



Segmentation of Stem Cell Microscopic Image Using Morphological Hessian Based Approach

Dr.R.Rajeswari¹, Dr.G.Nirmalapriya²

Professor, Department of ECE, Rajalakshmi Institute of Technology, Chennai, India¹

Professor, Department of ECE, Rajalakshmi Institute of Technology, Chennai, India²

Abstract: Cell segmentation is the first stage of a (semi)automatic stem cell tracking system. A vast amount of toxicological data can be obtained from the feature analysis of cells treated in vitro. But, this requires microscopic image segmentation of cells. The main aim of this paper is to propose a new strategy, namely Morphological Hessian Based Approach (MHB Approach), to segment cells that are partially overlapped with a large amount of curved edges. MHB approach is a machine learning based method, where loosely annotated images are first used to train and optimize parameters, and then the optimal parameters are inserted into a Normalized Cut segmentation process. Furthermore, segmentation results are compared using MHB Approach to another four classical and three state-of-the-art methods. The overall experimental result shows better results with usefulness and effectiveness of this method over the six methods under consideration.

Keywords: Image Segmentation, Stem Cells, Machine Learning, Supervised Normalized Cut, Morphological Hessian Based Approach

I. INTRODUCTION

Recent toxicology uses modification in individual cells as endpoint to determine and quantify harmful effects. One conceptualization is to use human migratory neural crest cells, a transient fetal stem cell type, to study developmental toxicity by investigating changes in cellular features observable by microscopy. These cells change their morphology upon treatment with chemicals and different chemicals induce different cell shapes. If two compounds induce the same cell shape, they interfere with the same biological process. Therefore, in the first stage, segment microscopic images of these cells then analyze cellular features in the later stage. As these cells are grown in relatively dense culture, they are in contact with each other and have a lot of curved edges (Fig. 1). This creates difficulty in their segmentation.

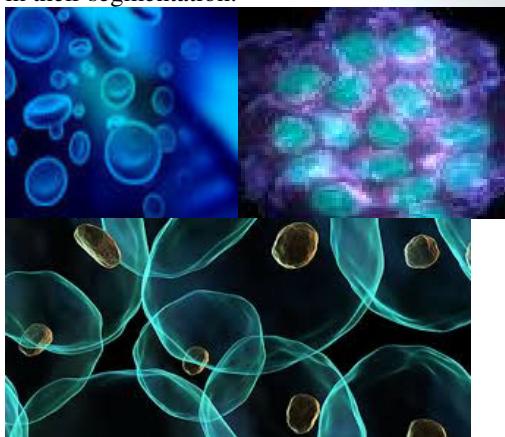


Fig. 1 Microscopy image of Stemcell

Existing segmentation methods such as edge detection or watershed transformation segments the main bodies of cells in a better way, but the efficiency decreases when the cells are dense or when they have many curved edges. For example, edge detection algorithms can localize edges, but when the density of objects is very high, many adjacent edges link together and lead to a wrong segmentation result. Watershed algorithm usually merges many unrelated regions and easily results in an over-segmentation of the image, it cannot exactly capture the curved edges of cells under biological treatment. Due to these two problems, a machine learning based segmentation method is used. The Supervised Normalized Cut Segmentation (SNCS), which builds a data-driven system using input examples, leading to a more accurate segmentation result. This method enables to segment cells even if they touch each other and have many curved edges. The SNCS approach is constituted by a training and segmentation step. In the first stage, training images are first manually segmented, then they are used in a supervised learning framework for optimizing parameters for the segmentation. In the second step, the trained parameters are inserted into a Normalized Cut (NC) approach [1] to obtain the segmentation results for test images.

II. RELATED WORK

2.1. Cell Segmentation: Various image segmentation methods have been applied to different kinds of cell images. Segmentation and classification of cancer cells is discussed in [2] and [3]. In [4] and [5], image segmentation methods are presented for the blood cell analysis. Other applications



include segmentation of muscle cells [6] and brain cells. However, there are very less work on stem cell segmentation, and the challenges of stem cell segmentation result from cell overlapping and their curved edges [8].

