Savić, A. G., Zivković, S., Jovanović, K. K., Duponchel, L., and Kopriva, I. (2015) Complete determination of plant tissues based only on auto-fluorescence and the advanced image analysis – study of needles and stamens. J. Chemometrics, 29: 521–527. Doi: 10.1002/cem.2735. - accepted version

Complete determination of plant tissues based only on auto-fluorescence and the advanced image analysis – study of needles and stamens

Aleksandar G. Savić[1], Suzana Zivković[2], Katarina K. Jovanović[3], Ludovic Duponchel[1], Ivica Kopriva[4]

[1] LASIR CNRS UMR 8516, Université Lille 1, Sciences et Technologies, 59655 Villeneuve d'Ascq Cedex, France

[2] *Institute for Biological Research "Siniša Stanković", University of Belgrade, Bulevar Despota Stefana 142, Belgrade, Serbia*

[3]*Institute for Oncology and Radiology of Serbia, Pasterova 14, Belgrade, Serbia*

[4] *Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, Croatia*

Corresponding author information:

Name: Aleksandar G. Savic

E-mail: asavic@imsi.rs , enzo1320@yahoo.com

Phone: +33-068076-9455

Institution:

LASIR CNRS UMR 8516, Université Lille 1, Sciences et Technologies, 59655 Villeneuve d'Ascq Cedex, France

Short Title: Determination of plant tissues, based on advanced image analysis

Keywords: image analysis, explicit feature maps, ℓ_0-norm constrained nonnegative matrix factorization, fluorescence microscopy

Table of Content Abstract

Complete determination of plant tissues based only on auto-fluorescence and the advanced image analysis, study of needles and stamens

Aleksandar G. Savić[1], Suzana Zivković[2], Katarina K. Jovanović[3], Sladjana Z. Spasić[1], Ivica Kopriva[4]

Finding the procedure for proper tissue determination is a challenging task in various fields of biology and medicine, unfortunately often affected by subjectivity of the histologist. We have introduced the method based on ℓ_0-norm constrained nonnegative matrix factorization and compared it with several algorithms on controlled, simulated set of images. The key advantage is the ability to extract much more components from the starting set of images (often small number) in comparison to other methods, providing more versatile set of parameters for tissue classification.

Savić, A. G., Zivković, S., Jovanović, K. K., Duponchel, L., and Kopriva, I. (2015) Complete determination of plant tissues based only on auto-fluorescence and the advanced image analysis – study of needles and stamens. J. Chemometrics, 29: 521–527. Doi: 10.1002/cem.2735. - accepted version

SUMMARY

Proper determination of tissues is one of the challenging problems is modern medicine and histology. Currently, interpretation of the results mainly depends on the experience of histologist, leading to high percentage of results misinterpretation. Bearing in mind potential application, we proposed the set of procedures that allow us to obtain precise, mathematically determined parameters for tissue discrimination. First, the method was tested on simulated set of images and compared with several other algorithms. As the set of experimentally obtained input data, auto-fluorescence images of needle cross-sections (*Picea omorika*) and stamens of common centaury (*Centaurium erythraea*), were used. Determination of cell types is based on inherent features of plant cells – autofluorescence. As each cell type consists of various fluorescent components in different quantities for each type of tissue, its integral emission spectrum can be used as the fingerprint for identification. Cross-sections were imaged using 4 sets of filters for detection of fluorescence (both excitation and emission). Such filter set is standard equipment for most fluorescence microscopes. One additional image was transmission image using the same optics. By applying ℓ_0-norm constrained nonnegative matrix factorisation in a space induced by explicit feature maps, it is possible to identify up to 11 tissues in needles and 5 in stamens (actual number of tissues). In comparison to other image analysis methods, the greatest advantage is the fact that number of extracted components significantly exceeds the number of initial images, while most other techniques can extract only as much components as the number of initial images.

1. INTRODUCTION

The progress in development of fluorescence probes for various metabolic parameters of the cell (concentration of ions, pH, cell currents, free radicals, immunostaining etc.) has provided significantly better insight for understanding of processes in cells, but also emphasized the need for proper identification of cells and tissues.

In the past two decades, there is a noticeable increase of interest in analysis of spectral data in the emerging new discipline in analytical chemistry – the chemometrics. The main concept is the idea that each composite spectrum of the specimen can be represented as the linear combination of pure components. Challenges are, however, numerous. The first is the fact that in most cases analysis starts as blind source separation problem (BSS). In such scenario spectra and abundances of the components as well as their number are unknown. Separation of pure components becomes even harder if the pure components are not statistically independent , but highly correlated . This is especially notable in fluorescence spectroscopy. The problem becomes even more difficult if spatial distribution of the components is taken into account . Thus the starting point is blind, and proper method should identify the exact number of pure components, taking into account their interdependence and estimate their spatial distribution.

Depending on the properties of input spectra various methods can be used for solution of the problem described above. Factor analysis, including related methods, is particularly suitable for fluorescence and FTIR spectra [1, 2]. Its drawback is the fact that if the number of components is n, the number of required spectra for the analysis is at least $2n+1$. Blind source separation (BSS) methods such as independent component

96 analysis (ICA), sparse component analysis (SCA) and/or nonnegative matrix
97 factorisation (NMF) have found application in analysis of electron paramagnetic
98 resonance (EPR), nuclear magnetic resonance (NMR), electroencephalography (EEG)
99 and spectroscopic modalities [3-5].

