to get ride of.

In this paper we modified graph-based segmentation by adding multiple stages of merging using each different criterion that include high-level knowledge. This approach allows overcoming the previously enumerated drawbacks. The next section describes the proposed methodology.

3 Multistage graph-based segmentation method

Thoracoscopic surgical images differ from natural scenes as they are acquired in a compact viewing area with limited illumination, and are mainly composed of tissues in similar red-like colors. Another difficulty of this kind of images is the specular reflections from moist tissues and metallic instruments that change unpredictably from frame to frame. The nature of this kind of images (noisy, low contrast and fuzzy boundary), requires image pre-processing, such as smoothing, contrast enhancement, etc. The pre-processing can enhance discriminating features, but it is still hard to obtain an acceptable segmentation from standard graph-based method, especially with thoracoscopic images.

We propose a method that integrates standard graph-based segmentation and additional graph-based segmentation stages, each with different criteria. It is composed of the following steps:

1. Image pre-processing using Gaussian smoothing, brightness and contrast enhancement, color space conversion and specular reflections removal.
2. Coarse standard graph-based segmentation with a roughly selected parameter K.
3. Region post-processing to remove very small regions considered to be spurious areas.
4. Multiple stage of graph-based segmentation with different higher-level criteria.

These steps are shown in a flow chart (Fig. 1), and they are described in details in the subsequent sections following a brief discussion on the selection of a basic segmentation criterion.

Figure 1

3.1 Segmentation criteria

Many image features can be used as criteria of homogeneity for image segmentation, such as color, intensity, texture, motion vector, elemental model, etc. For thoracoscopic surgical images, due to noise, poor illumination and similarity in colors and unwanted specular reflections, some are not usable. Because there is not much texture and no dominant background or foreground in thoracoscopic images, texture and motion vector are not reliable. Furthermore, the color differences among regions are small, and the shapes of the regions are variable, so color and shape model are not reliable criteria either. Intensity is considered to be the most suitable one as our segmentation criterion. Intensity, which is a commonly used criterion, is simple and yet efficient in many situations. It can suit our images as pixels are variations of red. Hence, color

information is not relevant in our case, and it is the intensity of the different shade of red that are of interest to us.

We will use intensity as basic criterion during segmentation and do some pre-processing to increase region differences before applying the segmentation algorithm.

3.2 Image pre-processing

By pre-processing some unwanted variations, noise can be reduced and desired features enhanced [18]. We apply the following procedures to pre-process our data:

- Color to intensity image conversion
- Gaussian smoothing
- Brightness and contrast enhancement
- Specular reflections removal

The image is first processed to obtain an intensity image. There are many ways to do the color space conversion [19]. We have chosen to use the *Rec 709* standard:

$$Intensity = 0.213 * Red + 0.715 * Green + 0.072 * Blue \quad \text{(EQ 1)}$$

It gives less weight to the red component, and hence increases the differences between the reddish regions. Then, Gaussian smoothing allows us to remove variations and noise. We use a 2-D convolution operator with a standard 5 x 5 kernel. We found experimentally that the 5 x 5 kernel is more suited then smaller or bigger kernel on our data as it filters noise and blur the image moderately. Enhancing brightness and contrast allows us to improve information visibility. We use a standard brightness/contrast adjustment algorithm [20]. Finally, pre-processing ends by removing spurious pixels that are actually very high intensity pixels caused by specular reflections from polished instruments and moist tissues in the surgical site. We analyze the image histogram to find those specular reflections, and use simple thresholding to truncate the intensity of these pixels. This is an important step, because pixels corresponding to specular reflections

have large weights and they may affect region merging results unexpectedly. The graph-based segmentation algorithm follows.

3.3 Multistage graph-based segmentation algorithm

Our multistage graph-based segmentation algorithm is composed of a coarse standard graph-based segmentation followed by multiple stages of graph-based segmentation using different high-level criteria adapted to our application.

