

Automatic Classification of Coating Epithelial Tissue

Claudia Mazo¹, Maria Trujillo¹, and Liliana Salazar²

¹ School of Computer and Systems Engineering

² Department of Morphology

Universidad del Valle,

Ciudadela Universitaria Melendez, Cali, Colombia

{claudia.mazo,maria.trujillo,liliana.salazar}@correounivalle.edu.co

Abstract. Histology images may be used in E-Learning systems to teach how morphological features and function of each organ contribute to its identification. Automatic classification of coating epithelial cells is an open problem in image processing. This problem has been addressed using morphological gradient, region-based and, shape-based method, among others. In this paper, coating epithelial cells are recognised and classified into: Flat, Cubic and Cylindrical. Epithelial cells are classified based on sphericity and projection. Information about sphericity is used to classify cells into cubic and a measure based in projecting cell nucleus into light region is used to classify into flat and cylindrical. Experimental validations are conducted according to expert criteria, along with manually annotated images, as a ground-truth. Experimental results revealed that the proposed approach recognised coating epithelial cells and classified tissues in a similar way to how experts have performed these classifications.

1 Introduction

Tissue samples are used to study biological structures present in living organisms. These samples contain information about cells distribution and organisation. Large amounts of pathologies are identified by analysing cell normality or abnormality, besides supporting diagnosis in daily medical practice of histologists, biologists, pathologists, and related disciplines [1].

Large repositories of images containing tissue structures are available due to developments in information gathering. As a result, automatic identification and classification of structures in a tissue become possible. Automatic image analysis is based on cells information. Cell analysis process consists of three steps: identifying cell nuclei, measuring morphological features, and classifying. Since, cell nucleus are a key part in identifying biological structures, much effort is dedicated to identify cell nucleus. A computer-aided identification of cell nuclei has been addressed using different techniques: morphological gradient [2], region-based methods [3], shape-based methods [4], minimax algorithm and thresholding [5], geometric active contours [6], binary graph cuts [7], K-means clustering and

tracking algorithm [8], fuzzy C-means clustering [9], and colour intensity [10]. However, these techniques require to set manually parameters. Moreover, histology images contain cell variations due to photoelectronic and thermal noise that make cell nuclei identification an open problem. Computer-aided identification of a cell is affected by some issues such as: similarity between cell nucleus from different tissues, image noise acquired by capture devices, cell nuclei size, and low definition in image areas, like edges, among others.

In this paper, an approach to automatically identify and classify coating epithelial tissue – into flat, cubic and cylindrical – using morphological information is presented. Experimental results shown that the proposed approach correctly classify the different types of epithelial cell nucleus. The rest of the paper is organised as follows. The proposed classification approach is presented in Section 2. Experimental validation is included in Section 3. Finally, conclusions are presented in Section 4.

2 A Classification Approach of Coating Epithelial Tissue

Automatic identification and classification of epithelial cell nucleus are tackled by image processing techniques and tissue morphological information. The input is an image and the automatic process is four-fold. First, an initial segmentation is obtained using the largest eigenvalue of the Structure Tensor algorithm [10] on the red colour channel. Obtained segmentation along with the red and the green colour channels are used as input into the K-means algorithm [11] to identify cell nucleus and light regions – tissue samples have shapeless spaces or holes, that generate regions with maximum luminance values in an image, called light regions. Red and green colour channels have relevant information on cell nuclei along with light regions recognition, due to the use of stains. Second, distances between light regions and cell nucleus are calculated, cell nucleus closest to light regions are identified as epithelial cell nucleus. Third, the sphericity is calculated with the Feret’s diameter in order to identify the cubic epithelium tissue. Fourth, a projection measure is calculated to differentiate between flat and cylindrical epithelium tissue.

2.1 Coating Epithelial Cell Identification

Let $I \in R^{MN}$ be a histological image of the cardiovascular system in RGB color space; $Q \in R^{MN}$ be a matrix such that each element is formed by 3 values (x, y, z) where x represents the intensity of the red color channel of I , y represents the intensity of the green color channel of I , and z represents the largest eigenvalue of structure tensor of I ; $H_k(t)$ be a cluster represented by a set of vectors in \mathbb{R}^3 in the t – th iteration; and $C_k(t) \in \mathbb{R}^3$ be a centroid k of the cluster $H_k(t)$.

