Computerized Medical Imaging and Graphics 34 (2010) 453–461

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: A quantitative evaluation

S. Di Cataldo*, E. Ficarra, A. Acquaviva, E. Macii

Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

Article history:
Received 26 March 2009
Received in revised form 2 December 2009
Accepted 9 December 2009

Keywords:
Immunohistochemistry
Tissue images
Nuclear segmentation
Morphological operators
Active contours

ABSTRACT

In this paper we address the problem of nuclear segmentation in cancer tissue images, that is critical for specific protein activity quantification and for cancer diagnosis and therapy. We present a fully automated morphology-based technique able to perform accurate nuclear segmentations in images with heterogeneous staining and multiple tissue layers and we compare it with an alternate semi-automated method based on a well established segmentation approach, namely active contours. We discuss active contours' limitations in the segmentation of immunohistochemical images and we demonstrate and motivate through extensive experiments the better accuracy of our fully automated approach compared to various active contours implementations.

1. Introduction

Bio-image processing leads to the development of diagnostic tools helping pathologists and genetists in the quantification of biological activities related to diseases as well as in the design of targeted therapies [1–3]. Immunohistochemistry (IHC) [4] is a well established imaging technique that exploits intensity of stains in tissue images to quantify the protein activity related to cancer development. Image processing techniques in this context are devoted to the accurate and objective quantification and localization of such activity in specific regions of the tissue such as cytoplasm, membranes and nuclei.

In this paper we address the problem of nuclear segmentation in immunohistochemical (IHC) tissue images, that is critical for further investigations on target protein activity [5]. The main motivations of our work are: (i) to overcome the limitations of the existing methods with a fully automated morphology-based technique; (ii) to provide a better segmentation accuracy compared to a widely used and trusted method in Computer Vision and medical image processing, namely active contours, which is not well-performing in IHC tissue image segmentation. A deep quantitative evaluation of the accuracy of our morphology-based methods and active contours is provided.

Nuclear segmentation in IHC tissue images is a challenging task due to the intrinsic complexity of tissue images and to the many sources of variability that affect IHC technique. Main challenges are related to the non-predictable size and shape non-uniformity induced by the pathological process and by the lack of homogeneity of nuclear regions both in terms of morphological and chromatic features. From the morphological viewpoint, problems for segmentation arise from the presence of overlapping cells and nuclei, which are extremely difficult to separate, as well as to the presence in the sample of other non-pathological structures (e.g. connective tissue structures, blood vessels, lymphocytes, etc.) which may lead to segmentation errors. From the chromatic viewpoint, nuclear regions are characterized by non-uniform stain intensity and color, thus preventing a trivial segmentation based on color separation. In fact, the superposition of tissue layers as well as the diffusion of the dyes on the tissue surface may bring the stains to contaminate the background or other cellular regions which are different from their specific target. Moreover, different portions of the same tissue area may be not equally enlightened and stained.

We present a fully automated nuclear segmentation technique that exploits morphological and chromatic characteristics of the tissue in order to recognize images with nuclear, cellular membrane and cytoplasm activations; this allows the method to adapt to the characteristics of the image without any user-interaction. In our technique we separate nuclei from background by exploiting local analysis of intensity distribution in the neighborhood of each cell, that depends on the expected size of the cell and then on image resolution; this minimizes the effect of noise as well as of uneven staining and inhomogeneous enlightenment; then we separate clusters through our enhanced watershed technique which

* Corresponding author. Tel. +39 011 564 7072; fax: +39 011 564 7099.
E-mail addresses: santa.dicataldo@polito.it (S. Di Cataldo), elisa.ficarra@polito.it (E. Ficarra), andrea.acquaviva@polito.it (A. Acquaviva), enrico.macii@polito.it (E. Macii).

0895-6111/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2009.12.008

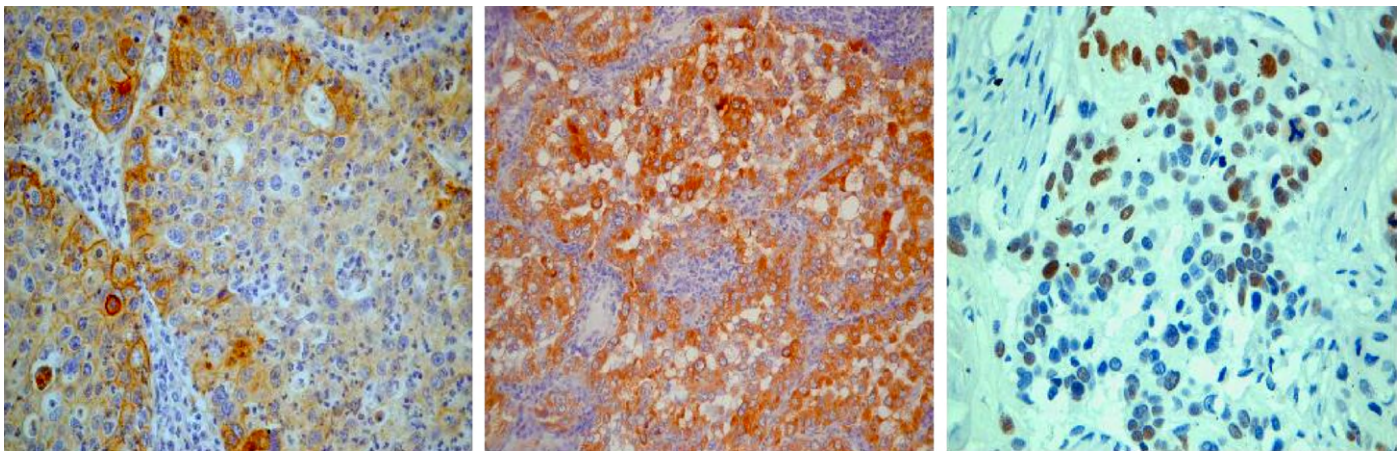

Fig. 1. IHC tissue images. From the left: membrane activations (400×); cytoplasm activations (200×); nuclear activations (800×).

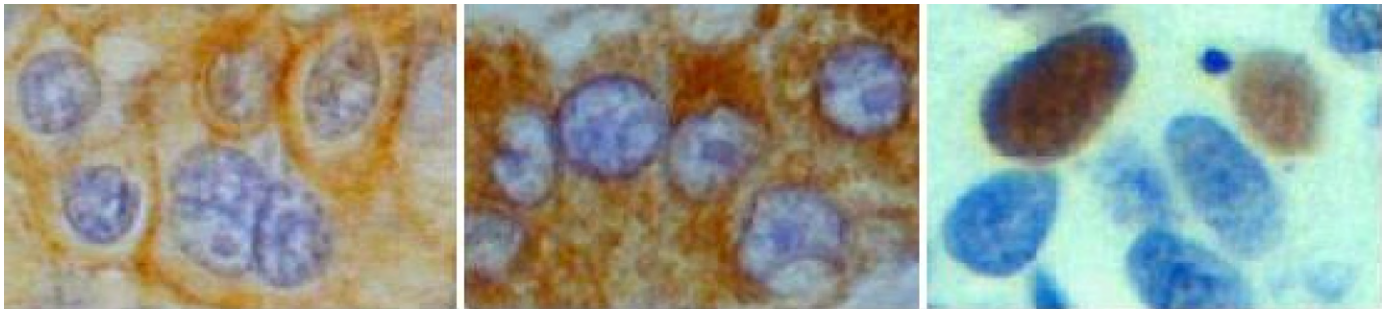

Fig. 2. Magnified particular of nuclei in IHC images. In images with membrane activity (left) or cytoplasm activity (center) nuclei are blue colored; in images with nuclear activity (right) they may be either blue or brown colored. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

merges oversplit nuclei by exploiting specific chromatic characteristics of the cells.

