Feature matching for multi-epoch historical aerial images

Lulin Zhang

École Doctorale MSTIC
Mathématiques & Sciences et Technologies de l'Information et de la Communication

THÈSE DE DOCTORAT

En vue de l'obtention du grade de :

Docteur de l'Université Gustave-Eiffel

Défendue par

Lulin ZHANG

Appariement des caractéristiques pour les images aériennes historiques multi-époques

Superviseur de thèse:
Marc PIERROT DESEILLIGNY, Ewelina RUPNIK

préparé à Univ. Gustave Eiffel/Lastig ACTE/IGN/ENSG

défendue le 20 mai 2022

Président : Livio De Luca

Jury :

Rapporteur:	Marie-Odile Berger	INRIA Nancy
Rapporteur:	El Mustapha Mouaddib	UPJV
Examinateur:	Denis Feurer	IRD
Examinateur:	Livio De Luca	CNRS
Membre invité:	Yann Klinger	IPGP
Membre invité:	Michele Santangelo	CNR-IRPI

MSTIC Doctoral School

Mathematics & Sciences and Technologies of Information and Communication

Ph.D Thesis

to obtain the title of

Doctorate of the Gustave Eiffel University

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 20, 2022

President : Livio De Luca

Jury :

Reviewer: Marie-Odile Berger	INRIA Nancy
Reviewer: El Mustapha Mouaddib	UPJV
Examiner: Denis Feurer	IRD
Examiner: Livio De Luca	CNRS
Invited Member: Yann Klinger	IPGP
Invited Member: Michele Santangelo	CNR-IRPI

Résumé

L'imagerie historique se caractérise par une haute résolution spatiale et des acquisitions stéréoscopiques. Elle constitue une ressource précieuse pour la détection des changements et la surveillance environnementale à long terme. Des millions d'images historiques ont été numérisées. Elles sont des témoins objectifs du temps et parfois la seule source visuelle restante de la forme historique du territoire. Cependant, l'énorme potentiel des images historiques diachroniques est supprimé en raison du goulot d'étranglement que constitue leur géoréférencement précis. Il s'agit d'un processus appelé ajustement de faisceau auto-calibré pour estimer les paramètres de calibrage de la caméra. Il faut un nombre suffisant de correspondances dans des paysages évolutifs, qui sont difficiles à obtenir automatiquement, en raison des changements de scène et des conditions hétérogènes d'acquisition des images.

Dans cette recherche, nous présentons des pipelines entièrement automatiques pour trouver des correspondances entre des images historiques prises à différents temps (c'est-à-dire, inter-époques), sans données auxiliaires nécessaires. En profitant de la géométrie 3D et de la stratégie grossier-à-précis, nous (1) enregistrons grossièrement les différentes époques en établissant un modèle de transformation globalement cohérent sur l'ensemble du bloc, et (2) nous apparions précisément les images inter-époques sous la direction du co-enregistrement grossier pour réduire l'ambiguïté. Six variantes de deux stratégies sont explorées pour l'étape de co-enregistrement grossier, et deux variantes pour l'étape d'appariement précis. Nos pipelines sont adaptés à diverses applications de surveillance environnementale. Cinq données représentatifs sont choisis pour les expériences, chacun représentant une application caractéristique. Avec les correspondances inter-époques récupérées, nous améliorons les orientations de l'image puis calculons les Digital Surface Models (DSMs) à chaque époque, et évaluons quantitativement les résultats avec les Difference of DSMs (DoDs) et le déplacement du sol dû à un séisme. Nous démontrons que notre méthode (1) peut géoréférencer automatiquement des images historiques diachroniques ; (2) peut atténuer efficacement les erreurs systématiques induites par des paramètres de caméra mal estimés ; et (3) est robuste contre les changements drastiques de la scène. Les pipelines proposés sont mis en œuvre dans MicMac, un logiciel de photogrammétrie libre et gratuit.

Mots clefs: Appariement des caractéristiques, Images historiques, Multi-époques, Estimation de la pose, Auto-étalonnage

Abstract

Historical imagery is characterized by high spatial resolution and stereoscopic acquisitions, providing a valuable resource for change detection and long-term environmental monitoring. Millions of historical images have been digitized. They are objective witness over time and sometimes the only remaining visual source of historical land-form. However, the huge potential of diachronic historical images is suppressed due to the bottleneck of their accurate geo-referencing. It involves a process called self-calibrating bundle adjustment to estimate the camera calibration parameters. Sufficient amount of matches under evolving landscapes are required, which are difficult to be obtained automatically, due to scene changes and heterogeneous image acquisition conditions.

In this research, we present fully automatic pipelines to finding matches between historical images taken at different times (i.e., inter-epoch), without auxiliary data required. By taking advantage of 3D geometry and rough-to-precise strategy, we (1) roughly co-register different epochs by establishing a globally consistent transformation model over the whole block, and (2) precisely match inter-epoch images under the guidance of rough co-registration to reduce ambiguity. Six variants out of 2 strategies are explored for rough co-registration stage, and two variants for precise matching stage. Our pipelines are suitable for diverse applications of environmental monitoring. Five representative sets of datasets are chosen for experiments, each one represents a characteristic application. With the recovered inter-epoch matches, we refine the image orientations followed by calculating Digital Surface Model (DSM)s in each epoch, and quantitatively evaluate the results with Difference of DSMs (DoD)s and ground displacement due to an earthquake. We demonstrate that our method (1) can automatically geo-reference diachronic historical images; (2) can effectively mitigate systematic errors induced by poorly estimated camera parameters; and (3) is robust against drastic scene changes. The proposed pipelines are implemented in MicMac, a free, open-source photogrammetric software.

Keywords: Feature matching, Historical images, Multi-epoch, Pose estimation, Self-calibration

Acknowledgments

I recollect the beginning of my PhD career as if it was yesterday. I resigned from a job that lasted 6 years in China and came to Paris in search of a new start, hoping to find a more research-oriented position in photogrammetry and computer vision. I still remember how excited I was when I received the offer of this PhD position. Now that I'm approaching the end of this meaningful journey, I can't help but feel how quickly time flies. Looking back on the past three years, mixed feelings well up in my mind, among them there are pleasure and delight, as well as regret, and most importantly gratitude. I am deeply grateful for the people I met.

First of all, sincere thanks go to my supervisors Marc Pierrot-Deseilligny and Ewelina Rupnik. They offered me this precious opportunity which met all my needs, and I thank them from the bottom of my heart for their recognition and trust in me. During my PhD research, they have given me a lot of guidance and help in my work, I learned a lot from them. Marc is a senior specialist in photogrammetry. He possesses profound experience in this domain, from which I benefited a lot. He is a first-rate helmsman who is good at controlling direction with ease. Ewelina is versatile, she is always there for me whenever I have a problem. She is capable of many unexpected skills that surprises me, both in theory and application. More importantly, both of them are rigorous scholars, they set up role models for me in my future research. Besides, Ewelina is an exemplary feminism to me and sets an instance for women pursuing their careers.

I would also like to express my special gratitude to Yann Klinger and Arthur Delorme in Institut de Physique du Globe de Paris (IPGP), Michele Santangelo in National Research Council, Research Institute for Hydrogeological Protection (CNR-IRPI), Han-Kyung Bae and Dong-Eun Kim in Korea Institute of Geoscience and Mineral Resources (KIGAM), Joaquín Muñoz-Cobo Belart in National Land Survey of Iceland. Thanks to all of them, I learned a lot about the possible applications for my research.

Heartfelt acknowledgment is made to Denis Feurer and Fabrice Vinatier for their contribution to the interior calibration of the Pezenas dataset, Benjamin Ferrand for the 2015 acquisition of Pezenas, Sebastien Giordano for the Fréjus dataset, Han-Kyung Bae and Dong-Eun Kim for the Kobe dataset, Michele Santangelo for the Alberona dataset, and Joaquín Muñoz-Cobo Belart for the Hofsjökull dataset. We are thankful to ANR project DISRUPT (ANR-18-CE31-0012–0) for supporting this work.

In particular, I deeply appreciate my husband, Teng Wu, who is also an academic in photogrammetry. He gave me a lot of useful advises in my work and life. He is also a very reliable partner, I got countless help and support from him. Although he is not very expressive, he is fully devoted to me and our family. I know that he

will support me unconditionally whenever I need him, he met all the expectations I had for a life partner. Thanks to the gift of life, we have two healthy and lovely children who brought hope and joy into our lives. Since having them, we've got a deeper sense of responsibility and a stronger motivation to become better examples for them. With a great sense of debt, I am grateful to my parents for raising me and supporting all the choices I made in my life. Thanks to my brother, who took over the responsibility of taking care of my parents for me after I left China. As an aged PhD student, it is not easy to balance work and life, I can't imagine pursuing my PhD without the support from all the family members I mentioned before.

Last but not least, during my research at Lastig, I met many warm-hearted friends: Manchun Lei, Yilin Zhou, Imane Fikri, Mohamed Ali Chebbi, Christophe Meynard, Jean-Michael Muller, Jean-Philippe Souchon, Olivier Martin, Lanfa Liu, Nathan Piasco, Emile Blettery, Raphael Sulzer, Stephane Guinard, Oussama Ennafii, Evelyn Paiz etc., who made my work environment full of friendliness and laughter. I am grateful to all the friends in my life who enriched me and I look forward to meeting them more often after the COVID subsides.

Contents

List of Acronyms

Chapter 1

Introduction en français

Contents

1.1 Motivation et objectifs

1.1.1 Pourquoi les images historiques sont-elles intéressantes

Les images aériennes historiques (c'est-à-dire analogiques ou d'archives) jouent un rôle important en fournissant des informations uniques sur l'évolution de la couverture terrestre. Ce sont des atouts précieux pour un grand nombre d'applications telles que (1) l'analyse des catastrophes naturelles (par exemple, tremblement de terre, glissement de terrain, volcan, inondation, avalanche, etc.), (2) la surveillance éco-environnementale (par exemple, forêt, atmosphère, glacier, eau, littoral, etc.), (3) l'expansion urbaine et (4) la pollution et la protection de l'environnement, etc.

Les images aériennes historiques ont été régulièrement acquises depuis les années 1920 par des agences cartographiques, militaires ou cadastrales du monde entier. Une quantité massive d'entre elles ont été numérisées et rendues accessibles par des services web [Giordano & Mallet 2019, USGS 2019, IGN 2019]. Par exemple, selon une enquête réalisée au début de 2017 en Europe [Giordano & Mallet 2019], il y a environ 50 millions d'images aériennes archivées en Europe, dont environ 37,8% sont numérisées. Les images sont de haute résolution spatiale, et sont acquises en configuration stéréoscopique, permettant la restitution 3D des territoires. Elles sont souvent accompagnées de métadonnées, comprenant dans la plupart des cas la focale de la caméra, la hauteur de vol, l'échelle et la taille physique du capteur, qui sont généralement enregistrées ou mentionnées sur les films. D'autres métadonnées telles que les plans de vol, les certificats d'étalonnage de la caméra ou les orientations ne sont pas couramment disponibles.

Lorsque les paramètres d'étalonnage de la caméra sont inconnus, ils doivent être évalués au moyen d'une procédure appelée ajustement du faisceau d'auto-étalonnage. Ground Control Point (GCP)s sont nécessaires, sinon des paramètres de caméra estimés de manière inexacte entraîneront des surfaces d'erreur systématiques appelées effet de dôme (c'est-à-dire effet de bol). En général, les GCPs proviennent (1) de mesures sur le terrain [Micheletti *et al.* 2015, Walstra *et al.* 2004, Cardenal *et al.* 2006], (2) d'orthophotos et de DSM récents [Nurminen *et al.* 2015, Ellis *et al.* 2006, Fox & Cziferszky 2008] et (3) d'images satellites récentes [Ellis *et al.* 2006, Ford 2013]. Le plus difficile est d'identifier les GCPs sur les images historiques, ce qui n'est pas facile en raison des inévitables changements de scène. Les GCP sont généralement mesurés manuellement à l'aide de photos récentes, mais cela reste monotone et laborieux. Il est urgent d'identifier automatiquement les points correspondants (c'est-à-dire les correspondances) sur des images historiques et récentes.
Lorsque les utilisateurs sont uniquement intéressés par la comparaison de différentes époques historiques, l'auto-calibrage peut être réalisé sans GCPs. Les correspondances entre différentes époques serviraient d'observations dans l'ajustement du faisceau pour éliminer les erreurs systématiques des surfaces. En conclusion, le goulot d'étranglement de l'auto-calibration des images historiques est la récupération des correspondances sur des images prises à des époques différentes (c'est-à-dire multi-époques).

1.1.2 Comment faire correspondre des images historiques multi-époques

Cependant, la comparaison d'images historiques multi-époques reste difficile, malgré le fait qu'il existe un grand nombre d'algorithmes de comparaison d'images dont l'efficacité a été prouvée sur des images modernes. Les raisons en sont les suivantes:

1. Les images multi-époques sont souvent acquises à différents moments de la journée et par différents temps et saisons, ce qui entraîne inévitablement des différences d'apparence.

