Stefan Patzke*, Christian Wöhler, and Jörg Thiem

Superpixel-Based Endmember Extraction for Hyperspectral Endometriosis Detection

https://doi.org/10.1515/cdbme-2025-0188

Abstract: Hyperspectral Imaging (HSI) is a promising tool for assisting medical diagnostics, as it enables precise tissue characterization and differentiation through detailed spectral analysis. However, HSI data analysis faces challenges in reliably identifying relevant regions (e.g., tumors, vascular structures) due to spectral variability, complicating universal algorithm development, especially with limited ground truth data. Focusing on endometriosis as a medical use case, we generate plausible reference spectra (endmembers) for affected tissue using HSI data, despite limited and weakly annotated datasets. Our processing pipeline includes Savitzky-Golay (SG) smoothing, Standard Normal Variate (SNV) standardization, Principal Component Analysis (PCA), superpixel segmentation and the Pixel Purity Index (PPI). Validation via Intersection over Union (IoU) achieves 0.94 accuracy in lesion detection.

Keywords: hyperspectral imaging, hsi, endometriosis, endmember extraction

1 Introduction

Hyperspectral Imaging (HSI) leverages detailed spectral data to enhance tissue analysis in medical applications, offering insights into tissue composition beyond conventional imaging techniques. By extracting spectral fingerprints, he discrimination of even subtle tissue alterations can be enhanced. However, its diagnostic utility is limited by spectral variability and the lack of comprehensive annotated datasets, posing challenges for robust feature extraction.

We examine endometriosis as a model case using ex-vivo HSI data acquired with a UV-VIS camera with a specral range of 330-800 nm and a spectral resolution of 2.8 nm (FWHM), measuring reflectance using a Spectralon reference standard. Endometriosis presents diagnostic challenges due to its variable appearance across different organs and frequent recurrence after treatment, often causing severe pain and infertility.

***Corresponding author: Stefan Patzke,** University of Applied Sciences and Arts Dortmund, Sonnenstr. 96, 44139 Dortmund, Germany, e-mail: stefan.patzke@fh-dortmund.de
Christian Wöhler, TU Dortmund University, 44227 Dortmund, Germany
Jörg Thiem, University of Applied Sciences and Arts Dortmund, 44139 Dortmund, Germany

This study aims to generate reference spectra (endmembers) for affected tissue to improve endometriosis lesion detection in hyperspectral images. The current work is limited to four weakly annotated images (see Figure 1), demonstrating the feasibility of the approach for initial analysis despite limited data.

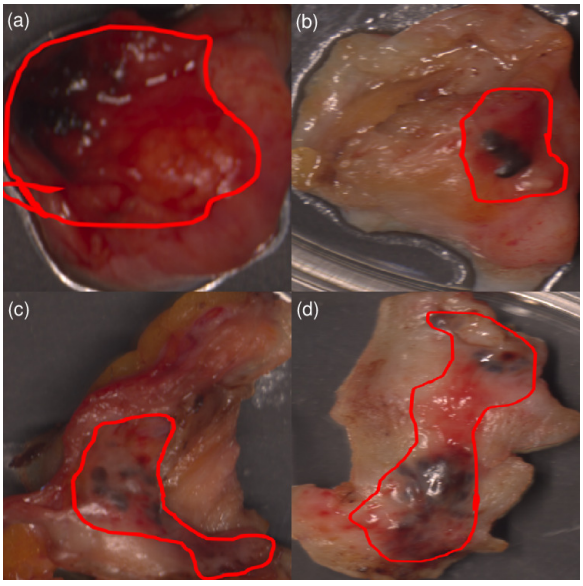

Fig. 1: RGB representations of all used images. The red marked areas show the regions of interest which include tissue affected by endometriosis.

