Comparison of algorithms for ultrasound image segmentation without ground truth

Karan Sikka[a] and Thomas M. Deserno[b]

[a]Dept. of Electronics and Communication, Indian Institute of Technology Guwahati,
Guwahati, India
[b]Dept. of Medical Informatics, RWTH Aachen University, Aachen, Germany

ABSTRACT

Image segmentation is a pre-requisite to medical image analysis. A variety of segmentation algorithms have been proposed, and most are evaluated on a small dataset or based on classification of a single feature. The lack of a gold standard (ground truth) further adds to the discrepancy in these comparisons. This work proposes a new methodology for comparing image segmentation algorithms without ground truth by building a matrix called region-correlation matrix. Subsequently, suitable distance measures are proposed for quantitative assessment of similarity. The first measure takes into account the degree of region overlap or identical match. The second considers the degree of splitting or misclassification by using an appropriate penalty term. These measures are shown to satisfy the axioms of a quasi-metric. They are applied for a comparative analysis of synthetic segmentation maps to show their direct correlation with human intuition of similar segmentation. Since ultrasound images are difficult to segment and usually lack a ground truth, the measures are further used to compare the recently proposed spectral clustering algorithm (encoding spatial and edge information) with standard k-means over abdominal ultrasound images. Improving the parameterization and enlarging the feature space for k-means steadily increased segmentation quality to that of spectral clustering.

Keywords: Image processing, evaluation, segmentation, ultrasound

1. INTRODUCTION

Medical image analysis utilizes image segmentation as a preliminary step, followed by classification or delineation of the region of interest (ROI). Owing to the lack of a universal segmentation algorithm, many have been proposed in the recent years. These algorithms are evaluated on different and often insufficiently described parameters and data sets based on the discretion of the researchers, making their comparison difficult. This in turn increases the complexities associated with selection of an algorithm for a particular application. Thus, validation of algorithms (performance) has been gaining importance over the years.[1]

The existing evaluation techniques can be broadly classified into intra-evaluation and inter-comparison techniques.[2] The later is a conventional approach and employs a reference image called ground truth for making a region-wise comparison with the segmented result from the algorithm under consideration. This quantitative estimation is based on figures of merit that consider the number of classified and mis-classified pixels. On the other hand, inter-evaluation techniques require no a priori information and evaluate an algorithm by calculating certain goodness measures[3] that are based on the characteristics of good segmentation as put forward by Haralick and Shapiro.[4] The first method has been used for appraising a particular algorithm while the other method to rank different algorithms.

Ultrasound images are particularly difficult to segment due to the presence of speckle noise, artifacts and shadowing.[5] Hence, a number of problems exist with the evaluation of segmentation algorithms for ultrasound

Send correspondence to:
Prof. Dr. Thomas M. Deserno
Dept. of Medical Informatics, RWTH Aachen University, 52057 Aachen, Germany
E-mail: deserno@ieee.org, Telephone: +49 241 80 88793, Fax: +49 241 80 33 88793

Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment,
edited by David J. Manning, Craig K. Abbey, Proc. of SPIE Vol. 7627, 76271C
© 2010 SPIE · CCC code: 1605-7422/10/$18 · doi: 10.1117/12.844504

images. A critical problem among these is non-availability of a ground truth, due to which simulated ultrasound images are often employed to obtain a quantitative evaluation.[6] It is not possible to draw any definite conclusions about an algorithm since the simulated images (phantoms) can never accurately take into account all the conditions (imaging and anatomical) that exist during acquisition of clinical data.[7] Even the standard method of validation based on estimating disparity with a manually delineated image (ground truth) from a single expert suffers from human bias. A number of approaches have been proposed to tackle this issue by utilizing multi-expert ground truth data.[7, 8]

However, this process has high cost in terms of time and human-input. The comparison of two algorithms is often based on visual inspection[9] of their segmented results. This assessment may also be unreliable owing to its subjective nature. The most general approach for evaluation is comparison based on a single feature.[10] This partial assessment based on a local ROI may not serve as an accurate performance measure since the algorithm may be biased towards that particular feature. These mentioned approaches are also used in combination to shadow individual shortcomings. For instance, the quantitative comparison of two algorithms can be based on simulated ultrasound images, while qualitative assessment is carried out through visual inspection of segmentation results from real images.[6]

