Université Gustave Eiffel

École Doctorale MSTIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

P H D T H E S I S

to obtain the title of

PhD of Science

of the Université Gustave Eiffel

Defended by

Lulin ZHANG

Feature matching for multi-epoch historical aerial images

Thesis Advisor:
Marc PIERROT DESEILLIGNY, Ewelina RUPNIK

prepared at Univ. Gustave Eiffel/Lastig ACTE/IGN/ENSG

defended on May xx, 2022

Jury :

Acknowledgments

Contents

List of Acronyms

Abbreviations

IGN IPGP Lastig ENSG

intra-epoch: from the same time inter-epoch: from different times DSM: Digital Surface Model DoD: Difference of DSMs GCP: Ground Control Point BBA: Bundle block adjustment

Chapter 1

Introduction

Contents

1.1 Motivation

Historical (i.e. analogue or archival) aerial images play an important role in providing unique information about evolution of land-covers. They are objective witness over time and sometimes the only remaining visual source of historical land-form. Therefore, they are valuable assets for a wide range of applications such as: (1) change detection, (2) spatial and urban planning, (3) long-term environmental monitoring including but not limited to forests, ice glaciers and coastlines, (4) analysis of natural disaster (e.g. earthquake, volcano eruption, landslide ect.) and the estimation of its future trends, etc.

Historical aerial images have regularly been acquired since the 1920's by mapping, military or cadastral agencies all over the world. A mass amount of them have been digitized and made accessible through web services [Giordano & Mallet 2019, USGS 2019, IGN 2019]. For example, according to a survey in Europe [Giordano & Mallet 2019], there are in total 50 million of aerial images archived, with around 37.8% of them digitized. The images are of high spatial resolution, and are acquired in stereoscopic configuration, allowing for 3D restitution of territories. They are often accompanied by metadata, in most cases including the camera focal length and the physical sensor size. Other metadata such as flight plans, camera calibration certificates or orientations are not commonly available.

When the camera calibration parameters are unknown, they should be evaluated by a process called self-calibrating bundle adjustment. Adequate Ground Control Points (GCPs) are required, otherwise inaccurately estimated camera parameters will lead to systematic error surfaces called dome effect (i.e. bowel effect). Generally, GCPs originate from (i) field surveys [Micheletti *et al.* 2015, Walstra *et al.* 2004, Cardenal *et al.* 2006], (ii) recent orthophotos and DEM [Nurminen *et al.* 2015, Ellis *et al.* 2006, Fox & Cziferszky 2008] and (iii) recent

satellite images [Ellis *et al.* 2006, Ford 2013]. The most challenging part is to identify the GCPs on the historical images, due to inevitable scene changes. GCPs are usually manually measured with the help of recent photos, however, it is still monotonous and time-consuming. There is an urgent need to automatically identify corresponding points (i.e. correspondences) on historical and recent images. However, recovering correspondences on images taken at different times (i.e. multi-epoch) remains challenging for the following reasons:

- Multi-epoch images are often acquired at different times of day and in different weathers or seasons, which unavoidably leading to appearance differences.
- The scene changes over time due to anthropogenic phenomenas (e.g. urban planning) or natural ones (e.g. earthquake), especially for large time gaps.
- Multi-epoch images often exhibit heterogeneous spatial resolutions, accompanied with different acquisition conditions (sensors, spectral channels, etc.).
- Historical images are often facing low radiometric quality, including low contrast, image noise, deterioration due to the aging of the films, or even scratches on the films.

1.2 Objective

1.3 Contribution

1.4 Organization of the thesis

Chapter 2

Literature review

Contents

2.1 Local feature matching

Local feature refers to finding a discriminative structure found in an image, such as a point, corner, blob, edge or image patch. It is often accompanied with a descriptor, which is a compact vector representing the local neighborhood.

According to the different data storage types, descriptors can be divided into two categories: floating-point descriptors and binary descriptors. The former is recorded in floating-point format, which has the advantage of being informative. It is widely used in various matching scenarios. The latter is stored in binary type, which guarantees faster processing while demanding less memory. It is particularly suitable for real-time and/or smartphone applications. Since our goal is to match multi-epoch images for high accuracy ground survey, we are interested in floating-point descriptors rather than binary ones.

According to whether machine learning techniques are applied, local features can be categorized as hand-crafted or learned. We will subsequently elaborate on the two categories of approaches.

2.1.1 Hand-crafted methods

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 methods

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]. Tian et al. introduced L2-Net [Tian *et al.* 2017] to learn high performance descriptor in Euclidean space via the Convolutional Neural Network (CNN). Afterwards Mishchuk et al. [Mishchuk *et al.* 2017] introduced a compact descriptor named HardNet, by applying a novel loss to L2Net [Tian *et al.* 2017]. DELF [Noh *et al.* 2017] is an attentive local feature descriptor based on CNN, which works particularly well for illumination changes. 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. D2-Net [Dusmanu *et al.* 2019] is a single neural network that works as a *dense* feature descriptor and a feature detector simultaneously, but their keypoints are less accurate compared to classical features since they are extracted on feature maps which have a resolution of 1/4 of the input resolution. ASLFeat [Luo *et al.* 2020] improves shape-awareness and localization accuracy by applying light-weight yet effective modifications on an improved D2-Net. R2D2 [Revaud *et al.* 2019] is a CNN architecture that learns *dense* local descriptors (one for each pixel) as well as two associated repeatability and reliability confidence maps. Contextdesc [Luo *et al.* 2019] is a unified learning framework that leverages and aggregates the cross-modality contextual information. D2D [Wiles *et al.* 2020] allows

dense features to be modified based on the differences between the images by conditioning the feature maps on both images. Different than the aforementioned feature extraction methods, SuperGlue [Sarlin *et al.* 2020] presents a new way of thinking about the feature matching problem. It matches two sets of pre-existing local features by adopting a flexible context aggregation mechanism based on attention to jointly find correspondences and reject non-matchable points.

Early learned methods (LIFT [Yi *et al.* 2016], L2-Net [Tian *et al.* 2017], HardNet [Mishchuk *et al.* 2017], DELF [Noh *et al.* 2017], SuperPoint [DeTone *et al.* 2018], LF-Net [Ono *et al.* 2018]) use only intermediate metrics (e.g., repeatability, matching score, mean matching accuracy, etc.) to evaluate the matching performance. Even though they 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]. Jin et al. [Jin *et al.* 2020] introduced a comprehensive benchmark for local features and robust estimation algorithms, focusing on the accuracy of the reconstructed camera pose as the primary metric. Using the new metric, SIFT [Lowe 2004] and SuperGlue [Sarlin *et al.* 2020] take the lead [Trulls *et al.* 2020].

2.2 Historical image processing

When it comes to inter-epoch historical images, however, directly applying SIFT or SuperGlue often results in inferior results due to large radiometric differences. In Figure **??** we showed an example where SIFT and SuperGlue failed on an inter-epoch image pair with drastic scene changes. It is understandable as (1) SIFT is not sufficiently invariant over time, while (2) SuperGlue is not invariant to rotations and it underperforms on larger images because it was presumably trained on small images.

Therefore, many previous researches bypassed the task of extracting inter-epoch correspondences by processing different epochs separately followed by an inter-epoch co-registration relying on Ground Control Points(GCPs). Between 10 and 169 GCPs are required in [Pinto *et al.* 2019], [Bożek *et al.* 2019], [Persia *et al.* 2020], [Micheletti *et al.* 2015], [Mölg & Bolch 2017]. GCPs are usually measured with the help of photointerpretation on recent orthophotos, however, it is still monotonous and time-consuming. Furthermore, it is difficult to find salient points that are stable over time.

