Spectral unmixing of multiply stained fluorescence samples

T. Pengo[1], **A. Muñoz-Barrutía**[1,2], and **C. Ortiz-de-Solórzano**[1,2]

[1]Cancer Imaging Laboratory, Centre for Applied Medical Research, Avda. Pio XII 55, 31008 Pamplona, Spain
[2]Departament of Electrical, Electronics and Automation Engineering, Universidad de Navarra (TECNUN), San Sebastian, Spain

The widespread use of fluorescence microscopy along with the vast library of available fluorescent stains and staining methods has been extremely beneficial to researchers in many fields, ranging from material sciences to plant biology. In clinical diagnostics, the ability to combine different markers in a given sample allows the simultaneous detection of the expression of several different molecules, which in turn provides a powerful diagnostic tool for pathologists, allowing a better classification of the sample at hand. The correct detection and separation of multiple stains in a sample is achieved not only by the biochemical and optical properties of the markers, but also by the use of appropriate hardware and software tools. In this chapter, we will review and compare these tools along with their advantages and limitations.

Keywords fluorescence; microscopy; spectral imaging

1. Introduction

The use of optical instruments to investigate the microscopical properties of nature has a long history and it has recently evolved into a science with a very high level of sophistication. Microscopy has been the essential tool for biology since its infancy, allowing the exploration of life with high level of detail. It has also taken full advantage of the progress in Physics by incorporating more and more advanced instruments with an increasing level of complexity. With the discovery of fluorescence, a new era began: it became possible to light up –previously invisible- structures of interest, thus achieving an unprecedented level of detail and specificity. Namely, when the mechanisms of fluorescence became better understood, researchers started using fluorescence as a mean to highlight (stain) specific subcellular structures or particular proteins of interest, and to report the expression of genes. Then, with the continuous development of staining protocols, it became possible to combine several biomarkers into an individual sample, simultaneously highlighting different structures or proteins within the same cell or tissue section. However, as more biomarkers of different colors are added to the sample, it becomes increasingly difficult to distinguish signals from each stain. A whole new research area in signal separation techniques has born to solve this problem, taking inspiration from the advances of hyperspectral imaging in geology, astronomy and chemometrics, where similar problems have already been tackled.

The technology to successfully separate the spatial distribution of each stain requires an adequate combination of hardware and software. The goal of this chapter is to summarize the principles of fluorescence, the hardware to acquire spectral fluorescence data and the software algorithms to complete the data separation. The chapter is organized accordingly: first we will introduce the mechanisms of fluorescence; then we will present an overview of different types of fluorescence acquisition hardware. Finally, we will review the most popular software algorithms applied to the acquired data in order to separate the signals from each biomarker.

1.1The phenomenon of fluorescence

Fluorescence is a phenomenon found in nature in a variety of forms, from deep-sea fish to minerals. Fluorescent molecules have the property of absorbing light and emitting it later, at a longer wavelength than that of the light they absorb. The first fluorescent protein to be isolated from a live organism was the Green Fluorescent Protein (GFP), extracted from the Aequorea victoria jellyfish [1, 2]. Until then, only synthetic proteins such as DAPI had been used. The advantage of extracting proteins from live organisms is that they can be sequenced and included into the genome of another organism without disrupting its function [3]. When illuminated with ultraviolet light, it absorbs the light and emitts green light after a short time delay. Many other fluorescent proteins were discovered since then and have been sequenced and commercialized to meet the particular needs of biologists. The need for using multiple colors arose as multiple structures had to be made visible at the same time in the same sample.

The phenomenon of fluorescence is due to a change in the energy states of the fluorescent molecule: whenever the molecule absorbs a photon of a particular wavelength range, its electron configuration changes to a state of higher energy, the excited state. This higher electron configuration is unstable so, after a few picoseconds during which the electrons rearrange themselves to a state with a slightly lower energy (relaxation), the electrons drop to the ground state. This energy drop is compensated by the emission of a photon of energy proportional to the difference between the two energy levels. The emitted photons are what we see as fluorescent light.

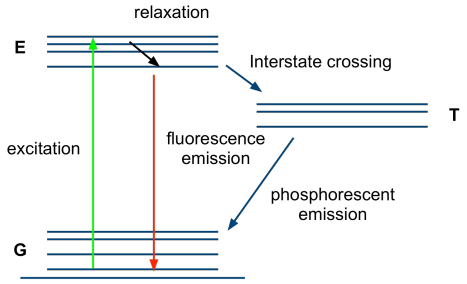

Figure 1 - Jablonski diagram. Shows the three energy states: the ground state (G), the excited state (E) and the triplet state (T). Upon excitation with, for example, green light, the molecule enters the excited state. In the excited state, it undergoes relaxation and looses some energy in reconfiguration. After a certain time, the molecule returns to the ground state simultaneously releasing its energy in the form of a photon. The photon is now of a lower energy and therefore has a longer wavelength. With a lower probability, the molecule can also perform 'interstate crossing', where the transitions to the ground state occur with much lower probability.