The visual examples presented in the scientific literature are often ambiguous, and the suggested improvements by novel approaches may also be due to sub-optimal parameterization of the established methods used for comparison. For instance, a clustering approach called spectral clustering that is based on graph theory is gaining popularity as an image segmentation algorithm,[11] and it was employed recently for ultrasound image segmentation.[9, 10] Analytically, it has been proved to be superior to k-means due to projection of data points in higher dimensional space, but applicability to segmentation of ultrasound is shown in the literature only for some (carefully) selected example images. Such an evaluation is not enough to draw any generalizations regarding the performance of algorithm, since the improved results may be attributed to the factors discussed earlier.

This paper presents a solution to tackle the discussed problems by introducing an innovative approach for comparing ultrasound image segmentation maps arising from different algorithms in absence of ground truth. The following section discusses the formulation of a matrix that is used to estimate two measures for computing similarity between the segmentation results. The introduced measures have been implemented on synthetic segmentation maps and also used to compare the segmentation results from spectral clustering and k-means (with modifications). Finally in Section 5, we discuss the effects of parametrization and feature space selection.

2. METHOD

The proposed distance metrics are based on the so-called region correlation matrix (RCM), which is defined first. We then define the measures, their combination, and prove the axioms of a quasi-metric. Spectral clustering of ultrasound is described as particular field of application for our methodology. Examples are based on ultrasound images of the bladder, after standardized preprocessing and field of view detection.

2.1 Region Correlation Matrix

RCM encodes the similarly information in the two maps. It is similar to a confusion matrix[2] and is unique for any two maps. Let $P(x, y) = g$, $g \in G$ denote an image, where G is the set of n bit gray scales with position coordinates x and y. $S(x, y) = l$, $l \in L$ denotes a segmentation of $P(x, y)$, where L is the set of labels for the partitioning map. Each pixel in a map S has a unique label, and the labels are consecutively used. Hence, $\forall l > 0, S(x, y) = l \Rightarrow \exists S(x, y) = l - 1$. We denote the region with $S_l = \bigcup_{S(x,y)=l}(x, y)$ and require a unique, non-overlapping partitioning with connected segments, i.e., $S_i \cap S_j = \emptyset, \forall i \neq j$ and $\bigcup_i S_i = P$.

Let S^1 and S^2 have been taken as segmentation results from two different algorithms to be compared. The number of segments A and B is given by

$$A = \max_{x,y} \left\{ S^1(x, y) \right\} \quad (1)$$

$$B = \max_{x,y} \left\{ S^2(x, y) \right\} \quad (2)$$

In case $A < B$, the RCM matrix M is given by

$$M_{IJ} = [M_{ij}] = \sum_x \sum_y \left\{ (S^1(x,y) = i) \cap (S^2(x,y) = j) \right\} \quad \forall i, j \in L \tag{3}$$

where I and J are the number of rows and columns of M, $I = \min\{A, B\}$ and $J = \max\{A, B\}$, respectively. When $A > B$, S^1 and S^2 are interchanged in (3). For $A = B$, S_1 is swept row-wise for comparing the member count and pixel position with S_2 till no new clusters are encountered or a difference is observed. Equation (3) is followed if the maximum member count corresponds to S^1 and vice-versa. If no difference is observed, the maps are identical.

The normalized version of the matrix M is denoted by $\tilde{M} = \frac{1}{N} M$, where the image size $N = \sum_j \sum_i M_{ij}$ is a global parameter that is independent of factors like region size in the segmentation map. $\tilde{M}$ contains the probabilities of overlap for different regions and is label independent.

2.2 Distance Measures

Based on $\tilde{M}$, we have defined two distance measures. The first measure E is called overlap index and measures the effective equivalence (overlap) for a region in the two maps S^1 and S^2

$$E = \sum_{i=1}^{I} \max_j \left\{ \tilde{M}_{ij} \right\} \tag{4}$$

There are two important implications:

1. E gives more weight to the overlap of dominant regions, and
2. mechanisms of the human visual system (HVS)[3] (intra-region uniformity and size of a region) are incorporated into the calculation of E to give results matching our intuition.

