Forgery Detection in Hyperspectral Document Images using Graph Orthogonal Nonnegative Matrix Factorization

Abderrahmane Rahiche, Mohamed Cheriet
Synchromedia Laboratory, École de Technologie Supérieure (ETS), Montreal, Canada
abderrahmane.rahiche.1@ens.etsmtl.ca, Mohamed.Cheriet@etsmtl.ca

Abstract

The analysis of inks plays a crucial role in the examination process of questioned documents. To address this issue, we propose a new approach for ink mismatch detection in Hyperspectral document (HSD) images based on a new orthogonal and graph regularized Nonnegative Matrix Factorization (NMF) model. Although some previous works have proposed orthogonality constraints to solve clustering problems in different contexts, the application of such constraints is not straightforward due to the sum-to-one assumption related to the physical nature of Hyperspectral images. In this work, we demonstrate that under some acquisition protocols, latent factors in HSD images can be constrained to be orthogonal. We also incorporate a graph regularized term to exploit the geometric information lost by the matricization of HSD images. Furthermore, we propose an efficient alternating direction method of multipliers based algorithm to solve the proposed method. Our empirical validation demonstrates the competitiveness of the proposed algorithm compared to the state-of-the-art methods. It shows a high potential to be used as a reliable tool for quick investigation of questioned documents.

1. Introduction

Forensic document analysis aims to answer questions about the authenticity of questioned documents that refer to documents whose authenticity is not certain. Forensic scientists use scientific methods to identify if any changes or modifications have been made to a document and determine whether it has been a subject of fakes, forgeries, or not.

Analysis of documents involves physical examinations of the ink and paper from which questioned documents are made. The scientific procedures used for inks examination can be broadly classified into destructive and non-destructive techniques. Destructive analysis methods, such as Thin Layer Chromatography (TLC) [2], includes sampling fragments and removing parts from documents for the purpose of destructive analysis. These methods cause in general permanent and irreversible changes to documents.

Non-destructive techniques, such as digital imaging and spectroscopy methods, allow the characterization and identification of materials without the need for destructive sampling. Indeed, Hyperspectral (HS) and Multispectral (MS) imaging modalities have demonstrated to be useful for the analysis of documents, and they are becoming a key tool for forensic questioned document examination. Spectral imaging has shown to be a powerful non-contact analysis technique for document analysis. It has opened new prospects for the detection, characterization, and identification of material composition. For instance, it has been used for the detection, visualization, identification and age estimation of forensic traces [11], artwork authentication [25], age estimation of old manuscripts [27], sketch layer separation [9], and ink mismatch detection [18].

Spectral imagery allows differentiation between different types of materials present on the scanned scene (or specimen) based on their spectral responses. However, one of the key issues of HSD image processing is the automatic identification and mapping of materials. NMF factorization technique is a powerful tool that can be used to extract meaningful information from high-dimensional data. NMF is widely used for remote sensing and data mining application. Due to its meaningful nonnegativity property, it has been used successfully to address the spectral unmixing problem, which aims to decomposing mixed pixels into a set of endmembers and abundance fractions.

In this work, we propose a novel approach for ink forgery detection in HSD images via a graph orthogonal nonnegative matrix factorization model. The advantages of this model are: i) it is independent of the number of spectral bands, ii) it exploits the spectral and spatial information of pixels as well, and iii) it performs unsupervised classification of inks, it does not require training data, and its performance is totally independent of the type of ink. Thus, the proposed method could be used as a reliable tool for quick investigation of questioned documents.

We summarize our contribution as follows: 1) we de-

velop a new NMF based method to address the issue of ink mismatch detection in HSD images. 2) Besides the nonnegativity of low-rank factors, we show that under some conditions, the latent components in HSD images can be constrained to be orthogonal (independent), and accordingly, we incorporate a manifold regularization. 3) We leverage these three properties to perform source (inks) separation. 4) We develop an ADMM based algorithm for the optimization of our model, and we validate our model on real HSD images.

The remainder of this paper is organized as follows. The following section presents related works and relevant studies published in this filed. Section 3 then presents the proposed approach and the optimization method adopted to solve the formulated problem. The obtained results are presented in Section 5 with the pertinent discussions. Section 6 closes the paper with a conclusion and future works.

2. Related Work

Several previous studies have addressed the issue of inks analysis using spectral imaging modalities, and their potential has been demonstrated for the analysis and the determination of forgery, fraud, and manipulation of documents. For instance, authors in [14] and [29] developed a HS imaging-based system for forensic analysis of document forgery. With the advancements in pattern recognition capabilities, automated analysis techniques continue to replace visual analysis of questioned documents.

Spectral image processing techniques used for inks discrimination and mismatch detection can be broadly categorized into supervised and unsupervised techniques. Unsupervised techniques exploit the rich spectral information obtained by HS imagery and make use of clustering techniques to analyze HS images as in [18, 19, 21]. HS unmixing methods have also been used to address the ink differentiation problem as in [1].

Unlike unsupervised techniques, supervised methods require training data to build discriminative models. Morales *et al.* [23] proposed an approach for ink analysis in handwritten documents and pen verification using hyperspectral analysis and Least Square SVM classification. Recently, the booming of deep learning techniques attracted the attention of researchers in this field. For instance, Jaleed Khan *et al.* [17] proposed a spatio-spectral hybrid CNN based approach for HSD authentication. Supervised methods, especially deep learning-based approaches, show great potential if only we train them with a huge amount of images. Their performances rely on training data, i.e., they could not deal with new spectral signatures of inks that they have not seen before during the learning process.

