Unsupervised decomposition of low-intensity low-dimensional multi-spectral fluorescent images for tumour demarcation

I. Kopriva,* A. Peršin

Department of Laser and Atomic Research and Development, Rudjer Bošković Institute, Bijenička cesta 54, P.O. Box 180, 10002 Zagreb, Croatia.

ikopriva@irb.hr, persin@irb.hr

Abstract

Unsupervised decomposition of static linear mixture model (SLMM) with ill-conditioned basis matrix and statistically dependent sources is considered. Such situation arises when low-dimensional low-intensity multi-spectral image of the tumour in the early stage of development is represented by the SLMM, wherein tumour is spectrally similar to the surrounding tissue. The original contribution of this paper is in proposing an algorithm for unsupervised decomposition of low-dimensional multi-spectral image for high-contrast tumour visualisation. It combines nonlinear band generation (NBG) and dependent component analysis (DCA) that itself combines linear pre-processing transform and independent component analysis (ICA). NBG is necessary to improve conditioning of the extended mixing matrix in the SLMM, while DCA is necessary to increase statistical independence between spectrally similar sources. We demonstrate good performance of the method on both computational model and experimental low-intensity red-green-blue fluorescent image of the surface tumour (basal cell

carcinoma). We believe that presented method can be of use in other multi-channel medical imaging systems.

Keywords: Unsupervised decomposition; Ill-conditioned static linear mixture models; Multi-spectral imaging; Tumour demarcation; Multi-channel medical imaging.

1.0 Introduction

Methods for tumour demarcation include examples with visualisation based on: fluorophores (Koenig et al., 2001; Bäumler et al., 2003), green fluorescent protein gene tumour transduction system (Hasegawa et al., 2000; Hoffman, 2002) or fluorescent nanoparticles (Tréhin et al., 2006). In the case of fluorophores, contrast visualisation is obtained under optimal combination of parameters such as concentration of photo-synthesiser, duration of treatment with photo-synthesiser and intensity of illuminating light. Under these conditions a few well-established methods can be used for tumour visualisation (Scott et al., 2000; Ericson et al., 2003; Fischer et al., 2001). However, variability in some of the parameters will cause fluctuation of the intensity level of the acquired fluorescent image, causing predefined threshold constants not being optimal any more. Here we propose a high-contrast tumour visualisation algorithm that is based on unsupervised decomposition of a low-dimensional multi-spectral fluorescent image. It exhibits a high level of robustness with respect to the fluctuation of the intensity level. That is achieved due to the unsupervised nature of the algorithm and the scale invariance property of the independent component analysis algorithm (ICA) (Jutten and Herault, 1991; Bell and Sejnowski, 1995; Hyvärinen et al., 2001; Cichocki and Amari, 2002) that forms the heart of the proposed method. However, spectral similarity between the tumour and surrounding tissue causes an ill-conditioning of the basis matrix as well as statistical dependence between the sources in the hypothetical linear mixture model of the

fluorescent image. Therefore, ICA fails to yield accurate demarcation. Here we show that by nonlinear band generation (NBG), (Ren and Chang, 20001; Du et al., 2004), and dependent component analysis (DCA), (Cichocki and Amari, 2002; Cichocki and Georgiev, 2003; Cichocki, 2007; Hyvärinen, 1998; Kopriva and Seršić, 2008), we can fix ill-conditioning problem and increase statistical independence between the sources, hence creating an environment for the ICA to work more accurately. This newly proposed method exhibits improved performance over number of state-of-the-art multi-channel blind decomposition algorithms on both computational and experimental multi-spectral images.

2.0 Methods

The unsupervised decomposition problem consists in finding the basis matrix $\mathbf{A} \in \mathrm{R}_{+}^{N \times M}$ and matrix of hidden components or sources $\mathbf{S} \in \mathrm{R}_{+}^{M \times T}$ given only the matrix of observed data $\mathbf{X} \in \mathrm{R}_{+}^{N \times T}$ such that the following static linear mixture model (SLMM) holds

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

Each row of **X** and **S** is a signal or 1D image representation, N is the number of observed signals, M is the number of hidden components (sources) and T is the number of samples. Because we are concerned with an unsupervised image decomposition problem, we have assumed **X**, **A** and **S** to be nonnegative.

2.1 Independent component analysis

The unsupervised decomposition problem (1), also known as a static blind source separation problem, is efficiently solved by ICA (Hyvärinen et al., 2001; Cichocki and Amari, 2002), provided that the sources are statistically independent, at most one source has Gaussian distribution, and the number of sources M is less than or equal to the number of observed signals N. Then a solution to the unsupervised decomposition problem (1) is obtained with scale and permutation indeterminacy, i.e.,

$$\hat{\mathbf{S}} = \mathbf{WX} \tag{2}$$

with $\mathbf{WA}=\mathbf{P\Lambda}$, where $\mathbf{W}$ represents the un-mixing matrix, $\mathbf{P}$ is a general permutation matrix and $\mathbf{\Lambda}$ is a diagonal matrix. This implies that ICA-based solution of the unsupervised decomposition problem is unique up to the ordering, scale and sign. Thus, ICA algorithms possess a scale invariance property that makes them attractive for use in tumour visualisation and demarcation from the multi-spectral fluorescent image, when the level of the fluorescence may vary from measurement to measurement for the reasons already discussed.

It is assumed by many ICA algorithms that sources are zero mean as well as that data are whitened (spatially uncorrelated). Most algorithms include mean removal and data whitening as a first phase in the algorithm development. Mean removal is achieved very simple through

$$\left\{\mathbf{x}_n \leftarrow \mathbf{x}_n - E\left[\mathbf{x}_n\right]\right\}_{n=1}^{N} \tag{3}$$

where $\mathbf{x}_n$ denotes rows of $\mathbf{X}$ and $E[\mathbf{x}_n]$ denotes mathematical expectation of $\mathbf{x}_n$. Data whitening is achieved through

$$\mathbf{X} \leftarrow \mathbf{\Lambda}^{-1/2}\mathbf{E}^{\mathrm{T}}\mathbf{X} \tag{4}$$

where $\mathbf{\Lambda}$ and $\mathbf{E}$ respectively stand for diagonal matrix of eigenvalues and matrix of eigenvectors of the sample data covariance matrix $\mathbf{R}_{\mathbf{XX}}=E[\mathbf{X}\mathbf{X}^{\mathrm{T}}]$. Since in the muti-spectral image decomposition problem source signals represent spatial maps of the materials resident in the image, we have always rescaled extracted sources to the [0,1] interval such that probability of the source being present (that is maximally 1 and minimally 0) can be assigned at each pixel level.

The strategy of the ICA algorithms in solving blind decomposition problem is to find linear transform $\mathbf{W}$ such that components of $\hat{\mathbf{S}}$ are as much statistically independent as possible. Depending on available type of *a priori* information about the sources various approaches are exploited by ICA algorithms. Number of ICA methods requires *a priori* information about the class of distributions to which source signals belong (Bell and Sejnowski, 1995; Pham, 1997; Choi et al., 2000). Such information is however not readily available. One alternation is to derive methods that are adaptive to the unknown source distributions. The representatives of such class of methods are kernel density methods (Xue et al., 2008; Principe and Xu, 1999; Principe et al., 2000). Main disadvantage of the kernel density based ICA methods is their computational complexity that is $O(T^2N^2)$ where T represents number of samples. In application domain considered in this paper T represents number of pixels and can take value of even 10^7. Hence, the computational complexity in such a case

would be huge. It has been pointed out recently (Xue et al., 2008) that computational complexity of the kernel density methods can be reduced by fast Fourier transform (Sliverman, 1982). Other approaches to reduce computational complexity of kernel density-based ICA methods were also proposed (Pham, 2003; Schwartz et al., 2004) yielding respectively computational complexity of $O(3^N T + N^2 T)$ and $O(NT\log T + N^2 T)$. If we consider multi-spectral image to be an RGB image, then N=3. Assuming $T \approx 4\times 10^6$ (that is today normal size of the RGB images from commercial digital cameras) we arrive at still very large numbers for computational complexity of the kernel density-based ICA methods. Due to computational complexity reasons we have relied in this manuscript on ICA methods that are also source distribution independent but solve blind decomposition problem through minimization of the statistical dependence between the sources up to the fourth order (FO) (Cardoso and Soulomniac, 1993) or through minimization of the statistical dependence between the sources of the second order but for different lags (Belouchrani et al., 1997; Stone, 2001).

Representative of the first group is FO cumulant based ICA algorithm JADE (Cardoso and Soulomniac, 1993) that stands for Joint Approximate Diagonalisation of the Eigen-matrices, where statistical independence is achieved through minimisation of the squares of the FO cross-cumulants between the components $\hat{\mathbf{s}}_m$

$$\mathbf{W} = \arg\min \sum_{i,j,k,l} \mathbf{off}\left(\mathbf{W}^{\mathrm{T}} \hat{C}_4(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_j, \hat{\mathbf{s}}_k, \hat{\mathbf{s}}_l)\mathbf{W}\right) \tag{5}$$

where **off**(**A**) is measure for the off-diagonality of a matrix defined as $\mathbf{off}(\boldsymbol{A}) = \sum_{1 \le i \ne j \le N} \left| a_{ij} \right|^2$. $\hat{C}_4(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_j, \hat{\mathbf{s}}_k, \hat{\mathbf{s}}_l)$ are sample estimates of the related FO cross-cumulants (Mendel, 1991; McCullagh, 1995) i.e.

$$\begin{aligned} &\hat{C}_4(\hat{\mathbf{s}}_i, \hat{\mathbf{s}}_j, \hat{\mathbf{s}}_k, \hat{\mathbf{s}}_l) = \\ &E\left[\hat{\mathbf{s}}_i \hat{\mathbf{s}}_j \hat{\mathbf{s}}_k \hat{\mathbf{s}}_l\right] - E\left[\hat{\mathbf{s}}_i \hat{\mathbf{s}}_j\right] E\left[\hat{\mathbf{s}}_k \hat{\mathbf{s}}_l\right] - E\left[\hat{\mathbf{s}}_i \hat{\mathbf{s}}_k\right] E\left[\hat{\mathbf{s}}_j \hat{\mathbf{s}}_l\right] - E\left[\hat{\mathbf{s}}_i \hat{\mathbf{s}}_l\right] E\left[\hat{\mathbf{s}}_j \hat{\mathbf{s}}_k\right] \end{aligned} \tag{6}$$

