

Automated Dynamic Cellular Analysis in Time-Lapse Microscopy

Shuntaro Aotake¹, Chamidu Atupelage², Zicong Zhang¹, Kota Aoki², Hiroshi Nagahashi², Daisuke Kiga¹

¹Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Japan

²Imaging Science and Engineering Laboratory, Tokyo Institute of Technology, Yokohama, Japan

Email: aotake.s.aa@m.titech.ac.jp, atupelage.s.aa@m.titech.ac.jp, z_zhang@sb.dis.titech.ac.jp, aoki.k.af@m.titech.ac.jp, nagahashi.h.aa@m.titech.ac.jp, kiga@dis.titech.ac.jp

Received 30 December 2015; accepted 10 March 2016; published 17 March 2016

Abstract

Analysis of cellular behavior is significant for studying cell cycle and detecting anti-cancer drugs. It is a very difficult task for image processing to isolate individual cells in confocal microscopic images of non-stained live cell cultures. Because these images do not have adequate textural variations. Manual cell segmentation requires massive labor and is a time consuming process. This paper describes an automated cell segmentation method for localizing the cells of Chinese hamster ovary cell culture. Several kinds of high-dimensional feature descriptors, K-means clustering method and Chan-Vese model-based level set are used to extract the cellular regions. The region extracted are used to classify phases in cell cycle. The segmentation results were experimentally assessed. As a result, the proposed method proved to be significant for cell isolation. In the evaluation experiments, we constructed a database of Chinese Hamster Ovary Cell's microscopic images which includes various photographing environments under the guidance of a biologist.

Keywords

High Dimension Feature Analysis, Microscopic Cell Image, Cell Division Cycle Identification, Active Contour Model, K-Means Clustering

1. Introduction

Cell phase detection is important in stem cell researches and drug planning. In general, biologists manually segment the cells. This process is time consuming and sometimes subjective. Therefore, high throughput cell segmentation is significant for assessing cell phases.

The approach of automated cell region extraction, often uses a property that there is a large intensity difference between cell regions and background ones, and separates them by a global thresholding [1], or segments them by using Otsu's method [2] [3]. However, the data used in this research has extremely slight intensity difference between the cell areas and the background, so those methods are not effective.

In the approach of automatic identification of cell division cycle, it is general to stain specimens for making some organs emit light or visualizing particular organs. Then several image analysis techniques are applied on

these images for their identification [4]-[7]. However, there is a problem when cells are stained, they are destroyed and their behavior would change. Also, another problem occurs that there exist cells that cannot be stained. Therefore, in this paper, we propose a high-accuracy and high-stability cell region extraction method, and also investigate how to identify cell division cycle of unstained cells to solve these problems.

This paper is organized as follows. Section 2 explains our proposed system composed of cell region extraction method and cell division cycle identification method. Section 3 describes and discuss about the experimental results. Finally, we conclude our paper in section 4.

2. Proposed Method

The proposed research is carried out in three major steps; 1) cell segmentation, (2) cell region division and 3) cell phase classification. The flow of proposed cell region classification method is shown in **Figure 1**. In the cell segmentation, we utilized a filter bank to obtain high-dimensional features of each pixel. In the cell region division, cell-concentrated areas are divided into individual cell areas by applying a level set method to get detail areas with higher accuracy. In the cell phase classification, we compute a multi-dimensional feature vector for each segmented region. Subsequently, we classify the phases by using a Random Forest classifier system.

2.1. Cell and Background Region Classification

First we utilize a filter bank to obtain high-dimensional feature vector for each pixel. Next, we apply k-means clustering to all pixels mapped in a high-dimensional feature space. After clustering in the feature space, the class that has the largest area is defined as a background. For other classes, we execute a post-processing composed of small region exclusion and morphological closing, and define the resultant image as cell regions.

In this research we examined the filter banks shown in **Table 1** in the experiments which are Standard Filter + Entropy Filter, Leung-Malik Filter Bank [8], Schmid Filter Bank [9], Selected Schmid Filter Bank, and Combination of Standard Filter, Entropy Filter and Selected Schmid Filter Bank. S-S filter is a subset of S filter that we selected from the Schmid filter bank so that we can obtain more features of the cells. Each of S-S filter set is represented by pair of parameters $(\sigma, \tau) = (4, 1), (6, 1), (8, 1), (10, 1), (10, 3)$. Fusion is a filter set that combined Std + Ent and S-S filter for obtaining extended features.

