

Segmentation and localisation of whole slide images using unsupervised learning

Hazem Hiary¹, Raja S. Alomari^{1,2}, Vipin Chaudhary²

¹King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan

²Department of Computer Science and Engineering, The University at Buffalo, SUNY, USA

E-mail: ralomari@buffalo.edu

Abstract: Digital pathology has been clinically approved for over a decade to replace traditional methods of diagnosis. Many challenges appear when digitising the whole slide scan into high resolution images including memory and time management. Whole slide images require huge memory space if the tissue is not pre-localised for the scanner. The authors propose a set of clinically motivated features representing colour, intensity, texture and location to segment and localise the tissue from the whole slide image. This step saves both the scanning time and the required memory space. On average, it reduces scanning time up to 40% depending on the tissue type. The authors propose, using unsupervised learning, to segment and localise tissue by clustering. Unlike supervised methods, this method does not require the ground truth which is time consuming for domain experts. The authors proposed method achieves an average of 96% localisation accuracy on a large dataset. Moreover, the authors outperform the previously proposed supervised learning results on the same data.

1 Introduction

Anatomic pathology is the study of organs and tissues to determine the causes and effects of particular diseases. It is one of two branches of pathology, the other is clinical pathology; the diagnosis of disease through the laboratory analysis of bodily fluids. Anatomic pathology is concerned with diagnosis of disease based on three levels of examinations: gross, microscopic and molecular. Examinations are performed on organs, tissues and whole bodies (autopsy) [1].

Digital anatomic pathology utilises high resolution imaging technology for diagnosis of particular diseases [1]. Many commercially available high resolution scanners evolved over the last decade. These scanners involve many steps to finally produce high resolution pathology images that resemble the traditional microscopic view of the whole slides. The quality of these high resolution images enhanced over time and helped replacing the traditional anatomic pathology. Hence, digital pathology has become an acceptable clinical standard for various diagnosis and prognosis tasks.

Digital pathology eliminated the necessity of physical slides sharing and thus eliminated all the shipping costs and related issues. Furthermore, it enabled case communication and remote diagnosis. Digital pathology has made a major revolution in diagnosis and prognosis quality and reliability. However, many issues became the focus including image storage and retrieval as these images are usually large and require huge storage infrastructure and special management.

In this paper, we segment the tissue from the whole slide (as shown in Fig. 1) image and localise it within a

bounding box. This step helps reducing between 60 and 80% from the final image size. The high resolution scanner scans only the segmented portion instead of the whole slide. However, it requires a reliable method for localising and segmenting the tissue. We propose a reliable and robust method for localisation and segmentation of the tissue from the whole slide images based on our previous work [2].

In our previous work [2], we used supervised learning (neural network) and performed segmentation by classification of image blocks. However, our supervised-learner requires training data (sample sets along with the ground truth) which is a tedious and time consuming step. Clinical labs have huge variability in their settings that require us to re-train our neural network to re-learn the setting of each lab. Preparation of training data is time consuming because it requires the ground truth each time we prepare the training data for a new lab settings. In this work, we use unsupervised learning and thus we eliminate the need for preparing the ground truth. Thus, our methods become more practical in various lab settings because we do not ask pathologists to provide ground truth regardless of the lab settings and the types of images produced in the lab.

In this segmentation problem, two outcomes are required: a set of focus points (\mathcal{F}_k^*) and a bounding box (\mathcal{B}_k^*) (or boxes). The set of focus points are required to lie on the actual tissue (\mathcal{T}) to guide the high resolution scanner to determine the tissue properties for subsequent scanning steps. On the other hand, the bounding box guides the scanner for the area of interest for subsequent scanning.

Focus points (\mathcal{F}_k^*) are required to be detected with minimal false positives (FP) because each FP point wastes about 200

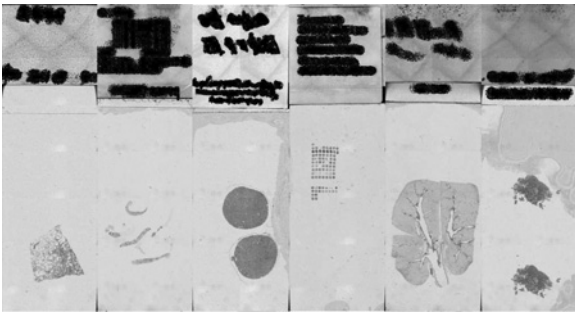


Fig. 1 Sample set of slide images

Upper portion is reserved for label/bar code (blanked out for anonymity) and usually takes the top one-third of the image

ms from the scanner time. On the other hand, the second output, the bounding boxes (\mathcal{B}_k^*), should contain every useful tissue location with minimal false negatives (FN) (2) so that the high resolution scan does not miss any tissue

$$\mathcal{F}_k^* = \arg \min_{\mathcal{F}_k} P(\mathcal{F}_k | \overline{\mathcal{T}}) \quad (1)$$

$$\mathcal{B}_k^* = \arg \min_{\mathcal{B}_k} P(\overline{\mathcal{B}_k} | \mathcal{T}) \quad (2)$$

where (\mathcal{T}) and ($\overline{\mathcal{T}}$) represent tissue and non-tissue locations, respectively. Hence, the ultimate goals are to: (i) minimise the focus points (\mathcal{F}_k) that lie on a non-tissue location ($\overline{\mathcal{T}}$) and (ii) minimise the missed tissue from bounding boxes (minimise the tissue (\mathcal{T})) area that is not included in a bounding box ($\overline{\mathcal{B}_k}$).