3.3.1 Stage one: Coarse graph-based segmentation (RSST)

For coarse standard graph-based segmentation, our method is based on an efficient variant of the RSST algorithm [17]. In RSST, an image is mapped onto a graph. A 2D image is described as an undirected graph *G = (V, E)*, in which pixels $v_i \in V$ are mapped to nodes in the graph, where the edges $e_k = (v_i, v_j) \in E$ reflects spatial relationships between neighboring pixels. The graph can be 4-connected or 8-connected. It means that a pixel can be linked to four or eight neighbors to form the graph edges. We have chosen to use four neighbors because it will result in a coarser segmentation that is just what we need for the following graph-based segmentation stages. Indeed, for this first coarse segmentation, we aim at obtaining images that are over-segmented, so higher-level criteria can be used to efficiently group the region into instrument and different type of tissues, including the cavity. Moreover, using four neighbors forms fewer edges which results in faster processing.

Each node and each edge has a corresponding weight. The weight of a node is the intensity (grey level) in our case, and the weight of an edge is the weight difference of linked nodes. During the segmentation process, similar neighboring pixels are grouped into regions, and then the weight of a region is updated as the average intensity of composing pixels. A segmentation *S* is a partition of *V* into regions such that each region *R* corresponds to a connected subset of the edges in *E*. It is conducted by recursively finding the least weight edge and merging the two

regions (which are originally pixels) connected by the edge. After each merge, the weights of affected regions are recomputed.

More specifically, RSST operates as follows. RSST recursively builds the shortest spanning tree (SST). The RSST starts with a mapping of an image onto a weighted graph. Each region or vertex initially contains only one pixel. The pixel intensity values of regions are used to evaluate vertex weights and link weights of the graph. A vertex weight is defined as the average intensity value of the corresponding region, while a link weight is evaluated by a cost function, which is basically a function of the vertex weights and the sizes of the connected regions. All links are then sorted according to their link weights, and stored in a heap. The linking process is an iterative process to form the SST. In the beginning of the cycle, a link with the smallest link weight in the graph is chosen to add in the SST. The chosen link is saved and the two connecting or merging regions are merged. The vertex weight of the newly merged region is updated, hence, all surrounding links need to be recalculated and all loop-forming links, also known as duplicated links, will be removed. Subsequently, all remaining links are sorted. Thus, the number of regions is progressively reduced from a pixel-by-pixel image, down to just one if desired. Those saved links form a spanning tree representation of the image. By noting the order in which the links are saved, a hierarchical representation of the original image is created.

The traditional RSST based segmentation algorithm [12] will recompute weights of related edges as well and then find the least weight edge to merge. Since sorting of edges is slow as there are a lot of edges (e.g. 168 960 edges in a 352 x 240 image), putting it in the computation cycle is really time-consuming. More efficient algorithm [17] will neither recompute weight of edges, nor re-sort edges, but use an internal difference criterion to evaluate if two regions can be merged.

Although the efficient algorithm makes the segmentation criterion and stop condition not as clear as the traditional one, it is really efficient (nearly real-time running on a Pentium IV 2.4GHz PC). The efficient algorithm is hundreds of times faster than the traditional one on segmenting a typical 352 x 240 video frame of a sequence of diskectomy procedure, and the quality of

segmentation result is quite similar. Actually, the exact level of similarity depends on the parameters used. In our case, it does not matter since we are aiming at obtaining a coarse segmentation that will be used as input to segmentation stages with more advance region merging criteria. We just need to select parameter K to obtained over-segmented image, as under-segmentation cannot be resolved with our algorithm. We use this efficient algorithm to generate an initial segmentation map as the input to other stages of graph-based segmentation.

