Target/Background Classification Regularized Nonnegative Matrix Factorization for Fluorescence Unmixing

Binjie Qin, Member, IEEE, Chen Hu, and Shaosen Huang

Abstract—Nonnegative matrix factorization (NMF) is usually applied to multispectral fluorescence imaging for fluorescence unmixing. Unfortunately, most NMF-based fluorescence unmixing methods fail to take advantage of spatial information in data. Besides, NMF is an inherently ill-posed problem, which gets worse in the sparse acquisition of multispectral data (from a small number of spectral bands) due to its insufficient measurements and severe discontinuities in spectral emissions. To overcome these limitations by exploiting the spatial difference between multiple-target fluorophores and background autofluorescence (AF), we propose improved normalized cut to automatically classify all multispectral pixels into target fluorophores and background AF groups. We then initialize NMF by extracting the endmember spectra of target/background fluorescent components in the two groups, and impose a $L_{1/2}$-norm partial sparseness constraint on merely the abundances of target fluorophores within hierarchical alternating least squares framework of NMF. Experimental results based on synthetic and in vivo fluorescence data show the superiority of the proposed algorithm with respect to other state-of-the-art approaches.

Index Terms—Fluorescence spectra, insufficient measurements, multispectral imaging, nonnegative matrix factorization (NMF), partial sparseness constraint, signal decomposition, spatial information, target/background classification.

I. Introduction

IN VIVO multispectral fluorescence imaging instrument has been widely used to measure and/or record cellular and subcellular biological processes in the life and medical sciences, such as drug discovery and disease diagnosis [1]. The vast majority of applications of in vivo fluorescence imaging are based on epi-illumination planar imaging, where the exciting source and detectors reside on the same side of the tissue and the measurements are acquired in reflectance mode. Given exciting light sources, different fluorophores labeled with fluorescent dyes can emit fluorescence photons from visible to near-infrared wavelengths to generate multispectral images. The multispectral images involve multispectral pixels represented by vectors, with each component being a measurement corresponding to the specific wavelengths. This fluorescence imaging instrument enables the simultaneous use of multiple fluorophores to detect and localize particular components of complex biomolecular assemblies in the in vivo sample. For most fluorophores, emission spectra are distinct, but often overlap and become indistinguishable in the mixed multispectral images. Hence, spectral unmixing (SUM) [2] is necessary in the multispectral fluorescence imaging instrument to decompose the mixed multispectral images D into a product of pure spectral signatures S, i.e., endmembers, and corresponding fractional abundance C, indicating the proportion of each endmember. If the endmember spectra S are identified [3] in advance, C can be easily estimated by the use of supervised SUM methods such as least squares method. However, the factory-provided reference endmember spectra used in the supervised SUM are uncertain and always require extensive calibration efforts for the endmember identification [3]. Therefore, the unsupervised SUM has been developed to simultaneously estimate the spectra and abundances without a priori knowledge about endmember spectra.

In designing, implementing, and assessing the fluorescence imaging instrument, there are some practical challenges that must be overcome, among which the so-called autofluorescence (AF) [1], [4] can be produced by some proteins such as collagens and other biological materials when they are excited by appropriate visible light in in vivo fluorescence imaging. Generally, AF originates from all possible background disturbances, which mainly include two kinds of sources [1], [4]: 1) the AF caused by the natural fluorescent molecules in tissue and food and 2) some instrument-based noise, shading, and leakage light from exciting filters. Therefore, AF stems from various sources covering large background areas, and has a dispersive spatial distribution. Furthermore, the AF wavelength ranging from 400 to 700 nm is overlapped with the emission spectra of most fluorophores. Due to these extensive overlaps occurring between the fluorophores and AF in the spatial and spectral distributions, it is difficult to blindly separate multiple fluorophores from AF when the AF is regarded as a constituent component by the current

Manuscript received July 8, 2015; revised November 13, 2015; accepted November 16, 2015. Date of publication January 25, 2016; date of current version March 8, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61271320 and Grant 60872102 and in part by the China Scholarship Council through the Small Animal Imaging Project under Grant 06-545. The Associate Editor coordinating the review process was Dr. Shervin Shirmohammadi.

The authors are with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: bjqin@sjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2016.2516318

unsupervised SUM methods. Alternatively, some hardware based methods subtract an AF estimate from observation data after using extra excitation Þlters or extra unlabeled samples to acquire bare AF images [1], [4]. To successfully practice these methods, we must carefully match the speciÞc Þlter sets with the spectral properties of both AF and ßuorophores. In many cases, none of the mentioned methods in the instrument design can fully remove the AF from the ßuorescence imaging.

As an unsupervised data decomposition (or blind source separation) technique, nonnegative matrix factorization (NMF) has been successfully applied to blindly separate several source signals in SUM [4], [5], biomedical source separation [6], [7], and nondestructive testing [8]. However, there are three shortcomings. First, NMF suffers from an ill convergence problem such that starting from different initial search points results in different values for the elements of C and S matrices. Especially, the ill convergence problem becomes worse when there are insufÞcient measurements and low spectral resolutions in the sparse acquisition of multispectral ßuorescence imaging data from a small number (e.g., 3Ð10) of spectral bands. However, this sparse acquisition can be fast and cost-effective in clinical applications. Therefore, different NMF studies have proposed appropriate initial values [9] and some additional regularization constraints (such as sparseness [10], [11] and smoothness constraints [9]) to ensure the optimal NMF solution. Second, current sparseness constraints [10], [11] for strengthening part-based representation do not discriminate sparse components from nonsparse components, limiting unmixing accuracy when only some special targets of interest are sparse while a speciÞc background component is nonsparse. In ßuorescence imaging, some sparseness-constrained methods [4], [12], [13] update the whole abundance matrix of all components so that the abundance matrices of ßuorescent targets and AF may interfere with each other in NMF. Third, NMF does not consider spatial information of neighboring pixels from speciÞc components to Þnd more intuitive and interpretable unmixing solution of abundance matrix C.

To overcome these limitations of NMF, this paper proposes unsupervised target/background classiÞcation regularized NMF (TBCR-NMF)[1] with partial sparseness constraint. The motivation is based on the following two facts. First, multiple ßuorophores tend to locally accumulate in speciÞc biological tissues so that their sparse spatial distributions are usually conÞned to relatively small areas, while background AF propagates at all directions and diffuses widely over large areas. This spatial distribution difference between multiple target ßuorophores and background AF is preserved across the whole spectral bands, although the sparse acquisition introduces sharp discontinuity in the spectral emissions across the multiple spectral bands. Second, the set of pixels in the multiple localized ßuorophores similarly exhibits high intensities within local patches and can be classiÞed into a single target group, while the set of pixels in the large background areas contains low intensity pixels that can be grouped as a single background group. This inherent target/background contrast is still preserved across the multiple spectral bands in the ßuorescence imaging.

With the above analysis facilitating the development of unsupervised target/background classiÞcation without any training samples [2], [14], we propose TBCR-NMF from the following two aspects. First, an unsupervised target/background classiÞcation is implemented as a preprocessing that extracts endmembers and corresponding abundances to optimally initialize NMF. In solving the local minimum problem of NMF, most initialization methods [15], [16] have not used the spatial information in the source spectral data and therefore cannot accurately identify the endmembers (and their corresponding abundance) for the near-optimal starting point for NMF. In this paper, initial target/background classiÞcation is helpful to discriminably extract the endmembers from the localized target regions and large background regions. Then, we initialize the abundance matrix C by Þxing the spectra S during the Þrst ten iterations of NMF. Second, the TBCR-NMF facilitates optimal ßuorescence unmixing by imposing partial sparseness constraints on the abundances of multiple target ßuorophores but not on the abundance of diffusive AF. In summary, classifying mixed multispectral data into two groups is a useful strategy for initializing and regularizing NMF, such that the target/background classiÞcation can transfer the classiÞed spatial structures [17] into the accurate and unique solution of NMF-based unmixed results. The most recent trend of utilizing the spatial information and sparsity for unmixing/classiÞcation of multispectral image has enabled the realization of some new computing model in multispectral/hyperspectral imaging [18]Ð[21].

