International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

Detection and Ranking of Cancer Cells by Automatic Cell Segmentation using Nuclear-to-Cytoplasm Ratio

Pooja Rajendra Prasad[1], M. Thiruchelvi[2], Durga. D. Thankachan
[1,3]*Student of Engineering,* [2]*Asst.Professor of the Department, Department of CSE, National College of Engineering, Tamil Nadu, India*

Abstract **– Conventional biopsy procedures required invasive tissue removal from a living subject, followed by time-consuming and complicated processes. A noninvasive *in vivo* virtual biopsy is highly required which has the potential to obtain tissue images without actually removing them. The proposed cell segmentation approach is based on the watershed-based approach and the concept of convergence index filter for automatic cell segmentation. By experimental analysis the proposed algorithm reveals high accuracy for cell segmentation and has a potential for noninvasive analysis of cell nuclear-to-cytoplasmic ratio (NC ratio), which is important in identifying or detecting early symptoms of diseases with abnormal NC ratios, such as skin cancers during clinical diagnosis via medical imaging analysis. Classification is done further by ranking the cells based on their feature characteristics modeling abnormality degrees. According to the procedure a tree is constructed using hierarchical clustering, and then arranging the cells in a linear order by using an optimal leaf ordering algorithm that maximizes the similarity of adjacent leaves.**

Keyword **– Cell segmentation, convergence index filter, nuclear-to-cytoplasmic ratio (NC ratio), third harmonic generation (THG), watershed transform.**

I. INTRODUCTION

Traditionally, the early detection of skin cancer is obtained by physical biopsy, which involves the removal of suspected tissue, preparation of abundant samples, and evaluation at the microscopic level.

The 1230 nm-based higher harmonic generation microscopy [1], which combines Second Harmonic Generation (SHG) and Third Harmonic Generation (THG)modalities, has been successfully and safely used for noninvasive *in vivo* virtual optical biopsy on skin and oral mucosa with a penetration depth of [7], for applications including skin age determination and early cancer diagnosis [6]- [8].

Microscopy based on THG has been shown to accurately evaluate cytological malignancy in skin cancer development, due to its capability to provide a clear boundary definition between the epidermis nuclei and cytoplasm [1], [7]. As demonstrated by [2], [9], the cell nuclear-to-cytoplasmic ratios (NC ratios) of epidermis are generally larger in skin cancer than in normal cells. So the NC ratio in the epidermis is a good indicator for identifying early symptoms of skin cancer in addition to revealing the type and stage of the developing disease. Traditional

NC ratio analysis, however, is performed by manual selection of cells from each THG image and subsequent calculation of their volume ratio. In addition to being time consuming, this procedure causes fatigue of the medical personnel and hence inconsistent accuracy of results caused by, for example, visual error.

There are many published algorithms for cell segmentation. Watershed transform [5] is one that uses the concept of morphological image processing and considers contextual information to produce stable segmentation results. Convergence index filters [4] using the distribution of the orientation of gradient vectors are local filters designed for edge or boundary enhancement in images with weak contrast and irregular noise, especially in biomedical images. Considering the tradeoff between algorithmic complexity and the precision of segmentation results in our application, we adopt the watershed-based approach and incorporate some principles of the sliding band filter to segment cells in vivo virtual biopsy images of human skin. We then propose an unsupervised non-parametric ordering procedure that uses a tree structure formed by hierarchical clustering. First, we present the features that are used for describing the segmented cells. Then, we describe the details of the ordering algorithm.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

In this paper, we focus on the design and development of an algorithm that automatically isolates and segments individual cell from THG virtual biopsy images, whose segmentation results improve on our previous work [3]. This will eventually improve the efficiency and accuracy of the analytical part of virtual biopsy procedures. Section II details the physical background of the acquired images. Section III presents the proposed algorithm for cell segmentation. Section IV shows the experimental results. Section V presents the conclusion and discussion.

