

Automatic cell segmentation in microscopic color images using ellipse fitting and watershed

M. Park, *Member, IEEE*, J.S. Jin, Y. Peng, P. Summons, D. Yu, Y. Cui,
S. Luo, F. Wang, L. Santos and M. Xu

Abstract— This paper presents an efficient and innovative method for the automated counting of cells in a microscopic image. The performance of watershed-based algorithms for the segmentation of clustered cells has been well demonstrated. The strength of our algorithm lies in the fact that it incorporates knowledge of color in the image. Our method uses the watershed transform with iterative shape alignment and is shown to be more accurate in retaining cell shape. We report a sensitivity of 97% and specificity of 96% when all color bands are used. Our methods could be of value to computer-based systems designed to objectively interpret microscopic images, since they provide a means for accurate cell segmentation.

I. INTRODUCTION

ANALYZING for abnormalities of cell images provides a basis for reducing deaths and morbidity from cancer through the detection of potentially cancerous cells, the provision of prompt advice and opportunities for follow-up investigation and treatment [1]. However, when conducted manually by human experts, quantitative population studies are difficult and time-consuming work. A computer-assisted microscopy system for automatic quantification is required to increase the accuracy and reduce the workload [2]. As image segmentation is the first step towards image understanding and image analysis, accurate cell segmentation is crucially important to guarantee correct results in computer-assisted microscopy. However, realization of cell segmentation methods is far from ideal as the images are often noisy and non-uniform, with objects on the image often overlapping each other or occurring in heaps [3].

There are two popular methods for cell segmentations: ellipse fitting and the watershed transform. Ellipse fitting is an attractive method for segmenting circle-like or ellipse-like objects. However, it requires robust contour pre-processing

Manuscript received February 5, 2010. This work was supported in part by the ARC LP0669645 and IntelliRAD.

Mira Park, Jesse S. Jin, Yu Peng, Peter Summons, Suhui Luo, Yue Cui, Min Xu and are with the School of Design, Communication and IT, University of Newcastle, Callaghan 2308 Australia (phone: 61-2-49854509; e-mail: Yue.Cui@studentmail.newcastle.edu.au).

W. S. Felix Wong is with the Department of Obstetrics and Gynaecology, School of Medicine, University of New South Wales, Sydney, NSW 2052, Australia.

L. Santos is with Department of Anatomical Pathology, Sydney South west Pathology service, Liverpool Hospital, Liverpool NSW 2170, Australia

to find the concave points of the contour. Even though there are many proposed improvements to the watershed transform, it requires a perfect thresholding algorithm to obtain a binary image and over-segmentation is still an issue if the objects in the binary image are not smooth enough.

This paper proposes an efficient method for image cell segmentation that uses both ellipse fitting and the watershed transform. This new method does not require the concave points for ellipse fitting and it also solves the over-segmentation problem by using the colors present in the image. Color image segmentation attracts more and more attention because color is directly and naturally attached to the regions of an image. Color images can provide more information and a color segmentation method might give better results than monochromatic segmentation methods [4]. Color segmentation raises problems related to the representation and the processing of color, such as edge-based and region-based segmentation [5], histogram analysis [6], pixel classification [7][8] and region-growing [9]. We use clustering techniques to identify homogeneous clusters in the RGB color space and then label each cluster as a different region. The homogeneity criterion is color similarity. We therefore propose an efficient method for the automated counting of cells in microscopic color images.

This paper describes the new method and presents the results of its evaluation on the microscopic images in a public database [3].

