

Received 29 November 2023, accepted 21 December 2023, date of publication 5 January 2024, date of current version 18 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3350514

RESEARCH ARTICLE

High Accurate Rephotographic Image Registration by Attention Masks: Enabling Intention-Driven Rephotographic Image Registration With Interactive Areas of Interest Masks

AXEL SCHAFFLAND¹ AND JULIUS SCHÖNING², (Senior Member, IEEE)

¹Institute of Cognitive Science, Osnabrück University, 49069 Osnabrück, Germany

²Faculty of Engineering and Computer Science, Osnabrück University of Applied Sciences, 49076 Osnabrück, Germany

Corresponding author: Axel Schaffland (axschaffland@uni-osnabrueck.de)

ABSTRACT Embarking on the journey of rephotography, capturing a contemporary image from the vantage point of a historical counterpart and registering them, is a formidable challenge. Traditional automated registration methods stumble in the face of this task, while manual methods, reliant upon painstakingly identified corresponding points, demand an investment of time, precision, and expertise. Often, only image fragments can be seamlessly registered due to changes in the scene, like new and removed buildings. Determining the areas of interest (AOI) for registration becomes a critical decision, placing users in the process's role as curators. This work proposes a new method combining state-of-the-art automatic deep learning-based registration methods with user-provided masks. Users draw masks around the AOI they want to register and exclude non-indented AOI from registration. Using AOI masks reduces the required time, painstaking identification of corresponding points, and knowledge needed for manual registration while giving the user control over the registration process by providing an intuitive way to embed which AOI is vital to register. This interactive method achieves excellent registration quality and positive user feedback compared to regular automated image registration methods. It can not replace manual registration completely. However, for many rephotography tasks, it significantly reduces the required effort. The deep learning-based automatic method already achieves a high acceptance rate i.e., a score of at least 4 out of 5 of 55%, which is a considerable improvement to standard automatic registration method with an acceptance rate of 12%. With the interactive AOI masks method, which combines user-drawn masks with the automatic deep learning-based method, the acceptance rate increases to 60% and is almost as good as manual registration with a rate of 65%.

INDEX TERMS Repeat photography, rephotography, areas of interest, high accurate image registration, interactive, image registration, HCI for image registration, deep features.

I. INTRODUCTION

Rephotography originated towards the close of the 19th century within the realm of glaciology in the Italian Alps. In this context, photographs captured from identical camera positions of a particular glacier, but at distinct moments in time, were utilized to quantify the glacier's movement.

The associate editor coordinating the review of this manuscript and approving it for publication was Giuseppe Desolda¹.

In the present day, rephotography finds application not only across various domains of research but also within the spheres of popular science and art [1], [2], [3]. The fundamental concept remains unaltered: capturing two or more images of a given scene from an identical camera stance but at different times. For accurate measurements and aesthetically pleasing outcomes, the captured images must undergo a registration process. This registration process entails aligning the images so that the objects within them occupy the exact

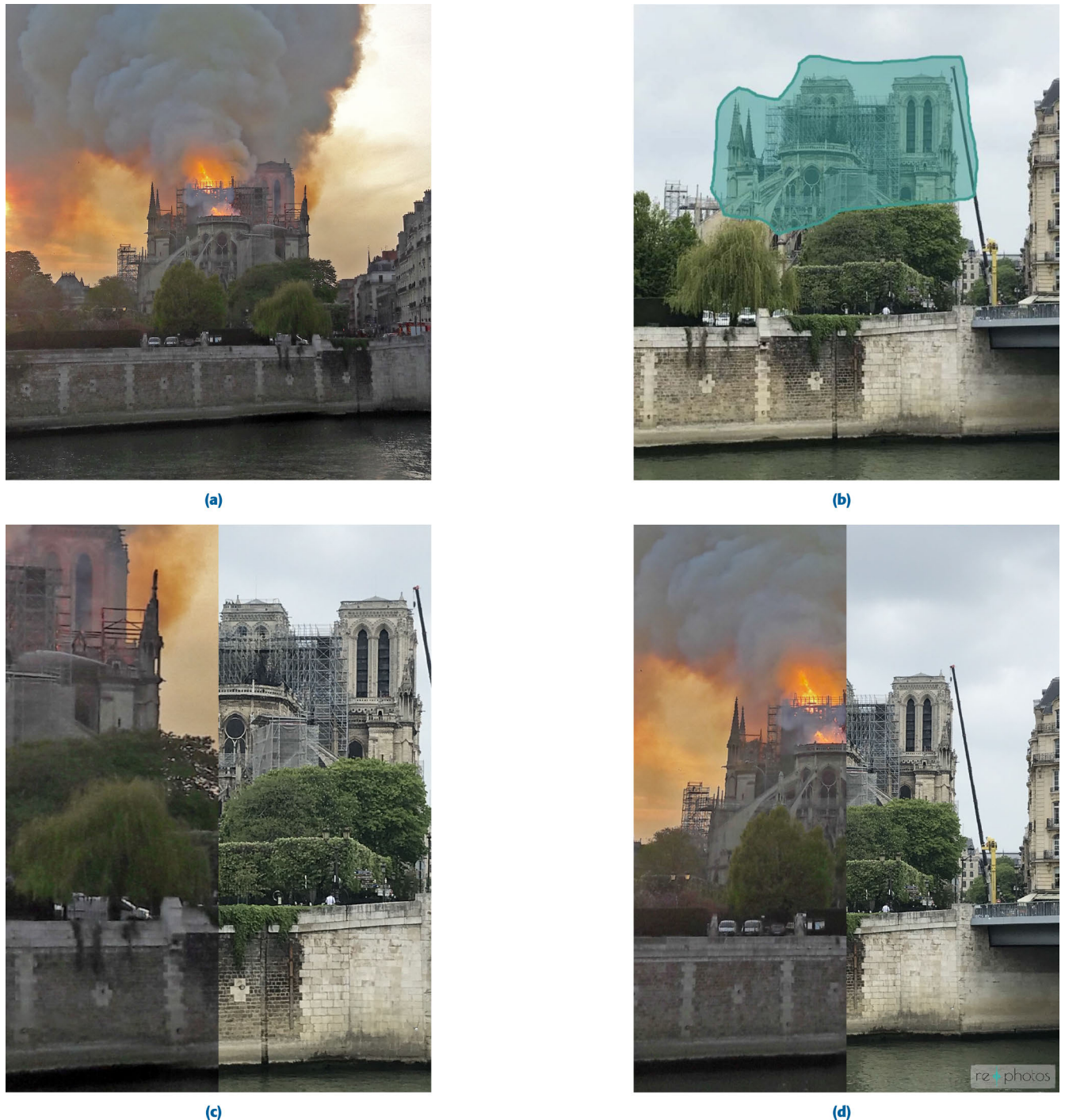


FIGURE 1. AOI Masks supports users in correctly registering this rephotography of Notre-Dame de Paris during and after the 2019 fire. (a) and (b) show the two original images. (b) the mask-based registration method; the interactive AOI mask is visible in green. (c) image registration result of unmasked automated registration. Note that the retaining wall in the foreground is perfectly registered; the cathedral itself, however, is not registered at all. (d) image registration result using AOI mask registration mask from (b) leads to accurate registration of the cathedral. The retaining wall is only roughly registered since the camera positions were different. Nonetheless, the area the user intended to register, namely the masked cathedral, is registered well, showing that users' intentions can be embedded into the registration process with AOI masks. Images (c) and (d) are cropped. (a): CC BY-SA 4.0 Waterced; (b): CC BY-SA 4.0 Louis H. G.

