Spatial-Spectral Representation for X-Ray Fluorescence Image Super-Resolution

Qiqin Dai, Emeline Pouyet, Oliver Cossairt, Marc Walton, and Aggelos K. Katsaggelos, *Fellow, IEEE*

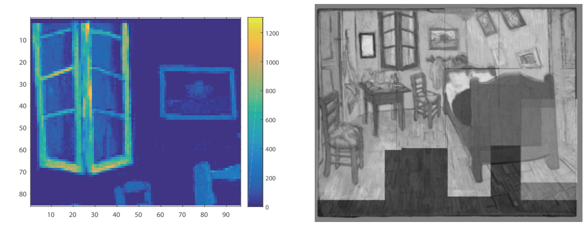
***Abstract*—X-Ray fluorescence (XRF) scanning of works of art is becoming an increasing popular non-destructive analytical method. The high quality XRF spectra is necessary to obtain significant information on both major and minor elements used for characterization and provenance analysis. However, there is a trade-off between the spatial resolution of an XRF scan and the Signal-to-Noise Ratio (SNR) of each pixel's spectrum, due to the limited scanning time. In this project, we propose an XRF image super-resolution method to address this trade-off, thus obtaining a high spatial resolution XRF scan with high SNR. We fuse a low resolution XRF image and a conventional RGB high-resolution image into a product of both high spatial and high spectral resolution XRF image. There is no gauruntee of a one to one mapping between XRF spectrum and RGB color since, for instance, paintings with hidden layers cannot be detected in visible but can in X-ray wavelengths. We separate the XRF image into the visible and non-visible components. The spatial resolution of the visible component is increased utilizing the high-resolution RGB image while the spatial resolution of the non-visible component is increased using a total variation super-resolution method. Finally, the visible and non-visible components are combined to obtain the final result.**

***Index Terms*—X-Ray fluorescence, Super-resolution, spatial-spectral.**

I. Introduction

OVER the last few years, X-Ray fluorescence (XRF) laboratory-based systems have evolved to lightweight and portable instruments thanks to technological advancements in both X-Ray generation and detection. Spatially resolved elemental information can be provided by scanning the surface of the sample with a focused or collimated X-ray beam of (sub) millimeter dimensions and analyzing the emitted fluorescence radiation, in a nondestructive in-situ fashion entitled Macro X-Ray Fluoresence (MA-XRF). The new generations of XRF spectrometers are used in the Cultural Heritage field to study the technology of manufacture, provenance, authenticity, etc, of works of art. Because of their fast non-invasive set up, we are able to study of large, fragile and location inaccessible art objects and archaeological collections. In particular, XRF has been extensively used to investigate historical paintings, by capturing the elemental distribution images of their complex layered structure. This method reveals the painting history from the artist creation to restoration processes [1], [2].

As with other imaging techniques, high spatial resolution and high Signal-to-Noise Ratio (SNR) are desirable for XRF scanning systems. However, the acquisition time is usually limited resulting in a compromise between dwell time, spatial resolution, and desired image quality. In the case of scanning large scale mappings, a choice may be made to reduce the dwell time and increase the step size, resulting in low SNR XRF spectra and low spatial resolution XRF images.



(a) (b)

Fig. 1. (a) XRF map showing the distribution of Cr Ka on a section of the ”Bedroom”, by Vincent Van Gogh, The Art Institute of Chicago, and (b) the automatic registration of 10 maps layered on top of the original resolution RGB image.

An example of an XRF scan is shown in Figure 1 (a). Channel 636 corresponding to Cr Ka elemental X-Ray lines was extracted from a scan of Vincent Van Gogh's “Bedroom” painted in 1889 (housed at The Art Institute of Chicago, acc # 1926.417). The image is color coded for better visibility. This is an image out of 4096 channels that were simultaneously acquired by a Bruker M6 scanning energy dispersive XRF instrument. The image has a low resolution (LR) of 96×85 pixels, yet still took $1-2$ hour to acquire it. Given the fact that the paining has dimensions 73.6×92.3 cm, at least 10 such patches are needed to capture the whole painting. Much higher resolution would be desirable for didactic purposes to show curators, conservators, and the general public. This makes the acquisition process highly impractical and therefore impedes the use of XRF scanning instruments as high resolution widefield imaging devices. In Figure 1 (b) we also show an automatic registration of all 10 averaged XRF maps (across all channels) layered on top of the original RGB image.

In this paper, we propose a super-resolution (SR) approach to obtain high resolution (HR) XRF images, with the aid of a conventional HR RGB image, as shown in Fig. 2. Our proposed XRF image SR algorithm can also be applied to spectral images obtained by any other raster scanning methods, such as Scanning Electron Microscope (SEM), Energy

Q. Dai, O. Cossairt, A. Kappeler and A. K. Katsaggelos are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, 60208 USA. E. Pouyet, M. Walton, F. Casadio are with the Northwestern University / Art Institute of Chicago Center for Scientific Studies in the Arts (NU-ACCESS), Evanston, IL, 60208 USA. e-mail: qiqindai2012@u.northwestern.edu; emeline.pouyet@northwestern.edu; ollie@eecs.northwestern.edu; marc.walton@northwestern.edu; fcasadio@artic.edu; aggk@eecs.northwestern.edu.

Manuscript received September 15, 2016.

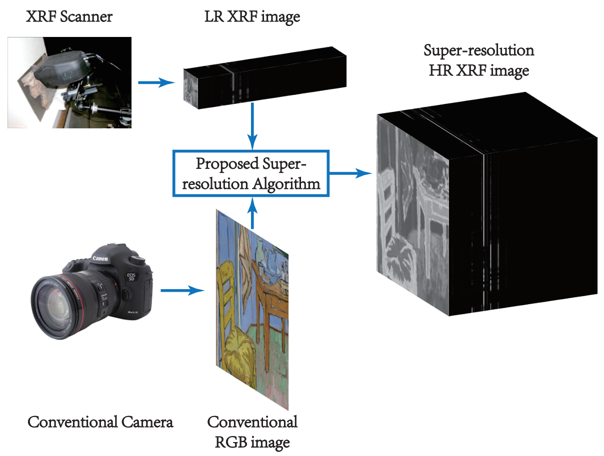


Fig. 2. XRF images have high spectral resolution but low spatial resolution, whereas the opposite is true for conventional RGB images. The LR XRF image and the HR RGB image are fused to obtain an HR XRF image.

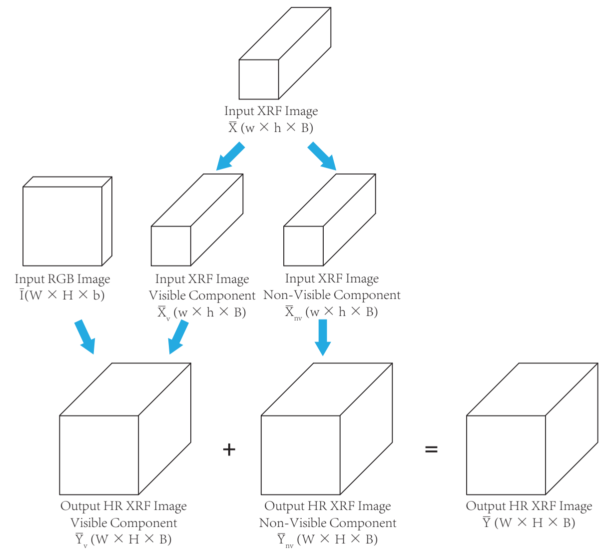


Fig. 3. Proposed pipeline of spatial-spectral representation for X-ray fluorescence image super-resolution. The visible component of input XRF image is fused with the input RGB image to obtain the visible component of HR XRF image. The non-visible component of the input XRF image is super-resolved to obtain the non-visible component of HR XRF image. The HR visible and non-visible component of output XRF image are combined to obtain the final output.

Dispersive Spectroscopy (EDS) and Wavelength Dispersive Spectroscopy (WDS). We model the spectrum of each pixel using a linear mixing model [3], [4]. Since there is no direct one-to-one mapping between the visible RGB spectrum and the XRF spectrum, because the hidden part of the paining is not visible in the conventional RGB image, but it can be captured in the XRF image [5], we model the XRF signal as a combination of the visible signal (on the surface) and the non-visible signal (hidden under surface), as shown in Fig. 3. For super-resolving the visible XRF signal we follow a similar approach to previous research in [6]–[12]. A coupled XRF-RGB dictionary is learned to explore the correlation between XRF and RGB signals. The RGB dictionary is applied to obtain the sparse representation of the HR RGB input image, resulting in an HR coefficient map. Then the XRF dictionary is applied on the HR coefficient map to reconstruct the HR XRF image. For the non-visible part, we increase its spatial resolution using a standard total variation regularizer [13], [14]. Finally, the HR visible and the HR non-visible XRF signals are combined to obtain the final HR XRF result. We do not explicitly separate the input LR XRF image into visible and non-visible parts in advance. Instead, we formulate the whole SR problem as an optimization problem. By alternatively optimizing over the coupled XRF-RGB dictionary and the visible / non-visible HR coefficient maps, the fidelity of the estimated HR output to both the LR XRF and HR RGB input signals is improved, thus resulting in a better SR output. Both synthetic and real experiments show the effectiveness of our proposed method, in terms of reconstruction error and visual sharpness of the SR result, compared to other methods, such as bicubic interpolation, the total variation only SR method [13], [14] and hyperspectral image SR methods [7]–[9].

