Segmentation of Sputum Cell Image for Early Lung Cancer Detection

N. Werghi, C. Donner, F. Taher and H. Al-Ahmad
Department of ECE
Khalifa University, Sharjah, UAE
Outline

- Motivation
- Research challenge and proposed approaches
- Desirable Outcomes and Deliverables
- Previous Work
- Thresholding classifier
- Bayesian classifier
- Mean shift segmentation
- Testing & Evaluation
- Conclusions & Future Work
Motivation

- Leading cause of cancer death throughout the world.
- 7.6 million deaths worldwide each year.
- Mortality from cancer is expected to be around 17 million worldwide in 2030.
- Medical image are used in detecting the cancerous cells.
- Very large amount of data are produced from medical imaging techniques, such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI) and Breast Thermograph.
Diagnosis

- Lung Cancer is difficult to detect early because symptoms only appear at advance stages.
- Physicians use several techniques to diagnose lung cancer, such as:
  - Chest Radiograph, a standard chest X-ray.
  - Computerized Tomography (CT) scans.
  - Sputum Cytology.
Detection

Diagnose
Sputum Cytology: Early detection

- Microscopic image of stained sputum cell
- Simple and safe procedure
- Morphological and chromatic properties of the nuclei of the sputum cells with respect to the cytoplasm
Research Challenges and proposed approaches.

- Background subtraction and isolation of the regions of interest.
- Cell segmentation: Extraction of the nuclei and the cytoplasm
- Classification: Twofold problem:
  - What shape descriptors must be extracted from the nuclei shape.
  - How to use these descriptors for discriminating between normal and abnormal cells.
Desirable Outcomes and Deliverables

- The eventual objective is to design and to develop an automatic diagnostic system for detecting lung cancer in its early stages based on the analysis of the sputum color images.
- The target is an overall successful classification rate above 95% where by false negative rate is reduced to its minimum level.
- The system should be sufficiently efficient to make a decision within an interactive time frame.
Computer Aided Diagnosis (CAD) System for Lung Cancer Detection

Digital camera

Light Microscope

Sputum Specimens Stained

Image Capture

Segmentation of Sputum Color Images

Analysis of the segmented Images

Formulation of Diagnostic rules

Testing and Evaluation
Diagnosis Process of CAD system

- Preprocessing-Sputum Cell Extraction
- Segmentation-Detection of Nuclei and Cytoplasm
- Feature Extraction-Geometric Feature, Chromatic Feature
- Classification
- Diagnosis Result
Previous work

Lung Cancer Detection Modalities

- Lung cancer Detection
- Sputum Analysis
- Other Modalities

Lung Cancer Detection Modalities

- Neural Network
- Fuzzy C-Mean Clustering
- Thresholding Techniques
- Edge detection
- Region detection

CT Scan

Region Growing
- Active Contour Model
- Thresholding
- K-mean clustering
- Bit-plane slicing technique
- Genetic optimization
- Neural Network

Other Modalities

Adaptive Threshold
- Maximum Likelihood
- Hybrid Fourier-wavelet de-convolution
- Genetic algorithms
- Similarity coefficients
Medical Data

- Database of 2D sputum color images, which contains normal and abnormal cases (Tokyo center of Cancer)
- These sputum samples have been stained, by using Papanicalaou standard staining methods.
- Some of the sputum nuclei cells are overlapping due to the dispersion of the cytoplasm in the staining process.
Sputum Color Images
Contribution

- Detection of sputum cell using a thresholding technique and a Bayesian classification framework.
  - Best color space
- Mean shift technique for the sputum cell segmentation.
Sputum Cell Detection

- Segmenting the sputum image into:
  - Sputum cell (region of interest).
  - Background
Sputum Cell Detection: Previous Work

- Sammouda 98, Taher 2010
- Heuristic threshold-based rules based on the chromatic disparity
- Threshold: Trial and error
Sputum Cell Detection Methods

• Two detection methods:
  – Threshold technique
  – Bayesian classification
Thresholding Filtering Algorithm

- The nuclei of the sputum cells are extracted using the following:

\[
\text{if } ((2 \cdot G(\chi, y) + \theta) < (R(\chi, y) + B(\chi, y))) \text{ then } (G(\chi, y) = 0)
\]

- To remove the debris cells the following equation is used:

\[
\text{If } \left( R(\chi, y) < (G(x, y) + \Theta) \right) \text{ then } (R(x, y) = 0)
\]
Threshold classifier Results
Sputum Cell Detection: Bayesian classification: 2 classes: background, Cell

- A pixel $x$ is considered part of the of sputum cell if

$$p(bg|x) < p(sp|x)$$

$$f(\lambda, p(sp), p(bg)) < \frac{p(x|sp)}{p(x|bg)}$$

$p(sp), p(bg) \leftarrow$ Training data

$P(x/sp), p(x.bg) \leftarrow$ normalized histograms

Choice of $\lambda$

- False negatives $\downarrow \Rightarrow \lambda > 1$
256-RGB Histogram Visualization

