



**HAL**  
open science

# Guided feature matching for multi-epoch historical image blocks pose estimation

Lulin Zhang, Ewelina Rupnik, Marc Pierrot-Deseilligny

► **To cite this version:**

Lulin Zhang, Ewelina Rupnik, Marc Pierrot-Deseilligny. Guided feature matching for multi-epoch historical image blocks pose estimation. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., 2020. hal-02563454

**HAL Id: hal-02563454**

**<https://hal.science/hal-02563454>**

Submitted on 5 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# GUIDED FEATURE MATCHING FOR MULTI-EPOCH HISTORICAL IMAGE BLOCKS POSE ESTIMATION

Lulin Zhang<sup>1</sup>, Ewelina Rupnik<sup>1</sup>, Marc Pierrot-Deseilligny<sup>1</sup>

<sup>1</sup>LASTIG, Université Gustave Eiffel, ENSG, IGN, F-94160 Saint-Mandé, France - (Lulin.Zhang, Marc.Pierrot-Deseilligny)@ensg.eu, Ewelina.Rupnik@ign.fr

Commission II, WG II/1

**KEY WORDS:** Feature matching, Analogue aerial images, Multi-epoch, Pose estimation

## ABSTRACT:

Historical aerial imagery plays an important role in providing unique information about evolution of our landscapes. It possesses many positive qualities such as high spatial resolution, stereoscopic configuration and short time interval. Self-calibration remains a main bottleneck for achieving the intrinsic value of historical imagery, as it involves certain underdeveloped research points such as detecting inter-epoch tie-points. In this research, we present a novel algorithm to detecting inter-epoch tie-points in historical images which do not rely on any auxiliary data. Using SIFT-detected keypoints we perform matching across epochs by interchangeably estimating and imposing that points follow two mathematical models: at first a 2D spatial similarity, then a 3D spatial similarity. We import GCPs to quantitatively evaluate our results with Digital Elevation Models (DEM) of differences (abbreviated as DoD) in absolute reference frame, and compare the results of our method with other 2 methods that use either the traditional SIFT or few *virtual* GCPs. The experiments show that far more correct inter-epoch tie-points can be extracted with our guided technique. Qualitative and quantitative results are reported.

## 1. INTRODUCTION

Historical (i.e. analogue) aerial imagery plays an important role in providing unique information about evolution of our landscapes. The imagery had been acquired by many countries all over the world and can be traced back to the beginning of 20<sup>th</sup> century (Cowley, Stichelbaut, 2012). In many cases high spatial resolution stereoscopic acquisitions were performed (Giordano, Mallet, 2019), allowing for 3D restitution of territories. Coupled with high temporal sampling, the imagery becomes also suitable for long-term environmental studies. As of now, many historical images collections have been digitized (Giordano, Mallet, 2019; USGS, 2019; IGN, 2019) and a number of attempts have been made to process the images with the *state-of-the-art* automated methods developed in photogrammetry and computer vision community over the last twenty years (Ayoub et al., 2009; Bakker, Lane, 2017; Chen, Tseng, 2016; Ellis et al., 2006; Feurer, Vinatier, 2018).

Apart from imagery itself, sometimes corresponding metadata is archived too. For instance, camera focal length and physical size of the images are commonly found, whereas flight plans, camera calibration certificates or orientations are rarely available. Because the images were scanned, and prior to scanning they may have been kept in unsuitable conditions, a self-calibration is always necessary.

Historical images are self-calibrated in a bundle block adjustment (BBA) routine, using tie-points and Ground Control Points (GCPs) as input observations. When dealing with single-epoch datasets, tie-points are efficiently recovered with modern extraction algorithms, such as SIFT (Lowe, 2004), while GCPs originate from (i) field surveys (Micheletti et al., 2015; Walstra et al., 2004; Cardenal et al., 2006), (ii) existing orthophotos and DEM (Nurminen et al., 2015; Ellis et al., 2006; Fox, Cziferzky, 2008), (iii) recent satellite images (Ellis et al., 2006; Ford,

2013). In many scenarios none of the auxiliary data is at disposal. When processing multi-epoch image blocks, ideally a joint self-calibrating bundle adjustment is carried out, including inter-epoch image observations. However, extracting inter-epoch tie-points remains challenging because the scene changes over time, and images are often of low radiometric quality violating the *brightness constancy* rule that all modern extraction algorithms rely upon.

In this work we propose a new approach to detecting inter-epoch tie-points. Using SIFT-detected key-points we perform matching across epochs by interchangeably estimating and imposing that points follow two mathematical models: at first a 2D spatial similarity, then a 3D spatial similarity. We do not rely on any auxiliary data, making it a generic approach. We adopt GCPs only to quantitatively evaluate our results. In the following we briefly describe existing feature matching methods, as well as recent approaches to georeferencing of historical images. In Section 3 we introduce our methodology, and in Section 4 the experiments as well as results are given. The terms *features* and *tie-points* are considered synonymous and used interchangeably throughout this publication.

## 2. RELATED WORK

### 2.1 Feature Extraction

Feature extraction refers to finding discriminative structure in images such as corner, blob, edge and so on, followed by a matching step. According to whether machine learning techniques are applied, the image features can be categorized as hand-crafted or learned.

