

# AUTOMATIC SEGMENTATION OF CELL NUCLEI IN 2D USING DYNAMIC PROGRAMMING

*Kaustav Nandy, Prabhakar R Gudla and Stephen J. Lockett*

Image Analysis Laboratory  
SAIC-Frederick/National Cancer Institute(NCI), Frederick, Maryland

**Abstract** - Automatic segmentation of cell nuclei is an essential task in a variety of biomedical applications. However, performance is variable, being typically around 95% correct detection of nuclei depending on the degree of clustering. Furthermore, work to assess and optimize the segmentation accuracy of each nucleus is in its infancy. In this study we developed an algorithm intended for segmentation of clustered nuclei. It integrates gradient magnitude and direction information to achieve robust detection of nuclei and utilizes dynamic programming for highly accurate delineation of detected nuclei. This is the first time that dynamic programming has been used for automatic segmentation of cell nuclei. We achieve precise segmentation of 97% of the detected nuclei.

**Index Terms**— segmentation, dynamic programming, gradient information

## 1. INTRODUCTION

Several cell biology applications require automated, robust and accurate segmentation of individual cells and cell nuclei in 2D cell culture assays. Measurements of the segmented objects are then used to quantitatively answer the underlying biological hypothesis. The types of measurement are application dependent and can range from simple (counting number of cells or cell nuclei) to complex (degree of elongation or spatial patterns of protein expressions etc.). The former includes applications like screening candidate pharmaceuticals to study the efficacy for cell killing or arresting proliferation and detecting rare events (e.g. stem cell detection). The latter includes applications such as calculation of DNA sequence copy number variation from cell to cell for detection of mutated cells [1] and analysis of the spatial positioning of DNA

sequences in individual cells [2].

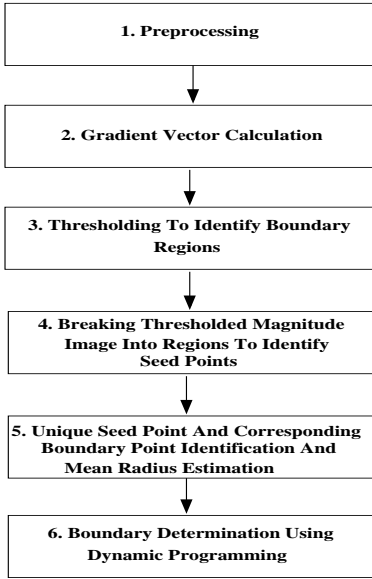
Segmentation of cells and cell nuclei from confocal microscope images is a widely studied topic and over the past decade several methods have been proposed for (semi-) automatic segmentation of cell nuclei in 2 and 3-dimensions. Popular techniques for cell nuclei segmentation are those using thresholding [3], gradient based methods [4], the watershed algorithm [5] [6] [7], level sets [8], dynamic programming [9] [10], active contours [11] [12] and surfaces and various other pattern analysis and machine learning algorithms [13]. The most accurate segmentation algorithms are those that optimally combine image information with *apriori* knowledge about the size and shape of the objects of interest. Overall, irrespective of the application, segmentation should be accurate, robust and automatic since it is the first step before quantitative analysis.

However, current segmentation algorithms achieve varying performance (from 90% to 99%) in terms of accurate detection of cell nuclei depending on the degree of clustering and noise variations. Although this level of performance is satisfactory for some applications, improvements in terms of automation is still warranted in order to increase throughput and reduce human verification. Additionally, assessment of the segmentation accuracy is seldom undertaken and ways to optimize the boundaries has not been rigorously investigated.