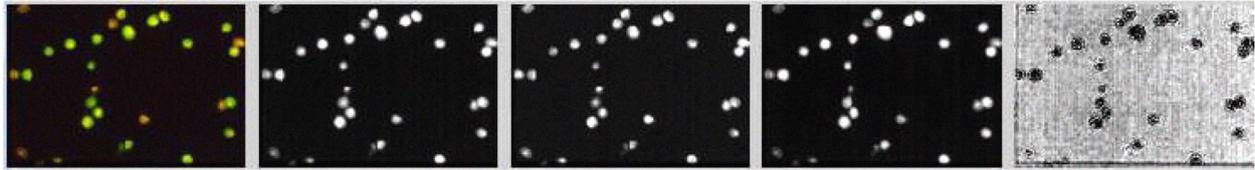
II. METHOD

A. Preprocessing

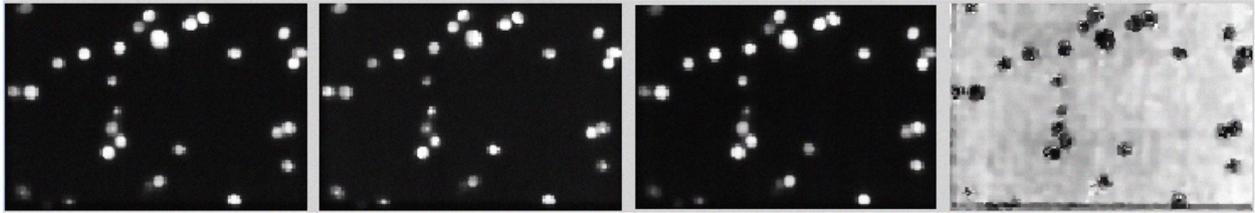
We first apply an adaptive nonlinear diffusion algorithm [10] to remove noise from the image. The nonlinear anisotropic diffusive process has the good property of eliminating noise whilst preserving the accuracy of edges. However, filtering depends on the threshold of the diffusion process and the threshold varies from image to image, and even from region to region within an image. This problem compounds with intensity distortion and contrast variation. We apply the Central Limit Theorem to automatically determine and set the threshold for the adaptive nonlinear diffusion algorithm.

B. Producing of contour segments

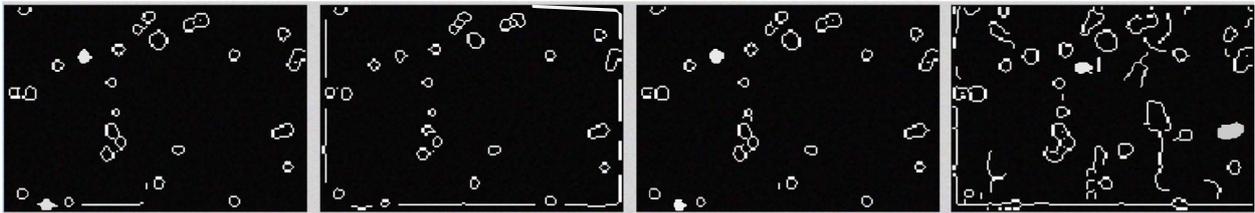
After preprocessing, the canny edge detector extracts the



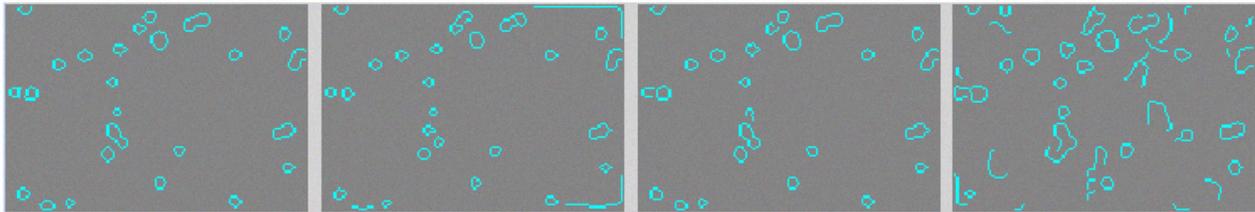
(a) From left: Original image, gray image, image in red band, image in green band and image in blue band



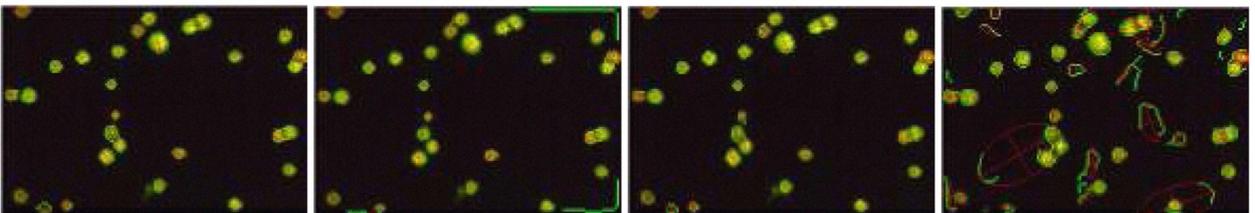
(b) Diffused image
 Gray red band green band blue band



