International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.3 (2014), pp.181-192
http://dx.doi.org/10.14257/ijsip.2014.7.3.15

Cell Contour Irregularity Feature Extraction Methods based on Linear Geometric Heat Flow Curve Evolution

Xu Chuanyun[1,2] and Zhang Yang[3]

[1]*College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400044, China*
[2]*College of Computer Science, Chongqing University, Chongqing 400044, China*
[3]*College of Computer and Information Science, Chongqing Normal University, Chongqing 400044, China*
thisisatest@Tom.com, 495461428@qq.com

Abstract

When the cytological smear method is used to check cervical lesions, the irregularity of the cells contour is essential for smear image interpretation and has the vital significance for research on computer aided diagnosis of cervical lesions. Aimed at measuring the regularity of cells outlines, based on the characteristic that the normal cell smears is generally circular, oval, convex polygon, we put forward a cell contour irregularity feature extraction and measurement method based on linear geometric heat flow curve evolution, in which the curve of cells outline is dealt with linear geometric heat flow curve evolution to totally convex (i.e. no curvature zero crossing) and compared with the original contour to extract indexes which can describe irregularity features of cell contour. Experimental results show that the method proposed in this paper can better describe the irregularity of cell contour and distinguish cells of different lesions.

Keywords: *Feature Extraction, Linear Geometric Heat Flow Curve Evolution, Cell Contour Irregularity*

1. Introduction

In cervical cytological smear image, normal cells are round, oval or polygonal with contour- regular cytoplasm and nucleus, but the diseased cells usually have deformity cytoplasmic and nuclear contour and convex & concave boundary [1]. The pathological study shows that the Cell contour irregularity degree has a high correlation with lesions, so the extraction of cell contour irregular features has an important significance to realize the automation interpretation of cervical cytology smear image.

The literature [2] used polygon approximation to represent the cell contour, and measured the parameters of polygonal approximation such as distortion points, end points, the mean curvature, curvature, mean edge length, edge variance to describe the irregular degree. However the cell shape is usually circular arc and few irregular linear boundaries, so the information loss about using the polygonal approximation to represent the contour shape is relatively large. The literature [3] used Gauss filtering method to smooth curves and determined the appropriate scale to measure irregularity with multiscale contour curvature model, which had computational complexity and large amount of calculation because there needed to track the extreme point of the curve in the smoothing process. The literature [4] measured the irregularity of profile by adding up the distance distribution of the focus of the closed contour curve to each point on the contour curve, which paid attention to the local

ISSN: 2005-4254 IJSIP

change curve but still tend to similar circular contour. The literature [5, 12] brought forward using local fractal dimension to measure the irregularity of profile curve, which obtained the multi-scale contour curve by using Gauss smoothing curve and computing the local fractal dimension of the smoothed curve.

Because the normal cell smears is generally circular, oval, convex polygon, we smooth contour by using linear geometric heat flow curve evolution until the profile is convex (*i.e.*, no curvature zero crossing), and describe the Irregular degree of cell contour with a series of metrics from the comparison of the cell contour after smoothing with the original cell outline.

2. Linear Geometric Heat Flow Evolution of Cell Contour Curve

In the cervical smears, the normal cells are round, oval, convex polygon, and the outline of the diseased cells usually has a concave significantly, showing irregular contour shape. Based on this, if we evolve the outline of normal and diseased cells to fully convexity evolution direction, the contour of pathological cell will have great changes as well as the contour of normal cells will be in smaller changes, so we can obtain the irregular degree of cell contour by calculating the varying size.

The linear geometric heat flow evolution of the curve [6] is to evolve the curve according to the heat flow equation, and has the following properties:

1) In the "limited time (iterations)", the curve will become fully convex, which means that there is no curvature zero crossing;
2) Meet the invariance - rotation, translation, uniform scaling;
3) The closure and connectivity of the curve would be unchanged;
4) The centroid position of closed planar curve would be unchanged.