The goal of this work is to develop an automatic segmentation of cell nuclei which achieves accurate boundary delineation in the presence of considerable background noise, shading graylevel variations and clustering of cell nuclei. We achieve this goal by combining gradient vector information for robust detection with dynamic programming (DP) [9] for accurate delineation. DP is a widely used and efficient search algorithm which finds the path between two points that globally maximize or minimize a cost function. When the cost function measures edge strength of an object, the resulting path is a highly accurate and robust delineation of the object. However the path is subject to the curvature constraint that the Euclidean distance to the end point must always decrease when proceeding from the beginning to the end. This prevents DP being used in a direct way to segment objects because the path cannot be a loop returning to the starting point. The solution to the problem is to transform the Cartesian image of the

---

This project was funded in whole or in part with federal funds from the National Cancer Institute (NCI), National Institutes of Health under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services and nor does mention of trade names, commercial products or organizations imply endorsement by the U.S. Government. We are thankful to Tom Mistelli (NCI, Bethesda) and Karen Meaburn (NCI, Bethesda) for providing the datasets. *Kaustav Nandy - nandyk@ncifcrf.gov, Prabhakar R Gudla - red-dyg@ncifcrf.gov and Stephen J. Lockett - slockett@ncifcrf.gov.*



**Fig. 1.** Flowchart illustrating steps for automatic segmentation

object into a polar image where the origin of the polar image is a point inside the object. This converts an object's boundary to a path from  $\theta = 0^\circ$  to  $\theta = 360^\circ$  which DP can find. The only constraint following the transformation is that the object must be point convex [9]. Mathematically we express the minimum of the cost function that defines the optimal boundary around the object as

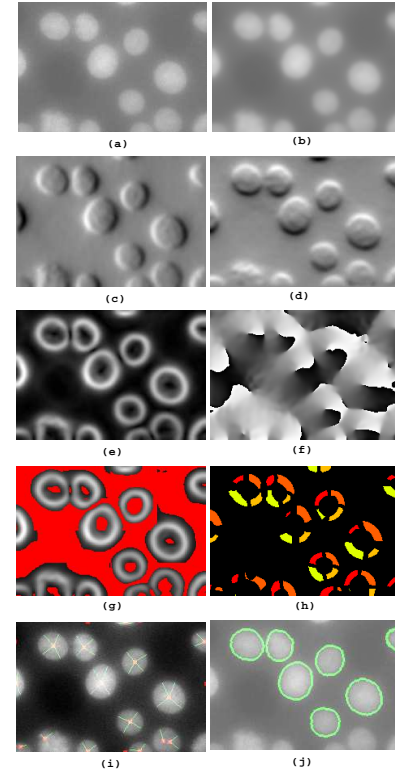
$$\min_{\{p\}} \left[ \sum_{\theta} I_{GM}(\theta, r_p(\theta)) \right] \quad (1)$$

where  $r_p(\theta)$  is the radial coordinate of the pixel on the path  $p$  at  $\theta$  and  $I_{GM}(\theta, r)$  is the intensity at  $(r, \theta)$  in the gradient magnitude image  $I_{GM}$ .

The organization of the paper is as follows. Section 2 describes and justifies the step by step procedure for segmentation. Section 3 presents the results of testing the method on datasets of DAPI-stained nuclei. The discussion, section 4 emphasizes the strengths of our method and indicates directions for further enhancement.

## 2. METHOD

Figure(1) is a flowchart that illustrates the steps followed in the procedure. Segmentation commences with image filtering to reduce noise (Box 1). This is followed by detection of multiple pairs of seed and boundary points corresponding to possible nuclei (Box 2,3 and 4). Each seed and boundary point pair is used by the DP to calculate an optimal region corresponding to candidate nucleus. Three or more regions that overlap in area by more than 90% are taken as the final



**Fig. 2.** (a) Original DAPI channel (b) DAPI channel after the preprocessing step (c) X-Derivative image (d) Y-Derivative image (e) Magnitude image (f) Gradient direction image (g) Thresholded magnitude image (h) Processed quadrant image after breaking up the large pieces (i) Grouped seed points identified and connected to the boundary points (j) Final segmented image