This localisation and segmentation step is necessary to save both time and memory during the high-resolution scanning step. One whole slide takes about 1 h to fully scan it without our localisation step. However, an average slide requires about 15 min after using our localisation step. On the other hand, the whole region ($44.449 \text{ mm} \times 25.4 \text{ mm} \approx 1129 \text{ mm}^2$) at a sampling rate of $0.46 \mu\text{m}/\text{pixel}$ ($20\times$) produces more than 6150 high-resolution images of dimensions 1024×768 pixels. Considering overlaps between these high-resolution images, which are used for the stitching stage, the final image dimension for one slide would be about $110\,000 \times 44\,000$ pixels which needs $110\,000 \times 44\,000 \times 3$ bytes/pixel ≈ 13.52 Gbytes. The resulting images, after using our localisation step, reduces the required space to an average of 3 Gbytes.

Image segmentation is a known problem in the medical imaging field. Various methods are proposed for various segmentation tasks for various medical imaging modalities [3–5]. Early trends in image segmentation were based on heuristics, edge detection, region growing and thresholding operators. The second trend utilises machine intelligence methods including supervised, unsupervised and semi-supervised learning techniques. The third trend in segmentation is based on prior knowledge incorporation in the forms of atlases [6], prior probabilities or models [7].

Many supervised learning methods have been applied in the literature such as [3, 8, 9] and including our previous work [2]. On the other hand, unsupervised learning has been also applied to many segmentation problems such as Hall *et al.* [10] who performed a comparison between a neural network and fuzzy clustering (unsupervised) techniques in segmenting MR images of the brain.

They trained both methods on normal and abnormal cases and found that each method has advantages in some aspects.

Zhang and Ji [11] proposed multi-layer Bayesian network model for fully automatic and interactive image segmentation. They studied the statistical dependencies and measurements of the image entities like: regions, vertices and edge segments for image segmentation and they provided the user with a new active input selection to do interactive image segmentation. For testing, they used Weizmann dataset and VOC2006 cow images and they improved the overall segmentation accuracy compared with similar methods. Fuzzy C-means algorithm has been used for image segmentation: Nadernejad and Sharifzadeh [12] employed bilateral filtering algorithm to obtain a pixonal image without unnecessary details of the image and avoid the over segmentation then they applied the fuzzy C-means algorithm to the obtained pixonal image, their approach was comparable to similar other techniques with better accuracy.

Chen *et al.* [13] introduced a framework for image segmentation using multiple-kernel fuzzy C-means algorithm and proved its significant flexibility in kernel selections. They tested their work on medical images and showed better performance than other kernel fuzzy C-means algorithms. Similarly, Sulaiman and Ashidi [14] used adaptive fuzzy K-means clustering algorithm that is applied on many types of images for image segmentation. Their results were visually tested and showed better quality compared to the other clustering techniques. Alpert *et al.* [15] proposed a probabilistic bottom-up aggregation technique and cue integration to provide hierarchical image segmentation. They performed their aggregation via the integration of intensity and texture distributions for image regions.

Song *et al.* [16] focused on segmenting medical images trying to solve some of the segmentation problems such as: restrain noise, keeping the edge property well and model mismatch. They used non-parametric mixture models with spatial information. They depended on the cosine orthogonal sequence with spatial information functions to design the non-parametric mixture model. They showed better performance on both MR brain images and CT scans for abdomen.

Tosun and Gunduz-Demir [17] developed a segmentation method of digital pathology tissue images. They applied grey-level run-length matrices to extract texture features. Their results showed a high accuracy on colon tissue images compared with other segmentation algorithms. Kong *et al.* [18] worked on cell segmentation and cell splitting for digital pathology images. For segmentation, they first classified the pixels into cell or extra-cellular group. Then they extracted the colour and texture using local ‘Fourier Transform’. For splitting, they classified the component of the segmentation map into touching-cell clump or a single non-touching cell depending on the distance between the radial-symmetry centre and the geometrical centre of these components. For testing, they used follicular lymphoma pathology images. Splitting error rate was 5.25% per image.

Liu *et al.* [19] improved the spatial spectral clustering technique to a non-local one and used it for image segmentation. They used the kernel K-means algorithm incorporated with the non-local spatial constraints. Then they applied spectral clustering to their non-local spatial matrix for image segmentation. They used synthetic and real images for testing and showed a high performance and reduced noise in their results.

Zhao *et al.* [20] was trying to solve sensitivity to noise resulting from ignoring the spatial information for the

image pixels. They presented a new fuzzy clustering algorithm with non-local adaptive spatial constraints. They utilised the non-local spatial information of the pixels to control the noise in the segmentation process. Their experimental results showed satisfying segmentation performance on noisy images.

In pathology images, many research efforts have been performed on many segmentation, diagnosis and prognosis tasks. Latson *et al.* [21] used basic fuzzy *C*-means clustering for nuclei segmentation from high resolution images whereas Petushi *et al.* [22] used adaptive thresholding. However, these methods do not extend for high variabilities in data [23]. Bamford and Lovell [24] used active contours for segmentation of the nuclei which succeeds when there is no overlapping nuclei.

