From Cell Image Segmentation to Differential Diagnosis of Thyroid Cancer

S.Ablameyko*, V.Kirillov**, D.Lagunovsky*, O.Patsko*, N.Paramonova*, M.Petrou***, O.Tchij

* Institute of Engineering Cybernetics, National Academy of Sciences, 6, Surganov Str, 220012, Minsk, Belarus
** Research and Clinical Institute of Radiation Medicine and Endocrinology, Philimonova str., 23, 220114 Minsk, Belarus
*** University of Surrey, Guildford, GU2 7XH, United Kingdom

Abstract

An approach for cytologic diagnosis of thyroid cancer with the help of automatic morphometry is proposed. This approach is based on a developed computer analyser of images that is aimed at: automated processing and binarization of colour images; automatic raster-to-vector transformation and formation of biological objects; morphometric assessment of biological objects by quantitative parameters characterizing the changes of cell nuclei; building of an expert system to aid the diagnosis of thyroid cancer.

1. Introduction

In differential cytologic diagnosis of thyroid tumors, the role of generally accepted criteria of malignancy, which are based on the whole complex of quantitative indices of cell abnormality, is limited. This leads to low frequency of thyroid cancer recognition at early stages and necessitates the development and adoption of new, more effective methods of oncological diagnosis.

One of the approaches to solving this problem is the transformation of qualitative indices of pathological changes in cells into a quantitative form with the help of a computerized morphometry method. Much attention has been given to this question in the scientific literature. In particular, it has been shown that the use of thyrocyte nuclear area values allows one to differentiate adenoma from cancer [1-2].

The first morphometry methods were based on manual techniques that are not very effective. Automated cytological image analysis techniques should give more benefits. The automated procedure should be able to analyse cytologic images, extract all features about cells, nuclei, aggregates and compute parameters to propose/verify diagnosis.

There are four main tasks/steps in this procedure: a) image segmentation, b) vectorisation and morphometric database building, c) karyometric parameter selection and building of the expert system, and d) thyroid cancer diagnosis.

Many attempts have already been made to develop algorithms for segmentation of biomedical images. They are based on a wide range of mathematical tools, namely: mathematical morphology, snakes, Fourier and Hough transforms, neural networks, and others [4-7]. An overview of cell image segmentation techniques can be found in [8]. However, due to the very complex nature of cytologic images, it is not always possible to select or develop automatic segmentation methods that could be applied to extract objects correctly. This is why alongside the automatic cell extraction algorithms we also developed interactive editing tools.

The next task is selection of parameters that can be computed for cells and nuclei from a segmented image and used for diagnosis. The first attempt to select parameters from medical point of view was made in [3].

In this paper, we propose an approach to cytologic diagnosis of thyroid cancer with the help of morphometry methods. This approach is based on a developed computer analyser of cytologic images.

2. Color Image Segmentation

We developed several algorithms for color cytological image segmentation.

2.1. Analysis of color images by setting bounds in the color space

This algorithm is based on the assumption that image areas corresponding to cell nuclei usually correspond to a definite subspace of the RGB color space. In this case, building such a subspace within the whole RGB space gives us a basis for image segmentation.

When the color coordinates of a pixel place it within the “nucleus” subspace, the pixel is assumed to be from a nucleus. Otherwise, the pixel is considered to be from the background. To obtain information for the construction of such a subspace (we assume it to be a parallelepiped) we
use a sample image with manually segmented nuclei. From the analysis of the sample image, the minimal and maximal values of the R, G and B components of the nuclei pixels were determined and were used to construct the subspace. Experiments showed that we can set the minimal values of these components to zero – it will not influence the result of segmentation. An example of cell segmentation is shown in Fig.1.

![Fig.1. Initial cytological image (a) and result of segmentation by cutting parallelepiped in color space (b).](image)

The advantage of this method is its simplicity and the use of “global” parameters (the same for all images).

2.2. Color analysis in the HSV space

In contrast with the previous algorithm, this algorithm uses only one color component for segmentation. Information about “pure” color is extracted from the image by transforming it into the HSV color space. Here, the H (Hue) component can be regarded as information about “pure” color, without brightness, etc. Once this component is obtained, the image is segmented by histogram analysis. One would expect these histograms to have two peaks, corresponding to the nuclei and the background. However, the resulting histogram has a very complex shape, which significantly varies from image to image. That is why threshold detection is difficult.