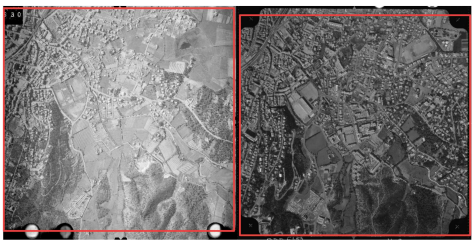
2. La scène change au fil du temps en raison de phénomènes anthropiques (par exemple, l'urbanisme) ou naturels (par exemple, un tremblement de terre), en particulier pour les grands écarts temporels.

3. Les images multi-époques présentent souvent des résolutions spatiales hétérogènes, accompagnées de conditions d'acquisition différentes (capteurs, canaux spectraux, etc).

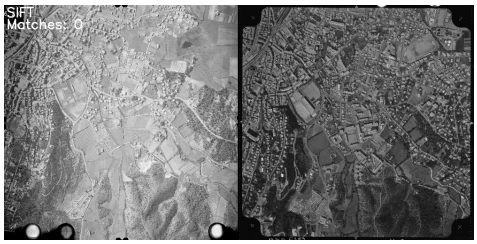
4. Les images historiques sont souvent confrontées à une faible qualité radiométrique, notamment un faible contraste, du bruit d'image, une détéri-

oration causée par le vieillissement des films, ou même des rayures sur les films.

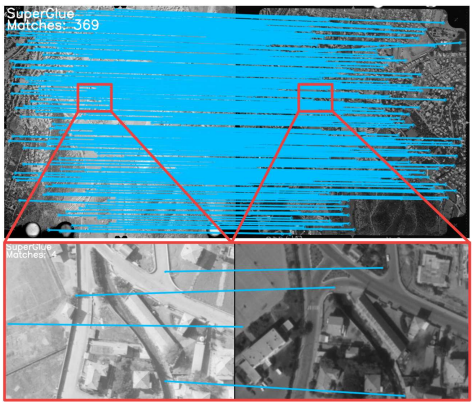
La simple application de méthodes d'appariement des caractéristiques (par exemple, SIFT [Lowe 2004] ou SuperGlue [Sarlin *et al.* 2020]) sur des paires d'images multi-époques donne souvent des résultats insatisfaisants. Un exemple est donné dans la Figure 1.1. Une paire d'images multi-époques est représentée avec des rectangles rouges indiquant la zone de chevauchement sur la Figure 1.1(a). Les images de gauche et de droite ont été prises au même endroit en 1954 et 1970 respectivement. La scène a changé de manière significative, beaucoup de nouveaux bâtiments sont apparus, les tons de couleur étaient très différents. Dans la Figure 1.1(b-d), les résultats de correspondance de SIFT, SuperGlue et le nôtre sont affichés pour comparaison. Comme on peut le voir, SIFT n'a trouvé aucune correspondance. SuperGlue a trouvé 369 correspondances, dont la plupart semblent bonnes, mais en regardant plus attentivement, les détails révèlent une faible précision de localisation. Notre méthode a trouvé 1463 correspondances avec une grande précision, grâce à l'aide (1) de la géométrie 3D et (2) de la stratégie diviser et conquérir (c'est-à-dire grossier-à-précis), qui sont détaillées dans les textes suivants.



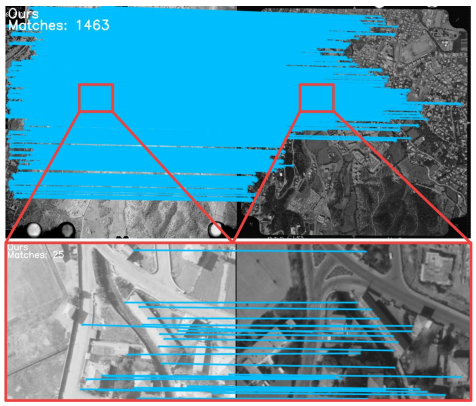
(a) Paire d'images multi-époques



(b) Résultat de SIFT (0 correspondances)



(c) Résultat de SuperGlue



(d) Résultat du nôtre

Figure 1.1: (a) Une paire d'images multi-époques avec des rectangles rouges indiquant la zone de chevauchement. (b-d) Résultat de la correspondance de SIFT, SuperGlue et le nôtre.

Avantages de la géométrie 3D Les images RGB sont largement utilisées pour pour l'appariement des images. Cependant, elles présentent les inconvénients suivants:
(1) Leur apparence change avec le temps (voir la Figure 1.2), et avec des angles de vue variables sur des surfaces non-Lambertiennes (voir la Figure 1.3). (2) Les autosimilitudes (par exemple, les modèles répétitifs) favorisent les fausses correspondances (voir la Figure 1.3).
Heureusement, la géométrie 3D telle que DSM compense parfaitement ces défauts. Comme on peut le voir sur la Figure 1.2, les images RGB sont très différentes car la scène a beaucoup changé. Cependant, les DSM correspondants sont similaires, ce qui est raisonnable, car le paysage 3D est plus stable dans le temps. De plus, les DSM sont plus distinctifs que les images RGB lorsqu'il s'agit de surfaces non-Lambertiennes et de motifs répétitifs, comme indiqué dans la Figure 1.3. Même si la géométrie 3D manque de textures et de détails par rapport à l'image RGB, elle sert de complément idéal. En outre, elle joue un rôle important en fournissant des informations 3D pour établir un modèle de transformation de Helmert 3D entre les époques afin (1) de déplacer différentes époques dans le même cadre de coordonnées et (2) de supprimer les fausses correspondances dans une routine RANSAC qui est plus fiable que les modèles de transformation 2D.

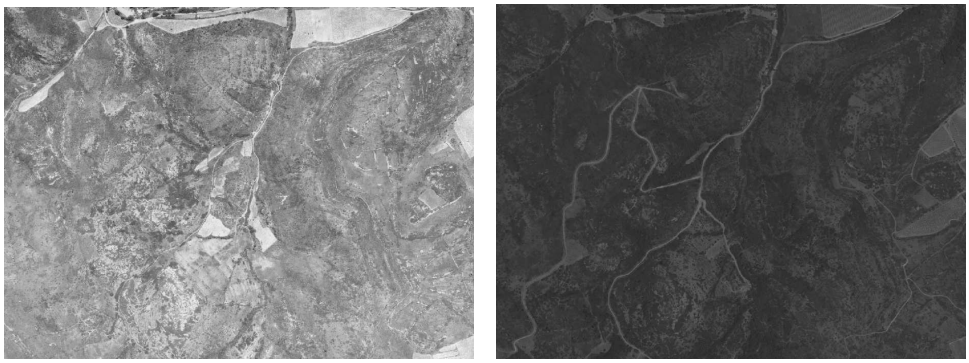
Diviser et conquérir Puisque la récupération de correspondances robustes et précises sur des paires d'images multi-époques est une tâche difficile, nous divisons la tâche en deux sous-tâches et les conquérons individuellement avec la stratégie grossier-à-précis. Cette stratégie est illustrée dans la Figure 1.4. Les deux sous-tâches sont les suivantes:

1. Co-enregistrement grossier, comme illustré sur la Figure 1.4(b). Son objectif est d'aligner grossièrement les paires d'images multi-époques en se concentrant sur la robustesse et en relâchant l'exigence de précision.

2. Appariement précis, comme illustré sur la Figure 1.4(c). Elle améliore les correspondances prédites par le résultat grossier du co-enregistrement en recherchant uniquement le voisinage local pour réduire l'ambiguïté.

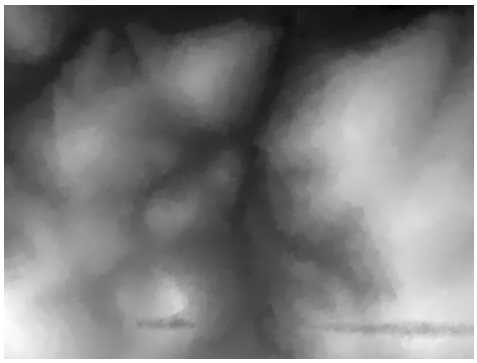
1.2 Contributions

Dans cette thèse, nous présentons des pipelines grossiers-à-précis pour l'appariement d'images multi-époques. Ils sont adaptés aux images aériennes, satellitaires et mixtes, ce qui ouvre la possibilité de géoréférencer des millions d'images historiques sans nécessiter de GCPs. Six variantes sont proposées pour l'étape de co-enregistrement grossier et deux variantes pour l'étape d'appariement précis. Chaque variante a sa propre caractéristique:

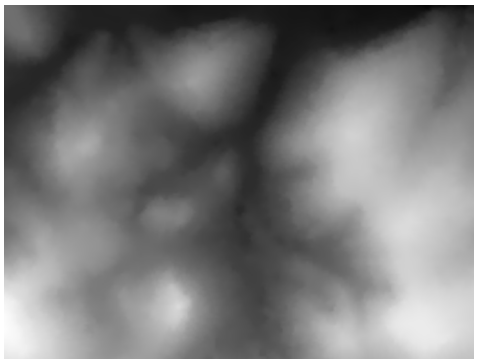


(a) Image RGB 1971

(b) Image RGB 2015

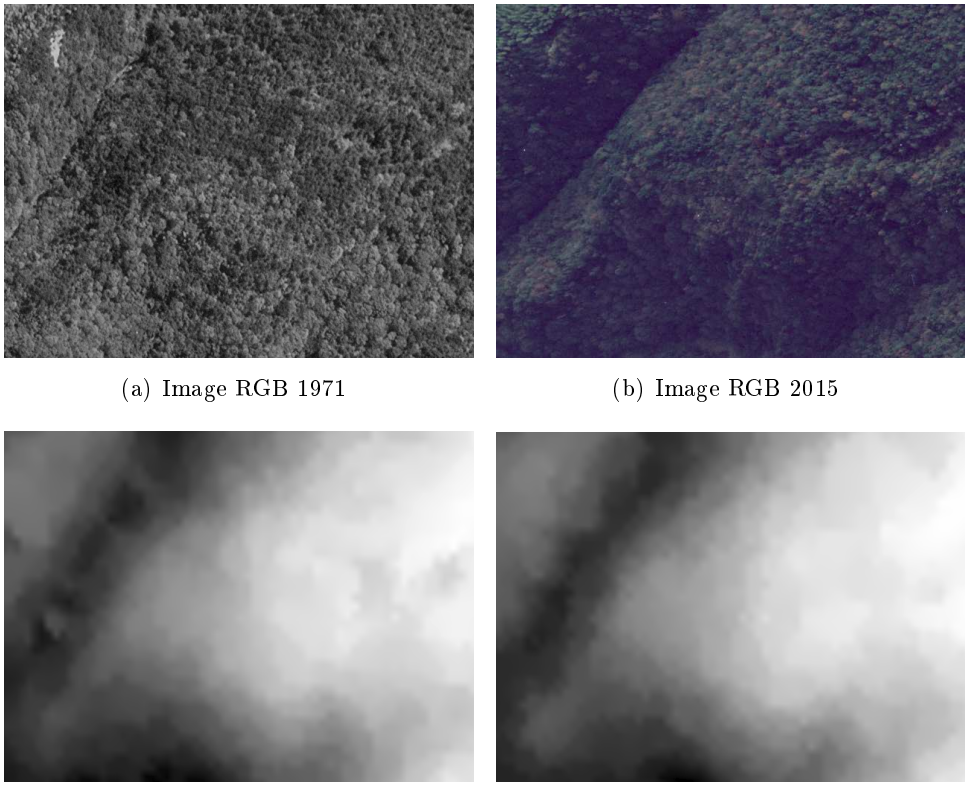


(c) DSM 1971



(d) DSM 2015

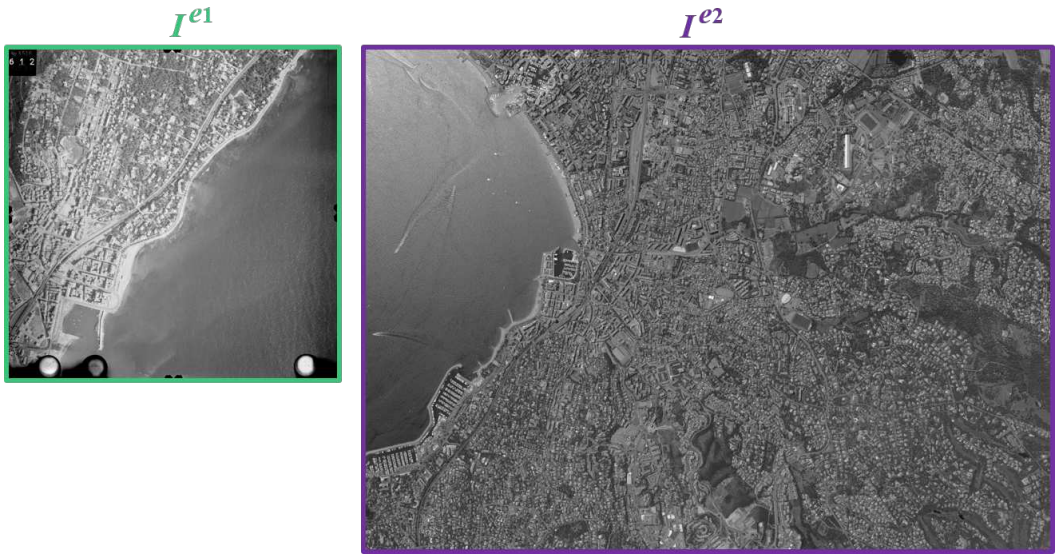
Figure 1.2: La même zone observée à différents moments. Les images RGB ont beaucoup changé alors que les DSMs sont restés stables au fil du temps.



