Solving photogrammetric cold cases using AI-based image matching : new potential for monitoring the past with historical aerial images

Ferdinand Maiwald, Denis Feurer, Anette Eltner

ISPRS Journal of Photogrammetry and Remote Sensing 206 (2023) 184–200

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Solving photogrammetric cold cases using AI-based image matching: New potential for monitoring the past with historical aerial images

Ferdinand Maiwald [a,*], Denis Feurer [b], Anette Eltner [a]

[a] Institute of Photogrammetry and Remote Sensing, TU Dresden, Helmholtzstraße 10, Dresden, 01069, Germany
[b] UMR LISAH, Univ. Montpellier, AgroParisTech, INRAE, IRD, Institut Agro Montpellier, 2 place Pierre Viala, Montpellier, 34060, France

ARTICLE INFO

Keywords:
Historical aerial images
Feature matching
Neural networks
Structure-from-motion
Digital surface model
Multi-temporal

ABSTRACT

With the ongoing digitization in archives, an increasing number of historical data becomes available for research. This includes historical aerial images which provide detailed information about the depicted area. Among the applications enabled by these images are change detection of land use, land cover, glaciers, and coastal environments as well as the observation of land degradation, and natural hazards. Studying the depicted areas and occurring 3D deformations requires the generation of a digital surface model (DSM) which is usually obtained via photogrammetric Structure-from-Motion (SfM). However, conventional SfM workflows often fail in registering historical aerial images due to their radiometric characteristics introduced by digitization, original image quality, or vast temporal changes between epochs. We demonstrate that the feature matching step in the Structure from Motion (SfM) pipeline is particularly crucial. To address this issue, we apply the two synergetic neural network methods SuperGlue and DISK, improving feature matching for historical aerial images. This requires several modifications to enable rotational invariance and leveraging the high resolution of aerial images. In contrast to other studies our workflow does not require any prior information such as DSMs, flight height, focal lengths, or scan resolution which are often no more extent in archives. It is shown that our methods using adapted parameter settings are even able to deal with quasi texture-less images. This enables the simultaneous processing of various kind of mono-temporal and multi-temporal data handled in a single workflow from data preparation over feature matching through to camera parameter estimation and the generation of a sparse point cloud. It outperforms conventional strategies in the number of correct feature matches, number of registered images and calculated 3D points and allows the generation of multi-temporal DSMs with high quality.

With the flexibility of the method, it enables the automatic processing of formerly unusable or only to be interactively processed data, e.g. aerial images where the flight route is unknown, or with difficult radiometric properties. This makes it possible to go back even further in time, where the data quality usually decreases, and enables a holistic monitoring and comparison of environments of high interest. The code is made publicly available at https://github.com/tudipffmgt/HAI-SFM.

1. Introduction

In environmental remote sensing, historical aerial images can be used in various fields such as change detection of land use and land cover (Picon-Cabrera et al., 2020), glaciers (Mölg et al., 2019; Andreassen et al., 2020), forests (Vastaranta et al., 2015; Berveglieri et al., 2018), coastal environments (Warrick et al., 2017) or for the observation of land degradation (Bolles and Forman, 2018), landslides (DeWitt and Ashland, 2023; Soldato et al., 2018) and natural hazards (Wang et al., 2021). Compared to satellite images, which are also capable of mapping large areas, historical aerial images provide the distinct advantage of extending temporal coverage by approximately 100 years, dating back to around 1858 (Albertz, 2009). Even after the advent of earth observation satellites around 1960 (CORONA program), historical aerial images often remain the only existing records capturing landscapes from around the world with remarkable detail (Pinto et al., 2019). With an increase of digitization in archives, automatic processing of these historical aerial images gains even more relevance.

However, automatic analysis and processing of historical aerial images presents various challenges. For instance, it is important to carefully document the metadata and the digitization process if high geometric accuracy wants to be achieved (Sevara, 2016). Currently, archives deal with challenges, such as physical preservation, integration

* Corresponding author.
E-mail address: ferdinand.maiwald@tu-dresden.de (F. Maiwald).

https://doi.org/10.1016/j.isprsjprs.2023.11.008
Received 26 June 2023; Received in revised form 6 November 2023; Accepted 7 November 2023
Available online 16 November 2023

of repositories, and metadata standards (Pinto et al., 2019; Giordano et al., 2018). Different digitization procedures achieve inconsistent and varying quality standards, potentially leading to varying digitization products between epochs, which complicate especially inter-epoch processing of historical aerial images.

Further issues arise during automatic processing. Metadata is often lost so that the flight area, camera parameters, and possible ground control points (GCPs) are unknown and sometimes, the fiducial marks are missing (Giordano et al., 2018). Furthermore, the presence of low overlap in flight strips and/or image pairs can greatly complicate the process of automatically searching for tie points. Additionally, the depicted area may provide a challenging environment, e.g., in forests due to movement of canopy between image pairs (Velasco et al., 2022), or over larger water bodies where usually no tie points are found (Rupnik et al., 2015; Carrivick and Smith, 2018).

Previous studies introduced different automatic approaches to align historical aerial images. Thereby, time-invariant line features (Nagarajan and Schenk, 2016) or patch based methods (Craciun and Bris, 2022) were developed, considering recent digital surface models (DSMs) and orthophotos as reference and thus simultaneously allowing for automatic georeferencing. Further, Giordano et al. (2018) propose a method for finding contemporary GCPs in historical aerial image data. Others use multi-temporal tie points which can be tracked over several epochs, originally introduced as Time-SIFT method (Feurer and Vinatier, 2018) and further generalized by Cook and Dietze (2019) and Blanch et al. (2021). The most recent studies use contemporary DSMs via a co-registration algorithm to georeference the SfM products of the historical data (Zhang et al., 2021; Knuth et al., 2023). Zhang et al. (2021) introduce a complete workflow for processing specific historical aerial image datasets including the novel use of the Artificial Intelligence (AI) based feature matching method SuperGlue (Sarlin et al., 2020). These works depict the complexity of processing historical aerial image data and while providing comprehensive toolsets such as using AI, DSM co-registration and multi-temporal tie points, a general processing strategy for diverse datasets is not yet available.

This is mainly due to the following persistent issues. Digitized historical (aerial) images often suffer from poor image quality resulting in close-to-zero correct feature matches (Maiwald, 2019; Zhang et al., 2021; Morelli et al., 2022). Further, even when applying learned feature matching methods, the texture quality and visual changes, particularly in forested areas, remain challenging. Another problem is the need for definition of hypotheses for every specific dataset. This includes assumptions such as that the topography is variable enough for automatic registration (Zhang et al., 2021; Craciun and Bris, 2022), GCPs in the study area (Giordano et al., 2018) exist, enough image information for finding distinctive features (Feurer and Vinatier, 2018), and flight information is provided (Knuth et al., 2023).

Novel learned feature matching methods in combination with SfM are often able to initially process data without any assumptions. However, they are still challenging to use on historical aerial images. That is because of their original training procedure on mainly terrestrial data and low image resolutions (Dusmanu et al., 2019; Sarlin et al., 2020; Tyszkiewicz et al., 2020; Sun et al., 2021), which also results in a lack of rotational invariance as reported in Tyszkiewicz et al. (2020). Retraining the networks would require labeled historical aerial image datasets, high computational resources and the openly available training protocol, which is sometimes not accessible (Sarlin et al., 2020).

Our approach provides a general solution for processing historical aerial images unlocking the potential of existing aerial image archives even for difficult datasets. The study deals particularly with the automatic detection and matching of tie points in challenging historical aerial image pairs. Finding these points enables a subsequent Structure-from-Motion (SfM) pipeline to model 3D geometry, including the estimation of camera parameters, in one single workflow. We are using *learned feature matchers* based on deep neural networks due to the limited efficacy of conventional feature matching techniques applied to historical datasets. Specifically, we focus on a combination of two AI-based image matching strategies, i.e., SuperGlue (Sarlin et al., 2020) and DISK (Tyszkiewicz et al., 2020), while presenting strategies for dealing with high-resolution aerial images and rotations between subsequent image pairs. As this workflow initially operates in image space only, it avoids the need to make further assumptions while leveraging the potential of learned feature matching methods. With the aim of providing a complete workflow, all feature matches and images are imported into a open-source SfM software where all camera parameters and 3D points are calculated simultaneously using bundle adjustment. The tie point accuracy and the quality of the resulting DSMs is metrically estimated for one of the processed datasets. The code is openly available on https://github.com/tudipffmgt/HAI-SFM including a small sample image dataset from the TIME benchmark (Farella et al., 2022).

2. Materials and methods

Our developed approach to match challenging historical aerial images is tested for two different study areas. The feature matching methods and their different parameter settings used on the datasets are explained in detail. Subsequently, a strategy for estimating the metric accuracy of our method is presented.

2.1. Data

2.1.1. Congo dataset (1961)

The first dataset contains images which could neither be processed in conventional SfM software nor with the adaptations shown in several other works (Feurer and Vinatier, 2018; Zhang et al., 2021; Knuth et al., 2023). Only limited information about the images is available. Flight plan, flight height, scan size, and the focal length of the camera were unknown. The site is located in the tropical rainforest and the images depict a network of rivers in Congo in Central Africa. All 31 images are taken in 1961 and the only external information is the fact that they are consecutively numbered from 176 to 192 (17 images) and from 222 to 231 (10 images) according to their positions in successive flight strips (Fig. 1).

An overview of the data is given in the upper part of Table 1.

The original images have a resolution of 11 400 × 11 408 pixels. Resampling them, using the available fiducial marks, resulted in a final resolution of 10 500 × 10 500 pixels. Resampling was done using the ReSampFid function of MicMac which detects the center of the fiducial marks and aligns the images respectively (Rupnik et al., 2017).

2.1.2. Occitanie dataset (1971–2001)

The second test site extends over an area of a 170 km^2 in the Mediterranean landscape in Occitanie in southern France (43°5N, 3°19E) (Fig. 2).

The area of interest is mainly covered by vineyards with forests in its uppermost regions. The area exhibited a strong change in land use because vineyards transformed during the 1980s, from goblet to trellised vineyards, accompanied by progressive land abandonment and urbanization during the past 50 years (Vinatier and Arnaiz, 2018). The fact that these images were scanned with photogrammetric scanners is important in regard of the production of geometrically correct outputs as noted by Sevara (2016). We used a sample of these data from four different epochs that allowed for a stereo coverage of the entire test site. The characteristics of the images are reported in the bottom part of Table 1.

In order to test the different SfM approaches as if camera data is unknown, all scanning information and interior camera orientation parameters were initially omitted. In the post-processing step, when estimating the accuracy of the multi-temporal DSMs, the camera parameters were used as described in Section 2.3.

Fig. 1. Sequential historical aerial images of the Congo dataset. Similar regions are marked with a red circle.

Table 1
Characteristics of the two datasets used in the experiments.

Congo dataset					
Epoch (year)	Focal length (mm)	Estimated flight height (m)	Images (#)	Scan size (μm)	GSD (cm)
1961	n.a.	n.a.	27	n.a.	n.a.
Occitanie dataset					
1971	152	2700	61	21	37
1981	153	4800	27	21	66
1990	153	5000	31	21	69
2001	153	4000	44	21	55

