DOI: 10.1111/phor.12496

ORIGINAL ARTICLE

Detecting change in graffiti using a hybrid framework

Benjamin Wild[1] | **Geert Verhoeven**[2] | **Rafał Muszyński**[3] | **Norbert Pfeifer**[1]

[1]Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria

[2]Department of Prehistoric and Historical Archaeology, University of Vienna, Vienna, Austria

[3]IPI-URC-IMEC-UGent, Ghent, Belgium

Correspondence
Benjamin Wild, Research Unit Photogrammetry—TU Wien, Vienna, Austria.
Email: benjamin.wild@tuwien.ac.at

Abstract

Graffiti, by their very nature, are ephemeral, sometimes even vanishing before creators finish them. This transience is part of graffiti's allure yet signifies the continuous loss of this often disputed form of cultural heritage. To counteract this, graffiti documentation efforts have steadily increased over the past decade. One of the primary challenges in any documentation endeavour is identifying and recording new creations. Image-based change detection can greatly help in this process, effectuating more comprehensive documentation, less biased digital safeguarding and improved understanding of graffiti. This paper introduces a novel and largely automated image-based graffiti change detection method. The methodology uses an incremental structure-from-motion approach and synthetic cameras to generate co-registered graffiti images from different areas. These synthetic images are fed into a hybrid change detection pipeline combining a new pixel-based change detection method with a feature-based one. The approach was tested on a large and publicly available reference dataset captured along the Donaukanal (Eng. Danube Canal), one of Vienna's graffiti hotspots. With a precision of 87% and a recall of 77%, the results reveal that the proposed change detection workflow can indicate newly added graffiti in a monitored graffiti-scape, thus supporting a more comprehensive graffiti documentation.

KEYWORDS
3D modelling, change detection, colour difference, cultural heritage, digital imaging, edge-aware smoothing, feature matching, graffiti

INTRODUCTION

Project INDIGO

Contemporary graffiti and the graffiti-scapes they constitute are inherently dynamic, both spatially and temporally. Once a graffito is created, it is often subject to fast removal by authorities or other graffitists who remove the earlier work(s) or cover it with their creation(s). Suppose one belongs to the constantly growing number of scholars who classify present-day graffiti as cultural heritage (e.g., Forster et al., 2012; Ronchi, 2009; Verhoeven, Carloni, et al., 2023). In that case, these dynamics can be viewed as problematic as they imply the constant loss of cultural heritage at a large scale. That is why the past decade has witnessed several advances to mitigate this steady loss of cultural heritage (Niemann, 2022; Novak, 2015; Wogrin, 2022). Despite the increasing consensus among scholars that contemporary graffiti are worthy of documentation, there is a lack of comprehensive research data to assess and quantify its true interdependencies with society. While graffiti are omnipresent and significantly influence the appearance of our cities, potentially serving as mirrors and magnifying glasses reflecting human society, their actual significance remains largely unknown due to this data deficiency.

From September 2021 to August 2023, the research project INDIGO (INventory and DIssemination Graffiti along the dOnaukanal) took place, partly as an attempt to research this issue. While acknowledging the inherent friction and ambivalence associated with modern graffiti, the project members recognised the heritage value in these contemporary but ephemeral forms of cultural expression, similar to how ancient graffiti are valued nowadays. INDIGO aimed to introduce scientific rigour in the documentation, digital archival and dissemination practices surrounding such modern graffiti (Verhoeven, Wogrin, et al., 2023). To that end, a range of bespoke hardware and software solutions were developed, such as a thesaurus to standardise graffiti terminology through a hierarchical controlled vocabulary (Schlegel et al., 2022), a small device to accurately record a camera's exterior orientation (i.e., its position and angular rotation) whilst photographing (Wieser et al., 2022) and the automated orthorectification software AUTOGRAF (Wild et al., 2022). All these developments focused on Vienna's Donaukanal (Eng. Danube Canal), a unique urban environment where graffiti—despite being permitted at only a few spots—enjoys widespread acceptance and occurs in a density like hardly anywhere else in the world (Ringhofer & Wogrin, 2018). The Donaukanal serves as a canvas for countless new graffiti creations every day, and INDIGO tried to monitor and digitally record approximately 13 km of this ever-changing graffiti-scape.

At least once per week, one of three INDIGO photographers went along the Donaukanal to document new graffiti with photographs and spectrometer measurements (see details in Verhoeven, Wogrin, et al. (2023)). During these tours, all graffiti created since the previous tour had to be recorded. The identification of those graffiti changes primarily relied on two techniques. First, the memory and expertise of project member Stefan Wogrin who has been documenting graffiti for his SprayCity archive (https://spraycity.at) for decades. Stefan's in-depth knowledge of the Donaukanal graffiti-scape enabled him to identify all major changes, much more than any other photographer could (which explains why Stefan became INDIGO's primary photographer). Second, the team actively monitored Instagram, a popular photo and video-sharing social network where many graffitists who frequent the Donaukanal regularly post their latest creations. Tracking these updates via a mobile GIS app (ESRI's ArcGIS Field Maps; https://www.esri.com/en-us/arcgis/products/arcgis-field-maps/overview) led to valuable information on numerous new creations. To enhance this social-media-based approach, INDIGO also introduced the #INDIGODonaukanal hashtag, promoted within the creator community. Photos tagged with this hashtag could be easily filtered and often depicted new and undocumented works. During the weekly tours, all this online gathered information complemented the memory-based method.

Despite their usefulness, both methods possessed significant drawbacks. First, they required substantial time as the photographer must traverse the entire research area, carefully inspecting the walls. This undertaking posed physical challenges due to the long distances and the often unfavourable weather conditions. Furthermore, it posed a mental challenge, as the photographer had to diligently observe the walls, actively searching for changes compared to the previous photo tour. However, the major limitation of this methodology was that despite these efforts, many new graffiti went unnoticed.

Graffiti were often inadvertently overlooked because they were too tiny or not striking enough. In other cases, graffiti stayed undocumented as they appeared and got covered in between two photo tours. While this ephemerality of graffiti is a core characteristic that likely contributes to its attraction (Blanché, 2015), the absence of comprehensive digital documentation hinders thorough and impartial graffiti research. In the case of INDIGO, the digital graffiti records inherently exhibit a bias toward flashy and larger creations, which runs counter to the project's objective of capturing the entirety of the graffiti-scape (Verhoeven et al., 2022). Therefore, the need arose to improve the accuracy and efficiency of finding graffiti-relevant changes.

Status quo in graffiti change detection

One approach for automating the identification of new graffiti along the Donaukanal could be the installation of cameras that cover the relevant graffiti areas. Previous studies on graffiti change detection (Angiati et al., 2005; Di Stefano et al., 2008; Tombari et al., 2008) have successfully employed this method, focusing on detecting the act of creating a graffito rather than explicitly identifying new creations. For instance, the study conducted by Angiati et al. (2005) employed a multi-step methodology that involved extracting video frames, reducing image noise, performing feature extraction and tracking features over multiple consecutive frames. Through this method, they successfully identified individuals in front of a frequently graffitied wall and could thus implicitly retrieve graffiti change with a reported high success rate. Although this approach holds potential for achieving near-real-time detection of new graffiti, it is not practical for most graffiti-scapes, including the Donaukanal, primarily for two reasons: the implementation of numerous permanent cameras poses significant logistical and financial challenges, and there are various, presumably insuperable, legal and ethical considerations surrounding such a large-scale public surveillance approach.

Instead of installing multiple permanent cameras, a more feasible alternative is to utilise a single camera system mounted on a mobile platform, enabling a fast and repeatable capturing of photos on which graffiti-relevant changes could be detected. Image-based change detection in all possible variations was essentially incentivised by the myriad of images produced by satellite remote sensing, a trend that started with NASA's Landsat programme in the 1970s. The change detection process commonly involves image preprocessing (e.g., co-registration and denoising), applying a specific change detection algorithm, and segmenting the change detection result to arrive at a binary change map (Asokan & Anitha, 2019). The choice of change detection algorithm varies greatly depending on the specific application and available image data. In their extensive review paper, Tewkesbury et al. (2015) suggest that this task can be categorised based on the unit of analysis (e.g., pixel, kernel or object) and the employed comparison technique (e.g., arithmetic differencing, change vector analysis or transformation-based). In recent years, also the use of deep learning approaches for change detection has become popular. While many studies have shown very promising results, deep learning approaches still require large labelled data samples, and their transferability to other, particular applications can be challenging (Shi et al., 2020).

Project INDIGO acquired substantial expertise in photographing and processing extensive collections of graffiti photos (Verhoeven, Wogrin, et al., 2023). Therefore, utilising photographs as the primary data source for change detection was a reasonable choice. Equipping a mobile platform, such as a bike, with one or more cameras would make it feasible to capture photographs of the entire research area in approximately 1 h. The orientation parameters of photos taken at different time intervals could be determined using an incremental structure-from-motion (SfM) method. With the exterior orientation known for each photo, various approaches can generate co-registered images from neighbouring photographs. Although implementing this photo acquisition and co-registration strategy is a non-trivial endeavour that demands substantial expertise, careful planning and appropriate hardware, prior research (Verhoeven, Wogrin, et al., 2023; Wild, Verhoeven & Pfeifer, 2023) suggested that the main challenge in developing a graffiti change detection framework lies in the automated identification of graffiti-related changes from such preprocessed photos. The following section sheds light on the challenges that had to be addressed while developing a change detection algorithm that could effectively differentiate graffiti-related changes from other image variations within two co-registered images.

Challenges in image-based change detection of graffiti

According to Radke et al. (2005), the task of (image-based) change detection can be defined as identifying the set of pixels that differ significantly between two images of the same scene acquired at different times. An essential constraint of this definition is that changes can and usually must be distinguished between “important” and “unimportant” (Radke et al., 2005). The expected result of such change detection is usually a binary change map representing the image regions that depict relevant (i.e., important) changes compared to a given reference epoch. For this study, important change is defined as graffiti-related change.

Figure 1 exemplifies that detecting significantly differing pixels or regions between two images is trivial. For example, finding purely radiometric differences between two images can be achieved by calculating the absolute differences in grey values between them. The rationale is that significant grey-value differences indicate radiometric changes in the two input images. However, it becomes apparent that these changes primarily result from variations in inter-epoch illumination conditions rather than alterations in the graffiti cover. In addition, small noise-like structures in the input image propagate to the difference image.

Ambient light is inherently variable due to changing cloud conditions, as well as the altitude and azimuth of the sun. The latter two are especially problematic since objects close to a graffito, such as trees or pillars, often cast shadows onto the walls, causing sharp transitions between illuminated and shaded areas. Especially at lower solar incidence angles, minor structures on the wall surface can cause many small shadows (see Figure 1a). They can be considered noise and usually do not hamper the human visual system (Arnheim, 1954), but they significantly complicate change detection (Bitelli et al., 2004). While varying illumination poses the largest challenge, there are other potential reasons for unimportant yet significant differences between image pairs:

- Co-registration errors negatively affect change detection results, especially in regions with strong textural changes. Albeit usually very small (in the order of a few pixels), they are prone to cause large image differences along sharp edges.
- Differing exterior camera orientations can cause variable occlusions or record spurious reflections in the final co-registered images.
- Kinematic image acquisition might lead to uncontrollable motion blur in the input photos. Reducing the speed of the camera-carrying platform and shortening the exposure time can largely prevent, or at least mitigate, this effect.

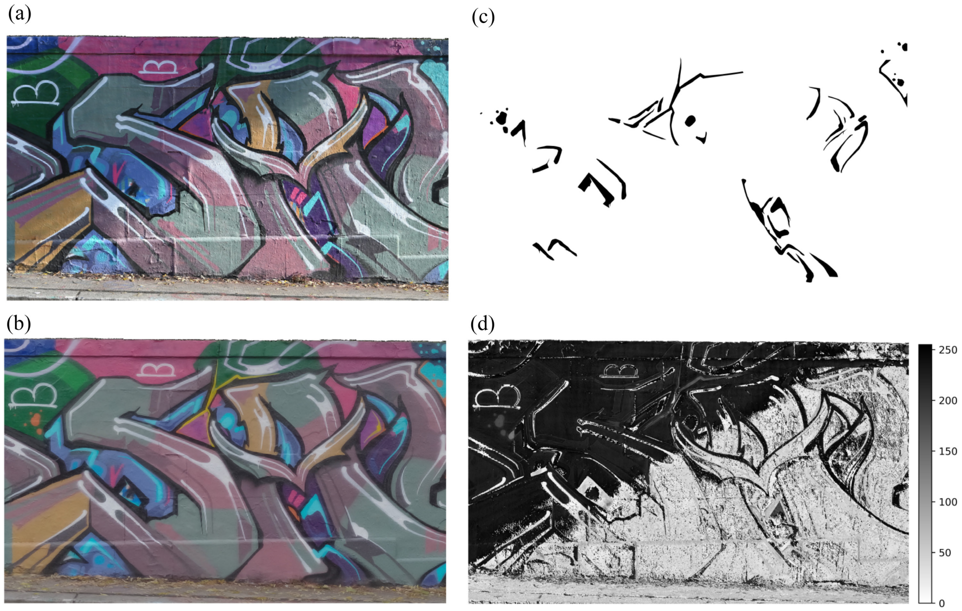


FIGURE 1 (a, b) Co-registered image pair of a scene depicting (parts of) a slightly altered graffito. (c) Manually generated change map of (a) and (b). (d) Absolute grey-value differences between (a) and (b).