The energy levels and the transitions between them can be visualized on the Jablonski diagram (see **Figure 1**). Each line of the diagram represents a different energy level, or state, of the electrons of molecule. A third state is also represented in the diagram (T). This state, the 'triplet state' or 'forbidden state', is reached with a certain low probability while the molecule is in the excited state. As it has a lower energy level with respect to the excited state and transitions to the ground state occur with low probability, it is a relatively long-lived state. Eventually, upon making the transition from the triplet state to the ground state a photon is emitted, in analogy to fluorescence. When photons are emitted as they drop from the triplet state, the phenomenon is called phosphorescence.

When in an excited state, the molecule is more likely to form covalent bonds with neighboring molecules, modifying the structure of the protein and loosing the ability to emit fluorescence (photobleaching). As the triplet state is a relatively long-lived state, it is more subject to this kind of molecular interactions, so care has to be taken not to overexpose the sample and thus increase the chance of molecules ending up in the triplet state. To avoid photobleaching, samples with fluorescent proteins are often prepared in a dark environment and kept in the dark during the useful lifetime of the fluorophore.

2. Multispectral Image acquisition

Fluorescence has been used in microscopy by exploiting the difference between the excitation and the emission wavelengths of fluorescent molecules. If we place an appropriate dichroic mirror between the sample and the detector (as seen in **Figure 2**), we are able to separate the illumination –excitation- light from the emitted light. A dichroic mirror reflects all light with wavelengths shorter than a certain nominal wavelength, while transmitting all the rest. If we design the mirror to have its nominal wavelength between the excitation and the emission peak of a fluorescent molecule, we are able to isolate the light emitted by the molecule from the light we used to excite it. The resulting images will have a strong contrast between the dark background and the emission of the fluorophores.

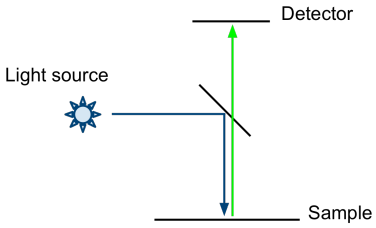

Figure 2 - The basic principle of fluorescence microscopy. Light emitted from a light source at a certain wavelength gets reflected to the sample by a dichroic mirror. The light emitted from the sample, having a longer wavelength, passes through the mirror and hits the detector.

The schematic setup for a typical widefield fluorescence microscope is shown in Figure 3**Figure 3**. The light emitted by the lamp is at first filtered through an excitation filter and reflected by the dichroic mirror onto the sample. The target fluorescent molecules present in the sample are excited and emit photons at a longer wavelength. The light is collected by the objective and, as these photons have a lower energy and longer wavelength, they are let through by the dichroic mirror. Finally, an emission filter is used to remove stray light before reaching the detector. Stray light may come from molecules which were not the target of the excitation but were excited due to some overlap of their absorption spectrum.

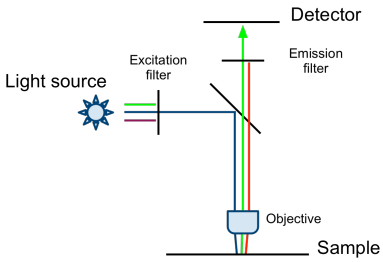

Figure 3 - The filter-based fluorescence microscope setup. The light coming from the light source is filtered to select the appropriate excitation band; once reflected onto the sample, it excites the fluorophores present in the sample. The fluorophores then emit fluorescence and pass through the dichroic mirror. Before hitting the detector the signal is cleaned eliminating all emission, which does not belong to the fluorophore under study.

The overall efficiency of the process is affected by components along the path of the light. Namely:

- **Light source**: lamps do not emit light with the same intensity at all wavelengths: although the total light power may seem sufficient, it may not be emitting strongly at the needed wavelength. Lasers emit coherent light at a few specific wavelengths.
- **Dichroic mirror**: the transmission and the reflection spectrum are not perfect, especially near the cut-off wavelength. Part of the light reflected by the sample (i.e. not fluorescent) may also make it through the mirror and cause (additional) background light.
- **Emission and excitation filters**: only recently filter technology has been able to produce near-step function transmission spectra, but until now around 10% of light in the nominal transmission band of the filters is lost.
- **Sample**: the main problem of the sample is autofluorescence, i.e. background fluorescence not generated by the stain but by the sample itself. Sources of autofluorescence include the mounting medium, fluorescence emitting tissue surrounding the structures under study, specific cell types containing fluorescent proteins like hemoglobin or impurities such as dust. Autofluorescence can mask the signal coming from the actual fluorophores. In addition to autofluorescence, the quantum efficiency of the fluorescent molecule is important, i.e. the average photons emitted per photons absorbed.
- **Detector**: each detector is able to translate photons into electrons with varying efficiency (i.e. quantum efficiency) which is also wavelength dependent. Sensitivity problems may arise in the ultraviolet or the infrared range.

