remote sensing

MDPI

Article

Giving Historical Photographs a New Perspective: Introducing Camera Orientation Parameters as New Metadata in a Large-Scale 4D Application

Ferdinand Maiwald [1,2,*,†], Jonas Bruschke [3], Danilo Schneider [4], Markus Wacker [5] and Florian Niebling [3]

[1] Institute of Photogrammetry and Remote Sensing, Technische Universität Dresden, 01062 Dresden, Germany
[2] Chair for Digital Humanities, The Friedrich Schiller University Jena, 07743 Jena, Germany
[3] Human-Computer Interaction, University of Würzburg, 97070 Würzburg, Germany
[4] Chair for Photogrammetry, HTW Dresden, 01069 Dresden, Germany
[5] Chair for Computer Graphics, HTW Dresden, 01069 Dresden, Germany
[*] Correspondence: ferdinand.maiwald@tu-dresden.de
[†] Current address: Helmholtzstraße 10, 01069 Dresden, Germany.

Abstract: The ongoing digitization of historical photographs in archives allows investigating the quality, quantity, and distribution of these images. However, the exact interior and exterior camera orientations of these photographs are usually lost during the digitization process. The proposed method uses content-based image retrieval (CBIR) to filter exterior images of single buildings in combination with metadata information. The retrieved photographs are automatically processed in an adapted structure-from-motion (SfM) pipeline to determine the camera parameters. In an interactive georeferencing process, the calculated camera positions are transferred into a global coordinate system. As all image and camera data are efficiently stored in the proposed 4D database, they can be conveniently accessed afterward to georeference newly digitized images by using photogrammetric triangulation and spatial resection. The results show that the CBIR and the subsequent SfM are robust methods for various kinds of buildings and different quantity of data. The absolute accuracy of the camera positions after georeferencing lies in the range of a few meters likely introduced by the inaccurate LOD2 models used for transformation. The proposed photogrammetric method, the database structure, and the 4D visualization interface enable adding historical urban photographs and 3D models from other locations.

Keywords: historical images; 4D-GIS; content-based image retrieval; Structure-from-Motion; camera orientation; feature matching

check for updates

Citation: Maiwald, F.; Bruschke, J.; Schneider, D.; Wacker, M.; Niebling, F. Giving Historical Photographs a New Perspective: Introducing Camera Orientation Parameters as New Metadata in a Large-Scale 4D Application. *Remote Sens.* **2023**, *15*, 1879. https://doi.org/10.3390/rs15071879

Academic Editors: Maurice Murphy, Garrett Keenaghan, Eimear Meegan, Simona Scandurra, Anthony Corns and Alain Chenaux

Received: 13 February 2023
Revised: 27 March 2023
Accepted: 28 March 2023
Published: 31 March 2023

1. Introduction

With an ever-increasing amount of historical urban photographs being digitized, numerous opportunities exist for using and presenting that data. In combination with three-dimensional (3D) models, maps, and elevation models, it becomes possible to link all data in time-dependent four-dimensional (4D) geographic information systems (GIS). Originally conceptualized as a tool to support scholars in their art and architectural history research in the project UrbanHistory4D [1], the so-called 4D browser (https://4dbrowser.urbanhistory4d.org, accessed on 1 February 2023) serves as a front-end and back-end application for representing different types of 4D data (Figure 1). This includes time-dependent maps, 3D building models, and photographs, which can be generated on a global scale.

While historical models of several cities and landmarks already exist or are created in various research projects, it is still difficult to integrate external image data or realize large-scale applications. That is mainly because working with historical photographs comes with several challenges, particularly when automatically finding photographs showing similar buildings or when used in a structure-from-motion (SfM) process, automatically generating historical 3D models. This is mostly caused by vast radiometric and geometric differences in image quality.

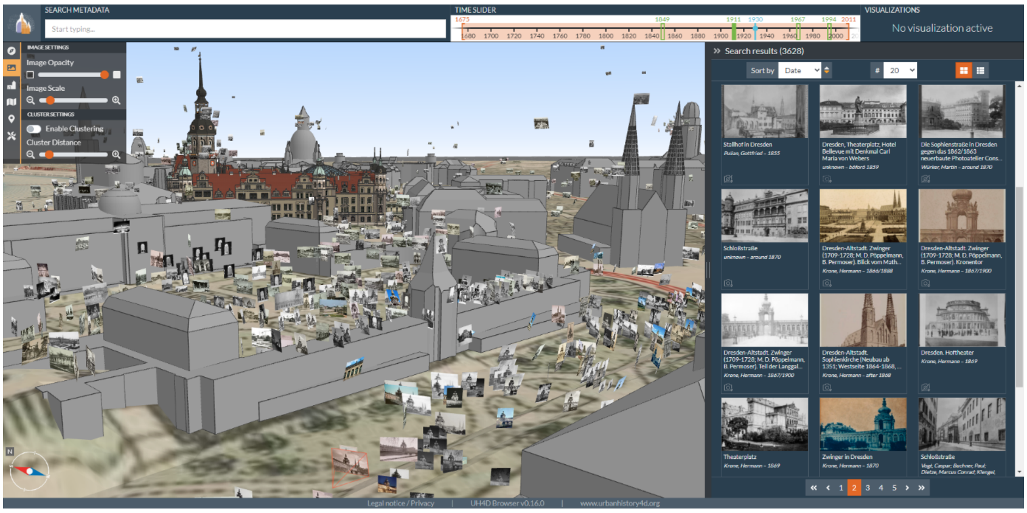

SEARCH METADATA
Start typing...
TIME SLIDER
1675
1849
1911
1930
1967
1994
2011
1680 1700 1720 1740 1760 1780 1800 1820 1840 1860 1880 1900 1920 1940 1960 1980 2000
VISUALIZATIONS
No visualization active
IMAGE SETTINGS
Image Opacity
Image Scale
CLUSTER SETTINGS
Enable Clustering
Cluster Distance
Search results (3628)
Sort by
Date
#
20
Stallhof in Dresden
Pulian, Gottfried – 1855
Dresden, Theaterplatz, Hotel Bellevue mit Denkmal Carl Maria von Webers
unknown – before 1859
Die Sophienstraße in Dresden gegen das 1862/1863 neuerbaute Photoatelier Cons...
Würker, Martin – around 1870
Schloßstraße
unknown – around 1870
Dresden-Altstadt. Zwinger (1709-1728; M. D. Pöppelmann, B. Permoser). Blick vom Math...
Krone, Hermann – 1866/1888
Dresden-Altstadt. Zwinger (1709-1728; M. D. Pöppelmann, B. Permoser). Kronentor
Krone, Hermann – 1867/1900
Dresden-Altstadt. Zwinger (1709-1728; M. D. Pöppelmann, B. Permoser). Teil der Langgal...
Krone, Hermann – 1867/1900
Dresden-Altstadt. Sophienkirche (Neubau ab 1351; Westseite 1864-1868, ...
Krone, Hermann – after 1868
Dresden. Hoftheater
Krone, Hermann – 1869
Theaterplatz
Krone, Hermann – 1869
Zwinger in Dresden
Krone, Hermann – 1870
Schloßstraße
Vogt, Caspar; Buchner, Paul; Dietze, Marcus Conrad; Klengel,
« ‹ 1 2 3 4 5 › »
N
Legal notice / Privacy | UH4D Browser v0.16.0 | www.urbanhistory4d.org

Figure 1. Graphical user interface of the 4D browser showing the 3D city model of Dresden, Germany, and spatially oriented photographs.

Another issue is that data are often processed and used for visualization purposes, but are mostly not accessible after a project is completed. We tackle that problem by efficiently storing camera parameters as metadata for every historical image following the close-range photogrammetry Guide to Good Practice of the Archaeology Data Service (https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/data-collection-and-fieldwork/close-range-photogrammetry/introduction/close-range-digital-photogrammetry-in-archaeology/, accessed on 15 March 2023). Therefore, the presented approach shows a bipartite pipeline that includes the initialization of a new dataset and the subsequent addition of new images as shown in Figure 2.

Initially, historical photographs need to be selected in order to be usable for SfM. For photogrammetric processing, mainly exterior views are relevant. However, close-up photographs that show only a small detail of the original building are excluded. Therefore, all images of one landmark are investigated using content-based image retrieval (CBIR). As this requires calculating a feature vector for every image when compared to a query, this feature vector is stored in the database, making it accessible if newly digitized images are added. In the second step, the top-ranked photographs (with the smallest distances) in relation to the query are used in a SfM process. Due to the possible variation in historic image quality, the conventionally used software Agisoft Metashape fails to register a large number of these types of images. In contrast, *learned feature-matching methods* are often able to find tie points between historical images. Hence, it is proposed to use SuperPoint [2] + SuperGlue [3] for feature-matching and COLMAP [4] for generating a sparse point cloud and estimating all camera orientation parameters. As SfM generates only a local model, the initial georeferencing has to be done by hand or by automatically finding similar points in the GIS and the sparse cloud. For all images that can be registered, the interior and exterior camera parameters are stored in the database.

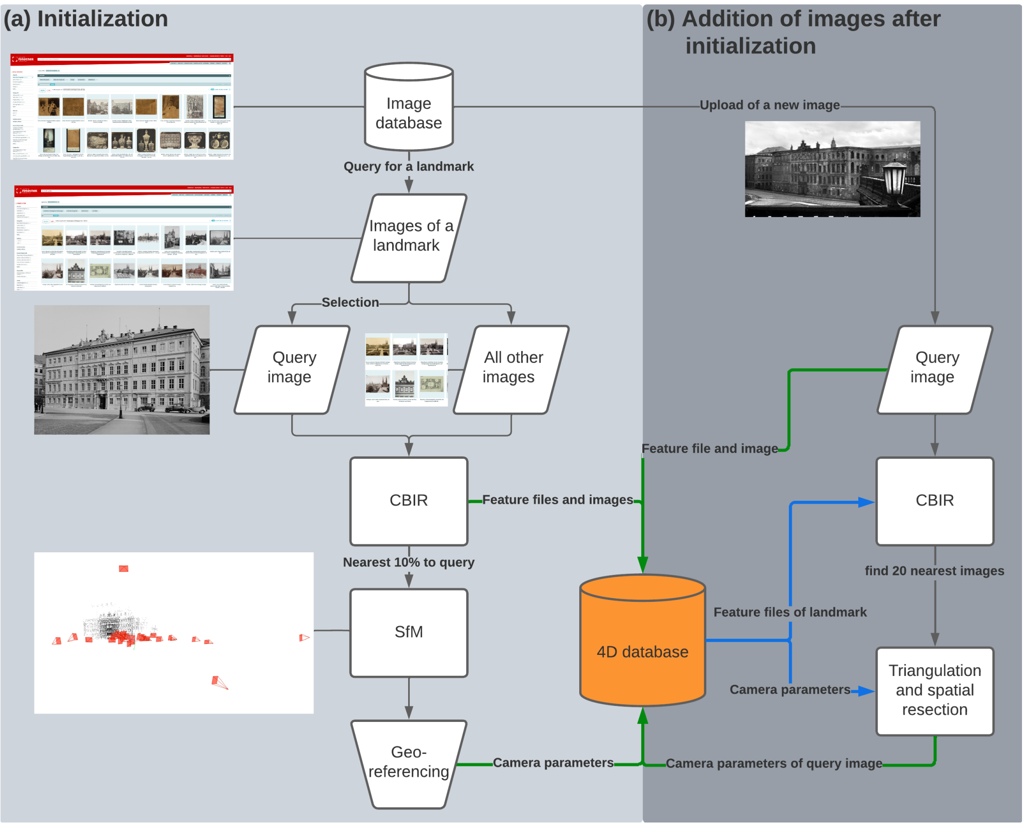


Figure 2. Schematic depiction of the presented bipartite pipeline for estimation of camera orientation parameters of historical photographs. The left hand shows the initialization of a new dataset, including the automatic selection of appropriate images using CBIR, subsequent SfM processing, and interactive georeferencing. The right hand shows the usage of the initialized data for estimating the camera orientation parameters for newly digitized images. The 4D database serves as a central part of storing and distributing all relevant information.