**2.1.1 Hand-crafted Features** In the early stage, Moravec detects corner feature by measuring the sum-of-squared-differences (SSD) by applying a small shift in a number of

directions to the patch around a candidate feature (Moravec, 1980). Based on this, Harris computes an approximation to the second derivative of the SSD with respect to the shift (Harris, Stephens, 1988). Since both Moravec and Harris are sensitive to changes in image scale, algorithms invariant to scale and affine transformations based on Harris are presented (Mikolajczyk, Schmid, 2004). Other than corner feature, SIFT (Scale-invariant feature transform) (Lowe, 2004) detects blob feature in scale-space, which is an entire pipeline including detection and description. It uses a difference-of-Gaussian function to identify potential feature points that are invariant to scale and orientation. SIFT is a milestone among hand-crafted features, and comparable with machine learning alternatives. RootSIFT (Arandjelović, Zisserman, 2012) uses a square root (Hellinger) kernel instead of the standard Euclidean distance to measure the similarity between SIFT descriptors, which leads to a dramatic performance boost. Similar to SIFT, SURF (Bay et al., 2006) resorts to integral images and Haar filters to extract blob feature in a computationally efficient way. DAISY (Tola et al., 2009) is a local image descriptor, which uses convolutions of gradients in specific directions with several Gaussian filters to make it very efficient to extract dense descriptors. KAZE (Alcantarilla et al., 2012) is an algorithm that detects and describes multi-scale 2D feature in nonlinear scale spaces. AKAZE (Alcantarilla et al., 2013) is an accelerated version based on KAZE.

**2.1.2 Learned Features** With the rise of machine learning, learned features have shown their feasibility in the image matching problem when enough ground truth data is available. FAST (Rosten, Drummond, 2006) uses decision tree to speed up the process of finding corner feature. LIFT (Learned Invariant Feature Transform) (Yi et al., 2016) is a deep network architecture that implements a full pipeline including detection, orientation estimation and feature description. It is based on the previous work TILDE (Verdie et al., 2015), the method of (Moo Yi et al., 2016) and DeepDesc (Simo-Serra et al., 2015). SuperPoint (DeTone et al., 2018) is a self-supervised, fully-convolutional model that operates on full-sized images and jointly computes pixel-level feature point locations and associated descriptors in one forward pass. LF-Net (Ono et al., 2018) is a deep architecture that embeds the entire feature extraction pipeline, and can be trained end-to-end with just a collection of images. Finally, D2-Net (Dusmanu et al., 2019) is a single convolutional neural network that works as a feature detector and a dense feature descriptor simultaneously.

Even though the learned features demonstrate better performance when compared to hand-crafted features on certain benchmark, it does not necessarily imply a better performance in terms of subsequent processing steps. For example, in the context of Structure from Motion (SfM), finding additional correspondences for image pairs where SIFT already provides enough matches does not necessarily results in more accurate or complete reconstructions (Schonberger et al., 2017).

## 2.2 Geo-referencing of Historical Images

Pose estimation describes the intrinsic and extrinsic parameters of an image and is classically solved with the SfM algorithms (Snavely et al., 2006; Pierrot-Deseilligny, Cléry, 2012; Schonberger, Frahm, 2016). If the image pose should be known in a reference coordinate system, i.e. be georeferenced, a 7-parameter transformation followed by a *posteriori* bundle block adjustment must be carried out.

Unlike in modern images where the image coordinate system overlaps with the planimetric camera coordinate system, in historical images the overlap is not maintained due to the scanning procedure. To account for this, an additional 2D transformation is estimated in the course of the processing (McGlone, 2013). (Giordano et al., 2018) demonstrates the importance of the self-calibration of historical images and its impact on 3D accuracy. Poorly modelled intrinsic parameters result in the known *dome*-like deformations that the authors eliminate with the help of automatic GCPs from exiting orthophotos.

Many cases reported in the literature calculate the image poses with SfM using features matched exclusively within the same epoch (Nurminen et al., 2015; Cardenal et al., 2006; Fox, Cziferszky, 2008; Walstra et al., 2004). In multi-epoch scenarios, the individual epochs should be defined in a common frame, be it the frame of a reference epoch or an absolute reference frame (i.e., a projection coordinate frame). Control points derived from recent orthophotos and DEM (Nurminen et al., 2015; Ellis et al., 2006; Fox, Cziferszky, 2008) or GPS survey (Micheletti et al., 2015; Walstra et al., 2004; Cardenal et al., 2006) serve to transform the individual epochs to the common frame. Alternatively, a coarse flight plan may provide an approximate co-registration (Giordano et al., 2018).

To increase relative accuracy between several epochs, (Cardenal et al., 2006; Korpela, 2006; Micheletti et al., 2015) manually insert inter-epoch tie-points. (Giordano et al., 2018) extracts tie-points between analogue images and recent images based on the method of (Aubry et al., 2014). (Feurer, Vinatier, 2018) joins multi-epoch images in a single SfM block based on SIFT-like algorithm (Lowe, 2004; Semyonov, 2011) by making the assumption that a sufficient number of feature points remain invariant across each time period. Identifying permanent points over a long time-span has not reached a maturity and research in this domain remains insignificant. Most of the mentioned approaches rely on manual effort, auxiliary input or limiting hypothesis.

