Multi-Modal Dictionary Learning for Image Separation With Application In Art Investigation

Nikos Deligiannis, *Member, IEEE,* João F. C. Mota, *Member, IEEE,* Bruno Cornelis, *Member, IEEE,* Miguel R. D. Rodrigues, *Senior Member, IEEE,* and Ingrid Daubechies, *Fellow, IEEE*

***Abstract*—In support of art investigation, we propose a new source separation method that unmixes a single X-ray scan acquired from double-sided paintings. In this problem, the X-ray signals to be separated have similar morphological characteristics, which brings previous source separation methods to their limits. Our solution is to use photographs taken from the front- and back-side of the panel to drive the separation process. The crux of our approach relies on the coupling of the two imaging modalities (photographs and X-rays) using a novel coupled dictionary learning framework able to capture both common and disparate features across the modalities using parsimonious representations; the common component captures features shared by the multi-modal images, whereas the innovation component captures modality-specific information. As such, our model enables the formulation of appropriately regularized convex optimization procedures that lead to the accurate separation of the X-rays. Our dictionary learning framework can be tailored both to a single- and a multi-scale framework, with the latter leading to a significant performance improvement. Moreover, to improve further on the visual quality of the separated images, we propose to train coupled dictionaries that ignore certain parts of the painting corresponding to craquelure. Experimentation on synthetic and real data—taken from digital acquisition of the Ghent Altarpiece (1432)—confirms the superiority of our method against the state-of-the-art morphological component analysis technique that uses either fixed or trained dictionaries to perform image separation.**

***Index Terms*—Source separation, coupled dictionary learning, multi-scale image decomposition, multi-modal data analysis.**

Figure 1. Panels from the *Ghent Altarpiece*: (left) panels of *Adam* and *Eve*, (centre) the respective paintings on the back, (right) corresponding X-ray images containing a mixture of components. ©KIK-IRPA

I. INTRODUCTION

BIG DATA sets—produced by scientific experiments or projects—often contain heterogeneous data obtained by capturing a physical process or object using diverse sensing modalities [2]. The result is a rich set of signals, heterogeneous in nature, but strongly correlated due to a common underlying phenomenon. Multi-modal signal processing and analysis is thus gaining momentum in various research disciplines ranging from medical diagnosis [3] to remote sensing and computer vision [4]. In particular, the analysis of high-resolution multi-modal digital acquisitions of paintings in support of art scholarship has proved a challenging field of research. Examples include the numerical characterization of brushstrokes [5], [6] for the authentication or dating of paintings, canvas thread counting [7]–[9] with applications in art forensics, and the (semi-) automatic detection and digital inpainting of cracks [10]–[12].

The *Lasting Support* project has focused on the investigation of the *Ghent Altarpiece* (1432), also known as *The Adoration of the Mystic Lamb*, a polyptych on wood panel painted by Jan and Hubert van Eyck. As one of the most admired and influential masterpieces in the history of art, it has given rise to many puzzling questions for art historians. Currently, the *Ghent Altarpiece* is undergoing a major conservation and restoration campaign that is planned to end in 2017. The panels of the masterpiece were documented with various imaging modalities, amongst which visual macrophotography, infrared macrophotography, infrared reflectography and X-radiography [12]. A massive visual data set (comprising over 2TB of data) has been compiled by capturing small areas of the polyptych separately and stitching the resulting image blocks into one image per panel [13].

X-ray images are common tools in painting investigation,

The work is supported by the VUB research programme M3D2, the EPSRC grant EP/K033166/1, and the VUB-UGent-UCL-Duke International Joint Research Group. A preliminary version of this work has been presented at the IEEE International Conference on Image Processing (ICIP) 2016 [1].

N. Deligiannis and B. Cornelis are with the Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels 1050, Belgium, and with iMinds, 9050 Ghent, Belgium (e-mail: {ndeligia, bcorneli}@etrovub.be). B. Cornelis was with the Department of Mathematics, Duke University, Durham, NC 27708 USA.

J. Mota and M. Rodrigues are with the Department of Electronic and Electrical Engineering, University College London, UK (e-mail: j.mota,m.rodrigues@ucl.ac.uk).

I. Daubechies is with the Departments of Mathematics and Electronic and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: ingrid@math.duke.edu).

since they reveal information about the composition of the materials, the variations in paint thickness, the support as well as the cracks and losses in the ground and paint layers. The problem we address in this paper relates to the outer side panels, namely, the panels showing near life-sized depictions of Adam and Eve, shown in Fig. 1. Due to X-ray penetration, the scans of these panels are a mixture of the paintings from each side of the panel as well as the wood panel itself. The presence of all these components makes the reading of the X-ray image difficult for art experts, who would welcome an effective approach to separate the components.

The task of separating a mixture of signals into its constituent components is a popular field of research. Most work addresses the blind source separation (BSS) problem, where the goal is to retrieve the different sources from a set of mixed signals, without information about the source signals or the mixing process. Several methods attempt to solve the BSS problem by imposing constraints on the sources' structure. Independent component analysis (ICA) [14], [15] commonly assumes that the components are independent non-Gaussian signals. Nonnegative matrix factorization is another approach to solve the problem, where it is assumed that the sources are nonnegative (or they are transformed to a nonnegative representation) [16]. In an alternative path, the problem has been cast into a Bayesian framework, where either the sources are viewed as latent variables [17], or the problem is solved by maximizing the joint posterior density of the sources [18]. Under the Bayesian methodology, spatial smoothing priors (via, for example, Markov random fields) have been used to regularize blind image separation problems [19]. These assumptions do not fit our particular problem as both components have similar statistical or structural properties.

Sparsity is another source prior heavily exploited in BSS [20], [21], as well as in various other inverse problems, such as, compressed sensing [22], [23], image inpainting [24], [25], denoising [26], and deconvolution [27]. Morphological component analysis (MCA), in particular, is a state-of-the-art sparsity-based regularization method, initially designed for the single-mixture problem [20], [28] and then extended to the multi-mixture case [29]. The crux of the method is the basic assumption that each component has its own characteristic morphology; namely, each has a highly sparse representation over a set of bases (or, *dictionaries*), while being highly non-sparse for the dictionaries of the other components. Prior work in digital painting analysis has employed MCA to remove cradling artifacts within X-ray images of paintings on a panel [30]. The cradling and painting components have very different morphologies, captured by different predefined dictionaries. Namely, complex wavelets [31] provide a sparse representation for the smooth X-ray image and shearlets [32] were used to represent the texture of the wood grain. Alternatively, dictionaries can be learned from a set of training signals; several algorithms have been proposed to construct dictionaries including the method of optimal directions (MOD) [33] and the K-SVD algorithm [34]. Both follow a similar sparse decomposition approach; however, they differ in the way the dictionary elements are learned. Recently, multi-mixture MCA has been combined with K-SVD, resulting in a method where dictionaries are learned adaptively while separating [35].

However, in our particular separation problem we have a simple mixture of two X-ray components that are morphologically very similar [see Fig. 1]. Hence, as we will show in the experimental section, simply using fixed or learned dictionaries is insufficient to discriminate one component from the other. Unlike prior work, in our setup we have access to high-quality photographic material from each side of the panel that can be used to assist the X-ray image separation process.

In this work, we elaborate on a novel method to perform separation of X-ray images from a single mixture by using images of another modality as side information. Our contributions are as follows:

- We present a new model based on parsimonious representations, which captures both the inherent similarities and the discrepancies among heterogeneous correlated data. The model decomposes the data into a sparse component that is common to the different modalities and a sparse component that uniquely describes each data type. Our model enables the formulation of appropriately regularized convex optimization procedures that address the separation problem at hand.
- We propose a novel dictionary learning approach that trains dictionaries coupling the images from the different modalities. Our approach introduces a new modified OMP algorithm that is tailored to our data model.
- We devise a novel method that ignores craquelure pixels—namely, pixels that visualize cracks in the surface of paintings—when learning coupled dictionaries. Paying no heed to these pixels avoids contaminating the dictionaries with high frequency noise, thereby leading to higher separation performance. Our approach bears similarities with inpainting approaches, e.g., [25]; it is, however, different in the way the dictionary learning problem is posed and solved.
- We devise a novel multi-scale image separation strategy that is based on a recursive decomposition of the mixed X-ray and visual images into low- and high-pass bands. As such, the method enables the accurate separation of high-resolution images even when a local sparsity prior is assumed. Our approach differs from existing multi-scale dictionary learning methods [25], [36], [37] not only by considering imaging data gleaned from diverse modalities but also in the way the multi-scale decomposition is constructed.
- We conduct experiments using synthetic and real data proving that the use of side information is crucial in the separation of X-ray images from double-sided paintings.

In the remainder of the paper: Section II reviews related work and Section III poses our source separation with side information problem. Section IV describes the proposed coupled dictionary learning algorithm. Section V presents our method that ignores cracks when learning dictionaries, and Section VI elaborates on our single- and multi-scale approaches to X-ray image separation. Section VII presents the evaluation of our algorithms, while Section VIII draws our conclusions.

II. Related Work

A. Source Separation

Adhering to a formal definition, MCA [20], [28] decomposes a source or image mixture $x = \sum_{i=1}^{\kappa} x_i$, with $x, x_i \in \mathbb{R}^{n\times 1}$, into its constituents, with the assumption that each x_i admits a sparse decomposition in a different overcomplete dictionary $\Phi_i \in \mathbb{R}^{n\times d_i}$, $(n \ll d_i)$. Namely, each component can be expressed as $x_i = \Phi_i z_i$, where $z_i \in \mathbb{R}^{d_i\times 1}$ is a sparse vector comprising a few non-zero coefficients: $\|z_i\|_0 = \#\{\xi : z_{i_\xi} \neq 0, \xi = 1,\ldots,d_i\} = s_i \ll d_i$, with $\|\cdot\|_0$ denoting the ℓ_0 pseudo-norm. The BSS problem is thus addressed as [20], [28]

$$(\hat{z}_1,\ldots,\hat{z}_\kappa) = \arg\min_{z_1,\ldots,z_\kappa} \sum_{i=1}^{\kappa} \|z_i\|_0 \;\text{ s.t. }\; x = \sum_{i=1}^{\kappa} \Phi_i z_i. \quad (1)$$

Unlike the BSS problem, informed source separation (ISS) methods utilise some form of prior information to aid the task at hand. ISS methods are tailored to the application they address (to the best of our knowledge they are applied only for audio mixtures [38], [39]). For instance, an encoding/decoding framework is proposed in [38], where the sources are mixed at the encoder and the mixtures are sent to the decoder together with side information that is embedded by means of quantization index modulation (QIM) [40]. Unlike these methods, we propose a generic source separation framework that incorporates side information gleaned from a correlated heterogeneous source by means of a new dictionary learning method that couples the heterogenous sources.

