Multimodal Target Detection by Sparse Coding: Application to Paint Loss Detection in Paintings

Shaoguang Huang, *Member, IEEE,* Bruno Cornelis, *Member, IEEE,* Bart Devolder, Maximiliaan Martens and Aleksandra Pižurica, *Senior Member, IEEE*

***Abstract*—Sparse representation based methods have demonstrated their superior performance in target detection tasks compared to more traditional approaches such as matched subspace detectors and adaptive subspace detectors. However, the existing sparsity-based target detection methods were mostly formulated for and validated on a single imaging modality (sometimes with multiple spectral bands). In many application domains, including art investigation, multimodal data, acquired by different sensors are readily available, and yet, efficient processing techniques for such data are still scarce. In this paper, we propose a sparsity-based multimodal target detection method that processes jointly the information from multiple imaging modalities in a kernel feature space, and making use of the spatial context. We develop our target detector such to be robust to errors in labelled data, which is especially important in applications like digital painting analysis, where pixel-wise manual annotations are unreliable. We apply the proposed method to a challenging application of paint loss detection in master paintings and we demonstrate its effectiveness on a case study with multimodal acquisitions of the *Ghent Altarpiece*.**

***Index Terms*—Sparse representation, target detection, paint loss, kernel, multiple imaging modalities.**

I. Introduction

DIGITAL painting analysis has made vast progress over the recent years, powered by a wide range of new image acquisition techniques [1]. Numerous tasks, such as characterization of painting style and forgery detection [2, 3], crack detection [4–7], authorship identification [8], classification of ancient coins [9], thread count analysis (of canvas supports) [10] and portraits [11], indexing of cultural heritage collections [12], colorization of historical art pieces [13], removal of canvas texture [14], source separation [15] and inpainting [16, 17], have demonstrated the great potential of digital image processing and machine learning in art investigation. Multimodal imaging is now routinely employed in order to support the technical study of art works [6, 15], their restoration, conservation, and even presentation. Consulting different modalities of the same object often aids in uncovering regions or patterns of interest that would otherwise remain unnoticed, offers new insights and support for specific decisions that are taken during restoration treatments [17].

Fig. 1. Examples of paint loss in the macrophotography after cleaning. Left: original painting. Right: enlarged paint loss. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

We address the problem of paint loss detection in digitized paintings and we formalize it as a particular instance of more general target detection from multimodal data. Our goal is thus to discriminate paint loss pixels, i.e. the *target*, from the non paint loss or *background* pixels in an automatic fashion. Paint losses in oil paintings are typically caused by abrasion or mechanical fracture and are often retouched or overpainted during numerous restoration campaigns. Modern conservation treatments typically require the removal of old varnish as well as old retouches and overpaint, revealing paint loss, such as in the examples shown in Fig. 1. The paint loss regions can vary significantly in size, from very tiny areas to larger holes or areas of missing paint, and typically have complex and irregular shapes. Detection of such paint loss areas is of great importance to conservators in estimating the extent of the damage within the painting, which is required for documentation purposes on one hand, but is also a crucial step in the virtual inpainting of the painting's digital counterpart. The latter can act as a simulation within a decision-making process before the actual restoration. Digitized scans of works of art are often taken in different modalities during treatment, as shown in Fig. 2. This allows painting conservators to locate various areas of interest, such as overpaint and retouchings, as well as paint losses, in a more reliable way. In general, locating these areas is a very tedious procedure, especially in larger paintings, and is often only done approximately or in relatively small areas. Despite its importance, the problem of automatic paint loss detection has received little attention in the literature so far. Besides our earlier preliminary results, reported in two conference abstracts [18, 19], we are not aware

This work was funded by the Fonds voor Wetenschappelijk Onderzoek (FWO) project: G.OA26.17N and received also funding from the Flemish Government (AI Research program).

S. Huang and A. Pižurica are with the Department of Telecommunications and Information Processing, TELIN-GAIM, Ghent University, 9000 Ghent, Belgium (e-mail: shaoguang.huang@ugent.be).

B. Cornelis is with the Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels 1050, Belgium.

B. Devolder is with the Princeton University Art Museum, Princeton, NJ 08544, USA.

M. Martens is with the Department of Art, music and theatre sciences, Ghent University, Belgium.

(a) (b) (c) (d) (e)

Fig. 2. Examples of imaging modalities acquired at different restoration stages of the *Ghent Altarpiece* depicting a detail of the panel *Prophet Zacharias*. (a) Macrophotography before cleaning. (b) Macrophotography after cleaning. (c) infrared macrophotography before cleaning. (d) Infrared reflectography after cleaning. (e) X-radiography before cleaning. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

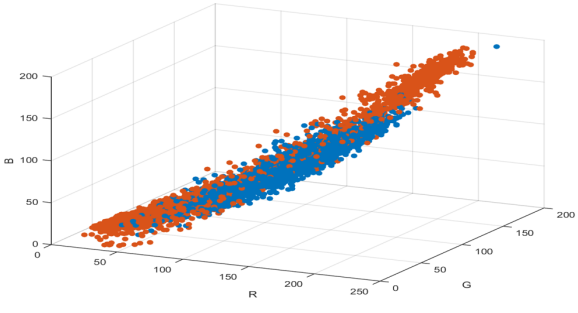


Fig. 3. Scatter plot of RGB values of randomly selected paint loss and background pixels in a macrophotograph after cleaning, shown in Fig. 2 (b) (red: paint loss; blue: background).

of any elaborate technical studies on this subject.

The detection of paint loss areas is particularly challenging due to their large variation in shape, size and intensity as well as their complex background, as illustrated in Fig. 2 (b). The scatter plot in Fig. 3 depicts the distribution of RGB values for randomly selected paint loss and background pixels extracted from the painting in Fig. 1. Clearly, it is not trivial to separate these two classes of pixels based on their color alone. We shall make use of multimodal data, such as the examples illustrated in Fig. 2, and our method will be formulated such to admit an arbitrary amount and type of imaging modalities, as they can differ from one painting to another.

Detecting paint loss areas can be seen as an instance of the more general target detection problem, where the purpose is to distinguish a particular target of interest from the background. More specifically, the target can be a signal in a communication channel, landmines in hyperspectral images, vehicles and pedestrians in videos or thermal images, etc. The background can consist of interfering communication channels or the union of different non-target background classes. When only the background information is known, the corresponding problem is called *anomaly detection*. For the task of paint loss detection, some prior information of target and background is typically available in the form of labelled samples that can be used to model and estimate the statistical characteristics of both classes. In the next section, we will refer to the general formulation of target detection. Generally, target detection can be formulated as choosing between the two competing hypotheses H_0 : "target absent" and H_1 : "target present". It is important to understand that in different applications the targets' appearance may have different correlations with the background. For the tasks of vehicle or pedestrian detection in video sequences, the target and the background are usually in different, yet relatively well separated subspaces, showing substantially different characteristics, and typically one cannot be represented by the other. On the contrary, in radar systems, where the background is a structured interference in the communication channel, the observed target signal is represented by a sum of the ideal target signal, background interference and noise. In other words, in the former type of the detection problem, the observed target replaces the background while in the latter one the observed target is a combination of an ideal target and the background. The application that we shall address is of the former type, and in particular relates to the detection of target pixels in images. Although the framework that we develop will be for compactness represented for this particular case (target replaces the background), the whole approach can be readily extended to the second case as well (where the target is a combination of the background and "pure" target pixels). The appropriate remarks in this respect will be given at the corresponding places in the paper.

Well-known statistical detectors include the Neyman-Pearson detector [20, 21], spectral matched filters [22], matched subspace detector [23], and adaptive subspace detectors [24]. These methods calculate a detection statistic, which is then compared to a threshold value to determine whether a target is present. Typically, the explicit distribution characteristics of the target and background are assumed and then likelihood ratio (LR) or the generalized likelihood ratio test (GLRT) are employed to develop the detectors. However, the assumptions with respect to the probability distribution and covariance matrix of both the target and background may not always be representative for real world data, leading to a deteriorated performance in practice.

Methods based on sparse representation, which are widely employed in applications such as face recognition [25], hy-

perspectral image (HSI) classification [26], HSI unmixing [27] and denoising [28], have been increasingly used in target detection too following the seminal work [29]. Recent representatives include [30–33]. The common underlying idea is that each test sample can be sparsely represented as a linear combination of a few prototype samples (commonly referred to as atoms) from a well-designed dictionary. When the dictionary atoms are selected from labelled training samples, the sparse coefficients of a test sample provide directly the information about the likely class labels. This approach is known as sparse representation classification (SRC) [25]. Sparsity-based target detection methods have been successfully employed in a wide range of tasks, including landmine detection [29] and vehicle detection [30, 31] from HSIs, defect detection in semiconductor units [32] and detection of underwater targets with sonar images [33, 34].

The target detection methods described above were mostly formulated and validated on a single imaging modality (typically with multiple spectral bands). Nowadays, multimodal data are routinely acquired by various sensors and leveraging their complementary information can help in discriminating better between target and background. Existing multimodal learning algorithms have demonstrated superior performance in clustering [35, 36], classification [37–39], image super-resolution [40], source separation [15] and event detection [41] tasks. In this paper we first introduce a general sparsity-based multimodal target detection method, which we then tailor for the task of paint loss detection. The information from multiple imaging acquisitions is projected to a high-dimensional kernel feature space, which facilitates discriminating between target and background and at the same time avoids the need of constructing dictionaries in the explicit projected feature space. In addition, we make use of the spatial context to model the spatial dependencies among neighbouring pixels. An important aspect of our work is dealing with erroneous labelled data. To increase the robustness to errors in labelled data, we propose a simple and elegant multi-pass scheme, which makes the detection more robust to imperfect training. As a case study, the proposed method is evaluated on a very high-resolution multi-modal data set comprising digital scans of the *Ghent Altarpiece*, acquired during the restoration/conservation treatment of this masterpiece. The main contributions of the paper can be summarized as follows:

1) A multimodal target detection method based on sparse coding is proposed that makes effective use of the rich information from multimodal data to discriminate better between the target and the background. An effective method based on a kernel function is devised to implicitly project the fused features to a high-dimensional feature space where the target and background classes can be better separated by extracting the intrinsic non-linear information from the complex data structure.
2) The spatial dependencies among the neighbouring pixels are exploited by applying a smoothing filter on the residual maps derived from a kernel sparse representation problem, improving thereby the accuracy and the reliability of the detection. While discontinuity-adaptive filtering has been widely employed in image enhancement, we show how it can significantly improve the performance of sparsity-based target detection, by filtering appropriately the class-specific residuals of the sparse approximation.
3) We design and thoroughly evaluate a multi-pass strategy in our multimodal target detection approach to alleviate the adverse effect of erroneous labeled training samples, which contaminate the constructed dictionaries. At the same time, this approach lowers the computational complexity, avoiding the need to store and process large dictionaries all at once.
4) The proposed method is tailored to the specific task of paint loss detection in paintings for which we extended the detection method to automatically discard cracks (as an unwanted pseudo-target class). To our knowledge, this is the first technical paper (apart from our earlier conference abstract) to address this important problem in digital painting analysis. The case study on the *Ghent Altarpiece* illustrates clearly the effectiveness of our method in this task and its superior performance compared to existing statistical and sparse-representation based target detection methods. We also prove a wider applicability of the proposed method to other detection tasks.

The rest of this paper is organized as follows. Section II reviews briefly the relevant classical target detectors. Section III presents the proposed target detection method in general, and Section IV describes an extended formulation to a specific task of paint loss detection. Experimental results and analysis with a case study on the *Ghent Altarpiece* are presented in Section V. This Section includes also a generalization study with a different detection task, namely target detection in a hyperspectral image. The conclusions are drawn in Section VI.