2.2. Segmentation Approaches: Medical image segmentation methods are mainly categorized into ‘boundary based’, ‘region based’ and ‘model based’ [9]. Boundary based methods are based on the edges between different regions, which are fast to compute and do not need priori information about image contents. For example, many edge detection algorithms (e.g. Sobel, prewitt) are based on the change of the gradient of gray levels [10]. In [11] a new contour detection and hierarchical image segmentation method is proposed, where local and global information is combined to improve the segmentation performance of the basic edge detection methods. However, a main disadvantage of boundary based methods is that they have a low efficiency for segmenting dense objects. Region based segmentation methods are based on the similarities between regions, including region growing, watershed and NC methods. The region growing method segments an image by connecting neighboring pixels based on a similarity criterion [12]. The watershed algorithm uses image morphology to segment images, and can recognize different regions in an image by using a morphological watershed transformation. The above three methods exhibit two problems: Images of dense objects are over segmented, and all methods have a low efficiency for capturing cells with curved edges. In contrast, the NC method computes a similarity matrix based on intervening contours and segments the image according to a distance criterion [1]. It shows a promising ability to solve the over-segmentation and curved edge problems. Markov Random Field (MRF) based segmentation method is a model based segmentation approach, which is a natural way to incorporate spatial correlations into a segmentation process [13]. The MRF is a stochastic process that specifies the local characteristics of an image, and it is combined with the given data to reconstruct the true image. Among all the methods mentioned above, the NC method can overcome the over-segmentation and curved edge problems robustly. However, NC is an unsupervised approach, which does not use any priori information of cell shapes, resulting in a high possibility of great deviation within the segmentation results. To reduce the risk of such deviation, the Morphological Hessian Based Approach, a supervised learning framework [14], where loosely annotated images are used for training and optimizing parameters of the NC approach is proposed. Hence, this supervised approach further improves the segmentation performance of the NC method, leading to an efficient method for stem cell segmentation.

III. SEGMENTATION METHOD

3.1. Normalized Cut Method

Consider an undirected weighted graph $G(V, E)$ that is to be partitioned into two disjoint subgraphs $G_A(A, E_A), G_B(B, E_B)$, where $A \cup B = V$. Partitioning can be achieved by removing the edges connecting A to B. The cost of partitioning G is called the cut cost and is defined by the total weight of the edges that have been removed:

$$Cut(A, B) = \sum_{u \in A, v \in B} w(u, v) \quad (1)$$

where the vertex set $V = \{v_1, v_2, \dots, v_n\}$, and where W is the weighted adjacency matrix of G.

To find an optimal partitioning of G, the main strategy is to minimize eq.(1) to find the minimum cut. As described in [17], however, such a minimum cut can be unnaturally biased towards partitioning in which one of the sub graphs has a single vertex. To yield more balanced partitioning, other related cut costs have been proposed, one of the most popular of which is the Normalized Cut [14]:

$$NCut(A, B) = \frac{cut(A, B)}{assoc(A, X)} + \frac{cut(B, A)}{assoc(B, X)} \quad (2)$$

Where

$$assoc(S, X) = \sum_{u_i \in S, v_j \in X} W_{i,j} \quad (3)$$

is the total connection from all vertices in S to all vertices in the graph G.

3.2 Segmentation using SNCS

SNCS is a graph based segmentation method which divides an image according to some criterion. In graph theory, a graph is represented as $G = (V, E)$, where V denotes nodes and E denotes edges connecting the nodes [15]. In image analysis, an image can be regarded as a graph, where V represents a set of all pixels as shown in Fig. 2(a), and E denotes a set of all edges connecting the pixels as shown in Fig. 2(b). G can be divided into two separate sets A and B, $A \cup B = V$, $A \cap B = \emptyset$ by simply removing edges connecting the two parts as shown in Fig. 2(c)

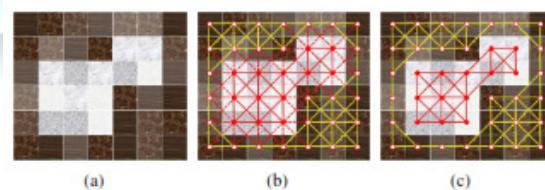


Fig. 2. SNCS segmentation process (a) Original image. (b) Bipartite graph. (c) Segmentation results