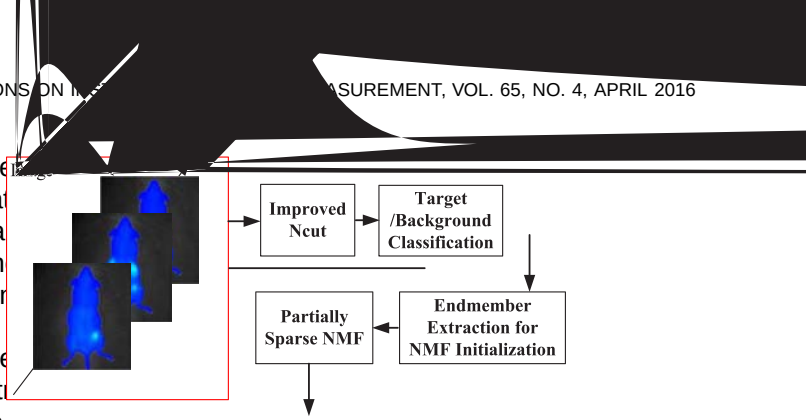
Traditional multispectral image classiÞcation [22] methods, such as unsupervised (e.g., K-means, kernel-based nonparametric method) and supervised (e.g., maximum likelihood, support vector machine), have considered the pixel-wise spectral dissimilarity between two pixels to group the image data into a Þnite number of discrete classes without using spatial dependence. To reduce the labeling uncertainty that exists when only the spectral information is used, recent research has introduced the spatial contextual information into the joint spectralÐspatial classiÞcation, which generally exploited the highly correlated regional information (entropy, variance, etc.) extracted from the standard (such as the crisp neighbor set employed by Markov random Þeld modeling) or adaptive neighbor system in the image. Rather than deÞning a crisp neighbor set containing insufÞcient neighboring samples for every pixel, image segmentation [22] is another approach to include spatial information in classiÞcation, enabling the large neighborhood deÞnition by partitioning an image into nonoverlapping large homogeneous regions. Many algorithms have been proposed to address image segmentation problem, such as region-growing algorithms, and watershed methods. In this paper, only binary image segmentation implementing target/background homogeneous region partition is desirable for the subsequent target/background classiÞcation. As an excellent binary image segmentation algorithm, the original normalized cut (Ncut) [23] is done by partitioning all graph nodes (i.e., pixels) of whole image into

[1]http://www.escience.cn/people/bjqin/research.html

two disjoint parts. Rather than focusing on local features and their neighboring consistencies in the image dat Ncut aims at extracting the global impression of a image. It is assumed to be capable of utilizing the distin global dissimilarities between target ßuorophores ar background AF in the whole ßuorescence images implement target/background segmentation. However, the are indeed more than two classes in the multispect ßuorescence images in the spe nce of multiple ßuorophore targets. It is possible that the spectral emissions of some ßuorescence targets are more similar to the AF spectral emission than other ßuorescence targets. In this case, Ncut-based method will lead to a wrong target/background classiÞcation by grouping some ßuorescence targets into the background AF group. To ensure accurate target/background classiÞcation, we modify classical Ncut method [23] to recursively repartition the large group of previous bipartition result if the number of groups segmented by Ncut is less than the number of endmembers in the ßuorescence imaging. As a result, all the pixels of the multispectral images are classiÞed into several groups, which further are simply merged into two main groups: the largest group is the background AF group and the rest of the smaller groups will join together into the target ßuorescence group. Based on the improved Ncut-based classiÞcation, this paper has the following two contributions for the ßuorescence SUM. First, by performing improved Ncut-based bipartitioning of target ßuorescences and background AF groups, we propose target/background classiÞcation to beneÞt the endmember identiÞcation from the target and background groups for accurately initializing NMF. Second, this target/background classiÞcation facilitates imposing $L_{1/2}$-norm [24], [25] partial sparseness constraint on the abundances of the target ßuorescent group but not on that of AF group in the NMF, which is based on hierarchical alternating least squares (HALS) framework. The remainder of this paper is organized as follows. Section II describes the idea and the details of the proposed ICR-NMF algorithm. Section III provides experimental results on synthetic andin vivo ßuorescence imaging data. The conclusion and discussion are given in Section IV.

II. Materials and Methods

The ßuorescence image data acquired with multispectral imaging instruments comprise four contiguous bands in this paper. A multispectral data set is usually stacked as an image cube and thus can be treated as a 3-D volumetric data set with two spatial axes (X and Y) and one spectral axis (λ), as illustrated in Fig. 1. From a data-ßow point of view, the ßowchart of the proposed algorithm can be characterized as the following (see Fig. 1). First, a multispectral image cube is iteratively segmented into multiple separated homogeneous regions using the improved Ncut algorithm. Second, we further group these regions into target and background groups by classifying the largest region into the background group while merging all other small regions into the target group. Third, endmember extraction methods are employed to extract the spectral signatures of the target ßuorescences and AF from the target and background regions for the NMF initialization.



where $\mathrm{assoc}(A, V) = \sum_{u\in A, t\in V} w(u, t)$ is the total connections from nodes in A to all nodes in the graph and $\mathrm{assoc}(B, V)$ is similarly defined. The Ncut grouping algorithm consists of the following steps. First, let H be an $N \times N$ diagonal matrix with h on its diagonal and N be the number of the nodes and $h(i) = \sum_j w(i, j)$, where the weight $w(i, j)$ is defined as $w(i, j) = e^{-(\|x_i - x_j\|^2/\sigma)}$, with x_i and x_j representing the normalized spectra of nodes i and j, which have zero-mean and unit variance, respectively. In addition, $\sigma = 0.1$ is a positive scaling factor determining the sensitivity of $w(i, j)$ to the spectrum difference between nodes i and j; Then, we solve $(H - W)v = \eta Hv$ for eigenvectors v with the smallest eigenvalues η, where W is $N \times N$ symmetric weight matrix with the element being $w(i, j)$. At last, use the second smallest eigenvector v_1 and the splitting value 0.4 to bipartition graph, i.e., the bipartition is implemented by grouping the ith node into A if the ith component of eigenvector v_1 is larger than 0.4, B otherwise.

For the multiple ßuorophores in in vivo ßuorescence imaging, there are more than two classes in the ßuorescence imaging. Therefore, the spectral emissions of some ßuorescence targets are more similar to those of AF than other ßuorescence targets, such that Ncut will lead to a wrong target/background classiÞcation by aggregating some ßuorescence targets into the background AF group. To avoid this misclassiÞcation, we need to modify the recursive two-way Ncut method. Considering that background AF (including various background noises) has a dispersive spatial distribution while multiple ßuorophores are locally accumulated at speciÞc locations, we assume that the pure background AF regions are larger than the target ßuorophore regions. Through Ncut-based bipartition, the larger group is either pure AF region or the regions that contain AF and some target ßuorophores. In the latter case, the larger group will be bipartitioned again until all target ßuorophores are separated from the AF region. Because the aim of the improved recursive Ncut method is to classify all pixels of the whole ßuorescence region into two classes, background AF and target ßuorescence groups, all the separated smaller regions except the largest AF region are Þnally combined together into the target ßuorescence group.