According to (4), E provides a measure of maximum overlap, which is important for post-segmentation procedures. For instance, a match between extracted features (e.g., texture, mean intensity) from two maps depends on the degree of ROI overlap. Since it considers all such ROIs, it is much more effective than the technique employing a particular ROI for comparison of two segmentation maps.

Since E does not take into account the non-equivalence and degree of division for a region, a second measure (fragment index (F)) has been introduced. To consider the extraneous (non-overlapping) regions, F employs a penalty factor $p_i \in \mathbb{N}$ that is equal to the number of non-zero elements in i^{th} row of $\tilde{M}$. It is defined as

$$F = \frac{\sum_{i=1}^{I} (p_i - 1)(1 - \max_j \{\tilde{M}_{ij}\})}{\sum_{i=1}^{I} (p_i - 1)} \tag{5}$$

In the mathematical formulation, F can be regarded as the expectation value of a random variable describing the probabilistic event of regions splitting into different numbers of extraneous regions. The weight factor also helps us to consider the HVS factor.

We have modified E to $\tilde{E} = 1 - E$ so that both $\tilde{E}$ and F yield the distance between the two segmentation maps being compared. The measures have been combined by taking a linear combination to obtain an overall distance

$$G = \frac{F\alpha + \tilde{E}\beta}{\beta + \alpha} \tag{6}$$

The parameters α and β are used to assign weights to the two measures in order to estimate a different overall measure G depending on the application domain and algorithm being compared. Based on the information that the overall distance between similar maps would be zero, percentage of similarity or dissimilarity can be obtained.

2.3 Quasi-Metric

A distance function D between features $a \neq b \neq c$ becomes a metric if it satisfies the following properties[12]

$$\begin{aligned} &\text{Reflexivity} && D(a,a) = 0 && (7)\\ &\text{Non-negativity} && D(a,b) > 0 && (8)\\ &\text{Symmetry} && D(a,b) = D(b,a) && (9)\\ &\text{Triangle inequality} && D(a,b) + D(b,c) \geq D(a,c) && (10) \end{aligned}$$

Any measure satisfying all the properties except the triangle inequality is called a quasi-metric.[13] In the following, we will analyze the measures defined in (4) and (5).

For perfect match, a complete overlap between the two maps, $\tilde{M}$ yields a diagonal square matrix. In the case of worst match, all the regions in first map split equally among regions present in other. Thus, all elements in $\tilde{M}$ are equal to $\frac{1}{AB}$. The regular match lies between the two extreme cases.

Since the matrix $\tilde{M}$ stores probabilities and is unique for any two segmentation maps, both the measures satisfy the non-negativity and symmetry axioms, (8) and (9), respectively. The modification of E to $\tilde{E}$ allows both $\tilde{E}$ and F to become 0 while comparing similar maps and hence satisfy the reflexivity axiom (7). Thus they can be characterized as quasi-metrics. This also holds for G in (6).

3. EVALUATION AND EXPERIMENTS

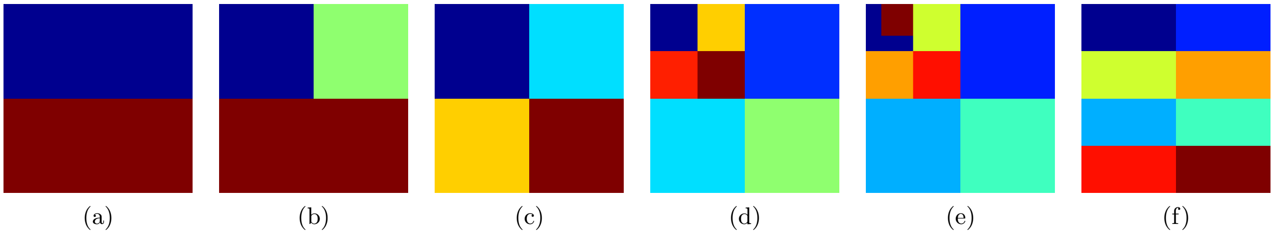
Experiments were performed on synthetic as well as clinical ultrasound data. In sthis section, we describe experimental setting and implementation.

