BioMedical Engineering OnLine

RESEARCH   **Open Access**

Histological image segmentation using fast mean shift clustering method

Geming Wu[1†], Xinyan Zhao[2†], Shuqian Luo[1] and Hongli Shi[1*]

* Correspondence: shl@ccmu.edu.cn
†Equal contributors
[1]School of Biomedical Engineering, Capital Medical University, Beijing, China
Full list of author information is available at the end of the article

Abstract

Background: Colour image segmentation is fundamental and critical for quantitative histological image analysis. The complexity of the microstructure and the approach to make histological images results in variable staining and illumination variations. And ultra-high resolution of histological images makes it is hard for image segmentation methods to achieve high-quality segmentation results and low computation cost at the same time.

Methods: Mean Shift clustering approach is employed for histological image segmentation. Colour histological image is transformed from RGB to CIE L*a*b* colour space, and then a* and b* components are extracted as features. To speed up Mean Shift algorithm, the probability density distribution is estimated in feature space in advance and then the Mean Shift scheme is used to separate the feature space into different regions by finding the density peaks quickly. And an integral scheme is employed to reduce the computation cost of mean shift vector significantly. Finally image pixels are classified into clusters according to which region their features fall into in feature space.

Results: Numerical experiments are carried on liver fibrosis histological images. Experimental results demonstrate that Mean Shift clustering achieves more accurate results than k-means but is computational expensive, and the speed of the improved Mean Shift method is comparable to that of k-means while the accuracy of segmentation results is the same as that achieved using standard Mean Shift method.

Conclusions: An effective and reliable histological image segmentation approach is proposed in this paper. It employs improved Mean Shift clustering, which is speed up by using probability density distribution estimation and the integral scheme.

Keywords: Clustering, Colour image segmentation, Mean shift, Histological image processing

Background

In recent years, with the increasing demands of quantitative analysis, digital image processing techniques attract more and more attention in histopathology [1,2]. They are considered to be more reliable than traditional manual assessment which heavily depends on the operator's experience and usually can not be reproduced. Image segmentation serves as a fundamental and key technique and is typically the first step in digital image analysis. In histological image analysis, it partitions a digitized histological image into multiple homogeneous or similar

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 2 of 12

regions which corresponding to different tissues or cellular components. As the result of image segmentation is taken as the input of the successive processing steps, it is essential and critical to the quality of final result of qualitative image analysis.

Histological image segmentation is much difficult due to the complex approach to make histological images. Histological specimens are fixed, processed, embedded, sectioned and then stained into different colours in different tissues. Coloured histological sections provide anatomical details for the diagnosis and histopathology, and bring out special features of different tissues at the microscopic level. To identify different tissues or cellular components, histological sections are segmented according to colour, shape or texture features after acquired with high-resolution digital camera, and then classified by commonly employing supervised methods [3-6]. Several methods based on digital image processing and pattern recognition techniques have been proposed to deal with histological image segmentation problem in the past years [7-12]. When no training data set is available in histological image analysis, feature vectors or points representing pixels in the histological image are usually extracted from their local properties. Then unsupervised techniques are used to label pixels into different clusters by separating feature vectors in the feature space. There are many existing approaches in literature which can be employed for segmenting histological image. These include clustering based methods (k-Means [13] and Mean Shift [14]), mixture models based schemes (i.e. Gaussian mixture models based Expectation-Maximization clustering, GMM-EM [15]) and state-of-the-art energy minimization based approaches (graph-cuts methods [16-18] and Markov Random Field based method [19]). However, histological image segmentation still faces several technique challenges. One of them is how to achieve high-quality segmentation results with low computation cost, especially for sequential histological image analysis. The size of acquired histological image is usually very large for better investigation of microstructures. It makes most of existing image segmentation algorithms, such as Mean Shift, very time-consuming and hard to be used in practice. Mean Shift is a non-parametric clustering approach which has no assumptions on the shape of the distribution and the number of clusters. So Mean Shift may achieve better segmentation results than model-based clustering schemes when it is used as a histological image segmentation method.

In this paper, we focus on pixel-level segmentation by colours in histological image with unsupervised method. A fast Mean Shift clustering approach is proposed and applied to segment histological images properly. The new method estimates the probability density distribution in advance and then separates the feature space into different regions by employing Mean Shift to find the density peaks. Feature points are divided into clusters according to the region that they fall into. And an integral scheme is employed to speed up the computation of mean shift vector. To apply the proposed method to histological image segmentation, CIE L*a*b* colour space is used as feature space.