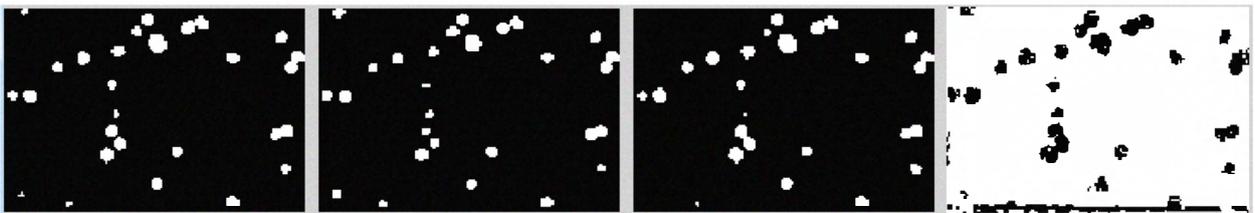
(c) Edge image
 Gray red band green band blue band



(d) Remove if the roundness of the edge segment is less than 0.5
 Gray red band green band blue band



(e) Ellipse fitting
 Gray red band green band blue band



(f) Binary image
 Gray red band green band blue band

Figure 1. Preprocessing of microscopic images of rare muscle fibres

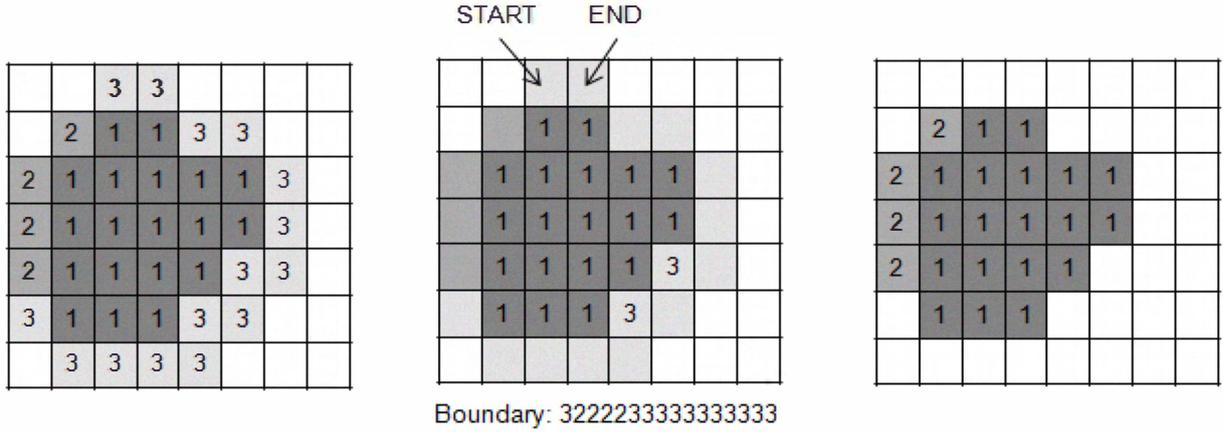


Figure 2. An example of *iterative shape alignment*

contour points are linked by an 8-neighbors' connection. The connected points are set as contour segments.

C. Ellipse fitting algorithm

The direct least square ellipse fitting method [11] is applied on the points in each contour segment obtained in step B. This tries to fit an ellipse to the points and incorporates an ellipticity constraint into the normalization factor by minimizing the algebraic distance between the best-fit ellipse and the points, subject to the constraint $4ac-b^2 = 1$. The contour segments are then classified into 3 categories: not ellipse, single cell and clustered cells. If a contour segment is not fitted by an ellipse, the contour segment is removed from the cell candidates list. If a contour segment is fitted by ellipse, the contour segment is classified as a single cell candidate or a clustered cells candidate. This classification is explained in the next section.

D. Generation of the classification rules

The system extracts features from the object (contour segment) fitted by an ellipse. The features are a major axis, a minor axis, phi, average intensity of the image object, standard deviation of intensity, eccentricity, solidity, etc. The C4.5 algorithm is used to reduce the number of features and select more relevant features. We first generate a tree using a set of data (called the *training set*). The resulting tree overfits the data as it infers more structure than is justified by the training set. This tree is then pruned to a smaller one using error-based pruning. Starting from the bottom of the tree each non-leaf sub-tree is examined. If replacement of this sub-tree with a leaf, or with its most frequently used branch, leads to lower predicted errors, then the tree is pruned accordingly. Finally, classification rules are generated from the resulting tree after the pruning process. This step classifies an ellipse into a single cell or clustered cells. The

single cell does not need further processing, but the clustered cells should be further separated into single cells.

