

# Review of Nuclei Detection, Segmentation in Microscopic Images

Rujuta O\* and Vyavahare AJ

Department of Electronics and Communication, Modern College of Engineering, Pune, Maharashtra, India

\*Corresponding author: Rujuta O, Department of Electronics & Communication, Student M.E. E &TC, Modern College of Engineering, Pune, Maharashtra, India, Tel: 020 2569 6064; E-mail: [rujuozarkar88@gmail.com](mailto:rujuozarkar88@gmail.com)

Received date: May 2, 2017; Accepted date: May 15, 2017; Published date: May 20, 2017

Copyright: © 2017 Rujuta O, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

## Abstract

Histological tissue section provides rich information about cellular structure. There are considerable amount of variations in cell structure, and cell organization that impede analysis of large histological dataset. This paper is a review of some recent state-of-art nucleus/cell segmentation approaches on different types of microscopic images. In the conventional diagnosis, pathologist analyze biopsies to make diagnostic assessment. Recently some automatic methods have been evolving in digital histopathology with growing application related to nuclear detection and segmentation. We and discussed and studied here various trends on nucleus detection, and segmentation.

**Keywords:** Microscopic analysis; Digital pathology; Nuclear detection; Nuclear segmentation.

## Introduction

In microscopic study of cell morphology, a tissue sample taken out from human body and prepared for viewing under microscope and then staining techniques can apply. However, this process different components of tissue with different stains can be dye. It is very difficult task to diagnose a disease after manually analyzation of biopsy slides. However it is a labor intensive and complicated work for pathologists. But due to recent advancement in digital pathology, the automatic recognition of pathology patterns in microscopic image has the potential to get valuable assistance to the pathologist. While doing various studies, automatic nuclei detection and segmentation is challenging task. The detection and segmentation of nuclei in cytological images can be done by well-separated nuclei and the absence of complicated tissue structures. But the segmentation of nuclei on histopathological images (tissue preserving its original structure) is more difficult because most of the nuclei are often part of histological structures having complex and irregular visual aspects. Currently there are many reviews on automated digital pathology and microscopic image analysis.

The main complications in nuclear segmentation are cellular structure (technical, biological) variation or heterogeneity. The techniques for automatic detection and segmentation are based on adaptive thresholding by morphological operator fuzzy clustering, level set based on gradient information and energy minimization function [1-5]. Color deconvolution after optimum thresholding, color and texture based hybrid analysis followed by clustering [6,7]. There are some application combined by above mentioned techniques i.e., iterative radial voting was used to estimate seeds for the location of nuclei and the model interaction between neighboring nuclei with multiphase level set, and in an initial segmentation of nucleus with graph cut is obtained by multi scale detection and result further refined with next iteration of same method [1,8-10]. Nuclear segmentation through color decomposition, using the same techniques developed for fluorescence microscopy [11]. Still it is a challenging to effectively address the systematical requirements of tumor histological

characterization. Thresholding and clustering re applicable for only constant chromatin content for nuclei in the image. But due to wide variation in chromatin contents problems occurs with overlapping and clumping of the nuclei, also due to tissue thickness, they cannot be segmented properly. The method proposed in focus iterative radial voting on delineation of overlapping nuclei, but seed detection can have failed in the presence of wide variation in the nuclear size, lead to fragmentation [6,12].

In this paper we review the different methods for nucleus detection, segmentation, feature extraction on H&E stained histopathological images. However nuclei detection is very important for cancer diagnosis and grading, as sometimes it is related to patient survival [13].