Known from the properties above, the linear geometric heat flow evolution of the curve can smooth concave and keep the centroid position of the curve unchanged. Based on these properties, we apply the linear geometric heat flow evolution to measuring the irregularity of the contour, and give the simple computing process of the linear geometric heat flow evolution on the contour.

Supposing that $C_0(p),\ p \geq 0$ is a closed cell contour sequence, the curve C_0 should be evolved according to the following partial differential equation:

$$\frac{\partial C(p,t)}{\partial t} = \mathbf{V} = \alpha(p,t)\mathbf{T} + \beta(p,t)\mathbf{N},\ \ C(p,0) = C_0(p) \tag{1}$$

where p is a point on curve, t is the time of evolution, vector $\mathbf{V}$ is the speed of evolution, $\mathbf{T}, \mathbf{N}$ are the tangent direction of curve and the unit vector on normal direction, α, β are the tangential direction of curve and the speed on tangential direction, C_0 is the curve before evolution. Because the change of curve shape is only in relation with the normal component β of speed $\mathbf{V}$ and has nothing to do with tangential component α, at the moment of study on the geometric curve evolution problem, we only need to consider normal rate, and simplify the general way of curve evolution into:

$$\frac{\partial C}{\partial t} = \beta\mathbf{N} \tag{2}$$

If we take the two order derivative of curve as normal rate of curve evolution, that is to use Formula (3) to replace β in Formula (2) [6,9]to get the linear heat equation of curve, we obtain Formula (4).

$$\beta = \frac{\partial^2 C}{\partial p^2} \tag{3}$$

$$\frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial p^2} \tag{4}$$

According to Formula (3), we can get the motion transmissibility of every point P on curve in the direction (x, y).

$$x_t = x_{pp} \ , \ y_t = y_{pp} \tag{5}$$

Where x_t, y_t are the increments of point (x, y) at the time t. Known from this formula, if the curve has a large asperity at the point (x, y), (x_{pp}, y_{pp}) would be relatively large, which means the evolution speed of x_t, y_t would be relatively large; on the contrary, if the point (x, y) is smooth, the evolution speed would be relatively slow. Especially if the point (x, y) is the point of inflexion on a curve, (x_{pp}, y_{pp}) would be 0, and the curve would not move.

The curvature of a curve on the plane can indicate the bending degree of curve, and the curvature value of a point on the curve C is equal to the reciprocal of radius of the osculating circle of this point. The curvature $\kappa(p,t)$ of points at any moment t of evolution curve could be calculated by Formula (6).

$$\kappa(p,t) = \frac{x_p(p,t)y_{pp}(p,t) - x_{pp}(p,t)y_p(p,t)}{(x_p(p,t)^2 + y_p(p,t)^2)^{3/2}} \tag{6}$$

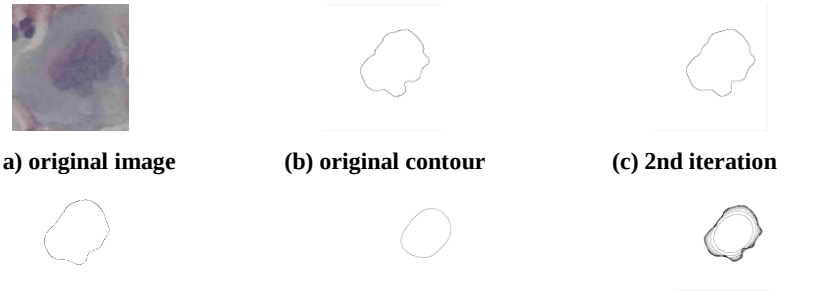
where $x_p(p,t), y_p(p,t)$ is the first order partial derivative of the point (x, y) of curve at the moment t, and $x_{pp}(p,t), y_{pp}(p,t)$ is the second order partial derivative of the point (x, y) of curve at the moment t.