II. Target detection based on sparse coding

Here we review briefly classical matched subspace detection, and then we turn to sparsity based detectors and their representatives. We refer to the detection of target pixels in images. For compactness, the indices denoting the spatial location of pixels are suppressed. Let $\mathbf{x} \in \mathbb{R}^m$ denote an input sample. In particular, in the following analysis this is a vector composed of pixel values in m image channels.

A. Matched Subspace Detectors

Matched Subspace Detectors (MSD) [23] assume two linearly independent subspaces: the background and the target subspace, and formulate the detection problem as follows:

$$\begin{aligned} H_0 &: \mathbf{x} = \mathbf{B}\boldsymbol{\beta}_0 + \mathbf{n}, \quad \text{target absent} \\ H_1 &: \mathbf{x} = \mathbf{T}\boldsymbol{\alpha}_1 + \mathbf{B}\boldsymbol{\beta}_1 + \mathbf{n}, \quad \text{target present} \end{aligned} \tag{1}$$

where $\mathbf{B}$ and $\mathbf{T}$ are two matrices that describe the background and target subspaces, (composed of the eigenvectors of the background and target covariance matrices), and $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are unknown coefficient vectors. The noise $\mathbf{n}$ is drawn from

the normal distribution $\mathcal{N}[0, \sigma^2\mathbf{I}]$ with unknown σ and thus the detection problem stated above becomes:

$$\begin{aligned} H_0 &: \mathbf{x} \sim \mathcal{N}[\mathbf{B}\beta_0, \sigma^2\mathbf{I}], \quad \text{target absent} \\ H_1 &: \mathbf{x} \sim \mathcal{N}[\mathbf{T}\alpha_1 + \mathbf{B}\beta_1, \sigma^2\mathbf{I}], \quad \text{target present} \end{aligned} \tag{2}$$

Under this model, the resulting GLRT yields

$$T_{MSD}(\mathbf{x}) = \frac{\mathbf{x}^T(\mathbf{I} - \mathbf{P_B})\mathbf{x}}{\mathbf{x}^T(\mathbf{I} - \mathbf{P_{TB}})\mathbf{x}} \tag{3}$$

where $\mathbf{P_B}$ and $\mathbf{P_{TB}}$ are the projection matrices associated with the background subspace and the target-and-background subspace, respectively:

$$\mathbf{P_B} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T \tag{4}$$

$$\mathbf{P_{TB}} = [\mathbf{T}, \mathbf{B}]([\mathbf{T}, \mathbf{B}]^T[\mathbf{T}, \mathbf{B}])^{-1}[\mathbf{T}, \mathbf{B}]^T. \tag{5}$$

When $T_{MSD}(\mathbf{x})$ is larger than a prescribed threshold δ, the test sample is labelled as target; otherwise, it is labelled as background.

Adaptive Subspace Detectors (ASD) [24] adapt the matched subspace detectors to unknown noise covariance matrices. The detection model is formulated as follows:

$$\begin{aligned} H_0 &: \mathbf{x} = \mathbf{v}, \quad \text{target absent} \\ H_1 &: \mathbf{x} = \mathbf{T}\boldsymbol{\theta} + \sigma\mathbf{v}, \quad \text{target present.} \end{aligned} \tag{6}$$

$\mathbf{T}$ is again the target matrix, whose columns are linearly independent vectors that span the target space. The scaled backround noise $\mathbf{v}$ under H_1 allows to model elegantly subpixel targets where part of the background is not masked by the target. Under these hypotheses, the GLRT of ASD results in the following detector:

$$T_{ASD}(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{T}(\mathbf{T}^T\mathbf{\Sigma}^{-1}\mathbf{T})^{-1}\mathbf{T}^T\mathbf{\Sigma}^{-1}\mathbf{x}}{\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}}, \tag{7}$$

where $\mathbf{\Sigma}$ is the estimated background covariance matrix. When the target subspace $\mathbf{T}$ is specified by the direction of a single vector $\mathbf{t}$, the particular version of ASD is called adaptive cosine estimator (ACE) [42]:

$$T_{ACE}(\mathbf{x}) = \frac{(\mathbf{t}^T\mathbf{\Sigma}^{-1}\mathbf{x})^2}{(\mathbf{t}^T\mathbf{\Sigma}^{-1}\mathbf{t})(\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x})}. \tag{8}$$

B. Sparse Representation-based Target Detectors

Similarly to the matched subspace detectors, sparse representation based target detectors (STD) [29, 30], also model a test sample to lie in a union of background and target subspaces. Unlike MSD, the STD model imposes no assumption about the particular target and background distributions. The STD model in [29] starts from a general assumption that if $\mathbf{x}$ is a background pixel, its spectrum approximately lies in a lower-dimensional subspace spanned by background training pixels $\{\mathbf{d}_i^b\}_{i=1,\dots N_b}$, and similarly if $\mathbf{x}$ is a target pixel, its spectrum approximately lies in a lower-dimensional subspace spanned by the target training pixels $\{\mathbf{d}_i^t\}_{i=1,\dots N_t}$. Formally:

$$\begin{aligned} H_0 &: \mathbf{x} = \mathbf{D}_b\boldsymbol{\alpha}_b + \mathbf{n}_1, \quad \mathbf{x} \text{ is a background pixel} \\ H_1 &: \mathbf{x} = \mathbf{D}_t\boldsymbol{\alpha}_t + \mathbf{n}_2, \quad \mathbf{x} \text{ is a target pixel,} \end{aligned} \tag{9}$$

where $\mathbf{D}_b \in \mathbb{R}^{m\times N_b}$ and $\mathbf{D}_t \in \mathbb{R}^{m\times N_t}$ are the background dictionary and the target dictionary, the columns of which are the background training samples and the target training pixels, respectively. $\boldsymbol{\alpha}_b \in \mathbb{R}^{N_b}$ and $\boldsymbol{\alpha}_t \in \mathbb{R}^{N_t}$ are sparse vectors, the entries of which are the abundances of the corresponding dictionary atoms, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are (arbitrarily distributed) noise vectors. Similarly to the MSD model, an unknown test sample is modelled to lie in the union of the two subspaces, and is thus represented by a linear combination

$$\mathbf{x} = \mathbf{D}_b\boldsymbol{\alpha}_b + \mathbf{D}_t\boldsymbol{\alpha}_t = \mathbf{D}\boldsymbol{\alpha} \tag{10}$$

where $\mathbf{x} = \mathbf{D}\alpha$, and $\mathbf{D} = [\mathbf{D}_t, \mathbf{D}_b]$ is a composite dictionary consisting of both the background and target samples. The sparse vector $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_b, \boldsymbol{\alpha}_t]$ is estimated as follows

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|^2 \quad s.t. \;\; \|\boldsymbol{\alpha}\|_0 < K_0, \tag{11}$$

where $\|\boldsymbol{\alpha}\|_0$ denotes the number of non-zero elements in $\boldsymbol{\alpha}$, and the constant K_0 is the maximum sparsity level. This recovery problem implicitly leads to a competition between the two subspaces [29], i.e., to a competition between the two hypotheses in (9). For the background pixels, ideally the estimated $\hat{\boldsymbol{\alpha}}_t$ part will be a zero vector and $\hat{\boldsymbol{\alpha}}_b$ a sparse vector. The opposite is true for the target pixels. Hence, the detector output is defined as the difference between the two residuals $r_b(\mathbf{x}) = \|\mathbf{x} - \mathbf{D}_b\hat{\boldsymbol{\alpha}}_b\|_2$ and $r_t(\mathbf{x}) = \|\mathbf{x} - \mathbf{D}_t\hat{\boldsymbol{\alpha}}_t\|_2$:

$$T_{STD}(\mathbf{x}) = r_b(\mathbf{x}) - r_t(\mathbf{x}) \tag{12}$$

The class label is then obtained by comparing the value of T_{STD} to a prescribed threshold δ. In case $\delta = 0$, the STD method reduces to the SRC method.

The approach of [30] extends the STD detector cleverly to a more general case where the target pixels are a mixture of "pure" target and background pixels. This situation arises in the case of subpixel target detection, which is common in hyperspectral imaging, due to the relatively small spatial resolution. In this case, the model in (9) is extended as

$$\begin{aligned} H_0 &: \mathbf{x} = \mathbf{D}_b\boldsymbol{\alpha}_b + \mathbf{n}_1, \quad \text{target absent} \\ H_1 &: \mathbf{x} = \mathbf{D}_b\boldsymbol{\gamma}_b + \mathbf{D}_t\boldsymbol{\gamma}_t + \mathbf{n}_2, \quad \text{target present} \end{aligned} \tag{13}$$

Correspondingly, the target residual r_t in (12) is obtained as $r_t(\mathbf{x}) = \|\mathbf{x} - \mathbf{D}\hat{\boldsymbol{\gamma}}\|_2$, and the sparse codes $\hat{\boldsymbol{\alpha}}_b$ and $\hat{\boldsymbol{\gamma}} = [\hat{\boldsymbol{\gamma}}_b, \hat{\boldsymbol{\gamma}}_t]$ are estimated separately solving the problems of the type (11). For details, the interested reader is referred to [30]. Since we are not dealing with subpixel classification in our application, we shall not discuss this model further. It should be noted, however, that the multimodal sparse detector that we develop in the next section, starting from the original STD model, can also equivalently be developed starting from the model in (13).

III. Kernel-domain Multimodal STD

We develop here a general multimodal target detection method based on sparse coding, which makes use of the information coming from multiple imaging modalities in a kernel feature space and takes into account the inherent spatial context of the input image. The overall method comprises two

parts: data preprocessing and target detection. In the following, we describe first the formation of the test samples from the input multimodal data, then we present the proposed target detection in a kernel feature space, and finally incorporate this method into a multi-pass scheme to mitigate the effect of unreliable label data.

A. Data preprocessing

Typically, multi-modal imaging data is acquired by different devices and the acquired images often have different resolutions. We are interested in target detection up to pixel-level accuracy. Thus pixel-perfect alignment of the different image modalities is required. We use for this purpose a joint photometric and geometric image registration technique [43] and denote by $\mathcal{X}^i \in \mathbb{R}^{M\times N\times d_i}(i = 1, 2, ..., n)$ the i-th available image modality, where M and N are the spatial dimensions of the data.

We concatenate all the imaging modalities $\mathcal{X}^i$ to form a 3-D data cube $\mathcal{X} \in \mathbb{R}^{M\times N\times m}$ with $m = \sum_{i=1}^{n} d_i$, and denote by $\mathbf{X} \in \mathbb{R}^{m\times MN}$ the corresponding data reshaped to a 2-D matrix. Each column in this matrix is a vector of pixel values at a given spatial location in all the modalities. To increase robustness to noise, we assign to each pixel also the values of pixels from a small $w \times w$ spatial window around it. These vectorized pixel values across all the modalities constitute a new sample $\mathbf{x} \in \mathbb{R}^d$, where $d = mw^2$, and we suppress the location index for compactness and clarity.

B. Target detection in kernel feature space

The target and background classes in the fused data space may be difficult to separate by linear detectors such as ACE, MSD and STD. An efficient way to approach this problem is to project the data into a higher-dimensional feature space by using a non-linear mapping function, where the structure of the non-linear data becomes more apparent and different classes become better separated. A non-linear mapping with a kernel function can effectively group data points within the same distribution and make them linearly separable [44–48]. We project the fused data sample $\mathbf{x}$ to a high-dimensional feature space, $\mathscr{F}$, by an implicit mapping function $\phi : \mathbb{R}^d \to \mathscr{F} \subset \mathbb{R}^{\hat{d}}$. The kernel function, $\boldsymbol{\kappa} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined as the inner product:

$$\boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle, \tag{14}$$

We utilize a commonly used radial basis function (RBF) with the Gaussian kernel:

$$\boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}) = \exp(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma}), \tag{15}$$

where $\sigma > 0$ is a parameter to control the width of the RBF.