Our morphology-based technique is compared to active contours, that is a well known and trusted technique in biomedical image processing [6–10]. However in the context of IHC tissue imaging remarkable intensity variations inside and outside the cellular regions to be segmented stress the limitations of active contours, calling for morphology-based approaches where the specific features of IHC images can be effectively expressed and exploited. Extensive experimental results show that our method is more accurate than various formulations of active contours in segmenting nuclei in IHC tissue images. Besides accuracy, another key metric to evaluate a bio-image processing tool is its autonomy with respect to operator input. Active contours is generally a semi-automated method that requires the operator to define a curve which the algorithm cripples iteratively to fit the boundary of the target region. On the contrary, the morphology-based technique we present in this paper provides a completely automated nuclear segmentation. To enable fully automated procedure, we also developed a pre-classification step to distinguish images showing nuclear protein activation from membrane and cytoplasm activation, so that a pathologist can give any set of tissue images as input to the tool which will recognize the type of activation and provide the quantification result as output.

A quantification technique for detection of protein activity in cytoplasm/membranes has been already presented in our previous paper [5]. In this work, we integrate the proposed nuclei segmentation technique in a fully automated tool for protein activity quantification in cancer tissues in all cell compartments (either membrane, cytoplasm or nuclei). Compared to our previous work [11], this paper provides a new set of experimental results on which a novel statistical method is applied in order to obtain more extensive comparison of morphology-based and active contours technique. Moreover we extend the examination to all the most widely known typologies of formulations of active contours present in literature, providing for each typology exhaustive discussion of its limitations in the segmentation of IHC tissue images.

The paper is organized as follows. Section 2 explains the morphological procedure. Section 3 discusses active contours based approaches. Section 4 describes the implementation details and Section 5 shows experimental results. Section 6 concludes the paper.

2. Proposed approach: morphology-based procedure

In this section we present our fully automated morphology-based procedure for nuclear segmentation in IHC tissue images. In this procedure well known morphological operators and novel techniques which exploit morphological and chromatic characteristics of the tissue are applied in order to solve the technical problems related to the segmentation of nuclei in IHC images that were already discussed in the previous section.

The accurate tracking of nuclear membranes is fundamental in IHC analysis. This is true not only for images with nuclear activations but also for images showing membrane as well as cytoplasm activations. In fact nuclear segmentation is usually the first step for the segmentation and quantification of protein activity of the other cellular compartments (i.e. cellular membranes and cytoplasm) [5] therefore a lack of accuracy of this step may alter IHC analysis in a substantial way.

Differently from recent methods in literature [12–15] our technique does not require any manual intervention by the operator,

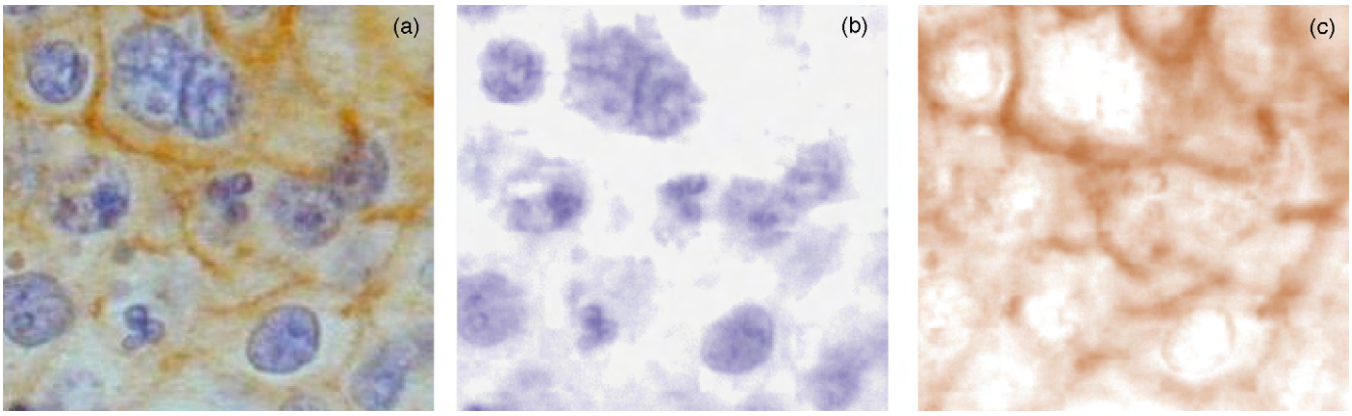

Fig. 3. Separation of blue (b) and brown stain (c) as obtained through color deconvolution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

thus overcoming the subjectivity and long time-consumption of manual analysis. Moreover, our method is a comprehensive procedure able to process images with nuclear as well as with membrane or cytoplasm activity without needing any *a priori* knowledge about the sample. This is possible thanks to a pre-classification step that enables the automated adaptation of the procedure to the type of the image.

The images considered in this paper are characterized by blue stain (Hematoxylin, H) for highlighting the tissue structure, which represents the background. Brown stain (Diaminobenzidine, DAB) reveals protein activity. Specific proteins are active in the cellular membranes of the pathological cells, others in the cytoplasm, others in the nuclei. As shown in Fig. 2, in images with membrane or cytoplasm activity nuclei are always blue colored since the brown stain reveals protein activity that is localized respectively in the cellular membrane or in the cytoplasm of the cells; in images with nuclear activity the protein activity is rather localized in the nuclei, so that nuclei may be either blue or brown colored.

In particular, nuclei that are negative to the target protein are blue, nuclei that are positive are brown.

It follows that a preventive discrimination of the location of protein activity as nuclear or non-nuclear (including in the latter either membrane or cytoplasm) is critical for fully automated nuclear segmentation in that it leads to infer the specific color characteristics of the nuclei and to automatically adapt the segmentation procedure to those characteristics without any *a priori* knowledge about the type of the image. The main steps of the procedure are the following.

2.1. Separation of stains

The original RGB image is separated into two monochromatic images containing respectively the contribution of the brown stain (i.e. DAB) and of the blue stain (i.e. H), as reported in the example of Fig. 3. For this purpose, a specific color deconvolution algorithm [16] is used, since it was shown to achieve better results than other color segmentation methods, especially in IHC applications [17]. Furthermore, it allows accurate separation not only of H and DAB but of all the standard histological stains (e.g. H–E, H AEC, etc.) as well as of any other stains, provided that their RGB vectors are experimentally determined. The color deconvolution plugin [18] was integrated to our algorithm.