The paper is organized as follows. Section II reviews the literature related to the proposed approach. We formulate the XRF image SR problem in Section III, while the proposed method is described in Section IV. In Section V, we provide the experimental results with both synthetic data and real data to evaluate the approach. The paper is concluded in Section VI.

II. Related work

While there is a large boby of work on SR for either conventional RGB images [15]–[19] or hyperspectral images [6]–[12], little work has been done for SR on XRF images. XRF SR poses a particular challenge because the acquired spectrum signal usually has low SNR. In addition, correlations among spectral channels need to be preserved for the interpolated pixels. Finally, the large number of channels (4096 channels in Fig. 1) leads to a computation challenge, since super-resolving each channel slice by slice is computational expensive.

The low spatial resolution limitations of hyperspectral images have led researchers in image processing and remote sensing to attempt to fuse them with conventional high spatial resolution RGB images. This image fusion [20] style SR can be seen as a generalization of pan-sharpening [21], [22], which enhances an LR color image by fusing it with a single-channel black-and-white ("panchromatic") image of higher resolution. Recently, matrix factorization has played an important role in enhancing the spatial resolution of hyperspectral imaging systems [6]–[9]. In [6], a sparse matrix factorization technique was proposed to decompose the LR hyperspectral image into a dictionary of basis vector and a set of sparse coefficients. The HR hyperspectral image was then reconstructed using the learned basis and sparse coefficients computed from the HR RGB image. The SR performance is improved by imposing spatio-spectral sparsity [7], physical constraints [8] and structural prior [9]. Bayesian approaches [10], [11] impose additional priors on the distribution of the image intensities and apply MAP inference. Non-parametric Bayesian dictionary learning is applied in [12] to obtain a spectral basis, and then obtain the HR image with Bayesian sparse coding.

In all these hyperspectral image SR methods [6]–[12], because the RGB spectrum is contained within the hyperspectral spectrum, the transformation from the hyperspectral signal to the RGB signal is linear and known. However, in XRF imaging, because the RGB spectrum (400 nm - 700 nm) is outside the XRF spectrum (0.03 nm - 6 nm, i.e., 0.2 KeV - 40 KeV), there is no direct transformation from the XRF signal to the RGB signal. Also the hidden part of the scanning object will be captured in the XRF image [5], while absent in the RGB image. According to our knowledge, no work has been done on XRF image SR, by modeling the input LR image as a combination of the visible and non-visible components, and increasing the spatial resolution of the visible component and non-visible component by fusing an HR conventional RGB image with implicit spectral transformation and using a standard total variation SR method, respectively. The physically grounded unmixing constraints in [8] on endmembers and abundances are extended in this paper to model the implicit transformation between the XRF spectrum and the RGB spectrum, as well as the visible / non-visible separation.

III. Problem Formulation

As shown in Fig. 3, we are seeking the estimation of an HR XRF image $\bar{Y} \in \mathbb{R}^{W\times H\times B}$ that has both high spatial and high spectral resolution, with W, H and B the image width, image height and number of spectral bands, respectively. We have two inputs: an LR XRF image $\bar{X} \in \mathbb{R}^{w\times h\times B}$ with lower spatial resolution $w \times h$, $w \ll W$ and $h \ll H$; and a conventional HR RGB image $\bar{I} \in \mathbb{R}^{W\times H\times b}$ with high spatial resolution, but a small number of spectral bands, $b \ll B$. The input LR XRF image $\bar{X}$ can be separated into two parts: the visible component $\bar{X}_v \in \mathbb{R}^{w\times h\times B}$ and the non-visible component $\bar{X}_{nv} \in \mathbb{R}^{w\times h\times B}$. We propose to estimate the HR visible component $\bar{Y}_v \in \mathbb{R}^{W\times H\times B}$ by fusing the conventional HR RGB input image $\bar{I}$ with the visible component of the input LR XRF image $\bar{X}_v$ and estimate the HR non-visible component $\bar{Y}_{nv} \in \mathbb{R}^{W\times H\times B}$ by using standard total variation SR methods.

To simplify notation, the images cubes are written as matrices, i.e. all pixels of an image are concatenated, such that every column of the matrix corresponds to the spectral response at a given pixel, and every row corresponds to the lexicographically ordered image in a specific spectral band. Accordingly, the image cubes are written as $Y \in \mathbb{R}^{B\times N_h}$, $X \in \mathbb{R}^{B\times N_l}$, $I \in \mathbb{R}^{b\times N_h}$, $X_v \in \mathbb{R}^{B\times N_l}$, $X_{nv} \in \mathbb{R}^{B\times N_l}$, $Y_v \in \mathbb{R}^{B\times N_h}$ and $Y_{nv} \in \mathbb{R}^{B\times N_h}$, where $N_h = W \times H$ and $N_l = w \times h$. We therefore have

$$X = X_v + X_{nv}, \tag{1}$$

$$Y = Y_v + Y_{nv}, \tag{2}$$

according to the visible / non-visible component separation model as shown in Fig. 3.

Let us denote by $y_v \in \mathbb{R}^B$ and $y_{nv} \in \mathbb{R}^B$ the one-dimensional spectra at a given spatial location of $\bar{Y}_v$ and $\bar{Y}_{nv}$, that is, representing a column of Y_v and Y_{nv}, according to the linear mixing model [23], [24], they can be described as

$$y_v = \sum_{j=1}^{M} d^{xrf}_{v,j}\alpha_{v,j}, \quad Y_v = D^{xrf}_v A_v, \tag{3}$$

$$y_{nv} = \sum_{j=1}^{M} d^{xrf}_{nv,j}\alpha_{nv,j}, \quad Y_{nv} = D^{xrf}_{nv} A_{nv}, \tag{4}$$

where $d^{xrf}_{v,j}$ and $d^{xrf}_{nv,j}$ represent respectively the endmembers for the visible and non-visible components, then $D^{xrf}_v \equiv [d^{xrf}_{v\ 1}, d^{xrf}_{v\ 2}, \ldots, d^{xrf}_{v\ M}] \in \mathbb{R}^{B\times M}$, $D^{xrf}_{nv} \equiv [d^{xrf}_{nv\ 1}, d^{xrf}_{nv\ 2}, \ldots, d^{xrf}_{nv\ M}] \in \mathbb{R}^{B\times M}$. $\alpha_{v,j}$ and $\alpha_{nv,j}$ are the corresponding per-pixel abundances. Equation 3 holds for a specific column y_v of matrix Y_v (say the k^{th} column). We take the corresponding $\alpha_{v,j,j=1,\ldots,M}$ and stack them into a $M \times 1$ column vector, this vector then becomes the k^{th} column of the matrix $A_v \in \mathbb{R}^{M\times N_h}$. In a similar manner we construct matrix $A_{nv} \in \mathbb{R}^{M\times N_h}$. The endmembers D^{xrf}_v and D^{xrf}_{nv} act as a basis dictionary to represent Y_v and Y_{nv} in a lower-dimensional space $\mathbb{R}^M$ and $rank\{Y_v\} \leq M,\ rank\{Y_{nv}\} \leq M$.

The visible and non-visible components of the input LR XRF image X_v and X_{nv}, respectively, are a spatially downsampled version of Y_v and Y_{nv}, respectively, that is

$$X_v = Y_v S = D^{xrf}_v A_v S, \tag{5}$$

$$X_{nv} = Y_{nv} S = D^{xrf}_{nv} A_{nv} S, \tag{6}$$

where $S \in \mathbb{R}^{N_h\times N_l}$ is the downsampling operator that describes the spatial degradation from HR to LR.

Similarly, the HR conventional RGB image I can be described by the linear mixing model [23], [24],

$$I = D^{rgb} A_v, \tag{7}$$

where $D^{rgb} \in \mathbb{R}^{b\times M}$ is the RGB dictionary. Notice that the same abundances matrix A_v is used in Equations 3 and 5. This is because the visible component of the scanning object is captured by both the XRF and the conventional RGB images. The matrix A_v encompasses the spectral correlation between the visible component of the XRF and the conventional RGB images.