While the sparsity detector STD described in Section II models the data linearly, we use the kernel function to first map the observed test sample into a high-dimensional feature space. The two competing hypotheses are now expressed as:

$$\begin{aligned} H_0 &: \phi(\mathbf{x}) = \phi(\mathbf{D}_0)\boldsymbol{\alpha}_0 + \phi(\mathbf{n}_0), \quad \text{background pixel} \\ H_1 &: \phi(\mathbf{x}) = \phi(\mathbf{D}_1)\boldsymbol{\alpha}_1 + \phi(\mathbf{n}_1), \quad \text{target pixel} \end{aligned} \tag{16}$$

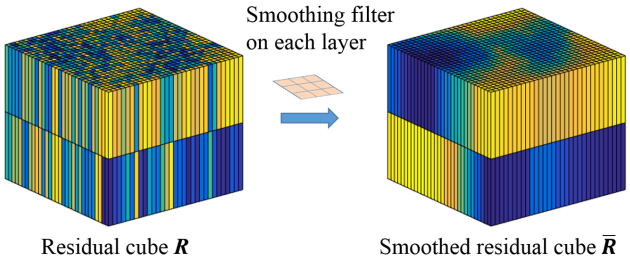


Fig. 4. Smoothing of the residual cube.

where $\phi(\mathbf{D}_0)$ and $\phi(\mathbf{D}_1)$ are the background and target subspaces in $\mathscr{F}$, respectively, and $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are the corresponding sparse coefficients, which are obtained by solving the following optimization problem:

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}=[\boldsymbol{\alpha}_0;\boldsymbol{\alpha}_1]}{\arg\min} \|\phi(\mathbf{x}) - \phi(\mathbf{D})\boldsymbol{\alpha}\|^2 \quad s.t. \ \|\boldsymbol{\alpha}\|_0 < K_0, \tag{17}$$

where $\phi(\mathbf{D}) = [\phi(\mathbf{D}_0), \phi(\mathbf{D}_1)] = [\phi(\mathbf{d}_0^1), ..., \phi(\mathbf{d}_0^{N_1}), \phi(\mathbf{d}_1^1), ..., \phi(\mathbf{d}_1^{N_2})]$. $\mathbf{D}_0$ is a sub-dictionary constructed by the selected training samples $\mathbf{d}_0^i \in \mathbb{R}^d$ from the background and $\mathbf{D}_1$ is another sub-dictionary constructed by the selected target training samples $\mathbf{d}_1^i \in \mathbb{R}^d$. We solve this problem by the kernel orthogonal matching pursuit (KOMP) algorithm [49]. Once the sparse coefficients are calculated, we obtain the class-specific residuals as follows:

$$\begin{aligned} r_i(\phi(\mathbf{x})) &= \|\phi(\mathbf{x}) - \phi(\mathbf{D}_i)\boldsymbol{\alpha}_i\|_2 \\ &= \langle \phi(\mathbf{x}) - \phi(\mathbf{D}_i)\boldsymbol{\alpha}_i, \ \phi(\mathbf{x}) - \phi(\mathbf{D}_i)\boldsymbol{\alpha}_i \rangle^{1/2} \\ &= (\boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}) - 2\boldsymbol{\alpha}_i^T \mathbf{K}_{\mathbf{D}_i} + \boldsymbol{\alpha}_i^T \mathbf{K}_{\mathbf{D}_i\mathbf{D}_i}\boldsymbol{\alpha}_i)^{1/2}, \end{aligned} \tag{18}$$

where $\mathbf{K}_{\mathbf{D}_i} = [\boldsymbol{\kappa}(\mathbf{d}_i^1, \mathbf{x}), \cdots, \boldsymbol{\kappa}(\mathbf{d}_i^{N_i}, \mathbf{x})]^T \in \mathbb{R}^{N_i}$ and $\mathbf{K}_{\mathbf{D}_i\mathbf{D}_i} \in \mathbb{R}^{N_i\times N_i}$ is a matrix with entries $\mathbf{K}_{\mathbf{D}_i\mathbf{D}_i}(s, t) = \boldsymbol{\kappa}(\mathbf{d}_i^s, \mathbf{d}_i^t)$. We collect all the residuals and denote $\mathbf{R} \in \mathbb{R}^{M\times N\times 2}$ as the residual cube, where $M \times N$ is the input image size.

Typically, the target and background are not scattered as isolated pixels in the image but tend to form homogeneous regions. This means that neighbouring pixels in local regions belong to the same class with high probability and thus have similar sparse coefficients. In order to capture this property, we apply a smoothing filter on each layer of the residual cube $\mathbf{R}_i \in \mathbb{R}^{M\times N}$ $(i = 1, 2)$. Specifically, we use a discontinuity-adaptive filtering technique based on weighted least squares (WLS) [50], which has been proved effective in enhancing digital photographs. In our setting, this means replacing the original residual cube $\mathbf{R}$ with its smoothed version $\bar{\mathbf{R}}$, as shown in Fig. 4. Formally, we solve the following optimization problem:

$$\min \sum_i \|\mathbf{R}_i - \bar{\mathbf{R}}_i\|_F^2 + \lambda(\|\mathbf{A}_x\mathbf{D}_x\bar{\mathbf{R}}_i(:)\|^2 + \|\mathbf{A}_y\mathbf{D}_y\bar{\mathbf{R}}_i(:)\|^2), \tag{19}$$

where the matrices $\mathbf{D}_x$ and $\mathbf{D}_y$ are the discrete differentiation operators in the horizontal and vertical directions, respectively, and $\mathbf{A}_x$ and $\mathbf{A}_y$ are the diagonal matrices that contain the

adaptive smoothing weights in the corresponding directions. The unique solution is given by

$$\bar{\mathbf{R}}_i(:) = (\mathbf{I} + \lambda \mathbf{L}_g)^{-1} \mathbf{R}_i(:), \tag{20}$$

where $\mathbf{L}_g = \mathbf{D}_x^T \mathbf{A}_x^2 \mathbf{D}_x + \mathbf{D}_y^T \mathbf{A}_y^2 \mathbf{D}_y$ is a Laplacian matrix. Following the general edge-adaptive filtering strategy of [51], smoothing weights are defined in terms of the discrete gradients of a reference image $\mathbf{L}$, which in our case leads to:

$$\mathbf{A}_x = \text{Diag}((|\mathbf{D}_x \mathbf{L}(:)|^a + \varepsilon)^{-\frac{1}{2}}) \tag{21}$$

$$\mathbf{A}_y = \text{Diag}((|\mathbf{D}_y \mathbf{L}(:)|^a + \varepsilon)^{-\frac{1}{2}}), \tag{22}$$

where ε is a small constant, the exponent a determines the sensitivity to the gradients of $\mathbf{L}$ and $\text{Diag}(\cdot)$ returns a diagonal matrix with the input vector as the main diagonal. Using the smoothed residual $\bar{\mathbf{R}}$, we construct the detector test as:

$$T(\mathbf{x}) = \bar{r}_0(\phi(\mathbf{x})) - \bar{r}_1(\phi(\mathbf{x})), \tag{23}$$

where $\bar{r}_i(\phi(\mathbf{x}))$ is the i-th class residual of $\mathbf{x}$ in $\bar{\mathbf{R}}$. We label the test sample by comparing the value of $T(\mathbf{x})$ with a prescribed threshold δ as follows:

$$class(\mathbf{x}) = \begin{cases} 1 & T(\mathbf{x}) > \delta \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

The proposed approach can be readily generalized to the case where target is a combination of background and "pure" target pixels. In this case the hypotheses formulation (13) will be converted to the kernel version analogously to (16). The resulting sparse codes $\hat{\boldsymbol{\alpha}}_b$ and $\hat{\boldsymbol{\gamma}}$ then follow from solving a sparse coding problem of the same form as (17) with the dictionaries $\mathbf{D}_b$ and $\mathbf{D}$, respectively, to be finally employed in calculating the class residuals for the target and background.

C. Majority voting

The statistics of the target and background classes are in many cases of practical interest estimated solely based on the available annotated data. Erroneous annotations will thus have an adverse effect on inferring discriminative characteristics of the target class compared to the background class, especially when the amount of training target samples is relatively small. Thus a precise data annotation is essential to the task of target detection, especially when only a few target pixels are present in the input image. However, manual annotations are rarely precise up to the pixel level. To solve this problem, instead of a single-pass detection, we develop a multi-pass scheme where we perform multiple detections based on different random selections of training samples followed by a majority voting strategy. This yields a more robust detection in the presence of incorrectly labelled pixels.

Let K denote the number of algorithm runs, each using a different portion of the labelled data set as training samples. This yields K detection results for each pixel. Let $N_c(\mathbf{x})$ denote the number of times that $\mathbf{x}$ was assigned to class c. We select the identity (class) of each pixel as:

$$identity(\mathbf{x}) = \arg\max_c \; P_e(Class(\mathbf{x}) = c) \tag{25}$$

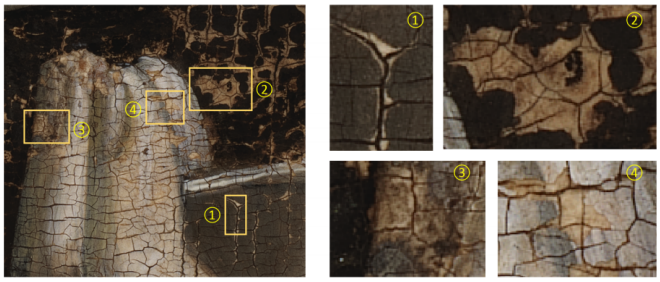


Fig. 5. Examples of cracks and paint loss in old master paintings. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

where $P_e(Class(\mathbf{x}) = c) = N_c(\mathbf{x})/K$ is the empirical probability for a pixel to belong to class $c \in \{Target, Other\}$.

Apart from improved robustness to erroneous annotation, the proposed majority voting scheme offers two other advantages. Firstly, this approach allows us to better deal with cases where labelled samples are not available for the input image, but we have access to K similar annotated images. The majority voting process over the K runs will effectively cope with outliers caused by larger differences in statistical properties of some of the annotated images compared to the input one.

Secondly, the proposed majority voting approach solves efficiently a trade-off between the expressiveness of the dictionary $\mathbf{D}$ (which increases with its size) and the complexity (which also increases when $\mathbf{D}$ is larger). An alternative is uniform random sampling of the training data as in [52–55], but this approach may result in a biased estimation of the characteristics of the involved classes. Our multi-pass scheme does not suffer from this problem.

IV. Kernel STD method for paint loss detection

In order to apply the developed general multimodal target detection method to paint loss detection, we introduce an extension to be able to distinguish paint loss from crack patterns. Cracks are treated differently than paint loss by restorers. Typically, paint loss will be documented and repaired but cracks, as an ageing sign of paintings, will be preserved without any treatment. Fig. 5 shows examples of crack and paint loss in old master paintings. Observe that they appear very similar. When applying directly the developed target detection method, these cracks will be detected as target rather than left in the background. To differentiate cracks from areas of paint loss, we can either perform post-processing with crack detection methods [4–7] or assign crack pixels to a particular class. We opt for a unified framework based on a sparse representation rather than relying on postprocessing. Therefore, we need to extend the previously described approach to account for an extra class.

