

Multi-temporal image analysis for preventive conservation of historical musical instruments

ALIREZA REZAEI, SATIE Laboratory, Paris-Saclay University, France

EMANUEL ALDEA, SATIE Laboratory, Paris-Saclay University, France

PIERCARLO DONDI, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy

SYLVIE LE HÉGARAT-MASCLE, SATIE Laboratory, Paris-Saclay University, France

MARCO MALAGODI, Department of Musicology and Cultural Heritage, CISRiC - Arvedi Laboratory of Non-Invasive Diagnostics, University of Pavia, Italy

Artworks need to be constantly monitored to check their state of conservation and to quickly spot the eventual presence of alterations or damages. Preventive conservation is the set of practices employed to reach this goal. Unfortunately, this results generally in a cumbersome process involving multiple analytical techniques. Consequently, methods able to provide a quick preliminary examination of the artworks (e.g., optical monitoring) seem very promising in order to streamline preventive conservation procedures. We are especially interested in the study of historical wood musical instruments, a kind of artwork particularly subject to mechanical wear since they are both held in museums and also occasionally played in concerts. Our primary goal is to early detect possible altered regions on the surface of the instruments and thus provide the experts some precise indications on where to apply more in-depth examinations to check for potential damages. In this work, we propose an optical monitoring method based on the a-contrario probabilistic framework. Tests were conducted on the “Violins UVIFL imagery” dataset, a collection of UV induced fluorescence (UVIFL) image sequences of artificially altered wood samples and violins. Obtained results showed the robustness of the proposed method and its capability to properly detect altered regions while rejecting noise.

CCS Concepts: • **Mathematics of computing** → **Hypothesis testing and confidence interval computation**; • **Computing methodologies** → **Visual inspection**.

Additional Key Words and Phrases: Cultural heritage, preventive conservation, historical musical instruments, a-contrario decision

ACM Reference Format:

Alireza Rezaei, Emanuel Aldea, Piercarlo Dondi, Sylvie Le Hégarat-Masclé, and Marco Malagodi. 2021. Multi-temporal image analysis for preventive conservation of historical musical instruments. *ACM J. Comput. Cult. Herit.* 0, 0, Article 1 (2021), 20 pages.

Authors' addresses: Alireza Rezaei, alireza.rezaei@universite-paris-saclay.fr, SATIE Laboratory, Paris-Saclay University, Bâtiment 660, rue Noetzlin, Gif-sur-Yvette, France, 91190; Emanuel Aldea, emmanuel.aldea@universite-paris-saclay.fr, SATIE Laboratory, Paris-Saclay University, Bâtiment 660, rue Noetzlin, Gif-sur-Yvette, France, 91190; Piercarlo Dondi, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, Pavia, Italy, piercarlo.dondi@unipv.it; Sylvie Le Hégarat-Masclé, sylvie.le-hegarat@universite-paris-saclay.fr, SATIE Laboratory, Paris-Saclay University, Bâtiment 660, rue Noetzlin, Gif-sur-Yvette, France, 91190; Marco Malagodi, Department of Musicology and Cultural Heritage, CISRiC - Arvedi Laboratory of Non-Invasive Diagnostics, University of Pavia, Corso Garibaldi 178, Cremona, Italy, marco.malagodi@unipv.it.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Association for Computing Machinery.

XXXX-XXXX/2021/0-ART1 \$15.00

<https://doi.org/>

1 INTRODUCTION

The main aim of preventive conservation procedures is to reduce the risk of alteration of historical artifacts and to avoid the natural aging of materials (such as pigments, organic binders, or protective layers), with the consequence of a general reduction of the requirements to carry out operations of restoration. A preventive conservation plan involves several different operations including the study of the environmental areas of buildings, microclimatic conditions, materials, airborne pollutants, and even the effect of the visitors in the museum [62].

Preventive conservation procedures have been carried out since the 1980s, and thus, nowadays, there is an extensive scientific literature concerning the evaluation of the indoor microclimatic conditions in the museums, ancient palaces, or depositories [5, 54, 57], and it is well known that the main factors that can generally accelerate the degradation processes of artworks are temperature, dampness, relative humidity, and pollutants. Bad conservative conditions can increase the risk of damage to the materials of the exposed artworks, leading, for example, to chromatic variations, modifications in the organic structures of the binders, or mechanical alterations of the stratigraphic system of the artifact. In the last years, the research for preventive conservation plans has aimed to develop new standards and protocols to control the indoor environmental parameters and, at the same time, to check day by day the conservation state of artworks and their original materials [11, 31, 41].

During the centuries, a lot of cultural assets, such as furniture or archaeological finds, have lost their original employment, and today they are preserved as historic and artistic exemplars from the past centuries. Nowadays, those kinds of objects are considered at the same level as other more “standard” artworks, like paintings or statues, and thus they are included in preventive conservation plans. However, in some cases, the use of those objects can still be performed, as for musical instruments in general and violins specifically. Musical instruments represent a particular class of artworks, on which conservation is not limited to their materials but also concerns the preservation of their acoustic quality [7, 21]. The use of these artifacts involves new problems concerning the conservation of materials: (i) dirt deposits on the musical instruments, depending on the climatic conditions such as the presence of pollutants, the dirt particle size, or the varnish surface roughness; (ii) direct contact with the player that may prompt dirt deposition and adhesion on the surface, due to the exposure to bad conditions during the performance; (iii) wearing of the varnishes and presence on the surface layers of sweat and acid compounds due to the direct contact with the violinist’s skin. Furthermore, the analysis of their surface is particularly challenging. First of all, the varnishes are generally highly reflective, thus, noisy reflections, that can be confused for alterations, are common during photo acquisition. Secondly, varnish wear can evolve in different ways depending on both the initial conditions of the surface and on the different substances present. Finally, the surface can be very complex and stratified due to multiple restorations having occurred during centuries (which is very common for historical violins).

Because of all those conditions, it became important to develop innovative non-invasive diagnostic procedures to monitor material transformation and to evaluate the conservation state of the musical instruments, especially regarding the wearing of the varnishes. The combined use of multiple analytical techniques has proved beneficial for handling these issues [20, 51]; however, this preventive conservation approach is very time consuming and can become tedious and prone to human errors if multiple instruments are being monitored at the same time. A more efficient procedure should consist of regular but rapid optical analyses of images of instruments to quickly identify the *possible* altered areas. The result of this analysis will indicate where a more thorough multi-modal analysis (i.e., the application of spectroscopic examinations) is needed. Image acquisition is very fast compared to other chemical-physical examinations, thus, ideally, using image processing, it would be possible to frequently examine in a limited time the state of conservation of one or more artworks, focusing the human attention only on those that are genuinely at risk.

In this work, we propose a probabilistic method for the early detection of alterations on varnished wooden surfaces of historical instruments, using multi-temporal image series acquired through UV induced fluorescence

photography (UVIFL), a well-known diagnostic technique largely employed in Cultural Heritage studies [29]. Our proposal is based on the a-contrario framework [13, 14] and it improves the method presented in [47, 48] which deals with the detection of changes between two single frames: a reference image and an up-to-date acquisition. In this work, we extend the detection framework in order to process a temporal sequence of changes and to account for their significance with respect to the spatial plus time dimensions. The main benefit of this approach is that the analysis is now focused on the evolution of the alterations in time, which in turn allows us to detect the wear more reliably. Specifically, we can successfully handle noise and artifacts that would have been detrimental to the change detection process if only two frames were used. This is important as an alteration generally varies slowly between frames in its early stages, and abrupt changes are still possible, but less likely.

A preliminary version of this multi-temporal approach was presented in [46]. Here, we both improve the theoretical bases and present an in-depth experimental analysis.

The remainder of the paper is structured as follows: Section 2 introduces the UVIFL photography and explains why it is useful for our purposes; Section 3 provides a brief summary of state-of-the-art image processing methods for damage and change detection, with a focus on Cultural Heritage applications; Section 4 presents the proposed approach; Section 5 discusses the experiments and the obtained results; finally, Section 6 draws the conclusions and proposes possible future steps.

2 UV INDUCED FLUORESCENCE (UVIFL) PHOTOGRAPHY

In the Cultural Heritage field, UVIFL photography is one of the most commonly employed non-invasive diagnostic techniques. It is based on the properties of some organic substances that react to UV radiation (generally in the UV-A range, between 315 and 400 nm), re-emitting radiations in the visible spectrum (400 - 700 nm) [29, 55]. This process allows us to see details which are not visible when using a standard illumination, such as, in the case of historical violins, substances commonly adopted as binders, pigments, adhesives, or material used for retouching [6].

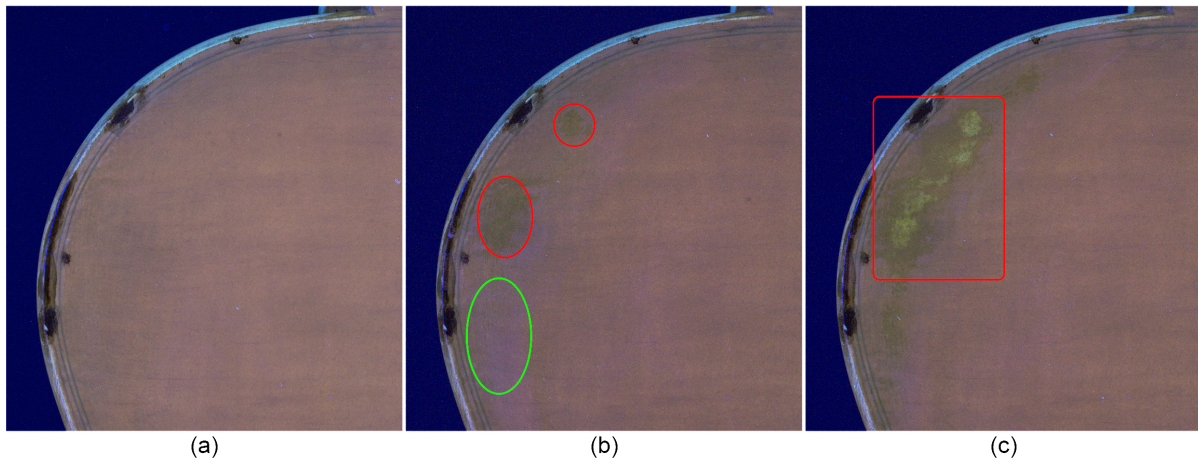


Fig. 1. Example of alteration growing on the top left side of a sample violin back plate (image brightness and contrast are manually increased for a better readability) – real alterations highlighted in red, noisy reflections in green: (a) the initial state, reference UVIFL image; (b) same region with some alterations in an early stage, i.e. having limited color variations that can be confused with noise; (c) same region with a large alteration in an advanced stage exhibiting very evident variations in both shape and color with respect to the reference image.

