

Cell classification by moments and continuous wavelet transform methods

Qian Chen¹
Yuan Fan^{1,2}
Lalita Udupa²
Virginia M Ayres¹

¹Electronic and Biological
Nanostructures Laboratory,

²Nondestructive Evaluation
Laboratory, College of Engineering,
Michigan State University, East
Lansing, MI, USA

Abstract: Image processing techniques are bringing new insights to biomedical research. The automatic recognition and classification of biomedical objects can enhance work efficiency while identifying new inter-relationships among biological features. In this work, a simple rule-based decision tree classifier is developed to classify typical features of mixed cell types investigated by atomic force microscopy (AFM). A combination of continuous wavelet transform (CWT) and moment-based features are extracted from the AFM data to represent that shape information of different cellular objects at multiple resolution levels. The features are shown to be invariant under operations of translation, rotation, and scaling. The features are then used in a simple rule-based classifier to discriminate between anucleate versus nucleate cell types or to distinguish cells from a fibrous environment such as a tissue scaffold or stint. Since each feature has clear physical meaning, the decision rule of this tree classifier is simple, which makes it very suitable for online processing. Experimental results on AFM data confirm that the performance of this classifier is robust and reliable.

Keywords: cell classification, atomic force microscopy, second moment, continuous wavelet transform

Introduction

Atomic force microscopy (AFM) has recently found widespread use in cellular investigations because of its ability to provide active and high resolution probing of specimens with minimal sample preparation and under nearly lifelike conditions. Recent studies include investigations of cell surfaces (Frankel et al 2006) and sub-surfaces (Pelling et al 2004), active response to environmental change (Canetta et al 2006), healthy and pathological cell determination by elasticity variation (Dulinska et al 2006; Rosenbluth et al 2006) and local macromolecular probing of individual receptor sites (Hinterdorfer and Dufrêne 2006).

The large amounts of newly available AFM-based cellular structure-function information drives a need to systematize both the experimental approach and interpretation of the results across multiple scales. To aid interpretation, image processing techniques are bringing new insights to biomedical research. Recent studies include investigations of supervised learning-based cell image segmentation (Mao et al 2006) and automated tracking of cancer cell nuclei (Chen et al 2006). The automatic recognition and classification of biomedical objects can enhance work efficiency while identifying new inter-relationships among biological features.

In the present work we develop a combined AFM and image processing approach for use in cellular investigations. Our example is recognition of major structure-function relationships across multiple scales in blood cells. We set a further goal of maintaining translation, rotation, and scale invariance as important for both correct biological interpretation and real-time automated investigation. We find that a combination of moment-based invariant features with continuous wavelet transform (CWT) analysis

Correspondence: Virginia M Ayres
Electronic and Biological Nanostructures
Laboratory, College of Engineering,
Michigan State University, East Lansing,
MI 48824, USA
Email ayresv@egr.msu.edu

can provide recognition of major features across multiple scales while retaining translation, rotation, and scale invariance for accurate interpretation of data.

Materials and methods

Cell samples

Blood samples (about 3 ml), from male Wistar rats (Charles River Laboratories, Inc, Wilmington, MA), were centrifuged at 300 RPM at 4 °C for 15 minutes. Small volumes (<1 ml) containing mainly neutrophilic leukocytes (white blood cells) and a small amount of erythrocytes (red blood cells) on top of the centrifuged samples were extracted with a pipette and placed on a glass cover slide. Further details are provided by Goolsby and colleagues (2003).

AFM experimental parameters

Atomic force microscope images of the mixed cell types, as shown in Figure 1, were obtained using a Veeco Instruments Nanoscope IIIa (Woodbury, NY) operated in Tapping Mode® in ambient air as shown in Figure 1. Other experimental parameters included: use of a J scanner with a maximum 125 × 125

square micron x-y scan range, silicon tips with a nominal 10 nm tip radius of curvature, and a scan rate of approximately 1 Hz. The cells adhered readily to the glass slides and there was no evidence of tip-induced damage to the samples. Each image was a 512 × 512 pixel raster scan with 3 x-y-z points per pixel.

Analysis techniques

Moment-based features and continuous wavelet transform-based analysis were integrated to realize a robust and reliable classifier for blood cell types. The overall analysis procedure consists of three steps as depicted in Figure 2.

Moment-based methods have been successfully used to extract the shape, more accurately the shape feature of the object under investigation (Prokop and Reeves 1992). In pattern recognition (Grimson 1991), a feature (Jain and Zongker 1997) is an individual property of the phenomena being observed. The choice of discriminating properties as features is key to any classifier being successful. For a method appropriate for analyzing typical properties of blood cells, we defined the shape feature using functions of the second central moment (Gonzalez and Woods 1992).