II. Physical Background

Biomedical studies and research has recently been conducted focusing on higher harmonic microscopy utilizing THG modality, which has emerged as a new optical virtual biopsy paradigm [2]. Obeying the law of conservation of energy, THG involves only virtual electron transitions. Hence, the possibility for photo damage and photo bleaching is reduced as this process deposits no energy in the interacted matter.

The relationship between the induced polarization Ṕ and the electric field Ẽ in nonlinear optics:

$$\tilde{P}(t) = \varepsilon_0 \chi^{(1)} \tilde{E}(t) + \varepsilon_0 \chi^{(2)} \tilde{E}(t)^2 + \varepsilon_0 \chi^{(3)} \tilde{E}(t)^3 + \cdots$$
$$= \tilde{P}^{(1)}(t) + \tilde{P}^{(2)}(t) + \tilde{P}^{(3)}(t) + \cdots,$$

Where ε_o represents the electric permittivity, and $\chi(i)$ denotes i^{th} order nonlinear susceptibility. The resolution of each image is 512 x 512 and is stored in gray-level 12-bit TIFF files, where pixel values are proportional to their third harmonic responses.

III. Proposed Algorithm

The proposed algorithm is categorized into mainly four parts: Nuclei segmentation, Cytoplasm segmentation, NC Ratio Evaluation and Ranking of cells. For nuclei segmentation, gradient watershed transform [5] based on the contextual information of pixels in images is used.. For cytoplasm segmentation, cytoplasm initialization would be performed based on the information of valid nuclei to obtain potential cytoplasms and corresponding refined cell boundaries. The NC ratios are obtained in the NC ratio evaluation stage. Finally ranking of the cells is done in the order of their abnormality degrees.

A. Nuclei Segmentation

The process of nuclei segmentation is performed using gradient watershed transform with marker-controlled strategy, blob detection, and consideration of shape descriptors to obtain accurate segmented nuclei. Nuclei segmentation is a relevant part in the proposed cell segmentation algorithm because the subsequent stage of cytoplasm segmentation references valid nuclei, which can be thought of as the initial condition of the optimization problem of whole cell segmentation in order to guide the algorithm in finding a feasible solution with high performance.

1) Watershed Transform with Marker-Controlled Strategy:

For considering contextual information in an image and identifying the regional minima, watershed transform is chosen here for nuclei segmentation. In observation of images, nuclei with lower intensity locate in the regional minima of images, which are highly correlated with homogeneous regions in the gradient map. Therefore, we utilize the gradient map with prior information for the boundaries of nuclei to help to extract and analyze the almost uniform and round nuclei of adjacent cells from the background. Sobel operator and Roberts cross-gradient operator is used to calculate gradient maps in the horizontal, vertical, and positively sloped diagonal, and negatively sloped diagonal directions.

Due to the existence of undesired regional minima and irrelevant noise in the gradient map, over-segmentation problem may exist if nuclei detection is performed using direct watershed transform of the gradient map. So a filter that should be implemented at the initial stage of nuclei detection to exclude the undesired regional minima and preserve the desired ones should be designed. Nuclei initialization would prevent over-segmentation problem by adopting the marker-controlled strategy followed by the technique of minima imposition existing in morphological image processing.

Two kinds of markers are needed: those that roughly mark the locations of potential nuclei and those that mark their corresponding cell boundaries. These two kinds of markers are defined respectively as

- *Internal markers:* the groups of connected pixels inside each region where the potential nucleus is to be segmented.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

- *External markers:* the groups of connected pixels ideally relative to the boundary of each cell.

External markers prevent over-flooding in the watershed transform and the internal markers are the initial points that guide the watershed transform to obtain the desired solution that fits the size and shape of the desired nuclei.

2) Nuclei Initialization:

Nuclei initialization is to obtain the potential nuclei and their corresponding cell boundary using blob detection followed by outlier removal as well as distance transform. The main block diagram of nuclei initialization is presented in Fig. 2.