B. Dictionary Learning

Dictionary learning factorizes a matrix composed of training signals $X = [x_1,\ldots,x_k] \in \mathbb{R}^{n\times k}$ into the product ΦZ as

$$\Big(\Phi, Z\Big) = \arg\min_{\Phi', Z'} \|X - \Phi' Z'\|_F^2 \;\text{ s.t. }\; \|z_i\|_0 \le s,\; i = 1,\ldots,k, \quad (2)$$

where $Z = [z_1,\ldots,z_k] \in R^{d\times k}$ contains the sparse vectors corresponding to the signals $X = [x_1,\ldots,x_k]$ and $\|\cdot\|_F$ is the Frobenius norm of a matrix. The columns of the dictionary Φ are typically constrained to have unit norm so as to improve the identifiability of the dictionary. To solve Problem (2), which is non-convex, Olshausen and Field [41] proposed to iterate through a step that learns the sparse codes and a step that updates the dictionary elements. The same strategy is followed in subsequent studies [33], [34], [42]–[44]. Alternatively, polynomial-time algorithms that are guaranteed to reach a globally optimal solution appear in [45], [46].

In order to capture multi-scale traits in natural signals, a method to construct multi-scale dictionaries was presented in [25]. The multi-scale representation was obtained by using a quadtree decomposition of the learned dictionary. Alternatively, the work in [36], [37] applied dictionary learning in the domain of a fixed multi-scale operator (wavelets). In our approach we follow a different multi-scale strategy, based on a pyramid decomposition, similar to the Laplacian pyramid [47].

There exist dictionary learning approaches designed to couple multi-modal data. Monaci *et al.* [48] proposed an approach to learn basis functions representing audio-visual structures. Alternatively, Yang *et al.* [49], [50] considered the problem of learning two dictionaries, D_x and D_y, for two families of signals, x and y, coupled by a mapping function $\mathcal{F}$ [with $y = \mathcal{F}(x)$]. The constraint was that the sparse representation of x in D_x is the same as that of y in D_y. The application targeted was image super-resolution, where x (resp. y) is the low (resp. high) resolution image. The study in [4] followed a similar approach with the difference that the mapping function was applied to the sparse codes, i.e., $z_y = \mathcal{F}(z_x)$, rather than the signals. Jia *et al.* [51] proposed dictionary learning via the concept of group sparsity so as to couple the different views in human pose estimation. Our coupled dictionary learning method is designed to address the challenges of the targeted source separation application and as such, the model we consider to represent the correlated sources is fundamentally different from previous work. Moreover, we extend coupled dictionary learning to the multi-scale case and provide a way to ignore certain noisy parts of the training signals (corresponding to cracks in our case).

III. Image Separation with Side Information

We denote by x_1^{ray} and x_2^{ray} two vectorized X-ray image patches that we wish to separate from each other given a mixture patch m, where $m = x_1^{\text{ray}} + x_2^{\text{ray}}$. Let y_1 and y_2 be the co-located (visual) image patches of the front and back side paintings. These patches play the role of *side information* that aids the separation. The use of side information has proven beneficial in various inverse problems [52]–[58]. In this work, we consider the signals $x_1^{\text{ray}}, x_2^{\text{ray}}, y_1, y_2 \in \mathbb{R}^n$ to obey (superpositions of) sparse representations in some dictionaries:

$$\begin{aligned} y_1 &= \Psi^c z_{1c} \\ y_2 &= \Psi^c z_{2c}, \end{aligned} \quad (3)$$

and

$$\begin{aligned} x_1^{\text{ray}} &= \Phi^c z_{1c} + \Phi v_1 \\ x_2^{\text{ray}} &= \Phi^c z_{2c} + \Phi v_2, \end{aligned} \quad (4)$$

where $z_{ic} \in \mathbb{R}^{\gamma\times 1}$, with $\|z_{ic}\|_0 = s_z \ll \gamma$ and $i = 1, 2$, denotes the sparse component that is common to the images in the visible and the X-ray domain with respect to dictionaries $\Psi^c, \Phi^c \in \mathbb{R}^{n\times\gamma}$, respectively. The parameter s_z denotes the sparsity of the vector z_{ic}. Moreover, $v_i \in \mathbb{R}^{d\times 1}$, with $\|v_i\|_0 = s_v \ll d$ denotes the sparse innovation component of the X-ray image, obtained with respect to the dictionary $\Phi \in \mathbb{R}^{n\times d}$. The common components express global features and structural characteristics that underlie both modalities. The innovation components capture parts of the signal specific to the X-ray modality, that is, traces of the wooden panel or even footprints of the vertical and horizontal wooden slats attached to the back of the paintings.

As the paintings are mounted on the same wooden panel, the sparse components that decompose the X-ray images via the dictionary Φ are expected to be the same, that is, we assume $v_1 = v_2 = v$. To motivate further this assumption, suppose that the dictionaries Ψ^c, Φ^c, and Φ have been learned (from training data), and we want to separate x_1^{ray} and x_2^{ray} from their sum m. This can be achieved by solving

$$\begin{array}{ll} \underset{z_{1c},z_{2c},v_1,v_2}{\text{minimize}} & \|z_{1c}\|_1 + \|z_{2c}\|_1 + \|v_1\|_1 + \|v_2\|_1, \\ \text{s.t.} & m = \Phi^c z_{1c} + \Phi^c z_{2c} + \Phi v_1 + \Phi v_2, \\ & y_1 = \Psi^c z_{1c}, \\ & y_2 = \Psi^c z_{2c}, \end{array} \tag{5}$$

where we relaxed the (nonconvex) ℓ_0-pseudo-norm to the (convex) ℓ_1-norm, denoted by $\|\cdot\|_1$. The components v_1 and v_2 are non-identifiable, because they are expressed in the same dictionary Φ. Specifically, if $(z_{1c}^\star, z_{2c}^\star, v_1^\star, v_2^\star)$ solves (5), so does $(z_{1c}^\star, z_{2c}^\star, \xi_1^\star, \xi_2^\star)$ with $\xi_1^\star + \xi_2^\star = v_1^\star + v_2^\star$. This means v_1 and v_2 are determined up to their sum. To resolve this ambiguity, we simply enforce $v_1 = v_2 = v$, yielding

$$\begin{array}{ll} \underset{z_{1c},z_{2c},v}{\text{minimize}} & \|z_{1c}\|_1 + \|z_{2c}\|_1 + 2\|v\|_1, \\ \text{s.t.} & m = \Phi^c z_{1c} + \Phi^c z_{2c} + 2\Phi v, \\ & y_1 = \Psi^c z_{1c}, \\ & y_2 = \Psi^c z_{2c}. \end{array} \tag{6}$$

Problem (6) boils down to basis pursuit, for which many solvers are available, e.g., [59]. For the purpose of simplicity, we assume that the ℓ_1-norm terms in (5) have equal weights, an approach which—as shown in Section VII—leads to high performance. Adding weights would provide more flexibility and may lead to better performance; however, the selection of weights requires careful hand-tuning.

We acknowledge the relation of our model with the sparse common component and innovations model that captures intra- and inter-signal correlation of physical signals in wireless sensor networks [55], [60]. Our approach is, however, more generic, since we decompose the signals in learned dictionaries rather than fixed canonical bases, as in [60].

IV. Coupled Dictionary Learning Algorithm

In the training step of our method, we learn coupled dictionaries, Ψ^c, Φ^c, Φ, by using image patches sampled from visual and X-ray images of single-sided panels, which do not suffer from superposition phenomena. The images were registered using the algorithm described in [61]. Let $Y = [y_1, \ldots, y_t], X = [x_1, \ldots, x_t] \in \mathbb{R}^{n \times t}$ represent a set of t co-located vectorized visual and X-ray patches, each containing $\sqrt{n} \times \sqrt{n}$ pixels. The DC value is extracted from each patch x_i, y_i, $i = 1, \ldots, t$, and it is treated as explained in Section VI. Hence, as per our model in (3) and (4), the columns of X and Y, which contain the texture of each X-ray and visual patch, are decomposed as

$$Y = \Psi^c Z, \tag{7a}$$

$$X = \Phi^c Z + \Phi V\,, \tag{7b}$$

where we collect their common components into the columns of the matrix $Z = [z_1, \ldots, z_t] \in \mathbb{R}^{\gamma \times t}$ and their innovation components into the columns of $V = [v_1, \ldots, v_t] \in \mathbb{R}^{d \times t}$. We formulate the coupled dictionary learning problem as

$$\begin{array}{ll} \underset{\substack{\Psi^c, Z \\ \Phi^c, \Phi, V}}{\text{minimize}} & \frac{1}{2}\|Y - \Psi^c Z\|_F^2 + \frac{1}{2}\|X - \Phi^c Z - \Phi V\|_F^2, \\ \text{s.t.} & \|z_\tau\|_0 \le s_z, \\ & \|v_\tau\|_0 \le s_v, \quad \forall \tau = 1, 2, \ldots, t. \end{array} \tag{8}$$

Problem (8) is bi-convex and—similar to related work [25], [34], [37]—we solve it by alternating between a sparse-coding step and a dictionary update step. Particularly, given initial estimates for dictionaries Ψ^c, Φ, and Φ^c—in line with prior work [34] we use the overcomplete discrete cosine transform (DCT) for initialization—we iterate on k between a sparse-coding step:

$$\begin{array}{rll} (Z^{k+1}, V^{k+1}) = & \underset{Z,V}{\arg\min} & \frac{1}{2}\left\| \begin{bmatrix} Y \\ X \end{bmatrix} - \begin{bmatrix} \Psi^{ck} & 0 \\ \Phi^{ck} & \Phi^k \end{bmatrix} \begin{bmatrix} Z \\ V \end{bmatrix} \right\|_F^2, \\ & \text{s.t.} & \|z_\tau\|_0 \le s_z, \\ & & \|v_\tau\|_0 \le s_v, \quad \forall \tau = 1, 2, \ldots, t, \end{array} \tag{9}$$

which is performed to learn the sparse codes Z, V having the dictionaries fixed, and a dictionary update step

$$(\Psi^{ck+1}, \Phi^{ck+1}, \Phi^{k+1}) = \arg \min_{\Psi^c, \Phi^c, \Phi} \frac{1}{2}\left\| \begin{bmatrix} Y \\ X \end{bmatrix} - \begin{bmatrix} \Psi^c & 0 \\ \Phi^c & \Phi \end{bmatrix} \begin{bmatrix} Z^{k+1} \\ V^{k+1} \end{bmatrix} \right\|_F^2. \tag{10}$$

which updates the dictionaries given the calculated sparse codes. The algorithm iterates between these steps until no additional iteration reduces the value of the cost function below a chosen threshold, or until a predetermined number of iterations is reached. In what follows, we provide details regarding the solution of the problem at each stage.

