INFO- 2228
Primljeno/Received:2020-03-12

UDK: 616.24:616-07
Preliminary Communication / Prethodno priopćenje

https://doi.org/ 10.32914/i.53.3-4.5

AUTOMATED COMMUNICATION SYSTEM FOR DETECTION OF LUNG CANCER USING CATASTROPHE FEATURES

AUTOMATIZIRANI KOMUNIKACIJSKI SUSTAV ZA DETEKCIJU RAKA PLUĆA KORIŠTENJEM KARAKTERISTIKA KATASTROFE

Ramaiah Arun[1],Shanmugasundaram Singaravelan[2]

Department of Computer Science and Engineering, PSR Engineering College, Sivakasi, India[1]; Department of Computer Science and Engineering, PSR Engineering College, Sivakasi, India[2]

Odjel za računalne znanosti i inženjerstvo, PSR Engineering College, Sivakasi, Indija[1]; Odjel za računarstvo i inženjerstvo, PSR Engineering College, Sivakasi, Indija[2]

Abstract

One of the biggest challenges the world face today is the mortality due to Cancer. One in four of all diagnosed cancers involve the lung cancer, where the mortality rate is high, even after so much of technical and medical advances. Most lung cancer cases are diagnosed either in the third or fourth stage, when the disease is not treatable. The main reason for the highest mortality, due to lung cancer is because of non availability of prescreening system which can analyze the cancer cells at early stages. So it is necessary to develop a prescreening system which helps doctors to find and detect lung cancer at early stages. Out of all various types of lung cancers, adenocarcinoma is increasing at an alarming rate. The reason is mainly attributed to the increased rate of smoking - both active and passive. In the present work, a system for the classification of lung glandular cells for early detection of Cancer using multiple color spaces is developed. For segmentation, various clustering techniques like K-Means clustering and Fuzzy C-Means clustering on various Color spaces such as HSV, CIELAB, CIEXYy and CIELUV are used. Features are Extracted and classified using Support Vector Machine (SVM).

Sažetak

Jedan od najvećih izazova s kojima se svijet danas suočava je smrtnost od raka. Jedan od četiri svih dijagnosticiranih karcinoma uključuje karcinom pluća, gdje je smrtnost visoka, čak i nakon tolikog tehničkog i medicinskog napretka. Većina slučajeva raka pluća dijagnosticira se u trećem ili četvrtom stadiju, kada se bolest ne može liječiti. Glavni razlog najveće smrtnosti zbog karcinoma pluća je nedostupnost sustava za „preskrining" koji može detektirati stanice raka u ranim fazama. Stoga je potrebno razviti sustav za predklinički pregled koji pomaže liječnicima da pronađu i otkriju rak pluća u ranim fazama. Od svih vrsta karcinoma pluća, adenokarcinom se povećava alarmantnom brzinom. Razlog se uglavnom pripisuje povećanoj stopi pušenja - i aktivnom i pasivnom. U ovom radu razvijen je sustav za klasifikaciju plućnih žljezdanih stanica za rano otkrivanje raka korištenjem više prostora u boji. Za segmentaciju koriste se razne tehnike klasteriranja na različitim prostorima boja kao što su HSV, CIELAB, CIEXYy i CIELUV. Značajke se izdvajaju i klasificiraju pomoću Support Vector Machine (SVM).

1. INTRODUCTION

Human body is made up of trillions of individual units called cells, in the usual case most cells make new cells to substitute older cells that age or become damaged. Sometimes this ordered process, of only making new cells while a person is growing, breaks down. When this happens cells can grow out of control and they form a lump or mass called a tumor (Cancer). Patients with symptoms of cough, hemoptysis or chest pain and radiologic abnormalities, whether or not indicative of lung cancer, there are several methods of further investigation. More specific diagnosis requires morphologic examination and analysis by cytological techniques or by biopsy has been pointed out by Koss & Melamed /**1**/. Samples obtained by the fiber optic bronchoscope also are very effective in the diagnosis and differential diagnosis of cancer. Brush specimens, obtained directly from the suspect lesion, often provide an excellent sample and exact information on the location of the disease.

