Short survey: adaptive threshold methods used to segment immunonegative cells from simulated images of follicular lymphoma stained with 3,3’-Diaminobenzidine&Haematoxylin

Lukasz Roszkowiak, Anna Korzynska, Dorota Pijanowska
Nalecz Institute of Biocybernetics and Biomedical Engineering
Ks. Trojdena 4 Str., 02-109 Warsaw, Poland
Email: lroszkowiak@ibib.waw.pl

Abstract—We perform a short survey of image thresholding methods for very specific task, and assess their performance comparison. We analyse performance of adaptive thresholding methods concerning segmentation of immunonegative cells of follicular lymphoma tissue samples stained with 3,3’-Diaminobenzidine&Haematoxylin. We use artificial images based on experimental images that greatly simulates real samples and simplifies process of evaluation. We chose 8 methods of adaptive threshold segmentation, with different approach. They were applied to 6 different monochromatic images derived from original RGB images, by splitting layers, conversion to Lab colour space and colour deconvolution. Evaluation of the results was performed with basic statistical measures as sensitivity and specificity along with Jaccard’s coefficient. We identify the thresholding algorithms with superior performance. Collected results will be used to design the new better method based on this approach.

I. INTRODUCTION

Tissue samples stained immunohistochemically are commonly used by pathologists to distinguish various types of cancer [1]. One of them is follicular lymphoma which is the second most common form of non-Hodgkin’s lymphoma [2]. Nowadays, pathologist doesn’t have to evaluate samples via microscope but it can be done by examination of digital images of samples. Unfortunately, even with this aid the human evaluation is irreproducible and prone to error [3]. Moreover, it tends to change from one expert to another, as well as in time for one expert. As analysis is mostly based on counting of immunopositive and immunonegative cells, computer evaluation would give many advantages, like acceleration and reproducibility of the process. Unfortunately, segmentation of such images is not an easy task. Most computer-based procedures for immunohistochemistry image analysis [4]–[7] have limited applicability due to numerous drawbacks.

For test images in this study we decided to use simulated (artificial) images based on experimental images of follicular lymphoma tissue sections stained with 3,3-diaminobenzidine (DAB) and contra-stained with haematoxylin (H). The main reason for this decision was having ‘gold standard’ reference image that can be used to evaluate results of segmentation methods.

Experimental as well as artificial images consist of brown objects among blue ones with bright background that sometimes has slight blue tint. Most of the objects have minor but visible texture. Unfortunately, each and every sample differs in many characteristics. There is a huge variability of shape (from round to elongated), size and colour of objects of interest as well as in the image generally; in colour intensity, range of colour and tone throughout image plane.

For this survey we chose 8 methods of adaptive threshold segmentation to test and evaluate their ability to segment immunonegative cells. We decided to apply segmentation algorithms to 6 different monochromatic images derived from original RGB images. Apart from simple dividing images into single layer images of separate RGB layers we used conversion to Lab colour space, as it was proved to be useful during segmentation in other applications [8]. Since b axis of Lab colour space represents yellow and blue at two ends of the axis, it could be very useful to find proper threshold value to segment blue objects of interest. Also we try to use L (luminescence) layer of Lab colour space as the objects should have contrasting value from the background. Another method we applied in this research is colour deconvolution [9]. We use Haematoxylin layer obtained with predefined colour vector in colour deconvolution algorithm. In summary, we compare results of segmentation performed on: (1) red layer of RGB; (2) green layer of RGB; (3) blue layer of RGB; (4) L layer of Lab colour space; (5) b layer of Lab colour space; (6) Haematoxylin layer after colour deconvolution.

To sum up, in this investigation, we present preliminary study concerning segmentation of immunonegative cells of follicular lymphoma tissue stained with DAB&H. It will lead to development of improved method of segmentation. In cooperation with our previous research [10] on segmentation of immunopositive cells we hope to achieve fully capable software to analyse digitized samples of follicular lymphoma tissue sections.

