

Histopathologic patterns of nervous system tumors based on computer vision methods and whole slide imaging (WSI)

Slawomir Walkowski^{a,*} and Janusz Szymas^b

^a*Poznan University of Technology, Poznan, Poland*

^b*University of Medical Sciences, Poznan, Poland*

Abstract. *Background:* Making an automatic diagnosis based on virtual slides and whole slide imaging or even determining whether a case belongs to a single class, representing a specific disease, is a big challenge. In this work we focus on WHO Classification of Tumours of the Central Nervous System. We try to design a method which allows to automatically distinguish virtual slides which contain histopathologic patterns characteristic of glioblastoma – pseudopalisading necrosis and discriminate cases with neurinoma (schwannoma), which contain similar structures – palisading (Verocay bodies).

Methods: Our method is based on computer vision approaches like structural analysis and shape descriptors. We start with image segmentation in a virtual slide, find specific patterns and use a set of features which can describe pseudopalisading necrosis and distinguish it from palisades. Type of structures found in a slide decides about its classification.

Results: Described method is tested on a set of 49 virtual slides, captured using robotic microscope. Results show that 82% of glioblastoma cases and 90% of neurinoma cases were correctly identified by the proposed algorithm.

Conclusion: Our method is a promising approach to automatic detection of nervous system tumors using virtual slides.

Keywords: Whole slide imaging, computer vision, pattern recognition, structural analysis, shape descriptors

1. Background

The challenge of making an automatic diagnosis based on virtual slides is a background of this work. This is a wide and complex issue to design an algorithm which makes a diagnosis based on digital images. Therefore, we focus on detecting specific diseases. Our work is based on WHO Classification of Tumours of the Central Nervous System. Our goal was to design, implement and test an algorithm which detects glioblastoma and neurinoma in virtual slides and discriminates between these cases. The method will classify a virtual slide as belonging to one of these classes: ‘Glioblastoma’, ‘Neurinoma’ and ‘Nothing’.

The last class is selected when neither glioblastoma nor neurinoma was detected. Fragments of exemplary virtual slides belonging to each class are shown in Fig. 1.

To classify these cases we use patterns – characteristic structures in the slides [1]. Multiple patterns can be used to detect glioblastoma and neurinoma. In our work we focus on typical pseudopalisading structures for glioblastoma and palisading structures for neurinoma. To be able to find these patterns and distinguish them – either by a pathologist or an algorithm – we need description of their features. They are gathered in Table 1.

2. Materials

Our method was tested on a set of 49 virtual slides. They were captured using robotic

*Corresponding author: Slawomir Walkowski, Poznan University of Technology, Poznan, Poland. Tel.: +1 425 753 8663; E-mail: E-mail: slawowski@gmail.com.

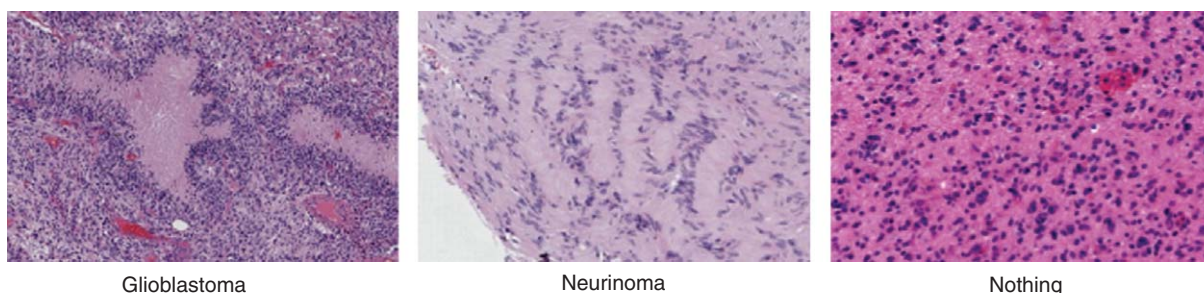


Fig. 1. Fragments of virtual slides belonging to three different classes: 'Glioblastoma', 'Neurinoma' and 'Nothing'.

