



Automatic ink mismatch detection for forensic document analysis



Zohaib Khan*, Faisal Shafait, Ajmal Mian

School of Computer Science and Software Engineering, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

ARTICLE INFO

Article history:

Received 9 December 2013

Received in revised form

22 December 2014

Accepted 3 April 2015

Available online 15 April 2015

Keywords:

Hyperspectral imaging

Multispectral document examination

Band selection

ABSTRACT

A key aspect of handwritten document examination is to investigate whether some portion of the text was modified, altered or forged with a different pen. This paper demonstrates the use of hyperspectral imaging for ink mismatch detection in a handwritten note. We propose a novel joint sparse band selection technique that selects informative bands from hyperspectral images for accurate ink mismatch detection. We have developed an end-to-end camera-based hyperspectral document imaging system and collected a database of handwritten notes which has been made publicly available. Algorithmic solutions are presented to handle specific challenges in camera-based hyperspectral document imaging. Extensive experiments show that the proposed band selection method selects the most informative bands and improves average accuracy up to 15%, compared to using all bands.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The human eye is sensitive to light in the visible range to distinguish materials based on their color [1]. However, humans are unable to distinguish between two apparently similar colors [2] due to the trichromatic nature of the human visual system. When a document is manipulated with the intention of forgery or fraud, the modifications are often done in such a way that they are hard to catch with a naked eye. The forger not only tries to emulate the handwriting of the original writer, but also uses a pen that has a visually similar ink compared to the rest of the note. Hence, analysis of inks is of critical importance in questioned document examination.

The outcome of ink analysis can potentially lead to the determination of forgery, fraud, backdating and ink age. Of these, one of the most important tasks is to discriminate between different inks which we term as *ink mismatch detection*. There are two main approaches to distinguish inks, destructive and non-destructive examination. Chemical analysis such as thin layer chromatography (TLC) [3] is a destructive test which separates a mixture of inks into its constituents via capillary action. However, TLC compromises the originality of a sample, of a forensic evidence. Furthermore, observing any noticeable differences in a chromatograph incurs considerable time.

An alternative approach is to employ spectral imaging to differentiate apparently similar inks. Spectral imaging captures subtle differences in the inks which is valuable for mismatch detection as shown in Fig. 1. A hyperspectral image (HSI) is a series

of discrete narrow-band images in the electro-magnetic spectrum. In contrast to a three channel RGB image, an HSI captures finer detail of a scene in the spectral dimension as shown in Fig. 2. Hyperspectral imaging has recently emerged as an efficient non-destructive tool for detection and identification of forensic traces such as bloodstain analysis and latent print analysis [4]. It is also useful to questioned document examination for ink discrimination, document age estimation and restoration of historical documents [5–8].

Brauns and Dyer [9] developed a hyperspectral imaging system for forgery detection in potentially fraudulent documents in a non-destructive manner. They prepared written documents with blue, black and red inks and later introduced alterations with a different ink of the same color. Using fuzzy clustering, the ink spectra were grouped into one or more cluster indicating their degree of association. They qualitatively showed that the inks can be separated into two different classes. The absence of quantitative results and slow imaging process collectively limit the applicability of their system to practical ink mismatch detection. In contrast, our approach is quantitative instead of qualitative-subjective analysis.

A relatively improved hyperspectral imaging system for the analysis of historical documents in archives was developed by Padoan et al. [10]. Their use of narrowband tunable light source reduced the chances of damage to a document due to excessive heat generated by a strong broadband white light source. However, its extremely slow acquisition time (about 15 min) [11] resulted in prolonged exposure. Therefore, the benefit gained by a tunable light source may be nullified and the productivity of the system in terms of the number of documents captured is reduced. Our proposed system captures hyperspectral images in only a fraction of that time using an *electronically tunable filter* which is fast, precise and has no moving parts.

* Corresponding author.

E-mail address: zohaib.khan@uwa.edu.au (Z. Khan).

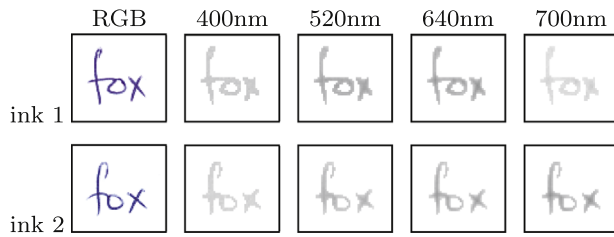


Fig. 1. The images highlight the discrimination of inks at different wavelengths offered by spectral imaging. A word written in two different blue inks is shown in this example. Observe that the two inks appear similar at short wavelength and gradually appear different at longer wavelengths. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

In commercial hyperspectral document imaging systems [4,12] the examiner needs to select a suspected portion of a note for ink mismatch analysis. The examiner has to search through hundreds of combinations of the different wavelengths to visually identify the differences in inks, which is laborious. For instance, to analyze a 33 band HSI in an exhaustive search, the total number of band combinations is of the order of $\approx 10^{10}$, which is not feasible in time critical scenarios. This procedure is not required in our proposed *automatic* document analysis approach. Note that acquisition of all bands is time consuming and limits the number of documents that can be scanned in a given time. Moreover, the resulting data is huge and some bands with low energy contain significant system noise. Therefore, it is desirable to select the most informative subset of bands, thereby reducing the acquisition time, and increase the accuracy by getting rid of the noisy bands. Since, hyperspectral images are densely sampled along the spectral dimension, the neighboring bands are highly correlated. This redundancy makes hyperspectral images a good candidate for sparse representation and band selection [13].

Hedjam and Cheriet [14] proposed a hyperspectral band selection algorithm based on graph clustering. They created a band adjacency graph in which the nodes represent the bands and the edges represent the similarity weights between bands. Markov clustering was then used to form distinct clusters of highly correlated bands. In contrast, Martinez et al. [15] proposed a hierarchical clustering structure which minimizes certain distance measures (KL-divergence, MI, etc.) to reduce redundancy in adjacent bands. In both approaches, clustering results in groups of highly correlated bands, and a new challenge of selecting a representative band which accounts for maximum information within each cluster arises. Unlike clustering based methods, sparse representation explicitly indicates a selection of bands which is representative of the maximum information in the complete hyperspectral data. One such approach is to optimize a subspace projection while simultaneously selecting features in a supervised manner [16].

We propose joint sparse PCA (JSPCA) that computes a PCA basis by explicitly removing the non-informative bands in an unsupervised manner. The joint sparsity ensures that all basis vectors share the same sparsity structure whereas the complete hyperspectral data can be represented by a sparse linear combination of the bands. Our work is most similar to Xiaoshuang et al. [17], who solved a linear system for sparse PCA with an $\ell_{2,1}$ constraint for joint sparsity. However, a major drawback of their closed form regression is the instability of solution at high joint sparsity constraint. In contrast, our proposed algorithm produces jointly sparse solutions due to a robust iterative optimization scheme. We demonstrate the joint sparse band selection (JSBS) algorithm for hyperspectral ink mismatch detection and show that it selects fewer bands with higher accuracy compared to the state-of-the-art.

This paper is a significant extension of our preliminary work on ink mismatch detection [18]. The main contributions of the work are:

- A novel *joint sparse PCA (JSPCA)* algorithm for dimensionality reduction and feature selection.
- A *joint sparse band selection (JSBS)* technique for automatic ink mismatch detection.
- An *efficient hyperspectral document imaging system* and collection of a *new public database* of writing ink hyperspectral images.¹

The rest of this paper is organized as follows. In Section 2, we present the proposed ink mismatch detection methodology. In Section 3, we describe the database specifications, acquisition and normalization. Section 4 provides details of the experimental setup, evaluation protocol, and analysis of the results. The paper is concluded with a discussion in Section 5.

2. Ink mismatch detection

Ink mismatch detection is based on the fact that the same inks exhibit similar spectral responses whereas different inks are spectrally dissimilar [18]. We assume that the spectral responses of the inks are independent of the writing styles of different subjects (which is a spatial characteristic). Thus, unlike works that identify hand writings by the ink-deposition traces [19], our work solely focuses on the spectral responses for ink discrimination. In the proposed ink mismatch detection framework, the initial objective is to segment handwritten text from the paper. The next task is to select features (bands) from the ink spectra by the proposed band selection technique. Finally, the class membership of each ink pixel is determined by clustering of the ink spectral responses using selected features.

Notations: In the following text, a scalar is denoted as lowercase letter (i), a vector as a lowercase letter in bold font (\mathbf{x}), and a matrix as an uppercase letter in bold font (\mathbf{X}). X_{ij} is the (i, j)th element of a matrix. \mathbf{x}^i is the i th row and \mathbf{x}_j is the j th column of a matrix. $\mathbf{X}_{\cdot s}$ is the submatrix obtained by indexing the columns of \mathbf{X} by the index set S ($\mathbf{X}_{\cdot s}$ indexes the rows). All vectors are treated as column vectors.

2.1. Handwritten text segmentation

Consider a three dimensional HSI $\mathbf{I} \in \mathbb{R}^{x \times y \times p}$, where (x, y) are the number of pixels in spatial dimension and p is the number of bands in the spectral dimension. The objective is to compute a binary mask $\mathbf{M} \in \mathbb{R}^{x \times y}$ which associates each pixel to the foreground or background. The ink pixels (text) make up the foreground and the blank area of the page is the background. A global image thresholding method, such as the Otsu [20] is ineffective because of the non-uniform illumination over the document (Fig. 3(a)). A local image thresholding method such as Sauvola and Pietikäinen [21] with an efficient integral image based implementation [22] more effectively deals with such illumination variations. The Sauvola's method generates a binary mask according to

$$M_{ij} = \begin{cases} 1 & \text{if } I_{ij} > \mu_{ij} \left(1 + \kappa \left(\frac{\sigma_{ij}}{r-1} \right) \right) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

¹ UWA Writing Ink Hyperspectral Image Database <http://www.csse.uwa.edu.au/~ajmal/databases.html>.