## 3. METHODOLOGY

Figure 1 exhibits the workflow of our algorithm proposed in this paper. It can be divided as two main parts: processing intra-epoch, and processing inter-epoch. The images are resampled to the geometry of the fiducial marks prior to processing. For the sake of simplicity we only exhibit the processing flow of two epochs, however, it can be easily extended to more epochs. In the remaining of this section, both intra-epoch and inter-epoch processing will be explained in detail.

### 3.1 Intra-epoch Processing

We process each single-epoch data individually as follow:

1. Intra-epoch tie-points are effectively obtained by extracting and matching feature points based on SIFT algorithm (Lowe, 2004);
2. Interior and relative orientations are performed with standard SfM algorithm implemented in MicMac software (Pierrot-Deseilligny et al., 2015);
3. Individual depth maps are obtained with semi-global dense matching available in MicMac (Pierrot-Deseilligny, Paparoditis, 2006).

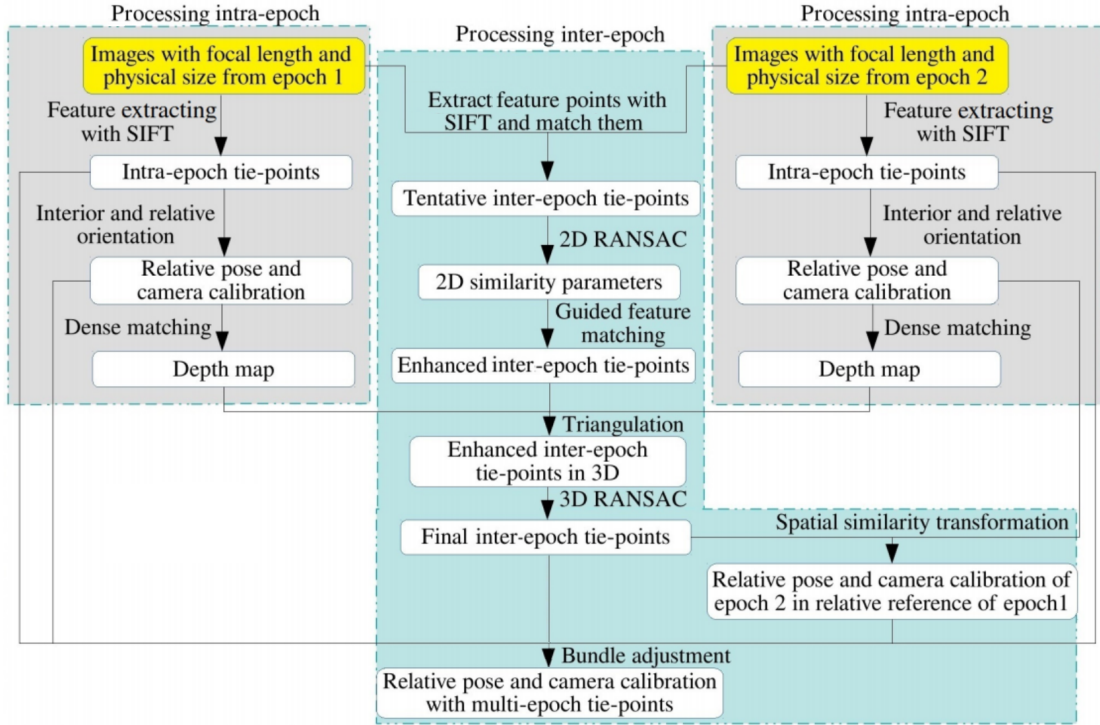


Figure 1. Our processing workflow. The inputs are marked in yellow, the blue area corresponds to the *inter-epoch* processing, and the gray is the *intra-epoch* processing. The input images had been resampled to the geometry of the fiducial marks prior to the processing.

### 3.2 Inter-epoch Processing

Using images of all epochs, we now extract inter-epoch tie-points. We propose the following 3-step guided matching strategy to increase the robustness of the automated tie-points:

**Get tentative inter-epoch tie-points:** extracted with SIFT and matched based on mutual nearest neighbor (i.e., Euclidean distance) between respective descriptors. To increase computational efficiency and discard noise, features are extracted on images down-sampled with a factor of  $1/6$ . No ratio test (Lowe, 2004) was applied since very few tie-points were otherwise retained. This obviously results in a higher number of false matches but as these are only our tentative matches, we accept that. When choosing which feature extraction algorithm to employ, we hesitated between SIFT as the *state-of-the-art* algorithm for hand-crafted features, and D2-Net – the *state-of-the-art* in learned features. We chose SIFT based on a conducted experiment (see Section 4.3) where it outperformed D2-Net.

**Get enhanced inter-epoch tie-points:** out of the set of all tentative matches we now want to identify and eliminate the false matches. We introduce a 2D similarity transformation (cf., Equation 1) to describe the relation between feature points in a pair of inter-epoch images. This assumption is satisfactory for nadir images, which is our primary focus. The calculation is embedded in a RANSAC framework, and at each iteration the 2D similarity parameters ( $\lambda, \theta, \Delta x, \Delta y$ ) are computed from minimal points in inter-epoch images. All points falling within  $N$  pixels from the prediction are considered inliers. In our experiment,  $N$  was set to 5% of the image diagonal, and RANSAC iterations to 1000.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \lambda \cdot \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \quad (1)$$

**Get final inter-epoch tie-points:** in the preceding step we got rid of a substantial number of outliers, however, we believe not all outliers could be identified. To obtain the final tie-points we introduce another filter, this time based on 3D similarity transformation:

$$\mathbf{P}' = \lambda \cdot \mathbf{R} \cdot \mathbf{P} + \mathbf{T} \quad (2)$$

Where the  $\mathbf{P}'$  are the 3D coordinates of a feature point in the reference frame of the intra-epoch reconstruction (e.g.,  $epoch_1$ ) and  $\mathbf{P}$  are its respective coordinates in  $epoch_2$ . The 3D coordinates are retrieved from the individual depth maps computed within the intra-epoch processing step. We now carry out the estimation of the scale  $\lambda$ , the translation vector  $\mathbf{T}$ , and the rotation matrix  $\mathbf{R}$  in a RANSAC routine. We set the number of RANSAC iterations to 1000, and consider tie-points within 10m of its predicted position as inliers.

