Published in final edited form as:
Comput Med Imaging Graph. 2012 September ; 36(6): 442–451. doi:10.1016/j.compmedimag.2012.05.003.

A General Framework for the Segmentation of Follicular Lymphoma Virtual Slides

Myriam Oger[1,2], **Philippe Belhomme**[2], and **Metin N. Gurcan**[1]

[1]Department of Biomedical Informatics, The Ohio State University, 3185 Graves Hall, 333 W. 10th Avenue, Columbus, OH 43210, USA

[2]GRECAN-EA 1772, Université de Caen Basse-Normandie, Avenue du Général Harris, 14076 Caen Cedex 5, France

Abstract

Follicular Lymphoma (FL) is one of the most common types of non-Hodgkin's Lymphomas in the world. Diagnosis of FLis based on morphological and immunohistochemical characteristics found on tissue sections. Our project's aim is to develop computer-aided analysis tools on virtual slide images (VSI) of lymphoid tissues with the purpose of improving the FL grading performed in malignant follicles. In this paper, we focus on the first step of our work, an automated system for detecting follicles in VSI of lymphoid tissues. To mimic the human expert process, the system works on low-resolution CD20 images and maps the follicle boundaries on high-resolution H&E images.

Keywords

image segmentation; virtual slide; registration; immunohistochemistry; follicular lymphoma; quality control

1. Introduction

Follicular lymphoma (FL) is a group of malignancies of lymphocyte origin that typically arise from the lymph nodes, spleen and bone marrow in the lymphatic system. FL images are characterized by follicular or nodular patterns of growth presented by follicle center B cells consisting of centrocytes and centroblasts. FL represents 22% of non-Hodgkins lymphomas in the world and accounts for 35% of all adult B-cell lymphomas and 70% of low-grade lymphomas in U.S. clinical trials. The grading system of the World Health Organization (WHO) divides FL into three histological grades based on the average number of centroblasts (CB) in ten random standard microscopic high power fields (HPF) [1] representing malignant follicles in H&E stained tissue. Grade 1 has up to 5 CBs/HPF, grade 2 has 6-15 CBs/HPF and grade 3 has more than 15 CBs/HPF. While grades I and II are considered indolent, with long average survival rates and no needs of chemotherapy, grade III is an aggressive disease; it is rapidly fatal if not immediately treated with aggressive chemotherapy. These differences underline the importance for precisely measuring an

Myriam Oger Department of Biomedical Informatics, The Ohio State University, 3185 Graves Hall, 333 W. 10th Avenue, Columbus, OH 43210, USA Tel: (+1) 614-366-2505 Fax: (+1) 614-688-6600 myriam.oger@gmail.com Philippe Belhomme GRECAN-EA 1772, Université de Caen Basse-Normandie, Avenue du Général Harris, 14076 Caen Cedex 5, France Tel: (+33) 2 33 01 46 28 Fax: (+33) 2 33 01 46 21 philippe.belhomme@unicaen.fr Metin N. Gurcan Department of Biomedical Informatics, The Ohio State University, 3185 Graves Hall, 333 W. 10th Avenue, Columbus, OH 43210, USA Tel: (+1) 614-366-2409 Fax: (+1) 614-688-6600 metin.gurcan@osumc.edu.

accurate histological FL grading to guide crucial clinical decisions of timing and type of chemotherapy. However, in a multi-site study, the inter- and intra-rater variability in FL grading by experts has been shown to lie between 61% and 73% [2]. Such a large disparity underscores the need for additional diagnostic tools and computer-based image analysis systems may be such tools.

In conjunction with the aims of the global MIRACLE project (Microscopic Image Processing, Analysis, Classification and Modelling Environment, FP7-PEOPLE-2009-IRSESMarie Curie PIRSES-GA-2009-247091: http://miracle.ee.bilkent.edu.tr/), the BMI Department at the Ohio State University and four European laboratories are currently developing a complete series of linked tools dedicated to FL grading. Aiding this work is the acceptance and application of Computer-aided diagnosis systems (CADs) in medical labs due to recent advancements in imaging technologies. We have been developing tools for computer-aided prognosis of neuroblastoma [3], [4] and grading of follicular lymphoma [5], [6] with promising results. The work described in [6] elaborates some tools for detecting follicles using H&E and IHC stained tissues. Based on these results, we have developed models to describe tissue histology for classification of FL grades [5]. The previous research in this area focused on morphometric analysis of FL images [7] by comparing diagnosis based on three different stains while [8] developed classifiers for sub typing FL. Recently, Neuman *et al* [9] developed image analysis tools for counting nuclei in IHC stained FL tissue images using color features and watershed segmentation.

Both pathologists' and computerized analysis of an FL case starts with identification of follicle regions from H&E-stained images. For the human analysis, ten representative HPFs need to be selected. These HPFs are generally selected to coincide with follicles, as the intra-follicular regions do not contain CB cells. Similarly, the computer analysis, although not limited to 10 HPFs, needs to identify all follicular regions because further analysis will be carried out only in those areas. However, in many cases, the identification of follicles on an H&E slide is challenging because the boundaries of the follicles are not immediately visible. Pathologists often select areas of the slide where these boundaries are more visible or review IHC-stained adjacent slides, where the follicles are better visible. Our computer-aided approach mimics pathologists' use of IHC information in a systematic way. We first detect the follicles in IHC slides and then use this information to detect the corresponding follicles in adjacent H&E slides. This process involves both segmentation of follicles as well as the registration of adjacent slides.

In this paper, we describe a new method for automatically detecting follicle areas in IHC slides based on our previous work [6], in which we followed a feature based clustering approach. Each pixel is represented by a feature vector of color and texture information. In order to capture the color information, the RGB images are converted to the HSV (hue-saturation-value) color space. The S (saturation) channel, which describes how pure the hue is with respect to the white reference, is used as the color information. The texture information is quantified by the homogeneity of the 9x9 neighborhood of each pixel using the co-occurrence matrix approach. The resulting feature vectors are classified using a k-means classifier with k=4 classes: 1) follicles (B-cells), 2) intra-follicular area (T-cells), 3) mixture (T- and B-cells) and 4) background. The resulting follicle candidate images are post-processed morphologically and then ellipses are fitted to the binary follicle candidates. However, fitting an ellipse onto a follicle is not always optimal since their shapes can significantly vary depending on tissue sections. We attempt here to improve the segmentation quality of follicular areas by providing reliable and accurate follicle boundaries while removing the objects that are suspected to be of poor quality. We use serial sections respectively stained according to the IHC (CD20) and H&E protocols. IHC images are highly contrasted, which is well suited for pre-segmenting follicle areas. The resulting

binary masks are then mapped on H&E images after a registration procedure, conducted according to the thin plate spline algorithm as explained in [10] and [11]. We utilized a novel approach to reduce the complications involved in the manual preparation of slides, namely tearing and folds that will appear on both IHC and H&E images. Our approach, based on inpainting [12], reduced the influence of the manual preparation during the segmentation process. Inpainting is a technique for reconstructing lost or deteriorated parts of images and videos either by filling in small image gaps, or generating large image regions by synthesizing texture. In our case, the small image gaps produced by the tears and folds are reduced by this technique.