The rest parts of this article are organized as follows: Section 2 describes the main ideas and schemes employed to speed up Mean Shift clustering approach in details. The framework of histological image segmentation based on the proposed

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 3 of 12

method is also given in this section. Section 3 shows numerical experimental results from live fibrosis histological images, and followed by the study conclusion in Section 4.

Methods

Feature extraction in L*a*b* color space

To obviously identify different tissues in histological specimens, they are stained into different colours. Digitized histological images are usually acquired and stored in RGB. CIE L*a*b* colour space is a non-linear transformation of RGB, and models all visible colours approximate to human vision. So the simple Euclidean distance in L*a*b* space could differentiate among colours perceptually. We convert histological images from RGB space to L*a*b* space and only extract a* and b* components as features for clustering to ignore variations in brightness. Thus histological image segmentation becomes a 2D clustering problem. Denote V_{ab} the feature space consisting of a* and b* components.

Fast mean shift clustering

Given a set of points $\{x_1, x_2, \ldots x_n\}$ in V_{ab}, mean shift vector (MSV) with uniform kernel is defined as follows

$$m_h(x) = \frac{1}{K_S} \sum_{x_i \in S_h(x)} (x_i - x), \quad \forall x \in V_{ab},$$ (1)

where S_h is a sphere with center x and radius h, and K_s is the number of points located in S_h. h is termed the window size.
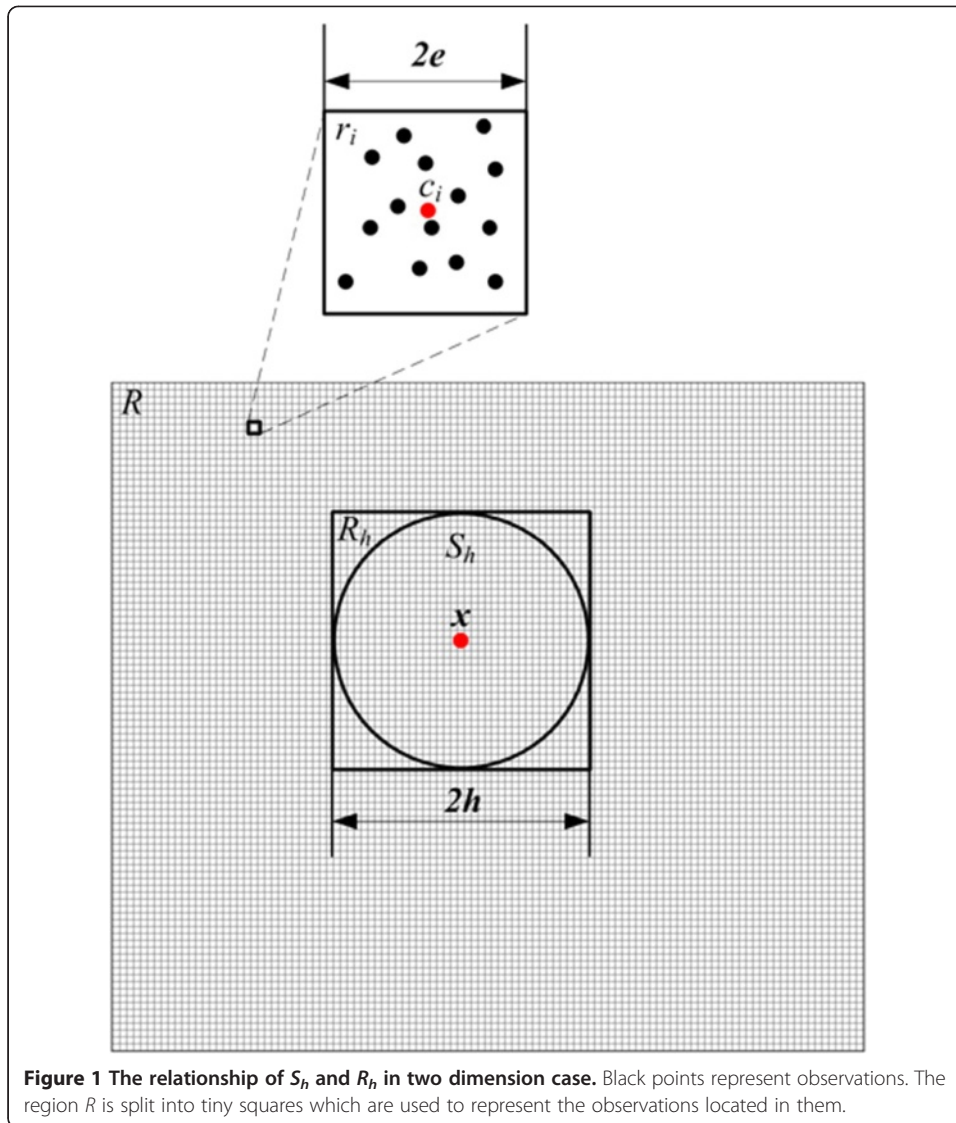
Standard Mean Shift clustering method employs an iterative gradient ascent procedure to estimate local density. For a given point x in the feature space, it sets $x \leftarrow x + m_h(x)$ and repeats this step until convergence. The stationary points of this procedure represent the modes of the underlying distribution. Points associated with the same stationary point are considered as members of the same cluster.