We acknowledge the financial support of the Polish National Science Centre grant, PRELUDIUM, 2013/11/N/ST7/02797.
II. METHODS

For this study, we decided to use previously created artificial images. Thus, we obtain the 'gold standard' images to evaluate the results. To process the images we have used previously tested methods with changed parameters appropriately for the new type of segmentation (presented in Table I). Evaluation of the results was performed with basic statistical measures as sensitivity, specificity along with Jaccard’s coefficient. All necessary computations were performed in MATLAB.

A. Artificial images synthesis

The process of creating and the usefulness confirmation of the artificial images is fully described in our previous work [10]. In short, it is done using the adjusted version of SIMCEP software [11], [12] and Camera Raw 4.1 module of Photoshop CS5. Lehmussola and co-workers developed SIMCEP to synthesize the full colour fluorescent microscopic images of nuclei or cells’ culture. With our modification it creates images of transmission light microscopy.

Our artificial images were created based on the model images experimentally collected in Hospital Verge de la Cinta in Tortosa, Spain. They were created to resemble their experimental counterparts as much as possible. The number of cells, their size, shape and distribution have been adjusted. We tried to simulate signal degradation, typical for the microscope and camera technical limitations, such as noise, vignetting and blurring.

The great advantage of this technique is that along with the artificial image we obtain the reference image of 'true' objects of interest. This lets us not only compare number of segmented objects and their approximate position but properly evaluate every pixel of the image.

B. Methods of segmentation

Since samples tend to show wide range of characteristics like colour, tone and intensity as well as contrast fluctuations, the locally adaptive thresholding methods seems to be most appropriate. Local threshold is calculated for every pixel with sliding window image processing. Threshold value is based on the intensity of the analysed pixel and its neighbourhood. All methods used in this study are chosen based on the survey [13], and are fully described in previous publication [10], therefore in the following part we present them briefly. All values of parameters used in segmentation algorithms are chosen experimentally and are presented in Table I.

Niblack [14] is the most basic adaptive threshold method; based on local variance.

Sauvola [15] is another local variance method and can be treated as modified version of Niblack’s method.

Bernsen [16] is based on local contrast. Threshold value is calculated as a mean of the minimum and maximum value in neighbourhood of the analysed pixel if the contrast value was high enough.

White [17], basically, if the pixel is considerably (depending on the bias value parameter) darker than its surrounding, it is considered as an object.

Palumbo [18] is using centre-surround scheme. The treated area is divided into near neighbourhood and 4 diagonal windows are far neighbourhood. The tested pixel is supposed to be treated as object when the central window contains the foreground object and the neighbouring windows are filled with background.

Yasuda [19] is local contrast method and consists of four steps. First two are preprocessing; increasing dynamic range in the image, followed by nonlinear smoothing. Then primary thresholding is done, with course marking of background based on local contrast. Finally, precise segmentation to classify rest of the pixels is performed.

Hybrids of Niblack and Sauvola methods are described in previous publication [10]. As the basic methods seemed unsatisfactory in terms of sensitivity and specificity we modified them by adding the contrast condition.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>VALUES OF PARAMETERS USED IN SEGMENTATION METHODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w )</td>
<td>( k )</td>
</tr>
<tr>
<td>Niblack</td>
<td>51</td>
</tr>
<tr>
<td>Hyb.Nib.</td>
<td>51</td>
</tr>
<tr>
<td>Sauvola</td>
<td>51</td>
</tr>
<tr>
<td>Hyb.Sau.</td>
<td>51</td>
</tr>
<tr>
<td>White</td>
<td>51</td>
</tr>
<tr>
<td>Bernsen</td>
<td>51</td>
</tr>
<tr>
<td>Palumbo</td>
<td>21</td>
</tr>
<tr>
<td>Yasuda</td>
<td>51</td>
</tr>
</tbody>
</table>

C. Methods of evaluation

To perform proper evaluation exact position of every object should be known. Fortunately, while we use artificial images we can use the black and white reference image of ‘true’ objects of interest as the ‘gold standard’. Taking into account the result of each segmentation method and ‘gold standard’ image, following measurements are possible: true positive (TP), true negative (TN), false positive (FP), false negative (FN). Based on these parameters, statistical measurement of the performance of segmentation methods can be calculated, such as sensitivity, specificity, and Jaccard’s coefficient.

