Sparse Mid-Infrared Spectra Enable Real-time and In-vivo Applications in Tissue Discrimination

Felix Fischer[1,*], *Karsten* Frenner[1], *and Alois M.* Herkommer[1]

[1]Institute of Applied Optics, University of Stuttgart, 70569 Stuttgart, Germany

Abstract. Differentiation of malign and benign tissue based on spectral information can be done by only using a small fractional amount of the original mid-infrared spectrum. An optimally selected arrangement of a few narrow-band quantum cascade lasers provides proficient signal-to-noise ratios and can drastically reduce the data acquisition time with constant discriminability, such that real-time applications will be possible in short-term and in-vivo diagnostics in the long-term.

1 Introduction

Most information in analyses of biological structures in human tissue is found in the broad mid-infrared (MIR) region [1]. In order to achieve reasonable signal-to-noise ratios (SNRs) throughout this complete spectrum, either the light source has to be utterly powerful, the detector has to be more efficient or the spectrum has to be averaged over multiple measurements [2]. Whereas the former yields an increased energy consumption and naturally more waste heat that has to be dealt with, the latter entails a longer acquisition time. Both of which is not ideal for preferential real-time and intra-operative tissue discrimination with spectral measurement technologies.

Thus, modern MIR spectrometers tend to use quantum cascade lasers (QCLs), that are more powerful light sources with a higher spectral irradiance than conventionally used thermal SiC sources [3]. When confined to narrower bands, e.g. in gas analyses of particular media [4], a real-time analysis of the spectrum is viable. Since most discriminative information in tissue spectra is also found in these distinct peaks (i.e. biomarkers) [5], the logical conclusion is that a single targeted approach to the relevant areas yields a drastic decline of the spectral acquisition time.

With the help of several different data-driven feature selection (FS) methods, we show that only a small fraction of the original MIR spectrum is necessary to reliably separate benign from malign urothelia. According to that, we introduce an optical design that enables the spatially resolved real-time discrimination of bladder tissue.

2 Wavelength Selection for MIR Spectra

The selection of relevant wavenumbers $\tilde{\nu}$ is based on the data that has been measured with the FT-IR spectrometer Spectrum Two from Perkin Elmer and an attenuated total reflection (ATR) accessory with a diamond crystal. A spectral range from 4,000-450 cm^{-1} is obtained with this combination. The spectra obtained are from multiple patients that have been diagnosed with bladder cancer. Due to non-macroscopic tumors in the resected bladder, it has not been possible to acquire tumorous spectra yet. For the sake of discriminating different bladder wall constituents, we decided to measure the urothelium and the detrusor muscle (DM) of benign bladder samples, see Fig. 1a and b.

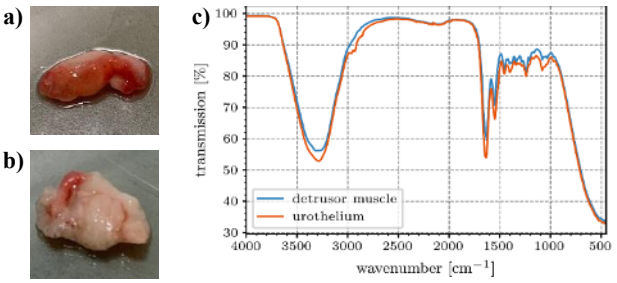

Fig. 1. Investigated samples of **a)** urothelium, **b)** detrusor muscle and **c)** mean spectra of the two investigated spectra. The correlation coefficient between both spectra is 99.75 %.

In order to select the most relevant wavenumbers for an optimal differentiation, we have chosen two different methods. The results of a principal component analysis (PCA) are compared to the results of a recently developed feature selection algorithm called FeaSel-Net [6], that is based on recursive elimination of the input signal in neural networks.

2.1. Feature Selection with PCA

Figure 2a shows the scores plot from PCA of the two different tissue types, where a clear separation between the two clusters is shown along the first component. Except from one sample of the urothelium class, all samples can be clearly detached using the y-axis as border. Even though the correlation between both mean spectra in Fig. 1c is nearly 100 %, it is still possible to

* Corresponding author: felix.fischer@ito.uni-stuttgart.de

transform the data into distinct clusters using PCA. The data has been standardized along the wavenumber axis during the pre-processing.

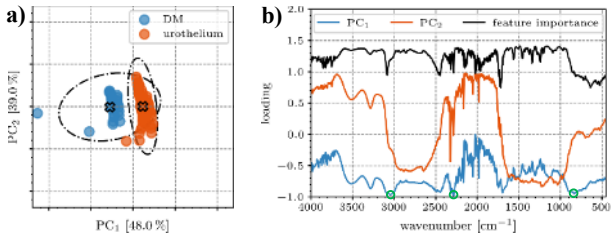

Fig. 2. Scores **a)** and loadings **b)** plot of a PCA showing the first two PCs that describe nearly 90 % of the data. A distinct separation of DM and urothelial tissue can be observed.