Mean shift vector estimation

To find out the mean shift vector at x, one has to compute the distances from all points to the specified centre x and selects out those located in S_h. Thus the computational cost of mean shift vector is very high and results in standard Mean Shift a very slow clustering approach.

Notice that all points are located in a square R in V_{ab}. Split R into tiny squares $\{r_1, r_2, \ldots, r_m\}$ whose sides have the length of $2e$. For each tiny square r_i, denote the frequency of points in it by w_i and the center by c_i. Denote the closest external square of S_h by R_h which consists of tiny squares (Figure 1 illustrates the relationship of S_h and R_h in V_{ab}). We replace R_h with R_h and approximate m_h at x with

$$\hat{m}_h(x) = \frac{1}{K_R} \sum_{x_i \in R_h(x)} (x_i - x).$$ (2)

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 4 of 12

Figure 1 The relationship of S_h and R_h in two dimension case. Black points represent observations. The region R is split into tiny squares which are used to represent the observations located in them.

Here, K_R is the number of observations located in R_h, and now h is half of the side length of R_h in what follows. Considering the definition of w_i, we can rewrite (2) as

$$\hat{m}_h(x) = \frac{\sum\limits_{x_i \in R_h(x)} x_i}{\sum\limits_{r_i \in R_h(x)} w_i} - x. \tag{3}$$

Given $e \ll h$, we have a further approximation

$$\hat{m}_h(x) \approx \frac{\sum\limits_{r_i \in R_h(x)} w_i c_i}{\sum\limits_{r_i \in R_h(x)} w_i} - x. \tag{4}$$

If w_i and c_i are computed in advance and stored in the frequency matrix w and the centre matrix c respectively, R_h corresponds to a sub-region of these matrices. Thus

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 5 of 12

for a given x, it is not needed to find out the points located in S_h by computing the distances from all points to the x. So the computation cost of $\hat{m}_h$ is much lower than that of m_h.
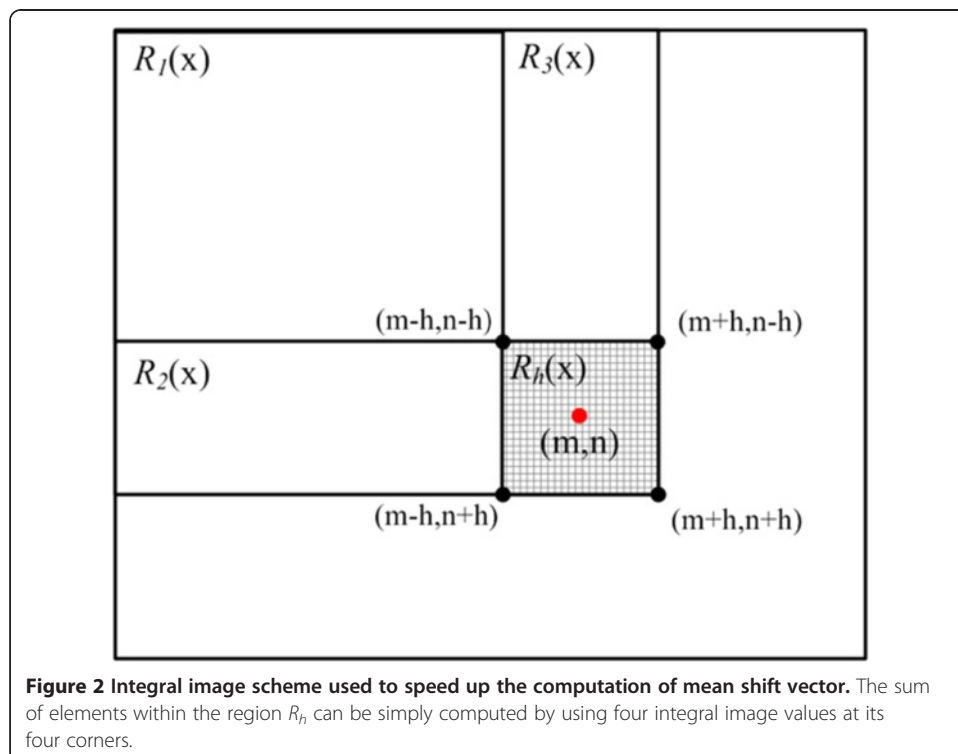
Using of integral image scheme