However, even if fluorescent materials produce characteristic color responses (e.g., a yellow fluorescence can be attributed to oils and a green one to protein substances such as hide glues [4]), the fluorescence phenomenon alone is not sufficient to unambiguously identify materials. In fact, the varnishes adopted in violin making include very heterogeneous mixtures of substances, that, added to the various restorations occurred during centuries, create very complex emission spectra. Hence, there is not a unique color-substance match. Moreover, UV radiation penetrates only the superficial layers of varnishes, thus the underlying ones cannot be detected by this technique. Given these limitations, UVIFL images are mainly employed as a preliminary analysis [17] to spot regions of interest where to apply more precise, but slower, spectroscopic techniques, such as X-Ray Fluorescence (XRF) [52] or Fourier Transform Infrared (FTIR) spectroscopy [28], to fully characterize the materials present.

In the current scenario of a constant monitoring, UVIFL photography can be particularly helpful to spot new superficial alterations, since, when the wear removes the outer layers of varnishes, the lower layers start becoming visible producing slightly different fluorescence colors. Thus, the occurrence of a color variation in a region while the surroundings remain unchanged is a clear hint of a possible alteration. Of course, it is important to take into account the possible presence of noisy reflections (due to the high reflective varnishes), that can sometimes occur even with a rigorous acquisition process, and that can be mistakenly interpreted as alterations, especially in the early stages, as illustrated in Figure 1.

3 PREVIOUS APPROACHES FOR DAMAGE AND CHANGE DETECTION

During the last two decades, computer vision has been more and more used in artwork health monitoring. A widely studied topic is the detection of cracks in paintings. Notable approaches in this field include the use of content-based analysis [1], the extension of morphological operations to hyper-spectral images [12] and the combination of elongated filters, multiscale morphology and K-SVD [10], subsequently improved using multimodal data [43]. Numerous other examples can be quoted from damage detection and identification in stone artifacts [8, 24, 38], to monitoring the decay of frescoes [26], to detection of damaged archaeological sites from space [9]. However, computer vision methods are still rarely used for wear detection in varnished wooden surfaces, such as those of historical musical instruments. Attempts have been made to detect areas with ingrained wear, that had peculiar chromatic characteristics [18], but this approach is useful only to monitor surface damages already in an advanced stage and not to spot new ones, that are typically smaller and more difficult to identify.

In this study, we assume that we have at our disposal an image time series, and that our objective is to detect wear evolution (apparition and/or spatial growing). Then, considering the first image in the series as a reference, we can cast wear detection as a change detection problem.

In recent years, change detection has been tackled with deep neural networks for a variety of applications such as remote sensing [30, 60], city [59], traffic [37] and medical [32] monitoring, or safety controls [56]. Typically, the feature extraction modules of these architectures are fine-tuned on a large enough set of samples of the target domain. However, even if machine and deep learning approaches have been also proposed and evaluated in the Cultural Heritage field, their adoption is not as widespread as in other sectors, mainly due to the limited availability of annotated data [22]. In our considered scenario, the relative scarcity of the data coupled with the high level of imprecision of the expert annotations favor alternative approaches in which prior information may be more readily integrated into the models. Luckily, the more traditional computer vision techniques have proved to be very useful and applicable in this context. For instance, recently, Dondi et al. [19] presented a segmentation approach based on histogram quantization and a genetic algorithm. Again, when extending the possible approaches to different application fields, the literature is vast, ranging from clustering to segmentation algorithms, and covering defect or anomaly detection methods as well. Let us thus better specify our problem.

Since wear areas are highly variable, we look for some approaches which can handle the difficulties in wear detection (unknown shape, unknown distribution, high amount of noise) while at the same time providing a criterion

for judging how likely a certain configuration of local variations would be an actual change. Both spatial and temporal consistencies could indicate a wear region. The probably most popular approaches to satisfy the mentioned requirements are the 3D data (width-height-time) clustering algorithms applied to image series, e.g. to monitor “gradual” changes in remote sensing [66]. However, the proposed methodologies are difficult to transpose to our case since these methods work with very long image sequences (spanning months and years, e.g. Santos et al. [53]) to cope with seasonal variations and they focus on semantic and balanced classes, e.g. Zhang et al. [64].

Apart from these rather traditional approaches, the a-contrario framework introduced by Desolneux et al. [13] has also the potential to comply with our problem specifications. Across the computer vision field, it has been used in a variety of problems, including the detection of changed areas across multi-temporal images. In early studies, Lisani and Morel [33] used a-contrario framework and spectral invariant features to detect meaningful changes between two satellite images of the same area taken at different times. In 2007, Rousseau et al. [50] proposed an a-contrario approach for change detection in three dimensional multi-modal medical images, such as Magnetic Resonance sequences. In 2010, a similar approach was proposed by Robin et al. [49] for sub-pixel change detection in a time-series of satellite images. Flenner and Hewer [23] further investigated this approach by using exchangeable random variables instead of independent and identically distributed (IID) assumption. As a later example, Liu et al. [34] applied a-contrario framework for comparing very high-resolution images of urban areas.

In the field of preventive conservation, the a-contrario framework has been used for the comparison of two single images to detect changed pixels, and then to group them either in a two-step process [45], or in a one-step process [48] by considering both their location and their gray-level value. While such approaches handle some of the aforementioned problems, they only work with two frames of the series at a time, missing some potentially very useful information present in the temporal evolution of the clusters, that would make the distinction between wear and noise more evident. The current paper aims to introduce the notion of time into the a-contrario framework via the definition of the noise model and the computation of the number of false alarms.

4 THE PROPOSED APPROACH

4.1 Basics in a-contrario framework

In the a-contrario framework, the words ‘a-contrario’ refer to the fact that detection is performed by contradicting a ‘naïve’ model that corresponds to the null hypothesis (H_0) in statistical decision theory. Then, to instantiate an a-contrario detection model, one has to define the two following elements: first, the naïve model that represents the noise in the absence of structures to detect, and second, a measurement of deviation relatively to the naïve model, from which the decision of a detection (or not) will be taken.

In image processing, two naïve models have been widely used, namely the Gaussian and Bernoulli models that stand for gray level and binary images, respectively. In the Bernoulli model, the pixels independently take values in $\{0, 1\}$ set. If we denote by p the Bernoulli distribution parameter (probability for a pixel to equal 1), this model is equivalent to assuming that the v pixels with value equal to 1 are uniformly distributed on the image spatial domain (of size $|\mathcal{I}|$ in pixel number) and $p = \frac{v}{|\mathcal{I}|}$:

DEFINITION 1 (NAÏVE MODEL $\mathcal{M}_0$). *The set of points with value 1 is a random set of v independent, uniformly distributed points over the image lattice $\mathcal{I}$.*

Note that according to this naïve model, the probability of observing a pattern of κ points by chance is given by the binomial distribution.

As second element of a-contrario detection, the Number of False Alarms (NFA) has been proposed in Desolneux’s seminal article [13]. It represents the average number of erroneous detections when considering a pattern (potential detection) of κ points, and it is simply computed from the tail of the binomial distribution (probability of a pattern

of at least κ points) and the number, called N_{test} , of considered patterns within the whole image:

$$NFA_0(\kappa, v, p) = N_{test} \sum_{i=\kappa}^v \binom{v}{i} p^i (1-p)^{v-i}, \quad (1)$$

where $\binom{k}{j}$ denotes the binomial coefficient equal to $\frac{k!}{j! \times (k-j)!}$ with $n! = n \times (n-1) \times \dots \times 2 \times 1$. In the previous equation, based on uniform distribution model $\mathcal{M}_0$, p is estimated by the ratio between the area of considered pattern with respect to the whole image area.

For wear monitoring, we had to adapt basic a-contrario approach in two ways: firstly, from 2D to 3D to introduce the time dimensionality, and secondly, focusing on patterns which correspond to point clusters.

4.2 From input data to 3D point cloud

The raw data are a series of n multi-temporal RGB images of varnished wooden samples and violins (see Section 5.1 for the dataset specifications); I_0 being the reference image.