The physically grounded constraints in [8] are shown to be effective, so we propose to impose similar constraints, by making full use of the fact that the XRF endmembers are XRF spectra of individual materials, and the abundances are proportions of those endmembers. Consequently, they are subject to the following constraints:

$$0 \leq D^{xrf}_{v,ij} \leq 1, \forall i,j \tag{8a}$$

$$0 \leq D^{xrf}_{nv,ij} \leq 1, \forall i,j \tag{8b}$$

$$0 \leq D^{rgb}_{ij} \leq 1, \forall i,j \tag{8c}$$

$$A_{v,ij} \geq 0, \forall i,j \tag{8d}$$

$$A_{nv,ij} \geq 0, \forall i,j \tag{8e}$$

$$\mathbf{1}^{\mathbf{T}}(A_v + A_{nv}) = \mathbf{1}^{\mathbf{T}}, \tag{8f}$$

where $D^{xrf}_{v,ij}$, $D^{xrf}_{nv,ij}$, D^{rgb}_{ij}, $A_{v,ij}$ and $A_{nv,ij}$ are the (i, j) elements of matrices D^{xrf}_v, D^{xrf}_{nv}, D^{rgb}, A_v and A_{nv}, respectively, $\mathbf{1^T}$ demotes a row vector of 1's compatible with the dimensions of A_v and A_{nv}. Equations 8a, 8b and 8c enforce the non-negative, bounded spectrum constraints on endmembers, Equations 8d and 8e, enforce the non-negative constraints on abundances, and Equation 8e enforces the visible component abundances and non-visible component abundances for every pixel to sum up to one.

IV. Proposed Solution

In order to solve the XRF image SR problem, we need to estimate A_v, A_{nv}, D^{rgb}, D^{xrf}_v and D^{xrf}_{nv} simultaneously. Utilizing Equations 1, 5, 6, 7 and 8, we can form the following constrained least-squares problem:

$$\min_{\substack{A_v, A_{nv}, D^{rgb},\\ D^{xrf}_v, D^{xrf}_{nv}}} \|X - D^{xrf}_v A_v S - D^{xrf}_{nv} A_{nv} S\|_F^2 + \|I - D^{rgb} A_v\|_F^2 + \lambda \|\nabla(D^{xrf}_{nv} A_{nv})\|_F^2 \tag{9a}$$

$$\text{s.t.} \quad 0 \le D^{xrf}_{v\ ij} \le 1, \forall i, j \tag{9b}$$

$$0 \le D^{xrf}_{nv\ ij} \le 1, \forall i, j \tag{9c}$$

$$0 \le D^{rgb}_{ij} \le 1, \forall i, j \tag{9d}$$

$$A_{v\ ij} \ge 0, \forall i, j \tag{9e}$$

$$A_{nv\ ij} \ge 0, \forall i, j \tag{9f}$$

$$\mathbf{1^T}(A_v + A_{nv}) = \mathbf{1^T}, \tag{9g}$$

$$\|A_v + A_{nv}\|_0 \le s, \tag{9h}$$

with $\|\cdot\|_F$ denoting the Frobenius norm, and $\|\cdot\|_0$ the ℓ_0 norm, i.e., the number of non-zero elements of the given matrix. The first term in Equation 9a represents a measure of the fidelity of the observed XRF data X, the second term the fidelity to the observed RGB image I and the third term in Equation 9a is the total variation (TV) regularizer. It is defined as

$$\begin{aligned}
&\|\nabla(D^{xrf}_{nv} A_{nv})\|_F^2 \\
&= \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} \|D^{xrf}_{nv} \bar{A}_{nv}(i, j, :) - D^{xrf}_{nv} \bar{A}_{nv}(i+1, j, :)\|_2^2 \\
&\qquad + \|D^{xrf}_{nv} \bar{A}_{nv}(i, j, :) - D^{xrf}_{nv} \bar{A}_{nv}(i, j+1, :)\|_2^2 \\
&= \|D^{xrf}_{nv} A_{nv} G\|_F^2
\end{aligned} \tag{10}$$

where $\bar{A}_{nv} \in \mathbb{R}^{W \times H \times M}$ is the $3D$ volume version of A_{nv} and $\bar{A}_{nv}(i, j, :) \in \mathbb{R}^M$ is the non-visible component abundance of pixel (i, j). $G \in \mathbb{R}^{N_h \times ((W-1)(H-1))}$ is the horizontal/vertical first order difference operator. To estimate the HR visible component abundance A_v, the HR RGB image I can provide spatial details. However, to estimate the HR non-visible component abundance A_{nv}, there is no additional spatial information, so we need the TV regularizer (Equation 10) to impose spatial smoothness on the non-visible component. The TV regularizer parameter λ controls the spatial smoothness of the reconstructed non-visible component, $Y_{nv} = D^{xrf}_{nv} A_{nv}$.

The constraint Equations 9e, 9f, 9g together restrict the abundances of visible and non-visible components, and also act as a sparsity prior on the per-pixel abundances, since they bound the ℓ_1 norm of the combined abundances $(A_v + A_{nv})$ to be 1 for all pixels. The last constraint Equation 9h is an optional constraint, which further enforces the sparsity of the combined abundance $(A_v + A_{nv})$.

The optimization in Equation 9 is non-convex and difficult to solve if we optimize over all the parameters A_v, A_{nv}, D^{rgb}, D^{xrf}_v and D^{xrf}_{nv} directly. We found it effective to alternatively optimize over these parameters. Also because Equation 9 is highly non-convex, good initialization is needed to start the local optimization. A similar approach as the coupled dictionary learning technique in [25], [26] is applied here to initialize these parameters.

A. Initialization

Let $I^l \in \mathbb{R}^{b \times N_l}$ and $A^l_v \in \mathbb{R}^{M \times N_l}$ be the spatially downsampled RGB image I and visible component abundance A_v, we have

$$I^l = IS, \tag{11}$$

$$A^l_v = A_v S. \tag{12}$$

Then the coupled dictionary learning technique in [25], [26] can be utilized to initialize D^{rgb} and D^{xrf}_v by

$$\begin{aligned}
\min_{D^{rgb}, D^{xrf}_v} \quad & \|I^l - D^{rgb} A^l_v\|_F^2 + \|X - D^{xrf}_v A^l_v\|_F^2 \\
& + \beta \sum_{k=1}^{N_l} \|A^l_v(:, k)\|_1, \\
\text{s.t.} \quad & \|D^{rgb}(:, k)\|_2 \le 1, \forall k, \\
& \|D^{xrf}_v(:, k)\|_2 \le 1, \forall k,
\end{aligned} \tag{13}$$

where $\|\cdot\|_1$ is the ℓ_1 vector norm, parameter β control the sparseness of the coefficients in A^l_v $A^l_v(:, k)$, $D^{rgb}(:, k)$ and $D^{xrf}_v(:, k)$ denote the k^{th} column of matrix A^l_v, D^{rgb}, and D^{xrf}_v, respectively. Details of the optimization can be found in [25], [26]. D^{rgb} and D^{xrf}_v are initialized using Equation 13 and D^{xrf}_{nv} is initialized to be equal to D^{xrf}_v. A_v is initialized by upsampling A^l_v computed in Equation 13, while A_{nv} is set to be zero at initialization.