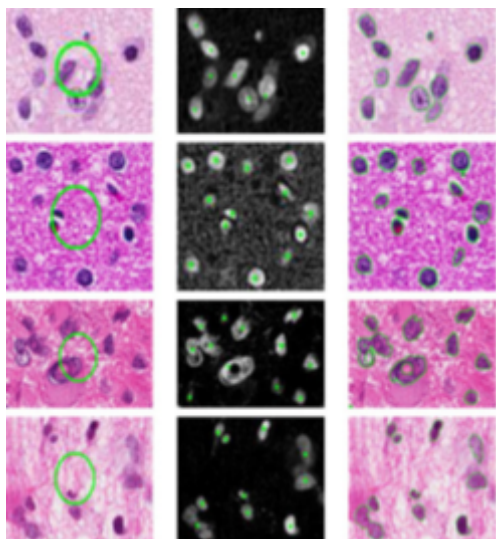
## Multireference level set (MRL)

Level set method is used to represent contour of object with gradient information by obtaining energy minimization function. Multi-reference Level set deals with the problems of technical and biological variation using local and global information from decomposed nuclear channel and manually annotated reference images. MRL approach developed from deep understanding and experimental outcomes obtained from simple color decomposition and thresholding techniques. Sometimes it miss or overestimates some of the nuclei in the image, i.e., nuclei with low chromatin contents are excluded. This becomes further complicated for detection of nuclear size and shape (e.g., the classic scale problem). It is clear that the incorporation of prior knowledge (e.g., manual annotation and validation by the pathologist) would be needed not only for validation, but also for constructing a model that captures wide variations in the nuclear staining, both within and across tissue sections. Correspondingly, our proposed approach combines prior knowledge, characterized by the Gaussian mixture models (GMM), and the nuclear staining information of the original image patch, extracted by color decomposition, within a level set framework. The obtained result is a binarized image of blobs (a single nucleus or a clump of nuclei), which are either validated or partitioned further through geometric reasoning.

**Algorithm 1:** Multireference level set for characterization of nuclear morphology [14].

- Nuclei detection using laplacian of Gaussian filter.
- Nuclear staining information represented by color deconvolution.
- Color normalization for reducing variation in image statistics.

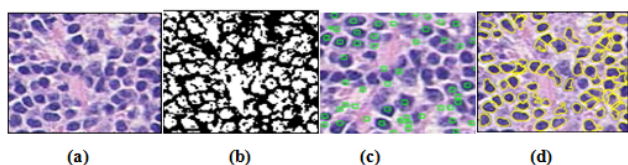
Integration of the nuclear detection by GMM and staining information for the final segmentation (Figure 1).



**Figure 1:** Segmentation of nuclei using MRL (a) original image patch with Level set initialization for MRL, (b) decomposed nuclear channel with Level set evolution, (c) final segmentation by MRL.

In this method, multiscale LoG response gives advance information into the system. Both obtained by LoG and nuclear staining information by color deconvolution combined to get better characterization of data, hence it more effective and robust.

**EMaGAC:** The Expected Maximization based Geodesic Active Contour [5] is able to overcome drawbacks related with probabilistic segmentation method which is need for defining representative dataset (Figure 2).



**Figure 2:** (a) Original BC histopathology image, (b) binarized lymphocyte nuclei, (c) contour initialization, (d) segmentation after splitting contour.

**Algorithm 2:** Steps involved in EMaGAC model:

- Obtain initial segmentation to get binarized nuclei using EM.
- Initialize and evolve contour from EM results.
- Distinguish objects using texture based clustering.

- Final segmentation after splitting contour enclosing multiple object.

EM algorithm avoids need for representative training data and reduces effect of variability of dataset on segmentation. Here Magnetostatic Active Contour is used as Geodesic Active Contour model. For initialization of segmentation centroid of the object detected (as seed points) to start the contour, it improves performance of active contour model for capturing active contour model. Then overlapping of object is resolved by splitting contour between high concavity points. The contour splitting is designed where high concavity points are connected by path through which relevant edge point within a contour. The edge path algorithm provide path by including single edge points at a time to make sure that splitting of contour represent an edge or overlapping boundary. And the last step is object classification state of setting based on texture based clustering to separate lymphocyte nuclei from other similar structure in the cell image as shown in Figure 2.

The EM algorithm can be a better option to improve results of automation by replacing requirement of representative training data which is difficult to define variability across images. Again the concavity detection algorithm increases segmentation accuracy followed by splitting contour by edge path algorithm. As compare to MRL, EMaGAC is sensitive to initialization hence provided by large variation in the given dataset.