2.1Filter-based systems

The most common optical setup for fluorescence microscopy involves the use of filter combinations designed to select the appropriate excitation and emission bands for the target fluorophore.

In widefield microscopes, the excitation filter, the dichroic and the emission filters are often arranged in a single component called a filter-cube. Each filter cube is tailored to a specific fluorescent molecule type (or in some cases to few fluorescent dyes) and a certain number of these filter cubes can be arranged in a rotatable turret, which can be rotated to position the desired filter-cube along the optical path.

In a confocal microscope, where the sample is usually excited by a laser, the optical setup differs slightly. In particular, filtering the excitation light is often performed by means of an acousto-optical tunable filter (AOTF). This type of filters contains a special crystal, which changes its refractive index when vibrating at particular frequencies. The crystal is tuned to different frequencies in order to deviate incoming light at different angles depending on the wavelength, effectively working as a wavelength selection filter. The dichroic mirror is specially designed to reject the specific wavelengths of the laser excitation. In these systems, the emission filters are usually decoupled from the dichroics: a single laser line, in fact, may excite multiple types of fluorescent molecules at the same time so, for the same laser/dichroic combination, the emission filters alone can be used sequentially to discriminate each single fluorochrome.

2.2 Spectral detectors

Spectral detectors are sensitive to light within a range of wavelengths with a certain wavelength resolution. One spectral technology involves the use of a liquid crystal tunable filter (LCTF) placed between the dichroic and the detector. These filters are narrow-band filters which can be tuned to select wavelength bands of the order of 5 to 10nm, within a range of a few hundred nanometers. Although initially they suffered from low light efficiency, recently the performance of these spectral detectors has improved considerably. The acquisition of the spectra is performed sequentially.

Another technique for spectral detection measures all wavelength bands in parallel by using a prism or a diffraction grating that disperses light rays at increasing angles according to its wavelength. Once dispersed, the light falls onto a

series of detectors (point or line detectors) that record light within a particular narrow wavelength band only. Commercial confocal systems normally have around 30 channels for a wavelength range of 400-700nm, thus using a nominal bandwidth of around 10nm per channel. The resolution may vary among systems.

2.3 Excitation scanning

The amount of light a fluorescent molecule emits also depends on the wavelength of the exciting light. By exciting a sample with light at different wavelengths (e.g. with different laser lines or excitation filters), the amount of emitted light can be modulated according to the absorption spectrum of the molecules. Therefore, to distinguish multiple fluorophores within a given sample, it can be helpful to excite it with light of different wavelengths and measure the resulting modulation of the emitted light.

Wavelength switching can be done very efficiently on LSM systems by means of AOTFs. These are fast enough to switch excitation lines between subsequent lines of the scan process, which decreases the additional time needed to switch between channels.

Excitation and emission scanning can be done at the same time, generating multiple excitation data points from which it is possible to robustly extract the different fluorescent components.

2.4 Space-domain Fourier-transform spectroscopy

This setup uses a Fourier-transform (FT) spectroscope. The FT spectroscope measures interference patterns of light at a particular wavelength for varying phase configurations. The interferometer generates a pattern as a function of phase: the interferogram. To calculate the actual spectrum, the interferogram is subject to a Fourier transform[4]. Commercial systems use a Sagnac-type interferometer, but setups with a Michelson interferometer are also possible. Both interferometers measure the phase difference between a reference and a probe ray of light, but they differ in the way they change the optical path length of the light. To be able to correctly Fourier-transform the interferogram, it is necessary to acquire the entire range of phases, so it is not possible to sample a subset of the spectrum as in other kinds of spectral analyzers.

3. Spectral unmixing

Regardless of the technology used to acquire the data, if two fluorophores have overlapping spectra, it is difficult -or even impossible- to acquire the signal from one fluorescent probe without also receiving interfering signal from the other. Spectral unmixing is the data processing technique used to recover the spatial distribution of each fluorophore from the mixed signals[5-7]. This section will review what software algorithms exist to perform spectral unmixing, starting with the way the emission of a mixture of fluorophores is modelled and then dealing with how to recover the mixing coefficients from the measured data.