B. Optimization Scheme

We propose to alternatively optimize over all the parameters in Equation 9a. First we optimize over A_v and A_{nv} by fixing all other parameters,

$$\begin{aligned}
\min_{A_v, A_{nv}} \quad & \|X - D^{xrf}_v A_v S - D^{xrf}_{nv} A_{nv} S\|_F^2 \\
& + \|I - D^{rgb} A_v\|_F^2 + \lambda \|\nabla(D^{xrf}_{nv} A_{nv})\|_F^2 \\
\text{s.t.} \quad & A_{v\ ij} \ge 0, \forall i, j \\
& A_{nv\ ij} \ge 0, \forall i, j \\
& \mathbf{1^T}(A_v + A_{nv}) = \mathbf{1^T}, \\
& \|A_v + A_{nv}\|_0 \le s.
\end{aligned} \tag{14}$$

PALM (proximal alternating linearized minimization) algorithm [27] is utilized to optimize over A_v and A_{nv} by a projected gradient descent method. For Equation 14, the following three steps are iterated for $q = 1, 2, \ldots$ until convergence:

$$V_v^q = A_v^{q-1} - \frac{1}{d_v} {D^{rgb}}^T (D^{rgb} A_v^{q-1} - I) \tag{15a}$$

$$\begin{aligned} V_{nv}^q = A_{nv}^{q-1} \\ &- \frac{1}{d_{nv}} ({D_{nv}^{xrf}}^T (D_{nv}^{xrf} A_{nv}^{q-1} S - (X - D_v^{xrf} A_v^{q-1} S)) S^T \\ &+ \lambda {D_{nv}^{xrf}}^T D_{nv}^{xrf} A_{nv} G G^T) \end{aligned} \tag{15b}$$

$$\{A_v^q, A_{nv}^q\} = prox_{A_v, A_{nv}} (V_v^q, V_{nv}^q), \tag{15c}$$

where $d_1 = \gamma_1 \| D^{rgb} {D^{rgb}}^T \|_F$, $d_2 = \gamma_2 \| D_{nv}^{xrf} {D_{nv}^{xrf}}^T \|_F$ are non-zero step size constants, and $prox_{A_v, A_{nv}}$ is the proximal operator that project V_v^q, V_{nv}^q onto the constraints of Equation 14. The proximal projection is computational inexpensive because it just sets negative entries of V_v^q and V_{nv}^q to zero and scales every column of V_v^q and V_{nv}^q simultaneously to equal one in ℓ_1 norm. Notice that in Equation 15a, only the gradient of the second term in Equation 14 is utilized to update V_v^q, because we want the visible component coefficients A_v to be determined by the RGB image I only, instead of being determined jointly by the RGB image I and the XRF image X.

Second, we optimize over D^{rgb} solving the following constrained least-squares problem:

$$\begin{aligned} \min_{D^{rgb}} \quad & \| I - D^{rgb} A_v \|_F^2 \\ \text{s.t.} \quad & 0 \leq D_{ij}^{rgb} \leq 1, \forall i, j. \end{aligned} \tag{16}$$

Likewise, Equation 16 is minimized by iterating the following steps until convergence:

$$E^q = {D^{rgb}}^{q-1} - \frac{1}{d_{rgb}} ({D^{rgb}}^{q-1} A_v - I) {A_v}^T \tag{17a}$$

$${D^{rgb}}^q = prox_{D^{rgb}} (E^q), \tag{17b}$$

with $d_{rgb} = \gamma_3 \| A_v {A_v}^T \|_F$ again a non-zero step size constant and $prox_{D^{rgb}}$ the proximal operator that projects E^q onto the constraint of Equation 16. The proximal operator this time is also computationally inexpensive since it just truncates the entries of E^q to 0 from below and to 1 from above.

Similarly, D_v^{xrf} is then optimized by solving

$$\begin{aligned} \min_{D_v^{xrf}} \quad & \| (X - D_{nv}^{xrf} A_{nv} S) - D_v^{xrf} A_v S \|_F^2 \\ \text{s.t.} \quad & 0 \leq D_{v\,ij}^{xrf} \leq 1, \forall i, j, \end{aligned} \tag{18}$$

using the following iteration steps:

$$\begin{aligned} U^q = {D_v^{xrf}}^{q-1} \\ &- \frac{1}{d_v^{xrf}} ({D_v^{xrf}}^{q-1} A_v S - (X - D_{nv}^{xrf} A_{nv} S)) S^T {A_v}^T \end{aligned} \tag{19a}$$

$${D_v^{xrf}}^q = prox_{D_v^{xrf}} (U^q), \tag{19b}$$

where $d_v^{xrf} = \gamma_4 \| A_v {A_v}^T \|_F$ is the non-zero step size constant and $prox_{D_v^{xrf}}$ is the proximal operator which project U^q onto the constraints of Equation 18. It is the same as the proximal operator in Equation 17b.

Finally, we optimize D_{nv}^{xrf} by solving the following problem,

$$\begin{aligned} \min_{D_{nv}^{xrf}} \quad & \| (X - D_v^{xrf} A_v S) - D_{nv}^{xrf} A_{nv} S \|_F^2 \\ & + \lambda \| \nabla (D_{nv}^{xrf} A_{nv}) \|_F^2 \\ \text{s.t.} \quad & 0 \leq D_{nv\,ij}^{xrf} \leq 1, \forall i, j. \end{aligned} \tag{20}$$

Likewise, the following two steps are iterated until convergence:

$$\begin{aligned} L^q = {D_{nv}^{xrf}}^{q-1} \\ &- \frac{1}{d_{nv}^{xrf}} ({D_{nv}^{xrf}}^{q-1} A_{nv} S - (X - D_v^{xrf} A_v S)) S^T {A_{nv}}^T \\ &- \lambda D_{nv}^{xrf} A_{nv} G G^T {A_{nv}}^T \end{aligned} \tag{21a}$$

$${D_{nv}^{xrf}}^q = prox_{D_{nv}^{xrf}} (L^q), \tag{21b}$$

where $d_{nv}^{xrf} = \gamma_5 \| A_n v A_n v^T \|_F$ again is a non-zero step size constant, $prox_{D_{nv}^{xrf}}$ is the proximal operator projecting L^q onto the constraints of Equation 20, which is the same proximal operator as the ones in Equations 17b and 19b. The complete optimization scheme is illustrated in Algorithm 1. According to Equations 2, 3, 4, the HR XRF output image Y can be computed by

$$Y = Y_v + Y_{nv} = D_v^{xrf} A_v + D_{nv}^{xrf} A_{nv}. \tag{22}$$

Algorithm 1. Proposed Optimization Scheme of Equation 9

input: LR XRF image X, HR conventional RGB image I.
1: Initialize ${D^{rgb}}^{(0)}$, ${D_v^{xrf}}^{(0)}$ and ${A_v^l}^{(0)}$ by Equation (13);
Initialize ${D_{nv}^{xrf}}^{(0)} = {D_v^{xrf}}^{(0)}$;
Initialize ${A_v}^{(0)}$ by upsampling ${A_v^l}^{(0)}$;
Initialize ${A_{nv}}^{(0)} = \mathbf{0}$;
$n = 0$;
2: **repeat**
3: Estimate ${A_v}^{(n+1)}$ and ${A_{nv}}^{(n+1)}$ with Equation 15;
4: Estimate ${D^{rgb}}^{(n+1)}$ with Equation 17;
5: Estimate ${D_v^{xrf}}^{(n+1)}$ with Equation 19;
6: Estimate ${D_{nv}^{xrf}}^{(n+1)}$ with Equation 21;
7: n=n+1;
8: **until** convergence
output: HR XRF image
$Y = D_v^{xrf} A_v + D_{nv}^{xrf} A_{nv}$.

V. Experimental Results

To verify the performance of our proposed SR method, we have performed extensive experiments on both synthetic and real XRF images. The basic parameters of the proposed SR method are set as follows: the number of atoms in the dictionaries D^{rgb}, D_{nv}^{xrf} and D_v^{xrf} is $M = 50$ for synthetic experiments and $M = 200$ for real experiments; $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 1.01$, which only affects the speed of convergence; parameter β in Equation 13 is set to 0.02 and λ in Equation 9 is set to 0.1. The optional constraint in Equation 9h is not applied here.

A. Error Metrics

As a primary error metric we use, the root mean squared error (RMSE) between the estimated HR XRF image Y and the ground truth image Y^{gt} is computed

$$RMSE = \sqrt{\frac{\|Y - Y^{gt}\|_F^2}{BN_h}}. \tag{23}$$

The peak-signal-to-noise ratio (PSNR) is reported as well,

$$PSNR = 20\log_{10}\frac{\max(Y^{gt})}{RMSE}, \tag{24}$$

where $\max(Y^{gt})$ denoting the maximum entry of Y^{gt}.

The spectral angle mapper (SAM, [28]) in degrees is also utilized, which is defined as the angle in $\mathbb{R}^B$ between an estimated pixel and the ground truth pixel, averaged over the whole image,

$$SAM = \frac{1}{N_h}\sum_{j=1}^{N_h}\arccos\frac{Y(:,j)^T Y^{gt}(:,j)}{\|Y(:,j)\|_2\|Y^{gt}(:,j)\|_2}. \tag{25}$$

B. Comparison Methods

In order to compare over results with the hyperspectral image SR method GSOMP [7], CSUSR [8] and NSSR [9], the linear degradation matrix P mapping the XRF spectrum to its corresponding RGB representation needs to be estimated first. Since these methods do not estimate this linear transformation,

$$I^l = PX, \tag{26}$$

where $I^l \in \mathbb{R}^{b\times N_l}$ is defined in Equation 11 and $X \in \mathbb{R}^{B\times N_l}$ is the input LR XRF image. Although this linear transformation model does not hold for XRF and its corresponding RGB images, we are trying to find the best approximation P so that we can apply the mentioned above hyperspectral image SR methods. The best approximation P can be computed by the following least-squares problem

$$\min_{P} \quad \|PX - I^l\|_F^2. \tag{27}$$

The Trust-Region-Reflective Least Squares algorithm [29] can be utilized to estimate P.