100 If the desired number of pure components is the same as the number of input
101 component mixtures (experimentally obtained spectra or images), some other
102 algorithms are suitable. Among them, the best known is independent component
103 analysis (ICA). ICA based methods are however incapable to solve underdetermined
104 BSS problems, often present in image analysis. Moreover, ICA algorithms assume
105 statistical independence between the components and that, due to sum-to-one
106 constraints as well as due to overlapping of the abundances, is violated in multispectral
107 and hyperspectral image analysis [6]. Furthermore, ICA does not utilize non-negativity
108 that should be considered in fluorescence spectroscopy because of its physical
109 background. SCA and sparseness constrained NMF algorithms can handle
110 underdetermined BSS problems but exhibit difficulties in discrimination between pure
111 components with similar spectral, concentrations or density profiles. Described
112 problems are partially addressed in algorithms popular in remote sensing and the
113 analysis of satellite images that are based on pure pixel assumption [7]. The
114 representatives of such methods are N-FINDR [8], simplex volume maximization
115 (SVMAX) [9] and vertex component analysis (VCA) [10]. However, all these
116 algorithms demand that number of components is less than or equal to number of
117 spectral bands and that is violated in fluorescence microscopy imaging of plant tissues.

118 Analysis of fluorescence images resembles in some manner to the paradigm for
119 analysis of emission fluorescence spectra. The main principle is to obtain the series of

emission spectra after irradiating the sample with different excitation wavelengths. As fluorophores [11] provide different quantum yields (and consequently emission intensities) when excited with determined wavelength, series of excitations will indicate their presence as the changes of integral spectrum (caused by changes in its relative contributions to integral spectrum). Fluorescence microscopes are usually equipped with several excitation/emission filter sets, providing in our case 4 fluorescent images and one transmission image.

To discriminate pure components with similar emission spectra we propose unsupervised (a.k.a. blind or automatic) decomposition of 5-channel image by means of ℓ_0-constrained NMF algorithm in a space induced by explicit feature maps (EFM). The nonlinear EFM mapping increases number of spectral bands and makes spectral profiles of the pure components less correlated. That in combination of sparseness and nonnegativity constraints enables separation of large number of spectrally similar pure components from small number of available spectra.

Another feature of the cell that can be used as the parameter for tissue determination is the shape of the cells. Images can be segmented according to textural features, using variety of methods, and then apply described technique to distinguish various biochemical changes (fluorescence properties) in tissues built from the same cells in mean of shape. As the quantitative measure of cell shape, the 2-D generalization of Higuchi's fractal dimension can be applied as recent novelty, but classical Haralick texture features (that include entropy, standard deviation and about 20 other parameters) can be applied as well [12]. Our tests were performed on whole, raw images. Applied ℓ_0-constrained NMF algorithm was so effective, that there was no need for any pre-processing based on textural features extraction as the pre-processing step. Without the

144 example obtained from real and complex enough measurements this improvement
145 remains presumable assumption.

146 The drawback of the ℓ_0-constrained NMF algorithm is also characteristic for other data
147 driven algorithms mentioned before - reproducibility of the results is influenced by
148 matching between the model and true experimental image [7] and, in case of the NMF
149 algorithms, by non-convexity of the related factorization problem. The former case
150 relates to pixel-to-pixel variability of spectral signature and presence of sensor noise [7].
151 Thus, if the algorithm is applied on several similar images the extracted components
152 among the samples are similar (especially several leading components) but not always
153 the same. This problem could be overcome by analysis of several images
154 simultaneously using simple augmentation of images and then simultaneous analysis of
155 all the images. These limits make more difficult further application of learning methods
156 that could additionally automate the process. Besides of that limitation, the
157 classification can be still done leaving only the final step of the interpretation of results
158 (which cluster represents which tissue) to the observer.

159 All the tissues (11, plus background in case of needle and 5 in case of stamen) were
160 successfully determined, based solely on chemical properties of the cells (presence of
161 fluorophores), without any pre-knowledge about the number of tissues, or its
162 fluorescence properties. It is important to notice that the algorithm did not artificially
163 made false positive identification of tissues that are not anatomically correct.

164 In order to additionally check this claim, we have tested the method on simulated set
165 of images and compared proposed method with the several other popular methods
166 including independent component analysis (ICA) and principal component analysis
167 (PCA) (on image set with extended dimensionality). This problem can be present in

real, experimental data sets because the same material (the same spectral properties) can be presented in various intensities, thus the suitable method must provide proper determination of qualitative properties of the image elements.

With the introduction of even more complex samples, ℓ_0-constrained NMF algorithm can be applied following any of numerous pre-processing techniques that can facilitate the analysis by selecting specified regions of interest that can be separately analysed. With such freedom in choice of analytical techniques, possibilities are limited by sample complexity and the ability of instruments to record larger set of input images. Such devices, micro-spectrometers, are still mainly custom-made but the development of experimental techniques is in constant progress.

2. MATERIALS AND METHODS

Fluorescence microscopy

Fresh needles of *P. omorika* were collected, tight cross-sections were made by a razor blade, put on microscopic glass, and examined using Zeiss Axio Observer Z1 microscope, equipped with AxioCamMR3 camera (8 bit per channel) (Zeiss, Göttingen, Germany).