Sparse-coding step. Problem (9) decomposes into t problems, each of which can be solved in parallel:

$$\begin{array}{rll} (z_\tau^{k+1}, v_\tau^{k+1}) = & \underset{z_\tau, v_\tau}{\arg\min} & \frac{1}{2}\left\| \begin{bmatrix} y_\tau \\ x_\tau \end{bmatrix} - \begin{bmatrix} \Psi^{ck} & 0 \\ \Phi^{ck} & \Phi^k \end{bmatrix} \begin{bmatrix} z_\tau \\ v_\tau \end{bmatrix} \right\|_F^2, \\ & \text{s.t.} & \|z_\tau\|_0 \le s_z, \\ & & \|v_\tau\|_0 \le s_v. \end{array} \tag{11}$$

To address each of the t problems in (11), we propose a greedy algorithm that constitutes a modification of the orthogonal matching pursuit (OMP) method [see Algorithm 1]. Our method adapts OMP [62] to solve:

$$\begin{array}{ll} \underset{w}{\text{minimize}} & \|b - \Theta w\|_2^2 \\ \text{s.t.} & \|w(\mathcal{I})\|_0 \le s_z, \\ & \|w(\mathcal{J})\|_0 \le s_v, \end{array} \tag{12}$$

where $w(\mathcal{I})$ [resp., $w(\mathcal{J})$] denotes the components of vector $w \in \mathbb{R}^{(\gamma+d)\times 1}$ indexed by the index set $\mathcal{I}$ (resp., $\mathcal{J}$), with $\mathcal{I} \cup \mathcal{J} = \{1, 2, \ldots, \gamma + d\}$, $\mathcal{I} \cap \mathcal{J} = \{\emptyset\}$. Each sub-problem in (11) translates to (12) by replacing: $b = \begin{bmatrix} y_\tau \\ x_\tau \end{bmatrix}$, $\Theta =$

Algorithm 1 Modified Orthogonal Matching Pursuit (mOMP)

Input: The vector $b \in \mathbb{R}^{(2n)\times 1}$, the matrix $\Theta = \begin{bmatrix}\theta_1 & \dots & \theta_{\gamma+d}\end{bmatrix}$, where $\theta_i \in \mathbb{R}^{(2n)\times 1}$, the set of indices $\mathcal{I}$, $\mathcal{J}$, and the sparsity levels s_z, $s_v \in \mathbb{N}$ of vectors z, v.

Initialization: Initialize the residual $r_0 = b$; set $s_w = s_z + s_v$; set counters $\ell_z = 0$, $\ell_v = 0$, $\Omega_0 = \emptyset$, and $T_0 = [\ ]$ (empty matrix).

Algorithm

1: **for** $i = 1, 2, \dots, s_w$ **do**
2: Compute the vector of ordered indices q_i such that $|\langle r_{i-1}, \theta_{q_i(1)}\rangle| \geq |\langle r_{i-1}, \theta_{q_i(2)}\rangle| \geq \dots \geq |\langle r_{i-1}, \theta_{q_i(\gamma+d)}\rangle|$.
3: Set $\mathcal{G} = \emptyset$ and $\mathtt{iter} = 0$.
4: **while** $\mathcal{G} = \emptyset$ **do**
5: $\mathtt{iter} \leftarrow \mathtt{iter} + 1$.
6: $\kappa = q_i(\mathtt{iter})$.
7: **if** $\kappa \in \mathcal{I}$ and $\ell_z < s_z$ **then**
8: Set $\mathcal{G} = \kappa$ and $\ell_z \leftarrow \ell_z + 1$.
9: **else**
10: **if** $\kappa \in \mathcal{J}$ and $\ell_v < s_v$ **then**
11: Set $\mathcal{G} = \kappa$ and $\ell_v \leftarrow \ell_v + 1$.
12: **end if**
13: **end if**
14: **end while**
15: Update: $\Omega_i = \Omega_{i-1} \cup \{\kappa\}$ and $T_i = \begin{bmatrix}T_{i-1} & \theta_\kappa\end{bmatrix}$.
16: Solve: $w_i = \arg\min_w \|b - T_i\, w\|_2$.
17: Update residual: $r_i = b - T_i\, w_i$.
18: **end for**

Output: $w \in \mathbb{R}^n$, with $w(\Omega_i) = w_i$ and $w(\Omega_i^c) = 0$.

$\begin{bmatrix}\Psi^{ck} & 0\\ \Phi^{ck} & \Phi^k\end{bmatrix}$, $w = \begin{bmatrix}z_\tau\\ v_\tau\end{bmatrix}$, $\mathcal{I} = \{1, \dots, \gamma\}$ and $\mathcal{J} = \{1, \dots, d\}$. The main difference with respect to OMP [62] is that our method partitions the indices of the columns of Θ in the sets $\mathcal{I}$, $\mathcal{J}$. As such, the vector q_i in Step 1 contains the indices of the columns of Θ, in decreasing order of correlation with the residual vector. The while loop in Step 4 differs from the corresponding step in OMP in that we select the column to be added taking into account the sparsity levels of the vectors z and v.

It is worth mentioning that we follow an OMP-based approach in the dictionary learning stage so as to benefit from a fast algorithm, while we adhere to the ℓ_1-norm minimization strategy for the source separation problem [see Problem (6)] with the goal to achieve high X-ray reconstruction quality[1].

Dictionary update step. Problem (10) can be written as

$$\underset{\Psi^c, \overline{\Phi}}{\text{minimize}} \quad \tfrac{1}{2}\left\|Y - \Psi^c Z^{k+1}\right\|_F^2 + \tfrac{1}{2}\left\|X - \overline{\Phi}\,\overline{V}^{k+1}\right\|_F^2, \tag{13}$$

where $\overline{\Phi} = \begin{bmatrix}\Phi^c & \Phi\end{bmatrix}$ and $\overline{V}^{k+1} = \begin{bmatrix}Z^{k+1}\\ V^{k+1}\end{bmatrix}$. Problem (13) decouples into two (independent) problems, that is,

$$\underset{\Psi^c}{\text{minimize}} \quad \tfrac{1}{2}\left\|Y - \Psi^c Z^{k+1}\right\|_F^2 \tag{14}$$

and

$$\underset{\overline{\Phi}}{\text{minimize}} \quad \tfrac{1}{2}\left\|X - \overline{\Phi}\,\overline{V}^{k+1}\right\|_F^2. \tag{15}$$

[1]It is known [63]–[65] that a basis pursuit strategy offers higher performance than OMP, while the latter admits simple, fast implementations [66].

Provided that Z^{k+1} and $\overline{V}^{k+1}$ are full row-rank, each of these problems has a closed-form solution, namely,

$$\Psi^{c\,k+1} = Y Z^{k+1^T}\left(Z^{k+1} Z^{k+1^T}\right)^{-1}$$

and

$$\overline{\Phi}^{k+1} = X \overline{V}^{k+1^T}\left(\overline{V}^{k+1}\overline{V}^{k+1^T}\right)^{-1}.$$

When Z^{k+1} and $\overline{V}^{k+1}$ are rank-deficient, (14) and (15) have multiple solutions, from which we select the one with minimal Frobenius norm. This is done by taking a thin (also called reduced) singular value decomposition [67] of $Z^{k+1} = G_{z^{k+1}}\Sigma_{z^{k+1}}U_{z^{k+1}}^T$ and $\overline{V}^{k+1} = G_{\overline{v}^{k+1}}\Sigma_{\overline{v}^{k+1}}U_{\overline{v}^{k+1}}^T$, and calculating

$$\Psi^{c\,k+1} = Y U_{z^{k+1}}\Sigma_{z^{k+1}}^{-1} G_{z^{k+1}}^T$$

and

$$\overline{\Phi}^{k+1} = X U_{\overline{v}^{k+1}}\Sigma_{\overline{v}^{k+1}}^{-1} G_{\overline{v}^{k+1}}^T.$$

V. Weighted Coupled Dictionary Learning

Visual and X-ray images of paintings contain a high number of pixels that depict cracks. These are fine patterns of dense cracking formed within the materials. When taking into account these pixels, the learned dictionaries comprise atoms that correspond to high frequency components. As a consequence, the reconstructed images are contaminated by high frequency noise. In order to improve the separation performance, our objective is to obtain dictionaries that ignore pixels representing cracks. To identify such pixels, we generate binary masks identifying the location of cracks by applying our method in [10]. Each sampled image patch may contain a variable number of crack pixels, meaning that each column of the data matrix contains a different number of meaningful entries. To address this issue, we introduce a weighting scheme that adds a weight of 0 or 1 to the pixels that do or do not correspond to cracks, respectively. These crack-induced weights are included using a Hadamard product, namely, our model in (7) is modified to

$$Y \odot \Lambda = (\Psi^c Z) \odot \Lambda \tag{16a}$$
$$X \odot \Lambda = (\Phi^c Z + \Phi V) \odot \Lambda. \tag{16b}$$

where the matrix Λ has the same dimensions as X and Y and its entries are 0 or 1 depending on whether a pixel is part of a crack or not, respectively. We now formulate the weighted coupled dictionary learning problem as

$$\begin{aligned}\underset{\substack{\Psi^c, Z\\ \Phi^c, \Phi, V}}{\text{minimize}} \quad & \tfrac{1}{2}\|(Y - \Psi^c Z)\odot\Lambda\|_F^2\\ & + \tfrac{1}{2}\|(X - \Phi^c Z - \Phi V)\odot\Lambda\|_F^2,\\ \text{s.t.} \quad & \|z_\tau\|_0 \leq s_z,\\ & \|v_\tau\|_0 \leq s_v, \quad \forall \tau = 1, 2, \dots t.\end{aligned} \tag{17}$$

Similar to (8), the solution for (17) is obtained by alternating between a sparse-coding and a dictionary update step.