Besides the above mentioned hyperspectral image SR methods, we also propose two baseline methods to compare against, since SR for XRF images is still an open problem. Baseline method $\#1$ only uses LR XRF image as input, increasing its spatial resolution by the same TV regularizer as in Equation 9, by solving

$$\begin{aligned}
&\min_{A,D^{xrf}} \|X - D^{xrf}AS\|_F^2 + \lambda\|\nabla(D^{xrf}A)\|_F^2 && \text{(28a)}\\
&\text{s.t.} \quad 0 \le D_{ij}^{xrf} \le 1, \forall i,j && \text{(28b)}\\
&\qquad A_{ij} \ge 0, \forall i,j && \text{(28c)}\\
&\qquad \mathbf{1^T}A = \mathbf{1^T}, && \text{(28d)}\\
&\qquad \|A\|_0 \le s, && \text{(28e)}
\end{aligned}$$

which is a special case of the proposed optimization problem in Equation 9. A detailed optimization scheme can be found in Appendix A. After solving for D^{xrf} and A, the HR output XRF image Y can be reconstructed by

$$Y = D^{xrf}A. \tag{29}$$

Baseline method $\#2$ does not model the input LR XRF image as a combination of visible and non-visible components, and increases its spatial resolution with a conventional HR RGB image, by solving

$$\begin{aligned}
&\min_{A,D^{xrf},D^{rgb}} \|I - D^{rgb}A\|_F^2 + \|X - D^{xrf}AS\|_F^2 && \text{(30a)}\\
&\text{s.t.} \quad 0 \le D_{ij}^{xrf} \le 1, \forall i,j && \text{(30b)}\\
&\qquad 0 \le D_{ij}^{rgb} \le 1, \forall i,j && \text{(30c)}\\
&\qquad A_{ij} \ge 0, \forall i,j && \text{(30d)}\\
&\qquad \mathbf{1^T}A = \mathbf{1^T}, && \text{(30e)}\\
&\qquad \|A\|_0 \le s, && \text{(30f)}
\end{aligned}$$

which is also a special case of the proposed optimization problem in Equation 9. Detailed optimization scheme can be found in Appendix B. After solving for D^{rgb}, D^{xrf} and A, the HR output XRF image Y can be reconstructed by Equation 29.

C. Synthetic Experiments

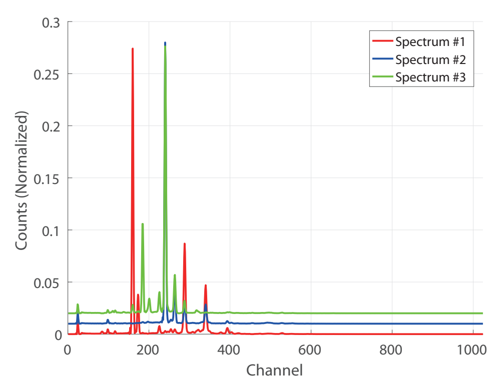
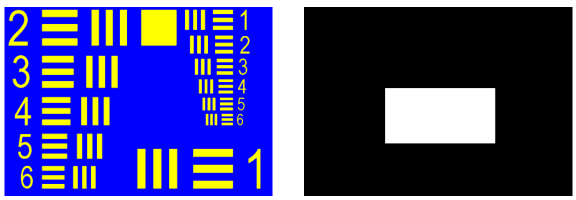


Fig. 4. Three noise free spectra used to synthesize the HR XRF image. Spectra # 2 and # 3 are shifted vertically (by 0.01 and 0.02, respectively) for visualization purposes.

We compare the SR results for different methods with a synthetic experiment first. We combined 3 noise free spectra (1024×1), an HR airforce target image ($345 \times 490 \times 3$) as the visible image and a rectangle image ($345 \times 490 \times 3$) as the non-visible image to simulate the ground truth HR XRF image Y^{gt} ($345 \times 490 \times 1024$). The 3 noise free spectra, HR airforce target image and the rectangle image are shown in Figs. 4, 5 (a) and 5 (b), respectively. In detail, we assume that the yellow foreground of the airforce target image corresponds to spectrum # 1, the blue background of the airforce image corresponds to spectrum #2 and the white foreground of the rectangle image corresponds to spectrum #3. The LR XRF image X ($69 \times 98 \times 1024$) was obtained by spatially down-sampling Y^{gt} by a factor of 5 in each dimension and adding Gaussian noise to it with SNR 35dB.



(a) (b)

Fig. 5. (a) The airforce image is utilized as the visible component. (b) The rectangle image is utilized as the non-visible component.

The RMSE, PSNR and SAM metrics were computed between the SR results of different methods described in Section V-B and the HR ground truth Y^{gt}. The default parameters of methods GSOMP [7], CSUSR [8] and NSSR [9] in their original paper were applied in our synthetic experiments. Optimal parameter λ of Baseline #1 method (Equation 28) and the proposed method (Equation 9) was found experimentally. To make fair comparisons, the number of atoms in the dictionary is set to be 50 for all methods (GSOMP [7], CSUSR [8], NSSR [9], Baseline # 2 and the proposed method) that utilize dictionaries. As shown in Table I, our proposed method has the smallest RMSE, highest PSNR and smallest SAM. By comparing Baseline #1 method with the proposed method, the benefit of utilizing an HR RGB image can be validated. By comparing Baseline #2 method with the proposed method, it can be seen that a better and more realistic model that assumes the XRF signal is a combination of visible component and non-visible component is beneficial to obtain better SR results. The traditional hyperspectral image SR methods (GSOMP [7], CSUSR [8] and NSSR [9]) rely on an accurate linear degradation model from hyperspectral to RGB signals. When the degradation model is not accurate, their performance is inferior to our proposed method. Baseline #2 can be considered an extension of CSUSR [8], in that we learn the coupled RGB and XRF dictionaries simultaneously and we do not utilize the linear degradation model, which is a more flexible approach and produces better SR performance.

Fig. 6 shows the average PSNR curves as a function of the

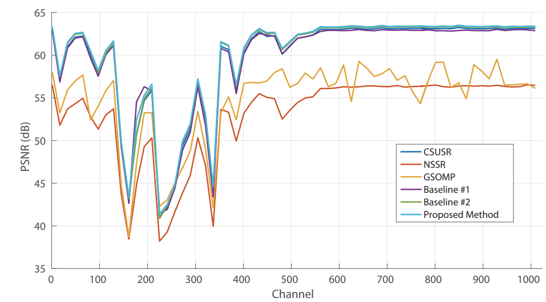


Fig. 6. The average PSNR curves as a function of the channels of the spectral bands for the SR method.

channels of the spectral bands for the test methods. It can be seen that the hyperspectral SR methods GSOMP [7] and NSSR [9] produce inferior SNR over all spectral channels. All other methods perform well for spectral bands outside the range $[100\ 400]$ and our proposed method constantly outperforms all other methods. For spectral bands in the range $[100\ 400]$, both methods' performance decreases because of the overlapping spectra, as shown in Fig. 4. Notice that Baseline #1 slightly outperforms the proposed method around channel #200, which is because there is a peak for both the non-visible (Spectrum #3 in Fig. 4) and visible component spectrum (Spectrum #1 in Fig. 4) around channel #200. The proposed method makes errors in separating these two peaks, resulting in worse performance than Baseline #1 which avoids explicit visible/non-visible decomposition.

We compare the visual quality of different SR methods on the region of interest of channel # 210 - 230 in Fig. 7. Because GSOMP, CSUR, and NSSR hyperspectral image SR methods [7]–[9] rely on an accurate linear degradation model from hyperspectral to RGB signals, SR results are poor. Baseline method #1 did not utilize the HR RGB image in SR and so failed to reconstruct fine details. Baseline method #2 assumes one-to-one mapping between RGB and XRF signals, thus it produced artifacts in the region where the visible and non-visible components overlap. Our proposed method produced the SR result closest to ground truth. Notice that the non-visible component (rectangle) is more blurry than the visible component (airforce target), since it is super-resolved by a TV regularizer and does not use any HR RGB image information.