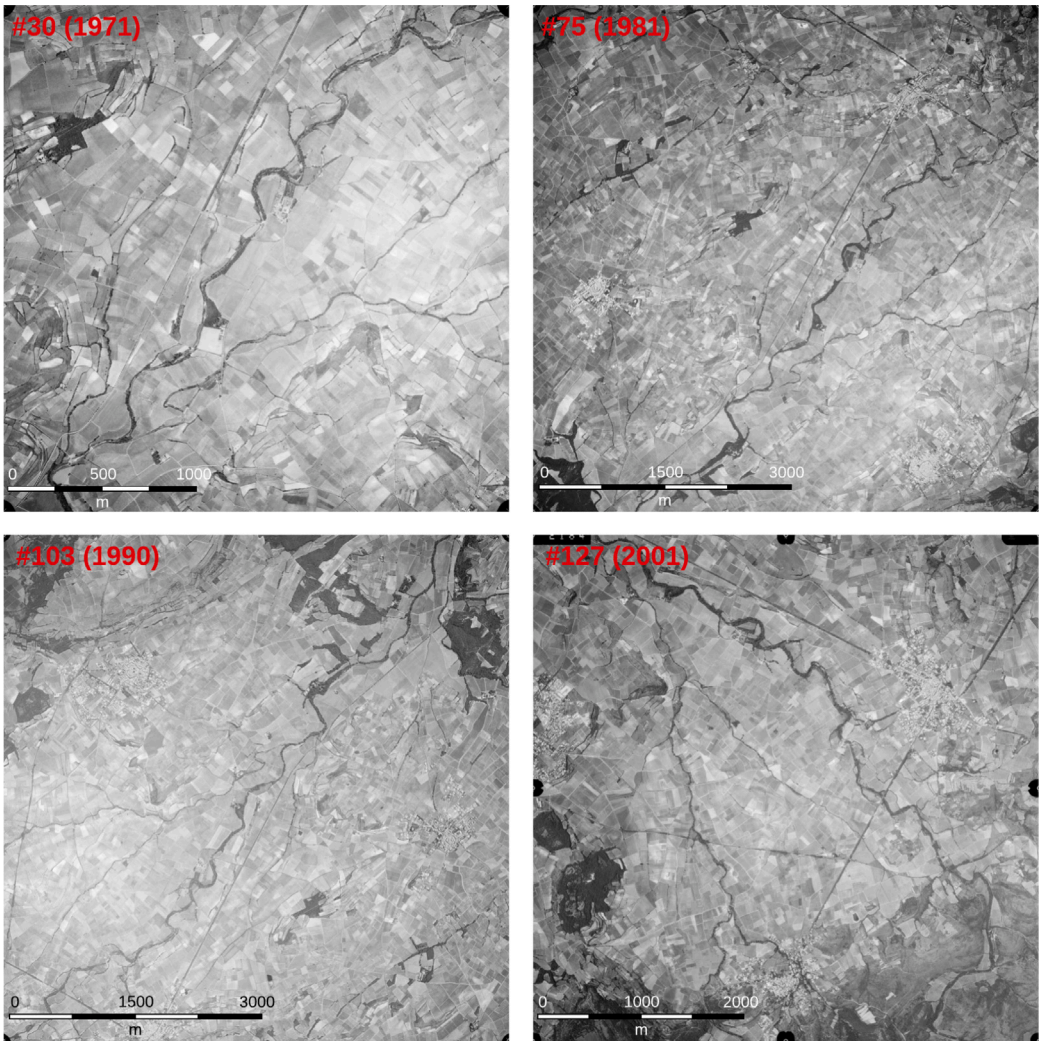

Fig. 2. Historical aerial image of every epoch of the Occitanie dataset showing the similar geographic region.

2.2. Methods

Two different methods were considered to process historical aerial images and especially to robustly find tie points between sequential image pairs. Both approaches have already shown good results on historical *terrestrial* images (Maiwald, 2022; Morelli et al., 2022; Maiwald et al., 2023). The first method combines the tie point extractor SuperPoint (DeTone et al., 2018) with the feature matching method SuperGlue (Sarlin et al., 2020) and will in the following be referred to as SuperGlue. The second method is DISK (Tyszkiewicz et al., 2020). Both methods are experimentally used to find mono-temporal tie points for the first dataset and mono-temporal and multi-temporal tie points for the second dataset. A short summary and explanation on the parameters and limitations are given in the following sub-sections.

After tie points are found with the AI-based approaches, bundle adjustment is performed in COLMAP to reconstruct the image geometry (Schönberger and Frahm, 2016). Thereby, the camera parameters and a sparse point cloud are calculated simultaneously.

2.2.1. SuperPoint+SuperGlue

SuperPoint serves as method for the feature detection stage (DeTone et al., 2018). SuperPoint generates reliable feature points using a synthetic pre-trained dataset to train the so-called MagicPoint convolutional neural network. Because the model does not perform sufficiently accurate on real images, DeTone et al. (2018) use Homographic Adaptation to transform the input image to multiple random homographic representations. Subsequently, the feature points are calculated with MagicPoint for all representations and are eventually aggregated to one final set of SuperPoint features.

Afterwards the extracted feature points are matched with SuperGlue. SuperGlue is a graph neural network designed for matching tie points by solving a differentiable optimal transport problem.

However, the methods come with several limitations especially regarding historical aerial images. The training procedure of SuperGlue is at the moment not openly available, which means that the user can only use the pre-defined configurations *indoor* and *outdoor* for feature matching. Sarlin et al. (2020) use scenes from the MegaDepth (Li and Snavely, 2018) dataset to train the outdoor configuration which however mainly consists of urban scenes. Thus, the parameter configuration might not be optimal for our historical aerial images.

We propose a *modified outdoor configuration* using a maximum edge length of 1600 pixels in SuperGlue in contrast to the default value of the outdoor configuration which is set to 1024 pixels. We use this value because it is recommended by Sarlin et al. (2020) to use images with a maximum image edge length of 1600 pixels. However, historical aerial images are usually digitized with a significantly higher resolution between approximately 10 000 × 10 000 pixels to about 26 000 × 26 000 pixels. Therefore, we propose a tile-based approach (t-ba), which splits the original images $I_i, i = 1, \ldots, N$ into $M \times N$ tiles with a maximum edge length of 1600 pixels. In order to reduce computation time, not every tile $I_i(M \times N)$ is matched with every tile $I_{i+1}(M \times N)$ and instead the overlap $O_{i,i+1}$ between images is considered (Fig. 3).

The overlap might be available from metadata. If this is not the case, we assume a maximum strip overlap in flight direction of 60% and a strip overlap across flight direction of 30%. Consequently, images are not matched to neighboring images but to corresponding tiles I_{i+2}, I_{i+3} or I_{i+4}. If every image should be considered for finding tie points, the overlap has to be set to 100% respectively.

SuperGlue is only rotational invariant up to approximately 45° (Tyszkiewicz et al., 2020). We use this property to automatically detect the beginning of the following flight strip. If there are less than 50 geometrically verified matches between sequential images with a matching confidence higher than 0.5, image I_{i+1} is rotated for 180° and feature matching is repeated (Fig. 4). If the repeated image matching results in a significantly higher number of matches, a new flight strip has been identified.

The found feature matches are stored in the Hierarchical Data Format (HDF) H5, which is readable and processable by the SfM software COLMAP.

2.2.2. DISK

DISK uses reinforcement learning for extracting and matching a set of local features in an end-to-end trainable approach. Features are extracted using a U-Net architecture and their optimal matching is obtained by using the policy gradient method. Again, training is done on the MegaDepth dataset and not modified during the experiments. Usually, the approach finds less correct matched features than SuperGlue but provides a higher absolute number of detected feature points (Maiwald et al., 2021; Morelli et al., 2022).

DISK deploys a set of adjustable parameters. Images can be used at full size but also internally downsampled to a pre-defined width and height that has to be a multiple of 16. In our experiments we used resolutions of 1024 × 1024, 1600 × 1600, and 3200 × 3200 pixels. Larger image sizes are not possible on the NVidia A100 GPU, used in this study, due to memory limitations. Furthermore, it is possible to set the maximum numbers of features that are extracted and the mode of feature extraction. Modes that can be used are non-maxima suppression (*nms*) or through training-time grid sampling technique (*rng*) as explained in Tyszkiewicz et al. (2020). Basically, nms allows finding multiple keypoints (or also none) per grid cell (with a prior defined size) while rng always provides exactly one keypoint per cell. In order to find an optimal solution for historical aerial images, the default approach and subsequently different combinations of parameter settings are used. Initially, we increase the image resolution parameter while retaining a constant number of maximum features per image. After the maximum image resolution is detected, we double the maximum number of detected features per experiment until the processable limit is reached. The resulting feature matches are again stored in H5 file format. As recommended, COLMAP is used for the geometric verification of the feature matches using its standard procedure (Schönberger and Frahm, 2016). That means, that the putative matches derived from DISK are filtered using the geometric properties of the Fundamental Matrix including outlier detection with Progressive Sampling Consensus (Chum and Matas, 2005).

2.2.3. COLMAP

COLMAP (https://github.com/colmap/colmap, accessed on 03 February 2023) is an open-source SfM software that can be modified enabling, e.g., the import of calculated feature matches (Schönberger and Frahm, 2016). In this study, COLMAP serves as a tool to simultaneously estimate interior and exterior camera orientation (if unknown) and to generate a sparse point cloud representing the 3D coordinates of the tie points. However, similar to the feature matching methods, COLMAP is mainly designed to reconstruct urban landscapes and less to process aerial images. Thus, some modifications need to be made. For instance, if the principal distance (focal length) f is known, it should be set as prior focal length to achieve valid results and if it is unknown it should still be initialized with a reasonable estimate, e.g. by comparing to other flight missions at that time and place. Note that in COLMAP the default initial value of $f = 1.25 \cdot \max(image_{width}, image_{height})$, which is not a suitable estimate for aerial images.

Using the default settings, COLMAP triangulates 3D points only if they are seen in three or more images, i.e., corresponding to the so-called (multi-view) feature tracks. We adapt that behavior by allowing two-view feature tracks during bundle adjustment to enable triangulation of 3D points seen in a minimum of two images considering the smaller overlap of historical aerial images. If desired the principal distance can be set to be fixed in the bundle adjustment, especially if a stable camera geometry can be assumed for the high quality aerial cameras developed for mapping tasks. This might avoid inaccurate reconstructions or dome effects.

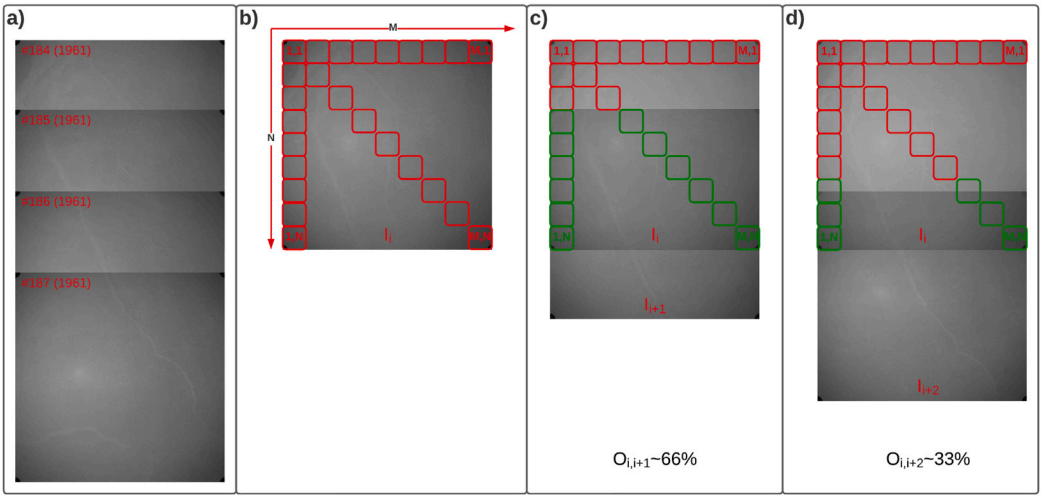

Fig. 3. Explanation of the tile-based approach (t-ba) illustrated for the Congo dataset. (a) Image sequence of four historical aerial images. (b) Due to their high resolution images are split into $M \times N$ tiles. (c) Only possibly overlapping tiles in green are matched with the subsequent image. (d) The second next image is matched with even less possible combinations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

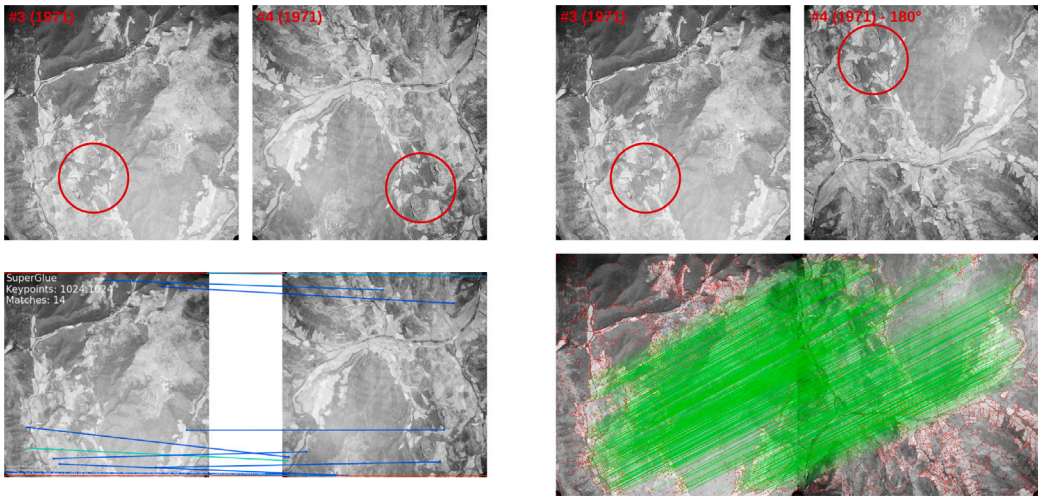

Fig. 4. Left: original image pair configuration where only incorrect feature matches with mostly low confidence (blue lines) can be found using SuperGlue (and DISK). Right: Image configuration with the subsequent image rotated by 180°. The image rotation enables finding a lot of correct feature matches depicted using green lines between the identified red tie points for SuperGlue (and DISK). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2.4. Proposed workflow using a combination of methods

For maximum efficiency a synergetic workflow is proposed using the contrary properties of both AI-based image matching methods (Fig. 5).