Sparse-coding step. Similar to (11), the sparse-coding problem decomposes into t problems that can be solved in parallel:

$$(z_\tau^{k+1}, v_\tau^{k+1}) = \underset{z_\tau, v_\tau}{\arg\min}\ \frac{1}{2}\Bigg\| \begin{bmatrix} y_\tau \\ x_\tau \end{bmatrix} \odot \begin{bmatrix} \lambda_\tau \\ \lambda_\tau \end{bmatrix} - \left(\begin{bmatrix} \Psi^{ck} & 0 \\ \Phi^{ck} & \Phi^k \end{bmatrix} \odot \left(\begin{bmatrix} \lambda_\tau \\ \lambda_\tau \end{bmatrix} \mathbf{1}^T \right) \right) \begin{bmatrix} z_\tau \\ v_\tau \end{bmatrix} \Bigg\|_2^2$$

$$\text{s.t.}\ \|z_\tau\|_0 \le s_z, \quad \|v_\tau\|_0 \le s_v, \quad \forall \tau = 1, 2, \dots t, \tag{18}$$

where we used λ_τ to represent column τ of Λ and $\mathbf{1}^T$ to denote a row-vector of ones with dimension equal to $\gamma + d$. To address each of the t sub-problems in (18), we use the mOMP algorithm described in Algorithm 1, as each sub-problem in (18) reduces to (12) by replacing: $b = \begin{bmatrix} y_\tau \\ x_\tau \end{bmatrix} \odot \begin{bmatrix} \lambda_\tau \\ \lambda_\tau \end{bmatrix}$, $\Theta = \begin{bmatrix} \Psi^{ck} & 0 \\ \Phi^{ck} & \Phi^k \end{bmatrix} \odot \left(\begin{bmatrix} \lambda_\tau \\ \lambda_\tau \end{bmatrix} \mathbf{1}^T \right)$, and $w = \begin{bmatrix} z_\tau \\ v_\tau \end{bmatrix}$.

Dictionary update step. The dictionary update problem is now written as

$$\underset{\Psi^c, \overline{\Phi}}{\text{minimize}}\ \frac{1}{2}\left\| Y \odot \Lambda - (\Psi^c Z^{k+1}) \odot \Lambda \right\|_F^2 + \frac{1}{2}\left\| X \odot \Lambda - (\overline{\Phi}\, \overline{V}^{k+1}) \odot \Lambda \right\|_F^2, \tag{19}$$

and it decouples into:

$$\underset{\Psi^c}{\text{minimize}} \frac{1}{2}\left\| Y \odot \Lambda - (\Psi^c Z^{k+1}) \odot \Lambda \right\|_F^2 \tag{20a}$$

$$\underset{\overline{\Phi}}{\text{minimize}} \frac{1}{2}\left\| X \odot \Lambda - (\overline{\Phi}\, \overline{V}^{k+1}) \odot \Lambda \right\|_F^2. \tag{20b}$$

We present only the solution of the first problem since the solution of the other follows the same logic. Specifically, we express the Frobenius norm in (20a) as the sum of t ℓ_2-norm terms, each corresponding to a vectorized training patch

$$\sum_{\tau=1}^{t} \| y_\tau \odot \lambda_\tau - (\Psi^c z_\tau) \odot \lambda_\tau \|_2^2. \tag{21}$$

By replacing the Hadamard product with multiplication by a diagonal matrix $\Delta_\tau = \text{diag}(\lambda_\tau)$, (21) can be written as

$$\sum_{\tau=1}^{t} \| \Delta_\tau y_\tau - \Delta_\tau \Psi^c z_\tau \|_2^2. \tag{22}$$

To minimize the expression in (22), we take the derivative with respect to the dictionary Ψ^c and set it to zero:

$$\frac{\partial}{\partial \Psi^c} \sum_{\tau=1}^{t} \| \Delta_\tau y_\tau - \Delta_\tau \Psi^c z_\tau \|_2^2 = 0$$

$$\Longrightarrow \sum_{\tau=1}^{t} \frac{\partial}{\partial \Psi^c} \left[(\Delta_\tau y_\tau - \Delta_\tau \Psi^c z_\tau)^T (\Delta_\tau y_\tau - \Delta_\tau \Psi^c z_\tau) \right] = 0$$

$$\Longrightarrow 2 \sum_{\tau=1}^{t} \frac{\partial}{\partial \Psi^c} y_\tau^T \Delta_\tau^T \Delta_\tau \Psi^c z_\tau = \sum_{\tau=1}^{t} \frac{\partial}{\partial \Psi^c} z_\tau^T \Psi^{cT} \Delta_\tau^T \Delta_\tau \Psi^c z_\tau$$

$$\Longrightarrow \sum_{\tau=1}^{t} \left(\Delta_\tau^T \Delta_\tau y_\tau \mathbf{1}^T \right) \odot (\mathbf{1} z_\tau^T) = \sum_{\tau=1}^{t} \left(\Psi^c z_\tau z_\tau^T \right) \odot ((\lambda_\tau \odot \lambda_\tau) \mathbf{1}^T). \tag{23}$$

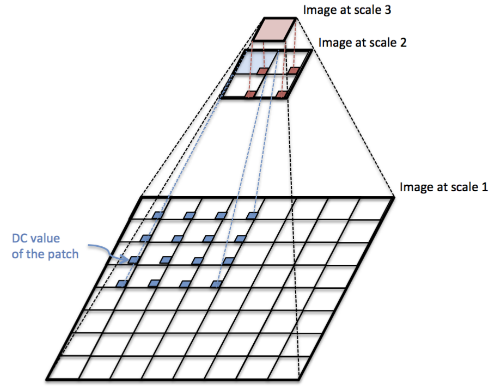


Figure 2. Schema of a 3-scale pyramid decomposition in the proposed multi-scale dictionary learning and source separation approaches.

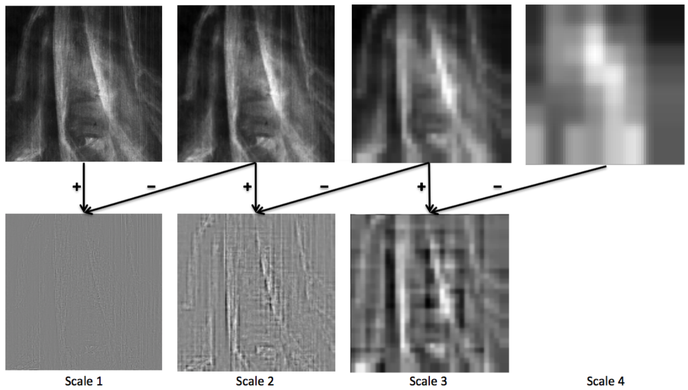


Figure 3. Example of a 4-scale pyramid decomposition of a mixed X-ray image. The original image resolution is 1024×1024 pixels. At scale 1, the image is split into non-overlapping patches of 8×8 pixels and the DC value of every patch is extracted, thereby generating the high-pass component. The aggregated DC values compose the low-pass component at scale 2, the resolution of which is 128×128 pixels. Dividing this component into non-overlapping patches of 4×4 pixels and extracting the DC value from every patch yields the high-pass band in scale 2. The procedure is repeated until finally the low-pass band at scale 4 has a resolution of 8×8 pixels.

Before proceeding with the method to solve (23), we recall that the entries of λ_τ are either 0 or 1. To avoid dividing by zero when solving (23), we have to update the rows of the dictionary matrix one-by-one. Specifically, for each row i of Ψ^c, we consider the matrix $A_i = \sum_{\tau \in S_i} z_\tau z_\tau^T$, where S_i is the support[2] of the i-th row of Λ, and z_τ is the τ-th column of Z. We also create a vector $c_i = \sum_{\tau \in S_i} Y(i, \tau) z_\tau$, where $Y(i, \tau)$ is the (i, τ)-th entry of Y. Provided that A_i is invertible, the i-th row of Ψ^c (which we denote by the row-vector ψ_i^c) will be given by

$$\psi_i^c = c_i A_i^{-1}. \tag{24}$$

If each z_τ is drawn randomly, A_i is invertible with probability 1 whenever the cardinality of S_i is at least equal to the number of columns of Ψ_c. Although in practice each z_τ is not randomly drawn, we still obtain an invertible A_i by guaranteeing that the number of training samples is large enough.

[2]The support S_i is defined by the indices where the i-th row of Λ is equal to 1.

VI. Single- and Multi-Scale Separation Methods

A. Single-Scale Approach

Given the trained coupled dictionaries, the source separation method described in Section III is applied locally, per overlapping patch of the X-ray image. Let the corresponding patches from the mixed X-ray and the two corresponding visual images be denoted as $m^{\boldsymbol{u}}$, $y_1^{\boldsymbol{u}}$, and $y_2^{\boldsymbol{u}}$, respectively. Each patch contains $\sqrt{n} \times \sqrt{n}$ pixels and has top-left coordinates

$$\boldsymbol{u} = (\epsilon \cdot u_1, \epsilon \cdot u_2), \quad 0 \leq u_1 < \left\lfloor \frac{H}{\epsilon} \right\rfloor, \quad 1 \leq u_2 < \left\lfloor \frac{W}{\epsilon} \right\rfloor,$$

where $\epsilon \in \mathbb{Z}_+$, $1 \leq \epsilon < \sqrt{n}$ is the overlap step-size, $\lfloor \bullet \rfloor$ is the floor function, and H, W are the image height and width, respectively. Prior to separation, the DC values $\mathrm{DC}_{m^{\boldsymbol{u}}}$, $\mathrm{DC}_{y_1^{\boldsymbol{u}}}$, and $\mathrm{DC}_{y_2^{\boldsymbol{u}}}$ are removed from the pixels in the patches $m^{\boldsymbol{u}}$, $y_1^{\boldsymbol{u}}$, and $y_2^{\boldsymbol{u}}$, respectively, and the residual values are vectorized. The solution of Problem (6) yields the sparse components $z_{1c}^{\boldsymbol{u}}, z_{2c}^{\boldsymbol{u}}$, and $v^{\boldsymbol{u}}$ corresponding to the patch with coordinates $\boldsymbol{u}$. The texture of each separated patch is then reconstructed following the model in (4), that is, $x_1^{\boldsymbol{u}} = \Phi^c z_{1c}^{\boldsymbol{u}} + \Phi v^{\boldsymbol{u}}$ and $x_2^{\boldsymbol{u}} = \Phi^c z_{2c}^{\boldsymbol{u}} + \Phi v^{\boldsymbol{u}}$. In certain cases, the v component may capture parts of the actual content; for example, vertical brush strokes can be misinterpreted as the wood texture of the panel. In this case, we can choose to skip the v component; namely, we can reconstruct the texture of the X-ray patches as $x_1^{\boldsymbol{u}} = \Phi^c z_{1c}^{\boldsymbol{u}}$ and $x_2^{\boldsymbol{u}} = \Phi^c z_{2c}^{\boldsymbol{u}}$. The DC values of the separated X-ray patches, are weighted according to the DC values of the co-located patches in the visual images, namely, $\mathrm{DC}_{x_1^{\boldsymbol{u}}} = \frac{\mathrm{DC}_{y_1^{\boldsymbol{u}}}}{\mathrm{DC}_{y_1^{\boldsymbol{u}}}+\mathrm{DC}_{y_2^{\boldsymbol{u}}}} \times \mathrm{DC}_{m^{\boldsymbol{u}}}$ and $\mathrm{DC}_{x_2^{\boldsymbol{u}}} = \frac{\mathrm{DC}_{y_2^{\boldsymbol{u}}}}{\mathrm{DC}_{y_1^{\boldsymbol{u}}}+\mathrm{DC}_{y_2^{\boldsymbol{u}}}} \times \mathrm{DC}_{m^{\boldsymbol{u}}}$, and then added back to the corresponding separated X-ray patches. Finally, the pixels in each separated X-ray are recovered as the average of the co-located pixels in each overlapping patch.