Since the images were not acquired simultaneously (by definition of the monitoring process), we first have to account for possible illumination changes and match as accurately as possible each time frame with the original image. Thus, the following pre-processing steps are performed:

- (1) *Illumination correction.* Despite setting the same illumination hardware setup and configuration for each image capture, the illumination variation from one acquisition session to another may be of the same order of magnitude as the color variations due to wear appearance or increase. Then, for each frame I_t , $t > 0$, for each color channel, we normalize the mean pixel values with the reference frame I_0 value, i.e. subtract the difference between means $\delta_t = \mathbb{E}[I_t] - \mathbb{E}[I_0]$, where $\mathbb{E}[\cdot]$ is the expectation operator approximated here by the average over image pixels, and then apply a ReLU operator to remove negative values. Note that, in our case, we keep this pre-processing as simple as possible, but, in less controlled environments, one may rather use some local contrast improvement methods, such as those of Pizer et al.[42] and Hitam et al.[27]. Since our approach is modular, it will not change the following steps.
- (2) *Spatial registration.* For this step, we rely on keypoint matching techniques. Since our captures have been performed on a rigid object in a controlled environment, we can assume that the transformation for each image pair may be approximated using small translation and scaling components. Then, focusing on SIFT [35] features well-known for their invariance to illumination and MLESAC robust estimation [58], an affine transformation between each time frame I_t and I_0 is estimated, I_t is then projected in I_0 geometry so that the alignment is achieved at pixel level.
- (3) *Reflection removal.* The UV reflection on the surface of the samples usually appears as purple/blue spots in the image. Note this is a specific problem of UV photography on varnished surfaces. It is simpler to compensate than the more well known reflections in visible light caused by glasses or lenses, that require complex corrections [3, 63]. Here, a simple heuristic is enough for our purposes. Therefore, to remove the most intrusive reflections in our RGB images, we filter out the pixels that have significantly higher blue values than green ($B \gg G$) or have higher red and blue than green ($B > G$ & $R > G$). Note that, depending on the application, these operations might or might not be necessary, and that the general aim is simply to increase the correlation between change and the investigated process (in our case, the wear).

Since we aim to consider both spatial and time domains for detecting the wear, for a time series of n color images (in addition to the reference one) of spatial size $w \times h$ pixels, the input volume would be $w \times h \times (n+1) \times 3$. Then, to reduce this data volume without harming the wear detection, we propose to extract the points likely to belong to a wear region, i.e., *points of interest*.

For this step, we could either rely on wear color features (if learned for instance) or on change detection with respect to reference image I_0 . However, both due to the lack of sufficient labeled data and in order to be robust

to the variability of the material, varnish, and pigment present on the instruments, we focus on change detection approaches (i.e., we assume that we are only interested by new wear with respect to I_0). Then, considering each pair of frames (I_t, I_0) , $t > 0$, we first compute the color difference map, called Δ_t , according to CIEDE2000 formula [36] normalized in the $[0, 1]$ range. Secondly, from each Δ_t , $t > 0$, we derive a binary image of the pixels of interest (likely to belong to a wear area). In our case, the method used is the one described in [48] but alternatives, from simple fixed thresholding to clustering or binary segmentation, could have been considered depending on the wear contrast compared to no-wear areas. This results in a set of binary images denoted $\{B_t\}_{t=1}^n$.

Considering the whole series $\{B_t\}_{t=1}^n$, we create the 3D point cloud $\mathcal{P}$ with two spatial dimensions defined by the image domain along with a third time dimension:

$$\mathcal{P} = \{(x, y, z), x \in \{1, 2, \dots, w\}, y \in \{1, 2, \dots, h\}, z \text{ such that } z/c_t \in \{1, 2, \dots, n\}\}, \quad (2)$$

where c_t is the time coefficient, a real valued constant that controls the importance of a distance in time compared to a distance in the image space. If $c_t = 1$, then we give the same importance to both. Then, the 3D volume of the cuboid from which $\mathcal{P}$ was extracted is $V_{\mathcal{P}}$:

$$V_{\mathcal{P}} = w \times h \times (nc_t). \quad (3)$$

The derived 3D point cloud $\mathcal{P}$ along with $V_{\mathcal{P}}$ are the input of the proposed clustering algorithm presented in the next subsection.

4.3 Clustering algorithm

Algorithm 1 details the clustering method that we propose to detect and rank clusters of points from the more to the less likely to correspond to a wear area. This algorithm represents a methodological contribution, here applied to wear detection, but which could be adapted to the detection of any other “object” of interest characterized by spatio-temporal consistency.

The proposed approach relies on the a-contrario detection that allows for evaluating the significance of any point clusters based on the point distance [16]. In this work, since Equation (2) has already integrated the desired balance between spatial and temporal aspect, we can simply consider the 3D Euclidean distance between the points of $\mathcal{P}$: for any pair of 3D points $(x_a, y_a, z_a), (x_b, y_b, z_b) \in \mathcal{P}$, $D_{a,b} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2}$. In Algorithm 1, distances are computed in the first **for** loop.

Then, the clusters are defined based on the distance as follows: for a given distance δ , a cluster C is a subset of $\mathcal{P}$ such that for any point in C there is another point in C at a distance lower than δ and there is no other point in $\mathcal{P} \setminus C$ at a distance lower than δ :

$$\begin{cases} \forall p \in C, & \exists p' \in C \text{ such that } D(p, p') \leq \delta, \\ \forall p'' \in \mathcal{P} \setminus C, & \nexists p' \in C \text{ such that } D(p'', p') \leq \delta. \end{cases} \quad (4)$$

Note that with such a definition, not all subsets of $\mathcal{P}$ are clusters. From now on, a “well-defined” cluster follows the previous definition. Now, in practice, the clusters are defined from the nodes of the minimum spanning tree mentioned in line 4 of the Algorithm 1. This tree is created by a merging process: starting from each point as a cluster, in each step we find the two closest clusters and connect them by a mutual node which stores the distance between them (which is the intrinsic distance of the merged cluster). This way, after the tree is created each subtree represents a cluster, and its root stores the value δ associated to cluster.

Then, the measure of significance of such a well-defined cluster involves the computation of two normalized volumes (defined from 3D closed shapes in $\mathbb{R}^3$). The first one represents the volume associated to cluster C . Since the points in C are connected up to distance δ , the dilation of every C point with a radius $\delta/2$ provides a 3D connected component in $\mathbb{R}^3$, called $\mathcal{V}(C)$. To ensure that $\mathcal{V}(C)$ is a closed area, we finally perform a morphological

Algorithm 1 Detecting and ranking clusters ; inputs: 3D point cloud $\mathcal{P}$ and original 3D volume $V_{\mathcal{P}}$; output: list of detected clusters $\mathbf{C}$ and their associated meaningfulness.

```

1: for each pair of points  $j$  and  $j'$  in  $\mathcal{P}$  do
2:   Compute 3D distance  $D(j, j')$ 
3: end for
4: Compute the minimum spanning tree for the points in  $\mathcal{P}$  based on  $D(j, j')$  so that each subtree in hierarchical
   representation stores in its root the intrinsic distance parameter  $\delta$  of the cluster associated to the subtree
5:  $i \leftarrow 1$ 
6: for each subtree  $T$  do
7:    $C_i \leftarrow$  cluster of points in  $T$ 
8:    $\delta \leftarrow$  value stored in the root of  $T$ 
9:    $\delta' \leftarrow$  value in the parent node of  $T$ 
10:   $\mathcal{V}(C_i) \leftarrow$  dilation of  $C_i$  with radius  $\delta/2$ 
11:   $\rho \leftarrow \delta' - \delta$ 
12:  Compute volumes  $V_{\underline{C}}$  and  $V_{\overline{C}}$  according to Eq. (5)
13:   $S_i \leftarrow S(C_i)$  computed using Eq. (6) and logarithm function
14:   $i \leftarrow i + 1$ 
15: end for
16:  $\mathcal{I} \leftarrow$  list of indices of the clusters sorted according to decreasing values of  $S_i$ 
17:  $\mathbf{C} \leftarrow \emptyset$ 
18: for each index  $j$  in  $\mathcal{I}$  do
19:    $C_j \leftarrow j^{th}$  cluster according to  $\mathcal{I}$ 
20:   if  $\forall C_l \in \mathbf{C}, C_j \cap C_l = \emptyset$  then
21:      $\mathbf{C} \leftarrow \mathbf{C} \cup \{C_j\}$ 
22:   end if
23: end for
24: return  $\mathbf{C}$  along with  $S_i$  values associated to  $\mathbf{C}$  elements

```

closing of this latter. The closing radius ρ is chosen as the maximum value which will not add any point to $\mathcal{V}(C)$, i.e. the difference between δ and the distance δ' to the closest point not in C : $\rho = \delta' - \delta$. Then, C volume is approximated by the 3D volume of $\mathcal{V}(C)$ after closing.

The second volume of interest is the one of the points not in C , i.e. the points in $\mathcal{P} \setminus C$. It is evaluated as the complementary with respect to the initial 3D cuboid, of the dilation of $\mathcal{V}(C)$ with radius ρ , i.e. up to its closest point in $\mathcal{P} \setminus C$, so that the empty space around C does not belong to the volume of the points in $\mathcal{P} \setminus C$. Then, the two volumes of interest are computed according to:

$$\begin{cases} V_{\underline{C}} = \frac{1}{V_{\mathcal{P}}} \times [\text{volume of } \mathcal{V}(C) \text{ after closing with radius } \rho], \\ V_{\overline{C}} = \frac{1}{V_{\mathcal{P}}} \times [\text{volume of } \mathcal{V}(C) \text{ after dilation with radius } \rho]. \end{cases} \quad (5)$$

Then, for any well-defined cluster $C \subseteq \mathcal{P}$ having $|C|$ points among the total number of points $|\mathcal{P}|$, we are able to compute the Number of False Alarms (NFA) criterion as:

$$NFA(C) \propto \sum_{j=|C|}^{|\mathcal{P}|} \binom{|\mathcal{P}|}{j} V_{\underline{C}}^j (1 - V_{\overline{C}})^{|\mathcal{P}|-j}. \quad (6)$$

This formula comes from the assumption that, in absence of any structure, points are uniformly distributed [15], in our case in the 3D space. Then, the significance of a cluster C is evaluated as:

$$S(C) = -\log(NFA(C)), \quad (7)$$

i.e., the higher this significance value, the more likely the cluster is not a false alarm, but a real wear area characterized by points both spatially close and time-consistent. In Algorithm 1, the significance values of all well-defined clusters are computed in the second **for** loop.