The linear heat equation of curve (4) is a linear partial differential method (PDE), which can be solved by use of the Fourier transform method [9]. The solution of the linear heat equation of curve is:

$$x(p,t) = x_0(p) * g(p,t) \ , \ y(p,t) = y_0(p) * g(p,t) \tag{7}$$

Where $g(p,t)$ is a Gaussian function, and its Gauss standard deviation σ is equal to $\sqrt{2t}$.

The figure 1 shows an instance of cell contour evolution.



a) original image **(b) original contour** **(c) 2nd iteration**

((d)16th iteration **(e) 128th iteration** **(f) iteration process**

Figure 1. Instance of Cell Contour Evolution

Seen from the Figure 1, going on with the evolution, the outline of cell becomes more and more irregular and smooth. After the 128th iteration, the contour of cell is fully convexity. Also seen from the Figure 1 (f), going on with the evolution, the outline of cell is also in constant contraction. If let the evolution proceeding, the contour would be a circular (then a point) at final.

3. Contour Amplification

Just as the analysis above, the contour of pathological cells is usually not smooth and irregular, but we can use linear geometric heat flow evolution to smooth curve. By comparing the contour after linear geometric heat flow evolution with the shape of the original cell, we can get the depression degree of the cell outline which reflects the irregular degree. Because contour evolution would lead to a contraction (reference the diagram (f) of Figure 1), there is no way to compare directly the contour after linear geometric heat flow evolution with the shape of the original cell. The solution is to enlarge firstly the contour after evolution to the size of the original cell (that is the same area) and then to compare these two shapes. Because the contour is irregular which makes the calculating of the amplification ratio very difficult, we take the radius ratio of the equivalent area circle as amplification ratio.

Defining the original contour $C = \{(x_1, y_1), (x_2, y_2),..(x_L, y_L)\}$, L is the length of contour, the contour after evolution is $C' = \{(x_1', y_1'), (x_2', y_2'),..(x_L', y_L')\}$, the surrounding area of C, C' are A_ϕ, A_ϕ', and the amplification ratio is:

$$r = \sqrt{\frac{A_\phi}{A_\phi'}} \tag{8}$$

The amplification contour C' need to ensure that the centroid of contour after evolution C_a' is same with the centroid of C'. Defining the centroid of C' to be (x_c, y_c) and the amplified coordinate of the point (x', y') on C' to be (x_a', y_a'), the computing process of (x_a', y_a') is as follows:

$$\begin{cases} x_m = x' - x_c \\ y_m = y' - y_c \end{cases} \tag{9}$$

$$\begin{cases} th = \arctan(y_m, x_m) \\ rr = \sqrt{x_m^2 - y_m^2} \end{cases} \quad (10)$$

$$rra = rr \times r \quad (11)$$

$$\begin{cases} x_{ma} = rra \times \cos(th) \\ y_{ma} = rra \times \sin(th) \end{cases} \quad (12)$$

$$\begin{cases} x_a' = x_{ma} + x_c \\ y_a' = y_{ma} + y_c \end{cases} \quad (13)$$

Where $a\tan()$ arctan is function and $\cos(), \sin()$ are cosine and sine function.

4. Calculation of Irregular Contour Degree

The cell outline would evolve from concave and convex to fully convexity, and this process is called changing the cell outline from irregularity to regularity. The cell outline of fully convexity is called regular outline which means there is no curvature zero crossing point. By comparing the original outline of cells with the regular outline, we can get the following index to measure irregularity.

4.1. The Total Area Ratio of Non-Overlapping Regions

The area of non-overlapping regions is defined to be the area of non-overlapping regions between the area surrounded by regular contour and the area surrounded by original contour (*i.e.*, cytoplasm or nucleus area). The total area ratio of non-overlapping regions is defined to be the ratio of the total area of non-overlapping regions and the area surrounded by original contour. The area of non-overlapping regions is shown as Figure 2.