3.3.2 Graph-based segmentation with higher-level criteria

The output of any graph-based segmentation method can be improved by simply merging similar neighboring regions together. The idea of region merging is similar to that of graph-based method, but it starts from regions instead of pixels, and can use different merging criteria at different iterations. Several region merging algorithms were proposed in the literature. Among those, Brox et al. [21] applied a multistage merging process based on Ward, Mean-Ward and border criteria following a watershed pre-segmentation. Xuan et al. [22] used a merge score based on grey-level similarity, region size and region connectivity in their MR brain image segmentation. Van Droogenbroeck et al. [23] explored texture features for their merging criteria. Compared to these methods, our algorithm is composed of four distinct merging stages which are based on graph-based segmentation. It incorporates in a single multistage method both the standard graph-based approach and higher-level criteria used in other region merging approaches. It allows us to process region in a unified graph-based manner starting from pixels and then adding progressively at different stages higher-level knowledge. Standard graph-based methods use a single criterion. Hence, our method enhances graph-based segmentation by using multiple stages with different criteria as performed in other region merging methods.

For stage two, three and four, a graph-based algorithm similar to stage one is used. At each stage, regions are merged iteratively from small to large until convergence, i.e. no more merging. Merging scores are computed and thresholded at each stage. Stop conditions are used to prevent under-segmentation. The merging scores are computed from the following region features: grey-

level similarity (as with standard graph-based merging), region size and common edge length. The formulation of the scores is based on what we believe are reasonable hypotheses in the context of our work, and it is distinct in different stage. The stop conditions are based on prior knowledge; in our case, they are minimum number of regions and minimum and maximum size of regions (percentage of the whole image). The stop conditions are kept unchanged at each stage.

The higher-level criteria used in the segmentation stages are describe in the following paragraphs.

Stage two: Intensity and common edge

Discrimination between adjacent areas with different intensity means and standard deviations can be made according to a variant of Fisher's criterion [24]:

$$F = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^4 + \sigma_2^4}}, \qquad \text{(EQ.2)}$$

where μ and σ^2 are the mean and the variance operators respectively. In other words, if two regions have good separation in their means, and low variance, then they can be discriminated. However, if the variance becomes high and the mean difference is low it is not possible to separate them. Note that we do not use the standard Fisher's criterion:

$$F' = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}, \qquad \text{(EQ.3)}$$

because it increases more the difference between the means. Since our images are composed of closely related shades of red and brightness and contrast have been adjusted, we do not wish to further increase the difference between the regions, which would limit merging possibilities.

Although Fisher's criterion is a good representation of similarity, we also want the two neighboring regions that are going to be merged to be intersecting each other as much as possible. By intersecting, we mean that they have some neighboring boundary pixels and these pixels form what we call a common edge. So we use the merging score formula:

$$\text{Score 1} = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^4 + \sigma_2^4}} \times \frac{P}{L}, \qquad \text{(EQ.4)}$$

where P is the perimeter of the smaller region and L is the common edge length, P and L are counted in number of pixels.

Given a region, we will find the smallest score among its neighbors, and then merge the two regions if the score is less than a threshold (determined through experiments) and the common edge is not too small ((P/L)<5, means 20% of the perimeter of the smaller region). Hence at this stage, we favor merging regions of close similarity as well as with large intersection.

Stage three: Size and common edge and intensity

Now that we have merged similar regions, suppose some adjacent ribbons of regions remain unmerged. Generally, these ribbons are not likely to be independent objects. In our case, we want them to be merged into large neighbors as long as they intersect each other adequately. The merging score should include a size term and a common edge term:

$$\text{Score 2} = \frac{S_1}{S_2} \times \frac{P}{L}, \qquad \text{(EQ.5)}$$

where S_1 is the size of the smaller region, S_2 is the size of the larger region, P is the perimeter of the smaller region and L is the common edge length. S_1, S_2, P and L are counted in number of pixels.

Given a region, we will find the smallest score among its neighbors, and merge two regions if the score is less than a threshold (determined through experiments) and the intensity difference is moderate (<25, determined through experiments with thoracoscopic videos). Hence at this stage we favor merging regions of large size difference as well as large intersection. Note that intensity is still used to avoid merging too dissimilar regions.

Stage four: Common edge then size, and intensity

This stage is a relaxation of stage three for further forming larger regions. First we disregard the size factor (S_1/S_2) and focus on the common edge term (P/L) to remove regions that are almost contained in other regions. Then, we disregard the common edge factor and focus on the size term to remove small regions adjacent to very large regions. At this stage intensity difference is still used as a condition to avoid merging dissimilar regions, but it is relaxed (<50, determined through experiments with thoracoscopic videos).