- Varying cameras or camera settings can cause significant image differences. While they can theoretically be avoided by consistently utilising one camera with the same settings, this might pose problems when the environmental conditions change between two epochs. For example, one likes to capture photos with a short exposure time to mitigate motion blur. However, the resulting high ISO values needed on cloudy days will yield different amounts of image noise than the lower ISO sensitivity settings needed on a sunny day.

The aim of this study was to design, implement and test an automatic change detection method to largely replace the burdensome process of manual change detection while still overcoming the above-mentioned challenges.

DATA AND METHODS

Figure 2 depicts the workflow employed in this study. The following sections will detail the individual steps.

Test area and image acquisition

A dedicated dataset was acquired to develop and assess the change detection workflow. This dataset covered INDIGO's test zone, a stretch of about 250 m in one of the most active graffiti areas, including one of the three legal "*Wienerwände*" walls along the Donaukanal (Figure 3b). Restricting the image acquisition to this much smaller but highly active area had the advantage that imagery depicting unchanged and changed regions could be acquired in short time intervals without much effort.

The collected data consists of 29 individual photo collections (Table 1). The photos were acquired on 11 days between 21 October 2022 and 1 December 2022 using two different camera set-ups. The first two datasets (ID1 and ID2) were captured using a Nikon Z 7II (45.4 MP) camera equipped with a Nikon NIKKOR Z 20 mm f/1.8 S lens. These photographs are part of INDIGO's second total coverage survey, executed to generate a detailed 3D

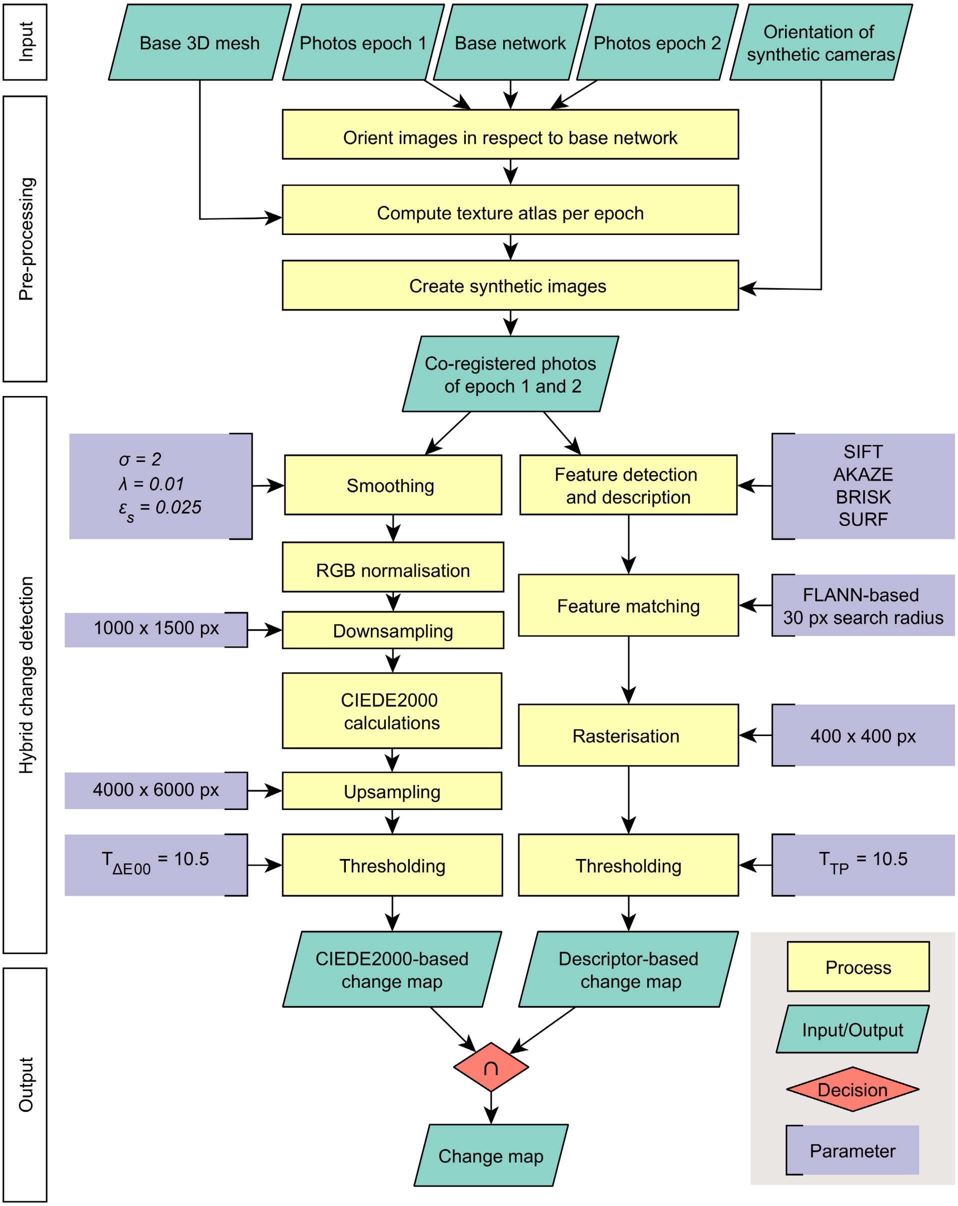
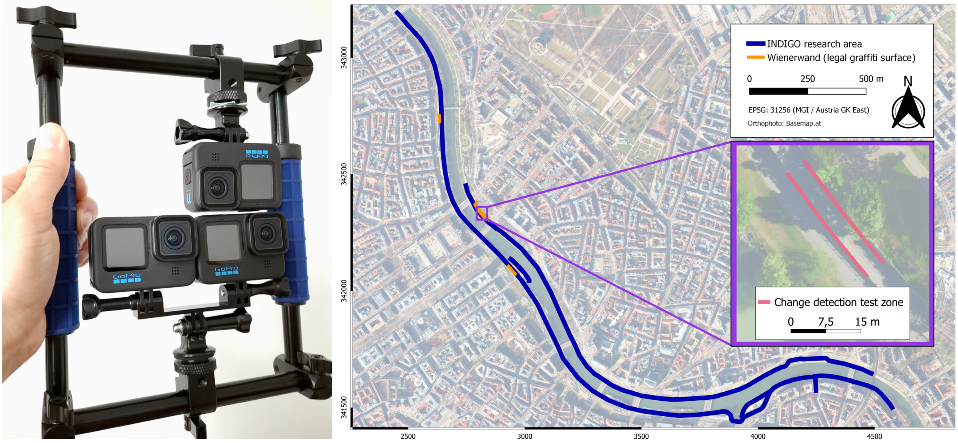


FIGURE 2 Flowchart depicting the main steps of the proposed hybrid change detection approach.

surface model of the entire research area. These photos were added to the change detection dataset to increase the graffiti and camera variability. The remaining 27 image datasets (ID3–ID29) were purposefully acquired for this change detection dataset with a triple GoPro set-up. This set-up consists of three GoPro cameras (two GoPro HERO10 Blacks and one GoPro HERO11 Black) mounted inside a frame (Figure 3a) with their optical axes mutually parallel and their centre of projection as close as possible. This set-up was used on 9 days, resulting in 27 (9 times 3) datasets.

At a walking pace of roughly 5 km/h, each GoPro camera captured two images per second using the built-in intervalometer, equating to a capture distance interval of about 0.7 m. The acquisition process for the Nikon Z 7II camera followed a stop-and-go pattern with a comparable baseline. Both camera set-ups predominantly featured



(a) (b)

FIGURE 3 (a) The three GoPro cameras mounted in a frame (left). (b) INDIGO's research area, along Vienna's Donaukanal with the designated change detection test zone depicted in the smaller inset. The legal graffiti surfaces (i.e., Wienerwände) are highlighted in orange.

a landscape orientation and had their optical axis perpendicular to the graffitied walls. Now and then, convergent images (i.e., with an inclined optical axis) and photographs for which the camera was rotated 180°, 90° clockwise or 90° anticlockwise around its optical axis were added (note that in the GoPro set-up, at least one camera was always upside down). Some of these photos also featured an increased object distance. All these variations were essential to strengthen the camera's network geometry (especially for elongated networks like those at the Donaukanal) and improve the estimation of camera orientations (Luhmann et al., 2016; Nocerino et al., 2014; Stamatopoulos & Fraser, 2014). Each of the 29 image datasets exhibits a high longitudinal overlap between the images (*c.*80%), a crucial prerequisite for the subsequent SfM pipeline.

It was explicitly sought to photograph in varying weather conditions to check the sensitivity of the change detection approach. Furthermore, different camera settings were dialled in for the three GoPro cameras (Table 1), which served two purposes: first, due to the strong influence of the camera setting on the resulting images, each of the three simultaneous acquisitions can be treated as a separate epoch for which the result (i.e., empty change maps) is known a priori. Having enough image pairs without any change is crucial for improving and validating the specificity of the developed algorithm. Second, having different combinations of camera settings and, to a lesser extent, weather conditions might give insights into their influence on image orientation estimation. This examination is valid since the GoPro's specific mounting yields a similar exterior orientation for all three cameras, so that variable cannot influence the analysis. However, this extensive research question will not be covered in this study.

Image co-registration via synthetic cameras

The cameras' orientation parameters were computed using an incremental SfM approach executed in Agisoft Metashape Professional 1.8.4. Incremental SfM initially estimates the interior and exterior camera orientations for a base network. Subsequently, each new image from the following epochs is incrementally oriented to this base network. This workflow was already successfully applied in several studies within the INDIGO project (e.g., Verhoeven, Wogrin, et al., 2023; Wild, Verhoeven, Muszyński et al., 2023). A requirement for this approach is that sufficient stable tie points can be established between the images of the different epochs. Therefore, it was crucial not to limit the photographic coverage to the graffiti-covered walls, as those could have undergone a complete

TABLE 1 The protocol for the 29 photo collections states the date, camera (and settings) and if shadows were present during the image acquisition. Note that photos acquired on the same day were shot simultaneously but with varying camera settings.

ID	Date	Camera	Sharpness	Colour	Shadows
1	21 October 2022	Nikon Z 7II	N.a.	N.a.	No
2	2 November 2022	Nikon Z 7II	N.a.	N.a.	Yes
3	12 November 2022	GoPro HERO10 Black (A)	High	Natural	Yes
4	12 November 2022	GoPro HERO10 Black (B)	High	GoPro	
5	12 November 2022	GoPro HERO11 Black	High	Flat	
6	13 November 2022	GoPro HERO10 Black (A)	High	Natural	Yes
7	13 November 2022	GoPro HERO10 Black (B)	High	Flat	
8	13 November 2022	GoPro HERO11 Black	High	GoPro	
9	14 November 2022	GoPro HERO10 Black (A)	High	Flat	No
10	14 November 2022	GoPro HERO10 Black (B)	High	GoPro	
11	14 November 2022	GoPro HERO11 Black	High	Natural	
12	17 November 2022	GoPro HERO10 Black (A)	Medium	Flat	No
13	17 November 2022	GoPro HERO10 Black (B)	Low	Flat	
14	17 November 2022	GoPro HERO11 Black	High	Flat	
15	19 November 2022	GoPro HERO10 Black (A)	Medium	Flat	No
16	19 November 2022	GoPro HERO10 Black (B)	Low	Flat	
17	19 November 2022	GoPro HERO11 Black	High	Flat	
18	22 November 2022	GoPro HERO10 Black (A)	Medium	GoPro	No
19	22 November 2022	GoPro HERO10 Black (B)	Low	GoPro	
20	22 November 2022	GoPro HERO11 Black	Low	Natural	
21	25 November 2022	GoPro HERO10 Black (A)	Low	GoPro	Yes
22	25 November 2022	GoPro HERO10 Black (B)	Medium	Natural	
23	25 November 2022	GoPro HERO11 Black	High	Flat	
24	27 November 2022	GoPro HERO10 Black (A)	Medium	Natural	No
25	27 November 2022	GoPro HERO10 Black (B)	High	Natural	
26	27 November 2022	GoPro HERO11 Black	Medium	Natural	
27	1 December 2022	GoPro HERO10 Black (A)	Medium	GoPro	No
28	1 December 2022	GoPro HERO10 Black (B)	Low	Natural	
29	1 December 2022	GoPro HERO11 Black	Medium	GoPro	

change in texture, but include elements like house façades, street lamps or bridge pillars whose geometry and texture are relatively invariant.