2.2. Cell Region Division

As for the cell region division, we apply the Chan-Vese model-based level set method [10] [11] to each cell region obtained in section 2.1 (hereafter we write it as a classified cell region). The specific processing operated is as follows: First, execute histogram-based image normalization to optimize the level set method and extract

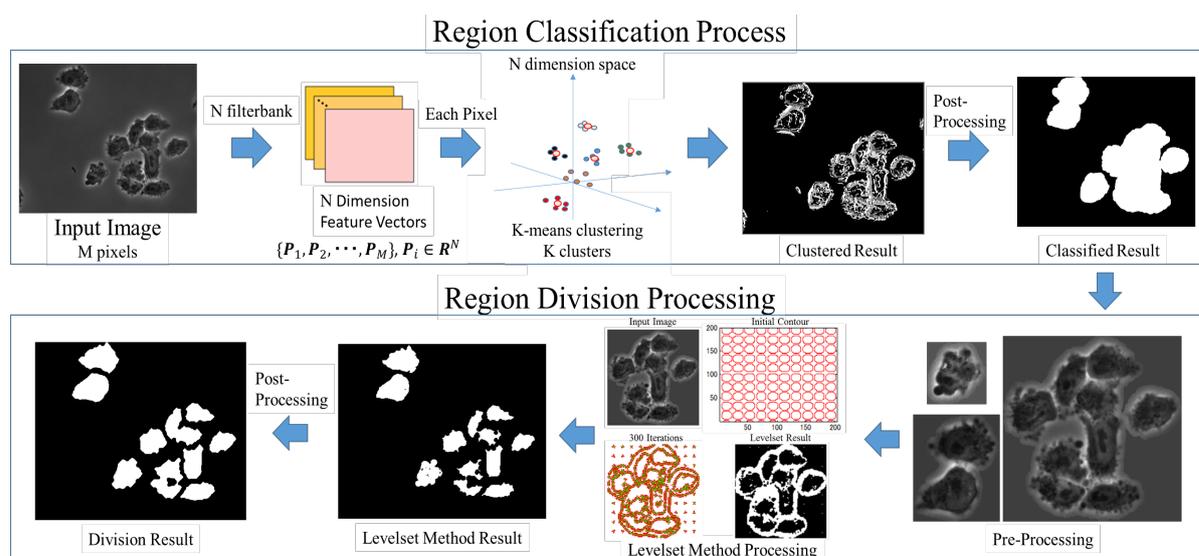


Figure 1. Flowchart of our proposed segmentation method.

classified cell regions. Then perform the level set on each classified cell region. For the initial contour, we placed small circles on grid lines in order to converge with less iteration and to gain higher accuracy. As for post-processing, we applied morphological processing and region detection by considering intensity. In this paper, we represent this method as Fusion + L.

2.3. Cell Phase Classification

The cell cycle has four phases, G1, S, G2 and M. Each phase has individual different tasks, and some of them show visible changes of the texture, but others could not be seen their differences. In this method, not only the texture but also biological knowledge are used for analyzing the cell features.

We extracted the nine features shown in **Table 2** from segmented cell regions. Each cell is characterized by a nine-dimensional feature vector. Random Forest is used to classify the cell phases.

3. Result and Discussion

3.1. Data Acquisition

For the evaluation of segmentation, we constructed a database that gives ground truth of cell regions by randomly choosing 50 images that include 450 Chinese hamster ovary cells (CHO-K1). These ground truth data were created manually under the guidance of a biologist. Original time-lapse images were captured with 40x magnification and recorded as 12 bit TIF format file (1280×1080 pixels), and ground truth data were recorded as BMP format (1280×1080 pixels) file. Test data are composed of different kinds of images taken in various recording conditions. Total number of files is 100 files, and the total file size is about 134 Mbytes. An example of snapshot image is shown in **Figure 2**.

On the other hand, for the cell phase classification evaluation, we used sequences of CHO-K1 (Chinese Hamster Ovary Cell) microscopic images of 11 hours long (1000 frames, 1.5 ~ 2 fps). These images were captured with 40x magnification and recorded in 12 bit TIF format (1280×1080 pixels). There are three cells in the video and each cell changes from G2 to G1 phase through M phase. As for training images, we adopted 2000 images that contain only 2 cells because of the simplification of cell segmentation. Details of the images are; 686 images for G2 phase, 70 images for M phase, and 1243 images for G1 phase, respectively. For simplicity of classification, we focused on just one-cell containing images and used 1000 test images. Details of the images are; 635 images for G2 phase, 27 images for M phase and 338 images for G1 phase. **Figure 3** shows a frame of time-laps data.