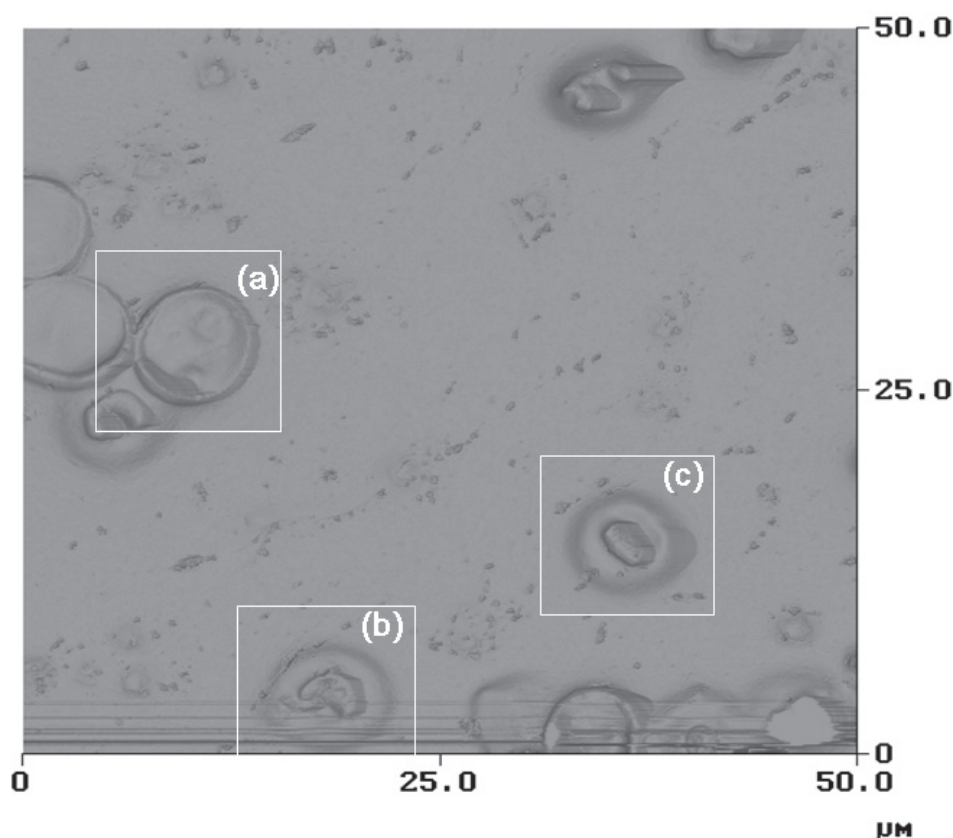


Figure 1 A typical atomic force microscopy mixed cell image. The cells shown in boxes are (a) an erythrocyte, (b) a neutrophilic leukocyte and (c) a possible monocyte leukocyte. All cells were globular objects about 5–6 microns in diameter. Copyright © 2003. Reproduced with permission from Goolsby B, Chen Q, Udupa L, et al 2003. Scanning probe microscopy with landmark referenced control for direct biological investigations. *J Nanosci Nanotech*, 3:347–50.

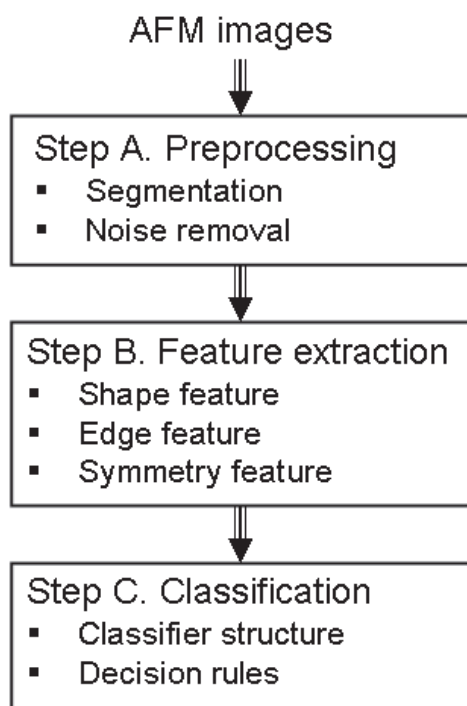


Figure 2 Schematic of overall approach.

Abbreviations: AFM, atomic force microscopy.

Wavelet analysis has been successfully used in both object segmentation (Unser 1995) and feature enhancement (Laine et al 1994). One property of the blood cells investigated in this paper is that they have objects which involve different scales. Since wavelet transform provides a multi-resolution analysis of the image, it can be used to detect objects of different scales. Another key property of the wavelet transform is its ability to characterize the local

regularity of a data set. Thus, by choosing the appropriate scale, we can get the relevant details from the images that are necessary to perform a particular recognition task across multiple scales. The resolution limit of the CWT multiscale recognition technique corresponds to the resolution within the image, which is nanometer scale for AFM (Wiesendanger 1994).

The classifier which we have developed has a tree structure. At each node, only a single feature is used to get efficient result. Each feature is normalized to be translation, scale, and rotation invariant, which makes the classifier robust.

Image analysis

Step A. Preprocessing

An atomic force microscope image of a cell is a map of the surface topography. The map is stored as pixelated data points (x, y, z). In a conventional AFM raster-scanned image, not all data points represent a cell. The overall measured image can be segmented into background (substrate) and cell (object) pixels which is superimposed with noise and artifacts due to operational parameters. Hence it is important to process the AFM image data to remove the background and the noise retain all cell features.