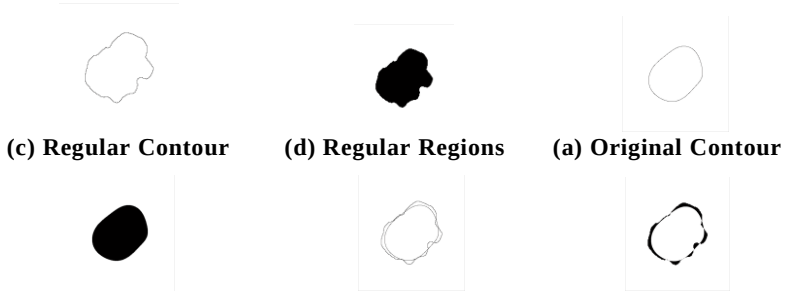


Figure 2. Non-overlapping Region

The calculation process of the total area of non-overlapping regions is shown as Formula (14).

$$\begin{aligned} \phi_{snol} &= \phi_c \oplus \phi_r \\ A_{\phi_{snol}} &= area(\phi_{snol}) \end{aligned} \quad (14)$$

where $A_{\phi_{snol}}$ is the total area of non-overlapping regions, $area$ () is the function to compute area, ϕ_c is the area surrounded by original contour, ϕ_r is the area surrounded by regular contour, and $\oplus$ is XOR operator.

The calculation formula of the total area ratio of non-overlapping regions (r_{snol}) is as follows:

$$r_{snol} = \frac{A_{\phi_{snol}}}{A_{\phi_c}} \times 100 \tag{15}$$

Where A_{ϕ_c} is the area surrounded by original contour, $A_{\phi_{snol}}$ is the area surrounded by regular contour. Larger r_{snol} , the bigger difference between regular contour and original contour, more irregular about the original contour.

According to the analysis above, the calculation method about the total area of non-overlapping regions is shown as follows:

Algorithm: Calculation about the total area of non-overlapping regions

Input: original contour $C = \{(x_1, y_1), (x_2, y_2),..(x_L, y_L)\}$ where L is the length of contour

Output: the total area of non-overlapping regions

① Calculating the area surrounded by original contour A_{ϕ_c}

② if the number of current evolution t < the maximum number of allowed evolution $tMax$

③ Calculating the evolution contour according to Formula (7)

④ $C_t^{'} = \{(x(t)_1^{'}, y(t)_1^{'}), (x(t)_2^{'}, y(t)_2^{'}),..(x(t)_L^{'}, y(t)_L^{'})\}$

⑤ Enlarging the evolution contour according to Formula (9) t o (13)

⑥ $C_{at}^{'} = \{(x_a(t)_1^{'}, y_a(t)_1^{'}), (x_a(t)_2^{'}, y_a(t)_2^{'}),..(x_a(t)_L^{'}, y_a(t)_L^{'})$

⑦ Calculating the curvature of each point $\kappa(p,t)$ of contour $C_{at}^{'}$ according to Formula (6)

⑧ if all the curvature of points $\kappa(p,t)$ >0, to execute step ⑧

⑨ else to execute step ②

⑩ Calculating A_{snol} according to Formula (14)

⑪ Calculating the total area of non-overlapping regions r_{snol} according to Formula (15)

⑫ End

The total area of non-overlapping regions r_{snol} has three fold invariance: rotation, translation, uniform scaling.

4.2. The Average Area of non-overlapping Regions

Parts of cell outline with big bump have great non-overlapping regions, and bigger the bump, larger non-overlapping regions, just as Figure 3 shown. In Figure 3, the contour in Figure (a) has the bigger bump as well as the contour in Figure (c) has the smaller bump, and

the single non-overlapping region in Figure (b) is obviously larger than Figure (d) by comparing Figure (b) with Figure (d). Therefore we put forward the average area of non-overlapping regions to measure this feature.