(a) Image RGB 1971 (b) Image RGB 2015

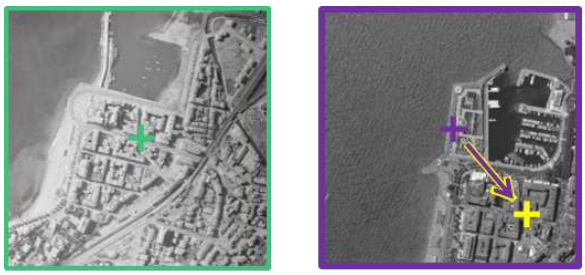
(c) DSM 1971 (d) DSM 2015

Figure 1.3: La même végétation observée à des moments différents. Réflexion non-lambertienne et autosimilitude présentes dans les images RGB, tandis que les DSMs restent distinctifs.



(a) Exemple d'une paire d'images inter-époques

(b) Co-enregistrement grossier



(c) Appariement précis

Figure 1.4: Stratégie grossier-à-précis. (a) Un exemple de paire d'images inter-époques à apparier. I^{e_1} et I^{e_2} représentent les images prises à $epoch_1$ et $epoch_2$ individuellement. (b) Illustration du co-enregistrement grossier entre I^{e_1} et I^{e_2}. En conséquence, I^{e_1} est grossièrement aligné avec I^{e_2}. (c) Illustration de l'appariement précis. Pour les points clés de I^{e_1} (croix verte), un emplacement est prédit dans I^{e_2} (croix violette) sur la base d'un co-enregistrement grossier, dont le voisinage local sera recherché pour trouver l'appariement précis (croix jaune).

1. Pour les variantes de co-enregistrement grossier: (1) celles basées sur l'idée d'appariement des DSMs conduisent généralement aux résultats les plus robustes ; (2) celles qui apparient les orthophotos pourraient servir d'alternatives dans les rares scénarios de terrain parfaitement plat où les DSMs ne fournissent pas d'informations utiles ; (3) les autres qui apparient les paires d'images originales conduisent souvent à des résultats moins satisfaisants, mais ce sont les seules options adaptées aux images terrestres.

2. Pour les variantes d'appariement précis: (1) *Patch* est basé sur des méthodes d'appariement par apprentissage, il donne généralement plus de correspondances car il est plus invariant dans le temps. (2) *Guided* est basé sur des méthodes artisanales, il est plus efficace en termes d'utilisation de la mémoire et des ressources CPU car il n'implique pas de rééchantillonnage des patchs, ce qui est nécessaire pour *Patch*.

Nos pipelines visent à libérer le potentiel des images historiques pour le suivi des conditions environnementales. Nous collaborons actuellement avec plusieurs instituts pour appliquer nos pipelines dans différentes applications, notamment:

1. Institut de Physique du Globe de Paris (IPGP) et Korea Institute of Geoscience and Mineral Resources (KIGAM) pour analyser les déformations de la croûte terrestre afin de comprendre les événements sismiques.

2. Conseil national de la recherche, Institut de recherche pour la protection hydrogéologique (CNR-IRPI) pour l'analyse de l'évolution des glissements de terrain en Italie.

3. Département des sciences de la terre et de l'environnement de l'université de Pavie pour l'analyse de l'évolution des badlands en Europe.

Nous avons également développé deux tutoriels complets accompagnés d'ensembles de données de test pour familiariser les utilisateurs avec nos pipelines implémentés dans MicMac[Zhang *et al.* 2021a] (plus de détails sont présentés dans l'annexe C):

1. Tutoriel d'appariement des images aériennes [Zhang *et al.* 2021e]

2. Tutoriel d'appariement d'images mixtes (c'est-à-dire d'images aériennes et satellitaires) [Zhang *et al.* 2021d]

Publications de l'auteur:

1. M Santangelo, **L Zhang**, E Rupnik, M Pierrot-Deseilligny, M Cardinali. Schéma d'évolution des glissements de terrain révélé par des MNS multitemporels obtenus à partir d'images aériennes historiques. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2022.

2. **L Zhang**, E Rupnik, M Pierrot-Deseilligny. Appariement des caractéristiques pour des images aériennes historiques multi-époques, 182, 176-189, 2021.

3. **L Zhang**, E Rupnik, M Pierrot-Deseilligny. Appariement des caractéristiques guidé pour l'estimation de la pose de blocs d'images historiques multi-époques. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020.

Nous fournissons également une vidéo [Zhang *et al.* 2021f], des diapositives [Zhang *et al.* 2021c] et le site web du projet [Zhang *et al.* 2021b] pour améliorer la visibilité de notre travail.

1.3 Organisation de la thèse

Cette thèse présente des pipelines entièrement automatiques pour l'appariement d'images multi-époques. Une brève présentation de *l'état de l'art* est donnée dans le **Chapitre** 3.

Dans le **Chapitre** 4, les applications ainsi que 5 données représentatifs sont présentés, qui sont ensuite utilisés pour tester nos pipelines.

Dans le **Chapitre** 5, six variantes de co-enregistrement grossier sont élaborées pour aligner grossièrement le bloc entier en construisant un modèle de transformation globalement cohérent entre les époques différentes.

Dans le **Chapitre** 6, deux variantes d'appariement précis sont introduites pour obtenir des appariements exacts sous la direction des orientations et des DSM qui sont grossièrement co-registrés.

Enfin, le **Chapitre** 7 présente les conclusions et les perspectives.

Chapter 2

Introduction

Contents

2.1 Motivation and objectives

2.1.1 Why are historical images interesting

Historical (i.e., analogue or archival) aerial images play an important role in providing unique information about evolution of land-covers. They are valuable assets for a wide range of applications such as (1) analyzing of natural disasters (e.g., earthquake, landslide, volcano, flood, avalanche, etc), (2) eco-environmental monitoring (e.g., forest, atmosphere, glacier, water, coastline, etc), (3) urban expansion and (4) environmental pollution and protection and so on.

Historical aerial images have been regularly 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 taken place at the beginning of 2017 in Europe [Giordano & Mallet 2019], there are approximately 50 millions of aerial images archived in Europe, 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, flight height, scale and the physical sensor size, which are usually saved or mentioned on the films. 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 procedure called self-calibrating bundle adjustment. GCPs are required, otherwise inaccurately estimated camera parameters will lead to systematic error surfaces

called dome effect (i.e., bowl effect). Generally, GCPs originate from (1) field surveys [Micheletti *et al.* 2015, Walstra *et al.* 2004, Cardenal *et al.* 2006], (2) recent orthophotos and DSM [Nurminen *et al.* 2015, Ellis *et al.* 2006, Fox & Cziferszky 2008] and (3) recent satellite images [Ellis *et al.* 2006, Ford 2013]. The most challenging part is to identify the GCPs on the historical images, which is not easy 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., matches) on historical and recent images.
When users are only interested in comparing different historical epochs, the self-calibration can be accomplished without GCPs. Matches between different epochs would serve as observations in bundle adjustment to eliminate the systematic errors in surfaces. In conclusion, the bottleneck of historical image self-calibration is recovering matches on images taken at different times (i.e., multi-epoch).

2.1.2 How to match multi-epoch historical images

However, matching multi-epoch historical images remains challenging, despite the fact that there exists a large number of image matching algorithms with their effectiveness proven on modern images. The reasons include:

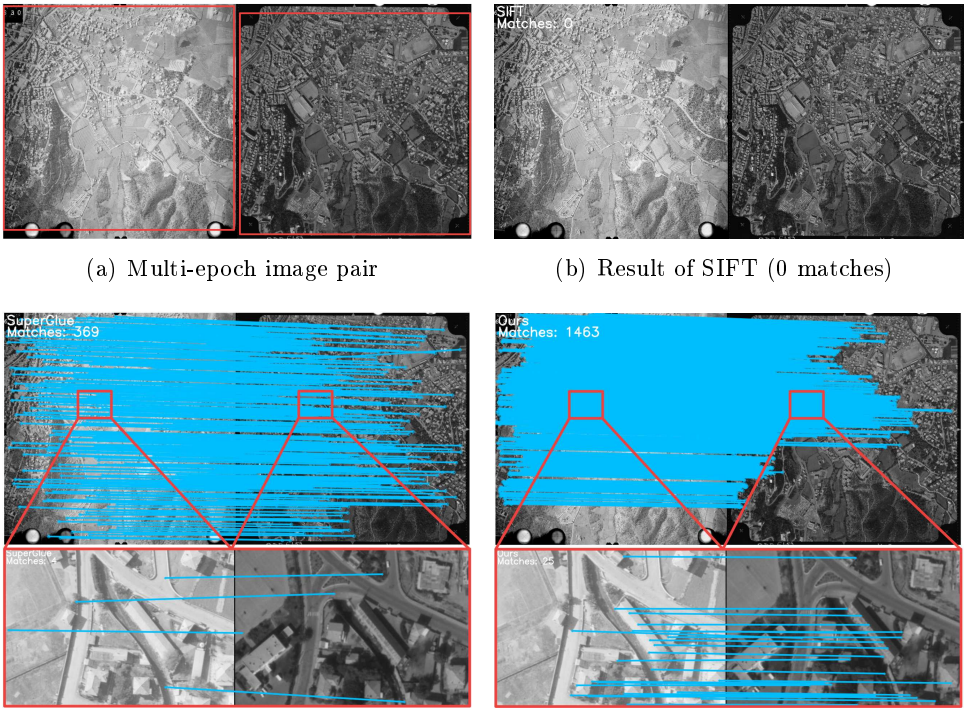
1. Multi-epoch images are often acquired at different times of day and in various weathers and seasons, which unavoidably leading to appearance differences.

2. The scene changes over time due to anthropogenic phenomena (e.g., urban planning) or natural ones (e.g., earthquake), especially for large time gaps.

3. Multi-epoch images often exhibit heterogeneous spatial resolutions, accompanied with different acquisition conditions (sensors, spectral channels, etc).

4. Historical images are often facing low radiometric quality, including low contrast, image noise, deterioration due to the aging of films, or even scratches on the films.

Simply applying *state-of-the-art* feature matching methods (e.g., SIFT [Lowe 2004] or SuperGlue [Sarlin *et al.* 2020]) on multi-epoch image pair often leads to unsatisfactory results. An example is given in Figure 2.1. A pair of multi-epoch images are demonstrated with red rectangles indicating the overlapping area in Figure 2.1(a). The left and right images are taken at the same place in 1954 and 1970 individually. The scene changed significantly, a lot of new buildings arose, the color tones were very different. In Figure 2.1(b-d), the matching result of SIFT, SuperGlue and Ours are displayed for comparison. As can be seen, SIFT failed to find any matches. SuperGlue recovered 369 matches, most of which seem good, but at a closer look the details reveal poor localization precision. Our method found 1463 matches with high accuracy, thanks to the help of (1) 3D geometry and (2) the divide and conquer (i.e., rough-to-precise) strategy, which are elaborated in the following texts.



(a) Multi-epoch image pair

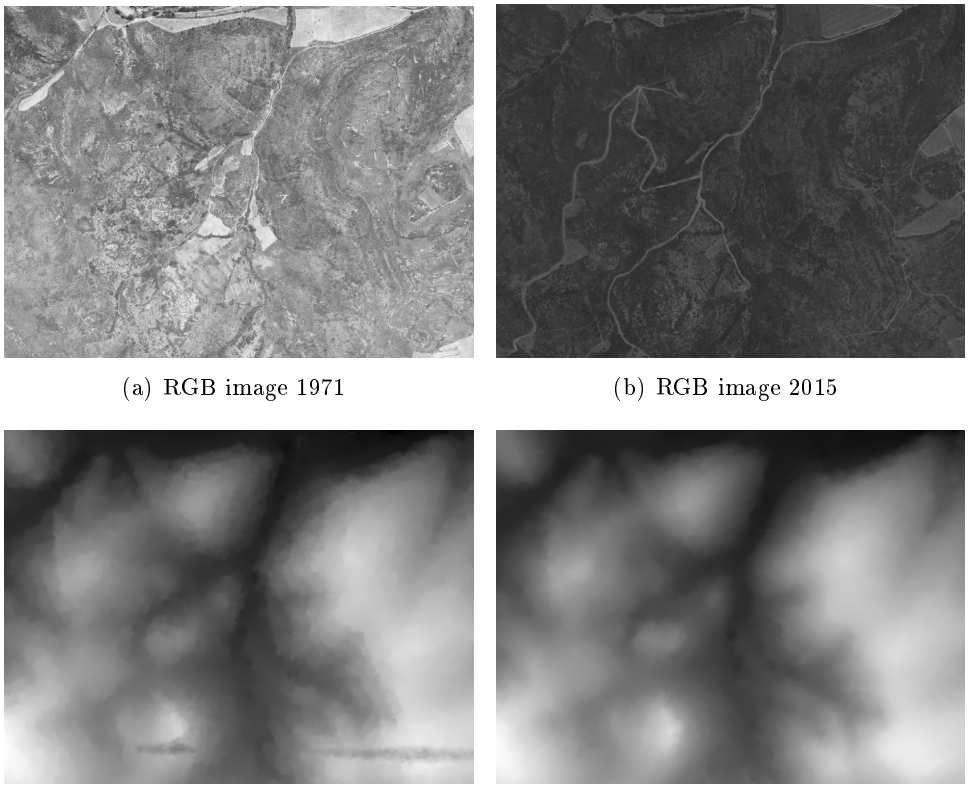
(b) Result of SIFT (0 matches)

(c) Result of SuperGlue

(d) Result of Ours

Figure 2.1: (a) A pair of multi-epoch images with red rectangles indicating the overlapping area. (b-d) Matching result of SIFT, SuperGlue and Ours.