The first step involves the use of a threshold (Gonzalez and Woods 1992) to identify and discriminate the cells from the background and any noise, as shown in Figure 3. A high threshold value was used to remove the noise pixels and a second low threshold value was used to eliminate the background pixels. The thresholding filter can be expressed as

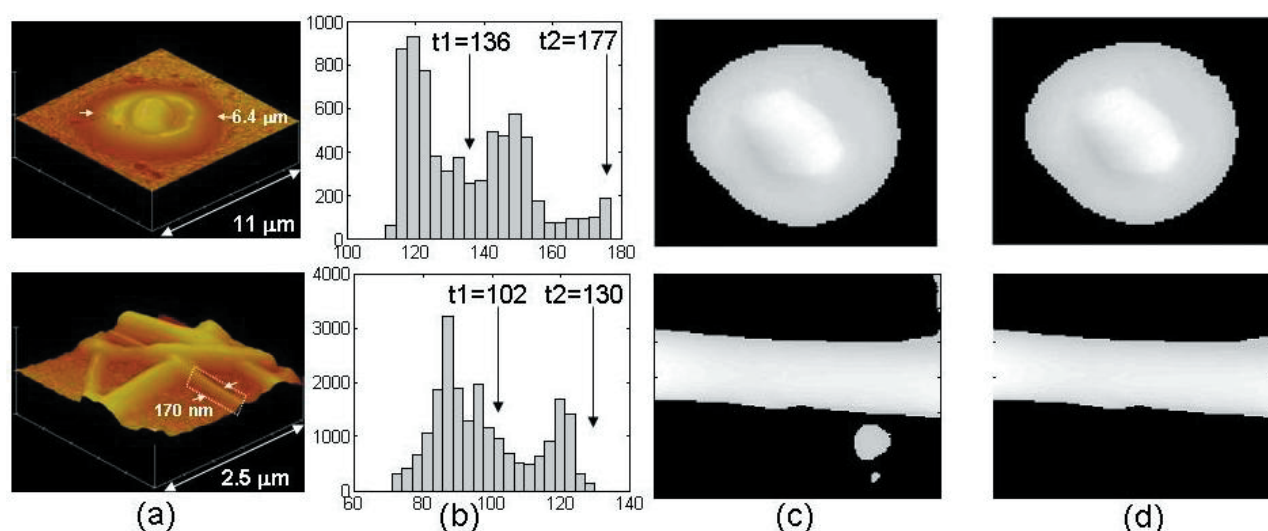


Figure 3 (a) Atomic force microscopy surface plots of original images of cell and tissue scaffold nanofibers examples; (b) Histograms to determine threshold ranges for cell and tissue scaffold nanofibers; (c) Result images after thresholding operation; (d) Result images after second filter operation.

$$g(x, y) = \begin{cases} 1 & \text{if } t_1 \leq f(x, y) \leq t_2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $f(x, y)$ is the original image, t_1 is the low threshold, t_2 is the high threshold, and $g(x, y)$ is the image after thresholding. The thresholds t_1 and t_2 were selected using a histogram based procedure (Tou and Gonzalez 1974) in which pixel intensities in the AFM is binned as shown in Figure 3 (b). After creating the histogram of original image, t_1 was chosen automatically by using Otsu's method (1978), which maximizes the between-class variance. t_2 was chosen manually by observing the histogram. If an image has considerable amount of high intensity noise pixels, a third peak may appear in the histogram in addition to the two peaks corresponding to background and object pixels. t_2 was chosen as a value below any third peak to eliminate such noise points. Results of the first thresholding operation illustrating the approach are given in Figure 3(a-c) for typical cell and also tissue scaffolding examples.

As seen in Figure 3(c), thresholding alone may not be sufficient to eliminate all of the background and the noise pixels. Therefore, after thresholding, the data was passed through a customized filter. In this filter, the area of each connected region was calculated, and those regions having small area were assumed to be noise regions and were removed. The results of thresholding combined with the second filtering operation were seen to be satisfactory for all data sets under investigation and are shown in Figure 3(d).

Step B. Feature extraction

Feature 1 (leukocyte versus erythrocyte)

Both leukocytes and erythrocytes have globular shapes. A distinguishing characteristic between them is that erythrocytes are anucleate. In previous work (Goolsby et al 2003), we have used a CWT method to successfully distinguish erythrocytes from leukocytes.

A CWT is a scale based two-dimensional transformation:

$$W(\sigma, \tau_1, \tau_2) = \frac{1}{\sigma} \iint f(x, y) \psi^* \left(\frac{x - \tau_1}{\sigma}, \frac{y - \tau_2}{\sigma} \right) dx dy \quad (2)$$

where σ is the scale parameter, τ_1 and τ_2 are the translation parameters, $*$ denotes the complex conjugation, and ψ is the mother wavelet function. We use a differential Gaussian function as the mother wavelet because this choice enhances edge information. A decision rule may be formulated based

on the observation that erythrocytes have only outer cell edges, whereas leukocytes have both inside nuclear edges and outer cell edges.

Examples of AFM images and their corresponding CWTs are shown in Figure 4(a) and 4(b). For each image, the mass center is calculated, which is the point where the pixel intensities are balanced for any line through that point. By plotting the pixel intensities along horizontal and vertical lines through the mass centers of the CWT transformed images as shown in Figure 4(c), it is clear that leukocytes and erythrocytes have a different number of peaks. The peaks correspond to edge pixels and hence the presence of a nucleus inside the cell is represented by the number of zero-crossings, which is the number of times that the pixel intensities cross the zero level. We define the zero level as the average value of the maximum and minimum pixel intensities along each line.

Two orthogonal lines are chosen for this analysis for robustness. We choose the number of zero-crossings as the feature which represents discriminating information for erythrocytes and leukocytes, and will refer to the feature thus defined as the edge feature.