Table 1

Description of pseudopalising and palising patterns

Pseudopalising (glioblastoma)	Palising (neurinoma)
Areas of higher nuclei density, adjacent to necrotic areas	Areas of higher nuclei density, forming linear pattern
Curves and meanders	Straight lines
Curves and bends, often closed and forming concave areas	Linear nuclei areas intermingled with cytoplasm

microscope Axioplan 2e (ZEISS) with objective Planapo 20×/0.75, equipped with Camera Axiocam HRc (ZEISS). Acquired 1300 × 1030 tiles with 15% overlap were stitched into single images (sizes about 8000 × 8000), resulting in 39 glioblastoma and 10 neurinoma cases.

Even if virtual slides are captured using higher magnification, pseudopalising and palising structures can be distinguished on lower zoom level. Therefore, virtual slides scaled to 25% of their original resolution were used by the algorithm. This is enough to find characteristic structures and allows the method to work faster.

3. Methods

The proposed algorithm is based on computer vision approaches: structural analysis and shape descriptors. It consists of a few common initial steps and two branches for handling pseudopalising and palising cases. Final result is calculated in the last step. A diagram for the algorithm is presented in Fig. 2 and its steps are described in the following paragraphs. A good description of structural analysis approaches can be found in OpenCV computer vision library documentation [2]. This library is also used in the implementation of the proposed method.

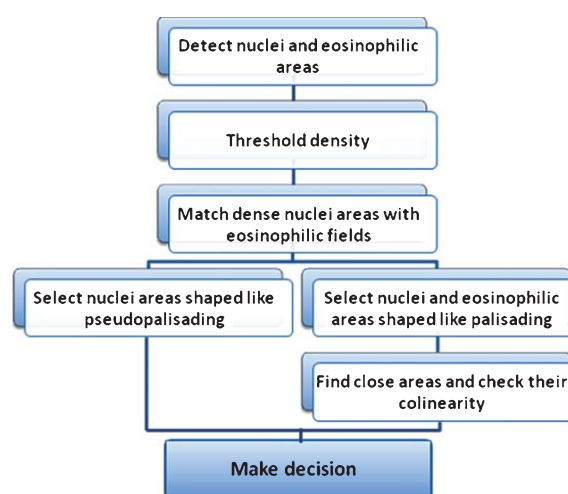


Fig. 2. Scheme of the algorithm proposed in this paper.

1. Detecting nuclei and eosinophilic areas. A virtual slide is converted to Hue-Saturation-Value color representation. It allows each pixel to be filtered according to specific HSV ranges and uses color information, which is more accurate than grayscale. This filtration leads to detecting nuclei and eosinophilic areas. We do not know the diagnosis in this and a few subsequent steps so we cannot say whether detected eosinophilic areas are necrotic areas or cytoplasm.

2. Thresholding density. Nuclei density is calculated as the average number of nuclei pixels in a square window around each pixel. Thresholding is done by selecting areas with density exceeding average density in the slide. Density of eosinophilic areas is calculated and thresholded similarly to nuclei density. However, for eosinophilic areas the result is additionally combined with inverted nuclei density to save only areas without nuclei. Examples of thresholded density of nuclei and eosinophilic areas are presented in Fig. 3.