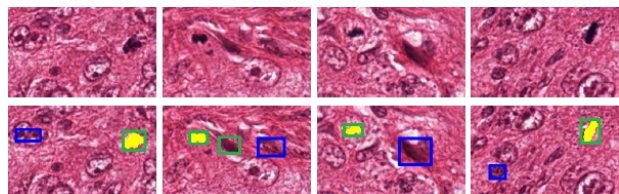
### Random forest

Random forest is the combination of tree predictors in which trees are dependent on the value of random vector sampled individually and the distribution of all trees in the forest. Generalization of forest error meet at a point to limit as no. of trees in the forest becomes large. It is depend on the strength of individual trees in the forest and mutual relation between them.

Random forest is one of the fast and accurate method for counting mitotic figures in from histo-pathological images. Because Detection and counting of mitotic cells from histo-pathological images are main steps in breast cancer diagnosis as mitosis signals cell division. Using regenerative random forest, the automatic feature selection can be perform. The regenerative random forest to detect mitotic in histo-pathological images, given to significance to each feature (dimension) of the feature vector based on importance of features. It also provides segmentation difficulty to each tree in the forest. The trees are then regenerated to make new population of trees (new forest) and only the more important features survive in new forest. Feature vector constructed from domain knowledge nucleus intensity feature, nuclear membrane feature and feature of possible stroma region surrounded to the cell. Classification performance improves by the use of domain knowledge. In H&E stained histo-pathological images, nuclear intensity pattern provides features between mitotic and non-mitotic cells. It has been observed that stroma or connecting tissues contain non-mitotic cells. Hence by evaluating texture feature of the region surrounding cells to find cell belong to stroma or not. After cell membrane breakdown during mitosis, the nuclear contour of mitotic cell become irregular compare to non-mitotic cell contour (Figure 3).

**Algorithm 3:** Steps involved in random forest:

- Nuclei segmentation.
- Feature extraction.
- Nuclei classification.



**Figure 3:** Sample results of mitotic cell detection. Top row shows original sub images, second row shows detected results, where ground truth shown by yellow colored region, true positive detection shown by green color boxes and false positive by blue color boxes in the second row.

### Morphometric analysis of GBM (Glioblastoma Multiforme)

Various nuclear segmentation methods computed as a typical tissue section (size 40k by 40k) could be take very much time, hence a pipeline structure for delineation of nuclei has been designed and its morphometric features with high quality output taken out.

**Algorithm 4:** Steps for delineating nucleus from H&E stained tissue section:

- Color mapping.
- Color deconvolution.
- Thresholding.
- Refinement.
- Validation.
- Nuclear cytoplasmic multidimensional profiling.

Algorithm gives steps involved in converting image into multidimensional representation. Color mapping step removes heterogeneity by normalizing H&E stain. color decomposition for reduces computational complexities. Next step thresholding based on analysis of histogram. Problem arises in thresholding is fast histogram based thresholding for refinement and validation. Refinement perform 1) Intensity spread over of each thresholded blob as Gaussian mixture to examine variation in background and adjacent nuclei are combined together and 2) It uses convexity extent to partition blob based on relative boundaries.

The multidimensional representation generated for each nucleus defines its signature and organization. Critical factors for pipeline computation are relative output, segmentation calibre, morphometric representation for subtyping, and genomic interdependence while computing pipeline. Segmentation, feature extraction, subtyping modules is tested; 1) For segmentation, no. of hand segmented images originates from tissue section of TCGA cohort; 2) Feature extraction tested against synthetic data with ground truth; 3) Subtyping evaluated by group similarity matrix. Morphometric analysis and cellular structure organization of tissue histology for large dataset, subtype can be identified are predictive outcome of medicinal protocol.