Here we introduce a crack compensation module. Next to the dictionary $\mathbf{D}_0$ (background) and $\mathbf{D}_1$ (paint loss) in (16), we now construct an additional sub-dictionary $\mathbf{D}_2$ for cracks, resulting in a new dictionary $\mathbf{D} = [\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2]$. Each sub-dictionary is constructed by the corresponding training sam-

Algorithm 1 Main algorithmic steps for paint loss detection

1: **Input**: An input matrix $\mathbf{X} \in \mathbb{R}^{d \times MN}$ containing a set of samples$\{\mathbf{x}_i\}_{i=1}^{MN}$ in the fused data from multiple imaging modalities; dictionary $\mathbf{D}$ and K_0.

2: **For** *j*=1:*MN*

3: Calculate sparse coefficients $\boldsymbol{\alpha}_j$ for the test sample $\mathbf{x}_j$ by solving (26)

4: **end**

5: Obtain $\mathbf{Z}$ by (29) and $\mathbf{R}^{'}$ by (18) and (27)

6: Calculate $\bar{\mathbf{R}}^{'}$ by (20) with the input $\mathbf{R}^{'}$ and obtain $\mathbf{M}^{'}$ by (24).

7: Obtain final detection map, $\mathbf{M}$, by (30).

8: **Output**: Detection map $\mathbf{M}$.

ples. The corresponding sparse coefficients $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0; \boldsymbol{\alpha}_1; \boldsymbol{\alpha}_2]$ are obtained by solving the following optimization problem:

$$\arg\min_{\boldsymbol{\alpha}} \|\phi(\mathbf{x}) - \phi(\mathbf{D})\boldsymbol{\alpha}\|^2 \quad s.t. \ \|\boldsymbol{\alpha}\|_0 < K_0. \tag{26}$$

The two sub-dictionaries $\mathbf{D}_0$ and $\mathbf{D}_2$ form together a new background dictionary, i.e., the classes other than the desired target. The corresponding residual error $r_0'(\phi(\mathbf{x}))$ is

$$\begin{aligned}
\bar{r}_0'(\phi(\mathbf{x})) &= \|\phi(\mathbf{x}) - \phi(\mathbf{D}_0)\boldsymbol{\alpha}_0 - \phi(\mathbf{D}_2)\boldsymbol{\alpha}_2\|_2 \\
&= \langle \phi(\mathbf{x}) - \phi(\mathbf{D}_0)\boldsymbol{\alpha}_0 - \phi(\mathbf{D}_2)\boldsymbol{\alpha}_2, \phi(\mathbf{x}) \\
&\quad - \phi(\mathbf{D}_0)\boldsymbol{\alpha}_0 - \phi(\mathbf{D}_2)\boldsymbol{\alpha}_2 \rangle^{1/2} \\
&= (\boldsymbol{\kappa}(\mathbf{x},\mathbf{x}) - 2\boldsymbol{\alpha}_0^T\mathbf{K}_{\mathbf{D_0}} - 2\boldsymbol{\alpha}_2^T\mathbf{K}_{\mathbf{D}_2} \\
&\quad + 2\boldsymbol{\alpha}_0^T\mathbf{K}_{\mathbf{D}_0\mathbf{D}_2}\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_0^T\mathbf{K}_{\mathbf{D}_0\mathbf{D}_0}\boldsymbol{\alpha}_0 \\
&\quad + \boldsymbol{\alpha}_2^T\mathbf{K}_{\mathbf{D}_2\mathbf{D}_2}\boldsymbol{\alpha}_2)^{1/2}
\end{aligned} \tag{27}$$

The residual error $r_1(\phi(\mathbf{x}))$ with respect to dictionary $\mathbf{D}_1$ is calculated as in (18) with $i = 1$. $\mathbf{R}^{'} \in \mathbb{R}^{M \times N \times 2}$ denotes the residual cube where $\mathbf{R}^{'}(i,j,1) = r_0'(\phi(\mathbf{x}_{i,j}))$ and $\mathbf{R}^{'}(i,j,2) = r_1(\phi(\mathbf{x}_{i,j}))$ and $\bar{\mathbf{R}}^{'}$ the smoothed version of $\mathbf{R}^{'}$. The corresponding detector becomes:

$$T(\mathbf{x}) = \bar{r}_0'(\phi(\mathbf{x})) - \bar{r}_1(\phi(\mathbf{x})). \tag{28}$$

Let $\mathbf{M}' \in \mathbb{R}^{M \times N}$ denote a paint loss map obtained by applying to each pixel the classification rule (24) with $T(\mathbf{x})$ from (28). Due to the smoothing operation that was applied to the residual cube, certain cracks crossing the paint loss regions may be labelled as paint loss. Therefore we refine the paint loss map using a binary crack map $\mathbf{Z} \in \mathbb{R}^{M \times N}$ obtained by

$$Z_{i,j} = \begin{cases} 1 & \arg\min\limits_{k=0,1,2} r_k(\phi(\mathbf{x}_{i,j})) = 2 \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

The final paint loss map is

$$\mathbf{M} = \mathbf{M}^{'} \odot \bar{\mathbf{Z}}, \tag{30}$$

where $\bar{\mathbf{Z}}$ is the complement of $\mathbf{Z}$ and $\odot$ is an element-wise multiplication. The proposed paint loss detection method is summarized in Algorithm 1.

V. Experimental Results and Analysis

In this section, we evaluate the proposed method in a real application, detecting paint loss in the *Ghent Altarpiece* during its ongoing restoration treatment.

A. Data set and experiment setting

The *Ghent Altarpiece*, also known as *The Adoration of the Mystic Lamb*, is by art scholars considered as one of the most influential masterpieces in the history of art [56]. This monumental polyptych on wood panels was painted by the brothers Hubert and Jan van Eyck in 1432. Digitized images of the *Ghent Altarpiece* are publicly available in very high resolution on the website *Closer to Van Eyck: Rediscovering the Ghent Altarpiece*[1], which is the result of an interdisciplinary research project, with the goal to investigate the structural condition of the *Ghent Altarpiece* and determine whether a full restoration of Van Eyck's polyptych was necessary [17]. During the following conservation campaign, several multimodal acquisitions at different stages of the project were captured to document the surface of the altarpiece, including macrophotography and infrared macrophotography, both at a resolution of 7.2 μm, infrared reflectography and X-radiography. Several images were selected to test the performance of our method, and all annotations on the images used in our experiments were made by painting conservators currently treating the *Ghent Altarpiece*.

We use the following methods as reference for comparison: MSD, ACE (a special case of ASD), SVM with the RBF kernel, STD, SRC and two recent methods: multimodal feature learning (MFL) [41] and matched subspace classifier (MSC) [34]. MFL learns discriminative features from multimodal data and feeds the learned features into an SVM classifier to identify the target. MSC learns separately a background dictionary and a target dictionary with the K-SVD algorithm to model the associated background and target subspaces, and formulates a sparsity-based detector based on the resulting dictionaries. For paint loss detection, we set the window size to 3×3 and δ to zero in our proposed method. These parameter values were optimized experimentally and their choice does not appear critical in a rather wide range (see Fig. 9). We utilize the macrophotography during treatment as the reference image $\mathbf{L}$ in (21) and (22), and set $\lambda = 0.4, a = 0.9$ and $\varepsilon = 0.0001$ empirically. For STD, we report the optimal results by tuning δ. When $\delta = 0$, STD reduces to SRC. We thus refer to our method with kernel function and spatial filtering as SRC-KF.

We adopt the overall accuracy (OA) and the Kappa coefficient (κ) as the quantitative measures, and we also report the Producer's and User's Accuracies. Let $n_{i,j}$ be the number of samples in class i that are labelled as class j by the detector. Producer's Accuracy and User's Accuracy for the i-th class are then $p_i = n_{i,i}/n_{i,+}$ and $u_i = n_{i,i}/n_{+,i}$, respectively. $n_{i,+} = \sum_j n_{i,j}$ is the number of samples in class i and $n_{+,i} = \sum_j n_{j,i}$ is the number of samples that are identified as class i. Producer's Accuracy indicates the class-specific

[1] http://closertovaneyck.kikirpa.be/

TABLE I
RESULTS OF DIFFERENT METHODS ON THE IMAGE PATCH OF *John the Evangelist*.

Methods		MSD	ACE	STD	SVM	SRC	MFL	MSC	SRC-KF
Producer's Accuracy	P	48.10	37.98	74.16	62.81	76.78	66.21	56.11	**82.05**
	N	83.49	91.15	89.00	**93.86**	87.39	92.04	93.70	92.61
User's Accuracy	P	46.53	56.65	66.77	**81.10**	64.45	74.18	72.60	76.84
	N	84.43	83.18	92.07	85.64	92.69	88.57	87.84	**94.56**
OA		75.39	78.98	85.60	84.60	84.96	85.27	85.10	**90.19**
κ		0.3119	0.3297	0.6078	0.6051	0.6012	0.6016	0.5391	**0.7289**

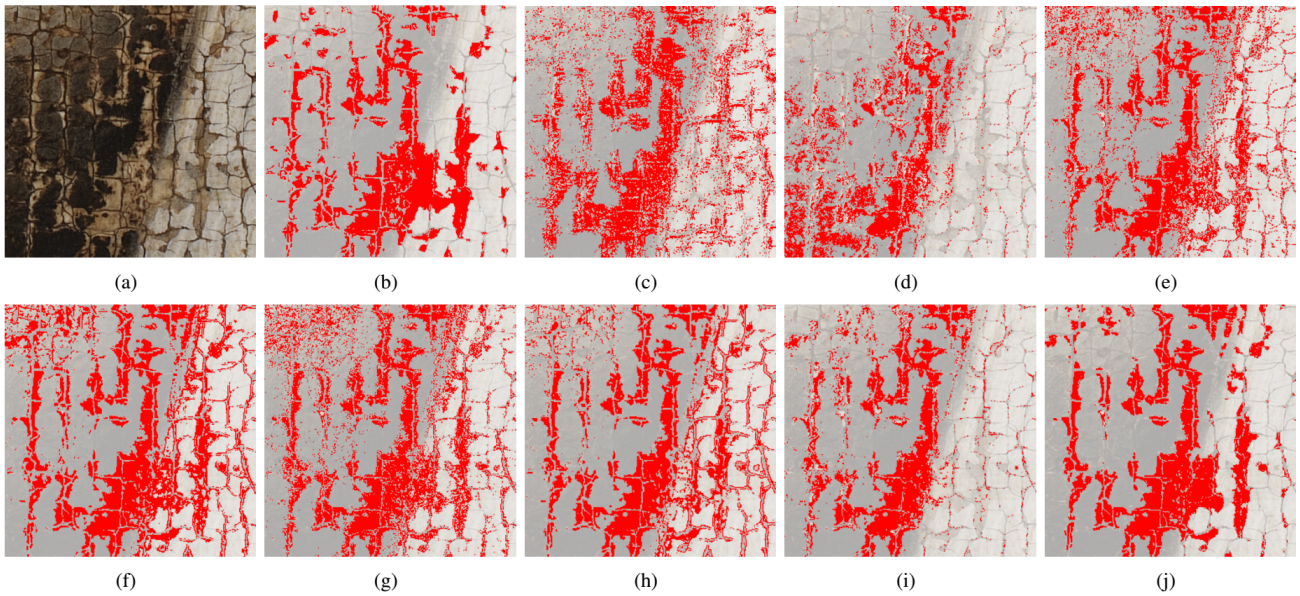


Fig. 6. Selected image patch from *John the Evangelist* in the *Ghent Altarpiece*. (a) Macrophotography after cleaning (b) Ground truth and detection maps of different methods obtained by (c) MSD, (d) ACE, (e) STD, (f) SVM, (g) SRC, (h) MFL, (i) MSC and (j) SRC-KF. Paint loss is marked in red. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