E. Separation of the clustered cells

The clustered cells are bounded by an ellipse, hence the region of the clustered cells should be obtained. We create a binary image for each color band to obtain the regions of interest.

Creation of a binary image

Color is perceived by humans as a combination of tristimuli R (red), G (green), and B (blue) which are usually called the three primary colors [12]. We use these color spaces separately to increase the accuracy of the segmentation, since we assume that each individual channel was conditionally independent. Otsu's threshold method is commonly used for the creation of a binary image, but we use the K-means algorithm as it is the most common clustering algorithm.

Our K-means algorithm takes a 2 dimensional image as input with $k = 2$. The steps in the algorithm are as follows:

Step 1. Compute the intensity distribution (also called the *histogram*) of the image intensities.

Step 2. Initialize the means with k intensities.

Step 3. Cluster the points based on the distance of their intensities from the mean intensities.

Generate a new partition by assigning each value to the nearest cluster center.

$$c_i := \arg \min_j \|x_i - \mu_j\|^2 \quad (1)$$

Step 4. Compute the new mean for each of the clusters.

$$\mu_i := \frac{\sum_{i=1}^m 1\{c_i = j\} x_i}{\sum_{i=1}^m 1\{c_i = j\}} \quad (2)$$

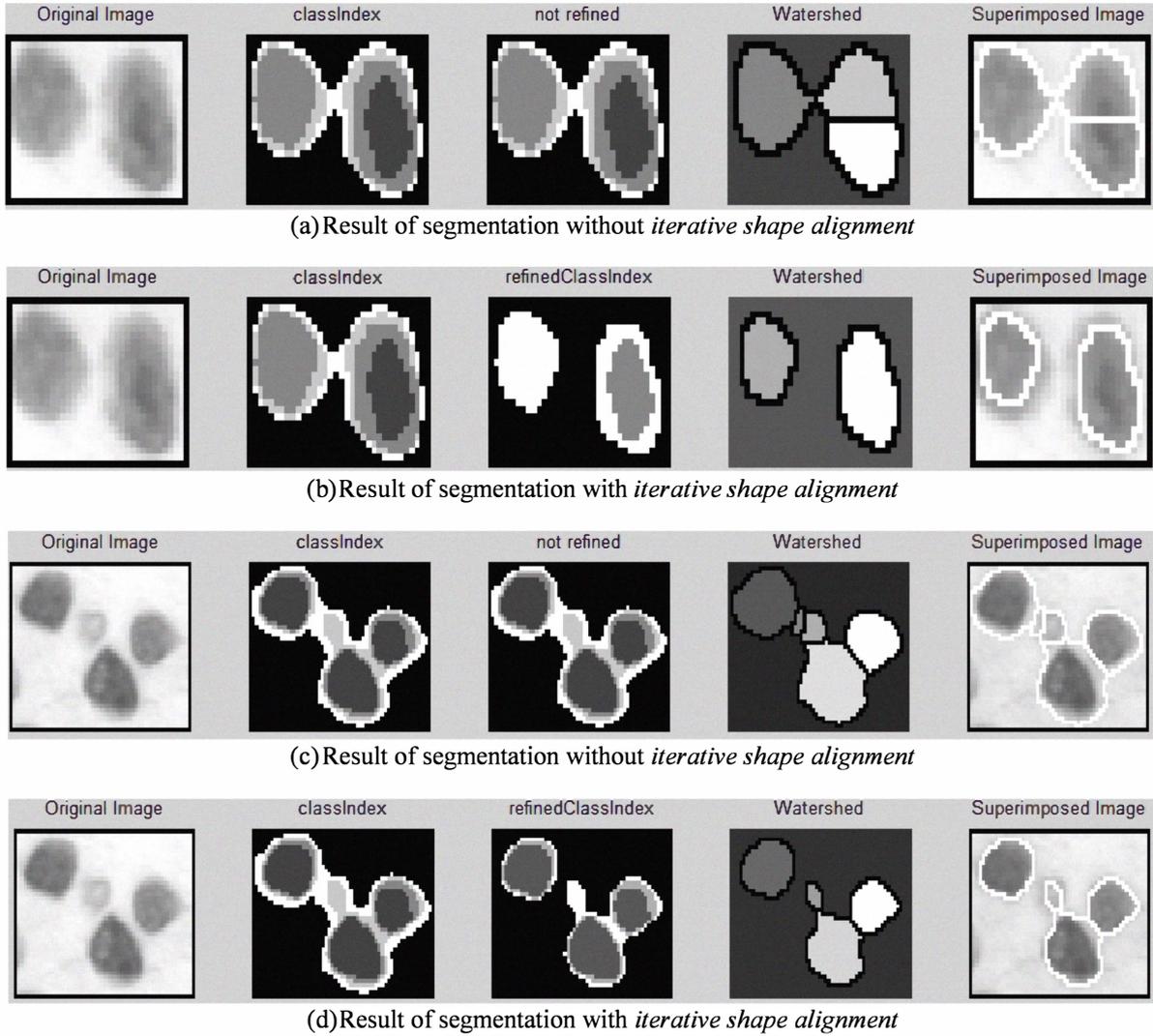


Figure 3. Results of clustered cells segmented by Watershed with, and without, *iterative shape alignment*

where k is a parameter of the algorithm (the number of clusters to be found), i iterates over the all the intensities, j iterates over all the means and μ_i are the mean intensities.

Step 5. If all cluster means do not change as the iteration step increases, the algorithm has converged and the procedure is terminated. Otherwise go to **Step 3**.