For initialising the cluster algorithm, initialisation parameters are set: $t = 0$, $C_1(0) = \{70, 30, 150\}$, $C_2(0) = \{160, 70, 30\}$, and $C_3(0) = \{110, 154, 0\}$.

$$Q_{ij} \in H_k(t) \rightarrow k = \underset{k \in \{1,2,3\}}{\operatorname{argmin}} \left| \sum_{u=0}^2 (Q_{ij}[u] - C_k(t)[u])^2 \right|, \quad (1)$$

$$C_k(t+1) = \frac{1}{|H_k(t)|} \sum_{i=1}^M \sum_{j=1}^N a_{ij} \cdot Q_{ij}, \quad (2)$$

where a_{ij} is 1 if $Q_{ij} \in H_k(t)$ and 0 in other case. Increase $t = t + 1$ and repeat (1) and (2) until $H_k(t+1) = H_k(t)$. Let $O = \{O_1, O_2, O_3\}$ be a binary image such that:

$$(O_k)_{ij} = \begin{cases} 1 & Q_{ij} \in H_k(t) \\ 0 & \text{else,} \end{cases} \quad (3)$$

where a value of k represents a different tissue in I , such that $k = 1$ contains pixels position associated to connective tissue, $k = 2$ contains pixels position associated to cell nucleus, and $k = 3$ contains pixels position associated to light regions.

A Flood-fill algorithm is applied to image regions associated to each O_k . The result is an image with a value within $[0, 3]$ such that pixels with 0 are background.

Definition: A region R is a set of pixels with same intensity value excluding the background.

According to the Definition, δ is a region that represents a cell nuclei and γ is a region that represent a light regions. Epithelial cells are characterised for being close to a light region. Distances [14] between cell nuclei and light regions depend on cell nucleus sizes, which is determined by the zoom used in a microscope and the resolution used to digitize images. In this work 40x zoom and a resolution of 2048x1536 pixels are used. Cells farther than 200.3pixels to light regions are not considered epithelial. Set of epithelial cells is defined as follows:

$$E = \{\delta_i | \exists p_i \in \delta_i \wedge \exists p_j \in \gamma_j : \operatorname{dist}(p_i, p_j) \leq 200.3\}, \quad (4)$$

where $\delta_i \in O_2$ is a cell nucleus, $\gamma_j \in O_3$ is a light region, p_i and p_j are coordinates (x, y) , and dist function is expressed by:

$$\operatorname{dist}(p_i, p_j) = \sqrt{(p_i - p_j)^2}. \quad (5)$$

2.2 Coating Epithelial Cell Classification

Epithelial cells have different functions according to their shape and orientation. They are classified into flat, cubic or cylindrical. Sphericity and position relative to light regions are used to classify automatically epithelial cells. Once epithelial

cells are classified, a tissue is classified based on the largest frequency of grouped cells.

Since cubic epithelial cells have sphere shape, a measure of sphericity, based on the Feret's diameter [13], is used to identify cubic epithelial cells. Flat and cylindrical cells have ellipse shapes. Flat cells are parallel to light regions whilst cylindrical cells are perpendicular to light regions. The line defined by the two extreme points in the widest diameter of a cell nuclei is called baseline. Baseline is used to calculate the *ProjectionRatio* of a cell nucleus. An epithelial cell is classified into flat or cylindrical based on the relative position to the nearest light region. The classification approach is described in detail as follows:

Cubic Epithelial Classification. Sphericity is measured as a ratio between two values the minimum and the maximum Feret's diameter. The Feret's diameter is the distance between two parallel tangents such that they are perpendicular to a fixed direction. The maximum Feret's diameter is the maximum possible value and the minimum Feret's diameter is the minimum possible value [13]:

$$Sphericity = \frac{FeretMinimum(c)}{FeretMaximum(c)}. \quad (6)$$

Cell nuclei with sphericity values larger than 0.8 are considered as sphere shape and cell nuclei are classified as cubic.

Flat and Cylindrical Epithelial Classification. Different methods were evaluated: slope [15] and Law of Cosines [16]. However, the proposed classification is based on orientation of cells with respect to a light region. The baseline length is the length of the line defined by the two connected points of the maximum Feret's diameter (q_1, q_2). The projection length is the length of the projection line of a cell over the closest light area, it is illustrated in Fig. 1.