An example of image segmentation by this algorithm is shown in Fig.2. The advantage of the algorithm lies in the use of “pure” color that makes it more universal and less dependable on color variations.

![Fig.2. Result of image segmentation by transformation in the HSV color system](image)

3. Morphometric database building

Vector contours are automatically extracted from the obtained binary image (noise filtration can be performed if necessary). Algorithms for noise reduction, contour extraction and vectorisation are based on one image scan and are described in [9].

One of the major tasks is the correct extraction of cell contours especially for overlapping cells. Cell contour splitting can be done at raster representation (by using for example methods of mathematical morphology) or at vector representation. We have chosen the second approach to avoid cell shape changes. The algorithm is based on 1) analysis of contour curvature and extraction of merging places, 2) contour splitting, 3) separate cell contour reconstruction (Fig.3a).

![Fig.3. Cell contour splitting by automatic (a) and interactive (b) techniques.](image)

Then, parameters for every cell are calculated. These include: cell area, perimeter, shape factor, diameter, (mean and max), elongation coefficient, circularity factor. The obtained feature vectors for each object alongside any semantic information are further used to build the morphometric database.

4. Karyometric parameter selection and expert system building

Morphometric study of operative material smears taken from patients with various thyroid pathology (all the
cases were histologically verified) showed that individual values of mean thyrocyte nuclear area varied rather widely. So, individual values of mean thyrocyte nuclear area in the group of malignant pathology varied within the interval of 91.1 – 145.4; while in the group with benign pathology, they ranged from 57.1-100.1 µm², i.e. within the interval of 91.1-100.1 µm² overlapping of this parameter values occurred. This uncertainty zone made up 10.2% of the total range of individual values of thyrocyte nuclear area for both groups of thyroid pathology.

Quantitative changes in the nuclear area of thyrocyte operative material in the three classes were most clearly manifested in histograms of area distribution. For the control group the histogram had the form of one column. For malignant pathology group, a dome-like distribution was observed (Fig.4). The area histogram for benign pathology group had an exponential form. The revealed regularities in the character of the distribution of the thyrocyte nuclear area can be expressed numerically as the ratio of frequency of occurrence in the 2nd and 1st histogram bins. This ratio was greater than 1 for malignant and less than 1 for benign cases. In the control group, this ratio was equal to 0. It should be noted that the number of filled bins in the histograms varied from 3 to 5 for malignant pathology, and from 2 to 4 for benign pathology. So, the cases whose number of filled histogram bins was 5, could be significantly related to the group of malignant pathology and those for which this number was 2 to benign. The uncertainty zone of this parameter was 50%.

Quantitative changes in the nuclear area of thyrocyte operative material in the three classes were most clearly manifested in histograms of area distribution. For the control group the histogram had the form of one column. For malignant pathology group, a dome-like distribution was observed (Fig.4). The area histogram for benign pathology group had an exponential form. The revealed regularities in the character of the distribution of the thyrocyte nuclear area can be expressed numerically as the ratio of frequency of occurrence in the 2nd and 1st histogram bins. This ratio was greater than 1 for malignant and less than 1 for benign cases. In the control group, this ratio was equal to 0. It should be noted that the number of filled bins in the histograms varied from 3 to 5 for malignant pathology, and from 2 to 4 for benign pathology. So, the cases whose number of filled histogram bins was 5, could be significantly related to the group of malignant pathology and those for which this number was 2 to benign. The uncertainty zone of this parameter was 50%.

Discriminant analysis allowed us to differentiate the studied groups more effectively. Density distribution curves of probable values of the discriminant functions for pairs of groups overlapped only insignificantly. For the pair malignant pathology-control the overlapped area was only 1%; for the pair benign pathology-control - 10% and for the pair malignant-benign pathology - 24%. The usefulness of the karyometric data in discriminating between the two forms of the disease was evaluated with the help of these discriminant functions. The results of thyrocyte nuclei classification by the decision rule for the pair malignant-benign disease showed that the proportion of cells classified as cancerous, according to the discriminant threshold ranged 65.3-99.3% for malignant disease and 2.7-62.3% for benign. Thus, cases with discriminant threshold more than 65.3% could be confidently classified as malignant and those with discriminant threshold less than 62.3% as benign.