2.2. Pre-classification of images

The image is pre-classified as image with nuclear or non-nuclear activity by analyzing the distribution of the brown stain in the image. As shown by Fig. 1, images with nuclear activity are characterized by a granular distribution of the brown dye: brown-colored regions have a well-defined round shape and nuclear-like dimension; in fact, the brown stain is always localized in the nuclei. On the contrary, in images with membrane/cytoplasm activity brown regions are less characterized in shape and dimension: nuclear-like regions may be present again, but they are not prevalent as in the former case.

Therefore the amount of nuclear-like brown regions can be effectively used to discriminate between nuclear and non-nuclear activity. In particular, in our technique the percentage of brown pixels belonging to regions with nuclear-like shape and dimension is adopted as classification feature.

First of all, our method detects the portions of tissue stained by brown. This is achieved through automated thresholding, i.e. by imposing an intensity threshold to the brown monochromatic component and selecting pixels whose intensities are lower than the threshold. The optimal threshold is obtained through the well established Isodata algorithm [19]. Then nuclear-like particles are highlighted on the base of their shape and dimension: regions with low circularity or very small area are selectively removed.

Finally the percentage of pixels of stained tissue belonging to nuclear-like regions is calculated. This is a measure of the amount of nuclear-like activations in the image. Nuclear-like activations below 20% classify the image in the category of non-nuclear protein activity: in fact protein activations are predominantly localized in the cellular membranes or in the cytoplasm; nuclear-like activations above 20% classify the image in the category of nuclear protein activity.

2.3. Segmentation of nuclear membranes

2.3.1. Binarization

Nuclei are separated from background through automated thresholding of the monochromatic images coming from stains' separation. Since intensity level may vary among different regions of the sample because of inhomogeneous illumination and inconsistent staining, that are typical problems of IHC imaging [20], a *local adaptive threshold* [21] dependent on the local distribution of the intensities is applied. This leads to a better accuracy than traditional global thresholding. The procedure is customized through a statistical analysis of local intensity distribution. This analysis aims to minimize the effects of unrepresentative pixel values due to noise. The size of neighborhood is selected taking into account the resolution of the image and the dimension of nuclei (for details see [22]). An example of intermediate results after binarization is reported in Fig. 4(b).

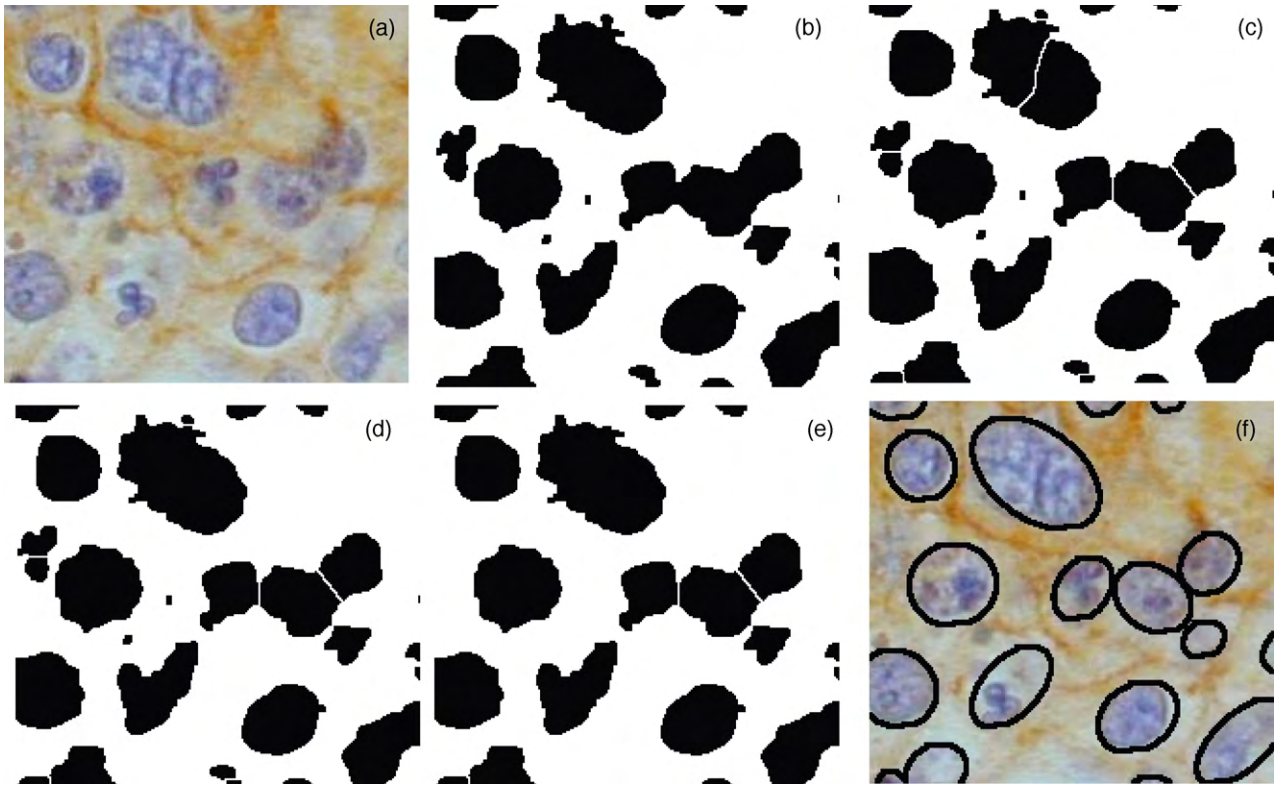

Fig. 4. Intermediate results of nuclear segmentation. Original image (a); binarization through local adaptive thresholding (b); watersheds (c); improved watersheds that exploits chromatic information within the nuclei to remerge over-segmentations (d); postprocessing (e); final nuclear segmentation (f).

For images with non-nuclear activity the only contribution of the blue stain is considered to distinguish nuclei from background. On the contrary, for images with nuclear activity both the contributions of the blue and the brown stains are binarized and merged through binary union.

2.3.2. Separation of clustered nuclei

Overlapped particles are separated through watershed algorithm [23]. As it is well known, intensity variations may lead watersheds to over-segmentation errors (i.e. to split individual nuclei in more than one particle). See Fig. 4(c) for an example. This problem arises especially in images with membrane/cytoplasm activations, due to the local intercontaminations of the two dyes used to highlight different cellular structures (i.e. nuclei and cellular membranes, or nuclei and cytoplasm, respectively). Unfortunately these intercontaminations occur quite often, being a consequence of the staining procedure.

In our approach over-segmentations are prevented through selective remerging of the split nuclei. First of all, the couples of particles split by watersheds are selected and the interposed area between each couple is scanned, thus computing the relative amount of blue and brown pixels. Couples of particles with a prevalence of brown pixels (i.e. belonging to cellular membrane or cytoplasm) in the interposed area are interpreted as two separated nuclei and left unchanged. On the contrary couples with a prevalence of blue pixels (i.e. belonging to nuclei) in the interposed area are interpreted as a single nucleus and remerged (see [22]). An example of intermediate results after selective remerging of over-segmentations is reported in Fig. 4(d).