Our previous works target many tasks in histology such as removal of stromal cells from histology images for breast cancer diagnosis where we utilise local binary patterns and a Gibbs model [25]. We also worked on nuclei counting for proliferation rate estimation from histology images which is based on segmentation of the nuclei via *K*-means clustering [26]. Recent efforts in utilising clustering have also been reported such as [27, 28].

2 Materials and methods

2.1 Data

We collect three hundred (300) images for building our dataset from our collaborating pathology labs. The image acquisition protocol is as follows: initially, the clinical technician prepares the biopsy and fixes it on a whole slide as shown in Fig. 2 [29]. It is inserted along with a set of other whole slides on a rack. The digital pathology scanner that scans these images in our data set contains one rack that has 64 slots. This rack is then inserted inside the machine in a completely constraint environment in terms of lightening conditions and other image acquisition parameters. The camera inside the digital pathology scanners captures the whole slide images (that we work on) for each whole slide.

These images are then displayed for the technician on a special workstation. The technician then, manually, draws

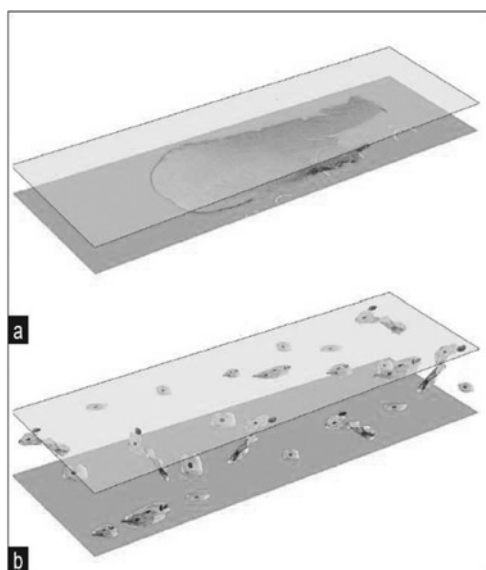


Fig. 2 Biopsy on a whole slide preparation [29]

the bounding box (or boxes in case of more than one tissue segments) surrounding the tissues. Then the technician runs a batching job to allow the high resolution scanners to scan each whole slide image. These high resolution scanners only scan the area within the bounding boxes. A stitching algorithm is then automatically used to provide one full high resolution image for each slide that has up to $20\times$ zoom levels. These high resolution images are the images the pathologist uses for diagnosis.

The whole slide images greatly vary in tissue type, size, colour and spread over the whole slide. The purpose of our work is to be able to present a generic robust localisation and segmentation for the tissue for the various types of tissues without the need to sort the slides. To show the variability in our dataset, we select a set of these as shown in Fig. 3. Despite the highly constraint image acquisition, the great variability in tissues remain a great challenge.

We perform the segmentation using unsupervised learning. We use *K*-means to cluster the data exist in whole slide images: tissue, detritus and background. Then we cluster new whole slide images in block-by-block fashion using the *K*-means learned cluster centers. This unsupervised learner, by definition, does not require annotated training data and thus eliminates the time and efforts required for generation of ground truth. This is more suitable in solving this problem when we are looking for more generic segmentation solution because of high variability between clinical labs settings. Fig. 4 shows the work flow of our proposed work. In this section, we illustrate our feature extraction, training and testing stages.

2.2 Feature extraction

Feature extraction reduces the problem dimension to the suitable dimensions that matter. These lower dimensions need great experimental and domain knowledge coordination to extract the suitable features for the problem in hand. For our problem, we study a set of features from various categories that our domain expert examines when performing the segmentation task manually.

Upon examining the whole dataset, we select features representing the pixel intensity, colour, texture and spatial distribution of the tissue on the slides. Combining all these aspects (dimensions) provides well-established discriminative power for the unsupervised learner. We point out that we utilise only the RGB colour model for feature extraction. Since the imaging environment is very constraint in terms of all possible parameters, utilising the RGB model was theoretically and practically justified. Thus transforming into other colour models such as $L^*a^*b^*$ or HSI was unnecessary because of the avoidable additional and unnecessary computational burden. We present our features in details in the following subsections.

2.2.1 Pixel intensity features: Intensity is a major role-player in determination of the tissue. It mainly captures both the unstained tissue such as Fig. 3e and the white and grey microscopic holes such as Fig. 3b where the tissue has no colour. Such tissue has no meaningful texture but intensity for discrimination. Furthermore, some white and grey microscopic holes appear inside the tissue that may not be captured by any other features such as the circular tissues (Fig. 3f). To capture this intensity, we utilise the mean grey intensity $\mu_{\text{grey}} = \frac{1}{W} \sum_i^W \text{grey}_i$ and variance of the intensity levels within the window $\sigma_{\text{grey}}^2 =$