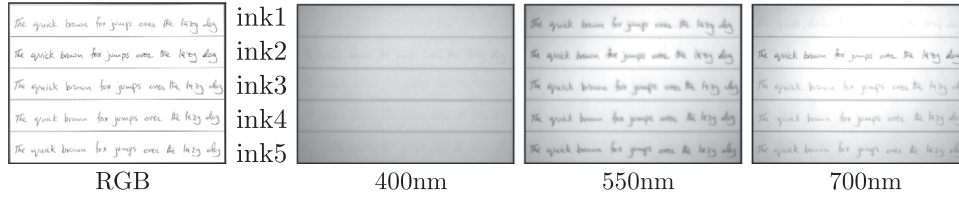


Fig. 2. A sample document captured in RGB with a flatbed scanner (left) and selected bands captured by our hyperspectral document imaging system (right). Notice how the photometric properties of the five inks vary when exposed to narrow bands of light.

where

$$\mu_{ij} = \frac{1}{w^2} \sum_{a=-(w-1)/2}^{(w+1)/2} \sum_{b=-(w-1)/2}^{(w+1)/2} I(i+a, j+b, c)$$

$$\sigma_{ij} = \sqrt{\frac{1}{w^2} \sum_{a=-(w-1)/2}^{(w+1)/2} \sum_{b=-(w-1)/2}^{(w+1)/2} I(i+a, j+b, c) - \mu_{ij}^2}$$

where (μ_{ij}, σ_{ij}) are the mean and standard deviation of a $w \times w$ patch centered at pixel (i, j) . The factors κ and r jointly scale the standard deviation term between $(0,1)$. The value of r is fixed to the maximum possible standard deviation of the patch which is 128 for an 8-bit image. We empirically found that a patch size $w \times w = 32 \times 32$ and $\kappa = 0.15$ give good segmentation results.

Independently applying thresholding to the p bands will result in p different masks which requires reduction to a single mask. Since the intensity of foreground pixels in different spectral bands is variable, the binarization results of each band are not identical. One solution is to merge multiple band binarization masks to get a single binary mask [23]. An alternative is to apply threshold to a single representative band and propagate the same mask to all the bands. This approach is valid if all bands are spatially aligned (which is true in case of stationary document images). We resort to the latter strategy since choosing a representative band for binarization is straightforward. For instance, the band with the highest contrast (640 nm band, $c=25$) can be chosen (Fig. 3(b)). We observed that selecting any other band in the range [620 nm, 660 nm] resulted in the same mask for various colored inks. Fig. 3(c) and (d) shows the results of Otsu and Sauvola binarization methods. Observe that Sauvola's method performs better, whereas Otsu's method returns an inaccurate mask due to non-uniform illumination.

2.2. Sequential forward band selection

Feature selection aims to find the feature subset that maximizes a certain performance criteria, generally accuracy [24]. In this section, we aim to find a subset of bands which maximizes ink mismatch detection accuracy (defined later in Section 2.4) by a sequential forward band selection (SFBS) technique. The procedure is described in Algorithm 1. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ be the matrix of n normalized spectral response vectors, $\mathbf{x} \in \mathbb{R}^p$, each corresponding to one ink pixel. In the first step, the mismatch detection accuracy of each of the p bands is computed individually. The band with the highest accuracy is added to the selected band set, S and removed from the remaining band set \mathcal{R} . From the remaining $p-1$ bands in \mathcal{R} , one band at a time is combined with S and the accuracy is observed on the new set. If the accuracy increases, then the added band which maximized accuracy combined with the previously selected bands is retained in S and removed from \mathcal{R} . This process continues until adding another band to S reduces

accuracy. In case all bands are added to S (which is a rare occurrence), the algorithm will automatically converge.

Algorithm 1. Sequential forward band selection.

Input: $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{y} \in \mathbb{Z}^n$
Initialize: $q \leftarrow 1$, $\text{converge} \leftarrow \text{false}$, $S \leftarrow \emptyset$, $\mathcal{R} \leftarrow \{1, 2, \dots, p\}$
for $j=1$ **to** p **do** ▷ Step 1: find best individual band
 $a_j \leftarrow \text{IMDaccuracy}(\mathbf{x}_j, \mathbf{y})$
end for
 $a^* \leftarrow \max_j(a_j)$, $j^* \leftarrow \arg\max_j(a_j)$ ▷ best accuracy
and band index
 $S \leftarrow S \cup j^*$, $\mathcal{R} \leftarrow \mathcal{R} \setminus j^*$ update band sets
while $q \leq p \wedge \neg \text{converge}$ **do** ▷ Step 2: sequentially add remaining bands
for $k=1$ **to** $p-q$ **do** ▷ loop over remaining bands
 $\mathcal{T} \leftarrow S \cup \mathcal{R}(k)$ ▷ add k th band to the temporary set
 $a_k \leftarrow \text{IMDaccuracy}(\mathbf{X}_{\cdot \mathcal{T}}, \mathbf{y})$ Section 2.4
end for
 $b^* \leftarrow \max_k(a_k)$, $k^* \leftarrow \arg\max_k(a_k)$
if $b^* > a^*$ **then** ▷ Step 3: check if accuracy improved
 $S \leftarrow S \cup k^*$, $\mathcal{R} \leftarrow \mathcal{R} \setminus k^*$ ▷ update band sets
 $q \leftarrow q+1$, $a^* \leftarrow b^*$
else ▷ accuracy did not improve
 $\text{converge} \leftarrow \text{true}$
end if
end while
Output: $S \in \mathbb{Z}^q$

2.3. Joint sparse band selection

Principal component analysis (PCA) is one of the most commonly used tools for hyperspectral data interpretation, dimensionality reduction and visualization [25]. PCA transforms spectral data from input feature space to an orthogonal feature space which corresponds to the direction of maximal variance of the spectral bands. It is particularly useful for reducing HS data dimensionality by projection to a subspace defined by the k most significant basis vectors ($k < p$). However, each principal component is obtained as a linear combination of *all* the bands making hyperspectral data interpretation difficult, and essentially requires all bands to be acquired.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ be the matrix of n normalized spectral response vectors, $\mathbf{x} \in \mathbb{R}^p$, each corresponding to an ink pixel. Assume that the sample mean $\bar{\mathbf{x}} \in \mathbb{R}^p$ has been subtracted from all n observations so that the columns of \mathbf{X} are centered. Generally, a PCA basis can be computed by singular value decomposition of

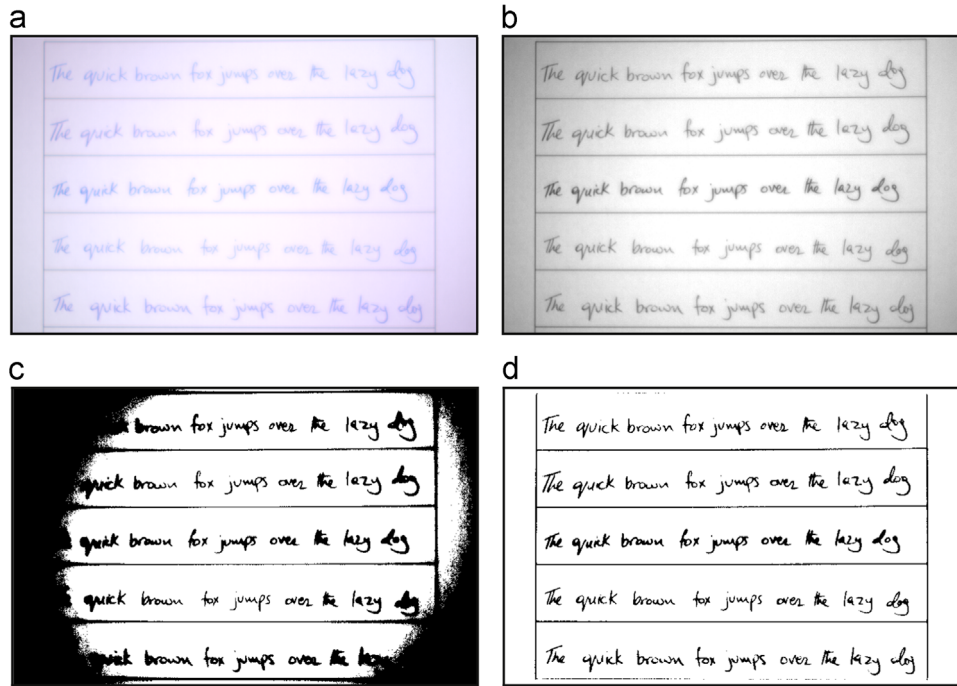


Fig. 3. Hyperspectral document image binarization. Notice the high energy in the center of the paper compared to the edges. Local image thresholding is far superior to global image thresholding because of invariance to non-uniform illumination.