In graph theory, the above process is called a ‘cut’:

$$Cut(A, B) = \sum_{u \in A, v \in B} w(u, v) \quad (4)$$

where $C(A, B)$ represents the cut of A and B, and $w(u, v)$ denotes the weight of the edge between u and v. The optimal segmentation of a graph is the one that minimizes $C(A, B)$,



which is called the ‘minimum cut’ of a graph [16]. However, the minimum cut criteria tends to cut small sets of isolated nodes in the graph [16]. To avoid this unnatural bias for segmentation, the SNCS criteria is defined following [1]:

$$N(A, B) = \frac{C(A, B)}{O(A, V)} + \frac{C(A, B)}{O(B, V)} \quad (5)$$

where $O(A, V)$ and $O(B, V)$ are the total weights of edges connecting pixels in A and B to all pixels in the graph, respectively. Under this definition, because $C(A, B)$ is a large percentage of both $O(A, V)$ and $O(B, V)$ for the small isolated nodes, the cutting bias problem is effectively solved. For a given graph $G = (V, E)$, the optimal partition is computed by [1]

$$\min_x N(x) = \min_y \frac{[y^T(P-Q)y]}{y^T P y} \quad (6)$$

$$(P - Q)y = \lambda P y \quad (7)$$

$$P^{-1/2}(P - Q)P^{-1/2}z = \lambda z \quad (8)$$

where x is an indicator vector, y is a non-zero vector, P denotes the total connections of the node, Q denotes the weight matrix of the connections, λ is the eigenvalue and $z = P^{1/2}y$. Therefore, computing the optimal partition can be regarded as a problem of seeking the eigenvectors with the eigen values of Eq. (8).

3.3. Initialization of SNCS:

When segmenting an image, the SNCS algorithm is initialized by an unsupervised clustering process, consisting of the following five steps [1]: (I) Given a weighted graph and a set of features to measure similarity, define the weight on each edge and summarize the information into Q and P . The weight on the edge should reflect the similarity between each two pixels. For a grayscale image of a cell, the brightness value of the pixels K and their spatial location R to define the graph edge weight connecting two nodes i and j are used as [1]:

$$w_{ij} = \frac{\|K_i - K_j\|_2^2}{\sigma_1^2} * \begin{cases} e^{\frac{\|R_i - R_j\|_2^2}{\sigma_2^2}} & \text{if } \|R_i - R_j\|_2^2 < r \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

are empirically set constants in an unsupervised NC system, and r is a predefined threshold. (II) Solve $(P - Q)x = \lambda P x$ for eigenvectors with the smallest eigenvalues. (III) Use the eigenvector with the second smallest eigenvalue to bipartition the image by finding the splitting point, so that $N(A, B)$ is minimised. (IV) Decide whether the current segmentation should be subdivided by checking the stability of the cut according to an empirical value σ_3 which is specified by Eq. (10). (V) Repartition the segmented parts if it is necessary.

$$\sigma_3 = \min(H(1)) / \max(H(1)) \quad (10)$$

where H represents the histogram of the eigenvector values and σ_3 is the ratio between the minimum and maximum values.

3.4. Training and Optimization of SNCS: In the above initialization step, SNCS relies on an unsupervised segmentation process, where the parameters σ_1 and σ_2 in Eq. (9), and σ_3 in step (IV) are predefined empirically. There tuning is key to achieve high-quality partitions. Thus, the SNCS approach optimizes the parameters using training images. Although meticulously manual annotation is a difficult and time-consuming work, we only use training images which are loosely annotated, so the total effort involved in the training process is reasonable (each cell costs about one minute to segment). Noticing that the proposed approach is an improved NC which is implemented in semi-supervised fashion rather than in a fully supervised fashion, the annotation work is relatively simple. First, training images are manually annotated by biologists according to their professional knowledge. Then, the parameters are optimised using the training images in a supervised learning framework as follows: First, we define an object function as follows

$$S(\hat{A}, \hat{B}) = \operatorname{argmax}_{A, B} [-N(A, B) + M(A, B)] \quad (11)$$

where \hat{A} and \hat{B} are the corresponding optimal pixel sets for A and B , $S(\hat{A}, \hat{B})$ denotes the final segmentation result of the image, $N(A, B)$ is the SNCS criteria in Eq. (2), and $M(A, B)$ is the similarity between A and B , which is defined as the Euclidean distance between their pixel brightness values. Second, we initialize the parameters σ_1 , σ_2 and σ_3 with priori knowledge. Thirdly, we implement the SNCS segmentation over all the training images using the initial parameters. Finally, we calculate the object function $S(\hat{A}, \hat{B})$ with respect to parameters, where a global convergence approach is used to guarantee that iterations converge to an optimal solution. If $S(\hat{A}, \hat{B})$ is not convergent, then the parameters returns to the previous step. We execute this process recursively until $S(\hat{A}, \hat{B})$ gets to a maximal value.