### 3.3 Combined Processing

With the help of the 3D similarity transformation we can move our multi-epoch acquisitions to a common reference frame. Finally, we perform a BBA to refine all the camera poses and camera calibrations. The intra- and inter-epoch features are the observations in our BBA. To analyze the results in a metric scale, the BBA is followed by another spatial similarity transformation that brings the arbitrary reference frame tied to a selected acquisition to a global reference frame. In our experiments, we derived this transformation thanks to several manually identified points in a recent orthophoto and a digital surface model over the area. If precise poses for one of the epochs were known, we fixed their parameters during the BBA and skipped the final spatial similarity transformation.



## 4. EXPERIMENTS

### 4.1 Test Site and Data

The test site is a  $420 \text{ km}^2$  rectangular area located in Pezenas in the Occitanie region in southern France. The area is mainly covered with vegetation, and several sparsely populated urban zones. We have at our disposal three sets of images acquired in June 1971, in June 1981 and in 2015 (cf., Table 1). The 2015 images were acquired with the IGN’s digital metric camera (Souchon et al., 2010). The area exhibited changes in scene appearance due to land-use changes in the 44-year period. Our ground truth (GT) is a reference 0.25m resolution DSM generated by the French Mapping Agency (IGN) in 2015.

	epoch 1971	epoch 1981	epoch 2015	
F [pix]	7589.06	7606.89	9967.5	9204.5
Size [mm]	230,230	230,230	47,35	50,36
GSD [m]	0.32	0.59	0.46	0.50
F. o.	60%	60%	60%	60%
S. o.	20%	20%	50%	50%
H [m]	2400	4500	4600	4600
Nb	57	27	308	74

Table 1. Details about 1971, 1981 and 2015 acquisitions. The 2015 acquisition is obtained with two sets of camera. GSD is the ground sampling distance, F.o. and S.o. are forward and side overlap, H is the flying height, Nb is the number of images.

### 4.2 Evaluation Method

As presented in Figure 1, we perform the BBA in a relative coordinate system embedded in a selected epoch. Subsequently, for evaluation purposes we transform the relative results to a reference frame of our GT with the help of several manually measured GCPs. We evaluate the results by computing DSM differences (DoD), between the GT DSM and a DSM resulting from our computations in respective years (cf. Figure 5). We compare *Our* method to two other approaches:

- $SIFT_{intra-inter}$ : intra- and inter-epoch tie-points are extracted and matched with SIFT;
- $SIFT_{intra}VGCP_{inter}$ : intra-epoch tie-points are extracted with SIFT but no inter-epoch tie-points are present; the individual epochs are co-registered with the help of *virtual* GCPs (i.e., points triangulated in one epoch and used as GCP in another).

### 4.3 Result

**Comparison between D2-Net and SIFT.** To decide which feature extracting algorithm to use in our guided matching, we performed an experiment based on D2-Net and SIFT. D2-Net is employed with the off-the-shelf trained model provided by the authors, and we use their multiscale detection version to achieve better performance with scale changes. We always apply the ratio test for SIFT, and for D2-Net we test scenarios with and without the ratio test. We used an inter-epoch image pair from 1971 and 1981. The images have a scale difference of factor 1.8 and are rotated by  $180^\circ$  with respect to each other. D2-Net failed at finding corresponding features. This is probably due to its sensitivity to severe rotation. Meanwhile, SIFT correctly identified several matching points as shown in Figure 2.

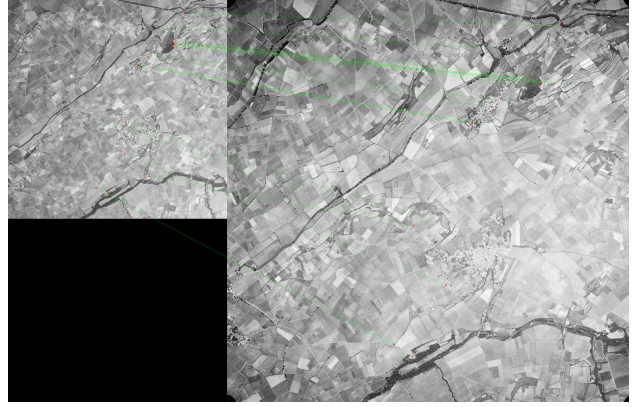


Figure 2. SIFT matching result on an inter-epoch image pair with rotation and scale change.