The additional advantage of using FO cumulants based ICA algorithm is its capability to suppress additive Gaussian noise based on the known property that FO cumulants are blind with respect to Gaussian noise (Mendel, 1991). Disadvantage of the use of FO statistics based methods is their sensitivity to outliers as well as requirement that FO cumulants for the source signals in consideration must exist. On the other side SO statistics based methods are more robust with respect to outliers. If the source signals have certain structure (temporal or spatial) it is possible to obtain demixing matrix **W** as the solution of the following joint diagonalization problem

$$J(\mathbf{W}) = \sum_{\tau \in S} \mathrm{off}\left(\mathbf{W} \bar{\mathbf{C}}_\tau^{\mathbf{s}} \mathbf{W}^{\mathrm{T}}\right) \tag{7}$$

where $\bar{\mathbf{C}}_\tau^{\mathbf{s}}$ are symmetrical one-lag covariance matrices of the sources

$$\bar{\mathbf{C}}_\tau^{\mathbf{s}} = E\left\{\mathbf{S}(t)\mathbf{S}(t-\tau)^{\mathrm{T}}\right\} + E\left\{\mathbf{S}(t-\tau)\mathbf{S}(t)^{\mathrm{T}}\right\}$$

Since by assumption source signals are statistically independent it applies that $\bar{\mathbf{C}}_\tau^{\mathbf{s}}$ must be a diagonal matrix. Thus, it follows from (7) that **W** is matrix of eigen-vectors

that jointly diagonilize set of matrices $\left\{\overline{\mathbf{C}}_{\tau}^{\mathbf{s}}\right\}_{\tau\in S}$. This is how the SOBI algorithm that stands for Second Order Blind Identification is formulated (Belouchrani et al., 1997).

For predictable signals it is further possible to obtain demixing matrix **W** as the solution of the generalized eigen-decomposition problem by maximizing predictability measure (Stone, 2001)

$$F\left(\left\{s_m(t_k)\right\}\right)=\log\frac{V\left(\left\{s_m(t_k)\right\}\right)}{U\left(\left\{s_m(t_k)\right\}\right)}=\log\frac{\sum_k^{k_{\max}}\left(\overline{s}_m(t_k)-s_m(t_k)\right)^2}{\sum_k^{k_{\max}}\left(\tilde{s}_m(t_k)-s_m(t_k)\right)^2} \tag{8}$$

where V reflects the extent to which $s_m(t_k)$ is predicted by a long term moving average $\overline{s}_m(t_k)$ and U reflects the extent to which $s_m(t_k)$ is predicted by a short term moving average $\tilde{s}_m(t_k)$. Since source signals in the problem considered in this paper are predictable (pixel values are locally correlated) we shall use this method, coined as "temporal" predictability maximization ICA algorithm, in blind multi-spectral image decomposition problem.

2.2 Dependent component analysis

The basic idea behind DCA (Cichocki and Amari, 2002; Cichocki and Georgiev, 2003; Hyvärinen, 1998; Kopriva and Seršić, 2008) is to find a linear transform H that can improve statistical independence between the sources but leave the basis matrix unchanged, i.e.,

$$H\left(\mathbf{X}\right)=H\left(\mathbf{AS}\right)\cong\mathbf{A}H\left(\mathbf{S}\right) \tag{9}$$

Since the sources in this new representation space will be less statistically dependent, any standard ICA algorithm can in principle be used to learn the mixing matrix **A** or demixing matrix **W**. Once they are estimated, the sources **S** can be recovered by applying **W** the multi-spectral image **X** in (1). Examples of linear transforms that have such a required invariance property and generate less dependent sources in the new representation space include: high-pass filtering (Cichocki and Georgiev, 2003), innovations (Hyvärinen, 1998) and wavelet transforms (Kopriva and Seršić, 2008). However, to get optimal performance out of DCA algorithm, chosen ICA algorithm should be *tuned* to the chosen statistical independence enhancement transform. One computationally efficient approach to solve the blind separation problem with statistically dependent sources is based on the use of innovations. The arguments for using innovations are that they are more independent from each other and more non-Gaussian than original processes. The innovation process is referred to as prediction error that is defined as:

$$\tilde{s}_m(t) = s_m(t) - \sum_{i=1}^{l} b_{mi} s_m(t-i), \quad m = 1, \cdots, M \tag{10}$$

where $s_m(t-i)$ is the i-th sample of a source process $s_m(t)$ at location $(t-i)$ and $\boldsymbol{b}_m$ is a vector of prediction coefficients. $\tilde{s}_m(t)$ represents the new information that $s_m(t)$ has but is not contained in the past l samples. It has been proved in (Hyvärinen, 1998) that if **X** and **S** follow the linear mixture model (1), their innovation processes $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{S}}$ (in matrix form) follow the same model as well, i.e.,

$$\tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}} \tag{11}$$

Because innovation basically removes predictable, or slow varying, part of the signal it is super-Gaussian and close to independent and identically distributed (i.i.d.) signal. This is especially true if order of the prediction error filter in (10) is high. Thus, FO cumulants are expected to be well defined for when high-order innovations are used. Therefore, it is justified in DCA algorithm to combine JADE ICA algorithm with high-order innovations based linear transform in order to achieve good performance. First order high pass filter can be thought as the first order innovation of the signal. In such a case significant part of the predictable signal component will still remain in its innovation. Therefore, it is justified in DCA algorithm to combine SOBI-like ICA algorithms (Molgedy and Schuster, 1994; Ziehe et al., 1998) with low-order innovations based linear transform to achieve good performance. Due to the reasons discussed, we shall apply these two DCA algorithms in the comparative performance analysis presented in section 3. At this place we would also like to explain why, by means of innovations or high-pass filtering, removing statistical dependence between the source pixels also increases statistical independence between the sources. Both high-pass filtering and innovations, they can be viewed as source adaptive high-pass filtering, remove low-frequency part of the source spectrum. It is low-frequency part of the source spectrum that is a cause for eventual statistical dependence among the sources. This empirical observation has been brought out in (Cichocki and Amari, 2002; Cichocki and Georgiev, 2003; Cichocki, 2007). It represents basis for construction of various DCA algorithms and explains why all of them are looking for high-frequency part of the source spectrum in order to learn a mixing matrix more

accurately by applying ICA algorithms on high-pass filtered version of the mixtures, see (9) and (11). Hence by removing low-frequency part of the source spectrum innovation process also removes eventual cause of statistical dependence among the sources.

2.3 Nonnegative matrix factorization

Alternatives to the ICA or ICA-based DCA include algorithms for nonnegative matrix factorisation (NMF), (Lee and Seung, 1999; Zdunek and Cichocki, 2007; Cichocki et al., 2008), which are also applicable to the SLMM (1) because the variables in the model are nonnegative. Unlike ICA, the NMF algorithms do not impose statistical independence and non-Gaussianity requirements on the sources. However they do generally require $N>>M$, which makes them not good candidate for unsupervised decomposition problems when $\mathbf{X}$ represents a low-dimensional multi-spectral image, such as a red-green-blue (RGB) image in which case N=3 can even be less than M. In addition to that, NMF algorithms are quite sensitive to the choice of initial conditions that often yields suboptimal performance. Various strategies are under investigation, including even ICA algorithms, to estimate good initial points for $\mathbf{A}$ and $\mathbf{S}$. The basic approach to NMF, that is described in general algorithmic form below, is alternating minimization of a chosen cost function (Zdunek and Cichocki, 2007; Cichocki et al., 2008).

Set Randomly initialize: $\mathbf{A}^{(0)}$, $\mathbf{S}^{(0)}$,

For k=1,2,…, until convergence **do**

$$\text{Step 1: } \mathbf{S}^{(k+1)} = \arg\min_{s_{mt}\geq 0} D\left(\mathbf{X}\middle\|\mathbf{A}^{(k)}\mathbf{S}\right)_{\mathbf{S}^{(k)}}$$

$$\text{Step 2: } \mathbf{A}^{(k+1)} = \arg\min_{a_{nm}\geq 0} \tilde{D}\left(\mathbf{X}\middle\|\mathbf{A}\mathbf{S}^{(k+1)}\right)_{\mathbf{A}^{(k)}}$$

End

In general, the cost function $D\left(\mathbf{X}\middle\|\mathbf{AS}\right)$ in Step 1 can be different than the cost function $\tilde{D}\left(\mathbf{X}\middle\|\mathbf{AS}\right)$ in Step 2, however, usually $D\left(\mathbf{X}\middle\|\mathbf{AS}\right) = \tilde{D}\left(\mathbf{X}\middle\|\mathbf{AS}\right)$. In (Lee and Seung, 1999) the algorithm was first applied to two different cost functions: squared Euclidean distance (Frobenius norm) and Kullback-Leibler divergence. Using a gradient descent approach to perform Steps 1 and 2, they obtained multiplicative algorithms. However, the multiplicative algorithms are known to be very slowly convergent and easily get stuck in local minima. Therefore, in (Zdunek and Cichocki, 2007) an algorithm was recently derived that is based on the use of second-order terms, Hessian, in the Taylor expansion of a cost function to speed up convergence. Specifically, the NMF algorithm used in the experiments in the cited paper combines quasi-Newton optimisation for basis matrix **A** and a fixed-point regularised least-square algorithm for **S**, with computer code provided in the appendix in (Zdunek and Cichocki, 2007). Excellent performance of this algorithm has been demonstrated in (Zdunek and Cichocki, 2007; Cichocki et al., 2008). This algorithm will be used in the comparative performance analysis in section 3. We shall refer to this algorithm through the rest of the paper as the SO NMF algorithm.