Table 1. Filter banks.

Name	Formal Name	Filter size (pixel)	Cluster Number
Std + Ent	Standard Division Filter and Entropy Filter	13×13 (Std) 29×29 (Ent)	4
LM Filters	Leung-Malik Filter Bank	13×13	7
S Filters	Schimid Filter Bank	25×25	4
S-S Filters	Selected Schmid Filter Bank	25×25	4
Fusion	Combination of Std + Ent and S-S Filters	-	6

Table 2. Features extracted from a cell for phase classification.

(i)	(ii)	(iii)	(iv)	(v)
Area	Area Difference	Moved Distance	Intensity Average	Intensity Variance
(vi)	(vii)	(viii)	(ix)	
Perimeter	Eccentricity	Normalized Area	Normalized Area Difference	

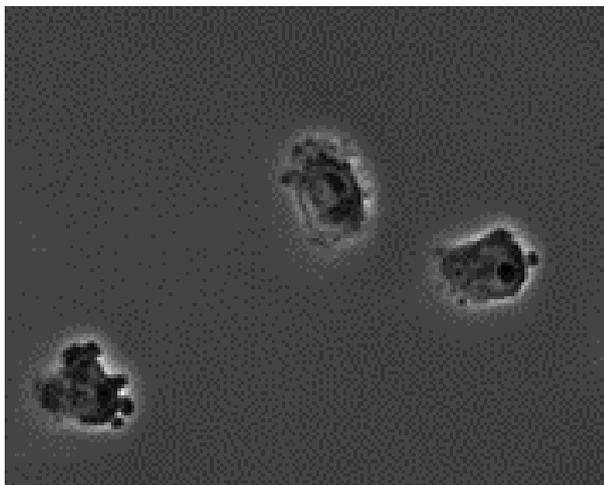


Figure 2. An image picked up from database.

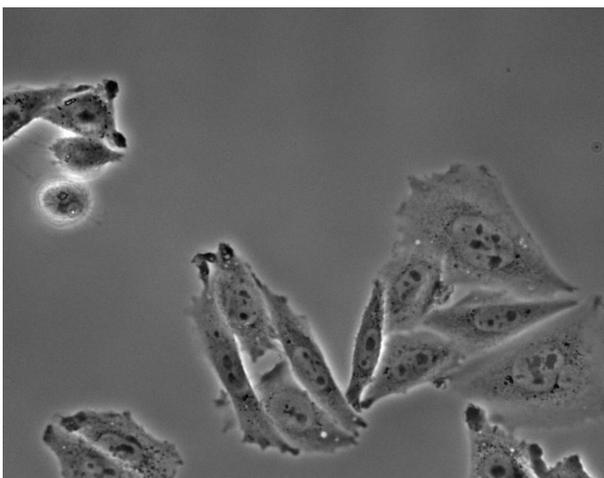


Figure 3. A frame of time-laps data.

3.2. Implementation

In this paper, we compared the segmentation performances of different methods. The number of clusters was decided according to the result of performance evaluation and the highest evaluation result was adopted. The properties of each filter and cluster numbers are shown in **Table 1**.

The post-processing applied after the clustering process are as follows: Firstly apply morphological closing operation by circular structural element with 9 pixel diameter, and eliminate small regions whose sizes are less than 5000 pixels. Then, after reversing black-and-white, again eliminate regions with less than 15,000 pixels. Finally, after re-reversing black-and-white, remained areas were extracted as cell regions.

In the process of cell region division, the following parameters are used: Intensity normalization for histogram optimization was done in the range of 540 ~ 1500. On level set method, the maximum iteration number is set to 300 and the parameter μ to 0.01. On morphological operation, we applied circular structural element with the following radius (R) on each operation: The opening operation applied on level set result ($R = 2$), the erode operation applied after the opening operation ($R = 20$), the dilate operation applied after the small region eliminating operation ($R = 2$) and the erode operation applied at last ($R = 10$).

As for cell phase classification evaluation, we evaluated all features shown in **Table 2**. Also we evaluated all combinations of these nine features which are totally 511 sets and picked up the set that gained the best average classification rate. That feature set is composed of (i), (iii), (iv), (v), (vi), and (ix). We represent it as a new feature (x).