The paper is organized as follows. In section 2, we present images and software used to illustrate our strategy and is focused on the segmentation process yielding to binary masks of follicular areas. Section 3 presents a visual display of various results and details the proposed method according to an expert's ground truth. Section 4 discusses is the place where points to be improved are discussed before finally concluding in section 5. The general flowchart of our method is presented in Figure 1. The red boxes correspond to the current work; the blue final box is another process out of the scope of this paper.

2. Materials and Methods

2.1. Virtual Slides

In order to achieve the overall follicle segmentation, tissues are stained according to two protocols. The first one is an immunohistochemical staining (IHC) revealing the surface of all mature B-cells (CD20). The second one is the classical Hematoxylin and Eosin staining (H&E) where Hematoxylin (blue) is linked to nuclear areas whereas Eosin (red) is linked to stroma and cytoplasm parts. CD20 images show a real good contrast and are then well suited for pre-segmenting follicles, even at low resolution. But since the FL grading is only conducted in H&E images, an intermediate registration step is necessary to be able to map the pre-defined follicle boundaries found in CD20 images onto H&E images.

Both CD20 and H&E images were digitized with an Aperio ScanScope XT (Aperio, San Diego, CA) at 40x magnification, before being sub-sampled by a factor of 64 (corresponding to 0.625X). Choosing such a large factor is justified in two ways. At first, hematologists are able to mark follicle boundaries on CD20 images. Secondly, this allows us to load the whole image in the RAM of a classical personal computer, thus avoiding the border effects encountered with tiled images. Later, the object boundaries will be over-sampled to fit in the original high-resolution images. At 40X, the resolution is 0.25 μm/pixel and a High Power Field (HPF) represents 0.18 mm^2 (analogous to a rectangle of size 2168×1353 px^2). At 0.625X resolution is 16 μm/pixel, so a HPF may be represented by a rectangle of size 34×21 px^2. To illustrate the proof of concept of our approach, six pairs of CD20-H&E images have been acquired, some of them revealing a non-uniform staining of the slides. Figure 2 shows an example of such a pair. It can be seen that at low resolution, the follicle regions are clearly more visible on CD20 images (Figure 2 (a)) than on H&E images (Figure 2 (b)). The average size of 40X images is 64300×68300 (12.3 GB if uncompressed) and 1005×1067 at 0.625X (3 MB).

2.2. Software and implementation details

Each step in the flowchart shown in Figure 1 corresponds to a script implemented in MATLAB© (The MathWorks Inc., Natick, MA) and makes use of its ‘Image Processing Toolbox.’ The algorithms, developed on a classical PC under Windows, are not optimized for speed. The execution times (around 40 minutes per image) are reduced substantially when the algorithms are implemented as parallelized C++ code, which is the subject of

another study. For the registration step, an extra-package also written in MATLAB is used [10]. The primary sub-sampling step is achieved by calling the DALTON tool, which is an executable code written in pure C++ language [13]. When slides are initially prepared and stained, and due to the manual operations involved, some tears and folds usually appear both on IHC and H&E images. The sizes of the tears/folds is quite large compared to follicles and a “line” crosses many of them. Instead of avoiding processing these follicles, especially in CD20 images, we have chosen to reduce their influence by applying an image restoration process based on inpainting [12].

2.3. Sub sampling and background segmentation

FL images are characterized by follicular or nodular patterns of growth presented by follicle center B-cells consisting of centrocytes and centroblasts. In IHC images, especially for the CD10 and CD20 staining, these follicles appear as separated ‘convex’ regions in which cellular density is higher than in the inter-follicular area.

The proposed algorithm uses low-resolution CD20 images to first pre-determine the follicle boundaries before mapping them on H&E images. For sub sampling CD20 and H&E images, the DALTON tool developed by the French partner in the MIRACLE project has been used [13]. This tool makes use of wavelet transforms for interpolating pixels and is fully parallelized (number of processing cores equal to the number of available cores minus one for the manager).

For each pair of CD20/H&E, the background segmentation is classically achieved thanks to a logical AND on their thresholded *RGB* components followed by hole filling and an area opening [14] of size 5000 (allowing to remove tissue smaller than 1.28 mm^2, that is about 7 HPF). The two binary masks obtained will then be used in the registration step.

2.4. Fold and tear removal

Folds and/or tears (F/T) are visible on all the 12 images we used for this study; they are inherent artifacts of slide preparation. Folds appear as quite large strong black lines more or less regular whereas tears correspond to thinner white lines, with a structure sometimes similar to branches. In some cases, both CD20 and H&E images in a pair have a F/T at the same location (as the serially sections are cut). In other cases, only one image in a pair has a F/T (as the tissue is laid on the slide). Even though the hematologist will not detect any centroblasts in such artifacts on H&E images, the actual presence of a F/T in CD20 images should not be an obstacle for detecting follicle boundaries. But their influence has to be reduced since algorithms based on object borders could be disrupted by the high contrast they involve.

In order to compute a binary mask of folds and tears, the same algorithm is finally conducted on complementary values. The darker pixels for folds (respectively the lighter for tears) are selected by thresholding the RGB components. A skeletonization by morphological thinning [15] is applied and for each skeleton, the number of multiple points is computed. All skeletons without multiple points are retained; they correspond to straight lines, which is the general shape of a F/T. All parts of skeletons for which the length between two multiple points is lower than 64 μm are removed. And finally, all skeletons having more than three multiple points are also removed; as they may correspond to inherent structures of the tissue. Figure 3(a) shows a sub part of a CD20 image with both folds and tears. The lighter areas in the right corner correspond to interfollicular tissue and are not considered as tears. Figure 3(b) shows the binary mask obtained from our algorithm. The “missing” tissue is finally reconstructed by image inpainting according to the MATLAB algorithm of [12] based on a neighborhood texture analysis (Figure 3(c)).

2.5. Registration

To provide the histological grading of a follicular lymphoma (FL), the pathologist has to manually count the average number of centroblasts (CB) in ten random standard microscopic high power fields (HPF) observed in an H&E image at 40X magnification. With a classical optical microscope, each HPF is a circle; its standardized size is 0.18 mm^2. For automated measurements, several parts of follicular areas can be assembled to obtain a 'virtual' HPF, provided that their cumulated area is equal to the HPF size. Even more, the virtual HPF may have an unusual shape such as a triangle, an ellipse, etc. But H&E images are not sufficiently contrasted for delimiting follicle boundaries, so that is why the most common way to solve this problem is to use IHC images such as CD20 staining. Of course, H&E and CD20 images have to come from a pair of serially cut sections so that the morphology between these sections remains pretty similar. Therefore, a registration step is required to be able to map the follicle boundaries found in CD20 images onto corresponding H&E images.

Once the serial tissue sections have been laid on slides before being stained, each slide will not match identically in terms of position and orientation. Moreover, the staining protocols may involve the slides to be placed a few seconds in a microwave, thus yielding to some shrinkage of tissue. The registration procedure has to take into consideration these constraints. Owing to the strong color difference between H&E and CD20 images, the registration algorithms based on intensity patterns via correlation metrics are not well suited. It is preferable to use feature-based methods finding correspondence between lines and contours [10], [11] and it is much faster; borders of the binary masks obtained in the first background segmentation will provide us these contours.