pixel coordinates. This registration procedure can be executed either automatically or interactively, with the possibility of incorporating additional input from users. One of the

biggest challenges in image registration for rephotography is incorporating the user's intentions into the registration process, i.e., identifying the object of interest (OOI) to

be registered accurately. Various reasons prevent perfect registration of the old and new images with rigid registration methods. One of these reasons is caused by different camera positions, from which the old and the new image were taken. Another reason are altered scene objects that are added, moved, or removed in the scene between the recording of the old and the new image. In the altered scene objects cases, only certain areas of interest (AOI) of the image can be registered. The decision of which AOI are essential for the registration depends entirely on the user's intentions. To illustrate the importance of AOI for rephotography, the following use case is assumed: An image pair of a tree in front of a glacier can interest both a biologist and a glaciologist. While for the biologist, the AOI is the tree, the glaciologist's AOI is in the glacier. Hence, for one, the tree should be registered, and for the other, the glacier.

Interactive non-rigid methods [4] may alleviate the limitation that only some areas of the images can be registered and may register several AOI. However, they may introduce artifacts, higher computational costs, and less intuitive and predictable results. Further, they may require complex user input. Direct manual methods, in which users first mark corresponding point pairs and perform the registration after, do not give users control during the registration process. A possible solution giving users more control during the registration process are the interactive rigid registration methods [5], [6]. Here, users select corresponding feature points in both images, while registration is directly computed based on the point pairs currently selected. These interactive methods allow for fine-grained control during the registration process. In the abovementioned use case, the biologist could select corresponding points at the bottom of the trunk of the tree and in its crown to register the tree, while the glaciologist chooses points in the bedrock around the glacier to register the glacier. While the registration results achieved with this method are very good, they require an accurate selection of corresponding feature point pairs by the users and thus can be expensive in time. On the other hand, automatic image registration methods often fail for this use case of rephotography. This failure is partially caused by large differences between the images on the pixel level evoked by different recording methods and small changes in the scene, like illumination, foliage, and weather. However, it is also caused by point pairs detected in unimportant areas, like the sky, or non-corresponding areas, like similar but not equal objects.

This paper proposes a new method to augment a deep automatic image registration process geared for challenging image pairs with user-provided AOI masks, providing information on the intended scene and objects to be registered, thus embedding the user's intentions into the registration process. We provide this unique combination of user-provided AOI masks with state of the art registration techniques as browser accessible web interface without the need of local software installation. We show that our combination outperforms state of the art deep registration alone and severely outperforms

classic registration techniques on a multimodal, multitemporal, and multipositional rephotographic dataset.

In the following, Section II will present rephotographic image registration with the user-provided AOI masks registration process, Section III explains the evaluation methods, Section IV evaluates the results of the registration method, and Section V discusses its advantages and disadvantages. Before concluding in Section VII, Section VI points out future research.

II. IMPLEMENTATION AND REGISTRATION METHODS

In short, the new method combines automatic image registration using deep learning with user-provided AOI masks. Users provide an AOI mask for OOI essential for the registration on only one image. By taking masked areas into account for the automatic image registration process the users' intentions can be embedded into the process.

A. IMPLEMENTATION

The AOI image registration procedure is implemented as a web service consisting of a server for computation and storage and the clients for user interaction via the web browser. Users can opt to use the automatic registration method, as well. The interactive rigid registration method [5], runs completely on the client without requiring a permanent connection to a server. In this case users can download the complete website as a local copy once, e.g., by the "Save as" function of the web browser, and she/he can use the interactive rigid registration without any internet connection, which is advantageous for on-site rephotography. A pure client side implementation is not possible for intention-driven image registration with AOI masks. The deep neural network used for image registration requires a special software and hardware constellation. Without an adequate GPU with correctly installed software, the AOI-based masked registration is so slow that the user interaction becomes infeasible. Even if the user's computer provides a proper software and hardware constellation, weights from the deep neural network used must be downloaded. This is especially invisible for use on mobile devices. Thus, the server-client architecture with a powerful back-end server is chosen to ensure that the users do not need to worry about soft- and hardware requirements and can use the registration method directly in their web browser. Another advantage of this architecture is the centralized storage of unregistered and registered images on the back-end. The server can be backed up regularly. If a group of rephotographers collaborates, their images are collected on the server and can be further organized, e.g., with a geographic information system (GIS) [7].

1) USER INTERACTION

Users start by selecting the old and new image; both are stored on the user's computers. During registration, the first image is the fixed image, i.e., the image which is not changed during the registration process, and the second image is the moving image which is transformed to align it to the first

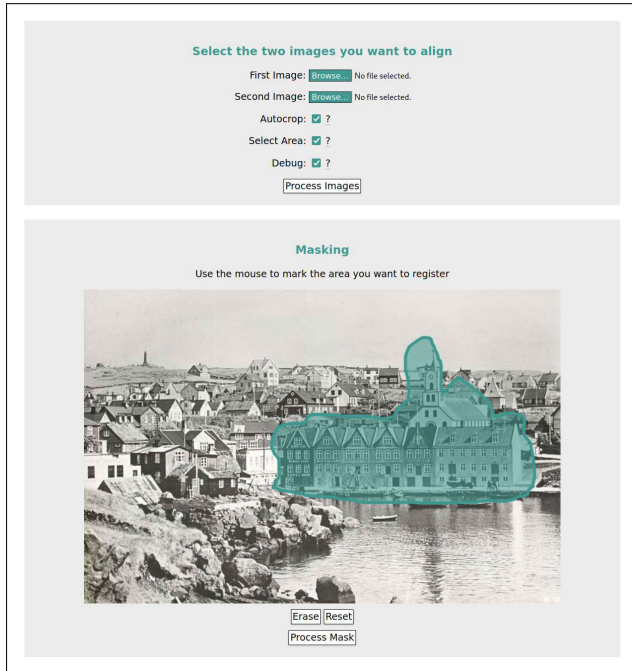


FIGURE 2. Web application for intention-driven rephotography during the registration process. The user has uploaded two images and is drawing the AOI mask into one image. Users can also use the automatic image cropping and debug information features. The later, providing a checkerboard overlay of the registered and unregistered images, as well as visualization of the matched point pairs and the transformation matrix. Users can add one or multiple AOI masks on the image, reset or erase them, and submit the mask and the images to the AI-based registration process running on the back-end server. *Image: No known rights.*