Step 1: Data preprocessing:

For faster preprocessing the images are downsampled to a maximum edge length of 1600 pixels using bilinear interpolation while maintaining the aspect ratio. Afterwards, our workflow proposes three different strategies for reliably finding feature matches in historical aerial images. This includes the *default approach* using a combination of SuperGlue and DISK, the computationally expensive but feature-rich *tile-based approach*, and the *approach for texture-less images* using DISK exclusively.

Step 2: Determination of image rotation:

To determine the image rotation, all strategies start with the use of the default SuperGlue feature matching method on the downsampled images because SuperGlue usually outperforms other learned matchers even on difficult datasets (Maiwald, 2022; Morelli et al., 2022). If more than 50 feature matches with a $match\_confidence > 0.5$ are found between sequential images, SuperGlue can be used to identify consecutive flight strips as explained in Section 2.2.1, Fig. 4. If SuperGlue is not able to find a sufficient number of feature matches, especially in textureless scenes like for the Congo dataset, DISK is used instead to determine the image rotation in the downsampled images.

Step 3: Feature matching: Using the correct rotation, feature matching can be applied on the fully sized historical aerial images. Therefore, we propose two different strategies. If DISK was used to find feature matches (because SuperGlue failed), it should also be applied on the images with full resolution (Fig. 5, DISK approach). In our experiments a parameter setting with an image resolution of 3200 × 3200 pixels and a maximum allowed number of feature matches $n = 32768$ provided the most reliable results. DISK has the advantage of being directly applicable to the original image resolution because it uses a convolution neural networks. This allows yielding precise tie point coordinates (in sub-pixel accuracy), which is why it is proposed as our default approach. However, for more equally distributed *inter-epoch* tie points, SuperGlue can also be used on historical aerial images with full resolution. In this case, the images have to be divided into separate tiles to enhance robustness and accuracy as explained in Section 2.2.1. Our approach keeps the a-priori determined tie-points of the downsampled images and the overlap information to enable a smart matching scheme of the different image tiles. Still, this approach is more computationally

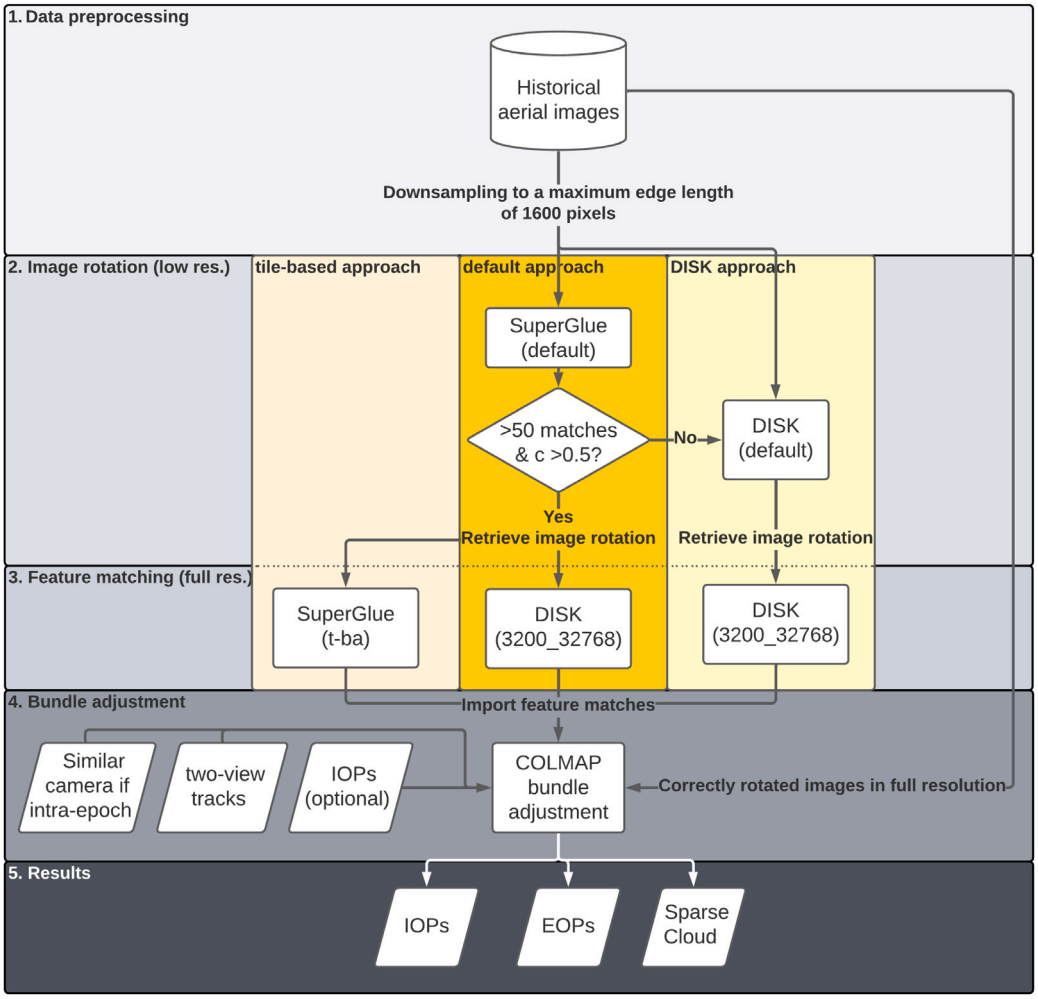

Fig. 5. Our proposed pipeline for processing historical aerial images. To accelerate initial processing, the images are downsampled first to determine the image rotation using SuperGlue as default method or DISK in the case of texture-less images. The rotated images are processed in full resolution using a specific DISK parameter setting with an image size parameter of 3200 × 3200 and a maximum number of extracted tie points of 32 768. An alternative is given using the tile-based approach for leveraging the high resolution of the original image material. The rotated images in full resolution, the derived feature matches, and interior camera parameters (if available) are used in the COLMAP bundle adjustment to determine interior and exterior camera parameters of all images as well as a sparse point cloud.

expensive because multiple tiles have to be matched to each other and all found tie points have to be merged into a single file for the respective image pair. As an example, a single photograph with an image resolution of 16 000 × 16 000 pixels is already divided into 100 separate tiles (with an image resolution of 1600 × 1600 pixels). Considering four successive images of one flight strip with 60% overlap this accumulates to a final number of 14 400 matching procedures as seen in Eq. (1) in comparison to the DISK matching process with only 5 procedures as calculated in Eq. (2).

$$\begin{aligned} pairs_{t-ba} &= P_{1-2} + P_{1-3} + P_{2-3} + P_{2-4} + P_{3-4} \\ &= 60 \cdot 60 + 60 \cdot 30 + 60 \cdot 60 + 60 \cdot 30 + 60 \cdot 60 \\ &= 14400 \end{aligned} \tag{1}$$

$$\begin{aligned} pairs_{DISK} &= P_{1-2} + P_{1-3} + P_{2-3} + P_{2-4} + P_{3-4} \\ &= 1 + 1 + 1 + 1 + 1 \\ &= 5 \end{aligned} \tag{2}$$

We are able to slightly reduce this high number of matching pairs by only matching the tiles that fall into the bounding box of pre-calculated matches in the downsampled image pairs.

Step 4: Bundle adjustment: The potentially rotated full-resolution photographs are imported in COLMAP and the image geometry is calculated using the feature matches. If interior camera orientation parameters (IOPs) are known a-priori, they will be imported. The bundle adjustment in COLMAP is performed allowing two-view tracks eventually resulting in the estimated interior and exterior camera orientation parameters (EOPs) for the historical aerial images and in a sparse point cloud.

The source-code to our introduced automatic workflow to match historical aerial images, including an experimental data sample is available at https://github.com/tudipffmgt/HAI-SFM. The final compilation of all tested and compared methods on the respective datasets can be seen in Table 2.

2.3. Accuracy assessment

Giving reliable estimates about the accuracy of historical data often proves to be difficult due to missing reference data also being visible in images of the past. Only the Occitanie dataset was explored to assess the accuracy of scaled 3D models because it was not possible to identify stable reference points in the Congo images. The area with the largest image overlap was considered to ensure a comparison across all four epochs (Fig. 6). Multiple procedures of accuracy assessment were carried out.

Reprojection error and tie point accuracy: The interior orientations, camera poses and sparse point clouds derived from our workflow were transferred from COLMAP to Metashape for further analysis.

Table 2
Methods and parameter settings used on the Congo and Occitanie dataset.

	SuperGlue	DISK
Congo dataset	Default	Default (1024 × 1024 pixels, 2048 features, nms)
	max. edge length 1600 pixels	1024 × 1024 pixels, 4096 features, nms
	1600 px + ReSampFid	1600 × 1600 pixels, 4096 features, nms
	1600 px + tile-based approach (t-ba)	3200 × 3200 pixels, 4096 features, nms
	1600 px + ReSampFid + t-ba	3200 × 3200 pixels, 4096 features, rng
		3200 × 3200 pixels, 8192 features, nms
		3200 × 3200 pixels, 16 384 features, nms
		3200 × 3200 pixels, 32 768 features, nms
		3200 × 3200 pixels, 65 536 features, nms
Occitanie dataset	Default	Default
	Proposed method using combination of SuperGlue and DISK	

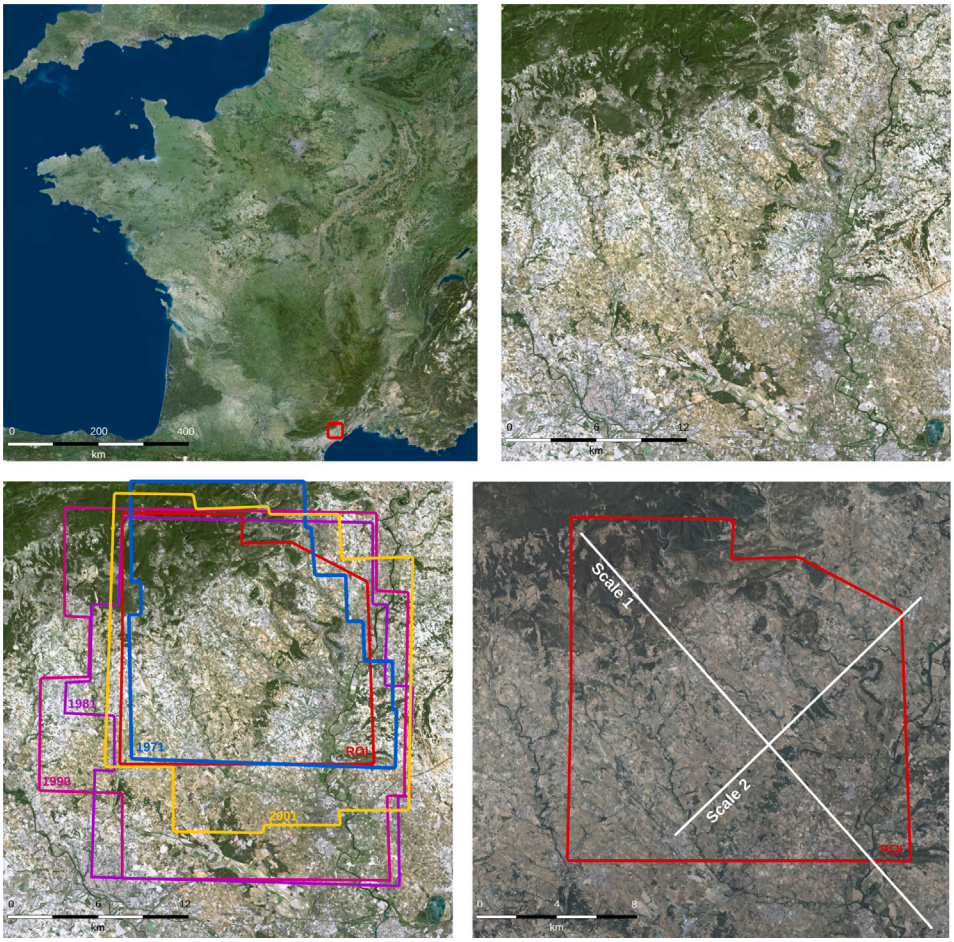

Fig. 6. Top-left: Location of region of interest (ROI) in France. Top-right: Example of one recent aerial image of the ROI. Bottom-left: Image overlap for all epochs and selected ROI. Bottom-right: Selected ROI transferred to a more detailed aerial view, including the display of the control scale bar (Scale 1) and the check scale bar (Scale 2) used for the accuracy assessment of scaled 3D models.
Image source: https://www.geoportail.gouv.fr.