Stamens of common centaury (*Centaurium erythraea* Rafn) were removed from the flowers and imaged without any particular preparation. The same optics was used for all recorded images with total magnification of 50 x. The exposition time was set to provide the optimal dynamic range. The applied excitation/emission filter sets for

fluorescence images were: 358/461 nm (DAPI), 492/518 nm (FAM), 488/510 nm (38 GFP) and 558/580 nm (DsRED). Additional transmission image was obtained by illuminating the sample with white light. No additional staining was performed; the imaging was based only on autofluorescence.

IMAGE ANALYSIS

Sparse nonnegative matrix factorization in explicit feature maps induced space

Unsupervised (a.k.a blind) approaches to multichannel image decomposition are commonly based on the linear mixture model (LMM) of multichannel image [13, 14]:

$$\mathbf{X}=\mathbf{AS} \tag{1}$$

where $\mathbf{X}\in \mathrm{R}_{0^+}^{N\times T}$ represents multi-spectral fluorescent image consisting of N spectral bands and $T=P\times Q$ pixels, $\mathbf{A}\in \mathrm{R}_{0^+}^{N\times M}$ represents mixing matrix or matrix of spectral responses and $\mathbf{S}\in \mathrm{R}_{0^+}^{M\times T}$ represents matrix of spatial distributions of the M pure components present in the image scene. Each row of **X** and **S** is a 1D image representation obtained from corresponding spectral image by some 2D→1D mapping called vectorization. Blind decomposition of pure components $\left\{\mathbf{s}_m \in R^{1\times T}\right\}_{m=1}^{M}$ using multichannel image matrix **X** only is also known as blind source separation (BSS)

problem [15]. Since we are concerned with an unsupervised image decomposition problem **X**, **A** and **S** are assumed to be nonnegative according to physical properties of fluorescence. In the case in which the number of pure components M is greater than number of fluorescent images N related BSS problem (1) is underdetermined. In addition it is ill-posed because matrix factorization suffers from indeterminacies: $\mathbf{X}=\mathbf{AS}=\mathbf{ADD}^{-1}\mathbf{S}$ for any square invertible matrix **D**. Hence, it has an infinite number of solutions. Meaningful solutions are characterized by the permutation and scaling indeterminacies in which case $\mathbf{D}=\mathbf{P}\Lambda$, where **P** represents permutation and Λ represents diagonal scaling matrix. Constraints are necessary to be imposed on **A** and **S** to obtain solution of (1) unique up to permutation and scaling indeterminacies only. For underdetermined BSS (uBSS) problems, of interest herein, the necessary constraint (in addition to nonnegativity) is sparseness of the pure components. In particular, when spatial resolution of the microscope is high enough, it is grounded to assume that at each pixel is present one pure component only. That is reflected to hard sparseness constraint expressed in term of ℓ_0-norm as $\ell_0=1$, whereas the ℓ_0 quasi-norm counts number of non-zero coefficients of pure components $\{\mathbf{s}_m\}_{m=1}^{M}$. Thus, nonnegativity and sparseness constraints can be combined to yield sparseness constrained NMF algorithm (sNMF). In particular, ℓ_0-constrained NMF (NMF_L0) can be used to estimate pure components $\{\mathbf{s}_m\}_{m=1}^{M}$ [16]. In examined samples however, one pixel corresponds to one type of the calls, but not necessarily one type of fluorophore, as the cells are highly complex. However, the differences are sufficient enough to fulfil the presumption.

MATLAB implementation of the NMF_L0 algorithm is available at [17]. Nevertheless, NMF as well as other BSS algorithms experience difficulties to separate pure

components with (very) similar spectral profiles [13, 7]. In such a case it is possible to perform pixel-wise nonlinear mapping of the multichannel image $\mathbf{X}$ [13]:

$$\mathbf{x}_t \in R_{0+}^{N\times 1} \mapsto \phi(\mathbf{x}_t) \in R_{0+}^{\bar{N}\times 1} \quad \forall t = 1,...,T \tag{2}$$

whereas $\bar{N} > N$ and perform sNMF in mapped space according to LMM:

$$\phi(\mathbf{X}) = \bar{\mathbf{A}}\bar{\mathbf{S}} \tag{3}$$

The linear mixture model (3) is constrained by nonnegativity: $\bar{\mathbf{A}} \geq \mathbf{0}$, $\bar{\mathbf{S}} \geq \mathbf{0}$ and sparseness $\|\bar{\mathbf{S}}\|_0$ that stands for number of nonzero coefficients of $\bar{\mathbf{S}}$. Very often sum-to-one constraint $\left\{\|\bar{\mathbf{s}}_t\|_1 = 1\right\}_{t=1}^{T}$ is used in hyperspectral image analysis [7]. However, as pointed out in Section III-B in [7], due to pixe-to-pixel spectral signature variability (the same material can at different locations have different spectra) and due to presence of artefacts sum-to-one constraint is often violated in practice. However, sparseness and nonnegativity constraints are more natural and often yield better unmixing results, see Section VI in [7]. That is why NMF_L0 method has been applied to (3).