3.5 Hessian Matrix Based Approach

Hessian Matrix based approaches have been widely used for vessel segmentation purposes in various cases. This approach finds the tubular structures in an image. It uses the cylindrical structure of the vessels and segments them with a line enhancement filter. Hessian matrix consists of the second order gradients of the Image. In this work the eigenvalue orientation of this matrix is the basis for the vessellness filter

$$H = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial x \partial z} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} & \frac{\partial^2 I}{\partial y \partial z} \\ \frac{\partial^2 I}{\partial x \partial z} & \frac{\partial^2 I}{\partial z \partial y} & \frac{\partial^2 I}{\partial z^2} \end{bmatrix} \quad (12)$$

Where I refers to the image and the values inside the matrix are second order gradients of the three dimensional image.



This process is repeated for each point to build a different matrix for each point with different scales. Using these values a vesselness value can be calculated according to Frangis article [8]

$$R_A = \frac{|\lambda_1|}{|\lambda_2|} \quad (13)$$

$$R_B = \frac{|\lambda_1|}{\sqrt{|\lambda_2 \lambda_3|}} \quad (14)$$

$$S = \|H\|_F = \sqrt{\sum_{i<3} \lambda_i^2} \quad (15)$$

$$V = \begin{cases} 0, \lambda_2 > 0 \text{ or } \lambda_3 > 0 \\ \left(1 - \frac{R_A}{e^{2z^2}}\right) e^{\frac{R_B}{2}} \left(1 - \frac{s^2}{e^{2z^2}}\right) \text{ Otherwise} \end{cases} \quad (16)$$

In equations (13,14,15) α , β and c values represents the weights RA, RB and S values calculated from different eigen values of the hessian matrix. Equation (16) gives a vesselness normalized value V for each point on the data. This vesselness value is mapped onto a [0-255] intensity interval using a log transformation, which highlights lower values. This result gives a probabilistic vesselness output for each point. The vesselness measure in Equation (16) is analyzed at different scales and the response of the filter will be maximum at a scale that approximately matches the size of the vessel is chosen as the appropriate vesselness value.

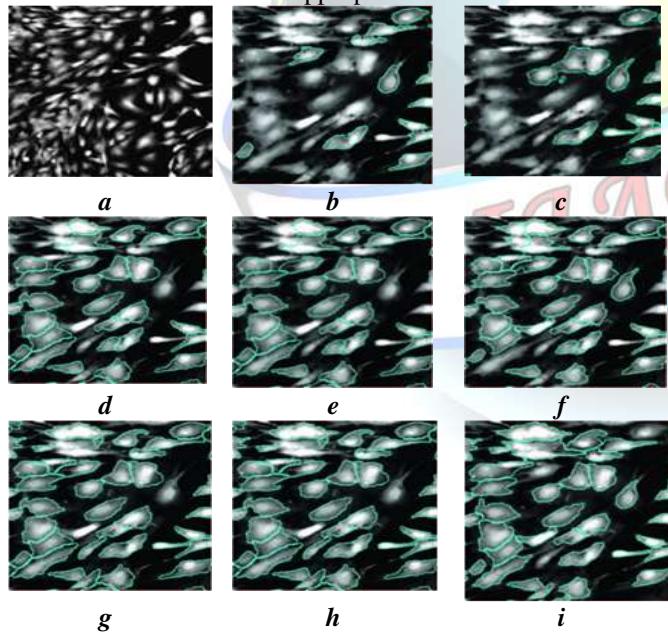


Fig 3. Outline results of eight different segment method. a) Original stem cell image. (b) Sobel Edge Detection. (c) Watershed. (d) MRF. (e) Contour Detection (f) Region Growing. (g) NC. (h) SNCS (i)Hessian approach.

4. EXPERIMENTAL RESULTS

4.1. Experimental Setting:

81 microscopic images where used for test, including approximately 6000 cells. Additional 81 corresponding nucleus images are used to count the actual segments of the image. In order to implement the proposed MHB approach, 28 ground truth images were used including about 1000 cells to train the system for optimizing the parameters. StemDB is a database designed to capture many different types of stem cell related information ranging from molecular profiling data to information about stem cell lines. StemDB was designed to specifically support pre-publication sharing of data within communities of researchers, such as those from collaborative European projects or researchers working within large-scale stem cell initiatives. In order to manage stem cell data effectively, StemDB provides fully integrated project management features and tools designed to capture personal and group collaborations.