Afterwards, the quality of our introduced SfM process was estimated using stable objects that were visible in all epochs. These points were measured manually in the images of all epochs and then compared to the corresponding projected image point. This allowed the visualization of the tie point accuracies and the calculation of the mean reprojection error across multiple epochs.

Multi-temporal comparison of scaled models: The model of the four epochs from Occitanie was scaled by introducing one control scale (Scale 1) into the adjustment of the camera alignment. During the alignment, the given flight protocol was also considered. For the independent estimation of the accuracy of the refined, i.e., scaled model, an additional check scale (Scale 2) was used. The 3D point coordinates to calculate the scales were obtained as UTM coordinates from the French geoportal (https://www.geoportail.gouv.fr/carte). The image measurement of the scales in the historical images were performed using the dataset from 2001 to ensure using points in the dataset with the smallest likelihood of change until recently. The a-priori accuracy of the control scale was set to 1 m. Scale 1 spanned the area from north-east to south-west with a length of 26.5 km. The check scale had a length of 17.6 km. The final error at Scale 2 (i.e., after the adjustment) was 0.49 m.

Four different historical point clouds were generated using solely images of the respective epoch, i.e., the originally merged model was separated by acquisition year. The dense point clouds were calculated in Metashape using *High* quality and *Mild* depth filtering settings. The 3D point clouds enabled the comparison of stable regions. Therefore,

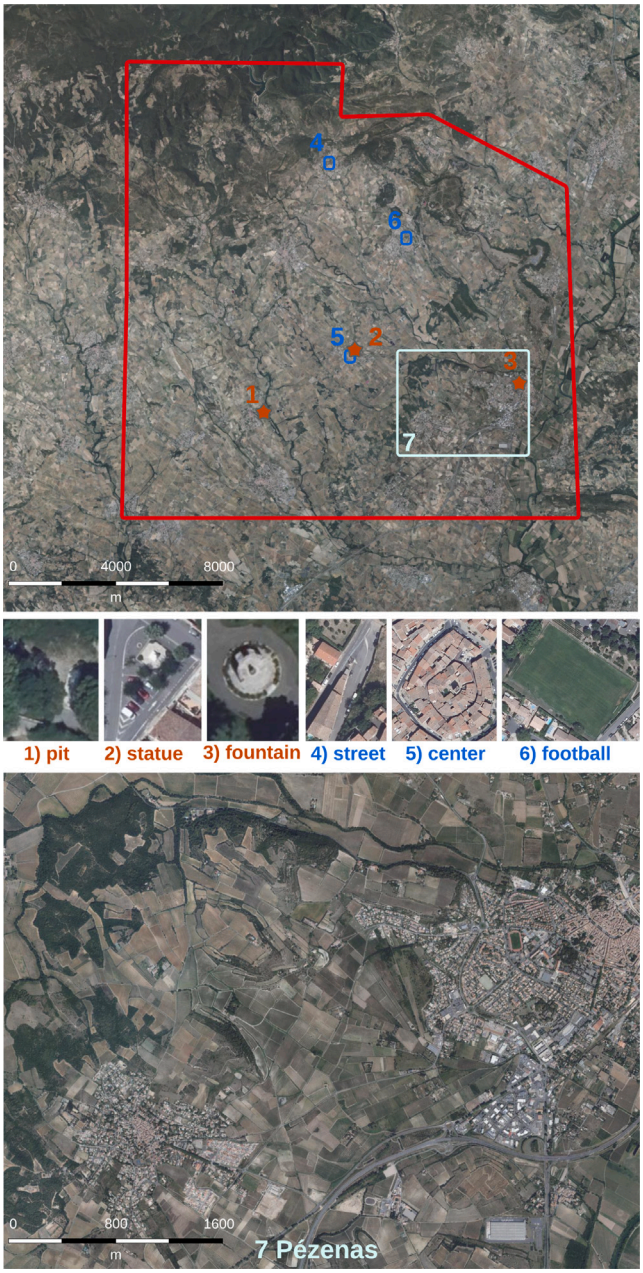

Fig. 7. Overview of the regions of interest used for the quality estimation considering tie point accuracy and reprojection error (1–3), accuracy estimation with multi-temporal DSMs (4–6), and qualitative use of the DSMs for change detection analysis close to the city Pézenas (7).

we calculated the cloud-to-cloud (C2C) distance for small DSM patches in CloudCompare 2.12.4. The most recent epoch was used as the reference (i.e., primary) for the comparison between all other epochs (i.e., secondary).

Qualitative change detection: The last step of accuracy assessment involved a test, whether the final dense point cloud products can be used for a fast detection of environmental changes in a region near the city of Pézenas. All selected areas of interest are shown in Fig. 7.

3. Results and discussion

We use the Congo dataset to introduce our developed approach as a proof-of-concept regarding the possibility of leveraging learned feature matching methods (trained on urban, terrestrial data) for historical aerial images. The Occitanie dataset is used to assess the capabilities of our workflow to match images taken at different points in time (inter-epoch) simultaneously with images taken during the same epoch (intra-epoch). This dataset is also used to provide insights about the accuracy of the aligned dense point cloud, which is relevant to assess the suitability of our workflow for multi-temporal change detection applications. The Occitanie dataset is the least challenging in our study. Hence, the standard configurations of the SuperGlue and DISK methods could be applied. Both datasets, Congo and Occitanie, provide different results and insights into the challenges when using our automatic approach of image matching to perform 3D reconstruction across time.

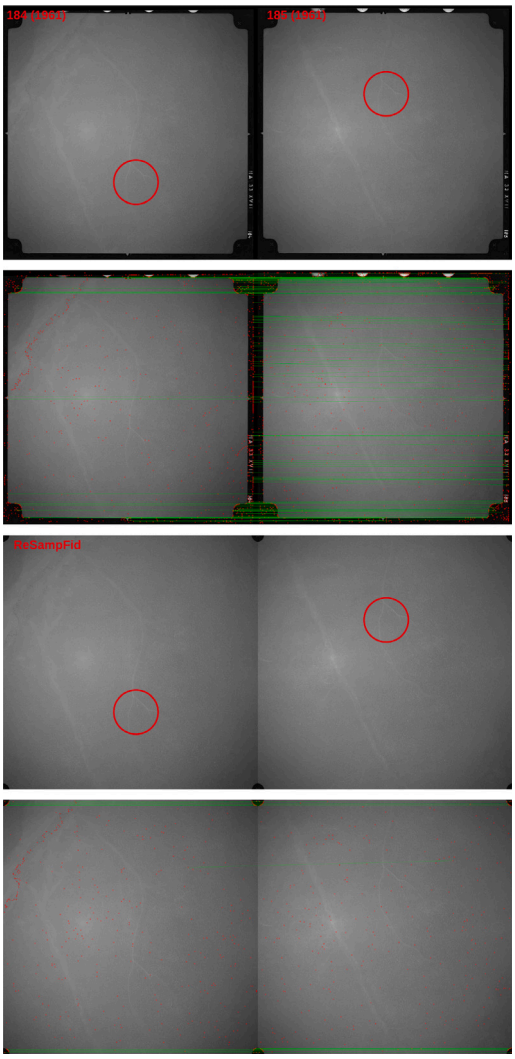

Fig. 8. SuperGlue feature matches for the Congo dataset. Top: Historical image pair with fiducial marks where similar regions are marked by a red circle and the matches are shown using a maximum edge length of 1600 pixels. Bottom: The aerial image with resampled fiducial marks is shown. All detected tie points are marked as red pixels and calculated feature matches are shown as green connected lines for both approaches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. Congo dataset: Intra-epoch matching of low-texture scenes

The Congo dataset was captured in a very challenging environment with quasi texture-less images, unknown camera parameters, and missing information about the study area (Fig. 8). The images show the canopy and a network of rivers in the tropical rainforest.

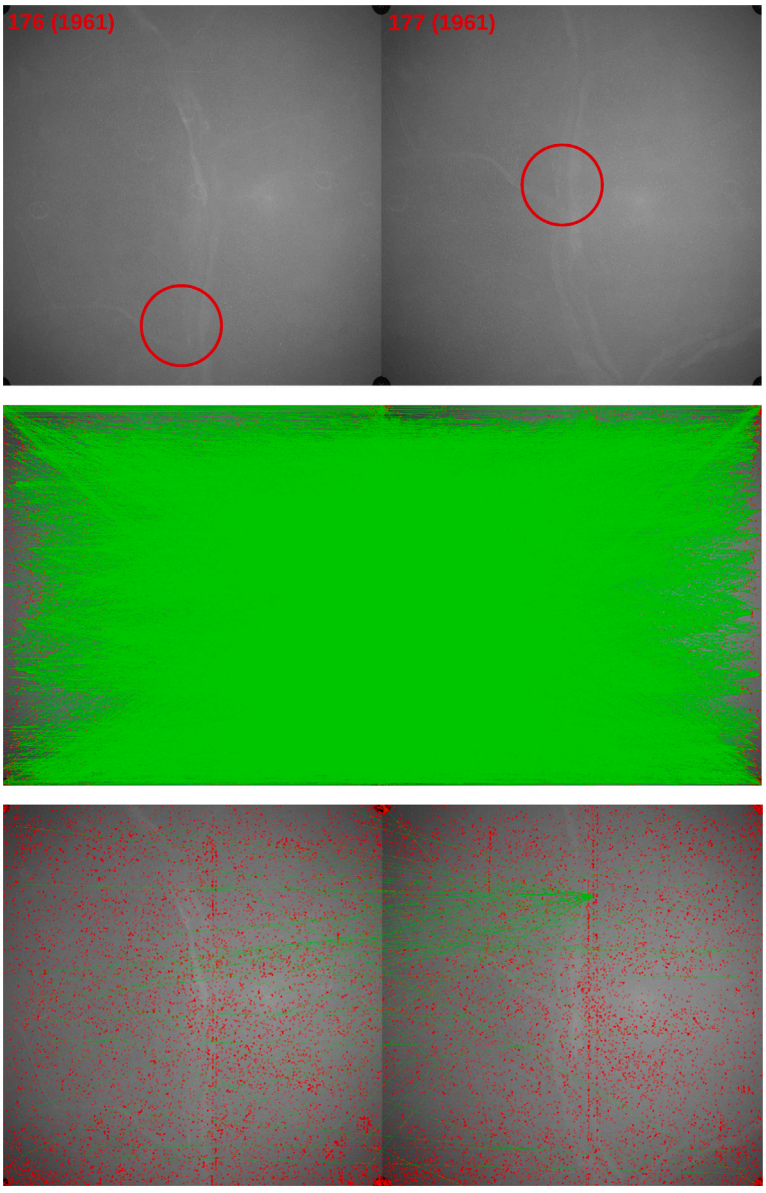

Fig. 9. Top: Historical image pair where similar regions are marked by a red circle for better readability. Middle: Initial tie points and matches found by SuperPoint and SuperGlue between all possible combinations of image tiles. Bottom: Geometrical verification of matches results in only 79 remaining incorrect matches.

3.1.1. Results for the SuperGlue approach

Results for matches between all image pairs using SuperGlue show that only the image frames are matched and that no correct tie points could be assigned. Resampling the images to a maximum edge length of 1600 pixels did not improve the results (Fig. 8-top) and removing the image borders with its fiducial marks using ReSampFid (Section 2.1) still did not solve the issue (Fig. 8-bottom).

To force SuperGlue to focus on other image regions, the original images are split into tiles with a maximum edge length of 1600 using the presented tile-based approach (Section 2.2.1). Although, this approach of using full image resolution provides a high number of image feature points extracted by SuperPoint, the subsequent feature matching with SuperGlue is not able to connect the image tiles correctly, still leading to insufficient results (Fig. 9).

Superglue performs well on feature-rich datasets (Zhang et al., 2021). However, the texture-less scene of the Congo dataset is too difficult to find reliable feature matches, also when testing multiple modifications.

3.1.2. Results for the DISK approach

The DISK approach is directly applied to the aerial images with resampled fiducial marks as often the border regions hampered feature matching resulting in a possibly lower accuracy also reported by Feurer and Vinatier (2018). In contrast to SuperGlue, the default setting is the usage of input images with a downsampled resolution of 1024 × 1024 pixels, a maximum allowed number of 2048 detected features per image and considering the non-maximum suppression (nms) mode. This also provides no correct feature matches, but when the maximum number of detected features is increased to 4096 the first correct matches occur.

To retrieve the optimal parameter settings for the historical aerial images of the Congo site, the parameters *image size* and *number of features n* are modified one after another. In regard of feature detection mode, mainly *nms* is used as recommended by Tyszkiewicz et al. (2020). Tests have also been done for multiple different parameter settings with the *rng* mode (only one is shown in the publication). However, the results were consistently worse, i.e. producing systematic outliers, when compared to the *nms* mode (Fig. 10).