3.1 Synthetic Segmentation Maps

Six synthetic segmentation maps have been employed to study the performance of the measures (Fig. 1). These maps have a simplified structure in order to allow the reader to correlate their values with visual observation. The images differ primarily in the number and orientation of the clusters present. Each image has been compared with all the other images present in the set to rank them on the basis of similarity. This experiment yields the response of the measure to either splitting (over-segmentation) or merging (under-segmentation) of different regions owing to difference in segmentation procedures. We have also considered the case of two images with similar count of clusters but different orientation to demonstrate the efficacy of the measures.

3.2 Application to Ultrasound Images

The introduced measures have been used to study the application of two clustering algorithms on ultrasound images of bladder and kidney (after pre-processing). The database has been taken from the IRMA framework[14] (Fig. 2a), which has been broadly applied for evaluation.[15,16] Since these ultrasound images lacked a well-defined ground truth, the measures (4) $\ldots$(6) were used to analyze the similarity of the results from the two segmentation algorithms under different conditions. The procedures for pre-processing, clustering algorithms and post-processing are discussed below.



(a) (b) (c) (d) (e) (f)

Figure 1: Segmentation maps with different clusters. (a): 2 clusters; (b): 3 clusters; (c): 4 clusters; (d): 7 clusters; (e) and (f): 8 clusters.

3.2.1 Pre-Processing

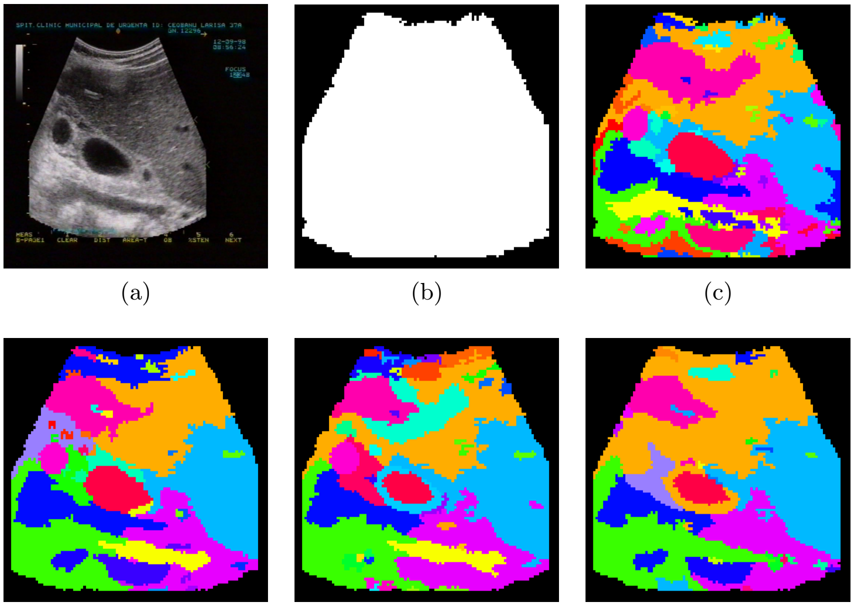
Ultrasound images were pre-processed to extract a field of view (mask image) and remove the pixel-decoded text information. First, the images were binarized based on an intensity threshold, followed by morphological erosion to remove the noisy pixels. Performing a dilation prior to the erosion prevented the diffusion of boundaries with the noisy pixels, which may be present near the edges. Finally connected components was executed to clear any remaining discontinuous region, with size below a threshold and merge with its background. This yielded the mask for an image containing the ROI and background. In order to reduce computation time, the background was cropped automatically (Fig. 2b) and these pixels were not included in the segmentation procedure.