(a) (b) (c) (d)

Figure 3. The Relationship between Irregularity and a Single Non-Overlapping Area

The average area of non-overlapping regions (r_{anol}) is defined to be the ratio between the average values of all non-overlapping regions and the area surrounded by original contour. In order to improve the robust of features, we should ignore some small area during the calculating process. The non-overlapping region is defined as:

$$\phi_{snol} = \phi_{nol\,1} \cup \phi_{nol\,2} \cup .. \cup \phi_{noli} \cup \phi_{nolj} .. \cup \phi_{nolN} \quad \phi_{noli} \cap \phi_{nolj} \in \Phi \quad and \quad i \neq j \tag{16}$$

$$A_{\phi_{anol}} = \frac{\sum_{\phi_{noli} \in \phi_{snol} \quad and \quad area\ (\phi_{noli}) > T_{\phi}} area\ (\phi_{noli})}{N_{T_{\phi}}} \tag{17}$$

Where $area\ ()$ the function to calculate area is, T_{ϕ} is the minimum area threshold, $N_{T_{\phi}}$ is the number of non-overlapping regions which are larger than T_{ϕ} . T_{ϕ} Is usually set as 0.01~0.05 times about the total area of non-overlapping regions and the average area ratio of non-overlapping regions are insensitive with T_{ϕ} .

The calculation formula of the average area ratio of non-overlapping regions (r_{anol}) is as follows:

$$r_{anol} = \frac{A_{\phi_{anol}}}{A_{\phi_c}} \times 100 \tag{18}$$

The bigger r_{anol} , the bigger bump, the more irregular original contour, the more obvious deformity of original contour.

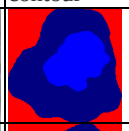
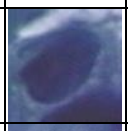
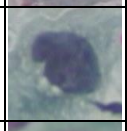
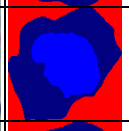
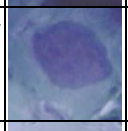
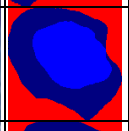
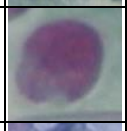
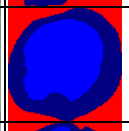
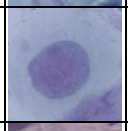
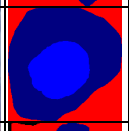
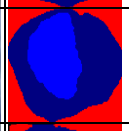
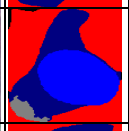
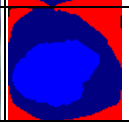
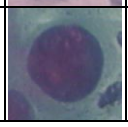
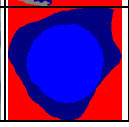
5. Experiment and Analysis

In order to validate the validity of cell contour irregularity feature extraction method based on linear geometric heat flow curve evolution, we designed three experiments.

5.1. Experiment 1: to Compare with Human Cognition

To validate the consistency between the method in this paper with human cognition, this experiment selected ten cell images from cervical cell image data set in Herlev [10] to calculate the irregular degree of its cell nucleus contour, and the results are shown in the Table 1.

Table 1. Irregular Cell Contour Experiment

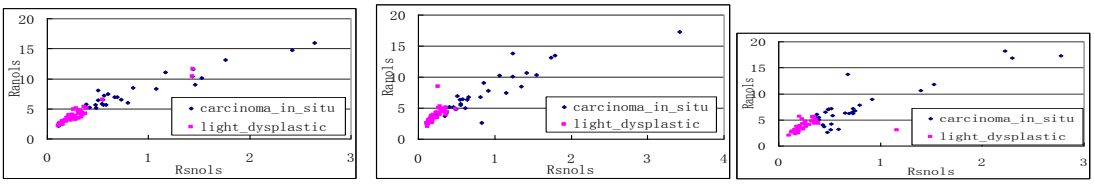
serial number	image	Nuclear contour	r_{snol}	r_{anol}	serial number	image	Nuclear contour	r_{snol}	r_{anol}
C10			9.481	1.032	C01			5.748	0.384
C05			8.882	1.238	C04			3.546	0.280
C08			7.996	0.884	C03			3.003	0.326
C02			6.600	0.600	C09			1.861	0.075
C06			6.292	0.543	C07			1.631	0.066

The Table 1 showed the computing result about the irregular degree of nuclear contours sorted by r_{snol} from small to large. Known from the table, the value of r_{snol} of nuclear contour of the cell C10 is biggest as well as the value of r_{snol} about C08 is smallest, which was got by the most regular nuclear contour of the cell C07 as well as the most irregular nuclear contour of the cell C08.