Feature 2 (monocyte leukocyte versus neutrophilic leukocyte)

The distinguishing characteristic between monocyte and neutrophilic leukocytes is the shape of their nuclei. In order to derive quantitative values for the features, the first step segments the image of nucleus from the overall image as described below.

Wavelet-based multi-resolution analysis clearly decomposes the overall image into objects at two different scales, namely, $\sigma 1$ – cell and $\sigma 2$ – nucleus. The objects at scale $\sigma 2$ provides the boundary of the nucleus which can then be filled in using morphological dilation operations (Gonzalez and Woods 1992) to isolate the final image of the nucleus alone as shown in Figure 5(f).

The nuclei of the two kinds of cells have different shapes. The differences in the shape can be quantified by calculating the second moment along the major and minor axes of the image which are orthogonal to each other. The major axis is found using Equation 3:

$$\theta = \arctan \left(\frac{2M_{xy}}{M_{xx} - M_{yy} + \sqrt{(M_{xx} - M_{yy})^2 + (2M_{xy})^2}} \right) \quad (3)$$

Where

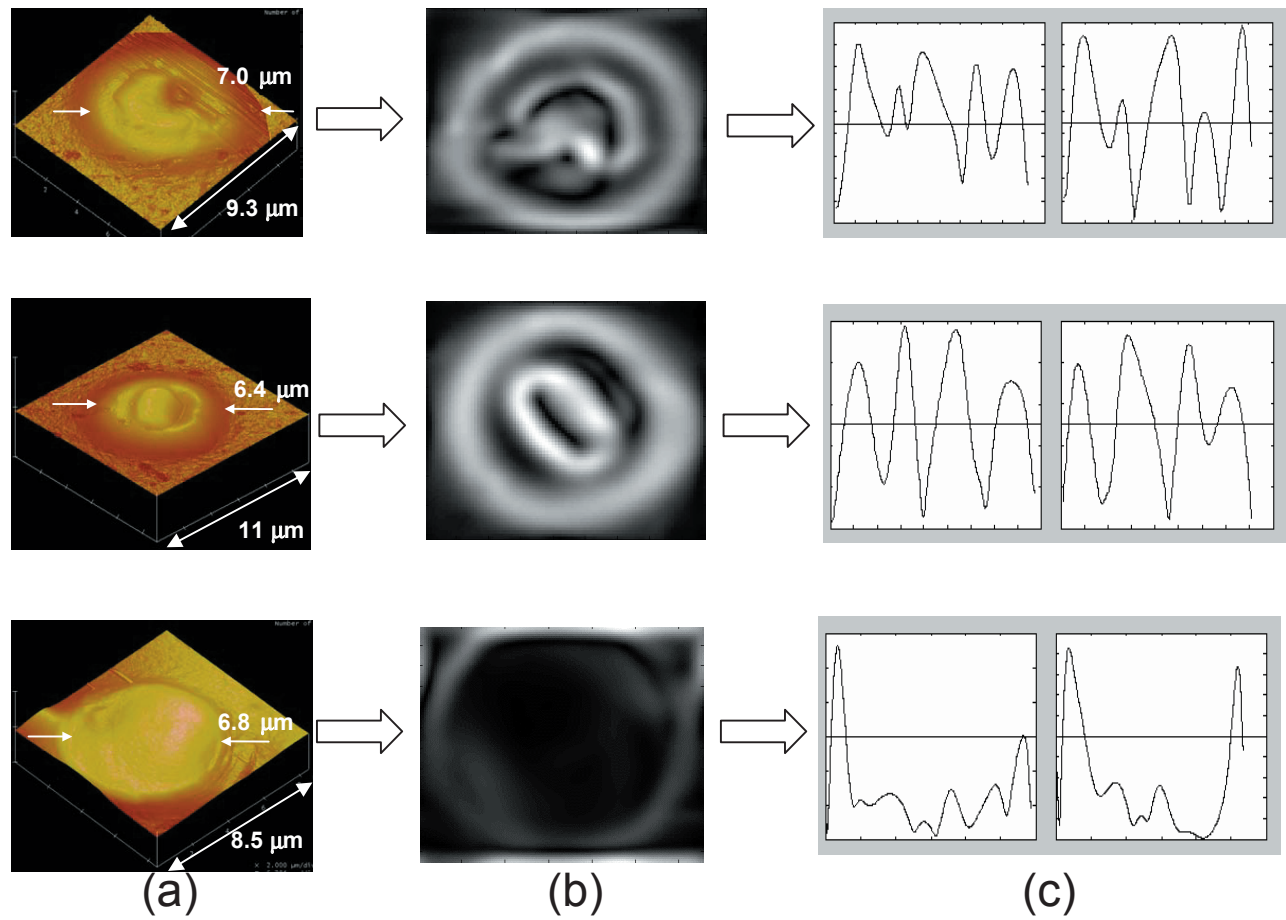


Figure 4 The definition of the edge feature is based on (a) atomic force microscopy images; (b) continuous wavelet transformed images. (c) Pixel intensities on horizontal and vertical lines clearly show the edge feature for each cell type.

$$M_{xx} = \sum_x \sum_y (x - \bar{x})^2 f(x, y) \quad (4a)$$

$$M_{xy} = \sum_x \sum_y (x - \bar{x})(y - \bar{y}) f(x, y) \quad (4c)$$

The second moment along the major axis can be calculated using the following equation.

