

Towards Improved Epilepsia Diagnosis by Unsupervised Segmentation of Neuropathology Tissue Sections using Ripley's- \hat{L} Features

Timm Schoening¹, Volkmar H. Hans², Tim W. Nattkemper¹

¹Biodata Mining Group, Faculty of Technology, Bielefeld University

²Institute of Neuropathology, Evangelisches Krankenhaus Bielefeld, Germany

tschoeni@cebitec.uni-bielefeld.de

Abstract. The analysis of architectural features in neural tissue sections and the identification of distinct regions is challenging for computer aided diagnosis (CAD) in neuropathology. Due to the difficulty of locating a tissue's origin and alignment as well as the vast variety of structures within such images an orientation independent (i. e. rotation invariant) approach for tissue region segmentation has to be found to encode the structural features of neural layer architecture in the tissue. We propose to apply the Ripley's- \hat{L} function, originating from the field of plant ecology, to compute feature vectors encoding the spatial statistics of point patterns described by selectively stained cells. Combining the Ripley's \hat{L} features with unsupervised clustering enables a segmentation of tissue sections into neuropathological areas.

1 Introduction

Visual inspection of microscopy images from neural tissue sections taken from human brains immediately show a variety of structure orientation and distribution of neural cells (Fig. 1a). Depending on the tissue's origin, the position and direction of the cut and the selected part of the tissue to be captured, several input factors determine the initial data to be analyzed. The images analyzed here contained captions of the cortical layers of the human brain. These layers are indexed by roman numbers. Beginning at the outside with layer I which is an almost empty region, five layers of different internal structure follow, ending with layer VI where the amount of neurons faints towards the white matter (Fig. 1b). By just looking at the images, be it the original microscopy outcome or a binarized or centroided one, it is possible to distinguish some major structures although it is a subjective task. Hence the challenge is to enable a computer to identify similar regions, even by means of minor dissimilarities, making it possible to get an objective, reproducible and understandable identification of tissue structures and distortions in the neural architecture. To gain an unbiased insight to the data, we chose to apply the Ripley's- \hat{L} function, which has only recently been introduced to the field of microscopy [1].

2 Materials and Methods

2.1 Imaging Protocol

Diagnostic samples were derived from neurosurgical resections at the Bethel Epilepsy Center, Bielefeld, Germany for therapeutic purposes, routinely processed for formalin fixation and paraffin embedded. Four μm thick microtome sections were immunostained using an antibody against NeuN, a marker labeling almost all mature cortical neurons. Positive staining of cells rendered them brown, leaving surrounding structures with a faint blue counterstain. Archival slides of 19 cases from the years 2009 and 2010 were selected for representing largely normal lateral temporal isocortex cut perpendicular to the brain surface. Slides were analyzed on a Zeiss Axioskop 2 plus microscope with a Zeiss Achroplan 2.5x/0.07 lens. After manually focusing and automated background correction, 1300×1030 pixels, 24 bit, true color RGB pictures were taken at standardized 3200 K light temperature in TIF format using Zeiss AxioVision 3.1 software and a Zeiss AxioCam HRc digital color camera (Carl Zeiss AG, Oberkochen, Germany). All images are fully anonymized and our work did not influence the diagnostic process or the patient's treatment.

2.2 Segmentation

The Ripley's- \hat{K} function [2] is defined for point patterns, so nuclei segmentation was applied to all images using a color channel based algorithm, that takes the average of the three channels of the RGB image at each pixel and compares this value to a given thresholding parameter. All values below this threshold were treated as background, all values above or equal were thought of as neural cells. The binary image for different thresholds was visually compared to the estimated result and a threshold value of about 70 has shown good outcomes. To get a point pattern, the binarized image was segmented, i.e. all connected pixels were combined to a cell region and the centroid of this region was computed as the cells representative point (RP). All RPs together create the nuclei point pattern of an image or an image region.