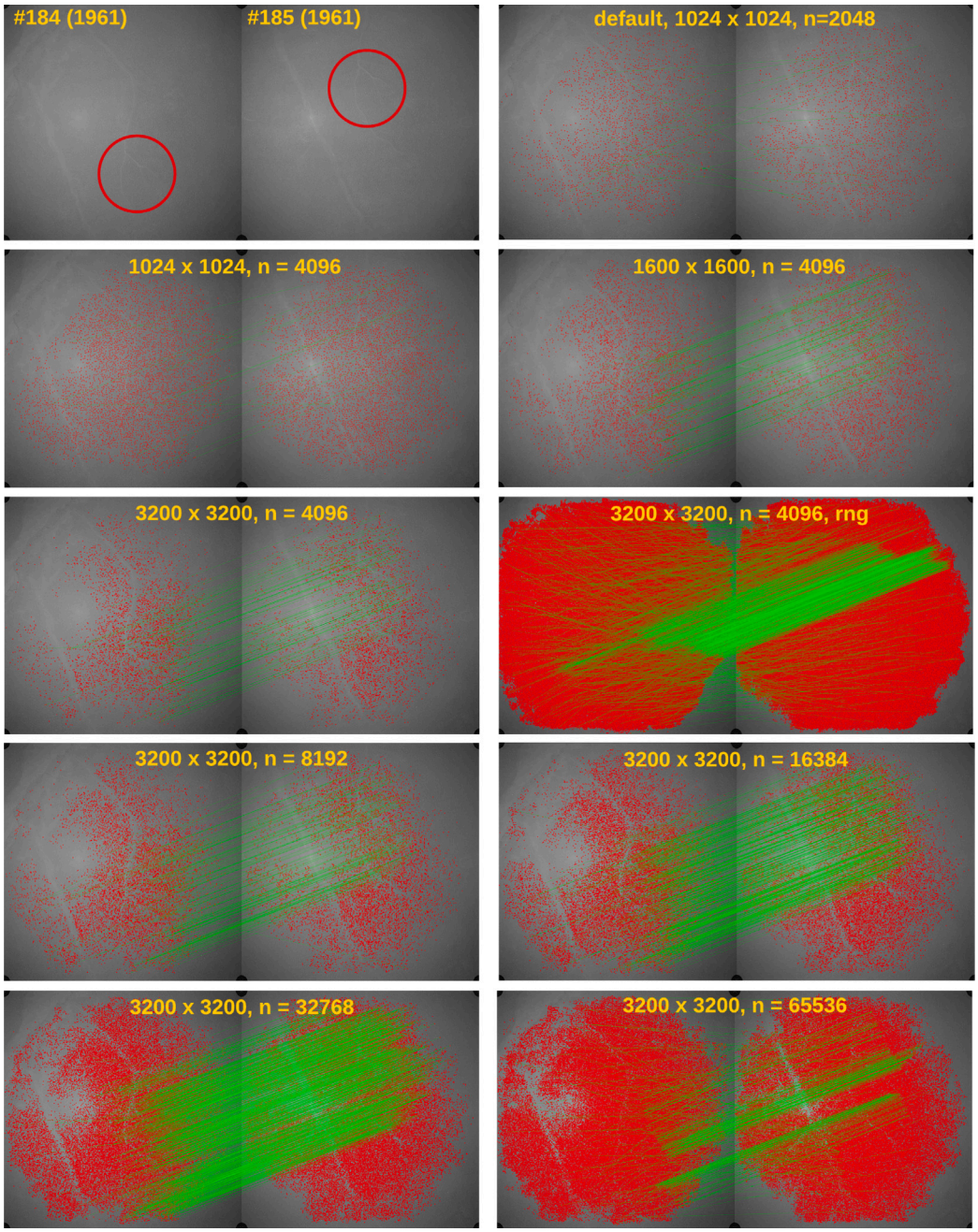

Fig. 10. DISK feature matches for the Congo dataset using varying sets of steering parameters. The best and most consistent results are generated using a configuration of 3200 × 3200 pixels image resolution and 32 768 extracted features. For all depicted tests, except one example, the *nms* mode is used.

Modifying the image resolution and number of extracted features leads to the following results. A larger image size initially increases the total number of correct feature matches. However, the increase of the image resolution from 1600 × 1600 pixels to 3200 × 3200 pixels (corresponds to the limit for the NVidia A100 GPUs) results in a approximately constant number of determined feature matches, when considering a constant maximum number of 4096 detected tie points.

Increasing the maximum number of detected features without limits results in an exponential growth of features until COLMAP can no longer read the h5 file and also the number of correct feature matches decreases (seen for $n = 65\,536$ and larger). According to the authors of DISK, this happens because the space of descriptors is saturated with too many detected features and the feature matching process obtains incorrect results because of too many false positives (Michał Tyszkiewicz, personal correspondence, 04 Feb 2023). Limiting the maximum number of extracted features to 32 768 provides more reliable results. The number of correct matches and the matching ratio $= \frac{\text{correct matches}}{\text{total matches}}$ in relation to the DISK configurations is shown in Fig. 11.

3.1.3. Results in COLMAP

Using this optimal configuration for DISK, the Congo dataset is processed in COLMAP to obtain the camera parameters and a sparse point cloud. It is recommended to only estimate one single camera model during intra-epoch reconstruction as we assume sufficient stability of the aerial cameras. Using multiple different camera models leads to strong variations of the estimated focal lengths and consequently of the flight heights (camera height). Additionally, it causes strong dome effects of the models due to overparameterization, especially considering the limited image overlap of the historical data (Eltner and Schneider, 2015).

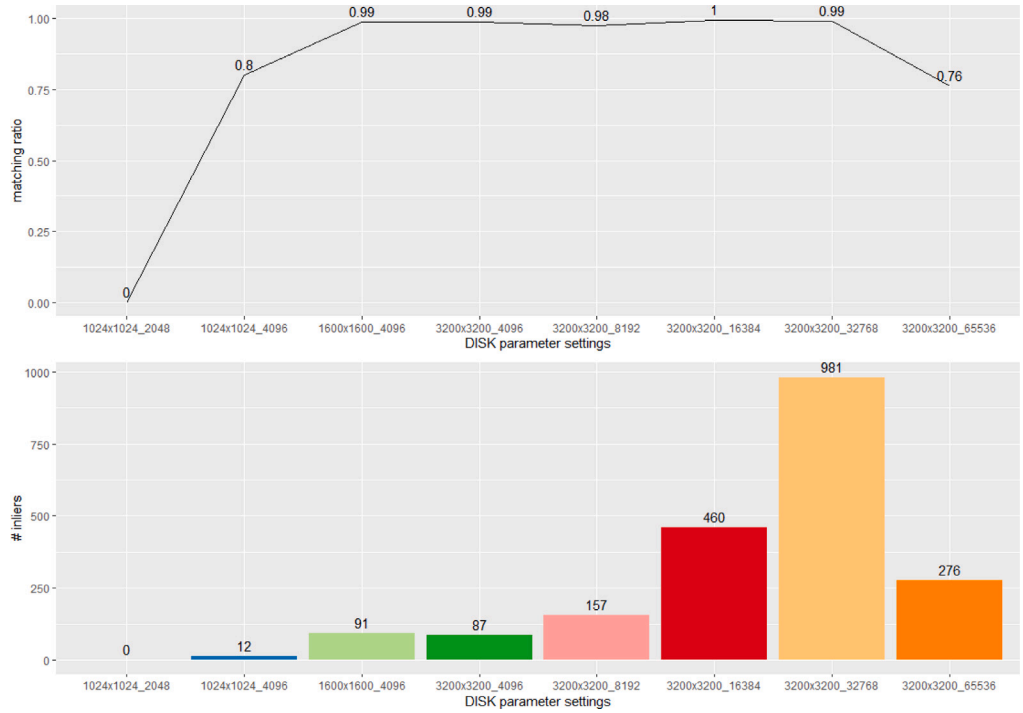

Fig. 11. DISK matching ratio and total number of inliers for the Congo dataset. The configuration 3200 × 3200 with 32 768 extracted tie points provides the highest number of 981 inliers with a matching ratio of 99.1%.

With our method all images could be used in a single SfM workflow, producing a DSM of this region. Furthermore, the flight route of two connected flight strips and one single flight strip not connected to the others was automatically reconstructed in two separate COLMAP models (Fig. 12).

3.2. Occitanie dataset: Intra-epoch and inter-epoch feature matching

Feurer and Vinatier (2018) and Zhang et al. (2021) already processed the Occitanie dataset and evaluated their results using the Time-SIFT method and the DSM-based approach respectively. However, it is of high interest if our adaptations and the use of learned feature matching method exclusively in image space are also capable of processing these mono-temporal and multi-temporal data in one single workflow. Thereby, we could show that the default method of SuperGlue is already capable of finding reliable inter-epoch and intra-epoch tie points (see comparison of all methods in Fig. 14).

Without further modifications, COLMAP generates several reconstructions for a different number of images. However, a single reconstruction using all images could not be calculated. As SuperGlue is not rotational invariant, some image blocks cannot be connected to others resulting in separated models. All images, except for two, are registered in four different models with a varying number of images (Fig. 13).

The same procedure is followed with the default DISK configuration. DISK is also able to find keypoints inter-epoch and intra-epoch wise. However, the default DISK approach found significantly less inter-epoch matches when compared to SuperGlue and especially our proposed method with adapted DISK parameter settings (Table 3 and Fig. 14).

The workflow in COLMAP leads to slightly worse results when using DISK based tie points. Although, DISK generates a similar amount of 3D points, these are not as evenly distributed throughout the images compared to the SuperGlue workflow (Figs. 13 and 15).

Both learned feature matching methods are able to match historical aerial images from the same epoch and between different epochs even in their default configurations. However, they revealed several limitations as already discussed previously (Section 2.2.1). Both methods are not fully rotational invariant and also require adjustments for processing the original image size.

Table 3
Comparison of the number of tie points for one inter-epoch (1981–1990) and one intra-epoch (1971–1971) image pair found by SuperGlue and DISK with default settings, and our method.

	Intra-epoch tie points	Inter-epoch tie points
SuperGlue	1248	369
DISK	798	195
Our method	7837	1616

Table 4
Comparison of the results of the default method and our proposed workflow on the Occitanie dataset consisting of 148 images. The largest model is compared regarding its total number of 3D points, the mean number of observations per image, and its mean reprojection error calculated by COLMAP.

	Largest model (# images)	3D points	Mean obs/img	repr. err. (px)
SuperGlue	81	98 500	3270.4	0.61
DISK	63	64 259	2203.6	0.25
Our method	148	1 006 539	17 173.1	1.11

With our proposed workflow using the combination of SuperGlue on downsampled images to determine the rotation and the subsequent DISK matching procedure applied to the full resolution images, we are able to estimate the camera parameters for all historical aerial images of the original Occitanie dataset in one single automatic run (Fig. 16).

A comparison of the final models reveals the increase in registered images, number of observations per images, estimated 3D points, and also a increase in the final reprojection error (Table 4). However, this is to be expected for larger datasets as a higher number of camera parameters needs to be optimized in the bundle adjustment procedure because we consider a separate single camera for every epoch. A mean reprojection error of 1.1 pixels can still be considered very accurate for a model with 148 historical images with generally lower image quality but does not yet give insights on the metric quality of the result.