$$M_{yy} = \sum_x \sum_y (y - \bar{y})^2 f(x, y) \quad (4b)$$

$$M_{min} = \sum_x \sum_y [(x - \bar{x}) \sin \theta - (y - \bar{y}) \cos \theta]^2 \quad (5)$$

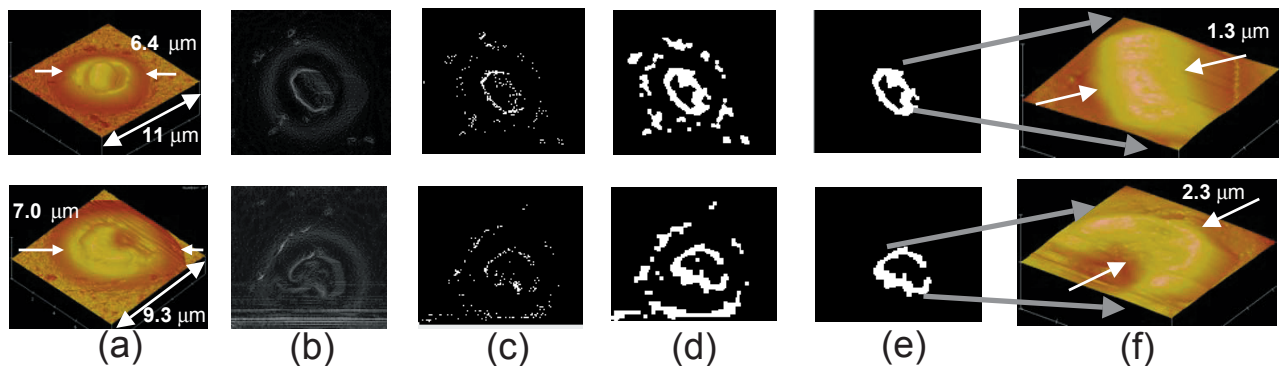


Figure 5 (a) Preprocessed atomic force microscopy images; (b) continuous wavelet transformed images; (c) Edge images detected based on continuous wavelet transformed images; (d) Enhanced edge images by dilation operation; (e) Edge images after noise removal; (f) Result images.

The second moment along the minor axis can be calculated using the orthogonality property by the following equation.

$$M_{max} = \sum_x \sum_y [(x - \bar{x}) \sin(\theta + 90^\circ) - (y - \bar{y}) \cos(\theta + 90^\circ)]^2 \quad (6)$$

By plotting the pixel intensities along the minor axis as shown in Figure 6, it is seen that the curve is almost symmetric respect to its center in the case of monotype leukocyte, and asymmetric in the case of neutrophilic leukocyte. The symmetry of the integrated area under the curve respect to dotted line as shown in Figure 6 was used as a feature to classify monotype leukocytes from neutrophilic leukocytes. We will refer to this as the symmetry feature. It is also clear that this feature is insensitive to translation, rotation, and scaling of the measured image.

Feature 0 (fibers versus cells)

In many applications it is necessary to identify cells in a stint, tissue scaffolding, or natural collagen environment. The approach described above can be easily extended to address the problem of distinguishing cells within a fibrous environment. An example of the analysis of tissue scaffolding

nanofibers (Chen et al 2004, 2005; Rutledge et al 2006) image data is presented in this section.

Tissue scaffold nanofibers and cells have distinguishable shapes. Moment-based methods may be used to determine a feature representing. A shape feature was defined as the ratio of second moments along the major axis (M_{min}) and minor axis (M_{max}):

$$ratio = \frac{M_{max}}{M_{min}} \quad (7)$$

The shape feature defined in Equation 7 has the properties of translation, scale and rotation invariance.

Step C. Classifier development

The classifier is a set of rules that can categorize the biomedical objects into one of a set of known types. Only one feature is used at each node in a decision rule (Duda et al 2000) as shown in Figure 7 to keep the classifier simple and fast.

Feature 0

At the first node, fibers are separated from cells based on the shape feature. The classification algorithm used is the well established K Nearest Neighbor (KNN) procedure (Duda et al 2000) where each object is assigned the class of its K nearest