Previous studies show some limitations to real applications. Unsupervised approaches are inspired by methods developed for remote HS imaging without taking into ac-count the physical properties of HS images of documents, which make them not well adapted for the analysis of HSD images. Supervised methods rely on labeled training data, that might be costly to obtain or not immediately available. Furthermore, the applicability of supervised classification systems is limited to scenarios where prior information about the type of inks used in a questioned document is available, which is not the case in real-life problems. Therefore, in this work, we develop a new NMF-based method to handle the inks discrimination issue. Unlike previous works, the proposed method takes into account the physical properties of HSD, does not rely on any trained model, and can work with an arbitrary number of channels.

3. Proposed method

The overall pipeline of our method is shown in Fig. 1. Given an input set of spectral images $(S_i)_{1 \leq i \leq m}$ of an arbitrary number of bands m, the input data cube is first reshaped into one data matrix, then factorized using the proposed graph orthogonal NMF model (GONMF-ADMM) that is described hereafter. Due to the nonnegativity property, the obtained parts are directly interpretable without any further post processing.

Let $\mathbf{Y} \in \mathbb{R}^{m \times n}$ be a data matrix of m features and n samples, and with only nonnegative entries. It is possible to approximate $\mathbf{Y}$ by the product of two low-rank nonnegative matrices $\mathbf{M} \in \mathbb{R}^{m \times k}$ and $\mathbf{A} \in \mathbb{R}^{k \times n}$, where $k = min(n, m)$. The factorization problem then writes: $\mathbf{Y} \approx \mathbf{MA}$.

The performance of the factorization can be measured using several well-known cost functions to measure the distance between the data matrix and the two factors product. For instance, the most used cost functions in the literature are the Euclidian distance [24], the Kullback-Leibler (KL) divergence, the Itakura-Saito (IS) divergence [13], and the Earth Mover's Distance (EMD) [28]. Generally, the factorization problem yields an optimization problem that minimizes the cost functions used. Furthermore, a regularization term could be added to the cost function to allow the obtention of low-rank factors with specific properties. Therefore, the NMF problem can be expressed by a general mathematical formulation by solving

$$\min_{\mathbf{M} \geq 0, \mathbf{A} \geq 0} D(\mathbf{Y}|\mathbf{MA}) + \lambda R(\mathbf{M}, \mathbf{A}), \qquad (1)$$

where D denotes a given cost function, R a regularization term, and λ is a parameter that controls the importance of the regularization.

Besides the nonnegativity constraint, the problem in 2 could be constrained with additional conditions depending on the problem in hand. For instance, in the field of HS image processing, the matrix $\mathbf{A}$ of abundances fractions should also obey the full additivity condition, i.e. the sum-to-one
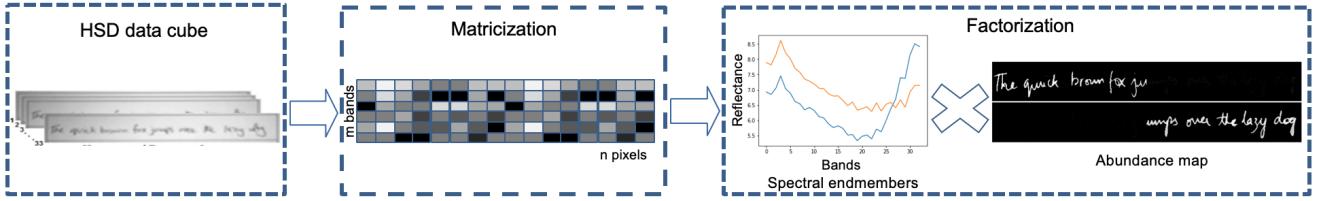
Figure 1: Pipeline of GONMF-ADMM based forgery detection in HSD images.

constraint defined as $\sum_{i=1}^{k} \mathbf{A}_{ij} = 1$ for all j, due to some physical assumptions related to HS data.

In this work, we are in particular interested in the orthogonality constraint that demonstrates effectiveness for data clustering. Nevertheless, the orthogonality constraint implies the linear independence of the rows of the abundance's matrix $\mathbf{A}$, by which each column is allowed to contain at most one non-zero value. Hence, it is not possible to use it with the presence of the sum-to-one constraint.

Therefore, to overcome this limitation, we relax the problem with the following assumptions: 1) spectral document images generally have a high resolution because they are scanned from very close distances, which implies that the size of pixels is small enough due to the small field-of-view, thus, it is not reasonable to expect that more than one material will occupy the area of one pixel; 2) the sum-to-one condition itself has been criticized in the literature [16]. Accordingly, by taking into account that the digitization process of documents is generally performed in a closed area and under controlled conditions allows us to do not consider this sum-to-one condition in our context.

In other hand, the matricization of HSD images involves flattening each HSD image and stacking them together to form the 2D data matrix required for matrix factorization. However, images flattening results in a loss of local structure between pixels. Thus, to take into account spatial correlations between pixels, we consider the following graph regularization term [6]:

$$R(\mathbf{A}) = tr(\mathbf{A}\mathbf{L}\mathbf{A}^T),$$

where tr denotes the matrix trace operator, $\lambda > 0$ is a regularization parameter, $\mathbf{L} \in \mathbb{R}^{n \times n}$ is the graph Laplacian matrix [8] associated to the data $\mathbf{Y}$.

Finally, by combining the nonnegativity, the orthogonality constraint, and the manifold regularization together, we formulate our graph orthogonal NMF (GONMF) model as follows:

$$\min_{\mathbf{M},\mathbf{A}} \frac{1}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_F^2 + \lambda tr(\mathbf{A}\mathbf{L}\mathbf{A}^T),$$

subject to: $\mathbf{M} \geq 0, \mathbf{A} \geq 0, \mathbf{A}\mathbf{A}^T = \mathbf{I}_k,$ $\quad(2)$

where $\lambda > 0$ is a regularization parameter and $\mathbf{I}_k$ is an identity matrix of k elements.