Certain attempts were made to extract inter-epoch correspondences. Giordano et al. [Giordano *et al.* 2018] extract feature correspondences between historical and recent images relying on HoG descriptors [Dalal & Triggs 2005]. The authors require flight plans as input, which are not commonly available as mentioned in Section 1. Feurer et al. [Feurer & Vinatier 2018], Filhol et al. [Filhol *et al.* 2019],

Cook et al. [Cook & Dietze 2019], Parente et al. [Parente *et al.* 2021] and Blanch et al. [Blanch *et al.* 2021] assume that a sufficient number of keypoints remain invariant across time and employ SIFT to extract inter-epoch feature correspondences. It remains questionable whether the method is capable of handling drastic scene changes. Zhang et al. [Zhang *et al.* 2020] extract inter-epoch correspondences from SIFT-detected keypoints based on the hypothesis that points follow 2D and 3D spatial similarity model. This method works in simple cases with few scene changes. Additionally, a stream of research works focuses on historical terrestrial images ([Maiwald & Maas 2021], [Beltrami *et al.* 2019], [Bevilacqua *et al.* 2019], [Maiwald 2019]) and historical video recordings ([Maiwald 2019]). However, their algorithms are not suitable to the aerial case.
This work is an extention of [Zhang *et al.* 2020]. Unlike in [Zhang *et al.* 2020], we introduce a rough co-registration between different epochs based on matching DSMs with SuperGlue, and use it to guide a precise matching. Our rough co-registration is robust under extreme scene changes because (1) SuperGlue utilizes context to enhance feature descriptors and (2) DSMs are generally stable over time. With the guidance of roughly co-registered orientations and DSMs, both SIFT and SuperGlue achieved good performance, as shown in our experiments.

2.3 Robust matching

The goal of robust matching is to tell apart inliers from outliers, and eliminate the latter from further processing.

Typically, an iterative sampling strategy based on RANSAC (Random Sample Consensus) [Fischler & Bolles 1981] relying on some mathematical model, such as homography [Sonka *et al.* 2014] or essential matrix [Sonka *et al.* 2014] is carried out to remove outliers. This is an important issue which was often not given sufficient attention. LMedS (Least Median of Squares) [Leroy & Rousseeuw 1987] is a meaningful groundwork before RANSAC, which is also commonly used to replace RANSAC. MLESAC (Maximum Likelihood SAC) [Torr & Zisserman 2000] adopts the same sampling strategy as RANSAC but chooses the solution that maximizes the likelihood instead of the number of inliers. PROSAC (Progressive Sample Consensus) [Chum & Matas 2005] chooses samples from progressively larger sets of top-ranked correspondences, which makes it significantly faster than RANSAC. DEGENSAC [Chum *et al.* 2005] is an algorithm for epipolar geometry estimation unaffected by planar degeneracy. It is widely used in the 2020 image matching challenge [Trulls *et al.* 2020]. USAC (Universal RANSAC) [Raguram *et al.* 2012] framework is a synthesis of the various optimizations and improvements that have been proposed to RANSAC. GC-RANSAC (Graph-Cut RANSAC) [Barath & Matas 2018] runs graph-cut algorithm in the local optimization step. MAGSAC [Barath *et al.* 2019] eliminates the need for a user-defined inlier-outlier threshold with marginalization.

Various deep learning methods have also been developed to handle the

erroneous matches. DSAC (the differentiable counterpart of RANSAC) [Brachmann *et al.* 2017] replaces the deterministic hypothesis selection by a probabilistic selection. CNe (Context Networks) [Moo Yi *et al.* 2018] trains deep networks in an end-to-end fashion to label the correspondences as inliers or outliers, known intrinsics are required as input, and a post-processing with RANSAC is often tasked. CNe was embedded into the framework of [Jin *et al.* 2020] to remove outliers, paired with DEGENSAC, PyRANSAC (a variant of DEGENSAC by disabling the degeneracy check, introduced in [Jin *et al.* 2020]) and MAGSAC. The results showed that with SIFT used to train CNe, about 80% of the outliers were filtered out. Nearly all classical methods benefited from CNe, but not the learned ones. Jin et al. [Jin *et al.* 2020] also stated that RANSAC should be tuned to particular feature detector and descriptor, and specific settings should be selected for a particular RANSAC variant.

In this research, we use RANSAC to estimate the 3D Helmert transformation between surfaces (i.e., DSMs) calculated in different epochs. Compared to the classical essential/fundamental matrix filtering, with less data (3 versus 5 points) we impose stricter rules on the sets of points. Lastly, we eliminate the remaining false correspondences by looking at their cross-correlation.

Chapter 3

Rough co-registration

Contents

3.1 Introduction

3.1.1 Motivation

3.1.2 Contribution

3.2 Methodology

Our goal is to improve robustness by building globally consistent transformation model over the whole block. In order to achieve this goal, we exploded 2 strategies:
(1) matching each potential image pair followed with global filtering based on 3D RANSAC;
(2) get a global image for each epoch first (DSM or orthophoto), apply matching and 2D RANSAC.

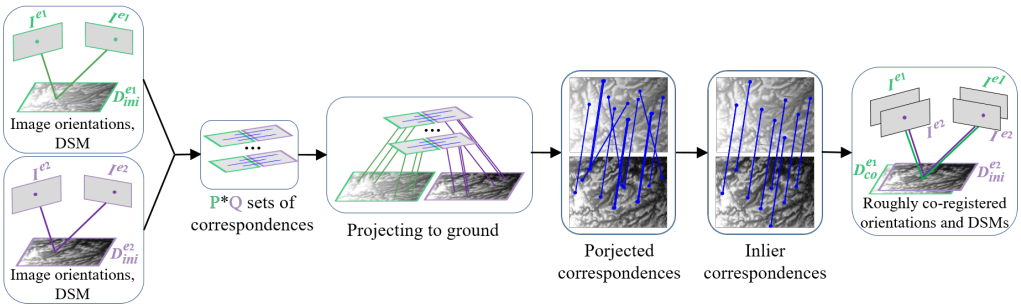


Figure 3.1: Rough co-registration by matching image pairs.

3.2.1 Strategy 1: Matching image pairs

3.2.1.1 SIFT

3.2.1.2 SuperGlue

3.2.2 Strategy 2: Matching DSMs/Orthophotos

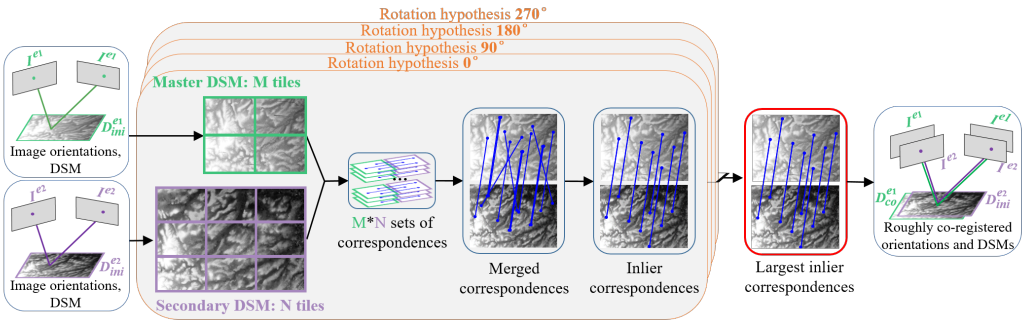


Figure 3.2: Rough co-registration by matching DSMs.