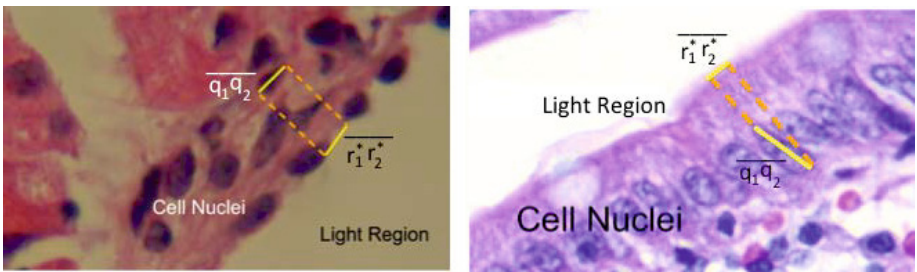


Fig. 1. illustration of cell projection

Let NLR be the set of the nearest light region coordinates, q_1 and q_2 be intersection points between the maximum Feret of a cell nuclei, r_1 and r_2 be intersection points between a NLR and q_1 and q_2 points.

$$r_1^* = \underset{r_1 \in NLR}{\operatorname{argmin}} \sqrt{(q_{1(x)} - r_{1x})^2 + (q_{1(y)} - r_{1y})^2}, \quad (7)$$

$$r_2^* = \underset{r_2 \in NLR}{\operatorname{argmin}} \sqrt{(q_{2(x)} - r_{2x})^2 + (q_{2(y)} - r_{2y})^2}. \quad (8)$$

The *ProjectionRatio* is calculated as:

$$\textit{ProjectionRatio} = \frac{\ell(\overline{r_1^* r_2^*})}{\ell(\overline{q_1 q_2})}, \quad (9)$$

where $\ell(\overline{r_1^* r_2^*})$ is the length of the line formed by r_1^* and r_2^* and $\ell(\overline{q_1 q_2})$ the length of the line formed by q_1 and q_2 .

Epithelial cells with a *ProjectionRatio* larger or equal value than 0.5 are classified as flat epithelium and epithelial cells with a *ProjectionRatio* smaller value than 0.5 are classified as cylindrical epithelium.

3 Experiments and Analysis of Results

In this section, results of the tests performed and the comparative analysis between the proposed method and the manual classification is carried out. In order to assess the proposed approach, epithelial tissue samples were processed with hematoxylin-eosin staining to highlight cell nucleus. A set of 100 histological images, composed by 34 images of flat epithelial tissue, 33 images of cubic epithelial tissue and 33 images of cylindrical epithelial tissue, were processed to validate the proposed approach. The sensitivity [17] is used as a performance measure.

Results of automatic classification of epithelial tissue were evaluated by a group of experts. Once cells are identified and classified into flat, cubic and cylindrical, the frequency of cells in each group is calculated. A tissue is classified according to the largest frequency of grouped cells. A subset of 30 manual and automatic classification results are included in Fig. 2.

A quantitative evaluation of the obtained results is presented in Table 1. The automatic identification results are compared with the manual identification using: True Positive calculated as epithelial tissue correctly identify; False Negative calculated as epithelial tissue incorrectly identify.

Table 1. Performance evaluation of tissues classification

Confusion Matrix	Flat	Cubic	Cylindrical	Total
True Positive	31	26	28	85
False Negative	3	7	5	15
Sensitivity	0.91	0.79	0.85	0.85

Results show that the proposed approach is able to recognize flat, cubic and cylindrical epithelial tissues with values of sensitivity greater than 0.79. The best score of sensitivity was obtained when classifying flat epithelial tissue while the worst score was obtained when classifying cubic epithelial tissue. The score obtained in cubic classification may be a consequence of joined cells since they look similar to other epithelial tissue.

<i>Image</i>	<i>Manual</i>	<i>Automatic</i>	<i>Image</i>	<i>Manual</i>	<i>Automatic</i>
	Flat	Flat		Flat	Flat
	Flat	Flat		Flat	Flat
	Flat	Flat		Flat	Flat
	Flat	Flat		Flat	Flat
	Cylindrical	Cylindrical		Flat	Flat
	Flat	Flat		Flat	Flat
	Flat	Flat		Cylindrical	Flat
	Cylindrical	Cylindrical		Cylindrical	Cylindrical
	Cylindrical	Cylindrical		Cubic	Cubic
	Cylindrical	Cylindrical		Cubic	Cubic
	Cylindrical	Cubic		Cubic	Cylindrical
	Cylindrical	Cylindrical		Cubic	Cubic
	Cylindrical	Cylindrical		Cubic	Cylindrical
	Flat	Flat		Cubic	Cylindrical
	Flat	Flat		Cubic	Cubic

Fig. 2. Results of manual and automatic classification of coating epithelial tissue

4 Conclusions

Coating epithelial tissue is automatically classified using image analysis. The problem is split into segmentation and classification. Three challenges are tackled for segmenting coating epithelial cells:

- cell's appearance variation produced by staining and low definition image areas such as borders
- cell's size and shape variations
- the presence of similar cell nucleus that do not belong to the epithelial tissue.