image. The user draws the AOI mask only in the first image. The first image is usually the old historical image, which is kept original. The new image is often of higher quality and resolution—another reason to transform the new image to prevent further loss of quality of the old image. Users can also select to autocrop the image, use the fully automatic registration process, and display debug information shown in Figure 2. Users then upload the images to the server. The server generates a unique ID for each image pair, which is used to store and retrieve all data belonging to this image pair. Users can now draw the mask in the *Masking* section if they decide to use the AOI masked registration process. Users can mask multiple AOIs in the first image by drawing around the area with the mouse, as shown in Figure 2. They can also reset and erase the mask, and finally upload it to start the registration process. The back-end processes and stores the masks drawn before starting the registration process, described in the following. After registration, the back-end stores and crops the registered images. The front-end then displays the *Registration process complete* section, shown in Figure 3. Users can view and compare the registered images using the morph slider animation. They can move the slider to reveal more of the old or new image. They can then download the registered images which are of the same size or use the linked interactive registration process to improve the registration quality.

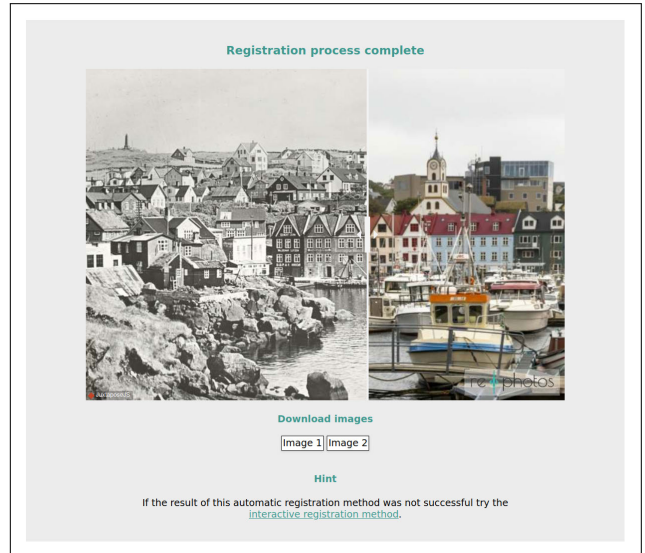


FIGURE 3. Web application showing the result after the AOI mask-based registration process. The user can verify the registration process by moving the slider revealing more of the old or the new image. If the results are satisfactory, the registered images can be downloaded. If results need further optimization, the linked interactive registration method can be used. *Left image: No known rights; Right image: CC BY-NC-SA: Axel Schaffland.*

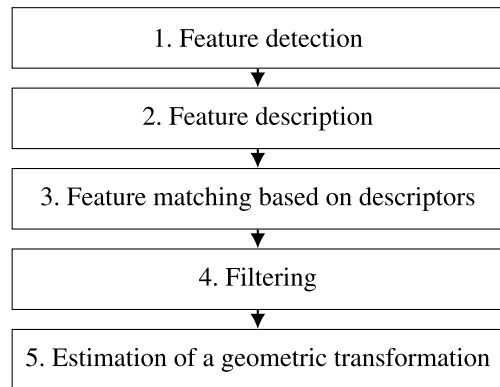


FIGURE 4. The five steps of the image registration pipeline. In our implementation Superpoint [8], an artificial neural network (ANN) for feature detection and description, and Superglue [9] an ANN for descriptor based matching, are used. Points are filtered by AOI-masks and the transformation is estimated via RANSAC [25].

B. AUTOMATIC REGISTRATION

Automatic image registration algorithms [10], [11] exist for various applications like medicine [12] and remote sensing [13]. Panorama stitching and focus stacking are straightforward application cases for which image registration works well since the images were recorded with the same camera from the same position without significant changes within the scene in a short time. For rephotography, multitemporal, multimodal, and multipositional image registration methods are required because the images are taken with different cameras from different positions at different points in time, usually several decades of the year lying between the captures. Classic feature and pixel-based image registration

methods fail because of pixel, scene, and content level differences caused by the different sensor technologies of the used cameras, capturing positions, scenery, and capturing times. Some registration methods geared towards rephotography exist [1]. These methods vary in their degree of automation and may require human intervention during the registration. They can be also directed to specific subsets of rephotography, e.g., architectural rephotography in which many straight lines are to be registered [4]. Various deep learning methods emerged in recent years both for general image registration [14], [15] and especially for medical image registration [16], replacing multiple parts within the traditional image registration pipeline shown in Figure 4. The registration pipeline takes two images as input. In the first step on both images locations of the features are detected. In the second step these locations are described taken the local area around the location into account. Features from both images are then matched into pairs based on their descriptors in the third step. In the fourth step features are filtered, here we use the AOI and discard all feature pairs outside the AOI. Lastly, an optimal transformation matrix is computed aiming to align the feature pairs. This matrix can be used to transform one of the input images to be registered to the other input image.

Providing a state-of-the-art web service for rephotography, we use two ANNs to replace step 1, 2, and 3 of the registration pipeline (Figure 4): (1) Superpoint [8] to simultaneously detect and describe feature points and (2) Superglue [9] for feature matching and filtering. Superpoint uses a fully convolutional encoder-decoder network with a shared encoder and two separate decoders for feature detection and description. While the Encoder is VGG-based [17] both decoders use one convolutional layer followed by non-learned upsampling. Superpoint is pre-trained on a synthetic dataset of geometric shapes with known feature point locations. Opposed to SIFT [18], with hand-coded invariance against scaling and rotation, Superpoint archives invariance by homographic adaption, a self-supervised learning method combining detector results over randomly transformed copies of the input image. Superglue uses an attentional graph neural network with alternating self- and cross-attention layers to transform the feature vectors of two images. The vectors combine the descriptors and the position. From the transformed feature vectors, an optimal partial assignment is computed by the Sinkhorn algorithm [19]. Various adaptations of the SIFT algorithm exist, addressing e.g., color [20], view [21], and illumination [22] invariance. However, none of these adaptations combine all these invariances [23] which are required for rephotographic image registration. For benchmarking the two abovementioned ANN methods, the regular methods SIFT [18] for feature detection and description as well as brute force matching for feature matching from their OpenCV implementation [24] are used. The geometric transformation matrix can be estimated with classic optimization methods provided by OpenCV with the point pairs found by SIFT respectively Superpoint and matched by the Brute-force