To curb the mortality rate of lung cancer it is utmost important to detect the cancer at an early stage. This is possible only by a mass screening of all men with a higher than average risk for developing lung cancer Chest X-ray along with sputum examination has been tried as a screening method to detect the presence of lung cancer. But it was not found useful. Imaging techniques combined with sputum cytology will be of great value to detect early lung cancers and its precursor lesions. Cells of squamous carcinoma vary considerably in shape and size, are typically found in a background of inflammation and necrosis, and often assume a most bizarre appearance has been pointed out by Koss & Melamed /**1**/.

In the previous work, a semi automated system for classification of lung glandular cell for early detection of cancer using multiple color spaces was developed. Next different kernels of SVM are used and then classified the glandular cells. In order to improve the efficiency, an automated lung cancer detection system by the analysis of glandular cells in sputum cytology images using scale space features is developed.

2. PROPOSED SYSTEM

The proposed methodology for classification of lung glandular cells as malignant and benign consists of cellular region detection, nuclear segmentation, artifact rejection, feature extraction and classification. The input high resolution image is preprocessed to find the region of interest to localize the cellular region using maximization of determinant of Hessian which considerably reduces the processing time required for analyzing the slide.

Segmentation of cellular region to nucleus, cytoplasm and background is done using K-means clustering. The sputum cytology images are characterized by the presence of various biological entities which are not required for cell analysis. These biological artifacts are eliminated based on morphology and color parameters prior to further analysis which otherwise adversely affect the system performance. Unusual nuclear characteristics show the presence of malignancy in cells. So catastrophe point based feature extraction from nuclear region is done. The catastrophe points occur when local maxima or local minima interact with a saddle points in scale space image, resulting in a pair wise annihilation. The catastrophe features extracted are used for classification of nuclei as benign or malignant using Support Vector Machine. The glandular cells are affected by reactive changes which alter the nuclear pattern mimicking malignancy in benign cells but the cytoplasm remains intact. So the cytoplasmic regions are further analyzed for the color intensity values and finally the cell is classified as benign or malignant.

The algorithm step involved in developing the automated system is as follows:

Step 1: Image Acquisition: Sputum Cytology Slide preparation

Step 2: Scale space – based cell region detection used for localizing the cellular region

a) Use Determinant of Hessian-based region localization to find Region of Interest (ROI)

Step 3: Cell Nuclei Segmentation: To segment nucleus, cytoplasm, and Background

a) Rudin–Osher–Fatemi (ROF) method for image denoising to remove biological noises
b) Segmentation using K – means Clustering with three cluster (nucleus, cytoplasm, Background)
Step 4: Artifact rejection:
a) Red Blood cells (RBC) are eliminated using color intensity value
b) Histiocytes are removed based on morphology
Step 5: Feature Extraction: Catastrophe point based feature extraction
a) Create Scale Space Stack
b) Find Maxima, Minima and Saddle in all Scales
c) Connect the critical points in each Scales
d) Find the top of Critical point connection
e)Find Catastrophe points
Step 6: Classification of Lung Glandular cells as Benign and Malignant
a) Nuclei classification using SVM
b) Cell classification using color intensity on cytoplasm
The proposed system is shown in the figure