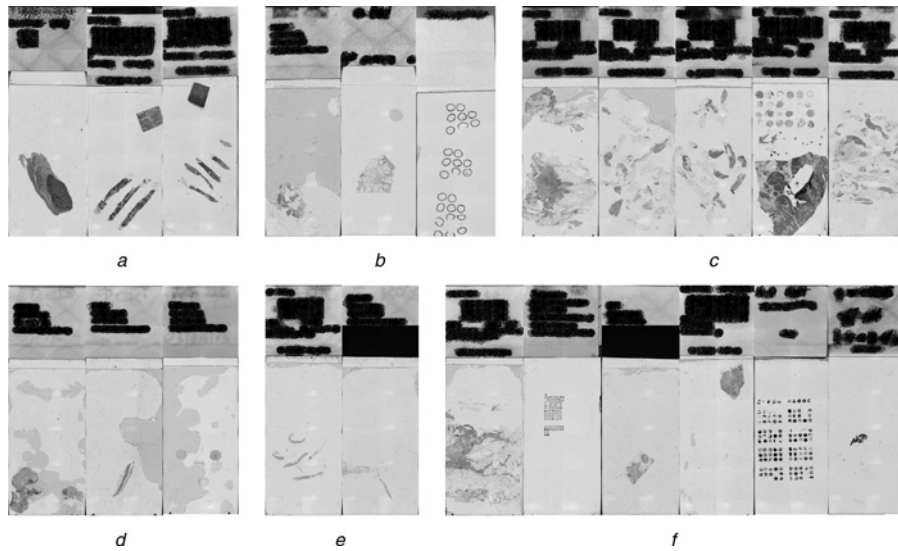


Fig. 3 Set of whole slide images from our dataset to show the huge variability in tissue size, type, colour, distribution and other aspects

- a Stained tissues
- b Microscopic holes
- c Distributed pieces
- d Detritus in middle
- e Unstained
- f Various shapes

$\sum_i^{\mathcal{W}} (\text{grey} - \mu_{\text{grey}})^2$ where \mathcal{W} refers to the window size (\mathcal{W}) and i is the pixel index within the window. These two features present the mean of the intensity of each cluster type as well as the variation of intensity levels within that type. To obtain the intensity from the colour images, we use the standard NTSC formula for the conversion

$$\text{grey} = 0.2989\mathcal{R} + 0.5870\mathcal{G} + 0.1140\mathcal{B} \quad (3)$$

where \mathcal{R} , \mathcal{G} and \mathcal{B} are the colour channels of the RGB colour model red, green and blue, respectively.

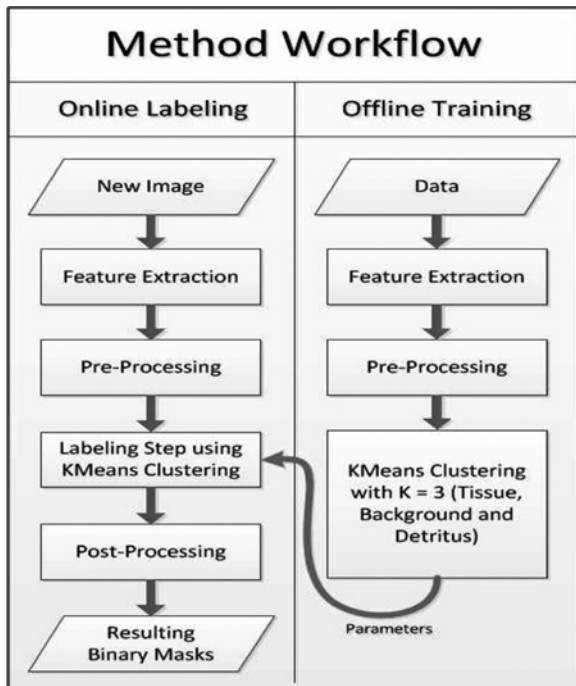


Fig. 4 Work flow of our unsupervised learner

2.2.2 Colour features: Great portion of the tissue is stained and thus it seems natural to utilise some suitable colours features. However, some of the detritus artifacts capture colour information because of light reflection through filters during image capturing. These are usually of pure colour frequencies which make them distinguished from stained tissue colours.

Tissue colour vary but mainly: pink, green, brown and blue as shown in Fig. 3. Thus, we select a set of colour features that do not represent specific colours. Instead, these features represent the variability inside the colour and discriminate them from the grey level regions. Hence, we utilise the variance in each colour channel: $\sigma_{\mathcal{R}}^2$, $\sigma_{\mathcal{G}}^2$, $\sigma_{\mathcal{B}}^2$, the variance between the peaks (highest value) in each colour channel σ_{Peaks}^2 , and the mean of the Hue value from the HSI (Hue, Saturation, Intensity) colour model μ_{hue} .

The variance within each colour channel gives a sense about whether this window (block) is stained and has colour (regardless of its colour) information or not. Stained tissue has higher variance in one or more of these channels. The variance is defined upon the mathematical variance within each window as

$$\sigma_{\mathcal{R}}^2 = \sum_i^{\mathcal{W}} (\mathcal{R}_i - \mu_{\mathcal{R}})^2 \quad (4)$$

$$\sigma_{\mathcal{G}}^2 = \sum_i^{\mathcal{W}} (\mathcal{G}_i - \mu_{\mathcal{G}})^2 \quad (5)$$

$$\sigma_{\mathcal{B}}^2 = \sum_i^{\mathcal{W}} (\mathcal{B}_i - \mu_{\mathcal{B}})^2 \quad (6)$$

where $\mu_{\mathcal{R}}$, $\mu_{\mathcal{G}}$ and $\mu_{\mathcal{B}}$ are the mean values for the red, green and blue channels, respectively. The window size ($\mathcal{W} = 21 \times 21$) is constant for work.

On the other hand, the variance between the peaks of the RGB channels is high for stained regions and low for

greyish regions, thus we utilise it for discriminating colour information

$$\sigma_{\text{Peaks}}^2 = \sum_{i=1}^3 (\mathcal{P}_{k_i} - \mu_{\mathcal{P}_k})^2 \quad (7)$$

where $\mu_{\mathcal{P}_k}$ is the mean value of the three peaks: \mathcal{R} , \mathcal{G} and \mathcal{B} .