Even though nonlinear mapping can have various forms we have applied approximate explicit feature map (EFM) obtained by factorization of Gaussian kernel $\kappa(\mathbf{x},\mathbf{y}) = \exp\left(-\|\mathbf{x}-\mathbf{y}\|^2/\sigma^2\right)$ where σ^2 denotes kernel bandwidth. To this end, the so called *kernel trick* [18] has been used: $\kappa(\mathbf{x},\mathbf{y}) = \langle\phi(\mathbf{x}),\phi(\mathbf{y})\rangle$ and $\langle\phi(\mathbf{x}),\phi(\mathbf{y})\rangle$ stands

for inner product of nonlinear mappings. By using multi-index notation obtained EFM is expressed as:

$$\phi(\mathbf{x}) = \mathrm{e}^{-\frac{\|\mathbf{x}\|_2^2}{\sigma^2}} \left\{ \frac{1}{\sigma^r} \sqrt{\frac{2^r}{\boldsymbol{\alpha}!}} \mathbf{x}^{\boldsymbol{\alpha}} \right\}_{|\boldsymbol{\alpha}|=r, r=0}^{\infty} \tag{4}$$

where $\boldsymbol{\alpha} \in \mathbb{N}_0^N$, $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_N$, $\boldsymbol{\alpha}! = \alpha_1! \ldots \alpha_N!$ and $\mathbf{x}^{\alpha} = x_1^{\alpha_1} \ldots x_N^{\alpha_N}$. Approximate EFM of order d is obtained for $0 \le r \le d < \infty$. In the experiments conducted herein d=2. Thus, N=5 channels image has been mapped onto $\bar{N} = \binom{N+d}{N} = 21$ channels image. Before mapping original image has been scaled such that $\{0 \le x_{nt} \le 1\}_{n,t=1}^{N,T}$ and variance of the kernel has been set to σ^2=1.

Simulated images

In order to check the reliability of proposed procedure in controlled conditions, the set of simulated images was analysed. Biological-like structure and the gradient image were drawn in Adobe Photoshop and the logic (true-false) maps were made for each type of the tissues. Resolution of images was 512 x 512 pixels in 4 separated channels. For more realistic appearance, the gradient image that simulates variations of quantum yields of fluorophores was multiplied element wise with simulated images. Values for all simulated tissues in various channels were simulated to resemble real excitation spectra.

This test is important because it proves whether the algorithm is capable to properly interpret and recognize the tissues built from the same fluorescence material. The calculations were done using Matworks Matlab R2014a, and Matlab based software FastICA. Spectral parameters used for simulation, starting images and the results of the analyses are listed in supplementary material.

Image preparation for figures

For colour rendering of initial fluorescence images, self developed Java based program BioCIE was applied [19]. It converts numerical value of wavelength (nm) to red, green and blue channel intensities and creates the RGB image. Conversion is done according to CIE convention. Advantage in comparison to standard method that considers grayscale images as channels is the ability to use any number of input images. Extracted pure components were considered as channels in RGB images, offering the possibility to accurately render the contributions of best suitable components as tissue properties.

3. RESULTS AND DISCUSSION

The test image comprises of 15 various tissues resembling the biological structures. Major difficulty for tested algorithm was the presence of gradient which can occur in

biological structures where the same building material can be present in various quantities in different areas of image. In the case of low dimensionality image set, it emerges to be difficult for all the algorithms to distinguish whether the structure is qualitatively the same or not. The results are presented in Figure 1.

The first class of algorithms based on factor analysis exhibited very low performances. The obtained results are almost the copies of initial images. Neither gradient nor the fine structure of the simulated tissues were recognized.

The ICA provided generally speaking good results, but the closer examination show that some of the structures that are different were recognized as one structure (Figure 1A, central region, numbers 14 and 15). The gradient was well recognized on some parts of the image, but emerged in false positive discrimination on the other sectors. According to this, we conclude that ICA algorithm was partially applicable.

The method we proposed in the paper provided the best result in this comparison, achieving to observe fine differences on simulated membranes successfully avoiding the gradient trap. Not perfect, with small misinterpretation in of the most similar tissues, but the most reliable. It should be kept in mind that the problem was underdetermined, and that the simulation was more difficult for solving (very strong gradient, very similar simulated spectra and number of tissues higher than expected in real experiments).