The images acquired with the Nikon Z 7II served as the base network. From the approximate 42 k Nikon Z7 II photos acquired during INDIGO's second total coverage, 11.5 k photographs were selected. These photos covered large parts of the graffiti-scape before and after the test zone and on the opposite bank of the Donaukanal. Their interior and exterior orientations were estimated using 179 graffiti-scape points (expressed in the MGI/Austria GK East coordinate reference system; EPSG: 31256) as constraints for the SfM bundle adjustment (see Verhoeven et al. (2022) for details). In that way, the photographs and graffiti-scape points covering the test zone were located in the middle of a much more extensive image network, likely leading to more accurate orientation values. Afterwards, a depth-map-based surface mesh was extracted from the *c.*3100 photos depicting the test zone. All vegetation, trash cans and street lamps were removed from this mesh in Geomagic Wrap 2021 (https://oqton.com/geomagic-wrap). The same software was used to fix mesh errors like small holes and non-manifold areas. Finally, a 25 m long section was cut out of this 250 m long mesh. Since that smaller mesh was still very facet-heavy, its geometry was drastically reduced using a 1 cm tolerance-based mesh decimation within Atangeo Balancer nPro (https://www.atangeo.com/buy/nPro). After running this decimated mesh again through Wrap's automated geometry fixing tools, an artefact-free and two-manifold triangular 3D surface mesh of slightly over 51 k facets was obtained (Figure 4a). From now on, this surface mesh will be called the change detection test zone mesh.

This change detection test zone mesh was reimported into Metashape Professional to compute photographic textures for each epoch (Figure 4b1, c1, d1). Before a 3D surface mesh can be textured, each of the mesh vertices needs associated 2D texture coordinates *u* and *v*. These so-called UV texture coordinates are computed by a UV parameterisation, or so-called “UV unwrap”, which essentially lays every mesh facet (cleverly or not-so-cleverly) out on the 2D texture map or texture atlas (Verhoeven, 2017). Afterwards, each pixel in the texture atlas gets an RGB value that depends on the photos observing that location in the 3D surface mesh (Figure 4b2, c2, d2). The resulting 2D RGB image can then be applied to the 3D mesh as a so-called diffuse texture. Although software for manual UV mapping exists, Metashape features an automatic and non-interactive “generic” parameterisation. Since the UV coordinates computed by this parameterisation can be used for every subsequent texturing operation, one 3D surface mesh with UV coordinates served as the basis for all diffuse texture atlases.

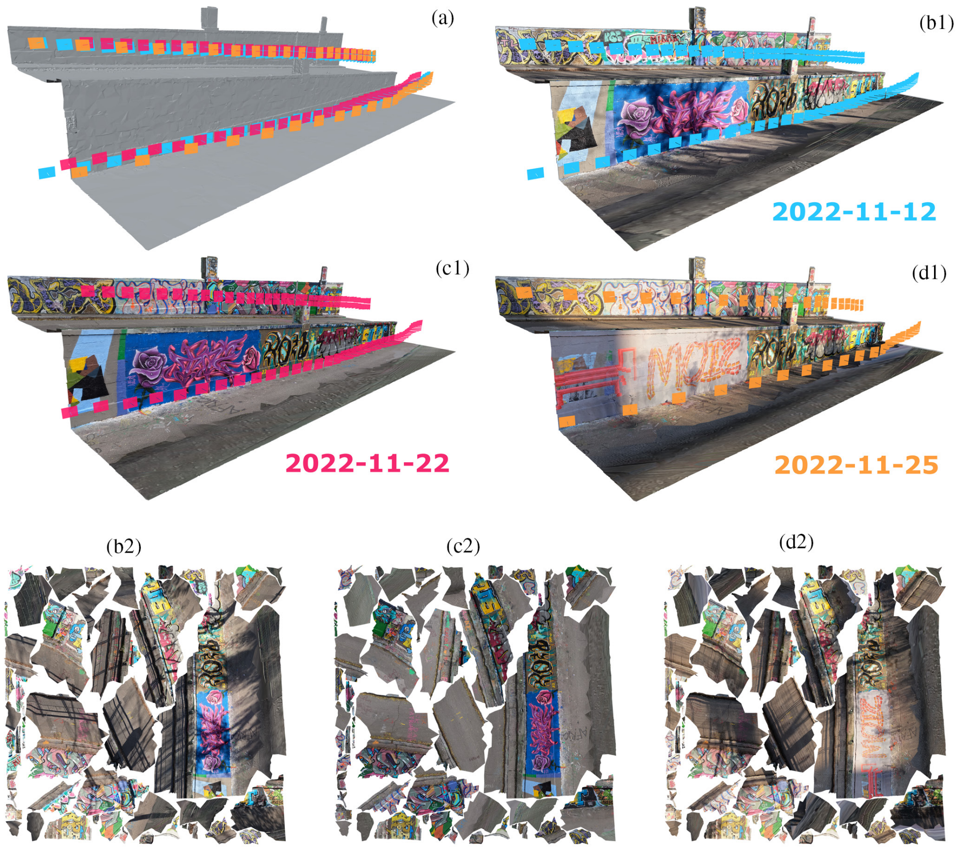


FIGURE 4 (a) The base surface mesh of the change detection test zone with the exterior camera orientations of three explanatory image datasets indicated by rectangles. The colours indicate different epochs. (b1, c1, d1) The base mesh textured using the photographs of the three different epochs (ID3, ID19 and ID22). (b2, c2, d2) The texture atlases used to generate (b1, c1 and d1), respectively.

After analysing various sizes, texture atlases of 16,384 × 16,384 pixels were computed for each of the 29 photo sets (Figure 4). Eventually, this approach yielded 29 3D models, which were geometrically identical but textured differently (i.e., their texture atlases' layout was identical, but their texture atlases' RGB values were not; compare Figure 4b2, c2, d2).

Theoretically, all the subsequent change detection approaches can be applied directly to the 29 co-aligned texture atlases. However, utilising texture atlases for change detection suffers from three main problems. First, each image is *c*.268 MP, which makes it hard to process these images effectively. Second, the layout and seams within and between texture atlases might negatively influence the change detection algorithm. Third, co-registered texture atlases require the geometry to remain unchanged over time. Although this was the case for the bespoke change detection dataset, it is an unattainable assumption for the Donaukanal, a societal hotspot where infrastructure is constantly built, removed and renewed. In such a case, an erroneous texture would occur if one would not update the geometry and texture UV parameterisation. However, updating the mesh geometry—and thus its UV texture parameterisation—destroys the texture atlases' co-alignment.

To circumvent these issues, so-called synthetic cameras were introduced. In total, 17 synthetic full-frame cameras with a distortion-free 50 mm lens were placed along the test zone at approximately equal distances to the walls (Figure 5). From each synthetic camera, a 24 MP image was rendered for every available textured 3D model. This was done using a custom Python script and yielded 493 (17 × 29) synthetic images. The ground sampling distance (GSD) of the derived synthetic images varies between 0.5 and 0.9 mm, depending on the distance between the synthetic camera and the 3D model. This GSD does, however, say little about the spatial resolution of these synthetic images, as this value also depends on the amount of spatial detail conveyed by the texture atlas. The latter is hard to determine because it is a function of the texture's pixel count, the size of the 3D mesh it covers and the UV parameterisation (which determines the amount of empty area in the texture map and its distortion). For the 25 m long test zone mesh used in this study, textures of 16,384 × 16,384 pixels proved to be the sweet spot. Whereas texture maps of 8192 × 8192 pixels resulted in photos with too little detail, texture atlases of 32,768 × 32,768 pixels showed minimal improvement in the synthetic images and were cumbersomely large.

In concluding this section, it is good to draw attention to a few crucial aspects of this workflow. First, when considering the scalability of this approach, it is possible to automate the synthetic camera placement by defining a line along which cameras are positioned at equal intervals, viewing approximately orthogonal to the wall. However, this automation may compromise the inherent advantages of our approach, namely the reduction of blind spots and the assurance that all relevant scene elements fall within the camera's field of view. Second, creating the base mesh and manually placing synthetic cameras is a one-time effort, so spending considerable time on these steps pays off. Third, the textures and synthetic images contain a few ghost-like depictions of people. Since photos were acquired via an intervalometer, avoiding passers-by appearing in the photographs was almost impossible. In two cases, even graffitists-at-work were included. Although one could implement a manual or automated people-masking step, this also runs the risk of getting holes in the texture. Lastly, errors in (interior and exterior) camera orientations and—to a lesser extent—the geometric surface model limit the accuracy in pixel alignment. However, overlaying images rendered from the same synthetic cameras shows that the average alignment error is in the order of a few pixels. While these errors might result in false positives along sharp textural transitions, they can be relatively easily mitigated by applying a morphological opening operation (see the "Change Map Merging" section).

Reference change map generation

To optimise and validate the developed change detection method, reference change maps (i.e., "ground truth") that depict all graffiti-related changes occurring between two epochs were generated. In this process, not only consecutive epochs were considered, but every possible combination was exploited to maximise the utility of the

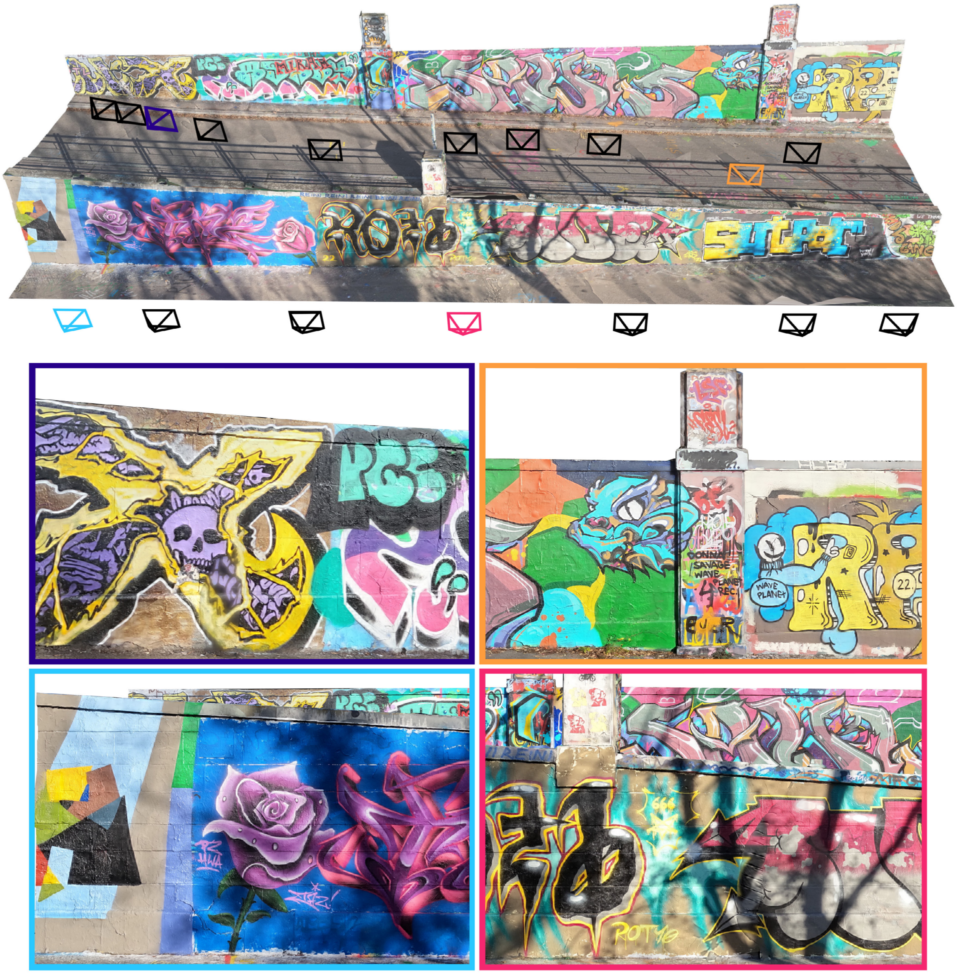


FIGURE 5 (Top) 3D surface mesh of the change detection test zone textured using the imagery acquired on 12 November with the GoPro HERO10 Black (epoch ID4). The manually placed synthetic cameras are symbolised by the black wireframe pyramids. (Below) The corresponding rendered images for the colourised synthetic cameras. To facilitate file organisation, these synthetic cameras were systematically numbered from 1 to 17. The numbering sequence commences at 1, located in the lowermost section of the first row, and proceeds until 17 in the upper right corner. For instance, the blue, pink, violet and orange rectangles correspond to synthetic cameras 1, 4, 10 and 16, respectively.

synthetic images. Given the 29 epochs for each of the 17 synthetic cameras, there are 13,804 (17×29×28) possible image pairs. This number is halved to 6902 when only considering unique pairs, where pair "AB" is equivalent to pair "BA" in terms of change.