3.2.2.1 SIFT

3.2.2.2 SuperGlue

3.3 Experiments

3.3.1 Datasets

Frejus, Pezenas, Kobe

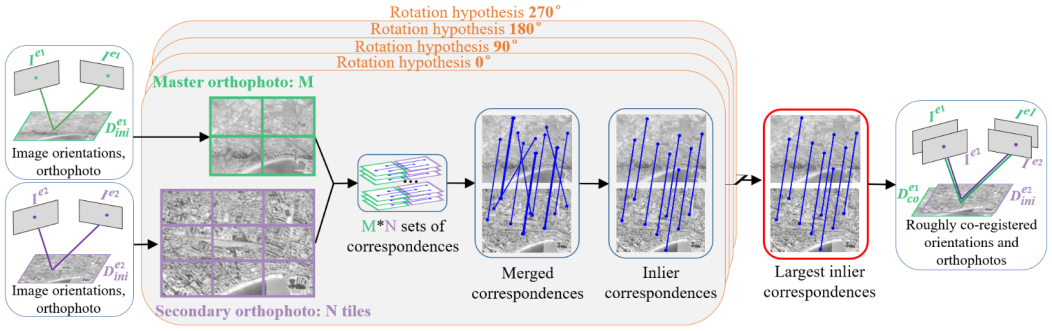


Figure 3.3: Rough co-registration by matching orthophotos.

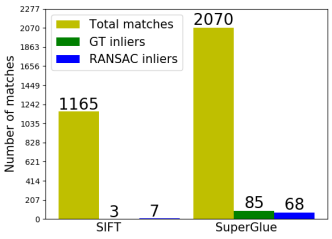
3.3.2 Evaluation

(1)Matches visualization
(2)Ground check points
(3)DoD

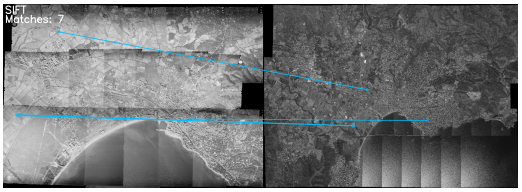
3.3.3 Comparison

3.3.3.1 Matches visualization

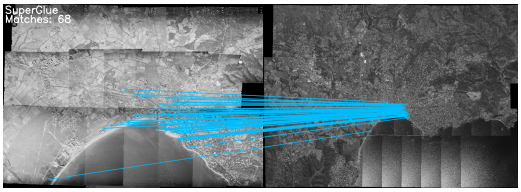
(a) Image pairs



(b) Match number (ImgPairs)



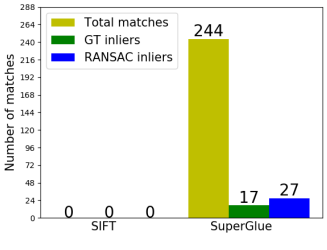
(c) $SIFT_{ImgPairs}$



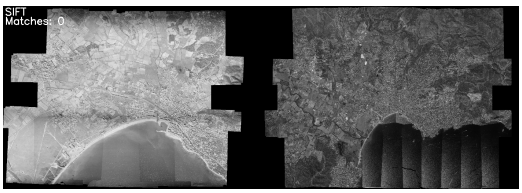
(d) $SuperGlue_{ImgPairs}$

Figure 3.4: Result of matching image pairs of Fréjus 1954 and 2014

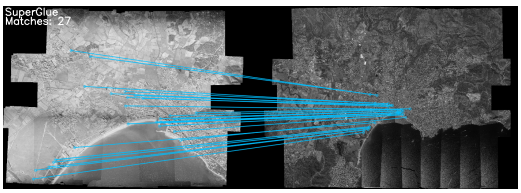
(a) Orthophotos



(b) Match number (ortho)

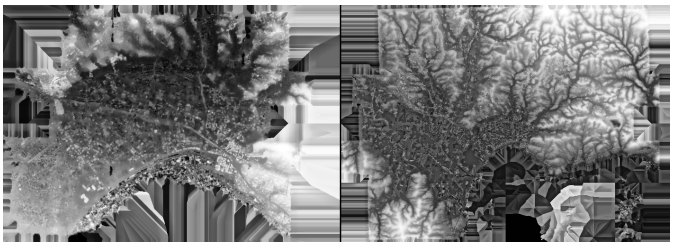


(c) $SIFT_{ortho}$

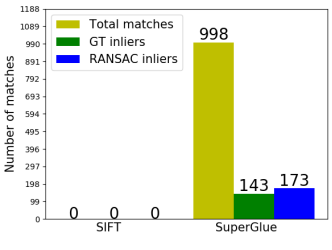


(d) $SuperGlue_{ortho}$

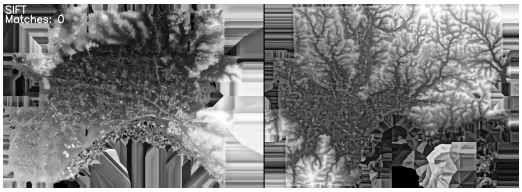
Figure 3.5: Result of matching orthophotos of Fréjus 1954 and 2014



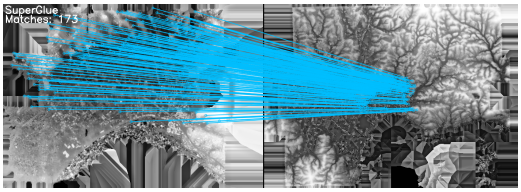
(a) DSMs



(b) Match number (DSM)

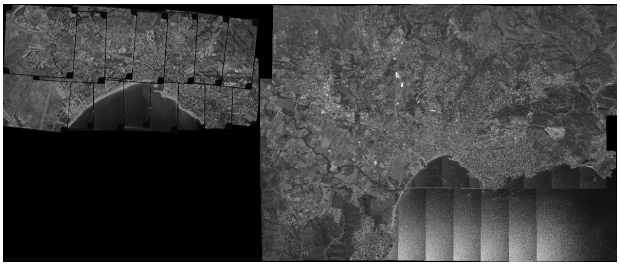


(c) $SIFT_{DSM}$

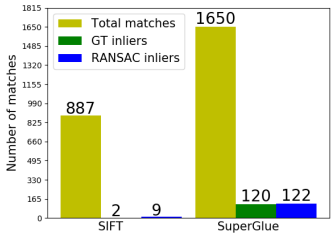


(d) $SuperGlue_{DSM}$

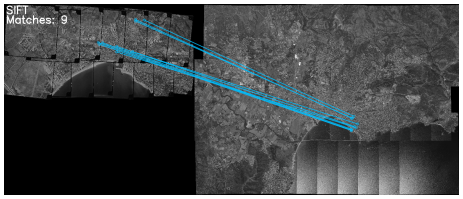
Figure 3.6: Result of matching DSMs of Fréjus 1954 and 2014



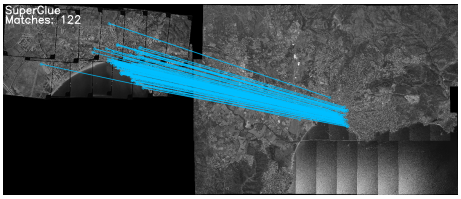
(a) Image pairs



(b) Match number (ImgPairs)