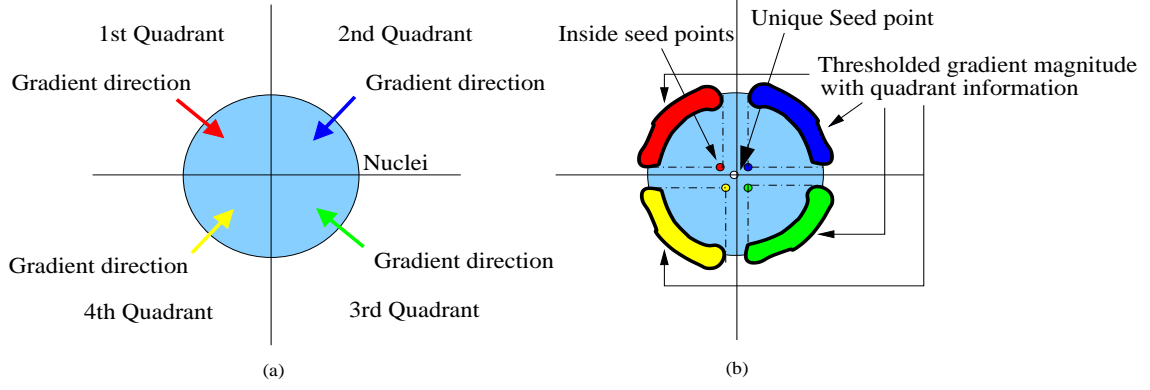
segmentation of a nucleus (Box 5 and 6). The subsequent subsections explain each step of the algorithm in more details.

### 2.1. Preprocessing

Microscope images are noisy and often have a lot of background variations making it difficult to analyze them. Hence it is necessary to clean up the images before segmentation. We apply Gaussian filtering to reduce the noise and then apply a corner preserving anisotropic filter to enhance the edges of the nuclei. Since the gradient calculation itself is very sensitive to high frequency noise, this procedure helps obtain a cleaner gradient vector image in the next step. Figure 2(a) shows the original DAPI channel and Figure2(b) shows the DAPI channel after the preprocessing step.

### 2.2. Gradient Vector Calculation

Our segmentation algorithm combines apriori information that nuclei are convex with image information that nuclei are high



**Fig. 3.** (a) Gradient directions in different quadrants (b) Gradient regions and the unique seed point

intensity objects on relatively dark backgrounds. This translates into approximate nuclear borders being associated with high gradient magnitudes directed towards the center of the nucleus. Therefore the first segmentation step is calculation of the gradient magnitude and direction of each pixel from the preprocessed DAPI image using a Gaussian derivative filter. Figure 2(c) shows the derivative in x-direction ( $\frac{\partial I}{\partial x}$ ) and Figure 2(d) in the y-direction ( $\frac{\partial I}{\partial y}$ ). The gradient magnitude and direction images are computed by combining the gradient images in x and y directions.

$$I_{GM} = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad (2)$$

$$I_{Dir} = \theta = \tan^{-1}\left(\frac{\frac{\partial I}{\partial y}}{\frac{\partial I}{\partial x}}\right) \quad (3)$$

where  $I$  is the preprocessed grayscale DAPI image.

Figure 2(e) and 2(f) show  $I_{GM}$  and  $I_{Dir}$  respectively.

### 2.3. Thresholding to identify boundary regions

In the gradient magnitude image the regions near the nuclei boundaries have very high intensity relative to background. However gradient magnitudes of both background and objects vary across the images. Consequently the boundary regions of the nuclei are extracted from the magnitude image by automatic, parameter free block thresholding that calculates thresholds that vary across the image [5]. This helps us to reliably extract the boundary regions even in the presence of considerable background intensity variations. The thresholded magnitude image is shown in Figure 2(g).

### 2.4. Breaking thresholded magnitude image into regions to identify seed points

The goal here is to identify several seed points near the center of each nucleus. This is done by first detecting contiguous

regions with gradient directions in quadrants with range  $0$  to  $\pi/2$ ,  $\pi/2$  to  $\pi$ ,  $\pi$  to  $3\pi/2$  and  $3\pi/2$  to  $2\pi$  radians (Figure 3(a)). In most cases, neighboring regions from the 4 quadrants correspond to single nuclei (Figure 3(b)). Larger regions however might correspond to more than one nucleus. Therefore large regions are automatically identified and split. This is done by first grouping all regions into 3 clusters based on size using K-means clustering [15]. Regions in the 2 larger clusters are split into smaller angular ranges. At this stage each region corresponds to a single nucleus in virtually all cases (Figure 2(h)). If we have a nearly circular isolated nucleus we will get four magnitude regions corresponding to the four quadrants in the phase image as shown in Figure 3(b). We take advantage of this pattern to identify the regions where a nuclei is present. Other nuclei may have more or fewer than 4 regions surrounding them, but this does not present a problem. In the next stage, an initial seed point inside a nucleus is determined from each region, using the maximum and minimum x and y coordinates of the region as the coordinates of the seed point (Figure 3(b)).