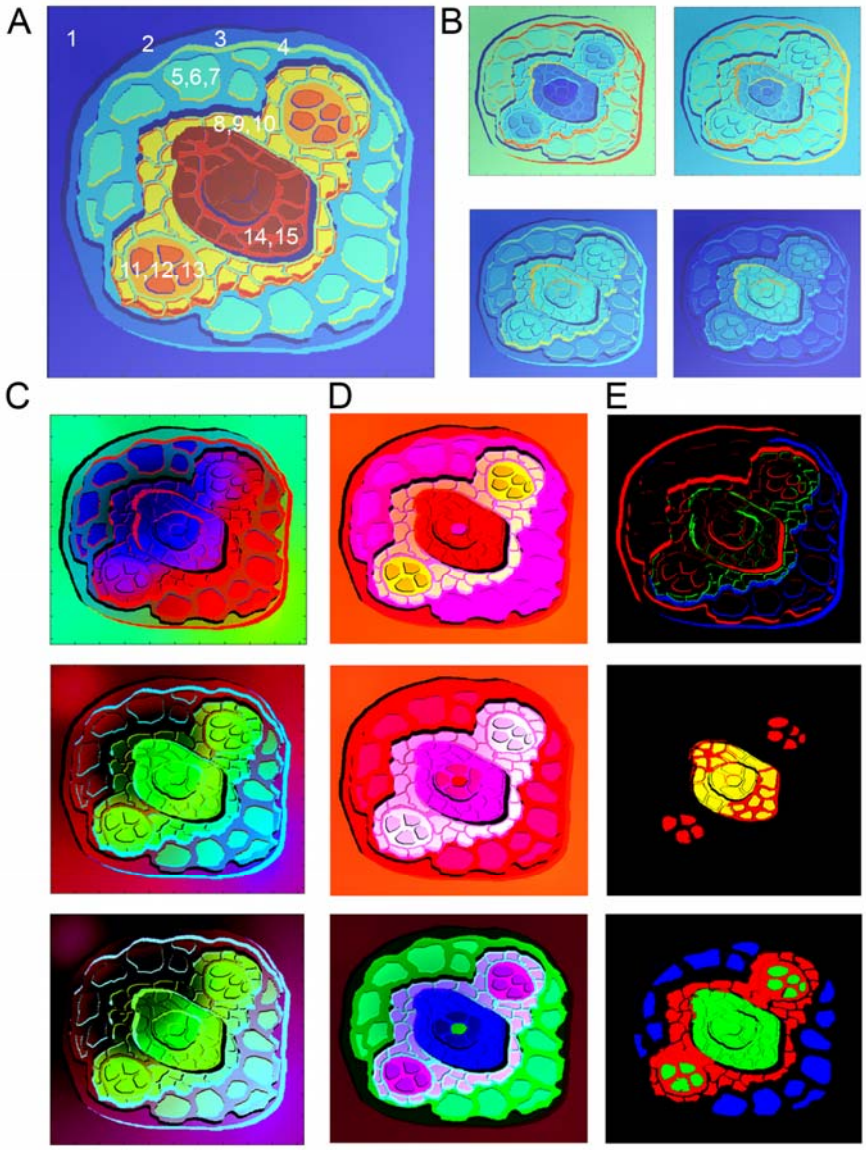


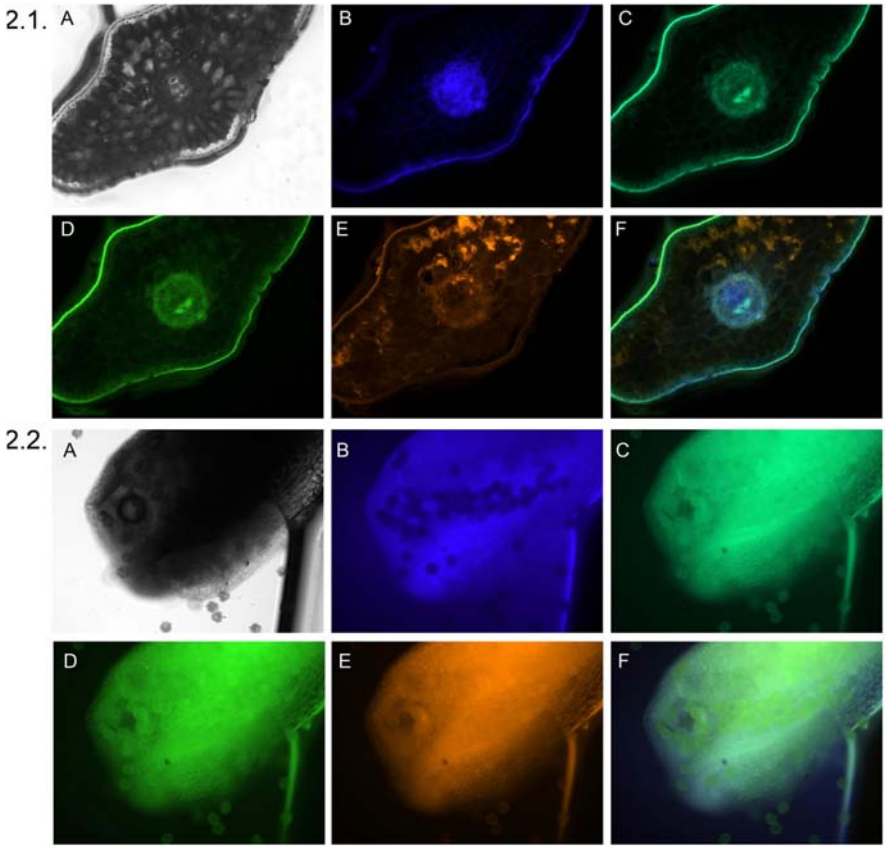
Figure 1. A – test image with results that are expected. B – 4 images that were used as the entry set of images. C – Results obtained using factor analysis, without rotation of factor scores, using orthogonal varimax method and oblique promax method (dimensionality extended by addition). D – results of ICA analysis, 4 components mixed as RGB channels (3 combinations). E – results obtained using the procedure described in our paper. Numbers on image A were used arbitrary to mark the simulated tissues and facilitate the discussion of results.

333 The results could be discussed from the spectroscopic point of view as well. The ICA
334 algorithm performed well if we speak about the images. But, as the number of extracted
335 components is only 4, while 15 tissues exist, the observed components are
336 misinterpreting the real spectra that correspond to specific tissues. Proposed method has
337 the additional advantage from this point of view as it extracts the number of components
338 much closer to the actual number of components.

339 After the methods proved their reliability on test set of images, the analysis was
340 performed on real experimental data.