To be able to compare the two feature extractors we removed the scale difference and rotated one image to the orientation of the second image. The matching results are presented in Figure 3. D2-Net detects and matches more points, however, their precision is inferior to that of SIFT (see Figure 3 bottom). As mentioned in (Dusmanu et al., 2019), D2-Net is a detection based on higher-level information which inherently leads to more robust but less accurate keypoints. At the same time, D2-Net features are also detected at lower resolution of the CNN features (feature maps have a resolution of one fourth of the input resolution). SIFT was the preferred algorithm because even though it finds few points, they are much more reliable than D2-Net, and they are robust to image rotation.

**Inter epoch matching and BBA.** In the BBA, all poses corresponding to 1971 and 1981 were considered as free parameters. Since the total number of intra-epoch tie-points is by far larger than the inter-epoch tie-points (cf. Table 2), we set the relative weight to balance the effect. We adopted the Fraser model (Fraser, 1997) to calibrate the cameras, and allowed image-dependent affine parameters. The remaining parameters were shared among all images. The poses of the 2015 acquisition were accurately known *a-priori* thereby treated as fixed during the combined BBA.

**Comparison between three methods.** Figure 4 illustrates the inter-epoch tie-points extracted with *Ours* and  $SIFT_{intra-inter}$  between 1971 and 1981, 1971 and 2015 as well as 1981 and 2015. Our guided matching strategy detected far more points than the classical SIFT pipeline in  $SIFT_{intra-inter}$  (cf. Table 2), especially for the most challenging scenario (i.e. the one with the longest time gap between 1971 and 2015). Figure 5 shows the computed DoDs and Table 3 reports the average and standard deviations computed on Z-coordinate difference. A *dome* artifact is present in DoDs corresponding to methods  $SIFT_{intra}VGCP_{inter}$  and  $SIFT_{intra-inter}$ . This kind of systematic error is known to originate from poorly modelled camera internal parameters (Giordano et al., 2018). In *Ours*, given the reasonable dense inter-epoch features and the calibrated 2015 acquisition, we are able to effectively remove this discrepancy. Even though our proposed method performs best of all, we still observe some systematic errors, e.g. the stripe-like pattern that coincides with the acquisition geometry, and the increasing differences towards the edge of the image block in 1971, where few inter-epoch tie-points were found.

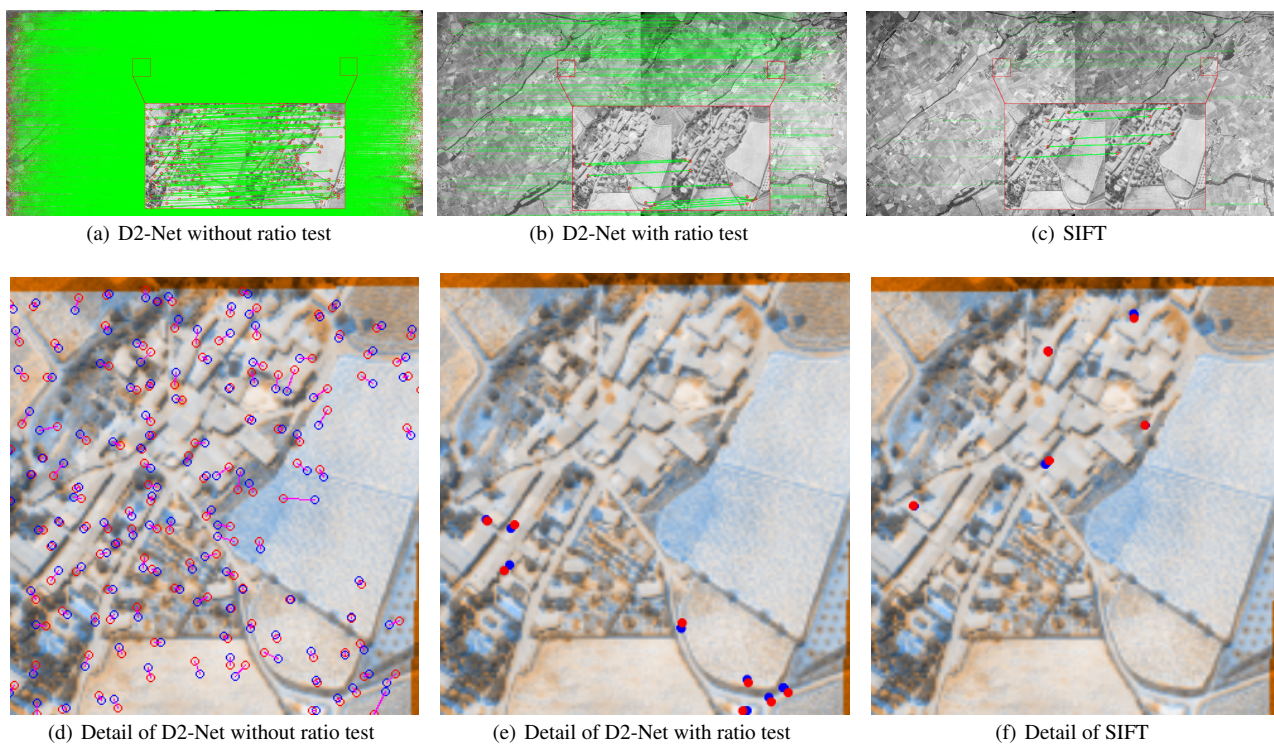


Figure 3. Matching result on inter-epoch image pair without scale change and rotation. Top: a global view, bottom: zoom over an area with the images transformed to a common geometry. Notice the offsets between the D2-Net corresponding points.