2.4 SLMM and multi-spectral imaging

The SLMM (1) is widely used in multi-spectral and hyper-spectral remote sensing, (Adams et al., 1993; Settle and Drake, 1993; Du et al., 2006; Du and Kopriva, 2008), where 3D image cube contains co-registered spectral images of the same scene. Within this application field, N represents the number of spectral bands; rows $\{\mathbf{x}_n\}_{n=1}^{N}$ of $\mathbf{X}$ represent spectral images, and columns of $\mathbf{X}$ represent multi-spectral

pixel vectors at particular spatial locations, $t \leq T$, in the image; T represents the number of pixels in the image, while column vectors $\{\mathbf{a}_m\}_{m=1}^{M}$ of the basis or mixing matrix represent spectral responses of the corresponding sources $\{\mathbf{s}_m\}_{m=1}^{M}$ that themselves are rows of **S** that represent spatial distributions of the sources. Assuming that **X** represents an RGB fluorescent image the number of spectral bands N equals 3. In order to relate the SLMM (1) to the tumour visualisation, we point out again that sources $\mathbf{s}_m$ represent spatial maps of the tumour, surrounding healthy tissue and possibly some other material resident in the multi-spectral fluorescent image, while corresponding column vectors $\mathbf{a}_m$ of the basis or mixing matrix represent their spectral responses. Following this interpretation we immediately see that spectral similarity of the sources $\mathbf{s}_m$ and $\mathbf{s}_n$ will affect the condition number of the basis matrix, because the corresponding column vectors $\mathbf{a}_m$ and $\mathbf{a}_n$ become close to collinear. This is elaborated in great details in the appendix where effects of the NBG transform on the condition number of the extended mixing matrix are evaluated analytically and numerically. Simulation example in section 3.1 based on the computational model of the RGB image also demonstrates that condition number of the mixing matrix is increasing from 11.7 to 117 when angle between two spectral vectors decreases from 10 degrees to 1 degree. In addition to deteriorate conditioning of the mixing matrix spectral similarity between the sources makes them statistically dependent. This is easily verified by assuming that sources $\mathbf{s}_m$ and $\mathbf{s}_n$ are spectrally very similar. Then $\mathbf{a}_n \cong c\mathbf{a}_m$, where c represents the intensity scaling factor. The contribution of these two sources at any pixel location t is $\mathbf{a}_m s_{mt} + \mathbf{a}_n s_{nt} \cong \mathbf{a}_m s_{mt} + \mathbf{a}_m c s_{nt}$, implying that $\mathbf{s}_m$ and $c\mathbf{s}_n$ are indistinguishable i.e. $\mathbf{s}_m$ and $\mathbf{s}_n$ are statistically dependent. Hence, two fundamental requirements imposed by the ICA algorithms on the SLMM fail when the sources become spectrally highly similar. We point out that spectral similarity among the sources also affects the performance of the NMF algorithms due to the ill-conditioning of the basis matrix. It is NBG transform that is necessary to improve

conditioning of the extended basis matrix. However, as it will be demonstrated in section 3, it is also DCA that is necessary to be used in combination with NBG transform in order to account for statistical dependence induced by spectral similarity between the sources.

2.5 New algorithm: NBG and DCA

Here we propose a novel solution for unsupervised decomposition of the SLMM (1) for the case of ill-conditioned basis matrix and statistically dependent sources, by combining NBG transform (Ren and Chang, 2000; Du et al., 2004), that increases the dimensionality of the original multi-spectral image **X** and decreases the condition number of the extended basis matrix, and DCA (Cichocki and Georgiev, 2003; Hyvärinen, 1998; Kopriva and Seršić, 2008) that increases the level of statistical independence between the spectrally similar sources. This combined use of two transforms is original contribution of this paper that is important for robust blind decomposition of low-dimensional multi-spectral images such as RGB image for example. It is demonstrated on computational model and experimental data that proposed algorithm yields high-contrast tumour maps when intensity of the acquired fluorescent image is fluctuating more than 10 times causing increased spectral similarity between tumour and surrounding tissue.

The NBG process was originally proposed in (Ren and Chang, 2000) with the aim of increasing the accuracy of the orthogonal subspace projection method in decomposition and classification of the multi-spectral images and also used for the same purpose in (Du et al., 2004), when the number of sources M to be classified exceeds the number of spectral bands N. The same limitation, $N \geq M$, also applies to the ICA and DCA algorithms in unsupervised decomposition of the SLMM (1).

According to (Ren and Chang, 2000) the basic idea behind the NBG approach arises from the fact that a second-order random process is generally specified by its first-order and second-order statistics. Looking at the original spectral images $\{\mathbf{x}_n\}_{n=1}^{N}$ as the first-order statistical images, a set of second-order statistical spectral images can be generated by capturing the correlation between the spectral bands. They provide useful second-order statistical information about spectral bands that is missing in the set of original spectral images. Theoretically, any nonlinear function can be used to generate set of artificial images with linearly independent spectral measurements. Previous experimental studies (Du et al., 2004) have shown that nonlinear function that enlarges or emphasizes discrepancy between original spectral measurements will help to improve classification performance, since the technique applied here uses spectral information. A simplest but effective choice is multiplication. When two original spectral images are multiplied together a new artificial spectral image is generated. Here, multiplication acts as matched filtering. When multiplicant and multiplier are equal, the product is the maximum. So multiplication can emphasize the spectral similarity between two spectral measurements of the same pixel, which is equivalent to emphasizing their dissimilarity or discrepancy. Multiplication can be also used for a single band. Then, it emphasizes a single spectral measurement itself, which is also equivalent to enlarging the spectral difference from other spectral measurements of this pixel. Thus, second-order statistics, which include auto-correlation, cross-correlation and nonlinear correlation, are used to create nonlinearly correlated spectral images and increase the dimensionality of the original image **X**. The set of auto-correlated spectral images is obtained as $\{\mathbf{x}_n^2\}_{n=1}^{N}$ while the set of

cross-correlated spectral images is obtained as $\{\mathbf{x}_n \mathbf{x}_m\}_{m,n=1,m\neq n}^{N}$. The dimensionality of the original image $\mathbf{X}$ is extended from N to $2N + \binom{N}{2}$. Thus the dimensionality of an RGB image is increased from 3 to 9. Although dimensionality increase is important, especially when the number of spectral bands is small, there is another important property of the NBG technique presented here for the first time. Namely, the NBG technique causes the two angularly close spectral vectors $\mathbf{a}_m$ and $\mathbf{a}_n$ to become more separated in the band-expanded version of the original image. We denote the band-expanded image as $\overline{\mathbf{X}}$. This is achieved provided that the two corresponding sources $\mathbf{s}_m$ and $\mathbf{s}_n$ do not have exactly the same intensity. Detailed demonstrations for 2D and 3D problems with cross-correlated and auto-correlated bands are presented in appendix. Thus the NBG technique significantly improves conditioning of the original basis or mixing matrix caused by spectral similarity between the sources. Since in a tumour demarcation problem we are interested in spatial localisation of the tumour class, we shall impose a special constraint on the sources in the SLMM (1): $\mathbf{s}_m \approx \{0,1\}$ i.e. we presume quasi-binary nature of the spatial maps of the sources, where 1 indicates source presence and 0 indicates source absence at the pixel level. However, the truth is that sources are continuous and not binary. But if blind decomposition is reasonably successful other sources present in each extracted source spatial map will be suppressed significantly. Thus, the extracted sources could be, at least as the first approximation, modeled as quasi-binary. This assumption is necessary to make analytical and numerical quantification of the effects of NBG transform mathematically tractable. The analysis itself is in presented in the appendix for 2D and 3D problems, where it is demonstrated that NBG transform really improves

conditioning of the extended mixing matrix in relation to the original mixing matrix. This is especially the case when the column vectors of the original mixing matrix become close to collinear (the angular separation between the column vectors is small), see Figures 7 to 9.

Another assumption that is necessary to carry on analysis discussed above is that sources do not overlap in the spatial domain: $s_{mt}s_{nt}=\delta_{mn}$, where t denotes pixel location and δ_{mn} represents Kronecker's delta symbol. This helps to get rid out of the cross-terms that show up in nonlinear band expansion process. This no-overlapping assumption is however not pure mathematical construct. It has justification in tumour demarcation problem due to the fact that the pixel footprint is small, usually less than a square mm, and it is highly unlikely that the tumour occupies the same pixel location with some other source resident in the multi-spectral image. Under these two assumptions (quasi-binary and no-overlapping sources) it is straightforward to show that the NBG technique applied to the SLMM (1) produces the following result:

$$x_{mt}x_{nt} \cong \sum_{k=1}^{M} a_{mk}a_{nk}s_{kt}$$

Thus in matrix notation model (1) is transformed into

$$\overline{\mathbf{X}} = \overline{\mathbf{A}}\mathbf{S} \tag{12}$$

Spectral similarity between the sources $\mathbf{s}_m$ and $\mathbf{s}_n$ makes them also statistically dependent. A few approaches that deal with the problem of statistically dependent sources are presented in (Cichocki and Amari, 2002; Cichocki and Georgiev, 2003;

Hyvärinen, 1998; Kopriva and Seršić, 2008, Kopriva, 2007) and references cited therein. As elaborated in section 2.2, the problem is fixed by finding a linear operator H with the property

$$H\left(\overline{\mathbf{X}}\right) = H\left(\overline{\mathbf{A}}\mathbf{S}\right) = \overline{\mathbf{A}}H\left(\mathbf{S}\right) \tag{13}$$

such that the transformed sources $H(\mathbf{S})$ are more statistically independent than the original sources $\mathbf{S}$. We have been already discussed the reasons for choosing high-order innovations transform with JADE ICA algorithm to form DCA algorithm, as well as to choosing first order innovation transform (first order high-pass filter) with the SOBI ICA algorithm to form another DCA algorithm. Coefficients of prediction-error filter $\boldsymbol{b}_m$ in (10) are efficiently estimated by means of the Levinson algorithm (Orfanidis, 1988). The filter is applied row-wise on the image $\overline{\mathbf{X}}$ and is obtained as an average of the prediction-error filters estimated for each image $\{\overline{\mathbf{x}}_n\}_{n=1}^{N}$. Depending on the order of the prediction error filter used, either JADE or SOBI ICA algorithms are applied to the SLMM (13) in order to learn more accurately the extended basis matrix $\overline{\mathbf{A}}$ or its inverse $\overline{\mathbf{W}}$ than would be possible from (12). Sources $\mathbf{S}$ are then recovered by applying $\overline{\mathbf{W}}$ to the band-expanded image given by (12).