3.1Model

The generally accepted model[5] for a mixture of fluorophore emissions is based on a linear combination of their abundance:

$$y_{i,j} = \sum_k a_{i,k} \cdot x_{k,j},$$

where $x_{i,j}$ indicates the abundance of fluorochrome k at position in the image j and $a_{i,k}$ represents the linear coefficient for fluorophore k in channel i. In this model, the linear coefficients depend on many different factors, including the spectra of all the devices along the optical path -emission spectrum of the lamp, spectrum of the emission and excitation filters, transmission filter of the dichroic, absorption and emission spectra of the fluorophore, etc- and the wavelength-dependent sensitivity of the detector. All of these different factors are summarized in the mixing coefficients.

The equation can be rewritten in matrix form as: Y=AX, where Y is a MxN, A is a MxK and X is a KxN matrix, being M the number of detected wavelength bands, N the number of detectors and K the number of fluorophores present in the mixture.

The uncertainty in the measurement is taken into account by including noise into the model. There are two noise models adopted in the literature. The first is the additive white Gaussian noise (AWGN) model[1], in which the equation is modified into Y = AX + R and R is a matrix formed by independently identically distributed (i.i.d.) zero-mean Gaussian variables. The second models the photon emission as a Poisson process, where $y_{i,j}$ becomes a Poisson-distributed stochastic variable with rate $\lambda = \sum_k a_{i,k} \cdot x_{k,j}$

[1] The underlying assumption is that each measurement is independent and Gaussian disributed. The gaussianity is due to the central limit theorem, which roughly states that whatever the distribution of a random variable (the causes of the error in this case), if we sum a great number of them the limit distribution of the sum is Gaussian.

3.2 Algorithms

The algorithms used to solve this problem can be divided into two groups. The first group assumes that the spectra of the components are known. The spectrum can be measured by recording the emission of the fluorophore in single-stained samples for each required excitation and emission pair [5]. This translates into a known A matrix and the problem is reduced to the solution of a system of equations. As the system of equations is often overdetermined (if each column of Y and X is treated as an independent system of equations, and the number of recorded channels is at least equal to the number of fluorophores to be found, then the number of equations is equal to or greater than the number of unknowns), it is usually solved in the least squares sense, with the additional constraint of non-negativity (each pixel is proportional to the number of recorded photons within the exposure time, so it cannot be negative). The second family of algorithms includes both the A and the X matrices as unknown (the number of unknowns is greater than the number of equations).

Recovering both matrices A and X simultaneously is usually approached by minimizing some energy functional or cost function, which is chosen according to the assumptions made on the model. For example, under maximum likelihood statistics, the log-likelihood estimator for the AWGN model translates to the Frobenius norm of the residual (Y-AX). On the other hand, if we assume the samples to follow a Poisson distribution, the log-likelihood function is analogous to the Kullback-Leibler divergence between Y and AX[8-11]. Solving the non-negative matrix factorization is then translated to the problem of optimizing the cost function[12].

3.2.1Non-negative Least Squares

If no constraint of non-negativity were needed, then a simple algorithm based on the QR decomposition would provide the optimal answer in the least-squares sense. On the other hand, if the solution has to be in the positive orthant (i.e. non-negative), there is no direct solution. The algorithm for the solution of the Y=AX equation with A known and constrained to positive solutions was first published by Lawson and Hanson in 1974[13]. The same algorithm is also present in MATLAB under the function *lsqnonneg*.

3.2.2 Blind algorithms

To simultaneously recover both the A and the X matrices we use what are called blind algorithms. They have received increasing attention due to their ability to extract patterns from the data itself without having to separately measure the spectrum of the various fluorophores. Here, a few approaches will be reviewed, but most of the discussion will be on non-negative matrix factorization, as it is the most promising method.

3.2.2.1Principal Component Analysis (PCA)

The overdetermined system of equations could in fact be approached using Principal Component Analysis (PCA) [14]. Using the singular value decomposition (SVD), we take the K strongest eigenvalues of the linear system and the corresponding eigenvectors and use these as a new base to represent the data. This new base is optimal in the least squares sense and should approximate the spectra of the various fluorophores. But as there is no constraint on the non-negativity of the abundances or of the spectrum, the results may be optimal in the mathematical sense but lack in physical significance.