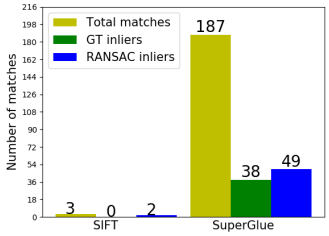
(c) $SIFT_{ImgPairs}$



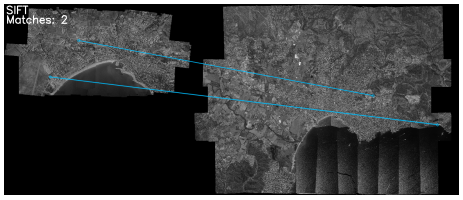
(d) $SuperGlue_{ImgPairs}$

Figure 3.7: Result of matching image pairs of Fréjus 1966 and 2014

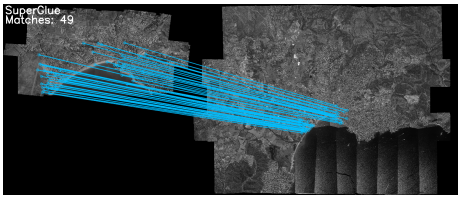
(a) Orthophotos



(b) Match number (ortho)

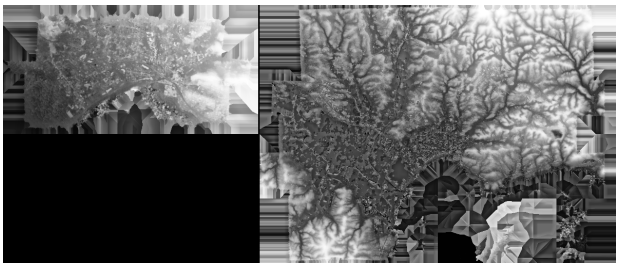


(c) $SIFT_{ortho}$

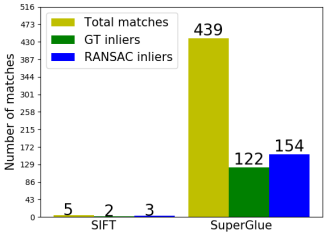


(d) $SuperGlue_{ortho}$

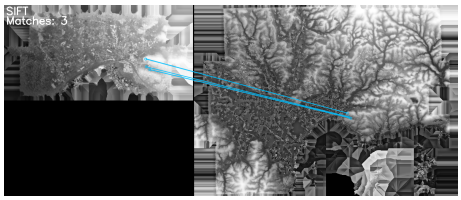
Figure 3.8: Result of matching orthophotos of Fréjus 1966 and 2014



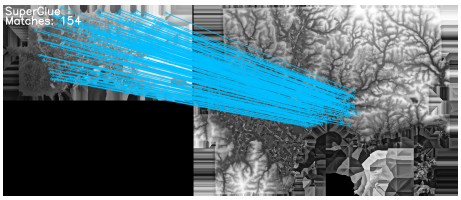
(a) DSMs



(b) Match number (DSM)

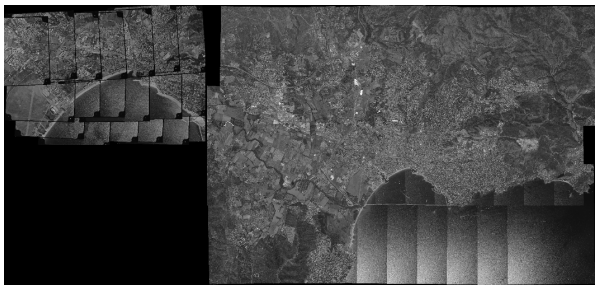


(c) $SIFT_{DSM}$

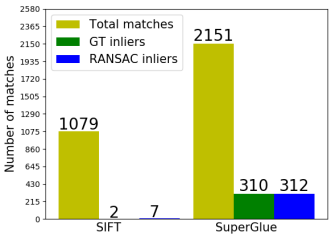


(d) $SuperGlue_{DSM}$

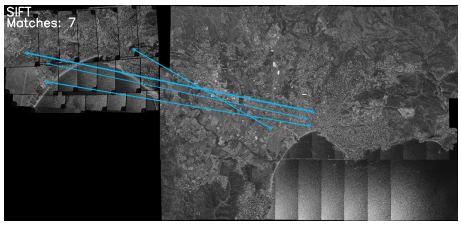
Figure 3.9: Result of matching DSMs of Fréjus 1966 and 2014



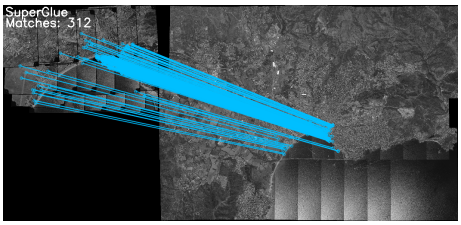
(a) Image pairs



(b) Match number (ImgPairs)

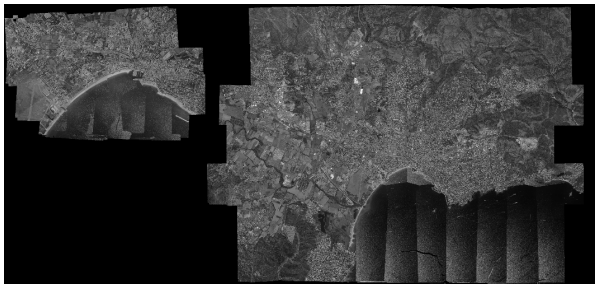


(c) $SIFT_{ImgPairs}$

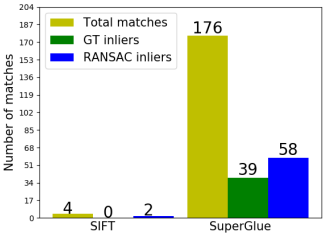


(d) $SuperGlue_{ImgPairs}$

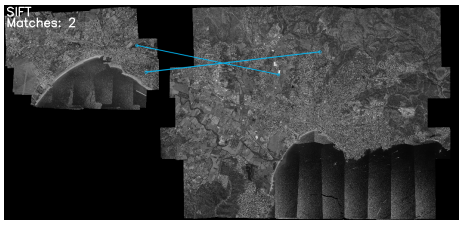
Figure 3.10: Result of matching image pairs of Fréjus 1970 and 2014



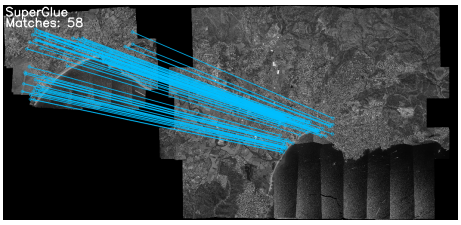
(a) Orthophotos



(b) Match number (ortho)

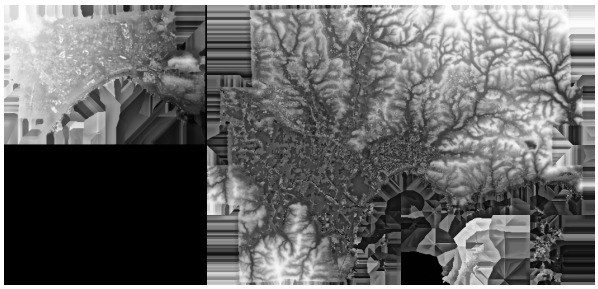


(c) $SIFT_{ortho}$

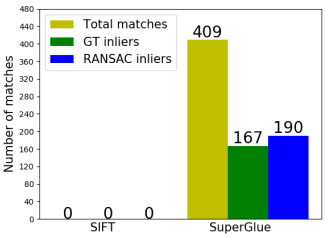


(d) $SuperGlue_{ortho}$

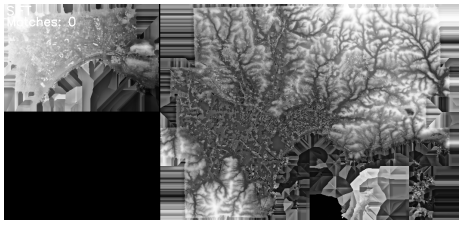
Figure 3.11: Result of matching orthophotos of Fréjus 1970 and 2014