Several areas of tissue can be obtained while extracting one section in the paraffin embedded block and those who are greater than 1.28 mm^2 are retained after background segmentation. In our approach, each area is processed separately; the centroid and major axis orientation is computed in both the H&E (reference) and CD20 (target) images and areas are first translated and rotated according to these values. It can be observed that the H&E could be used for the target and the CD20 for the reference, however, the objective is to perform measurements in the H&E images and as a result, the preservation of both images is required. Lastly, the registration algorithm provides a distortion map for each tissue area. A new CD20 image is then computed on which the following follicle segmentation step will be applied. Figure 4 illustrates the registration procedure with (Figure 4(a) and (b)) the control points of an H&E/CD20 pair before and after shape matching and, (Figure 4(c)), the new registered CD20 image.

2.6. Follicle Segmentation

In this work, segmentation of follicles is performed on CD20 images, where we are able to determine the follicle boundaries. The boundaries are transferred, in a second step, to segment the follicles on H&E images at a higher resolution.

2.6.1. Pre-segmentation in CD20 images—In CD20 stained tissues, follicle areas appear as brown objects (that means essentially red, with green in a smaller quantity) whereas the inter-follicular space appears as a lighter blue area. As a result, we based the color components for detecting follicles on blue and red channels. However, the RGB color space is known to be highly correlated; we thus apply decorrelation stretching [16]. First, the eigenvectors and eigenvalues of Σ_X (the covariance matrix of the initial data in the original RGB space) are calculated and the stretching vector S is computed as:

$$S=\sigma/\sqrt{\lambda_{i,j}} \quad i \neq j \Rightarrow \lambda_{i,j}=0 \tag{1}$$

Here, S is a diagonal matrix, σ is the desired standard deviation of the RGB channels in the output image and $\lambda_{i,i}$ are the eigenvalues of Σ_x. With D being the eigenvectors matrix of Σ_x, each pixel x in the original image is then transformed to y using:

$$y=DS\,D^t x \tag{2}$$

The three components in y are smoothed using median filtering (kernel size 3×3) and an adaptive equalization of histograms is applied. For clarification purposes, this new decorrelated color space is called $R_dG_dB_d$. We have chosen an adaptive equalization rather than a global one since the tissue preparation protocol often yields to staining gradients. Figure 5 shows the initial RGB components and the final R_d,G_d,B_d components for a CD20 image. As shown in Figure 5, the follicles appear more contrasted and homogeneous, especially in the red and blue channels.

Once the $R_dG_dB_d$ components are computed, R_d and B_d are thresholded according to the Otsu algorithm to generate two binary masks BM_R and BM_B [17]. From the binary masks, A_1 in (3) provides the most likely areas of follicles; A_4 in (6) is the union of most likely background and inter-follicular areas whereas A_2 and A_3 in (4) and (5) provide the pixels for which there exists an uncertainty. They can belong to follicles, background or inter-follicular areas and require further processing.

$$A_1=BM_R \cap BM_B \tag{3}$$

$$A_2=BM_R \cap BM_B^c \tag{4}$$

$$A_3=BM_R^c \cap BM_B \tag{5}$$

$$A_4=BM_R^c \cap BM_B^c \tag{6}$$

First, pixels in A_2 and A_3 that are close to pixels in A_1 (that is belonging to the morphological external gradient of A_1) are merged to the FL class if and only if their $R_dG_dB_d$ value is close to the mean $R_dG_dB_d$ value of the adjacent FL. The set S of aggregated points is defined in (7).

$$\begin{gathered}\exists S, \forall x \in \{\delta_B(A_1)\setminus A_1\} \cap \{A_2 \cup A_3\} \\ x \in S \Rightarrow x - \overline{N_{A_1,G}}(x) \leq \varepsilon\end{gathered} \tag{7}$$

In (7), the grid G used is the 8-connexity and B is the square structuring element of size 1. In our application, ε is fixed to 5 for each of the three $R_dG_dB_d$ components. Then S is used to reconstruct [18][19], the connected elements of A_2 ,followed by those of A_3. The morphological reconstruction of g from f is based on the dilation operator conducted until stability or idempotence according to (8). The follicular area, $\{FL\}$, is given by (9). {FL} at this point, still requires refinement.

$$R_g(f) = \delta_g^{(i)}(f) \tag{8}$$

$$\{FL\} = A_1 \cup R_{A_2}^{(i)}(S) \cup R_{A_3}^{(i)}(S) \tag{9}$$

Finally, the concave elements of $\{FL\}$ are separated using the watershed algorithm [20], the initial markers being provided by a distance function computed on the $\{FL\}$ binary mask. In order to limit the over-segmentation, all follicle areas, which are joined by a watershed line whose color intensity is close to the two neighbors, are merged. The objects smaller than 0.0256 mm^2 (approximately 15% of a HPF) are also removed. Figure 6 shows an enlarged part of the resulting mask. The 'ground truth' as delineated by an expert is shown in blue. The green lines correspond to the automated algorithm. Some differences appear for small parts of follicles that have not been retained by the expert but nevertheless are detected by the proposed algorithm.

2.6.2. Transfer in H&E images—Due to slide preparation protocols used for both CD20 and H&E images, their binary masks "tissue vs. background" may not be exactly stackable. In a same tissue area, some parts may have been shifted and/or rotated while the others are not. The final consequence usually results in a tear or a fold at this location. It is the main reason why we propose the inpainting process followed by a registration step to address shape matching [10]. The boundaries obtained by the process on CD20 are transferred on H&E images, where they can be refined.

2.7. Quality control of segmentation

In the proposed algorithm, many choices have to be made for tuning some of its inner parameters, such as the color components to use, some values for thresholding, sizes of structuring elements for morphological operators, minimal area of objects to be kept and so on. To validate the choices made, we compare the automated results with a ground truth provided by an expert. Of course this ground truth may be discussed since there is always an important inter-rater variability between experts and also an intra-rater variability. But in the context of a proof of concept, we only dispose of six binary masks of follicle areas, manually delineated from the CD20 images. By comparing these six binary masks with those provided by the automated system, four scalar values *p, q, r, s* as explained in Figure 7 can be computed.

Also known as the Jaccard index, the Jaccard similarity coefficient is a statistical measure of similarity between sample sets. For two sets, it is defined as the cardinality of their intersection divided by the cardinality of their union as in (10). Even though the Jaccard similarity is largely used in many applications, its value may not reflect the quality of a segmentation, as shown in Figure 8, where $J(A,B) = 0.5$ in both cases, A being the ground truth of the expert and B being the automated segmentation. Some other measures are available and are compared in [21].