2.3.3. Postprocessing

As already mentioned in the Introduction, the presence in the same sample of different types of tissues and cells adds a further element of complexity to the segmentation of IHC tissue images; in particular, lymphocytes that are usually smaller than nuclei but may appear very similar in shape and color may lead to segmentation errors. This problem is handled in our procedure through size analysis. Particles whose area is considerably lower than the average area of all the detected nuclei are not included in the final segmentation. See Fig. 4(e) for an example of intermediate results after postprocessing.

The boundaries of the detected nuclei are extracted and outlined, as reported in Fig. 4(f). The best-fitting ellipse of each boundary is calculated in order to smooth and regularize the curve in presence of noise and/or stain inhomogeneities.

3. Alternate approach: active contours

Our morphology-based method is compared in this work with active contours (also called *snakes*), a highly popular approach in Computer Vision that was successfully used in several applications including medical imaging [24,25]. In this section we describe how we applied active contours to IHC tissue image segmentation.

First of all, preliminary tests were carried out in order to select the most effective formulations of active contours for the segmentation of nuclei in IHC tissue images; then we performed extensive comparisons with our morphology-based approach.

In our tests active contours with shape priors, which integrate in the evolution equation prior shape knowledge about the segmented objects [26–30], obtained poor results and were disregarded. In fact nuclei in IHC pathological tissue may be affected by unpredictable size or shape variations induced by the tumor or by mechanical and thermal stress related to the preparation of the sample, thus calling for more flexible shape models.

We considered for our purpose both *parametric* [24,31,32] and *geometric* active contours [33–37]. Geometric approaches are renowned for being computationally more complex than the for-

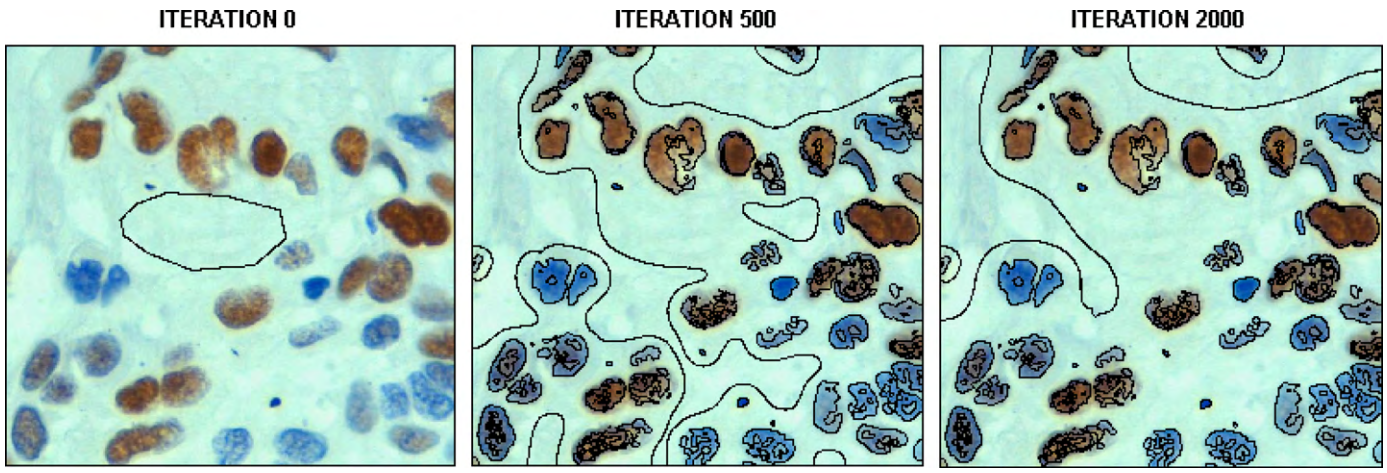

Fig. 5. Segmentation of nuclei with geometric active contours. As it is visible from the image, after 2000 iterations all the nuclei have been detected and bounded by the evolving curve.

mer due to their higher dimensional formulation, but also for offering a great flexibility in accounting for topological changes during the curve evolution and for being less dependent on the curve initialization thanks to the ability of level sets to split, merge and vanish across the image. This flexibility would allow to detect several nuclei in the sample by initializing only one or few curves in the image. For example, in Fig. 5 one single curve initialized in the center of the image was sufficient for the active contour to reach and detect all the nuclei in the slide. It is a state of the art approach with a variational level set formulation that is particularly intended for the segmentation of objects with intensity inhomogeneities and weak boundaries. Nevertheless geometric level sets tend to split in presence of heterogeneous staining and to merge together clustered nuclei, thus leading to segmentation errors that would require extensive postprocessing to be corrected. This would increase further the computational complexity of the segmentation procedure. Fig. 5 shows an example of image with nuclear activations, where nuclei are generally better imaged and then easier to segment compared to images with membrane and cytoplasm activations. In the same type of images parametric active contours [31] performed much better, as it is shown in the example of Fig. 6(a). Also experiments on images with membrane as well as cytoplasm activations showed a better performance of parametric over geometric active contours (see Fig. 6(b)). Therefore the tests we performed on real

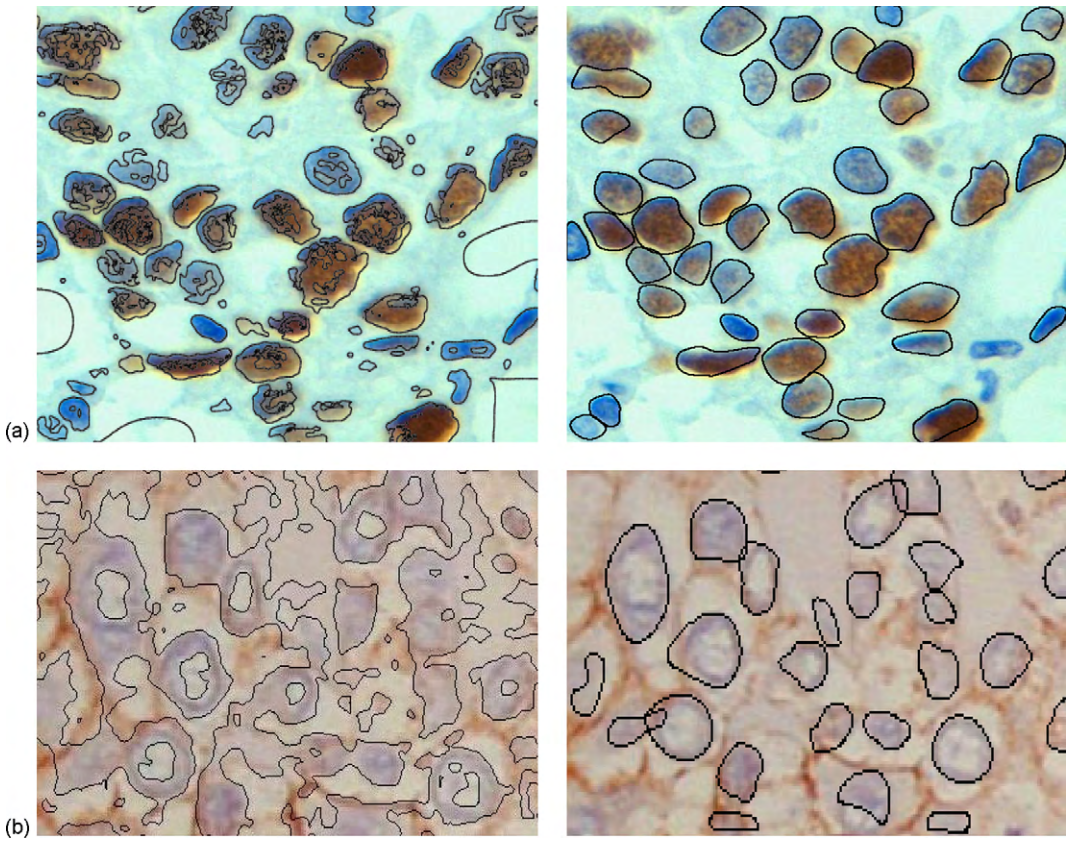

Fig. 6. Example of geometric and spline-based parametric active contours segmentation in images with nuclear (a) and membrane (b) activations.

