

# AUTOMATIC ALGORITHM FOR GEOREFERENCING HISTORICAL-TO-NOWADAYS AERIAL IMAGES ACQUIRED IN NATURAL ENVIRONMENTS

Daniela Craciun<sup>1,\*</sup>; Arnaud Le Bris<sup>1</sup>

<sup>1</sup> LASTIG, Univ. Gustave Eiffel, IGN, ENSG, F-94160 Saint-Mandé, France – {Daniela.Craciun, Arnaud.Le-Bris}@ign.fr

**KEY WORDS:** patch-based feature-less aerial image matching, Histogram of Gradients, Digital Surface Model (DSM), automated aerial multi-epoch georeferencing.

## ABSTRACT:

Automatic georeferencing for historical-to-nowadays aerial images represents the main ingredient for supplying territory evolution analysis and environmental monitoring. Existing georeferencing methods based on feature extraction and matching reported successful results for multi-epoch aerial images acquired in structured and man-made environments. While improving the state-of-the-art of the multi-epoch georeferencing problem, such frameworks present several limitations when applied to unstructured scenes, such as natural feature-less environments, characterized by homogenous or texture-less areas. This is mainly due to the lack of structured areas which often results in sparse and ambiguous feature matches, introducing inconsistencies during the pose estimation process. This paper addresses the automatic georeferencing problem for historical aerial images acquired in unstructured natural environments. The research work presented in this paper introduces a feature-less algorithm designed to perform historical-to-nowadays image matching for pose estimation in a fully automatic fashion. The proposed algorithm operates within two stages: (i) 2D patch extraction and matching and (ii) 3D patch-based local alignment. The final output is a set of 3D patch matches and the 3D rigid transformation relating each homologous patches. The obtained 3D point matches are designed to be injected into traditional multi-views pose optimisation engines. Experimental results on real datasets acquired over Fabas area situated in France demonstrate the effectiveness of the proposed method. Our findings illustrate that the proposed georeferencing technique provides accurate results in presence of large periods of time separating historical from nowadays aerial images (up to 48 years time span).

## 1. INTRODUCTION AND MOTIVATION

Generating historical 3D maps from aerial photogrammetric surveys acquired at different epochs represents a valuable information for a wide range of civilian and military applications. Detecting environmental changes across several epochs (Lucas and Gayer, 2021), (Feurer and Vinatier, 2018), supplying infrastructure for ecological transformation (Pinto et al., 2019), but also natural disaster prevention and management are several applications requiring automatic frameworks for producing accurately georeferenced historical Geographical Information Systems (GIS) encoded as orthoimages, Digital Surface Models (DSMs) or land cover maps.

Traditional georeferencing methods rely on operator's intervention for generating ground control points (GCP) (Pinto et al., 2019). While providing accurate solutions, such techniques are laborious at small scale and unfeasible when applied on massive datasets acquired over wide areas and coming from different epochs. Recently reported automatic frameworks exploit feature-based approaches (Feurer and Vinatier, 2018), (Zhang et al., 2021), being suitable for scenes representing structured environments. Therefore, introducing fully automatic techniques capable to provide accurate georeferencing of multi-epoch aerial imagery acquired in natural feature-less environments remains an open issue.

This paper addresses the automatic georeferencing problem for multi-views partially overlapped historical aerial images, acquired in unstructured natural and rural environments. Generally, the key issue standing behind the automation of the multi-epoch georeferencing process is represented by the capability

to extract homologous "features" or "landmarks" in presence of consistent environment changing which characterizes multi-epochs acquisition scenarios. This problem becomes more severe for images acquired in natural and rural environments which are generally less structured than built urban areas and where the absence of reliable detectable and trackable features over time can not be guaranteed. Natural environments are opposite to urban, man-made areas for which it is easier to extract homologous inter-date features (points or lines) in the neighbourhood of anthropic objects. In addition, as stated in (Giordano et al., 2018), the detected trackable features must be homogeneously dispersed over the entire 2D image space in order to ensure a stable multi-view pose optimization process via the bundle adjustment algorithm (Hartley and Zisserman, 2003), (Rupnik et al., 2017).

In order to address the aforementioned key issues, this paper presents a fully automatic patch-based method designed for accurately georeferencing historical aerial images acquired in natural environments. The present paper is organized as following: the next section summarizes existing georeferencing methods for historical datasets, emphasizing our main contributions. Section 3 presents the global overview of the proposed georeferencing algorithm, while highlighting the composing processing stages, each of which being detailed in the following two sections. Experimental results and their performance evaluation are presented in Section 6, while Section 7 draws the conclusions of our contribution and synthesizes future research directions.

## 2. RELATED AND PROPOSED WORK

Designing methods for supplying automatic georeferencing of aerial images acquired across different epochs requires to com-

\* Corresponding author

pute the optimal 3D rigid transformation lying between historical and recent images. Multi-epoch algorithms often proceed by first establishing homologous feature points between the historical and the recent images. In a second stage, image matches are injected into the pose computation process (Rupnik et al., 2017), (Hartley and Zisserman, 2003), (Moulon et al., 2016) and multi-view optimization via the bundle adjustment algorithm.

The main weakness of the existing multi-epoch georeferencing frameworks is represented by the difficulty to establish long-term feature matches based on which the pose computation can be performed. This paper focuses on the inter-epoch homologous image points computation stage which must be ensured in order to allow automatic georeferencing of multi-epochs aerial images acquired in natural environments.

Following the image matching strategy, it is possible to distinguish two main categories of methods. The first category relies on Ground Control Points (GCP) which are usually provided by an operator (Pinto et al., 2019) or supplied automatically (Giordano et al., 2018). While generating accurate results, manual GCP-based frameworks (Pinto et al., 2019) remain fastidious and difficult to be applied on large scale scenes. The second category of georeferencing methods is represented by feature-based approaches. Reported frameworks exploit SIFT descriptors (Feurer and Vinatier, 2018), (Zhang et al., 2021), line segments (Cléry et al., 2014) and deep learning approaches (Zhang et al., 2021). In (Feurer and Vinatier, 2018), authors combine SIFT extraction and matching (Lowe, 2004) techniques, with Structure from Motion (SfM) and multi-view bundle adjustment frameworks (Rupnik et al., 2017). In (Zhang et al., 2021), authors introduce an additional processing stage for filtering image matches provided by deep learning approaches, such as SuperGlue (Sarlin et al., 2020). This illustrates that deep learning techniques present several drawbacks when applied to the multi-epoch image matching problem. In addition, such algorithms require exhaustive learning which is not feasible in the context of multi-epoch images acquired over large-scale sceneries. While improving the state-of-the-art, feature-based method remain limited to structured and man-made environments.