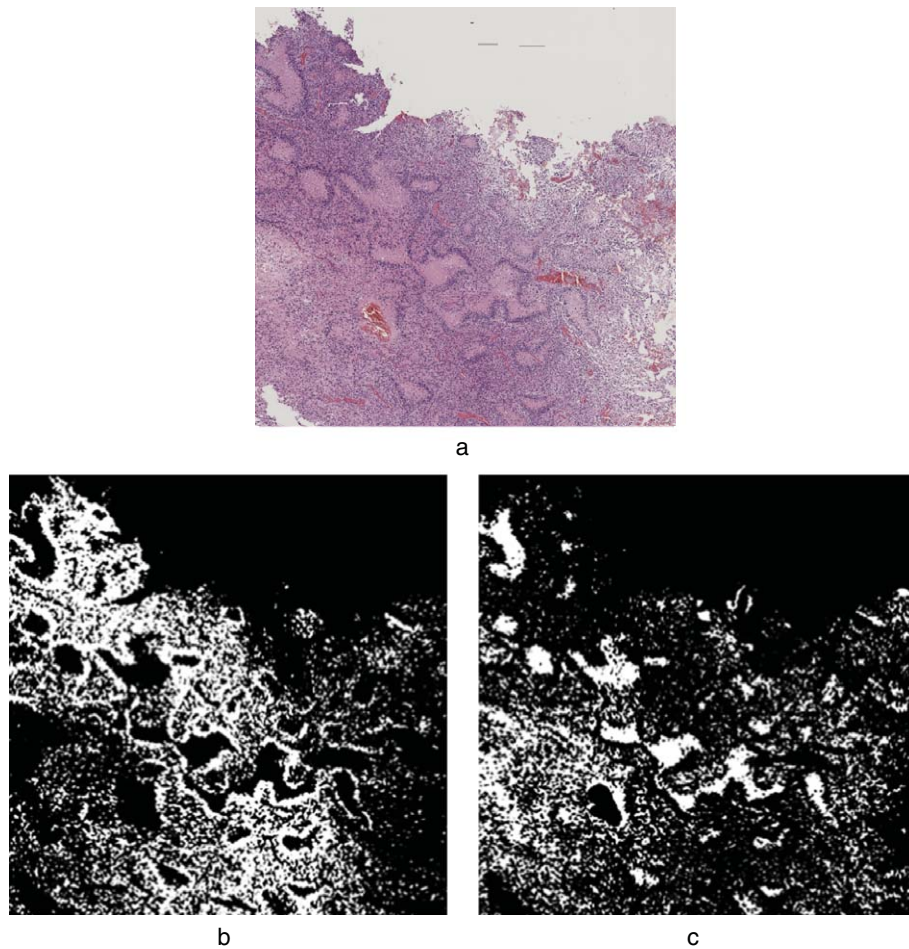


Fig. 3. Thresholding: (a) original virtual slide, (b) thresholded nuclei density, (c) thresholded density of eosinophilic areas.

Adjacent pixels which passed thresholding form many areas in the slide. They are the base of the patterns to be detected. Some auxiliary operations on these areas in the image are executed at this point and also later in the algorithm. They allow to filter out most false shapes of interest which should be skipped in subsequent steps. One of these operations is connecting adjacent areas which converts many close areas to one area. This allows to save shapes which were falsely disconnected by thresholding. A complementary operation to this one is a morphological operation of erosion which disconnects some improperly connected areas. Another operation is removing all small areas, which do not carry any significant information.

3. Matching nuclei adjacent to eosinophilic areas. Eosinophilic areas are dilated using a morphological operation and combined with dense nuclei areas. It is

done by using a binary operation 'AND' on two images. This operation creates specific nuclei patterns which may represent pseudopalisading or palisading shapes. They are examined in the next steps.