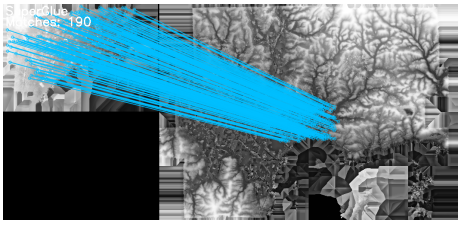
(a) DSMs



(b) Match number (DSM)



(c) $SIFT_{DSM}$



(d) $SuperGlue_{DSM}$

Figure 3.12: Result of matching DSMs of Fréjus 1970 and 2014

	$\|\mu\|$ [m]					
	ImgPairs		Orthophoto		DSM	
	SIFT	SuperGlue	SIFT	SuperGlue	SIFT	SuperGlue
$Frejus_{2014}^{1954}$	645.57	29.29	/	34.84	/	**10.72**
$Frejus_{2014}^{1966}$	1346.35	18.15	/	11.98	/	**10.43**
$Frejus_{2014}^{1970}$	2194.77	11.09	/	16.87	/	**10.23**
$Kobe_{1994}^{1991}$	/	**0.761**	0	1.2775	1.0465	1.5155

Table 3.1: Accuracy of 6 sets of co-registered orientations resulting from 6 methods, evaluated on 3 check points uniformly distributed in the block. Absolute average value $|\mu|$ is displayed for each method.

	μ [m]	σ [m]	$\|\mu\|$ [m]
$DoD_{Frejus1954}^{ImgPairs}$	5.70	6.32	6.62
$DoD_{Frejus1954}^{Orthophoto}$	2.19	6.46	4.55
$DoD_{Frejus1954}^{DSM}$	2.07	4.87	**3.83**
$DoD_{Frejus1966}^{ImgPairs}$	-1.36	3.82	2.90
$DoD_{Frejus1966}^{Orthophoto}$	-0.37	4.22	3.01
$DoD_{Frejus1966}^{DSM}$	-0.46	3.77	**2.68**
$DoD_{Frejus1970}^{ImgPairs}$	-5.04	5.09	5.70
$DoD_{Frejus1970}^{Orthophoto}$	-2.63	5.18	**4.39**
$DoD_{Frejus1970}^{DSM}$	-1.71	5.75	4.61

Table 3.2: Average value μ, standard deviation σ, and absolute average value $|\mu|$ of all the DoDs in Figure 3.13.

3.3.3.2 Ground check points

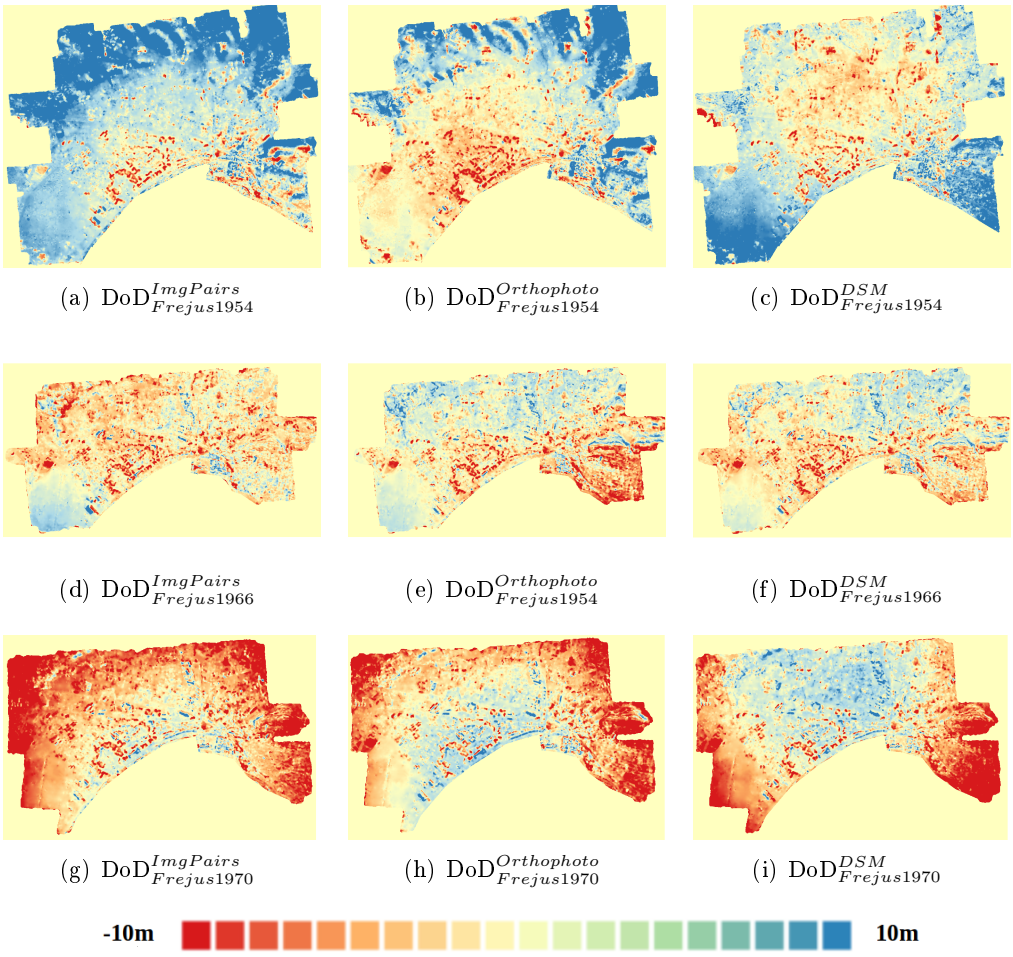
We manually measured 3 GCPs as check points to evaluate the roughly co-registered orientations resulted by 6 methods:

1. Match image pairs using SIFT;
2. Match image pairs using SuperGlue;
3. Match orthophotos using SIFT;
4. Match orthophotos using SuperGlue;
5. Match DSMs using SIFT;
6. Match DSMs using SuperGlue;

3.3.3.3 DoD

3.4 Conclusion

3.5 Discussion



(a) $\text{DoD}^{ImgPairs}_{Frejus1954}$ (b) $\text{DoD}^{Orthophoto}_{Frejus1954}$ (c) $\text{DoD}^{DSM}_{Frejus1954}$

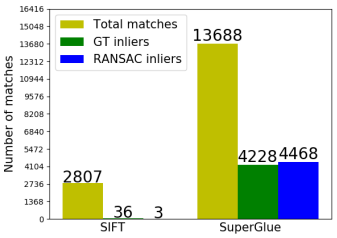
(d) $\text{DoD}^{ImgPairs}_{Frejus1966}$ (e) $\text{DoD}^{Orthophoto}_{Frejus1954}$ (f) $\text{DoD}^{DSM}_{Frejus1966}$

(g) $\text{DoD}^{ImgPairs}_{Frejus1970}$ (h) $\text{DoD}^{Orthophoto}_{Frejus1970}$ (i) $\text{DoD}^{DSM}_{Frejus1970}$