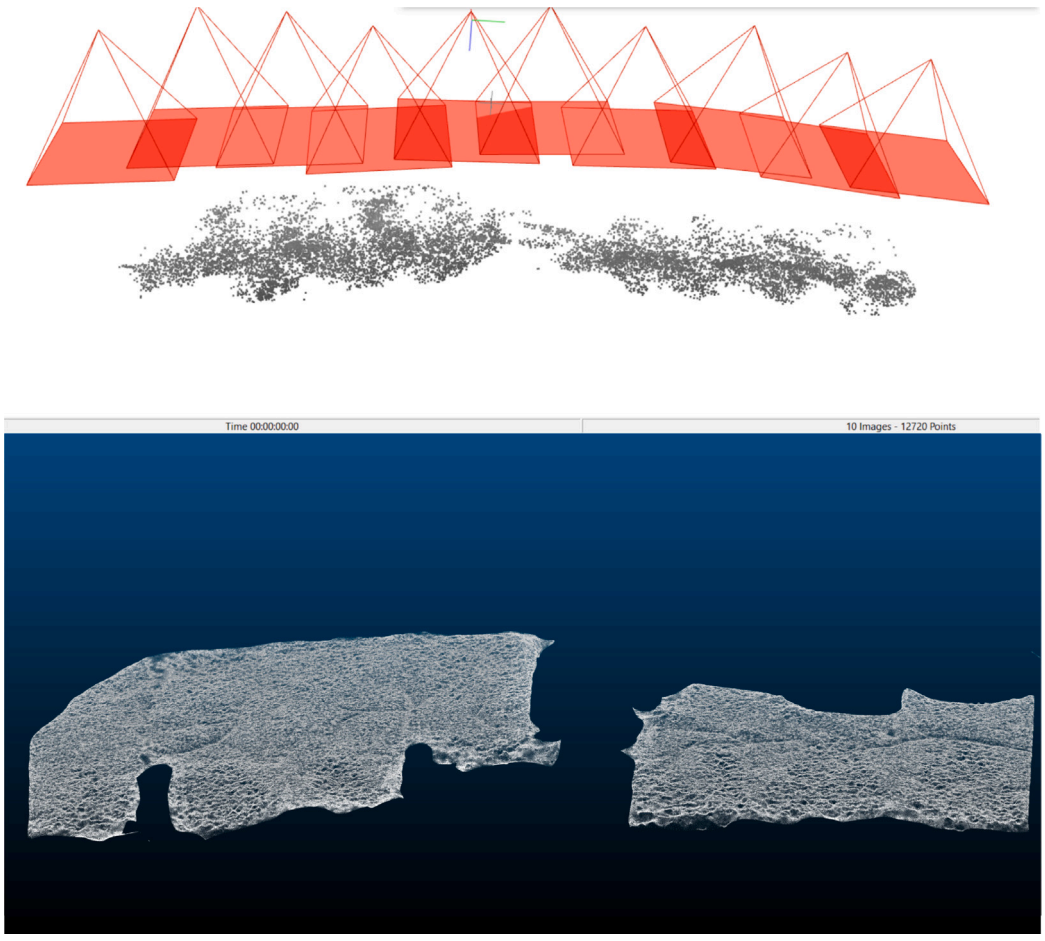

Fig. 12. Top: One flight strip of the Congo dataset including 10 images with a mean reprojection error of 0.84 pixels. The red frustum depict the camera orientation in relation to the estimated gray 3D points. Bottom: Dense point cloud created using OpenMVG (Moulon et al., 2016). OpenMVG falls short in providing a complete model due to the scarcity of reconstructed 3D points in the central region of the original sparse point cloud. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

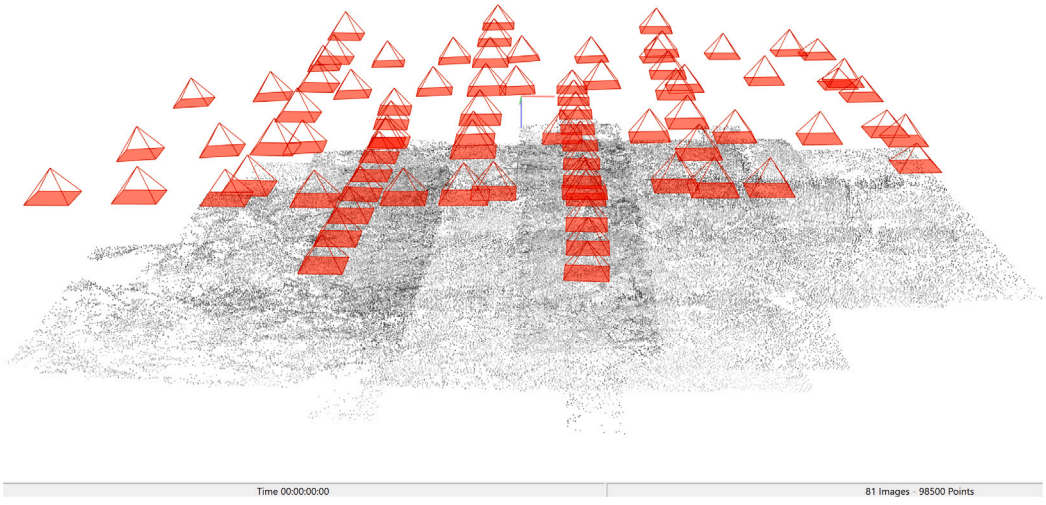

Fig. 13. Largest SfM model reconstructed using the default SuperGlue feature matching workflow. The model consists of 81 images from a total of 148. It includes aerial images of all four epochs and has a final mean reprojection error of 0.61 pixels.

3.2.1. Accuracies achieved for a scaled model

The previously presented results only show results in image space and in the local, arbitrary coordinate system defined by COLMAP. Here, we give insights into the tie-point accuracy, compare multi-temporal DSMs, and perform change detection using the Occitanie dataset.

To assess the final tie point accuracy, a few stable points were measured in the historical images (Fig. 17).

Note that conventionally used natural GCPs such as manholes or street corners were often not visible due to insufficient image quality or plant growth, or changes due to construction work during the 50 year

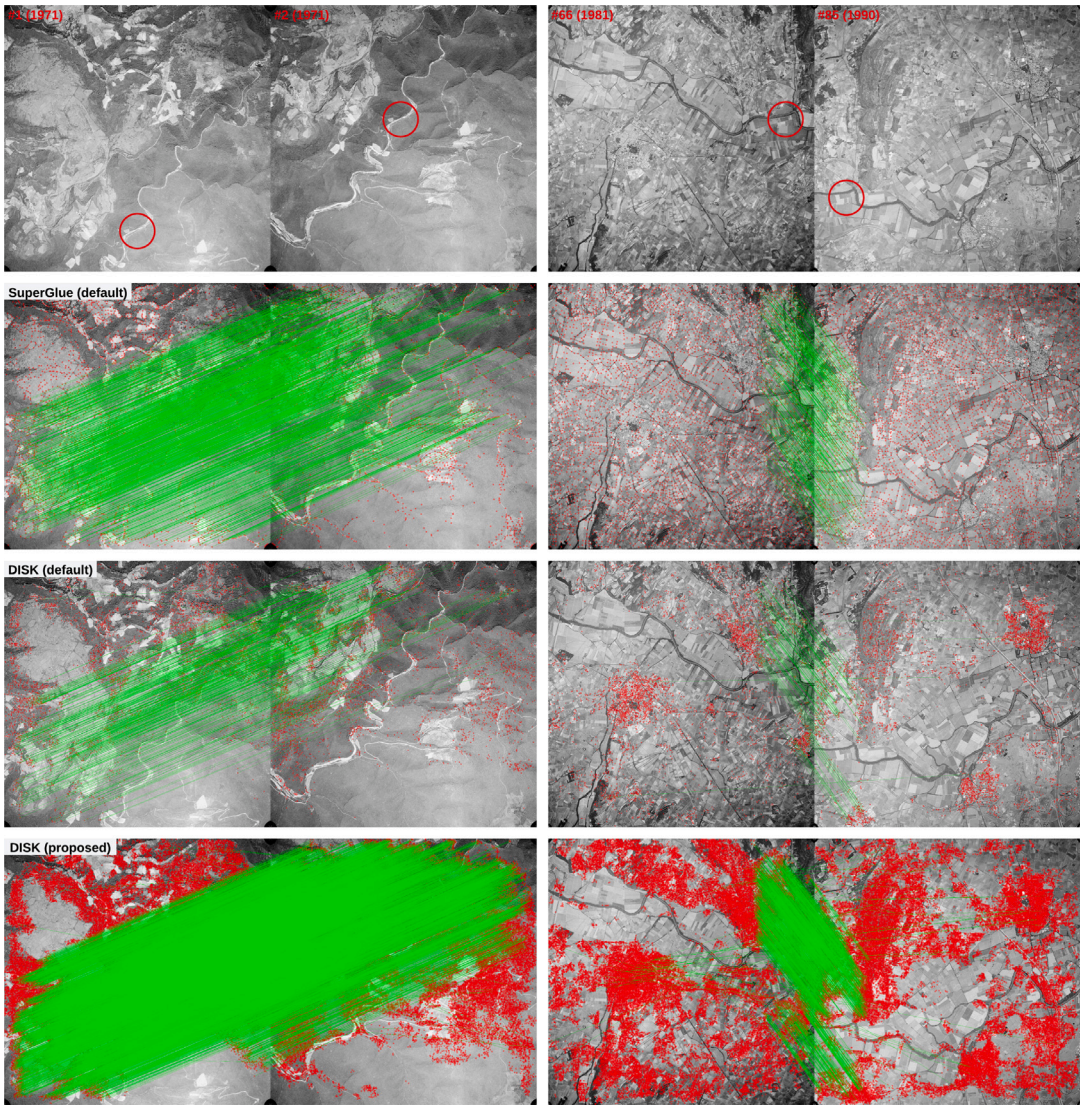

Fig. 14. Top-left: Intra-epoch image pair of 1971; Top-right: Inter-epoch image pair between epochs 1981 and 1990; Similar image regions are circled in red. Second row: Intra-epoch and inter-epoch SuperGlue feature matches are visualized by connected green lines. Third row: DISK feature matches. Fourth row: Feature matches resulting from our proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

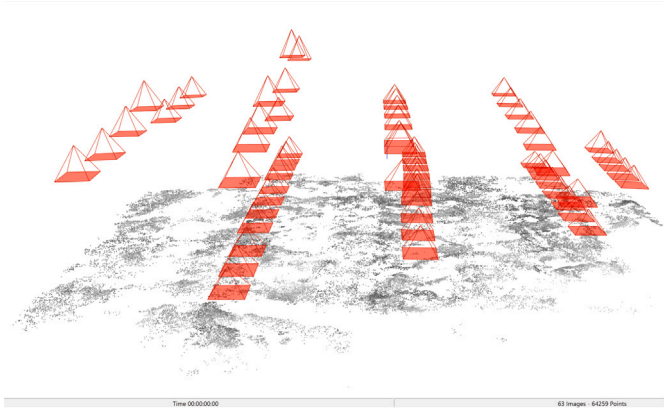

Fig. 15. Largest SfM model reconstructed using the default DISK feature matching workflow. The model consists of 63 images from a total of 148. It includes aerial images of the first three epochs with a final mean reprojection error of 0.25 pixels. A separate model is generated for the most recent epoch of 2001.

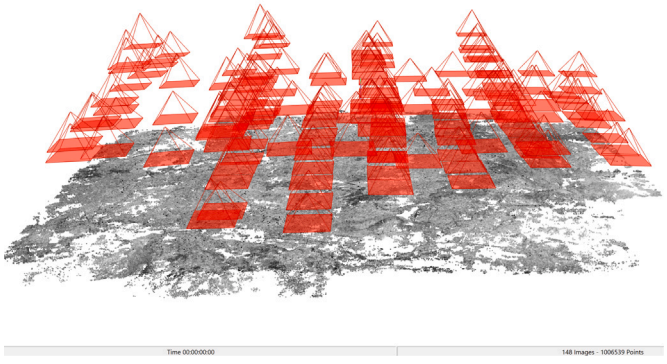

Fig. 16. All 148 images of the Occitanie dataset are included in one SfM model using the established workflow combining SuperGlue, DISK and COLMAP.

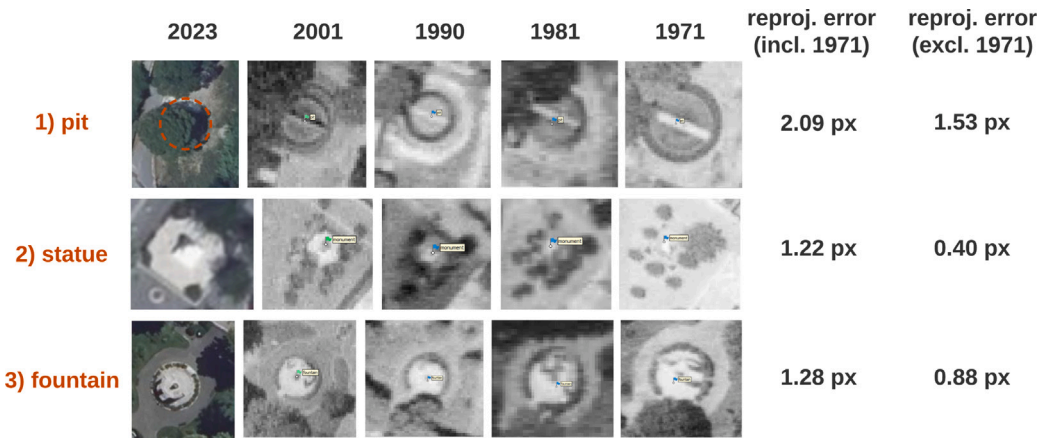

Fig. 17. Tie point accuracy for three selected stable structures. All tie points were selected in epoch 2001 and then projected into the images of the three other epochs of the aligned image block (see image patches). The mean reprojection error was afterwards calculated by manually correcting the marker positions including and excluding epoch 1971 (see columns).

observation period (i.e., from the most recent images used to extract the scale and to the earliest images dating back to the 1970's). Three different three-dimensional objects (*pit*, *statue*, *fountain*) were used for the tie point based accuracy assessment. While usually flat objects are advantageous to assess the reprojection error, such targets could not be identified in the dataset with sufficient quality.