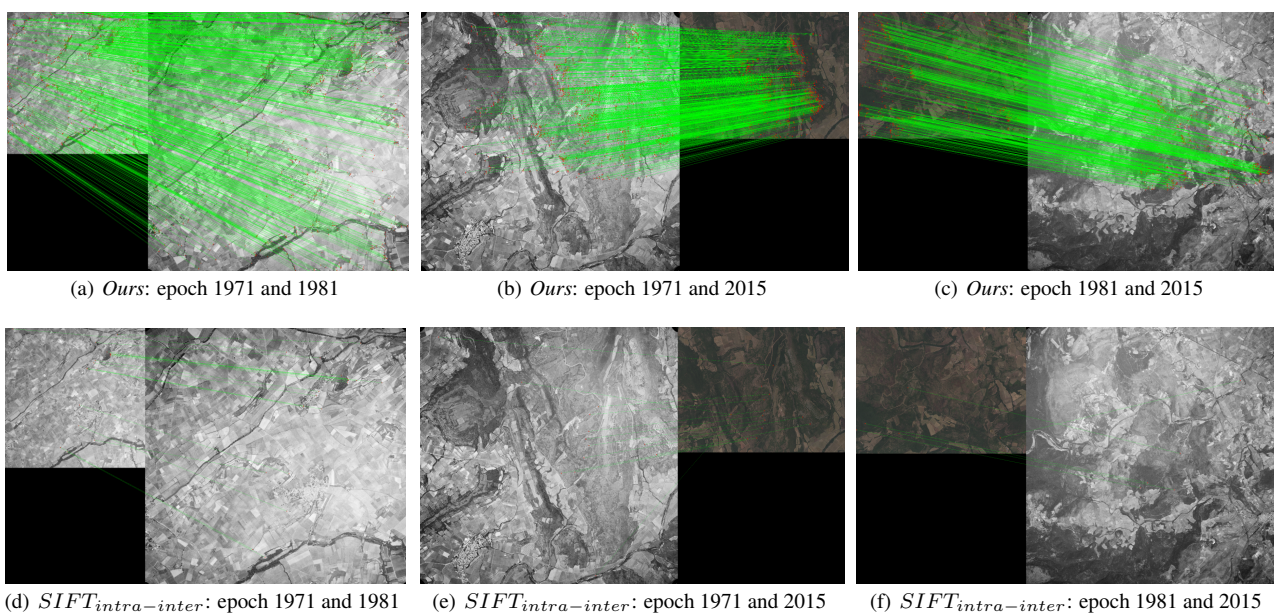


Figure 4. Inter-epoch tie-points.