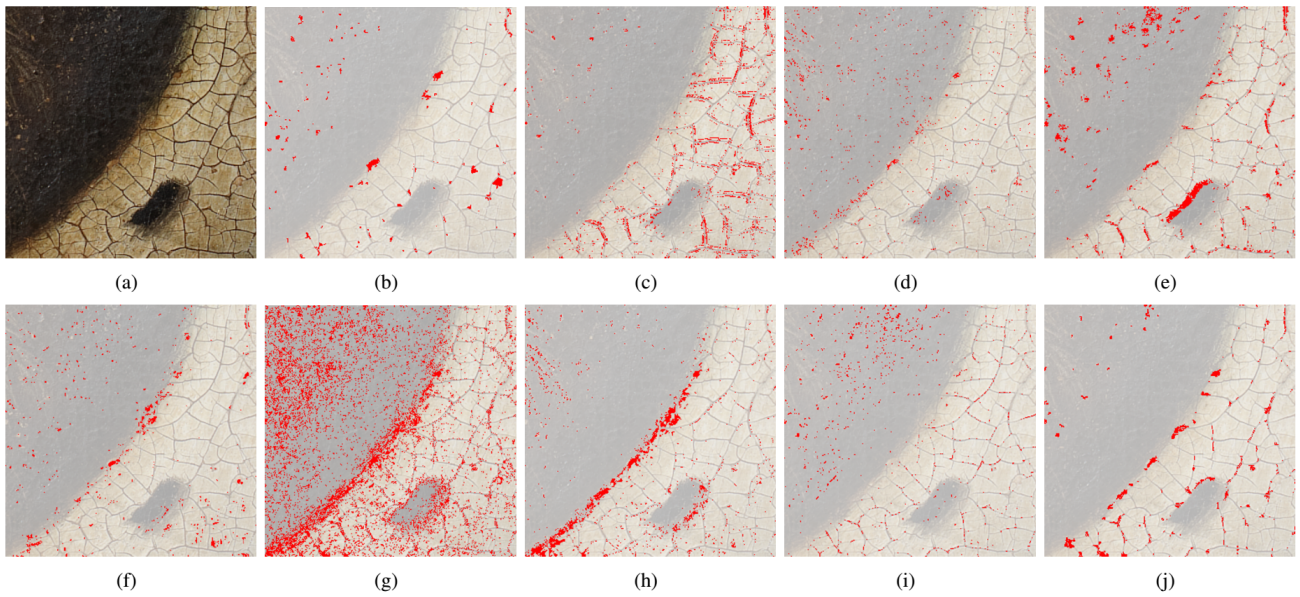


Fig. 7. Selected image patch from the panel *Prophet Zacharias* in the *Ghent Altarpiece*. (a) Macrophotography after cleaning (b) Ground truth and detection maps of different methods obtained by (c) MSD, (d) ACE, (e) STD, (f) SVM, (g) SRC, (h) MFL, (i) MSC and (j) SRC-KF. Paint loss is marked in red. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

TABLE II
RESULTS OF DIFFERENT METHODS ON THE IMAGE PATCH OF *Prophet Zacharias*.

Methods		MSD	ACE	STD	SVM	SRC	MFL	MSC	SRC-KF
Producer's Accuracy	P	12.83	19.19	25.61	7.85	**55.32**	7.37	32.89	46.19
	N	97.56	99.03	97.84	99.10	89.72	**99.28**	96.47	97.57
User's Accuracy	P	5.96	19.13	13.06	30.91	6.12	**42.71**	11.50	19.29
	N	98.95	99.04	99.10	91.49	**99.41**	93.15	99.18	99.35
OA		96.56	**98.09**	96.99	90.78	89.32	92.56	95.72	96.96
κ		0.0657	0.1818	0.1564	0.0975	0.0908	0.1034	0.1484	**0.2563**

accuracies and User's Accuracy reveals what percentage of the samples that are identified as one class is correctly identified as that class. OA and κ are computed as $\text{OA} = \frac{1}{n_t}\sum_i n_{i,i}$ and $\kappa = (\frac{1}{n_t}\sum_i n_{i,i} - \frac{1}{n_t^2}\sum_i n_{i,+}n_{+,i})/(1 - \frac{1}{n_t^2}\sum_i n_{i,+}n_{+,i})$, where $n_t = \sum_{i,j} n_{i,j}$ is the total number of test samples. Unless it is explicitly stated otherwise, all the reported results represent the average performance of ten runs. The training samples are randomly selected from the regions annotated by experts. We also report the results of digital inpainting starting from the different detection masks. The inpainting results allow additional insights into the actual quality of the detected paint loss maps.

B. Paint loss detection results

1) Experiment 1: An image patch of size 300×300 was selected from the panel *John the Evangelist*. We use three imaging acquisitions in this experiment: infrared macrophotography and macrophotography before and after cleaning. Fig. 6 (a) shows the macrophotography after cleaning and Fig. 6 (b) shows a manual expert labeling of the paint loss, which we refer to as the ground truth. The size of the paint loss areas as well as their color intensity vary significantly. The number of training samples for each class is set as 80. Table I reports the comparative results of different methods, where "P" means positive, i.e., paint loss, and "N" means negative, i.e., background. Visual results are shown in Fig. 6.

The proposed method SRC-KF yields the best performance in terms of OA and κ. The improvement in OA is around 4.6% relative to STD and more than 5.2% relative to SRC. MFL employs the representation coefficients as the input to SVM and shows an improved OA compared to using the original data. Although MSC adopts dictionary learning strategy to formulate the background and target subspaces, it performs comparable to SRC and STD, both of which use labeled training samples to construct the dictionary. It can be observed that the detection map of SRC-KF in Fig. 6 (j) contains less noise-like erroneous detections compared to the other methods and lies visually much closer to the ground truth in Fig. 6 (b).

2) Experiment 2: We select another image patch of size 300×300 from the panel *Prophet Zacharias* with the same imaging acquisitions as in Experiment 1. The macrophotography after cleaning, shown in Fig. 7, is more challenging than in the first experiment as the color intensity distribution of paint loss overlaps substantially with that of some painted areas. We set the number of training samples to 40 per class. The optimal results for each method are reported in Table II.

The proposed method SRC-KF achieves the best performance in terms of κ. ACE yields the best OA but the corresponding κ value is much lower compared to SRC-KF. In this experiment, since the area occupied by paint loss is relatively small compared to the background, the OA measure fails to provide a fair quantitative evaluation: even when all the paint loss is labelled as background, the OA can still be high. But the value of κ becomes zero. Compared to OA, the κ evaluation is more relevant in this case. The results in Table II show that SRC-KF yields a significant improvement over SRC in terms of κ. The detection maps of SVM, SRC and MFL in Fig. 7 (e)-(h) display large number of false positives, while MSC (Fig. 7 (i)) fails to detect even the largest paint losses, which demonstrates the difficulty of our task on this image. The proposed method SRC-KF still detects most of the paint losses that were indicated by the expert and the false detections align mostly with widened cracks, which also present the areas of missing paint.

3) Experiment 3: We select a 300×300 patch from the central panel *Adoration of the Mystic Lamb* with the same imaging acquisitions as in the previous two experiments. Fig. 8 (a) and (b) show the macrophotography during treatment and expert annotation of the paint loss. This dataset is very challenging as it shows very high similarities between target and some background areas in terms of color. The number of training samples is set to 40 per class.

We report the results in Table III and Fig. 8. Clearly, our method outperforms all the reference ones both in terms of OA and κ. Observe that STD, MSC and SRC-KF, which are based on sparse representation, perform better than the other methods. Although SRC can be viewed as a special case of STD with $\delta = 0$, its performance here is much worse, which can be attributed to mixed target and background subspaces. Our method, even with this fixed threshold δ, yields the result that adheres best with the ground truth in Fig. 8 (b) and is much less contaminated by sparse erroneous detections compared to all other methods.

4) Experiment 4: We train the methods on one panel and apply them to detect paint losses in another panel. In particular, we use 40 training samples per class from the dataset *Prophet Zacharias* from Experiment 2 and we run the detections on the dataset *John the Evangelist* from Experiment 1. Table IV shows the detection results. Our method yields the best performance both in terms of OA and κ. All the methods that employ sparse representation: STD, SRC, MFL, MSC and SRC-KF show better results compared to SVM. It is also observed that all the performances drop as expected

TABLE III
RESULTS OF DIFFERENT METHODS ON THE IMAGE PATCH OF THE *Adoration of the Mystic Lamb*.

Methods		MSD	ACE	STD	SVM	SRC	MFL	MSC	SRC-KF
Producer's Accuracy	P	48	38.09	58.52	45.36	70.17	50.16	50.65	**79.32**
	N	81.53	89.45	88.23	**91.81**	75.66	90.91	88.40	86.61
User's Accuracy	P	41.76	48.45	55.95	**72.65**	42.43	67.44	52.61	60.65
	N	86.04	85.13	89.33	77.73	90.88	82.58	87.57	**94.31**
OA		74.74	79.04	82.21	76.70	74.55	79.52	80.75	**85.14**
κ		0.2796	0.2944	0.4594	0.4113	0.3687	0.4431	0.3961	**0.5905**

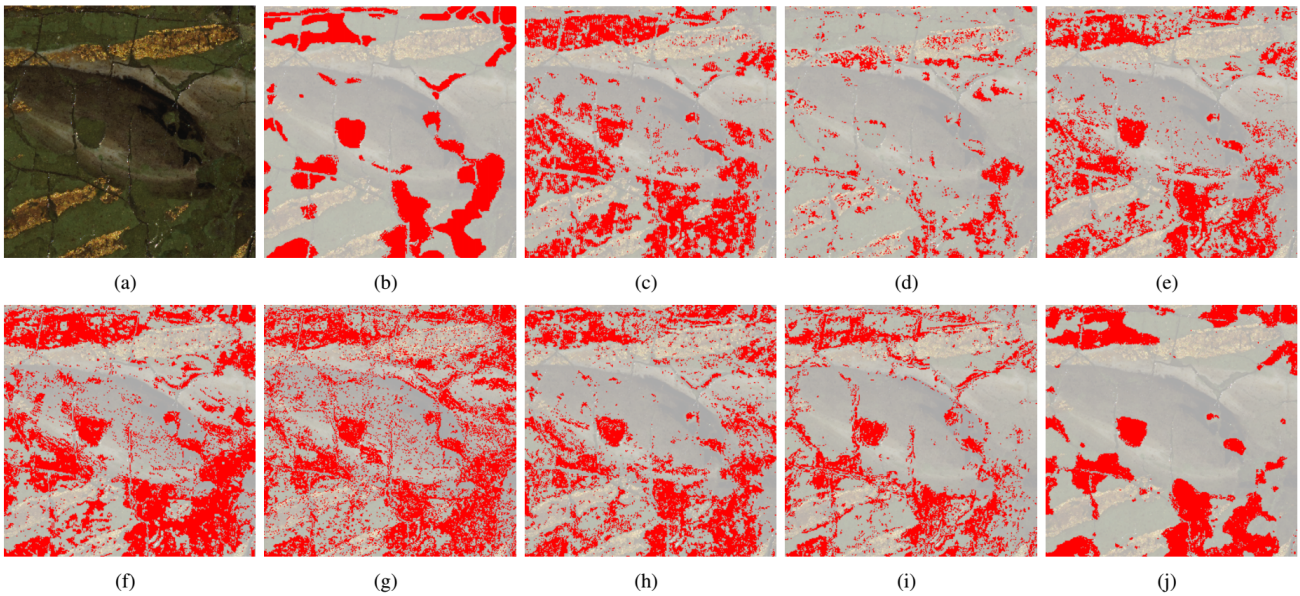


Fig. 8. Paint loss detection on a part of the central panel the *Adoration of the Mystic Lamb* in the *Ghent Altarpiece*. (a) Macrophotography after cleaning (b) Ground truth, and detection maps of (c) MSD, (d) ACE, (e) STD, (f) SVM, (g) SRC, (h) MFL, (i) MSC and (j) SRC-KF. Paint loss is marked in red. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

TABLE IV
RESULTS BY USING THE TRAINING SAMPLES FROM *Prophet Zacharias* TO DETECT PAINT LOSS IN *John the Evangelist.*

Methods	MSD	ACE	STD	SVM	SRC	MFL	MSC	SRC-KF
OA	72.79	68.83	81.62	63.41	81.36	80.33	81.52	**85.43**
κ	0.2188	0.0716	0.4569	0.067	0.3851	0.3461	0.3737	**0.5058**

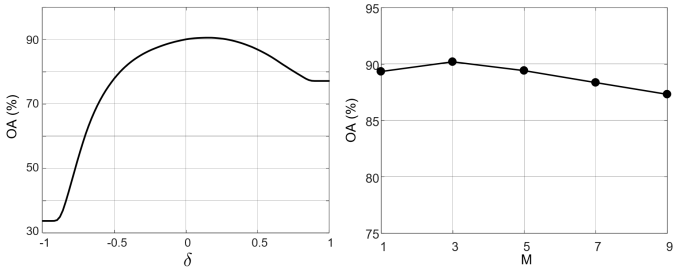


Fig. 9. The influence of the parameters δ (a) and w (b) on the overall accuracy for the dataset *John the Evangelist*.