In the literature, the orthogonality constraint has been addressed using different strategies. Ding *et al.* [10] developed an orthogonal nonnegative matrix tri-factorizations model by integrating the orthogonality as a penalty term into the objective function of his model, and solving the problem using the multiplicative update rules. In [26], the authors incorporated the orthogonality constraint as a quadratic penalty term and used an alternating scheme to update the different variables. They adopted an active-set based method to update the first variable and a projected gradient scheme for the second variable. Another strategy was used in [20], based on splitting the orthogonality constraint for general constrained problems.

In this paper, we adopt the splitting strategy and the alternating direction method to handle the orthogonality and the graph regularization in nonnegative matrix factorizations. The details of the developed optimization approach are presented in the next section.

4. Optimization approach

The objective function in Eq.2 is bi-convex, i.e. is not convex for both variables $\mathbf{M}$ and $\mathbf{A}$ simultaneously. Thus, like many approaches in the literature [24], we adopt an iterative alternating optimization scheme for minimizing Eq.2 over each variable.

After discarding the terms that are independent of the variable over which we want to optimize, the minimization problem with respect to each variable reduces to solving the following two sub-problems separately:

1. For $\mathbf{A}$ fixed, the optimal $\mathbf{M}$ can be computed by solving:

$$\mathbf{M} \leftarrow \arg\min_{\mathbf{M} \geq 0} \frac{1}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_F^2 \qquad (3)$$

2. For $\mathbf{M}$ fixed, the optimal $\mathbf{A}$ can be obtained by solving:

$$\mathbf{A} \leftarrow \arg\min_{\substack{\mathbf{A} \geq 0, \\ \mathbf{A}\mathbf{A}^T = \mathbf{I}_k}} \frac{1}{2}\|\mathbf{Y} - \mathbf{M}\mathbf{A}\|_F^2 + \lambda tr(\mathbf{A}\mathbf{L}\mathbf{A}^T) \quad (4)$$

Solving Eq. 3 is easy. Nevertheless, Eq. 4 is difficult to solve with the presence of the nonnegativity, the manifold regularization, and the orthogonality constraints altogether. Therefore, we adopt the Alternating Direction Method of Multipliers (ADMM) [5] scheme, which is a variant of

the augmented Lagrangian method, to solve the two subproblems in 3 and 4 efficiently. The pseudo-code given in Algorithm 1 summarizes the overall optimization scheme of the original problem in 2.

Algorithm 1 : Pseudo-code for GONMF-ADMM model

Input: $\mathbf{X}$, k, ρ_1, ρ_2, *Tol*, *niter*
Output: $\mathbf{M}$, $\mathbf{A}$
   Initialization: $\mathbf{M_0}$, $\mathbf{A_0}$
1: **for** $i = 1$ to *niter* **do**
2:    Update of $\mathbf{M}$ using Algorithm. 2.
3:    Update of $\mathbf{A}$ using Algorithm. 3.
4:    **if** $((r[i] - r[i-1])/r[i-1] < Tol)$ **then**
5:      break..
6:    **end if**
7: **end for**
8: **return** $\mathbf{M}$, $\mathbf{A}$

The optimization steps for each sub-problem are described in detail in the next sections.

Optimization step of M

By incorporating a positive indicator function $\imath(\mathbf{M})$ to handle the nonnegativity constraint of the matrix $\mathbf{M}$, that is defined as:

$$\imath(\mathbf{M}) = \begin{cases} 0, & \text{if} \quad \mathbf{M} \geq 0. \\ +\infty, & \text{otherwise.} \end{cases} \quad , \quad (5)$$

we can reformulate the subproblem in Eq. 3 as follows:

$$\min_{\mathbf{M} \geq 0} \frac{1}{2} \|\mathbf{Y} - \mathbf{MA}\|_F^2 + \imath(\mathbf{M}). \quad (6)$$

Moreover, by introducing an auxiliary variable $\mathbf{X} = \mathbf{M}$, we rewrite the problem in the following form:

$$\min_{\mathbf{M}, \mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{MA}\|_F^2 + \imath(\mathbf{X}),$$
$$\text{subject to:} \quad \mathbf{M} = \mathbf{X}, \mathbf{X} \geq 0. \quad (7)$$

The augmented Lagrangian for Eq. 7 is then given by:

$$\mathcal{L}_M(\mathbf{M}, \mathbf{X}, \mathbf{\Lambda_1}) = \frac{1}{2} \|\mathbf{Y} - \mathbf{MA}\|_F^2 + \imath(\mathbf{X}) + \frac{\rho_1}{2} \|\mathbf{M} - \mathbf{X} + \mathbf{\Lambda_1}\|_2^2, \quad (8)$$

where $\mathbf{\Lambda_1} \in \mathbb{R}^{k \times n}$ is a Lagrangian multiplier, and ρ_1 is the corresponding penalty parameter.

Based on ADMM paradigm, we can write down all the intermediate updates that take place to solve Eq. 8, and we derive the iterations listed here:

$$\begin{cases} \mathbf{M^{t+1}} := \arg\min_{\mathbf{M}} \frac{1}{2} \|\mathbf{Y} - \mathbf{MA}\|_F^2 + \frac{\rho_2}{2} \|\mathbf{M} - \mathbf{X} + \mathbf{\Lambda_1}\|_2^2, \\ \mathbf{X^{t+1}} := \arg\min_{\mathbf{X} \geq 0} \imath(\mathbf{X}) + \frac{\rho_2}{2} \|\mathbf{M} - \mathbf{X} + \mathbf{\Lambda_1}\|_2^2, \\ \mathbf{\Lambda_1^{t+1}} = \mathbf{\Lambda_1} + \mathbf{M} - \mathbf{X} \end{cases}$$
$$(9)$$

We note that we derived a specific optimization problem for each primal variable that only depends on the terms corresponding to that variable in the augmented Lagrangian $\mathcal{L}_M$. Next, for each primal variable, we solve the corresponding minimization problem.