The last part of Algorithm 1 (including the last **for** loop) aims at only keeping the most significant clusters that are not redundant, i.e., correspond to disjoint subsets of points. Indeed, our objective is to get a partition of $\mathcal{P}$. Therefore, from all the derived clusters C_i , only the ones not intersecting another well-defined cluster having a higher significance value are kept. This corresponds to keeping only the *maximal* clusters as defined in [14]. Then, the list of these *maximal* clusters sorted by their significance value is the output of the algorithm.

Finally, let us recall that Algorithm 1 does nothing else than detecting 3D sub-volumes (of unknown shape and unknown distribution) that are significantly denser than the whole 3D point cloud. Then, it is up to the user to provide a point cloud where the objects of interest exhibit such higher density. In our case, wear areas correspond indeed to areas that time-persistently exhibit many pixels having a significant color difference with respect to the reference image, and thus will be detected as dense in both spatial and temporal dimensions.

4.4 Wear area ranking

After running Algorithm 1, we get a sorted list of spatio-temporal clusters with the most significant cluster in the first rank, the second most significant disjoint cluster in the second rank, and so on. Significance is higher for more compact or dense clusters of pixels which indicate individually a local change with respect to the reference frame. Based on the assumptions that wear areas are linked to color change, and that compact or dense clusters emphasize time-persistent changes, we argue that the derived significance values provide a relevant ranking of the likelihood of the wear areas.

In the experiments presented in next section, from the ranked set C (output of Algorithm 1), we keep the three most meaningful clusters along with their respective rank that can be used in the evaluation. We also underline that, in contrast to clustering algorithms, providing such a ranking is a strength of the proposed method.

Finally, considering the temporal evolution of the clusters, we can further refine the results and identify more reliably the wear region(s). For example, a small but constantly growing cluster is more likely to be a worn-out area with respect to a large but stable cluster.

5 EXPERIMENTS

5.1 Dataset

For our experiments we used the “Violins UVIFL imagery” dataset¹ [19], a collection of multi-temporal UVIFL images, referring to both real instruments and artificially altered wood samples artificially altered in the laboratory to simulate the effect of playing. Note that the samples were still made by violin makers, using common substances employed for violin varnishing, as well as the correct type of wood, to have a realistic reference. We focused on the artificially created sequences since those referring to real historical violins contain only three sessions each, too few for performing a proper multi-temporal analysis. Each sequence consists in one to three images of the initial state of the sample and then a series of subsequent frames where the surface was progressively damaged, and the wear starts and grows. The used sequences are the following: WS01 (Figure 2, first line), a 20-frame sequence that displays the appearing and the growing of an alteration in an area with intact varnish; WS02 (Figure 2, second line), a short 9-frame sequence that shows the quick growing of an alteration on an intact area with a thin layer of

¹<https://vision.unipv.it/research/UVIFL-Dataset/>

varnish; SV01 (Figure 2, third line), a 20-frame sequence that shows the growing of an already worn-out region on a sample violin back plate; and SV02 (Figure 2, fourth line), a 35-frame sequence (added to the dataset for this work) showing a slowly growing alteration on the intact surface of a violin back plate with a thick layer of varnish.

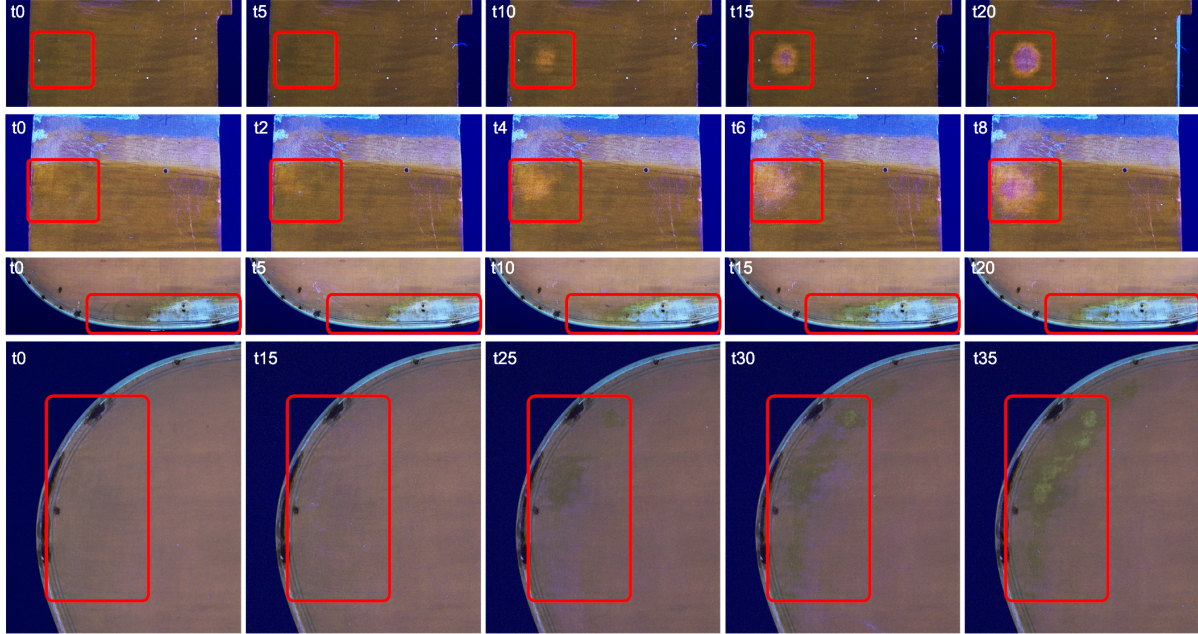


Fig. 2. Sample frames at different times from three sequences of the “Violins UVIFL imagery” dataset (image brightness and contrast are manually increased for a better readability) – in red the regions where the alterations will appear: first line, sequence WS01; second line, WS02; third line SV01; fourth line, SV02.

It must be stressed that the artificial “scratching” of the surface was slow and gradual, thus, in all the sequences, in the first frames no alteration occurs. This is particularly evident in the sequence SV02, where the varnish was so thick that no meaningful variation appeared before frame 19-20.

5.2 Results

For each frame in each sequence, we generate a binary change image with respect to the reference frame of that sequence by applying the method described in [47].

By accumulating the change detection results of all frames, we create a 3D point cloud for each sequence (Figure 3). The coefficient c_t is set to 2: This gives slightly more importance to spatial proximity over time proximity, penalizes the clusters which disappear in some frames, and keeps the size of the representation domain small enough for better performance. The time domain is the upward axis in all visualizations. From this figure, it clearly appears that there is a considerable amount of noise present in all point clouds. In addition, there are artifacts present on the left and right sides of all sequences, which are related to the UV reflections from the wooden sample.

The optical monitoring of a varnished wooden surface is an ongoing process meaning that the number of available frames and the accuracy of the detection increases as time goes on. At some point in time, the number of available frames becomes too large. To keep only the most relevant information, we thus maintain a sliding window of the last n frames. For each experiment on a specific sequence, n remains constant (in the following experiments $n = 10$).

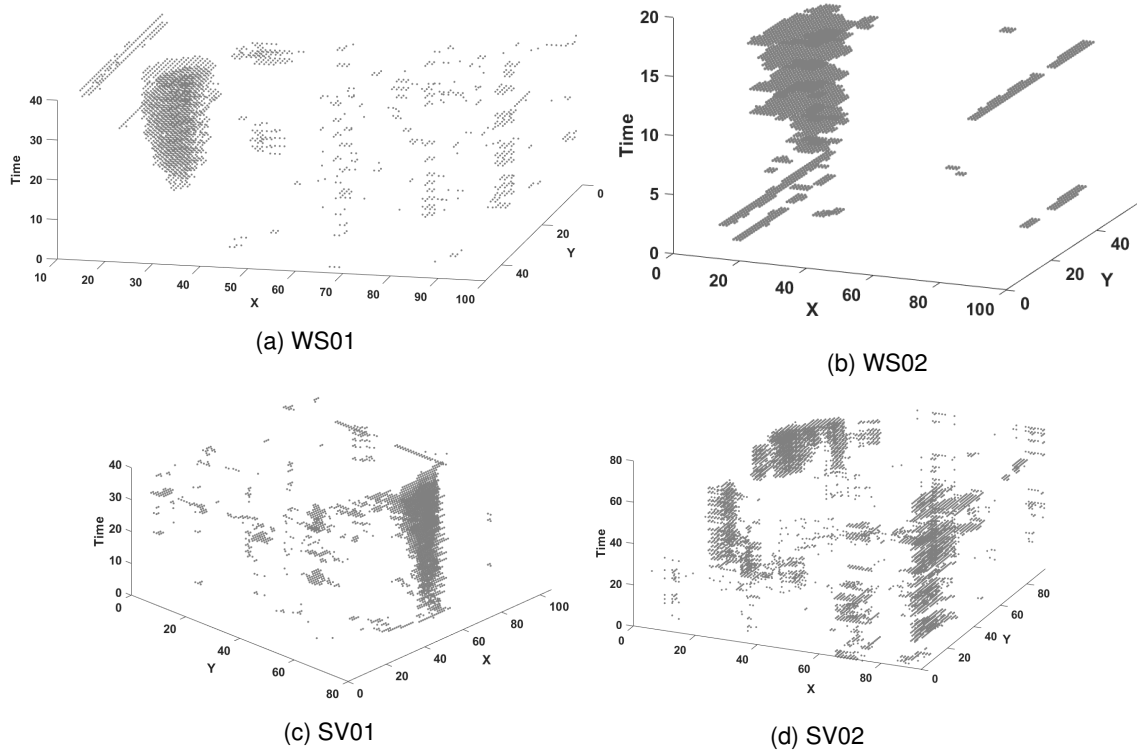


Fig. 3. Point clouds $\mathcal{P}$ derived from four different sequences, WS01 (a), WS02 (b), SV01 (c) and SV02 (d). These sequences have various length in terms of number of frames and exhibit different kind of alterations and artifacts (cf. Section 5.1). X and Y axes are in pixels, while the time dimension (Z axis) depends on the factor c_t (in these experiments $c_t = 2$).