3.2.2 Clustering Algorithms

The two clustering algorithms being compared are spectral clustering and k-means. The experiments used spectral clustering based on multiple eigenvector implementation[17] employing intensity, spatial and edge information (based on intervening contours[11]). Firstly, segmentation results from spectral clustering employing the three features have been compared with

1. k-means employing intensity information,
2. k-means employing both intensity and spatial information (using position coordinates with appropriate weight factor for each feature coordinate),
3. spectral clustering with different parameters (σ corresponding to each feature[17]),
4. mask image, and
5. single-label image (one region).

The last two comparisons are used to assess the performance of the two measures in case of under-segmentation and hence estimate an approximate working range.



(a) (b) (c)

(d) (e) (f)

Figure 2: Segmentation of ultrasound images. (a): original; (b) field of view obtained from pre-processing; (c): k-means clustering with intensity information; (d): k-means with intensity and spatial information; (e): spectral clustering; (f): spectral clustering with modified parameters.

Table 1: Results on synthetic images. $\tilde{E}$, F and similarity ranks are given on top, middle, and bottom rows.

Input	Fig. 1a	Fig. 1b	Fig. 1c	Fig. 1d	Fig. 1e	Fig. 1f
Fig. 1a	0 0 1	0.25 0.125 2	0.5 0.1667 3	0.5 0.2083 4	0.5 0.2143 5	0.75 0.3214 6
Fig. 1b	0.25 0.125 2	0 0 1	0.25 0.125 2	0.4375 0.1625 3	0.4375 0.1667 4	0.625 0.2292 5
Fig. 1c	0.5 0.1667 6	0.25 0.125 4	0 0 1	0.1875 0.1406 2	0.1875 0.15 3	0.5 0.1 5
Fig. 1d	0.5 0.2083 6	0.4375 0.1625 5	0.1875 0.1406 3	0 0 1	0.0278 0.0139 2	0.375 0.0938 4
Fig. 1e	0.5 0.2143 6	0.4375 0.1667 5	0.1875 0.15 3	0.0278 0.0139 2	0 0 1	0.375 0.0469 4
Fig. 1f	0.75 0.3124 6	0.625 0.2292 5	0.5 0.1 4	0.375 0.0938 3	0.375 0.0469 2	0 0 1

The first set of analysis has been complemented by providing similarity estimates of segmentation results from spectral clustering algorithm (with different features space) and corresponding case of k-means. These results have been employed to predict the correlation between the selection of features in different algorithms. Owing to the absence of a definite method for cluster center estimation, they were fed manually.

3.2.3 Post-Processing

All maps were subjected to post-processing for smoothing (neighborhood based) the clustered results and separating the disconnected labels. This resulted in better visualization, reduced dependency on initial cluster number, and uniformity of clustered results.

3.3 Implementation

The algorithms were implemented on Matlab version 7.3. Matlab employs a Fortran-based library ARPACK[18] for computation of eigenvectors of large matrices.

4. RESULTS

Table. 1 shows the distance measures and similarity ranks (based on percentage of similarity) for the synthetic maps of Fig. 1. The similarity order is in accordance with the visual observation. For three cases of Fig. 1a, the overlap parameter has the same value. In this situation, the correct order of similarity is estimated by F. The values of the measures are in accordance with the axioms defined earlier.

Fig. 2 shows an ultrasound image, the extracted field of view upon pre-processing, and the segmentation results from spectral clustering with different parameters and k-means with different feature coordinates. Table. 2 shows the percentage of similarity (estimated through overall distance (G) for comparing the segmentation maps from different algorithms, single-label image and mask image for ten test images. For the case of overall distance measure, we have taken $\alpha = 1$ and $\beta = 3$. The G between segmentation results from spectral clustering with similar parameters is 100 %, while the same measure returns a lower value on comparison with a different implementation (parameters) of spectral clustering. Secondly, the case of Fig. 2, the G between segmentation

Table 2: Percentage of similarity (Set 1).