### 2.5. Unique seed point, corresponding boundary point identification and mean radius estimation

In this step we calculate a unique seed point for each nucleus, corresponding boundary points and estimate the mean radius of nuclei in the image. A unique seed point is calculated by grouping together initial seed points that are closest to the same 4 regions corresponding to 4 different quadrants. The coordinates of the unique seed point are the mean coordinates of the grouped initial seed points. Next, for each unique seed point, its corresponding regions are used to calculate boundary points. For each region the gray weighted distance transform (GWDT) [14] is calculated and the pixel with the maximum value is taken as the boundary point. The overall mean radius  $\bar{r}$  for the nuclei in the image is calculated as the mean distance from each boundary point to its unique seed point. In a final step we find revised seed points corresponding to each

---

<u>INITIAL SEED POINT PLACEMENT</u>
<i>for (each magnitude piece)</i>
<i>Place inside seed point using gradient direction information and coordinates of the region</i>
<i>end</i>
<u>GROUPING SEED POINTS</u>
<i>for (each seed point)</i>
<i>Find the spatial neighbors and form groups having same neighbors</i>
<i>end</i>
<u>MEAN RADIUS ESTIMATION</u>
<i>for (each group)</i>
<i>Find boundary points using gray weighted distance transform</i>
<i>Find distance between seed point and boundary points</i>
<i>end</i>
<u>FIND OVERALL MEAN RADIUS</u>
<u>REALLOCATION OF SEED POINTS USING MEAN RADIUS</u>
<i>for (each magnitude piece)</i>
<i>Refine seed point using gradient direction information and mean radius information</i>
<i>end</i>

---

**Table 1.** Steps in identifying inside seed points and mean radius estimation by seed point grouping

region utilizing  $\bar{r}$ . This is done for each region by locating the pixel at a distance equal to  $\bar{r}$  from the maximum value in the GWDT and in the direction of the quadrant ( $\pi/4$ ,  $3\pi/4$ ,  $5\pi/4$  or  $7\pi/4$ ). Figure 2(i) shows the unique seed points in the sample image connected to the corresponding boundary points. Table 1 shows an overview of the method for identification of the revised seed points.

## 2.6. Boundary determination using Dynamic Programming

Dynamic programming (DP) is used to calculate multiple boundaries around each nucleus. Each boundary utilizes one of the previously determined boundary points and the corresponding revised seed point. The DP search is restricted to a distance  $4\bar{r}$  from the seed point. Figure (4) illustrates the method. To find a unique boundary around each nucleus, the overlap of the areas enclosed by the multiple boundaries is calculated. If the overlap of at least 3 boundaries is more than 90% the areas are merged using the binary 'OR' operation. This results in a very high confidence of a correct segmentation and nuclei with uncertain boundaries omitted. Final segmentation image is shown in Figure 2(j).