the data matrix:

$$\mathbf{X} = \mathbf{USV}^T \quad (2)$$

where $\mathbf{V} \in \mathbb{R}^{p \times p}$ are the PCA basis vectors (loadings) and \mathbf{S} is the diagonal matrix of eigenvalues. The basis \mathbf{V} is orthonormal such that $\mathbf{v}_i^T \mathbf{v}_j = 0 \forall i \neq j$ and $\mathbf{v}_i^T \mathbf{v}_j = 1 \forall i = j$. If \mathbf{X} is low rank, it is possible to significantly reduce its dimensionality by using the $k < p$ significant basis vectors. The projection of data \mathbf{X} upon the orthonormal basis \mathbf{V} gives the first k principal components (scores). Alternatively, PCA can also be formulated as a regression type optimization problem:

$$\underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{XAA}^T\|_F^2 \quad \text{subject to } \mathbf{A}^T \mathbf{A} = \mathbf{I}_k \quad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{A} \in \mathbb{R}^{p \times k}$ is an orthonormal basis $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$. Here, \mathbf{A} is equivalent to the first k columns of \mathbf{V} . Note that each principal component is derived from a linear combination of all p bands and α is typically non-sparse. In order to obtain a sparse PCA basis, a regularization penalty term may be included in the regression formulation (3). Inclusion of a sparse penalty may reduce the number of bands involved in each linear combination for obtaining the principal components. One way to obtain sparse basis vectors is by imposing the ℓ_0 constraint upon the regression coefficients (basis vectors):

$$\underset{\mathbf{A}, \mathbf{B}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{XBA}^T\|_F^2 + \lambda \sum_{j=1}^k \|\beta_j\|_0 \quad \text{subject to } \mathbf{A}^T \mathbf{A} = \mathbf{I}_k \quad (4)$$

where $\mathbf{B} \in \mathbb{R}^{p \times k}$ corresponds to the required sparse basis $\{\beta_1, \beta_2, \dots, \beta_k\}$. The ℓ_0 -norm regularization term penalizes the number of non-zero coefficients in β , whereas the loss term simultaneously minimizes the reconstruction error $\|\mathbf{X} - \mathbf{XBA}^T\|_F^2$. If λ is zero, the problem reduces to finding the ordinary PCA basis vectors, equivalent to (3). When λ is large, most coefficients of β_j will shrink to zero, resulting in sparsity as shown in Fig. 4(a).

Since each row of \mathbf{B} corresponds to a particular band, the sparsity within each row of \mathbf{B} may be used to find the relative importance of the bands. If all coefficients in a row turn out to be zero, the corresponding band will become irrelevant to the computation of

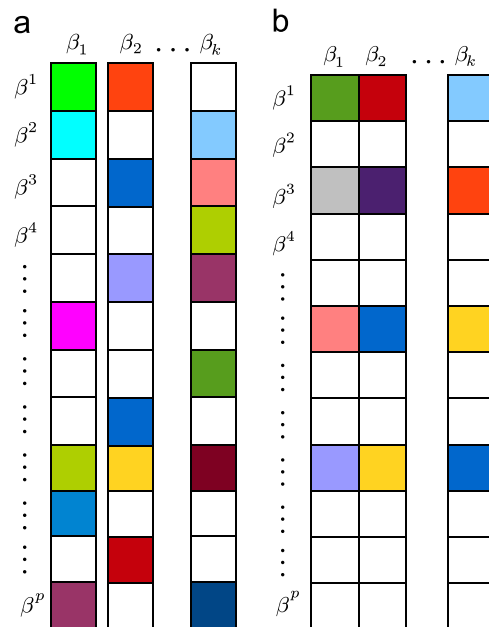


Fig. 4. The sparsity patterns are shown in pseudo-colors (white cells correspond to zero coefficients and shaded cells are non-zero coefficients). The simple sparsity is unconstrained with respect to the basis and therefore may end up using all bands. Joint sparsity penalizes the rows of the basis and explicitly uses a few bands.

the basis vectors. We therefore suggest a special type of sparsity which enforces most rows of \mathbf{B} to be zero, while the other rows may contain all non-zero coefficients. This type of sparsity among the basis vectors may be called joint sparsity (Fig. 4(b)). Note that the existing variants of sparse PCA [26–30] do not account for this type of joint sparsity, and therefore cannot be used for band selection. We propose a joint sparse PCA formulation which enforces the $\ell_{2,0}$ matrix norm on \mathbf{B} instead of the previously used ℓ_0 vector norm:

$$\underset{\mathbf{A}, \mathbf{B}}{\operatorname{arg. min}} \|\mathbf{X} - \mathbf{XBA}^T\|_F^2 + \lambda \|\|\beta^i\|_2\|_0 \quad \text{subject to } \mathbf{A}^T \mathbf{A} = \mathbf{I}_k \quad (5)$$

where β^i is the i th row of \mathbf{B} ($1 < i < p$). The minimization of ℓ_0 -norm over the ℓ_2 -norm of the rows of \mathbf{B} will force most of the rows to be null (for a sufficiently high value of λ). Although use of $\ell_{2,0}$ penalty gives a joint sparse basis, it makes the minimization non-convex and its solution is NP-hard. In general, the ℓ_0 norm minimization is relaxed to ℓ_1 norm [31] to reach an approximate solution. Therefore, we solve the following approximation to the joint sparse PCA formulation:

$$\underset{\mathbf{A}, \mathbf{B}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 + \lambda \sum_{i=1}^p \|\beta^i\|_2 \quad \text{subject to } \mathbf{A}^\top \mathbf{A} = \mathbf{I}_k \quad (6)$$

where the regularization term is often called $\ell_{2,1}$ norm of a matrix.

Although, the above formulation ensures a joint sparse PCA basis, simultaneous optimization of \mathbf{A} and \mathbf{B} makes the problem non-convex. If one of the two matrices is known, the problem becomes convex over the second unknown matrix. Hence, a locally convex solution of (6) can be obtained by iteratively minimizing \mathbf{A} and \mathbf{B} . Therefore, the joint sparse PCA formulation (6) is divided into two independent optimization problems. In the first optimization problem, \mathbf{A} is initialized with \mathbf{V} (from (2)) and the minimization under the joint sparsity constraint on \mathbf{B} is formulated as

$$\underset{\mathbf{B}}{\operatorname{argmin}} \|\mathbf{XA} - \mathbf{XB}\|_F^2 + \lambda \sum_{i=1}^p \|\beta^i\|_2 \quad (7)$$

which is similar to a multi-task regularized regression problem [32]:

$$\underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{Q} - \mathbf{XW}\|_F^2 + \psi(\mathbf{W}) \quad (8)$$

where $\mathbf{Q} = \mathbf{XA}$ is the response matrix, $\mathbf{W} = \mathbf{B}$ is the matrix of regression coefficients and ψ is any convex matrix norm. An optimization problem of the form (8) can be efficiently solved by proximal methods [33].

Theorem 2.1. *The regularization loss term $\|\mathbf{X} - \mathbf{XBA}^\top\|_F^2$ in (6) is equivalent to $\|\mathbf{XA} - \mathbf{XB}\|_F^2$ given $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_k$.*

Proof. The proof of this theorem is as follows from the minimization of the loss term $\|\mathbf{X} - \mathbf{XBA}^\top\|_F^2$

$$\mathbf{X} - \mathbf{XBA}^\top = \mathbf{C} \quad \mathbf{C} \in \mathbb{R}^{n \times p} \text{ is a residue matrix, } \mathbf{C} \neq \mathbf{0} \quad (9)$$

$$\mathbf{XA} - \mathbf{XBA}^\top \mathbf{A} = \mathbf{CA} \quad \text{multiplying by } \mathbf{A} \quad (10)$$

$$\mathbf{XA} - \mathbf{XB} = \mathbf{CA} \quad \text{since } \mathbf{A}^\top \mathbf{A} = \mathbf{I}_k \quad (11)$$

$$\mathbf{XA} - \mathbf{XB} = \mathbf{E} \quad \text{since } \mathbf{A} \text{ is fixed, } \mathbf{E} = \mathbf{CA} \text{ is also constant} \quad \square \quad (12)$$

Once the solution for \mathbf{B} is found by (7), the next step is to solve the second problem, i.e., optimizing with respect to \mathbf{A} . For a known \mathbf{B} the regularization penalty in (6) becomes irrelevant for the optimization with respect to \mathbf{A} . Therefore, the following objective function is required to be minimized:

$$\underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 \quad \text{subject to } \mathbf{A}^\top \mathbf{A} = \mathbf{I}_k \quad (13)$$

A closed form solution for minimizing (13) can be obtained by computing a reduced rank procrustes rotation [26].

Theorem 2.2. $\hat{\mathbf{A}} = \hat{\mathbf{U}}\hat{\mathbf{V}}^\top$ is the closed form solution of (13).

Proof. We first expand the Frobenius norm

$$\|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 = \operatorname{Tr}((\mathbf{X} - \mathbf{XBA}^\top)^\top (\mathbf{X} - \mathbf{XBA}^\top)) \quad (14)$$

$$\|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 = \operatorname{Tr}((\mathbf{X}^\top - \mathbf{AB}^\top \mathbf{X}^\top)(\mathbf{X} - \mathbf{XBA}^\top)) \quad (15)$$

$$\begin{aligned} \|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 &= \operatorname{Tr}(\mathbf{X}^\top \mathbf{X}) - \operatorname{Tr}(\mathbf{AB}^\top \mathbf{X}^\top \mathbf{X}) - \operatorname{Tr}(\mathbf{X}^\top \mathbf{XBA}^\top) \\ &\quad + \operatorname{Tr}(\mathbf{AB}^\top \mathbf{X}^\top \mathbf{XBA}^\top) \end{aligned} \quad (16)$$

$$\|\mathbf{X} - \mathbf{XBA}^\top\|_F^2 = \operatorname{Tr}(\mathbf{X}^\top \mathbf{X}) - 2\operatorname{Tr}(\mathbf{X}^\top \mathbf{XBA}^\top) + \operatorname{Tr}(\mathbf{B}^\top \mathbf{X}^\top \mathbf{XB}) \quad (17)$$