Each pixel is mapped to the one point in the feature space according to its features (the pixels' values). The feature points that having similar features will group into the same cluster.

The watershed transform [13] can be used to separate the clustered cells. The simple concept of watershed is that the troughs are filled with water in order to find the watershed ridge lines. However, the over-segmentation problem can

occur when we apply the watershed transform. We therefore align the object in order to reduce this problem [8].

Iterative shape alignment

A *classIndex* is created for the clustered cells using reducing colors before the *iterative shape alignment* starts. We fix the number of colors to four. We then iteratively eliminate undesired ones around the main segments. However, this elimination should not affect the real image objects. The final image segment is called the core object I_n . The steps involved are:

Step 1. Find undesired class pixels from the boundary of segment I_n .

$$B_n = \text{boundary}(I_n) \quad I_n \in \text{classIndex}, \text{ where } I_n \text{ is } n^{\text{th}} \text{ segment.}$$

$$B_n \subset \{1, 2, 3, 4\}$$

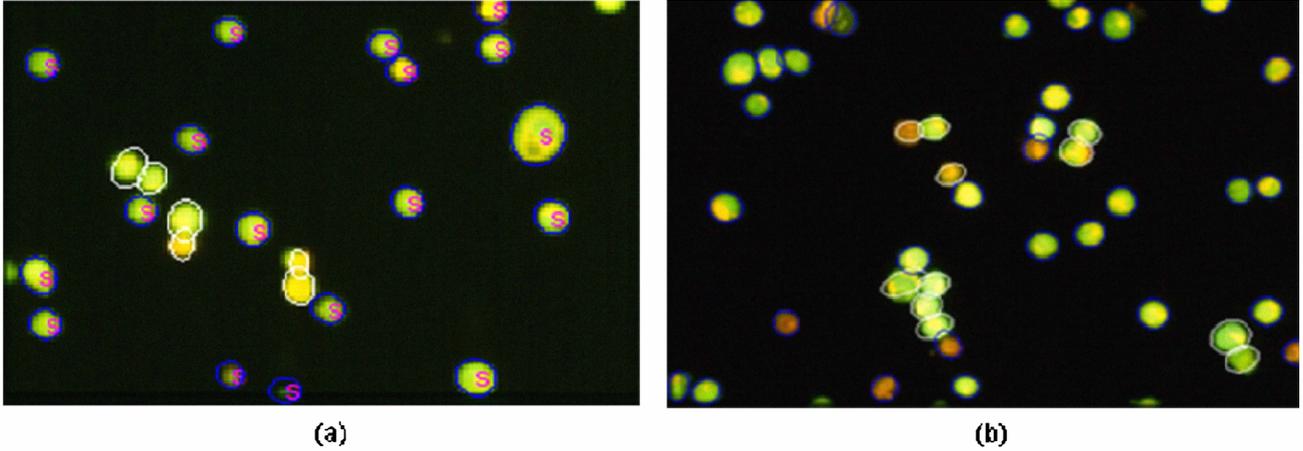


Figure 4. Examples of the final classified cells using ellipse fitting and Watershed with iterative shape alignment

$$C = \arg \max(\text{count}(B_n)) \text{ class } 1..4$$

Step 2. Remove: The pixels with value C are removed from segment I_n .

$$I_n' \leftarrow I_n(I_n \neq C) = 0;$$

Step 3. The area of I_n' is compared to the area of the original I_n segment. If the area is less than 80% of the original segment area then finish the process. Otherwise, set $I_n = I_n'$ and repeat **Step 1** and **Step 2**.