B. Multi-Scale Approach

Due to the restricted patch size in comparison to the high resolution of the X-ray image, the DC values $\mathrm{DC}_{m^{\boldsymbol{u}}}$ of all patches carry a considerable amount of the total image energy. In the single-scale approach, these DC values are common to the two separated X-rays, thereby leading to poor separation. To address this issue, we devise a multi-scale image separation approach. In contrast with the techniques in [25], [36], [37], the proposed multi-scale approach performs a pyramid decomposition of the mixed X-ray and visual images, that is, the images are recursively decomposed into low-pass and high-pass bands. The decompositions at scale $l = \{1, 2, \ldots, L\}$ are constructed as follows. The images at scale l—where we use the notation $M_l, Y_{1,l}, Y_{2,l}$, to refer to the mixed X-ray and the two visuals, respectively—are divided into overlapping patches $m_l^{\boldsymbol{u}_l}, y_{1,l}^{\boldsymbol{u}_l}$, and $y_{2,l}^{\boldsymbol{u}_l}$, each of size $\sqrt{n_l} \times \sqrt{n_l}$ pixels. Each patch has top-left coordinates

$$\boldsymbol{u}_l = (\epsilon_l \cdot u_{1,l}, \epsilon_l \cdot u_{2,l}), 0 \leq u_{1,l} < \left\lfloor \frac{H_l}{\epsilon_l} \right\rfloor, 0 \leq u_{2,l} < \left\lfloor \frac{W_l}{\epsilon_l} \right\rfloor,$$

where $\epsilon_l \in \mathbb{Z}_+$, $1 \leq \epsilon_l < \sqrt{n_l}$ is the overlap step-size, and H_l, W_l are the height and width of the image decomposition at scale l. The DC value $\mathrm{DC}_{m_l^{\boldsymbol{u}_l}}$ is extracted from each patch, thereby constructing the high frequency band of the image at scale l. The aggregated DC values comprise the low-pass component of the image, the resolution of which is $\left\lfloor \frac{H_l}{\epsilon_l} \right\rfloor \times \left\lfloor \frac{W_l}{\epsilon_l} \right\rfloor$ pixels. This low-pass component is then decomposed further at the subsequent scale $(l+1)$. The pyramid decomposition is schematically sketched in Fig. 2 and exemplified in Fig. 3.

The high frequency component of the mixed X-ray image at each scale l is separated patch-by-patch by solving Problem (6). Namely, the reconstructed high frequency component of each patch at scale l is calculated as $x_{1,l}^{\boldsymbol{u}_l} = \Phi_l^c z_{1c,l}^{\boldsymbol{u}_l} + \Phi_l v_l^{\boldsymbol{u}_l}$ and $x_{2,l}^{\boldsymbol{u}_l} = \Phi_l^c z_{2c,l}^{\boldsymbol{u}_l} + \Phi_l v_l^{\boldsymbol{u}_l}$; or alternatively, as $x_{1,l}^{\boldsymbol{u}_l} = \Phi_l^c z_{1c,l}^{\boldsymbol{u}_l}$ and $x_{2,l}^{\boldsymbol{u}_l} = \Phi_l^c z_{2c,l}^{\boldsymbol{u}_l}$. In this way, the DC information of the mixed X-ray is progressively separated while ascending through the scales. Note that the dictionary learning process can be applied per scale, yielding a triple of coupled dictionaries $(\Psi_l^c, \Phi_l^c, \Phi_l)$ per scale l. In practice, due to lack of training data in the higher scales, dictionaries are learned only from the low-scale decompositions and then copied to the higher scales. The separated X-ray images are finally reconstructed by following the reverse operation: descending the pyramid, the separated component at the coarser scale is up-sampled and added to the separated component of the finer scales.

Table I
Dictionary Identifiability of the Proposed Algorithm Based on Synthetic Data, Expressed in Terms of the Percentage of Retrieved Atoms for the Dictionaries in Model (7).

SNR [dB]	∞	40	35	30	25	20	15
Ψ^c	96%	95.18%	95.38%	95.65%	95.20%	90.42%	12.53%
Φ^c	96.78%	95.97%	96.53%	96.50%	95.48%	74.35%	0.27%
Φ	92.95%	91.90%	91.73%	91.27%	91.50%	88.25%	3.07%

VII. Experiments

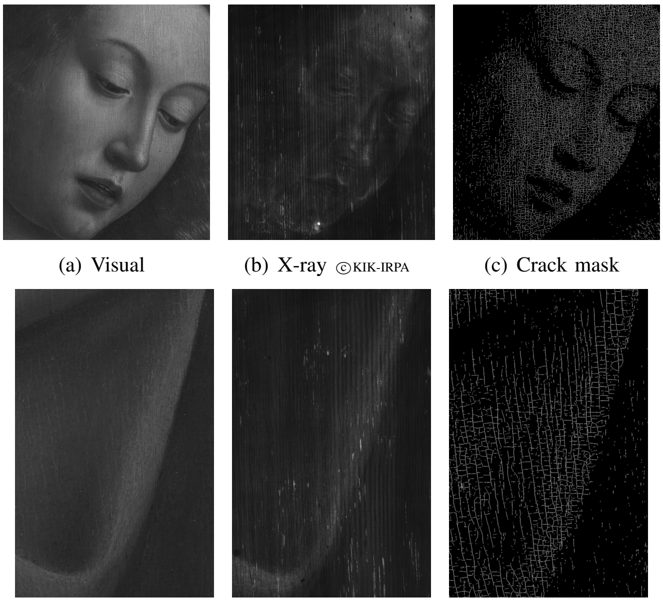
A. Experiments with Synthetic Data

As in [33], [34], we first evaluate the performance of our coupled dictionary learning algorithm—described in Section IV—and our source separation with side information method (see Section III) using synthetic data. Firstly, we generate synthetic signals, x, y, according to model (3), (4), using random dictionaries and then, given the data, we assess whether the algorithm recovers the original dictionaries. The random dictionaries Ψ^c, Φ, and Φ^c of size 40×60 contain entries drawn from the standard normal distribution and their columns are normalized to have unit ℓ_2-norm. Given the dictionaries, $t = 1500$ sparse vectors Z and V were produced, each with dimension $\gamma = d = 60$. The column-vectors z_τ and v_τ, $\tau = 1, 2, \ldots, t$, comprised of respectively $s_z = 2$ and $s_v = 3$ non-zero coefficients distributed uniformly and placed in random and independent locations. Combining the dictionaries and the sparse vectors according to the model in (7) yields the correlated data signals X and Y, to which white Gaussian noise with a varying signal-to-noise ratio (SNR) has been added.

To retrieve the initial dictionaries, we apply the proposed method in Section IV with the dictionaries initialised ran-

Table II
RECONSTRUCTION ERROR OF THE PROPOSED SOURCE SEPARATION WITH SIDE INFORMATION METHOD BASED ON SYNTHETIC DATA.

SNR [dB]	∞	40	35	30	25	20	15
x_1	1.66×10^{-5}	0.0094	0.0307	0.0734	0.1215	0.2125	0.5854
x_2	1.43×10^{-5}	0.0097	0.0222	0.0818	0.0857	0.2516	0.2781



(a) Visual (b) X-ray ©KIK-IRPA (c) Crack mask

(d) Visual (e) X-ray ©KIK-IRPA (f) Crack mask

Figure 4. Examples of images from single-sided panels of the *Ghent Altarpiece* and the corresponding crack masks.

domly and the maximum number of iterations set to 100—experimental evidence has shown that this value strikes a good balance between complexity and dictionary identifiability. To compare the retrieved dictionaries with the original ones, we adhere to the approach in [34]: per generating dictionary, we sweep through its columns and identify the closest column in the retrieved dictionary. The distance between the two columns is measured as $1 - |\delta_i^T \tilde{\delta}_j|$, where δ_i is the i-th column in the original dictionary Ψ^c, Φ^c, or Φ, and $\tilde{\delta}_j$ is the j-th column in the corresponding recovered dictionary. Similar to [34], a distance less than 0.01 signifies a success. The percentage of successes per dictionary and for various SNR values is reported in Table I. The results, which are averaged over 100 trials, show that for very noisy data (that is, SNR $\leq$ 15) the dictionary identifiability performance is low. However, for SNR values higher than 20 dB, the percentage of recovered dictionary atoms is up to 96.78%. The obtained performance is systematic for different dictionary and signal sizes as well as for different sparsity levels.

In a second stage, given the learned dictionaries, we separate signal pairs (x_1, x_2) from mixtures $m = x_1 + x_2$ by solving Problem (6) using the corresponding pair (y_1, y_2) as side information. The pairs are taken from the correlated data signals X and Y, to which white Gaussian noise with a varying SNR has been added. Table II reports the normalized mean-squared error between the reconstructed—defined by $\tilde{x_i}$—and the original signals, that is, $\frac{\|x_i - \tilde{x_i}\|_2^2}{\|x_i\|_2^2}$, $i = \{1, 2\}$. The results show that at low and moderate noise SNRs the reconstruction error is very low. When the noise increases, the recovery performance drops; this is to be expected as the noise affects both the dictionary learning and the generation of the mixtures.