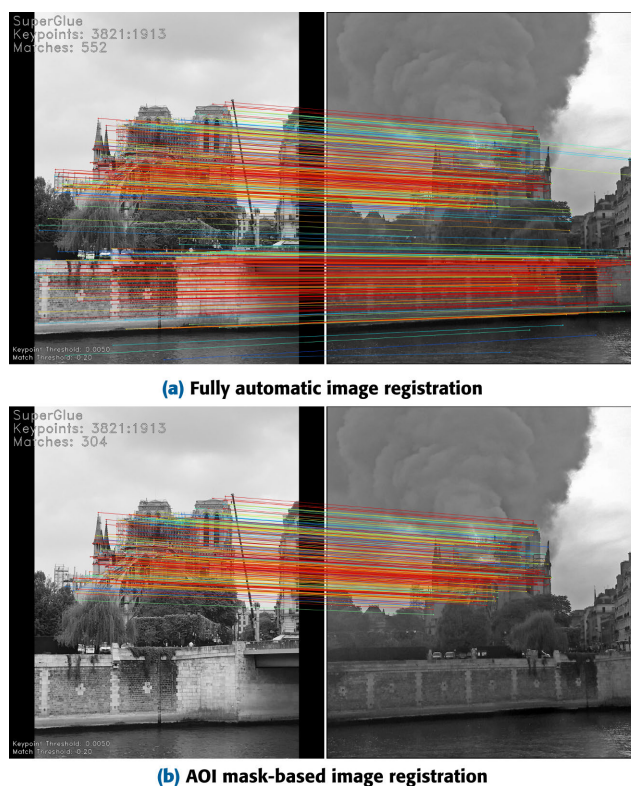


FIGURE 5. Point pairs for image registration. (a) shows all point pairs used for fully automatic image registration. (b) shows the result of our masked method filtering out all point pairs for which the point is outside the masked area in the left image. *Left images: CC BY-SA: Louis H. G.; Right images: CC BY-SA: Waterced.*

matcher respectively Superglue. These optimization methods minimize the reprojection error between all the points in the fixed image and all the corresponding transformed points in the moving image. The resulting optimal matrix is then used to transform the moving image. Finally, both images are cropped and returned to the user.

C. MASKING

The AOI masks can augment the fully automatic registration methods in different ways:

(1) Users draw AOI masks of the OOI in the old and new images. A point pair is only used for registration if the feature points in the old image are inside the old image's masked area and the new image's feature points are inside the new image's masked area. During the development, this way was refused since it doubles the user interactions required by the user. Further, it could wrongly reduce the number of point pairs used for registration: In an extreme case, users could mistakenly mask the left part of the OOI they want to register in the old image and mask the right part of the same OOI in the new image resulting in an empty intersection of masked areas and no point pairs for the image registration process.

(2) Users draw the AOI masks only in the old or the new image. This way user interaction is minimized compared to (1). In the presented web service, users draw the mask in the

TABLE 1. The scoring system to assess registration qualitatively.

Score	Name	Description
0	Failed	No resulting image or image distorted beyond recognition.
1	Very Bad	No areas registered or strong directly visible distortions.
2	Bad	Some areas are registered. Distortions notable under direct comparison.
3	Good	The important areas are registered. No distortions.
4	Very Good	The important areas are registered accurately. No distortions.
5	Perfect	The complete image is registered accurately. No distortions.

first image, usually the old image. Drawing the AOI masks in the historical image have advantages for the following reasons: (I) The rephotographic process starts with the old image and the intention to rephotograph an area of interest visible in the old image. From that, it follows naturally to mask this AOI in the old image. (II) The new image is often photographed with a smaller focus length to cover a larger part of the scene and to ensure that the complete area visible in the old image is also visible in the new image. Masking in the new image could result in masked areas not visible in the old image. (III) Some parts of the old image may be degraded over time and are no longer suitable for registration.

Nonetheless, there are reasons to mask the new image as well: Over time, the visibility of the AOI shrinks—e.g., by growing trees occluding the view on the OOI and by the decay of a building leading to the caved roof or leaving only ruins. Masking in the new image allows one to select the remaining AOI. On the other hand, trees may fall, and buildings may be completed leading to a larger AOI in the new image. Masking in the historical image is preferable. Thus, the users can draw the AOI masks on the first uploaded image, the fixed image, within the presented implementation. This enables the users to decide whether to upload the old or new image as the first image.

III. EVALUATION

The following section provides information on evaluating three aspects of the intention-driven rephotographic image registration method: (1) quantitative registration performance. (2) qualitative registration performance and (3) usability.

A. REGISTRATION EVALUATION

For both measuring qualitative and quantitative registration performance, 100 image pairs from the rephotography website *re.photos*¹ were randomly selected. Image pairs on *re.photos* website are scored by other users on a scale from 1 to 5. Ensuring a high registration quality and that the OOI are adequate for a rephotography benchmark, only

¹www.re.photos

image pairs with a score larger than 4 were taken into account. For these 100 image pairs, the original unregistered image pairs, in the following UNREGISTERED, the registered image pairs, in the following MANUAL, and the corresponding points marked by the users were retrieved. *re.photos* users registered the retrieved registered image pairs with a manual point pair-based registration method similar to our interactive registration method [5]. The 100 original image pairs (UNREGISTERED) were registered with the automatic registration method based on SIFT [18] (SIFT) and Superpoint [8] /Superglue [9] (ANN) and with the AOI mask-based registration methods extending either SIFT (SIFT+MASK) or Superglue (ANN+MASK). For all methods, the resulting images and transformation matrices were saved. UNREGISTERED served as lower baseline, corresponding to no registration and no user interaction, while MANUAL serves as upper baseline, being the most accurate and labor intensive method. To these two baselines the four registration methods SIFT, SIFT+MASK, ANN, and ANN+MASK were compared.

Thus, for each image pair, six versions exist: UNREGISTERED: the original as lower baseline, MANUAL: the manually registered images by *re.photos* users as upper baseline, SIFT: registered automatically using SIFT, SIFT+MASK: SIFT enhanced with the AOI masks, ANN: registered automatically with Superpoint and Superglue, ANN+MASK: ANN enhanced with the AOI masks. Additionally, for all image pairs the point pairs, set manually by *re.photos* users and the transformation matrices of all registration methods were added to the evaluation database.

1) QUANTITATIVE REGISTRATION EVALUATION

To compare registration quality the squared reprojection error (RPE) based on transformation matrices, applied to the *re.photos* point pairs, is used.