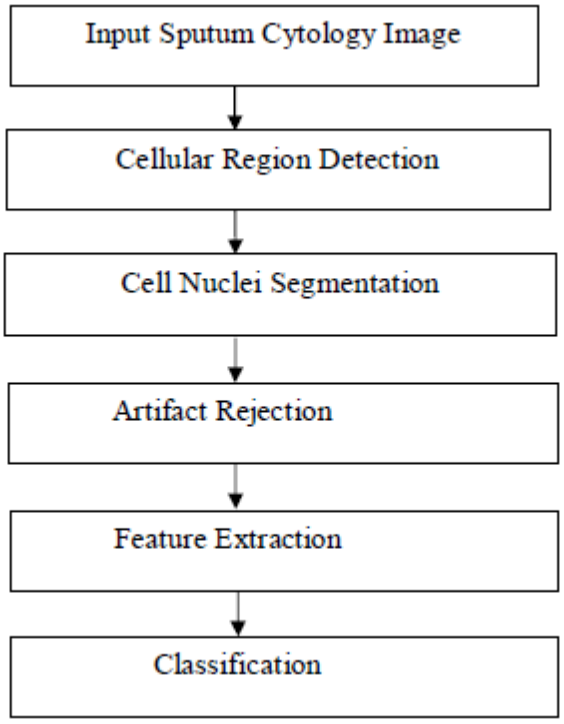


Figure 1: Sputum cytology image analysis system

2.1 Sputum Cytology

The surface of internal organs is usually covered by two types of cells (epithelium) squamous and columnar. The squamous cells are for giving protection to the organs while columnar or glandular cells are associated with secretions. In this work, adenocarcinoma is considered which the cancer affecting glandular epithelium. The mucus secreted by the goblet cells is distributed throughout the respiratory epithelium which traps dust particles which happen to be inhaled. The foreign particles entering the body are engulfed by a certain type of cells called phagocytes. The presence of phagocytic cells is typical of sputum cytology.

2.2 Scale Spaced Based Cell Region Detection

The high resolution (8 megapixels) input sputum cytology image is analyzed to find the cellular region. The cells in the sputum image are found scattered and scarcely populated with lot of background regions. Localizing the cellular region, subsequent processing can be restricted to those regions alone. As the first step, the input image is down sized to 408 X 306 to facilitate for faster computation. The sputum cytology

image consists of cellular materials of varying size, so the region localizing or cropping mechanism has to be size adaptive. Scale space based determinant of Hessian for finding the region and its corresponding crop size is used as given by Bay et al. /**2**/.

2.3 Cell Nuclei Segmentation

All cellular materials consist of two major parts- nucleus and cytoplasm. Nucleus contains chromosomes and is the main controlling part while cytoplasm acts as an envelope. For a benign cell, the cellular materials form a well defined pattern such as there is a uniform distribution of chromatin in the nuclei. The shape of nuclei is regular and usually follows elliptical or nearly elliptical shape. The presence of well distinguished cytoplasm is typical of benign cells. At the same time malignant cells are characterized by irregular nuclear texture due to DNA amplification. There may be multiple nuclei in same cell and are unusually large in size. Presence of scanty cytoplasm is another feature of malignant cell.

Segmentation is the process of grouping together regions following similar features such as color, texture, shapes etc. When segmenting, these regions results in over segmentation, which is to be eliminated. Otherwise the whole task of segmenting the image to nucleus, cytoplasm and background region is affected. To eliminate this, image is smoothened prior to segmentation. Smoothing has the adverse effect of removing edge regions which are the indication of regional boundaries; in this case it is extremely important. Here the edge preserving smoothing using total variation based Rudin-Osher-Fatemi(ROF) method and segmented using K- means clustering are used.

2.4 ROF Method for Image Denoising

The sputum cytology slides are degraded by several factors. The manual handling of slides may affect the quality of image acquired. Biological noises caused by mucus as well as non biological noises like dust can adversely affect the slide quality. ROF method effectively denoises the image while preserving sharp boundaries. The primary objective is to preserve the sharp boundaries which demarcate nucleus, cytoplasm and background. While preserving boundary it is also needed to smoothen other regions so that those regions form individual segments later, since smoothing increases the intensity of low intensity regions while decreasing that for high intensity regions.