Based on the above analysis, we propose an improved recursive Ncut method. First, to use the Ncut method, each 992×992 spectral image is decimated into a size of 100×100 pixels. Decreasing the number of graph nodes from near $N = 1\,000\,000$ to $N = 10\,000$ by this image subsampling can solve the large graph problem, which consumes too much memory and requires huge computational complexity in handling large-scale weight matrix W (with $N \times N$ elements) for the graphical representation and generalized eigenvalue computation. In our experiments, changing image size from 200×200 pixels to 100×100 pixels can obviously decrease Ncut computation time from 500 to 28 s, but does not have an adverse effect on target/background classiÞcation, because the Ncut method's graph-based generalized eigenvalue computation is less sensitive to the spatial information lost during subsampling than other local feature-based segmentation. Besides, even if small noisy misclassiÞcation occurs, it cannot affect the Þnal TBCR-NMF's performance, because we only require an approximate global target/background classiÞcation for further decomposition reÞnement by TBCR-NMF itself. Second, after the initial Ncut-based bipartition, only the large cluster of the bipartition result will be chosen for subsequent bipartition. Third, we recursively repartition the large cluster of previous bipartition result if the number of intermediate clusters segmented by Ncut is less than the number of endmembers in the ßuorescence imaging. Finally, the largest group is considered as the background AF group and all the rest of smaller groups are merged into the target ßuorescence group. The intermediate target/background classiÞcation result after improved Ncut segmentation is shown in Fig. 1, where the different colors mean the different intermediate clusters sequentially segmented by the improved Ncut. The Þnal target ßuorescence group is formed by grouping all the small clusters except the large cluster of AF region.

B. Endmember Extraction for NMF Initialization

To Þnd the optimal product CS that best approaches the mixed image data matrix $D \in \mathbb{R}_+^{N\times L}$ (N is the total pixel number in a single image and L is the spectral band number), HALS-based NMF [26]–[28] is adopted to perform sequential constrained minimization on a set of subobjective functions $F(C_{:k}, S_{k:}) = (1/2)\|R_k - C_{:k}S_{k:}\|_2^2$, where each column $C_{:k}$ of $C \in \mathbb{R}_+^{N\times K}$ represents the spatial distribution of one endmember component, K is the number of the endmember, and each row $S_{k:}$ of $S \in \mathbb{R}_+^{K\times L}$ represents the spectrum of a speciÞc endmember. For $k = 1, 2, \ldots, K$, $R_k = D - \sum_{i\neq k} C_{:i}S_{i:}$.

To initialize NMF, we use the pixels of two groups to determine the corresponding spectra for the different ßuorescence components. We assume that the Þrst $(K - 1)$ constituent components represent the ßuorophores and the last component describes the AF. AF's initialized spectrum $S_{K(:)}$ is set to the average spectrum of all AF pixels, while the spectra $S_1, S_2, \ldots, S_{K-1}$

the $L_{1/2}$-norm as

$$F(C_{:k}, S_{k:}) = \frac{1}{2}\|R_k - C_{:k}S_{k:}\|_2^2 + 2\theta\sum_{i=1}^{N}(C_{ik})^{1/2} \quad (2)$$

where the θ is a regularized parameter to balance the tradeoff between the approximation accuracy and the sparseness of the multiple ßuorophores' abundances. The gradient derivation of $F(C_{:k}, S_{k:})$ with respect to $C_{:k}$ is

$$\frac{\partial F(C_{:k}, S_{k:})}{\partial C_{:k}} = -(R_k - C_{:k}S_{k:})S_{k:}^{T} + \theta(C_{:k})^{-1/2} \quad (3)$$

where $C_{:k}^{-1/2}$ is given by the element-wise square root for each entry in the vector $C_{:k}$. By setting (3) to zero, we can get the updating rule of $C_{:k}$. However, it involves a rather high computation cost due to the computation of the term $C_{:k}^{-1/2}$. To circumvent this problem, we approximate $C_{:k}$ in the $\theta(C_{:k})^{-1/2}$ term by its estimation $\hat{C}_{:k}$ obtained from the previous update, rather than compute the term $C_{:k}$ directly. Hence, (3) takes a simpliÞed and more computationally efÞcient form

$$\frac{\partial F(C_{:k}, S_{k:})}{\partial C_{:k}} = -(R_k - C_{:k}S_{k:})S_{k:}^{T} + \theta\hat{C}_{:k}^{-1/2}. \quad (4)$$

By setting (4) to zero, the rule of updating $C_{:k}$ takes the following form:

$$C_{:k} = \max\left(\text{eps}, (R_k S_{k:}^{T} - \theta(C_{:k})^{-1/2})/\|S_{k:}\|_2^2\right) \quad (5)$$

where eps is a very small constant ($\approx 10^{-16}$) and prevents from dividing by zero. The rules of updating $S_{k:}$ for (2) is

$$S_{k:} = \max\left(\text{eps}, C_{:k}^{T}R_k/\|C_{:k}\|_2^2\right). \quad (6)$$

For $k = K$, the subobjective function has no sparseness constraint, and the corresponding updating rules are the same as the other parts of the HALS optimization.

For the convenience of parameter setting, we convert the regularized parameter θ of TBCR-NMF into a desired sparsity value φ [10], which represents the sparseness degree that we expect the abundances of multiple ßuorophores to reach. The sparsity value φ, being 0 for nonsparse results and 1 for extremely sparse results, can be deÞned as

$$\varphi(C_k) = \frac{\sqrt{N} - \left(\sum_{n=1}^{N}|c_{nk}|/\sqrt{\sum_{n=1}^{N}c_{nk}^2}\right)}{\sqrt{N} - 1} \quad (7)$$

where $C_k \in \mathbb{R}_{+}^{N\times 1}$ is the kth column of abundance matrix C, and c_{nk} is each element at C_k with $n = 1, 2, \ldots, N$. SpeciÞcally, for each ßuorophore's abundance that has a corresponding regularized parameter, we use a method similar to that in [31] to directly control the θ_k value: θ_k is initialized to 0.001, and after each iteration, the current sparsity φ is computed by (7) for the abundance C_k; then θ_k is increased by 5% if the current sparsity is less than the desired sparsity value φ; otherwise, θ_k is decreased by 5%.

The detailed pseudocode of partially sparse NMF algorithm is summarized in Algorithm 1. The algorithm computation is terminated when the absolute value of difference between the two adjacent objective functions is less than 10^{-4}, or the maximum number of iterations exceeds 1000.

Algorithm 1 Partially Sparse NMF

Input: Data matrix $D \in \mathbb{R}_{+}^{N\times L}$ and initial $C \in R_{+}^{N\times K}$

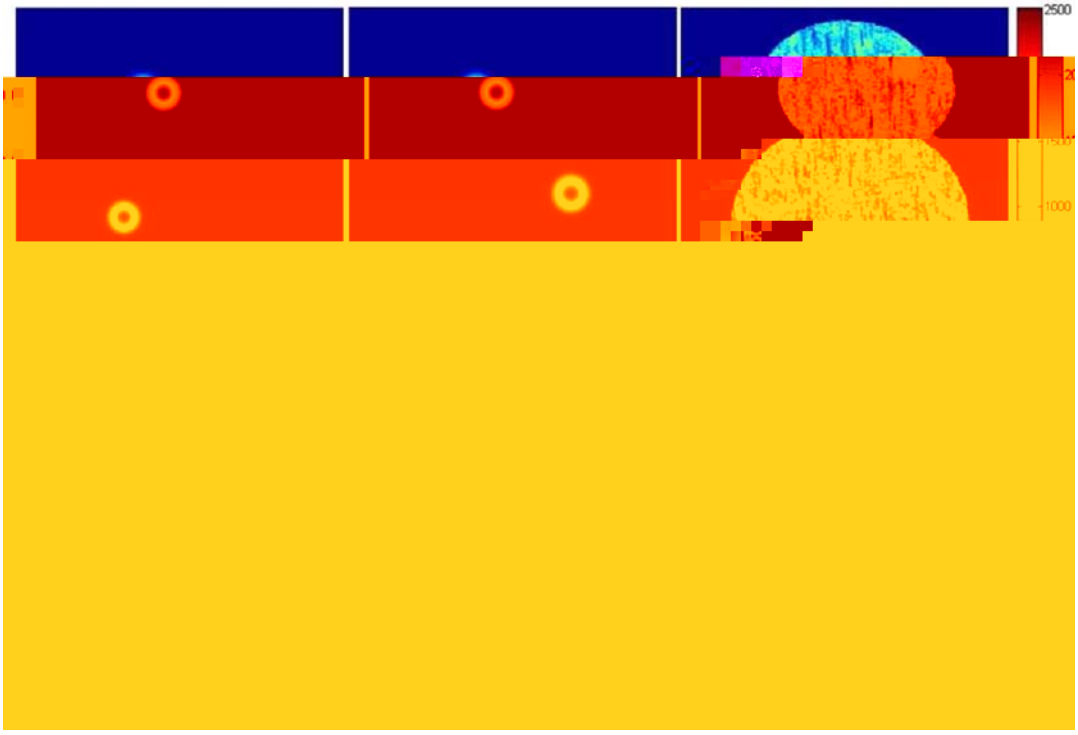


Fig. 2. Synthetic Data. (a)–(c) Abundances of AF488, AF555, and AF. (d) Corresponding emission spectra of AF488 (green line), AF555 (blue line), and AF (red line). (e) Mixed fluorescence image acquired at 555 nm. (f) Ground truth result obtained with improved Ncut target/background classification.

all endmember components are used to evaluate the overall performance of estimating spectral signatures and their corresponding abundances, respectively.