4 Results and analysis

Our segmentation method has been implemented in C++. It uses the RSST code developed by [17]. The method has been integrated into a software that can process thoracoscopic video files. The thoracoscopic videos were supplied by Sainte-Justine Hospital in Montreal. They correspond to operations performed on small pigs which are used by the surgeon for training. Images used to test our method are dynamically extracted from video files. Images size is 352 x 240 pixels. Images are first preprocessed, and then the segmentation method is applied. The quality of the result has been evaluated based on ground truth localization of the cavity. The ground truth images have been generated using a drawing software based on discussions and indications from a surgeon. Ground truth segmentations for sample images used in the paper are shown in Fig. 2.

Figure 2

Segmentation results are shown in Fig. 3, where results of single criterion graph-based segmentations are compared with our multistage graph-based segmentation method. The efficient RRST segmentation in Fig. 3a) correspond to the coarse segmentation obtained by our method at stage one. Analyzing results from Fig. 3 shows that a finer single criterion graph-based

segmentation could not improve the quality of the results as the tissues are over-segmented. Using a single criterion cannot produce the desired segmentation (see Fig. 2a) even if fine tuning the stop criterion. In fact, because of the single criterion used during the whole segmentation, graph-based segmentation could not remove spurious areas inside an object; or it would segment the whole image into few very large regions that are not corresponding to meaningful objects (see Fig. 3c). This is because higher-level knowledge needs to be added gradually to the merging criterion. Our multistage segmentation algorithm helped to partition the image into more meaningful structures or objects (block of tissues, cavity, and instrument), because the selected criteria favor merging small regions with their neighbors and finally segment the image into several subjectively meaningful large regions. Our experiments show that our method accepts merging regions with small color variations. Specular reflections do not affect the result either (see Fig. 3d).

Figure 3

Compared to ground truth segmentation of the cavity (Fig. 2a), a part is missing with our algorithm. In fact, visualization of the whole part of the video sequence is needed to identify correctly the cavity in this image. This is because of the incidence angle and the lost of depth perception. However, starting with a less problematic viewpoint, and by tracking the cavity in time with an active contour, we believe we can use motion vector from the pixel on the region boundary to identify the cavity correctly. Fig. 4 shows another result where the cavity identified by our algorithm is more closely related to ground truth segmentation. Still, it can be noted that bottom of the cavity is easier to identify than its edges because of illumination, which make the pixels at the boundary of the cavity more closely related to the surrounding tissues. Perhaps taking into account this fact using an illumination model for positioning the cavity boundary

could help to solve this difficulty. In any case, a major part of the cavity is correctly segmented. Studying its evolution in time should allow us to locate it more precisely.

Figure 4

Fig. 5 shows another set of images including intermediate results of our multistage segmentation algorithm. Note that the second segmentation stage did most of the work of forming meaningful regions (Fig. 5b), but there are still some small regions left because the compound score of similarity and intersection is within selected threshold. Then at the third segmentation stage (Fig. 5c), small regions are merged into large neighbors as long as their intensity differences are not too large. At the last stage (Fig. 5d), further merging happens when a region is almost contained in a neighbor region or it is adjacent to a relatively very large neighbor. As it can be observed from this result, our algorithm improves graph-based segmentation by varying the criteria at each stage and by adding gradually higher-level criteria.

Figure 5

Recall we used image pre-processing and multiple stages of graph-based segmentation in the whole process and some parameters need to be specified at each phase. The pre-processing methods are general except for histogram thresholding, which is application specific for removing specular reflections. Actually it is not necessary to use the thresholding for most natural scenes. In the coarse RSST segmentation only a K leading to a coarse segmentation needs to be specified. Our merging algorithm is not sensitive to the initial segmentation as long as it is not too fine (under-segmentation). In the following stages, we need to set the thresholds and other parameters coming from prior knowledge and experiment. In fact, parameters need to be selected according to the image characteristics, but once they are set, they can be applied to same kind of

images (In our case, complete sequences of laparoscopic images). Hence, we believe our method can be applied to other applications.