341 Transparency and fluorescence images (4) of leaf cross-sections and stamens were
342 used as the input set of data for the further analysis are presented in the Figure 2.
343 Brightness and contrast were set for best appearance in figure, while the raw images
344 were used for further analysis.

345



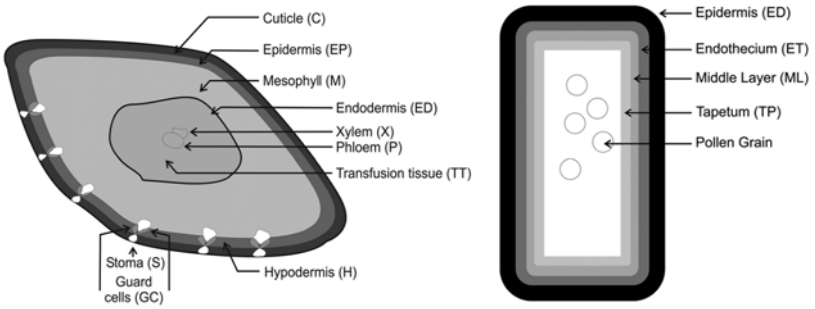
346
347

348 **Figure 2.** 2.1. – Cross-sections of Serbian spruce (*Picea omorika*) leafs. A –
349 transmission image, B-E – fluorescence images obtained by application of DAPI, FAM,
350 GFP and DsRED filter respectively. F – composite image made using fluorescence
351 images according to CIE convention. 2.2. – images of common centaury (*Centaurium*
352 *erythraea* Rafn) stamen captured using the same filter sets

353

354 Some of the tissues can be identified from the raw images, among them the layer of
355 cuticle, mesophyll cells, stomas and central vascular bundle. The resin artefact is visible
356 on red fluorescence image (Figure 2.1. E). However, the boundaries between the tissues
357 remain uncertain, and delicate structures of tissue organization are lacking. Schematic
358 view of leaf structure is presented in Figure 3.

359 In case of stamens, it is even more difficult to clearly determine which tissue types
360 are present. Only the pollen grains are easily distinguishable.
361
362



363
364

365 **Figure 3.** A – Schematic representation of *Picea omorika* leaf cross-section structure.
366 Cuticle (C), epidermis (EP), mesophyll (M), endodermis (ED), xylem (X), phloem (P),
367 transfusion tissue (TT), hypodermis (H), stoma cells (S), guard cells (GC). B –
368 Schematic representation of common centaury (*Centaurium erythraea* Rafn) stamen
369 structure: epidermis (ED), endothelium (ET), middle layer (ML), tapetum (TP) and
370 pollen grains.
371

372 After dimensionality extension and the application of ℓ_0-norm constrained NMF, 21
373 images were extracted. Three of them represented the background only (with artefacts
374 originating from particles and surrounding water), while the rest of images were used
375 for further analysis. Closer look to extracted images reveals presence of previously
376 hardly detectable structures such as individual components of central vascular bundle,
377 namely, xylem, phloem and transfusion tissue bordered by endodermis (Figure 4a).
378 Furthermore, stoma guard cells can be clearly distinguished from the surrounding
379 hypodermis tissue.

Histology of stamen is much simpler, and the images revealed: connective, recognized to be made from the similar material as epidermis of the anther, deeper layer that consists of endothelium, middle layer and tapetum, and the pollen grains, with even internal structures that can be observed on some grains.

Visually richer results can be easily obtained if the images of pure components were combined into a composite RBG image, where each channel represents one extracted image. Suitable combinations of images reveal all the tissues existed (Figure 4). Even more important, the same histological structure can be described by (different) multiple combinations of extracted images. This leads to a multidimensional space for accurate tissue segmentation, so that each tissue can be determined by several parameters in the linear combination of extracted images.

This step in the analytical procedure could sounds controversial as we pick only some combinations of pure components that support the classification of the tissues. In practice, such paradigm is very common, especially in diagnostics when one or few parameters define such important classes as health or illness. Reason for this also lies in the fluorescence properties of the cells. Most of the cells share various common fluorophores. Generally speaking, when we apply most of spectroscopic techniques on biological samples we observe complex mixtures of signals shared among all the cells various in origin.

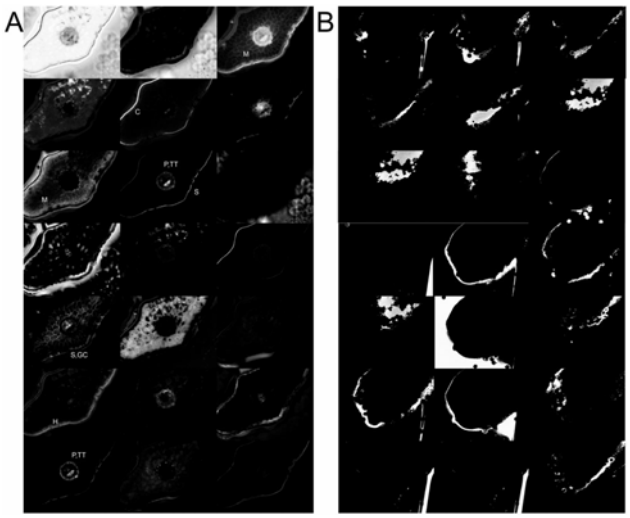
Each plant cell consists of complex polymers based on sugars followed by proteins and nucleic acids that have auto fluorescence. Large differences could appear if some type of cells contains special pigments. With such complex spectra it is impossible to obtain compounds which are pure in chemical meaning (one pure compound – one chemical compound), but instead proper analytical method will emphasise important

404 differences between very similar types of spectra. At this point the algorithms show its
405 full potential, as method should be sensitive enough to capture the smallest differences,
406 and yet robust enough not to produce the artefacts.