Image	Single label	Mask image	Spectral clustering			K-means		
			initial	modified	Δ	intensity only	spatial inform.	Δ
Fig. 2a	28.07	47.47	100	92.74	7.26	93.69	95.23	1.54
2	26.68	46.49	100	94.47	5.53	88.17	94.04	5.87
3	34.23	51.20	100	95.17	4.83	89.69	94.31	4.62
4	20.15	26.17	100	90.38	9.62	87.12	92.48	5.36
5	36.79	67.36	100	93.50	6.50	85.04	89.77	4.73
6	34.29	50.71	100	93.90	6.10	89.26	92.92	3.66
7	37.39	58.53	100	93.69	6.31	95.58	96.54	0.96
8	27.06	45.78	100	95.91	4.09	90.99	94.28	3.29
9	38.23	63.11	100	97.36	2.64	85.25	90.00	4.75
Mean					5.88			3.86

Table 3: Percentage of similarity (Set 2).

Image	Spectral clustering			
	intensity only (modified)		spatial information (initial)	
	k-means intensity	k-means spatial	k-means intensity	k-means spatial
Fig. 2a	93.88	91.98	86.71	89.59
2	93.94	89.77	87.68	91.15
3	94.09	90.65	89.38	92.07
4	92.18	89.91	77.39	80.88
5	92.37	92.75	85.17	96.50
6	96.12	89.89	90.47	95.14
7	91.94	91.21	96.66	94.54
8	93.58	90.18	90.13	93.57
9	94.98	90.55	85.54	97.26
Mean	93.66	90.76	87.68	92.30

maps from spectral clustering and k-means (intensity information) is 93.69 %, while between spectral clustering and k-means combining spatial and intensity information is 95.23 %. This increase in G with inclusion of spatial information in k-means has been experimentally observed for all cases being considered. Finally, the measure yields the lowest estimate on comparison with a single-label image.

Table. 3 shows the similarity percentages obtained on comparing the segmentation results from spectral clustering with modified features to the respective instance of k-means. It is evident from these observations that the similarity between segmentation maps with similar features is more than those with different features.

5. DISCUSSION

The inference that can be drawn from the previous observations is that the segmentation results are indeed affected by the feature space and modifications in runtime parameters. In the present case, the change in similarity between the results from two algorithms can be attributed to the tendency of spatial information, added later to k-means, to minimize the effect of noise present prior to clustering. This not only improves the uniformity in the results (visually evident in Fig. 2c as compared to Fig. 2d), but also increases its similarity to the results from spectral clustering, that already encodes spatial information. Moreover, as a result of different parametrization, the similarity in the results from the same algorithm is less than 100%. Our contention is further strengthened by the results in Table. 3 that highlight the correlation between the feature space and corresponding results from different segmentation algorithms.

Though spectral clustering has been known to give promising results, it has many disadvantages such as a high computation cost, parameter dependences,[19] limited size of input image due to large weight matrix, and convergence issues for the intermediate eigenvalue estimation.[20] Moreover, application of k-means in the last step[17] makes it susceptible to the disadvantages of k-means (convergence issue).

Based on the results and above discussions, we cannot conclude the superiority of spectral clustering over k-means for the case of ultrasound images. The fact that the evaluation methods of previous researches pertaining to this domain were inadequate (as discussed in Sec.1), further supports our contention. Although a number of independent researches have been conducted recently to counter the shortcomings of spectral clustering,[21, 22] they have still not been addressed effectively in the context of ultrasound images.

As mentioned earlier, the selection of an algorithm for a particular application has become a complex issue. Novelty of an algorithm cannot be the sole basis of decision, other factors like complexity, parameters involved and their optimization, and relevance of features should also be considered. Thus, construction of a cost function to rank different algorithms for an application remains a major challenge in the field of medical image segmentation.

6. CONCLUSION

We have devised a new methodology for comparing segmentation maps from different algorithms without the application of ground truth. The segmentation algorithms have been considered for ultrasound images, since these images usually lack ground truth and the inherent visual complexities along with limited colors for discrimination make visual assessment of the segmentation results impractical. Two measures, satisfying the axioms of a quasi-metric, have been introduced. They are supplemented by empirical and analytical proofs to demonstrate the consideration of human vision properties in their formulation. Thereafter, a comparison between novel spectral clustering and simple k-means is presented using abdominal ultrasound images. Our experiments show that using optimized parameters and features, standard k-means tends to match the results from spectral clustering. In conclusion, we cannot confirm superiority of spectral clustering over enhanced k-means when applied to ultrasound image segmentation.