Beside the need of finding long-term feature matches between multi-epoch images, an additional challenge is introduced by natural environments for which feature-based methods provide sparse and ambiguous image matches which are not uniformly dispersed over the image space (Giordano et al., 2018), therefore introducing inconsistencies into the multi-view optimisation procedure.

This paper introduces a two-step strategy for establishing reliable multi-epoch point matches for aerial images acquired in natural environments. The first stage focuses on finding 2D patch matches in the image space. This step overcomes the main issue introduced by instable features accumulated over time, such as forests or high-vegetation areas. The second stage is performed in the 3D space and starts with the ground extraction procedure, followed by the 3D patch-based local alignment process. This allows to generate a set of 3D homologous points and the associated 3D rigid transformation lying between each 3D patch match.

### 3. OVERVIEW OF THE PROPOSED METHOD

The present research work introduces a fully automatic technique for georeferencing historical images onto nowadays im-

ages. The algorithm first generates their DSMs and orthoimages, out of images (Giordano et al., 2018). The obtained products are coarsely georeferenced, out of the available image metadata.

The proposed georeferencing technique presented in this paper takes as input the coarsely georeferenced 2D orthoimages, corresponding to both epochs, together their associated DSMs and outputs the 3D poses together with homologous 3D points. The proposed georeferencing pipeline is composed of two stages. In the first stage, the algorithm performs patch extraction and matching in the 2D image space. The second stage associates the DSM corresponding to each 2D homologous patch and performs patch-based alignment in the 3D space. Figure 1 illustrates the workflow of the proposed DSM georeferencing technique which is summarized through the following description.

**2D Patch extraction and matching.** The first stage proceeds by extracting a set of patches in the historical image. The patch matching procedure computes homologous patches between the recent and the historical orthoimages through the use of the Histogram of Gradients (HOG) algorithm (Dalal and Triggs, 2005). For each patch extracted in the historical image, the associated HOG is computed. Its homologous patch is searched in the recent image by maximizing the Zero Normalized Cross Correlation (ZNCC) score between the HOGs associated to the historical patch and the recent patch, extracted in a searching area centered around the recent patch. The procedure outputs a set of 2D homologous patches which allows to estimate a 2D global translational motion model for the considered set of patches.

**3D Patch-based DSMs alignment.** The historical and the recent DSMs are associated to the 2D homologous patches in order to generate the 3D point cloud corresponding to each patch. The global 3D translation motion lying between 3D homologous patches is computed via the barycenter approach. The estimated translation is applied to the historical 3D point cloud, allowing to obtain a coarse 3D global alignment of the historical DSM with respect to the recent DSM.

In order to eliminate 3D points coming from unstable features accumulated over time (such as forest), a ground filtering procedure is performed based on curvature computation (Rusu and Cousins, 2011). This operation allows to discriminate points in two classes: ground and above-ground 3D points. The resulted 3D point clouds belonging to the ground are further injected into the local alignment process which is applied to each pair of 3D patch matches. The fine alignment procedure includes pose estimation based on the Iteratively Closest Point (ICP) algorithm (Besl and McKay, 1992), barycenter compensation and outlier rejection. The final output of the algorithm is a set of homologous 3D patches and their associated 3D rigid transformations. The resulted 3D homologous points are designed to be injected into traditional bundle adjustment engines (Hartley and Zisserman, 2003), (Rupnik et al., 2017) for multi-view pose estimation and optimization, which may include the use of other homologous features.

### 4. 2D PATCH MATCHING VIA HISTOGRAM OF GRADIENTS (HOG)

Let  $I_{old}$  and  $I_{new}$  be the historical orthoimage and its associated recent orthoimage. The local patch matching procedure