### Graph cut

Al-Kofahi [5] present a method based on LOG as generic blob detection incorporated with automatic and adaptive scale selection. The first step is to separate foreground pixel in nuclear channel from background pixel. For effective binarization graph cut algorithm required initialization, for that purpose seed detection processed by new generation of fast clustering algorithm to generate initial segmentation, which is then refined by graph cut segmentation. For *in vivo* tissue example, deparaffinized 5  $\mu$ m sections of formalin fixed and paraffin embedded breast tissue were rehydrated and stained with hematoxylin. For *in vitro* tissue sample 6  $\mu$ m section of OCT frozen block of K 1735 tumor cells were stained.

First step is to segment foreground pixel into nuclear channel from background pixels. For this purpose, a hybrid approach is made which incorporate initial binarization refinement using graph cut algorithm. The graph cuts binarization extracts connected clusters of nuclei that must be separated into individual nuclei. This identification of seed points, which can be effectively done by multiscale LOG filters. The aim of refinement of initial contour is to enhance delineation of touching nuclei. Hence it should preserve the shape of cell nucleus based on nuclear shape model and prevent two or more nuclei from being merged. Refinement step should allow multiple cell nuclei need to refine concurrently for better efficiency.

The result is taken on 25 representative images (15 *in vitro* images and 10 *in vivo* images) were inspected and type of segmentation error." should be altered to "The representation is taken on 25 sample images (15 *in vitro* images and 10 *in vivo* images) which were inspected and categorized based on the type of segmentation error.

Ref. and Dataset	Detection	Segmentation	Metrics/ Results
[14] 440 hematoxylin and eosin (H&E) stained GBM tumor section from 152 patients	Gaussian Mixture Model	Multireference Level Set	Precision=0.76 Recall=0.72 F-measure=0.74
[5] 62 HER2+ breast biopsy images	Gaussian Mixture Model	Geodesic Active Contour	Precision=0.60 Recall=0.85 F-measure=0.703
[15] Training data contains 35 images with 233 mitotic cells and test data contains 15 images with 103 mitotic cells from MITOS dataset.	Feature based domain knowledge	Random Forest	Precision=0.59 Recall=0.76 F-measure=0.664
[16] 344 large scale GBM tissue sections from TCGA for 133 patients.	Thresholding	Nuclear Morphology and cytoplasmic multidimensional profiling	Precision=0.60 Recall=0.85



			F-measure=0.703
[1] Analysis of 25 characteristic images (15 in vitro and 10 in vivo images) from cancer histopathology	LOG followed by Initial binarization by graph cut	Graph Cut	Precision=0.98 Recall=0.95 F-measure=0.97

**Table 1:** Nuclear detection segmentation methods.

## Nuclear Mask Partitioning

Final part is nuclei partitioning. However after level set evolution, we got binarized image of clumps of nuclei, next step is partition them into single nucleus. Generally nuclear shape is convex in shape therefore overlapping nuclei detected by concavities has to be separate out through geometric reasoning, which is explained by following steps:

**1) Detection of maximum point curvature:** As the contour of nuclear mask extracted, derivative are computed by derivative of Gaussian.

**2) Delauney triangulation:** DT is applied to all points of maximum curvature for making possible grouping. The conclusion of grouping then refined by removing edges based on triangulation rules.

**3) Geometric reasoning:** Properties of both obtained graph and shape of object combined for edge inference.

## Conclusion and Future Work

Since last some decades, considerable number of articles are published in field of microscopic and histopathology image for nuclei detection and segmentation. All of them are not applicable to the same dataset hence it is difficult to compare that which one is better than another. The multi-reference Level Set [14] approach applied to TCGA GBM cohort address problems of cellular organization by utilizing local and global information from manually annotated and reference images. EMaGACOR [5] provides a method for detection of lymphocytes in HER2+ BC histopathology images. This overcomes issues like detection sensitivity with random initialization and allow to solve problem of object overlap. The initialization contour from EM results enhancement of contour performance, eliminates need of training, and improves automation. The concavity detection scheme with edge path algorithm effectively resolves issue of overlapping objects. Using Random forest, intensity features of nucleus, texture feature of stroma, nuclear membrane feature are classified that perform automatic feature selection [15]. The regenerative process converges producing maximum classification accuracy during training also archive lower time complexity. [16] Morphometric analysis and cellular organization of tissue histology of large dataset identified as predictive outcome as a result of medicinal protocol. It is observed that histological subtyping shows intrinsic categories independent of supervised histological grades. The automatic segmentation algorithm by Al-kofahi [17] present a new method results in which seed detection processed by a new generation of fast clustering algorithm to generate initial segmentation that refined using graph cut segmentation algorithm. Hence a novel segmentation algorithm based on graph coloring and  $\alpha$ -expansions has been developed which is proven to be more robust and accurate over cancer histo-pathological images.