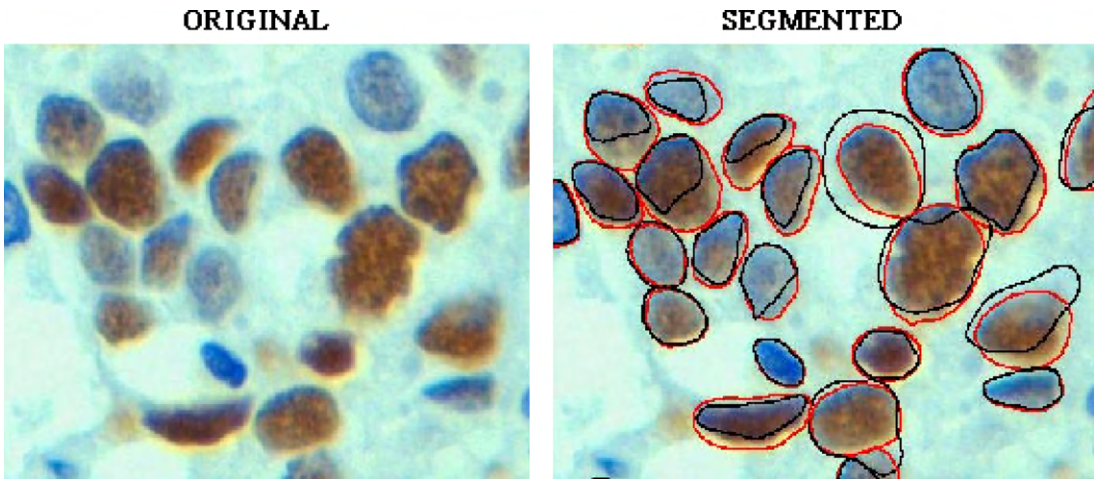

Fig. 7. Example of active contours' segmentation (manual initial boundaries in red, final converged active contours in black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

tissue images led us to prefer parametric formulation to geometric level set formulation.

Active contours may be classified according to the formulation of the image term guiding the evolving curve towards the target boundary as either *edge-based* snakes, based on local gradient information, or *region-based* snakes, based on global image information (e.g. statistical features). The former are generally more precise than the latter. However, region-based formulations are less sensitive to curve initialization and to noise and have less difficulties moving into concavities. Several attempts to integrate edge-based and region-based information [38,39] have also led to the development of *mixed* snakes.

In this work we made experiments with edge-based, region-based as well as with mixed energy active contours using the parametric spline-based formulations presented in [31], among the most theoretically valuable in literature. Its main prerogative is the formulation of image energy as linear combination of either edge-based (taking into account gradient magnitude as well as its direction) and region-based terms, thus inheriting the advantages of both [32]. The relative contribution of the two terms can be modulated through simple variation of the coefficients, so that a general unifying framework is provided which includes all the most widely used formulations of active contours. Further details are provided in [31,32,40].

We developed a semi-automated procedure that consists in feeding active contours with initial boundaries manually traced by a pathologist and with a monochromatic image that provides the information needed for the calculation of the image energy term. The active contours automatically converge to the final nuclear boundaries. The procedure allows the user to choose between three different active contours' formulations, respectively with *edge-based*, *region-based* and *mixed* image energy (i.e. equally weighted linear combination of edge and region-based terms). In case of sample with membrane/cytoplasm activity, i.e. with only blue colored nuclei (see Fig. 2), image energy is calculated considering the contribution of the only blue stain. On the contrary, in case of sample with nuclear activity, i.e. with either blue or brown colored nuclei, image energy is obtained by the contributions of both the dyes, therefore the two monochromatic stains are merged together through bitwise AND operation. The techniques for pre-classification of image and separation of stains are the same used for the morphology-based approach and described in Section 2.

As it is well known, one of the major drawbacks of active contours is that they are extremely sensitive to curve initialization, and they may show a lack of convergence far away from the target boundary [25]. Another major drawback is the lack of convergence caused by other attractors located close to the target. An example of such limitations is shown in Fig. 7, where the red curves are the initial boundaries traced by the operator and the black curves are the final boundaries obtained through active contours approach. In order to obtain the best performance achievable by active contours, in our experiments the operator was asked to trace the initial curves very close to the nuclear membranes. Results of the experiments are reported in Section 5.

The parameters of the active contours (e.g. knot spacing, etc.) were tuned by running experiments on real IHC images and are provided in our URL [41].

4. Implementation

The procedures were implemented in Java as plugins for ImageJ [42], a public domain software for image analysis and processing. We inherited the whole class hierarchy of ImageJ 1.38 API and the open-source plugins and macros for color deconvolution, local thresholding and spline-based snakes [18,43,40] and we implemented our own functions and classes. The parameters of both the methods were set by running experiments on real IHC images. See our URL [41] and [22] for details.

5. Experimental results and discussion

Our morphology-based method and the active contours approach were compared on the same real IHC images, on about 800 nuclei. These images showed lung cancer tissue stained by H-DAB and were acquired through high resolution confocal microscopy with three different enlargements (200×, 400× and 800×). Ten different histological samples were used to take the pictures: five of these samples showed activations of the target protein in the cellular membrane of the cancerous cells, two in the cytoplasm and the remaining three in the nuclei (see Fig. 1 for some examples).

The morphology-based method did not require any manual intervention by the user, since it is fully automated, whereas active contours had to be manually initialized by a skilled operator by tracing boundaries very close to the target nuclear membranes, thus leading the active contours to the best performance achievable.