Seen from the experiment results, the metrics method about irregularity in this paper is consistent with human cognition. What needs to specially explain is the cell C08. Although the general form of cells is round which means its compactness is high (the high compactness expresses it regular), r_{snol} is also rather high due to obvious concave. This makes clear that the measuring result of irregularity by using our method do not incline to the circular outline.

5.2. Experiment 2: the Scatter Analysis Between r_{snol} and r_{anol}

This experiment respectively selected 90 cell images and 120 cell images from the class "carcinoma_in_situ" and the class "light_dysplastic" in the cervical cell image data set in Herlev [10], and randomly divided them into A, B, C three groups in which the numbers of images about carcinoma_in_situ and the images about light_dysplastic are respectively 30 and 40. In this experiment, we calculated r_{snol} and r_{anol} by using the algorithm given in this paper, and draw the scatter diagram by taking r_{snol} as x-axis and r_{anol} as y-axis whose results are shown in Figure 4.



(1) Group A (2) Group B (3) Group C

Figure 4. Scatter Diagram of r_{snol} and r_{anol}

Seen from Figure 4, the irregular degree of cell outline of cancer lesion images and mild atypical hyperplasia has a significant aggregation.

The Table 2 showed the average values of each group of r_{snol} and r_{anol} .

Table 2. Average Values of r_{snol} and r_{anol}

	carcinoma_in_situ r_{anols}	light_dysplastic r_{snols}
A	0.931	0.243
B	0.923	0.232
C	0.841	0.244
	carcinoma_in_situ r_{snols}	light_dysplastic r_{snols}
A	8.074	3.612
B	7.941	3.791
C	7.996	3.695

Seen from the Table 2, the average values of irregular degree r_{snol} & r_{anol} of cell contour in cancer lesion images are larger than the average values of irregular degree r_{snol} & r_{anol} of cell contour in mild dysplasia images, which would be In line with "more serious the severity of cells, more obvious the deformation of cell morphology" in pathology interpretation standards of cervical cytology. Also seen from the Table 2, the average values of class "carcinoma_in_situ" & class "light_dysplastic" among three groups insides each class have high consistency as well as have obvious difference between classes.

5.3. Experiment 3: to Analyze the Ability of Classification

Extracting the irregular degree of cell contour is to realize computer aided interpretation of cervical smear, which makes the classification ability of the irregular degree very important. This experiment used the same data sets with the experiment 2, and took the group A as the sample of the classification learning as well as the group B & C as the test samples. This experiment used k-Nearest Neighbor (KNN) [11] to classify cell images, and took irregularity extraction methods in the reference [4] & reference [12] as our comparison algorithm whose results were shown in the Table 3.

Table 3. Classification Accuracy

Method	r_{anol}	r_{snol}	Reference[4]	Reference [12]
Group B	92.86	89.9	83.72	86.15
Group C	94.29	88.58	80.29	87.86

Seen from the Table 3, the metrics values of irregularity about classification accuracy in this paper was better than those in the reference [4] & reference [12]. The classification ability of r_{anol} was slightly stronger than that of r_{snol}, which was because the larger concavo convex region of the cell outline would influence the regularity of cell contour and the index r_{anol} excludes effects of small convex region to show better classification ability.