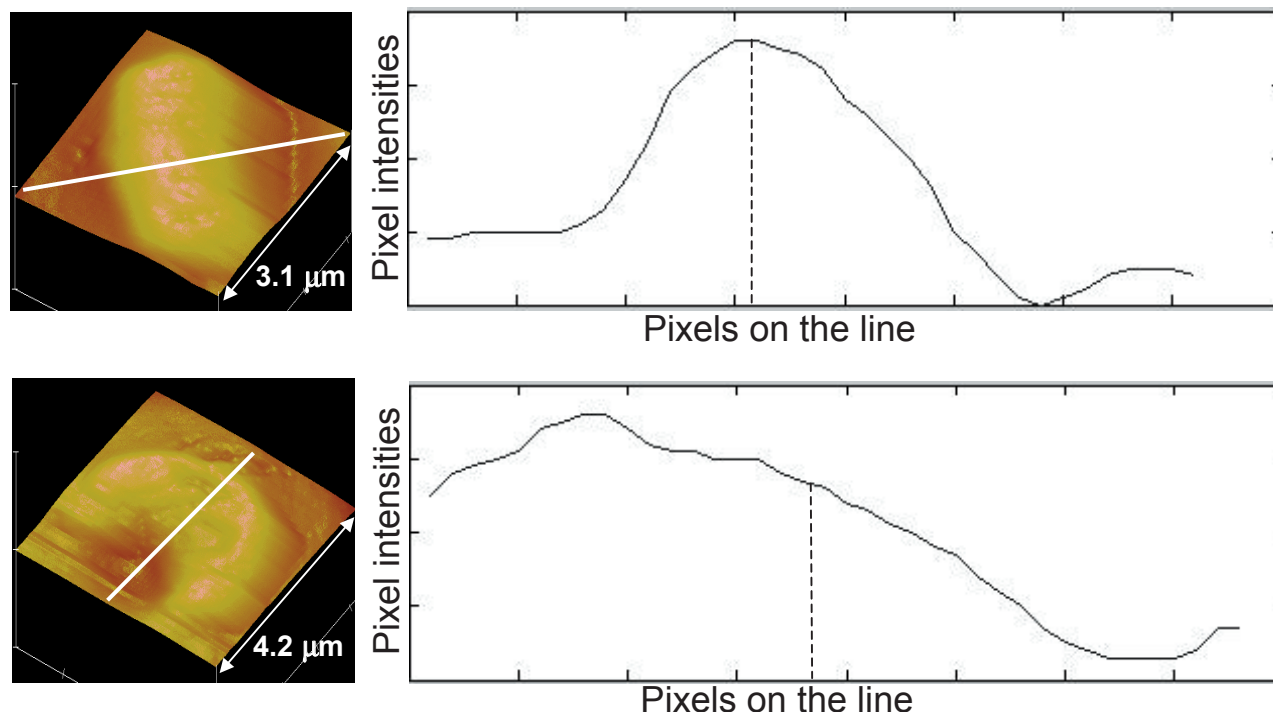


Figure 6 Minor axis and the pixel intensities on the minor axis for evaluation of the symmetry feature.

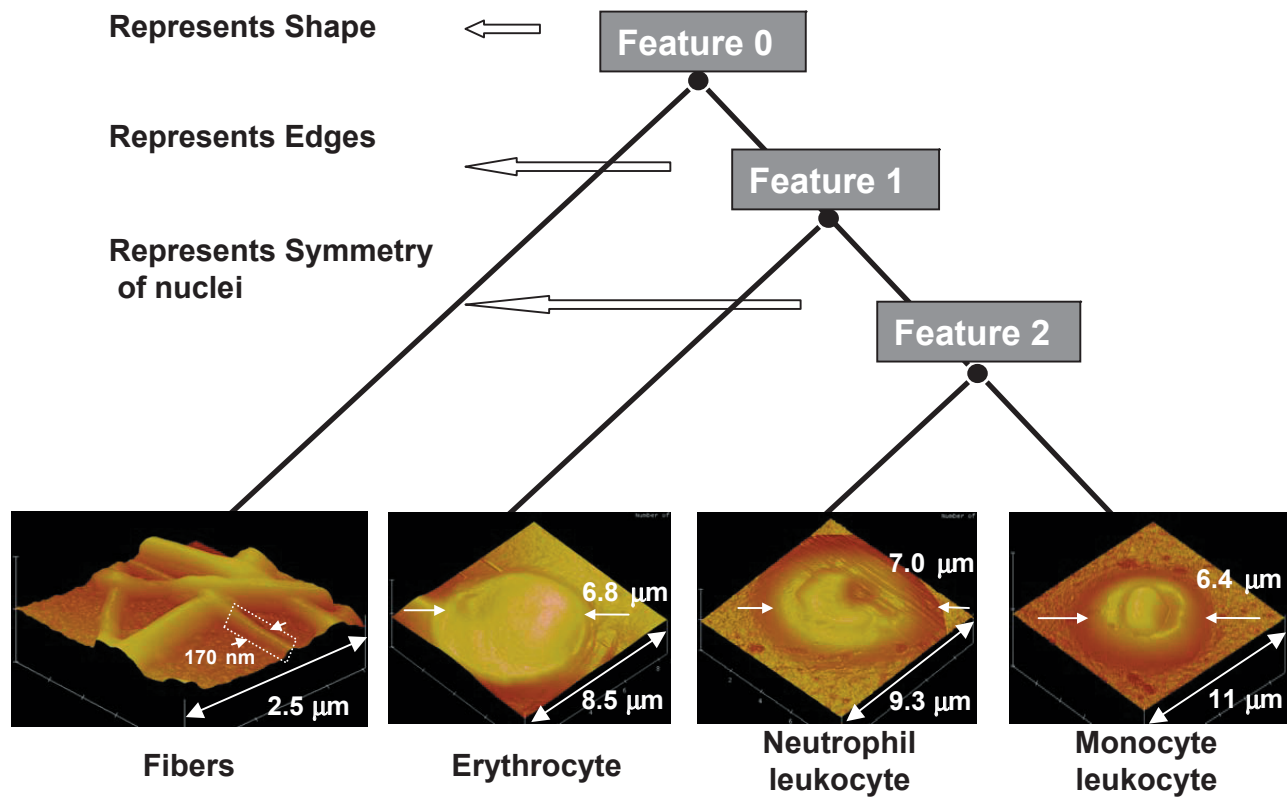


Figure 7 Classifier structure.