All four registration methods SIFT, SIFT+MASK, ANN, and ANN+MASK as well as the upper baseline MANUAL, use the same algorithm to compute a perspective transformation matrix, resulting in a total of five transformation matrices \mathbf{M}^j with $j : [1, 5]$, one matrix for each registration method. Direct comparison of transformation matrices is not meaningful. For a meaningful interpretation the matrices can be applied to ground truth point pairs, which should be registered perfectly by the different registration methods, i.e., the transformed points of the moving image should be projected onto the corresponding points of the fixed image. We use the point pairs from the upper baseline MANUAL as ground truth: Since image pairs from *re.photos* with a high user score were selected, the registration quality of these pairs should be high, meaning that the distance between the two points of a pair should be minimal in pixel coordinates between the fixed image and the transformed moving image. $\mathbf{p}_i'(x_i', y_i')$ is one of the N points from the fixed image and $\mathbf{p}_i(x_i, y_i)$ is the corresponding point from the moving image. The points \mathbf{p}_i from the moving image are transformed into $\hat{\mathbf{p}}_i^j$

with one of the transformation matrices \mathbf{M}^j of the different registration methods:

$$\check{\mathbf{p}}_i^j = \begin{pmatrix} \frac{\mathbf{M}_{1,1}^j x_i + \mathbf{M}_{1,2}^j y_i + \mathbf{M}_{1,3}^j}{\mathbf{M}_{3,1}^j x_i + \mathbf{M}_{3,2}^j y_i + \mathbf{M}_{3,3}^j}, \\ \frac{\mathbf{M}_{2,1}^j x_i + \mathbf{M}_{2,2}^j y_i + \mathbf{M}_{2,3}^j}{\mathbf{M}_{3,1}^j x_i + \mathbf{M}_{3,2}^j y_i + \mathbf{M}_{3,3}^j} \end{pmatrix} \quad (1)$$

This step is repeated for all point pairs and all five registration matrices. Then, the squared reprojection error (RPE) was computed for each image pair and for each of the transformation matrices \mathbf{M}^j of the different registration methods between the N points \mathbf{p}_i^j of the fixed image, and the N transformed points $\check{\mathbf{p}}_i^j$ of the moving image:

$$RPE = \sqrt{\frac{1}{N} \sum_i \|\mathbf{p}_i - \check{\mathbf{p}}_i^j\|^2} \quad (2)$$

Additionally, the error was computed for the UNREGISTERED images directly between the fixed points $\mathbf{p}_i^j(x'_i, y'_i)$ and the moving points $\mathbf{p}_i(x_i, y_i)$. To allow for comparison between different sized images, the images and points were normalized such that all images have a width of 1 and a height of $\frac{1}{\text{aspect ratio}}$. Five number summaries were reported, including the median of the RPE over the dataset instead of the mean RPE only, because one or few rogue transformation matrices create large outliers. In these cases, the actual value of the corresponding RSE would heavily influence the mean and render it meaningless—the resulting registration is unusable. Excluding these matrices from the error computation would render the evaluation meaningless. Transformations can be identified as outliers if their RPE is larger than $3 \times z$ -score. The z -score is computed on the mean and standard deviation of the upper baseline MANUAL only. Using this definition of outliers we observe 76 outliers for UNREGISTERED, 84 outliers for SIFT, 77 outliers for SIFT+MASK, 35 outliers for ANN, 35 outliers for ANN+MASK, and 3 outliers for MANUAL.

2) QUALITATIVE REGISTRATION EVALUATION

For the qualitative evaluation, a rephotography expert rated the registration quality of all the image pairs. To prevent score bias by the expert, the order of all versions of all image pairs was randomized. The Images were presented to the expert in randomized ordered and rated on a scale from 0 to 5 as listed in Table 1. Images were displayed on the screen using the slider method to allow the expert to evaluate registration quality at different positions of the image pair. This completely randomized evaluation approach avoids directly comparing the different methods on the same image pair and guessing the registration method.

B. USABILITY

New and expert rephotographers were asked to register their rephotographs and a set of test rephotographs with the web service to measure the method's usability. Rephotographers

were unconstrained in selecting the image pairs to cover various application cases. They were also free to choose between the ANN and ANN+MASK method, in selecting AOI, and in drawing the mask. After testing the registration methods they were asked for feedback.

IV. RESULTS

This section presents the quantitative and qualitative results of the registration methods in comparison to the upper and lower baseline as well as feedback by rephotographers on the developed intention-driven rephotographic image registration web service.

A. QUANTITATIVE REGISTRATION RESULTS

The RPEs of the different registration methods are listed in Table 2. SIFT results are worse than the UNREGISTERED original images, even when combined with masks as SIFT+MASK. ANN results are better and can be slightly improved by masks as ANN+MASK, while the upper baseline MANUAL, registration by re.photos users, has the lowest RPE. This indicates that for this rephotography dataset hand engineered SIFT feature detectors and descriptors are outperformed by trained neural network based feature detectors and descriptors. Further, ANN+MASK has a lower RPE than ANN, indicating that masking AOIs does indeed increase registration quality. A dataset with more image pairs, where only the AOI can be registered, would increase the difference between ANN and ANN+MASK, while a dataset consisting of image pairs which can be completely registered would decrease this difference. To bring the RPE into perspective the 5 number summary of the RPE was computed over all image pairs and registration methods with a score ≥ 4 , as presented in the following Section IV-B. Min: 0.000; Q1: 0.006; Med: 0.011; Q3: 0.021; Max: 0.243. This indicates that in order to reach a score of at least 4 the RPE has to be at least below 0.243. This is the case for a large part of image pairs registered with ANN and ANN+MASK, but only for a small part of image pairs registered with SIFT and SIFT+MASK.

Example image pairs together with the feature points are displayed in Figure 6. The registration results of the different methods and the lower and upper baseline are shown in Figure 7. In these examples ANN, ANN+MASK, and Manual are not easy to differentiate. ANN and ANN+MASK differ, since ANN finds an optimal registration for the whole image, while ANN+MASK finds an optimal registration for the AOI. The number of feature points is low for SIFT as visible in the first column of Figure 6. As visible in Table 2, masking will not increase the registration quality to an acceptable level for SIFT+MASK, possible due to the low number of point pairs. Thus, SIFT+MASK is excluded in Figure 7 and in Section IV-B.

B. QUALITATIVE REGISTRATION RESULTS

As summarized in Table 3, ANN+MASK, i.e., the combination of Superglue + AOI masks received higher average