D. Real Experiments

For our first real experiment, the real data was collected by a Bruker M6 scanning energy dispersive XRF instrument, with 4096 channels in spectrum. Studies from XRF image #3

Metric	GSOMP [7]	CSUSR [8]	NSSR [9]	Baseline #1	Baseline #2	Proposed
RMSE	3.42	0.70	3.85	2.03	0.59	**0.50**
PSNR	37.51	51.36	36.46	42.03	52.83	**54.12**
SAM	22.78	3.19	18.46	8.38	2.10	**2.00**

TABLE I
EXPERIMENTAL RESULTS ON SYNTHETIC DATA COMPARING DIFFERENT SR METHODS DISCUSSED IN SECTION V-B IN TERMS OF RMSE, PSNR AND SAM. BEST RESULTS ARE SHOWN IN BOLD.

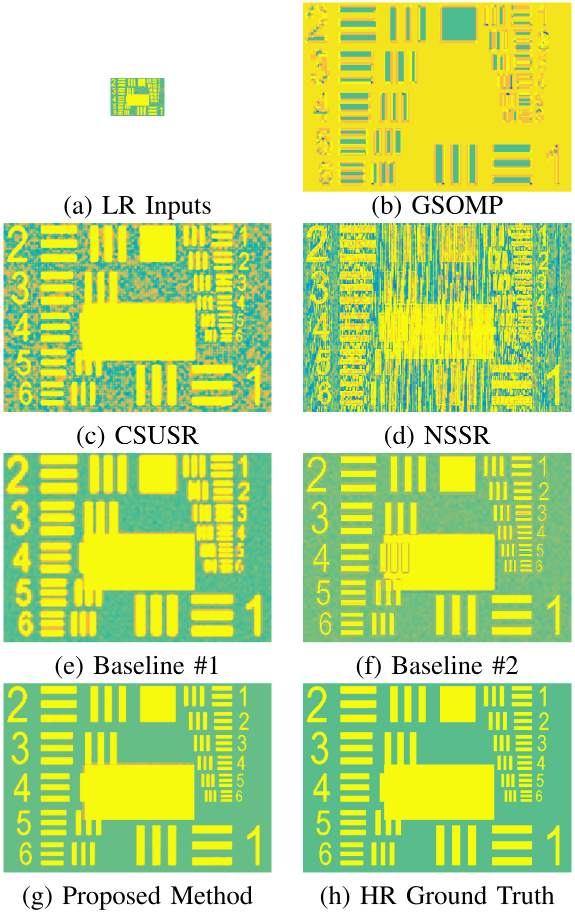


Fig. 7. Visualization of the SR result of the synthetic experiment. Region of interest of channel #210 - 230 is selected. (a) is the LR input XRF image. (b), (c), (d), (e), (f), (g) are the SR result of GSOMP [7], CSUSR [8], NSSR [9], Baseline #1, Baseline #2 and proposed method, respectively. (h) is the HR ground truth image.

scanned from Vincent Van Gogh's "Bedroom" (Fig. 1) are presented here.

We first validated that the proposed method in Equation (9) can accurately represent the XRF spectrum, and that the reconstructed spectral signal has a higher SNR compared to the original spectral signal.

As shown in Fig. 8, our proposed approach provides accurate reconstruction of the original signal. The XRF dictionary is trained from all spectral signals of the XRF image based on minimizing the Euclidean distance between the reconstructed and the original signals. As a result, noise is reduced, and the reconstructed signal has higher SNR compared to the original signal.

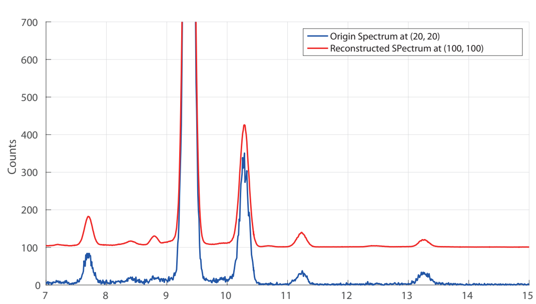


Fig. 8. The reconstruction of a spectrum using our proposed method. The reconstructed spectrum is shifted vertically (100 counts) for visualization purposes.

For our first real experiment, HR ground truth was not available to assess the quality of the reconstructed HR XRF images. This is because all XRF maps we had access to were low resolution and noisy. We compare the visual quality of different SR methods on the region of interest of channel # 611 - 657, corresponding to CrK XRF peak, in Fig. 9. Hyperspectral SR method GSOMP [7] produced a noisy output in (c), because it relies on an accurate degradation model from XRF signal to RGB signal. Hyperspectral SR methods CSUSR [8] and NSSR [9] update the XRF dictionary to ensure the fidelity to LR input, so they produce less noise as compared to GSOMP [7]. However, they either create non existing content in (d) or lose existing content in (e), in the towel regions. Baseline method #1 creates a blurry SR result, since it does not utilize an HR RGB image. Also it fails to resolve the fine detail in the towel region. Baseline method #2 produced visually satisfactory SR results, but failed to reconstruct the line between the wall and the floor. This is because of the one-to-one mapping assumption incorrectly maps brown pixels in the table and the line between the wall and the floor to the same XRF spectra. Our proposed method in (h) produces both a visually satisfactory result as well as strong similarity with the original LR input (a).

For our second real experiment, the real data was collected by a home-built X-ray fluorescence spectrometer (courtesy of Prof. Koen Janssens), with 2048 channels in spectrum. Studies from XRF image scanned from Jan Davidsz. de Heem's "Bloemen en insecten" (ca 1645), in the collection of Koninklijk Museum voor Schone Kunsten (KMKSA) Antwerp, are presented here. The original XRF image has dimension $680 \times 580 \times 2048$. We first spatially downsample the original XRF image by factor 5 and obtain the input LR XRF image with dimension $136 \times 116 \times 2048$. Then different SR methods are applied to increase the spatial resolution of the LR XRF image by factor 5. Notice that because the original HR XRF image is noisy and blurry, it is different from the HR ground truth. However, we can still use it as a reference to compute the RMSE, PSNR and SAM metrics to quantitatively compare the performance of different SR methods. We can also use it as a reference to visually compare different SR results with the original HR XRF image.

As shown in Table II, our proposed method provides the closest reconstruction compared to all other methods. The traditional hyperspectral image SR methods (GSOMP [7] and NSSR [9]) produce considerably greater reconstruction error. Baseline #2 does not assume linear transformation model from XRF spectrum to RGB and updates the XRF and RGB endmembers simultaneously, resulting in better SR results. Baseline method #1 produces SR results more similar to the original HR XRF image compared to Baseline method #2, since both SR results of Baseline method #1 and the original HR XRF image are blurry. Finally, our proposed method produces a result most similar to the input HR XRF image,

Metric	GSOMP [7]	CSUSR [8]	NSSR [9]	Baseline #1	Baseline #2	Proposed
RMSE	75.18	70.20	79.72	70.35	70.43	**69.83**
PSNR	42.70	55.66	49.70	56.06	54.93	**56.19**
SAM	32.60	12.30	25.81	11.60	12.98	**11.32**

TABLE II
EXPERIMENTAL RESULTS ON "BLOEMEN EN INSECTEN" COMPARING DIFFERENT SR METHODS DISCUSSED IN SECTION V-B IN TERMS OF RMSE, PSNR AND SAM. BEST RESULTS ARE SHOWN IN BOLD.

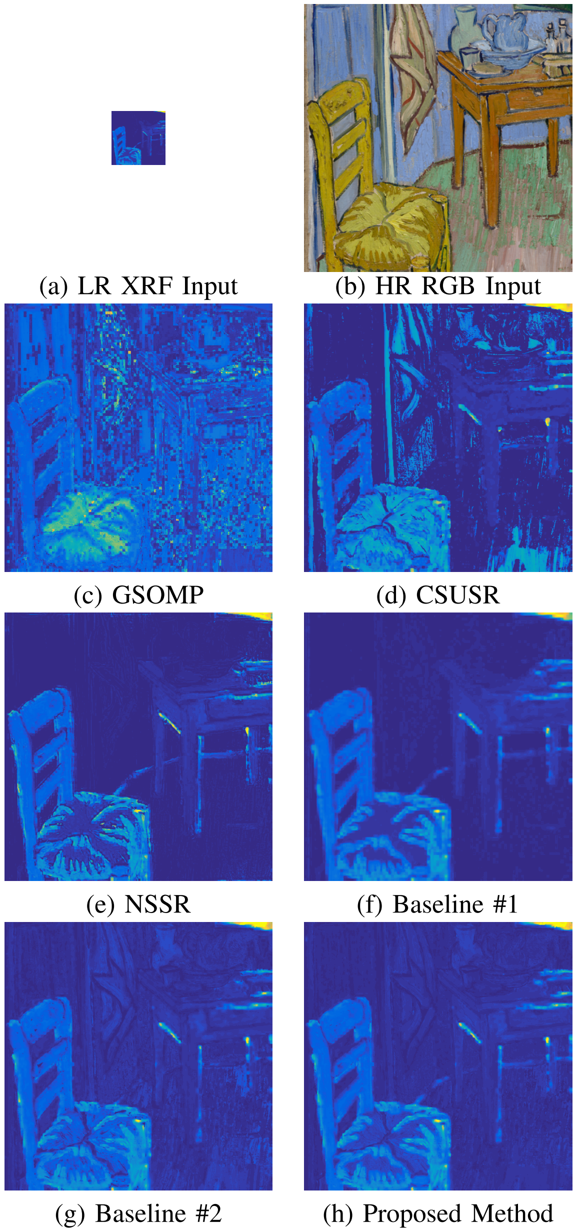


Fig. 9. Visualization of the SR result of the real experiment on the "Bedroom". Region of interest of related to CrK peak (channel #611 - 657) is selected. (a) is the LR input XRF image and (b) is the HR input RGB image. (c), (d), (e), (f), (g), (h) are the SR result of GSOMP [7], CSUSR [8], NSSR [9], Baseline #1, Baseline #2 and proposed method, respectively.

demonstrating the effectiveness of our proposed method.