We also found a good feature that successfully discriminates stained tissue from background and detritus by utilising the hue component of the HSI colour model. We utilise the mean value of the hue \mathcal{H}_i in each window

$$\mu_{\text{hue}} = \frac{1}{W} \sum_i^W \mathcal{H}_i \quad (8)$$

2.2.3 Texture features: We also utilise texture feature because many tissue types contain well-defined texture. On the other hand, both background and detritus do not seem to define any notable texture within its regions. We utilise the co-occurrence matrix \mathcal{C}

$$\mathcal{C} = \sum_{i,j}^{\text{Size}(\mathcal{C})} \mathcal{E}[(\mathcal{X}_i - \mu_i)(\mathcal{X}_j - \mu_j)] \quad (9)$$

where $\mu_i = \mathcal{E}(\mathcal{X}_i)$ is the expectation of the i th entry in the vector \mathcal{X} [30]. To obtain texture statistics including both correlation and energy

$$\text{Correlation} = \sum_{i,j}^{\text{Size}(\mathcal{C})} \frac{(i - \mu_i)(j - \mu_j)\mathcal{C}(i,j)}{\sigma_i \sigma_j} \quad (10)$$

$$\text{Energy} = \sum_{i,j}^{\text{Size}(\mathcal{C})} \mathcal{C}(i,j)^2 \quad (11)$$

2.2.4 Spatial features: This feature is novel for our application. It helps discriminating between the coloured detritus that extends from the edges of the slides to the inside and the stained tissues. The colour that appears on the detritus comes from the reflections of the filters and is different from the stain colour on the tissue. However, we find that spatial and location information is of great discrimination power between the stained tissue and the coloured detritus.

To that end, we suggest a heuristic-based preliminary segmentation to obtain this feature. Initially, we normalise a set of the colour features (σ_{Peaks}^2 , $\sigma_{\mathcal{R}}^2$, $\sigma_{\mathcal{G}}^2$, $\sigma_{\mathcal{B}}^2$, μ_{grey} and σ_{grey}^2) and accumulate them to create a probability map for the whole slide image. Then we apply the Otsu [31] threshold method to get a binary image as a preliminary segmentation. Then, we apply connected component analysis. Then, we produce a probability map where the pixels of the components that are connected to the border (within 50 pixels) are assigned lower probability value (0.001) whereas the rest are assigned a higher value (0.5). This border map is then used as a post-processing step after K -means is performed. This novel feature provides great help in elimination of the coloured detritus that extends to the inside of the whole slide images because these artefacts will obtain a lower probability value (0.001).

2.2.5 Training stage: Since we utilise unsupervised learning, we do not require a training stage. However, K -means clustering requires a seed point for each cluster and it is a common practice to feed the K -means with reasonable seed points for each cluster based on the given data. This will make sure that each time the K -means converges, clusters converge around the same labels, that is each of the three clusters (tissue, detritus and background) will have the same label each time K -means is run.

To learn the parameters of the K -means (initial seeds for clusters), we obtain a set of data points from the images for each cluster (tissue, background and detritus). Window size for each point is (20×20) which makes a (21×21) window for each data point. During this step, we check that each window consists of pixels within the same cluster type. This point selection step is necessary to provide K -means with fixed clusters seed points to start with before the convergence of the clusters. We manually select: 2140 data point for the tissue, 2140 data points for the detritus region, and 2140 data points for the background. Tissue region is the region where the actual tissue lie, detritus is particulate organic material that happens to be on the slides because of the lab conditions. Usually, it looks like a grey matter along the edges of the slides, and background is the white regions all around the slide.

This training step is only performed once and it can be eliminated if training data is not available unlike supervised learners [2]. As we discussed before, utilising unsupervised learning eliminates the need to re-train the system for each lab settings.

2.2.6 Testing stage: By testing stage, we mean the actual segmentation step where we apply the K -means on the whole slide image and see the outcome. We represent the outcome as a binary image overlayed on the original image. This binary image only passes the tissue pixels and suppresses all other pixels. To apply K -means on a new image, we initially divide the input image into equal size blocks (windows) and then extract the same feature vector ($V = \{v_i\}$) from each window. This feature vector is passed to the K -means for clustering. All pixels inside the window takes the same cluster outcome which is one of the three clusters. During the testing, we use the window size (5×5) which shows a fairly smooth outcome. The choice of (5×5) is completely empirical choice for the trade-off between the blocky-segmentation and segmentation accuracy. Post-processing steps include a median filter of (20×20) mask to smooth the resulting image followed by image filling [32] operation. Then we apply the spatial features (probability map) discussed earlier.

We performed many experiments to decide on the choice of this window size for training and testing. The trade-off between larger and smaller window sizes is between information availability (that led, in our case, to higher accuracy) and blocky segmentation outcome. Larger window size presents the classifier with more information whereas the smaller size window allows less information. On the other hand, larger window presents blocky outcome whereas smaller size presents smoother outcome. For example, when we chose (3×3) window size, classification accuracy was low despite the high smoothness that we observed whereas going to (15×15) gave us higher accuracy but the outcome was highly blocky despite the post-median filter. However, we point out that it is not the case if we go higher on window size, for example, (51×51) , that we have better classification accuracy because if

so, we can just apply K -means on the whole image as one unit which did not gave good results compared to what we obtained in our proposed method.