We employ the above method to estimate mean shift vector in a*b* colour space. For the convenience of description, we use two subscripts to index variants defined in the above section. Rewrite Eq. (4) as follows

$$\hat{m}_h(x) \approx \frac{\displaystyle\sum_{\substack{m-h\leq i\leq m+h \\ n-h\leq j\leq n+h}} w(i,j)c(i,j)}{\displaystyle\sum_{\substack{m-h\leq i\leq m+h \\ n-h\leq j\leq n+h}} w(i,j)} - x, \tag{5}$$

where (m,n) satisfies $\|c(m,n) - x\|_\infty \leq e$.

Notice that the sum operations in Eq. (5) are over all elements in a specified squared sub-region. So the integral image scheme can be employed to speed up the computation of $\hat{m}_h$. Integral image, also known as summed area table, was proposed for texture mapping in [20] and then widely used in pattern recognition and image processing [21,22]. It can be used to rapidly calculate summations over sub-regions of an image. The value at location (i,j) in the integral image is defined as the sum of all elements within the top and left side of (i,j). As Figure 2 shows, the sum of elements within the region R_h can be simply computed by using four integral image values at its four corners.

Figure 2 Integral image scheme used to speed up the computation of mean shift vector. The sum of elements within the region R_h can be simply computed by using four integral image values at its four corners.

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 6 of 12

To employ the integral image scheme, we define

$$W(k,l) = \sum_{i \le k, j \le l} w(i,j) \tag{6}$$

and

$$M(k,l) = \sum_{i \le k, j \le l} w(i,j)c(i,j). \tag{7}$$

Then we obtain

$$\sum_{\substack{m-h \le i \le m+h \\ n-h \le j \le n+h}} w(i,j)c(i,j) = M(m+h, n+h) + M(m-h, n-h) - M(m+h, n-h) - M(m-h, n+h) \tag{8}$$

and

$$\sum_{\substack{m-h \le i \le m+h \\ n-h \le j \le n+h}} w(i,j) = W(m+h, n+h) + W(m-h, n-h) - W(m+h, n-h) - W(m-h, n+h). \tag{9}$$

Thus mean shift vector can be approximately computed as following

$$\hat{m}_h(x) \approx \frac{M(m+h, n+h) + M(m-h, n-h) - M(m+h, n-h) - M(m-h, n+h)}{W(m+h, n+h) + W(m-h, n-h) - W(m+h, n-h) - W(m-h, n+h)} - x. \tag{10}$$

The above formula shows that the computation cost of mean shift vector is no longer related to the number of points and can be performed in constant time.

Histological image segmentation framework

Based on the above discussion, the framework of the proposed fast mean shift method (FMShift) for histological image segmentation is summarized as follows:

Algorithm: Fast mean shift clustering for histological image segmentation

Initialization:

Input color histological image I, e and h ($\gg e$).

Main:

Step 1: Transform I from RGB to L*a*b* color space, and extract a* and b* color components to build a feature space V_{ab}. Then separate the feature space into tiny squares with side length of $2e$, and compute the frequency matrix w of feature points and center matrix c.

Step 2: Compute W and M by using Eq. (6) and (7). Employ Mean Shift approach to separate the feature space into different regions by finding the density peaks.

Step 3: Divide pixels of histological image into clusters according to which region their corresponding feature point falls into.

Output:

Segmentation labels

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 7 of 12

Experiments and discussion

Quantitation of connective tissue and collagen is important in the assessment of fibrosis progression in chronic liver diseases. To evaluate our proposed method, a liver histological specimen from a Wistar rat, in which liver fibrosis was induced by albumin antigen-antibody complex, was used for the evaluation. With the use of Masson's trichrome staining, connective tissue and collagen were stained in light blue while smooth muscle in red and nucleus of hepatocytes in light dark. And blood vessels and sinusoids are in white. As shown in Figure 3, twenty histological images in 24-bit RGB colour were obtained by digitizing several sections from this specimen at different regions with an objective magnification of 20x. The size of each image is 1280×800.