4A. Selecting pseudopalisading areas. The goal in this step is to choose only these nuclei patterns which represent pseudopalisading shapes. This is done by passing the patterns through several filters using shape descriptors. Nuclei patterns are filtered by removing:

- filled shapes without big holes: let S be the area of a nuclei pattern, S_{circle} – area of the minimum circle which encloses the shape, S_{holes} – total area of holes inside the shape; then, patterns for which S/S_{circle} exceeds specific threshold and S_{holes}/S is below another threshold are filtered out,

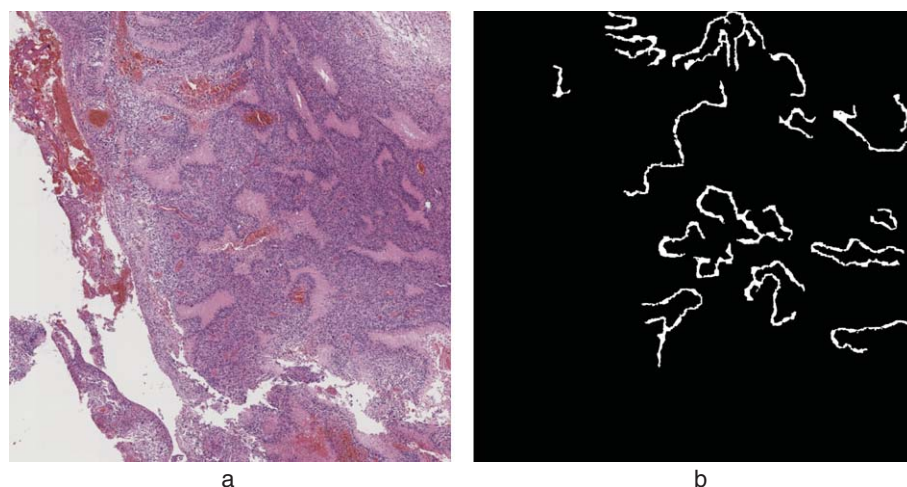


Fig. 4. Filtering pseudopalising shapes: (a) original virtual slide, (b) detected pseudopalising areas.

- shapes of high convexity and without big holes: if S_{convex} is the area of the convex hull of the shape, then areas with S/S_{convex} exceeding a threshold and S_{holes}/S below a threshold are removed,
- shapes with many big holes: if a single pattern has too many holes of significant sizes inside, it is removed,
- shapes without a big convex hull deficiency of high convexity and without a big hole: to pass this test a pattern must have at least one big convex hull deficiency of high convexity or one big hole – otherwise it is filtered out.

As a result of these steps, an image containing patterns selected as pseudopalising is obtained. An example of such an image is shown in Fig. 4.

4B. Selecting and matching palisading areas. In this branch we look for palisading shapes. Therefore, we select linear areas by checking shapes of both eosinophilic areas and nuclei patterns. Another sequence of filters is used here. They remove:

- filled shapes: areas for which S/S_{circle} exceeds a threshold are filtered out,
- areas of high concavity: this time shapes with S/S_{convex} ratio below a threshold are removed,
- areas with very low Pearson correlation coefficient of the shape contour: this procedure filters out areas which do not form linear shapes.

Linear shapes in neurinoma cases appear often side by side, which is a characteristic feature for palisading areas. Therefore, pairs of shapes close to each other are

selected in the algorithm and linear regression (least-squares method) is used to calculate their slopes. Then, close pairs of shapes with small differences in slopes are counted as collinear.

The results of this algorithm branch are a set of linear palisading areas and a set of pairs of close collinear palisading shapes. An example of detected shapes is shown in Fig. 5.

5. Making final decision. Final result returned by the algorithm is based on comparing weighted number of detected pseudopalising areas (n_{pseudo}) and pairs of collinear palisading areas (n_{palis}). Weight of the second measure is bigger because finding collinear shapes signifies higher confidence about detecting a neurinoma case:

- if $n_{pseudo} < 2$ and $n_{palis} < 1$ then decision is ‘Nothing’,
- else if $n_{pseudo} > 3n_{palis}$ then decision is ‘Glioblastoma’ (pseudopalising),
- else decision is ‘Neurinoma’ (palisading).