A. Synthetic Data

We use two fluorescence parts, Alexa Fluor 488 and Alexa Fluor 555 (AF488, AF555; Invitrogen, Carlsbad, USA), and one AF part to build a simulated phantom [4] in Fig. 2. The spectral signatures of fluorescence parts are the emission spectral curves of AF488 and AF555 at the spectral wavelengths from 480 nm to 650 nm with interval of 5 nm, while AF's spectral signature is a slowly varying curve in the same range [see Fig. 2(d), green line for AF488, blue line for AF555, and red line for AF]. The abundances of AF488 and AF555 consist of two parts: one part is pure fluorescent dye and the other part is mixed with AF488 and AF555 [see Fig. 2(a) and (b)]. The mixed part is at the top-left of the phantom, while the pure parts are at the bottom-left for AF488 in Fig. 2(a) and the bottom-right for AF555 in Fig. 2(b). According to the sparsity definition in (7), the true sparsity of the abundance of AF488 (or AF555) is 0.85. Finally, the total simulated phantom is obtained by adding the two fluorescent parts and the AF part together, and the abundance intensity ratio of AF to fluorophores (AF/F intensity ratio) is 0.3. Fig. 2(e) shows the mixed fluorescence image acquired at the spectral band of 555 nm wavelength. Fig. 2(f) shows the classification result of the synthetic data. The target fluorescence group is obtained by combining three different fluorescence regions (with different colors), which are sequentially isolated from the background AF group after three iterations of the improved recursive Ncut computation.

We test the above-mentioned algorithms' unmixing performances with the metrics of $\overline{\mathrm{SAD}}$ and $\overline{\mathrm{RMSE}}$, which have averages (bars) and standard deviations (error bars) resulting from the 20 runs of each algorithm. The algorithms' performances are dependent on the initial sparsity parameter φ (or the regularized parameter θ). Too small values of φ and θ cannot represent a reasonable sparsity of unmixed results while too large values of φ and θ will lead to excessive sparsity and inaccurate unmixed results. Considering the general sparsity of abundances for multiple fluorophores in preclinical applications, we select a series of φ values from 0.5 to 0.9 with an interval of 0.05 to demonstrate the effects of different initial sparsity values on the performance of S-NMF and TBCR-NMF. As for L_1-HALS, the θ value is 10^{-4}, 5×10^{-4}, 10^{-3}, 5×10^{-3}, 0.01, 0.05, 0.1, 0.5, and 1.0.

The performance metrics for different parameters θ and φ are shown in Fig. 3 when the AF/F intensity ratio is set to 0.3 with no noise added and with signal-to-noise ratio (SNR) being set to SNR = 15 dB. L_1-HALS's performance has been influenced by the regularized parameter θ. The $\overline{\mathrm{SAD}}$ and $\overline{\mathrm{RMSE}}$ achieved with L_1-HALS are the largest compared with other algorithms. For S-NMF and TBCR-NMF, their unmixing performances also have been influenced by the different values of initial sparsity parameter φ, which should be set to the true sparsity (0.85) of the abundance of multiple fluorophores in ideal situations. Therefore, too small or too large φ cannot produce good unmixed results for S-NMF and TBCR-NMF. In Fig. 3(a), when φ is smaller than 0.65 or larger

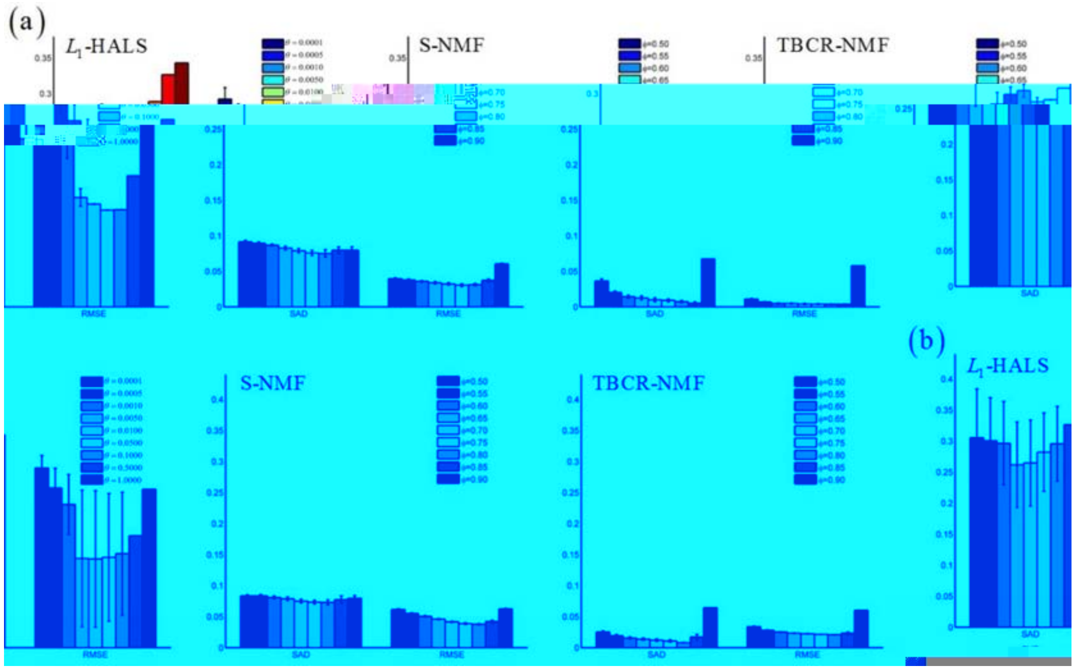


Fig. 3. Algorithm performances ($\overline{SAD}$ and $\overline{RMSE}$) for different values of parameter θ and φ when (a) no noise is added and (b) noise is added with SNR = 15 dB.

than 0.85, the $\overline{SAD}$ and $\overline{RMSE}$ become large and the unmixed results obtained with S-NMF and TBCR-NMF become worse. However, TBCR-NMF obtains the smallest $\overline{SAD}$ and $\overline{RMSE}$ among the three algorithms when the sparsity φ changes from 0.50 to 0.90 (or θ changes from 0.0001 to 1.0000). When the sparsity φ value is set to 0.90 (or 0.50), the poorest performances are achieved by all three algorithms, among which the TBCR-NMF is still the best.

The SNR in Fig. 3(b) is set to 15 dB, so that there is strong noise in the ßuorescence data and the performance of all algorithms degrades with increasing noise levels. The proposed TBCR-NMF still achieves the smallest $\overline{SAD}$ and $\overline{RMSE}$ when φ ranges from 0.5 to 0.9. The $\overline{RMSE}$ achieved by TBCR-NMF is relatively steady when φ ranges from 0.5 to 0.85. When the initial sparsity value φ exceeds the true sparsity (0.85), the values of $\overline{SAD}$ and $\overline{RMSE}$ obtained by TBCR-NMF increase obviously, but are still smaller than those obtained by other algorithms. Therefore, TBCR-NMF can achieve the best unmixed results when there is strong noise in the ßuorescence data.