## 3. EXPERIMENTS AND RESULTS

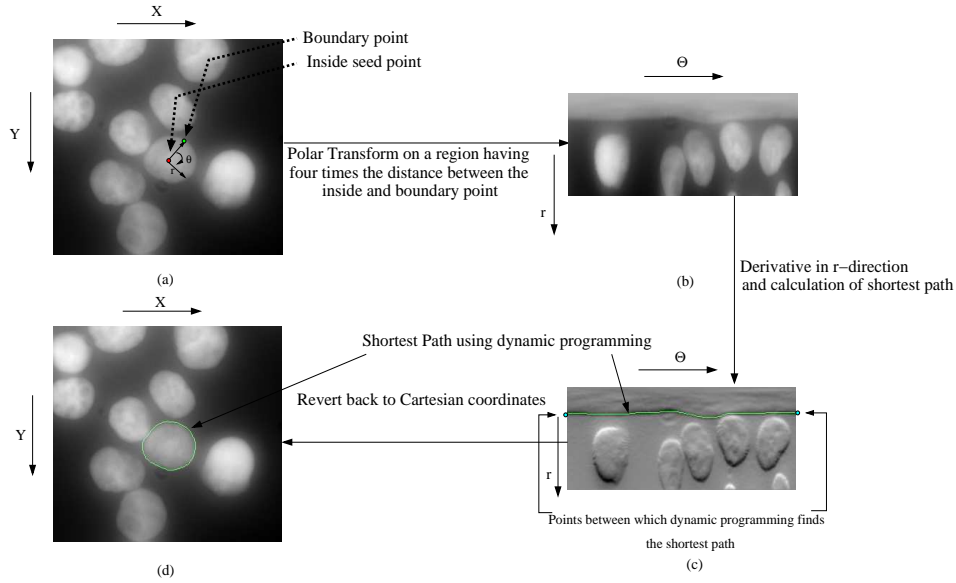
We tested the segmentation results on 8 datasets of images of cultured cell nuclei labeled with DAPI. Each dataset comes from a different sample and between 3 and 9 images were acquired from each sample, resulting in approximately 250 nuclei imaged per sample. In all the datasets the algorithm gave comparable results inspite of the fact that there were some significant differences between the datasets. In some of the

datasets the nuclei were distributed sparsely throughout the image while in some other datasets they were clustered together. None of the parameters in the algorithm had to be changed for segmentation in the 8 datasets. The method has the ability to segment the nuclei clusters to a high degree of accuracy. Figure 5 shows some of the representative images in the datasets with the segmentation results shown in the form of the yellow border around the nuclei. An unique feature of the algorithm is its ability to segment overlapping nuclei as indicated by the red boxes around a few examples in Figure 5.

For all the datasets we have manually counted the number of nuclei in each of the images and have compared it with the number of nuclei detected by our segmentation algorithm. The details about the datasets and the segmentation results are presented in Table 2. The average detection rate across all the datasets is 93.33%. The most important result is a high positive predictive value of 97.3% indicating that only 3% of detected objects are not nuclei. Hence we see that the algorithm has a very high detection rate across different datasets.

## 4. DISCUSSION AND CONCLUSION

In this paper we have presented an automatic method for segmentation of cell nuclei. The main idea of this method is to make it as simple and intuitive as possible so that it retains its generality and does not overfit the datasets that we have used in our experiments. In that way we can apply this method in a number of applications without loss of accuracy. Utilizing gradient vector and *a priori* shape information makes the method highly reliable for detecting clustered nuclei. Fur-



**Fig. 4.** (a) Original DAPI channel with an inside seed point and a boundary point demarkated (b) Region around the nuclei after polar transformation (c) Derivative image showing the shortest path having the minimum average intensity per unit length (d) Final segmentation in Cartesian coordinates.

Dataset Number	Manual count	Automatic count	Percentage count	PPV
1	272	231	84.93	96.6
2	225	203	90.22	98.5
3	242	221	91.32	99.6
4	208	206	99.04	99.5
5	262	252	96.18	97.7
6	250	234	93.6	92.0
7	215	213	99.06	97.7
8	350	323	92.29	97.0
Total / Average	2024	1883	93.3	97.3