In order to perform an exhaustive comparison, experiments were run with three active contours, with different formulations of the image energy functional, i.e. respectively edge-based, region-

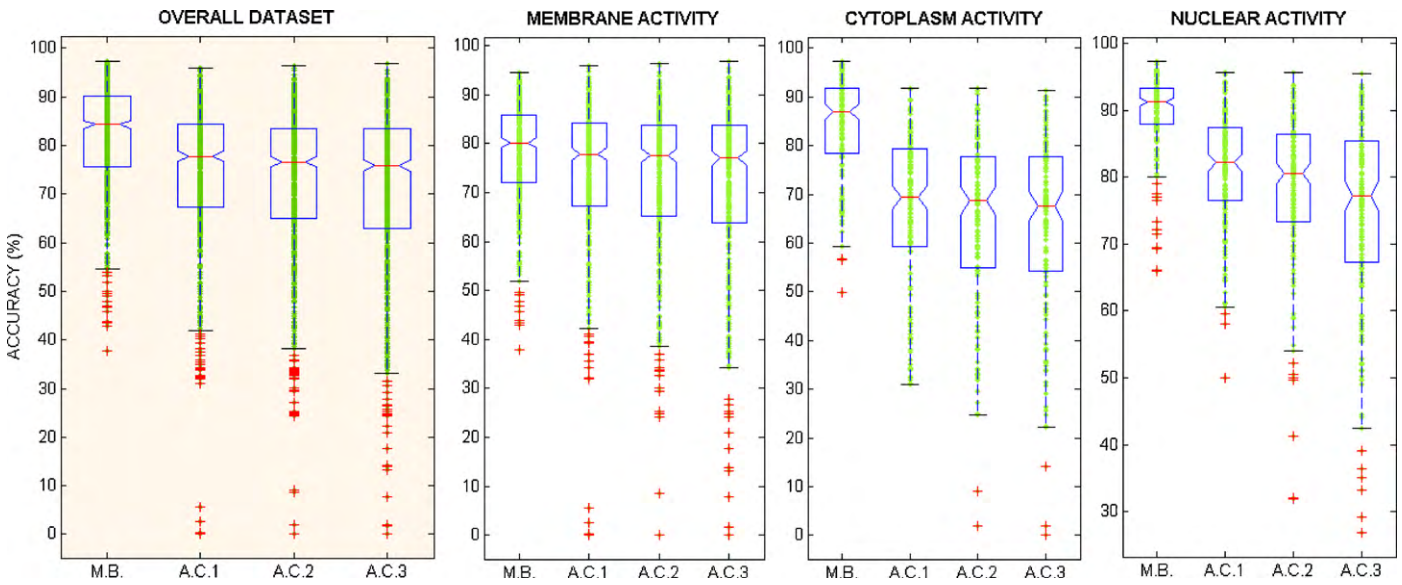

Fig. 8. Box and whisker plots of the accuracy values achieved by the proposed morphology-based method (M.C.) and respectively by gradient-based (A.C.1), mixed (A.C.2) and region-based (A.C.3) active contours in the overall validation dataset as well as in only images with membrane, cytoplasm or nuclear activity. The overall accuracy data are represented with green dots and superimposed to the plots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

based and mixed (with equally weighted linear combination of edge and region-based terms [41]).

We evaluated the accuracy of the segmentations performed by both the automated and the semi-automated methods through a very strict pixel-wise comparison with reference nuclei provided by manual operators; this procedure was repeated for each of the nuclei in our validation dataset (about 800). Then the tested methods were compared with each other on the basis of the distribution of the accuracy values achieved in the same validation dataset. In addition we performed a nucleus-by-nucleus paired comparison of the tested methods and we calculated the percentage of nuclei in which the best technique overcame the other methods; this provides a measure of the effective recurrence of the superiority of the best technique over the other approaches.

Segmentation accuracy was estimated as

$$\text{Accuracy}\,(\%) = 100 \cdot \left[1 - \frac{\sum(R\,\text{XOR}\,A)}{\sum(R\,\text{OR}\,A)}\right], \qquad (1)$$

where R is a binary image where pixels belonging to the reference manual nucleus are set to 1 and A is a binary image where pixels enclosed by the nuclear boundary provided by the automated (or semi-automated) procedure are set to 1. As shown by Fig. 9, $R\,\text{XOR}\,A$ returns the pixels that were misclassified by the automated method (i.e. non-nuclear pixels classified as nuclear and vice-versa) whereas $R\,\text{OR}\,A$ returns the totality of the evaluated pixels: therefore accuracy is the complementary of the percentage of misclassification.

Manual segmentation may lack reproducibility [12] due to the critical characteristics of IHC tissue images (e.g. small dimension of nuclei, superposition of cells and tissues, heterogeneity of staining, intercontamination of dyes, etc.). To improve the objectivity of our validation for each nucleus ten skilled operators were asked to outline manually the nuclear membrane. Thereafter, the reference nucleus used for the validation was defined by only those pixels enclosed by the most part of the ten manual boundaries.

The accuracy values achieved by each of the tested methods in the validation dataset are reported in the box and whisker plots of Fig. 8. We chose this widely used representation technique in descriptive statistics [44] since it is particularly useful to display differences between populations without making any assumptions of the underlying statistical distribution. The box have horizontal lines at the three data distribution's quartiles, i.e. the three values which divide the sorted data set of accuracy values into four equal parts, so that each part represents one fourth of the sampled population. The lowest horizontal line of the box detects the lower quartile of the accuracy distribution, that cuts off the lowest 25% of the accuracy values; the second line, shown in red, detects the second quartile or median accuracy value, that cuts accuracy distribution in half; the upper horizontal line detects the upper quartile

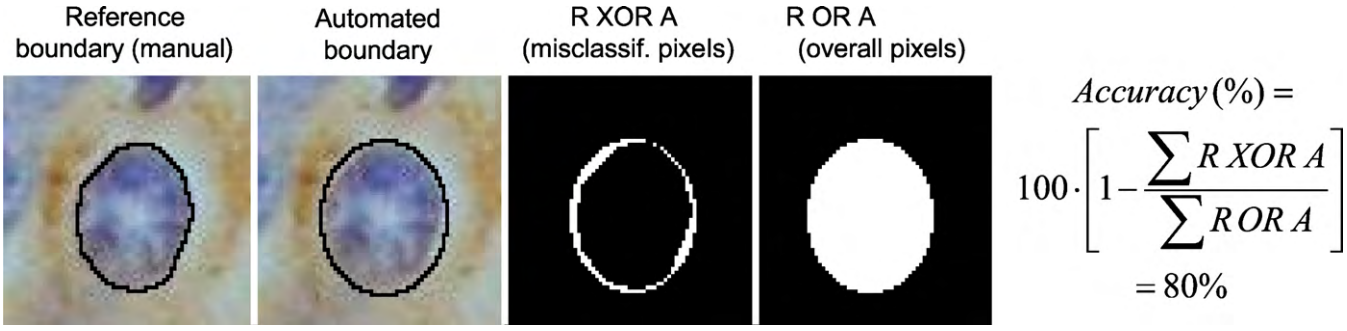

Fig. 9. Calculation of segmentation's accuracy. Example.

value of the accuracy distribution, that cuts off the lowest 75% (or the highest 25%) of the accuracy values. The so-called whiskers are vertical dotted lines extending from each end of the box to show the extent of the rest of the data. The overall accuracy data related to the about 800 tested nuclei are superimposed to the box and whisker plots and represented with green dots. As it is standard practice, data with values beyond the ends of the whiskers are considered outliers and represented with red crosses. Our results are available also in tabular form at URL [45].

As it is shown by Fig. 8, the median value of the accuracy achieved by our proposed method (detected by the red horizontal line of the box) is higher than the median values of the approaches based on active contours in all the considered datasets. In particular, the median accuracy value of the morphology-based method is around 85% and it overcomes median accuracy of edge-based, mixed and region-based active contours of about 10%. Also the comparisons of the lower and upper quartile values (detected respectively by lowest and the highest horizontal line of the box) confirm the superiority of the morphology-based approach over the active contours. Moreover, the extension of the distribution of accuracy values achieved by the morphology-based method (represented by the whiskers's extension) is always lower compared to active contours, thus showing less variability of performance.