2 Methodology

To generate endmember spectra, the HSI data undergoes the following processing steps. Initially, artifacts and noise are eliminated. Following this purification, the data set is standardized to ensure the integrity of the data. The dataset is then reduced in dimensionality, followed by superpixel segmentation to group pixels into homogeneous regions. Endmembers are extracted from these superpixels and spectral evaluation metrics are applied to assess the quality of the generated spectra.

2.1 Data Preprocessing

The preprocessing pipeline involves three sequential steps to enhance data quality. First, background and highly reflective

pixels are removed via threshold and visual inspection. Second, a Savitzky-Golay (SG) filter (2nd-order polynomial, 7-band filter window) reduces high-frequency noise while preserving spectral features, e.g. absorption peaks [4]. Third, Standard Normal Variate (SNV) normalization is applied to correct scattering effects and variations in illumination across samples by performing Equation 1 on each spectrum [5].

$$\mathbf{X_{SNV}} = \frac{\mathbf{X} - \mu}{\sigma} \tag{1}$$

Here, μ and σ represent the mean and standard deviation of the spectrum, respectively. This standardization of the data ensures cross-sample comparability.

Additionally, the dimensionality of the hyperspectral dataset is reduced by using Principal Component Analysis (PCA), retaining most of the information by transforming the data into orthogonal principal components (PCs) ordered by their variance contribution [6]. In this case, the first five PCs are kept, capturing over 95 % of the variance to represent the dataset in a lower-dimensional space. This step mitigates the curse of dimensionality [7], improves computational efficiency and minimizes spectral redundancy.

A Simple Linear Iterative Clustering (SLIC) algorithm then segments the hyperspectral images into spectrally homogeneous superpixels [8]. Each superpixel is represented by the average spectrum of its constituent pixels using Equation 2.

$$\bar{\mathbf{S}}_\mathbf{j}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} S_{i,j}(\lambda) \tag{2}$$

Here, $\bar{S}_j(\lambda)$ is the mean spectrum of over all N individual pixels in each superpixel denoted by j.

This reduces computational load, suppresses noise and enhances processing efficiency. The approach also improves interpretability, offering a structured basis for downstream analysis [8]. As an example, results for the image in Figure 1 (c) are shown in Figure 2 (a).

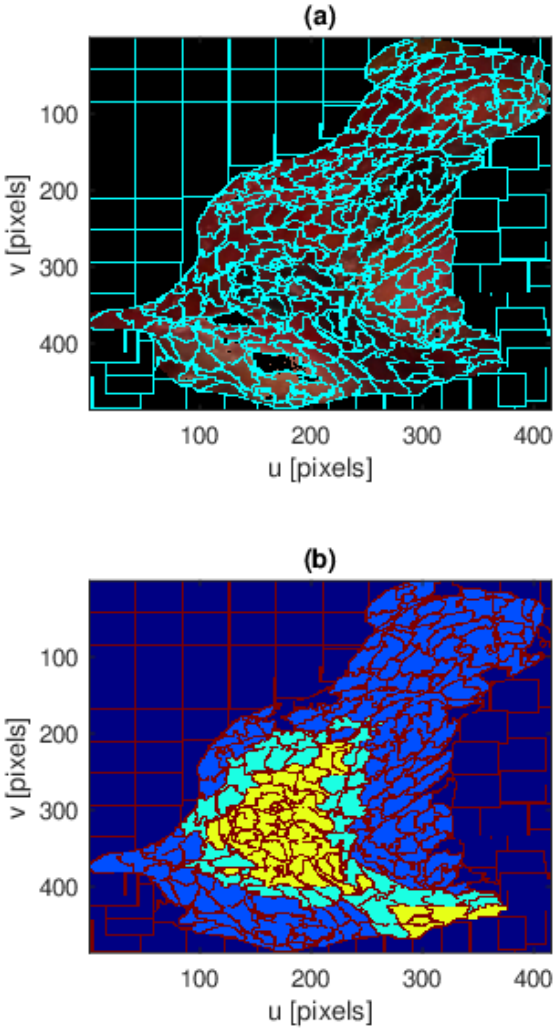

Fig. 2: Image examples of (a) superpixel segmented image and (b) markings of superpixels that lie outside (blue), on the border (cyan) and inside (yellow) the ROI corresponding to Figure 1 (c).