III. RESULTS

Segmentation was performed on 6 different layers derived from one image. We tried performing segmentation on all of RGB layers. We also tested performance on the Haematoxylin layer acquired by applying colour deconvolution. Additionally, we transformed images from RGB to Lab colour space, and performed segmentation on layers \( L \) and \( b \).

Performing segmentation on RGB layers is computationally most simple approach that does not involve any transformations. The mean values of sensitivity and specificity for every segmentation method on all images are presented in Table II. Comparing mean results of RGB channels only red and green layers give tolerable results. Segmentation of blue channel cannot give good results as values of object and background
are very similar for that layer, as can be seen in Figure 1. Hence, we discard blue channel from further analysis. The best of those three channels is segmentation of red channel. Transformation from RGB to Lab colour space is complex computational task but it may become helpful as results of segmentation on L and b layers are better in comparison to those of RGB layers. Objects of interest (blue immunonegative cells) in Luminescence layer (L) are significantly contrasted from the background. Unfortunately, so are other objects (brown immunopositive cells). Better results are achieved for b layer of Lab colour space. It seems that blue and brown objects can be well separated in this layer since the representation of b axis is yellow-blue. It is understandable because model brown colour consist only of red and green values (in RGB) as well as yellow which has the same components, while on the other end of the axis is blue colour. Mean value of sensitivity for these four layers is between 0.804 and 0.895 while specificity is between 0.912 and 0.930. Haematoxylin layer acquired by colour deconvolution algorithm stands out as best for performing segmentation. It has mean value of sensitivity 0.919 ± 0.047 and 0.915 ± 0.049 specificity, what is better than any other tested layer.

In Table III are presented values of sensitivity, specificity, and Jaccard’s coefficient for all tested methods on all proposed layers (channels) as mean performance on all test images.

As we stated before segmentation of red channel gives better results than green or blue channel. This also is confirmed while analysing separate methods instead of their mean value. For red channel best sensitivity is achieved by Sauvola method which has also worst specificity value. Contradictory is Palumbo method with best specificity and lowest sensitivity. White method has best value of Jaccard’s coefficient and quite good sensitivity and specificity values, both above 0.90. Another to consider is Bernsen method with second highest Jaccard’s coefficient and values of sensitivity and specificity above 0.91.

Layer L from Lab colour space presents information about Luminescence in the image. For this layer Sauvola and Palumbo methods have similar results as for red channel. Highest sensitivity and lowest specificity by Sauvola; highest specificity and lowest sensitivity by Palumbo. Best value of Jaccard’s coefficient is achieved by HybridSauvola method but it has lower value than for red channel of RGB.

Layer b from Lab colour space shows slightly different results. Best value of sensitivity has once again Sauvola method, and also worst specificity. HybridSauvola method seems to increase the specificity of Sauvola method to satisfactory level (0.937) but unfortunately simultaneously lowers the sensitivity value. White method has highest value of specificity but also lowest sensitivity. Palumbo method has acceptable results of sensitivity and specificity, 0.886 and 0.970 respectively, with best value of Jaccard’s coefficient.