M minimization step: The optimal value of $\mathbf{M}$ that minimize the corresponding sub-objective function has an analytical expression that is obtained by setting the gradient of the augmented Lagrangian with respect to $\mathbf{M}$ to zero, which yields:

$$\mathbf{M^{t+1}} = [\mathbf{YH}^T + \rho_1(\mathbf{X} + \mathbf{\Lambda_1})][\mathbf{HH}^T + \rho_1\mathbf{I}]^{-1} \quad (10)$$

X minimization step: The $\mathbf{X}$ minimization problem admits a closed-form solution given by the following proximal operator :

$$\mathbf{X^{t+1}} = \max(\mathbf{X} + \mathbf{\Lambda_1}, 0) \quad (11)$$

The Lagrange multiplier $\mathbf{\Lambda_1}$ is updated using the gradient ascent formula given in the last line of (9). As a stopping criterion, the normalized norm of the dual residual $r_1 = \mathbf{M} - \mathbf{X}$ must be smaller than a fixed tolerance.

The resulting ADMM algorithm is summarized in Algorithm 2:

Algorithm 2 : Update of M using ADMM

Input: $\mathbf{Y}$, $\mathbf{M}$, ρ_1, *Tol*, *niter*
   Initialization: $\mathbf{\Lambda_1} = 0$, $\mathbf{X} = \mathbf{M}$
1: **for** $i = 1$ to *niter* **do**
2:    $\mathbf{M} \leftarrow [\mathbf{YH}^T + \rho_1(\mathbf{X} + \mathbf{\Lambda_1})][\mathbf{HH}^T + \rho_1\mathbf{I}]^{-1}$.
3:    $\mathbf{X} \leftarrow \max(\mathbf{X} - \mathbf{\Lambda_1}, 0)$.
4:    $\mathbf{\Lambda_1} \leftarrow \mathbf{\Lambda_1} + \mathbf{M} - \mathbf{X}$.
5:    **if** $((\|r_1^i\| - \|r_1^{i-1}\|)/\|r_1^{i-1}\| < Tol)$ **then**
6:      break..
7:    **end if**
8: **end for**
9: **return** $\mathbf{M}$, r_1

Optimization step of A

Similarly to the first subproblem of Eq. 3, we incorporate a positive indicator function $\imath(\mathbf{A})$ to deal with the nonnegativity constraint of the matrix $\mathbf{A}$. Furthermore, we introduce two auxiliary variables, namely $\mathbf{Z}$ and $\mathbf{P}$, for splitting

the nonnegativity and the orthogonality constraints from $\mathbf{A}$, which yields:

$$\min_{\mathbf{A},\mathbf{Z},\mathbf{P}} \frac{1}{2}\|\mathbf{Y} - \mathbf{MA}\|_F^2 + \lambda tr(\mathbf{ALA}^T) + \imath(\mathbf{Z}),$$

$$\text{subject to:} \quad \mathbf{A} = \mathbf{Z}, \ \mathbf{Z} \geq 0, \tag{12}$$

$$\mathbf{A} = \mathbf{P}, \ \mathbf{PP}^T = \mathbf{I}_k,$$

where

$$\imath(\mathbf{Z}) = \begin{cases} 0, & \text{if } \mathbf{Z} \geq 0. \\ +\infty, & \text{otherwise.} \end{cases} \tag{13}$$

The underlying optimization problem (12) can be reformulated using the augmented Lagrangian method as follows:

$$\mathcal{L}_A(\mathbf{A},\mathbf{Z},\mathbf{P},\mathbf{\Lambda_2}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{MA}\|_F^2 + \lambda tr(\mathbf{ALA}^T)$$
$$+ \imath(\mathbf{Z}) + \frac{\rho_2}{2}\|\mathbf{A} - \mathbf{Z} + \mathbf{\Lambda_2}\|_2^2 \tag{14}$$
$$+ \frac{\rho_3}{2}\|\mathbf{A} - \mathbf{P} + \mathbf{\Lambda_3}\|_2^2,$$

where $\mathbf{\Lambda_2}, \mathbf{\Lambda_3} \in \mathbb{R}^{k \times n}$ are the Lagrangian dual variables associated with the linear constraints introduced in 12, and $\rho_2, \rho_3 > 0$ are the corresponding penalty parameters.

We apply the ADMM algorithm that allows splitting the optimization problem in Eq. 14 into three subproblems, from which we derive five sequential steps to update the involved primal and dual variables, which can be outlined here:

$$\begin{cases} \mathbf{A^{t+1}} := \arg\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{Y} - \mathbf{MA}\|_F^2 + \lambda tr(\mathbf{ALA}^T) \\ \qquad + \frac{\rho_2}{2}\|\mathbf{A} - \mathbf{Z} + \mathbf{\Lambda_2}\|_2^2 + \frac{\rho_3}{2}\|\mathbf{A} - \mathbf{P} + \mathbf{\Lambda_3}\|_2^2, \\ \mathbf{Z^{t+1}} := \arg\min_{\mathbf{Z}\geq 0} \imath(\mathbf{Z}) + \frac{\rho_2}{2}\|\mathbf{A} - \mathbf{Z} + \mathbf{\Lambda_2}\|_2^2, \\ \mathbf{P^{t+1}} := \arg\min_{\mathbf{P}} \frac{\rho_3}{2}\|\mathbf{A} - \mathbf{P} + \mathbf{\Lambda_3}\|_2^2, \ s.t.\,\mathbf{PP}^T = \mathbf{I}_k, \end{cases}$$

$$\begin{cases} \mathbf{\Lambda_2^{t+1}} = \mathbf{\Lambda_2} + \mathbf{A} - \mathbf{Z}, \\ \mathbf{\Lambda_3^{t+1}} = \mathbf{\Lambda_3} + \mathbf{A} - \mathbf{P}. \end{cases}$$

$$\tag{15}$$

In the next steps, for each primal variable, we solve the corresponding minimization problem. Notice that another choice of auxiliary variables and a splitting scheme will lead to different subproblems.