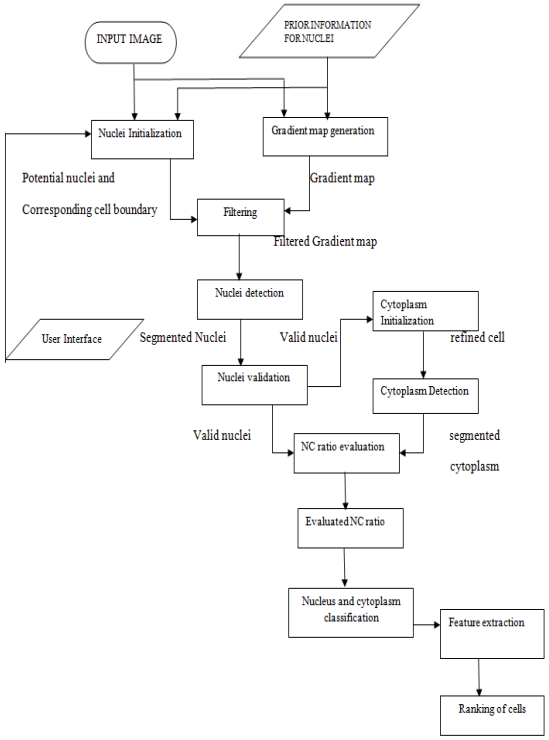

Fig.1. Block diagram of cell segmentation and NC ratio analysis.

The potential nuclei are marked with internal markers, and the corresponding cell boundaries are marked with external markers. Internal markers must meet the following criteria:

- Regions of internal markers must be surrounded by pixels of higher intensity.
- All pixels in each region of internal markers should form a connected component with homogeneous intensity.

Blob detection is performed with prior information of the nuclei to be segmented. Outliers of nuclei candidates are excluded by outlier removal to obtain potential nuclei, which are also called internal markers. Distance transform of internal markers is performed to obtain a distance map followed by cell boundary initialization to determine the potential cell boundaries, which are also called external markers.

a) Blob Detection:

The original image to be segmented is spanned into the scale-space to interpret the multi-scale representation of the image in order to extract structures of interest or feature points with the scale and spatial information concurrently.

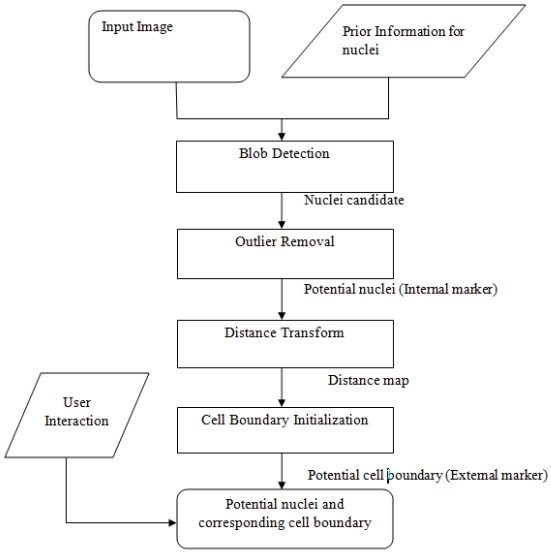

Fig. 2. Block diagram of nuclei initialization.

The input image is projected into the scale-space with Gaussian kernel ‘g’ and then the scale-space derivatives is utilized to extract the scale and spatial information of blobs concurrently in the images. The scale-space representation ‘L’ of a two-dimensional image ‘f’ is denoted by

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

$$L(.\ ,.\ ;\ \sigma\ ) = f(.\ ,\ .)*g(\ .\ ,.\ ;\ \sigma)$$

Where 'g' is the Gaussian operator denoted by

$$g(\ x,y;\ \sigma) = (1/2\pi\sigma^2)e^{(-(x^2+y^2)/2\sigma^2)}$$

σ – scale parameter
*-convolution operator

b) Outlier Removal:

Removing the possible outliers among the nuclei candidates requires applying local constraints. Valid nuclei must be surrounded by cytoplasm that has sufficient intensity and area. Local constraints adaptive to blob size are considered to exclude the outliers and preserve the valid nuclei that are completely surrounded by cytoplasm having sufficient intensity and area.

c) Distance Transform:

The detected blobs without outliers will now serve as the internal markers marking the rough position of potential nuclei. We calculate the Euclidean distance maps of the internal-marker map, which comprise the Euclidean distance between each pixel and the nearest pixel belonging to an internal marker. External markers roughly delineating cell boundaries will help form a complete marker map for the watershed transform with the marker-controlled strategy. The internal-marker map indeed roughly demarcates the potential nuclei, and the external-marker map roughly draws the cell boundary.

