Investigation of novel methods for hyperspectral data analysis in multiphoton microscopy

Maciej Barna[*#a], **Jakub Bogusławski**[a], **Grażyna Palczewska**[b,c], **Krzysztof Palczewski**[c], **Grzegorz Soboń**[a]

[a] *Laser and Fiber Electronics Group, Faculty of Electronics, Photonics and Microsystems, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*
[b] *Department of Medical Devices, Polgenix, Inc., Cleveland, OH, USA*
[c] *Department of Ophthalmology, Gavin Herbert Eye Institute, University of California, Irvine, Irvine, CA, USA*
[#] *corresponding author email: maciejbarna@gmail.com*
[*] *Presenting author*

1. Introduction

Multiphoton microscopy utilizes nonlinear interactions of light with a sample to create image contrast and is used in biomedical research [1]. The most popular modality is two-photon excited fluorescence (TPEF), where fluorophores are excited by quasi-simultaneous absorption of two lower-energy photons, in contrast to classical fluorescence microscopy, where one photon carries enough energy to excite the molecule [2]. Currently, the most commonly used setup includes a Ti:Sapphire laser, a scanning mechanism, and a photomultiplier tube (PMT) [2] for counting emitted photons on each scanning position, which is arranged in a spatial grid to create a grayscale, intensity-based image (Fig. 1a). However, when multiple fluorophores are present in the examined sample, they cannot be distinguished.

Hyperspectral microscopy solves this problem by expanding analytical capabilities by recording the fluorescence spectrum of each pixel using a sensitive spectrometer. Data recorded across multiple spectral channels are stored in a 3D tensor, with two dimensions for spatial positions and one for spectral channels [3]. A large data volume makes manual analysis highly inefficient. To solve this issue, automated analysis utilizing machine learning algorithms can be used. A few methods using algorithms such as K-Means [4] and non-negative matrix factorization (NMF) [5] were developed to perform tasks such as clusterization. These algorithms are robust and versatile, serving as a starting point for automated analysis. State-of-the-art methods, such as the NMF-based ImageJ plugin [5], work with full spectral data, improving processing speeds but the efficiency is still limited due to large data volumes.

A natural solution to these inefficiencies is to use data in phasor form as input, as gathering data directly in phasor form can lower operational costs and speed up the process [3]. A phasor plot is a polar plot contained within a unit circle, where each data point has only two values, reducing data volume compared to full spectral form [6]. Conversion is achieved with the phasor transform, the main idea of which is to integrate a product of fluorescence spectra registered for each spatial position and sine and cosine functions and normalize it to the integrated intensity on this position [6]. However, there is a concern that phasor form might not be able to replace full spectral data due to the possibility of data oversimplification in the process of phasor transformation. We show the equivalence of results obtained with both full spectral and phasor forms of data as the input, using NMF as the reference algorithm performing clusterization. Results obtained with our method are compared with results generated with the mentioned NMF-based ImageJ plugin.

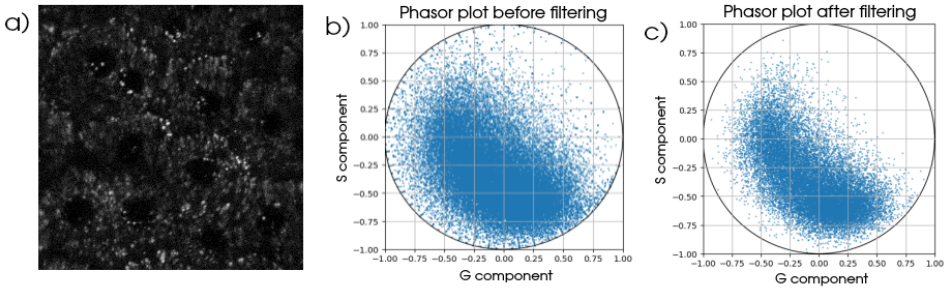

Fig. 1. **a)** Image created by integrating 3D tensor over spectral channels. Additional gamma correction was applied for better presentability. **b)** Phasor plot presenting all pixels of an image of size 256×256 (65,536 points). Data were not filtered before achieving this result. **c)** Phasor plot filtered with Poisson background separation. Data points labeled as background noise were removed, reducing the data set's size to 21,074 points (around 30%).

2. Methods and results

Data used in this analysis are TPEF images of mouse retina, obtained in albino $Abca4^{PV/PV}$ $Rdh8^{-/-}$ mouse without staining. The mouse is a genetic model of Stargardt disease [7]. The image was obtained in vivo using a multiphoton microscope (Leica TCS SP8). The excitation wavelength was 730 nm. The data set consists of a 3D tensor of shape 256×256×24. The registered spectral range cover 400 – 640 nm, which was recorded within 24 spectral channels with 10 nm per channel. Figure 1a shows an intensity-based image obtained after integrating photon counts in all wavelength channels. The image shows retinal pigment epithelium cells. At this excitation wavelength, two major fluorescence sources are expected: retinyl esters and A2E, a major

component of lipofuscin which is a by-product of the visual cycle; however, they cannot be distinguished on the intensity-based image. Figure 1b shows a phasor plot presenting unfiltered data, forming a large cloud of points. Further data analysis involves filtering and clusterization whose objective is to divide all data points into groups representing present fluorophores. The same data set, after filtration, is presented in Fig. 1c. As for filtering, Poisson background separation [8] was used to divide all pixels into *background noise* (noise only) and *signal* (useful signal mixed with noise) groups, where only signal pixels were used as input for NMF.