Although multistage of graph-based segmentation can significantly improve the segmentation results, it does not guarantee success all the time. Note that merging is based on the output of segmentation, so the final result may not correspond to ground truth if the initial coarse segmentation has under-segmented some regions. This happens especially on the thoracoscopic surgical images with instruments in the field of view (Fig. 6 and see Fig. 2c), because the reflection on the instrument makes its color or intensity very close to the surrounding tissues. We are considering using some other ways to refine the boundary, such as the contour modification method used in [15], or using a special filter to make the difference between the instruments and surrounding tissues large enough for a correct segmentation. In this case, we would have expected the cavity to be split in two. Then we could have use temporal information to merge both parts. Unfortunately, it is not the case as a part of the cavity has been merged with the instrument.

Figure 6

From the results of this experiment, we are planning to integrate temporal tracking using active contour. The results obtained in this paper will be use to initialize the active contour after a semi-automatic selection of the cavity region by the surgeon for a segmented region. By using temporal redundancy and tracking changes in the images, we should be able to identify the cavity to allow positioning. This hypothesis will be tested in future work.

4 Conclusion

In this paper we have introduced a multistage graph-based segmentation that adds progressively high-level criteria. The method first uses an efficient variant of the RSST algorithm to do an

initial coarse segmentation, followed by similar graph-based segmentation using high-level criteria to further merge regions into more meaningful structures or objects. Experimental results on laparoscopic images of a thoracic discectomy procedure show that our method can segment the cavity from other tissues, although not perfectly, for a variety of challenging images. In contrast to graph-based segmentation methods using a single criterion, our multi-criteria merging method shows significant improvements on final results in our experiments. We have also shown how at each stage the segmentation is improved by removing unwanted regions.

Our future work is to further develop the algorithm to track the cavity in the video sequence. Temporal information will be used to improve segmentation of the cavity by studying it in time. We will also investigate other ways to improve the segmentation results when they are not satisfactory, particularly when operating instruments are in the field of view of the endoscope. Furthermore, we need to look closely at illumination effects at the boundary of the cavity. The results of these works will then be applied to the 3D reconstruction of the cavity to estimate the relative distance between the instrument and the spinal cord in real time during the surgery.

References

[1] K.D.Kim, J.D.Babbitz and J.Mimbs, Imaging-guided costotransversectomy for thoracic disc herniation, Neurosurg Focus 9, Article 7, p.1-5, 2000.

[2] D.L.Pham, C.Xu and J.L.Prince, A Survey of Current Methods in Medical Image Segmentation, Annual Review of Biomedical Engineering, Annual Reviews, vol. 2, p. 315-337, 2000.

[3] D.R.Uecker, C.Lee, Y.F.Wang and Y.Wang, Automated Instrument Tracking in Robotically-Assisted Laparoscopic Surgery, Journal of Image Guided Surgery, Vol. 1, No. 6, p. 308-325, 1996.

[4] B.P.L.Lo, A.Darzi and G.Yang, Episode Classification for the Analysis of Tissue/Instrument Interaction with Multiple Visual Cues, MICCAI (1) 2003, p. 230-237, 2003.

[5] J.Boisvert, F.Cheriet and G.Grimard, Segmentation of Laparoscopic Images for Computer Assisted Surgery, SCIA03, p. 587-594, 2003.

[6] K.V.Asari, A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images, Medical Engineering & Physics, Volume 22, Issue 2, p. 89-96, 2000.

[7] U.Bockholt, W.Muller-Wittig, M.Chrupcala, H.Wang, G.Voss and A.Bisler, Classification and Segmentation of Malign Bladder Tissue in Endoscopic Images via Statistical Texture Analysis, Computer Graphik TOPICS, Vol. 15 Issue 5, p. 13-14, 2003.

[8] S.A.Karkanis, D.K.Iakovidis, D.E.Maroulis, D.A.Karras and M.D.Tzivras Computer Aided Tumor Detection in Endoscopic Video using Color Wavelet Features, IEEE Transactions on Information Technology in Biomedicine, vol. 7, no. 3, p. 141-152, 2003.