Before proceeding to the presentation of the results, we briefly comment on two important remaining issues: estimation of the unknown number of sources M, and prioritisation of the extracted sources by the measure of information contained in them. Determination of the number of sources is a very old problem in multivariate data analysis, and is also known as intrinsic dimensionality problem (Fukunaga and Olsen, 1971). Several methods for estimating the number of sources in a hyperspectral image have been tested in (Chang and Du, 2004). Malinowski's method (Malinowski, 1977a; Malinowski, 1977b) developed for determination of the number of factors in absorption spectroscopy, mass spectra, and chromatography has been demonstrated in (Chang and Du, 2004) to give a good result in estimating the number of sources

resident in the hyper-spectral image. We have used this method in estimating the number of sources in the band-expanded version of the multi-spectral image given by (5). When dealing with the original RGB image given by (1), estimating the number of sources was not an issue, because *a priori* knowledge and the fact that maximal number of sources is limited by the number of spectral bands, that is 3, narrowed down the number of sources to 1, 2 or 3. To prioritise the extracted sources we used the high-order statistics-based approximation of the negentropy measure (Wang and Chang, 2006) and Eq.(5), p. 115 in (Hyvärinen et al., 2001):

$$J(\mathbf{s}_m) \cong \frac{1}{12}\left[\kappa_m^3\right]^2 + \frac{1}{48}\left[\kappa_m^4 - 3\right]^2 \tag{14}$$

where $k_m^3 = (1/T)\sum_{t=1}^{T}(s_{mt})^3$ and $k_m^4 = (1/T)\sum_{t=1}^{T}(s_{mt})^4$ are sample means of third and fourth order statistics of $\mathbf{s}_m$.

3. Results

We now execute comparative performance evaluation of the presented method on the computational model of an RGB fluorescent image of a surface tumour as well as on experimental RGB fluorescent images of the surface tumour (basal cell carcinoma). For this purpose two experiments have been carried out. In first experiment we have recorded an image of a skin tumour after being treated for 4h with δ-5-aminolaevulinic acid. This is a photo-synthesiser that, through the process of biosynthesis, causes formation of the fluorophore protoporphyrin IX (PpIX), (Koenig et al., 2001). The tumour was illuminated with 405 nm light, which induces

fluorescence of the PpIX. The fluorescent image was recorded by a camera with an attached filter used to filter out the reflected 405 nm light. In order to simulate effects of variability of the parameters such as concentration of the photo-synthesiser, intensity of illumination light and duration of the treatment with the photo-synthesiser, on the variation of the fluorescent image intensity, we acquired a fluorescent RGB image after illumination with high-intensity light as well as with light with a weak intensity level. The high-intensity fluorescent image shown in Fig. 1 was used to extract spatial binary maps of the tumour and surrounding healthy tissue that served as a ground truth in the comparative performance evaluation.

In second experiment we have acquired a sequence of RGB fluorescent images of the patient with histologically verified diagnoses of superficial multicentric basal cell carcinoma in the lower right part of the back. As before, skin of the tumour was treated for 4h with δ-5-aminolaevulinic acid before images have been acquired and the tumour was illuminated with 405 nm light.The intensity of the illuminating light has been gradually varied over 10 times. Two RGB fluorescent images with illumination intensity that differs 10.91 times are shown in Figures 4a and 4b.

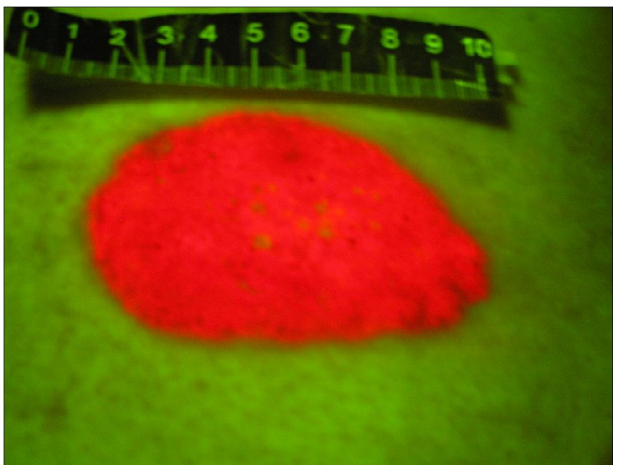

Figure 1. RGB fluorescent image of the skin tumour acquired after illumination with high-intensity light.

3.1 Simulation results

To carry out comparative performance analysis we have first created a computational model of an RGB image based on (1) in order to produce a synthetic fluorescent image with a controlled degree of spectral similarity between the tumour and surrounding healthy tissue. Thus, we have M=2 sources and by convention we choose source $\mathbf{s}_1$ to represent the tumour and source $\mathbf{s}_2$ to represent the surrounding healthy tissue. In the corresponding column vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ we set the components that correspond with the blue colour to be equal. Projections of both vectors into the red-green plane were separated by angle α, where the spectral response of the healthy tissue was 21 degrees apart from the green axis. Thus by changing the angle α we were able to control the degree of spectral similarity between tumour and healthy tissue. In addition to that, the intensity of the tumour was related to the intensity of the healthy tissue through $|\mathbf{a}_1| = c|\mathbf{a}_2|$. The intensity of the extracted tumour map was rescaled to [0, 1] interval such that interpretation of probability could be assigned to it. As a figure of merit we have selected probability margin, Δp, defined as the difference between the minimal probability that the tumour is present in the tumour region and the maximal probability that the tumour is present in the surrounding healthy tissue region. As seen from Fig. 2a, where Δp is plotted vs. α for signal-to-noise ratio (SNR) equal 30dB and c=1.1, the proposed method (it combines NBG and DCA with higher order innovations and JADE ICA algorithm) exhibits high contrast level even with extremely small values of angle α (α=0.01 degree) when the tumour and surrounding tissue are spectrally practically the same. The condition number of the corresponding mixing matrix was 11708. We point out that even when α=0 degrees and c=1.01 the separation margin achieved by the proposed algorithm was Δp=0.8624 with SNR=30dB. All other state-of-the-art methods under consideration

failed due to the rank deficiency problem of the basis matrix in the SLMM (1). We have also applied DCA algorithm directly (it combines higher order innovations and JADE ICA algorithm) to the original RGB image (1). The "temporal" predictability-based ICA algorithm (Stone, 2001) has been also applied directly to the original RGB image. Finally, original RGB image has been also decomposed by the SO NMF algorithm (Zdunek and Cichocki, 2007). The performance of the SO NMF and DCA algorithms is similar, and improves when the level of spectral dissimilarity increases. Once innovations-based pre-processing has increased statistical independence between the sources, ICA exhibits the same performance level as SO NMF, which is not sensitive to statistical dependence between the sources. However, both methods are sensitive to the ill-conditioning of the basis matrix and that is why proposed algorithm outperforms all other methods considered. Improved conditioning also gives the proposed algorithm a higher robustness level against additive noise, which is demonstrated in Fig. 2b where Δp is plotted vs. SNR for α=10 degrees and c=1.1. For the purpose of visual demonstration, Fig. 2c and Fig. 2d show synthetic RGB images for α=1 degree and α=10 degrees respectively with SNR=30dB and c=1.1. The visual contrast is very poor in Fig. 2c. Fig. 2e and Fig. 2f show tumour maps extracted from Fig. 2c by the proposed algorithm and the SO NMF algorithm after 2000 iterations, respectively. Note the very high contrast in Fig. 2e with Δp=0.866, while for Fig. 2f Δp=0.4648. The condition numbers of the corresponding basis matrices were 117 and 11.7. For the sake of comparison we mention that the condition numbers of the mixing matrices in the benchmark experiments used to test the SO NMF algorithm in (Zdunek and Cichocki, 2007) were between 4 and 10. It is noteworthy that computation time for Fig. 2e was around 30s, and for Fig. 2f around 330s in a MATLAB environment on a desktop computer with 3GHz clock speed and 4GB of RAM.