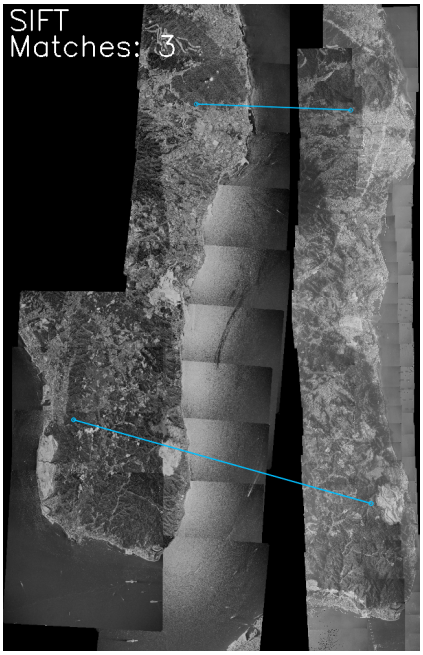
(j) DoD legend

Figure 3.13: DoDs of SuperGlue on dataset Fréjus

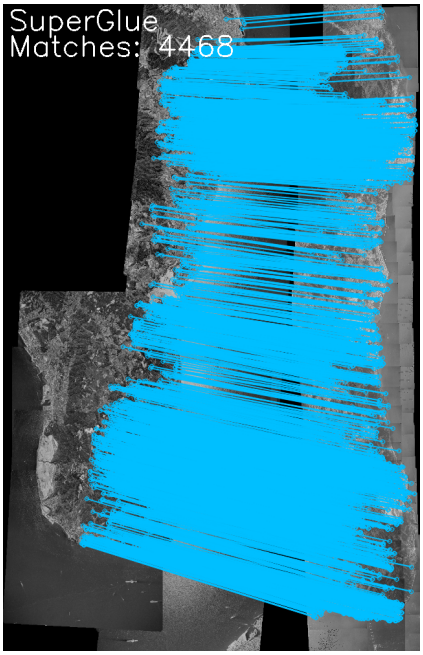
(a) Image pairs



(b) Match number (ImgPairs)

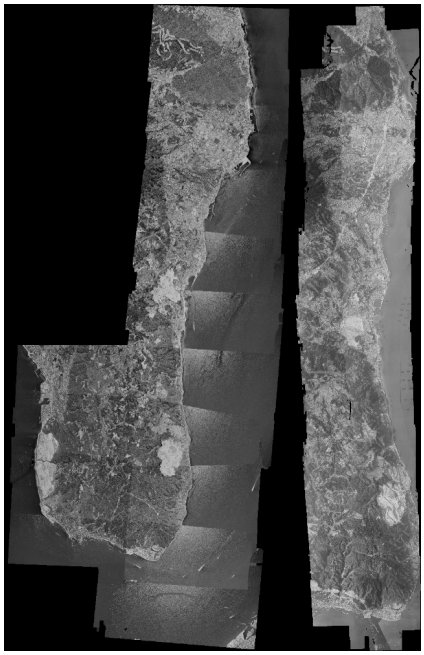


(c) $SIFT_{ImgPairs}$

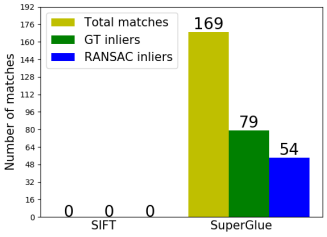


(d) $SuperGlue_{ImgPairs}$

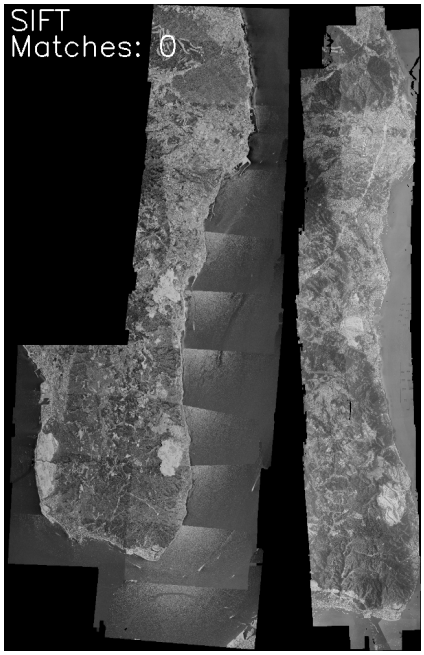
Figure 3.14: Result of matching image pairs of Kobe 1991 and 1995



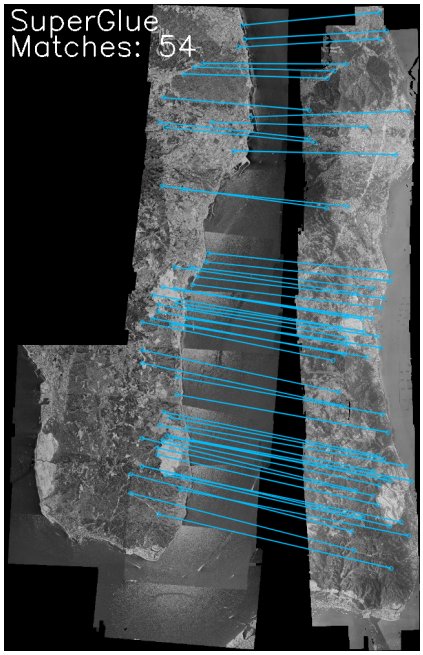
(a) Orthophotos



(b) Match number (ortho)



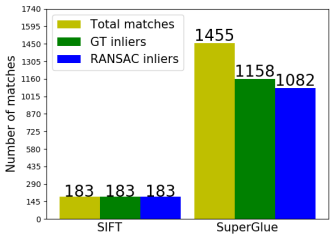
(c) $SIFT_{ortho}$



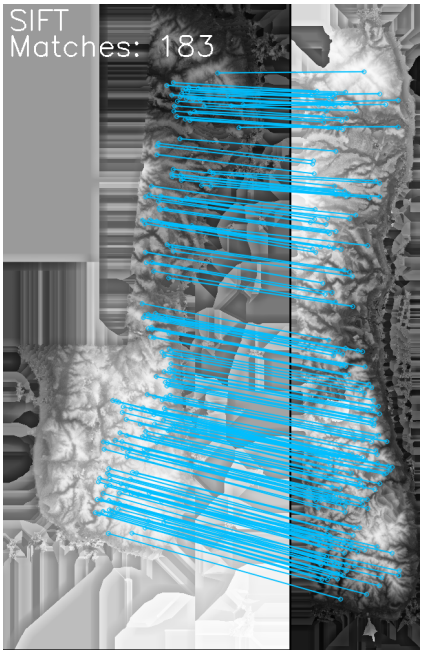
(d) $SuperGlue_{ortho}$

Figure 3.15: Result of matching orthophotos of Kobe 1991 and 1995

(a) DSMs



(b) Match number (DSM)



(c) $SIFT_{DSM}$

(d) $SuperGlue_{DSM}$

Figure 3.16: Result of matching DSMs of Kobe 1991 and 1995

Chapter 4

Precise matching

Contents

4.1 Introduction

4.1.1 Motivation

4.1.2 Contribution

4.2 Methodology

4.3 Experiments

4.4 Conclusion

4.5 Discussion

Chapter 5

Conclusion and Perspective

Appendices

Appendix A

Appendix Example

A.1 Appendix Example section

And I cite myself to show by bibtex style file (two authors) [**?**].

This for other bibtex stye file : only one author [**?**] and many authors [**?**].

Bibliography

[Alcantarilla *et al.* 2012] Pablo Fernández Alcantarilla, Adrien Bartoli and Andrew J Davison. *KAZE features.* In European Conference on Computer Vision, pages 214–227, 2012. (Cited on page 6.)

[Alcantarilla *et al.* 2013] P. F. Alcantarilla, J. Nuevo and A. Bartoli. *Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces.* In British Machine Vision Conf. (BMVC), 2013. (Cited on page 6.)

[Arandjelović & Zisserman 2012] Relja Arandjelović and Andrew Zisserman. *Three things everyone should know to improve object retrieval.* In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2911–2918. IEEE, 2012. (Cited on page 6.)