The reprojection errors excluding epoch 1971 range between 0.4 and 1.6 pixels. They coincide with or are even superior to the overall mean reprojection error of the model estimated in COLMAP. When epoch 1971 is also included into the error analysis, the reprojection error increases to values between 1.2 and 2.1 pixels. The lower quality of the 1971 dataset could also be observed in the DSMs, because the overall image quality of the aerial photographs from 1971 is considerable worse compared to the later epochs. Furthermore, the object *pit* is located closer to the border of the image block of the 1971 epoch, which is another potential reason for the lower performance because a decrease of accuracy can be expected due to lower image overlap. Nevertheless, our feature matching strategy was capable of finding many similar points between several epochs with an accuracy of approximately 2 pixels and better. This also shows that SuperGlue and DISK were able to operate in subpixel accuracy for distinctive objects.

To assess the quality of a scaled model, small patches of interest were compared within the region of interest (Fig. 18). The DSM comparison was obvious between the three epochs 2001, 1990, and 1981. However, the DSM of epoch 1971 revealed many incomplete regions and therefore mostly hindered an accuracy assessment in that regard. Two patches containing a *football* field and a town *center* seemed to be the most stable regions in our analysis.

It was possible to identify stable areas throughout the observation period. The mean differences of the C2C distances were mostly close to zero meters (Table 5). Again, the accuracy was lower for the incomplete models of epoch 1971. When analyzing the individual patches, it could be revealed that *center* and *football* field exhibited a normally distributed error for the height values with a standard deviation of 0.81 m, which is about twice the GSD. It should be noted that the patch *football* field was overgrown by grass in 1981 and therefore the C2C difference might not actually represent aligned DSM accuracy but vegetation heights. The *street* patch might have changed throughout the observation period – again highlighting the challenge to find stable areas – because average deviations are mostly negative.

To evaluate the quality of the DSMs for relative change detection between different epochs, a larger area of interest was observed (Fig. 19). Major changes in the range of several meters become easily visible and should be considered for further analysis. Tree growth, building construction and possibly erosion can be seen.

Table 5
Comparison of the mean difference and standard deviation of C2C distances considering only the Z-component.

Dataset	Epoch	Mean [m]	Median [m]	std. dev. [m]
Street	1971–2001	0.10	0.10	0.49
	1981–2001	−0.44	−0.22	1.33
	1990–2001	−0.55	−0.54	1.06
Center	1971–2001	−0.91	−0.89	0.97
	1981–2001	0.02	0.00	0.81
	1990–2001	0.10	0.02	0.82
Football	1971–2001	1.09	1.11	0.87
	1981–2001	0.51	0.53	0.62
	1990–2001	−0.02	−0.06	0.78

Considering the diverse multi-temporal data, the assessment of differences between scaled models highlighted that a sufficient accuracy could be achieved with our approach because the high number of tie points increased the quality of the geometry of the multi-temporal image block.

3.3. Implications and limitations

Working on the two different datasets using learned feature matching methods leads to several findings. Both, DISK and SuperGlue are able to find feature matches between historical aerial images of different epochs and within the same epoch as highlighted with the Occitanie dataset. The Congo dataset reveals that SIFT (not shown in the publication) and SuperGlue are not able to find feature matches in texture-less scenes of the tropical rainforest. Also, the default configuration of DISK results only in few correct feature correspondences. However, an extensive parameter testing enabled the finding of the optimal settings for historical aerial images.

Some limitations of our workflow have to be emphasized. Feature matching becomes more computationally expensive the more images are compared to each other. The number of exhaustive matching of image pairs is calculated as $(n^2+n)/2$ and hence increases significantly for large datasets, which is also the case for our presented tile-based approach. We intercept this increase slightly by including assumptions about the image overlap within and across flight strips and by using downsampled images to generate matches. It is recommended to use one or multiple GPUs for feature matching. As an example, the calculation of SuperGlue feature matches for one single image pair takes about 10 s on a common Intel i7-1165Gz @ 2.80 GHz while it reduces significantly to 0.1 s when using a single NVidia A100 GPU.

Due to the synergetic characteristics of SuperGlue and DISK, our workflow requires setting up both matching strategies, which can be

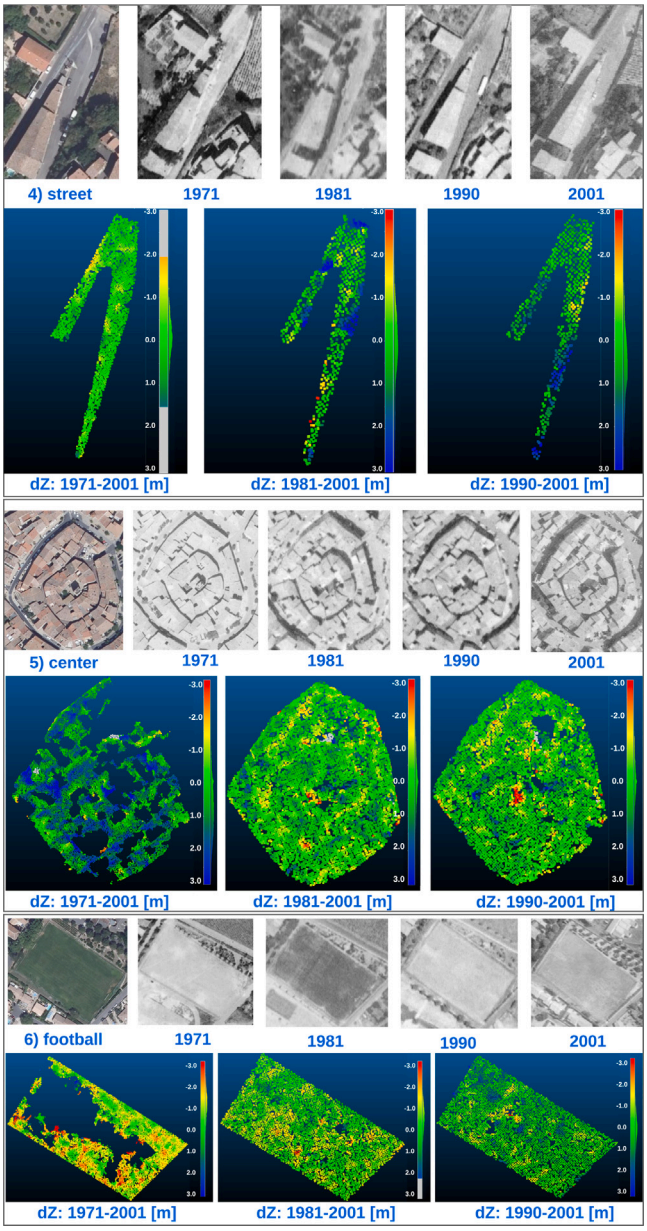

Fig. 18. Cloud-to-cloud distances between DSMs; epochs (1971, 1981, 1990) were compared to epoch 2001. The metric distances are shown color-coded in Z direction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

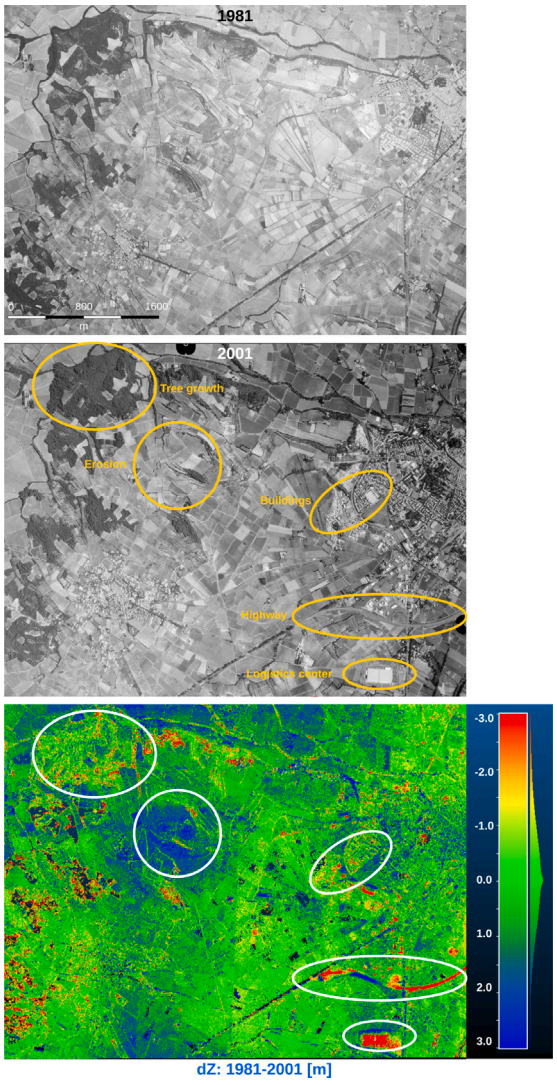

Fig. 19. Comparison of Z-difference of a larger area between 1981 and 2001. Larger regional changes are visualized with yellow ellipses in the aerial photograph of 2001. This includes natural phenomena such as tree growth but also man-made changes like the construction of a highway or logistics center. The changes are clearly visible in the distance image where positive changes larger than 3 m are visualized in red, and negative changes larger than 3 m in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

challenging. Therefore, we provide the complete pipeline as separate open-source modules performing the different steps one after another according to the user's needs.

Further, the applied learned feature matching methods were originally trained on terrestrial images. Although they provide good results on the aerial datasets of our case studies, the approaches might fail on other data and the general transferability from terrestrial to aerial perspective needs to be investigated more broadly. A future strategy could be the training of DISK with historical aerial images. However, this requires a (historical) ground truth dataset providing the images, depths and calibration protocols. One option for such a training procedure might be the newly established TIME benchmark dataset (Farella et al., 2022). In order to reach a comparable size of training data as large as the MegaDepth dataset with 135 scenes used for DISK, a similar amount of reconstructed historical aerial scenes would be necessary. With the 28 different scenes from the TIME benchmark and more historical data openly available, our workflow could possibly be used for generating this large amount of COLMAP reconstructions.

SuperGlue does not allow for such a training procedure because the architecture and corresponding weights of the neural network model are not openly available. However, first available open variants of SuperGlue might be alternatives (Viniavskyi et al., 2022).

Generation of DSMs becomes possible when using the proposed method with post-processing in Agisoft Metashape. It could be observed, that the comparison of DSMs between 2001, 1990 and 1981 of the Occitanie dataset lead to reasonable results with standard deviations in Z-direction of twice the GSD. However, for epoch 1971 with actually a lower GSD the final dense point cloud was more erroneous with visible holes and larger radiometric differences between overlapping areas. Consequently, the result for tie point accuracy and DSM comparison were worse than for the other epochs. This could not

be directly observed during feature matching but only after detailed analysis of the final sparse and dense point clouds.

We agree with Feurer and Vinatier (2018) and Knuth et al. (2023) who state that processing multiple epochs simultaneously is advantageous, as it may help to cope with the correlation between flying height and focal length. In the multi-temporal dataset less systematic errors occur, like e.g., the dome effect for the Congo flight strip. The multi-temporal feature matching enabled a reasonable convergence of the bundle adjustment with small final mean reprojection errors already in COLMAP without prior definition of camera parameters opposed to the Congo dataset.

4. Conclusions

Historical aerial images offer the potential to enable long-term observations of environments as early flight campaigns date back to the 19th century. However, this data often suffers from degradation, inadequate digitization quality, and missing flight information.

The study presents a comprehensive workflow for processing these historical aerial images and unlocking the potential of existing data lying in numerous archives. It especially focuses on mono-temporal and multi-temporal data with difficult radiometric properties and unknown camera parameters. We significantly improve the feature matching in the SfM workflow enabling the automatic detection of tie points between image pairs that could not yet be matched using existing strategies. Our workflow does initially not need any other information except that the processed data is an aerial image dataset. This will allow an extended observation period of different study areas and a monitoring of time-dependent environmental changes.

The results show a significant increase in tie points found in historical aerial image pairs and a low mean reprojection error of 1.1 pixels for the final 3D model determined by COLMAP. The accuracy investigation shows that this reprojection error can also be reproduced for several stable regions. After metrically scaling the point cloud, the comparison of multi-temporal DSMs reveal that our method is capable of producing reliable DSMs with a standard deviation of 0.8 meters in stable areas. Additionally, environmental changes in larger areas can be detected.