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{p}{p+q+r} \tag{10}$$

In order to evaluate the quality of the final segmentation proposed in this paper, the *sensitivity*, *specificity* and *conformity* are computed from True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) as represented by *p, q, r, s*. Figure 9 visually explains the definitions of *p, q, r, s*. For our medical application, TP are the well

detected pixels in follicles, FP and FN are the over-segmented and non-segmented follicle pixels, respectively, while TN are pixels in the background and/or in inter-follicular areas. Given these definitions, *sensitivity*, *specificity* and *conformity* (SSSC) are respectively defined by (11), (12), and (13).

$$sensitivity = \frac{p}{p+r} = \frac{TP}{TP+FN} \tag{11}$$

$$specificity = \frac{s}{s+q} = \frac{TN}{TN+FP} \tag{12}$$

$$conformity = 1 - \frac{q+r}{p} = 1 - \frac{FP+FN}{TP} \tag{13}$$

The *sensitivity* coefficient is a positive value reflecting how many pixels from the ground truth are correctly segmented, while the *specificity* coefficient, also positive, measures how many pixels outside the ground truth are correctly excluded. In the context of classification, *sensitivity* would be the recall for the class ‘follicle’ (also called True Positive Rate) whereas *specificity* would be the recall for the class ‘non-follicle’ (also called True Negative Rate). Finally, the *conformity* coefficient measures the ratio of mis-segmented pixels to the number of correctly segmented pixels; it may be also negative.

It is important to note that the *specificity* involves the computation of true negative pixels TN, mainly the background area, which is strongly related to the dimension of images and above all to the relative area fraction of the tissue. If $TN >> TP$, *specificity* $\rightarrow 1$; the best way to avoid this trap is to compute TN inside the bounding box of the tissue or inside its convex hull.

With a perfect segmentation, these four coefficients are set to 1. In the extraordinary case where the automated segmentation would provide the complementary mask of the ground truth, *sensitivity* and *specificity* would be set to 0 whereas *conformity* would generate a division by 0. With the examples shown in Figure 8, SSSC are {1/2; 1; 1; 0} for the left and {1; $(N\text{-}G_A)/N$; 0; 0} for the right, N being the number of True Negatives and G_A being the area of the ground truth. The higher value obtained when SSSC are multiplied should theoretically correspond to the best segmentation algorithm.

3. Results

For the six CD20 images, we dispose of a ground truth delineated by an expert. They have been used to tune the inner parameters of our algorithm by computing the SSSC coefficients. At the end of the boundaries pre-determination, the last question is to evaluate if splitting concave objects may be useful or not. This splitting is obtained by the watershed operator and followed by the suppression of watershed lines having a similar average intensity as compared with its two neighbors. Table 1 shows the results of SSSC coefficients without splitting. Table 2 shows the results of SSSC coefficients with splitting and merging.

As shown, the *sensitivity* and *specificity* values are practically homogeneous and relatively close to 1 in most cases (except perhaps for images 4 and 5). But *conformity* coefficients are very far from 1, except for images 1 and 3 and, in a relative measure, image 4. The main reason is that the automated algorithm has the ability to identify areas as a ‘follicle’ that the

expert has not even identified. Moreover, the delineated contours are regular, mainly convex, whereas the automated boundaries appear sometimes as fractal curves.

Taking into consideration the differences described, S^1S^2 product is computed in both Tables 1 and 2. The two last rows compute the product of data in the same column. In this trial, a perfect segmentation would have produced the set {6; 6; 6; 6; 36; 1296}. The influence of a reliable ground truth is obvious. Even though the difference is very low (+0.35% for S^1S^2 product and +2.4% for $S^1S^2S^3C$ product), the final splitting-merging step provides the best results and has been retained by the expert.

Table 3 presents the Jaccard coefficients computed for the six CD20 images with and without the final splitting-merging step. A perfect segmentation would also provide a sum of 6 in our trial but the intrinsic definition of $J(A,B)$ confirms that it is not well suited for this kind of application; the two sums are lower than 3.5. In Figure 10, curves of sensitivity, specificity, SS-product and the Jaccard coefficient are displayed for the retained algorithm. As depicted in Table 3, the Jaccard coeffient does not reflect the segmentation quality as its value for image 2 is one of the lowest values and decreases consistently with images 4-6. The Jaccard coefficient is instead, strongly correlated with the conformity coefficients, as is the ground truth provided by the expert.

The pre-determined boundaries obtained at the end of CD20 processing become the binary mask of the follicles on H&E images. Figure 11 shows the binary mask presenting the most likely areas of follicles in Figure 2(b) which we propose as an input to the ‘centroblast detection’ stage, out of the scope of this paper. Without the use of the CD20 twin image, this result would not have been reachable due tothe poor contrast encountered. The binary mask is, of course, enlarged to return back to the original 40X resolution. In this final stage, no ground truth is available so the refinement must be conducted without reference.

4. Discussion

This paper presents a methodology for achieving segmentation of follicle areas in virtual slides of follicular lymphomas, before applying a further step dedicated to full grading. It is a proof of concept, which should not be seen as the ultimate solution but as a general framework showing the feasibility to use both IHC and H&E images for accurate follicle segmentation. Many points have to be discussed and this paper tries to point out the main challenging issues to take into account in future work.

4.1. Virtual slide preparation

For this trial, the twelve images used come from the same laboratory, i.e. they have been prepared according to the same protocol and digitized on the same device. In a future context of a computer-aided diagnosis system (CADS), it is mandatory to become independent from the origin of the VS. Since it is impossible to control or systematize the preparation of slides in a laboratory, one can just hope that the standardized staining protocols are respected and fully automated to guarantee the constancy of quality. With respect to digitizers, they now embed an automatic color correction to standard using an ICC profile [22], which may be applied or ignored. Combined with a decorrelation of *RGB* color components, one can assume that input images are correctly presented. Indeed, a good CADS should also be able raising an alert when staining of input images seems to be far from the expected one.

4.2. Tuning of inner parameters

As in all applications dealing with object segmentation, many inner parameters exist. They are linked to image processing operators used all along the algorithm: threshold level for

background segmentation, size of a median filtering for smoothing, size of structuring elements with morphological operators used to clean binary masks and so forth. Even the sequence of operators themselves may be considered as an inner parameter; the number of possibilities is exponential. In the past, we employed some statistical parameter optimization techniques to select the set of parameters to be used in a CAD system [23] and in our future work, we will employ similar approaches for optimization. In this paper, one of the points we focus on is the way to correctly estimate a segmentation quality. Rather than the Jaccard index, we recommend computation of the *sensitivity*, *specificity* and *conformity* indexes and then refinement of the sequence of operators and/or inner parameter settings with the goal to improve their values. This method is also useful for the comparison of algorithms designed and developed to perform the same task.

4.3. Registration step

The registration step is an intrinsic part of our framework since we have to manage serial sections of tissue. For this trial, we used six pairs of IHC/H&E images for which the relative shape was preserved from one section to its twin neighbor. Due to the limited but non-homogeneous variations of orientation in our application, registration acts mainly as a non-uniform smoothing. In daily practice, however, our algorithm should be applied to image pairs where one image sharply contrasts its corresponding image, especially in cases where a large part is missing from one or both of the pairs and this piece is in a different location on each image. The algorithm used in this trial, based on shape matching, would thus act as a "morphing" process and results would drastically decrease.

Therefore, in a future work, we will have to avoid this trap by using another registration algorithm which can take into account both tissue boundaries and other local features by locally warping some areas of the target image. These algorithms belong to non-rigid transformations. The main challenges will then be to find corresponding points between IHC and H&E images; one solution should be to consider white areas such as blood vessels.