A semi-automated approach generated a reference change map for each unique pair. This process began by collecting all images generated from the same synthetic camera into folders. That arrangement resulted in 17 folders, one folder for every synthetic camera, each containing 29 chronologically sorted images. Because not every epoch exhibits graffiti changes compared to the previous epoch, a table was made to indicate epochs where changes with respect to the previous epoch occurred. Consecutive epochs where change occurred were highlighted, as illustrated by the coloured numbers in the "ID" column of Figure 6.

Subsequently, graffiti-scape changes between highlighted change epochs and their respective previous epochs were indicated with black polygons per image pair. This manual indication of change took place in the

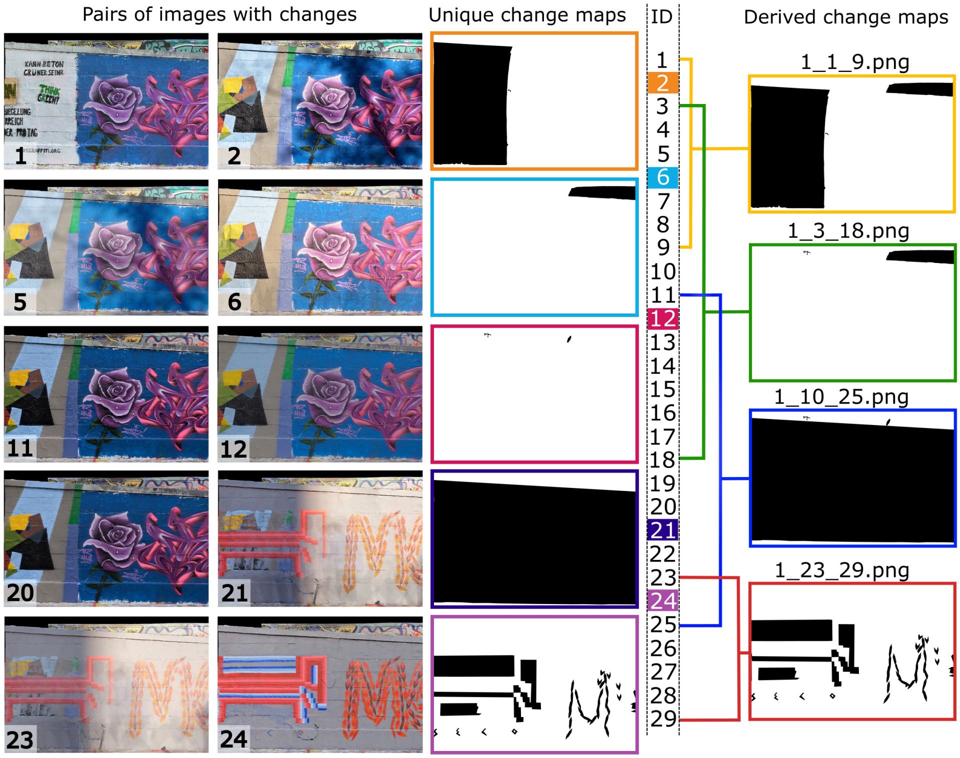


FIGURE 6 Exemplary depiction of the change map creation process for images rendered from the same synthetic camera. The first two columns depict image pairs between which graffiti-related change occurred. The numbers indicate the epoch IDs (cf. Table 1). Unique change maps (third column) were manually generated for epochs where change occurred with respect to the previous one. The coloured IDs in the fourth column indicate these epochs with change. The fifth column depicts selected change maps automatically derived from the five unique change maps. The coloured lines indicate for which epochs the change map was computed (e.g., the yellow-framed change map depicts the automatically derived graffiti-scape change between the epochs with IDs 1 and 9). The change map filenames reflect this organisational structure.

open-source vector graphics editor Inkscape (https://inkscape.org) and the image editor GIMP (https://www.gimp.org). As the first and third authors were involved in this process, a common statement to standardise the change map generation was elaborated:

> *Indicate all differences on the surface which originate from added graffiti between two epochs. This indication is done by generating polygons which cover all changed regions in the images. Because of the co-registration accuracy of the two images and potential blurriness, a positional accuracy of ca. 20 pixels is expected. The indication should start with the area-wise biggest changes and end with smaller changes and not take more than 30 min per image pair.*

This process resulted in a collection of 43 unique binary change maps. Based on these manually derived unique change maps, change maps for all other possible synthetic image pairs were automatically derived (Figure 6). This derivation was facilitated by a custom Python script that inputs the unique change maps plus corresponding IDs and outputs the change map of every possible combination. To streamline subsequent data processing, each change map was assigned a unique file name according to the following scheme: "X_Y_Z.png", where "X" is the number of the synthetic camera ranging from 1 to 17, "Y" is the ID of the reference epoch in the range 1–28 and

“Z” is the “new” epoch ID in the range 2–29. Lastly, 34 binary masks were manually generated to exclude irrelevant image parts: 17 masks indicated “no data” areas, with the other 17 identifying irrelevant image pixels from the ground or façades in the background. Image portions covered by these exclusion masks are thus also not considered in the performance analysis of the change detection approach. The entire change detection reference dataset (consisting of all 6902 synthetic image pairs, their corresponding change maps and the exclusion masks) is publicly available. For details, see the “Data Availability Statement.”

Hybrid change detection

As described earlier, implementing effective graffiti-aware image-based change detection comes with various challenges. After experimenting with existing solutions and available open-source software, which is sparse compared to the many methods developed, it became evident that relying solely on a single method would not be sufficient to handle the task's complexity. Drawing insights from prior research (Wild, Verhoeven & Pfeifer, 2023), adopting a hybrid approach and leveraging the strengths of two largely independent methods proved promising and was optimised for this study.

This hybrid change detection method will be detailed as follows: the “Pixel-based Change Maps” section introduces a new pixel-based change detection approach, while the “Descriptor-based Change Maps” section summarises a recently introduced descriptor-based method (Wild, Verhoeven & Pfeifer, 2023). These methods are then combined to create a hybrid change detection pipeline (see the “Change Map Merging” section, with additional details in the “Threshold Optimisation” section).

Pixel-based change maps

Image smoothing

As discussed in the “Status Quo in Graffiti Change Detection” section, a significant challenge in image-based graffiti change detection is the presence of noise-like structures on the surfaces bearing the graffiti. These structures can lead to small yet noticeable shadows under specific solar altitudes and azimuths. Mitigating this noise prior to applying pixel-based graffiti change detection methods would likely improve the change detection results. While conventional image smoothing techniques like the mean filter or Gaussian blurring effectively remove noise, they also blur all image edges, which is an undesirable side effect given the many sharp and thin lines many graffiti are composed of (e.g., Figures 2 and 5).

The present study addressed this issue by employing the edge-aware smoothing algorithm introduced by Xu et al. (2012). The algorithm uses a relative total variation metric to quantify spatial disparities along the image's x and y directions within designated windows. This approach enables the differentiation between an image's structural parts (mainly corresponding to larger image objects) and textural components (fine-scale details, typically with some periodicity) without requiring any prior assumptions about the regularity or symmetry of the underlying textural patterns (Xu et al., 2012). Consequently, this technique facilitates the smoothing of fine details while preserving the salient edges of the input image, allowing for a more refined subsequent analysis. Without going into detail, four parameters controlling the texture–structure separation and smoothing need to be defined (Xu et al., 2012):

- σ specifies the maximum size of texture elements and controls the algorithm's ability to separate texture from structure. The proposed range for σ is [0, 8]. Lower sigma values increase processing times and leave more spatial detail untouched.
- λ controls the smoothness of the result. High values result in smoother results and longer processing times. The proposed range for λ is [0.01, 0.03].

- ε_s controls the sharpness of the final results. Smaller values lead to sharper results but longer processing times. The proposed range for ε_s is [1e−3, 0.03].
- The number of iterations. Larger values yield longer processing times and more smoothing. However, smoothing increasingly diminishes per iteration.

Of those parameters, the scale parameter σ and the strength parameter λ are the most important to consider with respect to the size of the image. After running various combinations and quantifying their impact on the final results—reported in the "Pixel-based Results" section—the final parameter values used for this smoothing step were $\sigma = 2$, $\lambda = 0.01$, $\varepsilon_s = 0.025$ and two iterations. However, these parameters only provided satisfying smoothing results and acceptable running times when applied to images of 1000 × 1500 pixels (i.e., around 2–4 s on a fast processor using the MATLAB code provided by Xu et al. (2012)). Running the edge-aware smoothing on the 24 MP synthetic image of 4000 × 6000 pixels would lengthen that by a factor of 16 to over half a minute.

That is why a down- and an upsampling step were part of this smoothing operation. First, each synthetic 24 MP image was downsampled by a factor of 16 via a nearest neighbour interpolation. This 1.5 MP image was smoothed and subsequently resized to 24 MP, again via nearest neighbour interpolation. Nearest neighbour interpolation was chosen since the more ubiquitous bicubic interpolation can produce pixel values outside the original range. Although bilinear interpolation creates less jagged edges than nearest neighbour interpolation, it can make small features look fainter than nearest neighbour interpolation. However, both interpolations produce the same values for the large areas of quasi-identical pixel values which result from the smoothing algorithm. Although the influence of the down- and upscaling interpolation method on the change detection outcome remained unassessed in this study, the fact that these algorithms mainly treat edges and small features slightly differently makes it doubtful that another interpolation would have meaningfully influenced the results.

Illumination normalisation

Another main challenge is dealing with varying inter-epoch illumination differences, which can cause significant differences in the image space even when no graffiti-related change occurs. To achieve approximate illumination invariance in a fast manner, the smoothed synthetic images were RGB normalised via division of the individual red (R), green (G) and blue (B) image channels by the sum of all three channels for every pixel (Finlayson et al., 1998; Kampel et al., 2010):

$$R_n = \frac{R}{R+G+B}, G_n = \frac{G}{R+G+B}, B_n = \frac{B}{R+G+B}.$$

The resulting normalised red (R_n), green (G_n) and blue (B_n) pixel values are *rgb* chromaticity coordinates, and their sum always equals 1. Figure 7 depicts the RGB-normalised version of the same scene receiving direct sunlight (upper row) and diffuse illumination (lower row). The normalisation step effectively removed all shadows and retained the colour chromaticity. However, this normalisation does not account for the dominant colour of the illumination (or, in colorimetric terms, the illumination's correlated colour temperature). The upper right inset of Figure 7 illustrates this. Shade contains a larger proportion of blue light than direct sunlight. That is why shaded image regions feature bluer pixels after RGB normalisation.

CIEDE2000 colour differences

Once all images were smoothed and RGB normalised, pair-wise image changes were computed via CIEDE2000-based colour differencing. Colour differencing is used in applications that need to know how far specific colours deviate from their ideal values (e.g., a digital photograph vs. the real-world object, colour on screen and in print). When expressing colour differences, the CIE 1976 $L^*a^*b^*$ or CIELAB colour space is typically used, an approximately perceptually uniform colour space adopted in September 1975 by the Colorimetry Committee of the *Commission Internationale de l'Éclairage* (CIE; Eng. International Commission on Illumination) (McLaren, 1976).