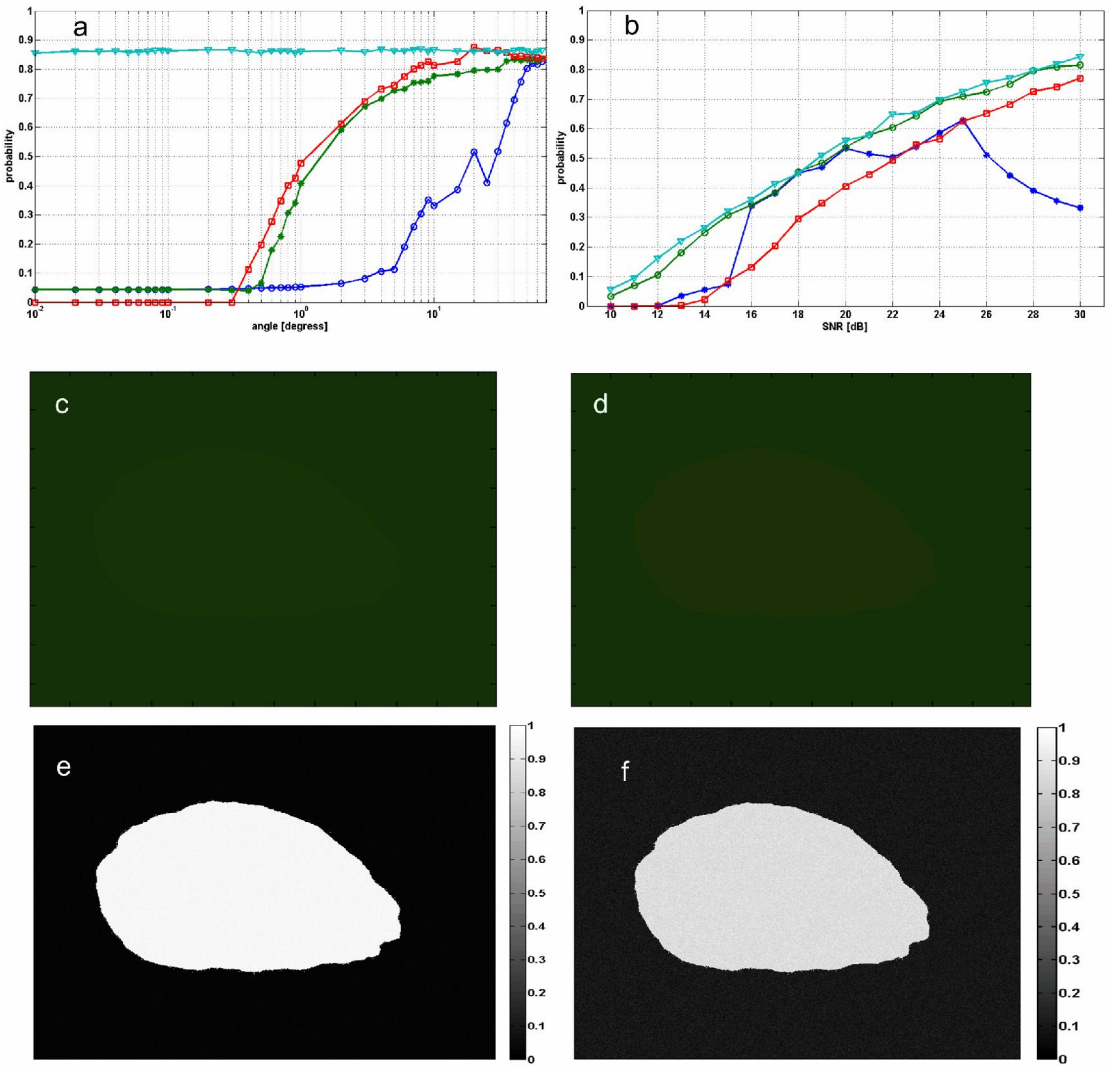

Figure 2. Computational model of fluorescent RGB image. (a) Probability margin Δp vs. angle α for the tumour map extracted by the proposed method (light blue triangles), by the SO NMF method (red squares), by DCA (green circles) and by "temporal" predictability maximization ICA algorithm (dark blue stars). SNR at the sensor level was 30dB. (b) probability margin Δp vs. SNR for α=10 degrees. (c) simulated RGB image for α=1 degree, SNR=30dB and c=1.1. (d) simulated RGB image for α=10 degrees, SNR=30dB and c=1.1. (e) tumour map extracted by the proposed algorithm from (c) with probability margin Δp=0.866. (f) tumour map extracted by the SO NMF algorithm from (c) with probability margin Δp=0.4648.

3.2 Results for experiment 1

We now present the results of the comparative performance analysis for the low-intensity version of the RGB fluorescent image shown in Fig. 1. The low-intensity image itself is shown in Fig. 3a. The receiver-operating-characteristic (ROC) curves selected for this experimental example as a figure of merit, that show probability of detection vs. probability of false alarm, are shown in Fig. 3b for the same algorithms as in Fig. 2. Note that the proposed algorithm exhibits significantly better performance in the region of small probability of false alarm than other methods in consideration. DCA, SO NMF and ICA performed similarly. Performance of the higher order statistics based ICA algorithms is inferior in relation to the second order statistics based method used here (Stone, 2001). Fig. 3c and Fig. 3d show tumour maps extracted from Fig. 3a by the proposed algorithm and by the SO NMF algorithm, respectively. The contrast of the tumour map in Fig. 3c is much better. That is confirmed in Fig. 3e and Fig. 3f, which show tumour demarcation lines calculated from Fig. 3c and Fig. 3d by Canny's edge extraction method with threshold set to 0.5. In relation to the experimental image shown in Fig. 3a, we make another comment: that its complexity was artificially increased by adding a ruler in to the scene. Because the ruler is not spectrally homogenous, it increased the number of sources resident in the image, probably to four, which exceeded the number of spectral bands of the original image. Despite that, and despite the fact that intensity was low, the proposed method performed well.

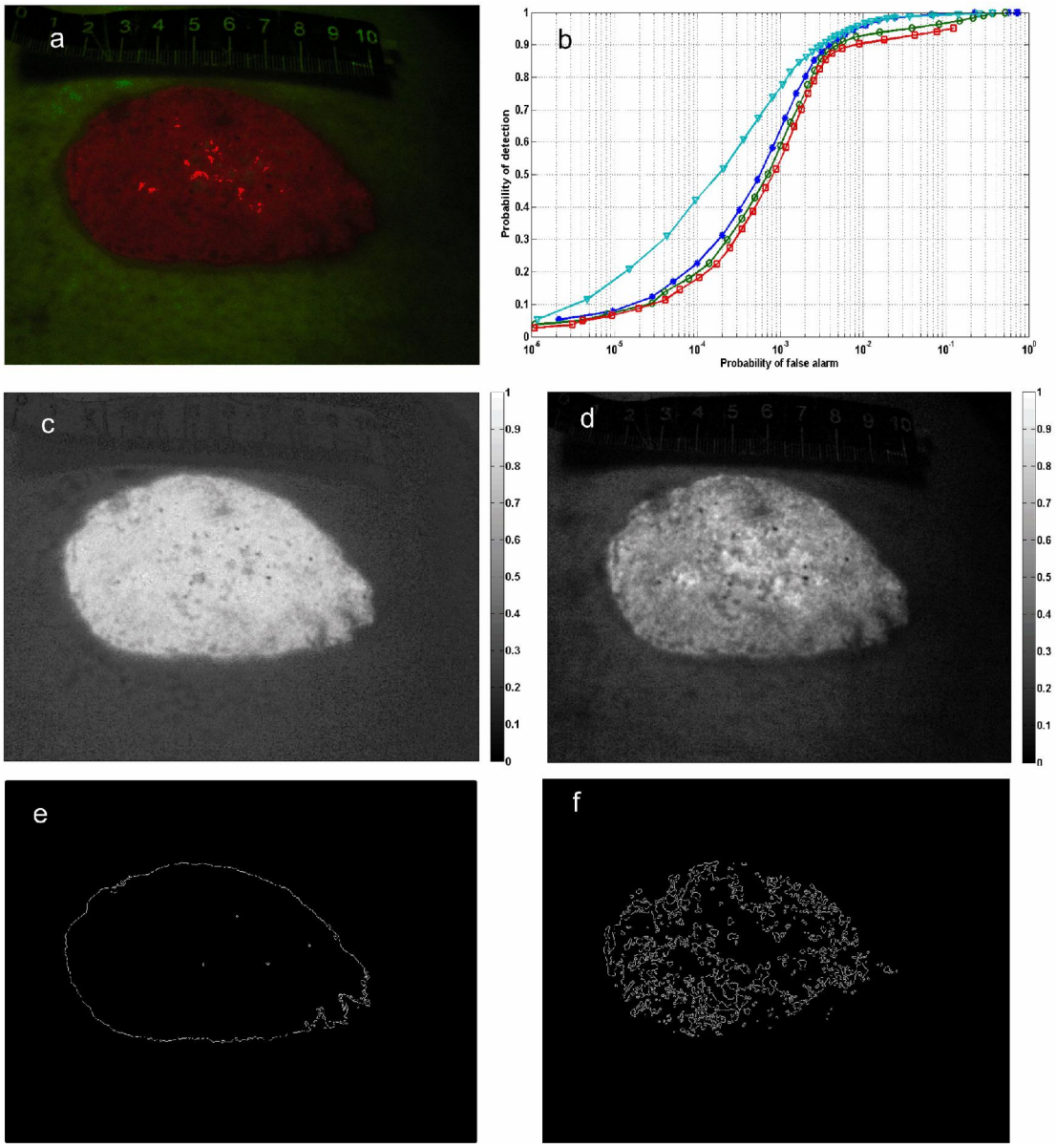

Figure 3. Experimental fluorescent image of skin tumour. (a) Low-intensity version of the RGB fluorescent image of the tumour shown in Fig. 1. (b) ROC curves with the legend described for Fig. 2a. (c) tumour map extracted by the proposed algorithm. (d) tumour map extracted by the SO NMF algorithm. Assigned probabilities in the grey scale colour are shown on vertical bars. (e) tumour demarcation line calculated by Canny's edge extraction method from (c). (f) tumour demarcation line calculated by Canny's edge extraction method from (d). The threshold used in the edge-extraction algorithm was set to 0.5.

3.3 Results for experiment 2

In order to quantify robustness of the proposed algorithm against intensity fluctuation we have acquired a sequence of RGB fluorescent images of the patient with histologically verified diagnoses of superficial multicentric basal cell carcinoma in the lower right part of the back. The intensity of the illuminating light has been gradually varied over 10 times. Two RGB fluorescent images with illumination intensity that differs 10.91 times are shown in Figures 4a and 4b. Figure 4c shows demarcation line of the tumour marked manually by red dots. Through biopsy it has been verified that there is no tumor outside of the region marked by red dots. Thus, Figure 4c can serve as a reference in comparative performance analysis of the tumor demarcation methods. Three best results are chosen and shown in Figure 5. It shows extracted spatial maps of the tumors rescaled to [0, 1] interval in the pseudo color scale, where dark blue color represents 0 (absence of the tumor) and red color represents 1 (presence of the tumor). First row shows maps extracted by "temporal" predictability-based ICA algorithm. Second row shows maps extracted by proposed algorithm where DCA part was composed of tenth order innovation process and JADE ICA algorithm. Third row shows maps extracted by proposed algorithm where DCA part was composed of first order high-pass filtering and SOBI ICA algorithm. First column shows maps extracted from RGB image shown in Figure 4a. Second column shows maps extracted from RGB image shown in Figure 4b, whereas intensity of the illumination was 10.91 times leas than in Figure 4a. The contrast between the tumor area and background area is the best preserved by proposed method with DCA algorithm composed of tenth order innovation process and JADE ICA algorithm. This statement is further supported by Figure 6, where length of the demarcation line in pixels is shown as a function of the relative intensity of the illuminating light. Relative intensity has been calculated as I_0/I_n where I_0 represented

maximal value of the intensity. Demarcation lines were calculated after edges were extracted from tumor maps by Canny's edge extraction method with a threshold varying in the interval [0.4, 0.6]. This small deviation of the threshold from the unbiased value of 0.5 was necessary to account for the quasi-binary nature of the sources. Note that high quality binary spatial maps of the tumor can be obtained from maps extracted by blind decomposition methods if some reasonably advance clustering algorithm is applied to them. Here we were applying Canny's edge extraction method with basically fixed threshold in order to emphasize good and robust performance of the proposed method against intensity fluctuation. Proposed method with DCA algorithm composed of tenth order innovation process and JADE ICA algorithm exhibited the best performance in term of the stability of demarcation line. Standard deviation for this case was estimated as 127.1 pixels. Version of the proposed algorithm when DCA was composed of first order high-pass filtering and SOBI ICA method yielded standard deviation of 263.8 pixels. "Temporal" predictability-based ICA algorithm when applied to RGB image directly yielded standard deviation of 327.8 pixels. When the same algorithm was applied to band expanded image it yielded standard deviation of 1157.6 pixels.