B. Experiments with Real Data

We consider eight image pairs—each consisting of an X-ray scan and the corresponding photograph—taken from digital acquisitions [12] of single-sided panels of the *Ghent Altarpiece* (1432). Furthermore, we are given access to eight crack masks (one per visual/X-ray image pair) that indicate the pixel positions referring to cracks (these masks were obtained using our method in [10]). Fig. 4 depicts two such pairs with the crack masks, one visualizing a face and the other a piece of fabric. An example X-ray mixture (of size 1024×1024 pixels) together with its two registered visual images corresponding to the two sides of the painting are depicted in Fig. 5.

Firstly, adhering to the single-scale approach, described in Section VI-A, we train a dictionary triplet, (Ψ^c, Φ^c, Φ), using our method in Section IV. We use $t = 46400$ patches, each containing 8×8 pixels, the dictionaries, Ψ^c, Φ^c, Φ, have a dimension of 64×256, and we set $s_z = 10$ and $s_v = 8$. The separated X-rays that correspond to the mixture in Fig. 5 are depicted in the first column of Fig. 6. We observe that our single-scale approach separates the texture of the X-rays; this is demonstrated by the accurate separation of the cracks. Still, however, the low-pass band content is not properly split over the images; namely, part of the cloth and the face are present in both separated images. Next, we apply the multi-scale framework, where we use $L = 4$ scales with parameters $\sqrt{n_l} = 8$, $l = \{1, 2, 3, 4\}$, $\epsilon_1 = 4$, $\epsilon_2 = 4$, $\epsilon_3 = 7$, and $\epsilon_4 = 8$. Dictionary triplets $(\Psi_\ell^c, \Phi_\ell^c, \Phi_\ell)$, each with dimension of 64×256, are trained for the first three scales and the dictionaries of the third scale are used for the forth. We use $t_1 = 46400$, $t_2 = 46400$ and $t_3 = 35500$ patches for scale 1, 2 and 3, respectively. The visualizations in the second column of Fig. 6 show that, compared to the single scale approach, the multi-scale method properly discriminates the low-pass frequency content of the two images (most part of the cloth is allocated to "Separated Side 1" while the face is only visible in "Separated Side 2"), thereby leading to a higher separation performance. Finally, we also construct dictionary triplets according to our weighted dictionary learning method in Section V. The remaining dictionary learning parameters are as before. It is worth mentioning that, in order to obtain a solution in (24), the number of training samples t needs to be higher that the total dimension of the dictionary. Namely, to update the columns of dictionary Ψ^c we need at least 16384 samples. Correspondingly, to update the rows of dictionary $\overline{\Phi}$ we need more than 32768 samples. The visual results in the third column of Fig. 6 corroborate that the quality of the separation is improved when the dictionaries are learned from only non-crack pixels. Indeed, with this configuration, the separated images are not only smoother but also the separation is more prominent.

It is worth mentioning that the results of our method, depicted in Fig. 6, are obtained without including the v component during the reconstruction; namely, we reconstructed each

Figure 5. Image set cropped from a double-sided panel of the altarpiece, on which we assess the proposed method; (a) and (d) photograph of side 1, (b) and (e) photograph of side 2; (c) and (f) corresponding X-ray image. The resolution is 1024×1024 pixels.

X-ray patch as $x_1 = \Phi^c z_{1c}$ and $x_2 = \Phi^c z_{2c}$. The visual results of our method when including the v component during the reconstruction are depicted in Fig. 7. These results are obtained with the same dictionaries that yield the result in the third column of Fig. 6. By comparing the two reconstructions, we can make the following observations. First, the v component successfully expresses the X-ray specific features, such as the wood grain, visualized by the periodic vertical stripes in the X-ray scan. The reconstruction of these stripes is much more evident in Fig. 7. Secondly, in this case, the v component also captures parts of the actual content that we wish to separate. For example, we can discern a faint outline of the eye in Fig. 7(a) as well as a fold of fabric appearing in Fig. 7(b).

We compare our best performing multi-scale approach (namely, the one that omits cracks when learning dictionaries) with the state-of-the-art MCA method [20], [28]. Two configurations of the latter are considered. Based on prior work [30], in one configuration we use fixed dictionaries, namely, the discrete wavelet and curvelet transforms are applied on blocks of 512×512 pixels. Inherently, the low-frequency content cannot be split by MCA and it is equally divided between both retrieved components. In the other configuration, we learn dictionaries with K-SVD using the same training X-ray images as in the previous experiment. One dictionary is trained on the X-ray images depicting fabric and the other on the images of faces. The K-SVD parameters are the same as the ones used in our method. Furthermore, the same multi-scale strategy is applied to the configuration of MCA with K-SVD learned dictionaries. The results are depicted in Fig. 8 and Fig. 9. Note that the third column in Fig. 8 and Fig. 9 are without and with taking the v component into account, respectively. It is clear that MCA with fixed dictionaries can only separate based on morphological properties; for example, the wood grain of the panel is captured entirely by curvelets and not by the wavelets. It is, however, unsuitable to separate painted content—it is evident that part of the cloth and face appear in both separated components. Furthermore, MCA with K-SVD dictionaries is also unable to separate the X-ray content. Nevertheless, we do observe that most cracks are captured by the face dictionary, as more cracks are present in that type of content. Unlike both state-of-the-art configurations of MCA, the proposed method separates the X-ray content accurately (the cloth is always depicted on "Separated Side 1" while the face is only visible in "Separated Side 2"), leading to better visual performance. These results corroborate the benefit of using side information by means of photographs to separate

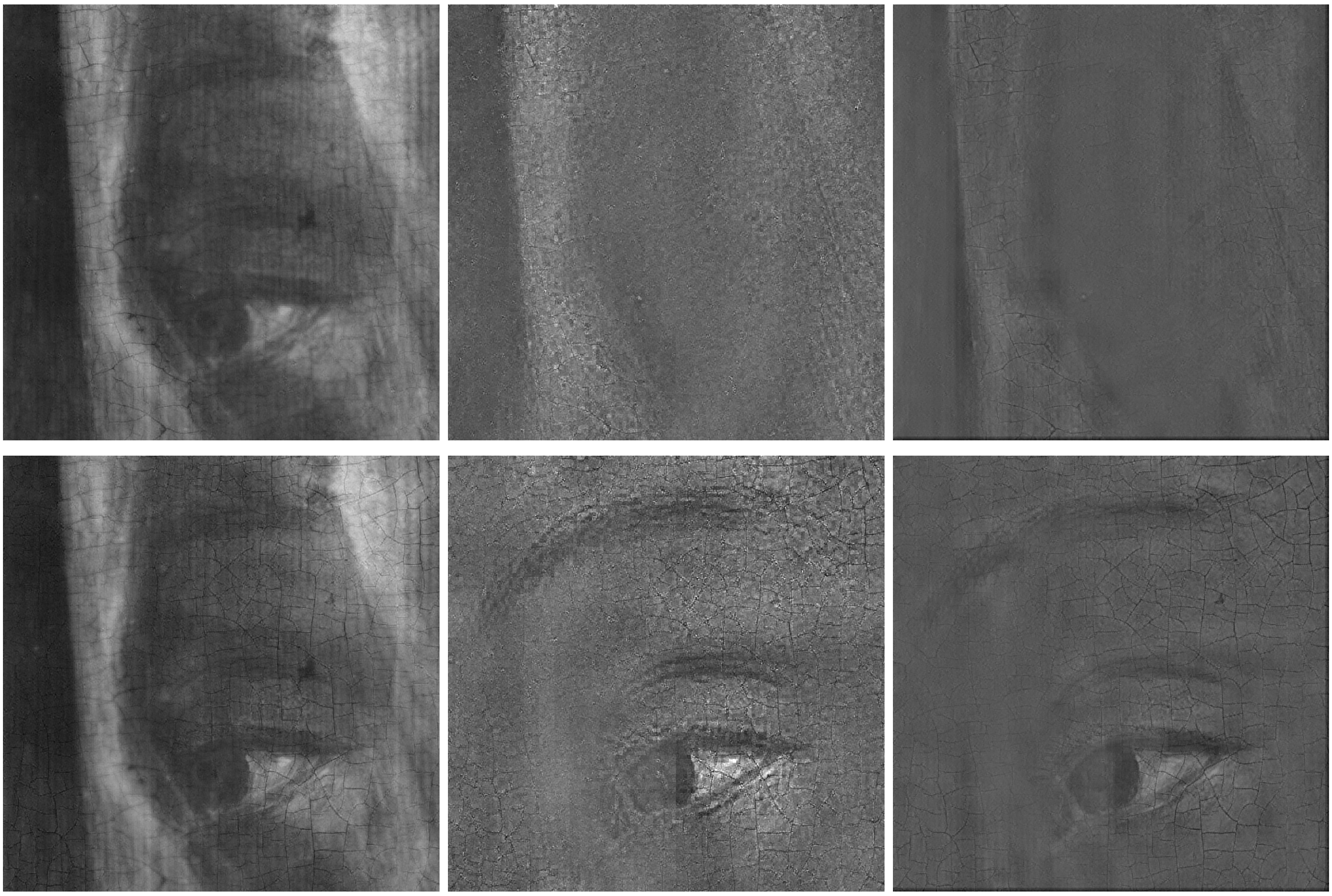


Figure 6. Visual evaluation of the different configurations of the proposed method in the separation of the X-ray image in Fig. 5(c); (first row) separated side 1, (second row) separated side 2. The configurations are: (first column) single-scale method (Section VI-A) with the coupled dictionary learning algorithm described in Section IV, (second column) multi-scale method (Section VI-B) with the coupled dictionary learning method from Section IV, (third column) multi-scale method (Section VI-B) with the weighted coupled dictionary learning method from Section V. The v component is not included in any of the configurations of the proposed method.

mixtures of X-ray images.