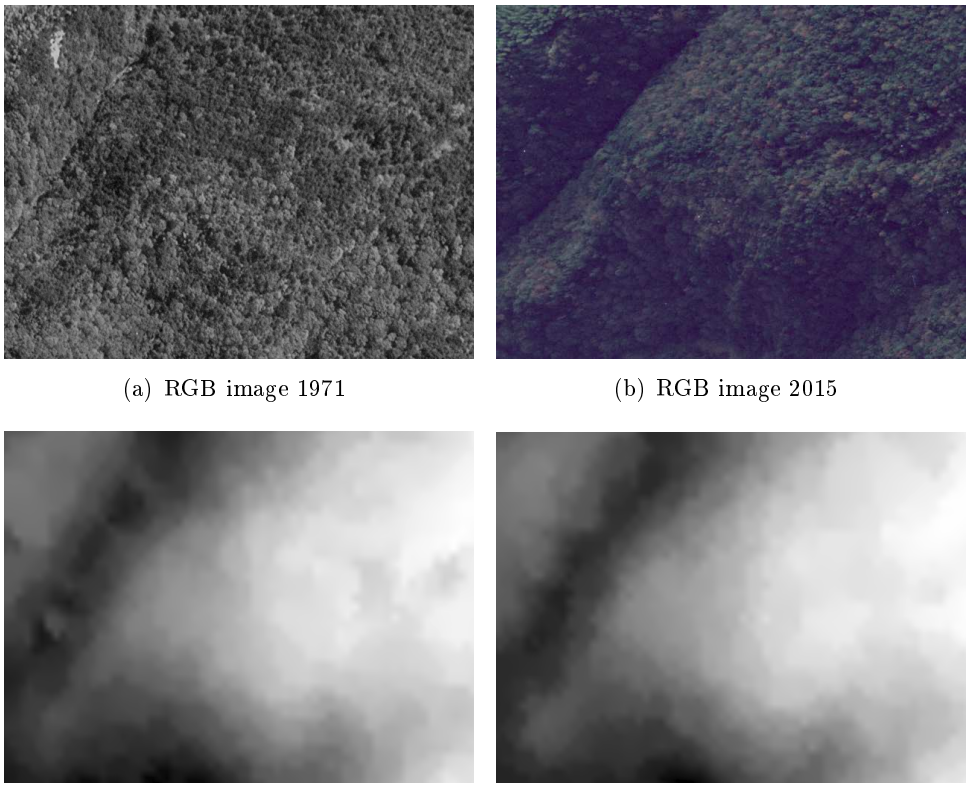
Advantages of 3D geometry RGB images are widely used for image matching. However, they have the following shortcomings:
(1) Their appearances change over time (see Figure 2.2), and over varying view angles on non-Lambertian surfaces (see Figure 2.3).
(2) Self similarities (e.g., repetitive patterns) favor false matches (see Figure 2.3).
Fortunately, 3D geometry such as DSM makes up for these shortcomings perfectly. As can be seen in Figure 2.2, the RGB images look very different because the scene changed a lot. However, the corresponding DSMs look similar, which is reasonable, as the 3D landscape is more stable over time. Besides, DSM is more distinctive than RGB image when it comes to non-Lambertian surfaces and repetitive patterns, as shown in Figure 2.3. Even though 3D geometry lacks textures and details compared to RGB image, it serves as an ideal supplement. Besides, it plays an important role in providing the 3D information for establishing 3D Helmert transformation model between epochs to (1) move different epochs into the same coordinate frame and (2) remove false matches in a RANSAC routine which is more reliable than 2D transformation models.



(a) RGB image 1971 (b) RGB image 2015

(c) DSM 1971 (d) DSM 2015

Figure 2.2: The same zone observed in different times. The RGB images changed a lot while the DSMs stayed stable over time.



(a) RGB image 1971 (b) RGB image 2015

(c) DSM 1971 (d) DSM 2015

Figure 2.3: The same vegetation observed in different times. Non-Lambertian reflection and self similarities present in RGB images, while the DSMs stay distinctive.

Divide and Conquer Since the task of recovering robust and precise matches on multi-epoch image pairs is difficult, we divide the task into two sub-tasks and conquer them individually with the rough-to-precise strategy. It is illustrated in Figure 2.4. The two sub-tasks includes:

1. Rough co-registration, as illustrated in Figure 2.4(b). Its goal is to roughly align the multi-epoch image pairs by focusing on robustness and relaxing the requirement for accuracy.
2. Precise matching, as illustrated in Figure 2.4(c). It refines the matches predicted by the rough co-registration result by searching only the local neighborhood to reduce ambiguity.

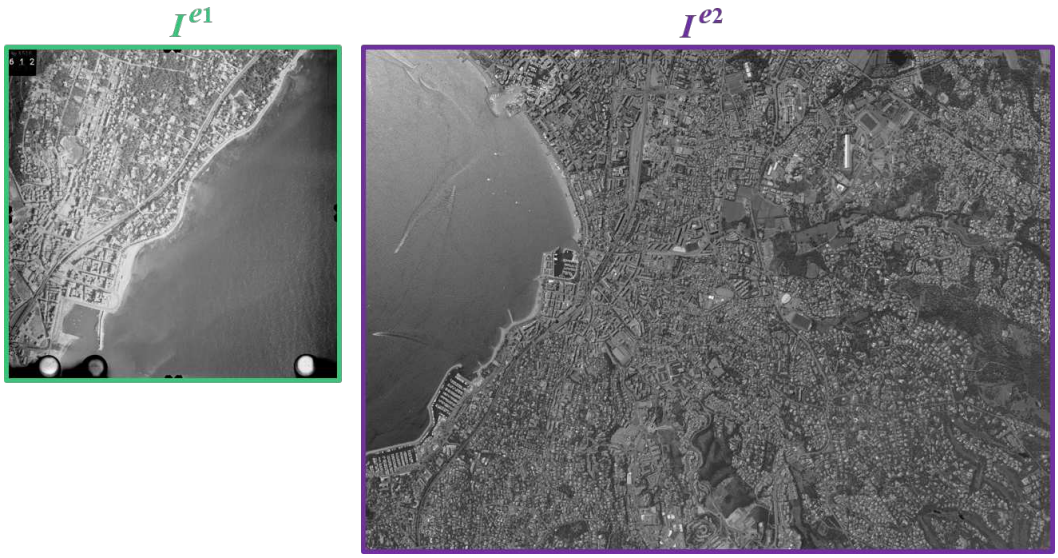
2.2 Contributions

In this thesis we present rough-to-precise pipelines for matching multi-epoch images. They are suitable for aerial, satellite and mixed images, which open the possibility of geo-referencing millions of historical images without requiring any GCPs. Six variants are provided for the rough co-registration stage and two variants for the precise matching stage. Each variant has its own characteristic:

1. For rough co-registration variants: (1) the ones based on the idea of matching DSMs generally lead to the most robust results; (2) the ones that match orthophotos could serve as alternates in rare scenarios of perfectly flat terrain where DSMs fail to provide useful information; (3) the others that match original image pairs often lead to less satisfactory results, but they are the only options suitable for terrestrial images.
2. For precise matching variants: (1) *Patch* is based on learned matching methods, it generally results in more matches as it is more invariant over time. (2) *Guided* is based on hand-crafted methods, it is more efficient in terms of the use of memory and CPU resources as it doesn't involve resampling patches, which is necessary for *Patch*.

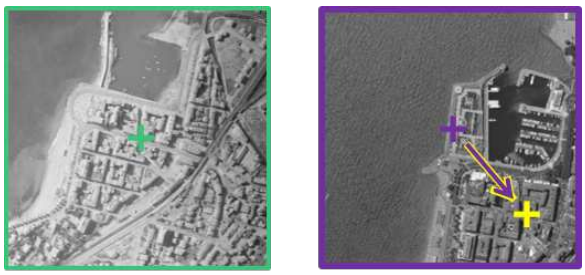
Our pipelines aim to unlock the potential of historical images for tracking environmental conditions. We are currently collaborating with several institutes to apply our pipelines in different applications, including:

1. Institut de Physique du Globe de Paris (IPGP) and Korea Institute of Geoscience and Mineral Resources (KIGAM) for analyzing deformations of the earth crust to understand the seismic events.
2. National Research Council, Research Institute for Hydrogeological Protection (CNR-IRPI) for analyzing landslide evolution in Italy.



(a) Example of an inter-epoch image pair

(b) Rough co-registration



(c) Precise matching

Figure 2.4: Rough-to-precise strategy. (a) An example of an inter-epoch image pair to be matched. I^{e_1} and I^{e_2} represents images take at $epoch_1$ and $epoch_2$ individually. (b) Illustration of rough co-registration between I^{e_1} and I^{e_2}. As a result, I^{e_1} is roughly aligned with I^{e_2}. (c) Illustration of precise matching. For keypoints in I^{e_1} (green cross), a location is predicted in I^{e_2} (purple cross) based on rough co-registration, whose local neighborhood will be searched to find the precise match (yellow cross).

3. Department of Earth and Environmental Sciences in University of Pavia for analyzing badland evolution in Europe.

We also developed two thorough tutorials accompanied with test datasets to familiarize users with our pipelines implemented in MicMac[Zhang *et al.* 2021a] (more details are introduced in Appendix C):

1. Tutorial of matching aerial images [Zhang *et al.* 2021e]
2. Tutorial of matching mixed images (i.e., aerial and satellite images) [Zhang *et al.* 2021d]

Publications of the author:

1. M Santangelo, **L Zhang**, E Rupnik, M Pierrot-Deseilligny, M Cardinali. Landslide evolution pattern revealed by multi-temporal DSMs obtained from historical aerial images. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2022.
2. **L Zhang**, E Rupnik, M Pierrot-Deseilligny. Feature matching for multi-epoch historical aerial images. ISPRS Journal of Photogrammetry and Remote Sensing, 182, 176-189, 2021.
3. **L Zhang**, E Rupnik, M Pierrot-Deseilligny. Guided feature matching for multi-epoch historical image blocks pose estimation. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020.

We also provide video [Zhang *et al.* 2021f], slides [Zhang *et al.* 2021c] and project website [Zhang *et al.* 2021b] to improve the visibility of our work.

2.3 Organization of the thesis

This thesis presents fully automatic pipelines to match multi-epoch images. A brief presentation of the *state-of-the-art* is given in **Chapter** 3.

In **Chapter** 4, applications as well as 5 sets of representative datasets are introduced, which are latter used to test our pipelines.

In **Chapter** 5, six rough co-registration variants are elaborated to roughly align the whole block by building a globally consistent transformation model between different epochs.

In **Chapter** 6, two precise matching variants are introduced to get accurate matches under the guidance of roughly co-registered orientations and DSMs.

Finally, in **Chapter** 7 conclusion and perspective are given.

Chapter 3

Literature review

Contents

3.1 Local feature matching

Local feature refers to 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 different data storage types, descriptors can be divided into two categories: floating-point 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 cartography, 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 subsequently elaborate on the two categories of approaches.

3.1.1 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 is able to outperform machine learning alternatives when matching conditions are favorable. 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.

3.1.2 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. introduce 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] introduce 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 image. ASLFeat [Luo *et al.* 2020] improves shape-awareness and localization accuracy by applying light-weight yet effective modifications on 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 matches and reject non-matchable points, leading to robust matching results even in challenging situations.

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 benchmarks, 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 matches for image pairs where SIFT already provides enough matches does not necessarily result in more accurate or complete reconstructions [Schonberger *et al.* 2017]. Jin et al. [Jin *et al.* 2020] introduce 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].

3.2 Robust matching

The goal of robust matching is to tell apart true matches (i.e., inliers) from false matches (i.e., 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 matches, 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 matches 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 state 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 we impose stricter rules on the sets of points. Lastly, we eliminate the remaining false matches by looking at their cross-correlation.

3.3 Pose estimation

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] based on local feature matches. The accuracy of matches plays an important role throughout the SfM process, since small inaccuracies in their locations can result in large errors in the estimated poses. Good matches will lead to better results on image orientation, camera calibration and 3D reconstruction [Lindenberger *et al.* 2021], [Sarlin *et al.* 2021], [Truong Giang *et al.* 2018].

Unlike in modern images where the image coordinate system overlaps with the 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 SfM procedure [McGlone 2013], which puts higher demands on the matches. [Giordano *et al.* 2018] demonstrates the importance of good matches in estimating the camera calibration and its great impact on the planimetric and altimetric accuracies of the resulted DSM. Systematic errors expressed as *dome* effect (i.e., a vertical doming of the surface) could appear in the DSMs when camera models are poorly estimated (i.e., inaccurately estimated lens distortion parameters) [Wackrow *et al.* 2008], [James & Robson 2014], which restricts the wider use

of DSMs.