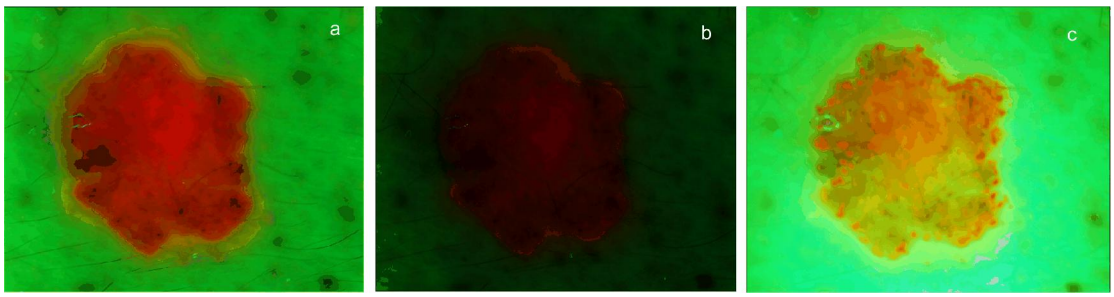

Figure 4. a) RGB image with maximal intensity; b) RGB image with 10.91 smaller intensity; c) RGB image with manually marked tumour demarcation lines (red dots).

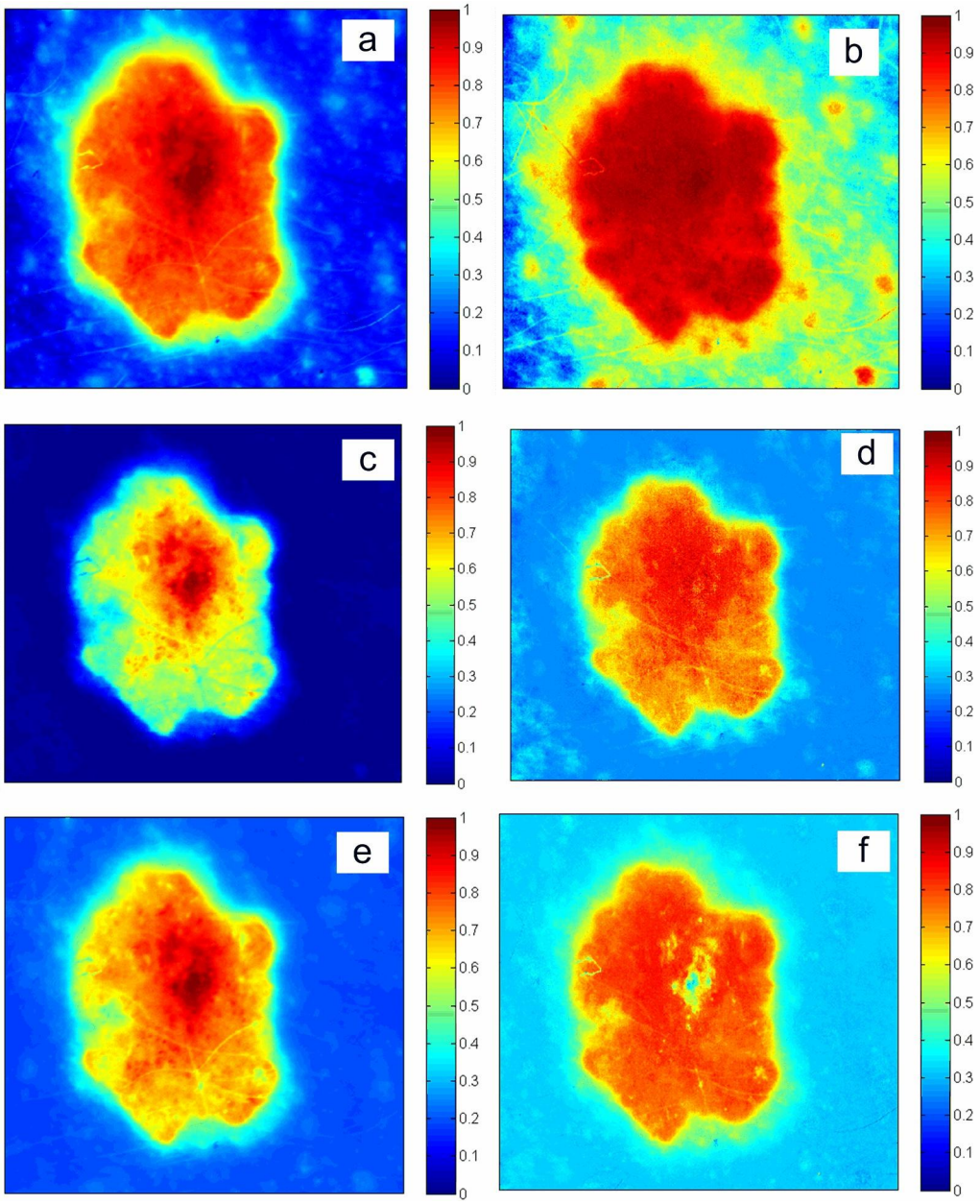

Figure 5. Spatial maps of the BCC extracted by 'temporal' predictability-based ICA algorithm (a and b), NBG algorithm with innovations-based preprocessing and JADE (c and d) and NBG algorithm with the first order high-pass filter-based preprocessing and SOBI (e and f) from RGB fluorescent images illuminated by light with maximal intensity (a, c and e) and by light with 10.91 times smaller intensity (b, d and f). Dark blue colour indicates that tumour is present with probability 0, while red colour indicates that tumour is present with probability 1.

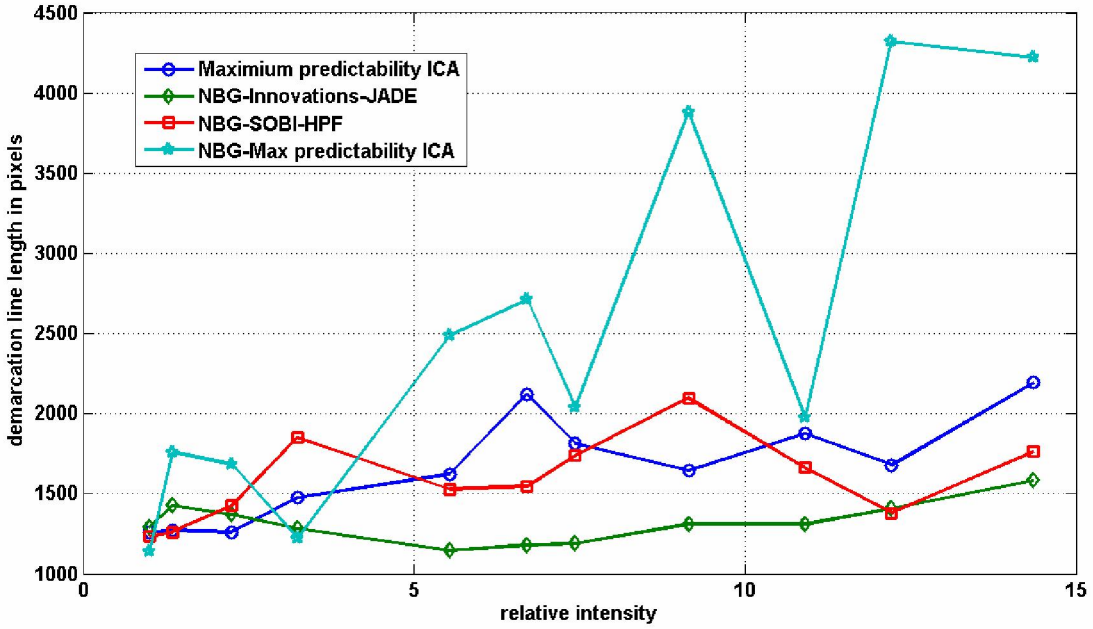

Figure 6. Length of the demarcation lines, in pixels, of the tumour spatial maps extracted from RGB fluorescent images illuminated by light with different intensity levels. Horizontal axis from left to right represents ratio between maximal intensity and intensity of each particular illumination. Legend: sky blue stars-'temporal' predictability-based ICA algorithm applied to band expanded RGB image; dark blue circles- 'temporal' predictability-based ICA algorithm applied on original RGB image; red squares- SOBI ICA algorithm applied on first order high pass filter-based pre-processed band expanded RGB image; green diamonds-JADE ICA algorithm applied on innovations-based pre-processed band expanded RGB image.

4.0 Discussion

We have presented an algorithm for high-contrast tumour visualisation and demarcation through unsupervised decomposition of a low-intensity low-dimensional (RGB) multi-spectral fluorescent image of the tumour. Coarse spectral resolution of the RGB-like images in combination with the low-intensity of the fluorescence yields

ill-conditioned SLMM with statistically dependent sources. This represents a great challenge for many state-of-the-art blind image decomposition algorithms. Proposed method is composed of two algorithms: NBG transform that improves conditioning of the extended basis matrix and DCA that increases statistical independence between the sources. We have demonstrated good performance of this method on experimental RGB fluorescent images of the BCC. High-contrast tumour maps were extracted when the intensity of the illuminating light was varied more than 10 times causing spectral similarity between the tumour and surrounding tissue. There are various situations in clinical practice that could benefit from these results: (*i*) concentration of the photo- synthesiser can be reduced; (*ii*) duration of treatment with the photo-synthesiser can be reduced; (*iii*) intensity of the illumination light can be reduced; (*iv*) tumour detection in the early stage of development should be possible. Reduced intensity of the illuminating light could be important, for example in surgery on brain tumours, which can last for several hours, and where high intensity of the illuminating light can cause damage to healthy tissue. Although the method is demonstrated for the visualisation and demarcation of a surface tumour, it ought to be equally applicable to visualisation and demarcation of other types of tumours, for example tumours at the cell level, where fluorescent nanoparticle markers are used to label the tumour cells (Tréhin et al., 2006), or where visualisation is based on a green fluorescent protein gene tumour transduction system (Hasegawa et al., 2000; Hoffman, 2002). Extraction of the tumour map from the spectrally similar surrounding tissue can efficiently be executed by the ICA algorithms alone and hyper-spectral imaging technology. However, achieving that with RGB images acquired by cheap commercial digital cameras is important from the affordability viewpoint. It should enable easier dissemination of the tumour visualisation technology presented herein. Application of the proposed method for real time tumour demarcation is straightforward and it is only a matter of the hardware platform chosen for the algorithm implementation.