The middle term in (17) arises due to the fact that the trace of a matrix and its transpose are equal. The last term has been simplified to $\mathbf{B}^\top \mathbf{X}^\top \mathbf{XB}$ due to orthogonality constraints on \mathbf{A} . Therefore, minimizing the loss function requires maximizing the trace of $\mathbf{X}^\top \mathbf{XBA}^\top$, since it is negative. Assume the SVD of $\mathbf{X}^\top \mathbf{XB}$ is $\hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top$, then

$$\operatorname{Tr}(\mathbf{X}^\top \mathbf{XBA}^\top) = \operatorname{Tr}(\hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top \mathbf{A}^\top) \quad (18)$$

$$\operatorname{Tr}(\mathbf{X}^\top \mathbf{XBA}^\top) = \operatorname{Tr}(\hat{\mathbf{V}}^\top \mathbf{A}^\top \hat{\mathbf{U}}\hat{\mathbf{S}}) \quad (19)$$

Eq. (19) is drawn from the cyclic nature of the trace of a product of matrices. Since, $\hat{\mathbf{S}}$ is a diagonal matrix, $\operatorname{Tr}(\hat{\mathbf{V}}^\top \mathbf{A}^\top \hat{\mathbf{U}}\hat{\mathbf{S}})$ is maximized when the diagonal of $\hat{\mathbf{V}}^\top \mathbf{A}^\top \hat{\mathbf{U}}$ is maximum. This is true when $\hat{\mathbf{V}}^\top \mathbf{A}^\top \hat{\mathbf{U}} = \mathbf{I}$. Therefore, after simplification, $\mathbf{A}^\top \hat{\mathbf{U}} = \hat{\mathbf{V}}$ or $\mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{V}}^\top$ is the closed form solution of (13). \square

The alternating minimization process is repeated until convergence or until a specified number of iterations is reached. The obtained joint sparse basis \mathbf{B} has exactly $q < p$ non-zero rows. A reduced band index set \mathcal{R} is computed which contains the indexes of the non-zero rows of \mathbf{B} . Then all possible subsets \mathcal{T} of the reduced band set \mathcal{R} are tested for ink mismatch detection. This is done by obtaining a reduced data matrix ($\mathbf{X}_{\cdot \mathcal{T}}$) by taking the columns of \mathbf{X} indexed by each set \mathcal{T} and computing mismatch detection accuracy. The subset \mathcal{T} which maximizes accuracy is chosen as the selected band set \mathcal{S} . Algorithm 2 summarizes the procedure for joint sparse band selection (JSBS).

Algorithm 2. Joint sparse band selection.

Input: $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{y} \in \mathbb{R}^n$, λ , ϵ , i_{\max}
Initialize: $i \leftarrow 1$, $\text{converge} \leftarrow \text{false}$, $\mathcal{S} \leftarrow \emptyset$, $\mathcal{R} \leftarrow \{1, 2, \dots, p\}$
 $\mathbf{USV}^\top \leftarrow \mathbf{X}$ \triangleright SVD of \mathbf{X}
 $\mathbf{A} \leftarrow \mathbf{V}_{\cdot \{1:k\}}$, $\mathbf{B} \leftarrow \mathbf{0}$ \triangleright Initialize \mathbf{A} with 1st k basis vectors and \mathbf{B} with zeros
while $i \leq i_{\max} \vee \neg \text{converge}$ **do**
 $\hat{\mathbf{B}} \leftarrow \min_{\mathbf{B}} \|\mathbf{XA} - \mathbf{XB}\|_F^2 + \lambda \|\mathbf{B}\|_{2,1}$ \triangleright Step 1: Find \mathbf{B} given \mathbf{A}
 $\hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top \leftarrow \mathbf{X}^\top \hat{\mathbf{X}}\hat{\mathbf{B}}$ \triangleright Step 2: Find \mathbf{A} given \mathbf{B}
 $\hat{\mathbf{A}} \leftarrow \hat{\mathbf{U}}\hat{\mathbf{V}}^\top$
if $\|\mathbf{B} - \hat{\mathbf{B}}\|_F < \epsilon$ **then** \triangleright Step 3: Check convergence
 $\text{converge} \leftarrow \text{true}$
else
 $\mathbf{B} \leftarrow \hat{\mathbf{B}}$, $\mathbf{A} \leftarrow \hat{\mathbf{A}}$ \triangleright Update \mathbf{A}, \mathbf{B}
 $i \leftarrow i + 1$
end if
end while
 $\mathcal{R} \leftarrow \{j \in \mathcal{R} : \|\beta^j\|_2 = 0\}$ \triangleright Step 4: Obtain reduced subset of bands
 $\mathcal{S} \leftarrow \operatorname{argmax}_{\mathcal{T}} \operatorname{IMDaccuracy}(\mathbf{X}_{\cdot \mathcal{T}}, \mathbf{y})$, $\forall \mathcal{T} \subseteq \mathcal{R}$ \triangleright Section 2.4
Output: $\mathcal{S} \in \mathbb{Z}^q$

2.4. Ink mismatch detection accuracy computation

In Sections 2.2 and 2.3, we proposed two algorithms for band selection from ink spectral responses. Using the selected bands, a reduced data matrix \mathbf{Z} is used for ink mismatch detection. We cluster the pixels belonging to g inks using k -means algorithm [34]. K -means minimizes the squared Euclidean distance between

the cluster centroid and its members by the following criteria:

$$\operatorname{argmin}_{\hat{C}} \sum_{i=1}^g \sum_{\mathbf{z}^j \in \hat{C}_i} \|\mathbf{z}^j - \bar{\mathbf{z}}_i\|^2, \quad (20)$$

where $\|\cdot\|^2$ is the squared error between the cluster member $\mathbf{z}^j \in \mathbb{R}^q$ and its centroid $\bar{\mathbf{z}}_i$. \mathbf{z}^j is the j th row of matrix \mathbf{Z} which is in the i th cluster \hat{C}_i . The total number of clusters is g which relates to the number of inks in \mathbf{Z} .

Let $\mathbf{Y} \in \mathbb{R}^{n \times g}$ be the ground truth class indicator matrix such that

$$Y_{ij} = \begin{cases} 1 & \text{if } \mathbf{z}^j \in C_i \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

where C_i is the ground truth cluster of ink i . Also, let $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times g}$ be the class indicator matrix predicted by k -means clustering

$$\hat{Y}_{ij} = \begin{cases} 1 & \text{if } \mathbf{z}^j \in \hat{C}_i \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

The mismatch detection accuracy of i th ink class is defined as the number of correctly labeled pixels of i th ink divided by the number of pixels labeled with i th ink in either the ground truth labels \mathbf{y}_i or the predicted labels $\hat{\mathbf{y}}_i$ [35]. The mismatch detection accuracy is computed as

$$\text{accuracy} = \max_g \frac{1}{g!} \sum_{i=1}^g \frac{T_i}{T_i + F_i + N_i} \quad (23)$$

where

$T_i = \mathbf{y}_i \wedge \hat{\mathbf{y}}_i \rightarrow$ no. of correctly labeled pixels of the i th ink

$F_i = \mathbf{y}'_i \wedge \hat{\mathbf{y}}_i \rightarrow$ no. of pixels incorrectly labeled as i th ink

$N_i = \mathbf{y}_i \wedge \hat{\mathbf{y}}'_i \rightarrow$ no. of incorrectly labeled pixels of i th ink

It is important to note that according to this evaluation metric, the accuracy of a random guess (e.g., in a two class problem) will be $1/3$. This is different to common classification accuracy metrics where the accuracy of a random guess is $1/2$. This is because the metric additionally penalizes false negatives which is crucially important in mismatch detection problem.

3. Writing ink hyperspectral image database

Cameras have recently emerged as an alternative to scanners for capturing document images. However, the focus has remained on mono-/tri-chromatic imaging. We outline and discuss the key components of our hyperspectral document imaging system, which offers new challenges and perspectives. We discuss the issues of spatial/spectral non-uniformity in data acquisition and propose solutions via pre and post-processing.

3.1. Acquisition setup

The hyperspectral document imaging system is illustrated in Fig. 5. The system consists of a monochrome machine vision camera with a native resolution of 752×480 pixels. In front of the camera is a focusing lens (1:1.4/16 mm). In order to capture hyperspectral images in discrete wavelength channels, a liquid crystal tunable filter (LCTF) is placed in the front of the lens. We chose the tunable filter due to its fast frequency tuning in the visible spectrum (< 50 ms). The filter bandwidth, measured in terms of the *full width at half maximum (FWHM)* is 7–20 nm which varies with the center wavelength. The scene is illuminated by halogen lamps on both sides of the document. A controller (computer) triggers cycles of filter tuning/image acquisition at a

high speed allowing for efficient image capture. The filter is tuned in the 400–720 nm range at steps of 10 nm so that a 33 band HSI is captured in less than 5 s. RGB scanned images at resolutions of 150, 300 and 600 dpi were also collected using a flatbed scanner for baseline comparison with HSI.

3.2. Database specifications

Using the described hyperspectral imaging setup, we have prepared a dataset that has a total of 70 hyperspectral images of hand-written notes by 7 subjects. All subjects were instructed to write the sentence, 'The quick brown fox jumps over the lazy dog', once in each ink on a white paper. The pens included 5 varieties of blue ink and 5 varieties of blank ink. It was ensured that the pens came from different manufacturers while the inks still appeared visually similar. All efforts were made to avoid prolonged exposure to ambient/daylight by keeping the samples under cover in dark. This is because inks might undergo slight transformation in their spectral properties due to chemical reactions induced by light. Such an occurrence might better distinguish two inks, but would bias the analysis. Moreover, all samples were collected in one session so that their effective age is the same.