3 Results

Our dataset consists of 300 images collected in two datasets of 150 images each. We use 85 images from one dataset of the two for extraction of the 2140 training data points for each type (tissue, background and detritus) as discussed in Section 2.2.5. Then we perform the segmentation via clustering on three sets: the 85 images, the 150 images from which training data was extracted and the rest 150 images as shown in Table 1.

To produce results for our collaborating pathologist to check for segmentation accuracy, we overlay the resulting binary mask on the tissue pixels. This produces images as shown in Fig. 5. Then, our collaborating pathologist examines each image along with the original image and decides on our segmentation as one of the following: correct segmentation, has some FN regions, or wrong segmentation (localisation error) as shown in Table 1.

A FP result contains some areas that are falsely labelled as tissue [positive automated result when the ground truth is non-tissue (negative)]. Similarly, a FN result contains some areas that are falsely labelled as non-tissue [negative automated result when the ground truth is tissue (positive)]. When the pathologist examines the result, if the mislabelled areas (FP or FN) are small (insignificant), the image is counted as FP or FN depending on the dominating type of the mislabelled areas. On the other hand, if the area (FP or

FN) is large (significant), the image is counted as an error (localisation error) as shown in Table 1.

We also compute the percentage of error segmentation by defining the localisation error ($\mathcal{L}_{\text{Error}}$) as

$$\text{Accuracy} = \left(1 - \frac{\mathcal{L}_{\text{Error}}}{N}\right) 100\% \quad (12)$$

where ($\mathcal{L}_{\text{Error}}$) is the number of images defined as error segmentation by the pathologist and N is the total number of images in the dataset. Table 1 shows the resulting error percentage for each dataset. We discuss these results in Section 4.

In medical image analysis, one major concern is the inter-operator reliability. The human factor, the experience, the environment, the data, the problem under investigation and other factors contribute in the variability in medical expert decisions. For the completeness of our results reporting, we study the inter-operator reliability for our specific problem from our dataset. We asked three pathologists to perform the segmentation manually. Each pathologist performs the segmentation using an interface that we provide. Images are displayed one-by-one and a graphical hand tool on a tablet is provided for the pathologist. Each pathologist independently performs the segmentation without any knowledge of each other and without having any external impact. The pathologist is asked to draw a contour surrounding the tissue only.

There are many statistical measures to study the inter-operator reliability (or variability). However, in this paper, we study the similarity between the three manual

Table 1 Results for the whole 300 images divided into two sets, 150 images each

Dataset	Size	False negatives	False positives	Localisation error	Localisation accuracy, %
training	85	7	3	2	96.9
testing (including the training)	150	13	9	7	95.3
testing (separate set)	150	10	7	6	96.0

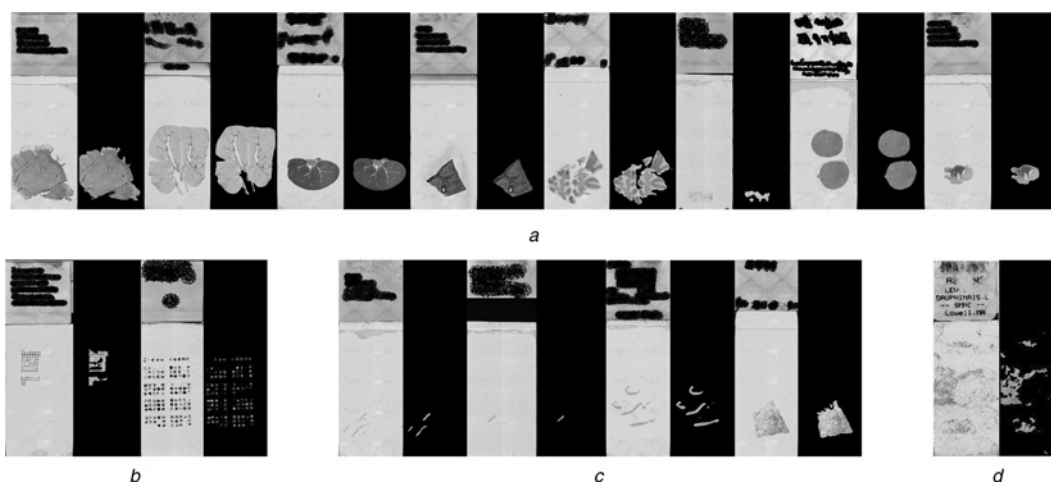


Fig. 5 Segmentation of various images from our dataset

a Correct localisation of tissues

b Contains false negative

c Localisation of various light stained cases

d Error

Segmentation smoothness and quality is indistinguishable from the supervised learning results

Table 2 Average dice percentage similarity between pathologists

Image subset	Main pathologist against pathologist 2, %	Main pathologist against pathologist 3, %
1–50	99.3	98.7
51–100	97.1	98.8
101–150	98.9	99.0
151–200	98.7	97.9
201–250	98.2	97.8
251–300	97.2	99.1
average	98.2	98.6

segmentations of the three pathologists by performing the ‘Dice’ measure to show the spatial similarity between the three segmentations.