An example of the process of iterative shape alignment is shown in Figure 2. Figure 3 shows a comparison of the normal watershed transform applied to the I_n' segments, with and without iterative shape alignment.

III. EXPERIMENTAL RESULTS

Our method was evaluated on the microscopic images in a public database [1]. The database included 98 microscopic images of oculopharyngeal muscular dystrophy (OPMD). OPMD is an adult-onset autosomal dominant disease that affects all muscles, especially those for eyelid elevation and for swallowing. Light microscopy shows vacuolisation in rare muscle fibres (Figure 1a). The cells were manually segmented, so they were used as the ground truth to evaluate the proposed algorithm. We tested our algorithm in each color space: gray image, image in red space, image in green space, image in blue space and a combination of all color spaces. Some examples of the resulting classifications are shown in Figure 4. In Figure 4(a), the 's' denotes a single cell which was segmented from the first segmentation phase and the white outlined cells are clustered cells, so they were segmented by the watershed transform. According to the evaluation process, the sensitivity is 97% and the specificity is 96% when all color bands are used (see Table 1).

Table 1. Sensitivity and Specificity

	gray	red	green	blue	combine
Sensitivity	0.66	0.98	0.96	0.95	0.97
Specificity	0.95	0.81	0.91	0.80	0.95

IV. CONCLUSIONS

We have outlined a proposed algorithm that is a new and an innovative approach to the problem of cell segmentation in microscopic color images. The strength of the algorithm lies in the fact that it incorporates knowledge of the color in the image. The performance of watershed-based algorithms for the segmentation of clustered cells has been demonstrated. The watershed transform with iterative shape alignment has been shown to be more accurate in retaining cell shape. Our method could be of value to computer-based systems designed to objectively interpret microscopic images, since it provides a means for accurate cell segmentation.

REFERENCES

- [1] Chan, S. W. K., Leung, K. S. and Wong, W. S. F., An Expert System for the Detection of Cervical Cancer Cells using Knowledge-based Image Analyzer, *Artificial Intelligence in Medicine*, 8:67-90, 1996.
- [2] S., E., Computer-Assisted image analysis of small cell lymphoma of the thyroid gland, *Computerized medical imaging and graph*, 22:479-488, 1998.
- [3] Kharna, N., H. Moghnieg, et al. (2007). Automatic segmentation of cells from microscopic imagery using ellipse detection. *IET Image Process* 1(1): 39-47.
- [4] Ohta, Y. I., Kanade, T. and Sakai, T., Color information for regions segmentation, *Computer graph image process*, 13:222-241, 1980.
- [5] Riseman, E. M. and Arbib, M. A., Computational techniques in the visual segmentation of static scenes, *Computer Vision Graphics Image Process*, 6:221-276, 1977.

- [6] Celenk, M., A color clustering technique for image segmentation, *Computer Vision Graphics Image Process*, 52:145-170, 1990.
- [7] Littman, E. and Ritter, E., Color image segmentation: A comparison of neural and statistical methods, *IEEE Transactions on Neural Networks*, 8:175-185, 1997.
- [8] Park M., Jin S.J., Xu M., Felix S.W, Luo S., Cui Y., Microscopic Image Segmentation Based on Color Pixels Classification, , The International ACM conference on Internet Multimedia Computing and Service, 61-77, 2009
- [9] Tremeau, A. and Borel, N., A region growing and merging algorithm to color segmentation, *Pattern Recognition*, 30(7):1191-1203, 1998.
- [10] Jin, J. S., Wang, Y. and Hiller, J., An Adaptive Nonlinear Diffusion Algorithm for Filtering Medical Images, *IEEE Transactions on Information Technology in Biomedicine*, 4(4):298-305, 2000.
- [11] Fitzgibbon A., P.M., Fisher R.B., 1999. Direct Least Square Fitting of Ellipses. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 21(5): 476-480. [doi:10.1109/34.
- [12] Cheng, H. D., Jiang, X. H., Sun, Y. and Wang, J., Color image segmentation: advances and prospects, *Pattern Recognition*, 34:2259-2281, 2001.
- [13] Roerdink, J. B. T. M. and Meijster, A., The watershed transform: algorithms and parallelization strategies, , *Fundamental Informatic*, 41:187-228, 2000