3.3. Spectral reflectance compensation

Common illumination sources generate lower spectral power in shorter wavelengths as compared to longer wavelengths. An observation of the LCTF transmittance in Fig. 6(b) is that the amount of light transmitted is a function of the wavelength such that, the shorter the wavelength λ , the lower the filter transmittance and vice versa. Finally, the sensor quantum efficiency is also variable with respect to the wavelength as shown in Fig. 6(c). Typically, each band of a hyperspectral image is captured with a fixed exposure time. With a fixed exposure setting, the captured hyperspectral image looks as shown in Fig. 6(d) which has low energy at shorter wavelengths and high energy at longer wavelengths of the spectrum.

A common approach to address this issue is to perform a white-black calibration [11]. Its main idea is to normalize each band with respect to the spectral reflectance of white/black reference patches. However, since LCTF transmittance gets very small (low signal-to-noise ratio) in near-ultraviolet spectral range, use of white/black

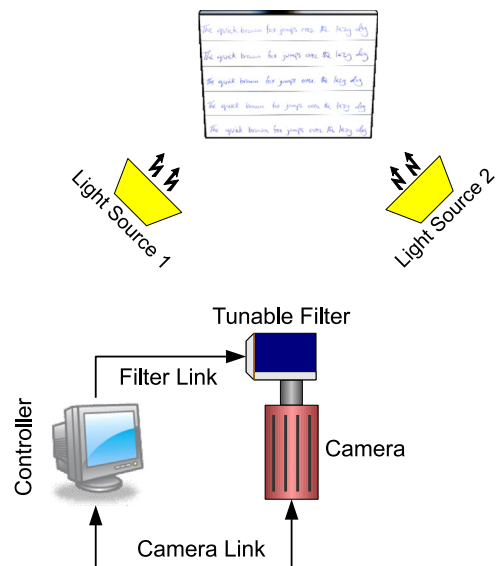


Fig. 5. An illustration of the proposed hyperspectral document image acquisition setup.

calibration amplifies noise in those bands. Instead, we use a variable exposure time imaging strategy. The exposure times are pre-computed for each band such that a reference white patch has a flat (uniform) spectral response across all bands. This enables simultaneous compensation for spectral variation due to illumination, filter and sensor. Using the exposures in Fig. 6(e), spectral uniformity is achieved in all bands as shown in Fig. 6(f). Note that in order to compensate for the limiting factors of this system, long exposure is required for bands in shorter wavelengths and short exposure is required in longer wavelengths.

3.4. Spatial illumination homogenization

In camera based document imaging, the use of a nearby illumination source induces an inhomogeneous scalar field over the document image as shown in Fig. 3(a). This means that there is a spatial non-uniformity in illumination energy. The result is that the pixels near the center of the image are brighter (high energy) as compared to the pixels farther away towards the edges (low energy). The effect can be minimized by using light diffuser [36,37] or specialized illumination geometry [11]. However, introduction of diffuser could result in low light intensity and decrease the signal-to-noise ratio of images.

We propose a post-capture normalization strategy which neither requires a diffuser nor an additional pre-calibration step. Let $\mathbf{I}_{ij} \in \mathbb{R}^p$ be the spectral response at the image pixel (i, j) . It is evident that illumination inhomogeneity is primarily a function of pixel coordinates (i, j) and does not depend on the spectral dimension. If each pixel is normalized by the magnitude of radiant

energy at that pixel, the effect of inhomogeneity can largely be reduced. Hence, normalizing the spectral response vector at each pixel to a unit magnitude will largely compensate for the effect of non-uniform illumination intensity.

$$\hat{\mathbf{I}}_{ij} = \frac{\mathbf{I}_{ij}}{\|\mathbf{I}_{ij}\|} \tag{24}$$

where $\hat{\mathbf{I}}_{ij}$ is the normalized spectral response. This assumption will hold for each (i, j) as long as \mathbf{I}_{ij} is not saturated. Thus, it is important to restrict the exposure to a level which does not saturate any pixel of the image. This requirement is effectively satisfied by the strategy presented in the previous section.

4. Experiments and analysis of results

Multi-ink handwritten notes are formed using single ink notes by individual words of each ink in equal proportion (see Appendix A for effect of disproportionate inks). If the number of inks in a note is unknown, g can theoretically lie in the range $[1, n]$. For the sake of this analysis, we fix $g=2$, i.e., we assume that there are two possible inks in the image. This is a practical assumption in questioned document examination where the original note is written with one pen and is suspected to be forged by a second pen. We also keep ink mismatch detection experiments for blue and black inks separate, as different colored inks can be easily distinguished simply by visual examination. In our analysis, five different inks, taken two at a time, result in ten ink combinations, for blue and black color each. In the following experiments, c_{ij} will denote the combination of ink i with ink j such that

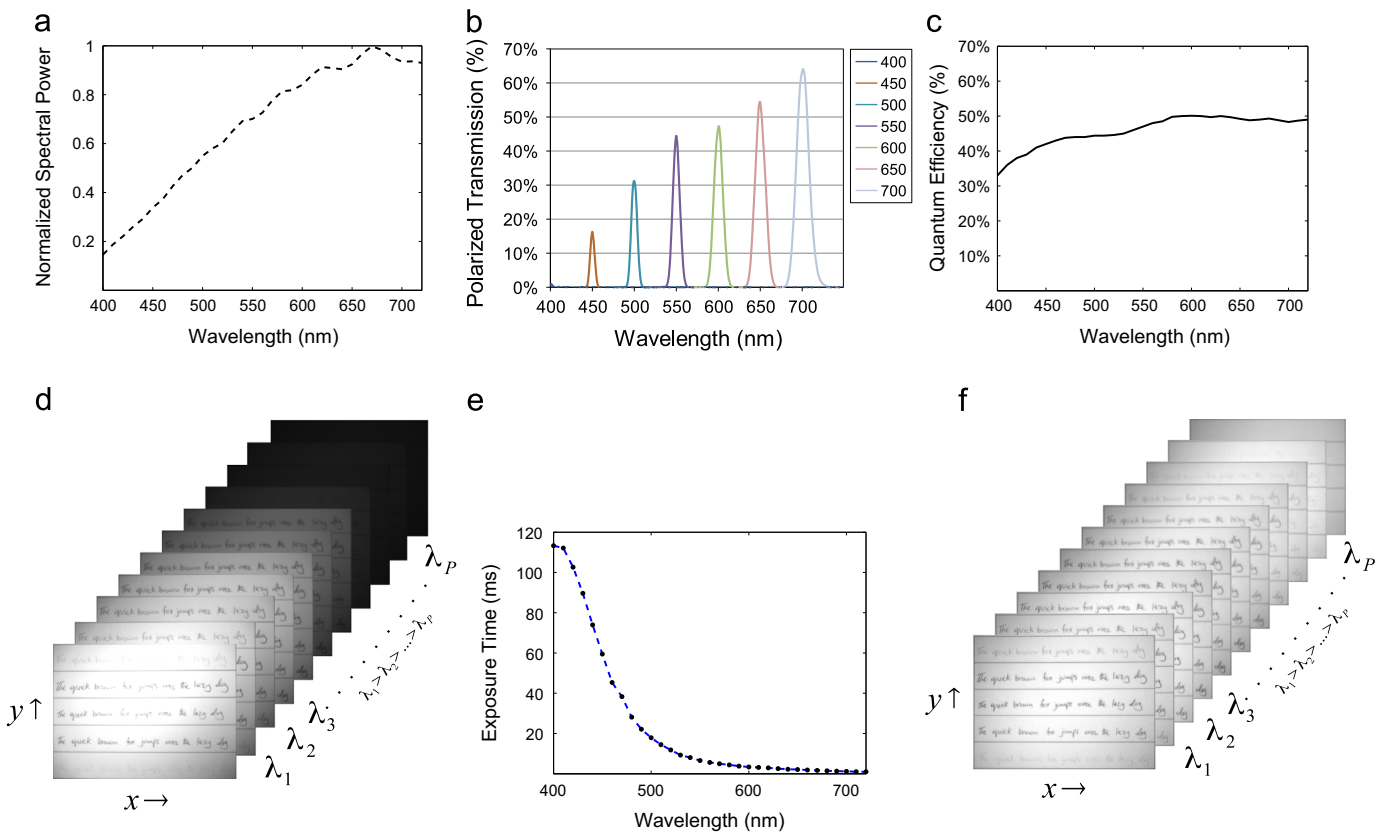


Fig. 6. (a) Spectral power distribution of the illuminant. (b) Transmission functions of the LCTF (only shown for seven different bands for conciseness). (c) Quantum efficiency of the CCD sensor. (d) Image captured with fixed exposure time. (e) Exposure times computed as a function of wavelength. (f) Image captured with variable exposure time. Observe that the illumination power, filter transmission and sensor quantum efficiency is compensated by variable exposure time.

$i = 1, \dots, 5$ and $j = i + 1, \dots, 5$. The mismatch detection accuracy is averaged over all samples for each ink combination c_{ij} .

We begin by analyzing the efficacy of the proposed hyperspectral document image spatial inhomogeneity correction. Fig. 7 shows two example handwritten notes in blue and black inks. The images are made by combining samples of ink 1 and ink 2 of blue and black inks, separately. The original images are shown in RGB for clarity. The ground truth images are labeled in different colors to identify the constituent inks in the note. The spatial inhomogeneity of the illumination can be observed from the center to the edges. The mismatch detection results on raw images indicate that the clustering is biased by the illumination intensity, instead of the ink color. After normalization of the raw HS images using (24), it is evident that the illumination inhomogeneity is highly suppressed. This results in accurate mismatch detection which closely follows the ground truth.