This process enables the convenient addition of newly digitized copies to the database. If a new image for a specific landmark (as indicated by metadata) is uploaded to the database, the image can be quickly compared via the proposed CBIR to all other photographs of that landmark since all feature vectors have already been calculated. The camera orientation parameters of the 20 closest images necessary for a successful SfM reconstruction are retrieved and initialized in COLMAP and feature matches need to only be calculated for 20 image pairs. As the global orientation of the 20 corresponding images is known, the new image is directly georeferenced using the triangulation of tie points and subsequent spatial resection. All estimated parameters of the newly uploaded image can be stored in the database.

2. Related Research

This section is split into two parts. The first part presents research that deals with the (geo)processing of historical photographs. The second part shows how other GIS research platforms deal with various data sources, such as 3D models and images.

2.1. Geoprocessing of Historical Photographs

Historical photographs are used in various types of applications. One major use-case is repeat photography (re-photography) which is the determination of the correct geometric alignment of a historical photograph taken at a certain place in comparison to a contemporary image of the same place. An exhaustive overview of the origins, available via online portals, and the present challenges are given in [5]. Georeferenced historical (aerial) images can be, e.g., used for the change detection of glaciers [6,7] or buildings [8], landslide investigation [9], or even the detection of unexploded ordnance [10].

In urban environments, terrestrial historical images are able to reveal the building changes [11], provide scene understanding [12], and can be used for measuring building heights [13] or virtual reconstruction of building parts [14,15]. In conclusion, historical photographs provide an extremely valuable tool for transferring knowledge of the past to shape the development of the future.

2.2. 3D/4D Research Platforms

Through the open availability and the simple accessibility of historical data, more platforms are emerging, allowing for the representation and visualization of various data sources. This includes, among others, maps, building models, images, and texts in combined virtual environments [16]. Most tools rely on web-based interfaces that combine different sources [17,18], enabling advanced searching [19] and user management [20]. Others develop comprehensive frameworks for creating Web3D Cultural Heritage applications [21]. A comprehensive overview can be found in [16,22]. Most related research presented in the literature focuses on specific objects or buildings and is geographically limited. The proposed workflow aims to overcome these limitations by efficiently estimating and storing camera orientation parameters, thereby retaining all derived information even after project completion.

3. Materials and Methods

This section provides details on the data used for the experiments, which consisted of two landmarks in the city of Dresden. The structure of the 4D browser, the connection to the database, and I/O operations are also explained. The main part of the section focuses on inputting a new image dataset into the database and initializing all database values. It explains how the images are filtered for their exterior views, how SfM is performed for the filtered images, and how all values can be transferred. Additionally, it shows how an existing initialized dataset can be extended easily by requesting all the existing data, enabling the quick processing of new photographs. Finally, the section outlines the limitations of the proposed workflow.

3.1. Data

The research area is focused on the city of Dresden, Germany, as it is possible to connect to the existing structure and data of the 4D browser developed in the UrbanHistory4D [23] project. It includes 3D city models of the urban center ranging from level of detail 2 (LOD 2) to LOD 3, with some of the models being texturized. However, most buildings do not provide a historically accurate depiction of their state, showing instead the contemporary state provided by the authorities. In the conducted experiments, the focus is on buildings that are not texturized, to provide a better depiction of the *visual* accuracy of the oriented historical photographs.

The research focuses on the Semperoper, which is a very famous building in Dresden and has 2172 hits when searching by metadata in the Deutsche Fotothek (https://www.deutschefotothek.de/, accessed on 30 November 2022). A query image for filtering exterior views is required for CBIR, so the focus is on the main façade (view from the south-east, Figure 3). To demonstrate that the approach is also applicable to smaller landmarks or even side streets, Taschenbergpalais is selected as a second building. It contains 584 hits in the Deutsche Fotothek (https://www.deutschefotothek.de/, accessed on 30 November 2022), with many views from far away. The query image shows the only unobstructed view from the north with only a few relevant hits in the Deutsche Fotothek (Figure 3).

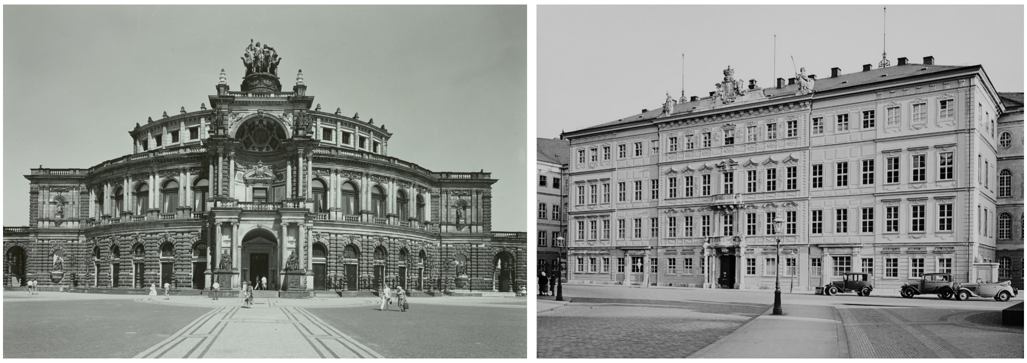


Figure 3. Query images used for the Semperoper (**left**) and Taschenbergpalais (**right**).

3.2. Structure of the 4D Browser and Its Database

The 4D browser was initially conceptualized and developed in the project UrbanHistory4D as a tool to support scholars in their art and architectural history research [1]. Images and photographs in particular are essential sources and key objects in this field of research. Unlike conventional media repositories where images can only be searched by metadata, the 4D browsers follow a spatial approach. The graphical user interface integrates a 3D viewport where the user is able to browse for spatialized images within a 3D city model without the necessity of having detailed knowledge about the object of interest and any metadata of corresponding images (Figure 1). Of course, a metadata search is still possible, but the user does not need to rely on it exclusively. With the time slider, temporal constraints can be set separately for the buildings, the images, and the maps. The filtering of images can also be achieved by interacting with the buildings or the 3D objects. The spatial approach also opens up new possibilities for gaining new knowledge, including taking the perspective of the photograph and understanding the photographer's situation [24], as well as using special visualizations to answer specific research questions [25].

On the server side, the 4D browser consists of a database and a backend application. The data are stored in the graph database *Neo4j* and largely follow the CIDOC conceptual reference model (CRM) [26], an ontology in the field of cultural heritage. All positional data are stored in WGS-84 world coordinates, i.e., latitude and longitude in decimal degrees (double precision), and altitude in meters. In the frontend, these world coordinates are converted into Cartesian coordinates (i.e., UTM) with a shift of the point of origin towards the scene contents for a true-to-scale visualization. The orientation is stored as three Euler angles: pitch angle ω, heading angle ϕ, and roll angle κ. Additionally, for spatially oriented images, the interior camera parameters, including the principal distance (ck) and principal point coordinates, are stored together with additional parameters such as distortion coefficients.

The backend application provides a RESTful API, serves binary files, and runs routines such as linking buildings and images by determining if a building is visible in the image. All resources can be queried and updated via HTTP requests. Some API routes are specifically used by the pipeline described in this paper. By this pipeline route, images can be queried by metadata or positions similar to the regular route, but the data are returned in a different format. Whereas the 4D browser frontend receives exterior and interior camera parameters as world coordinates and Euler angles, the pipeline routes use Cartesian coordinates and quaternions. For this purpose, latitude and longitude are converted into UTM coordinates in an OpenGL coordinate system. The origin of this coordinate system is then shifted to the position of the first image in the set of the queried images, such that the numbers are in a feasible numerical range. Together with the altitude, they form t_x, t_y, and t_z. The Euler angles are converted to quaternions q_w, q_x, q_y, and q_z to remove the constraint of the correct order of the angles. The interior camera parameters are set with respect to the image dimensions. To not lose the global reference, the set of queried images is returned with the world coordinates of the point of origin. When the pipeline finally pushes new or updated values for the set of images, these origin reference coordinates are used to convert the Cartesian values back to global WGS-84 coordinates.

3.3. Initialization of a New Dataset

This section describes the main workflow of this contribution for filtering historical images and the estimation of interior and exterior camera parameters. All relevant parameter decisions are explained in detail.

3.3.1. Layer Extraction Approach (LEA)

The initialization of a new dataset in the database requires a selection of suitable photographs. These are especially exterior views of the respective landmark as (a) they can be processed in a single photogrammetric workflow using SfM and (b) can be depicted without occlusions in the 4D environment. Thus, an automatic filtering method is required to robustly select relevant images. This excludes detailed views, other building states, public events, and paintings retrieved by conventional metadata search.

Thus, in prior research, we developed a content-based image retrieval workflow that successfully filtered historical images after a metadata search with one keyword. As an important part of the whole process, the pipeline, called the layer extraction approach (LEA), is briefly explained (Figure 4). For a more detailed investigation and justification of the parameterization, refer to [27].