2.5 Artifact Rejection

The sputum cytology images are subjected to high rate of artifacts both biological and physical. These artifacts have to be rejected for ensuring the overall efficiency of the system. The biological materials such as Red Blood Cells (RBC), histiocytes and nuclei located at the boundary of image are considered as artifact and are eliminated in this work. The nuclear region exhibits dark blue color as opposed to RBC which is reddish in color. Though the RBC appear in red, there is a significant contribution from other color channels also. It was observed that the color values of RBC range from {170,101,132} to {210,162,160}. To remove RBC, the mean red, blue, and green values in the segmented region are normalized. If the contribution of the red channel is above 38 % of all the color intensities, it is considered as RBC and removed.

The segmented result of nuclear region generally exhibits convex shape but histiocytes are seen as highly concave shapes and are eliminated based on morphology. This is done by finding the ratio between the segmented area and the area corresponding convex hull. If the ratio is less than 90 % then the region is considered as histiocyte and is removed from further analysis.

2.6 Catastrophe Point Based Feature Extraction

This is the most crucial part of the system where the scale space features are extracted. Scale space deals with the observation or study of image at various scales. It is about describing an image according to the size/scale of objects present in the image Catastrophe points or top points were successfully used for image matching and was pointed out by Platel

et al. /**3**/ and reconstruction was given by Kanters et al /**4**/.

2.7 Scale Space

In the physical world every object has the corresponding property which occurs at certain scale of observation only. The scale space theory says that the analysis of a particular object in an image should be done at the proper scale and was pointed out by /**5**/. Linear scale space was earlier thought to be the unique kernel which satisfies the following axioms and was pointed out by Florack et al ./**6**/.

1. Linearity, i.e., there is no knowledge of any kind
2. Spatial scale invariance, i.e., all objects whether small or large is treated equally
3. Spatial homogeneity, i.e., no preference is given to any particular location
4. Spatial isotropy, i.e., all orientations are treated equally

It was shown that if separability is not a criteria then there exist infinite linear scale spaces and was pointed out by Duits et al. /**7**/.

2.8 Deep Structure Analysis

The scale space stack is generated and stored for further analysis. One of the most important feature is the automatic detection of the scale of singularities like edges, blobs etc. was given by /**5**/. In general, deep structure analysis shows that the singularities decrease as we move up the scale space stack and was pointed out by Romney ,2003. This analysis leads to the well known catastrophe theory and has been pointed out by Florack & Kuijper /**8**/.

2.9 Finding Catastrophe Points

Maxima, minima and saddle points are found on all the scales using Blom's method. Here hexagonal neighborhood is considered for processing. In practical case this turns out to be taking the 8 neighborhood and discarding top right and bottom right neighbors for even rows and discarding top left and bottom left elements for odd rows. The hexagonal neighborhood such as Center pixel (black), hex neighborhood (medium gray), excluded pixels (light gray) from 8 neighborhoods for odd (left) and even (right)For these six neighbors the sign of the difference with center pixel is found. This leads to four cases as shown below:

1. No sign change, the given point is an extrema.
2. Two sign changes, the point is a regular point.
3. Four sign changes, the point is a saddle.
4. Six sign changes, the point is a monkey saddle.

2.10 Feature Extraction

Scale space stack is generated for 32 scales ranging from e1 to e3. The intermediate scales are chosen in such a way that the ratio between any two successive scales is the same /**5**/. The count for catastrophe points, both saddle-maxima and saddle-minima, occurring for each scale is taken and the feature vector is generated thus giving a size of 64 dimension vector. The feature extraction method is compared with other texture-based features like Gabor filter-based gray-level co-occurrence feature, local binary pattern (LBP) and Complex Daubecchies wavelet (CDW)-based features

2.11 Classification Technique

Classification forms the final stage in the proposed system. Here the individual cells are classified as benign or malignant. The lung cells being constantly in contact with external environment through the air we breathe are affected by reactive changes. The cells which had undergone reactive or inflammatory changes have unusual nuclei, with texture and size mimicking that of malignant cells. For this, the color property of cytoplasm is also taken. The cell classification is done in two stages. In stage one, the nuclei is classified as benign or malignant using SVM and in stages two, those nuclei which are classified as malignant is

analyzed along with cytoplasm for color intensity values.

