Blind Multi-spectral Image Decomposition by

3D Nonnegative Tensor Factorization

Ivica Kopriva[1,*] and Andrzej Cichocki[2]

[1]Division of Laser and Atomic Research and Development

Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

[2]Laboratory for Advanced Brain Signal Processing

Brain Science Institute, RIKEN 2-1Hirosawa, Wako-shi, Saitama, 351-0198, Japan

Abstract

α-divergence based nonnegative tensor factorization (NTF) is applied to blind multi-spectral image (MSI) decomposition. Matrix of spectral profiles and matrix of spatial distributions of the materials resident in the image are identified from the factors in Tucker3 and PARAFAC models. NTF preserves local structure in the MSI that is lost, due to vectorization of the image, with nonnegative matrix factorization (NMF)- or independent component analysis (ICA)-based decompositions. Moreover, NTF based on PARAFAC model is unique up to permutation and scale under mild conditions. To achieve this, NMF- and ICA-based factorizations respectively require enforcement of sparseness (orthogonality) and statistical independence constraints on the spatial distributions of the materials resident in the MSI, and that is not true. We demonstrate

efficiency of the NTF-based factorization in relation to NMF- and ICA-based factorizations on blind decomposition of the experimental MSI with the known ground truth.

OCIS codes: 100.6890; 100.3190; 150.6910; 100.2960; 170.3880.

Blind or unsupervised multi-spectral and hyper-spectral image (HSI) decomposition attracts increased attention due to its capability to discriminate materials resident in the MSI/HSI without knowing their spectral profiles [1,2]. However, most of blind decomposition schemes rely on two-dimensional (2D) representation of the MSI/HSI although it is inherently three-dimensional (3D). In this letter we represent a MSI/HSI as a three-way array or a 3D tensor $\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$ with elements $x_{i_1 i_2 i_3}$ where i_1=1,..., I_1, i_2=1,..., I_2, i_3=1, ..., I_3 and $\mathbb{R}_{0+}$ is a real manifold with nonnegative elements. Each index is called way or mode and number of levels on one mode is called dimension of that mode. MSI/HSI is a set of I_3 spectral band images with the size of $I_1 \times I_2$ pixels. Two ways of $\underline{\mathbf{X}}$ are for rows and columns and one way is for spectral band. This is standard notation that is adopted for use in multiway analysis, [3]. 2D representation of MSI has two disadvantages: (*i*) 3D tensor $\underline{\mathbf{X}}$ has to be mapped through 3-mode flattening, also called unfolding and matricization, to matrix $\mathbf{X}_{(3)} \in \mathbb{R}_{0+}^{I_3 \times I_1 I_2}$ whereas local structure of the image is lost; (*ii*) matrix factorization $\mathbf{X}_{(3)} = \mathbf{AS}$ employed by linear mixing models [1,2] suffers from indeterminacies because $\mathbf{ATT}^{-1}\mathbf{S} = \mathbf{X}_{(3)}$ for any invertible $\mathbf{T}$, i.e. infinitely many ($\mathbf{A}$, $\mathbf{S}$) pairs can give rise to $\mathbf{X}_{(3)}$. Meaningful solution of the factorization of $\mathbf{X}_{(3)}$ is characterized with $\mathbf{TT}^{-1} = \mathbf{P\Lambda}$ where $\mathbf{P}$ is permutation matrix and $\mathbf{\Lambda}$ is diagonal matrix. These permutation and scaling indeterminacies are standard for blind decompositions and are obtained by imposing sparseness

(orthogonality) constraints on $\mathbf{S}$ by NMF algorithms [4] and statistical independence constraints by ICA algorithms [1,5,6]. Orthogonality constraints imply that materials resident in the image do not occupy the same pixel footprint that is not correct assumption especially in airborne and spaceborne remote sensing. Statistical independence assumption is also not correct for MSI and HSI data especially when materials are spectrally similar what occurs in the case of the low-dimensional MSI with coarse spectral resolution [7]. Only very recently tensor factorization methods were employed to MSI/HSI analysis for the purpose of dimensionality reduction, de-noising, target detection and material identification [8-10]. For the purpose of MSI decomposition we adopt two widely used 3D tensor models: Tucker3 model [11] and PARAFAC/CANDECOMP model [12,13]. The Tucker3 model is defined as