StemDB also supports full text searching facilities that are integrated with the data access and security system. StemDB forms the data storage hub for the Innovative Medicine Initiative project, StemBANCC. Within StemBANCC StemDB will capture scientific information relating to the large scale generation and molecular characterization of 1500 iPS cell lines.

The parameters of SNCS are initialized according to the setting of NC in [1]. Furthermore, to compare the segmentation performance of SNCS, we tested six other segmentation approaches, including four classical methods: Sobel edge detection [10], watershed transformation, MRF [13] and region growing [12], and two state-of-the-art methods: contour detection [11] and NC [1] methods. Finally, two visible and four numerical comparisons are provided for evaluating the segmentation performance of MHB approach using the ground truth images. As the proposed method is a supervised method, the performance is closely related to the number and the variety of the training data. When performing the method in a large scale setting with few hundreds to thousands of cells with vastly varying morphology, a large number of ground truth images including various cells are required.

4.2. Experimental Results and Evaluation:

Some stem cell images with dense nucleus are shown in Fig. 3(a). When a segmented region (in colors) contains a nucleus (in black), it has a higher possibility to be correctly segmented. From Fig. 3(C) we can find that watershed



method has the best performance to detect the regions of different cell types. Following watershed transformation, SNCS, NC, region growing and contour detection have similar results. MRF based segmentation methods obtain the worst results. This comparison shows that MHB is robust to segment stem cells when the microscopic images show densely distributed cells. Fig. 3 shows the comparison of segmented regions (in blue outlines) and the original cells by coincident areas. If the segmentation is more accurate, the outlines are more close to the actual curved edges of the cells. From Fig. 3, we can find that SNCS and MHB obtain the best performance to describe the complicated and curved edges of cells. Sobel edge detection and contour detection have the second best performance. Lastly, watershed and region growing have similar worse results. MRF based segmentation method obtains the worst results, but the details of the cell images are most similar to the ground truth image. This comparison proves the usefulness of MHB for solving the segmentation problem for cells with curved edges. Furthermore, four numerical measures, similarity, sensitivity, specificity [17] and RA are used, to evaluate the segmentation results. Similarity indicates the overall consistency between segmented and ground truth images. It is a value in [0, 1]. If the similarity is close to one, segmentation is very similar to the ground truth. Sensitivity and specificity represent the consistency of foreground and background, respectively. RA is the ratio between obtained segmentation results and the existing number of cells, and it is defined as

$$RA = \frac{\text{Number of Segments}}{\text{Number of actual nucleus}} * 100 \quad (17)$$

The segmentation result is more accurate, if RA is close to 100%. The segmentation method is under-segmentation, if $RA < 100\%$, and over-segmentation, if $RA > 100\%$. A numerical comparison is shown in Table 1. Table 1 shows that MHB obtains the highest similarity, the third highest sensitivity, the best specificity, and the best RA. The RA of watershed algorithm is more than 100%. This leads to over-segmentation. Sobel edge detection and MRF segmentation leads to under-segmentation. The RA of MHB is closest to 100%, demonstrating that the proposed method can detect cells accurately in an image with dense cells. Furthermore, the overall result of these four measures shows that the MHB outperforms other methods. Thus, we prove that MHB outperforms traditional NC and SNCS methods due to its machine learning approach. [7] proposed a method in which the minimization is performed in a sequential manner by the fusion move algorithm that uses the QPBO min-cut algorithm. Multi-shape GCs are proven to be more beneficial than single-shape GCs. Hence, the segmentation methods are validated by calculating statistical measures. The false

positive (FP) is reduced and sensitivity and specificity improved by multiple MTANN. Fig.4 shows the comparison chart comparing the various parameters as similarity, sensitivity and specificity for the stem DB images. The comparison shows that Morphological hessian approach produces a better result in case of similarity, sensitivity and specificity.