**Table 2.** Table showing the experimental results on 8 datasets used

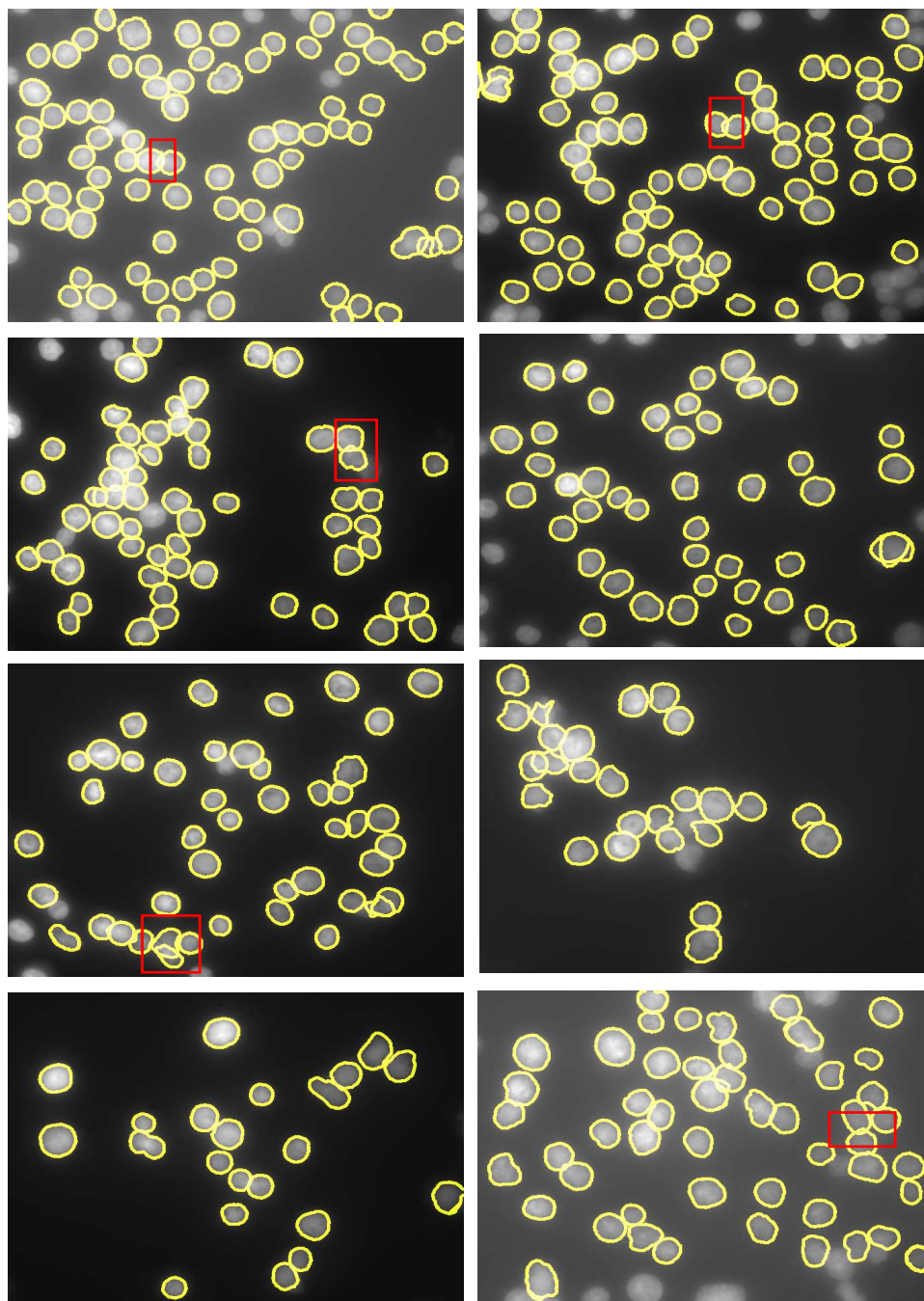
thermore dynamic programming results in optimal borders around objects that are highly robust against noise and compensates for the inaccuracies caused by automatic identification of inside seed points and boundary points. The method incorporates a number of other important features. It uses block thresholding to negate the effect of background variations to a considerable extent. The process of breaking up the gradient direction and magnitude into 4 quadrants based on angular direction instead of considering each pixel individually adds to the robustness of the system because it utilizes the image information in an integrated way. A unique property of our method is that we obtain multiple preliminary segmentations of each nucleus, which we combine into an optimal segmentation. This leaves us free to adjust the criteria for combining trial segmentations, for example the percentage of

overlapped area. By applying more stringent constraints we can be more selective about which segmentations are considered correct. This allows us to increase the PPV at the price of reducing the proportion of nuclei detected. This is valuable in many applications where highly reliable segmentation is beneficial while there are ample nuclei available for analysis.