k-Means++ [23], GMM-EM, Mean Shift, Kernel Graph-cuts (KGC) [18], hidden Markov Random Field with Expectation-Minimization algorithm (HMRF-EM) [19] and the proposed FMShift method had been implemented using MATLAB (The MathWorks, Inc., Natick, MA, USA), and were employed for histological image segmentation. a* and b* colour components of histological images were used as input features for previous three methods and FMShift while RGB for KGC and HMRF-EM. For the convenience of comparisons, the number of clusters k was given in prior. The clustering results of FMShift and Mean Shift were sorted by the number of assigned points in descent order. Then only the top k clusters were kept and others were reassigned into the nearest cluster. Considering that the objective of the segmentation is to identify fibrosis and vessels from other liver tissues, we set $k = 3$ in all experiments. In
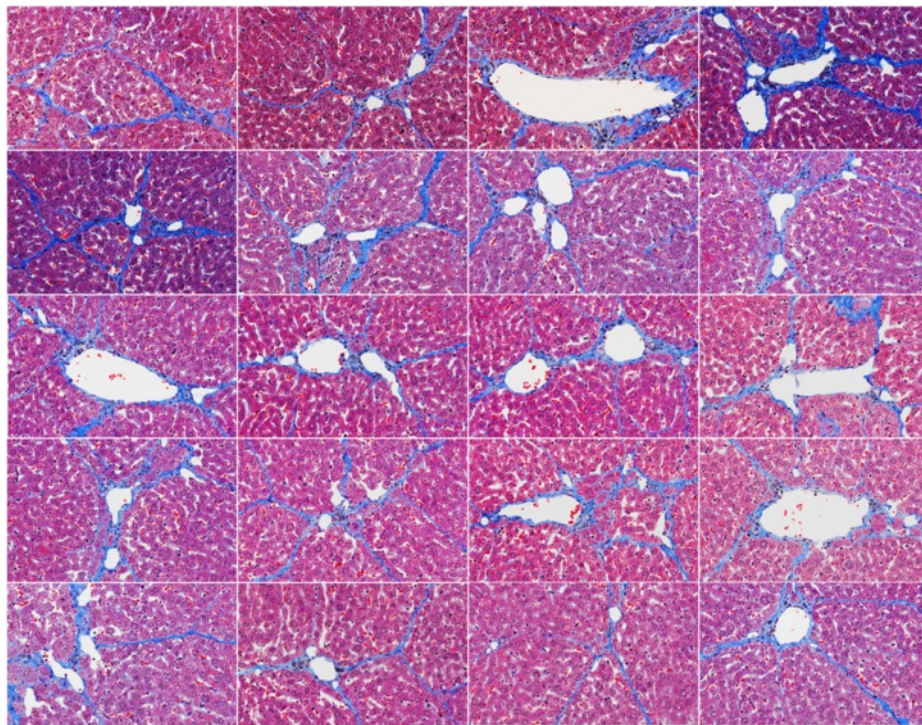

Figure 3 Original liver fibrosis histological images for performance evaluation. Twenty histological images obtained for performance evaluation by digitizing several sections from a liver histological specimen of a Wistar rat.

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 8 of 12

FMShift method, we set the parameter e to one thousandth of maximum range of a* and b* colour components, and h to 40 times e. Radial basis function with $\sigma = 0.5$ was used as the kernel function in KGC method. And the number of iterations of EM algorithm in HMRF-EM and GMM-EM is limited up to 100.

For quantitative comparison, we took manually segmentation results as references, and employed three performance metrics: Dice index, Rand index [24] and Variation of Information [25]. Dice and Rand indexes were used to measure the similarity between segmentations obtained by using different methods and their corresponding references. Denote S the segmentation result and R the reference. For a given pixel x_i of the segmented image, it is labelled with l_i and l'_i respectively in S and R. The Dice Index (DI) is defined as the size of the intersection divided by the average size of S and R as follows

$$DI(S,R) = \frac{2|S \cap R|}{|S| + |R|}. \tag{11}$$

And the Rand Index (RI) is defined as the ratio of the number of pairs of pixels which have a compatible label relationship between S and R, and is computed as follows:

$$RI(S,R) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} \left[\mathbf{I}\left(l_i = l_j \wedge l'_i = l'_j\right) + \mathbf{I}\left(l_i \neq l_j \wedge l'_i \neq l'_j\right) \right]. \tag{12}$$

where $\mathbf{I}$ is the identity function. The Variation of Information (VoI) [25] is an information-based performance metric and can be used to measure the distance between S and R. VoI satisfies

$$VoI(R,S) = H(R) + H(S) - 2I(R,S), \tag{13}$$

where $H(X)$ is the entropy of X, and $I(X,Y)$ is the mutual information between X and Y. Larger values of DI and RI and a smaller value of VoI mean higher segmentation accuracy.