As low AF/F ratio will highlight multiple localized ßuorophores from the background AF, it essentially makes the mixed spectral data sparser than the high AF/F ratio and the corresponding NMF problem will have sparser solutions than the high AF/F ratio. Therefore, the NMF performance is largely dependent on the AF/F ratio. For simulation experiments, the AF/F intensity ratio ranges from 0.1 to 0.9 with interval of 0.2. The φ and θ parameters are set to 0.8 and 0.01, respectively, to achieve the best unmixing performance for all algorithms.

Fig. 4(a) shows the different $\overline{SAD}$ and $\overline{RMSE}$ values for the different AF/F intensity ratios in the noiseless data. The initialization is Random method for L_1-HALS, Pure method for S-NMF, and Graph method for TBCR-NMF. Fig. 4(a) shows that the performances of L_1-HALS and S-NMF algorithms improve with decreasing AF/F intensity ratio. The unmixed results obtained with L_1-HALS are worst compared with other results. However, TBCR-NMF is not sensitive to the AF/F ratio and obtains the smallest $\overline{SAD}$ and $\overline{RMSE}$ values in all different AF/F intensity ratios.

Fig. 4(b) shows the performance dependence on the different initialization methods, when AF/F ratio, φ, and θ, are set to 0.3, 0.8, and 0.001, respectively. L_1-HALS is initialized with Random, Graph, and Pure methods, while S-NMF and TBCR-NMF are initialized with Graph and Pure methods. The $\overline{SAD}$ and $\overline{RMSE}$ values of L_1-HALS using different initializations are similar and the worst among the three algorithms due to the sparseness constraint being imposed on all abundances. S-NMFÕs performance improves with the Pure initialization having a p c lt F ti (w e e) 7 3 2 2 (p r m p o s) in

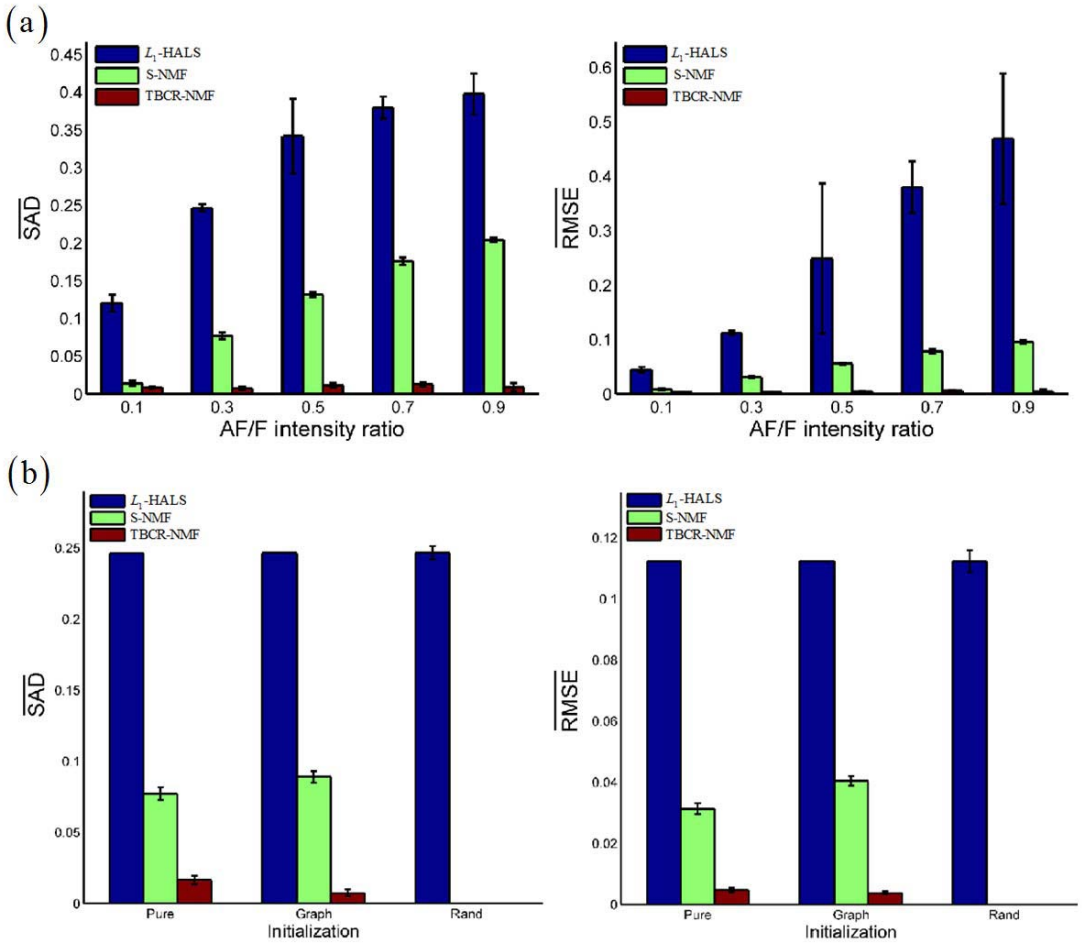


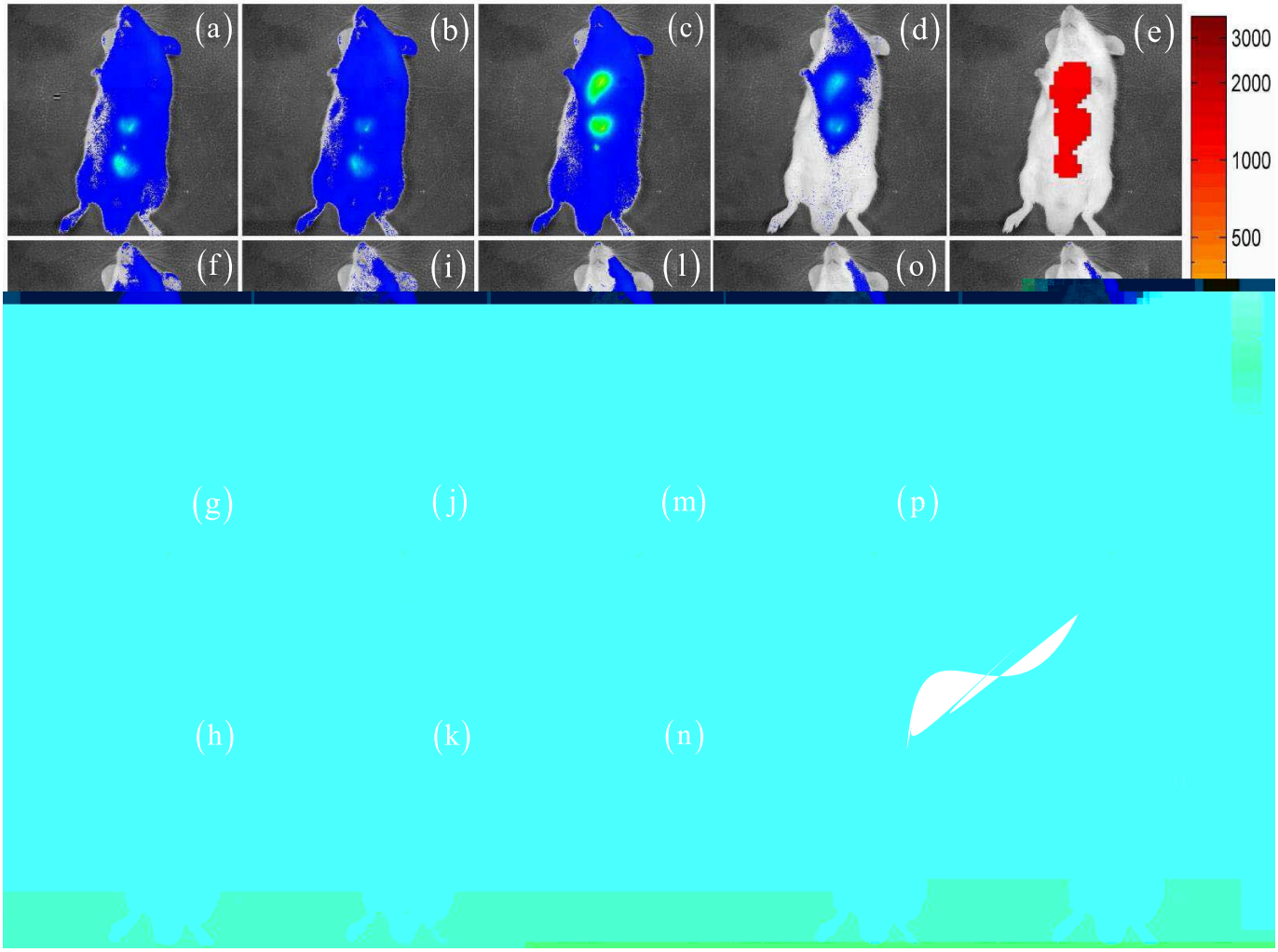
Fig. 4. (a) Algorithm performances for different values of AF/F intensity ratio. (b) Algorithm performances for different initialization methods.