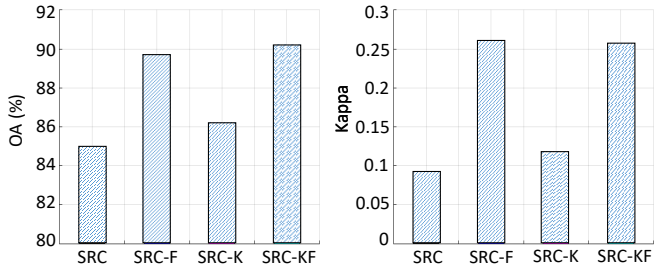


Fig. 10. Contributions of different parts in the proposed method on the OA (left) in *John the Evangelist* and κ (right) in *Prophet Zacharias*.

compared to when using the training data from the input image in Experiment 1. The performance decrease of sparsity-based detectors is much less than ACE and SVM.

C. Parameter study

We investigate empirically the influence of the threshold δ and the window size w on the detection performance. The experiments are conducted on the dataset *John the Evangelist*.

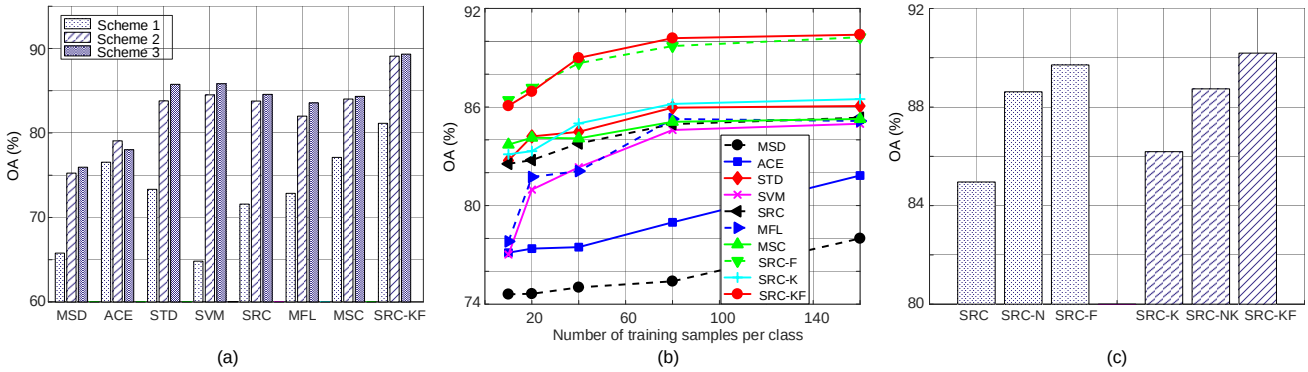


Fig. 11. (a) Overall accuracy with different sets of imaging modalities. (b) Effect of the training sample size on the overall accuracy. (c) Effect of the crack compensation on the overall accuracy.

We vary δ in the range -1 to 1 with a step size of 0.02, and w within $\{1, 3, 5, 7, 9\}$. The results are reported in Fig. 9. It can be seen that the values $0 \leq \delta \leq 0.03$ yield the best performance, which is also very stable in this whole range. The performance drops significantly only when δ exceeds 0.5 or when it becomes negative (for $\delta < -0.5$ the performance decreases severely). Setting $\delta = 0$ yields nearly optimal and stable detection performance. Regarding the window size, the performance is rather stable and $w = 3$ yields the highest OA. Incorporating spatial context from a relatively small window ($w = 3$) improves the detection accuracy compared to using the test pixel alone ($w = 1$). However, when w becomes too big, the performance deteriorates. For the painting scans in our dataset, $w = 3$ is the best choice.

D. The effect of spatial filtering

To investigate the effect of spatial filtering, we conduct a comparative evaluation of SRC, SRC-F, SRC-K and SRC-KF on two datasets. SRC-F and SRC-K are reduced versions of the proposed method: SRC-F employs spatial filtering but no kernel projection, and SRC-K incorporates the kernel projection but without spatial filtering. The results are shown in Fig. 10 in terms of OA and κ measures. Both of these indicate clearly the importance of spatial filtering. Specifically, in the analysed datasets, the use of spatial filtering led to more than 4% improvement in OA, and with kernel versions this improvement was nearly 5%. A similar conclusion can be drawn when inspecting the κ values in the right of Fig. 10. The use of spatial filtering doubles the kappa coefficient. This proves the huge benefit of the proposed spatial smoothing of the residuals in the sparse representation.

E. The effect of multiple imaging acquisitions

Here, we study the effect of using different sets of imaging acquisitions on the detection performance. The empirical analysis is conducted on part of *John the Evangelist* described under Experiment 1. Three schemes are described in Table V and their corresponding performance in terms of overall accuracy is reported in Fig. 11 (a). The number of training samples is set to 40 per class. The results show that for almost all the methods, the performance is improved with an increasing number of imaging acquisitions, except for ACE. This demonstrates clearly the benefit of using multiple imaging acquisitions/modalities in this task. The proposed SRC-KF consistently achieves the best performance in all the three schemes.

F. The effect of training sample size

To examine the effect of training sample size on the detection performance, we vary the number of training samples from 10 to 160 per class. Fig. 11 (b) reports the OA of the detection on the dataset *John the Evangelist*. Generally, the performance improves with larger training sample sizes for all the methods. SRC-F consistently achieves better results than SRC with an average improvement of 4%, which can also be observed for the kernel version. For most of the methods, the OA increases rapidly with the number of training samples up to about 80 samples per class, while for larger sample sizes the slope is less steep. This is different for MSD and ACE, where OA first increases gradually and then rises quickly with larger sample sizes.

G. The effect of crack compensation

This experiment investigates the contribution of the crack compensation strategy on the paint loss detection performance. The same dataset as in Experiment 1 is used. We denote by SRC-N the method which incorporates the spatial information by using a smoothing filter but without crack compensation, and by SRC-NK its kernel version. Table VI gives the taxonomy of all the analysed variants of the method with the corresponding abbreviations. The results in Fig. 11 (c) reveal that the improvement of SRC-KF over SRC follows from all the three key components: kernel projection, spatial smoothing and crack compensation strategy. While the biggest improvement in OA originated from using the spatial filtering (2.6%), a considerable improvement (1.5%) arose from the crack compensation. The crack compensation improved also the performance without kernel projection (for about 1.1%). This illustrates the effectiveness of the crack compensation in the modified model.

TABLE V
EXPERIMENTAL SETTINGS WITH DIFFERENT SETS OF IMAGING ACQUISITIONS.

Schemes	Number of acquisitions	Imaging acquisitions
1	1	macrophotography after cleaning;
2	3	infrared macrophotography and macrophotography before cleaning, macrophotography after cleaning;
3	5	infrared macrophotography, infrared reflectography, macro and X-radiography photography before cleaning, macrophotography after cleaning.

TABLE VI
SPECIFICATION OF DIFFERENT METHODS.

Methods	SRC	SRC-N	SRC-F	SRC-K	SRC-NK	SRC-KF
Smoothing filter	No	Yes	Yes	No	Yes	Yes
Kernel space	No	No	No	Yes	Yes	Yes
Crack compensation	No	No	Yes	No	No	Yes

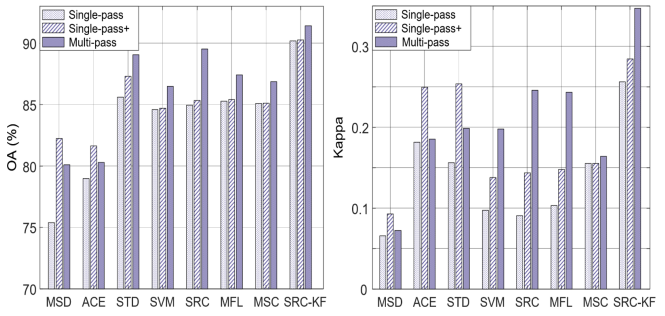


Fig. 12. Comparison between the results by "Single-pass", "Single-pass+" and "Multi-pass" in the data *John the Evangelist* (left) and *Prophet Zacharias* (right).

H. The effect of majority voting

Annotating manually paint loss in very-high resolution images of paintings at pixel precision is very difficult. Even for experts, some areas are ambiguous due to their color, low contrast, tiny cracks or scanning artefacts. In order to alleviate the adverse effects of improper annotations, which are inevitable, we introduced in Section III-C a multi-pass strategy with majority voting instead of the classic single detection approach. Here we evaluate this multi-pass strategy on parts of the panels *John the Evangelist* and *Prophet Zacharias* (multimodal datasets described in Section V-B). To avoid a biased performance caused by the random sampling, the averaged result over 10 simulations, denoted as "Single-pass", is compared to the result of "Multi-pass" for all the methods. In addition, we conduct the experiments with the "Single-pass+" scheme that employs at once all the training samples that are used by the "Multi-pass" scheme. The number of training samples is kept the same as in Section V-B. The results reported in Fig. 12 show that Single-pass+ always outperforms the Single-pass scheme regardless of the particular detection method. For most of the analysed methods, the multi-pass scheme still yields better performance, while it is also more efficient in terms of space complexity and avoids computational bottlenecks. For our approach, the multi-pass scheme yields a clear improvement over both Single-pass and Single-pass+ schemes on all the tested data sets.

I. Paint loss detection and inpainting in practice

In practice, when processing large images of paintings, we need to deal with scarce annotations from different places in the painting that a user labels at randomly picked locations. In this experiment, we are detecting paint loss in a larger region from the panel *Prophet Zacharias* by using the training data from two small image patches from other locations in that panel. The size of the test image is 869×667. Three imaging acquisitions listed as Scheme 2 in Table V are used. The macrophotography during treatment is shown in Fig. 13 (a). We use 40 training samples per class. The detection results are shown in Fig. 13. Clearly, SRC suffers from impulse noise in the detection map and also many cracks are incorrectly labelled as paint loss. These adverse effects are much less pronounced in our result.