4.4. Boundary transfer on H&E images

After CD20 processing and registration, we may encounter the following obstacles with the boundaries obtained: differences in boundary matching of follicles in H&E images and others where the boundaries will be shifted. Furthermore, folds and tears may appear on H&E images whereas a follicle has been detected at the same location in its twin CD20 image. Future work will consist of refinement of the pre-segmented boundaries using, for example, Active Contours (AC) [25] in H&E images. The initial active contours would be the boundaries obtained at the end of CD20 processing. The AC could avoid the problems of missing parts and overlap, such as folds and tears. This process also has the potential to merge follicular areas that were over-segmented on CD20 images, allowing us to remove areas of poor quality, small size or obscured by a hemorrhagic invasion, resulting in an accurate representation of the follicular areas.

4.5. Validation of results

With the new digital pathology environment, it is quite a challenging proposition to request a pathologist to delineate regions of interest on images of size 100k × 100k. Even a dedicated pathologist will take a very long time to accomplish this task. Additionally, as many studies have shown that the inter- and intra-rater variability in FL grading by experts ranges between 61% and 73% [2], thus making it very difficult, if not impossible, to compare automated results with a gold standard. Coefficients such as *sensitivity*, *specificity* and *conformity*, as opposed to utilizing the common Jaccard index, can help in refining sub-parts of a general algorithm. In this discussion, we propose reducing the expert workload by making use of stereology such as in [24] in order to ask the expert to only delineate regions

of interest for which a point of the grid is present. If the grid spacing fits the size of objects to be detected, the ground truth will be statistically reliable. By comparing only the corresponding elements, thanks to a morphological reconstruction, the SSSC coefficients will provide a good estimation of the global segmentation quality.

5. Conclusion

In this paper, we present a general framework for segmenting follicular areas before performing histological grading. Following the current practice of the pathologist, we start from a low resolution IHC image, in which the contrast is known to be correct, to generate a mask of the follicular boundaries. We then map these boundaries onto a corresponding registered H&E image. Our algorithm is built from modular elements that are easily modifiable and allow us to make changes to increase performance This allows us to control quality based on segmentation and to develop training sets correspondingly. Despite the huge size of virtual slide images, our algorithm is able to process them in one pass on a personal computer. Its outcome is a binary mask of likely follicular areas in which high power fields will be defined. It thus allows for the following follicular lymphoma grading system (i) to process the whole slide at high resolution, (ii) to obtain statistics on grading distribution and (iii) to assess the tumor heterogeneity. We also propose a general framework for the evaluation of segmentation results, which takes into account multiple facets of segmentation errors.

Acknowledgments

This work is supported in part by European Commission Seventh Framework Program with Grant No: 247091 (MIRACLE), in part by the Regional Council of Lower Normandy (France) and in part by Award Number R01CA134451 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the grant providers. The authors would like to thank Gerard Lozanski, MD for his guidance, insightful discussions and his help in the creation of the ground truth.

References

1. Jaffe, ES.; Harris, NL.; Stein, H.; Vardiman, JW. World health organization classification of tumours of haematopoietic and lymphoid tissues. IARC Press; 2001.
2. The Non-Hodgkin's Lymphoma Classification Project. A clinical evaluation of the international lymphoma study group classification of non-hodgkin's lymphoma. Blood. 1997; 89(11):3909–18. [PubMed: 9166827]
3. Gurcan MN, Kong J, Sertel O, Cambazoglu BB, Saltz J, Catalyurek U. Computerized pathological image analysis for neuroblastoma prognosis. Annual Symposium, American Medical Informatics Association. 2007:304–8.
4. Sertel O, Kong J, Shimada H, Catalyurek U, Saltz J, Gurcan MN. Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. Pattern Recognition. 2009; 42(6):1093–103. [PubMed: 20161324]
5. Sertel O, Catalyurek U, Lozanski G, Saltz J, Gurcan MN. Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. The Journal of Signal Processing Systems for Signal, Image, and Video Technology. 2009; 55(1):169–83.
6. Sertel, O.; Kong, J.; Lozanski, G.; Catalyurek, U.; Saltz, J.; Gurcan, MN. Medical Imaging 2008. Vol. 6915. Computer-Aided Diagnosis; Computerized microscopic image analysis of follicular lymphoma.; p. 691535-691535-11.
7. Swerdlow S, Pelstring R, Collins R. Morphometric analysis of follicular center cell lymphomas. Am J Pathol. 1990; 137(4):953–63. [PubMed: 2221019]
8. Firestone L, Preston K, Nathwani B. Continuous class pattern recognition for pathology, with applications to non-Hodgkin's follicular lymphomas. Pattern Recognition. 1996; 29(12):2061–78.

9. Neuman U, Korzynska A, Lopez C, Lejeune M. Segmentation of stained lymphoma tissue section images. Information Technologies in Biomedicine. 2010:101–13.
10. Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 24(4):509–22.
11. Simonson K, Drescher S, Tanner F. A statistics-based approach to binary image registration with uncertainty analysis. IEEE Pattern Analysis and Machine Intelligence. 2007; 29(1):112–25.
12. Criminisi A, Perez P, Toyama K. Object Removal by Exemplar-based Inpainting. IEEE Conference on Computer Vision and Pattern Recognition. 2003; 2:721–8.
13. Plancoulaine B, Fafin-Lefevre M, Herlin P. Sub-sampling spotting and cloning: a pre-processing strategy for image analysis of cell nuclei on virtual cytological slides. Image Analysis and Stereology. 2011 to be published.
14. Vincent, L. NATO Shape in Picture Workshop. Driebergen; The Netherlands: 1992. Morphological area openings and closings for grayscale images.; p. 197-208.
15. Meyer F. Skeletons and perceptual graph. Signal Processing. 1989; 16:335–63.
16. Karvelis P, Fotiadis D. A region based decorrelation stretching method: application to multispectral chromosome image classification. ICIP 2008, 15th IEEE International Conference on Image Processing. :1456–9.
17. Otsu N. A threshold selection method from gray-level histogram. IEEE Transactions on System Man Cybernetics. 1979; 9(1):62–6.
18. Serra, J. Image Analysis and Mathematical Morphology. Academic Press; London: 1982.
19. Coster, M.; Chermant, JL. Précis d'analyse d'images. 2ème ed. Presses du CNRS; Paris: 1989.
20. Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1991; 13(6):583–98.
21. Chang HH, Zhuang AH, Valentino DJ, Chu WC. Performance measure characterization for evaluating neuroimage segmentation algorithms. NeuroImage. 2009; 47:122–35. [PubMed: 19345740]
22. International Color Consortium (ICC). Specifications. http://www.color.org
23. Gurcan MN, Sahiner B, Chan HP, Hadjiiski LM, Petrick N. Selection of an optimal neural network architecture for computer-aided detection of microcalcifications – comparison of automated optimization techniques. Medical Physics. 2001; 28(9):1937–48. [PubMed: 11585225]
24. Herlin, P. Stereology on virtual slides. Aperio-Webinar; 2009. http://www.aperio.com/pathology-events/pdfs/Stereology-Webinar-Paulette-Herlin.pdf
25. Vese LA, Chan TF. A multiphase level set framework for image segmentation using the Mumford and Shah model. International Journal of Computer Vision. 2002; 50(3):271–93.