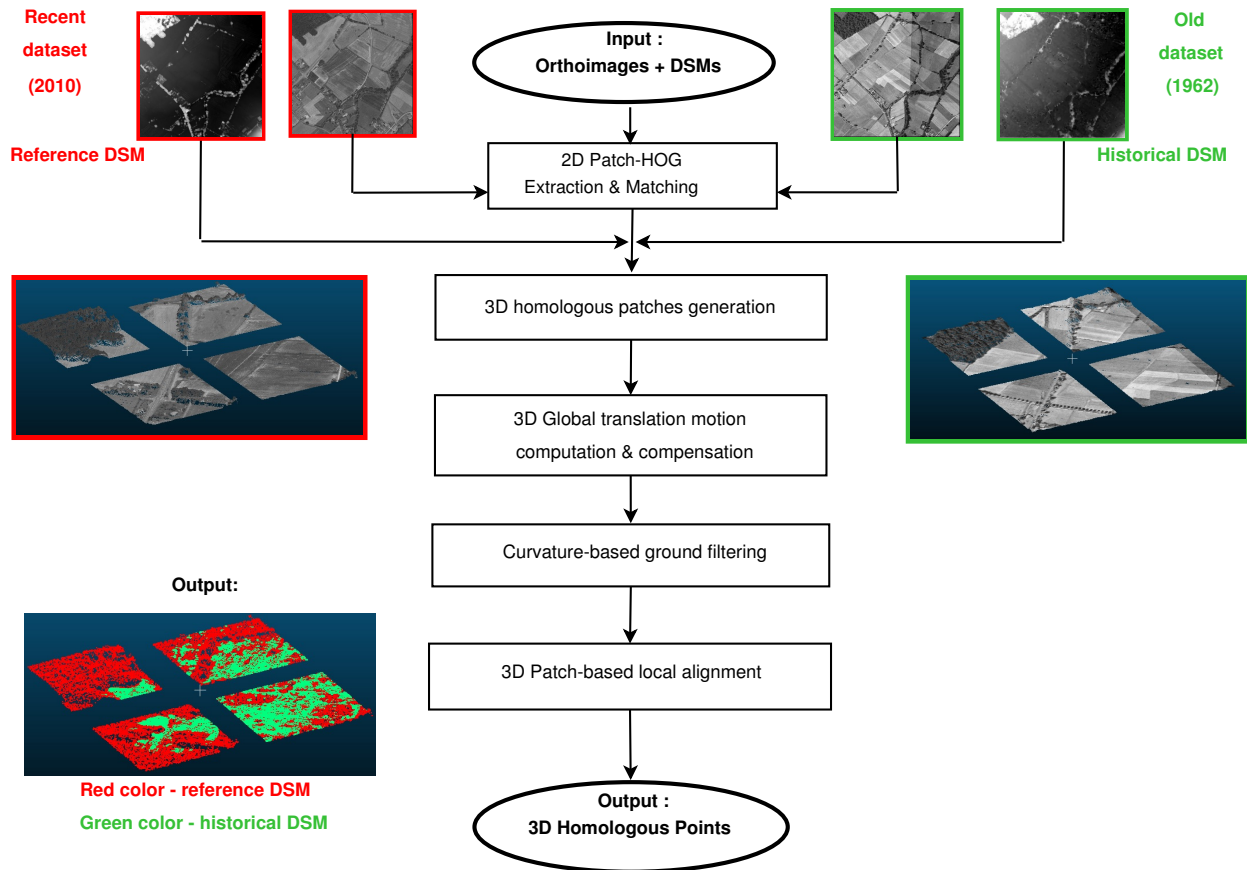


Figure 1. The workflow of the proposed automatic method for georeferencing archival aerial images acquired in unstructured environments. Example illustrated on the Fabas dataset (situated in France).

starts by extracting a list of  $N$  square patches in the  $I_{old}$  image,  $\mathbf{P}_{old} = \{\mathbf{P}_{old,i} \in I_{old}, i = 1, \dots, N\}$ , each patch being defined by a square neighborhood  $\mathcal{W}(\mathbf{u}_{old,i})$  centered around the image point  $\mathbf{u}_{old,i}$ . The size of each patch is given by the patch ray size, noted  $r$ , which provides  $(2r + 1) \times (2r + 1)$  pixels per patch. For each patch  $\mathbf{P}(\mathbf{u}_{old,i}) \in \mathbf{P}_{old}$  extracted in the image  $I_{old}$ , we search for its optimal patch match in the image  $I_{new}$  by exploring a windowed searching area, centered around  $\mathbf{u}_i$ , noted  $\mathbf{W}^{SA}(\mathbf{u}_i) \in I_{new}$ . The patch matching process is performed by maximizing the Zero Normalized Cross Correlation (ZNCC) score (noted  $\mathcal{Z}$ ) between the HOGs (Dalal and Triggs, 2005) associated to each patch.

Let  $\mathbf{P}_{new}^{SA} = \{\mathbf{P}_j \in I_{new}(\mathbf{u}_i), j = 1, \dots, M\}$  be the number of  $M$  patches found by exploring area  $\mathbf{W}^{SA}(\mathbf{u}_i) \in I_{new}$  with 1-pixel steps. Please note that the input orthoimages have the same spatial resolution. The best match  $\mathbf{P}_{new,i}$  is obtained by computing the similarity score,  $\mathcal{Z}$ , for each patch and by maximizing the score over the entire searching area space selected in the image  $I_{new}$ . This yields a list of 2D homologous patches  $\mathbf{P}(\mathbf{u}_{old,i}) \longleftrightarrow \mathbf{P}(\mathbf{u}_{new,i}), i = 1, \dots, N$  and the possibility to compute a 2D local translational motion for each match. Figure 2 illustrates the output of the patch matching procedure, displaying the corresponding index for each patch match.

### 5. 3D PATCH-BASED DIGITAL SURFACE MODELS ALIGNMENT

The historical and the recent DSMs are associated to the previously computed homologous 2D patches in order to generate

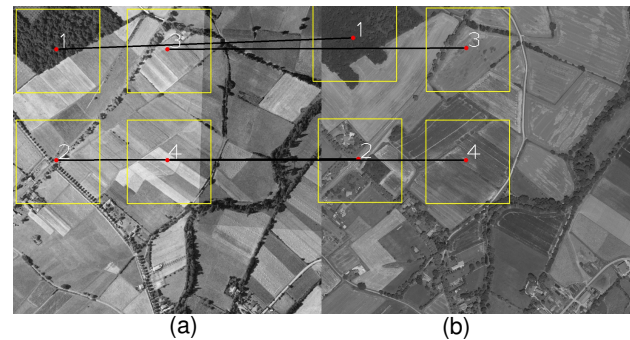


Figure 2. The output of the 2D patch matching procedure obtained for the dataset Fabas situated in France, with  $N = 4$  homologous patches: (a) historical orthoimage  $I_{old}$  (1962), (b) recent orthoimage  $I_{new}$  (2010); patch ray size  $r = 140$  pixels, encoding a surface area of  $252.9 \text{ m}^2$  per patch.

the 3D point clouds corresponding to each patch. Let  $\mathbf{P}_{3D,new} = \{\mathbf{p}_{new,i}(x, y, z), i = 1, \dots, N_{new}\}$  and  $\mathbf{P}_{3D,old} = \{\mathbf{p}_{old,i}(x, y, z), i = 1, \dots, N_{old}\}$  be the 3D point clouds obtained by computing the terrain projection. Figure 3 illustrates the 3D point cloud obtained by associating orthoimages and DSMs for recent and historical datasets. When analyzing Figures 3 (a) and 3 (b), it is possible to observe the initial positioning of the historical dataset  $\mathbf{P}_{3D,old}$  with respect to the reference 3D point cloud,  $\mathbf{P}_{3D,new}$ . It can be seen that the 2D coregistration is, at this stage already quite good, but the translation in the Z direction remains important. Figures 4 (a) and 4 (b) show the 3D homo-