This paper discussed five methods or approaches for nuclei detection and segmentation. Table 1 shows the summery of nuclear

detection segmentation methods with the results in the form of precision, recall, and F-measure. According to our observation, Graph cut method gives more precise output as compare to rest of all. Our future work will focus on designing robust nuclei detection and segmentation algorithm which can be applicable for maximum type's tumor.

## References

- AlKofahi Y, Lassoued W, Lee W, Roysam B (2010) Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng 57: 841-852.
- Phukpattaranont P, Boonyaphiphat P (2007) Color based segmentation of nuclear stained breast cancer cell images. ECTITrans Electr Eng Commun 5: 158-164.
- Ballaro B, Florena A, Franco V, Tegolo D, Tripodo C (2008) An automated image analysis methodology for classifying megakaryocytes in chronic myeloproliferative disorders. Med Image Anal 12: 703-712.
- Latson L, Sebek N, Powell K (2003) Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy. Anal Quant Cytol Histol 26: 321-331.
- Fatakawala H, Xu J, Basavanhally A, Bhanot G, Ganesan S, et al. (2010) Expectation maximization-driven geodesic active contours with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology. IEEE Trans Biomed Eng 57: 1676-1690.
- Chang H, Defilippis R, Tlsty T, Parvin B (2009) Graphical methods for quantifying macromolecules through bright field imaging. In Bioinf 25: 1070-1075.
- Datar M, Padfield D, Cline H (2008) Color and texture based segmentation of molecular pathology images using HSOMS. In Proc Int Symp Biomed Imag, pp: 292-295.
- Parvin B, Yang Q, Han J, Chang H, Rydberg B, et al. (2007) Iterative voting for inference of structural saliency and characterization of subcellular events. IEEE Trans. Image Process 16: 615-623.
- Nath S, Palaniappan K, Bunyak F (2006) Cell segmentation using coupled level sets and graph-vertex coloring. In Med Image Comput Computed-Assist Intervention-MICCAI, pp: 101-108.
- Chang H, Parvin B (2010) Multiphase level set for automated delineation of membrane-bound macromolecules. In Proc Int Symp Biomed Imag, pp: 165-168.
- Coelho L, Shariff A, Murphy R (2009) Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. In Proc Int Symp Biomed Imag 5193098: 518-521.
- Glotsos D, Spyridonos P, Cavouras D, Ravazoula P, Dadioti P (2004) Automated segmentation of routinely hematoxylin-eosin stained microscopic images by combining support vector machine, clustering and active contour models. Anal Quant Cytol Histol 26: 331-340.
- Alexe G, Dalgin GS, Scanfled D, Tamayo P, Mesirov JP, et al. (2007) High expression of lymphocyte-associated genes in node-negative HER2+ breast cancers correlates with lower recurrence rates. Cancer Res 67: 10669-10 676.
- Chang H, Han J, Spellman PT, Parvin B (2012) Multi-reference level set for the characterization of nuclear morphology in glioblastoma multiforme. IEEE Trans Biomed Engineering.
- Breiman L (2001) Random forests. Mach Learn 45: 5-32.

16. Chang H, Fontenay G, Han J, Cong G, Baehner F, et al. (2011) Morphometric analysis of TCGA glioblastoma multiforme. BMC Bioinf 12: 484.
17. Bunyak F, Hafiane A, Palanippan K (2011) Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level set. Adv Exp Med Biol 696: 413-424.