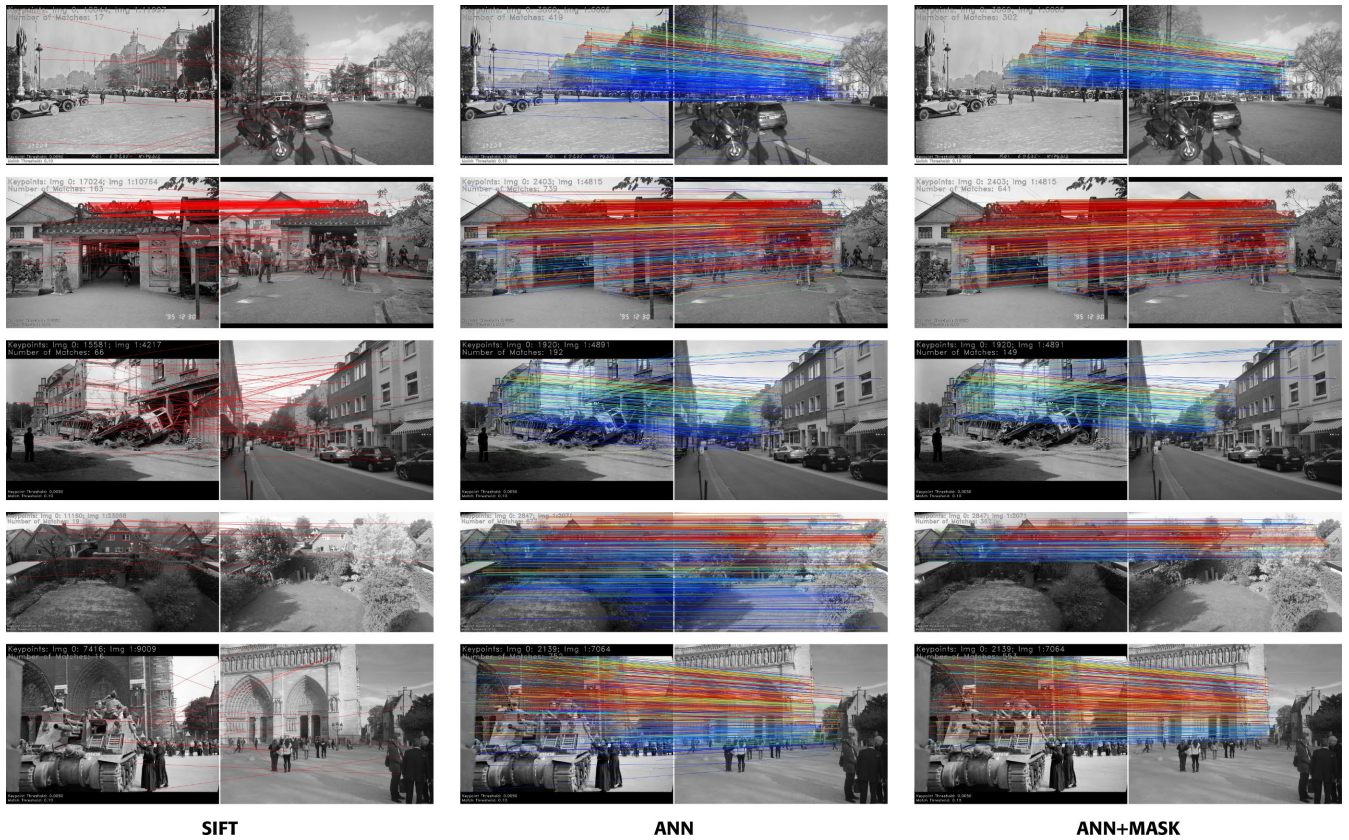


FIGURE 6. Point pairs used for the registration shown in Figure 7. SIFT detects and matches only few feature points, indicating that SIFT+MASK can not yield good registration results. ANN and ANN+MASK differ in the AOI-mask filtering of the feature points. Starting with a high number of points allows to filter out a large amount of them by the AOI-masks still leading good registration results. For images with few points detected by ANN, ANN+MASK can lead to worse results, when points inside the AOI are not sufficient for registration. Registration results for these image and point pairs are shown in Figure 7. 1st row: CC BY-SA: Bibliothèque Nationale de France, Nicolai Wolpert; 2nd row: CC BY: Ik T, CC BY-SA: Lena; 3rd row: CC BY-SA: Vestische Straßenbahnen GmbH, CC BY-NC-ND: Nicolai Wolpert; 4th row: CC BY-NC-ND: Oliver Vornberger; 5th row: Public Domain: Edward George Malindine, CC BY-NC-SA: Nicolai Wolpert.

TABLE 2. RPE of the image registration methods baselines. ANN+MASK performs best and is directly followed by ANN. MANUAL, the upper baseline achieves the best RPE but is the most labor intensive method. While SIFT+MASK performs better than SIFT, both have a high RPE even compared to the lower baseline UNREGISTERED. This corresponds with the users score and acceptance rate as shown in Table 3 and Figure 8: ANN and ANN+MASK perform almost as good as the upper baseline MANUAL but without or only with little human intervention. Best results in bold, worst result in italics.

RPE	UNREGISTERED	SIFT	SIFT+MASK	ANN	ANN+MASK(ours)	MANUAL
Min	0.000	0.000	<i>0.004</i>	0.002	0.003	0.000
Q1	0.059	<i>0.454</i>	0.319	0.010	0.014	0.001
Med	0.162	<i>0.969</i>	0.873	0.024	0.024	0.006
Q3	0.254	<i>1.585</i>	1.498	0.102	0.088	0.012
Max	0.646	<i>951.699</i>	22.050	16.588	3.169	0.063

TABLE 3. Qualitative comparison of the image registration methods by user scores and acceptance rate. SIFT scores are lower even compared to UNREGISTERED. ANN and ANN+MASK almost reach the score and rate of the upper baseline MANUAL, while ANN+MASK is higher scored compared to ANN, indicating that masks embedding AOIs increase registration quality. Best results in bold, worst result in italics.

	UNREGISTERED	SIFT	ANN	ANN+MASK(ours)	MANUAL
average score [0–5]	1.42	<i>0.67</i>	3.12	3.20	3.88
acceptance rate [%]	8	<i>12</i>	55	60	65

scores than ANN, i.e., Superglue alone. While these results are not on par with the upper baseline MANUAL, the manually registered images by re.photos users, they are

more time efficient, i.e., no or less user interaction is required. As shown in Figure 8, the average is influenced by failed registrations. SIFT results are worse than the lower



FIGURE 7. Registration results of the images and feature points shown in Figure 6. Registration results of the ANN and ANN+MASK method are comparable to the ground truth MANUAL method. SIFT fails in several cases resulting in worse images than UNREGISTERED. ANN and ANN+MASK differ, since ANN+MASK registers the AOI better but may register other areas worse compared to ANN. See Figure 6 for license information.

bound UNREGISTERED. This is in accordance with the quantitative registration results—compare Table 2. Analog to the image selection process for the creation of the database, see Section III-A, registration quality is deemed acceptable if the score is higher than or equal to 4. As indicated in the second row of Table 3 the baseline for this measure is 65%, ANN reaches 55%, while the inclusion of masks in ANN+MASK reaches 60% and closes half of the gap from ANN to the upper baseline MANUAL.