3.3. Evaluation Index

In order to evaluate the segmentation performances of target methods, we manually generated a ground-truth binary image for each frame of some videos in the database. The binary image gives a mask pattern for extracting cellular regions from its corresponding source image. The segmented regions obtained by each method can be categorized into 4 types of semantics, that is, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). By using these 4 semantics, we used recall, precision and F-measure as evaluation indices.

Also, in order to evaluate the segmentation accuracy when a cell is placed in high dense area, we created a new evaluation index named Accuracy of Segmented Number of Cells (ASNC). It is shown by Equation (1). SNUM represents the number of cells segmented, and the cell number of ground-truth is represented as GNUM.

$$\begin{aligned} & \text{Accuracy of Segmented Number of Cells (ASNC)} \\ & = \begin{cases} \left(\frac{SNUM}{GNUM} \right) \times 100 [\%] & (SNUM \leq GNUM) \\ \frac{GNUM - (SNUM - GNUM)}{GNUM} \times 100 [\%] & (SNUM > GNUM) \end{cases} \quad (1) \end{aligned}$$

3.4. Segmentation Result on Microscopy Cell Image

In our cell segmentation method, we focus on the acquisition of cell regions without losing as small as possible in order to pass appropriate data for cell region division method. So we especially paid attention to F-measure and recall in the evaluation.

Figure 4 shows the recall, precision and F-measure of each segmentation method. Focusing on the F-measure, the best result has become 0.736822 in case of using the S-S filter. Subsequently, both Std + Ent and Fusion can gain 0.717994 and 0.7112275, respectively. On the other hand, focusing on Recall, the best result has become 0.964766 in case of using S filter. Subsequently, superior results are 0.963716 of S-S filter and 0.9137323 of Fusion. The result by Std + Ent became noticeably low and was 0.830624.

3.5. Division Segmentation Result on Microscopy Cell Image

Figure 5 shows the accuracy of segmented number of the cell. Among the region classification methods, Std + Ent was the best result of 48.3%. Subsequently, both Fusion and S-S filter can gain 45.5% and 40.9%, respectively. Fusion + L gained overwhelmingly high result which is 65.8%. This shows that division processing is working effectively. As we can see on **Figure 5**, Fusion + L gained the highest result in F-measure which is 0.763. This shows that the method also extracts cell regions precisely.

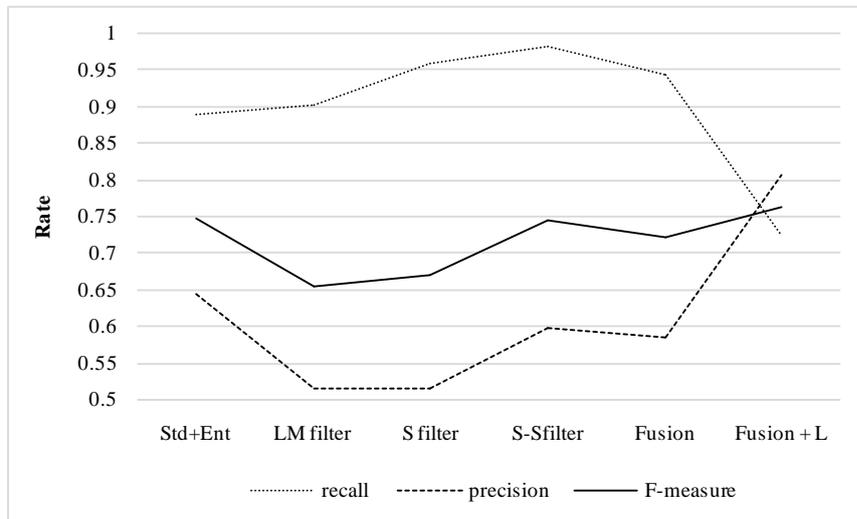


Figure 4. Result of recall, precision and F-measure.

3.6. Classification Result on Microscopy Cell Video

From the results of **Figure 4** and **Figure 5**, S-S filter has shown the best recall that might be able to extract the most significant information from the data, but from the stability point of view, Std + Ent has shown the best result on F-measure and ASNC. So in order to get balanced region classified result, it would be better to choose Fusion method for the segmentation.