The main idea of the conducted tests was to process data in both full spectral and phasor forms and compare the results with those obtained using the ImageJ NMF plug-in to test potentially excessive data loss from phasor reduction. Since the ImageJ plug-in implements the NMF algorithm, it was also implemented in this work. Figure 2a shows the ImageJ NMF results for full spectral data, with all pixels assigned to one of the fluorophores. Due to the presence of low-level noise, there are no neutral pixels. Green pixels (retinyl esters) are located mostly on the borders between the RPE cells, whereas red (A2E) are primarily located inside the cells. Figure 2b displays results from a custom NMF algorithm for full spectral data, with better background noise rejection, as there is no low-level noise. Full spectral data were transformed with phasor transform and used as an input for custom NMF. Figure 2c presents results of the clusterization using data in phasor form. The emission peak of A2E is located around 600 nm to 650 nm [9]. The group of points on the phasor plot corresponding to this peak is marked with red color. The second fluorophore, which is retinyl esters, has the emission peak around 500 nm to 530 nm [9], which is represented by a green color. Clusters marked with red and green colors are presented on integrated image in Fig. 2d. Custom NMF implementation yielded similar results for both data forms used. Over 90% of pixels had consistent labeling, with most differences in areas with similar fluorophore percentages or low-intensity noise. The ImageJ results had additional noise, which complicates comparison, however after removing the noise, similarity between results from ImageJ and custom NMF with phasor data form was around 70%.

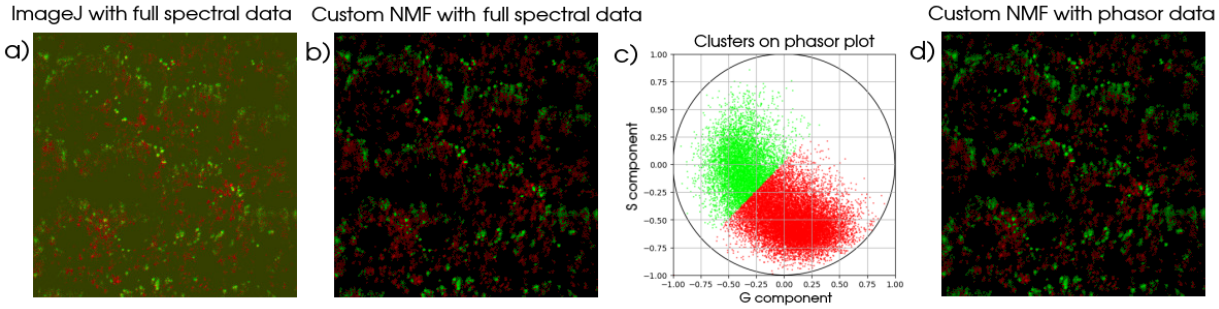

Fig. 2. **a)** Result obtained with ImageJ using full spectral data as input, **b)** Result obtained with custom NMF implementation using full spectral data as input, **c)** Phasor plot presenting results of NMF algorithm related to Fig. 2d, **d)** Result obtained with custom NMF implementation using phasor data as input. All methods used data filtered with Poisson background separation.

In conclusion, all 3 methods obtained similar results in terms of labeling pixels. There are many factors, like filtering methods, contributing to the differences in labeling, however it was shown, that phasor data format achieves data volume reduction without oversimplification, offering a reduction of needed resources without sacrificing the accuracy of data analysis, resulting in qualitatively comparable results.

3. Funding

National Science Centre, Poland (2021/43/D/ST7/01126)

4. References

[1] Bogusławski, J., Kwaśny, A., Stachowiak, D., and Soboń, G., "Increasing brightness in multiphoton microscopy with a low-repetition-rate, wavelength-tunable femtosecond fiber laser," Opt. Continuum **3**, 22 (2024).

[2] Zipfel, W.R., Williams, R.M., and Webb, W.W., "Nonlinear magic: multiphoton microscopy in the biosciences," Nat. Biotechnol. **21**, 1369–1377 (2003).

[3] Hedde, P.N., Cinco, R., Malacrida, L., Kamaid, A., and Gratton, E., "Phasor-based hyperspectral snapshot microscopy allows fast imaging of live, three-dimensional tissues for biomedical applications," Communications Biology **4**, 721 (2021).

[4] Vallmitjana, A., Torrado, B., and Gratton, E., "Phasor-based image segmentation: machine learning clustering techniques," Biomed. Opt. Express **12**, 3410 (2021).

[5] Neher, R.A., Mitkovski, M., Kirchhoff, F., Neher, E., Theis, F.J., and Zeug, A., "Blind source separation techniques for the decomposition of multiply labeled fluorescence images," Biophys. J. **96**, 3791–3800 (2009).

[6] Malacrida, L., Gratton, E., and Jameson, D.M., "Model-free methods to study membrane environmental probes: a comparison of the spectral phasor and generalized polarization approaches," Methods Appl. Fluoresc. **3**, 047001 (2015).

[7] Tanna, P., Rupert, S.W., Kaoru, F., and Michel, M., "Stargardt disease: clinical features, moleculargenetics, animal models and therapeutic options," Br. J. Ophthalmol. **101**, 25-30 (2016).

[8] Straasø, T., Müter, D., Sørensen, H.O., and Als-Nielsen, J., "Objective algorithm to separate signal from noise in a poisson-distributed pixel data set," J. of Appl. Crystal. **46**, 663–671 (2013).

[9] Palczewska, G., Bogusławski, J., Stremplewski, P., Kornaszewski, D.L., Zhang, J., Dong, Z., Liang, X.-X., Gratton, E., Vogel, A., Wojtkowski, M., and Palczewski, K., "Noninvasive two-photon optical biopsy of retinal fluorophores," Proc. Natl. Acad Sci. **117**, 22532-22543 (2020).