Last and possibly most interesting tested layer was Haematoxylin created with colour deconvolution. As it was for other

---

**TABLE II**

Mean values of sensitivity and specificity for all images applied to 6 different monochromatic images derived from the original RGB images

<table>
<thead>
<tr>
<th>layer</th>
<th>sensitivity</th>
<th>specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>R (RGB)</td>
<td>0.895 ± 0.043</td>
<td>0.914 ± 0.044</td>
</tr>
<tr>
<td>G (RGB)</td>
<td>0.827 ± 0.097</td>
<td>0.912 ± 0.048</td>
</tr>
<tr>
<td>B (RGB)</td>
<td>0.501 ± 0.219</td>
<td>0.878 ± 0.080</td>
</tr>
<tr>
<td>H (deconv.)</td>
<td>0.919 ± 0.047</td>
<td>0.929 ± 0.050</td>
</tr>
<tr>
<td>L (Lab)</td>
<td>0.804 ± 0.129</td>
<td>0.915 ± 0.049</td>
</tr>
<tr>
<td>b (Lab)</td>
<td>0.839 ± 0.071</td>
<td>0.930 ± 0.079</td>
</tr>
</tbody>
</table>

---

...
layers Sauvola has highest sensitivity and lowest specificity, while Palumbo has highest specificity and lowest sensitivity. The best results according to Jaccard’s coefficient are achieved by White method, and it has relatively good results of sensitivity and specificity, 0.930 and 0.964 respectively. Other methods worth considering using on this layer are Bernsen and HybridNiblack with second and third best values of Jaccard’s coefficient. Figure 2 shows comparison of 4 most efficient methods that perform segmentation on Haematoxylin layer. Black thick outline of ‘gold standard’ objects is covered with 4 colour overlays representing 4 methods of segmentation; Sauvola—red, Palumbo—green, Bernsen—blue, White—yellow.

Overall, best results are achieved for b layer of Lab colour space and Haematoxylin layer of colour deconvolution. Slightly higher values of specificity can be observed as results of segmentation b layer of Lab colour space. Also, slightly higher values of sensitivity are achieved for Haematoxylin layer. On the whole, according to Jaccard’s coefficient best of all methods is White method segmentation performed on Haematoxylin layer.

IV. DISCUSSION AND CONCLUSION

The main aim of this investigation is to analyse performance of adaptive thresholding methods and to gather knowledge how to design the new better method based on this approach. We evaluated performance of 8 methods of segmentation applied to: separate channels of RBG, L and b layer of Lab colour space, and Haematoxylin layer after colour deconvolution. Best results were achieved for b layer of Lab colour space and Haematoxylin layer of colour deconvolution, so only these two should be taken into further consideration. Comparing results of all 8 methods, we concluded that only 4 have future potential: Sauvola, Palumbo, Bernsen, and White. Hybrid methods are also worth mentioning, they tend to have better overall results than their unmodified counterparts based on the Jaccard’s coefficient.

Sauvola has best overall sensitivity, what means that there are not many false negative pixels, so there are least pixels representing ‘true’ objects not included into segmented image. The area of segmented objects of interest seems to be properly segmented; it does not change shape of segmented objects significantly. This method has problem with vast areas of clear background as there are less intensity variations. The extra segmented objects could be discarded from the results during validation phase or if compared to other segmentation method.

Quite the opposite, Palumbo in general has best specificity. The false positive pixels are minimal for this method. Unfortunately it does not segment full area of objects of interest and they have holes as a result. The main advantage of this method is that there are very little extra segmented pixels.

Bernsen has moderate results. It does not affect object size and the error in area detection is regularly located on border part of object. With too large window the method works slow and tends to misclassify small objects with low contrast, especially when they are located near objects with better contrast.

White method decreases the size of segmented objects. The local threshold level in the White method is dependent on bias which increase intensity of analysed pixel causes that the method perform well in images with high contrast between objects and background. Furthermore, in this investigation