Finally, we note that the general unsupervised decomposition method described above in the tumour demarcation and multi-spectral imaging context should be applicable to other types of the multi-channel medical imaging systems.

Acknowledgments

The work of I. Kopriva on algorithm development for unsupervised image decomposition was supported by the Ministry of Science, Education and Sport, Republic of Croatia Grant 098-0982903-2558. The work of A. Peršin on collecting data used in Fig. 1 and Fig. 3a was supported by the same Ministry through the project TP-05/0098-47.

Appendix. NBGP applied to 2D and 3D problems of spectrally identical sources

Let us assume that the SLMM (1) is specified by two spectral bands, N=2, and two sources, M=2. Let us also assume that the sources are spectrally identical and differ only in the intensity of their spatial distribution by some scale factor c. Under these assumptions the model (1) becomes

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & ca_{11} \\ a_{21} & ca_{21} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} \tag{15}$$

Clearly the rank of the basis matrix in (15) is 1, and the unsupervised decomposition problem does not have a solution. Let us now assume that the cross-correlated band has been generated. Under the quasi-binary constraint and no-overlapping assumption, the NBGP model (12) becomes

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_1\mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & ca_{11} \\ a_{21} & ca_{21} \\ a_{11}a_{21} & c^2a_{11}a_{21} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} \tag{16}$$

The cosine of the angle, β, between the two column vectors in (16) is given by

$$\cos\beta = \frac{a_{11}^2 + a_{21}^2 + c a_{11}^2 a_{21}^2}{\sqrt{a_{11}^2 + a_{21}^2 + a_{11}^2 a_{21}^2}\sqrt{a_{11}^2 + a_{21}^2 + c^2 a_{11}^2 a_{21}^2}} = \frac{p}{q} \tag{17}$$

After some elementary algebra we arrive at the following identity

$$q^2 - p^2 = (c-1)^2 \left(a_{11}^2 + a_{21}^2\right) a_{11}^2 a_{21}^2 \tag{18}$$

Evidently $q=p$, implying that column vectors remain collinear, occurs only when $c=1$. That is a trivial case where sources are not only spectrally identical but also have the same intensity level. In fact there is only one source in such a case. Otherwise, for $c\neq 1$ it follows that $q\neq p$ and $\cos\beta\neq 1$, implying that the two spectral vectors are not collinear any more.

The algebraic complexity of the analytical expressions grows very fast with the increase of the dimensionality of both original and expanded mixing matrices. Therefore, we continue to demonstrate numerically improvement of the conditioning of the extended mixing matrix through NBG transform for more complex 2D and 3D problems. We first consider 2×2 SLMM with the mixing matrix as follows

$$\mathbf{A} = \begin{bmatrix} 1 & \cos\varphi \\ 0 & \sin\varphi \end{bmatrix} \tag{19}$$

Here φ is the mixing angle that defines position of the mixing vector in 2D space (plane). Obviously, when $\varphi=0^0$ the two column (also called mixing) vectors are collinear, in which case the mixing matrix is singular and the condition number is infinite. Let us now assume that the cross-correlated band has been generated. Under

the quasi-binary constraint and no-overlapping assumption, the NBGP model (12) becomes of the form (16) with the extended mixing matrix as follows

$$\overline{\mathbf{A}} = \begin{bmatrix} 1 & \cos\varphi \\ 0 & \sin\varphi \\ 0 & \cos\varphi\sin\varphi \end{bmatrix} \tag{20}$$

Let us now assume that in addition to the cross-correlated spectral band the two auto-correlated spectral bands have been also generated. Under the quasi-binary constraint and no-overlapping assumption the extended mixing matrix becomes

$$\overline{\mathbf{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \cos\varphi & \sin\varphi & \cos\varphi\sin\varphi & \cos^2\varphi & \sin^2\varphi \end{bmatrix}^{\mathrm{T}} \tag{21}$$

We show in figure 7 condition numbers of the three matrices given by (19), (20) and (21) as a function of the mixing angle φ. The last transform resulting in 5×2 extended mixing matrix (21) yields small condition number even when mixing angle φ=0^0 in which case original mixing matrix (19) is singular with infinite condition number. It is equivalent to say that NBG transform separated two column vectors in the 5D space even though their projections on 2D subspace were collinear. We now carry on this numerical demonstration for the 3D problem that corresponds with the RGB multispectral imaging case elaborated experimentally in Section 3. The mixing matrix of the 3×2 problem is given as follows

$$\mathbf{A} = \begin{bmatrix} 1 & \cos\varphi\sin\theta \\ 1 & \sin\varphi\sin\theta \\ \sqrt{2} & \cos\theta \end{bmatrix} \tag{22}$$

In this case two mixing angles: azimuth φ and elevation θ are necessary to define position of the mixing vector in the 3D space. When the two mixing angles are $\varphi=\theta=45^0$ the two column vectors will be collinear and the condition number will be infinite. Let us now assume that three cross-correlated spectral bands and three auto-correlated spectral bands have been generated. Under the quasi-binary constraint and no-overlapping assumption the extended mixing matrix becomes

$$\overline{\mathbf{A}} = \begin{bmatrix} 1 & 1 & \sqrt{2} & 1 & \sqrt{2} & \dots \\ \cos\varphi\sin\theta & \sin\varphi\sin\theta & \cos\theta & \cos\varphi\sin\theta\sin^2\theta & \cos\varphi\sin\theta\cos\theta & \dots \end{bmatrix}$$

$$\begin{bmatrix} \dots & \sqrt{2} & 1 & 1 & 2 \\ \dots & \sin\varphi\sin\theta\cos\theta & \cos^2\varphi\sin^2\theta & \sin^2\varphi\sin^2\theta & \cos^2\theta \end{bmatrix}^{\mathrm{T}} \qquad (23)$$

We show in figure 8 condition numbers of the 3×2 mixing matrix given by (22) as function of the mixing angles φ and θ. Two avoid infinite condition number that occurs at $\varphi=\theta=45^0$ mixing angles were varied with the resolution of 2^0 in the interval $[30^0, 60^0]$. Condition numbers are denoted according the colour scheme on vertical colour bar. Figure 9 shows condition numbers for extended 9×2 mixing matrix (23). They are more than 10 times less than it is the case with the original mixing matrix (22). Even at the position $\varphi=\theta=45^0$ condition number of the extended mixing matrix (23) remains small. We can again conclude that NBG transform separated two column vectors in the 9D space even though their projections on 3D subspace were collinear. The above analysis could be in principle carried on for higher-than-3D problems wherein hypershperical coordinate system can be used to define position of the mixing (column) vectors in higher dimensional space. However, as already discussed we are especially interested in low-dimensional multispectral imaging technology where, due to the coarse spectral resolution, the problems associated with the spectrally similar materials are expected to be present more often.

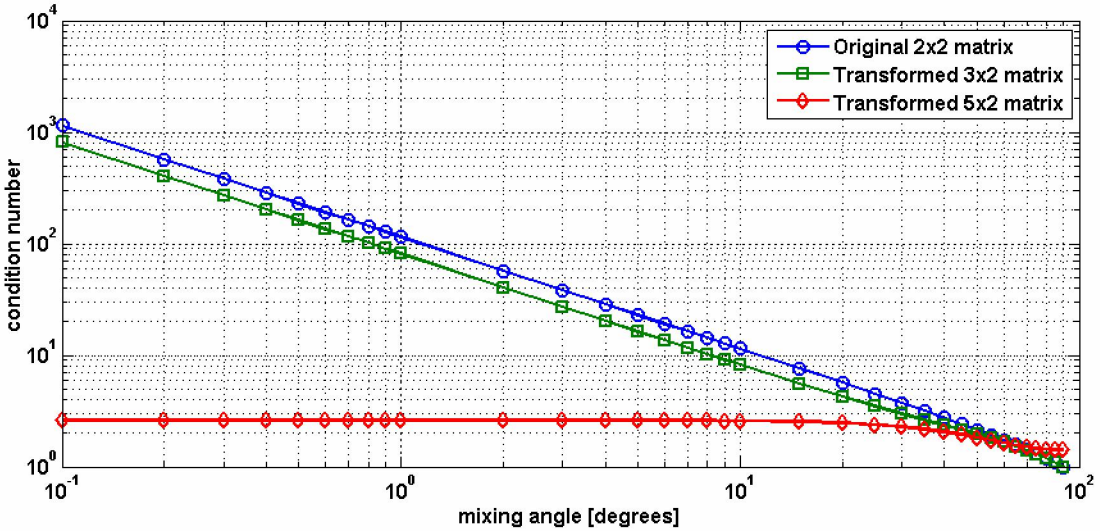

Figure 7. Condition numbers vs. mixing angle for: original 2x2 mixing matrix, Eq.(19), (blue circles); expanded 3x2 mixing matrix, Eq.(20), (green squares); expanded 5x2 mixing matrix, Eq.(21), (red diamonds).

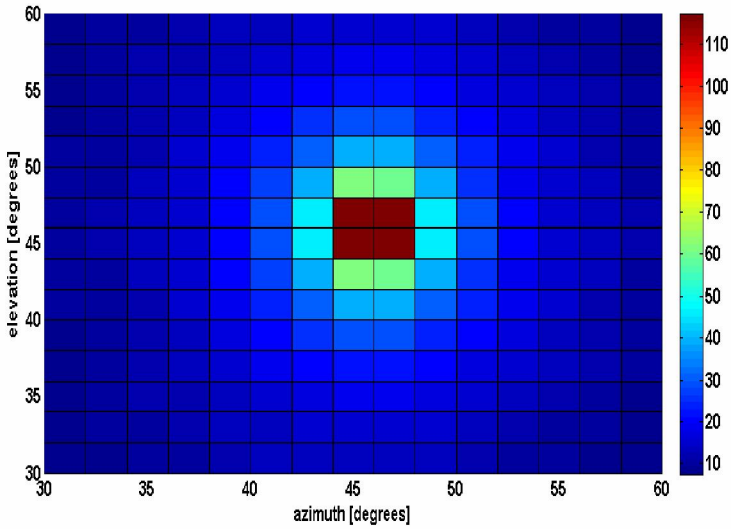

Figure 8. Condition numbers, according to the colour scheme on vertical colour bar, for 3x2 mixing matrix, Eq.(22), as a function of elevation and azimuth mixing angles. The mixing angles were varied in the two-degrees grid to avoid infinite condition number that would occur at 45 degrees of azimuth and elevation.