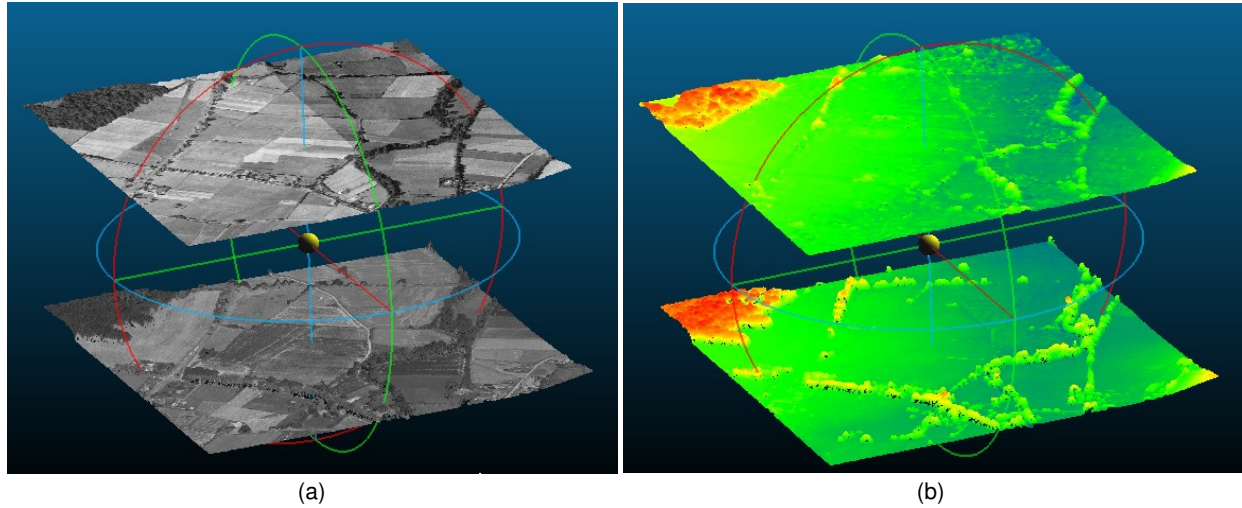


Figure 3. The 3D point clouds obtained by associating orthoimages and DSMs for the dataset Fabas. (a)  $\mathbf{P}_{3D,old}$  with  $N_{old} = 1Mpts$  situated in the upper side of the image,  $\mathbf{P}_{3D,new}$  with  $N_{new} = 1Mpts$  illustrated in the down side of the image; (b) input point clouds,  $\mathbf{P}_{3D,old}$ ,  $\mathbf{P}_{3D,new}$ , displayed with respect to elevation values.

logous patches obtained by associating the 3D coordinates to the 2D patch matches illustrated in Figure 2, generated by the 2D patch matching procedure described in Section 4.

### 5.1 3D Patch-based Global Coarse Alignment

In order to provide a coarse global alignment, the global 3D translation motion lying between the 3D homologous patches is computed via the barycenter approach by applying the following equation:

$$\bar{\mathbf{t}}_{3D} = \frac{\sum_{k=1}^n \mathbf{p}_{new,k} - \mathbf{p}_{old,k}}{n} \quad (1)$$

where  $n$  denotes the number of homologous 3D points belonging to all  $N$  patch matches, expressed as  $n = \sum_{i=1}^N \text{card}(\mathcal{P}_{old,i}^{3D}) = \sum_{i=1}^N \text{card}(\mathcal{P}_{new,i}^{3D})$ . The estimated translation,  $\bar{\mathbf{t}}_{3D}$ , is applied to the historical 3D point cloud  $\mathbf{p}_{old,k}$ ,  $k = 1, \dots, n$ , allowing to obtain a coarse 3D global alignment of the historical 3D point cloud with respect to the recent 3D point cloud, noted  $\hat{\mathbf{p}}_{new,k}$ ,  $k = 1, \dots, n$ .

$$\hat{\mathbf{p}}_{new,k} = \mathbf{p}_{old,k} + \bar{\mathbf{t}}_{3D} \quad (2)$$

In order to refine the obtained global alignment, a local 3D patch-based alignment procedure is designed to be further applied to each 3D patch match.

### 5.2 Curvature-based Ground Filtering

In order to eliminate 3D points coming from unstable features accumulated over time (such as forest) or from noisy DSMs, this paper proposes a ground filtering procedure based on curvature values computed for a given 3D point,  $\mathbf{p}_k$ ,  $k = 1, \dots, n$ , belonging to a 3D patch. This operation is performed on each 3D point cloud separately (historical and recent 3D patches) and allows to discriminate 3D points in two classes: ground and above-ground 3D points.

In the present research work, the curvature is estimated following the approach described in (Rusu and Cousins, 2011). For each 3D point  $\mathbf{p}(x, y, z)$ , a neighbourhood is selected around

the point. The method proceeds by estimating the plane normal and the curvature for each 3D point. The curvature, noted  $c(\mathbf{p})$ , is estimated through the use of the eigen values,  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , associated to the covariance matrix, being given by the following expression:

$$c(\mathbf{p}) = \frac{\lambda_1}{\sum_{l=1}^3 \lambda_l} \quad (3)$$

Let  $c(\mathbf{p}_k)$  be the curvature associated to a 3D point  $\mathbf{p}_k(x, y, z)$ . Being given the previously obtained 3D homologous patches, for each 3D point cloud  $\mathcal{P}_{old,i}^{3D}$  and  $\mathcal{P}_{new,i}^{3D}$ , the mean curvature is computed, noted  $\bar{c}_{old}$  and  $\bar{c}_{new}$ , respectively. Since 3D points belonging to the ground are described by low curvature values, the ground extraction is performed by rejecting 3D points with higher curvature values than the mean curvature,  $\bar{c}$ , computed for each 3D point cloud. The ground extraction is performed by associating a weight,  $w_k$ , to each 3D point  $\mathbf{p}_k(x, y, z)$  through the use of the following weighting procedure:

$$w_k = \begin{cases} w_k = 0 & \text{if } c(\mathbf{p}_k) > \delta \bar{c} \\ w_k = 1 & \text{otherwise.} \end{cases}$$

where, the parameter  $\delta$  denotes the threshold factor which allows to discriminate 3D points situated above the ground, corresponding to unstructured areas, such as forests and other high vegetation areas. Figure 4 (c) illustrates the result of the ground extraction procedure applied on the recent dataset Fabas,  $\mathcal{P}_{new,i}^{3D}$ ,  $i = 1, \dots, N$ . It is possible to observe that the 3D points representing the ground were correctly extracted. Moreover, 3D points situated above the ground (forest, high vegetation) were eliminated efficiently. The ground filtering procedure eliminates unstable features, therefore allowing to perform a consistent 3D local alignment for each patch in order to refine the global alignment.