$$\underline{\mathbf{X}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \mathbf{A}^{(3)} \tag{1}$$

where $\underline{\mathbf{G}} \in \mathbb{R}_{0+}^{J_1 \times J_2 \times J_3}$ is core tensor and $\left\{ \mathbf{A}^{(n)} \in \mathbb{R}_{0+}^{I_n \times J_n} \right\}_{n=1}^{3}$ are factors and $\times_n$ denotes n-mode product of a tensor with a matrix $\mathbf{A}^{(n)}$. The result of $\underline{\mathbf{G}} \times_n \mathbf{A}^{(n)}$ is a tensor of the same order as $\underline{\mathbf{G}}$ but the size J_n replaced by I_n. PARAFAC model is a special case of Tucker3 model when $\underline{\mathbf{G}}$ is superdiagonal tensor with all elements zero except those for which all indices are the same. Compared to PARAFAC, Tucker3 model is more flexible due to the core tensor $\underline{\mathbf{G}}$ which allows interaction between a factor with any factor in the other modes [14]. In PARAFAC model factors in different mode can only interact factorwise. However, this restriction enables uniqueness of tensor factorization based the PARAFAC model within the permutation and scaling indeterminacies of the factors under very mild conditions [15,16] without need to impose any

special constraints on them such as sparseness or statistical independence. Assuming that $J_1=J_2=J_3=J$ and $J \leq I_3$ uniqueness condition is reduced to $k_{\mathbf{A}^{(1)}} + k_{\mathbf{A}^{(2)}} + k_{\mathbf{A}^{(3)}} \geq 2J + 3$, where $k_{\mathbf{A}^{(n)}}$ is Kruskal rank of factor $\mathbf{A}^{(n)}$ [15,16]. Due to interaction between the factors there is no such theoretical guarantee on the uniqueness of tensor factorization based on Tucker3 model. However despite of this, Tucker3 model has been used successfully in HSI analysis for dimensionality reduction, de-noising and target detection [8,9]. To identify spatial distributions of the materials resident in the MSI/HSI we refer to standard linear mixture model used in MSI/HSI data analysis [1,2,10]:

$$\mathbf{X}_{(3)} \approx \mathbf{A}\mathbf{S} \tag{2}$$

where columns of $\mathbf{A} \in \mathbb{R}_{0+}^{I_3 \times J}$ represent spectral profiles of the J materials resident in the image while rows of $\mathbf{S} \in \mathbb{R}_{0+}^{J \times I_1 I_2}$ represent spatial distributions of the same materials. As already said, without additional constraints there are infinitely many decompositions satisfying model (2). From Tucker3 model (1) and linear mixture model (2) matrix of spectral profiles and tensor of spatial distributions of the materials are identified as

$$\mathbf{A} \approx \mathbf{A}^{(3)}$$

$$\underline{\mathbf{S}} \approx \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} = \underline{\mathbf{X}} \times_3 \left(\mathbf{A}^{(3)}\right)^{-1} \tag{3}$$

where $\underline{\mathbf{S}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times J}$. Second approximation for $\underline{\mathbf{S}}$ in (3) is less sensitive to numerical errors than first one due to the fact that only one reconstructed quantity, array factor $\mathbf{A}^{(3)}$, takes places into reconstruction of $\underline{\mathbf{S}}$. We can also express 3-mode flattened version of tensor $\underline{\mathbf{X}}$, this is matrix $\mathbf{X}_{(3)}$, in terms of 3-mode flattened core tensor $\underline{\mathbf{G}}$, this yields matrix $\mathbf{G}_{(3)} \in \mathbb{R}_{0+}^{J \times JJ}$, and array factors $\left\{\mathbf{A}^{(n)}\right\}_{n=1}^{3}$ as [17]:

4

$$\mathbf{X}_{(3)} \approx \mathbf{A}^{(3)} \mathbf{G}_{(3)} \left[\mathbf{A}^{(2)} \otimes \mathbf{A}^{(1)} \right]^{\mathrm{T}} \qquad (4)$$

where $\otimes$ denotes Kronecker's product. In direct comparison between (2) and (4) we arrive at:

$$\mathbf{A} \approx \mathbf{A}^{(3)}$$

$$\mathbf{S} \approx \mathbf{G}_{(3)} \left[\mathbf{A}^{(2)} \otimes \mathbf{A}^{(1)} \right]^{\mathrm{T}} = \left(\mathbf{A}^{(3)} \right)^{-1} \mathbf{X}_{(3)} \qquad (5)$$