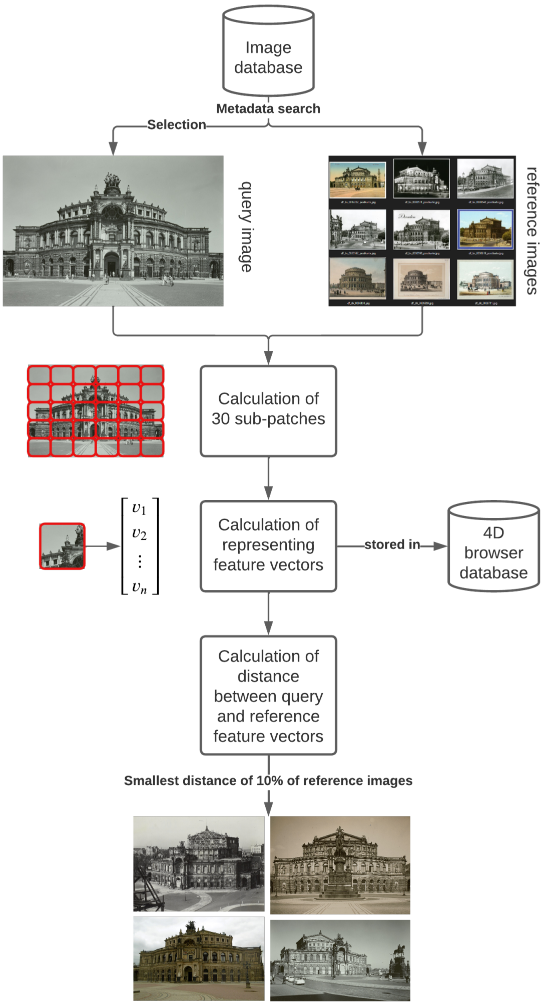


Image database
Metadata search
Selection
query image
reference images
Calculation of 30 sub-patches
v_1
v_2
⋮
v_n
Calculation of representing feature vectors
stored in
4D browser database
Calculation of distance between query and reference feature vectors
Smallest distance of 10% of reference images

Figure 4. General overview of the workflow in the layer extraction approach (LEA).

As a starting point, a specific keyword is selected to narrow the search in the image database and limit the computation time for the CBIR. In the presented research, the keywords are *Semperoper* and *Taschenbergpalais*, as shown in Section 4.

For these landmarks, all found images (reference images) are downloaded in the resolution provided by the repository with a maximum of 1600 pixels for the longer edge of the digitized image. Then, a query image has to be selected manually to show the scene of interest. For all images, including the query image, a representative feature vector is calculated. The approach is originally based on [28] and uses the pre-trained VGG16 convolutional neural network (CNN) [29] with modified upper layers for this image representation. The initial output in the last max-pooling layer is 512 feature maps with dimensions of 7×7, which are then reduced in dimension using spatial pooling with a kernel size of 7×7 and a stride of 0, resulting in the desired vector representation for every image of size $512 \times 1 \times 1 = 512$.

In order to improve the results of the retrieval, cropping the original images into subpatches with a separate feature vector leads to superior results [27]. We used the workflow described in [30] and cropped using the order parameter $L = 4$. This resulted in $\sum_{i=1}^{L} i^2 = 30$ subpatches.

In contrast to prior work, this feature vector is not deleted but serves as a unique representation of the respective image in the database. The vector holding the information on all subpatches is efficiently stored as a pickle file (.pkl) with 16 kilobytes (kB) and can be posted to (and requested from) the 4D browser database via HTTP. This allows using the vector representation of existing images when updating the database with new images (Section 3.4).

Generating a rank list for ordering the reference images from the closest to the furthest with respect to the query image requires the calculation of a single distance value, which is usually the L_2 normalized distance between two feature vectors. Since the images are divided into 30 subpatches, 30 different image representations exist, which are aggregated into a final distance number, as in [30]. The resulting rank list can now be used to show a predefined number of nearest images. It is of high relevance to use a meaningful number as this often defines whether the following SfM workflow fails or succeeds. In our experiments and with the given database *Deutsche Fotothek*, around 10% of images are significant and show (parts of) the expected query image. This implies using the 217 closest images for the Semperoper dataset and 58 images of the Taschenbergpalais dataset for the subsequent SfM procedure.

3.3.2. Estimation of Initial Interior Camera Orientation Parameters

The retrieved images from the previous step mostly depict the requested object from the query image. However, historical images may still show significant radiometric and geometric differences [31]. These range from day–night views to construction sites on relevant building parts.

Due to these reasons, it has already been determined that conventional SfM software often fails [31]. In particular, the estimation of the initial principal distance (focal length) and finding distinctive feature matches are seen as the main obstacles.

The proposed method uses vanishing point detection (VPD) for initializing the principal distance. The approach is based on [32] and refined in [12], where images are filtered out if they do not provide three significant vanishing points (VPs). Some additional modifications were made in the following. For initialization, a pinhole camera model is assumed so that parallel lines in object space are projected to a set of lines in image space. The advantage of the method on the presented data is that parallel lines often exist in man-made (urban) structures. In the following, all possible pairs of lines are intersected. With the assumption that the principal point $c(x, y)$ is in the image center, a polar coordinate system is spanned using that point as the origin (Figure 5-top). The three angles θ of the VPs can be easily determined using the polar coordinates in a histogram using a finite number of bins (Figure 5-bottom). Therefore, the infinite accumulator space from 0 to 2π of the angle θ is discretized into 720 bins (i.e., two bins for one degree).

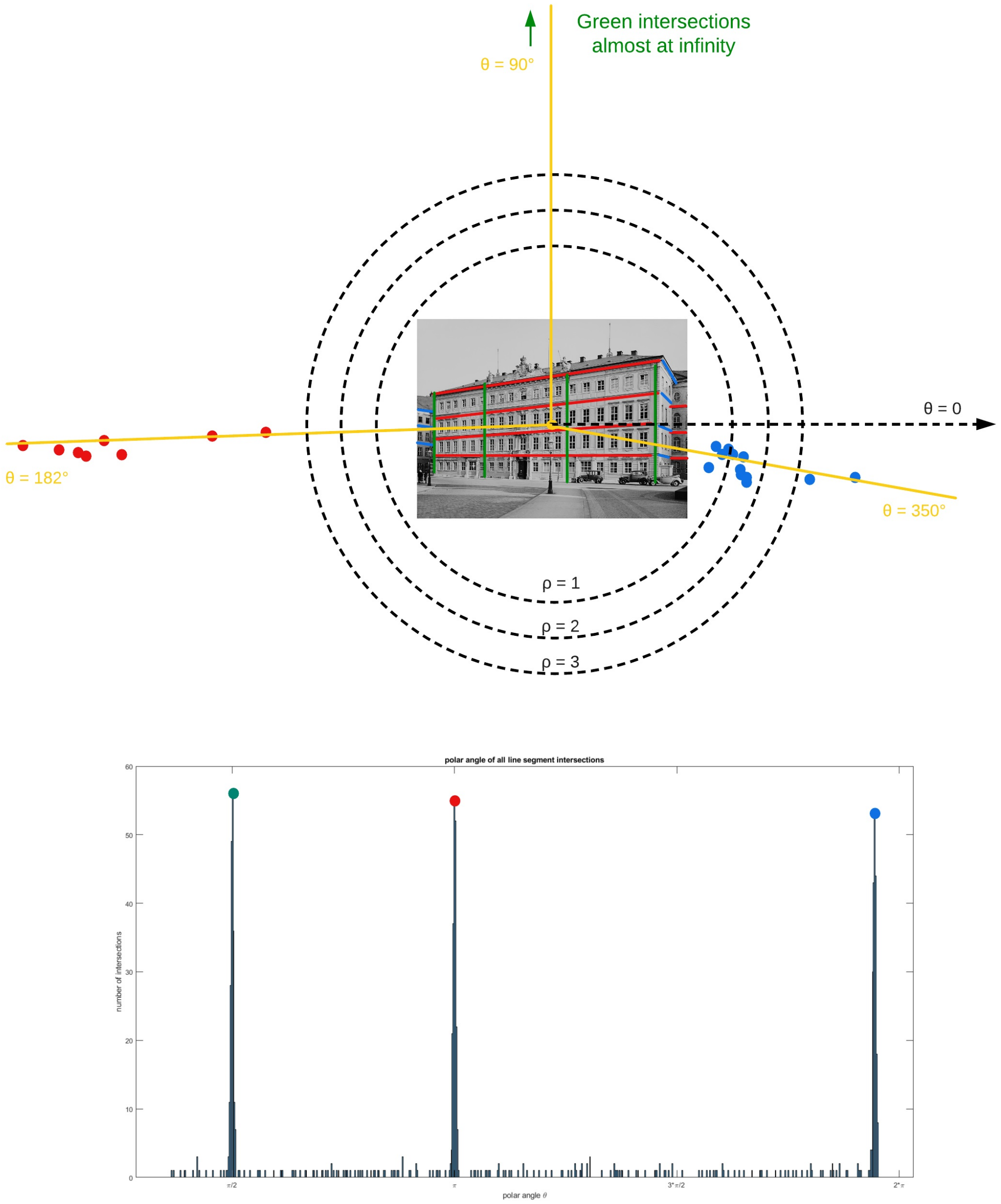


Figure 5. Schematic example of the principal distance estimation using VPs. Top: Example for the detected line segments with three main directions marked in blue, green, and red. Some selected line segment intersections are shown as colored dots. Bottom: Determination of the angle θ of three VPs using the distribution of all line segment intersections. The final angle θ is drawn as a line at the top of the figure.

The radius ρ for all VPs is determined in an equation system as a function of the principal distance as described by [32]. As the proposed approach only works if there are three significant VPs, we enhance the method and always determine four maxima during histogram analysis. A maximum is only considered significant if it is separated by 120 bins and if its prominence is >50% of the maximum prominence of all peaks. These values could be determined empirically as three significant VPs should be separated by at least 60 degrees in architectural scenes. If there are three significant maxima, the process continues and the principal distance is initially estimated. If there is no detection of the three significant maxima (i.e. three main directions of lines cannot be determined in the image), the standard initialization of COLMAP is used, which sets the principal distance to $1.25 \cdot \max(image_{width}, image_{height})$. The coordinates of the principal point are always assumed to be in the image center because there exists no metadata for this value.

Theoretically, the coordinates can be estimated within the bundle adjustment but this is only recommended after the initial reconstruction as the principal point estimation is an ill-posed problem [4]. It has been investigated that when considering c_x and c_y as unknowns in the bundle adjustment, the simultaneous estimation of the principal distance decreases in accuracy [33].