Figure 4 shows the segmentation results of four histological images selected from the dataset by using k-Means++, GMM-EM, KGC, HMRF-EM, Mean Shift and our proposed FMShift respectively. Fibrosis, vessels (including sinusoids) and other tissues are represented in light blue, white and lavender respectively. Table 1 illustrates the average accuracies with standard deviation over segmentation results obtained from twenty histological images using different methods. The performance of six segmentation methods is quantified by Rand Index and Variation of Information for global evaluations. And the segmentation results of fibrosis and vessels are also measured by Dice Index independently at the same time. Table 2 summaries the average computation times for each method. All the above results are obtained with a standard Windows computer equipped with a 2.4 GHz Intel Core i5 processor and 8 GB RAM.

It is shown in Figure 4 that the tissues adjacent to collagen tend to be wrong segmented as fibrosis using HMRF-EM and KGC. The possible reason is that the shape of collagen is thin and elongated. Graph cut based image segmentation method has the problems to segment them due to shrinkage bias, and HMRF-EM uses spatial information through the mutual influences of neighbours. So it results in small values of Dice
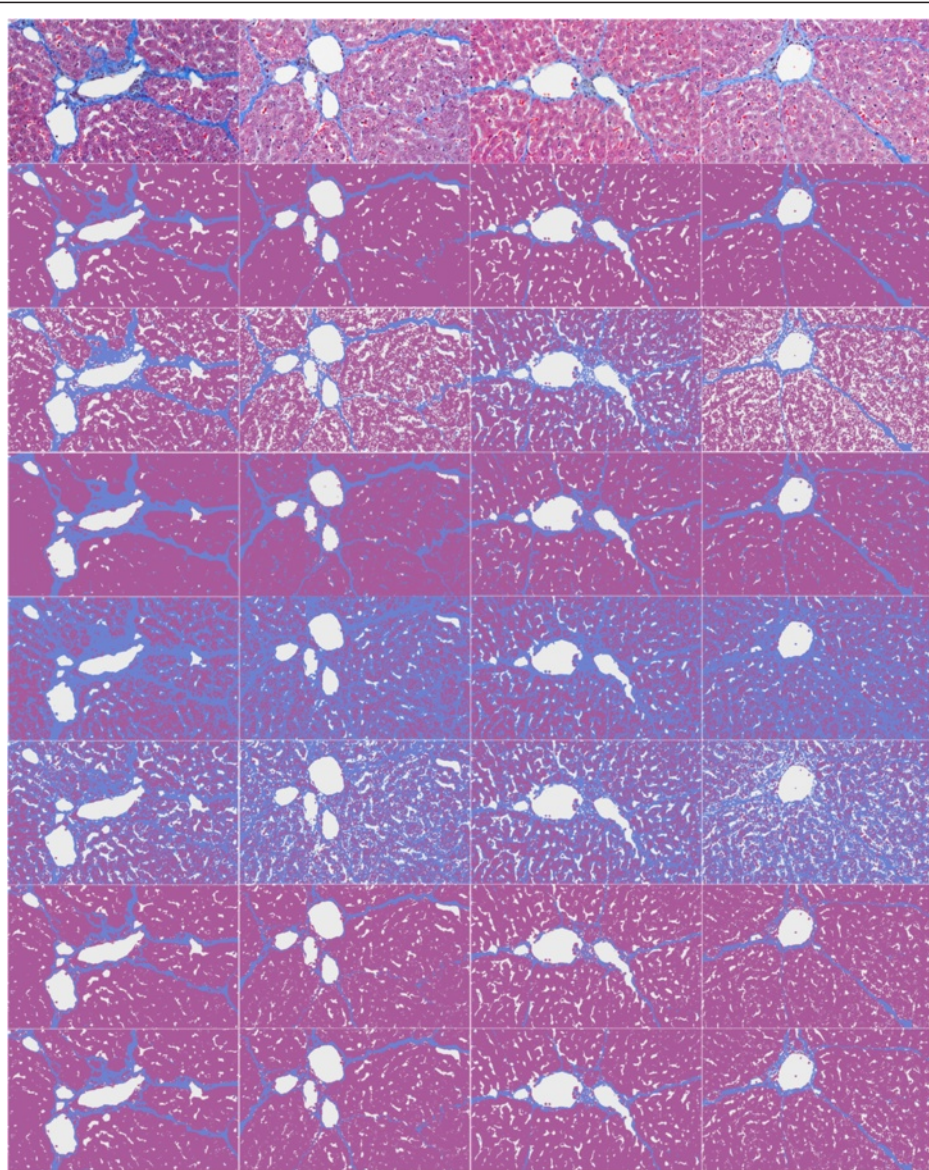
Wu et al. BioMedical Engineering OnLine (2015) 14:24