Mean accuracy values obtained respectively in samples with membrane, cytoplasm and nuclear activity are shown in the graph of Fig. 10: the morphology-based technique achieved the best mean accuracy at all times. Despite they were manually initialized very close to the target boundaries, all the three formulations of active contours tested in our experiments were overcome by our fully automated procedure. In particular, the morphology-based method achieved on average an 84% accuracy, overcoming mean accuracy of edge-based, mixed and region-based active contours of respectively 10%, 12% and 14%; these values were obtained by averaging the accuracy achieved in all the tested nuclei.

Nucleus-by-nucleus paired comparison of the tested methods confirmed the superiority of the morphology-based technique over the active contours: in fact we obtained that its accuracy was higher than edge-based, mixed and region-based active contours' accuracy respectively in 65%, 66% and 67% of the evaluated nuclei.

It must be noted that edge-based snakes performed slightly better than region-based snakes mainly because the initial curve

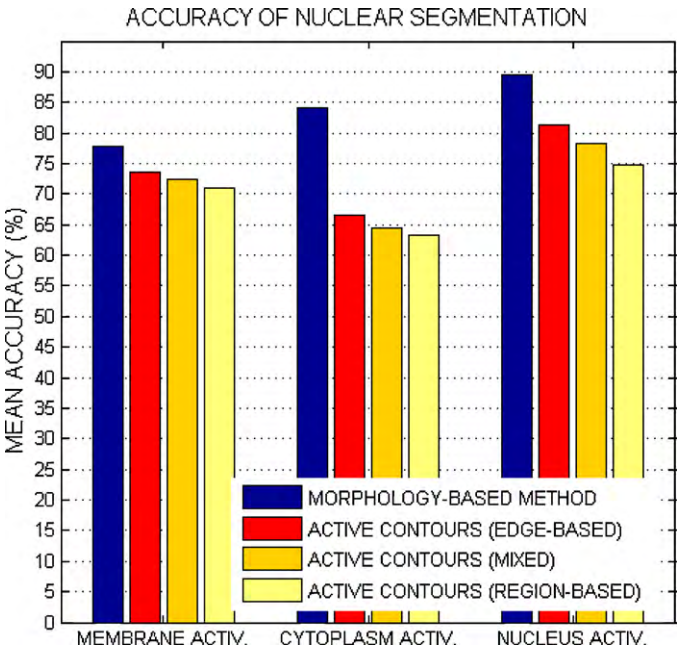

Fig. 10. Mean percentage accuracy achieved by the tested techniques in validation samples with membrane, cytoplasm or nuclear activity.

drawn by the skilled operator is very well fitting to the nuclear membrane. However, active contours still fail in detecting nuclear regions in presence of intensity gradients inside the region. Moreover, it is well known that edge-based snakes are the most sensitive to curve initialization [31], so they might have lacked convergence if the initial curve had not been so close to the nuclear membrane as it was in our experiments. Region-based active contours are less sensitive to gradients, however this technique also fails when nuclei are characterized by non-homogeneous staining. Moreover, because it is based on statistical characterization of background, its performance is worse when images contain heterogeneous regions with different distribution and density of nuclei. The main reasons for the superiority of the morphology-based approach we propose lies in the *binarization* step and separation of clustered

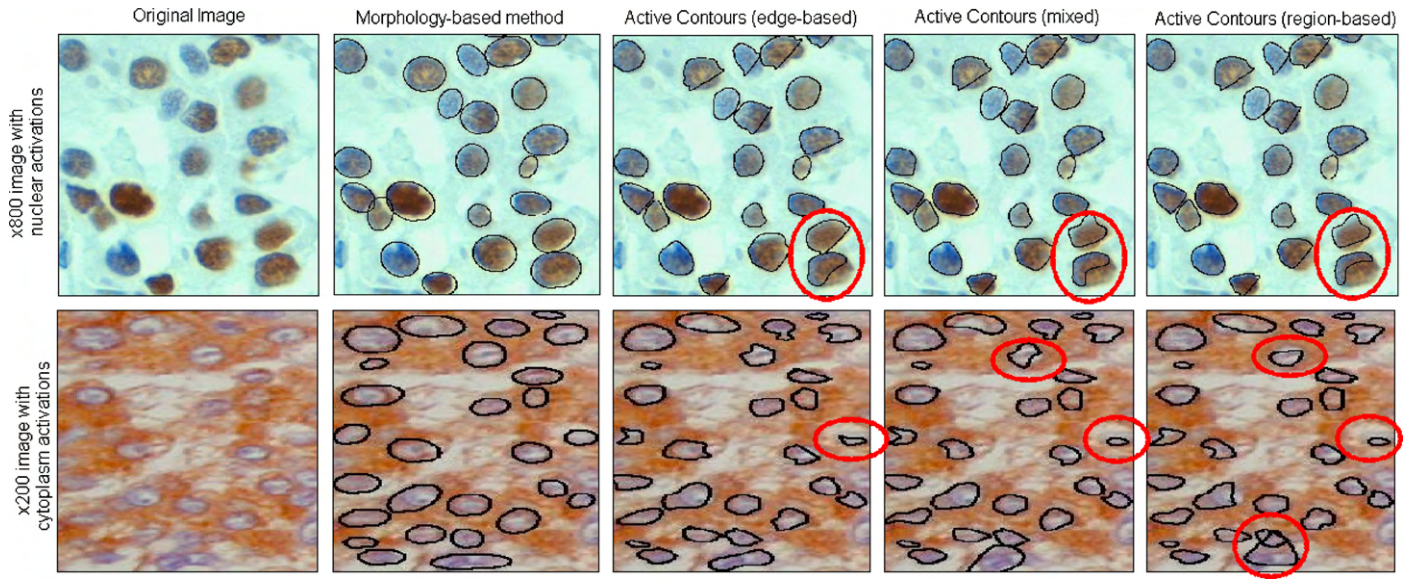

Fig. 11. Examples of nuclear segmentations performed by the tested methods. Most evident errors made by active contours are highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

nuclei. The *binarization* step handles the heterogeneity of staining and the consistent intensity variations within the target through the adaptive local thresholding algorithm. The local nature of this approach enables a better distinction between the colored nuclear regions and the background. As a result, regions are successfully identified even if they are characterized by intensity variations and heterogeneous staining. On the other side, the enhanced watershed algorithm exploits color information to merge oversegmented nuclei compared to active contours that take gray-scale images as input. RGB versions of active contours [6,10] suffer from similar limitations in that they cannot handle heterogeneous stained local regions. Ultimately we found out that active contours are in general not expressive enough to handle complex and heterogeneous chromatic information in IHC, where the targets cannot be distinguished by simple color/texture features. Therefore a correct segmentation requires high-level interpretation of the biological information carried by color, which is actually not provided by any available active contours formulation.

Fig. 11 shows examples of nuclear segmentations performed by the tested techniques (most evident errors made by active contours are highlighted in red).