## 5. REFERENCES

- [1] Chin K, Ortiz de Solorzano C, Knowles D, Jones A, Chou W, Rodriguez EG, Kuo WL, Ljung BM, Chew K, Myambo K, Miranda M, Krig S, Garbe J, Stampfer M, Yaswen P, Gray JW, Lockett SJ, *3D Image Analysis of Thick Breast Cancer Specimens show High Cell-*

- to-Cell Genetic Heterogeneity. Nature Genetics*, 36:984-988,2004
- [2] Meaburn KJ, Misteli T, Soutoglou E , *Spatial genome organization in the formation of chromosomal translocations, Semin Cancer Biol.*,2007 Feb;17(1):80-90
- [3] Beaver W, Kosman D, Tedeschi G, Bier E, McGinnis W, Freund Y, *Segmentation of nuclei in confocal image stacks using performance based thresholding, Proceedings of International Symposium on Biomedical Imaging 2007*, 53-56
- [4] Gang Li, Tianming Liu, Jingxin Nie, Lei Guo and Wong STC, *Segmentation of touching cells using gradient flow tracking, Proceedings of International Symposium on Biomedical Imaging 2007*, 77-80
- [5] Gudla Prabhakar R, Collins J, Meaburn KJ, Misteli T and Lockett S J, *HiFLO: a high-throughput system for spatial analysis of FISH loci in interphase nuclei, Proceedings of SPIE - Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues*, 6441, 2007
- [6] Gang Lin, Umesh Adiga, Kathy Olson, John F. Guzowski, Carol A. Barnes and Badrinath Roysam, *A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks, Cytometry Part A*, vol. 56A, no. 1, pp. 2336, 2003.
- [7] Xiaowei Chen, Xiaobo Zhou, and Stephen T. C. Wong, *Automated Segmentation, Classification, and Tracking of Cancer Cell Nuclei in Time-Lapse Microscopy, IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 4, 762-766, April 2006
- [8] Sophie Schpp, Abderrahim Elmoataz, Mohamed-Jalal Fadili and Daniel Bloyet, *Fast Statistical Level Sets Image Segmentation for Biomedical Applications, Lecture Notes in Computer Science*, Volume 2106/2001, pp 380, 2001
- [9] Daniel Baggett, Masa-aki Nakaya, Matthew McAuliffe, Terry P. Yamaguchi and Stephen Lockett, *Whole cell segmentation in solid tissue sections, Cytometry Part A*, Volume 67A, Issue 2 , Pages 137 - 143, 2005
- [10] Dean McCullough, Prabhakar Gudla, Karen Meaburn, Amit Kumar, Michael Kuehn and Stephen Lockett, *3D Segmentation of whole cells and cell nuclei in tissue using dynamic programming , Proceedings of International Symposium on Biomedical Imaging 2007*, 276-279
- [11] Min Hu, Xijian Ping and Yihong Ding, *Applying Fuzzy Growing Snake to Segment Cell Nuclei in Color Biopsy Images, Lecture Notes in Computer Science*, Volume 3314/2004, 672-677, 2004
- [12] Xiaoxu Wang, Weijun He, Metaxas D, Mathew R and White E, *Cell segmentation and tracking using texture-adaptive snakes, Proceedings of International Symposium on Biomedical Imaging 2007*, 101-104
- [13] Begelman G, Gur E, Rivlin E, Rudzsky M and Zalevsky Z, *Cell nuclei segmentation using fuzzy logic engine, Proceedings of International Conference on Image Processing*, Volume 5, 2937 - 2940, 2004
- [14] Mullikin JC, *The vector distance transform in two and three dimensions, CVGIP- Graph. Model. Im.*, 54(6), 526535, 1992.
- [15] Richard O. Duda, Peter E. Hart and David G. Stork, *Pattern Classification, Wiley Interscience*, 2<sup>nd</sup> Edition.



**Fig. 5.** Figure showing one sample segmented image from each of the datasets. The red boxes show correct segmentation of overlapping nuclei.