We now analyze how HSI can benefit ink mismatch detection compared to RGB images. Initially, we use all the bands of hyperspectral image for the analysis and use 300 dpi RGB images for comparison (see Appendix B for details). The mismatch detection accuracies by using HSI and RGB data are illustrated in Fig. 8. It can be seen that HSI significantly outperforms RGB in separating most ink combinations. This is evident in accurate clustering for blue ink combinations c_{12} , c_{14} , c_{25} , c_{35} and c_{45} . In case of the black inks, ink 1 is easily distinguishable from the other inks by using

HSI. This indicates that the RGB spectra is insufficient for distinguishing most ink combinations whereas for a few ink combinations, even HSI seems insufficient. These results invite further exploration of the most informative bands in HSI.

To study whether inks become more distinguishable in different regions of the spectrum, we plot the average normalized spectra of inks, in Fig. 9. These graphs are the outcome of average spectral response of each ink over all samples in the database. In order to evaluate the contribution of spectral ranges to ink discrimination, we perform separate experiments in sub-ranges, namely, low-visible (400 nm–500 nm), mid-visible (510 nm–590 nm) and high-visible range (600 nm–720 nm). These sub-ranges roughly correspond to the blue, green and red regions as no clear sub-categorization of the visible spectrum is defined in the current literature.

Fig. 10 shows the results of separate experiments in low-visible, mid-visible and high-visible range. For a majority of ink combinations, the high-visible range is most accurate, followed by the mid-visible and the low-visible range respectively. Observe, however, that the black ink combinations c_{34} , c_{35} and c_{45} are more distinguished in the low/mid-visible range which can be validated by referring to Fig. 9. Hence, it is likely that the inks are better distinguished in “specific bands” of the visible spectrum.

We now analyze the proposed band selection method and compare its performance to the related feature selection methods

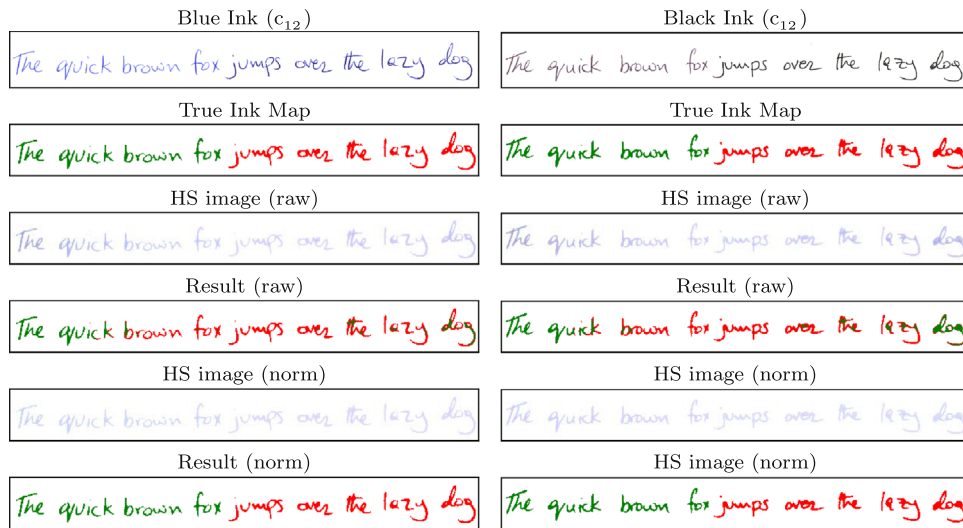


Fig. 7. An illustration of ink mismatch detection on a blue ink and a black ink image, acquired using adaptive exposure. The ground truth ink pixels are labeled in pseudo-colors (green: ink 1 and red: ink 2). The spatially non-uniform illumination pattern can be observed in raw HS images, with high energy in the center and low towards the edges. Normalization removes the illumination bias and greatly improves mismatch detection accuracy. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

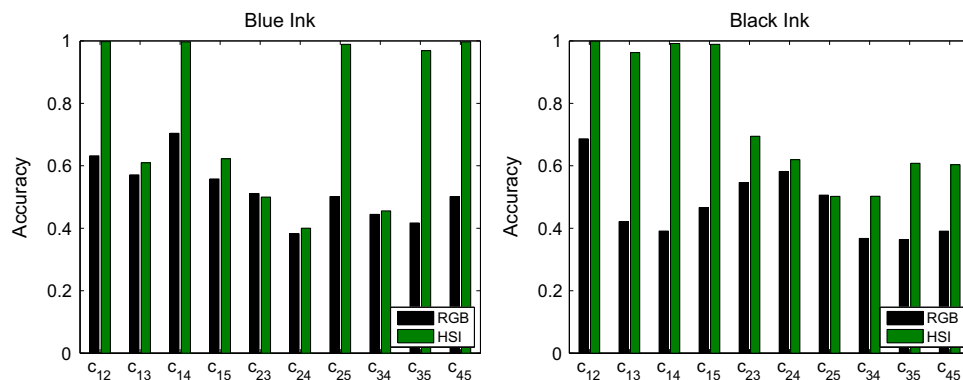


Fig. 8. Comparison of RGB and HSI based ink mismatch detection.

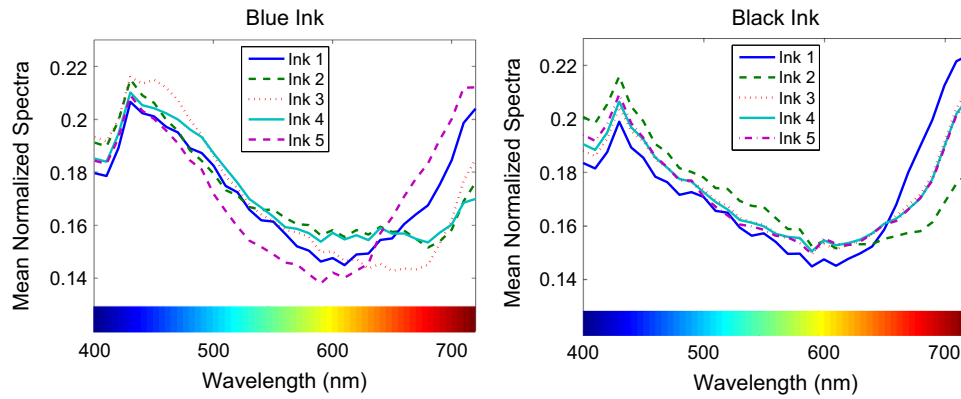


Fig. 9. Spectra of the blue and black inks under analysis. Note that at some ranges the ink spectra are more distinguished than others. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

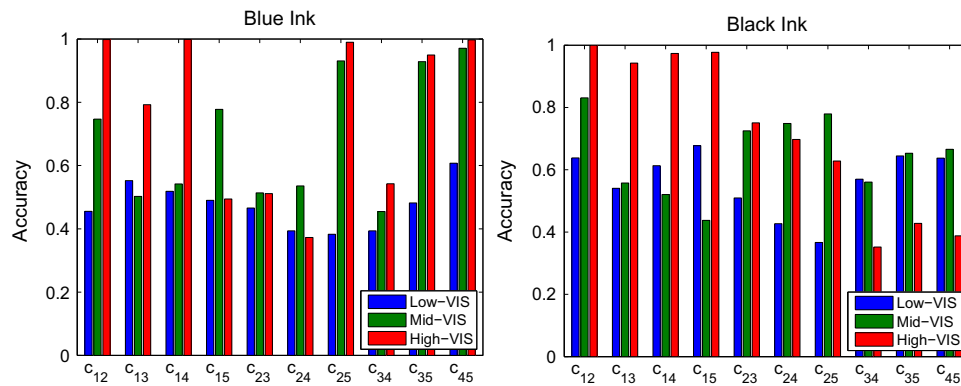


Fig. 10. HSI sub-spectral range analysis. Observe that the high-VIS range performs better than the mid-VIS and low-VIS ranges.

(SFBS [18] and L21-SPCA [17]). A leave-two-inks-out protocol is adopted to avoid bias of the selected features towards particular inks. In each fold, bands are selected from three ink combinations and tested on one ink combination while leaving the remaining seven ink combinations that contain either of the two test inks. For example, if the test ink combination is c_{12} , all ink combinations containing either ink-1 or ink-2 ($c_{12}, c_{13}, c_{14}, c_{15}, c_{23}, c_{24}, c_{25}$) are left out and the bands are selected from (c_{34}, c_{35}, c_{45}).

The accuracy on test data and the number of selected bands by each method is listed in Table 1. A higher accuracy by a band selection method indicates its selection of informative bands. On the contrary, a lower accuracy indicates missing out on informative bands (or selection of non-informative bands). Another important aspect of band selection is the number of bands that achieve a certain level of accuracy. Naturally, when band selection methods achieve similar accuracies, fewer selected bands are preferred. Moreover, it is likely for band selection methods to select same bands and achieve equal accuracies, in which case they are at par. Finally, the maximum number of bands selected by a method for any fold is a representative of the compactness of its selected feature space.

The accuracy using all bands (HSI-All) is useful for assessing the effectiveness of band selection methods. The SFBS method although improves over HSI-All for most ink combinations, but deteriorates for a few ink combinations which results in a low mean accuracy. The JSBS method results in the highest accuracy except for blue ink combinations c_{15} and c_{24} where SFBS is better (see Table 1). The L21-SPCA is closest in performance to JSBS in blue inks, however, JSBS has a clear advantage on all black ink combinations. Interestingly, black ink combination c_{12} is most

Table 1

Mismatch detection accuracies of all band selection methods on the test dataset using leave-two-inks-out strategy (both test inks have not been used in the training stage for band selection). The number of selected bands is reported in parenthesis. The accuracy of best performing method on each ink combination is emphasized in boldface.