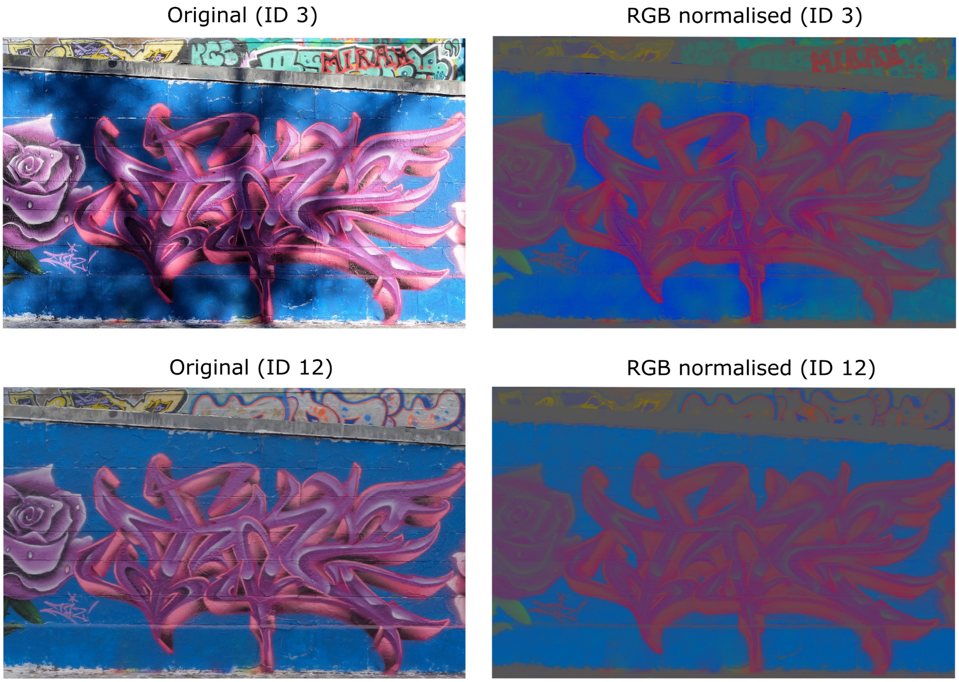


FIGURE 7 Results from RGB normalisation. The first column depicts input images recorded under strongly differing illumination conditions: direct sunlight resulting in strong shadows due to trees in front of the wall (upper left; ID3) and cloudy conditions (lower left; ID12). The corresponding RGB-normalised images are shown in the second column.

This three-dimensional colour space decomposes colour into its luminance and chrominance components: the former is represented by the lightness scale (L^*), while the latter is reflected in the a^* and b^* axes, respectively corresponding to the red-green opponent hues and the yellow-blue opponent hues (International Organization for Standardization, 2019). CIE $L^*a^*b^*$ is relatively perceptually uniform, meaning that every unit change in these three axes approximately corresponds to a just noticeable difference for the human visual system (Sharma, 2003).

A colour difference in CIE $L^*a^*b^*$ is expressed as ΔE. The symbol Δ means a change in variable E, which references the German word *Empfindung* or "sensation" in English. How this difference is calculated depends on the colour-difference formula, each providing subtly different variations of ΔE. The ΔE_{76} formula stems from 1975 (McLaren, 1976). However, since it quantifies the colour difference as the 3D Euclidian distance between the $L^*a^*b^*$ coordinates of the sample and the reference colour, it could not deal well with saturated colours for which CIELAB is no longer perceptually uniform. Instead of developing a new colour space, CIE addressed these perceptual non-uniformities via new weighted colour-difference formulae still based on the CIELAB colour space: CIE94 (Commission Internationale de L'Éclairage, 1995) and the currently CIE-recommended colour-difference formula CIEDE2000 (Commission Internationale de l'Éclairage, 2001), adopted as the joint ISO/CIE 11664–6:2022 International Standard (International Organization for Standardization, 2022). Compared to the CIE94 differencing standard, CIEDE2000 incorporated five corrections to more adequately correlate with human visual perception (International Organization for Standardization, 2022; Luo et al., 2001). Due to the complexity and length of the CIEDE2000 formulae, an explicit description is omitted for conciseness and clarity. However, all details can be found in the official CIE and ISO documents.

This change detection approach leverages CIEDE2000-based colour differencing. First, the smoothed and illumination-normalised synthetic images are converted from their sRGB colour space into the CIE $L^*a^*b^*$ colour space using a CIE D65 standard illuminant as a reference white point. Then, a pixel-wise ΔE_{00} or CIEDE2000 colour difference is computed between both images. Finally, turning these continuous colour-difference values into a binary change map necessitates a suitable threshold $T_{\Delta E_{00}}$. The procedure to find this threshold is covered in the "Threshold Optimisation" section.

Descriptor-based change maps

In addition to pixel-level change detection, descriptor-level analysis is another valuable and largely independent approach to identify changes. This method does not consider individual pixels but examines pixel neighbourhoods and their statistical characteristics. Different studies experimented with various feature descriptors as change indicator (e.g., Pillai et al., 2017; Seo et al., 2022; Wild, Verhoeven & Pfeifer, 2023). They found good performance of descriptor-based change maps, which generally contained less noise than pixel-based approaches but had problems detecting change on featureless objects. This study builds upon the methodology introduced by Wild, Verhoeven and Pfeifer (2023) which used the spatial density of matched descriptors between two images to filter unchanged image regions. The fundamental rationale behind this approach is that similar features at similar locations in two accurately co-registered images are strong indicators for unchanged regions around these detected features.

Wild, Verhoeven and Pfeifer (2023) provide a comprehensive explanation of this descriptor-based change detection method, so the following paragraph only summarises its three consecutive phases:

1. In both synthetic input images (which are not smoothed or normalised), feature points are extracted. Those points are usually clustered along distinctive textural or structural features, such as edges and corners. A fixed-length vector (i.e., descriptor vector) can describe each feature point. For this study, four well-established feature detectors and descriptors were used: SIFT (Lowe, 1999), SURF (Bay et al., 2006), AKAZE (Alcantarilla, 2013) and BRISK (Leutenegger et al., 2011).
2. The extracted descriptors are matched using FLANN (Fast Library for Approximate Nearest Neighbours), a collection of nearest neighbour search algorithms optimised for finding pairs of high dimensional descriptors with minimal distances. Falsely matched descriptors are filtered twofold: first, a Lowe ratio test (Lowe, 2004) is conducted to filter out matches based on insufficiently discriminative descriptor vectors. Specifically, the distances of the two closest matches are compared, and if their ratio (i.e., Lowe ratio) is smaller than a certain threshold, the match is removed. Here, the Lowe ratio was set to 0.7. Second, the nearly pixel-perfect co-registration is exploited by eliminating matched points with image coordinates which deviate by more than 30 pixels, corresponding to *c.*4 cm on the object.
3. The resulting matched feature points, commonly referred to as tie points, are used as a proxy for unchanged image regions. Therefore, the image is overlaid with a quadratic grid with a cell size of 400 pixels, resulting in 150 cells [(6000/400) + (4000/400)]. Each cell of the resulting raster is assigned the number of tie points within the respective cell boundaries. These 150 grid values are then upsampled using bilinear interpolation to achieve smoother transitions between the cells and match the input images' original pixel count (6000 × 4000 pixels). The resulting image depicts the tie point densities and by applying a threshold (T_{TP}) a binary change map can be derived.

Change map merging

Up to this point, two preliminary outcomes were obtained per synthetic image pair: the smoothed and RGB-normalised CIEDE2000 colour differences ΔE_{00} on a pixel-by-pixel basis, and the tie point density image. Both are solid indicators for change and no change, respectively. However, suitable thresholds must be established to convert these outcomes into a binary change map. This vital step is covered in the following section. After deriving the binarised versions of the pixel-based (CM_{pix}) and descriptor-based (CM_{des}) change maps, they were intersected via a logical AND (∩) operation to yield one merged binary change map (CM):

$$CM = CM_{pix} \cap CM_{des}.$$

This operation ensured that only image regions classified as “change” in both maps were considered graffiti-relevant alterations in the final change map. Finally, morphological opening was performed using a kernel size of 15×15 pixels to reduce potentially remaining noise and remove artefacts resulting from minor co-registration errors.

Threshold optimisation

As indicated above, thresholds must be applied to create binary change maps from the intermediate results. Specifically, a threshold for the CIEDE2000 differences ($T_{\Delta E_{00}}$) and for the tie point densities $\left(T_{TP}\right)$ must be set. A training dataset of 50 synthetic image pairs, selected randomly from all 6902 pairs, was used to determine the most suitable thresholds. The resulting threshold-specific change maps were systematically evaluated against the corresponding reference change maps using the following metrics:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F1\,\text{score} = \frac{2 * \text{PRECISION} * \text{RECALL}}{\text{PRECISION} + \text{RECALL}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Correct classification rate (CCR)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where TP/TN denote true positives/negatives and FP/FN denote false positives/negatives. These metrics were computed for every pair of threshold-specific and reference change maps. Afterwards, a threshold-specific median was calculated for each metric and used for threshold optimisation. In the end, $T_{\Delta E_{00}}$ was defined at 10 and T_{TP} was set at 10.5. The “Threshold Optimisation and Change Map Merging” section provides more details on how these results were obtained.

RESULTS AND DISCUSSION

This part provides more insight into the choice for (and influence of) the parameter values of the pixel- and feature-based change detection components (in the “Pixel-based Results” and “Descriptor-based Results” sections, respectively). Afterwards, the “Threshold Optimisation and Change Map Merging” section sheds more light on the combined threshold optimisation and merging of the change maps. Finally, the “Hybrid Change Detection Assessment” section assesses the performance of the entire hybrid change detection pipeline.

Pixel-based results

The impact of the edge-aware smoothing was visually assessed by comparing the smoothed results against the original input (Figure 8). The comparison shows its potential to enhance change detection outcomes by eliminating noise caused by intricate geometric features on the walls. The extent of smoothing varies considerably based on the image pixel count and the parameter values of the smoothing algorithm, mainly the scale parameter σ and the strength parameter λ. Figure 8 illustrates algorithm outputs using two different sets of parameter values. In the milder variant ("weak smoothing"), some smaller structures persist, whereas they are nearly entirely filtered out in the "strong smoothing" images. However, it becomes evident that "strong smoothing" also eliminates certain textural elements, such as thin lines, which would, in return, remain undetected by the change detection algorithm.

To determine favourable smoothing parameters for the current change detection task, we conducted a quantitative analysis comparing differently smoothed RGB-normalised CIEDE2000 change maps with the reference change maps generated for the randomly created sample dataset of 50 synthetic image pairs (i.e., the same 50 image pairs used in the "Threshold Optimisation" section as a training dataset). The results revealed that smoothing had a small yet positive impact on pixel-based change detection outcomes (Table 2). In this evaluation, the "weak smoothing" approach consistently yielded the best results, except for the specificity, which was slightly higher for the original input variant. While these differences are relatively small, they affirm that smoothing, on the whole, positively influenced change detection results. Furthermore, the visually enhanced interpretability of the results due to reduced noise levels, coupled with only marginal increases in processing times, supports the adoption of "weak smoothing" procedures for all subsequent analyses. For this analysis of smoothing parameters, the threshold for CIEDE2000 differences ($T_{\Delta E_{00}}$) was set to 21, which delivered visually pleasing results for the training dataset.

A visual examination of exemplary results (Figure 9) shows promising results regarding the RGB-normalisation following the smoothing operation. The smoothed RGB-normalised synthetic images (Figure 9b1, b2) exhibit a high degree of illumination invariance while preserving the chromatic characteristics of the graffiti well. Both traits

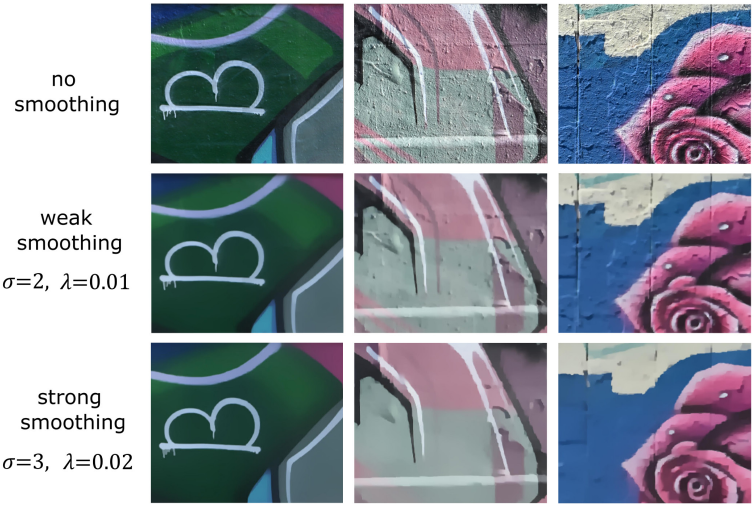


FIGURE 8 First row: three 1200 × 1000 pixel snippets extracted from different synthetic images. Second and third rows: the same snippets after applying edge-aware smoothing with different values for the scale parameter σ and the strength parameter λ. Both sets used $\varepsilon_s = 0.025$ for two iterations. In agreement with image smoothing, these snippets were downscaled 16 times before smoothing. The smoothed results were then upscaled to reach the initial pixel count again. Both resizing processes applied nearest neighbour interpolation.