Besides the L_1-HALS and S-NMF, we also use two NMF-based SUM algorithms: nonnegative matrix underapproximation (NMU)[2] [34] and NMF with L_0 constraint (L_0-NMF) [35] for comparison. NMU solves NMF problem with additional underapproximation constraint $CS \leq D$ which allows to obtain better part-based decompositions, while L_0-NMF introduces sparseness into all abundances via the L_0-norm constraint. All algorithms assume that the number of endmembers is 3, or $K = 3$. The TBCR-NMF, L_1-HALS, NMU, and L_0-NMF methods are not initialized with a priori knowledge of calibrated endmember spectra except the S-NMF method. The parameter θ for L_0-NMF, S-NMF, and TBCR-NMF is 0.75, the parameter θ for L_1-HALS is 0.005. To reduce computation cost, we use OtsuÕs [36] method for preprocessing to obtain the mouse body mask with whole ßuorescence regions of interest, whereby all algorithms are applied to the ßuorescence data within the mask only. For the best visual effect, all the observations and unmixed results (spatial distribution of all constituent components) are shown with rainbow pseudocolor and overlaid on the gray-scale photographic image of corresponding mouse.

We Þrst give two *in vivo* BALB/c mouse experimental results to validate the proposed methodÕs performance. All animal experiments in this paper were approved by our institutional review board. In experiment I, AF488 and Alexa Fluor 594 (AF594; Invitrogen, Carlsbad, USA) ßuorescent dyes are diluted to 0.1 μg$\cdot$ml^{-1}. AF488 is injected at the bottom of the body with 20 ng dye, while AF594 is injected near the neck with the same quantity, and a mixture of each dye with 10 ng is located at the middle portion of the body. These three injections are not exactly controlled with the same depth in tissues. Fig. 5(a)Ð(d) shows four raw ßuorescence images acquired at 542, 579, 624, and 716 nm spectral bands. The Þrst two images are excited at 474 nm and the last two images at 565 nm. The calibrated spectra of AF488, AF594, and AF [see Fig. 6(a)] are acquired at these four emission Þlters by precalibration *ex vivo* experiments in the same imaging conditions, while the spectrum of AF is the average spectrum acquired in some chosen regions of mouse with no ßuorescent dyes. Fig. 6(a) displays that AF488, AF594 and AF have overlapping emission spectra. Fig. 6(b) and (c) also shows the calibrated spectra of AF555, and AF, acquired *ex vivo* at the 525, 542, 579, and 624 nm spectral bands for the next two *in vivo* ßuorescence imaging experiments in the following section.

Fig. 5(e) shows the target/background classiÞcation result where the multiple ßuorophores are classiÞed as a target group (red color) and separated from the whole background

[2]https://sites.google.com/site/nicolasgillis/code


(a)
(b)
(c)
(d)
(e)
3000
2000
1000
500
(f)
(i)
(l)
(o)
(g)
(j)
(m)
(p)
(h)
(k)
(n)

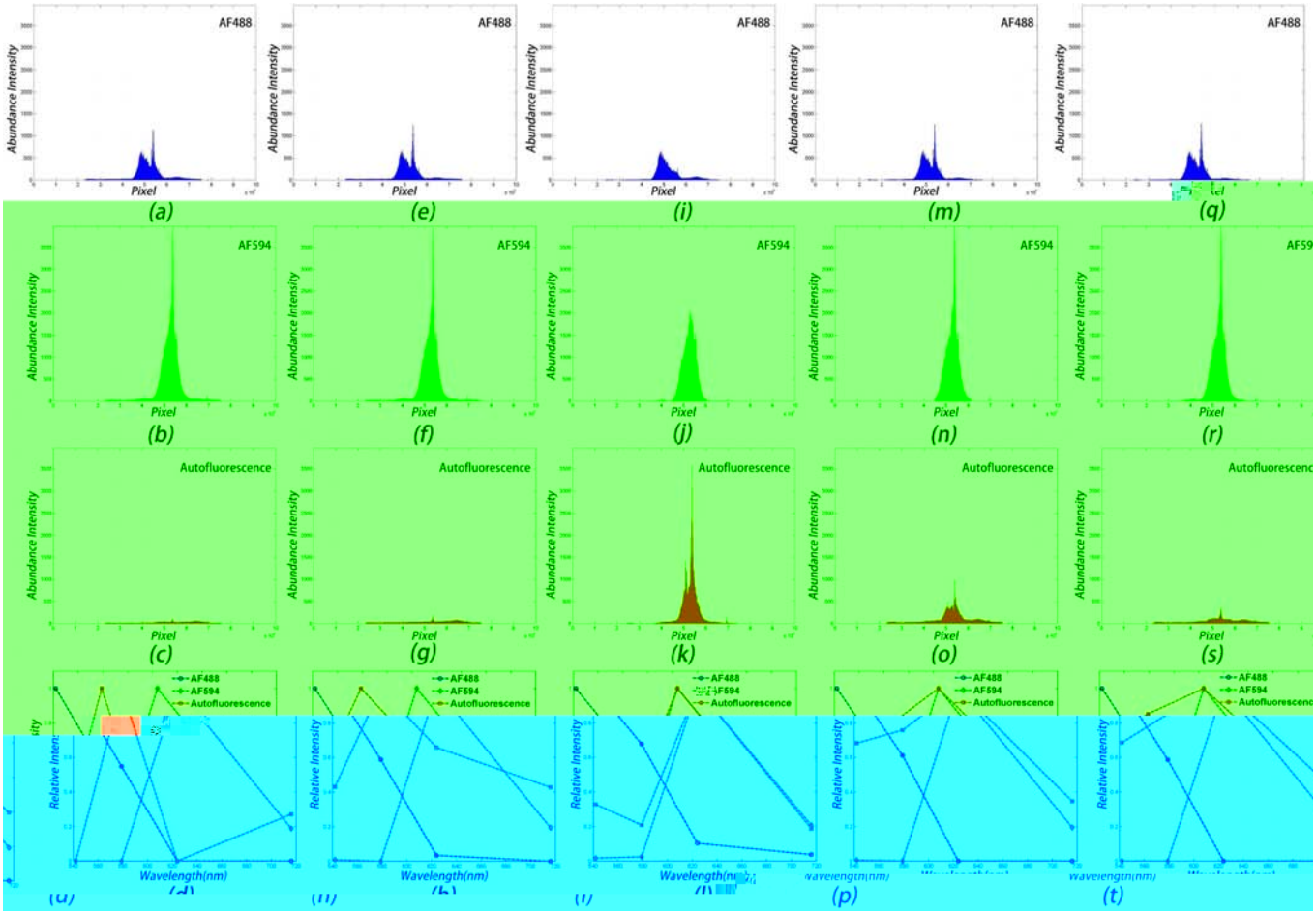


Fig. 7. Abundances (expressed as 1-D vectors) of AF488, AF594, AF, and their spectra from top row to bottom row for experiment I. The unmixed results obtained with (a)–(d) NMU, (e)–(h) L_0-NMF, (i)–(l) L_1-HALS, (m)–(p) S-NMF, and (q)–(t) TBCR-NMF.