A minimization step: The first subproblem is an unconstrained convex optimization problem that has an analytical expression that can be obtained by setting the first-order derivative of $\mathcal{L}_A$ with respect to $\mathbf{A}$ to zero.

After taking the first-order derivative and grouping similar terms together, the gradient is given by:

$$\frac{\partial \mathcal{L}_\mathbf{A}}{\partial \mathbf{A}} = (\mathbf{M}^T\mathbf{M} + (\rho_2 + \rho_3)\mathbf{I})\mathbf{A} + 2\lambda\mathbf{AL}$$
$$+ (\mathbf{M}^T\mathbf{Y} + \rho_2(Z - \mathbf{\Lambda_2}) + \rho_3(P - \mathbf{\Lambda_3})) = 0. \tag{16}$$

Eq. 16 has the form of the *Sylvester's* equation that is given by: $\mathbf{CA} + \mathbf{AD} = \mathbf{E}$, where:

$$\begin{cases} \mathbf{C} = \mathbf{M}^T\mathbf{M} + (\rho_2 + \rho_3)\mathbf{I}, \\ \mathbf{D} = 2\lambda\mathbf{L}, \\ \mathbf{E} = \mathbf{M}^T\mathbf{Y} + \rho_2(Z - \mathbf{\Lambda_2}) + \rho_3(P - \mathbf{\Lambda_3}), \end{cases} \tag{17}$$

which can be reformulated using the Kronecker sum and the vectorization operator in the form [15, Chapter 4]:

$$(\mathbf{C} \oplus \mathbf{D})Vect(\mathbf{A}) = Vect(\mathbf{E}), \tag{18}$$

where *Vect* symbol denotes a matrix vectorization operator, and the symbol $\oplus$ denotes the Kronecker sum of two matrices, which is defined by: $\mathbf{C} \oplus \mathbf{D} = \mathbf{I}_k \otimes \mathbf{C} + \mathbf{L} \otimes \mathbf{I}_n$, where the $\otimes$ symbol denotes the Kronecker product.

The final form in Eq. 18 is a least square problem that has a straightforward analytic solution given by:

$$Vect(\mathbf{A}) = (\mathbf{C} \oplus \mathbf{D})^{-1}Vect(\mathbf{E}) \tag{19}$$

Z minimization step: The second subproblem has a closed-form solution given by:

$$\mathbf{Z^{t+1}} = \max(\mathbf{A} + \mathbf{\Lambda_2}, 0) \tag{20}$$

P minimization step: The third subproblem is a constrained quadratic problem that has a closed-form solution [22, 20] given by:

$$\mathbf{P} = \mathbf{UI}_{k\times n}\mathbf{V}^T, \tag{21}$$

where $\mathbf{U} \in \mathbb{R}^{k\times k}$, $\mathbf{V} \in \mathbb{R}^{n\times n}$ are the left and the right singular vectors of the matrix $\mathbf{F} = \mathbf{A} + \mathbf{\Lambda_3}$ respectively, i.e. $\mathbf{F} = \mathbf{UDV}^T$.

Algorithm 3 outlines the needed iterations for solving Eq. 8 to update $\mathbf{A}$. As stopping criterion, we consider $r_2 = \max\{\|r_Z^i\| - \|r_Z^{i-1}\|)/\|r_Z^{i-1}\|, \|r_P^i\| - \|r_P^{i-1}\|)/\|r_P^{i-1}\|\}$, where $r_Z = \mathbf{A} - \mathbf{Z}$, and $r_P = \mathbf{A} - \mathbf{P}$.

5. Experimental results

For the sake of results reproducibility, we report in this section all the details about the used dataset, the evaluation metric, the hyper-parameters setting, and the obtained results. The implementation code will be available here [1].

5.1. Writing ink HSD image Database

Experiments are carried out on the HS multi-ink handwritten image database [2], which is a dataset of real handwritten notes scanned with an HS imaging system of 33

---

[1] https://github.com/arahiche/GONMF-ADMM
[2] http://www.csse.uwa.edu.au/%7Eajmal/databases.html

Algorithm 3 : Update of A using ADMM

Input: $\mathbf{Y}$, $\mathbf{M}$, $\mathbf{A}$, ρ_2, ρ_3, *Tol*, *niter*

   Initialization: $\boldsymbol{\Lambda_2} = 0$, $\boldsymbol{\Lambda_3} = 0$, $\mathbf{P} = \mathbf{A}$, $\mathbf{Z} = \mathbf{A}$

1: **for** $i = 1$ to *niter* **do**
2:    Calculate $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{E}$ using Eq. 17, and update $\mathbf{A}$ by solving:
     $Vect(\mathbf{A}) \leftarrow (\mathbf{C} \oplus \mathbf{D})^{-1} Vect(\mathbf{E})$, (see Eq. 19).
3:    $\mathbf{Z^{t+1}} \leftarrow \max(\mathbf{A} + \boldsymbol{\Lambda_2}, 0)$.
4:    Calculate $\mathbf{F} = \mathbf{A} + \boldsymbol{\Lambda_3}$ and update $\mathbf{P}$ with :
     $\mathbf{P} = \mathbf{U} \mathbf{I}_{k \times n} \mathbf{V}^T$, (see Eq.21).
5:    $\boldsymbol{\Lambda_2} \leftarrow \boldsymbol{\Lambda_2} + \mathbf{A} - \mathbf{Z}$.
6:    $\boldsymbol{\Lambda_3} \leftarrow \boldsymbol{\Lambda_3} + \mathbf{A} - \mathbf{P}$.
7:    **if** $(r_2^i < Tol$ **then**
8:       break..
9:    **end if**
10: **end for**
11: **return** $\mathbf{A}, r_2$

---

bands. The same notes were written by 7 subjects using 10 different inks, including 5 varieties of blue inks and 5 types of black inks. As described in [18, 19], different combinations of writing ink images were produced by merging portions from images of samples written by each subject with different inks in different ratios. We note that blue and black ink samples were not inter-mixed because their differentiation is trivial, even with a naked eye. Also, authors choose not to mix inks among different subjects to avoid any possible bias on the results. Hereafter in this section, we use the symbol C_{ij} to indicate a two-inks HSD sample composed of a proportion from two inks i and j.