3.2.2.2 Blind Signal Separation (BSS) and Independent Component Analysis (ICA)

The Blind Signal Separation (BSS) model[15] is similar to the mixture model: given the vector source s(t), the output is given by the input multiplied by a mixing matrix A: x(t)=A s(t). The goal of the signal separation is to find a separating matrix B such that y(t) = B x(t) is an estimate of the sources s(t). The theoretical framework presented by Cardoso [15] studies different methods by analyzing the corresponding estimation of the error, which he calls cost functions. The two main families of methods he studies are the maximum likelihood method and the mutual information method, adopting one or the other depending on whether or not an assumption is made on the distribution of the sources. If we assume some distribution on s(t), then the appropriate approach is the Maximum Likelihood estimation, which minimizes the Kullback-Leibler (K-L) divergence between the estimated sources y(t) and the assumed distribution of the sources s(t). If no previous knowledge is available or assumed, then we are left with the statistical independence of the sources, which is represented by a cost function based on the Mutual Information (MI). The author proposes to solve the BSS problem by gradient descent on the appropriate cost function and derives update rules both for an online (sample-by-sample) and an offline (all samples together) method. Although the derivative of the MI can be difficult to compute, it can be approximated using fourth order cumulants. As he points out, it is interesting to note that the order of the moments is four and not two, which indicates that decorrelation (whitening), which would refer to second-order moments is not a sufficient assumption to separate the sources. Statistical independence and non-correlation are in fact two different constraints and each leads to a different solution.

Both methods, as described in Cardoso's work, are to be used when the number of sources equals the number of measured signals. Extensions to overdetermined cases have also been proposed in later works.

An important common assumption of these methods is that at each point in time t, the variable at s(t) is statistically independent from s(t+dt) for any non-zero dt. This implies that whenever some dependence exists within the same source, in principle the method cannot be applied. Furthermore, the BSS model works well only if the signals are non-Gaussian: if the signals have Gaussian shape, in fact, the problem becomes unsolvable. Independent Component Analysis (ICA) (a particular kind of BSS) stresses this point by solving the BSS problem by maximizing the non-gaussian nature of the sources.

Another limitation of the method is that although not impossible, it does not explicitly take into account the positiveness of the sources.

3.2.2.3 Positive Matrix Factorization (PMF)

Positive Matrix Factorization (PMF) was proposed initially by Paatero [16, 17] for the separation of chemical mixtures, and adopts a least squares formulation. The authors argue that PCA assumes incorrect (unrealistic) weighting for the matrices, so his proposal is a cost function which takes into account different standard deviations for each observed value. They derive an alternating least squares algorithm that in turn updates the mixing matrix and the components matrix by least squares. The weighted least squares cost function is expanded to take into account four terms corresponding to the non-negativity of each of the two matrices and two penalization terms aimed at avoiding rotations and scale ambiguities. The weights of these two terms are adjusted during the algorithm in order to lower their importance near convergence, where the cost function is less likely to have local minima. In order to speed up the convergence of the algorithm, the authors also propose a second-order Newton-like method.

3.2.2.4 Non-negative Matrix Factorization (NMF)

Most of the literature on Non-negative Matrix Factorization (NMF) refers to the seminal work of Lee and Seung [8]. Although the authors had already published some previous work on convex and conic optimization [12], their non-negative matrix factorization method really started increasing adoption when it was published in a letter to Nature [8]. The target application of the work was the representation of a database of faces in terms of linear combination of (positive) features. The basic idea is to formulate the solution of the equation as a matrix factorization problem, with the addition of multiplicative update rules. As the name indicates, the correction is performed at each step by multiplying each component by a positive number. They are introduced to ensure that if the initial estimation is positive, then the estimation at each following step remains positive.

The main problem of this approach is the unresolved ambiguity of the solution. It is possible to both rotate and scale the solution without seeing any change in the cost function. Furthermore, even when the rotation and scaling ambiguities are taken into account, the cost function still presents multiple local minima, which render the solution dependent on the initialization[18].

Nevertheless, the method was seminal in its simplicity and effectiveness.

3.2.2.5 Regularized Non-negative Matrix Factorization

There have been multiple approaches to the problem of the sensitiveness to initialization, mainly involving a regularization of the cost function and a consequent reduction in the amount of local minima.

Feng [19] proposed to favour sparsity of the mixing matrix and expressiveness and orthogonality of the basis components. This was enforced through two additional weighted terms in the cost function: adding the Frobenius norm on the basis matrix and subtracting the Frobenius norm on the unmixed components. The weights were algebraically eliminated while deriving the update rules.

Li [20] evolved the algorithm proposed by Feng and chose the Kullback-Leibler divergence as the reconstruction error measure and the same sparseness constraint. The update rules of their algorithm are multiplicative.