As the presented work only deals with two datasets (and the TIME benchmark dataset integrated in the open-source code), we intend to test on further historical aerial images in the future. With DISK especially working well for the presented image datasets and providing an openly available training procedure, it is planned to integrate historical aerial images in the training of the neural network. This requires calculating approximately 100 COLMAP reconstructions from (historical) aerial image dataset, mirroring the training process applied to the original terrestrial datasets, which can then be used in the custom training procedure.

CRediT authorship contribution statement

Ferdinand Maiwald: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization. **Denis Feurer:** Conceptualization, Resources, Data curation, Writing – review & editing, Supervision. **Anette Eltner:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors express their gratitude to Fabrice Vinatier for providing information and images of the Occitanie dataset and to Maxime Réjou-Méchain for providing the Congo dataset. Special thanks go to Elisa Mariarosaria Farella for supporting with the TIME benchmark. The provided sample images are part of the TIME dataset realized with the support of EuroSDR and different European Mapping Agencies. In particular, the reported images are part of the Norway dataset, kindly provided by Mikko Sippo. The authors gratefully acknowledge the GWK's support for this project by providing computing time through the Center for Information Services and HPC (ZIH) at TU Dresden on HRSK-II.

References

Albertz, J., 2009. Einführung in die Fernerkundung, fourth ed. WBG (Wissenschaftliche Buchgesellschaft), Darmstadt.

Andreassen, L.M., y, H.E., llmoen, B.K., Belart, J.M.C., 2020. Glacier change in Norway since the 1960s – an overview of mass balance, area, length and surface elevation changes. J. Glaciol. 66 (256), 313–328. http://dx.doi.org/10.1017/jog.2020.10.

Berveglieri, A., Imai, N.N., Tommaselli, A.M., Casagrande, B., Honkavaara, E., 2018. Successional stages and their evolution in tropical forests using multi-temporal photogrammetric surface models and superpixels. ISPRS Journal of Photogrammetry and Remote Sensing 146, 548–558. http://dx.doi.org/10.1016/j.isprsjprs.2018.11.002.

Blanch, X., Eltner, A., Guinau, M., Abellan, A., 2021. Multi-epoch and multi-imagery (MEMI) photogrammetric workflow for enhanced change detection using time-lapse cameras. Remote Sens. 13 (8), 1460. http://dx.doi.org/10.3390/rs13081460.

Bolles, K.C., Forman, S.L., 2018. Evaluating landscape degradation along climatic gradients during the 1930s dust bowl drought from panchromatic historical aerial photographs, United States great plains. Front. Earth Sci. 6, http://dx.doi.org/10.3389/feart.2018.00153.

Carrivick, J.L., Smith, M.W., 2018. Fluvial and aquatic applications of structure from motion photogrammetry and unmanned aerial vehicle/drone technology. WIREs Water 6 (1), http://dx.doi.org/10.1002/wat2.1328.

Chum, O., Matas, J., 2005. Matching with PROSAC — progressive sample consensus. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, pp. 220–226. http://dx.doi.org/10.1109/cvpr.2005.221.

Cook, K.L., Dietze, M., 2019. Short communication: A simple workflow for robust low-cost UAV-derived change detection without ground control points. Earth Surf. Dyn. 7 (4), 1009–1017. http://dx.doi.org/10.5194/esurf-7-1009-2019, URL https://esurf.copernicus.org/articles/7/1009/2019/.

Craciun, D., Bris, A.L., 2022. Automatic algorithm for georeferencing historical-to-nowadays aerial images acquired in natural environments. Int. Arch. Photogr. Remote Sens. Spat. Inf. Sci. XLIII-B2-2022, 21–28. http://dx.doi.org/10.5194/isprs-archives-xliii-b2-2022-21-2022.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-supervised interest point detection and description. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, pp. 337–33712. http://dx.doi.org/10.1109/cvprw.2018.00060.

DeWitt, J.D., Ashland, F.X., 2023. Investigating geomorphic change using a structure from motion elevation model created from historical aerial imagery: A case study in northern lake michigan, USA. ISPRS Int. J. Geo-Inf. 12 (4), 173. http://dx.doi.org/10.3390/ijgi12040173.

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-net: A trainable CNN for joint description and detection of local features. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 8084–8093. http://dx.doi.org/10.1109/cvpr.2019.00828.

Eltner, A., Schneider, D., 2015. Analysis of different methods for 3D reconstruction of natural surfaces from parallel-axes UAV images. Photogramm. Rec. 30 (151), 279–299. http://dx.doi.org/10.1111/phor.12115.

Farella, E.M., Morelli, L., Remondino, F., Mills, J.P., Haala, N., Crompvoets, J., 2022. The eurosdr time benchmark for historical aerial images. Int. Arch. Photogr. Remote Sens. Spat. Inf. Sci. XLIII-B2-2022, 1175–1182. http://dx.doi.org/10.5194/isprs-archives-XLIII-B2-2022-1175-2022.

Feurer, D., Vinatier, F., 2018. Joining multi-epoch archival aerial images in a single SfM block allows 3-D change detection with almost exclusively image information. ISPRS J. Photogr. Remote Sens. 146, 495–506. http://dx.doi.org/10.1016/j.isprsjprs.2018.10.016.

Giordano, S., Le Bris, A., Mallet, C., 2018. Toward automatic georeferencing of archival aerial photogrammetric surveys. ISPRS Ann. Photogr. Remote Sens. Spat. Inf. Sci. IV-2, 105–112. http://dx.doi.org/10.5194/isprs-annals-IV-2-105-2018.

Knuth, F., Shean, D., Bhushan, S., Schwat, E., Alexandrov, O., McNeil, C., Dehecq, A., Florentine, C., O'Neel, S., 2023. Historical structure from motion (HSfM): Automated processing of historical aerial photographs for long-term topographic change analysis. Remote Sens. Environ. 285, 113379. http://dx.doi.org/10.1016/j.rse.2022.113379.

Li, Z., Snavely, N., 2018. MegaDepth: Learning single-view depth prediction from internet photos. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2041–2050. http://dx.doi.org/10.1109/cvpr.2018.00218.

Maiwald, F., 2019. Generation of a benchmark dataset using historical photographs for an automated evaluation of different feature matching methods. Int. Arch. Photogr. Remote Sens. Spat. Inf. Sci. XLII-2/W13, 87–94. http://dx.doi.org/10.5194/isprs-archives-XLII-2-W13-87-2019.

Maiwald, F., 2022. A window to the past through modern urban environments — Developing a photogrammetric workflow for the orientation parameter estimation of historical images (Ph.D. thesis). Technische Universität Dresden, http://dx.doi.org/10.13140/RG.2.2.19627.52004.

Maiwald, F., Bruschke, J., Schneider, D., Wacker, M., Niebling, F., 2023. Giving historical photographs a new perspective: Introducing camera orientation parameters as new metadata in a large-scale 4D application. Remote Sens. 15 (7), http://dx.doi.org/10.3390/rs15071879.

Maiwald, F., Lehmann, C., Lazariv, T., 2021. Fully automated pose estimation of historical images in the context of 4D geographic information systems utilizing machine learning methods. ISPRS Int. J. Geo-Inf. 10 (11), 748. http://dx.doi.org/10.3390/ijgi10110748.

Mölg, N., Bolch, T., Walter, A., Vieli, A., 2019. Unravelling the evolution of zmuttgletscher and its debris cover since the end of the little ice age. Cryosphere 13 (7), 1889–1909. http://dx.doi.org/10.5194/tc-13-1889-2019.

Morelli, L., Bellavia, F., Menna, F., Remondino, F., 2022. Photogrammetry now and then – from hand-crafted to deep-learning tie points –. Int. Arch. Photogr. Remote Sens. Spat. Inf. Sci. XLVIII-2/W1-2022, 163–170. http://dx.doi.org/10.5194/isprs-archives-xlviii-2-w1-2022-163-2022.

Moulon, P., Monasse, P., Perrot, R., Marlet, R., 2016. OpenMVG: Open multiple view geometry. In: International Workshop on Reproducible Research in Pattern Recognition. Springer, pp. 60–74.

Nagarajan, S., Schenk, T., 2016. Feature-based registration of historical aerial images by area minimization. ISPRS J. Photogramm. Remote Sens. 116, 15–23. http://dx.doi.org/10.1016/j.isprsjprs.2016.02.012.

Picon-Cabrera, I., Garcia-Gago, J.M., Sanchez-Aparicio, L.J., Rodriguez-Gonzalvez, P., Gonzalez-Aguilera, D., 2020. On the use of historical flights for the urban growth analysis of cities through time: The case study of avila (Spain). Sustainability 12 (11), 4673. http://dx.doi.org/10.3390/su12114673.

Pinto, A.T., Gonçalves, J.A., Beja, P., Honrado, J.P., 2019. From archived historical aerial imagery to informative orthophotos: A framework for retrieving the past in long-term socioecological research. Remote Sens. 11 (11), 1388. http://dx.doi.org/10.3390/rs11111388.

Rupnik, E., Daakir, M., Deseilligny, M.P., 2017. MicMac – a free, open-source solution for photogrammetry. Open Geospat. Data Softw. Stand. 2 (1), http://dx.doi.org/10.1186/s40965-017-0027-2.

Rupnik, E., Jansa, J., Pfeifer, N., 2015. Sinusoidal wave estimation using photogrammetry and short video sequences. Sensors 15 (12), 30784–30809. http://dx.doi.org/10.3390/s151229828.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. SuperGlue: Learning feature matching with graph neural networks. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 4938–4947, doi:arXiv:1911.11763, arXiv:1911.11763v2.

Schönberger, J.L., Frahm, J.-M., 2016. Structure-from-motion revisited. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 4104–4113. http://dx.doi.org/10.1109/cvpr.2016.445.

Sevara, C., 2016. Capturing the past for the future: an evaluation of the effect of geometric scan deformities on the performance of aerial archival media in image-based modelling environments. Archaeol. Prospect. 23 (4), 325–334. http://dx.doi.org/10.1002/arp.1539.

Soldato, M.D., Riquelme, A., Bianchini, S., Tomàs, R., Martire, D.D., Vita, P.D., Moretti, S., Calcaterra, D., 2018. Multisource data integration to investigate one century of evolution for the agnone landslide (molise, southern Italy). Landslides 15 (11), 2113–2128. http://dx.doi.org/10.1007/s10346-018-1015-z.

Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. LoFTR: Detector-free local feature matching with transformers. CVPR.

Tyszkiewicz, M.J., Fua, P., Trulls, E., 2020. DISK: Learning local features with policy gradient. 1–15 arXiv:2006.13566.

Vastaranta, M., Niemi, M., Wulder, M.A., White, J.C., Nurminen, K., Litkey, P., Honkavaara, E., Holopainen, M., Hyyppä, J., 2015. Forest stand age classification using time series of photogrammetrically derived digital surface models. Scand. J. Forest Res. 31 (2), 194–205. http://dx.doi.org/10.1080/02827581.2015.1060256.

Velasco, R.F., Lippe, M., Tamayo, F., Mfuni, T., Sales-Come, R., Mangabat, C., Schneider, T., Günter, S., 2022. Towards accurate mapping of forest in tropical landscapes: A comparison of datasets on how forest transition matters. Remote Sens. Environ. 274, 112997. http://dx.doi.org/10.1016/j.rse.2022.112997.

Vinatier, F., Arnaiz, A.G., 2018. Using high-resolution multitemporal imagery to highlight severe land management changes in mediterranean vineyards. Appl. Geogr. 90, 115–122. http://dx.doi.org/10.1016/j.apgeog.2017.12.003.

Viniavskyi, O., Dobko, M., Mishkin, D., Dobosevych, O., 2022. OpenGlue: Open source graph neural net based pipeline for image matching. http://dx.doi.org/10.48550/ARXIV.2204.08870.

Wang, X., Liu, L., Hu, Y., Wu, T., Zhao, L., Liu, Q., Zhang, R., Zhang, B., Liu, G., 2021. Progressive advance and runout hazard assessment of a low-angle valley glacier in east kunlun mountains from multi-sensor satellite imagery analysis. Nat. Hazards Earth Syst. Sci. http://dx.doi.org/10.5194/nhess-2021-57.

Warrick, J.A., Ritchie, A.C., Adelman, G., Adelman, K., Limber, P.W., 2017. New techniques to measure cliff change from historical oblique aerial photographs and structure-from-motion photogrammetry. J. Coast. Res. 33 (1), 39. http://dx.doi.org/10.2112/jcoastres-d-16-00095.1.

Zhang, L., Rupnik, E., Pierrot-Deseilligny, M., 2021. Feature matching for multi-epoch historical aerial images. ISPRS J. Photogr. Remote Sens. 182, 176–189. http://dx.doi.org/10.1016/j.isprsjprs.2021.10.008.