Let  $\mathbf{P}_G = \{\mathbf{p}_k(x, y, z) | w_k = 1, k = 1, \dots, n_G\}$  be a 3D point cloud representing the ground, where  $n_G$  denotes the total number of 3D points. When analyzing the obtained result, it can be observed that 42.8% of the total amount of the 3D points represent the ground. The homologous 3D points assigned to the ground surface are further employed for registering locally each



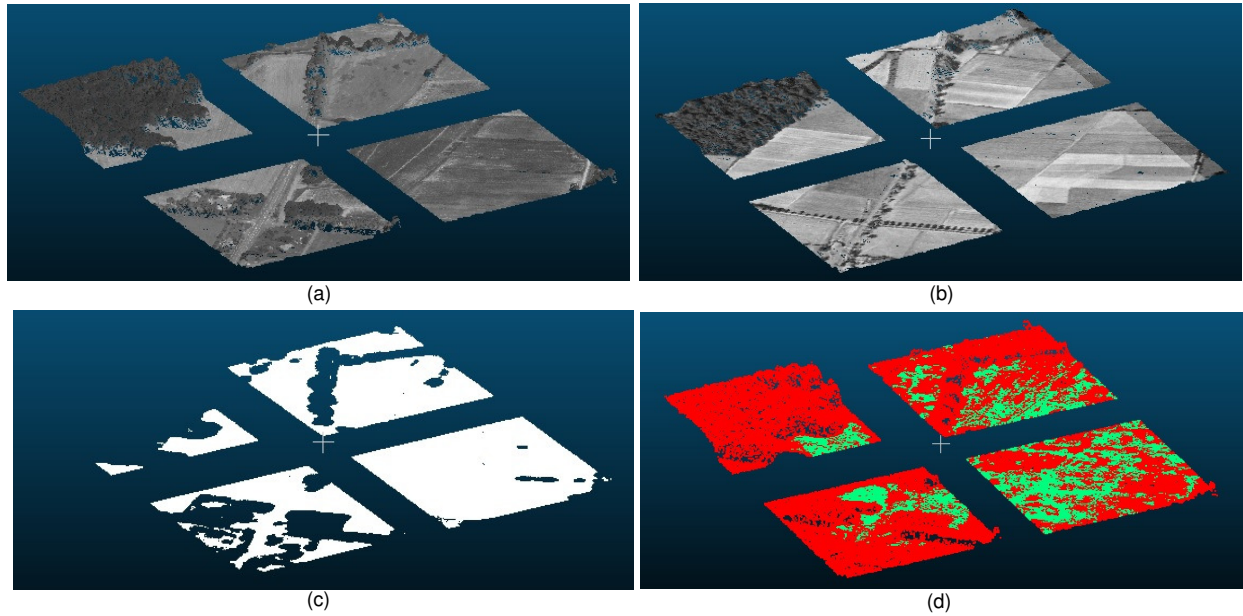


Figure 4. Experimental results obtained on the dataset Fabas. Figures (a), (b):  $N = 4$  homologous 3D patch matches, Figure (a)  $\mathcal{P}_{new,i}^{3D}, i = 1, \dots, N$ : reference (recent) 3D point cloud, (b)  $\mathcal{P}_{old,i}^{3D}, i = 1, \dots, N$ : historical 3D point cloud; total overlap:  $n = 285962$  points; Figure (c): the result of the curvature-based ground filtering procedure obtained by running the algorithm on the reference (recent) 3D point cloud illustrated in Figure(a),  $\mathcal{P}_{new,i}^{3D}, i = 1, \dots, N$ ; before ground extraction:  $n = 285962$  3D points; ground filtering parameter  $\delta = 0.3$ ; after ground extraction:  $n_G = 122517$  3D points representing the ground, i.e. 42.8% of the total amount of homologous points; Figure(d): superposed registered 3D point clouds: red color - reference (recent) DSM, green color - historical DSM.

3D patch match.

### 5.3 3D Patch-based Local Fine Alignment

Let us note with  $\mathbf{P}_{new,G}$  and  $\mathbf{P}_{old,G}$  the 3D point clouds representing the ground for the recent and the historical dataset, respectively. In order to refine the 3D global patch alignment, each 3D patch match is injected into the local alignment procedure which includes rigid pose estimation via the ICP (Besl and McKay, 1992) algorithm, barycenter compensation and outlier rejection.

Let  $\mathbf{P}_{new,G} = \{\mathcal{P}_{new,i}^{3D}, i = 1, \dots, N\}$  and  $\mathbf{P}_{old,G} = \{\mathcal{P}_{old,i}^{3D}, i = 1, \dots, N\}$  be the list of  $N$  homologous 3D patches representing the ground. For each 3D patch match,  $(\mathcal{P}_{new,i}^{3D} \longleftrightarrow \mathcal{P}_{old,i}^{3D})$ , the local 3D patch alignment procedure estimates iteratively the optimal transformation  $\hat{\mathbf{T}}_i = [\hat{\mathbf{R}}, \hat{\mathbf{t}}]$  which best aligns the 3D points belonging to each patch by minimizing the Residual Mean Square Error (RMSE), noted  $\mathbf{Q}_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]}$ , expressed in the following equation:

$$\mathbf{Q}_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]} = \frac{1}{n_G} \sum_{k=1}^{n_G} \|\mathbf{p}_{new,k} - \hat{\mathbf{R}}^{-1} \mathbf{p}_{old,k} - \hat{\mathbf{t}}\| \quad (4)$$

The accuracy of the obtained 3D matches is evaluated in terms of RMSE computed between the reference and the optimally aligned homologous 3D points. The local alignment procedure is performed individually for each patch  $i = 1, \dots, N$ . For simplicity reasons, let us now drop the subscript  $i$  through the following description. Let  $\hat{\mathbf{T}} = [\hat{\mathbf{R}}, \hat{\mathbf{t}}]$  be the optimal rigid transformation obtained by applying the ICP algorithm to each 3D patch match.

The second stage of the proposed 3D patch-based pose estimation procedure exploits the estimated rigid transformation  $\hat{\mathbf{T}} =$

$[\hat{\mathbf{R}}, \hat{\mathbf{t}}]$  to perform barycenter compensation between 3D points belonging to the overlapping area. The barycenter compensation is performed by registering the translation computed via the barycenter approach.