Since the ‘Dice’ measure is defined for two segmentations, we compared the segmentation of the main pathologist (who performed the accuracy check for each case described in Table 1) against the segmentation of the other two pathologists independently. Dice (DSC) is commonly used for spatial overlap similarity between two sets of segmentation of the same anatomy [33]. Its value ranges between 0 (no overlap) and 1 (perfect agreement). We express ‘Dice’ as percentages [34]

$$DSC = \frac{2(\mathcal{A} \cap \mathcal{G})}{(\mathcal{A} \cap \mathcal{G} + \mathcal{A} \cup \mathcal{G})} 100\% \quad (13)$$

where \mathcal{A} is the segmentation results from our main pathologist whereas \mathcal{G} is the corresponding segmentation from the other pathologist. We perform this twice; once for each of the two pathologists as shown in Table 2. The spatial segmentation between the two segmentations against the segmentation of the main pathologist is almost perfect. This indicates that the segmentation of the three pathologists for (i) this specific segmentation problem and (ii) using our dataset of whole slide images has almost perfect agreement between pathologists. On average over the 300 images, there was 98.2% and 98.6% spatial segmentation similarity between the main pathologist and the other two pathologists, respectively. This experiment validates considering only one pathologist opinion for performing the previous experiment. Moreover, it was not practical to have all three pathologists perform the previous experiment because we did not want them to meet in one room and know each other’s decision. We setup this experiment to reserve the independence in segmentation decision between each of the three pathologists.

4 Discussions

Since the unsupervised learning does not require training, it provides the convenience for our problem because we deal with various types of tissue that vary in various aspects. Supervised learning requires re-training for each lab setting whereas unsupervised eliminates the need for training. However, usually, this comes on the burden of losing portion of the accuracy. In our case, we produced comparable accuracy to our previous work that utilised supervised learning, thanks to the many empirically motivated features we utilise.

In this section, we discuss some issues related to our results. As shown in Table 1, only two images were marked as error segmentation out of the 85 training images, seven

out of the 150 images set (that contain the 85), and seven out of the other 150 images. We find that the training stage did not affect the accuracy result as the 95.3 and the 96% are comparable given the sample size of 150. These are also very similar to our previous results with supervised learning where we obtained about 94% accuracy for a set of 150 images. This shows the benefit of utilising unsupervised learning without loss of the accuracy.

Furthermore, these three experiments show that if we present our system with new images, and without any form of training, we will obtain comparable accuracy to the case when we train. This shows the great robustness of our selected features which are our main contribution in this work.

Fig. 5 shows a set of images from our dataset along with the resulting segmentation out of our proposed system. These images were also shown in our previous work after applying the supervised learning. We present them here for the sake of comparison. Upon the comments from our domain expert, the results were indistinguishable from the previous results in terms of segmentation smoothness and accuracy. However, we gain the huge benefit of the elimination of the training step and thus elimination of the need for the ground truth generation which is tedious and expensive to do.

5 Conclusion

We proposed utilising unsupervised learning for pathology image segmentation. Our method relies on feature selection that is motivated by the clinical setting and the way the pathologist judges the segmentation. We propose using a set of features for various aspects including colour, texture, intensity and location. We prepared a dataset of 300 images from various collaborating labs for the various tissue types. Our method utilises unsupervised learning which does not require training and thus, there is no need to fine-tune the parameters for each lab setting. Moreover, our method produced highly robust correctness results comparable to supervised learning. We presented our results in many settings including training of our method and without training of the method and we concluded with 95.5% correctness accuracy.

6 References

- 1 Lefkowitz, J.: ‘Anatomic pathology’ (Saunders, New York City, USA, 2006)
- 2 Alomari, R., Allen, R., Sabata, B., Chaudhary, V.: ‘Localization of tissues in high-resolution digital anatomic pathology images’. Proc. SPIE Medical Imaging 2009: Computer-Aided Diagnosis, FL, USA, 2009, vol. 7260, doi:10.1117/12.811430
- 3 Bezdek, J., Hall, L., Clarke, L.: ‘Review of MR image segmentation techniques using pattern recognition’, *Med. Phys.*, 1993, **20**, (4), pp. 1033–1048
- 4 Pham, D., Xu, C., Prince, J.: ‘A survey of current methods in medical image segmentation’, *Annu. Rev. Biomed. Eng.*, 2000, **2**, pp. 315–338
- 5 Withey, D., Koles, Z.: ‘Medical image segmentation: methods and software’. Proc. NFSI and ICFBI’07, Hangzhou, China, 2007, pp. 140–143
- 6 Cuadra, M., Pollo, C., Bardera, A., Cuisenaire, O., Villemure, J.: ‘Atlas-based segmentation of pathological MR brain images using a model of lesion growth Cuadra’, *IEEE Trans. Med. Imaging*, 2004, **23**, (10), pp. 1301–1314
- 7 Van Ginneken, B., Frangi, A., Staal, J., Romeny, B., Viergever, M.: ‘Active shape model segmentation with optimal features’, *IEEE Trans. Med. Imaging*, 2002, **21**, (8), pp. 924–933
- 8 Unay, D., Gosselin, B.: ‘Artificial neural network-based segmentation and apple grading by machine vision, Proc. IEEE Int. Conf. on Image Processing (ICIP’05), Genoa, Italy, 2005, vol. 2, pp. 630–633