3.3.3. Structure-from-Motion for Historical Photographs

After initialization of the interior camera parameters for the historical images, all possible image pairs are matched using SuperPoint [2] + SuperGlue [3] as it provides the most robust results for historical terrestrial images [27,34]. SuperPoint uses homographic adaptation to create multiple representations of a single image and the most significant features for every representation are accumulated in the original image. This method works for all kinds of images and also performs well under illumination and viewpoint changes [2], which are visible in historical photographs [27]. SuperGlue uses a graph neural network to match the derived features which are now represented in self- and cross-attentional aggregations [3]. It recently outperforms conventionally used methods, such as SIFT [35], on all kinds of different datasets [3,27,34,36]. However, the method is not rotationally invariant, which is fortunately irrelevant for digitized images of urban scenes.

The proposed method extracts a maximum of 4096 SuperPoint features in images, which are resized to a maximum of 1600 pixels (edge length). This can be done in the hierarchical localization toolbox (hloc) using the feature extraction method *superpoint_max* [37].

This toolbox is also used to initialize all values for the determined principal distances of the cameras and all feature matches in a single COLMAP project. COLMAP is an open-source SfM software solution allowing, i.e., the import of different feature-matching methods [4]. In order to generate a sparse point cloud, all exterior and interior camera parameters are optimized in a single bundle adjustment procedure. It is recommended to use two-view tracks for historical photographs, which means that the camera parameters are estimated even if feature matches between one single image pair (=a feature track between only two images) only exist.

Without reference data, the final reconstruction in the SfM process is in a local coordinate system. To transfer the exterior orientation (i.e., the camera poses) into the 4D browser, reference data of the 4D browser's global coordinate system is required. Georeferencing is usually accomplished using points of superior accuracy, such as ground control points (GCPs) or terrestrial laser scanning data (TLS). However, TLS data are often not readily available and do not usually cover entire cities. Thus, to provide a realistic scenario that is feasible for other cities, georeferencing is done using the LOD 2 and LOD 3 models provided by the authorities, which may result in lower accuracy.

This last step is still conducted interactively by selecting similar points in the sparse cloud and the 3D city model. It is important to select enough points with a good distribution over the model. In the experiments, the results converged when using a minimum of 15 points spread all over the landmark. This enables building a transformation matrix that maps the points of the sparse cloud (in the local coordinate system) to the points of the 3D model (in the global coordinate system). To calculate the transformation matrix, least squares estimation is used [38].

Transformation of the exterior camera parameters from COLMAP into the 4D browser is not a trivial task; the detailed presentation of the necessary steps is shown in the Appendix A. COLMAP uses the OpenCV camera coordinate system, which exports the camera poses as projections from the world to the camera coordinate system as quaternions defined using the Hamilton convention $R(q_w, q_x, q_y, q_z)$ and the translation vector $T(X, Y, Z)$. These values allow building a rotation matrix for every camera in the OpenCV camera coordinate system.

The 4D browser uses the OpenGL camera coordinate system, which uses a different representation of the rotation matrix. Thus, the OpenCV rotation matrix needs to be adjusted to match the representation. When both rotation matrices are in OpenGL representation, the exterior orientations of the cameras of COLMAP can be transferred to the 4D browser. The resulting rotation matrix allows the determination of the camera quaternions as $R(q_w, q_x, q_y, q_z)$.

Finally, this allows storing all determined values in the 4D browser via an HTTP POST request. This includes the width, height, feature file (pkl), t_x, t_y, t_z, q_w, q_x, q_y, q_z, principal distance (ck), principal point coordinates ($c(x, y)$), and radial distortion (k) (Figure 6).

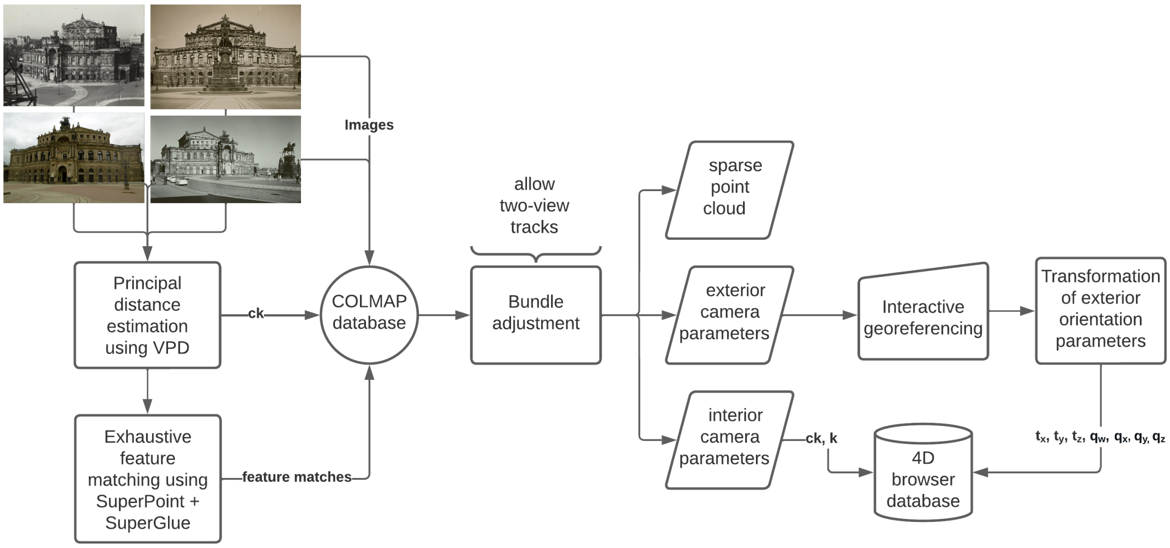

Images
Principal distance estimation using VPD
ck
COLMAP database
allow two-view tracks
Bundle adjustment
sparse point cloud
exterior camera parameters
Interactive georeferencing
Transformation of exterior orientation parameters
interior camera parameters
ck, k
4D browser database
t_x, t_y, t_z, q_w, q_x, q_y, q_z
Exhaustive feature matching using SuperPoint + SuperGlue
feature matches

Figure 6. Workflow for the estimation of the interior and exterior camera parameters of historical images.

3.4. Data Extension

The previously described workflow estimates the camera orientation parameters for around 10% of all images (for one landmark) in the database of the SLUB. However, there exists the possibility that not all images showing the exterior of the landmark are covered and/or that new images are uploaded to the repository due to digitization. In the following, this new image, which shall be added to the database of the browser, is referred to as a *query image*.

With the metadata of the query image, it is possible to determine which building is shown in the photograph. Using the existing values for cameras in the database now avoids the need for repetitive georeferencing. Functionality in the 4D browser uses ray casting to determine whether the building is already depicted in a previously spatialized image. It is checked whether the landmark is visible in at least 20 images. If this is true, all .pkl files from images showing that landmark are retrieved from the database. The newly uploaded image serves as the query image for LEA, and the nearest 20 images are determined as described in Section 3.3.1. The calculated feature file is stored in the 4D browser database.

These images are used in a subsequent SfM procedure using SuperPoint and SuperGlue. However, since the camera orientation parameters of the 20 images are already available, these can be used for faster calculation of the camera parameters of the query image. The 20 images serve as a reference model and are stored as a COLMAP reconstruction file using the inverted transformation from OpenGL to OpenCV camera model. For all possible image pairs between the query image and the other images, the SuperGlue feature matches are calculated. With known camera orientations, the feature matches can be triangulated into 3D spaces. The 3D points that are seen by the query image can be used to determine the camera parameters using the spatial resection provided in *hloc*. There is a possibility of performing a subsequent bundle adjustment, optimizing the camera parameters for all images, which was not necessary when only using one image. Finally, the exterior and interior camera orientation parameters of the query image are uploaded to the database and the image can be used to establish the camera orientation of new images.

3.5. Limitations

Both of the described processes have their limitations, which will be critically discussed in the following. Section 3.3 states that user input is necessary to define a query image for a landmark. This requires the user to know the data and specifically the landmark that is meant to be initialized. An automatic method for the selection of the query images could be query expansion [39]. All images for one landmark would be initially clustered, and it would be possible to select one image that has the highest correlation to all other images.

Furthermore, the CBIR yields the nearest 10% of the query image. That implies that if *more* than 10% of images in the database (Deutsche Fotothek) of the specific landmark are valid, these will not be considered for the subsequent SfM workflow. As this is just a minor issue when the database is known, it becomes a problem if the database and its quality are unknown. In the proposed workflow, a strict (distance) threshold for "good" images could not be determined using LEA. This issue could possibly also be solved using clustering for the result list (see above). Thus, instead of a fixed transfer of 10% of images, the most relevant cluster of images could be transferred In the opposite case, where *less* than 10% of the images show the actual landmark, these are reliably filtered out in the subsequent SfM procedure, by, e.g., increasing the number of minimum matches between image pairs (Section 4).

Another concurrent issue in the CBIR process is that metadata are initially used for pre-processing historical images. However, photographs may not correctly be tagged with metadata and, therefore, would not be retrieved by LEA. Additionally, the initial metadata search yields photographs from all periods when the building existed, including construction sites or former buildings in the same location. This can hinder fully automatic reconstruction via SfM (Section 4.2), but with appropriate parameter settings, most of these issues can be prevented. Another option would be to omit the metadata search and browse the entire database via CBIR for similarities. However, this would be highly computationally intensive and would have to be repeated for newly added images.

The initial georeferencing using LOD 2 or LOD 3 building models is not optimal and possibly results in a lower transformation accuracy. It is obvious that a LOD 2 model does not provide the geometric accuracy of a TLS and does not perfectly match the historical sparse point cloud. TLS data would also allow using automatic georeferencing by transforming the point clouds onto each other using, e.g., the iterative closest point (ICP) algorithm. However, the presented approach fits with the *real-world scenario* of the application where usually the authorities of a city provide the 3D city models. Additionally, for the provided application and the overlay of the historical photograph, visual accuracy is more important than geometric accuracy. It can be assumed that the absolute accuracy of the camera position is around a few meters and the accuracy of the rotational values is around a few degrees [40], which is consistent with the visual results presented in the next section.

4. Results

In this section, all results for the two investigated datasets are shown. This includes the results of the CBIR, the SfM workflow, and the final visual representation in the 4D browser. The results can be directly visited via https://4dbrowser.urbanhistory4d.org, accessed on 1 February 2023. The 3D scene of Dresden has to be opened. Via the metadata search, it is possible to filter for Taschenbergpalais or Semperoper. Images showing the respective views of the two buildings are oriented using the proposed workflow. Double-clicking on a single image allows one to jump into the perspective of the photographer and get an estimate of the visual accuracy. The underlying 3D model can be made visible by using the *Image Opacity* slider on the top left of the web page.