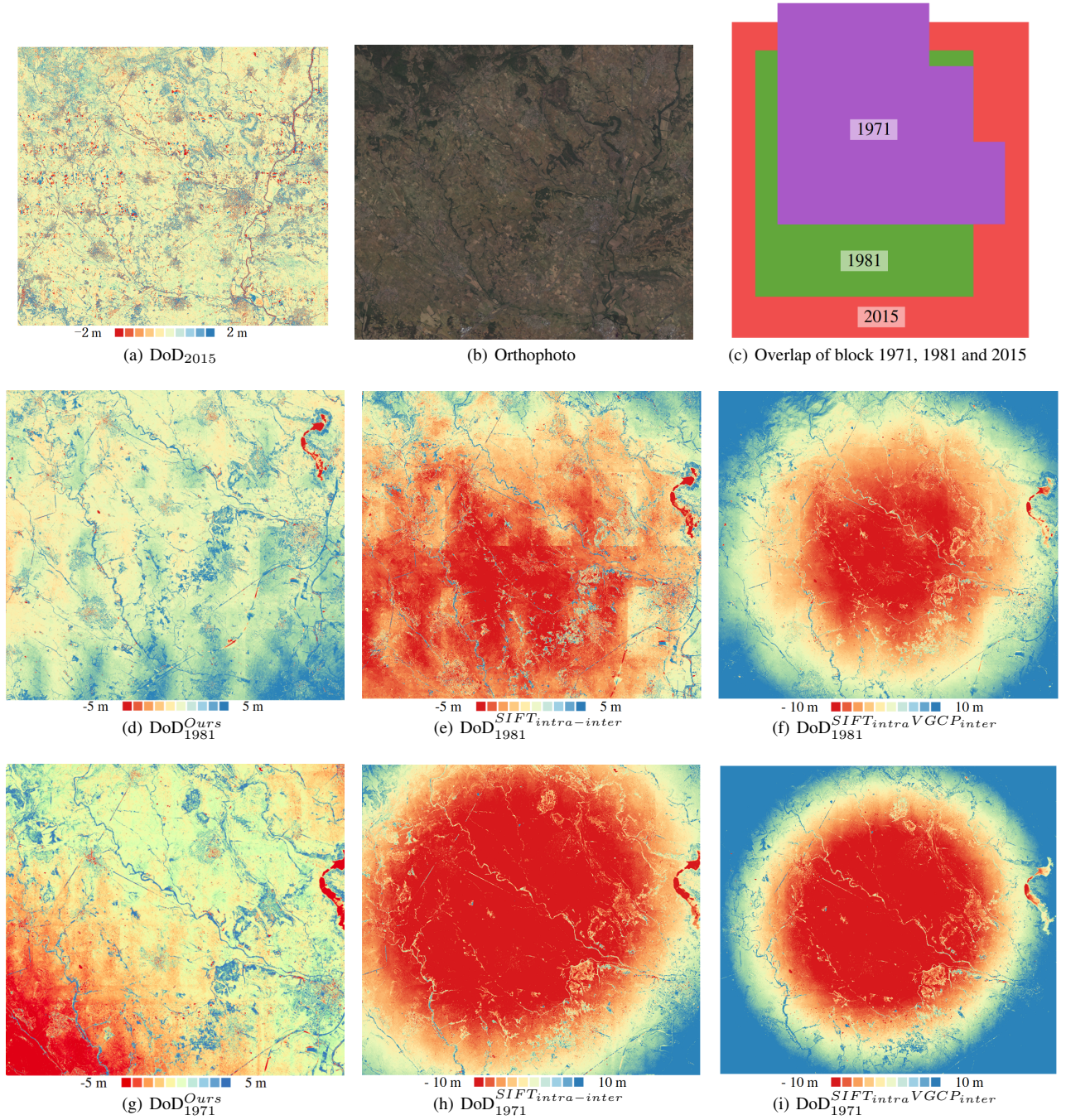


Figure 5. (a) DoD between GT DSM and 2015 DSM, (b) Orthophoto of the corresponding area, (c) Overlap of block 1971, 1981 and 2015, (d-f) DoD between GT DSM and 1981 DSM with *Ours*,  $SIFT_{intra-inter}$  and  $SIFT_{intraVGCP_{inter}}$ , (g-i) DoD between GT DSM and 1971 DSM with *Ours*,  $SIFT_{intra-inter}$  and  $SIFT_{intraVGCP_{inter}}$ .

	epoch 1971	epoch 1981	epoch 2015
epoch 1971	<b>448165</b> 448165	<b>160031</b> 9075	<b>28759</b> 341
epoch 1981	/	<b>631290</b> 631290	<b>61546</b> 2031
epoch 2015	/	/	<b>18363596</b> 18363596

Table 2. Number of extracted tie-points with *Ours* (in bold), and with *SIFT<sub>intra-inter</sub>*.

	$\mu$ [m]	$\sigma$ [m]	$ \mu $ [m]
DoD <sub>2015</sub>	0.36	1.56	0.81
DoD <sub>1981</sub> <sup><i>Ours</i></sup>	1.10	2.36	<b>1.45</b>
DoD <sub>1981</sub> <sup><i>SIFT<sub>intra-inter</sub></i></sup>	-1.10	3.45	2.94
DoD <sub>1981</sub> <sup><i>SIFT<sub>intra</sub>VGCP<sub>inter</sub></i></sup>	-1.53	8.00	6.98
DoD <sub>1971</sub> <sup><i>Ours</i></sup>	0.45	2.75	<b>1.61</b>
DoD <sub>1971</sub> <sup><i>SIFT<sub>intra-inter</sub></i></sup>	-7.55	7.67	9.62
DoD <sub>1971</sub> <sup><i>SIFT<sub>intra</sub>VGCP<sub>inter</sub></i></sup>	0.30	15.36	12.51

Table 3. Average  $\mu$ , standard deviation  $\sigma$ , and absolute average  $|\mu|$  of the Z-coordinate in the 7 DoDs in Figure 5.

## 5. CONCLUSION

A new approach to detecting inter-epoch tie-points in historical images has been proposed. Our validation datasets consisted of three epochs: 1971, 1981 and 2015. The proposed method outperformed the classical SIFT and D2-Net extraction algorithms in detecting many more and reliable points. This resulted in better pose estimation outcomes. The approach is generic as no auxiliary data is required, and it is independent of limiting initial assumptions. We also performed an experiment in which we show that (i) SIFT often provides less tie-points than D2-Net but they are more precise; (ii) D2-Net is sensitive to image rotation.

Future work might concentrate on further improving robustness of the proposed method, in order to handle more challenging conditions, such as larger temporal gap. This could for example be done by automatically searching for a more suitable scale factor to downsample the input image to perform the tentative matching, or resorting to line features when necessary.

## ACKNOWLEDGEMENTS

This work was supported by ANR project DISRUPT (ANR-18-CE31-0012-0). We thank Denis Feurer and Fabrice Vinatier for their contribution to the interior calibration of the pezenas dataset. We would also like to thank Sebastien Giordano for the ground truth DSM and Benjamin Ferrand for the 2015 acquisition.

## REFERENCES

Alcantarilla, P. F., Bartoli, A., Davison, A. J., 2012. Kaze features. *European Conference on Computer Vision*, 214–227.

Alcantarilla, P. F., Nuevo, J., Bartoli, A., 2013. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *British Machine Vision Conf. (BMVC)*.

Arandjelović, R., Zisserman, A., 2012. Three things everyone should know to improve object retrieval. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2911–2918.

Aubry, M., Russell, B. C., Sivic, J., 2014. Painting-to-3D model alignment via discriminative visual elements. *ACM Transactions on Graphics (ToG)*, 33(2), 1–14.

Ayoub, F., Leprince, S., Avouac, J.-P., 2009. Co-registration and correlation of aerial photographs for ground deformation measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), 551–560.

Bakker, M., Lane, S. N., 2017. Archival photogrammetric analysis of river–floodplain systems using Structure from Motion (SfM) methods. *Earth Surface Processes and Landforms*, 42(8), 1274–1286.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features. *European conference on computer vision*, 404–417.