Again, numerically more accurate approximation of $\mathbf{S}$ is obtained from second part of (5). Various cost functions can be used as discrepancy measure between tensor and its model. In this letter we employed α-divergence because it is adaptable to noise statistics [17] and because it has been demonstrated in [18] that it outperforms NTF based on least square error function [19]. We refer to appendix and [17] for α-divergence based update NTF algorithms, eq.(13) to (15). We have compared performance of α-NTF, second order NMF [20] and dependent component analysis (DCA) algorithms on real world example: blind decomposition of the experimental weak-intensity fluorescent red-green-blue (RGB) image of the skin tumor shown in Figure 1b. For the purpose of tumor demarcation it is of interest to extract spatial map of the tumor as accurately as possible. DCA algorithm combines JADE ICA algorithm [6] and innovation transform based preprocessing [21] to enhance statistical independence among materials present in the MSI and improve accuracy of the ICA. The MSI contains three spectral bands and three materials: tumor, surrounding tissue and the ruler added to the scene to give perspective about the size of the tumor. High-intensity image of the same tumor is shown in Figure 1a. It was used for estimation of the binary spatial maps of tumor and surrounding tissue necessary for the estimation of the receiver-operating-characteristic (ROC) curves. Spatial maps of the tumor extracted from Figure 1b are shown in Figures 1c to 1e. Implementation of Tucker 3 α-NTF algorithm was based on MATLAB Tensor Toolbox provided at [22]. Tensor of spatial

5

distributions of materials was identified by means of second part of (3). SO NMF and DCA algorithms were based on 2D MSI representation (2). According to ROC curves shown in Figure 2 α-NTF algorithm exhibited best performance i.e. yielded largest area under the ROC curve. State-of-the-art intensity based image segmentation algorithm, that is the level set method [23], has been also applied to the gray scale version of Figure 1b. The result in Figure 1f shows evolution curve after 1000 iterations. Due to weak boundaries the method failed to converge. α-NTF, SO NMF and DCA algorithms were implemented in MATLAB on a 2.4 GHz Intel Core 2 Quad Processor Q6600 based desktop computer with 4GB RAM. Computation times are given respectively as: 4783s, 30s and 3.6s. In the implementation of the innovation-based DCA algorithm a 10th order linear prediction filter has been used.

Appendix - elements of Tucker3 α-NTF algorithm

Multiplicative update rules for core tensor $\underline{\mathbf{G}}$ and factors $\left\{\mathbf{A}^{(n)}\right\}_{n=1}^{3}$ in (1) are obtained by minimizing α-divergence as:

$$\underline{\mathbf{G}} \leftarrow \underline{\mathbf{G}} \odot \left\{ \frac{\left(\underline{\mathbf{X}}/\underline{\hat{\mathbf{X}}}\right)^{.\alpha} \times_1 \mathbf{A}^{(1)^{\mathrm{T}}} \times_2 \mathbf{A}^{(2)^{\mathrm{T}}} \times_3 \mathbf{A}^{(3)^{\mathrm{T}}}}{\underline{\mathbf{E}} \times_1 \mathbf{A}^{(1)^{\mathrm{T}}} \times_2 \mathbf{A}^{(2)^{\mathrm{T}}} \times_3 \mathbf{A}^{(3)^{\mathrm{T}}}} \right\}^{.\frac{1}{\alpha}} \tag{A1}$$

$$\mathbf{A}^{(n)} \leftarrow \mathbf{A}^{(n)} \odot \left\{ \frac{\left[\left(\underline{\mathbf{X}}/\underline{\hat{\mathbf{X}}}\right)^{.\alpha}\right]_{(n)} \mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}}{\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}} \right\}^{.\frac{1}{\alpha}} \tag{A2}$$

where $\odot$ denotes element-wise multiplication and / denotes element-wise division. In (A2) **1** denotes a vector whose every element is one. Numerator in (A2) is calculated as