5.2. Evaluation metric

To measure the quality of ink differentiation, we considered the same accuracy metric adopted in [12, 18, 19, 1], which is defined as:

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \qquad (22)$$

where TP, FP, and FN denote the True positives, the False positives and the False negatives prediction respectively. This metric calculates the number of correctly labeled pixels of ink_i divided by the total number of pixels of that ink. Furthermore, only pixels that correspond to inks are considered. Therefore, pixels that correspond to the background (paper) are excluded from this measure.

5.3. Hyperparameters setting

For the initialization of our GONMF-ADMM model, we observed that a deterministic initialization works better than a stochastic (random) initialization. Thus, we used a deterministic initialization based on SVD-decomposition.

Therefore, each image decomposition was run only one single time, this because the orthogonality constraint leads to unique NMF decomposition. The hyperparameters values of our model are chosen using the grid search method. In our experiments, the following sets of values $\{\rho_1 = 10^{-4}, \rho_2 = 0.1, \rho_3 = 10^3, \lambda = 500\}$, and $\{\rho_1 = 10^{-3}, \rho_2 = 10^{-5}, \rho_3 = 10^3, \lambda = 500\}$ found to be optimal, by which our GONMF-ADMM model achieves highest accuracies. As stopping criterions, we set the maximum number of iterations to $niter = 50$ for the outer-loop and $niter = 40$ for the inner-loop respectively, and the tolerance $tol = 10^{-5}$ for the three algorithms, where $r = \max\{r_1, r_2\}$. For the rank selection, which is still an open research problem, we manually set the number of inks. We note that, in this study, we did not investigate the automatic rank selection issue. However, other existing techniques can be used to estimate the number of inks present on each image, such as the Elbow method, and the principal eigenvalues of the singular value decomposition (SVD). Finally, the setting of the graph regularization is as follows. The number of nearest neighbors is set to 10. The weight matrix is calculated using the Heat kernel weighting defined by [3]: $W_{ij} = e^{\frac{\|x_j - x_i\|^2}{\sigma}}$, where we set $\sigma = 1$.

5.4. Results and discussion

First, in Fig. 2, we qualitatively compare the inks isolation results obtained by the unmixing based approach (HySim+MVES) proposed by [1] and our GONMF-ADMM approach in terms of abundance maps. In [1], authors combined the HS subspace identification by minimum error algorithm (HySime) [4], used to estimate the number of inks (endmembers) present in the image, and the minimum volume enclosing simplex (MVES) algorithm [7] for HSD images unmixing.

(a) HySime + MVES (taken from [1])   (b) GONMF-ADMM
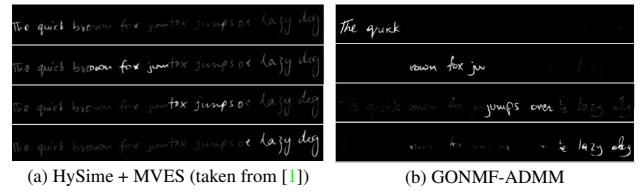
Figure 2: Qualitative comparison of abundance maps obtained by: (a) the unmixing MVES method [1], and (b) the proposed GONMF-ADMM algorithm. The two samples belong to different writers but contain the same type of ink, which do not have any effect on the final result.

As shown in Fig. 2, the proposed GONMF-ADMM approach allows better inks separation. Each one of the four inks (endmembers) present in the original HSD image is well located and isolated in the output images. In contrast, the output of the unmixing MVES algorithm shows that inks are still present as a mixture. We can clearly see that inks

are overlapping in output abundance images.

The quantitative evaluation given in Fig. 3 illustrates the accuracy achieved by our approach for the segmentation of HSD images produced by each writer. In this scenario, all samples contain a combination of two inks. Exceptionally for quantitative assessment, the output images are thresholded to obtain the required image format for comparison with the ground truth images.
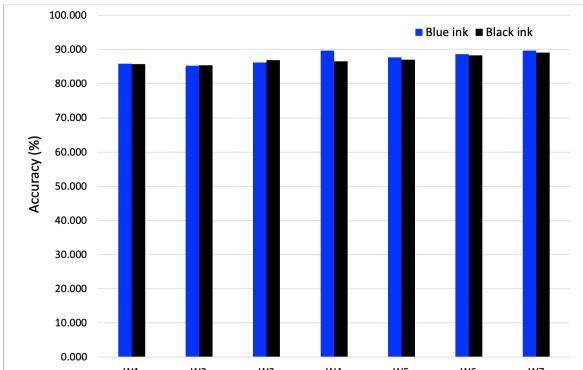

Figure 3: Qualitative evaluation of the ink mismatch detection task with two inks combinations. W_k denotes a dataset of samples of the same writer.

A second qualitative illustration is given in Fig. 4, in which we qualitatively compare the quality of the multi-inks segmentation results obtained by our method with the ground-truth images provided by authors of HS dataset. Experiments are carried out on images with different inks combinations and proportions.