[9] J. Roubert, Automatic Guidance of a Laparoscope Using Computer Vision, Master thesis, Lund University, November 2002.

[10] T.Vlachos and A.G.Constantinides, A graph-theoretic approach to colour image segmentation and contour classification, Proc. Int. Conf. Image Processing and its Applications, p. 298-302, 1992.

[11] J.Shi and J.Malik, Normalized Cuts and Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, p. 888-905, 2000.

[12] O.J.Morris, M.J.Lee and A.G.Constantinides, Graph theory for image analysis: An approach based on the shortest spanning tree, in Proc. Inst. Elect. Eng., vol. 133, April 1986, p. 146–152.

[13] J. Mulroy, Video content extraction: Review of current automatic segmentation algorithms, Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS97), June 1997.

[14] S.Cooray, N.O'Connor, S.Marlow, N.Murphy and T.Curran, Semi-Automatic Video Object Segmentation using Recursive Shortest Spanning Tree and Binary Partition Tree, Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2001), May 2001.

[15] R.Piroddi and T.Vlachos, Object-based Segmentation of Moving Sequences using Multiple Features, IEEE Proc. Digital Signal Processing, Vol.2 , p. 547- 550, 2002.

[16] E.Tuncel and L.Onural, Utilization of the Recursive Shortest Spanning Tree Algorithm for Video-Object Segmentation by 2-D Affine Motion Modeling, IEEE transactions on circuits and systems for video technology, vol. 10, No. 5, p. 776-781 , 2000.

[17] P.F.Felzenszwalb and D.P.Huttenlocher, Efficient Graph-Based Image Segmentation, *IJCV(59)*, No. 2, p. 167-181, September 2004.

[18] M.Sonka, V.Hlavac and R.Boyle, Image Processing, Analysis and Machine Vision, 2nd edition, PWS Boston, 1998.

[19] Color space FAQ, http://www.neuro.sfc.keio.ac.jp/~aly/polygon/info/color-space-faq.html

[20] R.C.Gonzalez, R.E.Woods, Digital image processing, 2nd edition, Prentice Hall, 2002

[21] S.Chandran and K.K.Madheshiya, A Fast Segmentation Algorithm Revisited Proc. of Indian Conference on Computer Vision, Graphics, and Image Processing, December 2002.

[22] T.Brox, D.Farin and P.H.N.de With, Multi-stage region merging for image segmentation, in Proceedings of the 22nd Symposium on Information Theory in the Benelux, p.189-196, 2001.

[23] J.Xuan, T.Adali and Y.Wang, Segmentation of Magnetic Resonance Brain Image: Integrating Region Growing and Edge Detection, ICIP-C 95, pp. 544-547, 1995.

[24] M. Van Droogenbroeck and H. Talbot, Segmentation by adaptive prediction and region merging, In Digital Image Computing Techniques and Applications, Volume II, DICTA 2003, p. 561-570, 2003.

[25] T.Y.Young and K.S.Fu. Handbook of Pattern Recognition and Image Processing, Academic Press, 1986.

Summary

Thoracic diskectomy is a video-assisted thoracoscopic surgery (VATS) which mainly consists of removing the intervertebral disc at some levels of the spine to treat symptomatic thoracic herniation causing spinal cord and nerve root compression. During this procedure, instruments and an endoscope are inserted through small incisions on the patient's body and the surgeon performs the operation by viewing the images acquired by the endoscope on a video monitor. The endoscope acquires image sequences of the instruments and the surgical site while the surgeon is removing the intervertebral disc. The advantages of this procedure are minimal inconvenience to patients by reducing the exposure of anatomical organs and the blood loss.

