Eigenvalue-Based Preprocessing for Tissue Extraction From Pathology Image slides

Do-Hyun Lim[1]

Department of Computer Science, Hanyang University, Seoul, Republic of Korea

ldh991014@hanyang.ac.kr

Introduction

In various biological fields, tissue samples are frequently collected from subjects and analyzed as image slides. These pathology image slides often contain not only the tissue of interest but also extraneous elements, including background regions and markings, which can compromise analysis accuracy.

To accurately isolate tissue regions, precise preprocessing is essential. By extracting each element's features using simple algorithms, we can effectively distinguish tissue from unwanted elements.

Objective

Extract eigenvalues and eigenvectors from patches containing each element to analyze their features.

Set an appropriate threshold to extract only tissue.

Method

```
Algorithm 1 Extracting tissue from Whole Image Slide
1:  Input: RGB Slide
2:  Calculate the mean and RGB Covariance Matrix.
3:      Let mean be the mean of the RGB channels.
4:      Let Σ be the covariance matrix of the RGB channels.
5:  Calculate the eigenvalues of the covariance matrix Σ.
6:      Let λ₁, λ₂, λ₃ be the eigenvalues of Σ, where λ₁ ≥ λ₂ ≥ λ₃.
7:  Calculate the overall mean of the RGB channels.
8:      Let mean_alpha = (255−mean)/255.
9:  Define the background condition:
10:     Background Judge: (240 > mean) Or (mean > 50)
11: Define the first eigenvalue condition:
12:     Eig_1 Judge: (λ₁ > 100) Or (λ₁ > mean_alpha × α)
13: Define the second eigenvalue condition:
14:     Eig_2 Judge: (λ₂ > 1.8) Or (λ₂ > mean_alpha × β)
15: Check the final condition:
16:     Output: True if (Background Judge) And (Eig_1 Judge) And (Eig_2 Judge);
        otherwise, False
```

Figure 1. Pseudo Code for Patch Extraction

Background Judge: Average which is too large or too small, consider it background or marking