The scores, displayed in Table 3, substantiate Table 2 and indicate that ANN+MASK performs better than ANN alone. This slight difference may suggest that there are not many images in the dataset for which masks are required—informally, a noticeable enhancement in registration quality was observed in the image pair of Figure 1. The MANUAL method, as conducted for the re.photos portal, yields the highest score, as anticipated, due to the time-intensive nature of this approach. Further analysis, depicted in Figure 8, reveals that the score for the ANN and the ANN+MASK method follows a U-shaped distribution, with some image

pairs yielding subpar outcomes and others achieving favorable results when compared to the upper baseline MANUAL. This suggests that while the ANN methods can substitute MANUAL methods for certain image pairs, there are cases where manual or interactive registration remains necessary. Registration using SIFT receives a lower score compared to the original UNREGISTERED image pairs, evident in both Table 3 and Figure 8, underscoring the challenges of image registration on this dataset. ORIGINAL images predominantly receive a score of 1, as anticipated and consistent with Table 1; scores for MANUAL are roughly evenly distributed among the top three scores. Since only images with a re.photos score of at least 4 were selected, this indicates that the re.photos score is not solely stringent and might consider content and other factors.

C. USABILITY

Users further stated that registration results with our method were impressive, and that they were able to register image with this method which could not be registered by the

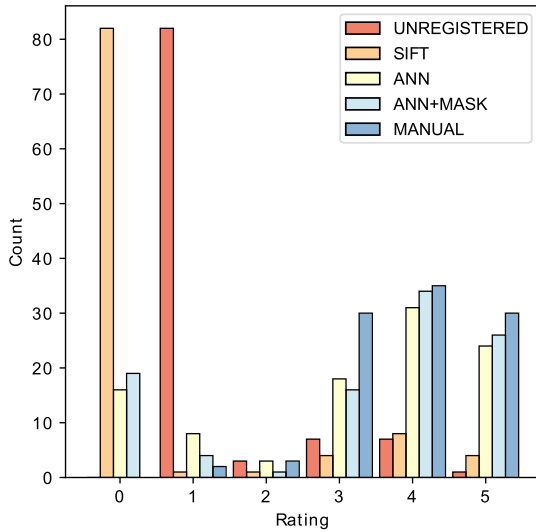


FIGURE 8. Scores of the ANN and the ANN+MASK method follows an u-shaped distribution with some image pairs failing and some archiving similar or better results compared to the MANUAL method. SIFT scores are worse than UNREGISTERED scores.

MANUAL method. One person planned to register image pairs registered with the MANUAL method again with our method to achieve better results. Others claimed that the combination of easy use combined with high registration quality is a game changer for rephotographic image registration. Several users spoke in favor to combine our method with the MANUAL method for fine-tuning and image pairs for which our method fails. Overall, our registration method received positive feedback. Following additional user feedback on our web interface, criticizing the lack of guidance and explanation of our method, an video was added to guide through the registration process and explain the difference between ANN and ANN+MASK. Some users preferred to see both images, while or before drawing the mask, in order to identify if an AOI is visible in both images. However, screen space is a limited resource, displaying both images next to each other at a lower scale, makes mask drawing more difficult. A possible solution would be the option to easily swap between displaying either the fixed or the moving image. Alternatives to freehand drawing of masks were also requested. For this reason other input methods were examined, see the following Section IV-D.

D. MASK DRAWING METHOD

As described above, mask are created by the user, by freehand drawing a line around the area to be masked. Other input method could be used as well: (1) Users can mark the corner points of a polygon, eliminating the need of exact freehand drawing and resulting in an approximation of the freehand form. (2) Users can span a rectangle by pressing the mouse button at one corner and then dragging the mouse to the opposite corner where the mouse button is released. The registration quality of the second method was tested by computing the bounding box of the user provided masks

TABLE 4. The method used to draw the mask has only a small influence on the RPE. The method, i.e., spanning a rectangle or drawing a form freehand, can be chosen by user preferences and Human Computer Interaction (HCI) criteria.

RPE	SIFT		ANN	
	freehand	rectangle	freehand	rectangle
Min	0.004	0.004	0.003	0.003
Q1	0.319	0.287	0.014	0.012
Med	0.873	0.818	0.024	0.024
Q3	1.498	1.355	0.088	0.118
Max	22.050	15.878	3.169	6.238

TABLE 5. RANSAC improves registration quality for ANN but not for SIFT compared to Least Squares. Possibly, because the number of SIFT descriptor matched feature points is already low, such that removal of additional points increases the RPE.

RPE	SIFT		ANN	
	RANSAC	Least Squares	RANSAC	Least Squares
Min	0.000	0.025	0.002	0.004
Q1	0.454	0.533	0.010	0.018
Med	0.969	0.765	0.024	0.043
Q2	1.585	1.100	0.102	0.319
Max	951.699	111.345	16.588	60.917

and using these rectangular bounding boxes as masks for the registration.

As indicated by Table 4 the rectangular masks have a slightly lower RPE for both SIFT and ANN. Results were computed as in Section IV-A. From the image registration standpoint both method are usable. User feedback suggest that the rectangle method is known from other applications and easier to use. Moreover, multiple rectangles can be used to generate more complex masks. These observation suggest that method (1), a polygon spanned by user marked points, will lead to similar registration quality results and could be used, too, if preferred by the users.

E. TRANSFORMATION COMPUTATION METHOD

Additionally, two methods used to compute the perspective transformation matrix were compared: (1) The regular least squares method using all detected feature points and (2) RANSAC [25] as robust method. For both the OpenCV implementation was used. The findings in Table 5 show that RANSAC improves the registration quality for ANN but not for SIFT. All results presented in the previous sections use RANSAC. The reason, why RANSAC performs worse for SIFT, may be in the limited number of points detected and descriptor matched, when using SIFT. Without a sufficient number of high quality matched point pairs masking does not increase registration quality. Further, the confidence parameter of the transformation computation method were varied but no notable improvement in registration quality was found when compared with the default value of 0.995.

V. DISCUSSION

As argued in Section I, automatic image registration for rephotographs is impossible when only some AOI of the OOI can be registered. It depends on the user's intentions for which OOI the images will be registered. The AOI masked-based methods require only little user input to embed the users' intentions into the registration process. However, drawing an AOI mask in one of the images is faster and requires less accuracy than manually marking corresponding point pairs in both images, as it is necessary for other interactive registration methods. Based on the currently published state-of-the-art in rephotography, no other intuitive and faster ways to embed intentions into the registration process are available to the users. As always, a compromise has to be found, weighting the resulting registration accuracy against the required time and effort spent by the user for the registration process: On the left side of the scale are previous interactive registration methods like MANUAL, requiring slow and precise input by the user but resulting in high-quality registrations. On the right side are automatic registration methods like ANN, requiring no user input but may result in unacceptable registration results. ANN+MASK resides in the middle of the scale.

A suitable workflow to register one or several rephotographic compilations could work from right to left through the different methods on the scale. In the first round, the images are registered automatically. Pairs not meeting the registration standards can then be registered with AOI mask-based registration. If the results do not meet the registration standard, these remaining images can be registered with the more expensive interactive registration methods. Thus, only the minimum required time and effort is spent for each image pair.