Some challenges still need to be overcome to facilitate the adoption of this procedure for the treatment of scoliosis, because the surgeon needs a lot of training as he loses depth perception since the surgical site is viewed indirectly via the monocular endoscope. Furthermore, context depth cues are scarce as the lens of the endoscope is typically very close to the organs being imaged. For these reasons, computer vision techniques can assist surgeons by enhancing thoracoscopic images to avoid any damages to the spinal cord. To help surgeons, the goal of this collaborative work with Ste-Justine Children's Hospital in Montreal is to localize the boundary of the operated disc cavity, to allow its 3D reconstruction and positioning relative to operating instruments and a 3D model of the spine obtained from MRI images acquired before surgery.

To localize the boundary, our method first uses an efficient variant of the RSST algorithm to do an initial coarse segmentation, followed by similar graph-based segmentation using high-level criteria to further merge regions into more meaningful structures or objects. Experimental results on laparoscopic images of a thoracic discectomy procedure show that our method can segment the cavity from other tissues, although not perfectly, for a variety of challenging images. In contrast to graph-based segmentation methods using a single criterion, our multi-criteria merging method shows significant improvements on final results in our experiments. We have also show how at each stage the segmentation is improved by removing unwanted regions.

Legends

Fig. 1. Flow chart of the proposed method.

Fig. 2. Ground truth segmentation of cavity for sample test laparoscopic images. a) Image1, b) Image2, c) Image3.

Fig. 3. For Image1. a) Efficient RSST segmentation, K=100. b) Efficient RSST segmentation, K=8000. c) Efficient RSST segmentation, K=12000. d) Our algorithm from the segmentation of a). Region contours are superimposed on original image.

Fig. 4. For Image2. a) Efficient RSST segmentation, K=100. b) Efficient RSST segmentation, K=8000. c) Efficient RSST segmentation, K=12000. d) Our algorithm from the segmentation in a). Region contours are superimposed on original image.

Fig. 5. For Image1. Multistage segmentation. (a) Efficient RSST segmentation, K=100. (b) Segmentation stage two. (c) Segmentation stage three. (d) Segmentation stage four – final result.

Fig. 6. False segmentation of a laparoscopic image with an instrument for Image3. (a) Efficient RSST segmentation, K=100. (b) Multistage segmentation based on a).

Figure 1

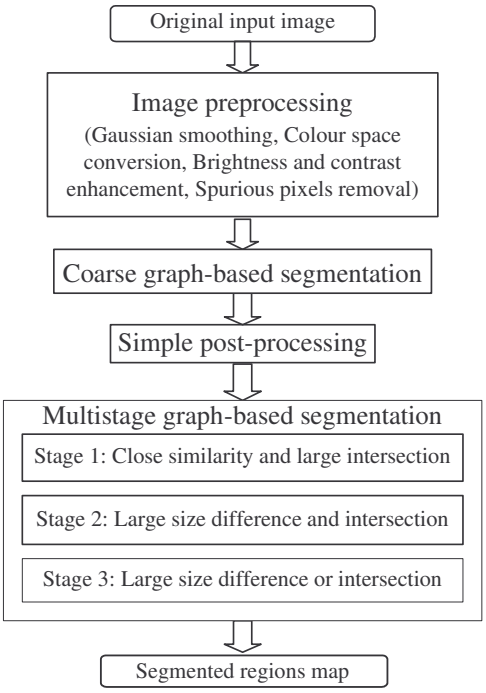

Original input image
Image preprocessing
(Gaussian smoothing, Colour space conversion, Brightness and contrast enhancement, Spurious pixels removal)
Coarse graph-based segmentation
Simple post-processing
Multistage graph-based segmentation
Stage 1: Close similarity and large intersection
Stage 2: Large size difference and intersection
Stage 3: Large size difference or intersection
Segmented regions map