6. Conclusion

In order to measure the irregularity of cells contour, we have put forward a cell contour irregularity feature extraction and measurement method based on linear geometric heat flow curve evolution. This paper firstly introduced the heat flow curve of linear geometric evolution method of cell contour, and then gave two irregular metric index: the total area ratio of non-overlapping area (r_{snol}) and the average area ratio of non-overlapping area ratio (r_{anol}) by comparing the contour after the evolution and the original contour, and finally designed three experiments to prove the effectiveness of the proposed irregular measure method. The experimental results showed that the measuring results of the proposed irregular measure method are basically consistent with human cognition, and the classification ability of proposed metrics is relatively better than the existing irregularity measurement methods.

In the future work, we would apply the extracted irregularity of cell contour in the computer aided interpretation of cervical smear and make a further optimization on the classification ability and robustness of Feature extraction algorithm in this paper.

Acknowledgements

This research is supported by the National Natural Science Foundation of China under Grant No. 61202348, the National Natural Science Foundation of China under Grant No. 61173184, the Key Science and Technology Project of Chongqing under Grant No. CSTC2012GGYYJS10027 and the Natural Science Foundation Project of Chongqing under Grant No. CSTC2012jjA1549.

References

[1] D. Solomon and R. Nayar, "The System Definition, Standard and Annotation for Cervical Cytology Bethesda Report (Second Edition)", People's Medical Publishing House, **(2009).**
[2] L. zhiyong and L. Guangming, "Extraction Algorithm of Cell Image Feature based on Polygon Fitting", Computer Simulation, vol. 26, no. 11, **(2009),** pp. 168-171.
[3] T.K. Lee, D.I. McLean and M.S. Atkins, "Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions", Medical Image Analysis, vol. 7, no. 1, **(2003),** pp. 47–64.
[4] K. Taouil and N. B. Romdhane, "Automatic segmentation and classification of skin lesion images", Proceedings of the 2nd International Conference on Multimedia Application, IEEE Press, **(2006),** 1-12, Penang, Malaysia.
[5] L. Ma, W. Xu and L. Zhu, "Description of boundary irregularity on multi-scale local FD for melanomas", Proceedings of the 3rd International Conference on Bioinformatics and Biomedical Engineering, **(2009),** pp. 1–4, Beijing, China.
[6] F. Mokhtarian and A. K. Mackworth, "A theory of multi scale, curvature-based shape representation for planar curves", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 8, **(1992),** pp. 789-805.
[7] P. Baochuan, "Research on the Early Diagnosis of Tumor based on Computer Vision", Huazhong University of Science and Technology, **(2009).**
[8] F. Jingping, "Study on Methods of Cervical Cell Image Segmentation and Recognition", Jinan University, **(2010).**

[9] W. Dakai, H. Yuqing and P. Jinye, "Partial Differential Equation in Image Processing", Science Press, **(2008).**
[10] J. Jantzen, J. Norup, G. Dounias and B. Bjerregaard, "Pap-smear benchmark data for pattern classification", In Proceedings of the NiSIS, Nature inspired Smart Information Systems (NiSIS), EU co-ordination action, Albufeira, Portugal, **(2005),** pp. 1-9.
[11] T.M. Cover and P.E. Hart, "Nearest neighbor pattern classification", IEEE Transactions on Information Theory, vol. 13, no. 1, **(1967),** pp. 21–27.
[12] M. Li, "Melanoma Irregular Contour Description Method based on Multi-scale Local Fractal Dimension", Chinese Journal of Image and Graphics. vol. 15, no. 5, **(2010),** pp. 736-741.

Authors

XU Chuanyun, he has received his Master degree in Software engineering in 2006 from Chongqing University, China. He is currently pursuing the Ph.D. degree in computer science at College of Computer Science, Chongqing University. His research interest includes pattern recognition and medical image processing.

ZHANG Yang, she has received the M.S. degree in 2007 and in the Ph.D. degree in 2011 from Chongqing University, China. He is working at College of Computer and Information Science, Chongqing Normal University, Chongqing, China. Her research interest includes image processing and software engineering.