The Structure Tensor algorithm captures edges of cells, even though edges are ill-defined and different stains may be used. Regarding different sizes and shapes of cells, they are not an obstacle for segmenting since information is handled at pixels level using edges — provided by the structure tensor algorithm —, the red and the green colour channels. Proximity to light areas is used to identify cell nucleus which belong to the epithelial tissue.

Once coating epithelial cells are segmented, classification into flat, cubic and cylindrical involves several challenges:

- size of cells may vary
- flat and cylindrical cells have similar shapes
- cells may lose their particular shape due to tissue cutting.

Variations on cell's size are overcome by measuring sphericity since this feature is preserved regardless size. The projection of cell nuclei into light areas is used to distinguish between flat and cylindrical. Finally, a tissue is classified using the mode, under the assumption of unimodal distribution, which is robust against tissue cutting effects.

Automatic identification may help to solve one of the greatest problems of pathological diagnosis – lack of reproducibility –, which is common since the diversity of views among pathologists.

Failure rate of the proposed method is due to union or proximity of cells nucleus which are taken as a single cell nuclei. A joined cells separation process will be required in order to improve results.

References

1. Gartner, L.P., Hiatt, J., Strum, J.M.: Cell Biology and Histology, 6th edn. Lippincott Williams & Wilkins (2010)
2. Nedzved, A., Ablameyko, S., Pitas, I.: Morphological Segmentation of Histology Cell Images. *Computer Vision and Pattern Recognition* 1, 500–503 (2000)
3. Pham, D.L., Xu, C., Prince, J.L.: A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 335–338 (2000)
4. Tsai, A., Yezzi Jr., A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E., Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging* 22, 137–154 (2003)

5. Swank, P.R., Greenberg, S.D., Winkler, D.G., Hunter, N.R., Spjut, H.J., Estrada, R., Taylor, G.R.: Nuclear segmentation of bronchial epithelial cells by minimax and thresholding techniques. A comparison. *Anal Quant Cytol.* 5(3), 153–158 (1983)
6. Harandi, N.M., Sadri, S., Moghaddam, N.A., Amirfattahi, R.: An automated method for segmentation of epithelial cervical cells in images of Thin Prep. *Journal of Medical Systems* 34(5), 1043–1058 (2012)
7. Eramian, M., Daley, M., Neilson, D., Daley, T.: Segmentation of epithelium in H&E stained odontogenic cysts. *Journal of Microscopy* 244(3) (2011)
8. Tonkin, J., Rees, P., Brown, M., Errington, R., Smith, P., Chappell, S., et al.: Automated cell identification and tracking using nanoparticle moving-light-displays. *Journal PLoS One* 7, 1–8 (2012)
9. Muhimmah, I., Kurniawan, R.: Indrayanti: Automatic Epithelial Cells Detection of Pap smears images using Fuzzy C-Means Clustering. *International Proceedings of Chemical, Biological & Environmenta.* 29, 122–127 (2012)
10. Bibo, L., Chunli, M., Hui, W.: Pixel level image fusion based on linear structure tensor. In: 2010 IEEE Youth Conference on nformation Computing and Telecommunications (YC-ICT), pp. 303–306 (2010)
11. Kanungo, T., Mount, D.M., Nathan, S.N., Christine, D., Piatko, S.R., Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 881–892 (2002)
12. Nosal, E.M.: Flood-fill algorithms used for passive acoustic detection and tracking. In: *New Trends for Environmental Monitoring Using Passive Systems*, pp. 1–5 (2008)
13. Merkus, H.G.: *Particle Size Measurements: Fundamentals, Practice, Quality*, vol. 15. Springer (2012)
14. Deza, E., Deza, M.: *Encyclopedia of Distances*, pp. 94–98. Springer, Berlin (2009)
15. Eves, H.: *Mathematical Circles Revisited: A Second Collection of Mathematical Stories and Anecdotes*. Prindle, Weber and Schmidt (1972)
16. Heath, L., Thomas, L.: *Euclids Elements* Thomas. Green Lion Press (2012)
17. Xiao-Qin, Z., Ng, W., Yeung, D.: Statistical Sensitivity Measure of Single Layer Perceptron Neural Networks to Input Perturbation. *Machine Learning and Cybernetics*, 3101–3105 (2006)