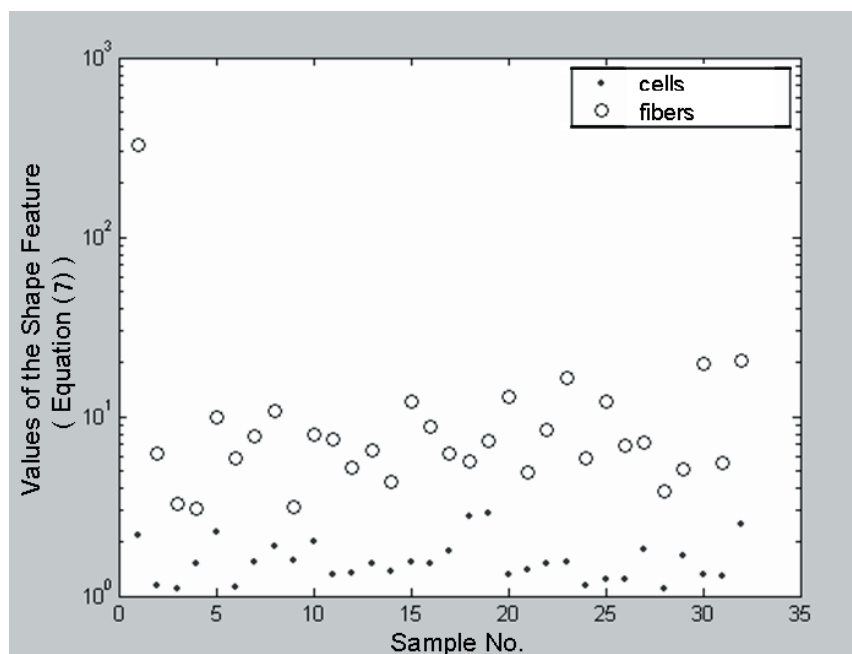


Figure 8 The value of the shape feature can be used to distinguish cells (solid dots) from fibers (open circles).

Table 1 Summary of classifier results I

		Assigned	
		Fibers	Cells
True	Fibers	32	0
	Cells	3	29

neighbors. Consequently, objects with similar values of the shape feature are clustered into the same class.

Feature 1

At the second node, erythrocytes are separated from leukocytes based on the edge feature. Since leukocytes have nuclei and erythrocytes do not have a nucleus, the decision rule is formulated as the following:

- If $n \geq 6$, object is a leukocyte
- If $n \leq 4$, object is an erythrocyte
- If $4 < n < 6$, object is unknown

Feature 2

At the third node, monocyte leukocytes are separated from neutrophil leukocytes based on the symmetry feature. For the two leukocytes shown in Figure 5, the value of the symmetry feature of the monocyte leukocyte was 1.9016 and that of the neutrophilic leukocyte was 3.5541. The threshold value for this feature can be derived using

training data and incorporated into the decision rule for classification.

Results of classifier application

A database of 32 cell-type and 32 nanofiber-type biomedical objects were used to test the performance of the classifier presented here. The segmented objects were classified as “cells” and “fibers”. Figure 8 shows that the cells can be separated from the fibers by extracting the shape feature.

The entire database containing 64 images was randomly split into training set (32 images) and test set (32 images). The classifier was trained using the training data and the performance was evaluated using the test data. This procedure was repeated 10 times to obtain the average error rate and the standard deviation of the error rate of the 10 test data sets. The result based on the KNN classifier is given in Table 1.

The classifier is tested further to classify 6 leukocytes and 6 erythrocytes. Table 2 shows the classification results based on edge feature obtained using the CWT method.

Only a single possible monocyte leukocyte was found in all cell samples and is shown in Figure 1. For the group of cells in Figure 1, the cell images were manually segmented and the sub-images were passed into our classifier. The value of the symmetry feature of the monocyte leukocyte was 1.9016 while that of the neutrophilic leukocyte were

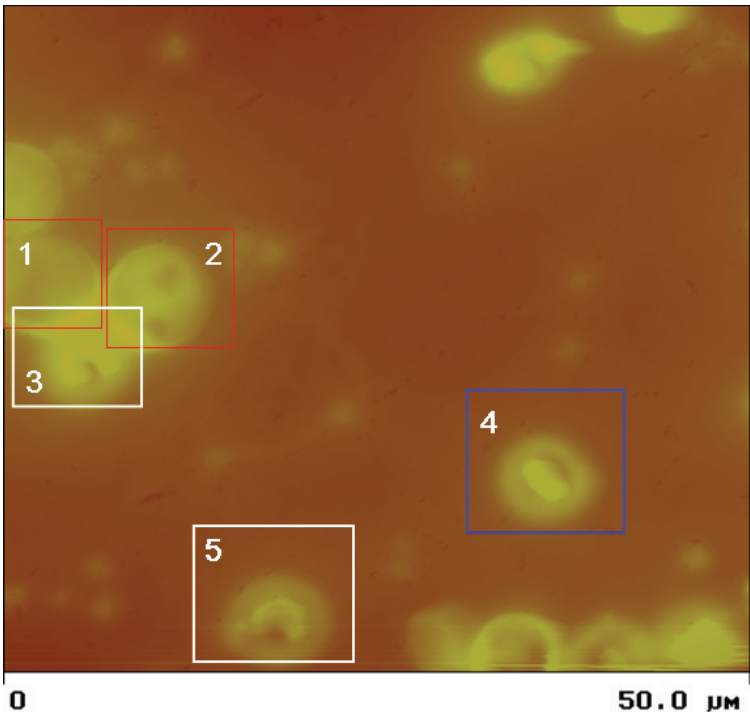


Figure 9 Classification results: those sub-images that are assigned to erythrocytes are shown in red boxes, sub-images that are assigned to neutrophilic leukocytes are shown in white boxes, and sub-images that are assigned to monocyte leukocytes are shown in blue boxes.