Hoyer introduced the concept of non-negative sparse coding as a variation of linear sparse coding that enforces non-negativity. In practice, the cost function is composed of a Frobenius norm on the residual error and the L_1 norm as the regularization. In a later work [21], he further developed the algorithm to use the ratio of the L_1 to the L_2 norm as a measure of sparseness. He argues that the sparseness constraint may be applied to either the base components or the linear mixing coefficients, according to the requirements of the application. The choice is made at the algorithmic level: at each iteration, a projective gradient descent is applied to the matrix where the sparseness is required, while the multiplicative rules are used to update the other matrix. The projected gradient descent works by first descending the gradient of the reconstruction error and then performing an iterative projection until the sparseness constraints are met (these are that L_2 remain unchanged and L_1 is reduced, see original article for details [21]).

A first attempt to generalize the regularized NMF methods and include them all under a single family was done by Dhillon and Sra [22], who derived multiplicative update rules for the class of Bregman divergences with a variety of regularization functions. The Bregman divergence is non-symmetric and is defined along with a convex function φ.

Special cases of the choice of the convex function reduce the corresponding Bregman divergence to the Frobenius norm ($\varphi=1/2\ x\hat{}2$) or the K-L divergence ($\varphi=x \log x - x$).

A. Cichocki derived a different family of algorithms by deriving multiplicative update rules for the Amari α-divergence [23]. This particular discrepancy measure includes as special cases the Kullback-Leibler divergence ($\alpha=0$ and $\alpha=1$), the Hellinger distance ($\alpha=1/2$) and the χ^2-distance ($\alpha=2$). It does not include the Frobenius norm. In his work, he demonstrated the convergence of the algorithm in a similar fashion as Lee and Seung demonstrated the convergence of theirs, by using an auxiliary function.

3.2.3 Parallel factor analysis and Non-negative Tensor Factorization

If the employed imaging setup allows the use of multiple excitations for each detector channel (see section 3.2 on excitation-side scan), the mathematical model can be changed slightly by adding a modulation coefficient representing the dependence on the absorption spectrum of the fluorophore. The advantage of this approach is the additional information resolves some of the ambiguities of NMF, so that it is possible to design the algorithms that are able to solve this problem in a more robust manner. There exist many possible ways of solving the augmented problem, but they can all be grouped under the name of non-negative tensor factorization[24-26].

3.3 Application

To get the best results, the most appropriate method must be used according to the imaging technique: the relevance of each unmixing algorithm family for each image acquisition has been summarized in Table 1.

Table 1 - The table summarizes what algorithms can be applied to what imaging technique. For each single/multiple excitation/emission combination, both the narrowband (e.g. LCTF) and the wideband variants are given, where applicable.

Unmixing algorithm / Imaging technique		NNLS	Blind two-matrix decomposition (ICA, PMF, PCA, NMF)	Blind three matrix decomposition (NTF)
Filter – cube		Yes	Yes	No (no advantage)
Single excitation, multiple emission	Wideband em.	Yes	Yes	No (no advantage)
	Narrowband em.	Yes	Yes[2]	No (no advantage)
Multiple excitation, single emission	Wideband em.	Yes (absorption side spectra needed)	Yes	No (no advantage)
	Narrowband em.	Yes (absorption side spectra needed)	Yes[2]	No (no advantage)
Multiple excitation, multiple emission	Wideband em.	Yes (both absorption and emission spectra needed)	No	Yes
	Narrowband em.	Yes (both absorption and emission spectra needed)	No	Yes
FT spectroscopy (single excitation)		Yes (absorption spectra needed)	Yes	No (no advantage)

The application of the algorithms requires a series of steps:

1. Each multi-channel pixel in the image represents a column of the input matrix Y, where each element of the column vector is the intensity of that pixel at each channel (see model).
2. For the NNLS method, we need the spectra of each fluorophore. Each spectrum is arranged in a column of the spectra matrix A and each element of the column vector is the relative intensity of the fluorophore for that particular channel (spectrum of the fluorophore). Each column vector is normalized to have sum 1[3].
3. For NMF, an initial estimation has to be given for both the spectra and the abundances. The better the initial estimation, the better will the final result, especially if no regularization is used. Random values can be used to initialize the abundance matrix X, but the initialization for the spectrum matrix A should be chosen with more

[2] Although from a technical point of view the NMF algorithm is applied to narrowband or wideband data in the same way, the results from a narrowand system allow a greater redundancy and a better robustness of the method.

[3] Note that this normalization assumes that the channel bandpasses are not overlapping. In case the constraint is violated, i.e. the channels do indeed overlap, the abundances loose their quantitative meaning.

care. An example is to initialize each column with a Gaussian function centered at the channel containing the expected peak emission signal for that fluorophore [24].