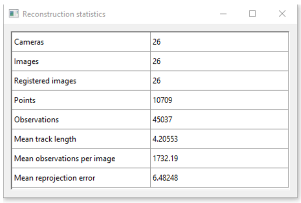
4.1. Taschenbergpalais

LEA yields 58 photographs for Taschenbergpalais. Since the absolute number of photographs is quite limited, a manual check of the results becomes possible. The unobstructed view from the north is only sparsely available in the database, but the majority of results show exactly this perspective. The results can be split into relevant photographs, relevant drawings, and irrelevant data, as shown in Table 1.

Table 1. Distribution of the final results for Taschenbergpalais yielded by LEA.

Relevant Photographs	Relevant Drawings	Other Photographs	Other Drawings
22	8	23	5

As far as all 584 hits in the Deutsche Fotothek were observed, LEA did not omit any relevant hits that should be included in the SfM process. The majority of the 584 images (including drawings and plans) show the building's interior or an extremely destroyed building state after 1945, which is not relevant for the reconstruction. That means less than 4% of the metadata search hits show the desired perspective, emphasizing the use of CBIR followed by SfM. For all images, it is desirable to use only the relevant photographs for the SfM reconstruction. However, SuperPoint + SuperGlue is even capable of matching geometrically correct drawings to photographs. Thus, it is expected to estimate the camera parameters of 22 to 30 images/drawings. Of course, in the case of drawings, the camera cannot be connected to a physical model but can be seen as a virtual camera for which the parameter errors are minimized during bundle adjustment. The final reconstruction in COLMAP shows 26 registered images, including 3 drawings (Figure 7).



Cameras	26
Images	26
Registered images	26
Points	10709
Observations	45037
Mean track length	4.20553
Mean observations per image	1732.19
Mean reprojection error	6.48248

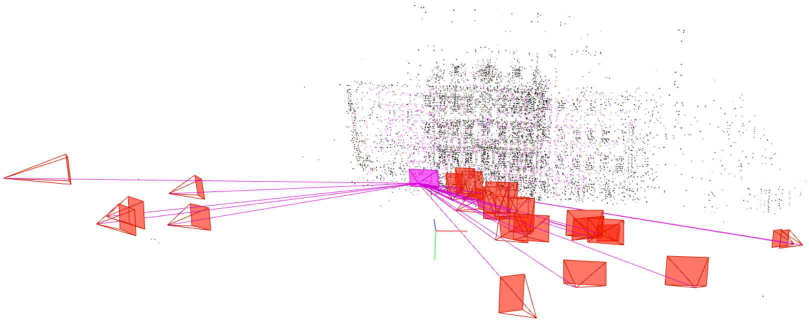


Figure 7. Final reconstruction in COLMAP showing the estimated camera positions of 26 photographs and drawings. The sparse point cloud consists of approximately 10,000 points.

An overview of the visual appearance and quality of images is shown in Figure 8. One image is connected to the scene, but is positioned incorrectly as it shows the building from a different perspective (Figure 8-bottom left).

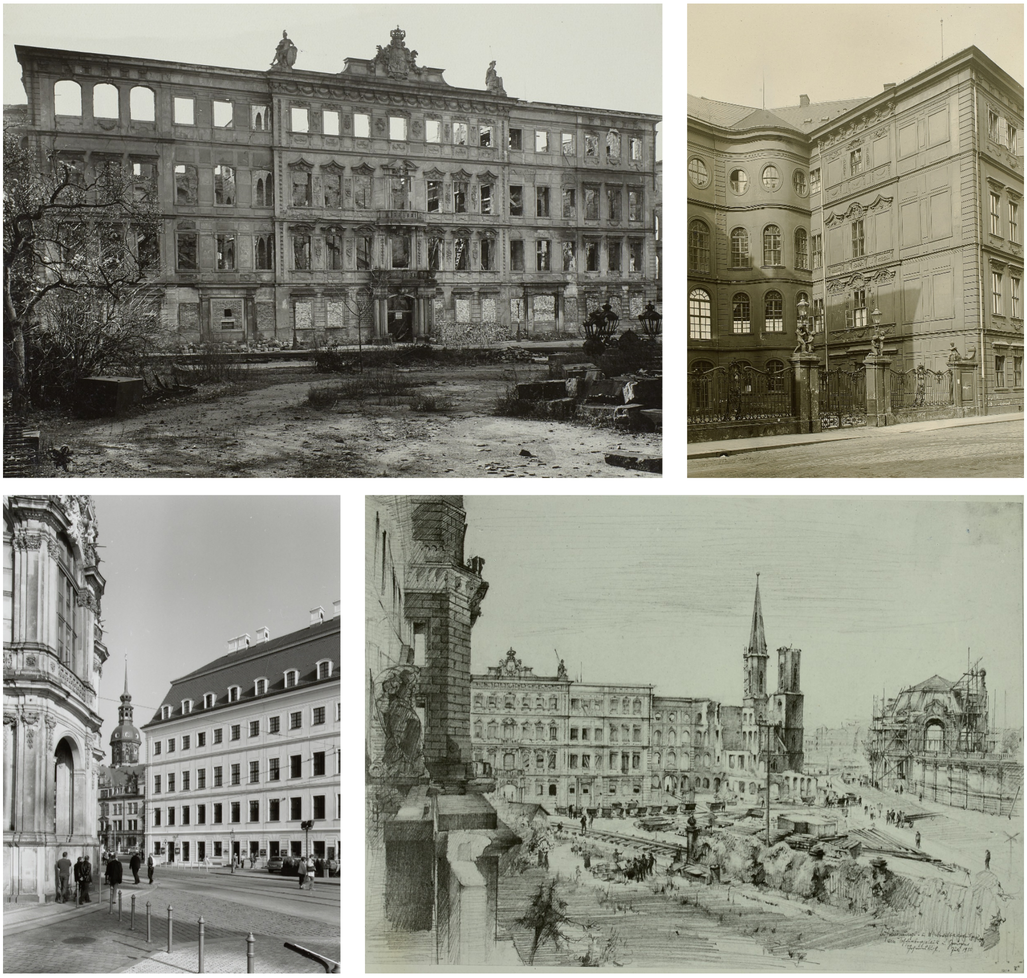


Figure 8. Four images of Taschenbergpalais that are retrieved by LEA. The automatic SfM reconstruction in COLMAP reveals the correct camera orientation for three images, while the bottom left image is incorrectly matched and oriented due to the repeating pattern of the building.

Nonetheless, all camera parameters are transferred to the 4D browser after georeferencing using the sparse cloud and a LOD 3 model of Taschenbergpalais by selecting 15 corresponding points (Figure 9-left). The accuracy of the transformation between the 3D model and sparse cloud yields σ_0 = 1.12 m with a maximum deviation in XYZ of 3.52 m. It was not possible to achieve higher accuracy, even when selecting more points or changing the distribution of similar points.

Double-clicking on an image allows one to jump into its perspective by using the calculated camera parameters. Using the *Image Opacity* slider on the top left toolbar allows visually estimating the accuracy of the approach with the underlying 3D model (Figure 9-right).

This depiction allows us to assume that the camera parameters are estimated with sufficient quality for visualization purposes. However, in object space (on the 3D model), deviations lie in a range of a few to several meters, especially visible on the eaves. This can be attributed to errors of a few degrees in the camera orientation angles or errors in the decimeter range for the camera position, which could have occurred during the transformation process due to the generalized LOD 3 model and the quality of the historical sparse point cloud.

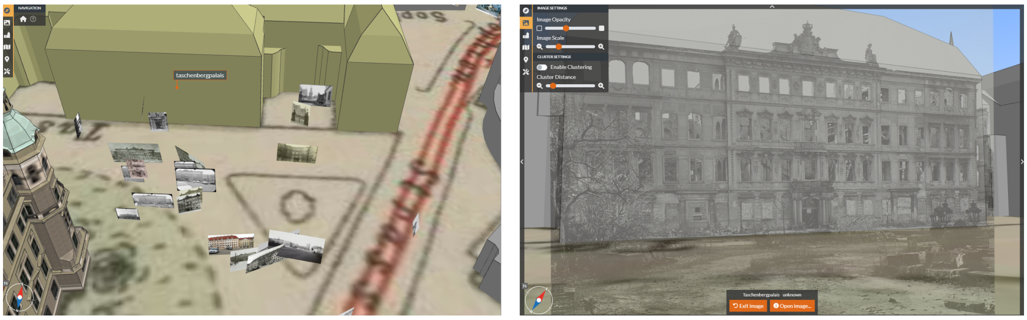


Figure 9. (**Left**): Final reconstruction results as depicted in the 4D browser. The images are visualized with their thumbnail images in the 3D space. (**Right**): Reconstructed perspective of one historical photograph blended with the underlying 3D model of Taschenbergpalais.

Nonetheless, the obtained results are used to estimate the camera orientation parameters of an image that has been added to the database. The image shows a similar perspective of the Taschenbergpalais taken from a nearby building, and the closest 20 images are retrieved by LEA using the pre-calculated .pkl files that are stored in the 4D browser. The feature matching is then performed, the tie points are triangulated, and the camera parameters of the new image are calculated via spatial resection (Figure 10).

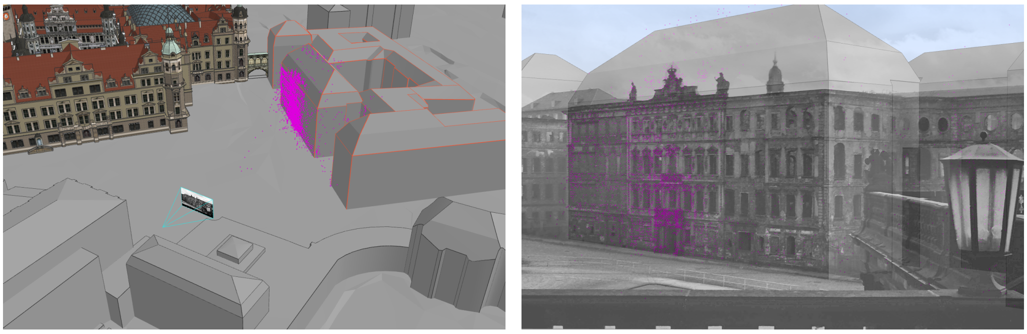


Figure 10. (**Left**): The overview in the 4D browser of the new image and the visualized tie points used for spatial resection. (**Right**): The perspective given by the camera parameters of the newly estimated image [40].

As expected, the image shows similar inaccuracies on the eaves as these were already introduced by the existing reconstruction. However, the results show that it is possible to determine the proper location and orientation of the camera in an automatic way.