The third stage of the local alignment process is the outlier rejection step based on the residual errors expressed as:

$$r_k = \|\mathbf{p}_{new,k} - \hat{\mathbf{R}}^{-1} \mathbf{p}_{old,k} - \hat{\mathbf{t}}\| \quad (5)$$

3D points with a residual error  $r_k > \bar{r}$  (where  $\bar{r}$  denotes the mean residual error) are rejected by updating the weights  $w_k, k = 1, \dots, n_G$ , associated to homologous 3D points for each patch, as following:

$$\hat{w}_k = \begin{cases} \hat{w}_k = 0 & \text{if } r_k > \bar{r} \\ \hat{w}_k = 1 & \text{otherwise.} \end{cases}$$

This allows to obtain a set of inliers matches  $n_{in,i}$  for each 3D patch  $i = 1, \dots, N$ . The resulted matches are further injected in the ICP procedure for pose refinement. The goal is to minimize the criterion

$$\mathbf{Q}_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]} = \frac{1}{n_{in}} \sum_{k=1}^{n_{in}} r_k \quad (6)$$

where  $r_k$  are the weighted residual errors defined by:

$$r_k = \hat{w}_k \|\mathbf{p}_{new,k} - \hat{\mathbf{R}}^{-1} \mathbf{p}_{old,k} - \hat{\mathbf{t}}\| \quad (7)$$

The final output of the 3D patch-based local alignment is a set of 3D point matches  $(\mathbf{p}_{new,k} \longleftrightarrow \mathbf{p}_{old,k}), k = 1, \dots, n_{in}$ , belonging to each patch and their associated rigid transformations lying between each patch,  $\hat{\mathbf{T}}_i = [\hat{\mathbf{R}}, \hat{\mathbf{t}}]_i, i = 1, \dots, N$ , where  $N$  denotes the number of patch matches. Figure 5 illustrates the result of the iterative pose estimation process generated by the

ICP algorithm, for a single patch match,  $\mathcal{P}_{old,i}^{3D} \longleftrightarrow \mathcal{P}_{new,i}^{3D}$ , with  $i = 4$ . The accuracy of the algorithm is measured in terms of fitness score (Rusu and Cousins, 2011), noted  $f_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]}$ , defined as the mean of squared distances between each 3D point in the reference patch (recent DSM) to its closest point belonging to the target (historical DSM). The current ICP implementation (Rusu and Cousins, 2011) employs a 3D point cloud subsampling stage which allows to decrease the runtime of the estimation procedure, while ensuring pose correctness. The proposed patch-based pose estimation algorithm alternates an ICP estimation cycle (100 iterations) with barycenter correction and outliers rejection. By analyzing Figure 5, it can be observed that the algorithm starts the estimation process with the initial fitness score of  $f_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]} = 2.41(\text{m})$  and converges to the minimum fitness score  $f_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]} = 0.03(\text{m})$  in the 3rd cycle, during the last 100 iterations.

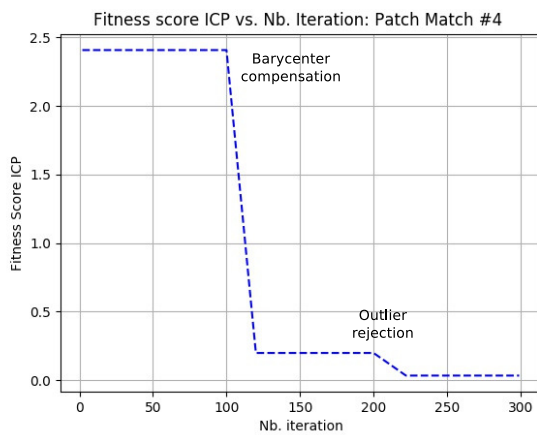


Figure 5. The output of the patch-based pose estimation procedure generated for the Fabas dataset: iterative pose estimation process obtained for the patch match #4,  $n_G = 54592$  3D points (representing the ground),  $n_{in} = 33668$  inliers.

Figure 6 illustrates the output of the local 3D patch-based alignment generated for the dataset Fabas. By visually inspecting Figures 6 (a) and 6 (b), it is possible to observe that the obtained patch matches are coherent in terms of radiometry values. In order to evaluate the accuracy of the generated matches in terms of geometry, the altimetric residual values are computed, noted DoD (Difference of DEMs), which are expressed in the following equation:

$$DoD_{[\hat{\mathbf{R}}, \hat{\mathbf{t}}]}(k) = \mathbf{p}_{new,k}(z) - \hat{\mathbf{p}}_{new,k}[\hat{\mathbf{R}}, \hat{\mathbf{t}}](z) \quad (8)$$

where,  $\hat{\mathbf{p}}_{new,k}[\hat{\mathbf{R}}, \hat{\mathbf{t}}] = \hat{\mathbf{R}}^{-1} \mathbf{p}_{old,k} - \hat{\mathbf{t}}$ ,  $k = 1, \dots, n_{in}$  denote the historical 3D points registered with respect to the reference (recent) dataset. Figure 6 (c) displays the altimetric residual values obtained for the dataset Fabas. A more detailed analysis of the obtained result is presented in the following section. In order to illustrate the final alignment result in terms of geometry, Figure 4(d) displays the initial reference 3D patches (recent DSM) superposed with the optimally aligned patches (historical DSM) generated via the 3D local alignment procedure. By visually inspecting the obtained result, it is possible to observe that the 3D local alignment procedure provides a geometrically correct alignment for which the qualitative evaluation is presented in the following section.

## 6. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

This section presents the experimental results together with the performance evaluation in terms of accuracy and computation time. The first part is dedicated to a description of the dataset, while the second part focuses on experimental results and their evaluation.

**Historical and Recent Datasets.** Experimental results are performed on the historical dataset acquired in 1962 over the Fabas area, situated in France, which is essentially composed of natural and rural areas. Land-cover changes are mainly due to forest evolution and to agricultural parcels clustering. The reference dataset is represented by recent orthoimages and DSMs, acquired in 2010 and extracted from the French national image databases. A detailed description of the dataset can be found in (Giordano et al., 2018). Experimental results performed on several datasets extracted from the Fabas study area allowed us to conclude on the effectiveness of the proposed approach. In order to provide a consistent analysis of the output generated by each composing stage of the algorithm, the experimental results illustrated throughout this paper exploit the same dataset extracted from the Fabas survey.

**Accuracy evaluation.** The first measure for qualifying the accuracy of the proposed approach is the Difference of Digital Elevation Models (DEMs), noted DoD, expressed in Equation 8. Figure 7 illustrates the DoDs (elevation residual values) obtained for each homologous patch  $\mathcal{P}_{new,i}^{3D} \longleftrightarrow [\mathcal{P}_{old,i}^{3D}, \mathbf{T}_i]$ ,  $i = 1, \dots, N$ , indicating that the proposed method generates accurate registration results in terms of elevation values. As illustrated previously in Figure 6 (c), from Figure 7 (b) it can be observed that the maximum altimetric residual values are defined over the interval  $[-1, 1]\text{m}$ , being obtained for the patch match #2. Moreover, when analyzing the corresponding 3D output for each patch match displayed in Figure 7, one can observe that high altimetric residual values correspond to the patch matches #1 and #2, from which a high amount of 3D points were eliminated through the curvature-based ground filtering procedure. For these patches, the remaining points describe a non-uniform area. In addition, it is also possible to observe that for the patch matches #3 and #4, which describe a dominant flat area, lower altimetric residual values were obtained.