4. In the case of tensor factorization, we have three matrices to initialize. As the tensor factorization is more robust, the matrices need not to be initialized with a great degree of precision. A random matrix is often sufficient.

An example of application of spectral unmixing is shown in Figure 4. The sample contains NIH-3T3 fibroblasts cells, where nucleid acids were stained with EtBr, filamentous actin with Alexa Fluor 532 and tubulin with Alexa Fluor 488. The images were acquired using an LSM510 Meta confocal microscope by Zeiss, exciting with a 488nm laser line and captured at three wavelengths (508, 550, 657) with a bandwidth of 21.4. As can be seen from the second image in the first row, the overlap between the spectra of AF532 and AF488 makes it difficult to separate the phalloidin and the tubulin, or to estimate the concentration of the two proteins.

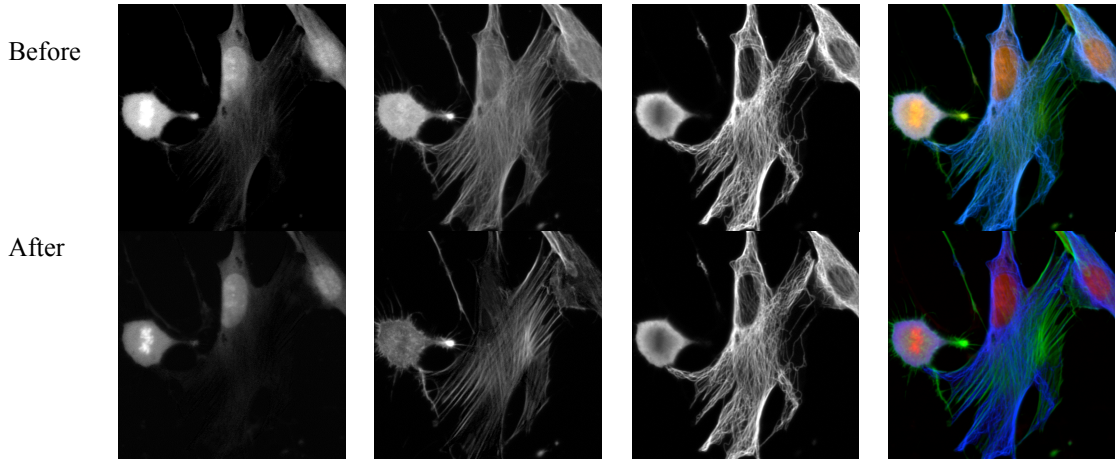

Figure 4 – A sample image series from a confocal section of NIH-3T3 cells stained with EtBr for double stranded nucleic acids, Alexa Fluor 532 for phalloidin and Alexa Fluor 488 for tubulin. The images were acquired with a confocal microscope at three emission windows (508, 527 and 657) of 21.4 nm each. The first row shows the three channels (508,527 and 657 respectively) along with a color composition following the mentioned channels in red green and blue respectively, while the second row shows the estimated concentrations of the three dyes (EtBr, AF532, AF488) along with a color composition.

The images were unmixed using the freely available plugin PoissonNMF for ImageJ [27], which implements the NMF algorithm presented in [24]. The NMF algorithm is based on the K-L divergence and regularized with a sparseness constraint on the unmixed components (named “segregation bias”) and a weighted orthogonality constraint on the mixing matrix (see the reference for additional details). The results of the unmixing are shown in the second row of Figure 4. After the unmixing procedure, the overlap of the spectra of the AF532 and the AF488 has been corrected and the images now represent the abundance of each of the fluorophores.

Conclusions

Summarizing the material presented in this chapter, spectral unmixing in microscopy is a powerful technique to identify fluorophore abundances in multiply-stained samples. In order to properly master the technique, however, it is useful to have a basic understanding of the underlying principles. Furthermore, to take full advantage of the sophisticated hardware used to acquire the data, proper analysis techniques need to be employed. This chapter has reviewed the spectral imaging techniques most widely adopted in microscopy, both from a hardware and from a software point of view. Although sufficient information is given for an intuition of the technologies, we refer the reader to the literature for a better coverage of each of the concepts or algorithms.

Acknowledgements. This project was partially funded by the "UTE Project CIMA", the subprogram for unique strategic projects of the Spanish Science and Innovation Ministry (MCIIN PSS-010000-2008-2), and projects MICINN TEC2005-04732 and MICINN DPI2009-14115-C03-03. TP holds a predoctoral fellowship of the Spanish Ministry for Science Innovation. COS and AMB both hold a Ramón y Cajal fellowship granted by the Spanish Ministry of Science and Innovation (MICINN). COS is also funded by a Marie Curie International Reintegration Grant (MIRG-CT-2005-028342)..