Figure 2

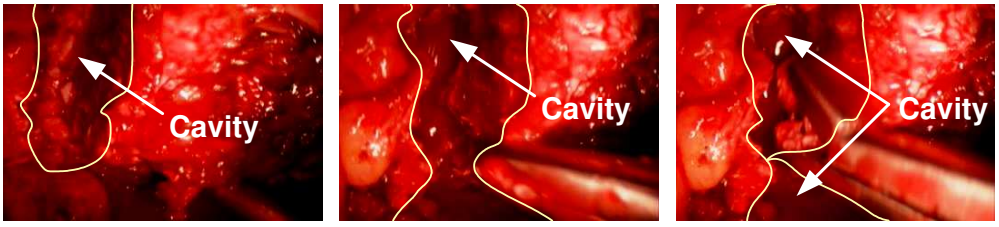


Figure 3

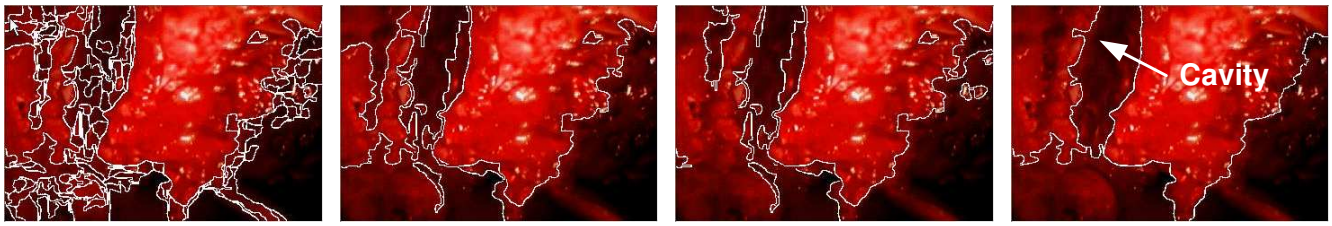


Figure 4

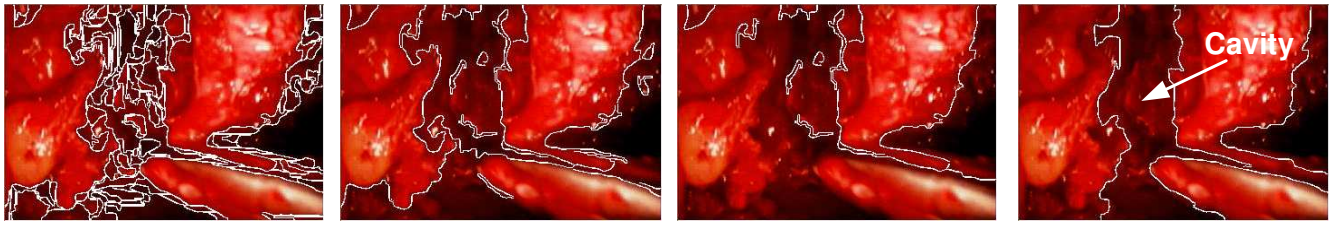


Figure 5

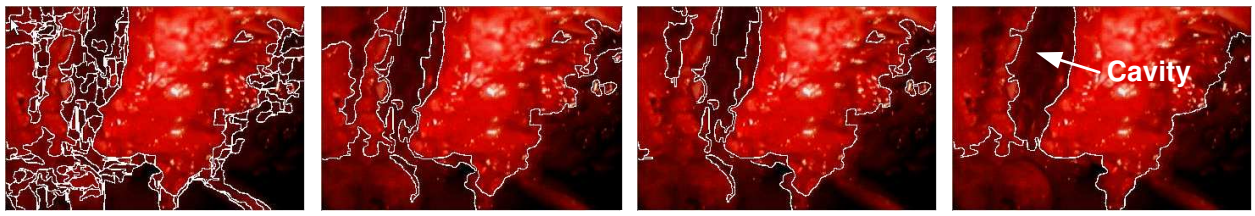
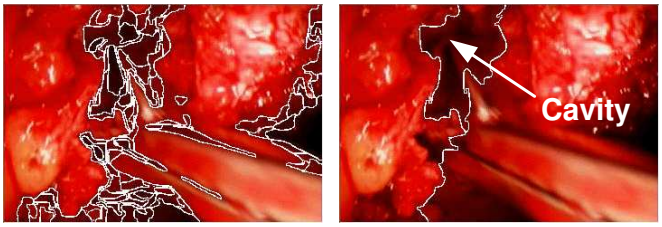


Figure 6



a) b)