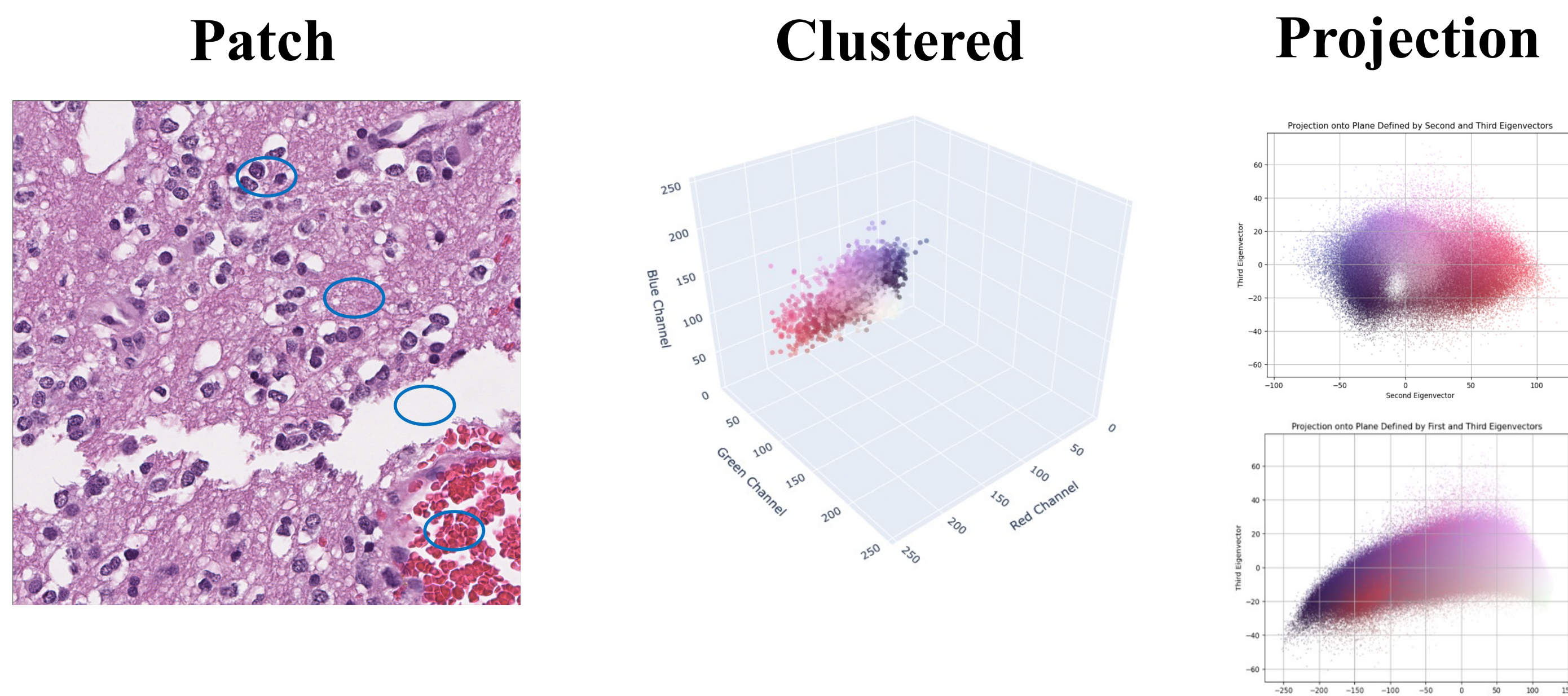| Patch	Clustered	Projection

Figure 2. Project the RGB points of an image onto a plane represented in three dimensions, with the first and second eigenvectors as normal vectors, respectively.

We found that eigenvectors classify features well.

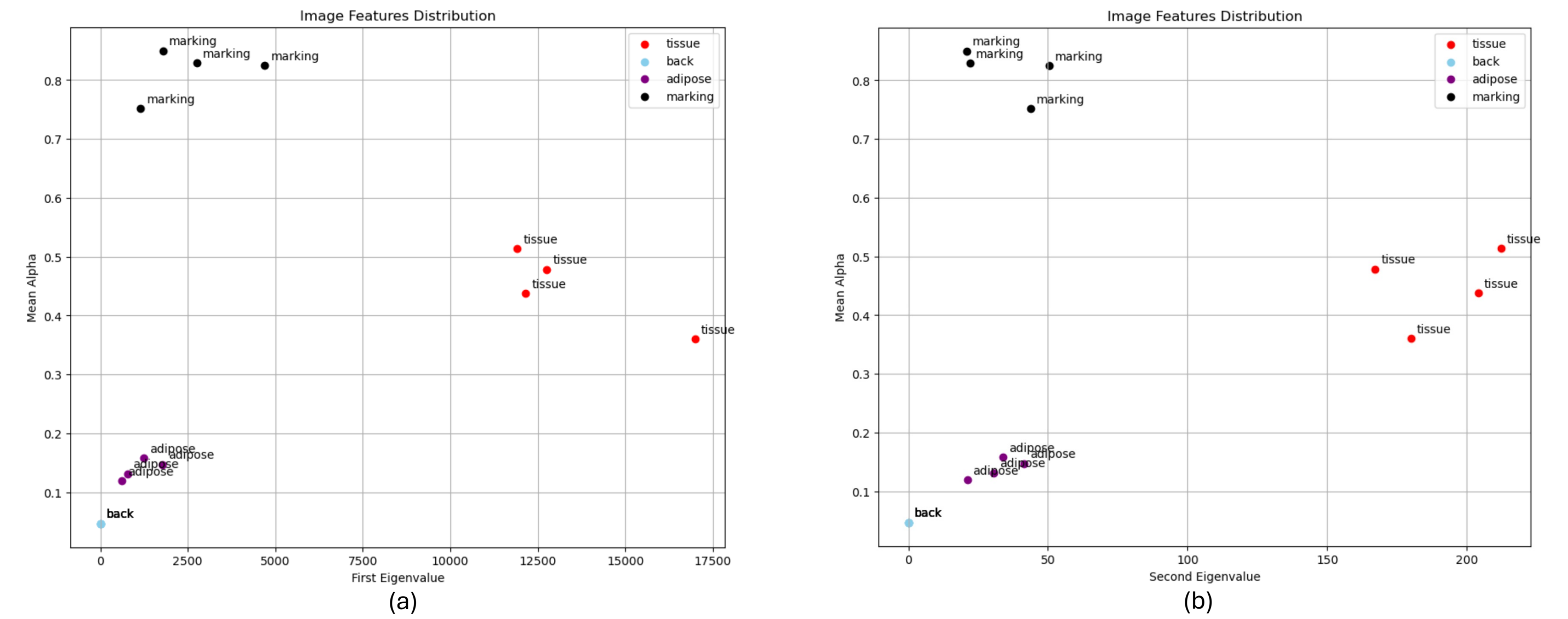We also found that the difference in the size of the eigenvalues for each feature was significant.