407 Keeping that in mind, combining multiple pure components to composite images
408 becomes more intuitive, as this procedure allows construction of long true-false string
409 that can describe and classify even the tissues that share almost entirely correlated
410 spectra.

411

412



413
414

415 **Figure 4.** Pure components separated by NMF_L0 algorithm on dimensionality
416 expanded multichannel image of needle (A), and the pure components of stamen (B).
417 Abbreviations were used to emphasize some of the extracted structures. Check Figure 3
418 for more details.

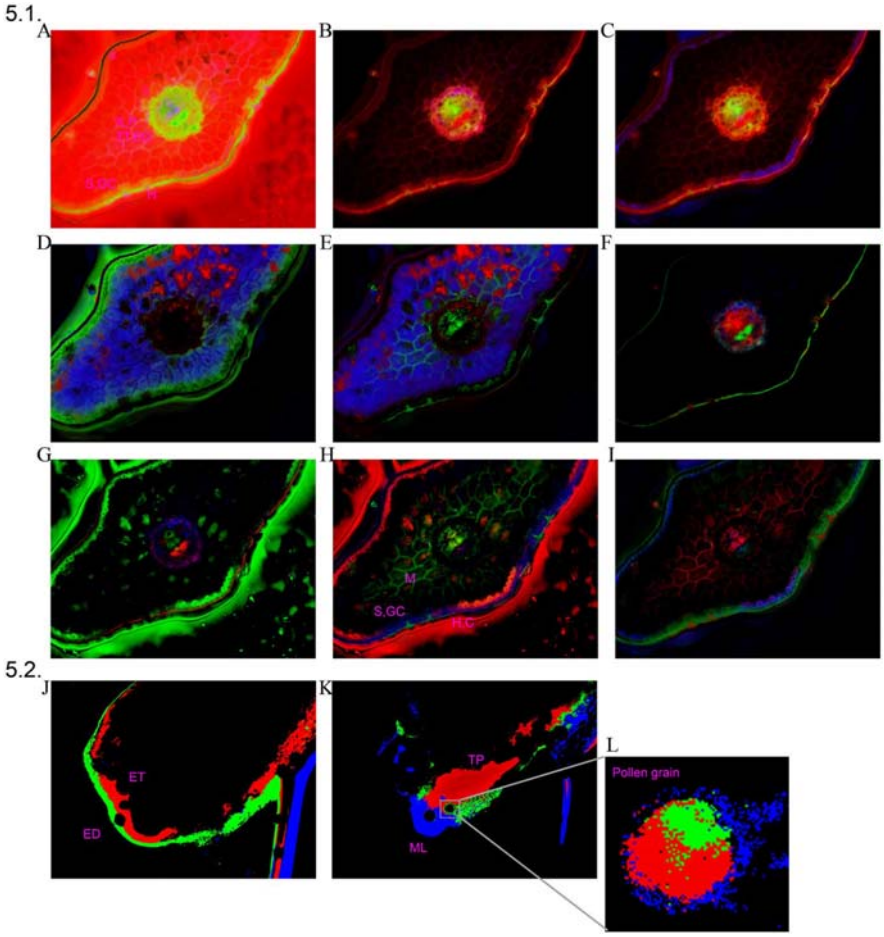


Figure 5. 5.1. A – Composite RBG images obtained by combining some of the extracted components from Figure 4, as channels. Each combination (A-I) uncovers cell types and describes it (check Figure 2 for abbreviations). **A** = C+H+ED+TT+X+P+S+GC, **B** = C+E+H+M+ED+X+S, **C** = C+E+H+M+ED+TT+X+P+S+GC, **D** = C+H+M+GC, **E** = C+M+X+S+GC, **F** = C+ED+P+TT+S, **G** = C+H+ED+X+P, **H** = C+E+H+M+TT+X+S+GC, **I** = C+E+H+M+TT+X+P+S+GC. 5.2. – **J** = ED+ET, **K** – ML+TP, **L** – pollen grain.

Examples from the Figure 5 are just a small fraction of many combinations with a strong biological meaning chosen because of visual effect. However, these 9 images presented on figure 5a reveal all 11 types of cells in needle with very high level of confidence based on anatomical meaning of extracted structures. Resin canals and some

artefact elements are presented in background. Histology of stamen is less complex, well presented on set of reconstructed composite images in figure 5b. On some grains, even internal structures can be observed, Figure 5l, blue – exine and intrine, red – middle part of the pollen grain and green shows two structures, generative nucleus (larger) and pollen tube nucleus (smaller). The reason why the complete determination of tissues in stamen was more difficult even the number if tissues was smaller, is the fact that the fluorescence properties among various tissues are less expressed, as the tissues in stamen do not have so different biological roles, as in needles.