7. ACKNOWLEDGMENT

The present work was completed as a part of research internship for international undergraduate students. The authors wish to thank DAAD, the German Academic Exchange Service for sponsoring the program.

REFERENCES

[1] Udupa JK, LeBlanc VR, Zhuge Y, et al. A framework for evaluating image segmentation algorithms. Comput Med Imaging Graph. 2006;30(2):75–87.

[2] Zhang YJ. A survey on evaluation methods for image segmentation. Pattern Recognit. 1996;29(8):1335–46.

[3] Zhang H, Fritts JE, Goldman SA. Image segmentation evaluation: A survey of unsupervised methods. Comput Vis Image Underst. 2008;110(2):260–80.

[4] Haralick RM, Shapiro LG. Image segmentation techniques. Computer Vis Graph Image Process. 1985;29(1):100–32.

[5] Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. IEEE Trans Med Imaging. 2006;25(8):987–1010.

[6] Yu JH, Wang YY, Chen P, Xu HY. Two-dimensional fuzzy clustering for ultrasound image segmentation. IEEE/ACM Trans Comput Biol Bioinform. 2007;p. 599–603.

[7] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004 July;23(7):903–21.

[8] Gordon S, Lotenberg S, Long R, Antani S, Jeronimo J, Greenspan H. Evaluation of uterine cervix segmentations using ground truth from multiple experts. Comput Med Imaging Graph. 2009;33(3):205–16.

[9] Chang-ming Z, Guo-chang G, Hai-bo L, Jing S, Hualong Y. Segmentation of ultrasound image based on texture feature and graph cut. Comput Sci Softw Eng. 2008;1:795–8.

[10] Archip N, Rohling R, Cooperberg P, Tahmasebpour H, Warfield SK. Spectral clustering algorithms for ultrasound image segmentation. In: Proc MICCAI. vol. 8; 2005. p. 862–9.

[11] Malik J, Belongie S, Leung T, Shi J. Contour and texture analysis for image segmentation. Int J Computer Vis. 2001;43(1):7–27.

[12] Restle F. A metric and an ordering on sets. Psychometrika. 1959;24(3):207–20.

[13] Agarwala R, Bafna V, Farach M, Narayanan B, Paterson M, Thorup M. On the approximability of numerical taxonomy (fitting distances by tree metrics). In: SODA '96: Proceedings of the seventh annual ACM-SIAM symposium on Discrete algorithms. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; 1996. p. 365–72.

[14] Lehmann TM, Wein BB, Dahmen J, Bredno J, Vogelsang F, Kohnen M. Content-based image retrieval in medical applications: a novel multistep approach. Proc SPIE. 1999;3972:312–320.

[15] Deselaers T, Muller H, Clough P, Ney H, Lehmann T. The CLEF 2005 automatic medical image annotation task. Int J Computer Vis. 2007;74(1):51–8.

[16] Deselaers T, Deserno T, Muller H. Automatic medical image annotation in ImageCLEF 2007: overview, results, and discussion. Pattern Recognit Lett. 2008;29(15):19881995.

[17] Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems 14. MIT Press; 2001. p. 849–56.

[18] Lehoucq RB, Sorensen DC, Yang C. ARPACK users' guide: Solution of Large-scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. United States of America: Society for Industrial and Applied Mathematics (SIAM); 1998.

[19] von Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007;17(4):395–416.

[20] von Luxburg U, Belkin M, Bousquet O. Consistency of spectral clustering. Ann Stat. 2008;36:555.

[21] Yu SX, Shi J. Multiclass Spectral Clustering. In: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. IEEE Computer Society; 2003. p. 313.

[22] Song Y, Chen WY, Bai H, Lin CJ, Chang EY. Parallel Spectral Clustering. In: ECML PKDD '08: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases - Part II. Springer-Verlag; 2008. p. 374–89.