- 9 Estevez, P., Flores, R., Perez, C.: 'Color image segmentation using fuzzy min-max neural networks'. Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN'05), Montreal, Quebec, Canada, 2005, vol. 5, pp. 3052–3057
- 10 Hall, L., Bensaid, A., Clarke, L., Velthuizen, R., Silbiger, M., Bezdek, J.: 'A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain', *IEEE Trans. Neural Netw.*, 1992, **3**, (5), pp. 672–682
- 11 Zhang, L., Ji, Q.: 'A Bayesian network model for automatic and interactive image segmentation', *IEEE Trans. Image Process.*, 2011, **20**, (9), pp. 2582–2593
- 12 Nadernejad, E., Sharifzadeh, S.: 'A new method for image segmentation based on Fuzzy C-means algorithm on pixonal images formed by bilateral filtering, signal, image and video Processing', doi: 10.1007/s11760-011-0274-0, 2011
- 13 Chen, L., Chen, P., Lu, M.: 'A Multiple-Kernel Fuzzy C-means algorithm for image segmentation', *IEEE Trans. Syst. Man Cybernet. B, Cybernet.*, 2011, **41**, (5), pp. 1263–1274
- 14 Sulaiman, S., Ashidi, N.: 'Adaptive Fuzzy-K-means clustering algorithm for image segmentation', *IEEE Trans. Consum. Electron.*, 2009, **55**, (4), pp. 2145–2153
- 15 Alpert, S., Galun, M., Brandt, A., Basri, R.: 'Image segmentation by probabilistic bottom-up aggregation and cue integration', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (2), pp. 315–327
- 16 Song, Y., Liu, Z., Chen, J., Zhu, F., Xie, C.: 'Medical image segmentation based on non-parametric mixture models with spatial information', *Signal Image Video Process.*, 2012, **6**, (4), pp. 569–578
- 17 Tosun, A., Cigdem, G.: 'Graph run-length matrices for histopathological image segmentation', *IEEE Trans. Med. Imaging*, 2011, **30**, (3), pp. 721–732
- 18 Kong, H., Gurcan, M., Belkacem-Boussaid, K.: 'Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting', *IEEE Trans. Med. Imaging*, 2011, **30**, (9), pp. 1661–1677
- 19 Liu, H., Jiao, L., Zhao, F.: 'Non-local spatial spectral clustering for image segmentation', *Neurocomputing*, 2010, **74**, pp. 461–471
- 20 Zhao, F., Jiao, L., Liu, H., Gao, X.: 'A novel fuzzy clustering algorithm with non local adaptive spatial constraint for image segmentation', *Signal Process.*, 2011, **91**, (4), pp. 988–999
- 21 Latson, L., Sebek, B., Powell, K.: 'Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy', *Anal. Quant. Cytology Histology*, 2003, **25**, (6), pp. 321–331
- 22 Petushi, S., Garcia, F., Haber, M.: 'Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer', *BMC Med. Imaging*, 2006, **6**, (14), doi:10.1186/1471-2342-6-14
- 23 Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J.: 'Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology'. Proc. IEEE Int. Symp. Biomedical Imaging (ISBI'08), Paris, France, 2008, pp. 284–287
- 24 Bamford, P., Lovell, B.: 'Unsupervised cell nucleus segmentation with active contours', *Signal Process.*, 1998, **71**, (2), pp. 203–213
- 25 Alomari, R., Ghosh, S., Chaudhary, V., Al-Kadi, O.: 'Local binary patterns for stromal area removal in histology images'. Proc. SPIE Medical Imaging 2012: Computer-Aided Diagnosis, San Diego, CA, USA, 2012, vol. 8315, pp. 831524
- 26 Alomari, R., Al-Lahham, H., Chaudhary, V., Hiary, H.: 'Automated proliferation rate estimation from Ki-67 histology images'. Proc. SPIE Medical Imaging 2012: Computer-Aided Diagnosis, San Diego, CA, USA, 2012, vol. 8315, pp. 83152A
- 27 Selvan, A., Saatchi, R., Ferris, C.: 'Improving medical image perception by hierarchical clustering based segmentation'. Proc. Int. Conf. Information Technology and Applications in Biomedicine, Lamaca, Cyprus, 2009, pp. 1–6
- 28 Xinwu, L.: 'A new volume segmentation algorithm for medical image based on improved K-means', *Int. J. Digit. Content Technol. Appl.*, 2011, **5**, (6), pp. 1–7
- 29 Lee, R., McClintock, D., Laver, N., Yagi, Y.: 'Evaluation and optimization for liquid-based preparation cytology in whole slide imaging', *J. Pathology Inform.*, 2011, **2**, pp. 46
- 30 Wasserman, L.: 'All of statistics: a concise course in statistical inference' (Springer, USA, 2004)
- 31 Otsu, N.: 'A threshold selection method from gray-level histograms', *IEEE Trans. Syst. Man Cybernet.*, 1979, **9**, (1), pp. 62–66
- 32 Soille, P.: 'Morphological image analysis: principles and applications' (Springer-Verlag, Berlin New York, 1999), pp. 173–174
- 33 Dice, L.: 'Measures of the amount of ecologic association between species', *Ecology*, 1945, **26**, (3), pp. 297–302
- 34 Babalola, K., Patenaude, B., Aljabar, P., *et al.*: 'Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI'. Proc. 11th Int. Conf. Medical Image Computing and Computer-Assisted Intervention – Part I (MICCAI'08), Springer-Verlag, Berlin, Heidelberg, 2008, pp. 409–416