$$\left[\left(\underline{\mathbf{X}}/\hat{\underline{\mathbf{X}}}\right)^{.\alpha}\right]_{(n)}\mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}=\left[\left(\underline{\mathbf{X}}/\hat{\underline{\mathbf{X}}}\right)^{.\alpha}\times_{m\neq n}\mathbf{A}^{(m)^{\mathrm{T}}}\right]_{n}\mathbf{G}_{(n)}^{\mathrm{T}}$$

where $\mathbf{G}_{(n)}$ represents n-mode flattened version of the core tensor $\underline{\mathbf{G}}$. Denominator in (A2) is computed as

$$\mathbf{1}^{\mathrm{T}}\mathbf{G}_{\mathbf{A}}^{(n)^{\mathrm{T}}}=\left[\underline{\mathbf{G}}\times_{m\neq n}\mathbf{1}^{\mathrm{T}}\mathbf{A}^{(m)}\right]_{(n)}^{\mathrm{T}}$$

where $\underline{\mathbf{G}}\times_{m\neq n}\mathbf{1}^{\mathrm{T}}\mathbf{A}^{(m)}$ denotes m-mode products between core tensor $\underline{\mathbf{G}}$ and matrices $\mathbf{1}^{\mathrm{T}}\mathbf{A}^{(m)}$ for all $m=1, ..., N$ and $m \neq n$.

References:

1. C.-I. Chang, S. -S. Chiang, J. A. Smith, and I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," IEEE Trans. Geosci. Remote Sens. **40**, 375 (2002).

2. T. Tu," Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," Opt. Eng. **39**, 897 (2000).

3. H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," J. Chemometrics **14**, 105 (2000).

4. A. Cichocki, R. Zdunek, and S. Amari," Nonnegative Matrix and Tensor Factorization," IEEE Signal Proc. Mag. **25**, 142 (2008).

5. A. Cichocki, and S. Amari, *Adaptive Blind Signal and Image Processing* (Wiley, 2002).

6. J. F. Cardoso, and A. Soulomniac, " Blind beamforming for non-Gaussian signals," Proc. IEE F **140**, 362 (1993).

7. J. M. P. Nascimento, and J. M. Dias Bioucas, " Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?," IEEE Trans. Geosci. Remote Sens. **43**, 175 (2005).

8. N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and Dimensionality Reduction Using Multilinear Tools for Hyperspectral Images," IEEE Geosci. Remote Sens. Lett. **5**, 138 (2008).

9. N. Renard, and S. Bourennane, "Improvement of Target Detection Methods by Multiway Filtering," IEEE Trans. Geosci. Remote Sens. **46**, 2407 (2008).

10. Q. Zhang, H. Wang, R. J. Plemons, and V. P. Pauca, "Tensor methods for hyperspectral data analysis: a space object material identification study," J. Opt. Soc. Am. A **25**, 3001 (2008).

11. L. R. Tucker, "Some mathematical notes on three-mode factor analysis," Psychometrika **31**, 279 (1966).

12. J. D. Carrol, and J. J. Chang, "Analysis of individual differences in multidimensional scaling via N-way generalization of Eckart-Young decomposition," Psychometrika **35**, 283 (1970).

13. R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an exploratory multi-mode factor analysis," UCLA Working Papers in Phonetics **16**, 1 (1970).

14. E. Acar, and B. Yener, "Unsupervised Multiway Data Analysis: A Literature Survey," IEEE Trans. Knowl. Data Eng. **21**, 6 (2009).

15. J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions," Linear Algebra Appl. **18**, 95 (1977).

16. N. D. Sidiropoulos, and R. Bro, "On the uniqueness of multilinear decomposition of *N*-way arrays," J. of Chemometrics **14**, 229 (2000).

17. Y. D. Kim, A. Cichocki, and S. Choi, "Nonnegative Tucker Decomposition with Alpha-Divergence," presented at the 2008 IEEE ICASSP International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, U.S.A., March 30-April 4, 2008, pp. 1829-1832.

18. A. H. Phan, and A. Cichocki, "Fast and Efficient Algorithms for Nonnegative Tucker Decomposition," Lect. Notes Comput. Sci. **5264**, 772 (2008).

19. C. A. Anderson, and R. Bro, "The *N*-way Toolbox for MATLAB," Chemometrics Intell. Lab. Syst. **52**, 1 (2000).

20. R. Zdunek, and A. Cichocki, "Nonnegative matrix factorization with constrained second-order optimization," Sig. Proc. **87**, 1904 (2007).