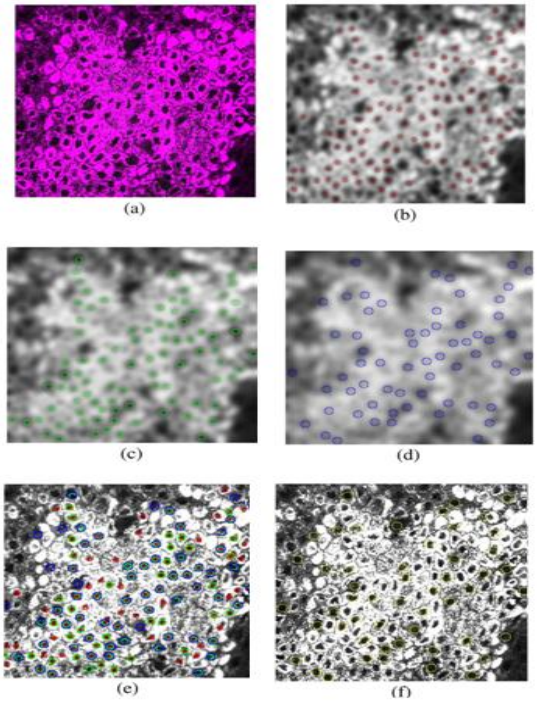

Fig. 3. The scale-space representation at various scales. (a) Input image.(b)–(d) Detected blobs on the scale-space representation at scale σ=5 , 7, and9, respectively. (e) Super-imposition of detected blob maps at scale σ=5 , 7,and 9. (f) Final detected blobs to be considered as nuclei candidates.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

3) Morphological Image Processing:

Now we design a filter using morphological minima imposition to remove the undesired regional minima in the gradient map and preserve the desired ones to resolve the over-segmentation problem of the watershed transform. The internal and external markers are used to perform minima imposition on the gradient map to obtain a filtered gradient map, whose regional minima only occur at the location of internal and external marker pixels.

4) Nuclei Detection and Validation:

The watershed transform is used again to calculate the watersheds of the filtered gradient map and obtain the segmented nuclei from original skin cell images. To help ensure that the nuclei have been accurately segmented, we also consider their shape. Compactness, which indicates irregularity associated with cancer cells, is utilized in this stage of nuclei validation and is defined as

$$\text{Compactness}=A/P^2$$

Where 'A' represents the area of the object and 'P' represents the perimeter of the object.

B. Cytoplasm Segmentation

A convergence index filter with intuitive parameter setting is adopted based on valid nuclei. Since it considers gradient vectors instead of intensity of images, a convergence index filter is suitable for low-contrast and noisy microscopy images. Such a filter makes unnecessary the pre-processing to enhance contrast and remove irregular noise in biomedical images, while preserving the information needed for clinical diagnosis.. Convergence index filters consider contextual and locality information to have confident segmentation results and reduce uncertainty of segmentation resulting from low contrast and noise in images.

1) Cytoplasm Initialization:

Two constraints, R_{min}, and R_{max} represent the minimum distance from the inner boundary and the maximum distance from the outer boundary to the center P of the support region on each orientation, respectively as in fig.4. These two constraints are adaptive to the shape and position of each valid nucleus.

The boundary of each valid nucleus can be thought of as a minimum boundary R_{min} of cytoplasm on that cell that makes sure that the candidates of cytoplasmic region are not present in the nuclear region.