Figure 6 shows the accuracy of classification on each feature set. Among nine features, (iv) Intensity Average has gained the best classification rate. Each G2, M, G1, and average identification accuracy were 21.25%, 85.19%, 38.21%, and 48.21%. Subsequently, both (i) Area and (vi) Perimeter has gained 48.01% and 39.13% on average identification accuracy, respectively. These two features are classifying all phases with high accuracy. Focusing on the best feature set (x), it gained 76.48% on average identification accuracy. Each G2, M, and G1 classification accuracy were 73.39%, 92.59%, and 63.46%.

If we focus on the classification result, it is clear that area, intensity and perimeter were effective features for cell phase identification. The intensity shows the texture information of the cell, furthermore the area and perimeter include the shape information of the cell. Therefore, we believe that texture and shape of cell are the effective features for analyzing the cell.

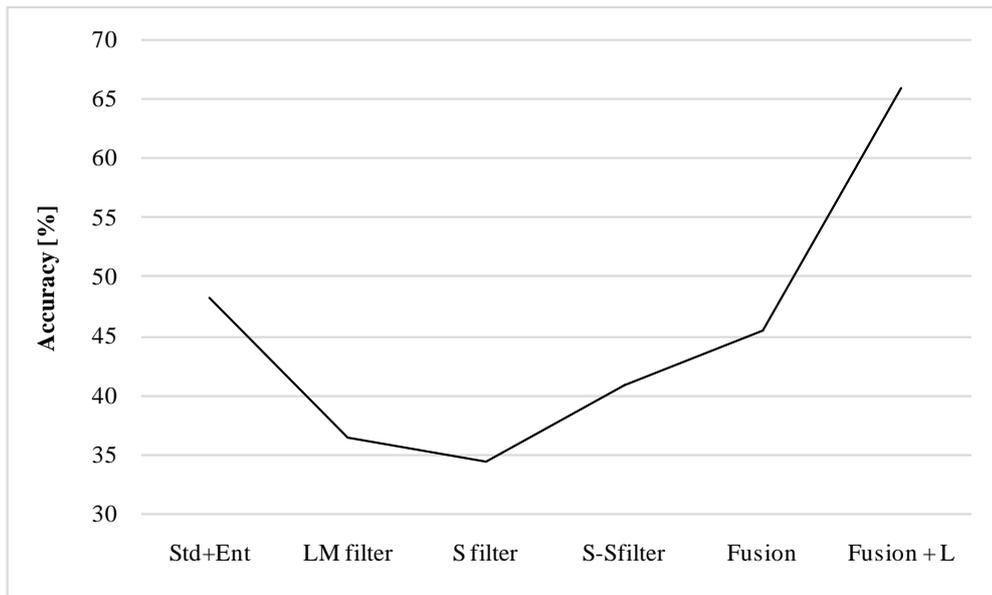


Figure 5. Accuracy of number of segmented cell.

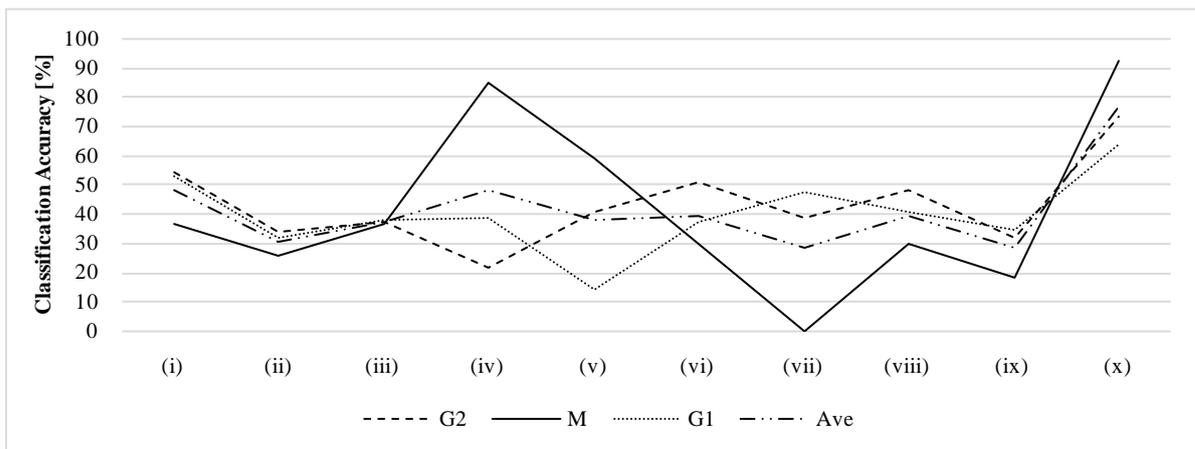


Figure 6. Accuracy of classifier on each feature and feature set.