Page 9 of 12

Figure 4 Segmentation results of four liver fibrosis histological imageS by using k-Means, GMM-EM, HMRF-EM, KGC, Mean Shift, and FMShift. Four columns are corresponding to four histological images. Fibrosis, vessels and other tissues are represented in light blue, white and lavender respectively. From top to bottom are original histological images, manual segmentations, and segmentations using k-Means, GMM-EM, HMRF-EM, KGC, Mean Shift and the proposed FMShist method.

Table 1 Comparison of segmentation accuracies

Method	Dice index		Rand index	Variation of information
	Fibrosis	**Vessels**		
k-Means	0.67 ± 0.24	0.52 ± 0.24	0.69 ± 0.08	1.25 ± 0.20
GMM-EM	0.60 ± 0.13	0.68 ± 0.22	0.76 ± 0.10	1.03 ± 0.29
HMRF-EM	0.28 ± 0.14	0.76 ± 0.22	0.62 ± 0.10	1.35 ± 0.24
KGC	0.23 ± 0.10	0.69 ± 0.19	0.55 ± 0.08	1.62 ± 0.16
Mean Shift	0.87 ± 0.05	0.84 ± 0.09	0.91 ± 0.03	0.53 ± 0.12
FMShift	0.86 ± 0.05	0.84 ± 0.07	0.91 ± 0.02	0.54 ± 0.12

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 10 of 12

Table 2 Comparison of average computation times

Method	k-Means++	GMM-EM	HMRF-EM	KGC	Mean Shift	FMShift
Avg. times (sec.)	4.2	117.0	954.2	14.8	732.3	6.6

Indexes for fibrosis segmentation by using HMRF-EM and KGC methods and also decreases the Rand Indexes as listed in Table 1. At the same time, k-Means method can successfully detect spherical clusters with similar size but is not a good scheme for others. This limitation makes it easy to assign points to wrong clusters. Thus the tissues adjacent to sinusoids are segmented wrong by k-Means++ as shown in Figure 4. GMM-EM is based Gaussian mixture models and is sensitive to the distribution of feature points. However, this requirement is usually hard to meet in the case of histological image segmentation. So GMM-EM does not achieve high Dice and Rand indexes as shown in Table 1.

The quantitative results listed in Tables 1 and 2 show that the speed of FMShift method is comparable to that of k-Means++ and much faster than other schemes including standard Mean Shift clustering approach while the segmentation accuracies obtained by FMShift and Mean Shift are almost the same but much better than that of other methods. By estimating the probability density distribution in advance and employing the integral image scheme, FMShift reduces the computation cost of standard Mean Shift clustering significantly. And because histological images are stained manually, the range of colour components is restricted. This makes the frequency matrix w sparse and also speeds it up to find the density peaks in FMShift. The above reasons make FMShift a very fast approach even in handling large-scale histological image segmentation problems. At the same time, FMShift is Mean Shift scheme based. So it has no demands on the shape of underlying distribution and thus achieves high accuracies in liver histological image segmentation.

Conclusions

We have developed a histological image segmentation approach by employing improved Mean Shift clustering. To eliminate illumination variations, colour histological image is transformed into CIE L*a*b* colour space, and then a* and b* components are extracted as features for clustering. The clustering approach consists of three steps. In the first step, the probability density distribution is estimated by splitting the effective feature space located with observations into tiny squares and computing the frequencies of observations occurring in each square. The second step is to separate the feature space into different regions by employing Mean Shift scheme to finding density peaks. Then all observations are assigned into different clusters according to which square they fall into in the last step. And an integral image scheme is used to speed up the computation of mean shift vector at the same time. By employing the probability density estimation and integral scheme, the computation cost of standard Mean Shift clustering method is significantly reduced while keeping the accuracy the same. From the results of numerical experiments on liver fibrosis histological images, we have the conclusion that the proposed method is a fast and reliable approach for color image segmentation, especially for large-scale histological image segmentation.