2.2 Endmember Selection

The superpixels being used are further restricted to those enclosed by the annotated regions shown in Figure 1 to ensure the resulting endmember spectra represent affected tissue. An example of this step is visualized in Figure 2 (b). Hence, only the yellow marked superpixels within this image are used.

The Pixel Purity Index (PPI) algorithm is then applied to these averaged spectra. PPI identifies pure pixels by projecting spectral data onto random vectors and counting extreme occurrences, with the highest-scoring pixels selected as endmembers [9]. According to the geometric rule of thumb $n \leq p + 1$, for a five-dimensional space ($p = 5$), a maximum of $n = 6$ endmembers are defined to form a simplex capturing spectral diversity [10]. Therefore, the number of expected endmembers is set to six.

2.3 Evaluation

For Evaluation, the Spectral Angle Mapper (SAM) algorithm is used, calculating the similarity between pixel spectra t and endmember spectra r by

$$\alpha = \cos^{-1}\left(\frac{\mathbf{t}\cdot\mathbf{r}}{|\mathbf{t}|\cdot|\mathbf{r}|}\right), \tag{3}$$

where lower scores α indicate higher similarity [11].

Afterwards, the Intersection over Union (IoU) is calculated using segmentation maps, which are generated by as-

signing each pixel to the endmember with the lowest SAM score, corresponding to the most suitable endmember. IoU determines the ratio of the overlap area between the predicted endmember and the label-area to the corresponding area of their union [12]. A weighted variant (wIoU) incorporates SAM scores for pixel weighting instead of binary segmentation. To emphasize differences, the wIoU is additionally standardised to the interval [0,1], continuously referred to as rescaled wIoU.

3 Results

By restricting the used superpixels to those enclosed by the areas labeled in Figure 1, the total number of superpixels is reduced from 1209 to 335, serving as input for the PPI algorithm, producing six endmember spectra. Consequently, the endmembers potentially indicate tissue affected by endometriosis.

Tab. 1: Evaluation scores corresponding to the significance of each endmember.

Endmember No.	1	2	3	4	5	6
IoU	0.80	0.52	0.26	0.21	0.87	0.94
wIoU	0.43	0.36	0.34	0.33	0.39	0.46
rescaled wIoU	0.75	0.21	0.07	0.00	0.44	1.00

Table 1 shows the results of all IoU variants used for each endmember across all images viewed. The IoU values indicate that segments overlap most for endmembers 1, 5 and 6, while endmember 4 has the least overlap. This is further refined via the wIoU and rescaled wIoU, showing endmembers 1 and 6 have the highest agreement.

Figure 3 presents the score maps for all images and endmembers, with scores inverted for visualization to show a higher score as a better match. Endmembers 1 and 6 show the highest scores within the labeled and therefore endometriosis affected areas (see Figure 1). In comparison, although endmember 5 also scores high values in these areas, they do not differ as much from those of the surrounding tissue. Thus, endmembers 1 and 6 are best for highlighting affected regions due to their high sensitivity in score differences.

Examining the spectra of the extracted endmembers in Figure 4, there are significant discrepancies in local extrema, especially at wavelengths around 350 nm, 420 nm, 630 nm and 750 nm. These areas may be particularly significant, warranting further analysis in future research.