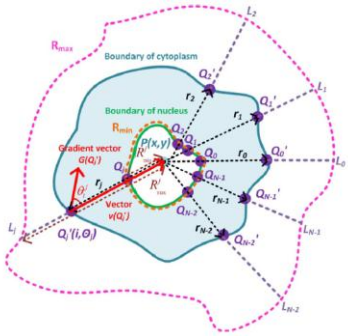

Fig.4. Support region of proposed local filter.

We adopt the concept of distance transform to generate the refined distance map of valid nuclei to delineate the maximum boundary of potential cytoplasmic region for each cell for determining R_{max}. This avoids the overlap of segmented cells and also resolves issues of cell segmentation of adjacent cells with indistinct boundaries in multivariate biomedical images.

2) Cytoplasm Detection:

The support region of the proposed local filter can be thought of as the union of N line segments Q_jQ_j', which represent cytoplasmic width radiating from cell center on each orientation in Fig.4.

The output of the proposed local filter applied to the pixel of interest P of Cartesian coordinate (x,y) is defined as the average of convergence indices on N line segments in 2-D discrete space

$$\text{Filter}(x,y)=\frac{1}{N}\sum_{j=0}^{N-1}\text{Filter}_j(x,y)$$

with

$$\text{Filter}_j(x,y)=\frac{1}{r_j-R_{\min}^j}\max_{R_{\min}^j<r_j\le R_{\max}^j}\sum_{i=R_{\min}^j+1}^{r_j}\cos\theta_i^j$$

$$\cos\theta_i^j=\cos\left(\theta_j-\phi(i,\theta_j)\right)$$

$$\theta_j=\frac{1}{N}2\pi j,\ where\ j=0,1,2,\cdots,(N-1)$$

$$\phi(i,\theta_j)=\tan^{-1}\left(\frac{G_y(i,\theta_j)}{G_x(i,\theta_j)}\right)$$

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

C. Cell Size and NC Ratio Evaluation

Developing stage of some diseases, skin aging and other quantifiable physical factors can be identified by abnormalities in cellular size and nuclear size. . For example, cellular and nuclear size in the layers of basale cells in forearm skin has been found to increase with age.

A protocol has been developed to obtain accurate NC ratios. Although NC ratio is defined as a volume ratio of nucleus to cytoplasm, it can be approximated by an area ratio of nucleus to cytoplasm.

D. Classification and Ranking of Cells

Different from many other unsupervised clustering approaches, the proposed procedure does not make any assumption about the distribution of the groups and does not require any information regarding the number of groups in the data. We pose the grouping problem as the ranking of cells according to their abnormality degrees. We propose an unsupervised non-parametric ordering procedure that uses a tree structure formed by hierarchical clustering. First, we present the features that are used for describing the segmented cells. Then, we describe the details of the ordering algorithm.

1) Feature Extraction:

Changes of cells can be associated with cell characteristics like size, color, shape, and texture of nucleus and cytoplasm. We describe each cell by using different features related to these characteristics. Then, a set of features is extracted as:

- *Nucleus area:* The number of pixels in the nucleus region.
- *Nucleus brightness*: The average intensity of the pixels belonging to the nucleus region.
- *Cytoplasm area:* The number of pixels in the cytoplasm part of a cell region divided by the number of cells in that cell region. We assume that the total cytoplasm is shared equally by the cells in a cell region.
- *Cytoplasm brightness:* Calculated similar to the nucleus brightness. However, overlapping cells are associated with the same value of the cytoplasm brightness.
- *Nucleus/cytoplasm ratio:* It is given by the ratio of the nucleus area to the cell area which is calculated as the sum of the nucleus and cytoplasm area.

2) Ranking of Cells:

Hierarchical clustering is used to produce a grouping of cells according to the features described above. Hierarchical clustering constructs a binary tree in which each cell corresponds to an individual cluster in the leaf level, and the two most similar clusters merge to form a new cluster in the subsequent levels. We use the Euclidean distance for computing the pair-wise feature distances and the average linkage criterion to compute the distance between two clusters.