We also evaluate the detection results by running digital inpainting with different detection maps and comparing the results with the physical restoration. Fig. 13 shows the inpainting results obtained with the method [57], for the detection maps of SRC and SRC-KF, in comparison with the physical treatment by restorers. Clearly, the inpainting result on the SRC mask (Fig. 13 (d)) shows excessive smoothing on the face and beard, and loss of the detailed textures of the painting. This is due to excessive false detections in the SRC detection mask which are filled by other image patches during the inpainting. This is also the reason why the majority of cracks are missing. The inpainting on the SRC-KF mask, shown in Fig. 13 (e), preserved the textures and crack pattern nicely, in close resemblance to the physical treatment by restorers. The color-tone difference is due to the different imaging conditions. The results in this experiment show the importance of accurate paint loss detection for virtual restoration of paintings.

J. Generalization analysis

We also evaluate our method in a different detection task on a public hyperspectral image in the field of remote sensing. The dataset known as Hyperspectral Digital Imagery Collection Experiment (HYDICE) *Urban* image was captured by the HYDICE sensor during a flight campaign over Copperas Cove, near Fort Hood, TX, USA. It has a spatial size of 307×307 and contains 210 bands. After removing the bands 1-4, 76, 87, 101-111, 136-153 and 198-210, which are seriously polluted by the

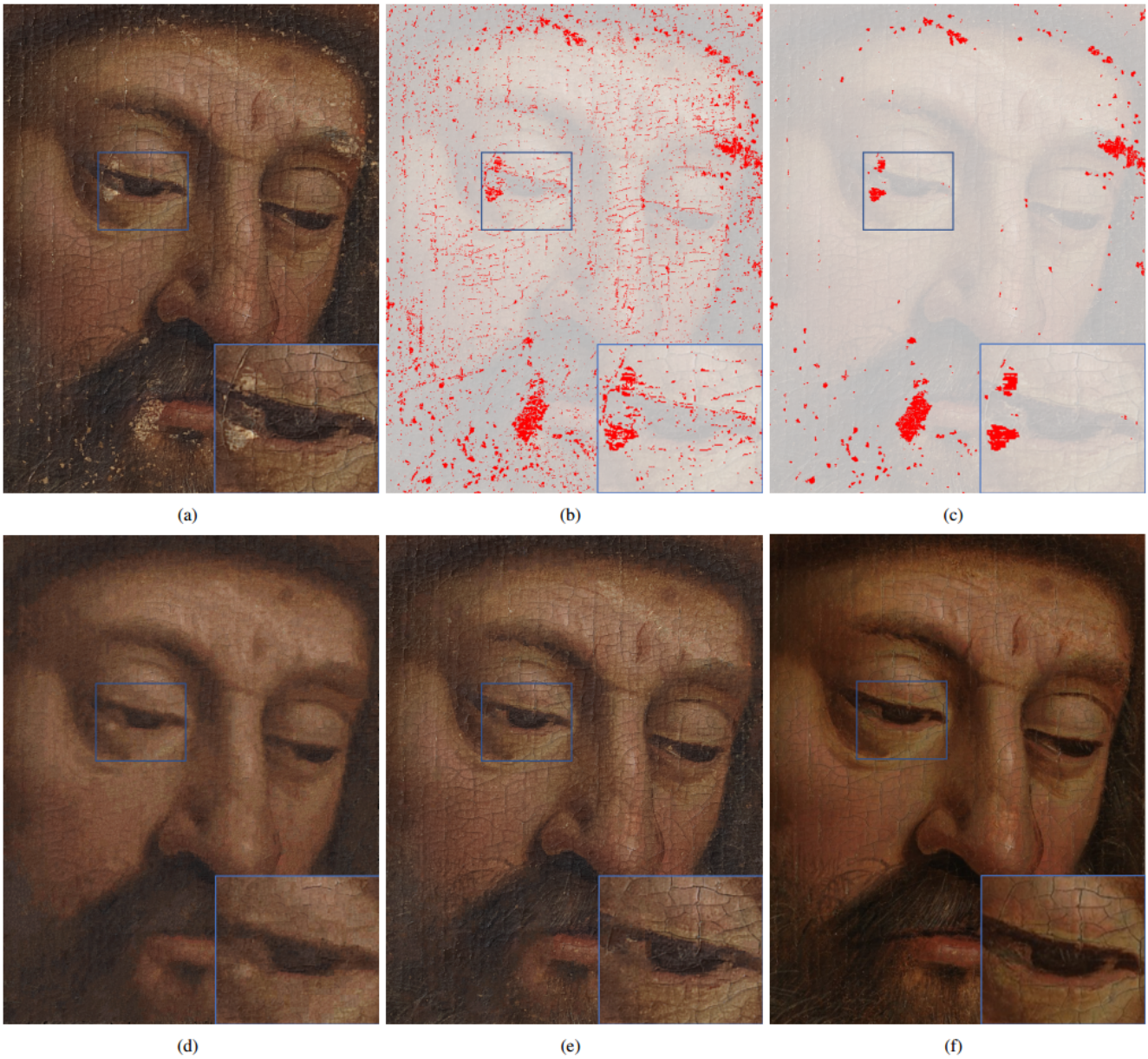


Fig. 13. (a) Input image from the panel *Prophet Zacharias* in the *Ghent Altarpiece*; detection maps obtained by (b) SRC and (c) SRC-KF; inpainting results (d) with SRC detection mask and (e) with SRC-KF detection mask; (f) Physical treatment by restorers. The right bottom corner in all the figures shows the details in the blue box. Image copyright Ghent, Kathedrale Kerkfabriek; photo courtesy of KIK-IRPA, Brussels.

TABLE VII
RESULTS OF DIFFERENT METHODS ON *HYDICE Urban* IMAGE.

Methods		MSD	ACE	STD	SVM	SRC	MFL	MSC	SRC-KF
Producer's Accuracy	P	66.73	54.41	80.57	74.09	73.91	77.16	77.78	**89.42**
	N	92.37	90.79	98.65	**99.30**	99.18	98.04	98.55	98.79
User's Accuracy	P	55.85	46.87	92.54	**96.26**	95.59	89.24	92.08	93.24
	N	94.66	92.98	95.90	93.70	93.83	95.24	95.15	**97.80**
OA		88.81	86.03	95.39	94.08	94.10	94.34	94.68	**97.11**
κ		0.5369	0.4225	0.8321	0.7993	0.7971	0.7928	0.8096	**0.8927**

atmosphere and water absorption, the remaining 162 bands are used in the experiments. Fig. 14 (a) shows in a false-color representation a part of this image with a size $150 \times 160 \times 162$ that we used as the test data. It contains seven distinct land-cover classes: "roof", "parking lot", "grass", "trees", "sparse vegetation", "asphalt road" and "concrete road" [58], with its spatial arrangement (ground truth detection map) shown in Fig. 14 (b). In this study, we focus on the detection of asphalt

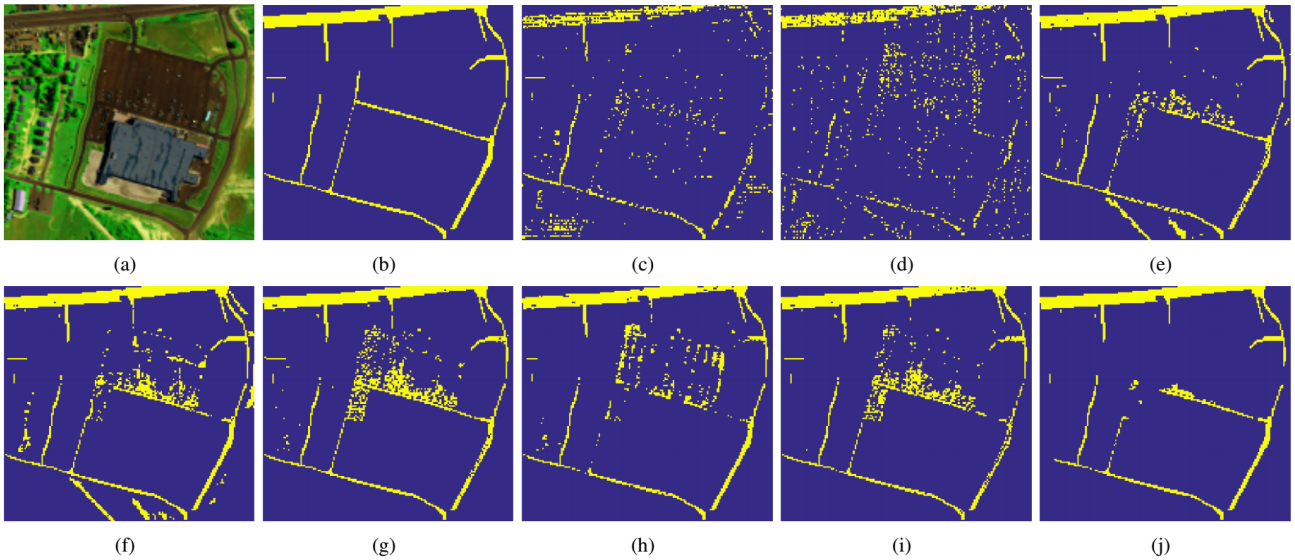


Fig. 14. Detection results on the *HYDICE Urban* image. (a) False color image (b) Ground truth of asphalt road and detection maps obtained by (c) MSD, (d) ACE, (e) STD, (f) SVM, (g) SRC, (h) MFL, (i) MSC and (j) SRC-KF.

road. Due to the high dimensionality of this data, we set $w = 1$ and directly utilize spectral signatures as the input for all the methods. The number of target and background training samples is set to 40. The results are reported in Table VII and Fig. 14. Our method yields the best detection performance in terms of the quantitative performance measures and visual assessment. MSD and ACE obtain worse results compared to others, both quantitatively and visually. Fig. 14 shows that all other reference methods also identified wrongly parts of the background as the target. The detection result of our method agrees best with the ground truth detection map, showing the effectiveness of the proposed approach in this detection task.

VI. Conclusion

In this work, we address the challenging problem of paint loss detection in digitized paintings. We first propose a generic multimodal target detection method based on sparse representation, where spatial features of various imaging modalities are fused in a kernel feature space by means of a kernel function. The spatial context is further exploited by applying a smoothing filter on the representation residuals. In addition, a majority voting strategy is introduced that overcomes the unavoidable problem of imperfect annotations. We tailor this generic target detection method to the problem of paint loss detection in paintings and we introduce a crack compensation strategy to meet the requirements posed by the conservation practice. Overall, this improves both the detection accuracy and inpainting performance. Experimental results on images acquired during the ongoing restoration-conservation treatment of the *Ghent Altarpiece* indicate the effectiveness of the proposed method, and its potential to support the conservation practice. A generalization study, consisting of target detection on remote sensing data, demonstrates the broad applicability of the proposed method.

References

[1] P. Abry, A. G. Klein, W. A. Sethares, and C. R. Johnson, "Signal processing for art investigation," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 14–16, July 2015.

[2] 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, July 2008.

[3] L. Platiša, B. Cornelis, T. Ružic, A. Pižurica, A. Dooms, M. Martens, M. D. Mey, and I. Daubechies, "Spatiogram features to characterize pearls in paintings," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2011, pp. 801–804.

[4] I. Giakoumis, N. Nikolaidis, and I. Pitas, "Digital image processing techniques for the detection and removal of cracks in digitized paintings," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 178–188, 2006.

[5] B. Cornelis, T. Ružic, E. Gezels, A. Dooms, A. Pižurica, L. Platiša, J. Cornelis, M. Martens, M. D. Mey, and I. Daubechies, "Crack detection and inpainting for virtual restoration of paintings: The case of the ghent altarpiece," *Signal Process.*, vol. 93, no. 3, pp. 605–619, 2013.