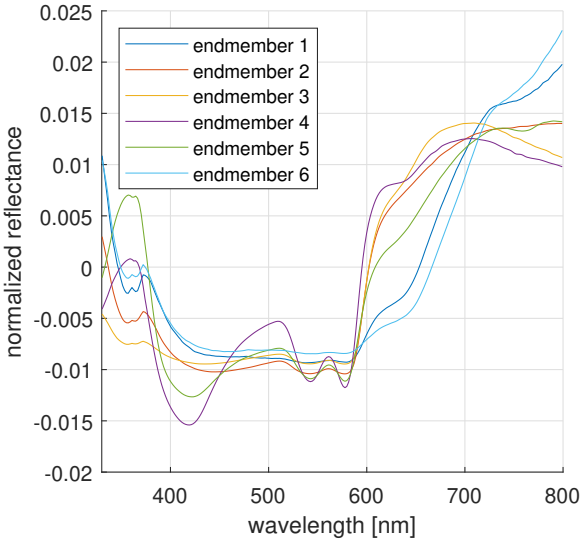

Fig. 4: Spectrum of each resulting endmember.

4 Conclusion and future work

In conclusion, extracting endmembers using the presented methodology highlights affected tissue areas in the sample data. Future studies will include additional data acquired and examined in collaboration with clinical partners. Spectral analysis and AI techniques will synthetically expand the dataset through Diffusion Models, optimizing algorithms to differentiate affected from healthy tissue. While pixel-wise labeling is not yet part of the process, it may be used for select examples in future work to refine algorithms. The ultimate goal is to develop a robust classifier for subtle endometriosis lesions, integrated with an HSI endoscope to assist surgeons by providing enhanced visualization (e.g. as an overlay of affected areas) during minimally invasive procedures. This is crucial for treating barely visible lesions to prevent recurrences and persistent pain, enhancing therapy quality and ensuring complete removal of endometriosis lesions.

Acknowledgment: We express our deepest gratitude to the Marienkrankenhaus in Schwerte (Germany) and the Klinikum Dortmund (Germany) for their invaluable support in data collection, which was essential to the success of this study.

Author Statement

Research funding: This work is funded by the Federal Ministry of Education and Research (BMBF, Funding number: 13FH097KX0). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical ap-