[Barath & Matas 2018] Daniel Barath and Jiří Matas. *Graph-cut RANSAC.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6733–6741, 2018. (Cited on page 8.)

[Barath *et al.* 2019] Daniel Barath, Jiri Matas and Jana Noskova. *Magsac: marginalizing sample consensus.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 10197–10205, 2019. (Cited on page 8.)

[Bay *et al.* 2006] Herbert Bay, Tinne Tuytelaars and Luc Van Gool. *Surf: Speeded up robust features.* In European conference on computer vision, pages 404–417, 2006. (Cited on page 6.)

[Beltrami *et al.* 2019] C Beltrami, D Cavezzali, F Chiabrando, A Iaccarino Idelson, G Patrucco and F Rinaudo. *3D Digital and Physical Reconstruction of a Collapsed Dome Using SFM Techniques From Historical Images.* International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2019. (Cited on page 8.)

[Bevilacqua *et al.* 2019] MG Bevilacqua, G Caroti, A Piemonte and D Ulivieri. *Reconstruction of lost architectural volumes by integration of photogrammetry from archive imagery with 3D models of the status quo.* International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2019. (Cited on page 8.)

[Blanch *et al.* 2021] Xabier Blanch, Anette Eltner, Marta Guinau and Antonio Abellan. *Multi-Epoch and Multi-Imagery (MEMI) Photogrammetric Workflow for Enhanced Change Detection Using Time-Lapse Cameras.* Remote Sensing, vol. 13, no. 8, page 1460, 2021. (Cited on page 8.)

[Bożek *et al.* 2019] Piotr Bożek, Jaroslaw Janus and Bartosz Mitka. *Analysis of changes in forest structure using point clouds from historical aerial photographs.* Remote Sensing, vol. 11, no. 19, page 2259, 2019. (Cited on page 7.)

[Brachmann *et al.* 2017] Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold and Carsten Rother. *Dsac-differentiable ransac for camera localization.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6684–6692, 2017. (Cited on page 9.)

[Cardenal *et al.* 2006] Javier Cardenal, Jorge Delgado, Emilio Mata, Alberto González and Ignacio Olague. *Use of historical flight for landslide monitoring.* Proceedings of the Spatial Accuracy, pages 129–138, 2006. (Cited on page 3.)

[Chum & Matas 2005] Ondrej Chum and Jiri Matas. *Matching with PROSAC-progressive sample consensus.* In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1, pages 220–226. IEEE, 2005. (Cited on page 8.)

[Chum *et al.* 2005] Ondrej Chum, Tomas Werner and Jiri Matas. *Two-view geometry estimation unaffected by a dominant plane.* In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 1, pages 772–779. IEEE, 2005. (Cited on page 8.)

[Cook & Dietze 2019] Kristen L Cook and Michael Dietze. *A simple workflow for robust low-cost UAV-derived change detection without ground control points.* Earth Surface Dynamics, vol. 7, no. 4, pages 1009–1017, 2019. (Cited on page 8.)

[Dalal & Triggs 2005] Navneet Dalal and Bill Triggs. *Histograms of oriented gradients for human detection.* In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1, pages 886–893. Ieee, 2005. (Cited on page 7.)

[DeTone *et al.* 2018] Daniel DeTone, Tomasz Malisiewicz and Andrew Rabinovich. *Superpoint: Self-supervised interest point detection and description.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 224–236, 2018. (Cited on pages 6 and 7.)

[Dusmanu *et al.* 2019] Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii and Torsten Sattler. *D2-Net: A Trainable CNN for Joint Detection and Description of Local Features.* In 2019 IEEE Conference on Computer Vision and Pattern Recognition, pages 8092–8101, 2019. (Cited on page 6.)

[Ellis *et al.* 2006] Erle C Ellis, Hongqing Wang, Hong Sheng Xiao, Kui Peng, Xin Ping Liu, Shou Cheng Li, Hua Ouyang, Xu Cheng and Lin Zhang Yang. *Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery.* Remote Sensing of Environment, vol. 100, no. 4, pages 457–473, 2006. (Cited on pages 3 and 4.)

[Feurer & Vinatier 2018] Denis Feurer and F Vinatier. *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, vol. 146, pages 495–506, 2018. (Cited on page 7.)

[Filhol *et al.* 2019] S Filhol, A Perret, L Girod, G Sutter, TV Schuler and JF Burkhart. *Time-Lapse Photogrammetry of Distributed Snow Depth During Snowmelt.* Water Resources Research, vol. 55, no. 9, pages 7916–7926, 2019. (Cited on page 7.)

[Fischler & Bolles 1981] Martin A Fischler and Robert C Bolles. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.* Communications of the ACM, vol. 24, no. 6, pages 381–395, 1981. (Cited on page 8.)

[Ford 2013] Murray Ford. *Shoreline changes interpreted from multi-temporal aerial photographs and high resolution satellite images: Wotje Atoll, Marshall Islands.* Remote Sensing of Environment, vol. 135, pages 130–140, 2013. (Cited on page 4.)

[Fox & Cziferszky 2008] Adrian J Fox and Andreas Cziferszky. *Unlocking the time capsule of historic aerial photography to measure changes in Antarctic Peninsula glaciers.* The Photogrammetric Record, vol. 23, no. 121, pages 51–68, 2008. (Cited on page 3.)

[Giordano & Mallet 2019] Sébastien Giordano and Clément Mallet. *Archiving and geoprocessing of historical aerial images: current status in Europe, Official Publication No 70.* In European Spatial Data Research, 2019. (Cited on page 3.)

[Giordano *et al.* 2018] S Giordano, A Le Bris and C Mallet. *Toward automatic georeferencing of archival aerial photogrammetric surveys.* ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2, pages 105–112, 2018. (Cited on page 7.)

[Harris & Stephens 1988] Chris Harris and Mike Stephens. *A combined corner and edge detector.* In In Proc. of Fourth Alvey Vision Conference, pages 147–151, 1988. (Cited on page 5.)

[IGN 2019] IGN. *remonterletemps.* https://remonterletemps.ign.fr/, 2019. (Cited on page 3.)

[Jin *et al.* 2020] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi and Eduard Trulls. *Image Matching across Wide Baselines: From Paper to Practice*. 2020 IEEE Conference on Computer Vision and Pattern Recognition, 2020. (Cited on pages 7 and 9.)

[Leroy & Rousseeuw 1987] Annick M Leroy and Peter J Rousseeuw. *Robust regression and outlier detection*. Wiley, 1987. (Cited on page 8.)

[Lowe 2004] David G Lowe. *Distinctive image features from scale-invariant keypoints*. International journal of computer vision, vol. 60, no. 2, pages 91–110, 2004. (Cited on pages 6 and 7.)

[Luo *et al.* 2019] Zixin Luo, Tianwei Shen, Lei Zhou, Jiahui Zhang, Yao Yao, Shiwei Li, Tian Fang and Long Quan. *Contextdesc: Local descriptor augmentation with cross-modality context*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2527–2536, 2019. (Cited on page 6.)

[Luo *et al.* 2020] Zixin Luo, Lei Zhou, Xuyang Bai, Hongkai Chen, Jiahui Zhang, Yao Yao, Shiwei Li, Tian Fang and Long Quan. *Aslfeat: Learning local features of accurate shape and localization*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6589–6598, 2020. (Cited on page 6.)

[Maiwald & Maas 2021] Ferdinand Maiwald and Hans-Gerd Maas. *An automatic workflow for orientation of historical images with large radiometric and geometric differences*. The Photogrammetric Record, 2021. (Cited on page 8.)