2.12 Nuclei Classification Using SVM

SVM is an example of supervised learning method in which we provide the feature set with corresponding label of the class which it belongs. During training stage, the SVM constructs a hyper plane that separates the two input classes. The hyper plane is selected in such a way that the separation between the two classes is maximum. SVM has the advantage that the classification depends on the number of support vectors only and not on the dimension of the input vectors.

SVM allows for a training error in which to maximize the margin some misclassification is allowed which results in better classification during testing. For high dimensional data, linear separation may not be possible to achieve. To facilitate nonlinear mapping, kernel trick is applied in which the input data point is mapped to a higher dimension using dot product of input features.

In this work, radial basis function (RBF) kernel is used. The training set consists of manually cropped 100 benign and 100 malignant cells. The testing set, separate from the training set, consists of 100 benign and 100 malignant images with cell count varying from 5 to 90. By trial and error, it was found that sigma value of one is optimal for RBF kernel using the current set of features.

3. RESULTS AND DISCUSSION

The region localization result, segmentation results, artifact rejection results, feature extraction results and classification results of the processing steps are discussed in the following sections.

3.1 Region Localization Results

The input microscopic sputum cytology image is of resolution 3264 X 2448, which takes considerable amount of time to process. From the experimental dataset of 100 benign and 100 malignant images, 2571 regions were detected. The average area saved for benign and malignant set of images is 74.23% and 91.27% respectively, which is shown in Table 1.

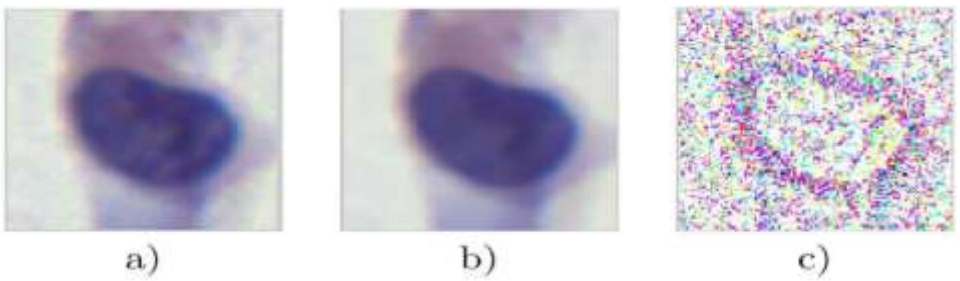
Image Set	Area saved%
Benign	74.23
Malignant	91.27

Table 1: Average Area Saved by Preprocessing

3.2 Segmentation Result

The detected region of interest images go to the segmentation module where the morphological and color features of the detected regions are analyzed. The input noisy image is denoised using ROF method. Total variation ROF smoothening is done on three color channels of Red, Green and Blue. Based on the experimental results, it is found that no significant change happens to the image after 125 iterations. In Figure 2, set one shows the segmentation result using K-means clustering, set two, three and four shows the segmentation result for artifact, benign and malignant cases using Chan-Vese method with an initial circular mask of radius 10 pixels, 25 pixels, and 60 pixels, respectively. Set five of Figure 2(a), (b), (c) shows the experimental result with a mask of 36 circles of radius 8 pixels each.

Figure 2: Experimental result with a mask of 36 circles of radius 8 pixels each.



3.3 Artifact Rejection Results

Artifact rejection module rejects RBC, histiocytes and partial nuclei regions. 2571 regions are detected in the image from which 690 nuclear regions are filtered and the rest are eliminated from further analysis. Sample result of histiocyte and RBC.