2.3 Ripley's- \hat{L} Function

The original Ripley function was developed as the Ripley's- \hat{K} function which analyses the distribution of n points in a given area A . It centers an imaginary circle of a given radius at each point of the observed pattern and counts the number of other points found within this circle. This number is set into relation to the expected amount of points within the circle determined by the size of the complete region ("study region" A) and the total number of points (n) and then summed up for all points i . The \hat{K} value

$$\hat{K}(r) = \frac{A}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_{ij}(r)}{C_i(r)} \quad (1)$$

is computed for all radii within a given interval (i.e. from $r = 1$ to $r = r_{\max}$), with δ in the numerator is 1 for point distances $d(i,j) < r$, else it is 0. The denominator is the fraction of the circle area with radius r at point i within the study region so the point counts for points i close to the study region borders are adjusted [3]. A drawback of the \hat{K} function are its fast growing values for increasing radii. Therefore the \hat{L} function was introduced which is a normalized representation and has values around zero

$$\hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}} - r \quad (2)$$

\hat{L} values above zero for a given radius indicate more, values below zero indicate less points within this radius than expected.

2.4 Confidence Envelopes

To evaluate the significance of an observed point pattern, a reference point pattern is required. Therefore usually the assumed pattern creation process is simulated to create a set of further point patterns [4]. There is a variety of possible point processes, the easiest of which is complete spatial randomness which was used here. We simulated a similar study region and randomly distributed the same amount of points as in the original pattern within this region. We then again calculated the \hat{L} function and repeated this procedure nineteen times to yield 95 percent confidence envelopes by taking the highest and lowest calculated values for each radius as envelope limits.

2.5 Feature Set

Each captured 1300×1030 pixel image was cut into overlapping tiles. Each tile was 150×150 pixel and the overlap in each direction was 140 pixels creating a set of 116×89 tiles. Each tile was represented by one Ripley's- \hat{L} feature vector, yielding 10,324 feature vectors per image. To avoid edge effects the tiles had to be completely inside the image, leaving a 70 pixel border at each side which was not analyzed. We then computed standard point pattern features like density (λ), mean point distance (\bar{d}) and deviation of mean distance (σ). The density gave good approximations for the underlying brain structure as can be expected. But being a one-dimensional measure, it was not able to identify semantic differences between the layers, e.g. layers I and VI (Fig. 1, B) both feature low density but resemble the far inner/outer layer and show distinctive structures at the transitions towards their neighboring layers. To compute Ripley's- \hat{L} based features, we computed the \hat{L} function (with $r_{\max} = 20$) on each tile's point pattern, so $A = 150 \times 150$ in (1), and calculated features from the resulting \hat{L} function like the amount of null points (N), area above (I_a), below (I_b) and outside (I_o) confidence envelopes. In order to achieve a compact representation of the \hat{L} function shape, we approximated the \hat{L} -function by a polynomial of

degree m and used the polynomial coefficients as features. However in our initial experiments we found that we had to include all the \hat{L} functions numerical values to get a more comprehensive description of its slope. Since we needed to take the confidence envelopes into account here, we calculated a shifted version of the \hat{L} function, that is zero for all radii at which the function is within the confidence envelope and the difference between the function and the envelope otherwise. As a result of several experiments we found, that a joint feature vector \mathbf{v} , containing density (λ) and *all* numerical values of the original as well as the shifted \hat{L} function, produced the best results and so we joined those features to a 41-dim feature space.

2.6 Clustering, Segmentation and Visualization

The feature vectors from all tiles of all images were fused to one training set ($19 \cdot 10,324 = 196,156$ items). Each vector was normalized to $|\mathbf{v}| = 1$ before clustering. Batch k -means was applied to fit $k=20$ prototype vectors $\mathbf{v}_{\alpha}, \alpha=1..20$ to regions of high feature vector density. The clustering was stopped if at most 50 iterations were performed or less than 0.1 % of the assignments of feature vectors to their best matching unit (BMU) changed during one iteration. Afterwards each pixel was mapped to its BMU to achieve a first segmentation i.e.