4. Results

Results of testing the correctness of the proposed method are presented in Table 2. They show that 82% glioblastoma cases and 90% neurinoma cases were correctly classified by the proposed method. In 4 of 49 cases the algorithm detected false disease (neurinoma instead of glioblastoma and vice versa) and in

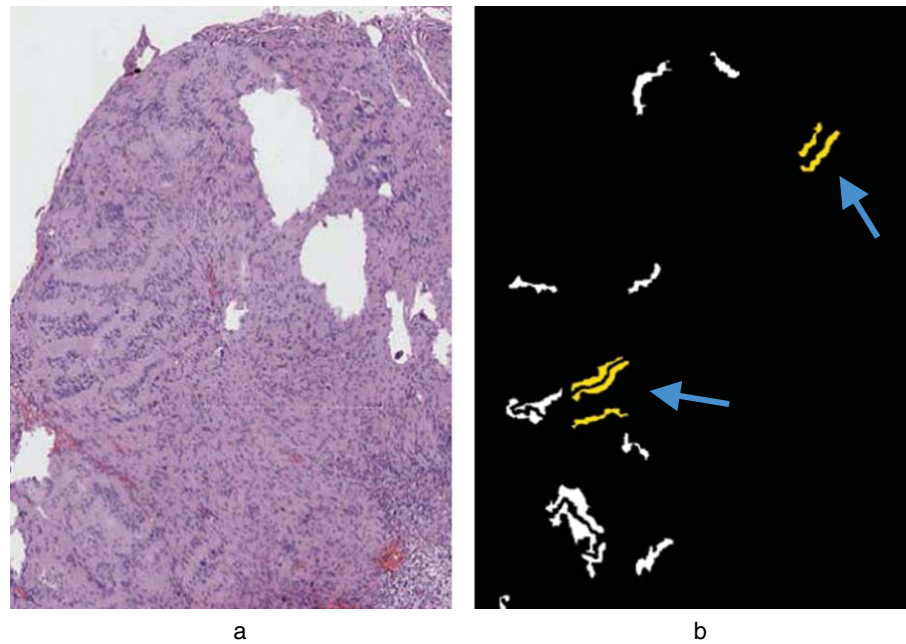


Fig. 5. Filtering palisading shapes: (a) original virtual slide, (b) detected collinear palisading areas marked by arrows.

Table 2
Results of testing the correctness of the proposed method

<i>Detected</i> <i>Actual</i>	Glioblastoma (pseudopalising)	Neurinoma (palising)	Nothing
Glioblastoma (pseudopalising)	32	3	4
Neurinoma (palising)	1	9	0

4 other cases the algorithm returned ‘Nothing’ while there were actually pseudopalising areas in the slide.

Time results of the same tests showed that a single virtual slide image can be classified within a few or a dozen of seconds. Average times for subsequent steps are presented in Table 3.

5. Conclusions

Structural analysis and shape descriptors are promising approaches in detecting glioblastoma and neurinoma cases in virtual slides. Most cases were correctly identified by the proposed method. The algorithm works relatively fast so it could be used in an automated pre-screening of the scanned virtual slides. However, some difficult cases which need to be reviewed by a pathologist were often improperly

Table 3
Time results of tests of the proposed method

Process	Avg. time
Loading and resizing	2.6 s
Pseudopalising detection	5.0 s
Palising detection	2.8 s
Total	10.4 s

classified by the proposed algorithm. Therefore, returning results with plausibility values and images processed by the method would be probably more useful for initial assessment of the slides.

In future work more parameters can be automatically adjusted in the algorithm. Another possibility is to improve existing and introduce new shape descriptors for better discrimination and avoiding incorrect

decisions. Also other patterns in the given classes of tumors can be introduced and number of test cases in each group can be increased. This would lead to further development of the proposed method and better evaluation of its possibilities.

References

- [1] A. Perry and D.J. Brat, *Practical Surgical Neuropathology: A Diagnostic Approach*, Churchill Livingstone, 2010.
- [2] W. Garage, OpenCV – Open Source Computer Vision, 2011. <http://opencv.willowgarage.com/>