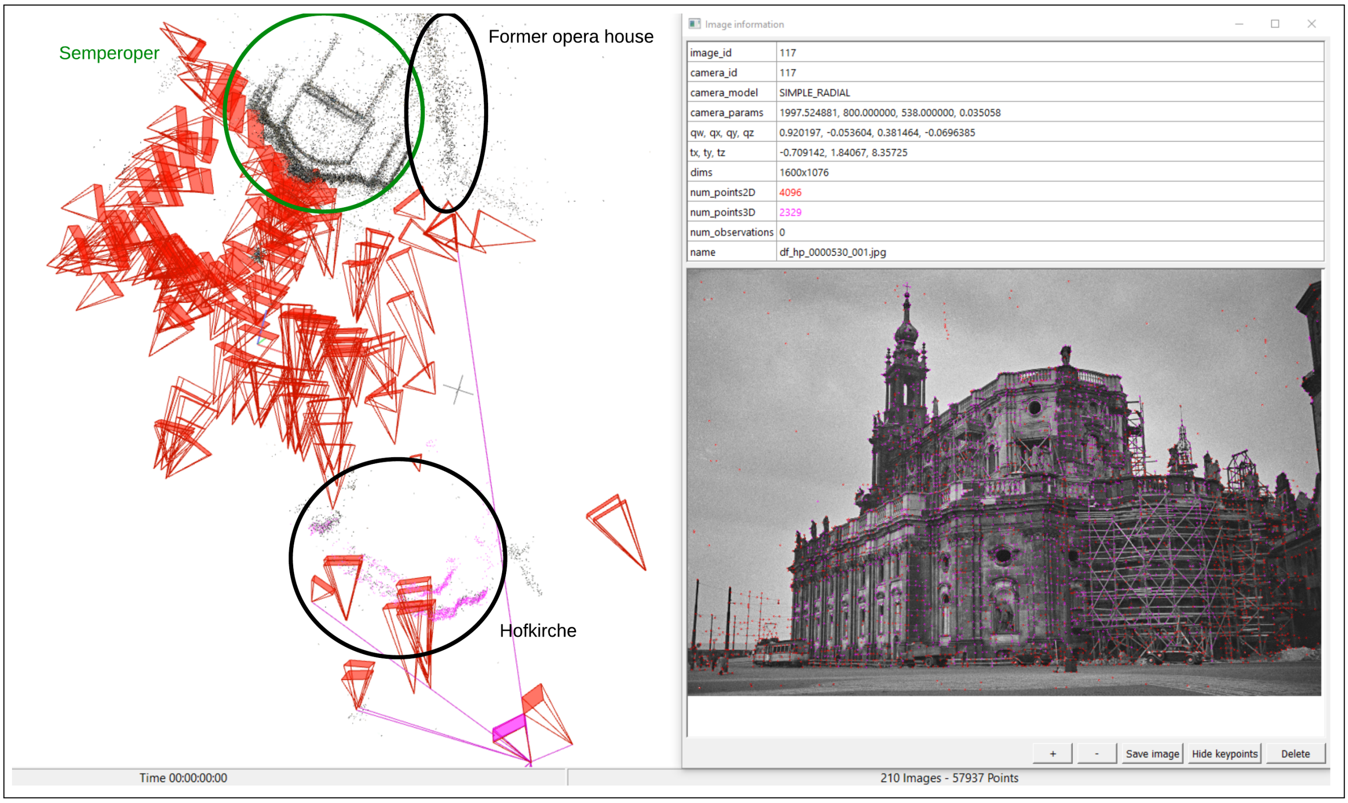
4.2. Semperoper

The Semperoper dataset reveals if the workflow is also able to deal with a larger number of images. LEA yields 217 photographs for the Semperoper. The result list is also checked for comparison with Taschenbergpalais, as shown in Table 2. However, a check of all 2172 digitized images is not performed.

Table 2. Table showing the distribution of final results for the Semperoper yielded by LEA. LEA also yields images of the Hofkirche, which is situated across the Semperoper. These images are also tagged in the metadata via the *Semperoper* keyword but are not meant to be used for its reconstruction.

Relevant Photographs	Drawings/Postcards	Other Photographs (Hofkirche)	Former Opera House	Stereo Photos
178	8	7	22	2

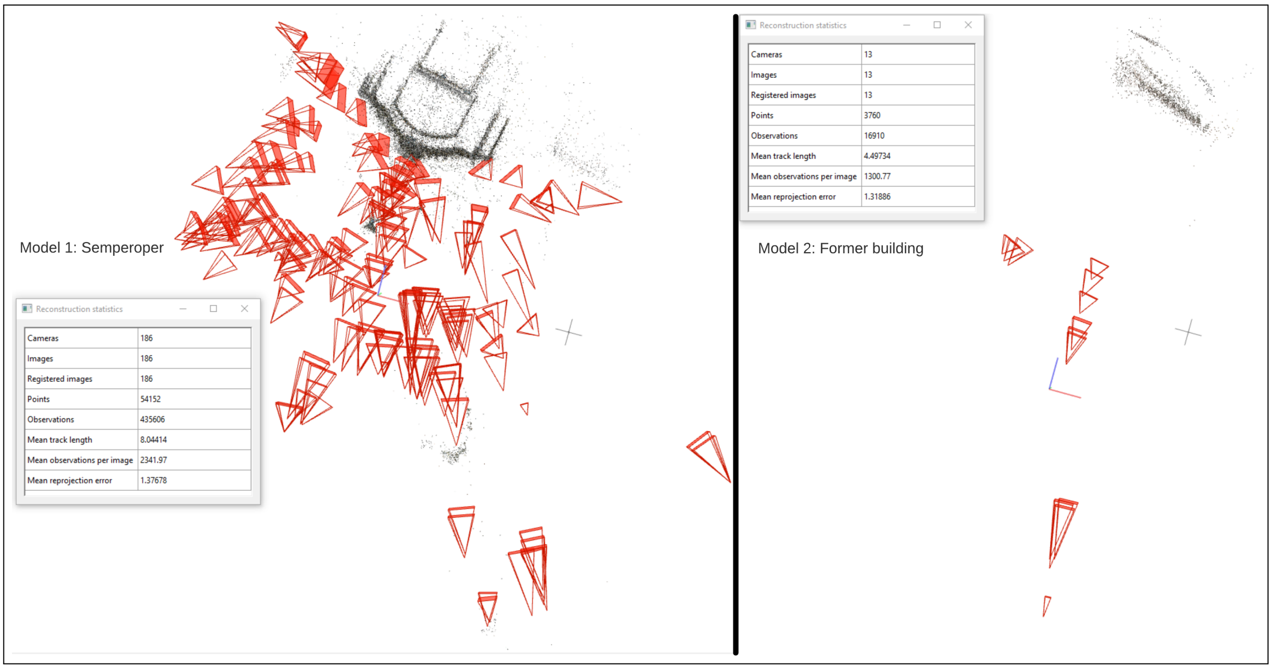
As for the other dataset, it is expected that all 178 relevant photographs can be oriented in the automatic SfM workflow using SuperPoint + SuperGlue feature matching. However, the first result shows inaccuracies in the reconstruction introduced by a small number of mismatches between some image pairs (Figure 11). While the images depicting the Semperoper in its contemporary state are positioned correctly, COLMAP also merges seven images of the Hofkirche (located across from the Semperoper building) and 15 images of the former opera house into the reconstruction, likely introducing errors to the other photographs. Additionally, these buildings are not located in the correct global positions.



image_id	117
camera_id	117
camera_model	SIMPLE_RADIAL
camera_params	1997.524881, 800.000000, 538.000000, 0.035058
qw, qx, qy, qz	0.920197, -0.053604, 0.381464, -0.0696385
tx, ty, tz	-0.709142, 1.84067, 8.35725
dims	1600x1076
num_points2D	4096
num_points3D	2329
num_observations	0
name	df_hp_0000530_001.jpg

Figure 11. Initial COLMAP reconstruction of all Semperoper images retrieved by LEA. The green area marks the Semperoper at the correct position while images of Hofkirche (as seen from the Semperoper building) and the former opera house (black areas) introduce errors to the reconstruction. The orientation of the Hofkirche building is incorrect due to sparse image material accidentally retrieved by LEA and repetitive structures.

These errors can be prevented by either (a) increasing the matching threshold of SuperGlue, or (b) increasing the minimum number of matches between image pairs used in the bundle adjustment of COLMAP. To avoid discarding relevant matches found by SuperGlue, option (b) was chosen. By increasing the minimum number of matches between image pairs from the standard value of 15 to 200, the final reconstruction is correct without the previously created artifacts (Figure 12). Using a similar procedure for Taschenbergpalais removes the previously incorrectly positioned image (shown in Figure 8). Another advantage is that the former opera house is reconstructed separately by COLMAP. Since this previous building was destroyed in a fire in 1869, only a few low-quality images exist.



Model 1: Semperoper

Cameras	186
Images	186
Registered images	186
Points	54152
Observations	435606
Mean track length	8.04414
Mean observations per image	2341.97
Mean reprojection error	1.37678

Model 2: Former building

Cameras	13
Images	13
Registered images	13
Points	3760
Observations	16910
Mean track length	4.49734
Mean observations per image	1300.77
Mean reprojection error	1.31886

Figure 12. (**Left**): Correct COLMAP reconstruction of all Semperoper images retrieved by LEA with an increased number of minimum matches between image pairs. (**Right**): Reconstruction of the former opera house of 1869 automatically created in the bundle adjustment in COLMAP.

The final results of the reconstruction show that SuperPoint + SuperGlue is again able to match all relevant photographs and drawings retrieved by LEA. Visual control of the camera poses reveals that the camera orientation parameters seem to be estimated correctly.

Again, the cameras are georeferenced. As the study area is larger, 22 corresponding points are selected in the LOD 2 model and the sparse point cloud. The transformation accuracy is comparable to the Taschenbergpalais dataset with a σ_0 = 1.84 m and a maximum deviation in XYZ of 7.25 m. Considering the size of the Semperoper and the distance between the photographer and the building, this seems to be an acceptable range, though it becomes clear that georeferencing with a LOD 2 model can often be inaccurate. The resulting camera frustums in the browser and an example are shown in Figure 13.