3.4 Historical image processing

Compared to modern digital images, historical images are accompanied with particular characteristics such as poor radiometric quality and deformation during scanning. Therefore, aligning multi-epoch historical images by directly applying *state-of-the-art* feature matching methods often leads to unsatisfactory results. In Figure 2.1 we showed an example where SIFT and SuperGlue failed to recover good matches 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 larger than 45° 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 matches by processing different epochs separately followed by an inter-epoch co-registration relying on 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] and [Mölg & Bolch 2017]. Manually measuring GCPs are laboursome and tedious. Furthermore, it is difficult to find salient points that are stable over time.

Certain attempts were made to extract inter-epoch matches. Giordano et al. [Giordano *et al.* 2018] extract feature matches 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 2. Feurer et al. [Feurer & Vinatier 2018] joins multi-epoch images in a single SfM block based on SIFT-like algorithm by making the assumption that a sufficient number of feature points remain invariant across each time period. Their methods are widely used in the subsequent works [Filhol *et al.* 2019], [Cook & Dietze 2019], [Parente *et al.* 2021] and [Blanch *et al.* 2021]. It remains questionable whether the method is capable of handling drastic scene changes. Additionally, a stream of previous works focus 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.

In this work, we propose a rough-to-precise strategy to recover inter-epoch matches, without requiring any GCPs or auxiliary data.

Chapter 4

Applications and Datasets

Our pipelines open the possibility for millions of historical images to come into play in diverse applications of geoscience, including but not limited to disaster analysis, eco-environmental monitoring, urban expansion and so on. In order to reveal the potential of our pipelines, we choose five representative datasets for experiments:

1. **Fréjus**: city zone in France which exhibited notable urban expansion between 1954 and 2014. It contains 4 epochs of aerial images.
2. **Pezenas**: mountain area in France mainly covered with vegetation that demonstrated large-scale landuse changes between 1971 and 2015. Four epochs are available, including 3 aerial epochs and a satellite epoch.
3. **Kobe**: dataset over an island in Japan that chronicled an earthquake with 2 sets of aerial images taken before and after the event (i.e., 1991 and 1995).
4. **Alberona**: rural area in Italy that witnessed a landslide with 2 epochs taken before and after the event (i.e., 1954 and 2003).
5. **Hofsjökull**: ice cap area in Iceland with only one aerial epoch. Instead of aiming at inter-epoch images alignment, in this dataset we are interested in studying extremely challenging case in matching intra-epoch images for glacier monitoring.

The datasets contain both aerial and satellite images, details of the former are listed in Table 4.1 and 4.2, while that of the latter are displayed in Table 4.3. All the historical aerial images are scanned from films followed by resampling to the geometry of the fiducial marks. The resampling procedure is illustrated in Figure 4.1.
The images of each dataset are demonstrated in Figure 4.2, 4.4, 4.5, 4.6 and 4.7 individually.

Fréjus. The dataset Fréjus is mainly covered with buildings along with scattered farmlands, except a half-moon-shaped bay located in south. It is a 15 km^2 rectangular area located in Fréjus, a commune in southeastern France. We have four sets of aerial images acquired in 1954, 1966, 1970 and 2014. The epoch 2014 was acquired with the Institut national de l'information géographique et forestière (IGN)'s digital metric camera [Souchon *et al.* 2010], its orientations are both in global

(a) Original image and the fiducial mark (b) Resampled image

Figure 4.1: Illustration of resampling historical image to the geometry of the fiducial marks. (a) Original image with enlarged detail of the fiducial mark on the bottom right corner displayed. (b) Image resampled based on the 4 fiducial marks in the corner.

reference frame and precise. Therefore it is treated as Ground Truth (GT) during our processing (in other words, their parameters will be fixed during the Block Bundle Adjustment (BBA)). The area exhibits drastic scene changes in the 60-year period, as can be seen in Figure 4.3, where evolution of a subregion is displayed.

Pezenas. The dataset Pezenas is mainly covered with vegetation and several sparsely populated urban zones. It is a 420 km^2 rectangular area located in Pezenas in the Occitanie region in southern France. We have at our disposal three sets of aerial images acquired in 1971, 1981 and 2015, and one set of satellite images acquired in 2014. Both the epoch 2014 and 2015 are treated as GT. In this dataset we are interested in matching historical epochs (1971 and 1981) with aerial GT and satellite GT individually. The area exhibits changes in scene appearance in the 44-year period.

Kobe. The dataset Kobe witnessed the well-known Kobe earthquake in January 1995. It is a 90 km^2 area of irregular shape located in the north of Awaji Island, Japan. We have two sets of aerial images: pre-event acquired in 1991 and post-event acquired in 1995. It is mainly covered with mountain area and narrow urban zones along the sea. There are neither GT epochs nor GCPs, therefore we measured 2 points on Google map to scale the result to metric units. In this dataset we are interested in localizing the earthquake fault.

Alberona. The dataset Alberona is characterized by the diffuse presence of clay rich lithologies, with the wide presence of slow moving landslides. It is a 90 km^2 rectangular area located in southern Italy, near the village of Alberona (Puglia region). It is a rural, poorly inhabited, mainly agricultural and wooded area. A slow moving slide-earthflow has been detected there since the 1950s. We have two sets of aerial images: pre-event acquired in 1954 and post-event acquired in 2003. Images were scanned with non photogrammetric scanner with 800 dpi. The films were poorly preserved before scanning, which present some scratches and dust, a typical feature for images that were not preserved for photogrammetric purposes. There are no GT epochs but 7 GCPs which could be used to move the results from relative coordinate system to absolute one. In this dataset we are interested in localizing the landslide.

Hofsjökull. The dataset Hofsjökull is a snow-covered area located in Hofsjökull in central Iceland. Unlike other datasets described previously, Hofsjökull consists of only one epoch, as in this dataset we are only interested in matching challenging intra-epoch image pairs. It contains several archival aerial images acquired in the year 1960, provided by the National Survey of Iceland. They were scanned with a photogrammetric scanner Wehrli RM-6, in 16micron/px and 12 bit, in order to digitize as much information as possible appearing in the films. In Figure 4.7 we displayed 6 consecutive images in the same flight strip, with snow-covered area gradually expanding. We are interested in matching the most challenging image pair (i.e., image 5 and 6, as they are fully snow-covered with very limited context), whose overlapping zone is labeled with red rectangles.

	Epoch	Image type	PixSz [μm]	Focal length [pix]	Focal length [mm]	Physical ImgSz [mm]	Digital ImgSz [pix]
Fréjus	1954	Historical	21.17	23350	502	300 × 300	13932 × 13902
	1966	Historical	21.17	10230	210	180 × 180	8758 × 8759
	1970	Historical	21.17	10230	210	180 × 180	8766 × 8763
	2014	Modern	6.8	18281	124.3	99.28 × 72.42	14600 × 10650
Pezenas	1971	Historical	21.17	7600	160	230 × 230	10600 × 10600
	1981	Historical	21.17	7600	160	230 × 230	10600 × 10600
	2015	Modern	6.8	9967.5	67.8	47 × 35	6950 × 5175
			6.8	9204.5	62.6	50 × 36	7325 × 5350
Kobe	1991	Historical	20	7662	152.66	230 × 230	10600 × 10600
	1995	Historical	20	7662	152.66	230 × 230	10600 × 10600
Alberona	1954	Historical	31.75	4760	153.0	230 × 230	7113 × 7109
	2003	Historical	31.75	4650	152.8	230 × 230	6689 × 7065
Hofsjökull	1960	Historical	16	9656	154.49	230 × 230	14014 × 14009

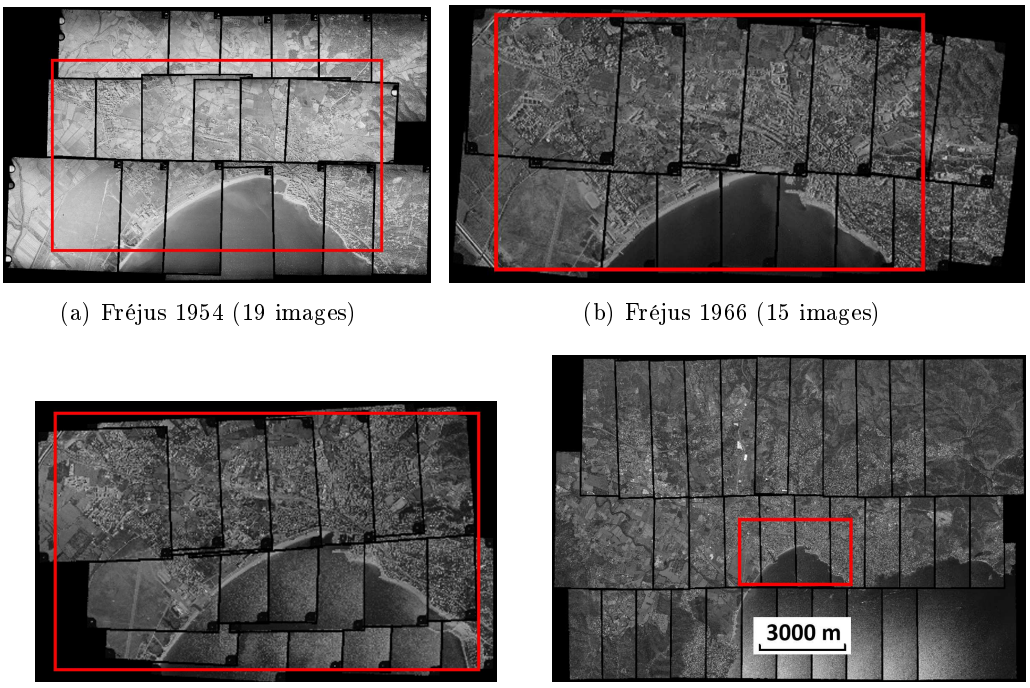
Table 4.1: Aerial dataset details of Fréjus, Pezenas, Kobe, Alberona and Hofsjökull. The 2015 acquisition of Pezenas is obtained with two sets of camera. PixSz means pixel size, ImgSz stands for image size. Digital image size of historical epoch is based on images resampled to the geometry of the fiducial marks.

	Epoch	GSD [m]	H [m]	Scale	Forward overlap	Side overlap	Image number	Flightline
Fréjus	1954	0.11	2530	1:5000	60%	20%	19	West-Est
	1966	0.17	1780	1:8000	60%	30%	15	West-Est
	1970	0.17	1770	1:8000	60%	30%	19	West-Est
	2014	0.35	6500	1:5000	60%	30%	33	West-Est
Pezenas	1971	0.32	2400	1:1500	60%	20%	57	West-Est
	1981	0.59	4500	1:3000	60%	20%	27	West-Est
	2015	0.46	4600	1:7000	60%	50%	308	West-Est
		0.5	4600	1:7000	60%	50%	74	West-Est
Kobe	1991	0.5	3800	1:25000	65%	35%	15	Northeast-Southwest
	1995	0.18	1100	1:7000	65%	65%	83	Northeast-Southwest
Alberona	1954	1.0	6000	1:4000	65%	/	3	North-South
	2003	0.85	4850	1:3000	65%	30%	7	West-Est
Hofsjökull	1960	0.57	5480	1:3500	60%	/	6	North-South

Table 4.2: Continuation of Table 4.1. GSD is the ground sampling distance, H is the flying height.

	Master image	Secondary image
Constellation	Pleiades	Pleiades
GSD [m]	0.5	0.5
Acquired date	12/06/2014	12/06/2014
Number of lines	38468	37710
Number of pixels per line	34108	33392
Cloud cover	3.9%	4.0%
Snow cover	0%	0%

Table 4.3: Satellite dataset details of Pezenas. It consists of 2 images, which is indicated as master and secondary image individually. GSD means the ground sampling distance.



(a) Fréjus 1954 (19 images)

(b) Fréjus 1966 (15 images)

(c) Fréjus 1970 (19 images)

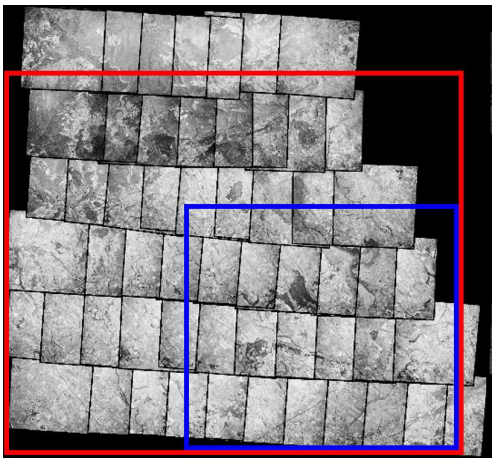
(d) Fréjus 2014 (33 images)

Figure 4.2: Images demonstration of 4 aerial epochs in **Fréjus** (i.e., 1954, 1966, 1970 and 2014), image number of each epoch is displayed in the parenthesis of each sub headline. The overlapping zone between all the epochs is indicated with red rectangles. Graphic scale is demonstrated on epoch 2014 in (d).