256- RGB Histogram visualization for the sputum and non-sputum pixels
Bayesian classification results

RAW IMAGES | GROUND TRUTH | $\lambda=2$ | $\lambda=7$
Testing and Evaluation

- From our collection of sputum color images, ground truth images were made manually.
- Ground truth images are used in the comparison with the output images from classifications algorithm.
- Different evaluation criteria such as precision, accuracy and specificity are used to evaluate the results.
Evaluation Criteria

- For the performance measurement, the True positive (TP's), False positive (FP's), True negative (TN's) and False negative (FN's) are calculated:
  - **TP's**: pixels, that were correctly classified as sputum pixels.
  - **FP's**: pixels, that were mistakenly classified as sputum pixels.
  - **TN's**: pixels, that were correctly classified as non-sputum pixels.
  - **FN's**: pixels, that were mistakenly classified as non-sputum pixels.
The evaluation Criteria

Precision = \( \frac{TP}{TP + FN} \)

Specificity = \( \frac{TN}{TN + FP} \)

Accuracy = \( \frac{(TP+TN)}{(TP+TN+FP+FN)} \)

- Color spaces: RGB, HSV, YCbCr, Lab
- Histogram resolution: 16, 32, 64, 128, 265
Colors sapce performance  Resolution 256
Colors sapce performance Resolution 64

![Graph showing true detection rate vs. false detection rate with different color spaces: rgb, ycbcr, hsv, and lab. The graph illustrates the performance of each color space under varying false detection rates.]
Colors sapce performance  Resolution 16
RGB performance: 16, 32, 64, 128, 256
ROC-Curve for thresholding classifier

ROC curve of RGB space for the thresholding classifier
Accuracy: resolution 64
Accuracy: resolution 128
Accuracy: resolution 256
Results

- HSV > RGB > YCbCr > Lab
- Resolution > 64

<table>
<thead>
<tr>
<th>Performance Measurements</th>
<th>Thresholding Algorithm</th>
<th>Bayesian Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>82%</td>
<td>89%</td>
</tr>
<tr>
<td>Specificity</td>
<td>99%</td>
<td>99%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>98%</td>
<td>98.7%</td>
</tr>
<tr>
<td>TP</td>
<td>82%</td>
<td>89%</td>
</tr>
<tr>
<td>FN</td>
<td>18%</td>
<td>11%</td>
</tr>
<tr>
<td>FP</td>
<td>31%</td>
<td>28%</td>
</tr>
</tbody>
</table>
Cell Segmentation

- Cell $\rightarrow$ nucleus and cytoplasm
- Previous method
  Hopfield NeuralNetwork (HNN)
  classes: 3
    background (black), nucleus, and cytoplasm
- Issue: variability of the nucleus and the cytoplasm
  number of classes $> 3$
Cell Segmentation

- Alternatives
  - Active contours
    - Issue: nucleus: non homogenous
    - Instability
  - Mean shift-based segmentation
Mean shift

- Non-parametric iterative technique
- Density function in the feature space
  - Two variants
    - Feature space: color features
    - Feature space: color and spatial features
Mean shift algorithm

1. Segment the feature space into regions (modes)
2. Choose the initial location of the modes.
3. Compute the new locations of the modes by updating them using gradient-based minimization of the density function
4. Repeat step 3 until convergence (shift step tends to zero).
5. Merge the neighbouring modes and their associated pixels.
Mean Shift-based Segmentation

- Gray level segmentation
- Histogram equalization
- Mean shift
- Refinement
- Modes merging
Mean Shift-based Segmentation
## Results

<table>
<thead>
<tr>
<th>Performance</th>
<th>HNN</th>
<th>Gray Mean Shift</th>
<th>Gray-Space Mean Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>37.5%</td>
<td>59.41%</td>
<td>60.61%</td>
</tr>
<tr>
<td>Specificity</td>
<td>61.64%</td>
<td>80.69%</td>
<td>86.60%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>63.84%</td>
<td>81.01%</td>
<td>85.51%</td>
</tr>
</tbody>
</table>
Conclusions & Future work

- We presented methods for the extraction and segmenting the sputum cells for the purpose of lung cancer early detection.
- Thresholding technique and Bayesian classification are presented for cells detection.
- Bayesian classifier outperforms the thresholding classifier, It allows elegant and methodological determination of the classification parameter.
- Mean Shift will be used for sputum cell segmentation, its outperforming the HNN technique.
- As future work: continue the next stage of the work, the elaboration and extraction of appropriate descriptors of the nucleus and the cytoplasm.
References


References


Thank you