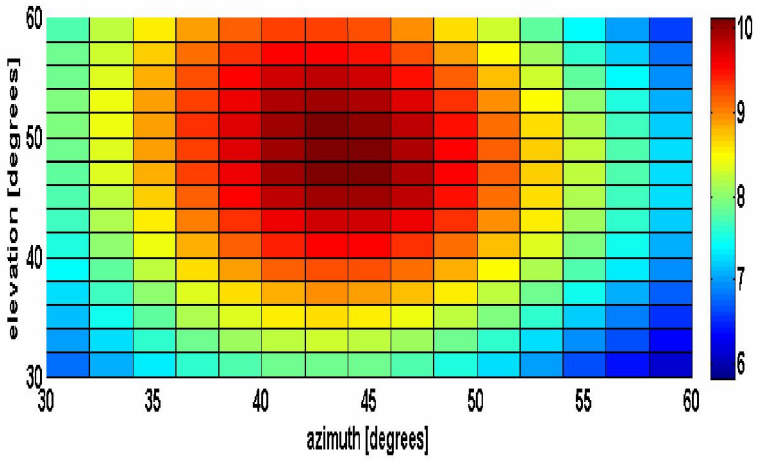

Figure 9. Condition numbers, according to the colour scheme on vertical colour bar, for 9x2 mixing matrix, Eq.(23), as a function of elevation and azimuth mixing angles.

References

Adams, J.B., Smith, M.O., Gillespie, A.R, 1993. Image spectroscopy: interpretation based on spectral mixture analysis. In: Pieters, C.M., Englert P.A. (Ed.), Remote Geochemical Analysis: Elemental and Mineralogical Composition, Cambridge University Press, Cambridge, Massachusetts.

Bäumler, W., Abels, C., Szeimies, R.M., 2003. Fluorescence diagnosis and photodynamic therapy in dermatology. Med. Laser Appl. 18, 47-56.

Bell, A., Sejnowski, T. J., 1995. An information maximization approach to blind separation and blind deconvolution. Neural Comput. 7, 1129-1159.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.F., Moulines, E., 1997. A blind source separation technique based on second order statistics. IEEE Trans. Sig. Proc. 45, 434-444.

Cardoso, J.F., Soulomniac, A., 1993. Blind beamforming for non-Gaussian signals. Proc. IEE F 140, 362-370.

Chang, C. I., Du, Q., 2004. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. IEEE Trans. on Geosc. and Remote Sensing 42, 608-619.

Choi, S., Cichocki, A., Amari, S., 2000. Flexible independent component analysis. J. VLSI Signal Process. 26 (1-2), 25-38.

Cichocki, A., Amari, S., Adaptive Blind Signal and Image Processing, John Wiley, New York, 2002.

Cichocki, A., Georgiev, P., 2003. Blind source separation algorithms with matrix constraints. IEICE Trans. Fund. Electron. Commun. Comput. Sci. E86-A, 522-513.

Cichocki, A., 2007. General Component Analysis and Blind Source Separation Methods for Analyzing Multichannel Brain Signals. In: Statistical and Process Models of Cognitive Aging, Notre Dame Series on Quantitative Methods, Editors: M.J Wenger and C. Schuster, Mahwah, NJ: Erlbaum.

Cichocki, A., Zdunek, R., Amari, S., 2008. Nonnegative Matrix and Tensor Factorization. IEEE Sig. Proc. Magazine **25**, 142-145.

Du, Q., Kopriva, I., Szu, H., 2004. Independent Component Analysis for Classifying Multispectral Images with Dimensionality Limitation. Int. J. of Inf. Acq. 1**,** 201-216.

Du, Q., Kopriva, I., Szu, H., 2006. Independent component analysis for hyperspectral remote sensing. Optical Engineering 45, 017008:1-13.

Du, Q., Kopriva, I., 2008. Automated Target Detection and Discrimination Using Constrained Kurtosis Maximization. IEEE Geoscience Remote Sensing Letters 5, 38-42.

Ericson, M.B., Sandberg, C., Gudmundson F., Rosén, A., Larkö O., Wennberg A.M., 2003. Fluorescence contrast and threshold limit: implications for photodynamic

diagnosis of basal cell carcinoma. Journal of Photochemistry and Photobiology B: Biology 69, 121-127.

Fischer F., Dickson, E.F., Pottier, R.H., Wieland, H., 2001. An affordable, portable fluorescence imaging device for skin lesion detection using a dual wavelength approach for image contrast enhancement and aminolaevulinic acid-induced protoporphyrin IX, part I: Design, spectral and spatial characteristics. Lasers Med Sci 16, 199-206.

Fukunaga, K., Olsen, D. R., 1971. An algorithm for finding intrinsic dimensionality of data. IEEE Trans. on Computers **C-20**, 176-183.

Hasegawa, S., Yang, M., Chishima, T., Miyagi, Y., Shimada, H., Moossa, A.R., Hoffman, R.M., 2000. In vivo tumor delivery of the green fluorescent protein gene to report future occurence of metastasis. Cancer Gene Therapy 10, 1336-1340.

Hoffman, R. M., 2002. Green fluorescent protein imaging of tumor growth, metastasis, and angiogenesis in mouse models. The Lancet Oncology 3, 546-556.

Hyvärinen, A., Karhunen, J., Oja, E. Independent Component Analysis, Wiley Interscience, New York, 2001.

Hyvärinen, A., 1998. Independent component analysis for time-dependent stochastic processes. In: Proceedings of the international conference on artificial neural networks (ICANN'98), Skovde, Sweden, pp. 541-546.

Jutten, C., Herault, J., 1991. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Proc. 24, 1-10.

Koenig, F., Knittel, J. & Stepp, H. Diagnostic cancer in vivo. *Science* **292**, 1401-1403 (2001).

Kopriva, I., 2007. Approach to Blind Image Deconvolution by Multiscale Subband Decomposition and Independent Component Analysis. J. Opt. Soc. Am. A 24, 973-983.

Kopriva, I., Seršić, D., 2008. Wavelet packets approach to blind separation of statistically dependent sources. Neurocomputing **71**, 1642-1655.

Lee, D.D., Seung, S.H., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788-791.

Malinowski, E. R., 1977a. Theory of error in factor analysis. Anal. Chem. 49, 606-612.

Malinowski, E. R., 1977b. Determination of the number of factors and experimental error in a data matrix. Anal. Chem. 49, 612-617.

Mendel, J. M., 1991. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. Proc. IEEE 79, 278-305.

McCullagh, P. 1995. Elementary theory of cumulants in Tensor Methods in Statistics, Chapman & Hall, London, pp. 24-46.

Molgedey, L., Schuster, H.G., 1994. Separation of mixture of independent signals using time delayed correlations. Phys. Rev. Lett. 72, 3634-3636.

Orfanidis, S. J., 1988. Optimum Signal Processing – An Introduction, 2nd ed. MacMillan Publishing Comp., New York.

Pham, D.T., 1997. Blind separation of mixtures of independent sources through a quasimaximum likelihood approach. IEEE Trans. Sig. Proc. 45, 1712-1725.

Pham, D.T., 2003. Fast algorithm for estimating mutual information, entropies and score functions. In: Proceedings of the Fourth International Conference on

Independent Component Analysis and Blind Signal Separation (ICA'2003), Nara, Japan, pp.17-22.

Principe, J.C., Xu, D., 1999. Information theoretic learning using Reny's quadratic entropy. In: Proceedings of the first international conference on independent component analysis and blind signal separation (ICA'99), Aussois, France, pp. 407-412.

Principe, J.C., Xu, D., Fisher, J.W., 2000. Information-Theoretic Learning in Unsupervised Adaptive Filtering-Volume I Blind Source Separation, J. Wiley, pp. 265-319.

Ren, H., Chang, Ch-I., 2000. A generalized orthogonal subspace projection approach to unsupervised multispectral image classification. IEEE Trans. on Geosc. and Remote Sensing 38, 2515-2528.

Schwartz, S., Zibulevsky, M., Schechner, Y.Y., 2004. ICA using kernel entropy estimation with NlogN complexity. Lecture Notes in Computer Science 3195, 422-429.

Scott, M.A., Hopper C., Sahota A., Springett R., Mcllroy B.W., Bown S.G., MacRobert A.J., 2000. Fluorescence photo diagnostics and photo bleaching studies of cancerous lesions using ratio imaging and spectroscopic techniques. Lasers Med Sci 15, 63-72.

Settle, J.J., Drake, N.A., 1993. Linear mixing and estimation of ground cover proportions. Int. J. Remote Sensing 14, 1159-1177.

Silverman, B.W., 1982. Kernel density estimation using the fast Fourier transform. Appl. Stat. 31, 93-99.

Stone, J.V., 2001. Blind source separation using temporal predictability. Neural Comput. 13, 1559-1574.

Tréhin, R., Figueiredo, J. L., Pittet, M.J., Weisslender, R., Josephson, L., Mahmood, U., 2006. Fluorescent nanoparticle uptake for brain tumour visualization. Neoplasia **4**, 302-311.

Wang, J., Chang, C. I., 2006. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. IEEE Trans. on Geosc. and Remote Sensing 44, 1586-1600.

Xue, Y., Wang, Y., Yang, J., 2008. Independent component analysis based on gradient equation and kernel density estimation. Neurocomputing doi:10.1016/j.neucom.2008.08.014.

Zdunek, R., Cichocki, A., 2007. Nonnegative matrix factorization with constrained second-order optimization. Signal Proc. 87, 1904-1916.

Ziehe, A., Muller, K.R., Nolte, G., Mackert, B.M., Curio, G., 1998. TDSEP-an efficient algorithm for blind separation using time structure. In: Proceedings of the international conference on artificial neural networks (ICANN'98), Skovde, Sweden, pp. 675-680.