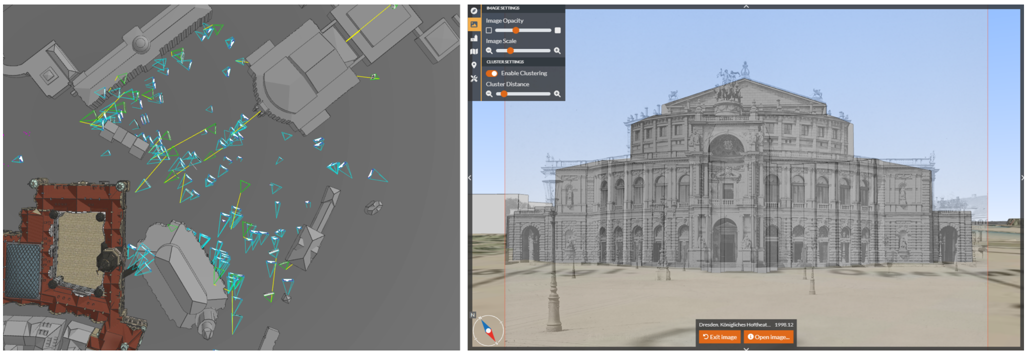


Figure 13. (**Left**): Final reconstruction results as depicted in the 4D browser. (**Right**): Perspective view of one image using the opacity slider to reveal the underlying 3D model of the Semperoper.

4.3. Discussion

For the presented results, no absolute accuracy is given at the moment. This is mainly due to the fact that it is almost impossible to generate historical reference data. Consequently, the evaluation criteria are the mean reprojection error of the COLMAP reconstruction, the σ_0 of the geometric transformation between LOD2 model and sparse point cloud, and the visual aspects of the 4D application.

Both experiments have led to several conclusions on the efficiency of the method and best practices for further datasets. The developed CBIR approach, in combination with an initial metadata search, serves as an efficient and accurate method for retrieving all exterior views of landmarks by selecting a suitable query image. Even if several false positives are retrieved, these are filtered out by the subsequent SfM approach in COLMAP, and usually all camera parameters can be estimated in a single SfM model. In addition to the mentioned datasets, the method has been tested on further landmarks in the vicinity of Dresden, and false positives have been identified and counted (Figure 14).

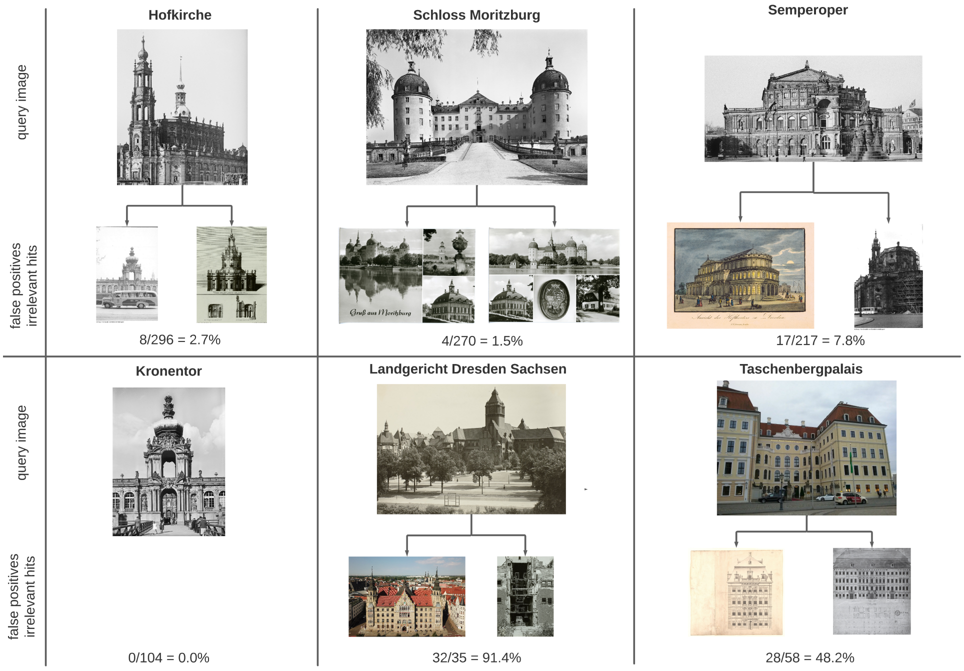


Figure 14. Metadata search tag and query image for six different landmarks in the vicinity of Dresden. Below every landmark, two examples and the absolute number of false positives retrieved by LEA are given. An especially high number of false positives can be seen for the smaller Landgericht and Taschenbergpalais datasets. These occur because there are no other relevant hits in the image repository.

All retrieved images for the respective landmarks are processed in the specified SfM workflow and the resulting characteristics of the reconstruction can be seen in Table 3.

Table 3. Results of the SfM processing in COLMAP using the SuperGlue feature matching. For all datasets, almost all relevant photographs can be oriented in a single reconstruction with reasonable mean reprojection errors. For the Landgericht dataset with only 3 relevant images, reconstruction is not possible.

Dataset	**Relevant Photographs**	**Reconstructed Cameras**	**Sparse Points**	**Mean Reprojection Error**
Hofkirche	296	235	56,359	1.47
Schloss Moritzburg	270	227	26,698	1.36
Semperoper	186	186	57,937	1.35
Kronentor	104	104	34,130	1.49
Landgericht Dresden Sachsen	3	NA	NA	NA
Taschenbergpalais	30	26	10,709	6.48

The required computing time depends on the number of processed reference images. According to empirical studies on the datasets presented above, a minimum of 20 reference images is needed for generating a reasonable SfM model and accurate camera orientation parameters. Significant depth information on the landmark is needed to be seen in several of the historical photographs, especially for successful modeling of the principal distance.

In the following, the operating system used and the computation time of the standard workflow for processing around 1000 images in the base repository are shown. LEA and the feature-matching SuperGlue are computed on an HPC system with a single NVidia V100 GPU with 32 GB VRAM. SfM with COLMAP, database transfer, and orientation of single new images are run on a local machine with an i7-1165G CPU with a clock speed of 2.80 GHz. For the Kronentor dataset, which has approximately 1040 images in the original database, the CBIR LEA initially takes 10 min to retrieve the 104 most relevant images. These images are passed to SuperGlue, which takes 25 min for feature point calculation and matching. The subsequent bundle adjustment in COLMAP takes 12 min, resulting in a total processing time of 47 min for 1000 database images. Single newly uploaded images can be processed in a few minutes due to the existing CBIR feature files and SuperGlue feature matches.

At the moment, the presented approach is mainly tested on specific urban landmarks for which several historical are provided. As an empirical value, a SfM reconstruction could almost always be obtained if 20 or more images of the same building view are available. It is necessary that some of the images provide depth information of the scene to retrieve reliable reconstructions. Using these existing reconstructions of a landmark (the saved camera parameters in global coordinates) finally allows adding images of side streets or non-famous buildings at a larger scale, which has not been done before.

5. Conclusions

The presented methodology allows for the automatic determination of camera parameters for historical photographs. CBIR allows for the reliable selection of relevant images for the subsequent SfM workflow by selecting one representative query image of a landmark. Using SuperPoint + SuperGlue feature-matching enables finding similar points between image pairs with large radiometric and geometric differences as well as filtering out images that do not belong to the building. For larger image datasets, it is recommended to increase the minimum number of matches between single image pairs in order to avoid noisy 3D models generated by COLMAP. The workflow was tested and evaluated on two buildings with different appearances, different LODs of the 3D model, and different image numbers in the repository. In contrast to many other works, all calculated data are stored in the database of the 4D browser. This allows a quick search of similar images by using the pre-calculated feature vector of the CBIR. Additionally, as the camera parameters of images in the database are known, newly uploaded images can be used as a query for the CBIR and its camera parameters can easily be determined in the global coordinate system using a standard photogrammetric procedure.

The restrictions of the presented pipeline, at the moment, include limited accuracy, which is probably introduced by using LOD 2 and LOD 3 models for the transformation of the camera poses. The deviations are especially visible at the edges and eaves of the 3D building model. While these inaccuracies are of minor relevance in the depicted application, they might not be acceptable for augmented reality applications. In the next step, these findings are planned to be verified by using TLS data with superior accuracy for transformation.

Nonetheless, the presented research is an important step toward the automatic generation and storage of historical city models and camera parameters in urban cultural heritage scenarios.

Author Contributions: Conceptualization, F.M., J.B., D.S., M.W. and F.N.; methodology, F.M. and J.B.; software, F.M. and J.B. ; validation, F.M. and J.B.; formal analysis, F.M. and J.B.; investigation, F.M. and J.B.; resources, F.M. and J.B.; data curation, F.M. and J.B.; writing—original draft preparation, F.M. and J.B.; writing—review and editing, F.M., J.B., D.S., M.W. and F.N. ; visualization, F.M. and J.B.; supervision, D.S., M.W. and F.N.; project administration, F.N.; funding acquisition, F.N. All authors have read and agreed to the published version of the manuscript.

Funding: The research undertaken in the project (HistKI) was funded by the German Federal Ministry of Education and Research under grant number 01UG2120A.

Data Availability Statement: All image data used in this study can be found at https://www.deutschefotothek.de/, accessed on 30 November 2022. The data can also be accessed via the presented 4D browser application under https://4dbrowser.urbanhistory4d.org/, accessed on 30 November 2022.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

3D	three-dimensional
4D	four-dimensional
GIS	geographic information system
SfM	Structure-from-Motion
CBIR	content-based image retrieval
LOD	level of detail
SLUB	The Saxon State and University Library Dresden
LEA	layer extraction approach
CNN	convolutional neural network
pkl	pickle file extension
kB	kilobytes
VPD	vanishing point detection
VP	vanishing point
hloc	hierarchical localization toolbox
TLS	terrestrial laser scanning

Appendix A

The appendix contains information on how to convert the local COLMAP reconstruction in OpenCV notation into the global coordinate frame of the 4D browser using OpenGL notation. A COLMAP reconstruction consists of three different files containing information on the images (*images.bin*), cameras (*cameras.bin*), and sparse point cloud (*points3D.bin*). For the Helmert transformation of all available data, only the camera file is of relevance. The exterior orientation of a camera in COLMAP is represented as a projection from the world coordinate system to the camera coordinate system. The pose of a single camera is determined by the rotation, using quaternions defined in the Hamilton convention $R(q_w, q_x, q_y, q_z)$, and the translation vector $T(X, Y, Z)$. To derive the projection center, the quaternions have to be normalized and multiplied with the negative translation vector (Equation (A1)).

$$\text{projection center}(X, Y, Z) = R(\hat{q_w}, \hat{q_x}, \hat{q_y}, \hat{q_z}) \cdot -T(X, Y, Z) \tag{A1}$$

Transformation of the rotational component of a single camera requires the derivation of the rotation matrix R_{OpenCV} from the original quaternions (Equation (A2)).

$$R_{OpenCV} = \begin{bmatrix} 1 - 2q_y q_y - 2q_z q_z & 2q_x q_y - 2q_z q_w & 2q_x q_z + 2q_y q_w \\ 2q_x q_y + 2q_z q_w & 1 - 2q_x q_x - 2q_z q_z & 2q_y q_z - 2q_x q_w \\ 2q_x q_z - 2q_y q_w & 2q_y q_z + 2q_x q_w & 1 - 2q_x q_x \end{bmatrix} \tag{A2}$$

The resulting rotation matrix has to be further processed. First, the second and third rows of the matrix have to be negated to follow the definition of the camera coordinate system in OpenGL (cf. Figure A1). Secondly, the resulting rotation matrix has to be transposed to match OpenGL's order of elements in the rotation matrix (column-major instead of OpenCV's row-major definition), as shown in Equation (A3).