6. Conclusions

In this paper we addressed the problem of nuclear segmentation in cancer tissue images comparing the effectiveness of morphology-based technique and active contours approach. Extensive experimental results show that our fully automated morphology-based technique is more accurate in IHC tissue images than various formulation of state-of-the-art semi-automated active contours, since it is less sensitive to intensity and color variations within the target region as well as overlapped nuclei that deviate the active contours far from the target. Our future work will include the extension of our morphology-based technique to other types of biomedical images as well as to 3D applications.

Acknowledgments

The authors would like to acknowledge Dr. Marco Volante and Dr. Ida Rapa of Hospital S. Luigi, Orbassano, Torino.

References

[1] Hengerer A. From genomics to clinical molecular imaging. Proc IEEE 2005;93(4).
[2] Chen W. Unsupervised tissue microarray analysis for cancer research and diagnostics. IEEE T Inf Technol B 2004;8(2):89–96.
[3] Taneja TK. Markers of small cell lung cancer. World J Surg Oncol 2004;2(10).
[4] Jrgensena JT. Pharmacodiagnostics and targeted therapies a rational approach for individualizing medical anticancer therapy in breast cancer. Oncologist 2007;12(4):397–405.
[5] Ficarra E. Computer-aided evaluation of protein expression in pathological tissue images. In: Proceedings of CBMS'06. 2006. p. 413–8.
[6] Yang L. Unsupervised segmentation based on robust estimation and color active contour models. IEEE T Inf Tech B 2005;9(3):475–86.
[7] Mukherjee DP. Level set analysis for leukocyte detection and tracking. IEEE T Image Process 2004;13(4):562–72.
[8] Elmoataz A. Using active contours and mathematical morphology tools for quantification of immunohistochemical images. Signal Process 1998; 71(2):215–26.
[9] Garrido A. Applying deformable templates for cell image segmentation. Pattern Recogn 2000;33(5):821–32.
[10] Ling Pi. Color image segmentation for objects of interest with modified geodesic active contour method. J Math Imag Vis 2006;27(1):51–7.
[11] Di Cataldo S. Segmentation of nuclei in cancer tissue images: contrasting active contours with morphology-based approach. In: Proceedings of BIBE2008. 2008. p. 1–6.
[12] Lejeune M. Quantification of diverse subcellular immunohistochemical markers with clinicobiological relevancies: validation of a new computer-assisted image analysis procedure. J Anat 2008;212(6):868–78.
[13] Chang H. Graphical methods for quantifying macromolecules through bright field imaging. Bioinformatics 2009;25(8):1070–5.
[14] McCullough DP. Segmentation of whole cells and cell nuclei from 3-D optical microscope images using dynamic programming. IEEE Trans Med Imag 2008;27(5):723–34.
[15] Diaz G. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. J Biomed Inform 2009;42(2):296–307.
[16] Ruifrok AC. Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 2001;23(4):291–9.
[17] Ruifrok AC. Comparison of quantification of histochemical staining by Hue-Saturation-Intensity (HSI) transformation and color deconvolution. Appl Immunohistochem Mol Morphol 2004;11(1):85–91.
[18] Landini G. Software; October 2007. http://www.dentistry.bham.ac.uk/landinig/software/software.html.
[19] Ridler TW. Picture thresholding using an iterative selection method. IEEE T Syst M Cy, SMC-8 1978:630–2.
[20] Nattkemper TW. Automatic segmentation of digital micrographs: a survey. In: Proceedings of MEDINFO'04, vol. 11(Pt 2). 2004. p. 847–51.
[21] Milstei N. Image segmentation through adaptive thresholding. Technion – Israel Inst Technol 1998:1–38.
[22] Di Cataldo S. Selection of tumor areas and segmentation of nuclear membranes in tissue confocal images: a fully automated approach. In: Proceedings of BIBM'07. 2007. p. 390–8.
[23] Roerdink J. The watershed transform: definitions, algorithms and parallelization strategies. Fund Inform 2001;41:187–228.
[24] Kass M. Snakes: active contours models. Int J Comp Vis 1987;1(4):321–31.
[25] McInerney T. Deformable models in medical image analysis: a survey. Med Imag Anal 1993;15(6):580–91.
[26] Leventon ME. Statistical shape influence in geodesic active contours. Proc CVPR'00 2000;1:316–23.
[27] Chen Y. Using prior shapes in geometric active contours in a variational framework. IJCV 2002;50(3):315–28.
[28] Rousson M. Shape priors for level set representations. In: Proceedings of ECCV. 2002. p. 78–92.
[29] Huang F. Face contour detection using geometric active contours. In: Proceedings of WCICA'04. 2002. p. 2090–3.
[30] Bresson X. A variational model for object segmentation using boundary information and shape driven by the Mumford–Shah functional. IJCV 2006;68(2):145–62.
[31] Jacob M. A unifying approach and interface for spline-based snakes. Proc SPIE MI 2001;4322:340–7.
[32] Jacob M. Efficient energies and algorithms for parametric snakes. IEEE T Imag Process 2004;13(9):1231–44.
[33] Caselles V. Geodesic active contours. In: Proceedings of 5th International Conference on Computer Vision. 1995. p. 694–9.
[34] Chan T. Active contours without edges. IEEE T Imag Process 2001;10:266–76.
[35] Osher S. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. J Comput Phys 1988;79:12–49.
[36] Li C. Minimization of region-scalable fitting energy for image segmentation. IEEE T Imag Process 2008;17(10):1940–9.
[37] Xu C. On the relationship between parametric and geometric active contours. In: Proceedings of Asilomar'00. 2000. p. 483–9.
[38] Paragios N. Unyfing boundary and region-based information for geodesic active tracking. In: Proceedings of IEEE CVPR. 1999. p. 300–5.
[39] Zhu SC. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE T PAMI 1996;18:884–900.
[40] Jacob M. http://bigwww.epfl.ch/jacob/software/SplineSnake.
[41] http://didattica-online.polito.it/eda/download/bio_eda_guest/CMIG-data/Parameters.pdf.
[42] Rasband WS. ImageJ. Bethesda, MD, USA: U.S. National Institutes of Health. http://rsb.info.nih.gov/ij/.
[43] Couture R. http://www.aecom.yu.edu/aif/instructions/imagej/macros/local-threshold-unsharp-masking.txt.
[44] Tukey JW. Exploratory data analysis. MA: Addison-Wesley; 1977.
[45] http://didattica-online.polito.it/eda/download/bio_eda_guest/CMIG-data/results.xls.

Santa Di Cataldo is PhD student at Politecnico di Torino. Her research concerns bioimaging and automated algorithms for systems biology.

Elisa Ficarra is Assistant Professor at Politecnico di Torino. Her research concerns bioimaging, algorithms for gene expression analysis, gene classification, clustering and networks, clinical genomics and systems biology.

Andrea Acquaviva is Assistant Professor at Politecnico di Torino. His main interests are distributed embedded software design and wireless sensor networks. His research also includes bioimaging and systems biology.

Enrico Macii is Full Professor at Politecnico di Torino. His main interests are computer-aided design of digital integrated circuits and systems, with particular emphasis on logic synthesis, optimization, testing, and formal verification. His research also includes bioinformatics, bioimaging and systems biology.