To simulate a practical optical monitoring process and to analyze different detection results at any given point in time, we start the experiment with the first two frames for each sequence; then, we run the algorithm repeatedly adding one more frame each time until we reach n frames. From then on, we keep only the last n frames; adding a new frame and removing the oldest one in each step.

Each run outputs a list of clusters sorted according to their meaningfulness value. Since each run uses frames situated later in time, the rank of each cluster within the whole set of clusters and its associated meaningfulness value evolve depending on the nature of the cluster. In general, assuming that the wear region expands over time, our expectation is that the wear cluster starts in lower ranks and steadily rises to the first rank. Similarly, we expect that the significance of the wear region increases over time. On the opposite side, the meaningfulness value for the noise and artifact clusters should remain nearly constant or change randomly. Another difference between the wear region and the rest is that it should be present in every frame after its first appearance. For other non-wear areas, it is possible that they divide into two or more clusters along the time axis. In that case, we only take into account the one with the highest meaningfulness.

Figures 4, 5, 6 and 7 illustrate the detection results for each sequence through time, i.e., based on the current last frame used as input. In each case, the top three clusters were shown. It is noticeable that, firstly, only clusters

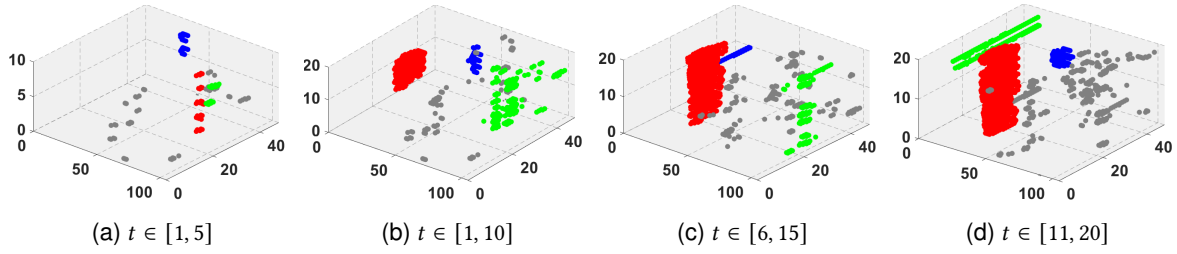


Fig. 4. Evolution of the detection by using the later time frames t in the sequence WS01. Color code gives the rank according to significance: red first, green second, blue third. The time domain is the upward axis.

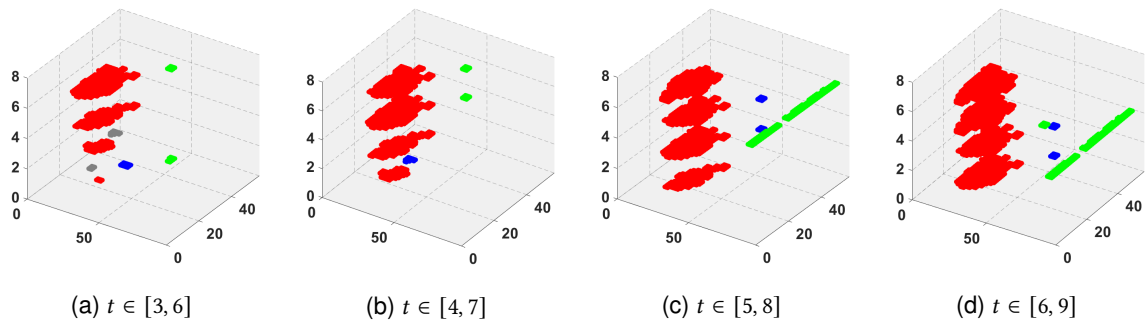


Fig. 5. Evolution of the detection by using the later time frames t in the sequence WS02. Same conventions as Figure 4.

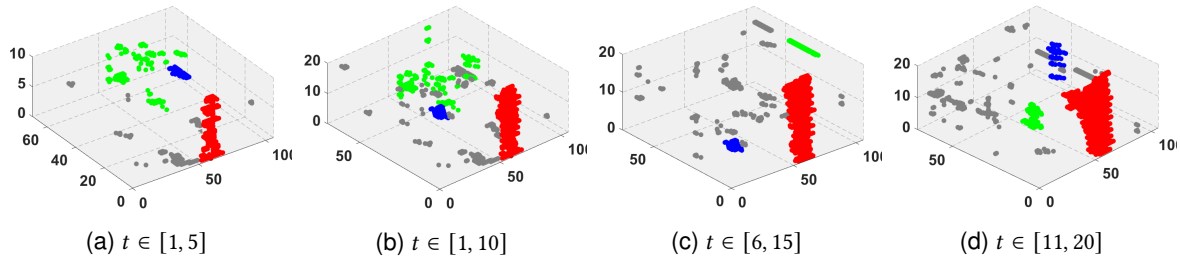


Fig. 6. Evolution of the detection by using the later time frames t in the sequence SV01. Same conventions as Figure 4.

which are consistent were detected. Minor artifacts and small noises were ignored, a fact which is a very desirable behavior when multi-temporal information is available. Secondly, the region of interest which is the wear area has been chosen by the end of the sequence as the most significant cluster.

For a more in-depth analysis, we can follow the meaningfulness evolution in time of both the wear and the noise artifacts for each sequence. Figure 8 shows the cluster meaningfulness (or significance, Equation (7)) versus the number of the last time frame. Specifically, for a sequence of n_{max} images, a time window set equal to t_w , and x-axis value equal to $n_f \in \{2, n_{max}\}$, Algorithm 1 is run with the subset of images indexed between $\max(1, n_f + 1 - t_w)$

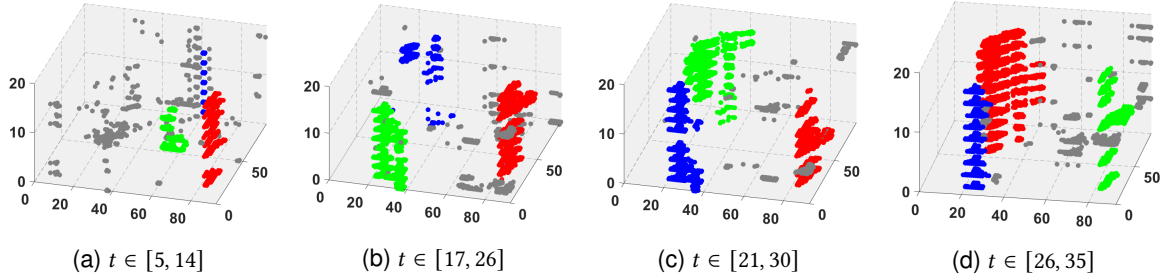


Fig. 7. Evolution of the detection by using the later time frames t in the sequence SV02. Same conventions as Figure 4.

and n_f as input; then, y-axis shows the highest meaningfulness values of the clusters at the output of the algorithm. Note also that the ad-hoc parameter t_w may be specific to each sequence, since it depends on the image acquisition frequency with respect to the instrument use and it is a compromise between temporal redundancy (to filter artifacts), sensitivity to early wear and the algorithm memory load. In this study, it was set equal to 10 for the longest sequences (WS01, SV01 and SV02) and to 4 for the shortest one (WS02).

Observing Figure 8a, we can infer that in sequence WS01 the meaningfulness of the most significant noise cluster (the orange line) remains very low and does not change throughout the experiment. This is expected because the noise regions are fairly small and do not grow over time. The cluster indicating the wear region (the blue line) starts from frame seven when the wear appears and consistently increases over time. This is in line with our expectation about an “ideal” wear that keeps growing. In practice, the wear may not expand over multiple acquisitions which will show itself as a plateau. However, it will never shrink or disappear so the values should remain increasing overall.

This trend can also be seen in Figure 8b for the sequence WS02, the shortest one. In this case, we used a smaller window of only 4 frames. The meaningfulness of the noise clusters remain more or less constant while the wear increases in significance after it appears.

In the same manner, we can interpret Figure 8c for sequence SV01. The wear cluster (the blue line) appears from the beginning and always has the highest meaningfulness. This latter increases in each step as the wear areas grow through time. This fast growth is coherent with the characteristics of the sequence, that shows the worsening of an already present worn-out region. The two distinct artifact clusters (the gray and orange lines) present in the sequence have meaningfulness values which change randomly from one step to another and sometimes disappear altogether. This is again in line with our expectation of a typical artifact region.

Finally, Figure 8d shows the results for sequence SV02, the most complex one. In the first 10 frames, we have two artifact clusters that appear to have increasing meaningfulness. This can lead to a false positive detection until we have enough frames to see their actual behavior. Artifact 2 (the gray line) is proven to be an artifact after frame 6. The same applies to artifact 1 (the orange line), after frame 12. On the other hand, the wear cluster (the blue line) has a meaningfulness value which continues to increase from its appearance around frame 20 until the end. More importantly, by the time the wear appears, we already dismissed the other two clusters as artifacts; therefore, automatically, it becomes the sole candidate wear region.

Comparing Figures 8b and 8d, we can see the behavior of the two extreme cases, the very thin and very thick varnish, respectively. It is interesting to notice, that, even if in the second case the wear takes clearly more time to occur, the behavior after the alteration appearance is similar in both cases, with a quick growing of the wear region.