4. CONCLUSION

Described procedures of image analysis offer numerous possibilities for improvement of virtually all imaging procedures that capture the same object using different instrumental settings. Besides the fluorescence microscopy, it is applicable in light microscopy, MRI (if different sequences were applied) or satellite images for environmental sciences. As the development of methods is in constant progress and more spectroscopic techniques become coupled with the microscopy, it is expected that the demand for procedures for analysis of multidimensional images will be in constant growth.

ℓ_0-norm constrained nonnegative matrix factorization in space induced by explicit feature map was proven to be the method of choice for BSS based analysis of images with good (high) spatial resolution and highly correlated spectra of the pure components. From only 5 starting images, algorithm needed to be able to find the exact

456 number of components that exceeds the number of input images by factor of more than
457 2. Next, it needed to distinguish among the pure components that are not independent
458 but highly correlated, sharing the same spectral form, as majority of fluorophores are
459 present in each pixel on image, only in different ratios. It means that often only the
460 contribution of individual fluorophores affects the final result as the pixel brightness
461 intensity. And finally, task should be proven for tissue discrimination in anatomically
462 right manner.

463 We also suggested that ℓ_0-norm constrained nonnegative matrix factorization in
464 mapped space can be applied as one module in whole analytical procedure that can be
465 applied not only on raw images, but instead on images with selected ROI. Such
466 approach could include image segmentation, clustering based techniques, textural
467 parameters, filters or fractal based techniques. Future work will be focused on
468 application of described procedure on various image sets originating from different
469 instruments. Special attention will be given to micro-spectroscopy, including
470 fluorescence and Raman micro-spectroscopy.

471

472

473 **List of abbreviations:** cuticle (C), epidermis (EP), mesophyll (M), endodermis (ED),
474 xylem (X), phloem (P), transfusion tissue (TT), hypodermis (H), stoma cells (S), guard
475 cells (GC), epidermis of stamen (ED), endothelium (ET), middle layer (ML), tapetum
476 (TP) and pollen grains.

477

478

Competing Interests: The authors declare that they have no competing interests, neither financial nor non-financial

Acknowledgments

Research was supported by grant 9.01/232 "Nonlinear component analysis with applications in chemometrics and pathology" funded by the Croatian Science Foundation, Federation of European Biochemical Societies (FEBS) fellowship, Serbian Ministry of Science, Education and Technological development, projects III41026 and III43010.

References

1. Stedmon CA, Bro R. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnol. Oceanogr. Methods.* 2008; **6**:5–779.

2. Yang H, Irudayaraj J, Paradkar MM. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food. Chem.* 2005; **93**(1):25–32.

3. Oja E, Hyvarinen A. Independent component analysis: algorithms and applications. *Neural. netw.* 2000; **13**(4–5):411–430.

4. Georgiev P, Theis F, Cichocki A. Blind source separation and sparse component analysis of overcomplete mixtures. In: *Acoustics, Speech, and Signal Processing, Paper presented at the proceedings of the IEEE International Conference on (Volume: 5) (Canada), Montreal. 2004.* IEEE. 493–496.

5. Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data. Anal.* 2007, **52**(1):155–173.

6. Nascimento JMP, Bioucas Dias JM. Does independent component analysis play a role in unmixing hyperspectral data? *Geoscience and Remote Sensing, IEEE Transactions on*. 2005, **43**(1):175–187.

7. Bioucas-Dias A, Plaza N, Dobigeon M, Parente Q, Gader P, Chanussot J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, *IEEE J. Sel. Topics Appl. Earth Observ.* 2012, 5:354–379.

8. Winter Michael E. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation,* International Society for Optics and Photonics. 1999, 266-275.

9. Chan TH, Ma WK, Ambikapathi A, Chi CY. A simplex volume maximization framework for hyperspectral endmember extraction. *Geoscience and Remote Sensing, IEEE Transactions on*. 2011, **49**(11):4177–4193.

10. Nasciment JMP, Bioucas-Dias JM. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing*, *IEEE Transactions on.* 2005, **43**: 898-910.

11. J. R. Lakowicz,Principles of Fluorescence Spectroscopy(Springer, 2006).

12. Miyamoto, E., & Merryman, T. (2005). Fast calculation of Haralick texture features. *Human computer interaction institute, Carnegie Mellon University, Pittsburgh, USA. Japanese restaurant office*.

13. Kopriva I, Cichocki A. Blind decomposition of low-dimensional multi-spectral image by sparse component analysis. *J. Chemom.* 2009, **23**(11):590–597.

14. Kopriva I, Peršin A. Unsupervised decomposition of low-intensity low-dimensional multi-spectral fluorescent images for tumour demarcation. *Med. Image. Anal. 2009,* **13**:507–518.

530 15. Comon P, Jutten C. Handbook of Blind Source Separation: Independent Component
531 Analysis and Separation. In: *Comon P, Jutten C, editors*. Academic Press: Oxford, UK
532 2010.
533 16. Peharz R, Pernkopf, F. Sparse nonnegative matrix factorization with ℓ^0-constraints.
534 *Neurocomputing*. 2012, **80**:38–46.
535 17. The Robert Peharz Website **[**https://www.spsc.tugraz.at/tools/nmf-l0-sparseness-
536 constraints]
537 18. Schölkopf B, Smola A. Learning with kernels. MIT Press, 2002.
538 19. Link to download BioCIE http://download.cnet.com/BioCIE/3640-2054_4-
539 12153354.html