[Maiwald 2019] F Maiwald. *Generation of a Benchmark Dataset Using Historical Photographs for an Automated Evaluation of Different Feature Matching Methods*. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2019. (Cited on page 8.)

[Micheletti *et al.* 2015] Natan Micheletti, Stuart N Lane and Jim H Chandler. *Application of archival aerial photogrammetry to quantify climate forcing of alpine landscapes*. The Photogrammetric Record, vol. 30, no. 150, pages 143–165, 2015. (Cited on pages 3 and 7.)

[Mikolajczyk & Schmid 2004] Krystian Mikolajczyk and Cordelia Schmid. *Scale & affine invariant interest point detectors*. International journal of computer vision, vol. 60, no. 1, pages 63–86, 2004. (Cited on page 5.)

[Mishchuk *et al.* 2017] Anastasiia Mishchuk, Dmytro Mishkin, Filip Radenovic and Jiri Matas. *Working hard to know your neighbor's margins: Local descriptor learning loss*. In Advances in Neural Information Processing Systems, pages 4826–4837, 2017. (Cited on pages 6 and 7.)

[Mölg & Bolch 2017] Nico Mölg and Tobias Bolch. *Structure-from-motion using historical aerial images to analyse changes in glacier surface elevation.* Remote Sensing, vol. 9, no. 10, page 1021, 2017. (Cited on page 7.)

[Moo Yi *et al.* 2016] Kwang Moo Yi, Yannick Verdie, Pascal Fua and Vincent Lepetit. *Learning to assign orientations to feature points.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 107–116, 2016. (Cited on page 6.)

[Moo Yi *et al.* 2018] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann and Pascal Fua. *Learning to find good correspondences.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2666–2674, 2018. (Cited on page 9.)

[Moravec 1980] Hans Moravec. *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, 1980. (Cited on page 5.)

[Noh *et al.* 2017] Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand and Bohyung Han. *Large-scale image retrieval with attentive deep local features.* In Proceedings of the IEEE international conference on computer vision, pages 3456–3465, 2017. (Cited on pages 6 and 7.)

[Nurminen *et al.* 2015] Kimmo Nurminen, Paula Litkey, Eija Honkavaara, Mikko Vastaranta, Markus Holopainen, Päivi Lyytikäinen-Saarenmaa, Tuula Kantola and Minna Lyytikäinen. *Automation aspects for the georeferencing of photogrammetric aerial image archives in forested scenes.* Remote Sensing, vol. 7, no. 2, pages 1565–1593, 2015. (Cited on page 3.)

[Ono *et al.* 2018] Yuki Ono, Eduard Trulls, Pascal Fua and Kwang Moo Yi. *LF-Net: learning local features from images.* In Advances in Neural Information Processing Systems, pages 6234–6244, 2018. (Cited on pages 6 and 7.)

[Parente *et al.* 2021] Luigi Parente, Jim H Chandler and Neil Dixon. *Automated Registration of SfM-MVS Multitemporal Datasets Using Terrestrial and Oblique Aerial Images.* The Photogrammetric Record, vol. 36, no. 173, pages 12–35, 2021. (Cited on page 8.)

[Persia *et al.* 2020] Manuela Persia, Emanuele Barca, Roberto Greco, Maria Marzulli and Patrizia Tartarino. *Archival Aerial Images Georeferencing: A Geostatistically-Based Approach for Improving Orthophoto Accuracy with Minimal Number of Ground Control Points.* Remote Sensing, vol. 12, no. 14, page 2232, 2020. (Cited on page 7.)

[Pinto *et al.* 2019] Ana Teresa Pinto, José A Gonçalves, Pedro Beja and João Pradinho Honrado. *From archived historical aerial imagery to informative orthophotos: A framework for retrieving the past in long-term socioecological research.* Remote Sensing, vol. 11, no. 11, page 1388, 2019. (Cited on page 7.)

[Raguram *et al.* 2012] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas and Jan-Michael Frahm. *USAC: a universal framework for random sample consensus.* IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pages 2022–2038, 2012. (Cited on page 8.)

[Revaud *et al.* 2019] Jerome Revaud, Cesar De Souza, Martin Humenberger and Philippe Weinzaepfel. *R2d2: Reliable and repeatable detector and descriptor.* In Advances in Neural Information Processing Systems, pages 12405–12415, 2019. (Cited on page 6.)

[Rosten & Drummond 2006] Edward Rosten and Tom Drummond. *Machine learning for high-speed corner detection.* In European conference on computer vision, pages 430–443, 2006. (Cited on page 6.)

[Sarlin *et al.* 2020] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz and Andrew Rabinovich. *Superglue: Learning feature matching with graph neural networks.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4938–4947, 2020. (Cited on page 7.)

[Schonberger *et al.* 2017] Johannes L Schonberger, Hans Hardmeier, Torsten Sattler and Marc Pollefeys. *Comparative evaluation of hand-crafted and learned local features.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1482–1491, 2017. (Cited on page 7.)

[Simo-Serra *et al.* 2015] Edgar Simo-Serra, Eduard Trulls, Luis Ferraz, Iasonas Kokkinos, Pascal Fua and Francesc Moreno-Noguer. *Discriminative learning of deep convolutional feature point descriptors.* In Proceedings of the IEEE International Conference on Computer Vision, pages 118–126, 2015. (Cited on page 6.)

[Sonka *et al.* 2014] Milan Sonka, Vaclav Hlavac and Roger Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014. (Cited on page 8.)

[Tian *et al.* 2017] Yurun Tian, Bin Fan and Fuchao Wu. *L2-net: Deep learning of discriminative patch descriptor in euclidean space.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 661–669, 2017. (Cited on pages 6 and 7.)

[Tola *et al.* 2009] Engin Tola, Vincent Lepetit and Pascal Fua. *Daisy: An efficient dense descriptor applied to wide-baseline stereo.* IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pages 815–830, 2009. (Cited on page 6.)

[Torr & Zisserman 2000] Philip HS Torr and Andrew Zisserman. *MLESAC: A new robust estimator with application to estimating image geometry.* Computer vision and image understanding, vol. 78, no. 1, pages 138–156, 2000. (Cited on page 8.)

[Trulls *et al.* 2020] Eduard Trulls, Yuhe Jin, Kwang Moo Yi, Dmytro Mishkin, Jiri Matas and Pascal Fua. *Image Matching Challenge 2020.* https://vision.uvic.ca/image-matching-challenge/, 2020. (Cited on pages 7 and 8.)

[USGS 2019] USGS. *earthexplorer.* https://earthexplorer.usgs.gov/, 2019. (Cited on page 3.)

[Verdie *et al.* 2015] Yannick Verdie, Kwang Yi, Pascal Fua and Vincent Lepetit. *TILDE: a temporally invariant learned detector.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5279–5288, 2015. (Cited on page 6.)

[Walstra *et al.* 2004] Jan Walstra, JH Chandler, N Dixon and TA Dijkstra. *Time for change-quantifying landslide evolution using historical aerial photographs and modern photogrammetric methods.* International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35, no. B4, 2004. (Cited on page 3.)

[Wiles *et al.* 2020] Olivia Wiles, Sebastien Ehrhardt and Andrew Zisserman. *D2D: Learning to find good correspondences for image matching and manipulation.* arXiv preprint arXiv:2007.08480, 2020. (Cited on page 6.)

[Yi *et al.* 2016] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit and Pascal Fua. *Lift: Learned invariant feature transform.* In European Conference on Computer Vision, pages 467–483, 2016. (Cited on pages 6 and 7.)

[Zhang *et al.* 2020] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Guided feature matching for multi-epoch historical image blocks pose estimation.* In ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., 2020. (Cited on page 8.)