Fold	HSI-All, % Acc.	SFBS, % Acc. (# bands)	L21-SPCA, % Acc. (# bands)	JSBS, % Acc. (# bands)
Blue inks				
c_{12}	99.8	99.5 (4)	99.8 (4)	99.8 (4)
c_{13}	61.0	98.2 (5)	99.5 (3)	99.5 (3)
c_{14}	99.7	99.9 (5)	99.9 (6)	99.9 (2)
c_{15}	62.3	83.3 (6)	59.4 (1)	73.6 (2)
c_{23}	50.0	54.3 (6)	59.0 (1)	59.0 (1)
c_{24}	40.0	58.0 (4)	44.1 (2)	44.2 (2)
c_{25}	98.9	98.6 (4)	99.1 (2)	99.1 (2)
c_{34}	45.5	45.6 (4)	91.6 (3)	93.0 (3)
c_{35}	96.9	95.9 (5)	98.9 (2)	99.0 (3)
c_{45}	99.7	99.8 (3)	99.8 (3)	99.8 (3)
Mean (max)	75.4	83.3 (6)	85.1 (6)	86.7 (4)
Black inks				
c_{12}	99.9	100.0 (6)	100.0 (2)	100.0 (1)
c_{13}	96.2	80.4 (6)	94.5 (2)	98.7 (4)
c_{14}	99.2	96.0 (5)	99.0 (2)	99.6 (5)
c_{15}	98.9	97.5 (5)	99.2 (3)	99.7 (6)
c_{23}	69.5	84.7 (7)	91.6 (1)	91.6 (1)
c_{24}	61.9	79.9 (10)	89.0 (1)	89.0 (1)
c_{25}	50.2	82.8 (6)	86.1 (2)	86.1 (2)
c_{34}	50.3	58.7 (3)	60.9 (2)	69.3 (2)
c_{35}	60.8	67.1 (8)	77.2 (4)	84.4 (3)
c_{45}	60.4	60.6 (4)	70.7 (5)	72.2 (5)
Mean (max)	74.7	81.8 (10)	86.8 (5)	89.0 (6)

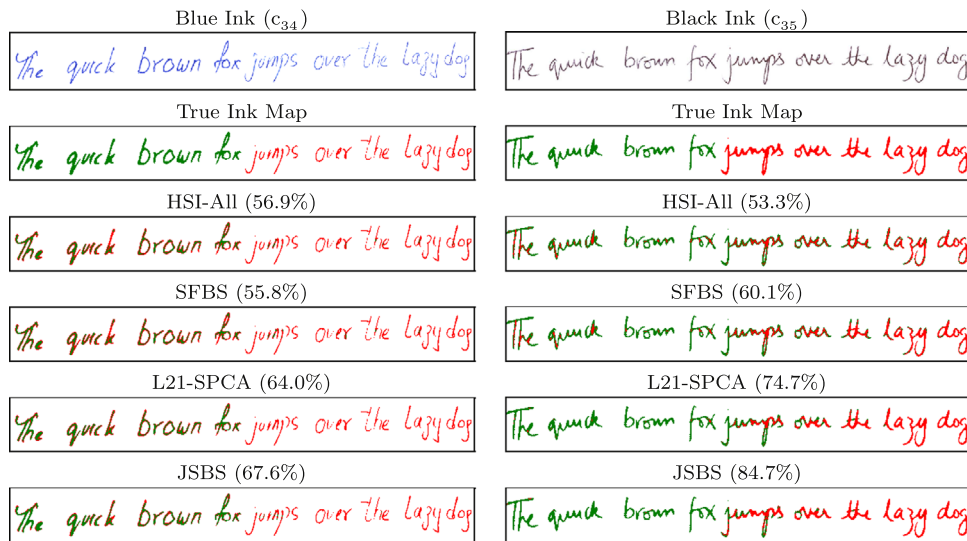


Fig. 11. Ink mismatch detection result on example test images. We purposefully selected two hard cases so that the capability of band selection based ink mismatch detection is visually appreciable. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

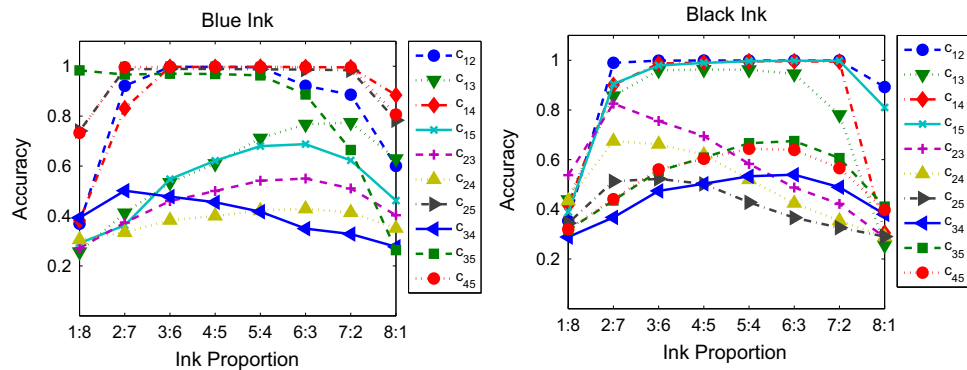


Fig. A1. The effect of varying ink proportion on mismatch detection accuracy.

easily distinguished by all methods as their spectra are highly distinguishable. In contrast, the relatively low accuracy of all methods on black ink combinations c_{34} , c_{35} and c_{45} is imminent from their highly similar spectra (see Fig. 9).

An important point to note is that the SFBS method selects a fairly high number of bands compared to L21-SPCA and JSBS. This is one of the major drawbacks of sequential feature selection which is forced to accumulate features until no further gain in accuracy is observed. In contrast, sparse coding based band selection is spontaneous and is able to select lower number of features as demonstrated by L21-SPCA and JSBS. Overall, the average accuracy of the proposed JSBS is better than L21-SPCA, SFBS and HSI-All for ink mismatch detection.

Fig. 11 shows ink mismatch detection results of different methods on example blue and black ink combinations. For the blue ink example, several boundary pixels of ink 3 are falsely grouped with ink 4, implying low accuracy by HSI-All, SFBS and L21-SPCA. However, this effect is least prominent in JSBS which groups majority of ink pixels according to the ground truth. Although the black ink example is visually hard to distinguish than blue ink, it yields relatively accurate results. Few words which are confused by HSI-All, SFBS and L21-SPCA are associated to the correct ink by JSBS. The presented JSBS method shows a remarkable performance gain for the case of black ink discrimination.

5. Discussion

Hyperspectral document imaging has immense potential for forensic document examination. We demonstrated the benefit of hyperspectral imaging in automated ink mismatch detection. The non-informative bands were effectively reduced by the proposed joint sparse band selection technique. Band selection gives an insight into how a customized multispectral imaging device with a smaller number of bands can be designed for ink mismatch detection, for instance, by combining an imaging sensor with a mechanical filter wheel [38]. Such devices may hugely benefit from the proposed findings in the selection of an optimal combination of filters which can be custom built for desired wavelengths.

One way to further improve mismatch detection accuracy would be to take spatial neighborhoods into account. Recall that the spectral responses of ink have been separated on a pixel-by-pixel basis. An interesting research direction would be to explore spatial pixel neighborhoods at sub-stroke level and develop algorithms that are sensitive to local spectral reflectance variations within a stroke. Another promising direction would be to explore spectral unmixing [39] in cases where two inks are superimposed, for instance, in the case of over-writing. We hope that the results presented in this paper along with the exciting new challenges would trigger more research efforts in the direction of automated hyperspectral document analysis.

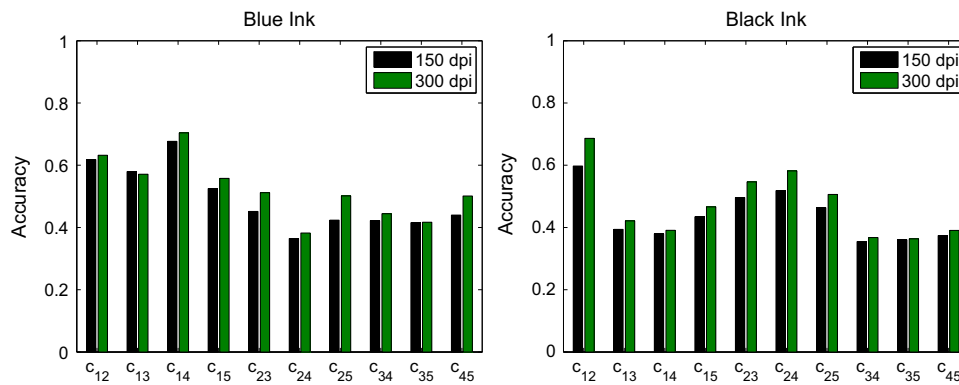


Fig. B1. Ink mismatch detection accuracy from RGB images. Note that neither black, nor blue inks could be reliably discriminated.

Conflict of interest

None declared.

Acknowledgments

This research was funded by the ARC Grant DP110102399 and the UWA Grant 00609 10300067.

Appendix A. Disproportionate inks

In this work, we used equal ink proportions (i.e., false writing to original writing) for conducting the experiments. Fig. A1 shows the effect of varying ink proportions on mismatch detection accuracy. It was experimentally realized by disproportionately combining words of different inks in a note. It was observed that highly disproportionate cases are less distinguishable which can be seen for ink proportions 1:8 and 8:1. For highly distinguishable ink combinations, the trend of accuracy in relation to ink proportion is generally symmetric.