The visual quality of different SR methods on the region of interest of channel #582 - 602, corresponding to $Pb\ L\eta$ XRF emission line, is compared in Fig. 10. Notice that the two long rectangles in the origin HR XRF image (h) are the stretcher bars under the canvas, which is not visible on the RGB image. Hyperspectral SR methods CSUSR [8] and GSOMP [7] in (c) and (d) produce noisy results and produce visible artifacts in many regions again. Baseline method #1 in (e) improves SNR compared to the origin HR XRF image. However, its SR result is blurry and fails to resolve the details on the flowers. Baseline method #2 in (f) utilizes HR RGB image as input, so its SR result is sharp and many details are resolved. However, because it does not model the non-visible component of the XRF image, it fails to precisely reconstruct the two hidden stretcher bars. Also when compared to the origin HR XRF image (h), it produces many artifacts, such as the textures of the flower in the middle, edges and stems. Our proposed method in (g) successfully reconstructs the non-visible stretcher bars. The reconstructed stretcher bars are blurry compared to other objects, because it does not utilize any information from the HR RGB image. More details are resolved by our proposed method. When compared to the origin HR image (h), we can conclude that those resolved details have high fidelity to the original HR image (h). The SNR is also improved by our proposed method.

VI. CONCLUSION

In this paper we presented a novel XRF image SR framework based on fusing an HR conventional RGB image. The XRF spectrum of each pixel is represented by an endmembers dictionary, as well as the RGB spectrum. We also decomposed the input LR XRF image into visible and non-visible components, making it possible to find the non-linear mapping from RGB spectrum to XRF spectrum. The non-visible component is super-resolved using a standard total variation regularizer. The HR visible XRF component and HR non-visible XRF component are combined to obtain the final HR XRF image. Both synthetic and real experiments show the effectiveness of our proposed method.

ACKNOWLEDGMENT

This work was supported in part by a grant from the Northwestern University and Art Institute of Chicago Center for Scientific Studies in the Arts (NU-ACCESS). The authors would like to thank Prof. Francesca Casadio from Art Institute of Chicago for supplying the "Bedroom" data and Prof. Koen Janssens from University of Antwerp for supplying the "Bloemen en insecten" data.

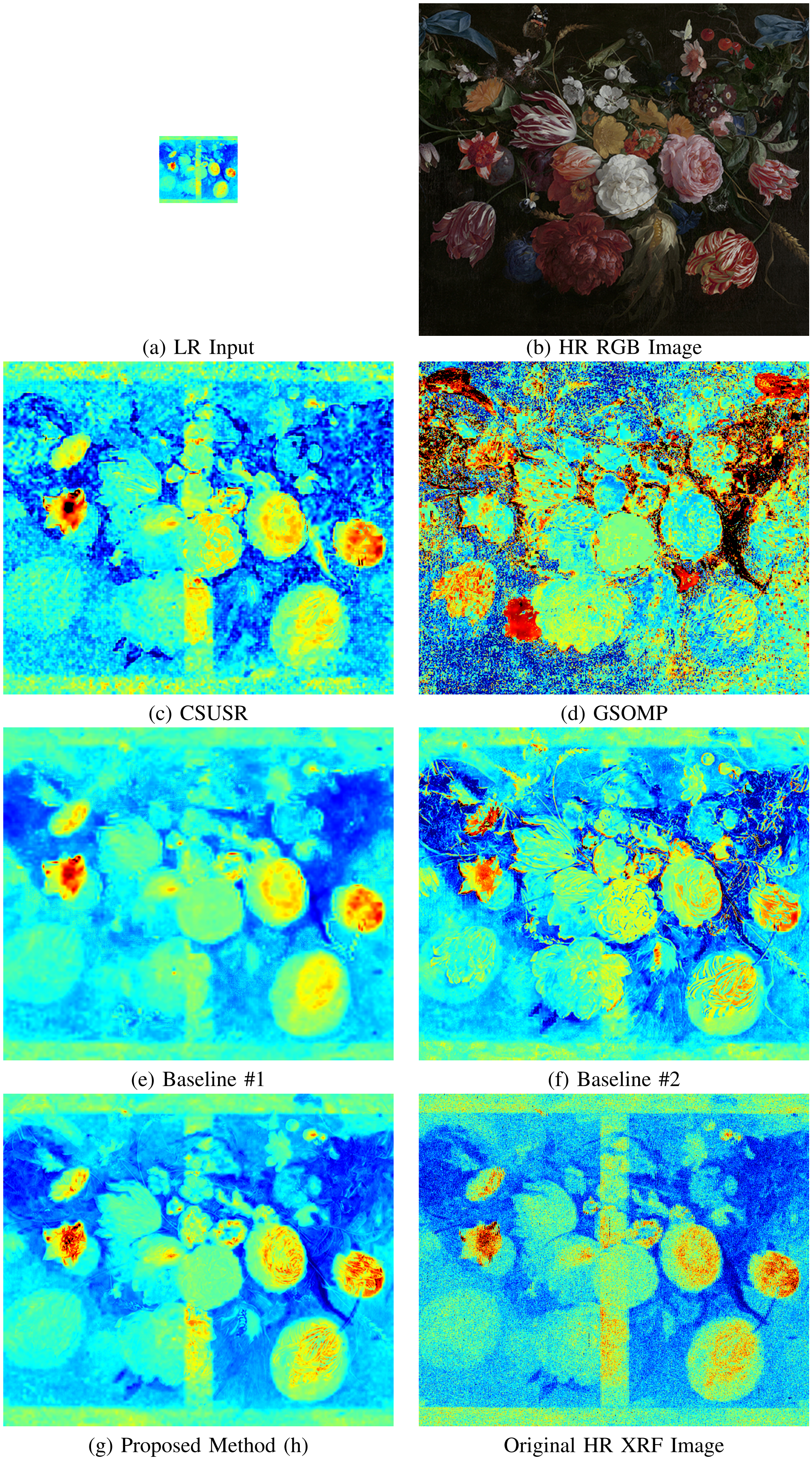


Fig. 10. Visualization of the SR result of the DeHeem real experiment on the "Bloemen en insecten". Region of interest of related to $Pb\ L\eta$ XRF emission line (channel #582 - 602) is selected. (a) is the LR input XRF image and (b) is the HR input RGB image. (c), (d), (e), (f), (g) are the SR result of CSUSR [8], GSOMP [7], Baseline #1, Baseline #2 and proposed method, respectively. (h) is the original HR XRF image. Readers are suggested to zoom in in order to compare the details of different results.

References

[1] M. Alfeld, J. V. Pedroso, M. van Eikema Hommes, G. Van der Snickt, G. Tauber, J. Blaas, M. Haschke, K. Erler, J. Dik, and K. Janssens, "A mobile instrument for in situ scanning macro-xrf investigation of historical paintings," *Journal of Analytical Atomic Spectrometry*, vol. 28, no. 5, pp. 760–767, 2013.

[2] A. Anitha, A. Brasoveanu, M. Duarte, S. Hughes, I. Daubechies, J. Dik, K. Janssens, and M. Alfeld, "Restoration of x-ray fluorescence images of hidden paintings," *Signal Processing*, vol. 93, no. 3, pp. 592–604, 2013.

[3] C. M. Pieters and P. A. Englert, *Remote geochemical analysis, elemental and mineralogical composition*, 1993, vol. 1.

[4] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 7, pp. 1392–1409, 2001.