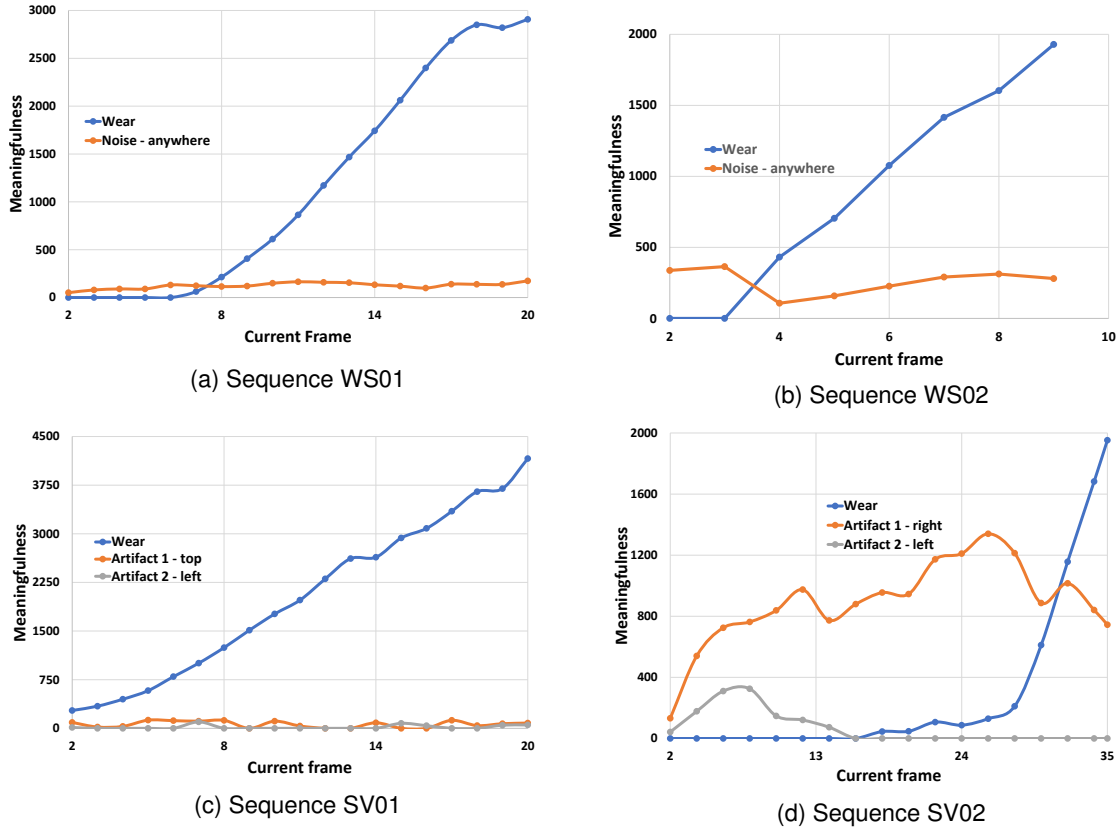


Fig. 8. Evolution of the meaningfulness value of sample clusters – output of Algorithm 1 in different runs of the algorithm, versus the considered “final” frame number, considering sequence WS01 (a), WS02 (b), SV01 (c) and SV02 (d). These sequences have various length in terms of number of frames and exhibit different kind of alterations and artifacts (cf. Section 5.1) For sequence WS02, in each run of Algorithm 1, at most the last 4 frames were used. For the rest, the last 10 frames were used.

Overall, the study of the evolution of meaningfulness values helps us to quickly identify the wear region as the correct area of interest even if it is smaller than the artifacts present on the surface, in full agreement with the preventive conservation principles. Compared to previous works leveraging single change maps [47, 48], the current analysis based on the temporal dynamics of the changes is able to pinpoint more reliably the emergence of a wear, also reducing the false positive detection. Figure 9 illustrates the improvements made by the multi-temporal analysis with respect to the single change map detection. Comparing the results achievable by using the method described in [48] (second row), that considers only two frames at time, with the outcomes of our current proposal (third row), we can notice how the multi-temporal analysis clearly improves the change detection by removing all the artifacts which do not grow over time and/or are not consistently present in every frame.

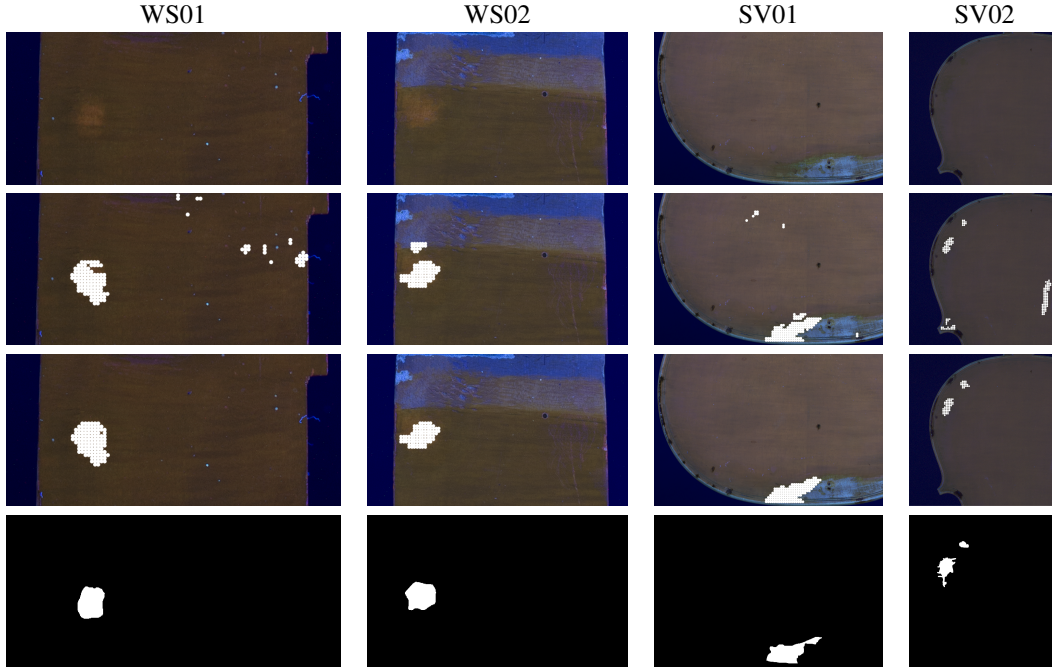


Fig. 9. Comparison, on sample frames of each sequence, between a wear detection performed considering only two frames at a time and the proposed multi-temporal approach: first row, original UVIFL image; second row, detected clusters as in [48]; third row, detected clusters applying the multi-temporal analysis; fourth row, ground truth showing the actual worn-out regions.

5.3 Comparison with other clustering approaches

Finally, we made a quantitative comparison between our proposed multi-temporal 3D clustering algorithm and several other clustering methods, namely: Agglomerative hierarchical clustering with complete linkage [40], Kmeans++ [2], robust spectral clustering [65], Expectation Maximization for Gaussian Mixture models (EM GM) [39], GBKmeans [44] and clustering by local gravitation [61]. We made sure to select algorithms from different families of clustering methods, considering both well-established approaches and some newly proposed ones. We apply these methods on the point cloud $\mathcal{P}$ (Eq. 2) generated from the input binary images. If the algorithm requires the total number of clusters, we set it to 4. Each method would cluster the point cloud $\mathcal{P}$ into several clusters; since none of them have a built-in feature for ranking the output clusters, to perform a fair comparison among the algorithms, we evaluated every cluster they produce with our metric and selected the best performing one. As metric for the comparison, we chose the F-score that is one of the most popular metrics [25] to evaluate the result of a binary classification of a detection output with respect to a ground truth. Some examples of binary maps corresponding to detection and binary ground truth are given in Figure 9. F-score is computed from the numbers of True Positives (TP), False Positives (FP) and False Negatives (FN). Specifically, F-score is the harmonic

mean of precision and recall defined as:

$$\begin{aligned} \text{F-score} &= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \\ \text{with Precision} &= \frac{TP}{TP + FP} \text{ and } \text{Recall} = \frac{TP}{TP + FN}. \end{aligned} \quad (8)$$

Table 1. The average $\pm$ standard deviation of F-score for the 3D clustering of each sequence using the proposed algorithm as well as six other clustering methods. First and second-best results are highlighted in green and light green respectively. The last column shows the run time values for each algorithm using a test point cloud with 2000 points (the test was done using a PC with a 3.1GHz CPU and 16GB of RAM).

Algorithm	WS01	WS02	SV01	SV02	Runtime
Agglomerative [40]	0.657 $\pm$ 0.088	0.624 $\pm$ 0.109	0.707 $\pm$ 0.128	0.820 $\pm$ 0.124	0.06s
Kmeans++ [2]	0.694 $\pm$ 0.128	0.572 $\pm$ 0.110	0.709 $\pm$ 0.129	0.727 $\pm$ 0.273	0.04s
Robust Spectral [65]	0.699 $\pm$ 0.108	0.735 $\pm$ 0.063	0.716 $\pm$ 0.076	0.564 $\pm$ 0.270	207.30s
EM GM [39]	0.756 $\pm$ 0.126	0.700 $\pm$ 0.061	0.810 $\pm$ 0.083	0.630 $\pm$ 0.263	0.70s
GBKmeans [44]	0.630 $\pm$ 0.116	0.556 $\pm$ 0.125	0.585 $\pm$ 0.144	0.646 $\pm$ 0.193	2.09s
Local gravitation [61]	0.773 $\pm$ 0.119	0.743 $\pm$ 0.048	0.796 $\pm$ 0.081	0.703 $\pm$ 0.137	2.44s
NFA (Ours)	0.784 $\pm$ 0.119	0.809 $\pm$ 0.061	0.812 $\pm$ 0.113	0.858 $\pm$ 0.130	84.46s

For each sequence, after computing the F-score for every frame containing any wear, we calculated the average and the standard deviation. The results are summarized in Table 1. As we can see, our proposal performs better than all the other state-of-the-art solutions in all four sequences, by having the highest average F-score while maintaining a low standard deviation. It should also be noticed that the second-best algorithm is different for each sequence, meaning that there is no alternative method with consistent performance in all the considered conditions. Although this comparison is performed at pixel level, at object level the benefit of our approach will be even more visible. Indeed F-score aims to measure the method's ability to detect *all* the wear pixels, but it does not directly reflect the algorithm's capability to early detect worn-out regions. In fact, although it is important to be able to identify as many wear pixels as possible, in a monitoring process we can afford to lose or mislabel some boundary pixels if we properly spot the correct position of the wear clusters and avoid the noise. Thus, at pixel level we prefer a high Precision (to guarantee noise avoidance), while at cluster level a high Recall (to not miss any wear cluster). This is highlighted in Figures 8 and 9, where we can see that in all four sequences the proposed algorithm was able to properly identify the wear position and distinguish it from any noise cluster(s), in a few frames from its appearance.