Fig. 4 shows the effectiveness of our approach for the separation between inks regardless of the number of inks or the proportion of inks present in the image. All samples show that our approach is able to cluster pixels of different inks together successfully. Moreover, segmentation's performance is stable and writer-independent. Despite the fact that some samples contain a very unbalanced ratio of inks, the performance achieved by our approach is high.

Finally, we compare our approach's performance on two inks mismatch detection task against four state-of-the-art methods. Table 1 illustrates the results of our GONMF-ADMM method alongside the results of the K-means based clustering method for HS images (HSI-All) [18], and three bands-selection based techniques, namely, the ℓ_{21} sparse principal components analysis (L$_{21}$−SPCA), the sequential forward band selection (SFBS), and the joint sparse band selection approaches (JSBS) reported in [19].

As shown in Table 1, our approach performs better for most cases. A slight improvement is obtained for the combinations C_{12}, C_{13}, and C_{14} of the blue inks, and the combinations C_{12}, C_{13}, C_{14}, and C_{15} of the black inks. The improvement is significant for the combinations C_{15}, C_{23}, and C_{24} of the blue inks. However, for the combinations

Table 1: Qualitative comparison of two inks segmentation results. Average accuracies obtained for all writers.

Ink	Fold	HSI-All	SFBS	ℓ_{21}-SPCA	JSBS	GONMF-ADMM
Blue	C_{12}	99.8	99.5	99.8	99.8	**99.9**
	C_{13}	61.0	98.2	99.5	99.5	**99.8**
	C_{14}	99.7	**99.9**	**99.9**	**99.9**	**99.9**
	C_{15}	62.3	83.3	59.4	73.6	**99.7**
	C_{23}	50.0	54.3	59.0	59.0	**88.3**
	C_{24}	40.0	58.0	44.1	44.2	**77.9**
	C_{25}	98.9	98.6	**99.1**	**99.1**	76.7
	C_{34}	45.5	45.6	91.6	**93.0**	69.8
	C_{35}	96.9	95.9	98.9	**99.0**	84.8
	C_{45}	99.7	**99.8**	**99.8**	**99.8**	78.9
Black	C_{12}	99.9	**100.0**	**100.0**	**100.0**	**100.0**
	C_{13}	96.2	80.4	94.5	98.7	**99.8**
	C_{14}	99.2	96.0	99.0	99.6	**99.9**
	C_{15}	98.9	97.5	99.2	99.7	**99.8**
	C_{23}	69.5	84.7	**91.6**	**91.6**	85.7
	C_{24}	61.9	79.9	**89.0**	**89.0**	75.4
	C_{25}	50.2	82.8	**86.1**	**86.1**	75.8
	C_{34}	50.3	58.7	60.9	69.3	**70.5**
	C_{35}	60.8	67.1	77.2	84.4	**84.5**
	C_{45}	60.4	60.6	70.7	72.2	**78.9**
Mean	**Blue**	75.4	83.3	85.1	86.7	**87.6**
	Black	74.7	81.8	86.8	**89.0**	87.0

C_{25} in blue inks and C_{23}, C_{24}, and C_{25} in black inks the accuracy declined. A possible reason could be that the hyperparameter setting is not optimal for these combinations, and therefore, more fine-tuning is required to obtain better results.

Unlike features selection based approaches, i.e., ℓ_{21}-SPCA, SFBS, SBFI, JSBS, and the unmixing (MVES) method, in general, our GONMF-ADMM approach does not require any further post-processing after the decomposition. If the chosen hyperparameters are optimal, then the proposed GONMF factorization will be able to extract the latent components targeted with high accuracy.

6. Conclusion

Ink mismatch detection plays an important role in the routine examination of questioned documents. HS imagery provides valuable information about the different materials that might exist in the same document scene and can not be distinguished visually. Graph orthogonal nonnegative matrix factorization model combines both spectral and spatial information and allows unsupervised analysis and clustering of different materials in HSD images. This work opens a new trend for forgery detection in questioned documents using nonnegative matrix factorization based methods and spectral imagery. In future work, we will address the issue of automatic estimation of the number of inks and the hyperparameters selection.

Acknowledgment

(a) HSD sample of blue ink (C_{12}) (b) HSD sample of black ink (C_{12})
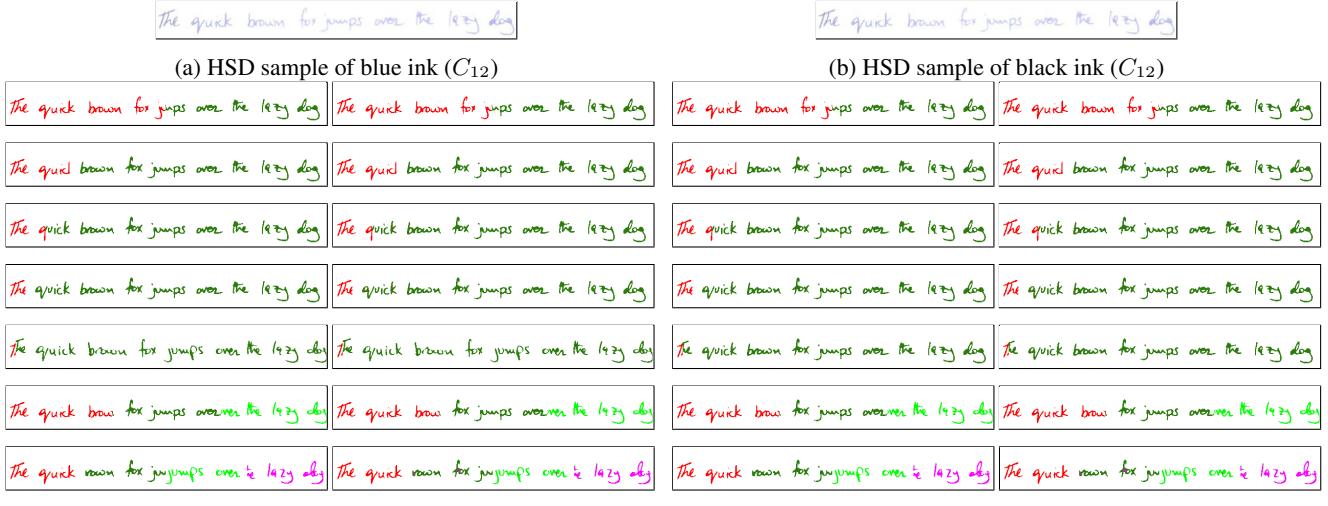
(c) Segmentation results of blue ink with different combinations. (d) Segmentation results of black ink with different combinations.