VI. FUTURE RESEARCH

Two areas are of interest for future research. (1) Automatic registration methods can be further adapted for the registration of images, taken with contemporary cameras, and historical images, not limited to analog photographs but also prints, drawings, paintings, and other works of art. Databases of historical and contemporary images exist [7] but are missing for works of art and contemporary images. However, development of registration methods would profit from these databases. (2) Embedding of intentions could be further improved by automatic image segmentation, providing different masks the user can simply select. Image segmentation could also be combined with text or voice command controlled AIs. However, this method could be slower and more prone to errors than simply drawing a mask into the image.

VII. CONCLUSION

This paper presents the highly accurate rephotographic image registration by AOI masks and automatic image registration methods adapted to rephotography. Registration for rephotography is challenging; users often want or need

to decide which areas are to be registered. Taking this need into account an improved method is offered as a web service to incorporate user intentions with masks in the registration process. Additionally, a new interface to a completely automatic deep learning-based registration method is provided. Both methods are combined in one application, allowing users to choose between the automated method and selecting the AOI they intend to register.

Compared to traditional feature-based automatic registration methods, deep learning-based methods achieve significant improvements, increasing the acceptance rate by 43 *p.p.* from 12 % for hand engineered features detectors and descriptors to 55 % for trained features detectors and descriptors. This increased acceptance rate results in a reduced number of images requiring manual registration. The AOI mask-based registration method raises the acceptance rate further to 60 %, while traditional labor intensive manual registration achieves 65 %, indicating that for 35 % of the images, achieving acceptable image registration is challenging even with manual registration methods based on hand-selected feature points. Compared to manual registration, AOI mask-based registration requires less user interaction, precision, and knowledge. Users no longer need to identify and precisely mark corresponding points in both images but create a rough freehand AOI mask in only one image. For many use cases the results of the automatic method are acceptable. For use cases in which only one OOI can be registered the AOI masks can embed the users' intentions in the registration process, leading to good registration results in this use cases. This intuitive user interaction based on AOI masks also opens up possibilities for in-field mobile image registration, e.g., for climate change research or disaster management [26], as well as directly verifying the correctness of the camera position when taking the rephotograph and, in the future, providing guidance for locating this camera position.

REFERENCES

- [1] A. Schaffland and G. Heidemann, "Heritage and repeat photography: Techniques, management, applications, and publications," *Heritage*, vol. 5, no. 4, pp. 4267–4305, Dec. 2022.
- [2] R. H. Webb, *Repeat Photography: Methods and Applications in the Natural Sciences*. Washington, DC, USA: Island Press, 2010.
- [3] G. F. Rogers, H. E. Malde, and R. M. Turner, *Bibliography of Repeat Photography for Evaluating Landscape Change*. Salt Lake City, UT, USA: Univ. Utah Press, 1984.
- [4] R. Chen and C. Gotsman, "Generalized as-similar-as-possible warping with applications in digital photography," in *Proc. 37th Annu. Conf. Eur. Assoc. Comput. Graph.* Eindhoven, The Netherlands: Eurographics Association, 2016, pp. 81–92.
- [5] A. Schaffland, O. Vornberger, and G. Heidemann, "An interactive web application for the creation, organization, and visualization of repeat photographs," in *Proc. 1st Workshop Structuring Understand. Multimedia Heritage Contents*, New York, NY, USA, Oct. 2019, pp. 47–54.
- [6] A. Schaffland, T. H. Bui, O. Vornberger, and G. Heidemann, "New interactive methods for image registration with applications in repeat photography," in *Proc. 2nd Workshop Structuring Understand. Multimedia Heritage Contents*. New York, NY, USA, Oct. 2020, pp. 41–48.
- [7] A. Schaffland and J. Schöning, "Faroe islands rephotography image registration dataset," *Data Brief*, vol. 47, Apr. 2023, Art. no. 108979.

- [8] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 337–33712.
- [9] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jun. 2020, pp. 4937–4946.
- [10] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.
- [11] A. A. Goshtasby, *Theory and Applications of Image Registration*. Hoboken, NJ, USA: Wiley, 2017.
- [12] J. V. Hajnal and D. L. Hill, *Medical Image Registration*. Boca Raton, FL, USA: CRC Press, 2001.
- [13] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, *Image Registration for Remote Sensing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [14] K. Kuppa, S. Banda, and T. R. Barige, "An overview of deep learning methods for image registration with focus on feature-based approaches," *Int. J. Image Data Fusion*, vol. 11, no. 2, pp. 113–135, Jan. 2020.
- [15] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, Jan. 2021.
- [16] H. R. Boveiri, R. Khayami, R. Javidan, and A. Mehdizadeh, "Medical image registration using deep neural networks: A comprehensive review," *Comput. Electr. Eng.*, vol. 87, Oct. 2020, Art. no. 106767.
- [17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, New York, NY, USA, Sep. 1999, pp. 1150–1157.
- [19] R. Sinkhorn, "Diagonal equivalence to matrices with prescribed row and column sums," *Amer. Math. Monthly*, vol. 74, no. 4, pp. 402–405, Apr. 1967.
- [20] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1978–1983.
- [21] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, Jan. 2009.
- [22] J. Križaj, V. Štruc, and N. Pavešic, "Adaptation of sift features for face recognition under varying illumination," in *Proc. 33rd Int. Conv. MIPRO*, 2010, pp. 691–694.
- [23] J. Wu, Z. Cui, V. S. Sheng, P. Zhao, D. Su, and S. Gong, "A comparative study of SIFT and its variants," *Meas. Sci. Rev.*, vol. 13, no. 3, pp. 122–131, Jun. 2013.
- [24] G. Bradski, "The OpenCV library," *Dr Dobb's J. Softw. Tools*, vol. 25, pp. 120–125, Nov. 2000.
- [25] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [26] A. Schaffland, J. Schöning, O. Vornberger, and G. Heideman, "Climate change combat and disaster management with re-photos, the web portal for rephotography," in *Proc. IEEE Int. Humanitarian Technol. Conf. (IHTC)*, Dec. 2022, pp. 69–72.



AXEL SCHAFFLAND received the B.Sc. and M.Sc. degrees in cognitive science from Osnabrück University, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the Institute of Cognitive Science, with a focus on rephotography and image registration techniques. He is a Research Assistant with the Osnabrück University of Applied Science on intelligent transport systems. His research interests include didactics of artificial intelligence and computer science unplugged.



JULIUS SCHÖNING (Senior Member, IEEE) received the Ph.D. degree in computer vision from the Institute of Cognitive Science, Osnabrück University. He was a Systems Architect and a Researcher focused on self-driving vehicles with ZF Friedrichshafen AG. He is currently a Full Professor with the Osnabrück University of Applied Sciences, with a focus on AI-in-the-loop. He is also a Software Requirement Engineer for agricultural HMIs and has studied mechatronics as well as intelligent embedded microsystems.

...