TABLE 2 Results from the pixel-based change detection with different parameters used for the edge-aware smoothing. All metrics are based on the 50 randomly selected synthetic image pairs.

Smoothing	Parameters used	F1-score [%]	Recall [%]	Precision [%]	CCR [%]	Specificity [%]
No smoothing	N.a.	65	73	73	85	89
Weak smoothing	$\sigma = 2; \lambda = 0.01$	69	76	78	87	88
Strong smoothing	$\sigma = 3; \lambda = 0.02$	67	74	75	88	88

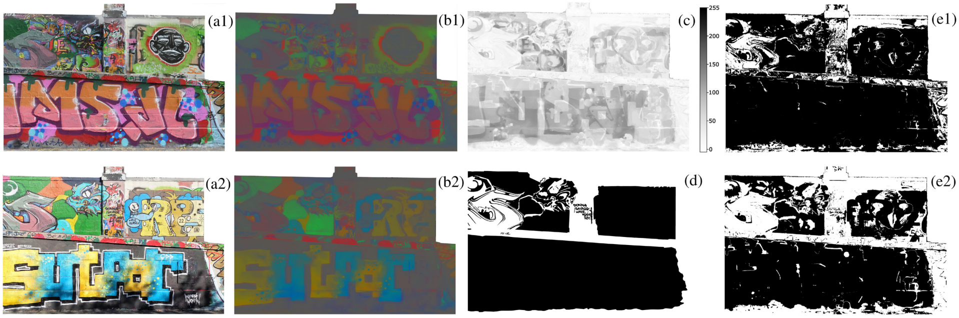


FIGURE 9 Pixel-based processing chain example. (a1, a2) Input image pair (epoch ID1 and ID3). (b1, b2) Respective smoothed and RGB-normalised variants. (c) Colour differences based on CIEDE2000. (d) Reference change map. (e1, e2) Pixel-based change maps derived using two different thresholds: $T_{\Delta E_{00}} = 10$ (e1) and $T_{\Delta E_{00}} = 21$ (e2).

suggest that this straightforward colour space transformation can effectively overcome illumination variances. However, in addition to the method's inability to deal with different illumination colours (see the "Illumination Normalisation" section), another weakness of the RGB-normalisation was identified: pixels with similar shares of R, G and B (i.e., white, grey and black pixels) become hardly distinguishable in the RGB-normalised colour space and are thus likely to be overlooked (Kender, 1976). This phenomenon is exemplified in Figure 9a1. In this synthetic image, the black facial elements in the upper right corner exhibit similar RGB-normalised values to the white and grey backgrounds.

Nevertheless, the resulting CIEDE2000 differences (depicted in Figure 9c) are still relatively large at that location due to the vivid colours used in the "new" graffito. However, this change might have gone unnoticed if the initial black graffito had been covered with white, grey or black paint. Achromatic areas like these can adversely impact the change detection results. Colours with lower chromaticity often serve as a graffiti base layer, thus constituting a significant portion of a graffito. One might argue that failing to detect changes in the graffito background is less critical, as it typically does not affect the potential meaning conveyed by the graffito. However, graffiti creators and scholars may not support this argument universally. From the perspective of graffiti documentation, failing to identify graffito background changes due to RGB normalisation remains acceptable, as new graffiti can usually still be identified, albeit not to their full extent.

Furthermore, Figure 9e1, e2 shows that the binary pixel-based change maps vary substantially with the CIEDE2000 difference threshold $T_{\Delta E_{00}}$. The choice of $T_{\Delta E_{00}}$ is a trade-off between detecting too much change and failing to detect changes. This trade-off is expressed in terms of the recall, precision and F1-score introduced in the "Threshold Optimisation" section. A suitable value for $T_{\Delta E_{00}}$ was found with the help of the previously mentioned training set of 50 synthetic image pairs and their reference change maps. In this procedure, 16 different (integer) thresholds, ranging from 9 to 24, were used to binarise the CIEDE2000 difference images, yielding 16 binary change maps for each synthetic image pair. The median metrics for each threshold were computed using the generated change maps and the reference data (see Figure 9d for an example). These metrics clearly showed the

TABLE 3 Metrics obtained from the 50-pair sample dataset for the pixel-based change detection component.

Method	Threshold	F1-score [%]	Recall [%]	Precision [%]	CCR [%]	Specificity [%]
RGB-norm ΔE^*	$T_{\Delta E_{00}} = 21$	69	77	77	87	88

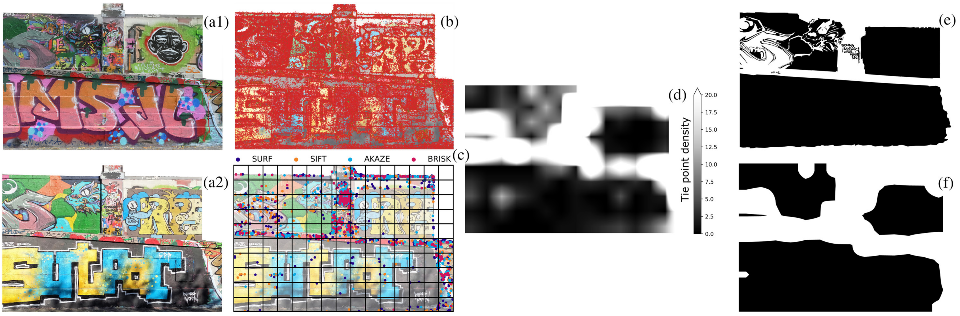


FIGURE 10 Descriptor-based processing chain example. (a1, a2) Input image pair (epochs ID1 and ID3). (b) Image A1 overlaid with extracted feature points (using SURF, SIFT, AKAZE and BRISK). (c) Matched and filtered tie points overlaid with a grid for tie point density calculations. (d) Tie point densities derived from panel (c) using linear interpolation. (e) Reference change map between a1 and a2. (f) Descriptor-based change map using a tie point density threshold (T_{TP}) of 11.

expected dependencies: with increasing $T_{\Delta E_{00}}$, fewer pixels were falsely detected as change (increasing precision), but at the cost of overlooking actually changed pixels (decreasing recall). For this study, the optimal threshold $T_{\Delta E_{00}}$ was defined to be at the intersection of recall and precision. For the training dataset, this occurred at $T_{\Delta E_{00}} = 21$, where both precision and recall reached 77% (Table 3).

Descriptor-based results

The effectiveness of the descriptor-based approach is most pronounced for objects exhibiting substantial textural variations, resulting in the extraction of numerous feature points (Figure 10). Two synthetic images depicting graffiti comprised of small text elements can be matched relatively easily and robustly, resulting in a dense distribution of tie points and, consequently, a higher likelihood of detecting "no change" areas. Nonetheless, these regions pose the inherent risk of concealing real changes, as demonstrated in the upper middle part of Figure 10c. This limitation can be mitigated by adjusting either the grid cell size or reducing the threshold T_{TP}. Since such adjustments may lead to an increased number of false-positive detections, selecting an appropriate threshold is critical. However, such a threshold selection—and the evaluation of this method—only makes sense in conjunction with the pixel-based approach because the descriptor-based change detection method alone does not facilitate identifying potential changes in image regions with little to no texture.

Threshold optimisation and change map merging

The "Pixel-based Results" and "Descriptor-based Results" sections clarified that thresholds must be applied to create binary change maps from the intermediate pixel-based and descriptor-based change detection results. Even though a suitable threshold $T_{\Delta E_{00}}$ for the pixel-based component of this graffiti change detection approach was defined in the "Pixel-based Results" section, this value discarded any descriptor-based information. In other words, optimally merging the complementary pixel- and descriptor-based results into one binary change map *CM*

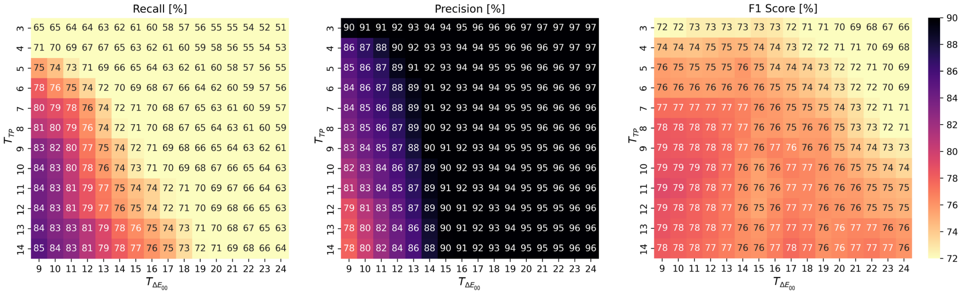


FIGURE 11 The tested tie points (T_{TP}) and CIEDE2000 ($T_{\Delta E_{00}}$) thresholds with their corresponding assessment values. These assessment values represent metric-specific median values computed via the thresholds-specific hybrid change maps and their reference change maps for the 50 image pairs of the sample dataset.

will necessitate a new threshold for the pixel-based change detection component CM_{pix} and enable the determination of a suitable threshold for the descriptor-based component CM_{des}.

Identical to the procedure outlined in the "Pixel-based Results" section, 16 integer CIEDE2000 thresholds from 9 to 24 were used to generate pixel-based change maps for each of the 50 synthetic image pairs in the training dataset (see the "Threshold Optimisation" section). For the descriptor-based approach, 12 change map threshold values ranged from 3 to 14. Merging all possible combinations of both change detection components yielded 192 (i.e., 16×12) binary hybrid change maps for each of the 50 image pairs. These change maps were individually evaluated against their corresponding reference change map using the precision, recall, F1-score, specificity and CCR metrics. Afterwards, metric-specific medians were computed and listed in a large matrix for each of the 192 threshold combinations (see Figure 11). In this way, the optimal thresholds turned out to be 10 or 11 for the tie point density limit and 10 for the CIEDE2000 colour difference. With $T_{TP}=10.5$ and $T_{\Delta E_{00}}=10$, the resulting combined change maps' precision and recall intersected at 83% and F1-score peaked at 79%. Figure 11 also revealed two dominant patterns:

1. Increasing T_{TP}, increases recall and decreases precision.
2. Increasing $T_{\Delta E_{00}}$ has the opposite effect, with decreasing recall and increasing precision.

The threshold evaluation also clearly showed the benefit of the descriptor-based method. With a very low T_{TP}, recall and F1-score dropped substantially. For example, with $T_{TP} = 3$, the maximum F1-score reached 73%, and the recall is only 65%. Selecting thresholds inherently involves a trade-off between detecting excess changes and failing to detect changes. This study's primary objective was to balance these two competing objectives. As more experience is gained, consideration may be given to adjusting threshold values to emphasise identifying true positives, even if it comes at the cost of accepting more false negatives or vice versa.

Hybrid change detection assessment

Assessing the entire image-based graffiti change detection method occurred in two different ways. First, the hybrid change detection pipeline with all optimal parameter values and thresholds was applied to all 6902 pairs of synthetic images. The reference change maps for each pair enabled quantitative assessment of these results via the metrics detailed in the "Threshold Optimisation" section. Second, the values of these performance metrics were compared to those obtained for graffiti change detection via Iteratively-Reweighted Multivariate Alteration Detection (IR-MAD; Nielsen, 2007; Nielsen et al., 1998), another frequently used change detection method (Tewkesbury et al., 2015). IR-MAD detects change by robustly identifying uncorrelated information indicative of

TABLE 4 Optimised thresholds, average runtime (on a 16-core INTEL i9-12900K processor) and metric-specific median results (with their median absolute deviations) for the three tested change detection algorithms. All metrics are based on the entire synthetic image set of 6902 pairs.

Method	Threshold(s)	F1-score [%]	Recall [%]	Precision [%]	CCR [%]	Specificity [%]	Average runtime per pair [s]
Hybrid	$T_{\Delta E_{00}} = 10; T_{TP} = 10.5$	80 ± 13	77 ± 11	87 ± 13	95 ± 5	98 ± 2	16.3 [11.6 (pixel-based) + 4.7 (descriptor-based)]
IR-MAD	Dynamic (*K*-means)	51 ± 25	69 ± 17	81 ± 17	64 ± 14	66 ± 14	17.1
IR-MAD	$T_{\text{IR-MAD}} = 28$	48 ± 32	81 ± 19	81 ± 17	98 ± 1	99 ± 0	7.2

change through canonical correlation analysis on multi-band images. The canonical correlations are subtracted to identify potential regions of change highlighted as MAD components, and then a threshold is applied to binarise the result. For this comparison, two sets of binary change maps were created via different thresholds:

- The first set of IR-MAD change maps were derived with a dynamic per-image threshold determined by *K*-means clustering ($K = 2$ to represent change vs. no change).
- The second set relied on a static threshold, whose optimal value was determined via the procedure detailed in the "Pixel-based Results" section. For this threshold optimisation, the same 50 image pairs training dataset was used as for the optimisation of the hybrid approach.