Figure 4: An illustration of ink mismatch detection on blue and black ink images with different combinations and proportions. a) and b) raw HSD images, ground truth images are shown in the left column of (c) and (d), respectively, and the segmented images of blue and black inks are in the right column of (c) and (d). A pseudo-colors coding is used to represent the pixels of each variant of ink (ink 1 in red, ink 2 in green, ink 3 in bright green, and ink 4 in pink).

References

[1] Asad Abbas, Khurram Khurshid, and Faisal Shafait. Towards automated ink mismatch detection in hyperspectral document images. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1229–1236. IEEE, 2017. 2, 6

[2] VN Aginsky. Forensic examination of slightly soluble ink pigments using thin-layer chromatography. *Journal of Forensic Science*, 38(5):1131–1133, 1993. 1

[3] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*, pages 585–591, 2002. 6

[4] José M Bioucas-Dias and José MP Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008. 6

[5] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011. 3

[6] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1548–1560, 2010. 3

[7] Tsung-Han Chan, Chong-Yung Chi, Yu-Min Huang, and Wing-Kin Ma. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Transactions on Signal Processing*, 57(11):4418–4432, 2009. 6

[8] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997. 3

[9] Amir Abbas Davari, Armin Häberle, and Christian Riess. Sketch layer separation in multi-spectral historical document images. In *International Symposium on Digital Humanities: Book of Abstracts*, pages 29–32, 2016. 1

[10] Chris Ding, Tao Li, Wei Peng, and Haesun Park. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 126–135, 2006. 3

[11] GJ Edelman, E Gaston, TG Van Leeuwen, PJ Cullen, and MCG Aalders. Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic science international*, 223(1-3):28–39, 2012. 1

[12] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136, 2015. 6

[13] Cédric Févotte and Jérôme Idier. Algorithms for nonnegative matrix factorization with the β-divergence. *Neural computation*, 23(9):2421–2456, 2011. 2

[14] Douglas Goltz, Michael Attas, Gregory Young, Edward Cloutis, and Maria Bedynski. Assessing stains on historical documents using hyperspectral imaging. *Journal of cultural heritage*, 11(1):19–26, 2010. 2

[15] Roger A Horn and Charles R Johnson. *Topics in matrix analysis*. Cambridge university press, 1994. 5

[16] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2014–2039, 2011. 3

[17] Muhammad Jaleed Khan, Khurram Khurshid, and Faisal Shafait. A spatio-spectral hybrid convolutional architecture for hyperspectral document authentication. In *2019 15th In-*

ternational Conference on Document Analysis and Recognition. IEEE, 2019. 2

[18] Zohaib Khan, Faisal Shafait, and Ajmal Mian. Hyperspectral imaging for ink mismatch detection. In *2013 12th International Conference on Document Analysis and Recognition*, pages 877–881. IEEE, 2013. 1, 2, 6, 7

[19] Zohaib Khan, Faisal Shafait, and Ajmal Mian. Automatic ink mismatch detection for forensic document analysis. *Pattern Recognition*, 48(11):3615–3626, 2015. 2, 6, 7

[20] Rongjie Lai and Stanley Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014. 3, 5

[21] Zhipei Luo, Faisal Shafait, and Ajmal Mian. Localized forgery detection in hyperspectral document images. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 496–500. IEEE, 2015. 2

[22] Jonathan H Manton. Optimization algorithms exploiting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635–650, 2002. 5

[23] Aythami Morales, Miguel A Ferrer, Moises Diaz-Cabrera, Cristina Carmona, and Gordon L Thomas. The use of hyperspectral analysis for ink identification in handwritten documents. In *2014 International Carnahan Conference on Security Technology (ICCST)*, pages 1–5. IEEE, 2014. 2

[24] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994. 2, 3

[25] Adam Polak, Timothy Kelman, Paul Murray, Stephen Marshall, David JM Stothard, Nicholas Eastaugh, and Francis Eastaugh. Hyperspectral imaging combined with data classification techniques as an aid for artwork authentication. *Journal of Cultural Heritage*, 26:1–11, 2017. 1

[26] Filippo Pompili, Nicolas Gillis, P-A Absil, and François Glineur. Two algorithms for orthogonal nonnegative matrix factorization with application to clustering. *Neurocomputing*, 141:15–25, 2014. 3

[27] Abderrahmane Rahiche, Rachid Hedjam, Somaya Almaadeed, and Mohamed Cheriet. Historical documents dating using multispectral imaging and ordinal classification. *Journal of Cultural Heritage*, 2020. 1

[28] Roman Sandler and Michael Lindenbaum. Nonnegative matrix factorization with earth mover's distance metric for image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1590–1602, 2011. 2

[29] Carolina S Silva, Maria Fernanda Pimentel, Ricardo S Honorato, Celio Pasquini, José M Prats-Montalbán, and Alberto Ferrer. Near infrared hyperspectral imaging for forensic analysis of document forgery. *Analyst*, 139(20):5176–5184, 2014. 2