Table 1 also shows the running times of each algorithm when applied to a test point cloud containing 2000 points. Such a number of points represents a typical case for a multi-temporal image sequence. However, point clouds with initially higher number of points can be simplified either by resizing the input sequence or by down-sampling the point cloud itself, depending on available computational resources. The test was done using a PC with a 3.1GHz CPU and 16GB of RAM. All algorithms were implemented in MATLAB R2020a. As we can see, the running times can be divided to four categories: very high values (Robust Spectral [65]), high values (NFA), low values (GBKmeans [44], Local gravitation [61]) and very low values (Agglomerative [40], Kmeans++ [2], EM GM [39]). Then, the running time of the proposed algorithm is among the highest. However, it remains in order of few minutes which, for the considered application, seems completely reasonable. We do not need a real-time response, accuracy in the detection is more important than speed. In addition, the main loop of the algorithm (lines 6 to 15 in

Algorithm 1), which is the most time consuming section, can be run in parallel. This is currently done on the CPU; therefore, a GPU implementation would drastically shorten the run time.

6 CONCLUSIONS

In this paper, we introduced a probabilistic method for detecting new alterations on varnished surfaces starting from a noisy multi-temporal series of images. The proposed method can be used in preventive conservation, to perform continuous monitoring of relics susceptible to wear (e.g., historical musical instruments still played today). The goal is to help experts and restorers in their work, with early and quick identification of possible altered areas, on which they can perform further in-depth examinations, such as spectroscopic and chemical analysis.

Different from other statistical learning-based and deep learning methods, our algorithm, rooted in the Gestalt theory, does not need precise annotations, and can properly work even with little data. These are both important properties in the Cultural Heritage fields, where limited availability of data and a lack of proper annotations are common.

Tests conducted on the “Violins UVIFL imagery” dataset showed good performance compared to the other state-of-the-art algorithms. It should also be stressed that, even if we focused on the case of historical violins, our method can be adapted to work on other kinds of relics with few modifications, mainly in the preliminary process of the input images.

Future works will involve the start of long monitoring (more than one or two years) of selected historical musical instruments, as soon as the concert season will restart at the same frequency as during the pre-pandemic period. Long monitoring will allow us to further refine our approach to deal with even more challenging conditions, such as additional noise due to occasional cleaning of the instruments (that can “mask” the presence of wear) or a higher probability of acquisition mistakes (e.g., due to different people performing the photos at different times).

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the French-Italian Galileo PHC partnership (project 44391VL/G20-160, *MUSical Instrument Conservation with Optical Monitoring (MUSICOM)*).

REFERENCES

- [1] Fazly Salleh Abas and Kirk Martinez. 2003. Classification of painting cracks for content-based analysis. In *Proc. SPIE 5011, Machine Vision Applications in Industrial Inspection XI*, Vol. 5011. SPIE, 149–161.
- [2] David Arthur and Sergei Vassilvitskii. 2006. *k-means++: The advantages of careful seeding*. Technical Report. Stanford.
- [3] Alessandro Artusi, Francesco Banterle, and Dmitry Chetverikov. 2011. A survey of specular removal methods. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 2208–2230.
- [4] Giovanna Bitossi, Rodorico Giorgi, Marcello Mauro, Barbara Salvadori, and Luigi Dei. 2005. Spectroscopic Techniques in Cultural Heritage Conservation: A Survey. *Applied Spectroscopy Reviews* 40, 3 (2005), 187–228. <https://doi.org/10.1081/ASR-200054370>
- [5] Susan Bradley. 2005. Preventive Conservation Research and Practice at the British Museum. *Journal of the American Institute for Conservation* 44, 3 (2005), 159–173. <https://doi.org/10.1179/019713605806082248>
- [6] Brigitte Brandmair and Peter Stefan Greiner. 2010. *Stradivari Varnish: Scientific Analysis of His Finishing Technique on Selected Instruments*. Serving Audio.
- [7] Voichita Bucur. 2016. *Handbook of materials for string musical instruments*. Springer.
- [8] Maria Mercedes Cerimele and Rossella Cossu. 2007. Decay regions segmentation from color images of ancient monuments using fast marching method. *Journal of Cultural Heritage* 8, 2 (2007), 170 – 175. <https://doi.org/10.1016/j.culher.2007.01.006>
- [9] Daniele Cerra, Simon Plank, Vasiliki Lysandrou, and Jiaojiao Tian. 2016. Cultural Heritage Sites in Danger—Towards Automatic Damage Detection from Space. *Remote Sensing* 8, 9 (Sep 2016), 781. <https://doi.org/10.3390/rs8090781>
- [10] Bruno Cornelis, Tijana Ruzic, Emile Gezels, Ann Dooms, Aleksandra Pizurica, Ljiljana Platisa, Jan Cornelis, Maximiliaan Martens, Marc De Mey, and Ingrid Daubechies. 2013. Crack detection and inpainting for virtual restoration of paintings: The case of the Ghent Altarpiece. *Signal Processing* 93, 3 (2013), 605 – 619.
- [11] CSN EN 16242 2011. *Conservation of cultural heritage - Procedures and instruments for measuring humidity in the air and moisture exchanges between air and cultural property*. Standard. European Standard.

- [12] Hilda Deborah, Noel Richard, and Jon Yngve Hardeberg. 2015. Hyperspectral crack detection in paintings. In *2015 Colour and Visual Computing Symposium (CVCS)*. 1–6. <https://doi.org/10.1109/CVCS.2015.7274902>
- [13] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. 2000. Meaningful alignments. *International journal of computer vision* 40, 1 (2000), 7–23.
- [14] Agnès Desolneux, Lionel Moisan, and J-M Morel. 2003. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 4 (2003), 508–513.
- [15] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. 2004. Gestalt theory and computer vision. In *Seeing, Thinking and Knowing*. Springer, 71–101.
- [16] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. 2007. *From gestalt theory to image analysis: a probabilistic approach*. Vol. 34. Springer Science & Business Media.
- [17] Piercarlo Dondi, Luca Lombardi, Claudia Invernizzi, Tommaso Rovetta, Marco Malagodi, and Maurizio Licchelli. 2017. Automatic Analysis of UV-Induced Fluorescence Imagery of Historical Violins. *Journal on Computing and Cultural Heritage* 10, 2, Article 12 (2017), 13 pages. <https://doi.org/10.1145/3051472>
- [18] Piercarlo Dondi, Luca Lombardi, Marco Malagodi, and Maurizio Licchelli. 2016. Automatic identification of varnish wear on historical instruments: The case of Antonio Stradivari violins. *Journal of Cultural Heritage* 22 (2016), 968–973.
- [19] Piercarlo Dondi, Luca Lombardi, Marco Malagodi, and Maurizio Licchelli. 2019. Segmentation of multi-temporal UV-induced fluorescence images of historical violins. In *New Trends in Image Analysis and Processing – ICIAP 2019 (Lecture Notes in Computer Science, Vol. 11808)*. Springer International Publishing, 81–91.
- [20] Giusj Valentina Fichera, Michela Albano, Giacomo Fiocco, Claudia Invernizzi, Maurizio Licchelli, Marco Malagodi, and Tomaso Rovetta. 2018. Innovative Monitoring Plan for the Preventive Conservation of Historical Musical Instruments. *Studies in Conservation* 63, sup1 (2018), 351–354. <https://doi.org/10.1080/00393630.2018.1499853>
- [21] Giacomo Fiocco, Claudia Invernizzi, Silvia Grassi, Patrizia Davit, Michela Albano, Tommaso Rovetta, Chiaramaria Stani, Lisa Vaccari, Marco Malagodi, Maurizio Licchelli, et al. 2021. Reflection FTIR spectroscopy for the study of historical bowed string instruments: Invasive and non-invasive approaches. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 245 (2021), 118926. <https://doi.org/10.1016/j.saa.2020.118926>
- [22] Marco Fiorucci, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, and Stuart James. 2020. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognition Letters* 133 (2020), 102–108. <https://doi.org/10.1016/j.patrec.2020.02.017>
- [23] Arjuna Flenner and Gary Hower. 2011. A Helmholtz principle approach to parameter free change detection and coherent motion using exchangeable random variables. *SIAM Journal on Imaging Sciences* 4, 1 (2011), 243–276.
- [24] Daniele Gelli, Riccardo March, Paolo Salonia, and Domenico Vitulano. 2003. Surface analysis of stone materials integrating spatial data and computer vision techniques. *Journal of Cultural Heritage* 4, 2 (2003), 117 – 125.
- [25] Cyril Goutte and Eric Gaussier. 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval*. Springer, 345–359.
- [26] Massimiliano Guarneri, Alessandro Danielis, Massimo Francucci, Mario Ferri De Collibus, Giorgio Fornetti, and Arianna Mencattini. 2014. 3D remote colorimetry and watershed segmentation techniques for fresco and artwork decay monitoring and preservation. *Journal of Archaeological Science* 46 (2014), 182 – 190.
- [27] Muhammad Suzuri Hitam, Ezmahamrul Afreen Awalludin, Wan Nural Jawahir Hj Wan Yusof, and Zainuddin Bachok. 2013. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In *2013 International Conference on Computer Applications Technology (ICCAT)*. 1–5. <https://doi.org/10.1109/ICCAT.2013.6522017>
- [28] Claudia Invernizzi, Giusj Valentina Fichera, Maurizio Licchelli, and Marco Malagodi. 2018. A non-invasive stratigraphic study by reflection FT-IR spectroscopy and UV-induced fluorescence technique: The case of historical violins. *Microchemical Journal* 138 (2018), 273 – 281. <https://doi.org/10.1016/j.microc.2018.01.021>
- [29] Koen Janssens and René Van Grieken. 2004. *Non-destructive micro analysis of cultural heritage materials*. Vol. 42. Elsevier.
- [30] Lazhar Khelifi and Max Mignotte. 2020. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *IEEE Access* 8 (2020), 126385–126400.
- [31] RP Kramer, HL Schellen, and AWM Van Schijndel. 2016. Impact of ASHRAE’s museum climate classes on energy consumption and indoor climate fluctuations: Full-scale measurements in museum Hermitage Amsterdam. *Energy and Buildings* 130 (2016), 286–294. <https://doi.org/10.1016/j.enbuild.2016.08.016>
- [32] Matthew D Li, Ken Chang, Ben Bearce, Connie Y Chang, Ambrose J Huang, J Peter Campbell, James M Brown, Praveer Singh, Katharina V Hoebel, Deniz Erdoğmuş, et al. 2020. Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *NPJ digital medicine* 3, 1 (2020), 1–9.
- [33] Jose Luis Lisani and J-M Morel. 2003. Detection of major changes in satellite images. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, Vol. 1. IEEE, I–941.
- [34] Gang Liu, Yann Gousseau, and Florence Tupin. 2019. A Contrario Comparison of Local Descriptors for Change Detection in Very High Spatial Resolution Satellite Images of Urban Areas. *IEEE Transactions on Geoscience and Remote Sensing* 57 (2019), 3904 – 3918. Issue