References

[1] H. Morise, O. Shimomura, F. H. Johnson and J. Winant. Intermolecular energy transfer in the bioluminescent system of aequorea. Biochemistry.1974. 13;12; 2656-2662.

[2] O. Shimomura, F. H. Johnson and Y. Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. Journal of Cellular and Comparative Physiology.1962. 59;3; 223-239.

[3] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward and D. C. Prasher. Green fluorescent protein as a marker for gene expression. Science.1994. 263;5148; 802-805.

[4] H. Tsurui, H. Nishimura, S. Hattori, S. Hirose, K. Okumura and T. Shirai. Seven-color fluorescence imaging of tissue samples based on fourier spectroscopy and singular value decomposition. J Histochem Cytochem.2000. 48;5; 653-62.

[5] T. Zimmermann. Spectral imaging and linear unmixing in light microscopy. Adv Biochem Eng Biotechnol.2005. 95;245-65.

[6] T. Zimmermann, J. Rietdorf and R. Pepperkok. Spectral imaging and its applications in live cell microscopy. FEBS Lett.2003. 546;1; 87-92.

[7] T. Zimmermann, J. Rietdorf, A. Girod, G. V and R. Pepperkok. Spectral imaging and linear un-mixing enables improved FRET efficiency with a novel GFP2-YFP FRET pair. FEBS Lett.2002. 531;2; 245-9.

[8] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. Nature.1999. 401;6755; 788-91.

[9] D. D. Lee and H. S. Seung, Algorithms for non-negative matrix factorization. Proceedings of the Neural Information Processing Systems: Natural and Synthetic (NIPS'01).Vancouver (BC), Canada. December 3-8, 2001. 13;556-562.

[10] R. Kompass. A generalized divergence measure for nonnegative matrix factorization. Neural Computation.2007. 19;3; 780-791.

[11] A. Cichocki, H. Lee, Y. Kim and S. Choi. Non-negative matrix factorization with α-divergence. Pattern Recognition Letters.2008. 29;9; 1433-1440.

[12] D. D. Lee and H. S. Seung. Unsupervised learning by convex and conic coding. Advances in Neural Information Processing Systems 9.1997. 9;515--521.

[13] C. L. Lawson and R. J. Hanson, Eds., Solving Least Squares Problems. USA:Society for Industrial and Applied Mathematics; 1995.

[14] K. Fukunaga, Ed., Introduction to Statistical Pattern Recognition. Academic Press; 1990.

[15] J. Cardoso. Blind signal separation: Statistical principles. Proceedings of the IEEE.1998. 86;10; .

[16] P. Paatero. Least squares formulation of robust non-negative factor analysis. Chemometrics and Intelligent Laboratory Systems.1997. 37;1; 23-35.

[17] M. Juvela, K. Lehtinen and P. Paatero. The use of positive matrix factorization in the analysis of molecular line spectra. 1996. 280;616-626.

[18] F. J. Theis, K. Stadlhammer and T. Tanaka, First results on uniqueness of sparse non-negative matrix factorization. Florence, Italy, September 4-8, 2006.

[19] T. Feng, S. Z. Li, H. Shum and H. Zhang, Local non-negative matrix factorization as a visual representation. Proceedings of the 2nd International Conference on Development and Learning (ICDL'02).2002. Cambridge, MA, USA; June 12-15 2002; 178-183.

[20] S. Z. Li, Xin Wen Hou, Hong Jiang Zhang and Qian Sheng Cheng, Learning spatially localized, parts-based representation. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001).2001. Kauai, Hawaii, USA. December 8-14 2001. 1;207-212.

[21] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. 2004. 5;1457-1469.

[22] I. S. Dhillon and S. Sra, Generalized nonnegative matrix approximations with bregman divergences. 2005. Vancouver, Canada; 283-290.

[23] A. Cichocki, R. Zdunek and S. -. Amari. Nonnegative matrix and tensor factorization [lecture notes]. IEEE Signal Processing Magazine.2008. 25; 25;1; 142-145.

[24] R. A. Neher, M. Mitkovski, F. Kirchhoff, E. Neher, F. J. Theis and A. Zeug. Blind source separation techniques for the decomposition of multiply labeled fluorescence images. Biophysical Journal.2009. 96;9; 3791-3800.

[25] R. Bro. PARAFAC. tutorial and applications. Chemometrics and Intelligent Laboratory Systems.1997. 38;2; 149-171.

[26] J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra and its Applications.1977. 18;2; 95-138.

[27] PoissonNMF ImageJ plugin. http://www.mh-hannover.de/cellneurophys/poissonNMF/; 2010;June 20; .