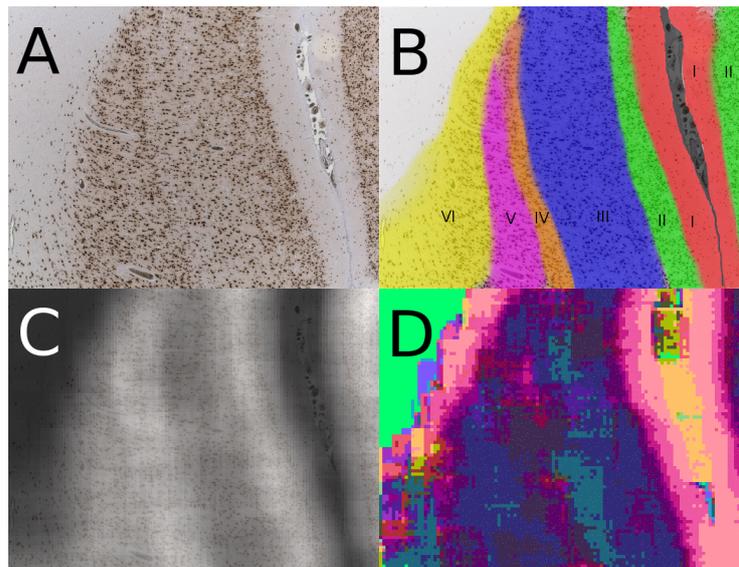


Fig. 1. A: an original input image, beginning on the left with white matter, then layers VI to I, attached to the right is the next cerebral gyrus, again beginning with layer I and II; B: a hand labeled image with highlighted regions and an arbitrary color coding; C: The original image as a background with a density overlay; D: The clustering result where less occupied regions are dark blue whereas the filled layers are purple and pink at sharp edges towards empty regions.

pixel labeling result. To visualize and evaluate this segmentation result, each prototype v_α was assigned to a RGB color $(r,g,b)_\alpha$. To preserve topology in the mapping from the feature to the color space, we assigned RGB components using projections of the prototype vectors onto the eigenvectors belonging to the three highest eigenvalues of the feature set (Fig. 1d). We reran the experiments with the same settings to confirm the cluster results and found that small differences can occur due to the randomization effects during confidence envelope calculation.

3 Results

The segmented result (Fig. 1d) shows a layered structure comparable to the density image (Fig. 1c). However the color scale compared to density of the prototypes shows that the prototypes describe a manifold of a higher dimension i.e. the features encode spatial features more complex than density. At least three eigenvectors, belonging to the three highest eigenvalues are necessary to encode 95 percent of the data variance. The less occupied regions are separated into the white matter and layer I. Layers II to VI show similar results as both feature pattern distributions of moderate to high density and are only discerned by their relative position to each other. The resulting structure shows good overlap with the hand labeled image (Fig. 1, B) although there are differences especially for layers V and VI where the subjective coloring shows a smaller region V but the data driven approach assigns more tiles to the comparable region.

4 Discussion

The outcome of our proposed method used as an initial attempt is encouraging for further optimization. Beginning with binarization and centroid finding, moving on to feature selection and normalization, ending up with clustering and result visualization, every component of this project could be improved. On the other hand every part of it is simple and easily understandable which is important to avoid explanation problems of black-box solutions. Our results are promising to provide in the future a computer aided tool for visualizing subtle developmental brain abnormalities associated with human epilepsy.

References

1. Mattfeld T, Eckel S, Fleischer F, et al. Statistical analysis of labelling patterns of mammary carcinoma cell nuclei on histological sections. *J Microscopy*. 2009;235:106–18.
2. Ripley BD. The second-order analysis of stationary point processes. *J Appl Prob*. 1976;13:255–66.
3. Goreaud F, Pelissier R. On explicit formulas of edge effect correction for Ripley’s K-function. *J Veget Sci*. 1999;10:433–8.
4. Wiegand T, Moloney KA. Rings, circles and null-models for point pattern analysis in ecology. *Oikos*. 2004;104:209–29.