C. Experiments on Simulated Mixtures

Due to the lack of proper ground truth data, we generate simulated X-ray image mixtures in an attempt to assess our method in an objective manner. To this end, we utilised the X-ray images from single-sided panels, depicting content similar to the mixture in Fig. 5(c) and (f). We generated mixtures by summing these independent X-ray images[3] and then we assessed the separation performance of the proposed method vis-à-vis MCA either with fixed or K-SVD trained dictionaries. For this set of experiments, patches of size 256×256 pixels were considered and the parameters of the different methods were kept the same as in the previous section. Table III reports the quality of the reconstructed X-ray components by means of the peak-signal-to-noise-ratio (PSNR) and structural similarity index metric (SSIM) [68]. It is clear that the proposed method outperforms the alternative state-of-the-art methods both in terms of PSNR and SSIM performance. Compared to MCA with fixed dictionaries, the proposed method brings an improvement in the quality of the separation by up to 1.26dB in PSNR and 0.0741 in SSIM for “Mixture 3”. The maximum gains against MCA with K-SVD trained dictionaries are 1.41dB and 0.0953 for “Mixture 3” again. While we realize that PSNR and SSIM are not necessarily the right image quality metrics in this scenario, they do demonstrate objectively the improvements that our method brings over the state of the art.

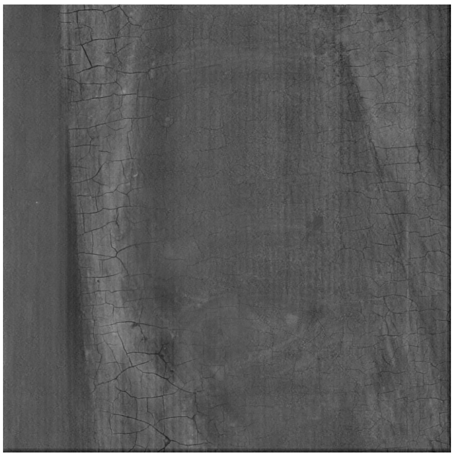
[3]We divided the sum by two to bring the mixture to the same range as the independent components.

VIII. Conclusion

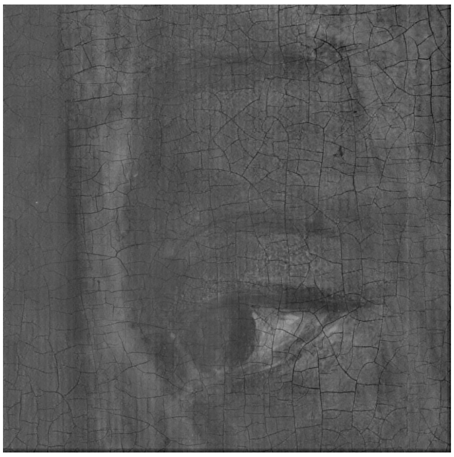
We have proposed a novel sparsity-based regularization method for source separation guided by side information. Our method learns dictionaries, coupling registered acquisitions from diverse modalities, and comes both in a single- and multi-scale framework. The proposed method is applied in the separation of X-ray images of paintings on wooden panels that are painted on both sides, using the photographs of each side as side information. Experiments on real data, consisting of digital acquisitions of the *Ghent Altarpiece* (1432), verify that the use of side information can be highly beneficial in this application. Furthermore, due to the high resolution of the data relative to the restricted patch size, the multi-scale version of the proposed algorithm improves the quality of

Table III
OBJECTIVE QUALITY ASSESSMENT OF THE X-RAY SEPARATION PERFORMANCE OF DIFFERENT METHODS ON SIMULATED MIXTURES.

		Mixture 1		Mixture 2		Mixture 3		Mixture 4		Mixture 5	
	Image	PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM
MCA fixed	X-ray 1	25.69	0.7941	30.87	0.9003	27.28	0.7915	27.99	0.7972	26.96	0.8473
	X-ray 2	25.50	0.8134	30.73	0.8818	27.15	0.8198	27.86	0.8628	26.78	0.8068
MCA trained	X-ray 1	26.04	0.8245	31.07	0.8381	28.13	0.7703	27.56	0.7783	27.24	0.8258
	X-ray 2	25.83	0.8485	31.15	0.8189	27.23	0.6966	27.41	0.8464	27.05	0.7927
Proposed	X-ray 1	**26.21**	**0.8583**	**31.91**	**0.9072**	**28.54**	**0.8656**	**28.31**	**0.8266**	**27.34**	**0.8592**
	X-ray 2	**26.00**	**0.8759**	**31.75**	**0.8892**	**28.36**	**0.8859**	**28.16**	**0.8921**	**27.14**	**0.8329**



(a)



(b)

Figure 7. Visual evaluation of the proposed multi-scale method in the separation of the X-ray image in Fig. 5(c); (a) separated side 1, (b) separated side 2. The reconstructions include the X-ray specific v component.

the results significantly. We also observed experimentally that omitting the high frequency crack pixels in the dictionary learning process results in smoother and visually more pleasant separation results. Finally, the superiority of our method, compared to the state-of-the-art MCA technique [20], [21], [35], was validated visually using real data and objectively using simulated X-ray image mixtures.

ACKNOWLEDGMENT

Miguel Rodrigues acknowledges valuable feedback from Jonathon Chambers.

REFERENCES

[1] N. Deligiannis, J. F. C. Mota, B. Cornelis, M. R. D. Rodrigues, and I. Daubechies, “X-ray image separation via coupled dictionary learning,” in *IEEE Int. Conf. Image Process. (ICIP)*. [Available: arXiv:1605.06474], 2016, pp. 1–5.

[2] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[3] D. Zhang and D. Shen, “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease,” *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

[4] S. Wang, L. Zhang, Y. Liang, and Q. Pan, “Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2216–2223.

[5] C. R. Johnson, E. Hendriks, I. J. Berezhnoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang, “Image processing for artist identification,” *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 37–48, 2008.

[6] N. van Noord, E. Hendriks, and E. Postma, “Toward discovery of the artist’s style: Learning to recognize artists by their artworks,” *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 46–54, July 2015.

[7] D. Johnson, C. Johnson Jr., A. Klein, W. Sethares, H. Lee, and E. Hendriks, “A thread counting algorithm for art forensics,” in *IEEE Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, Jan 2009, pp. 679–684.

[8] H. Yang, J. Lu, W. Brown, I. Daubechies, and L. Ying, “Quantitative canvas weave analysis using 2-D synchrosqueezed transforms: Application of time-frequency analysis to art investigation,” *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 55–63, July 2015.

[9] L. van der Maaten and R. Erdmann, “Automatic thread-level canvas analysis: A machine-learning approach to analyzing the canvas of paintings,” *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 38–45, July 2015.

[10] B. Cornelis, T. Ružić, E. Gezels, A. Dooms, A. Pižurica, L. Platiša, J. Cornelis, M. Martens, M. De Mey, and I. Daubechies, “Crack detection and inpainting for virtual restoration of paintings: The case of the Ghent Altarpiece,” *Signal Process.*, 2012.

[11] B. Cornelis, Y. Yang, J. Vogelstein, A. Dooms, I. Daubechies, and D. Dunson, “Bayesian crack detection in ultra high resolution multimodal images of paintings,” *arXiv preprint, ArXiv:1304.5894*, 2013.

[12] A. Pizurica, L. Platisa, T. Ruzic, B. Cornelis, A. Dooms, M. Martens, H. Dubois, B. Devolder, M. De Mey, and I. Daubechies, “Digital image processing of the Ghent Altarpiece: Supporting the painting’s study and conservation treatment,” *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 112–122, 2015.

[13] B. Cornelis, A. Dooms, J. Cornelis, F. Leen, and P. Schelkens, “Digital painting analysis, at the cross section of engineering, mathematics and culture,” in *Eur. Signal Process. Conf. (EUSIPCO)*, 2011, pp. 1254–1258.

[14] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2004, vol. 46.

[15] A. Hyvärinen, “Gaussian moments for noisy independent component analysis,” *IEEE Signal Process. Lett.*, vol. 6, no. 6, pp. 145–147, 1999.

[16] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman, “Static and dynamic source separation using nonnegative factorizations: A unified view,” *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 66–75, 2014.

[17] J. W. Miskin, “Ensemble learning for independent component analysis,” in *Advances in Independent Component Analysis*, 2000.

[18] D. B. Rowe, “A Bayesian approach to blind source separation,” *Journal of Interdisciplinary Mathematics*, vol. 5, no. 1, pp. 49–76, 2002.