Figure 3. (a) Plot with first eigenvalue as X-axis and mean_alpha as Y-axis (b) Plot with second eigenvalue as X-axis and mean_alpha as Y-axis

Eig_1 Judge: First, Detach the backgrounds

Eig_1 Judge & Eig_2 Judge: Then, Among other remaining features, marking becomes linearly separable from tissue and adipose.

With these three judges, we can completely extract only tissues.
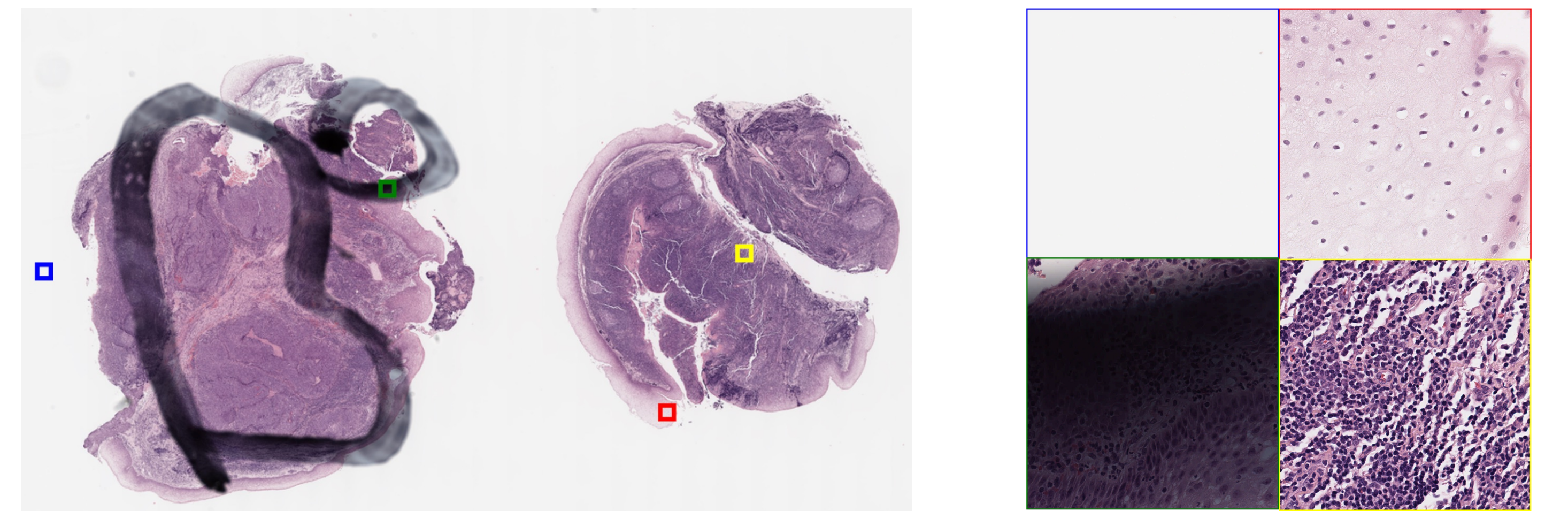
Results

Figure 4. Project the RGB points of an image onto a plane represented in three dimensions, with the first and second eigenvectors as normal vectors, respectively.

Our algorithm-based experiments demonstrate that it is good at extracting tissue-only patches from whole slide images and can quickly separate the tissue from the background in an image, which is advantageous when processing large amounts of data at once.

Conclusion

We successfully demonstrated a straightforward and effective approach for distinguishing tissue regions from unwanted elements such as background and markings in pathology image slides. By leveraging eigenvalue-based criteria and simple preprocessing steps, our method isolates tissue regions with high accuracy. Using eigenvector directions as the basis for feature extraction, we were able to segment image patches based on their characteristic variance patterns, which efficiently separated tissue from extraneous elements.

Our method's simplicity and robustness make it particularly advantageous for large-scale applications, allowing for rapid and accurate extraction of tissue-only data. This approach holds promise for improving analysis efficiency in biological and medical image processing by enabling a more focused examination of relevant regions.