When looking at the loadings for the transformation in Fig. 2b, it can be seen that the PC1 and PC2 loading vectors are in the same order of magnitude. When calculating the sum of absolutes of the loadings, it seems that the wavenumber range between 1,600-900 cm^{-1} is slightly more important than the rest of the spectrum. This coincides with the difference between the mean spectra that occurs in Fig. 1c. Since PC1 is attributed with the discriminability though, only the loadings from PC1 are considered. The green circles in Fig. 2b indicate the three most important wavenumbers in this component. The exact wavenumber values are given in Table 1.

2.2. Feature Selection with Neural Networks

As a non-linear evaluation of the wavenumbers, we use FeaSel-Net [6]. During the training process of a classifier neural network, the feature importance

$$\mathcal{I}_f(\mathbf{X}_f) = -\frac{1}{\mathrm{n_s}} \cdot \sum_{i=1}^{n_\mathrm{s}} \sum_{\mathbf{x} \in \mathbf{X}_f} \hat{\mathbf{y}}\,(\mathbf{x}) \cdot \log \mathbf{y}(\mathbf{x}) \qquad (1)$$

is evaluated, where the original dataset $\mathbf{X}$ with $n_\mathrm{f} = 888$ and $n_\mathrm{s} = 79$ samples is masked at each feature f. Subsequently, the impact on the classification performance is analysed and features are discarded, if the impact of masking is negligible. With a pruning rate of $\pi = 0.2$, a convergence with $n_f' = 3$ distilled features occurs after 25 epochs. Inherent randomness in neural networks make it indispensable to statistically evaluate multiple selections. The wavenumbers that are chosen most often are given in Table 1. We additionally trained another neural network with only the two stated wavenumbers and calculated the mean and standard deviation of the overall classification accuracy (OCA) for ten individual optimization processes. Both methods clearly surpassed random wavenumber selections.

Table 1. Relevant biomarkers and the OCA for both methods.

FS Method	$\tilde{\nu}_1$ [cm^{-1}]	$\tilde{\nu}_2$ [cm^{-1}]	OCA
PCA based	3,048	2,288	97.5 ± 2.0 %
FeaSel-Net	1,696	1,740	96.9 ± 2.2 %
Random	3,592	2,260	78.7 ± 5.9 %

3 Multi-spectral Spectrometer

With the few wavelengths selected in section 2 we propose a multi-spectral approach for the illumination of tissue samples inside an ATR measurement system. As depicted in Fig. 3, an echellette grating is used to guide all distributed feedback (DFB) QCLs onto the optical axis. The QCLs are chosen such that they correspond to our previous findings. With well-timed shutters, the sample is illuminated sequentially.

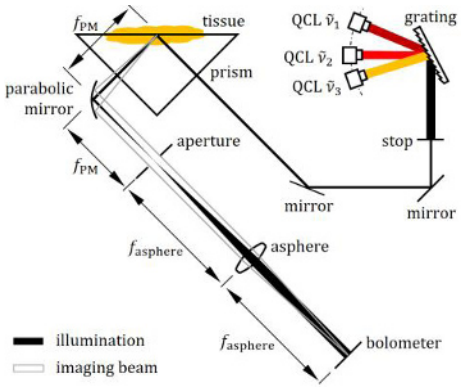

Fig. 3. Possible setup for a multi-spectral ATR spectrometer. The tissue is illuminated by multiple DFB-QCLs and the attenuated reflections are imaged using a microscope.

In this configuration, the lasers are imaged on the focal plane array (FPA) bolometer as a reference at first. As soon as there is close contact with the tissue, a part of the laser light is absorbed depending on the tissue's absorption coefficient at the interface. A magnification of the internal reflective element (IRE) surface is made possible using an $4f$ beam expander.

The work described in this paper was conducted in the framework of the Graduate School 2543/1 "Intraoperative Multisensory Tissue Differentiation in Oncology" (project A2) funded by the German Research Foundation (DFG – Deutsche Forschungsgemeinschaft).

References

1. V.F. Kalasinsky, Appl. Spectrosc. Rev., *Biomedical Applications of Infrared and Raman Microscopy*, **31**, 193 (1996)
2. G.W. Ewing, Rev. Sci. Instrum, *Signal-to-Noise Enhancement in Infrared Absorption Spectrophotometry*, **42**, 169 (1971)
3. M. J. Weida, B. Yee, Eds. Proc. SPIE, *Quantum cascade laser-based replacement for FTIR microscopy* (2011)
4. R. Ghorbani, F.M. Schmidt, Appl. Phys. B, *Real-time breath gas analysis of CO and CO2 using an EC-QCL*, **123**, 144 (2017)
5. D.I. Ellis, W.B. Dunn, J.L. Griffin, J.W. Allwood, R. Goodacre, Pharmacogenomics J., *Metabolic fingerprinting as a diagnostic tool*, **8**, 1243 (2007)
6. F. Fischer, A. Birk, K. Frenner, A.M. Herkommer, *FeaSel-Net: A Recursive Feature Selection Callback in Neural Networks* (2022)