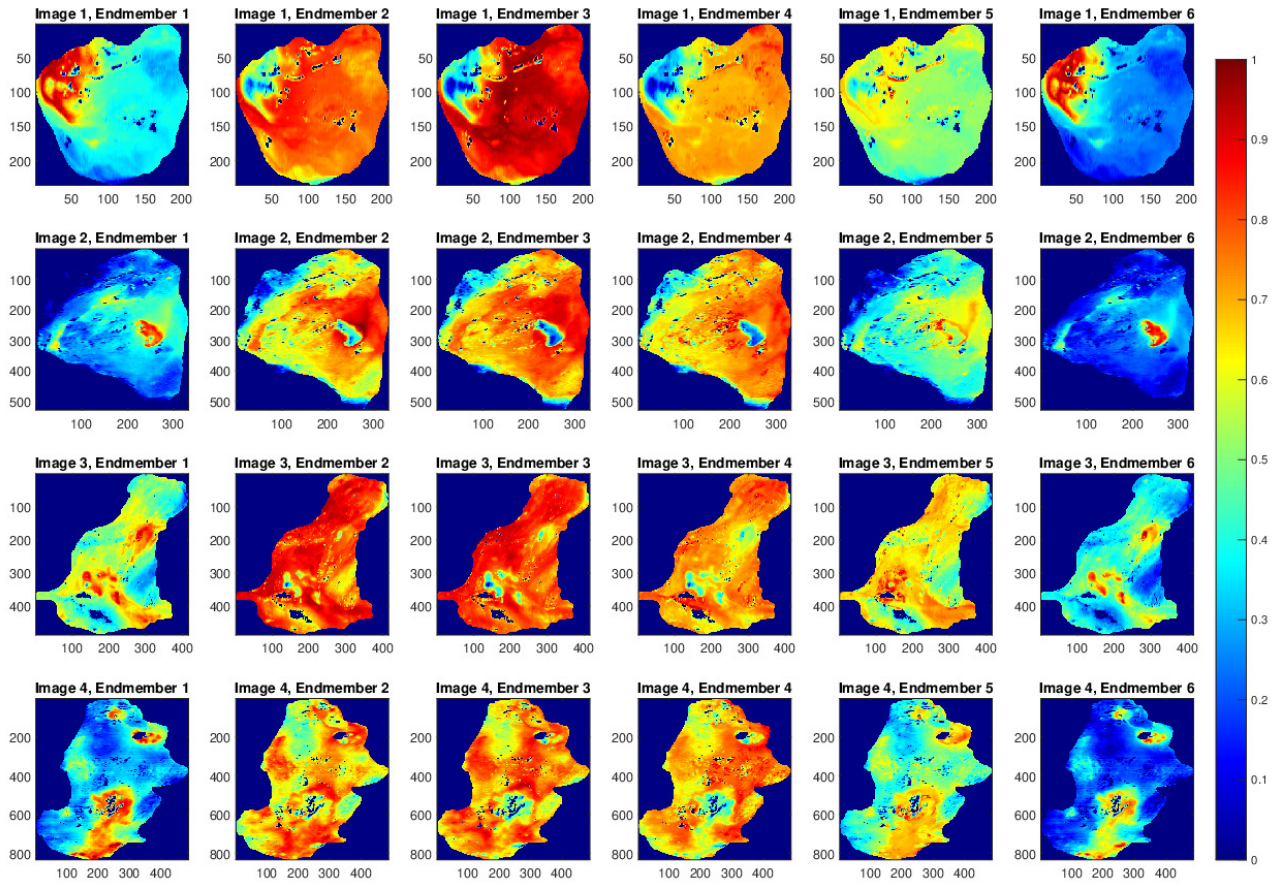

Fig. 3: Score-Maps of for each endmember and image configuration. A higher score corresponds to a greater match between the pixel and the respective endmember.

proval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

References

[1] Patzke S, Wöhler C, Thiem J. Hyperspectral Image Data for Diagnosis of Endometriosis. Proceedings Of The 11th International Conference On Bioinformatics Research And Applications 2025;30-34.

[2] Tanriverdi F, Schuldt D, Thiem J. Dual Snapshot Hyperspectral Imaging System for 41-Band Spectral Analysis and Stereo Reconstruction. Advances In Visual Computing 2011;3-13.

[3] Georgiev G, Butler J. Long-term calibration monitoring of Spectralon diffusers BRDF in the air-ultraviolet. Appl. Opt. 2007;46:7892-7899.

[4] Schafer R. What Is a Savitzky-Golay Filter? [Lecture Notes]. IEEE Signal Processing Magazine 2011;28:111-117.

[5] Zeaiter M, Rutledge D. Preprocessing Methods. Comprehensive Chemometrics 2009;121-231.

[6] Marsland S. Machine learning (2 ed.). Chapman and Hall/CRC, Philadelphia, PA, 2014.

[7] Ayesha S, Hanif M, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion 2020;59:44-58.

[8] Xu X, Li J, Wu C, Plaza A. Regional clustering-based spatial preprocessing for hyperspectral unmixing. Remote Sensing Of Environment 2018;204:333-346.

[9] Boardman JW, Kruse FA, Green RO. Mapping target signatures via partial unmixing of AVIRIS data., Technical Report, California, USA, 1995.

[10] Boardman JW. Analysis, understanding and visualization of hyperspectral data as convex sets in n-space. SPIE Conference Proceedings, Imaging Spectrometry, SPIE Publication 2480, Orlando, FL, 1995.

[11] Kruse F, Lefkoff A, Boardman J, Heidebrecht K, Shapiro A, Barloon P, Goetz A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote Sensing Of Environment 1993;44:145-163.

[12] Csurka G, Larlus D. What is a good evaluation measure for semantic segmentation?. IEEE Trans. Pattern Anal. Mach. Intell. 2013;26.