There are some missing parts of AF488 in the middle portion of BALB/c mouse in Fig. 5(l) for L_1-HALS, which also falsely makes the unmixed background AF [Fig. 5(n)] appear sparser and brighter than it actually is. The TBCR-NMF and S-NMF can separate the fluorescence targets from AF in Fig. 5(o)–(q) and Fig. 5(r)–(t), respectively. However, the unmixed results obtained with TBCR-NMF are smoother and clearer than S-NMF.

The unmixed results of abundances (expressed as 1-D vectors) and endmember spectra are illustrated in Fig. 7(a)–(d) for NMU, Fig. 7(e)–(h) for L_0-NMF, Fig. 7(i)–(l) for L_1-HALS, Fig. 7(m)–(p) for S-NMF and Fig. 7(q)–(t) for TBCR-NMF, respectively. The TBCR-NMF algorithm obtains more accurate unmixed fluorescence abundances compared with the other algorithms. All algorithms have the highest abundance intensities that correspond to the true pixel positions of fluorescence targets. The abundances of AF488 and AF594 are wide and contain the unwanted AF parts that are not removed with NMU in Fig. 7(a) and (b) and L_0-NMF in Fig. 7(e) and (f). This AF remainder also can be confirmed by AF's abundances [Fig. 7(c) and (g)] obtained with both algorithms. The L_1-HALS [Fig. 7(k)] and S-NMF [Fig. 7(o)] have abnormal (too large) values in the abundances of AF, except that the TBCR-NMF [Fig. 7(s)] has slowly varying abundances of AF. More importantly, the spectra estimated with S-NMF

TABLE I

SAD AND $\overline{\text{SAD}}$ RESULTS ON THE *In Vivo* EXPERIMENTS I AND II (THE SMALLER VALUES MEAN BETTER RESULTS. THE NUMBER IN BOLD REPRESENTS THE BEST PERFORMANCE)

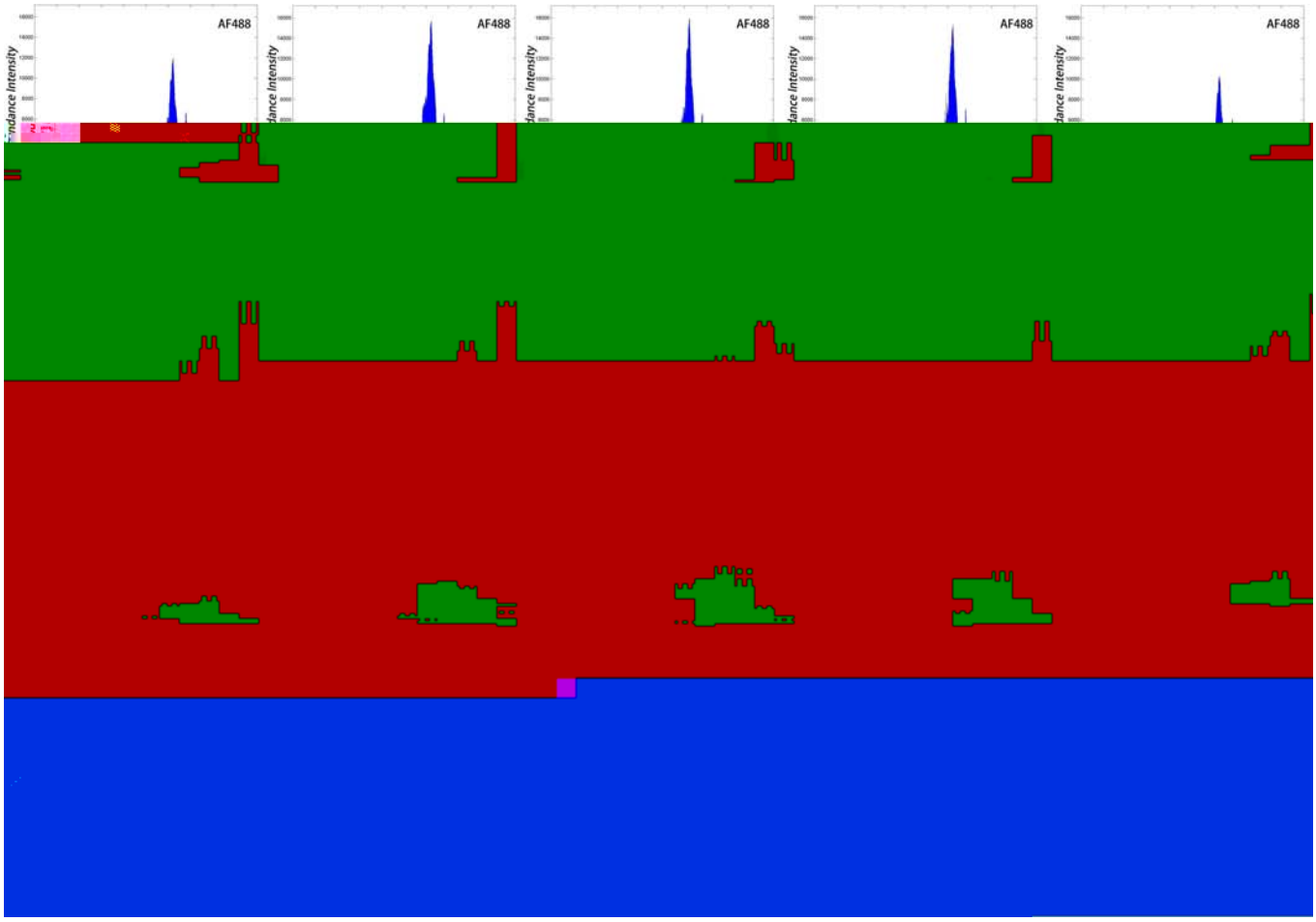


Fig. 9. Abundances (expressed as 1-D vectors) of AF488, AF555, AF, and their spectra from top row to bottom row for experiment II. The unmixed results obtained with (a)–(d) NMU, (e)–(h) L_0-NMF, (i)–(l) L_1-HALS, (m)–(p) S-NMF, and (q)–(t) TBCR-NMF.

The unmixed abundances (expressed as 1-D vectors) and endmember spectra are illustrated in Fig. 9(a)–(d) for NMU, Fig. 9(e)–(h) for L_0-NMF, Fig. 9(i)–(l) for L_1-HALS, Fig. 9(m)–(p) for S-NMF and Fig. 9(q)–(t) for TBCR-NMF, respectively. The TBCR-NMF achieves the most accurate unmixed results compared with other algorithms. Particularly, the first row in Fig. 9 displays that AF488 abundance intensities obtained with other algorithms still contain AF parts that are not fully removed. The second row in Fig. 9 shows that the AF parts have made contributions to the AF555 abundances with the NMU, L_0-NMF and L_1-HALS algorithms except the TBCR-NMF and S-NMF. The S-NMF has abnormal (too large) values for the AF abundance in Fig. 9(o), while TBCR-NMF can get slowly varying abundances of AF in Fig. 9(s). In general, the unmixed abundances of AF488 and AF555 from the TBCR-NMF algorithm [Fig. 9(q) and (r)] are more sparse than other algorithms. Moreover, the spectra obtained with NMU in Fig. 9(d), L_0-NMF in Fig. 9(h), and L_1-HALS in Fig. 9(l) are clearly different from the calibrated spectra in Fig. 6(b). However, the spectra obtained with S-NMF in Fig. 9(p) and TBCR-NMF in Fig. 9(t) are more accurate than other algorithms.