Cardenal, J., Delgado, J., Mata, E., González, A., Olague, I., 2006. Use of historical flight for landslide monitoring. *Proceedings of the Spatial Accuracy*, 129–138.

Chen, H., Tseng, Y., 2016. Study of automatic image rectification and registration of scanned historical aerial photographs. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B8, 1229–1236.

Cowley, D. C., Stichelbaut, B. B., 2012. Historic aerial photographic archives for European archaeology. *European Journal of Archaeology*, 15(2), 217–236.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 224–236.

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. *arXiv preprint arXiv:1905.03561*.

Ellis, E. C., Wang, H., Xiao, H. S., Peng, K., Liu, X. P., Li, S. C., Ouyang, H., Cheng, X., Yang, L. Z., 2006. Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery. *Remote Sensing of Environment*, 100(4), 457–473.

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.

Ford, M., 2013. Shoreline changes interpreted from multi-temporal aerial photographs and high resolution satellite images: Wotje Atoll, Marshall Islands. *Remote Sensing of Environment*, 135, 130–140.

Fox, A. J., Czifszky, A., 2008. Unlocking the time capsule of historic aerial photography to measure changes in Antarctic Peninsula glaciers. *The Photogrammetric Record*, 23(121), 51–68.

Fraser, C. S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote sensing*, 52(4), 149–159.

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.

- Giordano, S., Mallet, C., 2019. Archiving and geoprocessing of historical aerial images: current status in europe, official publication no 70. *European Spatial Data Research*.
- Harris, C., Stephens, M., 1988. A combined corner and edge detector. *In Proc. of Fourth Alvey Vision Conference*, 147–151.
- IGN, 2019. remonterletemps. <https://remonterletemps.ign.fr/>.
- Korpela, I., 2006. Geometrically accurate time series of archived aerial images and airborne lidar data in a forest environment. *Silva Fennica*, 40(1), 109.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110.
- McGlone, C., 2013. *Manual of photogrammetry*. American Society for Photogrammetry and Remote Sensing (ASPRS).
- Micheletti, N., Lane, S. N., Chandler, J. H., 2015. Application of archival aerial photogrammetry to quantify climate forcing of alpine landscapes. *The Photogrammetric Record*, 30(150), 143–165.
- Mikolajczyk, K., Schmid, C., 2004. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1), 63–86.
- Moo Yi, K., Verdie, Y., Fua, P., Lepetit, V., 2016. Learning to assign orientations to feature points. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 107–116.
- Moravec, H., 1980. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical Report CMU-RI-TR-80-03, Carnegie Mellon University, Pittsburgh, PA.
- Nurminen, K., Litkey, P., Honkavaara, E., Vastaranta, M., Holopainen, M., Lyytikäinen-Saarenmaa, P., Kantola, T., Lyytikäinen, M., 2015. Automation aspects for the georeferencing of photogrammetric aerial image archives in forested scenes. *Remote Sensing*, 7(2), 1565–1593.
- Ono, Y., Trulls, E., Fua, P., Yi, K. M., 2018. Lf-net: learning local features from images. *Advances in Neural Information Processing Systems*, 6234–6244.
- Pierrot-Deseilligny, M., Cléry, I., 2012. Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-5/W16, 269–276.
- Pierrot-Deseilligny, M., Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(1/W41), 1–5.
- Pierrot-Deseilligny, M., Rupnik, E., Girod, L., Belvaux, J., Maillat, G., Deveau, M., Choqueux, G., 2015. MicMac, Apero, Pastis and Other Beverages in a Nutshell. *MicMac documentation*, 4.
- Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. *European conference on computer vision*, 430–443.
- Schonberger, J. L., Frahm, J.-M., 2016. Structure-from-motion revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4104–4113.
- Schonberger, J. L., Hardmeier, H., Sattler, T., Pollefeys, M., 2017. Comparative evaluation of hand-crafted and learned local features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1482–1491.
- Semyonov, D., 2011. Algorithms used in photoscan. *Agisoft PhotoScan Forum*, 3.
- Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F., 2015. Discriminative learning of deep convolutional feature point descriptors. *Proceedings of the IEEE International Conference on Computer Vision*, 118–126.
- Snavely, N., Seitz, S. M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3d. *ACM Siggraph 2006 Papers*, 835–846.
- Souchon, J.-P., Thom, C., Meynard, C., Martin, O., Pierrot-Deseilligny, M., 2010. The IGN CAMv2 System. *The Photogrammetric Record*, 25(132), 402–421.
- Tola, E., Lepetit, V., Fua, P., 2009. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE transactions on pattern analysis and machine intelligence*, 32(5), 815–830.
- USGS, 2019. earthexplorer. <https://earthexplorer.usgs.gov/>.
- Verdie, Y., Yi, K., Fua, P., Lepetit, V., 2015. Tilde: a temporally invariant learned detector. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5279–5288.
- Walstra, J., Chandler, J., Dixon, N., Dijkstra, T., 2004. Time for change-quantifying landslide evolution using historical aerial photographs and modern photogrammetric methods. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(B4).
- Yi, K. M., Trulls, E., Lepetit, V., Fua, P., 2016. Lift: Learned invariant feature transform. *European Conference on Computer Vision*, 467–483.