The morphometrically revealed regularities of pathological changes in follicular cell nuclei of the operative material can be combined to form a set of quantitative parameters characterizing malignant and benign thyroid diseases. Table 1 gives this set of parameters including mean thyrocyte nuclear area, numerical characteristics of scattered histograms of their area, regression equation coefficient and the discriminant threshold. A set of boundary values of the karyometric parameters associated with malignant pathology given in this table constitutes an expert system for thyroid carcinoma diagnosis.

However, an expert system must have a unique output that performs the diagnosis of malignant pathology with different values for a, b and c for the different groups. The values of these coefficients can be used as class discriminators with P<0.05 for a and b, and P<0.01 for c.

Fig. 4. Histogram of distribution of nuclear areas.

Construction of scattered diagrams of thyrocyte nuclear perimeter and area of operative material (Fig.5) shows that the clusters of points characterizing the control, malignant and benign groups, considerably overlapped one another. The set of these experimental points was described well by the second order regression dependence with correlation coefficient more than 0.990. The regression equations had the form $y=ax^2 + bx +c$, with different values for a, b and c for the different groups.
with a certain degree of confidence. Each of the parameters used by the expert system was assigned a weight according to its ability to distinguish between the different nosologic forms according to the formula:

$$k_i = \frac{A_i + Sn_i + Sp_i}{\sum_{j=1}^{n}(A_j + Sn_j + Sp_j) \times 100}$$

where $k_i$ is the weighting coefficient of the $i$-th karyometric parameter; $A$ its accuracy; $Sn$ its sensitivity; $Sp$ its specificity and $n$ is the total number of karyometric parameters (in our case, $n=6$).

The output of the expert system is a diagnostic index defined as the weighted sum of the values of the parameters that fall within the range of values for malignancy as stored by the expert system.

5. Results and discussion

Thyroid cells of autopsy, operative and biopsy material were used for system testing. 60 touch smears with histologically verified diagnosis – 10 for each nosological form of a disease and 10 controls have been taken for system training. Then, 82 aspirate smears: papillary carcinoma – 27, follicular carcinoma – 10, follicular adenoma - 12, autoimmune thyroiditis – 12, nodular colloid – 11 and diffuse toxic goiter – 10 have been used to test the system performance. Operation and aspiration materials were obtained from different patients and at different clinics.

Aspiration cytologic material from each case was processed computationally and its karyometric parameters were evaluated. Their values were then compared with the table data of the expert system (Table 1) and the values of those falling within the limits for malignancy were linearly combined using the corresponding weights to form the diagnostic index. Of the 37 malignant cases, for 11 the diagnostic index was 50, for 7 cases - 70, for 3 - 80, for 8 - 91 and for 8 - 100, i.e. it ranged from 50 to 100%. At the same time, for all 45 cases of benign pathology, the diagnostic index was equal to 0.

6. Conclusion

We presented a complete system for the diagnosis of malignant thyroid cancer. The output of the system is a diagnostic index that expresses the confidence in diagnosing malignancy. Using a set of karyometric parameters allows us to reach high reliability of the system in distinguishing malignancy from benign pathology.

Acknowledgment

This research was supported by the International Science and Technology Center (Project B-323).

References


Table 1. Expert System for Thyroid Cancer Diagnosis

<table>
<thead>
<tr>
<th>Thyroid pathology group</th>
<th>Mean area, µm²</th>
<th>Histograms of area distribution</th>
<th>Regression equation coefficients</th>
<th>Discriminant threshold, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean area, µm²</td>
<td>Ratio of frequencies</td>
<td>Number of classes</td>
<td>b</td>
</tr>
<tr>
<td>Malignant</td>
<td>&gt;100.1</td>
<td>&gt;1</td>
<td>5</td>
<td>&lt;0.263</td>
</tr>
</tbody>
</table>