[5] M. Alfeld, W. De Nolf, S. Cagno, K. Appel, D. P. Siddons, A. Kuczewski, K. Janssens, J. Dik, K. Trentelman, M. Walton *et al.*, "Revealing hidden paint layers in oil paintings by means of scanning macro-xrf: a mock-up study based on rembrandt's an old man in military costume," *Journal of Analytical Atomic Spectrometry*, vol. 28, no. 1, pp. 40–51, 2013.

[6] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2329–2336.

[7] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 63–78.

[8] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3586–3594.

[9] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," 2016.

[10] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *Image Processing, IEEE Transactions on*, vol. 13, no. 9, pp. 1174–1184, 2004.

[11] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3176–3180.

[12] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3631–3640.

[13] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation super resolution using a variational approach," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 641–644.

[14] A. Marquina and S. J. Osher, "Image super-resolution by tv-regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.

[15] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang, "Bilevel sparse coding for coupled feature spaces," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2360–2367.

[16] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3467–3478, 2012.

[17] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[18] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 184–199.

[19] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian super resolution," *Image Processing, IEEE Transactions on*, vol. 20, no. 4, pp. 984–999, 2011.

[20] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 6, pp. 1391–1402, 2005.

[21] A. Garzelli, F. Nencini, L. Alparone, B. Aiazzi, and S. Baronti, "Pansharpening of multispectral images: a critical review and comparison," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International*, vol. 1. IEEE, 2004.

[22] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 10, pp. 3012–3021, 2007.

[23] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354–379, 2012.

[24] N. Keshava and J. F. Mustard, "Spectral unmixing," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 44–57, 2002.

[25] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[26] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.

[27] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[28] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," 1992.

[29] T. F. Coleman and Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, no. 4, pp. 1040–1058, 1996.

[30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 689–696.

Appendix A
Optimization scheme for Baseline #1

Let $A^l \in \mathbb{R}^{M \times N_l}$ be the spatially downsampled abundance A. The dictionary learning technique in [30] can be applied to initialize D^{xrf} and A^l by solving

$$\begin{aligned} \min_{D^{xrf},A^l} \quad & \|X - D^{xrf}A^l\|_F^2 + \beta \textstyle\sum_{k=1}^{N_l} \|A^l(:,k)\|_1, \\ \text{s.t.} \quad & \|D^{xrf}(:,k)\|_2 \le 1, \forall k. \end{aligned} \tag{31}$$

D^{xrf} is initialized using Equation 31 and A is initialized by upsampling A^l computed in Equation 31.

Similar to the optimization scheme of our proposed method (Equation 14), Equation 28 can be alternatively optimized. First we optimize over A by fixing D^{xrf},

$$\min_{A} \quad \|X - D^{xrf}AS\|_F^2 + \lambda \|\nabla(D^{xrf}A)\|_F^2 \tag{32a}$$
$$\text{s.t.} \quad A_{ij} \ge 0, \forall i,j \tag{32b}$$
$$\mathbf{1^T}A = \mathbf{1^T}, \tag{32c}$$
$$\|A\|_0 \le s, \tag{32d}$$

PALM is utilized to optimize over A. For Equation 32, the following two steps are iterated until convergence:

$$\begin{aligned} V^q = A^{q-1} & \\ & - \frac{1}{d}({D^{xrf}}^T(D^{xrf}A^{q-1}S - X)S^T \\ & + \lambda {D^{xrf}}^T D^{xrf} AGG^T) \end{aligned} \tag{33a}$$
$$A^q = prox_A(V^q), \tag{33b}$$

where $d_2 = \gamma_2 \|D^{xrf}{D^{xrf}}^T\|_F$ are non-zero step size constants, and $prox_A$ is the proximal operator that project V^q onto the constraints of Equation 32.

We then optimize over D^{xrf} solving the following constrained least-squares problem:

$$\begin{aligned} \min_{D^{xrf}} \quad & \|X - D^{xrf}AS\|_F^2 \\ \text{s.t.} \quad & 0 \le D_{ij}^{xrf} \le 1, \forall i,j, \end{aligned} \tag{34}$$

using the following iteration steps:

$$\begin{aligned} U^q = {D^{xrf}}^{q-1} & \\ & - \frac{1}{d^{xrf}}({D^{xrf}}^{q-1}AS - X)S^T A^T \end{aligned} \tag{35a}$$
$${D^{xrf}}^q = prox_{D^{xrf}}(U^q), \tag{35b}$$

where $d^{xrf} = \gamma_4 \|AA^T\|_F$ is the non-zero step size constant and $prox_{D^{xrf}}$ is the proximal operator which project U^q onto the constraints of Equation 34.

The complete optimization scheme is demonstrated in Algorithm 2.

Algorithm 2. Proposed Optimization Scheme of Equation 28

input: LR XRF image X.
1: Initialize ${D^{xrf}}^{(0)}$ and ${A^l}^{(0)}$ by Equation (31);
Initialize $A^{(0)}$ by upsampling ${A^l}^{(0)}$;
$n = 0$;
2: **repeat**
3: Estimate $A^{(n+1)}$ with Equation 33;
4: Estimate ${D^{xrf}}^{(n+1)}$ with Equation 36;
5: n=n+1;
6: **until** convergence
output: HR XRF image
$Y = D^{xrf}A$.

Appendix B
Optimization scheme for Baseline #2

For Equation 30, A, D^{xrf} and D^{rgb} can be initialized by Equation 13. We then alternatively optimize the unknowns in Equation 30. We first update A based on the RGB image by fixing all other parameters,

$$\min_{A} \quad \|I - D^{rgb}A\|_F^2 \tag{36a}$$
$$\text{s.t.} \quad A_{ij} \ge 0, \forall i,j \tag{36b}$$
$$\mathbf{1^T}A = \mathbf{1^T}, \tag{36c}$$
$$\|A\|_0 \le s, \tag{36d}$$

utilizing the following iteration steps:

$$V^q = A^{q-1} - \frac{1}{d}{D^{rgb}}^T(D^{rgb}A^{q-1} - I) \tag{37a}$$
$$A^q = prox_A(V^q), \tag{37b}$$

where $d = \gamma_1 \|D^{rgb}{D^{rgb}}^T\|_F$ is non-zero step size constants, and $prox_A$ is the proximal operator that project V^q onto the constraints of Equation 36.

We then update D^{rgb}

$$\begin{aligned} \min_{D^{rgb}} \quad & \|I - D^{rgb}A\|_F^2 \\ \text{s.t.} \quad & 0 \le D_{ij}^{rgb} \le 1, \forall i,j. \end{aligned} \tag{38}$$

by the following iteration steps:

$$E^q = {D^{rgb}}^{q-1} - \frac{1}{d_{rgb}}({D^{rgb}}^{q-1}A - I)A^T \tag{39a}$$
$${D^{rgb}}^q = prox_{D^{rgb}}(E^q), \tag{39b}$$

with $d_{rgb} = \gamma_3 \|AA^T\|_F$ again a non-zero step size constant and $prox_{D^{rgb}}$ the proximal operator that projects E^q onto the constraint of Equation 38.

Finally we update D^{xrf}

$$\begin{aligned} \min_{D^{xrf}} \quad & \|X - D^{xrf}AS\|_F^2 \\ \text{s.t.} \quad & 0 \le D_{ij}^{xrf} \le 1, \forall i,j, \end{aligned} \tag{40}$$

using the following iteration steps:

$$U^q = {D^{xrf}}^{q-1} - \frac{1}{d^{xrf}}({D^{xrf}}^{q-1}AS - X)S^T A^T \tag{41a}$$

$$ {D^{xrf}}^q = prox_{D^{xrf}}(U^q), \tag{41b}$$

where $d^{xrf} = \gamma_4 \|AA^T\|_F$ is the non-zero step size constant and $prox_{D^{xrf}}$ is the proximal operator which project U^q onto the constraints of Equation 40.

The complete optimization scheme is summarized in Algorithm 3.

Algorithm 3. Proposed Optimization Scheme of Equation 30

input: LR XRF image X, HR conventional RGB image I.
1: Initialize ${D^{rgb}}^{(0)}$, ${D^{xrf}}^{(0)}$ and ${A^l}^{(0)}$ by Equation (13);
Initialize $A^{(0)}$ by upsampling ${A^l}^{(0)}$;
$n = 0$;
2: **repeat**
3: Estimate $A^{(n+1)}$ with Equation 37;
4: Estimate ${D^{rgb}}^{(n+1)}$ with Equation 39;
5: Estimate ${D^{xrf}}^{(n+1)}$ with Equation 41;
6: n=n+1;
7: **until** convergence
output: HR XRF image
$Y = D^{xrf}A$.