TABLE III

RESULTS OF IMAGE SEGMENTATION.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>r_j</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>r_j</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>r_j</th>
</tr>
</thead>
<tbody>
<tr>
<td>layer</td>
<td>R from RGB</td>
<td></td>
<td>G from RGB</td>
<td></td>
<td>H from RGB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Niblack</td>
<td>0.8869</td>
<td>0.8948</td>
<td>0.7153</td>
<td>0.8462</td>
<td>0.8910</td>
<td>0.6775</td>
<td>0.6100</td>
<td>0.8386</td>
<td>0.4505</td>
</tr>
<tr>
<td>Hyb.Nib.</td>
<td>0.8689</td>
<td>0.9421</td>
<td>0.7845</td>
<td>0.8460</td>
<td>0.9385</td>
<td>0.7437</td>
<td>0.6078</td>
<td>0.9078</td>
<td>0.5079</td>
</tr>
<tr>
<td>Sauvola</td>
<td>0.9357</td>
<td>0.8144</td>
<td>0.6571</td>
<td>0.9064</td>
<td>0.8043</td>
<td>0.6264</td>
<td>0.6937</td>
<td>0.7054</td>
<td>0.4169</td>
</tr>
<tr>
<td>Hyb.Sau.</td>
<td>0.9357</td>
<td>0.9111</td>
<td>0.7787</td>
<td>0.9061</td>
<td>0.9034</td>
<td>0.7452</td>
<td>0.6912</td>
<td>0.8649</td>
<td>0.5351</td>
</tr>
<tr>
<td>Bernsen</td>
<td>0.9148</td>
<td>0.9372</td>
<td>0.7989</td>
<td>0.8626</td>
<td>0.9377</td>
<td>0.7541</td>
<td>0.5248</td>
<td>0.8826</td>
<td>0.4226</td>
</tr>
<tr>
<td>White</td>
<td>0.9024</td>
<td>0.9433</td>
<td>0.8000</td>
<td>0.8494</td>
<td>0.9444</td>
<td>0.7541</td>
<td>0.5283</td>
<td>0.9370</td>
<td>0.4634</td>
</tr>
<tr>
<td>Palumbo</td>
<td>0.7988</td>
<td>0.9409</td>
<td>0.7090</td>
<td>0.6053</td>
<td>0.9478</td>
<td>0.5359</td>
<td>0.0611</td>
<td>0.9502</td>
<td>0.0549</td>
</tr>
<tr>
<td>Yasuda</td>
<td>0.8951</td>
<td>0.9222</td>
<td>0.7595</td>
<td>0.7937</td>
<td>0.9251</td>
<td>0.6741</td>
<td>0.2912</td>
<td>0.9379</td>
<td>0.2550</td>
</tr>
</tbody>
</table>

Where, \( r_j \) - Jaccard’s coefficient, deconv. - colour deconvolution.
we adjusted the window size and values of the parameters especially for the tested images. The idea is to find criteria for automatic adjustment of parameters to fit to processed image. We should consider developing appropriate preprocessing phase that could equalize images [20] before they are segmented as it may improve results and make it easier to adjust parameter values.

As a result of this investigation we believe that there are three ways to improve segmentation of tested methods. First, globally, results of different methods could be merged using logical operands and probability analysis. For example, the results of the most accurate in object localization method could be used as seed points for the other methods with precise local area selection. Regrettably, the computational cost and time of computation could be quite high. Second, it seems object oriented analysis could give good results. Separate objects segmented by the best method could be evaluated by shape or colour features. Unfortunately, the immunonegative cells have wide range of these features, shape form rounded to elongated and colour from light blue to very dark. Also background has often blue tint that could distort the evaluation. Third, development of new method that would merge advantages of best tested methods. Since all four best methods are based on different factors, combining the benefits of variance method, centre-surround scheme, contrast based method, and bias method would not be easy but should give best results. This methodology could be most efficient regarding time and computation cost. As sliding window operations of adaptive thresholding methods tend to be slow it would be best to limit the number of whole image processing as much as possible.

A. Future works

In future work, we plan to apply tested methods of segmentation to experimental images. If the results will not be satisfactory we plan to develop new method of segmentation based on the advantages of best of tested methods.

ACKNOWLEDGMENT

We would like to thank Molecular Biology and Research Section, Hospital de Tortosa Verge de la Cinta, Institut dInves-

REFERENCES