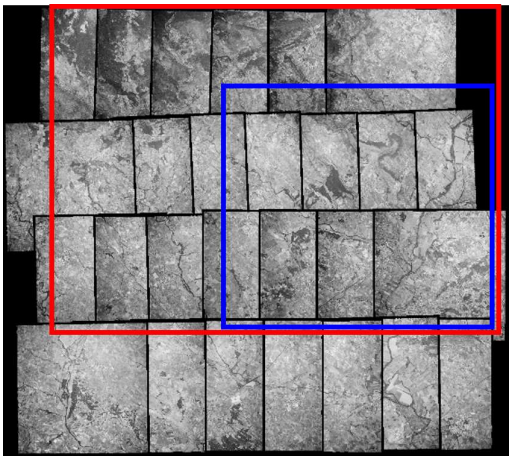
(a) Subregion of Fréjus 1954

(b) Subregion of Fréjus 2014

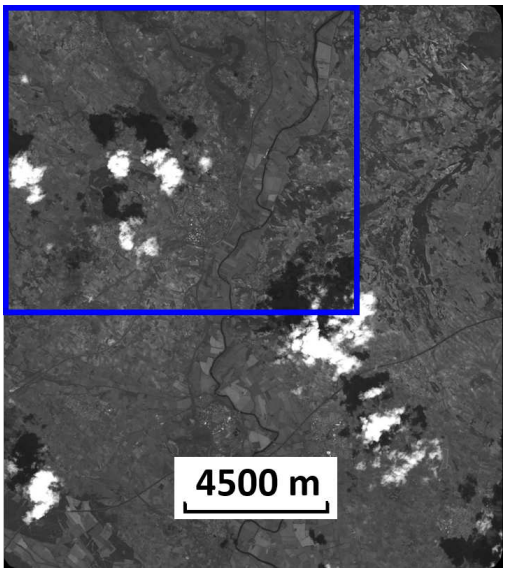
Figure 4.3: Evolution of a subregion in Fréjus.



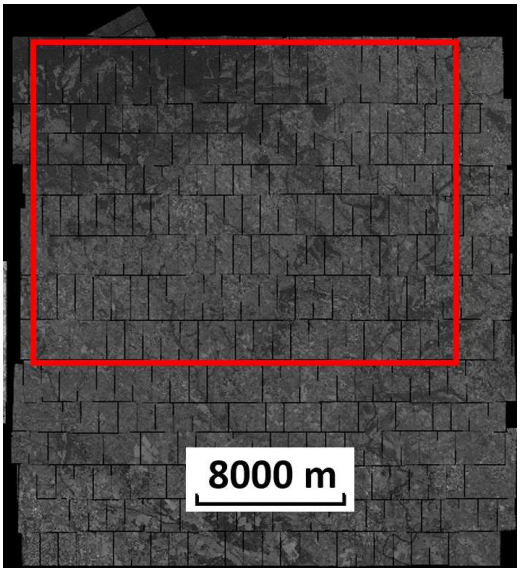
(a) Pezenas 1971 (57 images)



(b) Pezenas 1981 (27 images)

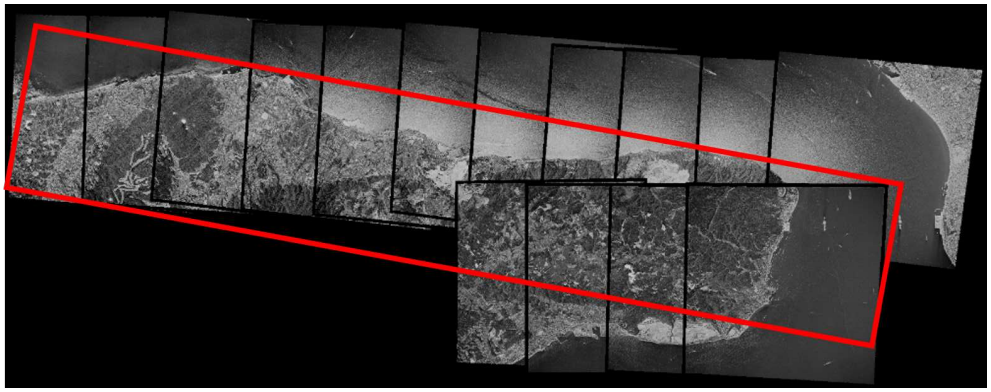


(c) Pezenas 2014 (2 satellite images)

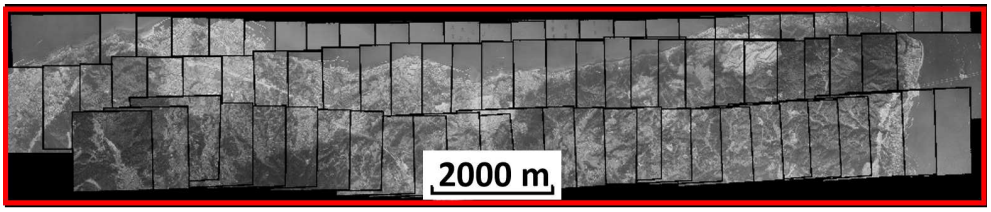


(d) Pezenas 2015 (382 images)

Figure 4.4: Images demonstration of 3 aerial epochs (i.e., 1971, 1981 and 2015) as well as satellite epoch (i.e., 2014) in **Pezenas**, image number of each epoch is displayed in the parenthesis of each sub headline. There are 2 historical aerial epochs (1971 and 1981) and 2 GT epochs (2014 the satellite epoch and 2015 the aerial epoch) in this dataset. The overlapping zone between the historical epochs and the 2014 satellite epoch is indicated with blue rectangles, while that between historical epochs and the 2015 aerial epoch is in red rectangles. Graphic scales are demonstrated on epoch 2014 and 2015 in (c) and (d).

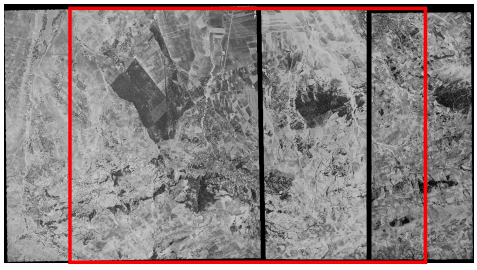


(a) Kobe 1991 (15 images)

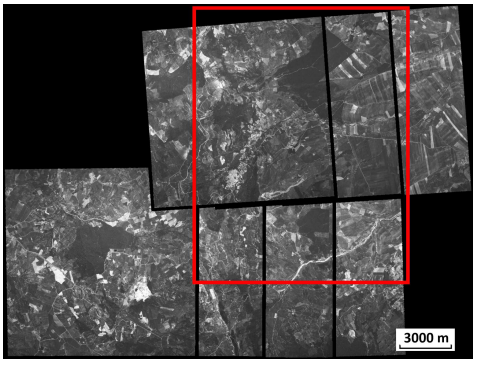


(b) Kobe 1995 (83 images)

Figure 4.5: Images demonstration of 2 aerial epochs (i.e., 1991 and 1995) in **Kobe**, image number of each epoch is displayed in the parenthesis of each sub headline. The overlapping zone between different epochs is indicated with red rectangles. Graphic scale is demonstrated on epoch 1995 in (b).

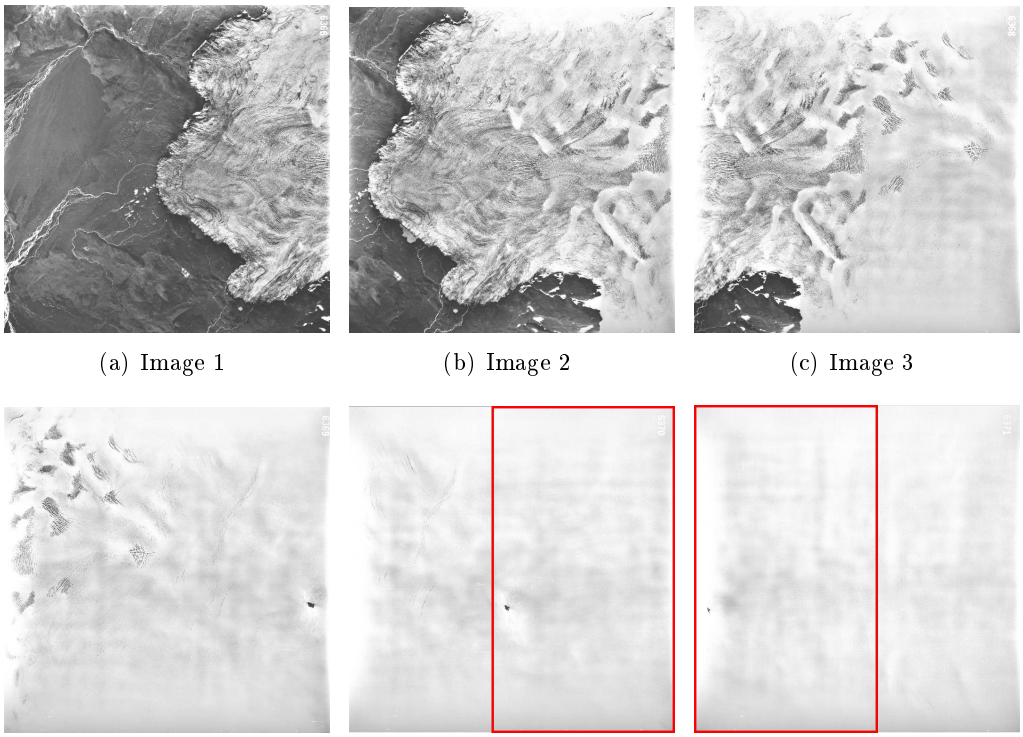


(a) Alberona 1954 (3 images)



(b) Alberona 2003 (7 images)

Figure 4.6: Images demonstration of 2 aerial epochs (i.e., 1954 and 2003) in **Alberona**, image number of each epoch is displayed in the parenthesis of each sub headline. The overlapping zone between different epochs is indicated with red rectangles. Graphic scale is demonstrated on epoch 2003 in (b).



(a) Image 1 (b) Image 2 (c) Image 3

(d) Image 4 (e) Image 5 (f) Image 6

Figure 4.7: Images demonstration of epoch 1960 in **Hofsjökull**. The overlapping zone of the most challenging image pair (i.e., image 5 and 6) is labeled with red rectangles.

Chapter 5

Rough co-registration

Contents

5.1 Introduction

5.1.1 Motivation and objective

The goal of rough co-registration is to perform rough alignment between image pairs, so that they can be later used to guide the precise matching by narrowing down the search space in 2D image geometry. It plays a fundamental role as wrongly aligned results would lead to deviation from the right search space. Therefore robustness is the most critical target in rough co-registration. In the meantime, it doesn't need to be very accurate as the precision would be improved in the precise matching part via searching local neighborhood. Therefore, we can reasonably sacrifice precision when necessary to trade for robustness.

The task of rough alignment could be accomplished by matching every possible inter-epoch image pair, followed by a RANSAC routine to recover the best transformation model between each image pair. However, inter-epoch image pairs often demonstrate different appearances due to scene changes and heterogeneous acquisition conditions, which often leads to limited inlier ratio, in other words,

failure of the later RANSAC procedure. Therefore, instead of aligning a group of inter-epoch image pairs separately, we come up with an idea to improve robustness by aligning the whole block integrally to build a globally consistent transformation model. Afterwards, we can move different epochs to the same coordinate frame, providing orientations and DSMs in the same geographic system for later processing (i.e., precise matching and BBA). Besides, we take advantage of 3D geometry which could be easily obtained within each epoch to boost the inter-epoch matching performance.
As a consequence, we come up with 2 strategies for multi-epoch rough co-registration: (1) *ImgPairs* (i.e., matching image pairs) and (2) *Ortho* or *DSM* (i.e., matching orthophotos or DSMs). The former strategy comes up first, but its performance is less satisfactory. This motivated us to conceive and explore the latter strategy. Even though *ImgPairs* is generally outperformed by *Ortho* or *DSM*, it still has its own strengths in certain cases such as viability for terrestrial images.

5.1.2 Contributions

Our main contribution is complete and fully automated pipelines for rough co-registration between inter-epoch image pairs. The pipelines are composed of the following key ingredients:

1. improving matching robustness by building globally consistent transformation model;
2. introducing the idea of matching DSMs to obtain robust matches even under drastic scene changes, as the 3D landscape often stays globally stable over time;
3. introducing RANSAC in 3D for *ImgPairs*: each three matches projected to DSM serve to compute a 3D Helmert transformation model between epochs, and most importantly provide a 2D constraint on all image pairs;
4. introducing *4 rotation hypotheses* to make SuperGlue suitable for matching images with rotation larger than 45°;
5. introducing *one-to-many tiling scheme* to scale-up the deep learning methods for feature matching;
6. improving the performance of matching inter-epoch images with SIFT by (1) using downsampled images and (2) skipping ratio test.

5.2 Methodology

To roughly co-register different epochs to the same frame, we propose 2 strategies:

1. *ImgPairs*: matching each inter-epoch image pairs followed by global filtering over the whole block;

2. *Ortho* or *DSM*: generating global image for each epoch (i.e., orthophotos or DSMs) and performing matching only once.

Please notice our pipelines are generic, different feature matching methods can be readily applied. At present we adopt either SIFT or SuperGlue in our pipeline as they are currently the *state-of-the-art*, but they can be replaced when better matching methods arise in the future.
Our pipelines are able to match both aerial and satellite images. For aerial images, they are supposed to be accompanied with focal lengths and physical sensor sizes, which are usually available, as mentioned in Section 2.