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 11 of 12

In this paper, we only use two dimensional color features of histological images and discuss how to accelerate Mean Shift method in the two-dimension case. Our future work includes extending fast Mean Shift scheme to high dimensional case and taking more features, such as shape and texture, into consideration for better performance.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The work presented here was carried out in collaboration between all authors. GMW, XYZ, SQL and HLS defined the research theme, and GMW designed the methods. XYZ carried out the experiments, and GMW analyzed the data, interpreted the results and wrote the paper. All authors read and approved the final manuscript.

Author details

[1]School of Biomedical Engineering, Capital Medical University, Beijing, China. [2]Liver Research Center, Beijing Friendship Hospital, Capital Medical University, Beijing, China.

References

1. Madabhushi A. Digital pathology image analysis: Opportunities and challenges. Imaging Med. 2009;1:7–10.
2. He L, Long LR, Antani S, Thoma GR. Histology image analysis for carcinoma detection and grading. Comput Methods Programs Biomed. 2012;107(3):538–56.
3. Tabesh A, Teverovskiy M, Pang HY, et al. Multifeature prostate cancer diagnosis and Gleason grading of histological images. IEEE Trans Med Imaging. 2007;26(10):1366–78.
4. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng. 2010;57(4):841–52.
5. Janssens T, Antanas L, Derde S, Vanhorebeek I, Berghe GV, Grandas FG. CHARISMA: An integrated approach to automatic H&E-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting. Med Image Anal. 2013;17:1206–19.
6. Xu Y, Zhu JY, Chang EIC, Lai MD, Tu ZW. Weakly supervised histopathology cancer image segmentation and classification. Med Image Anal. 2014;18:591–604.
7. Saraswat M, Arya KV. Automated microscopic image analysis for leukocytes identification: A survey. Micron. 2014;65:20–33.
8. Yang L, Meer P, Foran DJ. Unsupervised segmentation based on robust estimation and colour active contour models. IEEE Trans Inform Technol Biomed. 2005;9:475–86.
9. Lezoray O, Cardot H. Cooperation of colour pixel classification schemes and colour watershed: a study for microscopic images. IEEE Trans Image Process. 2002;11(7):783–9.
10. Mohapatra S, Patra D, Satpathy S. Automated leukemia detection in blood microscopic images using statistical texture analysis. In: Proceedings of International Conference on Communication, Computing & Security. New York, USA: ACM; 2011. p. 184–7.
11. Ko BC, Gim J, Nam J. Automatic white blood cell segmentation using step-wise merging rules and gradient vector flow snake. Micron. 2011;42:695–705.
12. He L, Long LR, Antani S, Thoma GR. Distribution Fitting-based Pixel Labeling For Histology Image Segmentation, Proc. SPIE 7963, Medical Imaging. Florida, USA: Computer-Aided Diagnosis, SPIE; 2011. p. 79633D.
13. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society. Series C. 1979;28(1):100–8.
14. Comaniciu D, Meer P. Mean Shift: A Robust Approach Toward Feature Space Analysis, IEEE Trans. Pattern Analysis and Machine Intelligence. 2002;24(5):603–19.
15. Xu L, Jordan MI. On Convergence Properties of the EM Algorithm for Gaussian Mixtures. Neural Comput. 1996;8(1):129–51.
16. Shi J, Malik J. Normalized Cut Method for Image Segmentation, IEEE Trans. Pattern Analysis and Machine Intelligence. 2000;22(8):888–905.
17. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell. 2001;23(11):1222–39.
18. Ben Salah M, Mitiche A, Ben Ayed I. Multiregion Image segmentation by Parametric Kernel Graph Cuts. IEEE Trans Image Process. 2011;20(2):545–57.
19. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov Random Field model and the expectation–maximization algorithm. IEEE Trans Med Imaging. 2001;20(1):45–57.
20. Franklin C. Summed-area tables for texture mapping, SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques. ACM, New York, USA. 1984;18(3):207–12.

Wu *et al. BioMedical Engineering OnLine* (2015) 14:24

Page 12 of 12

21. Viola P, Michael J. Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001;1:511–8.
22. Wang J, Guo YW, Ying YT, Liu YL, Peng QS. Fast non-local algorithm for image denoising, Proceeding of IEEE International Conference on Image Processing. 2006. p. 1429–32.
23. Arthur D, Vassilvitskii S. k-means++: The advantages of of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. New York, USA: ACM; 2007. p. 1027–35.
24. Rand WM. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc. 1971;66(336):846–50.
25. Meilǎ M. Comparing clusterings – an information based distance. J Multivar Anal. 2007;98:873–95.