Disproportionate ink mismatch detection is challenging because it is inherently an unbalanced clustering problem. This requires sophisticated clustering algorithms and is the focus of our future work. Note that our current approach can still be used to detect disproportionate inks using a sliding window. At some point, the window will contain equal proportions of two inks and the detection algorithm will show a peak at that point.

Appendix B. The effect of RGB image resolution on ink mismatch detection

In this work, we resorted to 300 dpi RGB images, as it is the most widely used resolution in document analysis systems. Fig. B1 shows the average accuracy at two commonly used RGB resolutions (150 and 300 dpi) for all ink combinations. Since, discrimination of inks relies primarily on the spectral dimension compared to the spatial dimension, the spatial resolution of RGB images has a minor impact on the degree of accuracy. Our experiments with up to 600 dpi scanning did not produce any significant improvement.

References

- [1] E.H. Land, J.J. McCann, et al., Lightness and retinex theory, *J. Opt. Soc. Am.* 61 (1) (1971) 1–11.
- [2] K.R. Gegenfurtner, Cortical mechanisms of colour vision, *Nat. Rev. Neurosci.* 4 (7) (2003) 563–572.
- [3] V. Aginsky, Forensic examination of slightly soluble ink pigments using thin-layer chromatography, *J. Forensic Sci.* 38 (5) (1993) 1131–1133.
- [4] ChemImage (www.chemimage.com/).
- [5] S. Joo Kim, F. Deng, M.S. Brown, Visual enhancement of old documents with hyperspectral imaging, *Pattern Recognit.* 44 (7) (2011) 1461–1469.
- [6] R. Hedjam, M. Cheriet, Historical document image restoration using multi-spectral imaging system, *Pattern Recognit.* 46 (8) (2013) 2297–2312.
- [7] G. Edelman, E. Gaston, T. van Leeuwen, P. Cullen, M. Aalders, Hyperspectral imaging for non-contact analysis of forensic traces, *Forensic Sci. Int.* 223 (2012) 28–39.
- [8] F. Hollaus, M. Gau, R. Sablatnig, Enhancement of multispectral images of degraded documents by employing spatial information, in: *Proceedings of the International Conference on Document Analysis and Recognition*, 2013, pp. 145–149.
- [9] E.B. Brauns, R.B. Dyer, Fourier transform hyperspectral visible imaging and the nondestructive analysis of potentially fraudulent documents, *Appl. Spectrosc.* 60 (8) (2006) 833–840.
- [10] R. Padoan, T.A. Steemers, M. Klein, B. Aalderink, G. de Bruin, Quantitative hyperspectral imaging of historical documents: technique and applications, in: *ART Proceedings*, 2008.
- [11] M.E. Klein, B.J. Aalderink, R. Padoan, G. De Bruin, T.A. Steemers, Quantitative hyperspectral reflectance imaging, *Sensors* 8 (9) (2008) 5576–5618.
- [12] foster + freeman (<http://www.fosterfreeman.com/index.php>).
- [13] A. Chakrabarti, T. Zickler, Statistics of real-world hyperspectral images, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2011, pp. 193–200.
- [14] R. Hedjam, M. Cheriet, Hyperspectral band selection based on graph clustering, in: *Proceedings of the International Conference on Information Science, Signal Processing and their Applications*, 2012, pp. 813–817.
- [15] A. Martínez-Usó, F. Pla, J.M. Sotoca, P. García-Sevilla, Clustering-based hyperspectral band selection using information measures, *IEEE Trans. Geosci. Remote Sens.* 45 (12) (2007) 4158–4171.
- [16] Q. Gu, Z. Li, J. Han, Joint feature selection and subspace learning, in: *Proceedings of the International Joint Conference on Artificial Intelligence*, vol. 22, 2011, p. 1294.
- [17] S. Xiaoshuang, L. Zhihui, G. Zhenhua, W. Minghua, Z. Cairong, K. Heng, Sparse principal component analysis via joint $l_{2,1}$ -norm penalty, in: *AI 2013: Advances in Artificial Intelligence*, 2013, pp. 148–159.
- [18] Z. Khan, F. Shafait, A. Mian, Hyperspectral imaging for ink mismatch detection, in: *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, 2013, pp. 877–881.
- [19] A. Brink, J. Smit, M. Bulacu, L. Schomaker, Writer identification using directional ink-trace width measurements, *Pattern Recognit.* 45 (1) (2012) 162–171.
- [20] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* 9 (1) (1979) 62–66.
- [21] J. Sauvola, M. Pietikäinen, Adaptive document image binarization, *Pattern Recognit.* 33 (2) (2000) 225–236.
- [22] F. Shafait, D. Keysers, T.M. Breuel, Efficient implementation of local adaptive thresholding techniques using integral images, *Doc. Recognit. Retr.* XV (2008) 681510–681510–6.
- [23] B. Gatos, I. Pratikakis, S.J. Perantonis, Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information, in: *19th International Conference on Pattern Recognition*, 2008, ICPR 2008, 2008, pp. 1–4.
- [24] L.C. Molina, L. Belanche, Á. Nebot, Feature selection algorithms: a survey and experimental evaluation, in: *Proceedings of the International Conference on Data Mining*, 2002, pp. 306–313.
- [25] Jolliffe, Ian. *Principal Component Analysis*. Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd., 2005 (ISBN: 9780470013199).
- [26] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, *J. Comput. Graph. Stat.* 15 (2) (2006) 265–286.
- [27] I.T. Jolliffe, N.T. Trendafilov, M. Uddin, A modified principal component technique based on the lasso, *J. Comput. Graph. Stat.* 12 (3) (2003) 531–547.

- [28] A. d'Aspremont, L. el Ghaoui, M.I. Jordan, G.R. Lanckriet, A direct formulation for sparse PCA using semidefinite programming, *SIAM Rev.* 49 (3) (2007) 434–448.
- [29] H. Shen, J.Z. Huang, Sparse principal component analysis via regularized low rank matrix approximation, *J. Multivar. Anal.* 99 (6) (2008) 1015–1034.
- [30] M. Journée, Y. Nesterov, P. Richtárik, R. Sepulchre, Generalized power method for sparse principal component analysis, *J. Mach. Learn. Res.* 11 (2010) 517–553.
- [31] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B (Methodol.)* (1996) 267–288.
- [32] J. Mairal, R. Jenatton, G. Obozinski, F. Bach, Network Flow Algorithms for Structured Sparsity, arXiv preprint [arXiv:1008.5209](https://arxiv.org/abs/1008.5209).
- [33] R. Jenatton, J. Mairal, G. Obozinski, F. Bach, Proximal methods for sparse hierarchical dictionary learning, in: Proceedings of the International Conference on Machine Learning, 2010, pp. 1–8.
- [34] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, *ACM Comput. Surv. (CSUR)* 31 (3) (1999) 264–323.
- [35] M. Everingham, L. van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL Visual Object Classes Challenge, 2012.
- [36] A.R. Cortés, Multispectral Analysis and Spectral Reflectance Reconstruction of Art Paintings (Ph.D. thesis), Télécom ParisTech, 2003.
- [37] F. Hollaus, M. Gau, R. Sablatnig, Multispectral image acquisition of ancient manuscripts, in: Progress in Cultural Heritage Preservation, 2012, pp. 30–39.
- [38] G. Bianco, F. Bruno, M. Muzzupappa, Multispectral data cube acquisition of aligned images for document analysis by means of a filter-wheel camera provided with focus control, *J. Cult. Herit.* 14 (3) (2013) 190–200.
- [39] N. Akhtar, F. Shafait, A. Mian, Futuristic greedy approach to sparse unmixing of hyperspectral data, *IEEE Trans. Geosci. Remote Sens.* 53 (4) (2015) 2157–2174.

Zohaib Khan received the BE degree in Avionics with distinction from the National University of Sciences and Technology, Pakistan (2008) and won the award for best performance in academics and the gold medal for best project in engineering. He received the PhD degree in Computer Science at The University of Western Australia, Australia (2014) as a recipient of the prestigious International Postgraduate Research Scholarship (2010–2014). He is currently working as a Research Associate at the School of Computer Science and Software Engineering, The University of Western Australia. His research interests mainly include image processing, pattern recognition with a focus on multi-modal imaging.

Faisal Shafait is working as an Assistant Professor in the School of Computer Science & Software Engineering at The University of Western Australia. Formerly, he was a Senior Researcher at the German Research Center for Artificial Intelligence (DFKI), Germany and a visiting researcher at Google, California. He received his Ph.D. in computer engineering with the highest distinction from Kaiserslautern University of Technology, Germany in 2008. His research interests include machine learning and pattern recognition with a special emphasis on applications in document image analysis and recognition. He has co-authored over 100 publications in international peer-reviewed conferences and journals in this area. He is an Editorial Board member of the International Journal on Document Analysis and Recognition (IJ-DAR), and a Program Committee member of leading document analysis conferences including ICDAR, DAS, and ICFHR. He is also serving on the Leadership Board of IAPR's Technical Committee on Computational Forensics (TC-6).

Ajmal Mian completed his PhD from The University of Western Australia in 2006 with distinction and received the Australasian Distinguished Doctoral Dissertation Award from Computing Research and Education Association of Australasia. He received two prestigious fellowships; the Australian Postdoctoral Fellowship in 2008 and the Australian Research Fellowships in 2011. He received the UWA Outstanding Young Investigator Award in 2011, the West Australian Early Career Scientist of the Year award in 2012 and the Vice-Chancellors Mid-Career Research Award in 2014. He has secured four Australian Research Council grants worth over \$2 Million. He is currently in the School of Computer Science and Software Engineering at The University of Western Australia. His research interests include computer vision, 3D shape analysis, hyperspectral image analysis, pattern recognition, machine learning and multimodal biometrics.