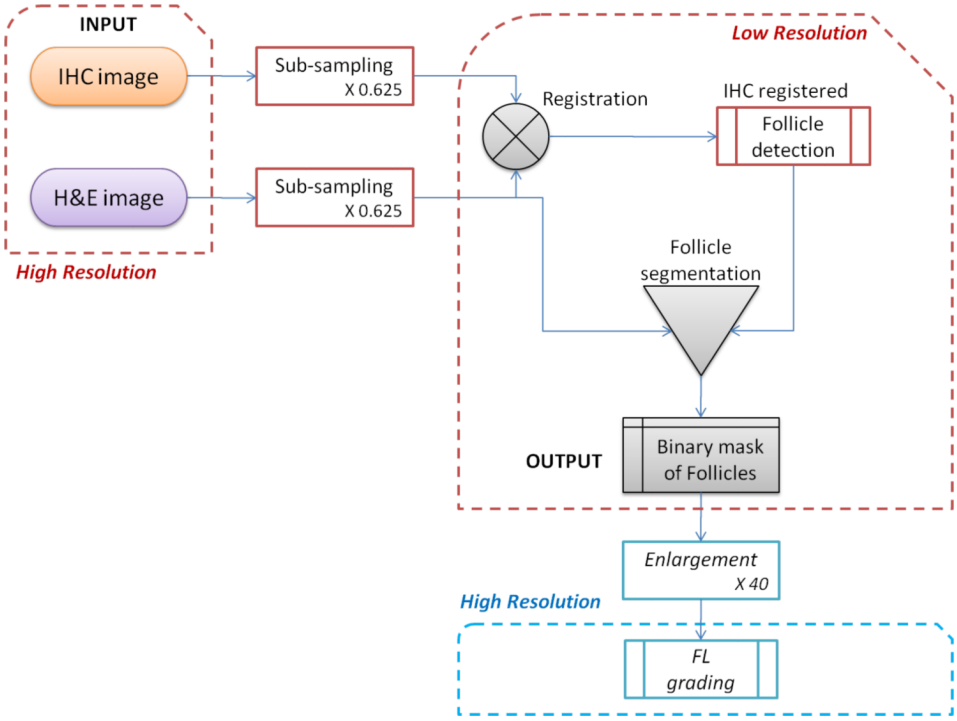

FIGURE 1.
Flowchart of the process.

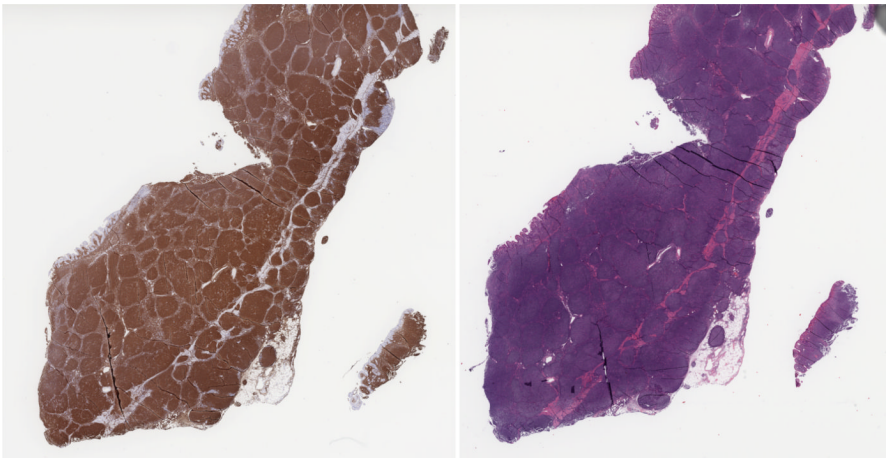

FIGURE 2.
(a) CD20 stained tissue. (b) H&E stained tissue.

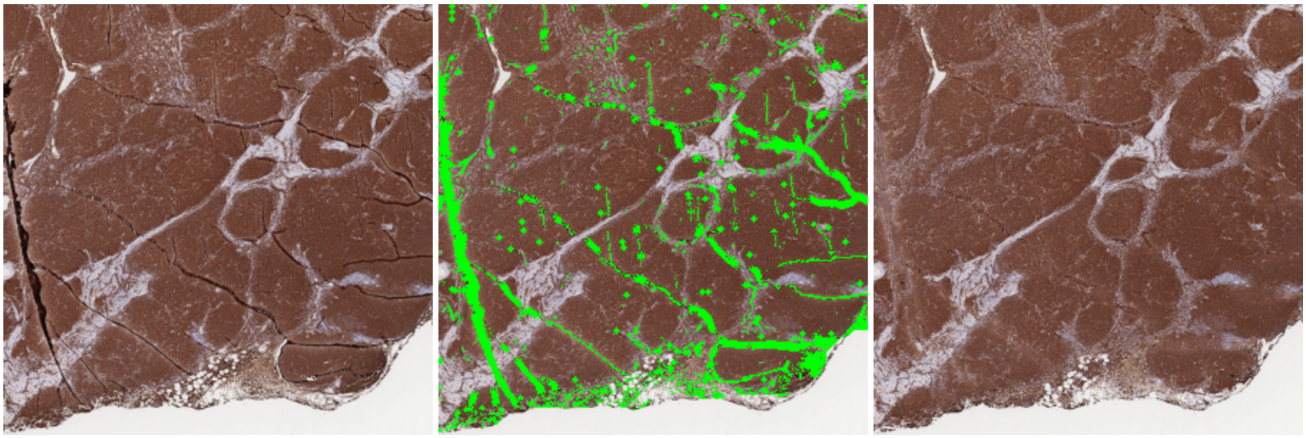

FIGURE 3.
(a) Folds and tears on a CD20 image. (b) Binary mask obtained by skeletonization. (c) Reconstruction by image inpainting.

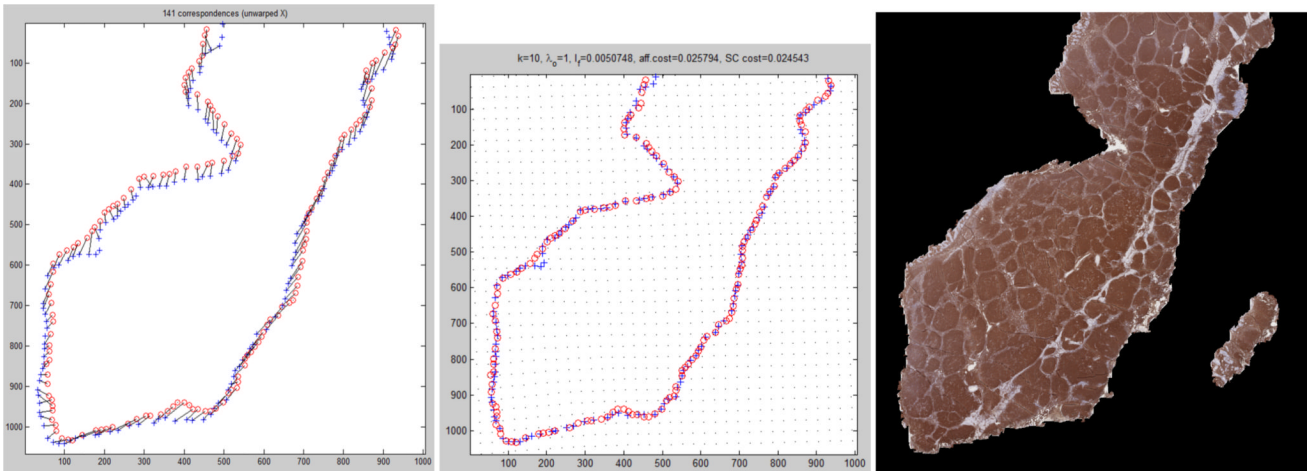

FIGURE 4.