Table 1
Numerical comparison of Segmentation results (in [%])

Methods	Similarity	Sensitivity	Specificity	RA
<i>Sobel Edge Detection</i>	89.8	97.6	60.7	59.8
<i>Watershed</i>	91.3	34.3	80.0	150.6
<i>MRF</i>	85.1	44.7	72.5	45.9
<i>Contour Detection</i>	90.4	44.0	88.3	138.8
<i>Region Growing</i>	91.4	23.2	93.2	112.7
<i>NC</i>	92.3	55.1	94.2	108.3
<i>SNCS</i>	92.4	54.3	92.0	117.8
<i>Morphological Hessian approach</i>	92.8	56.5	98.2	102.8

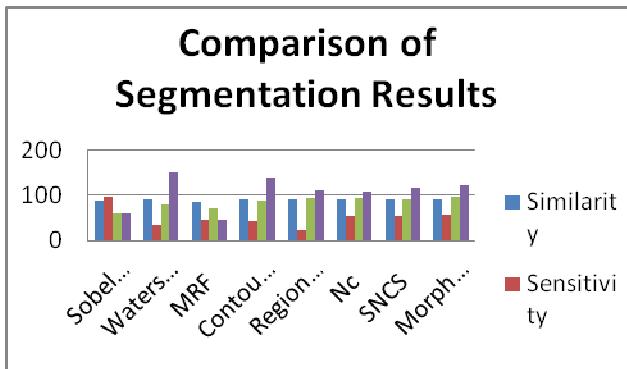


Fig. 4 Comparison of Segmentation Results

5. CONCLUSION AND FUTURE WORK

In this paper, MHB method is proposed for segmenting individual stem cells from an image, which takes advantage of the justice of clustering criterion in NC, SNCS algorithm and the robustness of supervised learning processes. The experimental results show a good performance of the proposed method. In the future, MHB approach in a clustering process is used in segmentation that results to aid taxonomic tasks in relation to stem cells [18]. Additionally, annotating ground truth images and segmenting cells in a large scale setting will be used in future.

6. REFERENCES

- [1] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [2] X. Chen, X. Zhou, and S. Wong, "Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 4, pp. 762–766, 2006.
- [3] F. Yang, M. A. Mackey, and F. Lanzini, "Cell segmentation, tracking, and mitosis detection using temporal context," in *Medical Image Computing and Computer Assisted Intervention*, 2005, pp. 302–309.
- [4] K. Jiang, Q. Liao, and S. Dai, "A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering," in *International Conference on Machine Learning and Cybernetics*, 2003, pp. 2820–2825.
- [5] S. Farnoosh and S. Zainina, "A framework for white blood cell segmentation in microscopic blood images using digital image processing," *Biological Procedures Online*, vol. 11, no. 1, pp. 196–206, 2009.
- [6] A. K. Jain, S. P. Smith, and E. Backer, "Segmentation of muscle cell pictures: a preliminary study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 232–242, 1980.
- [7] Christo Ananth, G.Gayathri, M.Majitha Barvin, N.Juki Parsana, M.Parvin Banu, "Image Segmentation by Multi-shape GC-OAAM", American Journal of Sustainable Cities and Society (AJSCS), Vol. 1, Issue 3, January 2014, pp 274-280.
- [8] M. Shen, B. Zimmer, M. Leist, and D. Merhof, "Automated image processing to quantify cell migration," in *Bildverarbeitung fuer die Medizin 2013*, 2013, pp. 152–157.
- [9] K. Fu and JK. Mui, "A survey on image segmentation," *Pattern Recognition*, vol. 13, no. 1, pp. 3–16, 1981.
- [10] I. Sobel, "History and definition of the Sobel operator," 2014.
- [11] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.
- [12] J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating colouredge extraction and seeded region growing," *IEEE Transactions On Image Processing*, vol. 10, pp. 1454–1466, 2001
- [13] Huawu Deng and David A Clausi, "Unsupervised image segmentation using a simple mrf model with a new implementation scheme," *Pattern recognition*, vol. 37, no. 12, pp. 2323–2335, 2004.
- [14] Y. T. Yang, B. Fishbain, D. S. Hochbaum, E. B. Norman, and E. Swanberg, "The supervised normalised cut method for detecting, classifying, and identifying special nuclear materials," *Informs Journal on Computing*, pp. 1–14, 2013.
- [15] J. A. Bondy and U. S. R. Murty, *Graph Theory*, Springer, Berlin, 2008
- [16] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1101–1113, 1993.
- [17] C. Li, K. Shirahama, and M. Grzegorzek, "Environmental microbiology aided by content-based image analysis," *Pattern Analysis and Applications*, vol. 19, no. 2, pp. 531–547, 2016.
- [18] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297