We adopt the following naming conventions: (1) I^{e_1} and I^{e_2}: images acquired in $epoch_1$ and $epoch_2$; (2) O^{e_1} and O^{e_2}: orientations of $epoch_1$ and $epoch_2$; (3) Op^{e_1} and Op^{e_2}: orthophotos of $epoch_1$ and $epoch_2$; (4) D^{e_1} and D^{e_2}: DSMs of $epoch_1$ and $epoch_2$.
Prior to inter-epoch rough co-registration, we process each epoch individually to recover the relative orientations and DSM within the same epoch. It is a standard photogrammetry or SfM pipeline and can be accomplished with lots of solutions (e.g., MicMac [Pierrot-Deseilligny & Cléry 2012], COLMAP [Schonberger & Frahm 2016], OpenMVG [Moulon *et al.* 2016], Theia [Sweeney 2015], etc.). The one used in our experiment is MicMac. It is performed within each $epoch_i$ individually as follows:

1. Extract intra-epoch matches between images I^{e_i} with SIFT [Lowe 2004];

2. Based on the sequential SfM to compute interior and relative orientations ($O^{e_i}_{ini}$) for aerial images, or to refine the Rational Polynomial Coefficient (RPC) for satellite images;

3. Based on image orientations $O^{e_i}_{ini}$, perform semi-global dense matching [Pierrot-Deseilligny & Paparoditis 2006] between images I^{e_i} to get DSM ($D^{e_i}_{ini}$) in their arbitrary coordinate frames.

4. Orthorectify the images to get orthophotos ($Op^{e_i}_{ini}$) if strategy *Ortho* is applied.

5.2.1 Strategy 1: Matching image pairs (*ImgPairs*)

The strategy *ImgPairs* is the first attempt we made for rough co-registration. The workflow is displayed in Figure 5.1(a). For the sake of simplicity, only 2 epochs are present in our processing flows, however, it can be easily extended to more epochs. The matching procedure (i.e., the magenta rectangle in Figure 5.1(a)) with SIFT or SuperGlue is slightly different, so we elaborate both in Figure 5.1(b) and (c) individually for better understanding. We introduce *4 rotation hypotheses* when feature matching method that is not invariant to rotations lager than 45° (e.g.,

SuperGlue) is applied.

Four rotation hypotheses. It normalizes rotation to achieve invariance, which is similar to ASIFT [Morel & Yu 2009] by exploring the space of possible deformation, but adapted as it only explores 2D rotation.
The *4 rotation hypotheses* works as follows (cf. Figure 5.1(d)):

1. Rotate the secondary image by 90° four times;
2. Match each rotated image with the master image;
3. Keep the rotation hypothesis with the largest number of matches.

Workflow of *ImgPairs*. Assuming the numbers of images in $epoch_1$ and $epoch_2$ are P and Q individually, the strategy *ImgPairs* works as follows:

1. Match P×Q inter-epoch image pairs respectively, giving rise to P×Q sets of matches $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$ ($\mathbf{K}^{e_i}$ represents keypoints in image I^{e_i}).
2. Sample matches $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$ iteratively to compute the 2D similarity transformation RANSAC model:

$$\begin{bmatrix} K_x^{e_2} \\ K_y^{e_2} \end{bmatrix} = \lambda \cdot \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \cdot \begin{bmatrix} K_x^{e_1} \\ K_y^{e_1} \end{bmatrix} + \begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix}. \tag{5.1}$$

   where λ is the scale factor, θ is the in-plane rotation angle and $\begin{bmatrix} \Delta_x, \Delta_y \end{bmatrix}^T$ is the translation vector. Matches within T_r of its predicted position (i.e., $|\mathbf{K}^{e_2} - (\lambda \cdot \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \cdot \mathbf{K}^{e_1} + \Delta)| < T_r$) are considered as inliers. The resulted inliers are referred to as $M(\widetilde{\mathbf{K}}^{e_1}, \widetilde{\mathbf{K}}^{e_2})$ ($\widetilde{\mathbf{K}}^{e_i}$ represents keypoints in image I^{e_i}).
   This step aims to reduce matches into a reasonable number when SIFT is applied, otherwise the subsequent global filtering would become prohibitive. It is not necessary if SuperGlue is applied as it simultaneously performs filtering during the matching procedure.
3. Project $\widetilde{\mathbf{K}}^{e_i}$ from SIFT or $\mathbf{K}^{e_i}$ from SuperGlue onto DSM $D_{ini}^{e_i}$ with the help of orientations $O_{ini}^{e_i}$, resulting in 3D points $\mathbf{KG}^{e_i}$.
4. Sample matches $M(\mathbf{KG}^{e_1}, \mathbf{KG}^{e_2})$ iteratively to compute the 3D Helmert transformation RANSAC model:

$$\begin{bmatrix} KG_x^{e_2} \\ KG_y^{e_2} \\ KG_z^{e_2} \end{bmatrix} = \lambda \cdot \mathbf{R} \cdot \begin{bmatrix} KG_x^{e_1} \\ KG_y^{e_1} \\ KG_z^{e_1} \end{bmatrix} + \begin{bmatrix} \Delta_x \\ \Delta_y \\ \Delta_z \end{bmatrix}. \tag{5.2}$$

where λ is the scale factor, $\mathbf{R}$ is the rotation matrix and $\left[\ \Delta_x, \Delta_y, \Delta_z\ \right]^T$ is the translation vector. Matches within T_r of its predicted position (i.e., $|\mathbf{KG}^{e_2} - (\lambda \cdot \mathbf{R} \cdot \mathbf{KG}^{e_1} + \Delta)| < T_r$) are considered as inliers.

This strategy is adapted from our early attempt to match different epochs by matching P×Q inter-epoch image pairs separately without the later step of building a globally consistent transformation model. In [Zhang *et al.* 2020] we accomplished it by estimating a 2D similarity model for each image pair and using it to guide precise matching. Obviously it is less robust as the P×Q 2D similarity models might not be consistent. However, the idea of using 2D similarity model to guide matching is more generic as it doesn't require initial orientations $O_{ini}^{e_i}$ and DSMs $D_{ini}^{e_i}$. It is a viable and maybe even the only possible approach when it is impossible to acquire orientations $O_{ini}^{e_i}$ and DSMs $D_{ini}^{e_i}$. An example is demonstrated in section 5.3.4.

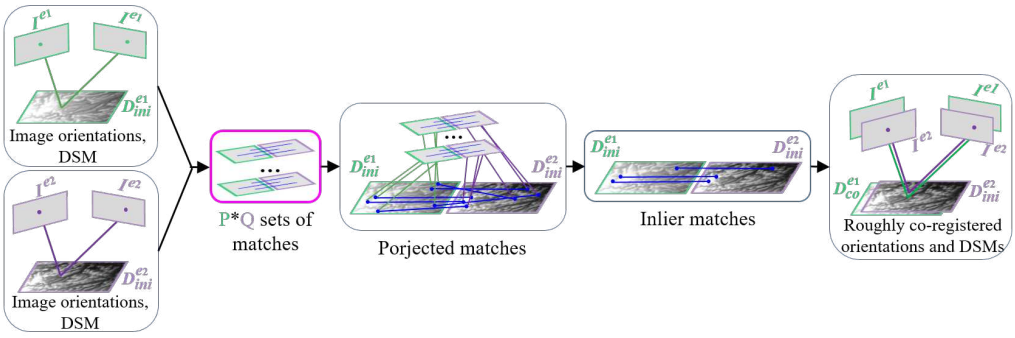
5.2.2 Strategy 2: Matching Orthophotos/DSMs (*Ortho* or *DSM*)

Another strategy is to match orthophotos or DSMs. The detailed workflows are displayed in Figure 5.2(a) and Figure 5.3(a) individually. Different than matching P×Q image pairs in strategy *ImgPairs*, we only need to match one pair of DSMs/orthophotos. The DSMs are typically floating-point images, in order to apply feature matching methods directly on them without adjusting the implementation of SIFT or SuperGlue, we further describe the conversion of DSM to a gray-scale raster. Additionally, we propose a *one-to-many tiling scheme* to maximize the performance of feature matching with learned methods.

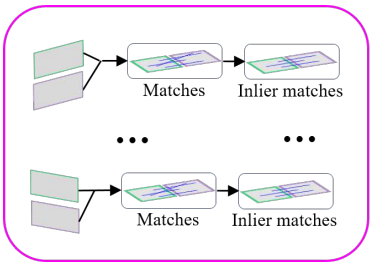
Matching DSMs/orthophotos has the following merits: (1) redundancy caused by the forward and side overlapping areas is removed; (2) it enables a follow-up search for globally consistent inliers directly without the need to project matches onto ground; (3) it decreases the combinatorial complexity caused by rotation ambiguity of P×Q images; (4) when matching DSMs, robust matches can be expected even under extreme scene changes, as 3D landscape generally provide stable information over time.

Converting DSM to grayscale image. DSMs are 2.5D rasters recorded in floating-point format. It is complicated to apply feature matching methods on them directly, as most of the methods are implemented for RGB images. Therefore, we convert DSM beforehand to [0–255] range grayscale images as follows:

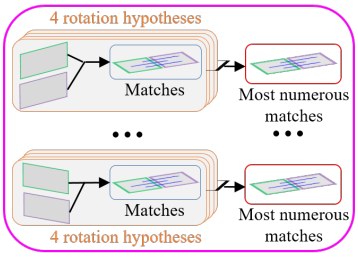
1. Calculate the standard deviation of the DSM elevation;
2. Pixels with elevations larger than double the standard deviation are considered outliers and therefore ignored;
3. Transform the inlier pixels to the range of [0-255], resulting in grayscale image.



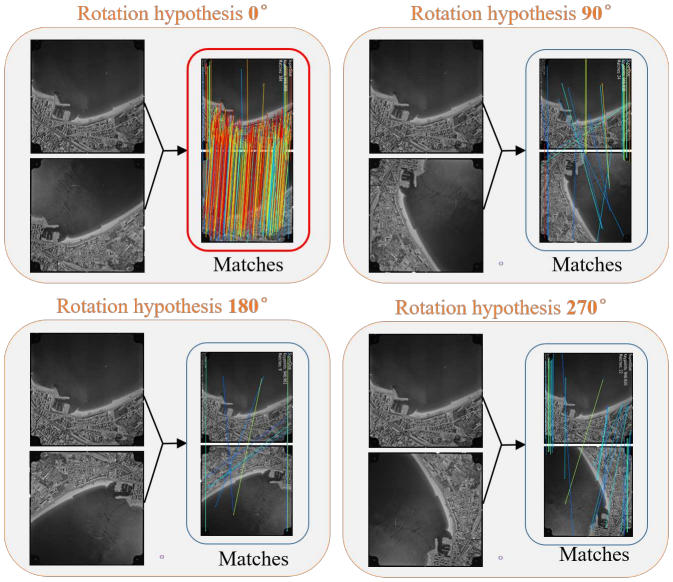
(a) Workflow of *ImgPairs*



(b) Match image pairs with SIFT

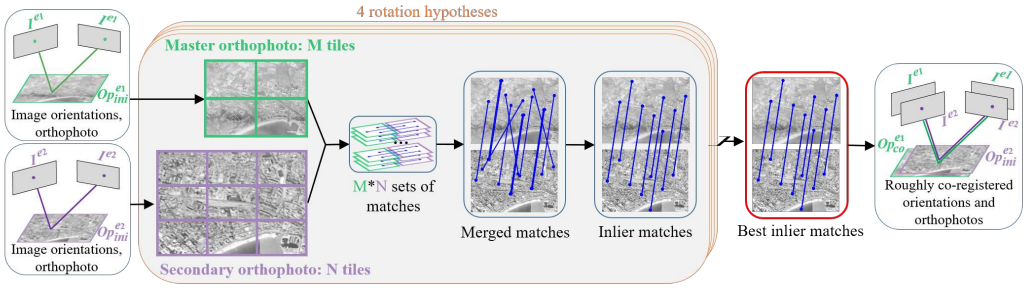


(c) Match image pairs with SuperGlue

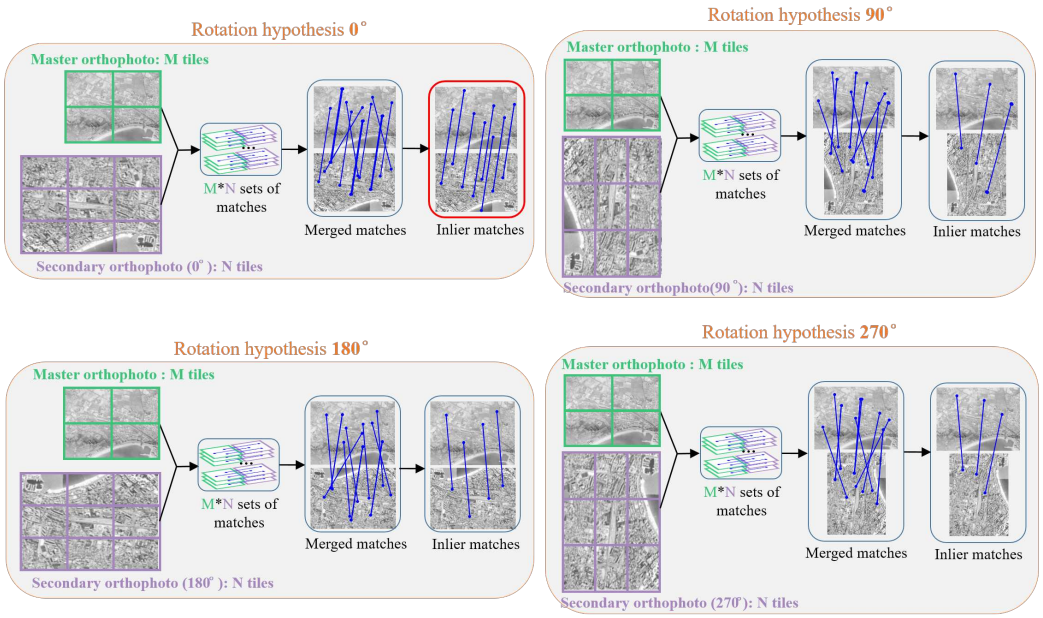


(d) Four rotation hypotheses

Figure 5.1: Rough co-registration by matching image pairs (i.e., *ImgPairs*). (a) Whole workflow. Each inter-epoch image pair is matched individually, followed by projecting the matches onto ground to find globally consistent inliers. (b) Match image pairs with SIFT, which involves matching followed by 2D similarity RANSAC to find inliers. (c) Match image pairs with SuperGlue, which involves matching combined with *four rotation hypotheses*. (d) Four rotation hypotheses. We rotate the secondary image by 90 ° four times to match with master image and keep the best one with the largest number of matches (red rectangle).