- 6.
- [35] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 2 (2004), 91–110.
- [36] M Ronnier Luo, Guihua Cui, and B Rigg. 2001. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 26, 5 (2001), 340–350.
- [37] Murari Mandal and Santosh Kumar Vipparthi. 2021. An Empirical Review of Deep Learning Frameworks for Change Detection: Model Design, Experimental Frameworks, Challenges and Research Needs. *IEEE Transactions on Intelligent Transportation Systems* (2021), 1–22.
- [38] Anna Maria Manfredini, Valentina Baroncini, and Cristiana Corsi. 2012. An integrated and automated segmentation approach to deteriorated regions recognition on 3D reality-based models of cultural heritage artifacts. *Journal of Cultural Heritage* 13, 4 (2012), 371 – 378. <https://doi.org/10.1016/j.culher.2012.01.014>
- [39] Geoffrey J McLachlan and Kaye E Basford. 1988. *Mixture models: Inference and applications to clustering*. Vol. 38. M. Dekker New York.
- [40] Daniel Müllner. 2011. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378* (2011).
- [41] Rocío Ortiz and Pilar Ortiz. 2016. Vulnerability Index: A New Approach for Preventive Conservation of Monuments. *International Journal of Architectural Heritage* 10, 8 (2016), 1078–1100. <https://doi.org/10.1080/15583058.2016.1186758>
- [42] Stephen M Pizer, R Eugene Johnston, James P Ericksen, Bonnie C Yankaskas, and Keith E Muller. 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*. IEEE Computer Society, 337–338.
- [43] Aleksandra Pizurica, Ljiljana Platisa, Tijana Ruzic, Bruno Cornelis, Ann Dooms, Maximiliaan Martens, Helene Dubois, Bart Devolder, Marc De Mey, and Ingrid Daubechies. 2015. Digital Image Processing of The Ghent Altarpiece: Supporting the painting’s study and conservation treatment. *IEEE Signal Processing Magazine* 32, 4 (2015), 112–122. <https://doi.org/10.1109/MSP.2015.2411753>
- [44] Mustafa Jahangoshai Rezaee, Milad Eshkevari, Morteza Saberi, and Omar Hussain. 2021. GBK-means clustering algorithm: An improvement to the K-means algorithm based on the bargaining game. *Knowledge-Based Systems* 213 (2021), 106672.
- [45] Alireza Rezaei, Emanuel Aldea, Piercarlo Dondi, Marco Malagodi, and Sylvie Le Hégarat-Masclé. 2019. Detecting alterations in historical violins with optical monitoring. In *Proceedings of the 14th International Conference on Quality Control by Artificial Vision (QCAV)*, Vol. 11172. 1117210–1 – 1117210–8. <https://doi.org/10.1117/12.2521702>
- [46] Alireza Rezaei, Sylvie Le Hégarat-Masclé, Emanuel Aldea, Piercarlo Dondi, and Marco Malagodi. 2021. Analysis of multi-temporal image series for the preventive conservation of varnished wooden surfaces. In *16th International Symposium on Visual Computing*.
- [47] Alireza Rezaei, Sylvie Le Hégarat-Masclé, Emanuel Aldea, Piercarlo Dondi, and Marco Malagodi. 2021. One step clustering based on a-contrario framework for detection of alterations in historical violins. In *25th International Conference on Pattern Recognition (ICPR2020)*. 9348 – 9355.
- [48] Alireza Rezaei, Sylvie Le Hégarat-Masclé, Emanuel Aldea, Piercarlo Dondi, and Marco Malagodi. 2022. A-contrario framework for detection of alterations in varnished surfaces. *Journal of Visual Communication and Image Representation* 83 (2022), 103357.
- [49] Amandine Robin, Lionel Moisan, and Sylvie Le Hégarat-Masclé. 2010. An a-contrario approach for subpixel change detection in satellite imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 11 (2010), 1977–1993.
- [50] Francois Rousseau, Sylvain Faisan, Fabrice Heitz, Jean-Paul Armspach, Yves Chevalier, Frederic Blanc, Jerome de Seze, and Lucien Rumbach. 2007. An a contrario approach for change detection in 3D multimodal images: application to multiple sclerosis in MRI. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2069–2072.
- [51] Tommaso Rovetta, Claudia Invernizzi, Giacomo Fiocco, Michela Albano, Maurizio Licchelli, Monica Gulmini, Gregg Alf, Alessandro Rombolà, and Marco Malagodi. 2019. The case of Antonio Stradivari 1718 ex-San Lorenzo violin: History, restorations and conservation perspectives. *Journal of Archaeological Science: Reports* 23 (2019), 443–450. <https://doi.org/10.1016/j.jasrep.2018.11.010>
- [52] Tommaso Rovetta, Claudia Invernizzi, Maurizio Licchelli, Fausto Cacciatori, and Marco Malagodi. 2018. The elemental composition of Stradivari’s musical instruments: new results through non-invasive EDXRF analysis. *X-Ray Spectrometry* 47, 2 (2018), 159–170. <https://doi.org/10.1002/xrs.2825>
- [53] Lorena Alves Santos, Karine Ferreira, Michelle Picoli, Gilberto Camara, Raul Zurita-Milla, and Ellen-Wien Augustijn. 2021. Identifying spatiotemporal patterns in land use and cover samples from satellite image time series. *Remote Sensing* 13, 5 (2021), 974.
- [54] Herb Stovel. 1998. *Risk preparedness: a management manual for world cultural heritage*. ICCROM.
- [55] Barbara H. Stuart. 2007. *Analytical techniques in materials conservation*. John Wiley & Sons.
- [56] Mirco Sturari, Marina Paolanti, Emanuele Frontoni, Adriano Mancini, and Primo Zingaretti. 2017. Robotic platform for deep change detection for rail safety and security. In *2017 European Conference on Mobile Robots (ECMR)*. IEEE, 1–6.
- [57] Garry Thomson et al. 1968. *Contributions to the London Conference on museum climatology, 18-23 September 1967*. International Institute for Conservation of Historic and Artistic Works.

- [58] Philip HS Torr and Andrew Zisserman. 2000. MLESAC: A new robust estimator with application to estimating image geometry. *Computer vision and image understanding* 78, 1 (2000), 138–156.
- [59] Ashley Varghese, Jayavardhana Gubbi, Akshaya Ramaswamy, and P Balamuralidhar. 2018. ChangeNet: A deep learning architecture for visual change detection. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 0–0.
- [60] Sagar Verma, Akash Panigrahi, and Siddharth Gupta. 2021. QFabric: Multi-Task Change Detection Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1052–1061.
- [61] Zhiqiang Wang, Zhiwen Yu, CL Philip Chen, Jane You, Tianlong Gu, Hau-San Wong, and Jun Zhang. 2017. Clustering by local gravitation. *IEEE transactions on cybernetics* 48, 5 (2017), 1383–1396.
- [62] Heidi Wirilander. 2012. Preventive conservation: A key method to ensure cultural heritage s authenticity and integrity in preservation process. *E-conservation Magazine* 6, 24 (2012), 165–176.
- [63] Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. 2018. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *Proceedings of the european conference on computer vision (ECCV)*. 654–669.
- [64] Zheng Zhang, Ping Tang, and Thomas Corpetti. 2016. Satellite image time series clustering via affinity propagation. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2419–2422. <https://doi.org/10.1109/IGARSS.2016.7729624>
- [65] Xiatian Zhu, Chen Change Loy, and Shaogang Gong. 2014. Constructing robust affinity graphs for spectral clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1450–1457.
- [66] Zhe Zhu. 2017. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 130 (2017), 370–384.