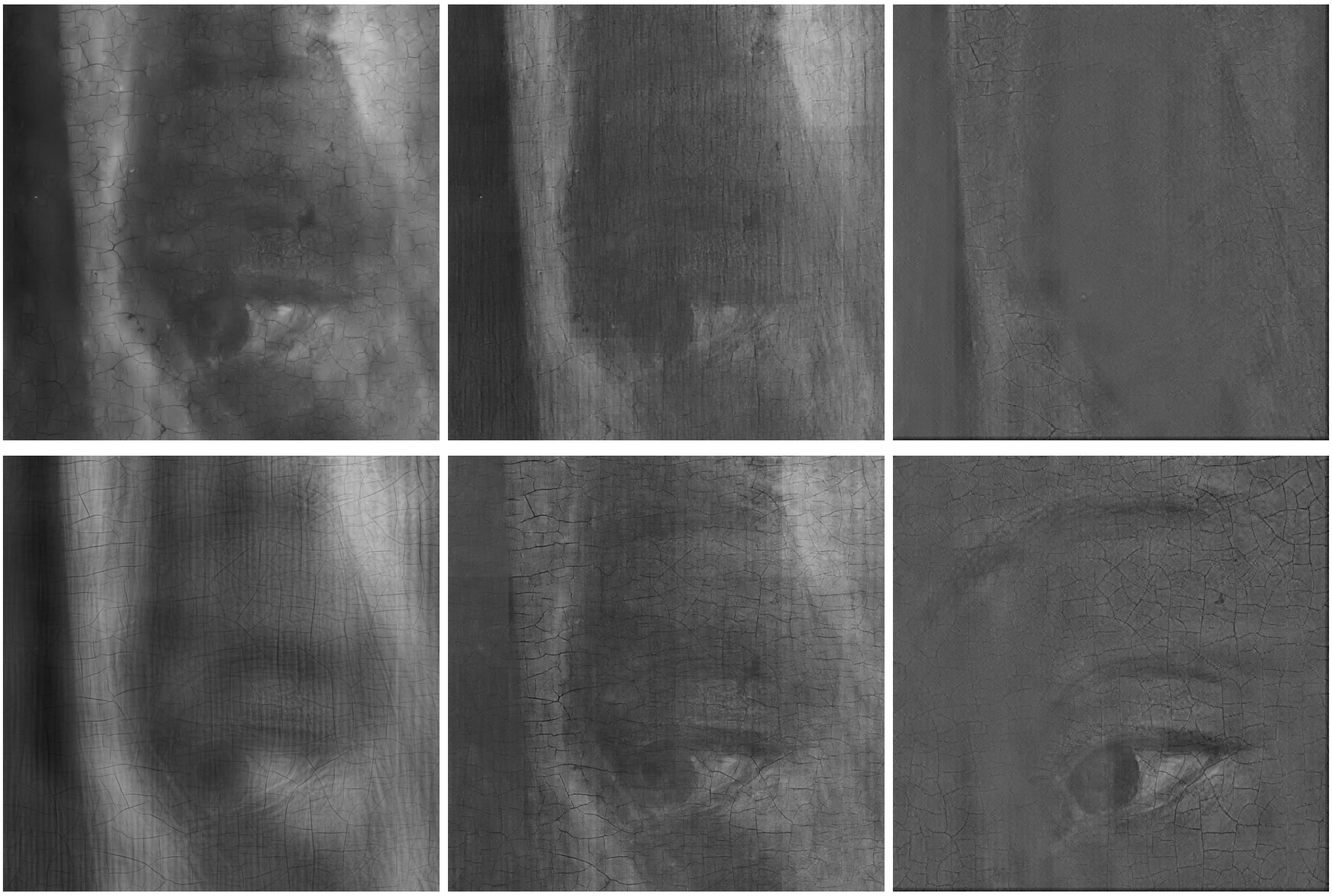


Figure 8. Visual evaluation of the proposed multi-scale method in the separation of the X-ray image in Fig. 5(c); (first row) separated side 1, (second row) separated side 2. The competing methods are: (first column) MCA with fixed dictionaries [30], (second column) multi-scale MCA with K-SVD, (third column) Proposed method without including the v component.

[19] K. Kayabol, E. Kuruoğlu, and B. Sankur, "Bayesian separation of images modeled with MRFs using MCMC," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 982–994, 2009.

[20] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2662–2674, 2007.

[21] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[22] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[23] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[24] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 127–144, 2014.

[25] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.

[26] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[27] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math*, vol. 57, p. 14131541, 2004.

[28] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component separation," *Advances in Imaging and Electron Physics*, vol. 132, pp. 287–348, 2004.

[29] J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 409–412, 2006.

[30] R. Yin, D. Dunson, B. Cornelis, B. Brown, N. Ocon, and I. Daubechies, "Digital cradle removal in X-ray images of art paintings," in *IEEE Int. Conf. Image Process. (ICIP)*, Oct 2014, pp. 4299–4303.

[31] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *Proc. 8th IEEE DSP Workshop*, vol. 8, 1998, p. 86.

[32] K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM journal on Mathematical Analysis*, vol. 39, no. 1, pp. 298–318, 2007.

[33] K. Engan, S. O. Aase, and J. Hakon-Husoy, "Method of optimal directions for frame design," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 1999, pp. 2443–2446.

[34] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[35] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Blind separation of image sources via adaptive dictionary learning," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 2921–2930, 2012.

[36] R. Yan, L. Shao, and Y. Liu, "Nonlocal hierarchical dictionary learning using wavelets for image denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4689–4698, 2013.

[37] B. Ophir, M. Lustig, and M. Elad, "Multi-scale dictionary learning using wavelets," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1014–1024, 2011.

[38] A. Liutkus, J. Pinel, R. Badeau, L. Girin, and G. Richard, "Informed source separation through spectrogram coding and data embedding." *Signal Process.*, vol. 92, no. 8, pp. 1937–1949, 2012.

[39] S. Gorlow and S. Marchand, "Informed separation of spatial images of stereo music recordings using second-order statistics," in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*. IEEE, 2013, pp. 1–6.

[40] B. Chen and G. W. Wornell, "Quantization index modulation: A class

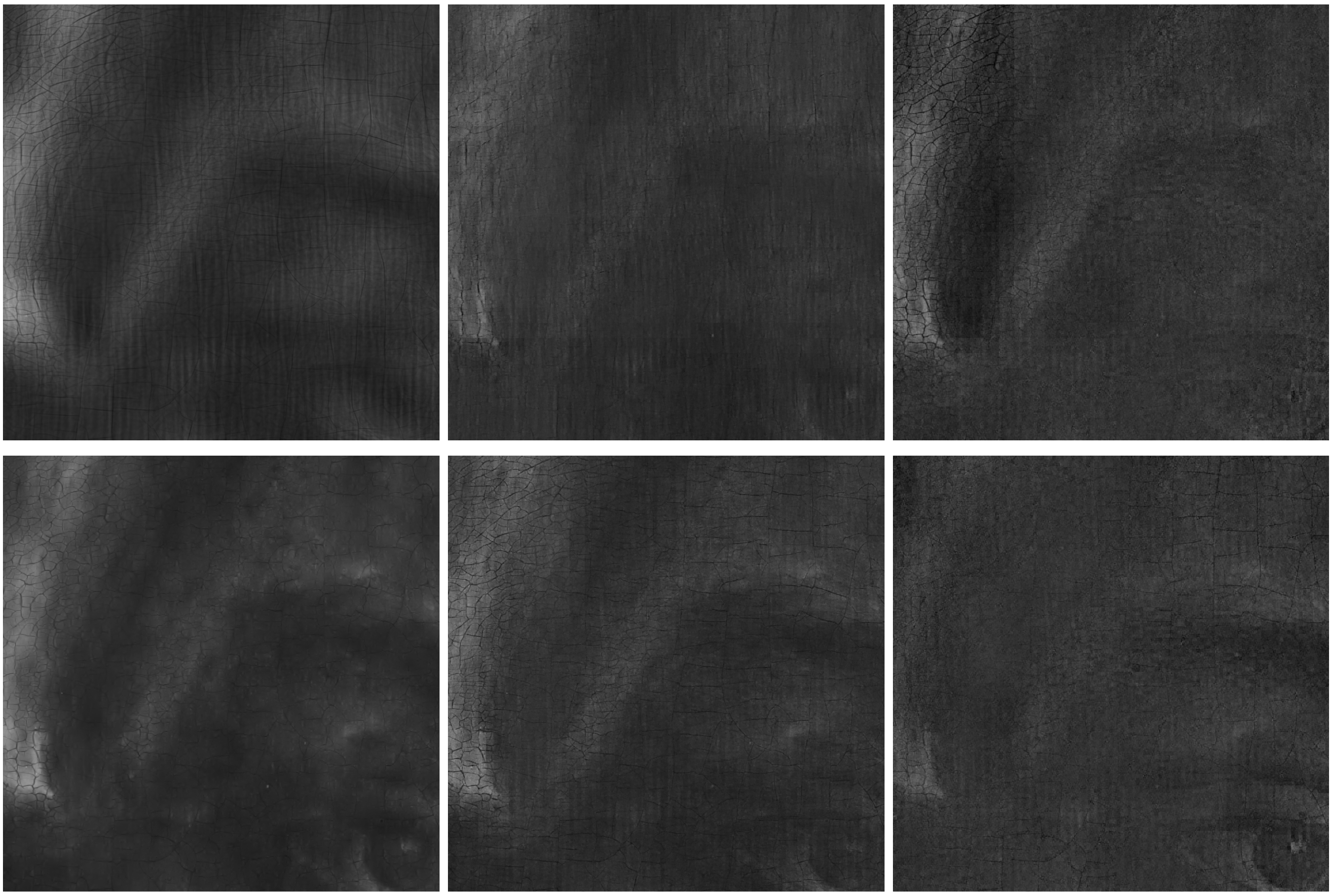


Figure 9. Visual evaluation of the proposed multi-scale method in the separation of the X-ray image in Fig. 5(f); (first row) separated side 1, (second row) separated side 2. The competing methods are: (first column) MCA with fixed dictionaries [30], (second column) multi-scale MCA with K-SVD, (third column) Proposed method including the v component.

of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, 1999.

[41] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[42] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, 2003.

[43] W. Chen, I. Wassell, and M. R. Rodrigues, "Dictionary design for distributed compressive sensing," *IEEE Signal Process. Lett.*, vol. 22, no. 1, pp. 95–99, 2015.

[44] H. Zayyani, M. Korki, and F. Marvasti, "Dictionary learning for blind one bit compressed sensing," *IEEE Signal Process. Lett.*, vol. 23, no. 2, p. 187191, 2016.

[45] D. A. Spielman, H. Wang, and J. Wright, "Exact recovery of sparsely-used dictionaries," in *Conference on Learning Theory (COLT)*, 2012.

[46] S. Arora, R. Ge, and A. Moitra, "New algorithms for learning incoherent and overcomplete dictionaries," *JMLR: Workshop and Conference Proceedings*, vol. 35:1-28, 2014.

[47] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, 1983.

[48] G. Monaci, P. Jost, P. Vandergheynst, B. Mailhe, S. Lesage, and R. Gribonval, "Learning multimodal dictionaries," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2272–2283, 2007.

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

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

[51] Y. Jia, M. Salzmann, and T. Darrell, "Factorized latent spaces with structured sparsity," in *Advances in Neural Information Processing Systems*, 2010, pp. 982–990.

[52] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2010.

[53] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv preprint arXiv:1408.5250*, 2014.

[54] ——, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 512–516.

[55] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional noisy features in the presence of side information," *IEEE Trans. Inf. Theory*, 2016. (in press).

[56] J. Scarlett, J. S. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 427–439, 2013.

[57] E. Zimos, J. F. C. Mota, M. R. D. Rodrigues, and N. Deligiannis, "Bayesian compressed sensing with heterogeneous side information," in *IEEE Data Compression Conf. (DCC)*, 2016.

[58] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *IEEE Int. Symp. Inf. Theory (ISIT)*. IEEE, 2009, pp. 483–487.

[59] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[60] M. B. Wakin, M. F. Duarte, S. Sarvotham, D. Baron, and R. G. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Proc. Neural Inform. Process. Syst., (NIPS)*, 2005.

[61] I. Arganda-Carreras, C. O. S. Sorzano, R. Marabini, J. M. Carazo, C. O. de Solorzano, and J. Kybic, "Consistent and elastic registration

of histological sections using vector-spline regularization," in *Computer Vision Approaches to Medical Image Analysis*, May 2006, pp. 85–95.

[62] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[63] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[64] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[65] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *ACM-SIAM Symposium on Discrete Algorithms*, 2003, pp. 243–252.

[66] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[67] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[68] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.