An advantage of IR-MAD is its resistance to linear scaling, which makes it robust to different illumination and sensor conditions (Nielsen, 2007). This assessment used the IR-MAD implementation from Chen Hongruixuan's GitHub repository (Hongruixuan, 2021).

Table 4—which provides all relevant metrics for the three change detection approaches—reveals a high overall performance with, compared to the results obtained with IR-MAD, relatively low variability for the newly developed hybrid approach. Compared to the metric values for the training dataset (Figure 11), there is a slight decline in recall, a marginal increase in F1-score and a few percentages higher precision. However, the values for these metrics are all very close (certainly given their variabilities) to those obtained during threshold tuning, indicating the representativity of the training data set. The very high CCR and specificity values confirm the method's general applicability for the task at hand, further underscored by the visualisation of the exemplary results (Figure 12).

However, there are also a few working points. The visual appearance of the results is sometimes not directly reflected in the obtained metrics, which might indicate a relatively good change map that turns out to look not very pleasing (e.g., Figure 12f4). Due to the relatively coarse initial tie-point-based change map, the current method also often overlooked smaller graffiti, such as tags (e.g., Figure 12f1). In contrast, this hybrid approach performed very well on completely unchanged scenes (e.g., Figure 12f3), mirrored in the achieved specificity values (Tables 4 and 5).

Comparison with the established IR-MAD method showed that the hybrid method outperformed IR-MAD when applying *K*-means dynamic thresholding (Table 4). All metrics were substantially worse for this IR-MAD binarisation variant, and their variabilities were higher. This phenomenon can be attributed to the inherent limitation of *K*-means-based thresholding when applied to completely unchanged image pairs, as illustrated in Figure 12d3. However, with an optimised IR-MAD threshold ($T_{\text{IR-MAD}} = 28$)—established via the same threshold optimisation procedure used for the pixel-based approach—one can observe a similar performance between IR-MAD and the proposed hybrid method, particularly regarding recall, precision, CCR and specificity. It is important to note that the IR-MAD results exhibited a significantly higher variability, as evidenced by the large median absolute deviations for recall, precision and F1-score. This substantial variability also accounts for the notable difference between F1-score and recall/precision.

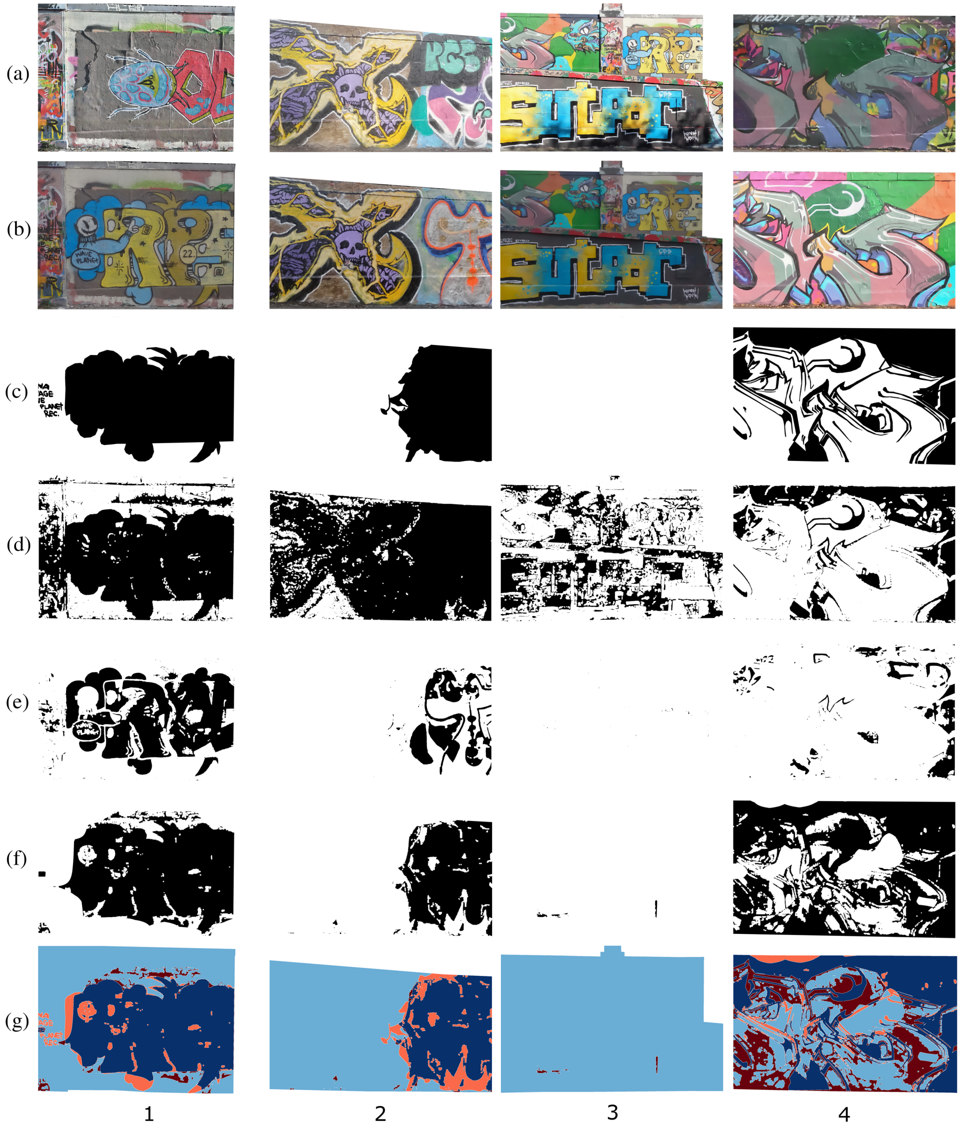


FIGURE 12 Four exemplary results (columns 1–4) with the respective synthetic input pairs (a and b), reference change maps (c), IR-MAD results with *K*-means thresholding (d), IR-MAD results with a static threshold of $T = 28$ (e), hybrid change detection results (f) and colourised comparison of the hybrid results with the reference map (g); (light) blue denotes true (negatives)/positives and (light) red denotes false (negatives)/positives.

Regarding computation times, it is worth noting that the introduced hybrid approach (16.3 s per image pair) exhibits higher computational demands than the IR-MAD variant with a static threshold (7.2 s per image pair). This difference is primarily due to the hybrid method's additional steps, like smoothing and feature point extraction plus matching, in addition to the computational complexity of the CIEDE2000 formulas. However, it is essential to emphasise that the processing times of the proposed hybrid approach remain well within acceptable limits for the designated task and research area.

To put this into perspective, consider the following rough estimation: each image pair covers approximately 5 m of graffiti-covered wall, and there are approximately 13 km of such surfaces in the Donaukanal study area

TABLE 5 Obtained metrics from the four examples depicted in Figure 12. For each example the best and worst metric values are highlighted in blue and red.

Method	Example #	F1-score [%]	Recall [%]	Precision [%]	CCR [%]	Specificity [%]
Hybrid	1	84	88	81	83	77
IR-MAD (*K*-means)		79	66	99	81	99
IR-MAD ($T_{\mathrm{IR-MAD}} = 28$)		87	95	81	85	75
Hybrid	2	86	76	98	91	99
IR-MAD (*K*-means)		49	33	99	75	99
IR-MAD ($T_{\mathrm{IR-MAD}} = 28$)		57	95	41	49	23
Hybrid	3	N.a.	N.a.	N.a.	99	99
IR-MAD (*K*-means)		N.a.	N.a.	N.a.	66	66
IR-MAD ($T_{\mathrm{IR-MAD}} = 28$)		N.a.	N.a.	N.a.	99	99
Hybrid	4	72	87	61	69	55
IR-MAD (*K*-means)		**75**	61	96	81	98
IR-MAD ($T_{\mathrm{IR-MAD}} = 28$)		29	17	99	63	99

(Verhoeven et al., 2022). Accounting for redundancies estimated at around 10% within the image pairs, this translates to 2860 pairs or approximately 13 h of processing for generating a complete change map dataset for the entire research area (assuming invariant surface geometry so the base surface mesh could be used). However, this estimate does not encompass the essential photo acquisition and co-registration phase (i.e., image orientation and synthetic image generation), for which runtimes heavily depend on the number of acquired images and the incremental SfM implementation. Assuming a good base network of photos, slightly over 1 h of image acquisition with 3 GoPro cameras mounted on a bike helmet (yielding about 24 k photos; 2 photos per second result in 8000 photos/camera), and 2 h for transport plus data management, all synthetic images could likely be generated in about a day on a potent computer. When combining both processing pipelines, it seems reasonable to suggest that change in the 13 km long graffiti-scape could thus be monitored by a dedicated team on a two-day, maximally three-day basis.

CONCLUSIONS AND OUTLOOK

This study first introduced an efficient and reasonably automated procedure to yield co-registered synthetic graffiti images from a collection of overlapping photographs. This procedure was successfully applied within a test zone, a small albeit highly active graffiti area along Vienna's Donaukanal. Binary reference change maps were generated for all synthetic images, culminating in a comprehensive change detection reference dataset of 6092 co-registered synthetic image pairs with corresponding reference change maps. Importantly, this entire dataset has been made publicly available.

This reference dataset was used to develop and validate a novel hybrid image-based change detection algorithm. The proposed approach synergises a pixel- and descriptor-based technique to detect relevant graffiti changes on urban surfaces while effectively filtering out graffiti-unrelated image disparities. The comparison of the generated reference data with the output of the proposed and threshold-optimised method yields a very high CCR (95%) and specificity (98%). Recall (77%), precision (87%) and F1-score (80%) further demonstrating the efficacy of the method in discriminating changes in graffiti cover from unrelated image variations. In a comparative evaluation against the well-established IR-MAD change detection method, the hybrid approach consistently demonstrated similar performance while exhibiting notably reduced variability.

Achromatic image regions still present a significant challenge for the developed hybrid method because they are indistinguishable in RGB-normalised images. In the future, the authors hope to address this issue by initially identifying these spectrally neutral areas. Furthermore, smaller new graffiti (such as tags) are sometimes falsely filtered out by the descriptor-based method, introducing a particular detection bias toward more extensive graffiti. To improve the method's performance in this area, one could think of dynamic gridding, which uses smaller grid cells when many feature points (and thus potential tie points) are detected and larger cells for low feature point densities.

The surface geometry of the test zone was relatively simple, with mostly straight walls and a few slightly protruding elements. In the future, more complex surfaces such as staircases and bridges could cause problems in creating and texturing the 3D surface model, either because the digital geometry cannot be satisfactorily constructed (e.g., when the photos suffer from too many occlusions) or the excessive parameterisation needed to generate a texture atlas. The results could vary from slightly warped or fuzzy textures to completely untextured areas. To mitigate the resulting risk of increasing false positives, one could apply exclusion masks for these regions, similar to the ground and no-data masks generated for the reference dataset. Since these areas are usually difficult to access without ladders or other aids, changes are less likely to happen. That said, leaving such areas unmonitored undoubtedly introduces a slight but specific bias in the subsequent graffiti documentation.

In addition to the technical challenges, there are significant ethical considerations to address. Perhaps the most pressing ethical concern involves documenting subversive content. Complicating matters, it is often extremely challenging to determine whether certain graffiti are problematic because they may utilise codes or languages unfamiliar to those documenting them. While large-scale dissemination has not yet occurred, it is planned (Verhoeven et al., 2022), prompting members of the INDIGO project to seek guidance from experts in research ethics. These consultations have provided valuable insights into how to approach this issue. One potential strategy is to add a content warning for clearly homophobic, Nazistic, racist or otherwise problematic content and prevent their searchability via tags. Additionally, a mechanism for reporting content that may not be immediately identifiable as problematic to the documentation team could be implemented. Furthermore, there is the additional ethical consideration of documenting largely illegal activities, but this issue exceeds the scope of this paper.

In summary, this approach holds substantial promise for streamlining the monitoring and image-based documentation of extensive graffiti-scapes like the one found along Vienna's Donaukanal. The question remains whether the identified thresholds will perform similarly when applied to entirely different urban environments or scenes larger than the test zone. Nonetheless, the automated character and scalability of the proposed hybrid approach bodes well for its future application on a broader scale when combined with efficient photo acquisition.