The second evaluation measure employed for analyzing the accuracy of the local alignment procedure is the RMSE measure computed between the reference patch and the locally aligned patch. Table 1 illustrates the residual mean error obtained for each patch match, expressed in Equation 6. When analyzing the results, one can observe that the mean residual errors,  $\bar{r}_{in}$ , are consistent with their corresponding DoD values illustrated in Figure 7. More precisely, when comparing the accuracy obtained for matches #1 and #2 with the accuracy corresponding to matches #3 and #4, it can be observed that more accurate results are obtained for matches #3 and #4 which describe flat sceneries. Moreover, by analyzing the altimetric residual values illustrated in Figure 7, it is possible to conclude that the obtained results confirm the residual values,  $\bar{r}_{in}$ , obtained for matches #1 and #2, which represent non-uniform areas and for which higher values were obtained.

**Computation time.** The proposed algorithm is implemented in C/C++ and exploits OpenCV (Bradski, 2000) and PCL (Rusu and Cousins, 2011) libraries for 2D images and 3D point cloud processing, respectively. The algorithm runs on a desktop computer equipped with 32 Gb of RAM and an Intel processing

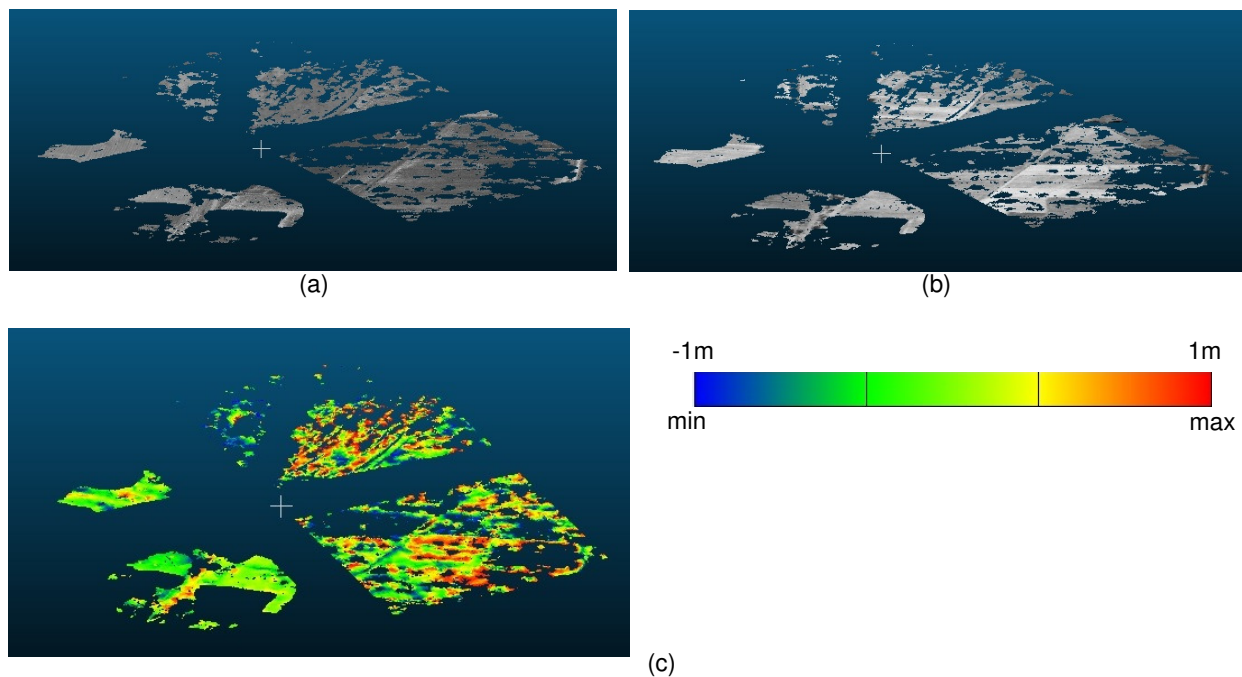


Figure 6. The output of the local 3D patch-based alignment procedure obtained for the dataset Fabas. (a) reference 3D point cloud, (b) historical 3D point cloud, (c) DoD (Difference of DEMs): elevation residual differences.

| Index Patch Match ( $i$ ) | $n_G$ | $\bar{r}$ (m) | $n_{in}$ | $\bar{r}_{in}$ (m) |
|---------------------------|-------|---------------|----------|--------------------|
| #1                        | 8444  | 25.77         | 5340     | 0.35               |
| #2                        | 25997 | 10.15         | 12621    | 0.84               |
| #3                        | 33484 | 5.65          | 18849    | 0.18               |
| #4                        | 54592 | 9.17          | 33668    | 0.15               |

Table 1. Accuracies obtained by running the patch-based local alignment procedure on the dataset Fabas, with  $N = 4$  homologous patches;  $n_G$ : the number of 3D points representing the ground (input),  $n_{in}$ : the number of inliers output by the pose estimation procedure.

unit running at 3.5 GHz. Table 2 summarizes the runtime of each processing stage composing the algorithm. When analyzing different processing stages, it can be observed that the most time consuming stage of the algorithm is the 3D fine alignment phase. Several optimization schemes are possible in order to improve the runtime. In terms of algorithm, 2D ICP techniques allow to cut down the combinatory when computing a first pose estimation which will decrease the searching area space for the final 3D pose refinement. When applied to the multi-view georeferencing context, parallelization schemes allow to decrease the computation time considerably.

| Processing stage              | runtime   |
|-------------------------------|-----------|
| 2D Patch extraction+matching  | 30 sec    |
| Coarse 3D-Global alignment    | 0.5 sec   |
| Curvature-based filtering     | 10 sec    |
| Local 3D-patch fine alignment | 11.69 min |

Table 2. Computation time obtained for the dataset Fabas, with  $N = 4$  homologous patches.