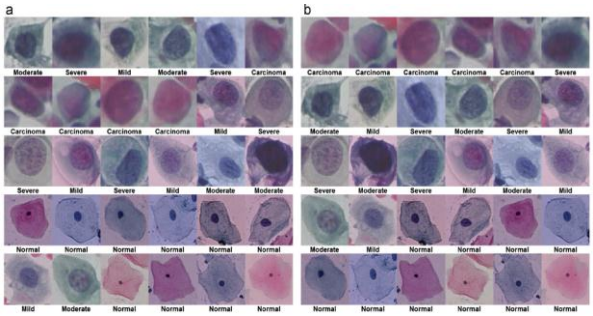

Fig.5. Example ordering of cells. (a) Initial ordering of the cells. (b) Final ordering obtained by applying the optimal leaf ordering algorithm.

Fig. 5(a) shows the cells and their class names corresponding to this ordering. The dysplastic cells are found at the beginning of the ordering and the normal cells are grouped at the end of the list. However, the sub-classes of dysplastic cells are not accurately ordered according to their dysplasia degree, and the group of normal cells at the end of the list contains some dysplastic cells as well. The main reason is that many heuristic orderings only consider local associations and do not consider any global consistency.

IV. Experimental Results

Automatic cell segmentation and NC ratio evaluation were performed using the proposed algorithm on about 600 THG virtual biopsy images of the stratum basale layer of human forearm skin from 31 healthy volunteers.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

The evaluated NC ratios and cell sizes were discussed and interpreted by a dermatologist and a radiologist. Profile of the experimental cell segmentation, including the evaluated NC ratios, cell sizes, and position, is given in Table 1. For evaluating the performance of classification, we obtain the ground truth ranking U where we know the rank Ui of each cell qi, i= 1, . . . ,I. An example set of cells q1, q2, q3, q4, q5, q6, q7, q8 belonging to three different classes are given in Table 2. The ranks of the cells with the same class label should be the same so we assign all of these cells to the mean of their initial ranks and obtain the ground truth ranking U.

Table 1:
Profile of each segmented cell

Cell Index	Cell Area (pixels)	Nuclear Area (pixels)	Cytoplasmic Area (pixels)	NC Ratio	Position (row,column)
1	924	229	695	0.329496403	(476 , 16)
2	649	160	489	0.327198364	(289 , 36)
3	901	228	673	0.338781575	(453 , 66)
4	1076	316	760	0.415789474	(422 , 91)
5	836	201	635	0.316535433	(451 , 91)
...	...	...	...	...	...
49	862	222	640	0.346875	(143 , 364)
50	972	319	653	0.488514548	(242 , 374)
51	943	296	647	0.457496136	(442 , 370)
52	722	217	505	0.42970297	(452 , 394)
53	1226	303	923	0.328277356	(222 , 398)
54	1221	408	813	0.501845018	(460 , 419)
55	1357	400	957	0.417972832	(267 , 428)
56	1341	452	889	0.508436445	(193 , 464)
57	1323	195	1128	0.17287234	(143 , 467)
58	1066	265	801	0.330836454	(157 , 490)
Total	56250	15732	40518	-	-
Average	969.8276 (pixels)	271.2414 (pixels)	698.5862 (pixels)	0.388271879	-
Average	53.2742 (μm^2)	14.8997 (μm^2)	38.3745 (μm^2)	-	-

Then, suppose that our algorithm ranks these cells in the order of q1, q5 ,q2, q4, q3, q7,q6, q8.

Table 2
An example ranking scenario. Eight cells are assumed to belong to three classes. The ground truth ranking U and the algorithm's ranking V are calculated according to the scenario described in the text.