Registration using H&E (reference) and CD20 (target). (a) Control points from the boundaries of the two initial binary masks. (b) Shape matching of control points. (c) Reconstructed CD20 image and its segmented background (to be compared to Fig 2(a)).

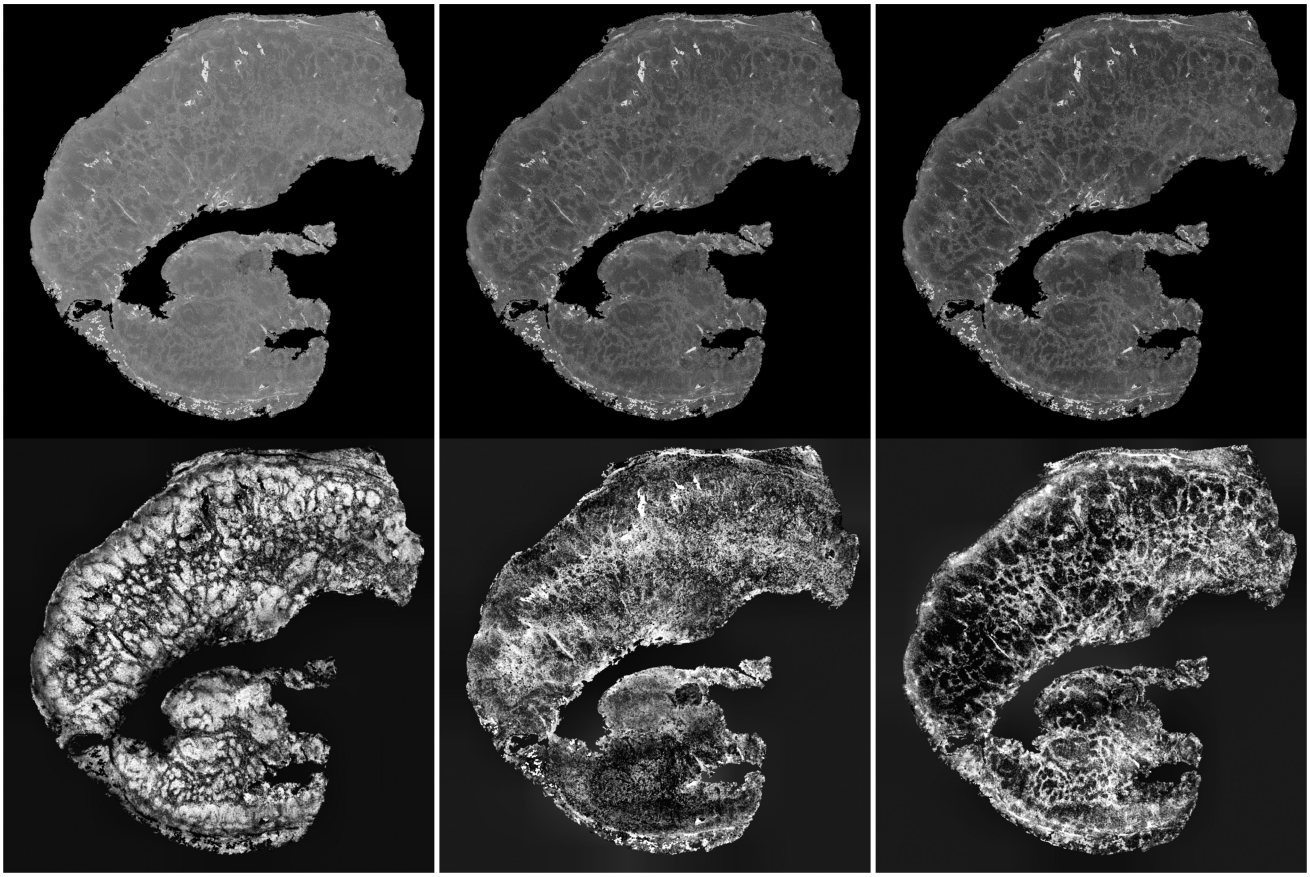

FIGURE 5.
CD20 image before and after decorrelation stretching. (a),(b),(c) Respectively R,G,B original components. (d),(e),(f) Respectively Rd,Gd,Bd components after decorrelation, median filtering and adaptive equalization of histograms.

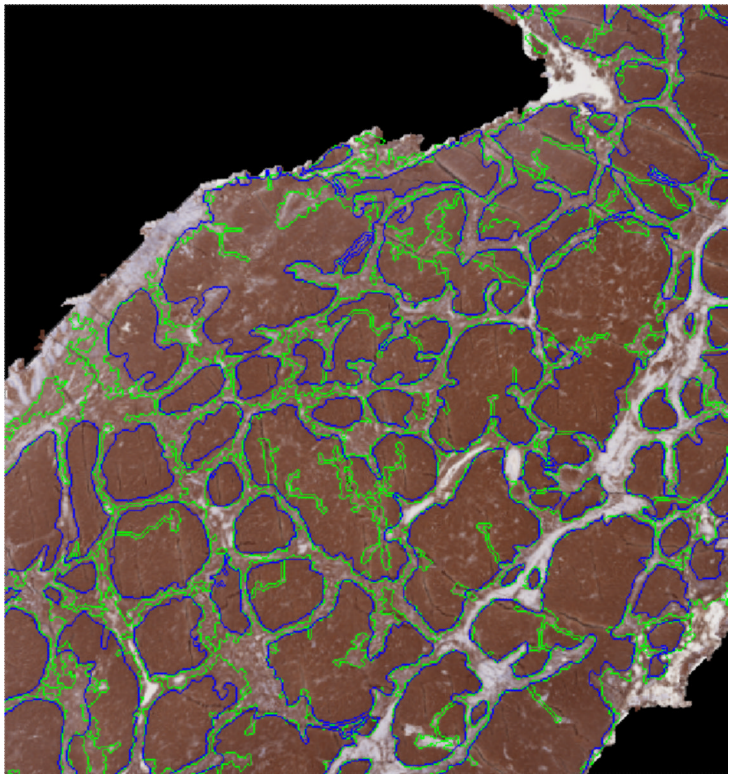

FIGURE 6.
Pre-segmentation on a CD20 image. Automated (green) vs. Expert ground truth (blue).