## 7. CONCLUSIONS AND FUTURE RESEARCH WORK

The presented workflow addresses the archival images georeferencing problem for natural (or rural) feature-less environments through the use of a two-step patch matching and alignment

procedure. The first stage outputs a coarse 3D global alignment which initializes the pose refinement stage performed locally for each patch. Experimental results and a qualitative evaluation performed on several epochs allowed us to conclude on the effectiveness of the presented method. Our findings demonstrate that the proposed georeferencing technique provides accurate results in presence of large periods of time separating historical from nowadays aerial images (up to 48 years time span). One of the main research perspective is directed toward the multi-view georeferencing stage via the bundle adjustment procedure, including the design of runtime optimization schemes. Beside the multi-epoch image georeferencing application, the proposed method is able to supply automatic marker-less pose estimation for several applications undertaken in rural and/or natural environments, overcoming therefore the difficulty of establishing homologous features.

## ACKNOWLEDGEMENTS

This work was supported by the French National Research Agency under the grant ANR-18-CE23-0025.

## REFERENCES

- Besl, P., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2).
- Bradski, G., 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Cléry, I., Pierrot-Deseilligny, M., Vallet, B., 2014. Automatic georeferencing of a heritage of old analog aerial photographs. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information*, II-3, 33–40.



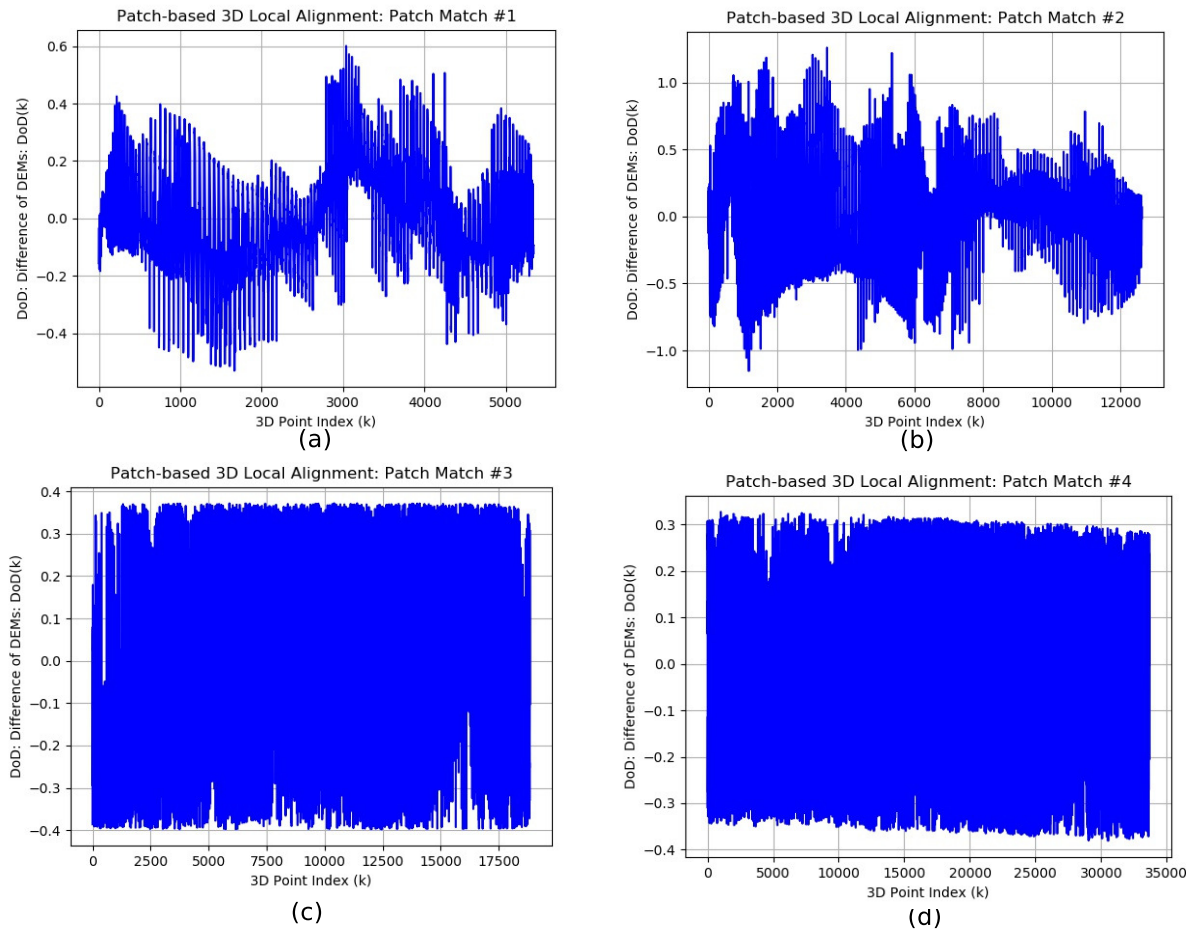


Figure 7. Experimental results obtained on the dataset Fabas: output generated by the local fine alignment procedure performed for each patch match, with  $N = 4$  homologous patches: (a) - (d): DoD (Difference of DEMs) altimetric residual values corresponding to each inlier index,  $k = 1, \dots, n_{in}$ , for each patch.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*.

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 Journal of Photogrammetry and Remote Sensing*, 146, 495–506.

Giordano, S., Le Bris, A., Mallet, C., 2018. Toward Automatic Georeferencing of Archival Aerial Photogrammetric Surveys. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2, 105–112.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. *Cambridge University Press*.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *Journal of Computer Vision*, 2(11), 91–110.

Lucas, A., Gayer, E., 2021. Decennial geomorphic transport from archived time series digital elevation models: a cookbook for tropical and alpine environments. *IEEE Geoscience and Remote Sensing Magazine*.

Moulon, P., Monasse, P., Perrot, R., Marlet, R., 2016. Open-MVG: Open multiple view geometry. *International Workshop*

*on Reproducible Research in Pattern Recognition*, Springer, 60–74.

Pinto, A. T., Gonçalves, J. A., Beja, P., Pradinho Honrado, J., 2019. From Archived Historical Aerial Imagery to Informative Orthophotos: A Framework for Retrieving the Past in Long-Term Socioecological Research. *Remote Sensing*, 11(1388), 47–56.

Rupnik, E., Daakir, M., Pierrot Deseilligny, M., 2017. MicMac – a free, open-source solution for photogrammetry. *Open Geospatial Data, Software and Standards*, 2(1).

Rusu, R. B., Cousins, S., 2011. 3D is here: Point Cloud Library (PCL). *IEEE International Conference on Robotics and Automation (ICRA)*.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. SuperGlue: Learning Feature Matching With Graph Neural Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, L., Rupnik, E., Pierrot-Deseilligny, M., 2021. Feature matching for multi-epoch historical aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 182, 176–189.