21. A. Hyvärinen, "Independent component analysis for time-dependent stochastic processes," in Proc. Int. Conf. Art. Neural Net. (ICANN'98), Skovde, Sweden, 1998, pp. 541-546.

22. B. W. Bader, and T. G. Kolda, *MATLAB Tensor Toolbox version 2.2.* http://csmr.ca.sandia.gov/~tkolda/TensorToolbox.

23. Ch. Li, Ch. Xu, Ch. Gui, and M.D. Fox, "Level Set Evolution Without Re-initilaization: A New Variational Formulation," in: IEEE International Conference on Computer Vision and Pattern Recognition, 2005, 430-436.

Figure Captions:

Figure 1 (color online). Experimental fluorescent MSI (RGB) image of skin tumor: a) high-intensity version; b) weak-intensity version. Spatial maps of the tumor extracted from Figure 1b by means of: c) α-NTF algorithm [18] with α=0.1; d) SO NMF algorithm [5]; e) DCA algorithm [7 and 21]. f) evolution curve calculated by level set method on gray scale version of Figure 1b after 1000 iterations. Dark red color indicates that tumor is present with probability 1, while dark blue color indicates that tumor is present with probability 0.

Figure 2 (color online). ROC curves calculated for spatial maps of the tumor shown in Figures 1c to 1e: red squares - α-NTF algorithm based on Tucker3 model with α=0.1; blue stars - DCA algorithm [12]; green triangles - SO NMF algorithm [43].
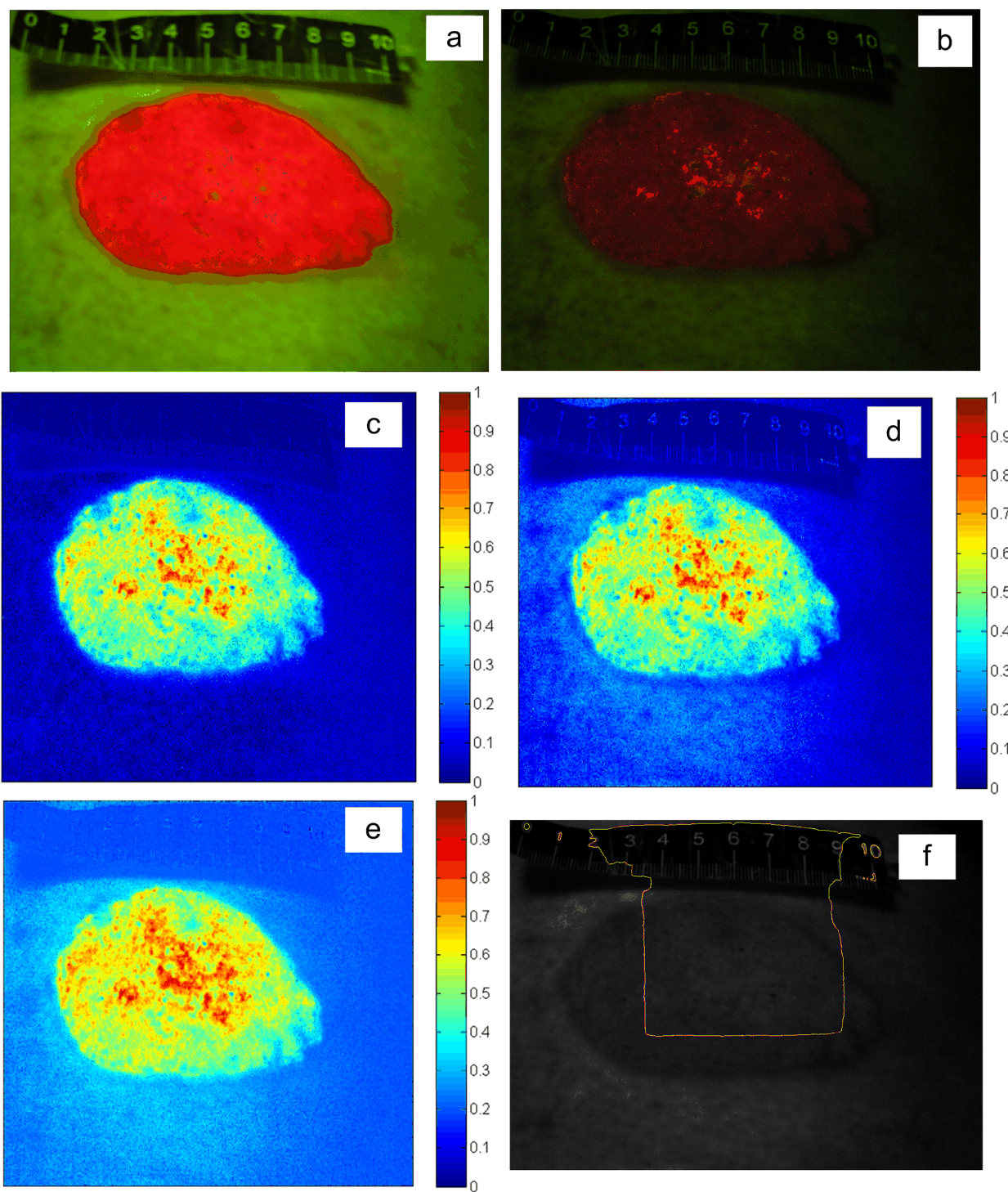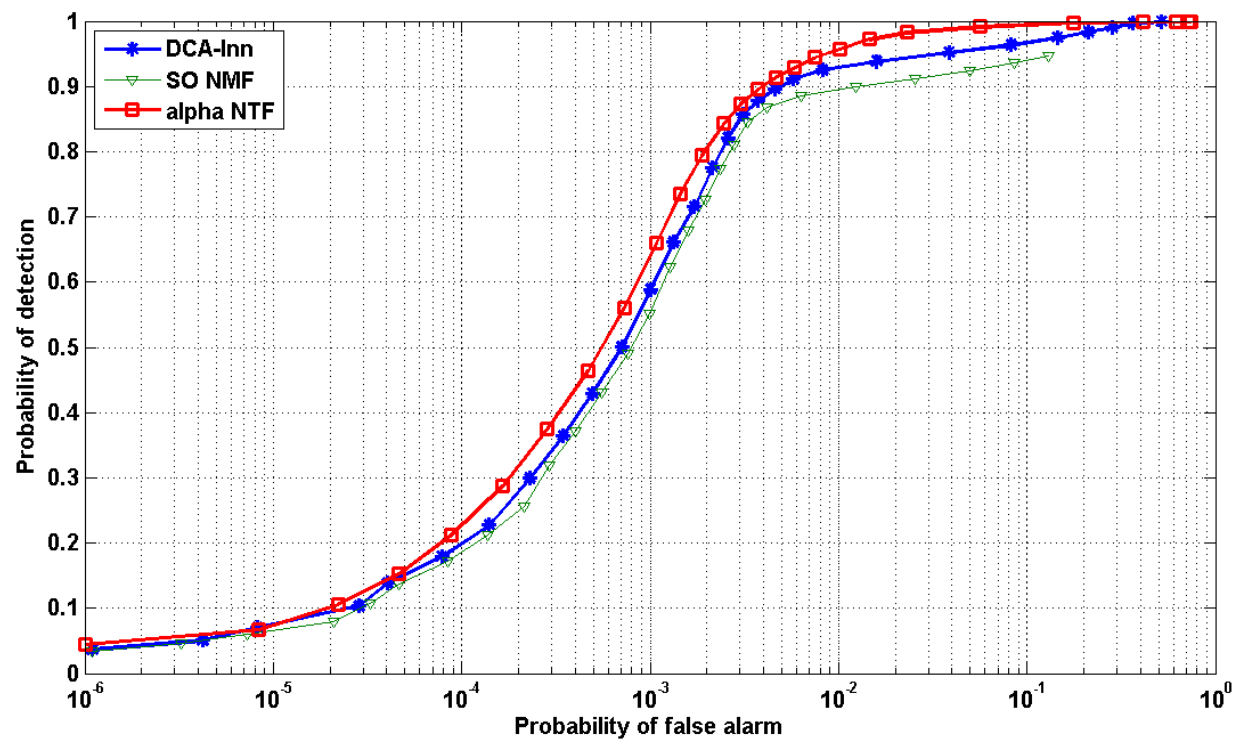
Fig. 1, OL, Kopriva & Cichocki

Fig. 2, OL, Kopriva & Cichoki