4. Conclusions

In this paper, we proposed a high-accuracy and high-stability cell region extraction method of unstained cells in the microscope image. This was done by classifying the cell and background regions through pixel clustering in the high-dimensional space of feature descriptors constructed by filter banks. Additionally, the cell region division to individual ones by Chan-Vese model-based level set method plays an important role. We also proposed a cell division cycle phase identification method which obtains cell features (area, movement, intensity, etc.) from segmented cell regions and construct classifiers using these features to identify each cell division cycle phase. In the evaluation experiment, we focused on F-measure, recall and accuracy of segmented number of cell in order to investigate the accuracy of the cell region extraction. To perform these experiments, we constructed a database of Chinese Hamster Ovary Cell (CHO-K1) that gives the correct cell region under the guidance of a biologist.

This research is still under developing. As mentioned in section 3.6, we have recognized that texture and shape of cell must be effective feature for phase identification. Hence, we are now developing and evaluating the following two ideas: 1) By using the Bag-of-Feature method, we construct the histogram that represents the texture information and calculate the texture difference between the frames as a new feature. 2) By getting the shape of the cell, we calculate the difference of the shape between the frames as a new feature.

References

- [1] Chen, X. and Wong, S.T.C. (2005) Automated Dynamic Cellular Analysis in High Throughput Drug Screens. *IEEE International Symposium on Circuits and System*, Kobe, 23-26 May 2005, **5**, 4229-4232. <http://dx.doi.org/10.1109/ISCAS.2005.1465564>
- [2] Nobuyuki, O. (1980) An Automatic Threshold Selection Method Based on Discriminant and Least Squares Criteria. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **J63-D**, 349-356.
- [3] Xiaobo, C., Xiaobo, Z., Stephen, T.C. and Wong (2006) Automated Segmentation, Classification, and Tracking of Cancer Cell Nuclei in Time-Lapse Microscopy. *IEEE Transactions on Biomedical Engineering*, **53**, 762-766. <http://dx.doi.org/10.1109/TBME.2006.870201>
- [4] Petra, P., Horst, P. and Bernd, M. (2002) Mining Knowledge for HEP-2 Cell Image Classification. *Artificial Intelligence in Medicine*, **26**, 161-173. [http://dx.doi.org/10.1016/S0933-3657\(02\)00057-X](http://dx.doi.org/10.1016/S0933-3657(02)00057-X)
- [5] Loris, N. and Alessandra, L. (2008) A Reliable Method for Cell Phenotype Image Classification. *Artificial Intelligence in Medicine*, **43**, 87-97. <http://dx.doi.org/10.1016/j.artmed.2008.03.005>
- [6] Nicholas, A.H., Radosav, S.P., Kelly, H. and Rohan, D.T. (2007) Fast Automated Cell Phenotype Image Classification. *BMC Bioinformatics*, **8**, 110. <http://dx.doi.org/10.1186/1471-2105-8-110>
- [7] Loris, N., Alessandra, L., Lin, Y.-S., Hsu, C.-N. and Chung, C.L. (2010) Fusion of Systems for Automated Cell Phenotype Image Classification. *Expert Systems with Applications*, **37**, 1556-1562. <http://dx.doi.org/10.1016/j.eswa.2009.06.062>
- [8] Jitendra, M., Serge, B., Thomas, L. and Jianbo, S. (2001) Contour and Texture analysis for Image Segmentation. *International Journal of Contour Vision*, **43**, 7-27. <http://dx.doi.org/10.1023/A:1011174803800>
- [9] Lrystian, M. and Cordelia, S. (2001) Indexing Based on Scale Invariant Interest Points. *Computer Vision ICCV Proceedings of the Eighth IEEE International Conference*, 7-14 July 2001, **1**, 525-531. <http://dx.doi.org/10.1109/ICCV.2001.937561>
- [10] Chan, T.F. and Vese, L.A. (2001) Active Contours without Edges. *IEEE Transactions on Image Processing*, **10**, 266-277. <http://dx.doi.org/10.1109/83.902291>
- [11] Pascal, G. (2012) Chan-Vese Segmentation. *Image Processing On Line*, **2**, 214-224. <http://dx.doi.org/10.5201/ipol.2012.g-cv>