[6] B. Cornelis, Y. Yang, J. T. Vogelstein, A. Dooms, I. Daubechies, and D. Dunson, "Bayesian crack detection in ultra high resolution multimodal images of paintings," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, 2013, pp. 1–8.

[7] S. Kaur and A. Kaur, "Restoration of historical wall paintings using improved nearest neighbour algorithm," *Int. J. Eng. Comput. Sci.*, vol. 3, no. 12, 2014.

[8] P. Abry, S. G. Roux, H. Wendt, P. Messier, A. G. Klein, N. Tremblay, P. Borgnat, S. Jaffard, B. Vedel, J. Coddington, and L. A. Daffner, "Multiscale anisotropic texture analysis and classification of photographic prints: Art scholarship meets image processing algorithms," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 18–27, 2015.

[9] H. Anwar, S. Zambanini, M. Kampel, and K. Vondrovec, "Ancient coin classification using reverse motif recognition: image-based classification of Roman Republican coins," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 64–74, 2015.

[10] L. van der Maaten and R. G. 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, 2015.

[11] R. Srinivasan, C. Rudolph, and A. K. Roy-Chowdhury, "Computerized face recognition in renaissance portrait art: A quantitative measure for identifying uncertain subjects in ancient portraits," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 85–94, 2015.

[12] D. Picard, P. Gosselin, and M. Gaspard, "Challenges in content-based image indexing of cultural heritage collections," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 95–102, 2015.

[13] F. Lozes, A. Elmoataz, and O. Lézoray, "PDE-based graph signal processing for 3-d color point clouds: Opportunities for cultural heritage," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 103–111, 2015.

[14] B. Cornelis, H. Yang, A. Goodfriend, N. Ocon, J. Lu, and I. Daubechies, "Removal of canvas patterns in digital acquisitions of paintings," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 160–171, 2017.

[15] N. Deligiannis, J. F. C. Mota, B. Cornelis, M. R. D. Rodrigues, and I. Daubechies, "Multi-modal dictionary learning for image separation with application in art investigation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 751–764, 2017.

[16] T. Ružic, B. Cornelis, L. Platiša, A. Pižurica, A. Dooms, W. Philips, M. Martens, M. D. Mey, and I. Daubechies, "Virtual restoration of the Ghent Altarpiece using crack detection and inpainting," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2011, pp. 417–428.

[17] A. Pižurica, L. Platiša, T. Ružic, B. Cornelis, A. Dooms, M. Martens, H. Dubois, B. Devolder, M. D. 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.

[18] S. Huang, W. Liao, H. Zhang, and A. Pižurica, "Paint loss detection in old paintings by sparse representation classification," in *Proc. International Traveling Workshop on Interactions Between Sparse Models and Technology (ITWIST)*, 2016, pp. 62–64.

[19] S. Huang, L. Meeus, B. Cornelis, B. Devolder, M. Martens, and A. Pižurica, "Paint loss detection via kernel sparse representation," in *Proc. Image Processing for Art Investigation (IP4AI)*, 2018, pp. 24–26.

[20] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*. McGraw-Hill Boston, 2000.

[21] H. L. Van Trees, "Detection, estimation, and modulation theory, optimum array processing (part iv)," *Wiley Interscience*, 2002.

[22] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.

[23] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, 1994.

[24] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, 2001.

[25] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[26] S. Huang, H. Zhang, and A. Pižurica, "A robust sparse representation model for hyperspectral image classification," *Sensors*, vol. 17, no. 9, p. 2087, 2017.

[27] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, 2017.

[28] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, 2016.

[29] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, 2011.

[30] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1346–1354, 2015.

[31] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, 2016.

[32] B. M. Haddad, S. Yang, L. J. Karam, J. Ye, N. S. Patel, and M. W. Braun, "Multifeature, sparse-based approach for defects detection and classification in semiconductor units," *IEEE Trans. Autom. Sci. Eng.*, 2016.

[33] M. R. Azimi-Sadjadi, N. Klausner, and J. Kopacz, "Detection of underwater targets using a subspace-based method with learning," *IEEE J. Ocean. Eng.*, 2017.

[34] J. J. Hall, M. R. Azimi-Sadjadi, S. G. Kargl, Y. Zhao, and K. L. Williams, "Underwater unexploded ordnance (uxo) classification using a matched subspace classifier with adaptive dictionaries," *IEEE J. Ocean. Eng.*, vol. 44, no. 3, pp. 739–752, 2019.

[35] Z. Yang, Q. Xu, W. Zhang, X. Cao, and Q. Huang, "Split multiplicative multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5147–5160, 2019.

[36] D. Hu, F. Nie, and X. Li, "Deep multimodal clustering for unsupervised audiovisual learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 9248–9257.

[37] M. Strese, C. Schuwerk, A. Iepure, and E. Steinbach, "Multimodal feature-based surface material classification," *IEEE Trans. Haptics*, vol. 10, no. 2, pp. 226–239, 2016.

[38] A. Jafari, A. Ganesan, C. S. K. Thalisetty, V. Sivasubramanian, T. Oates, and T. Mohsenin, "Sensornet: A scalable and low-power deep convolutional neural network for multimodal data classification," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 1, pp. 274–287, 2018.

[39] B. Xue and N. Tong, "Real-world isar object recognition using deep multimodal relation learning," *IEEE Trans. Cybern.*, 2019.

[40] P. Song, X. Deng, J. F. Mota, N. Deligiannis, P.-L. Dragotti, and M. Rodrigues, "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries," *IEEE Trans. Comput. Imag.*, 2019.

[41] Y. Gwon, W. Campbell, K. Brady, D. Sturim, M. Cha, and H. Kung, "Multimodal sparse coding for event detection," *arXiv preprint arXiv:1605.05212*, 2016.

[42] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 427–438, 2005.

[43] H. Q. Luong, B. Goossens, A. Pižurica, and W. Philips, "Joint photometric and geometric image registration in the total least square sense," *Pattern Recognit. Lett.*, vol. 32, no. 15, pp. 2061–2067, 2011.

[44] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Int. Conf. Artif. Neural Netw.* Springer, 1997, pp. 583–588.

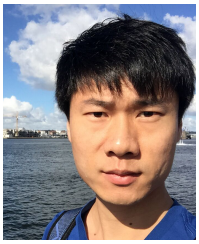
[45] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, 2001.

[46] J. Yang, A. F. Frangi, J.-y. Yang, D. Zhang, and Z. Jin, "Kpca plus lda: a complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 230–244, 2005.

[47] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy c-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573–584, 2013.

[48] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel methods on riemannian manifolds with gaussian rbf

kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2464–2477, 2015.
[49] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, 2013.
[50] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graphics*, vol. 27, no. 3, pp. 1–10, 2008.
[51] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 646–653, 2006.
[52] X. Chen and D. Cai, "Large scale spectral clustering with landmark-based representation," in *Proc. AAAI Conf. Artif. Intell.*, 2011, pp. 313–318.
[53] R. Chitta, R. Jin, T. C. Havens, and A. K. Jain, "Approximate kernel k-means: Solution to large scale kernel clustering," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2011, pp. 895–903.
[54] D. Yan, L. Huang, and M. I. Jordan, "Fast approximate spectral clustering," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2009, pp. 907–916.
[55] X. Peng, L. Zhang, and Z. Yi, "Scalable sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 430–437.
[56] M. De Mey, M. Martens, and C. Stroo, "Vision and material: Interaction between art and science in Jan van Eyck's time," *KVAB Press*, 2012.
[57] T. Ružic and A. Pižurica, "Context-aware patch-based image inpainting using Markov random field modeling," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 444–456, 2015.
[58] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 166–180, 2019.



Shaoguang Huang (S'17-M'20) received the M.S. degree in telecommunication and information system from Shandong University, Jinan, China, in 2015 and the Ph.D. degree in computer science engineering from Ghent University, Belgium, in 2019. He is currently a Post-Doctoral Research Fellow with the Group for Artificial Intelligence and Sparse Modelling (GAIM), Ghent University, Belgium.

His area of interests includes image processing, sparse representation, clustering, hyperspectral image analysis and machine learning.

Bruno Cornelis received the Masters degree of Industrial Engineer in Electronics from Erasmus Hogeschool Brussel (EHB), Belgium, in 2005 and the Masters degree of Electrical Engineering from Vrije Universiteit Brussel (VUB), Belgium, in 2007. He obtained a Ph.D. degree in Applied Sciences from VUB in 2014. During his Ph.D with the the Department of Electronics and Informatics he investigated the use of various image processing tools in support of art scholarship. His research interests include statistical data analysis and sparse representations in computational imaging applications. In the period of 2014-2015 he was a visiting assistant professor at the Mathematics Department of Duke University. Since 2016 he is a guest professor with the department of Electronics and Informatics at VUB.

Bart Devolder received his M.A. in painting conservation in 2002 at the Royal Academy of Fine Arts Antwerp, Belgium. He held internships at the Akademia Sztuk Pieknych Krakow, Poland, the Royal Institute for Cultural Heritage (KIK-IRPA), Brussels and at the Muse du Louvre, Paris. He received a fellowships from the Straus Center for Conservation at the Harvard University Art Museums (2003-2004) and he was the Andrew W. Mellon Fellow in Painting Conservation at the National Gallery of Art, Washington D.C. (2004-2007).

Bart has worked for the Kimbell Art Museum & Amon Carter Museum of American Art in Fort Worth, TX first as Assistant Conservator of Paintings (2007-2010) and later as Associate Conservator of paintings (2010-2012).

Before joining the Princeton University Art Museum in the summer of 2018, he was the On-site Coordinator/ Painting Conservator for the restoration of the Ghent Altarpiece by the brothers van Eyck (2012-2018).

Bart has studied, published and lectured on a wide variety of topics ranging from Fayum portraits, Early Netharlandish Canvaspaintings, the representation of gold brocades in Netherlandish paintings to the methods and techniques of Cubist paintings. He is also particularly interested in the newer applications of computer sciences to the field of studying old master paintings.

Bart likes to use the understanding of the way artworks are created as a catalyst to interact with people and students from different backgrounds and disciplines.

Maximiliaan Martens received a PhD in Art History from the University of California at Santa Barbara (1992). He is full professor in Art History at Ghent University and a member of the Royal Flemish Academy of Belgium for Science and Arts (KVAB). As a specialist in Flemish art of the fifteenth and early sixteenth century, he authored several books and articles on Jan van Eyck, Hans Memling, Pieter Bruegel the Elder and others and supervised numerous PhD-dissertations.

His main research interests are artistic production in an urban environment, technical art history and the history and theory of art conservation.

Aleksandra Pižurica (SM'15) received the Diploma in electrical engineering from the University of Novi Sad, Serbia, in 1994, the Master of Science degree in telecommunications from the University of Belgrade, Serbia, in 1997, and the Ph.D. degree in engineering from Ghent University, Belgium, in 2002. She is a Professor in statistical image modeling with Ghent University. Her research interests include the area of signal and image processing and machine learning, including multiresolution statistical image models, Markov Random Field models, sparse coding, representation learning, and image and video reconstruction, restoration, and analysis.

Prof. Pižurica served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2012 – 2016), Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2016 – 2019) and an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016 – 2020). She was also the Lead Guest Editor for the EURASIP Journal on Advances in Signal Processing for the Special Issue "Advanced Statistical Tools for Enhanced Quality Digital Imaging with Realistic Capture Models" (2013). The work of her team has been awarded twice the Best Paper Award of the IEEE Geoscience and Remote Sensing Society Data Fusion contest, in 2013 and 2014. She received the scientific prize "de Boelpaepe" for 2013–2014, awarded by the Royal Academy of Science, Letters and Fine Arts of Belgium for her contributions to statistical image modeling and applications to digital painting analysis.