Cells	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
Class labels	1	1	1	2	2	2	3	3
Initial ranking	1	2	3	4	5	6	7	8
Ground truth ranking *U*	2	2	2	5	5	5	7.5	7.5
Algorithm ranking *V*	2	2	5	5	2	7.5	5	7.5

Since we aim to order the cells according to their abnormality degrees, we can hypothesize that our method labels the first three cells, namely q1,q5,q2, asclass1, the next three cells, namely q4,q3,q7, as class2, and the last two cells, namely q6,q8, as class3 because the classes 1,2 and 3 are known to have three, three, and two images, respectively, in the ground truth.

When we calculate the cell rankings based on these class associations, we obtain the ranking result V shown in Table 2 where each cell qi has a corresponding rank Vi ,i =1, . . . ,I. Finally, we measure the agreement between the ground truth ranking U and our ranking result V statistically using the Spearman rank- order correlation coefficient and the kappa coefficient.

V. CONCULSION

Here a computer-aided design for automatic cell segmentation, NC ratio analysis and ranking of cells was proposed. The experimental results show that the proposed method provides objective segmentation results with high efficiency and consistent accuracy. The evaluated NC ratio values are very close to the results of manual cell selection, indicating that the proposed work considering not only the performance of analysis procedure but also practical criteria as well as clinical requirement has significant potential for biomedical imaging analysis and medical values in a variety of applications. It can help medical doctors to noninvasively and immediately identify early symptoms of diseases, especially fatal diseases like cancer, that involve abnormal NC ratios. It can also quantify several physical factors correlated with NC ratios or cell size, such as skin aging. The ranking enabled modeling abnormality degrees.

REFERENCES

[1] C.-K. Sun, "Higher harmonic generation microscopy," Adv. Biochem.Eng./Biotechnol., vol. 95, pp. 17–56, 2005.

[2] E. Vega-Memije, N. Martinez de Larios, L. M. Waxtein, and L. Dominguez-Soto, "Cytodiagnosis of Cutaneous Basal and Squamous Cell Carcinoma," Int. J. Dermatol., vol. 39, no. 2, pp. 116–120, 2000.

[3] H.-H. Lin, M.-R. Tsai, C.-F. Chen, S.-Y. Chen, Y.-H. Liao, G. G. Lee, and C.-K. Sun, "Cell segmentation and NC ratio analysis of Third Harmonic Generation virtual biopsy images based on marker-controlled gradient watershed algorithm," Proc. IEEE Int. Symp. Circuits and Systems, pp. 101–104, 2012.

[4] H. Kobatake and S. Hashimoto, "Convergence index filter for vector fields," IEEE Trans. Image Process., vol. 8, no. 8, pp. 1029–1038, 1999.

[5] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithmbased on immersion simulations," IEEE Trans. Pattern Anal. Machine Intell., vol. 13, pp. 583–598, 1991.

[6] M.-R. Tsai, D.-B. Shieh, P.-J. Lou, C.-F. Lin, and C.-K. Sun, "Characterization of oral Squamous Cell Carcinoma based on higher-harmonic generation microscopy," J. Biophoton., vol. 5, no. 5–6, pp. 415–424,2012.

[7] S.-Y. Chen, S.-U. Chen, H.-Y.Wu, W.-J. Lee, Y.-H. Liao, and C.-K.Sun, "In vivo virtual biopsy of human skin by using noninvasive higher harmonic generation microscopy," IEEE J. Sel. Topics Quantum Electron., vol. 16, no. 3, pp. 478–492, 2010.

[8] Y.-H. Cheng, C.-F. Lin, and C.-K. Sun, "A novel intravital multi-harmonic generation microscope for early diagnosis of oral cancer," presented at the Optical Biopsy X Conf., Photonics West, San Francisco, CA, USA, 2013, paper 8577-25.

[9] Y. Oram, O. Turhan, and E. A. Din, "Diagnostic value of cytology in basal cell and Squamous Cell Carcinomas," Int. J. Dermatol., vol. 36, no. 2, pp. 156–157, 1997.

[10] Z. Bar-Joseph, D.K. Gifford, T.S. Jaakkola, Fast optimal leaf ordering for hierarchical clustering, Bioinformatics 17 (2001) S22–S29.