Table 2 Summary of classifier results II

		Assigned	
		Leukocytes	Erythrocytes
True	Leukocytes	5	1
	Erythrocytes	0	6

3.5541 and 4.0153. The experiments shows that the cells of Figure 1 could be correctly auto-classified as indicated in Figure 9.

Discussion

This paper presents an automated algorithm for the analysis of cell and tissue images. The overall procedure consists of three steps, namely, preprocessing and segmentation, feature extraction, and classification. Each feature extracted represents the most significant difference between the two types of objects to be classified and a simple rule-based decision tree classifier with one feature used at each node offers a fast and simple technique for image analysis. The experiments result shows that the classifier has low error rate with simple structure and runs fast.

The integration of image processing techniques are with AFM techniques will offer the possibility of a new area of biomedical research (Udpa et al 2006). By combining the results of automated image analysis with the AFM system, appropriate lithography commands could be used to dynamically control the movement of the tip in the x-y plane which in turn can help accomplish more efficient scanning of the sample. Such a system can be of much value for the effective and intelligent use of scanning probe microscopy systems.

References

- Canetta E, Adya AK, Walker GM. 2006. Atomic force microscopic study of the effects of ethanol on yeast cell surface morphology. *FEMS Microbiology Letters*, 255:308–12.
- Chen Q, Udpa L, Schindler MS, et al. 2004. Scanning probe recognition microscopy investigation of cells on scaffolding. *Mat Res Soc Symp Proc*, EXS-1:143–5.
- Chen Q, Fan Y, Ayres VM, et al. 2005. Scanning probe recognition microscopy investigation of the elastic properties of tissue scaffolding. *Mater Res Soc Symp Proc*, 838E.
- Chen X, Zhou X, Wong STC. 2006. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Trans Biomed Eng*, 53:762–6.
- Duda RO, Hart PE, Stork DG. 2000. Pattern classification (2nd ed). New York: Wiley Interscience.
- Dulinska I, Targosz M, Strojny W, et al. 2006. Stiffness of normal and pathological erythrocytes studied by means of atomic force microscopy. *J Biochem Biophys Methods*, 66:1–11.
- Frankel DJ, Pfeiffer JR, Surviladze Z, et al. 2006. Revealing the topography of cellular membrane domains by combined atomic force microscopy/fluorescence imaging. *Biophys J*, 90:2404–13.
- Gonzalez RC, Woods RE. 1992. Digital image processing. Reading, MA: Addison-Wesley.
- Goolsby B, Chen Q, Udpa L, et al. 2003. Scanning probe microscopy with landmark referenced control for direct biological investigations. *J Nanosci Nanotech*, 3:347–50.
- Grimson WEL. 1991. Object recognition by computer: the role of geometric constraints. Cambridge, MA: MIT Pr.
- Hinterdorfer P, Dufrêne YF. 2006. Detection and localization of single molecular recognition events using atomic force microscopy. *Nature Meth*, 3:347–55.
- Jain A, Zongker D. 1997. Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell*, 19:153–8.
- Laine AF, Schuler S, Fan J, et al. 1994. Mammographic feature enhancement by multiscale analysis. *IEEE Trans Med Imaging*, 13:725–40.
- Mao KZ, Zhao P, Tan P. 2006. Supervised learning-based cell image segmentation for P53 immunohistochemistry. *IEEE Trans Biomed Eng*, 53:1153–63.
- Otsu N. 1978. A threshold selection method from gray-scale histogram. *IEEE Trans Syst Man Cybern*, 8:62–6.
- Pelling AE, Sehati S, Gralla EB, et al. 2004. Local nanomechanical motion of the cell wall of *Saccharomyces cerevisiae*. *Science*, 305:1147–50.
- Prokop RJ, Reeves AP. 1992. A survey of moment-based techniques for unoccluded object representation and recognition. *Comput Vis Graph Image Process*, 54:438–60.
- Rosenbluth MJ, Lam WA, Fletcher DA. 2006. Force microscopy of nonadherent cells: a comparison of leukemia cell deformability. *Biophys J*, 90:2994–3003.
- Rutledge SL, Shaw HC, Yowell LL, et al. 2006. Self assembly and correlated properties of electrospun carbon nanofibers. *Diamond Relat Mater*, 15:1070–4.
- Tou JT, Gonzalez RC. 1974. Pattern recognition principles. Reading, MA: Addison-Wesley.
- Udpa L, Ayres VM, Fan Y, et al. 2006. Deconvolution of atomic force microscopy data for cellular and molecular imaging. *IEEE Sig Proc Mag*, 23:73–83.
- Unser M. 1995. Texture classification and segmentation using wavelet frames. *IEEE Trans Image Proc*, 4:1549–60.
- Wiesendanger R. 1994. Scanning probe microscopy and spectroscopy. Cambridge, England: Cambridge Univ Pr.