ACKNOWLEDGEMENTS

INDIGO was funded by the Heritage Science Austria programme of the Austrian Academy of Sciences (ÖAW). The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

DATA AVAILABILITY STATEMENT

The 6902 co-registered synthetic image pairs and their corresponding change maps are freely and publicly available at the TU Data Repository under 10.48436/ayj4e-v4864 (Wild, Verhoeven, Muszyński, et al., 2023). Please note that for this dataset, two different licences apply. Detailed licensing information can be found in the repository description.

ORCID

Benjamin Wild https://orcid.org/0000-0002-0769-3415
Geert Verhoeven https://orcid.org/0000-0003-4825-9604
Rafał Muszyński https://orcid.org/0000-0002-1676-8458
Norbert Pfeifer https://orcid.org/0000-0002-2348-7929

REFERENCES

Alcantarilla, P.F. (2013) Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: *British machine vision conference, Bristol*. Available from: https://doi.org/10.5244/C.27.13

Angiati, D., Gera, G., Piva, S. & Regazzoni, C.S. (2005) A novel method for graffiti detection using change detection algorithm. In: *IEEE conference on advanced video and signal based surveillance, 2005*. Como, Italy: IEEE, pp. 242–246. Available from: https://doi.org/10.1109/AVSS.2005.1577274

Arnheim, R. (1954) *Art and visual perception: a psychology of the creative eye*. Berkeley: University of California Press. Available from: https://psycnet.apa.org/record/1955-03680-000 [Accessed 17 April 2024].

Asokan, A. & Anitha, J. (2019) Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics*, 12, 143–160. Available from: https://doi.org/10.1007/s12145-019-00380-5

Bay, H., Tuytelaars, T. & Van Gool, L. (2006) SURF: Speeded Up Robust Features. In: *Computer vision—ECCV 2006: 9th European conference on computer vision, Graz, Austria*. Berlin: Springer, pp. 404–417. Available from: https://doi.org/10.1007/11744023_32

Bitelli, G., Camassi, R., Gusella, L. & Mognol, A. (2004) Image change detection on urban area: the earthquake case. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35, 692–697.

Blanché, U. (2015) Street art and related terms. *SAUC—Street Art and Urban Creativity*, 1(1), 32–39. Available from: https://doi.org/10.25765/sauc.v1i1.14

Commission Internationale de L'Éclairage. (1995) *Industrial colour-difference evaluation*. Vienna, Austria: CIE, p. 22.

Commission Internationale de l'Éclairage. (2001) CIE 142-2001 improvement to industrial colour-difference evaluation. Vienna, Austria: CIE. ISBN: 978 3 901906 08 4.

Di Stefano, L., Tombari, F., Lanza, A., Mattoccia, S. & Monti, S. (2008) Graffiti detection using two views. In: *The eighth international workshop on visual surveillance—VS2008*. Available from: https://inria.hal.science/inria-00325654 [Accessed 17 April 2024].

Finlayson, G.D., Schiele, B. & Crowley, J.L. (1998) Comprehensive colour image normalization. In: *Computer vision—ECCV'98: 5th European conference on computer vision, Freiburg, Germany*. Berlin: Springer, vol. 5, pp. 475–490. Available from: https://api.semanticscholar.org/CorpusID:8142201 [Accessed 17 April 2024].

Forster, A.M., Vettese-Forster, S. & Borland, J. (2012) Evaluating the cultural significance of historic graffiti. *Structural Survey*, 30(1), 43–64. Available from: https://doi.org/10.1108/02630801211226637

Hongruixuan, C. (2021) Structure extraction from texture via relative total variation—Matlab code. *Change Detection Repository (Github)*. Available from: https://github.com/ChenHongruixuan/ChangeDetectionRepository [Accessed 22 February 2024].

International Organization for Standardization. (2019) Colorimetry—part 4: CIE 1976 L* a* b* colour space. Geneva: ISO/CIE. Available from: https://www.iso.org/standard/74166.html [Accessed 17 April 2024].

International Organization for Standardization. (2022) *Colorimetrypart 6: CIEDE2000-colour-difference formula*. Geneva: ISO/CIE, pp. 11664–11666. Available from: https://www.iso.org/standard/82662.html [Accessed 17 April 2024].

Kampel, M., Wildenauer, H., Blauensteiner, P. & Hanbury, A. (2010) Improved motion segmentation based on shadow detection. In: *Progress in computer vision and image analysis*, pp. 519–533. Singapore: World Scientific. Available from: https://doi.org/10.1142/9789812834461_0028

Kender, J.R. (1976) *Saturation, hue, and normalized color: calculation, digitization effects, and use*. Pittsburgh, Pennsylvania: Department of Computer Science, Carnegie-Mellon University. Available from: https://apps.dtic.mil/sti/citations/ADA037772 [Accessed 17 April 2024].

Leutenegger, S., Chli, M. & Siegwart, R.Y. (2011) BRISK: binary robust invariant scalable keypoints. In: *2011 International conference on computer vision*. Barcelona: IEEE, pp. 2548–2555. Available from: https://doi.org/10.1109/ICCV.2011.6126542

Lowe, D.G. (1999) Object recognition from local scale-invariant features. In: *Proceedings of the seventh IEEE international conference on computer vision*. Kerkyra, Greece: IEEE, vol. 2, pp. 1150–1157. Available from: https://doi.org/10.1109/ICCV.1999.790410

Lowe, D.G. (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 91–110. Available from: https://doi.org/10.1023/B:VISI.0000029664.99615.94

Luhmann, T., Fraser, C. & Maas, H.-G. (2016) Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 37–46. Available from: https://doi.org/10.1016/j.isprsjprs.2015.10.006

Luo, M.R., Cui, G. & Rigg, B. (2001) The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, 26(5), 340–350. Available from: https://doi.org/10.1002/col.1049

McLaren, K. (1976) XIII—the development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula. *Journal of the Society of Dyers and Colourists*, 92(9), 338–341. Available from: https://doi.org/10.1111/j.1478-4408.1976.tb03301.x

Nielsen, A.A. (2007) The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Transactions on Image Processing*, 16(2), 463–478. Available from: https://doi.org/10.1109/TIP.2006.888195

Nielsen, A.A., Conradsen, K. & Simpson, J.J. (1998) Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: new approaches to change detection studies. *Remote Sensing of Environment*, 64(1), 1–19. Available from: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=1220 [Accessed 17 April 2024].

Niemann, S. (2022) INGRID—archiving graffiti in Germany. *Document | Archive | Disseminate Graffiti-Scapes*, 1, 231–238. Available from: https://doi.org/10.48619/indigo.v0i0.712

Nocerino, E., Menna, F. & Remondino, F. (2014) Accuracy of typical photogrammetric networks in cultural heritage 3D modeling projects. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5, 465–472. Available from: https://doi.org/10.5194/isprsarchives-XL-5-465-2014

Novak, D. (2015) Photography and classification of information. *SAUC—Street Art and Urban Creativity*, 1(1), 13–25. Available from: https://doi.org/10.25765/sauc.v1i1.22

Pillai, G.V, Gupta, N. & Ari, S. (2017) Descriptors based unsupervised change detection in satellite images. In: *2017 International conference on communication and signal processing (ICCSP)*. Chennai, India: IEEE, pp. 1629–1633. Available from: https://doi.org/10.1109/ICCSP.2017.8286666

Radke, R.J., Andra, S., Al-Kofahi, O. & Roysam, B. (2005) Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing*, 14(3), 294–307. Available from: https://doi.org/10.1109/TIP.2004.838698

Ringhofer, A. & Wogrin, S. (2018) Die Kunst der Straße—Graffiti in Wien. *Wiener*, 428, 46–53. Available from: https://wiener-online.at/2018/05/02/die-kunst-der-strasse-graffiti-in-wien/ [Accessed 17 April 2024].

Ronchi, A.M. (2009) eCulture: cultural content in the digital age. Berlin: Springer Science & Business Media. Available from: https://doi.org/10.1007/978-3-540-75276-9

Schlegel, J., Carloni, M., Wogrin, S., Graf, A.M. & Verhoeven, G.J. (2022) Making a mark—towards a graffiti thesaurus. *Document | Archive | Disseminate Graffiti-Scapes*, 203–219. Available from: https://doi.org/10.48619/indigo.v0i0.710

Seo, J., Park, W. & Kim, T. (2022) Feature-based approach to change detection of small objects from high-resolution satellite images. *Remote Sensing*, 14(3), 462. Available from: https://doi.org/10.3390/rs14030462

Sharma, G. (2003) Color fundamentals for digital imaging. In: *Digital color imaging handbook*. London: CRC Press, pp. 1–114. Available from: https://doi.org/10.1201/9781420041484

Shi, W., Zhang, M., Zhang, R., Chen, S. & Zhan, Z. (2020) Change detection based on artificial intelligence: state-of-the-art and challenges. *Remote Sensing*, 12(10), 1688. Available from: https://doi.org/10.3390/rs12101688

Stamatopoulos, C. & Fraser, C. (2014) Automated target-free network orienation and camera calibration. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II–5, 339–346. Available from: https://doi.org/10.5194/isprsannals-II-5-339-2014

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A. & Fisher, P.F. (2015) A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, 1–14. Available from: https://doi.org/10.1016/j.rse.2015.01.006

Tombari, F., Di Stefano, L., Mattoccia, S. & Zanetti, A. (2008) Graffiti detection using a time-of-flight camera. In: *Advanced concepts for intelligent vision systems: 10th international conference, ACIVS 2008, Juan-les-Pins, France, 20–24 October 2008. Proceedings 10*. Berlin: Springer, pp. 645–654. Available from: https://doi.org/10.1007/978-3-540-88458-3_58

Verhoeven, G. (2017) Computer graphics meets image fusion: the power of texture baking to simultaneously visualise 3D surface features and colour. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W2, 295–302. Available from: https://doi.org/10.5194/isprs-annals-IV-2-W2-295-2017

Verhoeven, G., Wild, B., Schlegel, J., Wieser, M., Pfeifer, N., Wogrin, S. et al. (2022) Project INDIGO: document, disseminate & analyse a graffiti-scape. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46, 513–520. Available from: https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-513-2022

Verhoeven, G.J., Carloni, M., Schlegel, J., Wild, B. & Wogrin, S. (2023) Finding listeners for walls that speak. *Proceedings of the GoINDIGO2022 international graffiti symposium*, 1, 6–15. Available from: https://doi.org/10.48619/indigo.v0i0.699

Verhoeven, G.J., Wogrin, S., Schlegel, J., Wieser, M. & Wild, B. (2023) Facing a chameleon—how project INDIGO discovers and records new graffiti. *Document | Archive | Disseminate Graffiti-Scapes*, 1, 63–85. Available from: https://doi.org/10.48619/indigo.v0i0.703

Wieser, M., Verhoeven, G. & Wild, B. (2022) Acquiring centimetre-accurate camera coordinates in project INDIGO. Available from: https://doi.org/10.5281/zenodo.7109573

Wild, B., Verhoeven, G.J., Muszyński, R. & Pfeifer, N. (2023) INDIGO change detection reference dataset. Available from: https://doi.org/10.48436/ayj4e-v4864

Wild, B., Verhoeven, G.J. & Pfeifer, N. (2023) Tracking the urban chameleon—towards a hybrid change detection of graffiti. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 285–292. Available from: https://doi.org/10.5194/isprs-annals-X-M-1-2023-285-2023

Wild, B., Verhoeven, G.J., Wieser, M., Ressl, C., Schlegel, J., Wogrin, S. et al. (2022) AUTOGRAF—AUTomated orthorectification of GRAFfiti photos. *Heritage*, 5(4), 2987–3009. Available from: https://doi.org/10.3390/heritage5040155

Wogrin, S. (2022) Spraycity.at—graffiti archive and online map. *Document | Archive | Disseminate Graffiti-Scapes*, 1, 239–249. Available from: https://doi.org/10.48619/indigo.v0i0.713

Xu, L., Yan, Q., Xia, Y. & Jia, J. (2012) Structure extraction from texture via relative total variation. *ACM Transactions on Graphics (TOG)*, 31, 1–10. Available from: https://doi.org/10.1145/2366145.2366158

How to cite this article: Wild, B., Verhoeven, G., Muszyński, R. & Pfeifer, N. (2024) Detecting change in graffiti using a hybrid framework. *The Photogrammetric Record*, 39, 549–576. Available from: https://doi.org/10.1111/phor.12496