The SAD and the average SAD values of three unmixed endmembers for in vivo experiment II are shown in Table I. We can see that the TBCR-NMF gets the smallest SAD (0.0961) and the best unmixing performance.

D. In Vivo Experiments III and IV

In this section, using another two in vivo BALB/c mouse experiments, we further compare the proposed method with three recently published unmixing methods which also utilize the spatial information in the multispectral images. Specifically, besides the S-NMF that has good unmixing performance, we also use the following methods for performance comparison: the beta compositional model based spatial-spectral (BCM-spatial) algorithm [18], sparse unmixing via variable splitting augmented Lagrangian and total variation (SparseTV) algorithm [19], and regularized simultaneous forward-backward greedy (RSFoBa) algorithm [20]. The BCM-spatial method assumes beta-distributed endmembers and identifies pixels with similar proportion values to the pixel under unmixing by identifying the K-nearest spatial-spectral neighbors. The SparseTV algorithm [19], two

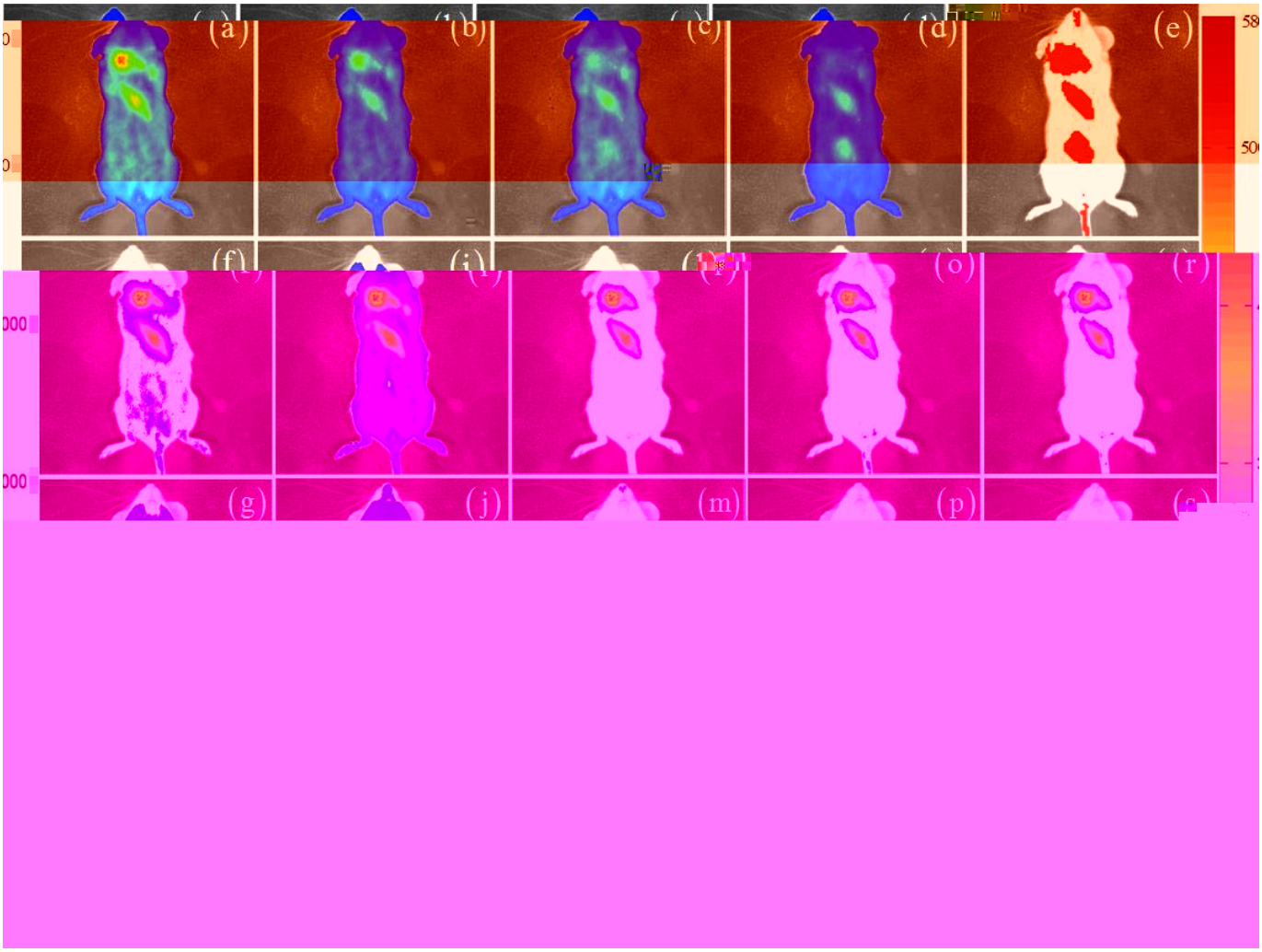


Fig. 10. (a)–(d) Raw fluorescence (AF488 and AF546) images for experiment III acquired at the 525, 542, 579, and 624 nm emission filters respectively, the first two images are excited at 474 nm and the last two images at 500 nm. (e) Classification results. The different unmixed results obtained with (f)–(h) BCM-spatial, (i)–(k) SparseTV, (l)–(n) RSFoBa, (o)–(q) S-NMF, and (r)–(t) TBCR-NMF.

includes the total variation regularization to the classical sparse regression formulation to exploit the spatial-contextual information present in the multispectral images. The RSFoBa

TABLE II

[illegible] IV (THE SMALLER VA[illegible]TTER RESULTS.

[illegible]

T. Tian-heng, and T. Li-ming for contributing to the project, and Dr. P. Jin-liang for supporting the experiments in ßuorescence imaging. The authors declare that they have no competing interests.

REFERENCES

[1] F. Leblond, S. C. Davis, P. A. ValdŽs, and B. W. Pogue, ÒPre-clinical whole-body ßuorescence imaging: Review of instruments, methods and applications,Ó J. Photochem. Photobiol. B, Biol., vol. 98, pp. 77Ð94, Jan. 2010.

[2] F. L. Pe–a, J. L. Crespo, and R. J. Duro, ÒUnmixing low-ratio endmembers in hyperspectral images through Gaussian synapse ANNs,Ó IEEE Trans. Instrum. Meas., vol. 59, no. 7, pp. 1834Ð1840, Jul. 2010.

[3] J. Plaza, E. M. T. Hendrix, I. Garc'a, G. Mart'n, and A. Plaza, ÒOn endmember identiÞcation in hyperspectral images without pure pixels: A comparison of algorithms,Ó J. Math. Imag. Vis., vol. 42, pp. 163Ð175, Feb. 2012.

[4] A.-S. Montcuquet, L. HervŽ, F. Navarro, J.-M. Dinten, and J. I. Mars, ÒIn vivo ßuorescence spectra unmixing and autoßuorescence removal by sparse nonnegative matrix factorization,Ó IEEE Trans. Biomed. Eng. vol. 58, no. 9, pp. 2554Ð2565, Sep. 2011.

[5] R. Zdunek, ÒRegularized nonnegative matrix factorization: Geometrical interpretation and application to spectral unmixing,Ó J. Appl. Math. Comput. Sci., vol. 24, no. 2, pp. 233Ð247, 2014.

[6] J. Galeano, S. Perez, Y. Montoya, D. Botina, and J. Garz—n, ÒBlind source separation of ex-vivo aorta tissue multispectral images,Ó Biomed. Opt. Exp., vol. 6, no. 5, pp. 1589Ð1598, 2015.

[7] J. Xu et al., ÒSparse non-negative matrix

Chen Hu received the B.S. degree in biomedical Engineering from XiÕan Jiaotong University, XiÕan, China, in 2011, and the M.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2014.

Shaosen Huang received the B.S. degree and the M.S. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011 and 2013, respectively.