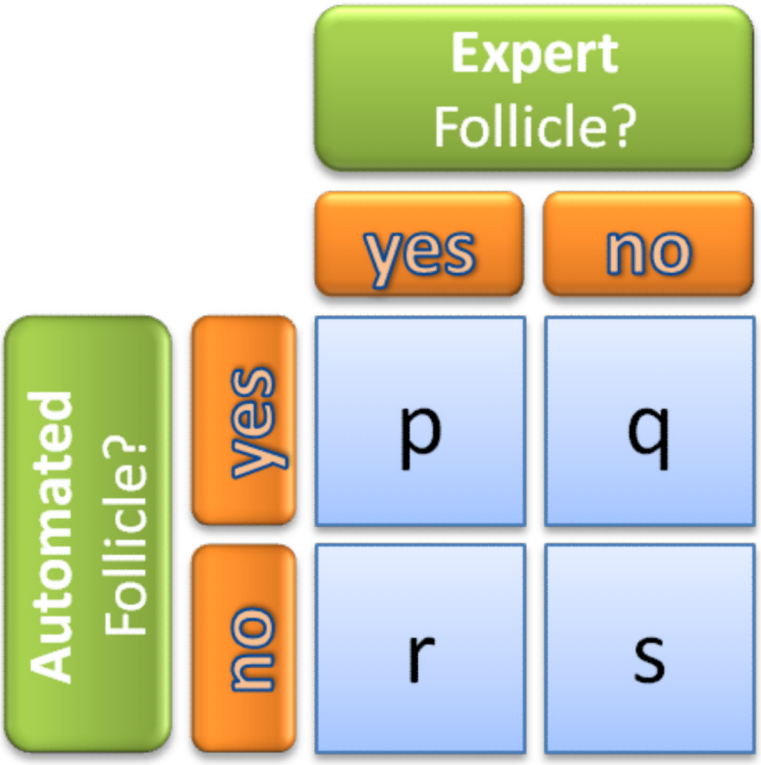

FIGURE 7.
Comparing manual and automated segmentation. Four statistical values.

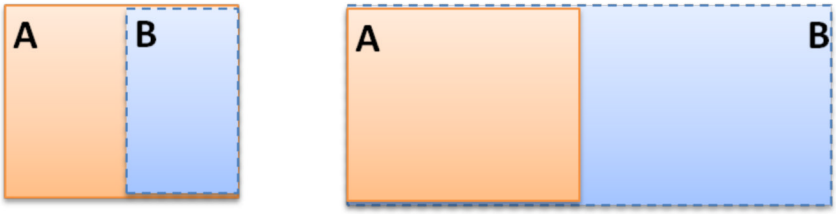

FIGURE 8.
Limits of Jaccard index: Two very different segmentations yielding to the same value for J(A,B).

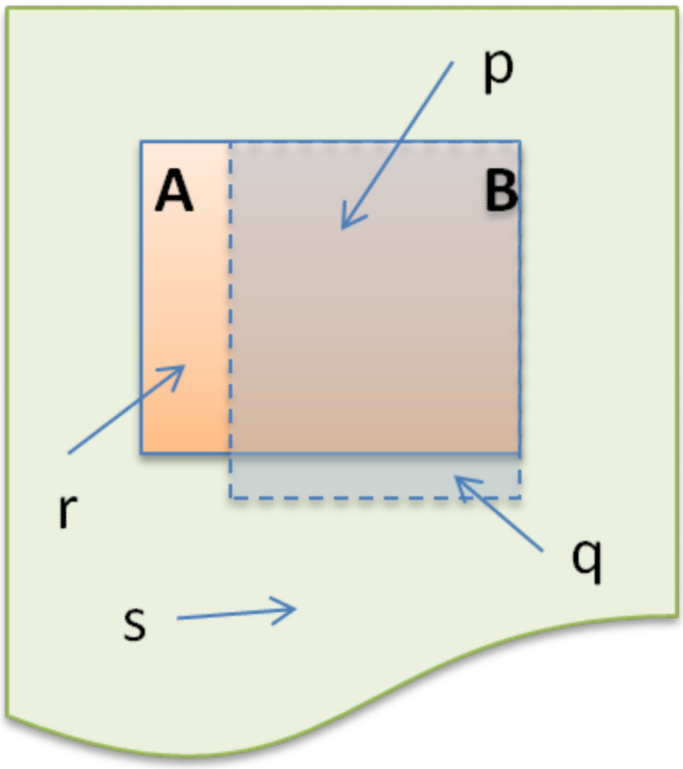

FIGURE 9.
Visual definition of p,q,r,s. A is the ground truth, B the automated segmentation.

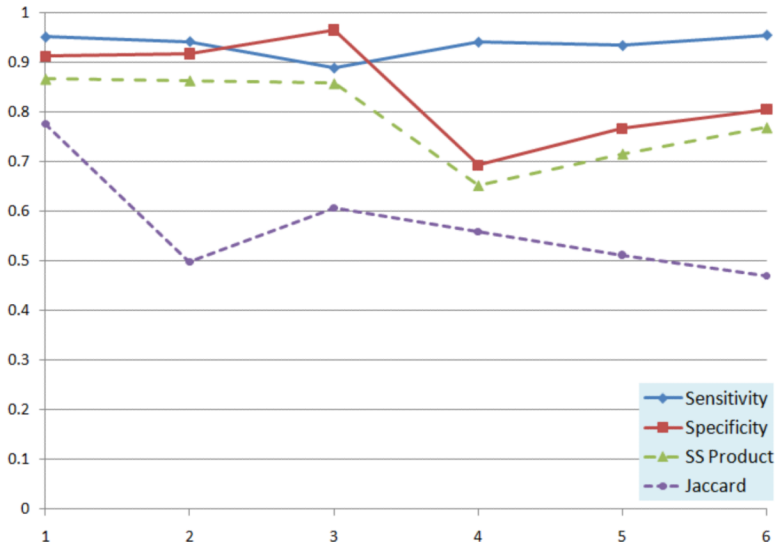

FIGURE 10.
Quality control of the final segmentation in CD20 images. Sensitivity, specificity and their product are compared with the Jaccard coefficient.

FIGURE 11.
Final segmentation as a binary mask of follicles.

Table 1

SSSC coefficients without final splitting/merging procedure.

	CD20 image number						
	1	**2**	**3**	**4**	**5**	**6**	**Sum**
Sensitivity[1]	0.942	0.944	0.888	0.945	0.924	0.950	5.593
Specificity[2]	0.910	0.918	0.964	0.677	0.773	0.811	5.053
Conformity	0.693	0.008	0.330	0.182	0.051	-0.103	1.161
S^1S^2 Product	0.857	0.867	0.856	0.640	0.714	0.770	28.3

Table 2

SSSC coefficients with final splitting/merging procedure.

	CD20 image number						
	1	**2**	**3**	**4**	**5**	**6**	**Sum**
Sensitivity[1]	0.951	0.942	0.889	0.941	0.934	0.955	5.612
Specificity[2]	0.912	0.917	0.965	0.693	0.767	0.805	5.059
Conformity	0.712	-0.008	0.353	0.210	0.048	-0.127	1.188
S^1S^2 Product	0.867	0.864	0.858	0.652	0.716	0.769	28.4

Table 3

Jaccard coefficients for the two different methods.

Jaccard Coefficient	CD20 image number						
	1	2	3	4	5	6	Sum
Without splitting	0.765	0.502	0.599	0.550	0.513	0.476	3.405
With splitting	0.776	0.498	0.607	0.559	0.512	0.470	3.422