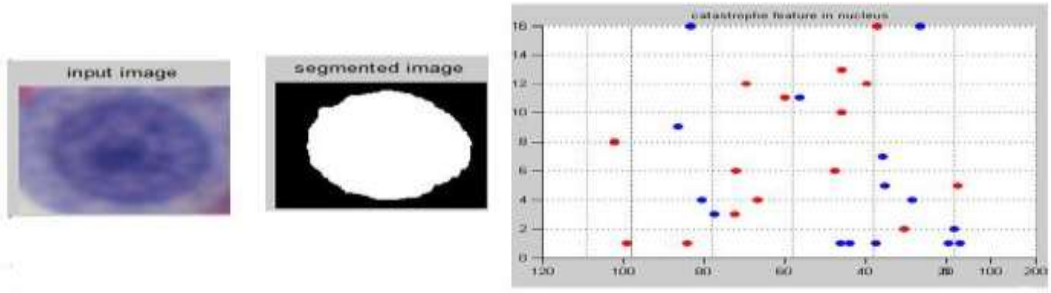


Figure 3: Sample Malignant case 1 with segmented image and catastrophe points.

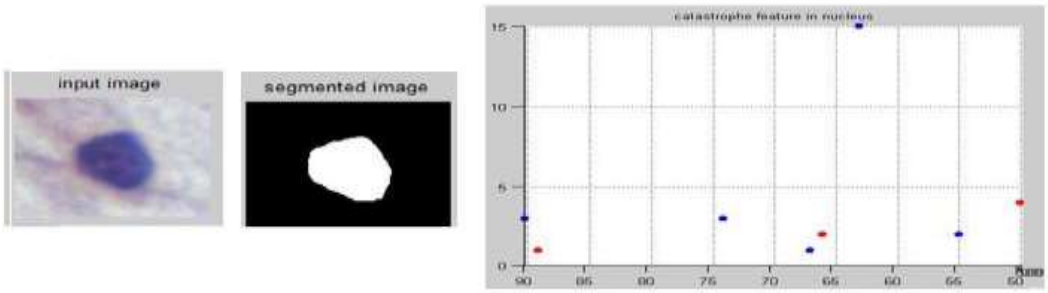


Figure 4: Sample Benign case 2 with segmented image and catastrophe points.

Notes

/1/ Koss L.G., Melamed, M.R. (2006), Koss's diagnostic cytology and its hystopathologic bases, Lippincott Williams and Wilkins, New York.

/2/ Bay, H., Ess, A., Tuytelaars, T., Van Gool, L. (2008), 'Speeded-Up Robust Features (SURF)', Computer Vision and Image Understanding,vol. 110, pp. 346-359.

/3/ Platel, B., Kanters, F.M.W., . Florack, L.M.J. & Balmachnova, E.G. (2004), 'Using multiscale top points in image matching', Proceedings.

/4/ Kanters, F., Lillholm, M., Duits, R., Janssen, B., Platel, B., Florack, l., Haar Romeny, B. (2005,)'On Image Reconstruction from Multiscale Top Points, in Lecture Notes in Computer Science, vol. 3459, pp. 431-442.

/5/ Lindeberg, T. (2000), 'Scale-space theory: A basic tool for analysing structures at different scales', Journal of applied statistics, vol. 21, no. 2, pp. 224-270.

/6/ Florack, L.M.J., Romeny, B., Koenderink, J.J., Viergever, M.A. (1994),,'Linear scale-space", Jounal of Mathematical Imaging and Vision', vol. 4, no. 4, pp. 325-351.

/7/ Duits, R., Florack, L.M.J., De Graaf, J., Romeny, B. (2004), 'On the Axioms of Scale Space Theory", Jounal of Mathematical Imaging and Vision, vol. 20, no. 3, pp. 267-298.

/8/ Florack, L.M.J, Kuijper, A, (2000), 'The Topological Structure of Scale- Space Images', Journal of mathematical imaging and vision, vol. 12, no. 1, pp. 65-79.