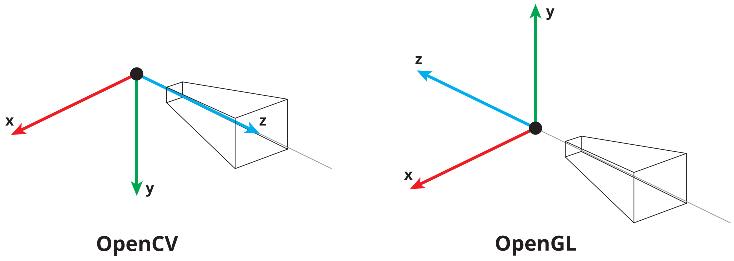


Figure A1. Conversion between coordinate systems: In OpenCV, the camera looks along the positive z-axis and the y-axis points downward. In OpenGL, the camera looks along the negative z-axis, and the y-axis points upwards.

$$R_{OpenGL} = (R_{OpenCV} \cdot \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix})^T \tag{A3}$$

The local camera pose is now determined by the homogeneous synthesis of the rotation and translation component (Equation (A4)).

$$pose_{local} = \begin{bmatrix} R_{local} & & & T_{local} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A4}$$

The transformation matrix can be derived from point correspondences in both coordinate systems. This geometric transformation in 3D space is called the Helmert transformation, seven-parameter transformation, or similarity transformation. Least-squares estimation is used to find this transformation matrix between these correspondences [38].

To derive the global camera poses, the local pose is simply multiplied by the transformation matrix (Equation (A5)).

$$\begin{bmatrix} R_{global} & & & T_{global} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{transformation} & & & T_{transformation} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{local} & & & T_{local} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A5}$$

References

1. Friedrichs, K.; Münster, S.; Kröber, C.; Bruschke, J., Creating Suitable Tools for Art and Architectural Research with Historic Media Repositories. In *Digital Research and Education in Architectural Heritage. UHDL 2017, DECH 2017*; Münster, S., Friedrichs, K., Niebling, F., Seidel-Grzesińska, A., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 117–138. [CrossRef]
2. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 337–33712. [CrossRef]
3. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching with Graph Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 4938–4947. [CrossRef]
4. Schönberger, J.L.; Frahm, J.M. Structure-from-Motion Revisited. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113. [CrossRef]
5. Schaffland, A.; Heidemann, G. Heritage and Repeat Photography: Techniques, Management, Applications, and Publications. *Heritage* **2022**, *5*, 4267–4305. [CrossRef]
6. Gaetani, C.I.D.; Ioli, F.; Pinto, L. Aerial and UAV Images for Photogrammetric Analysis of Belvedere Glacier Evolution in the Period 1977–2019. *Remote Sens.* **2021**, *13*, 3787. [CrossRef]
7. Knuth, F.; Shean, D.; Bhushan, S.; Schwat, E.; Alexandrov, O.; McNeil, C.; Dehecq, A.; Florentine, C.; O'Neel, S. Historical Structure from Motion (HSfM): Automated processing of historical aerial photographs for long-term topographic change analysis. *Remote Sens. Environ.* **2023**, *285*, 113379. [CrossRef]
8. Nebiker, S.; Lack, N.; Deuber, M. Building Change Detection from Historical Aerial Photographs Using Dense Image Matching and Object-Based Image Analysis. *Remote Sens.* **2014**, *6*, 8310–8336. [CrossRef]
9. Deane, E.; Macciotta, R.; Hendry, M.T.; Gräpel, C.; Skirrow, R. Leveraging historical aerial photographs and digital photogrammetry techniques for landslide investigation—A practical perspective. *Landslides* **2020**, *17*, 1989–1996. [CrossRef]
10. Meixner, P.; Eckstein, M. Multi-Temporal Analysis of WWII Reconnaissance Photos. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B8*, 973–978. [CrossRef]

11. Khalil, O.A.; Grussenmeyer, P. 2D & 3D Reconstruction workflows from archive images, case study of damaged monuments in Bosra Al-Sham city (Syria). *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W15*, 55–62. [CrossRef]
12. Maiwald, F.; Maas, H.G. An automatic workflow for orientation of historical images with large radiometric and geometric differences. *Photogramm. Rec.* **2021**, *36*, 77–103. [CrossRef]
13. Farella, E.M.; Özdemir, E.; Remondino, F. 4D Building Reconstruction with Machine Learning and Historical Maps. *Appl. Sci.* **2021**, *11*, 1445. [CrossRef]
14. Beltrami, C.; Cavezzali, D.; Chiabrando, F.; Idelson, A.I.; Patrucco, G.; Rinaudo, F. 3D Digital and Physical Reconstruction of a Collapsed Dome using SFM Techniques from Historical Images. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W11*, 217–224. [CrossRef]
15. Stellacci, S.; Condorelli, F. Remote survey of traditional dwellings using advanced photogrammetry integrated with archival data: The case of Lisbon. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *XLIII-B2-2022*, 893–899. [CrossRef]
16. Muenster, S. Digital 3D Technologies for Humanities Research and Education: An Overview. *Appl. Sci.* **2022**, *12*, 2426. [CrossRef]
17. Manferdini, A.M.; Remondino, F. Reality-Based 3D Modeling, Segmentation and Web-Based Visualization. In *Digital Heritage*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 110–124. [CrossRef]
18. Nishanbaev, I. A web repository for geo-located 3D digital cultural heritage models. *Digit. Appl. Archaeol. Cult. Herit.* **2020**, *16*, e00139. [CrossRef]
19. Gominski, D.; Gouet-Brunet, V.; Chen, L. Connecting Images through Sources: Exploring Low-Data, Heterogeneous Instance Retrieval. *Remote Sens.* **2021**, *13*, 3080. [CrossRef]
20. Schaffland, A.; Vornberger, O.; Heidemann, G. An Interactive Web Application for the Creation, Organization,and Visualization of Repeat Photographs. In Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia Heritage Contents, Nice, France, 21 October 2019. [CrossRef]
21. Fanini, B.; Ferdani, D.; Demetrescu, E.; Berto, S.; d'Annibale, E. ATON: An Open-Source Framework for Creating Immersive, Collaborative and Liquid Web-Apps for Cultural Heritage. *Appl. Sci.* **2021**, *11*, 11062. [CrossRef]
22. Champion, E.; Rahaman, H. Survey of 3D digital heritage repositories and platforms. *Virtual Archaeol. Rev.* **2020**, *11*, 1. [CrossRef]
23. Münster, S.; Maiwald, F.; Bruschke, J.; Kröber, C.; Dietz, R.; Messemer, H.; Niebling, F. Where Are We Now on the Road to 4D Urban History Research and Discovery? *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *VIII-M-1-2021*, 109–116. [CrossRef]
24. Schindler, G.; Dellaert, F. 4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections. *J. Multimed.* **2012**, *7*, 124–131. [CrossRef]
25. Bruschke, J.; Wacker, M.; Niebling, F., Comparing Methods to Visualize Orientation of Photographs: A User Study. In *Research and Education in Urban History in the Age of Digital Libraries. UHDL 2019*; Niebling, F., Münster, S., Messemer, H., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 129–151. [CrossRef]
26. Bekiari, C.; Bruseker, G.; Doerr, M.; Ore, C.E.; Stead, S.; Velios, A. *Definition of the CIDOC Conceptual Reference Model v7.1.1*; The CIDOC Conceptual Reference Model Special Interest Group: Berlin, Germany, 2021 . [CrossRef]
27. Maiwald, F.; Lehmann, C.; Lazariv, T. Fully Automated Pose Estimation of Historical Images in the Context of 4D Geographic Information Systems Utilizing Machine Learning Methods. *ISPRS Int. J. -Geo-Inf.* **2021**, *10*, 748. . [CrossRef]
28. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 512–519. [CrossRef]
29. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
30. Razavian, A.S.; Sullivan, J.; Carlsson, S.; Maki, A. Visual instance retrieval with deep convolutional networks. *ITE Trans. Media Technol. Appl.* **2016**, *4*, 251–258. [CrossRef]
31. Maiwald, F. Generation of a Benchmark Dataset Using Historical Photographs for an Automated Evaluation of Different Feature Matching Methods. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *XLII-2/W13*, 87–94. [CrossRef]
32. Li, B.; Peng, K.; Ying, X.; Zha, H. Simultaneous Vanishing Point Detection and Camera Calibration from Single Images. In *Advances in Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 151–160. [CrossRef]
33. de Agapito, L.; Hayman, E.; Reid, I. Self-Calibration of a Rotating Camera with Varying Intrinsic Parameters. In Proceedings of the British Machine Vision Conference, Southampton, UK, 14–17 September 1998 ; Nixon, M.; Carter, J., Eds.; British Machine Vision Association: Southampton, UK, 1998; pp. 105–114. [CrossRef]
34. Morelli, L.; Bellavia, F.; Menna, F.; Remondino, F. Photogrammetry Now And Then—From Hand-Crafted To Deep-Learning Tie Points. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *XLVIII-2/W1-2022*, 163–170. [CrossRef]
35. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
36. Jin, Y.; Mishkin, D.; Mishchuk, A.; Matas, J.; Fua, P.; Yi, K.M.; Trulls, E. Image Matching Across Wide Baselines: From Paper to Practice. *Int. J. Comput. Vis.* **2020**, *129*, 517–547. [CrossRef]
37. Sarlin, P.E.; Cadena, C.; Siegwart, R.; Dymczyk, M. From Coarse to Fine: Robust Hierarchical Localization at Large Scale. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 12708–12717. [CrossRef]
38. Umeyama, S. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 376–380. [CrossRef]

39. Chum, O.; Mikulik, A.; Perdoch, M.; Matas, J. Total recall II: Query expansion revisited. In Proceedings of the CVPR 2011, Springs, CO, USA, 20–25 June 2011. [CrossRef]
40. Maiwald, F. 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, Dresden, Germany, 2022. Available online: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-810852 (accessed on 12 February 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.