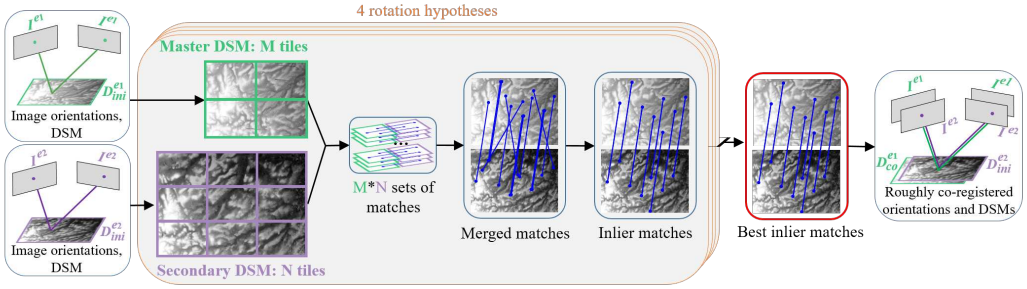


(a) Workflow of *Ortho*

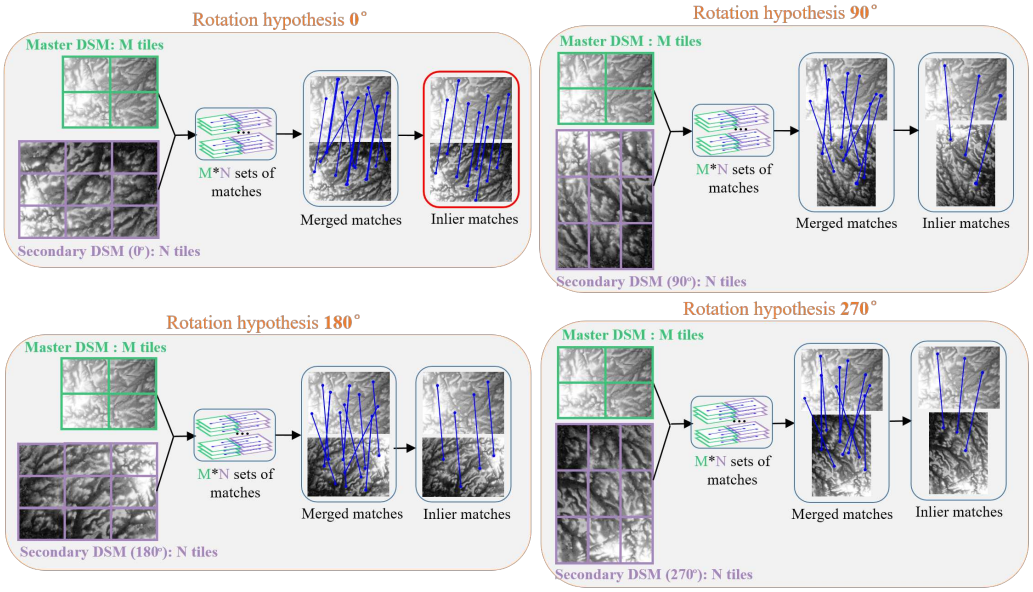


(b) Four rotation hypotheses combined with *one-to-many tiling scheme*

Figure 5.2: Rough co-registration by matching orthophotos (i.e., *Ortho*). (a) Whole workflow. Orthophotos are matched, followed by projecting the inlier matches onto ground to build 3D Helmert transformation model. (b) Four rotation hypotheses combined with *one-to-many tiling scheme*. We rotate the secondary orthophoto by 90 ° four times to match with master orthophoto and keep the best one with the largest number of RANSAC inliers (red rectangle). *One-to-many tiling scheme* is applied during each hypothesis, with both orthophotos croped into tiles followed by matching all the tile pairs and merging the matches.



(a) Workflow of *DSM*



(b) Four rotation hypotheses combined with tiling scheme

Figure 5.3: Rough co-registration by matching DSMs (i.e., *DSM*). (a) Whole workflow. DSMs are matched, followed by projecting the inlier matches onto ground to build 3D Helmert transformation model. (b) Four rotation hypotheses combined with tiling scheme. We rotate the secondary DSM by 90 ° four times to match with master DSM and keep the best one with the largest number of RANSAC inliers (red rectangle). Tiling scheme is applied during each hypothesis, with both DSMs cropped into tiles followed by matching all the tile pairs and merging the matches.

4. Apply Wallis filter on the grayscale image to get rid of uneven illumination, resulting in more informative image.

One-to-many tiling scheme. DSMs/orthophotos are usually large images as they have larger extent than original images. Learned matching methods often underperform on large images as they are either trained on small images in order to run real-time or with limited spatial resolution of CNN feature maps. To make up for the deficiency, we propose an *one-to-many tiling scheme*, which is performed as follows (Figure 5.2(a) and Figure 5.3(a)):

1. Crop master and secondary images into M and N tiles of size $SZ_{one-to-many}$;
2. Apply matching on M×N tile pairs respectively;
3. Merge the matches and perform RANSAC based on 2D similarity transformation to remove outliers.

The *one-to-many tiling scheme* can be combined with *4 rotation hypotheses*, as shown in Figure 5.2(b) and Figure 5.3(b).

Workflow of *Ortho* and *DSM*. The matching DSMs/orthophotos strategy works as follows:

1. Transform DSMs to grayscale images if the strategy *DSM* is applied.
2. Match DSMs/orthophotos, giving rise to one set of matches $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$ ($\mathbf{K}^{e_i}$ represents keypoints in DSM D^{e_i} or orthophoto Op^{e_i}).
3. Sample matches $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$ iteratively to compute the 2D similarity transformation RANSAC model:

$$\begin{bmatrix} K_x^{e_2} \\ K_y^{e_2} \end{bmatrix} = \lambda \cdot \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \cdot \begin{bmatrix} K_x^{e_1} \\ K_y^{e_1} \end{bmatrix} + \begin{bmatrix} \Delta_x \\ \Delta_y \end{bmatrix}. \tag{5.3}$$

   where λ is the scale factor, θ is the in-plane rotation angle and $\begin{bmatrix} \Delta_x, \Delta_y \end{bmatrix}^T$ is the translation vector. Matches within T_r of its predicted position (i.e., $|\mathbf{K}^{e_2} - (\lambda \cdot \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \cdot \mathbf{K}^{e_1} + \Delta)| < T_r$) are considered as inliers.
4. Project the inlier matches onto DSM $D_{ini}^{e_i}$ to fit the best 3D Helmert transformation parameters.

5.3 Experiments

As described in the previous section, we provide 3 pipelines out of 2 strategies to perform rough co-registration:

1. *ImgPairs*: match image pairs;
2. *Ortho*: match orthophotos;
3. *DSM*: match DSMs.

For each pipeline, we employ either SIFT or SuperGlue as the feature matching method, giving rise to 6 variants:

1. $SIFT_{ImgPairs}$;
2. $SuperGlue_{ImgPairs}$;
3. $SIFT_{Ortho}$;
4. $SuperGlue_{Ortho}$;
5. $SIFT_{DSM}$;
6. $SuperGlue_{DSM}$;

We test the 6 variants on all the multi-epoch datasets which are elaborated in Chapter 4: Fréjus, Pezenas, Kobe and Alberona, except that we skip the variants $SIFT_{ImgPairs}$ and $SuperGlue_{ImgPairs}$ for satellite images in Pezenas as there are only 2 images with the same extent.

Additionally, we provide experiments where we test the influence of the SIFT parameters (image downsampling factor, ratio test, RANSAC, etc.) in Section 5.3.2 and the effectiveness of *one-to-many tiling scheme* in Section 5.3.3. In Section 5.3.4 we demonstrate a real case study where the basic 2D similarity model outperforms the more sophisticated 3D Helmert transformation model. Finally, the 6 variants are compared in Section 5.3.5.

5.3.1 Implementation details

To improve efficiency, all input images are downsampled by a factor of 3 beforehand, except for dataset Alberona as it consists of very few images. To calculate the DSMs and orthophotos, we further downsample the images by a factor of 8, which amounts to a total downsampling factor of 24 with respect to the input images (total downsampling factor of 8 for Alberona). For example, the images in Fréjus 1970 are downsampled from [8766, 8763] to [365, 365] for calculating DSMs and orthophotos.

As the goal of rough co-registration is to get robust rather than precise matches, a low resolution DSM/orthophoto is good enough and keeps the computational cost low.
However, the downsampling factor of Fréjus 2014 is set to be 12 instead of 24, as Fréjus is mainly covered with buildings and the GSD of Fréjus 2014 is too limited to tell details from DSM with low resolution.
For each dataset, one epoch (generally the most recent epoch) is chosen as the reference epoch E_r, the others would be treated as free epoch E_f. The rough co-registration is applied between each free epoch E_f and the reference epoch E_r. As a result, all free epochs E_f would be moved to the frame of epoch E_r.

For the procedure of RANSAC to build (1) 3D Helmert transformation, we empirically set the number of iteration to 2000, and T_r to 50m; (2) 2D similarity transformation, we set the number of iteration to 1000, and T_r to 15 pixels. For the *one-to-many tiling scheme*, the tile size $SZ_{one-to-many}$ is set to be 1280×960 pixels to balance performance and efficiency. All the image/tile pairs entering SuperGlue are downsampled to 640×480 pixels, as it is the default parameter provided by the author and guarantees the best performance.

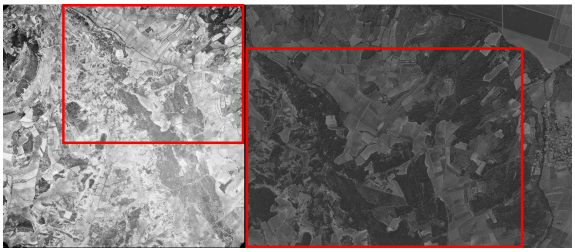
5.3.2 Comparison between $SIFT_{Adapted}$ and $SIFT_{Default}$

In this section, two different sets of SIFT parameters are compared, which are referred to as $SIFT_{Default}$ and $SIFT_{Adapted}$:

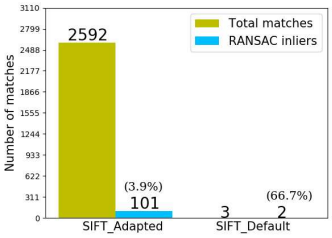
1. $SIFT_{Default}$: Extract SIFT keypoints on the original images, followed by mutual nearest neighbor matching combined with ratio test.
2. $SIFT_{Adapted}$: Downsample the input images with a factor of 3 and extract SIFT keypoints, match them by mutual nearest neighbor without ratio test, followed by applying RANSAC based on 2D similarity transformation model to remove outliers.

Results on matching image pairs. For pipeline $ImgPairs$, we choose a pair of images from dataset Pezenas consists of images taken at 1971 and 2015 individually. The results are displayed in Figure 5.4. As can be seen, $SIFT_{Adapted}$ recovers 101 good matches out of 2592 total matches, however, $SIFT_{Default}$ finds only 3 matches in total, even though 2 of them are correct matches, it is impossible for the RANSAC procedure to screen the correct ones.

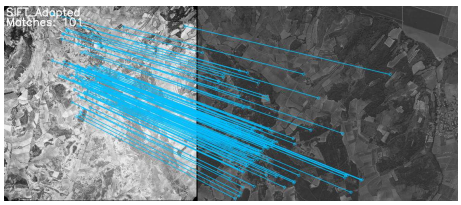
Results on matching orthophotos. For pipeline $Ortho$, we choose orthophotos from Pezenas 1981 and 2015 individually. The results are displayed in Figure 5.5. As can be seen, $SIFT_{Adapted}$ recovers 44 good matches out of 855 total matches, while $SIFT_{Default}$ finds 8 total matches which are all wrong.



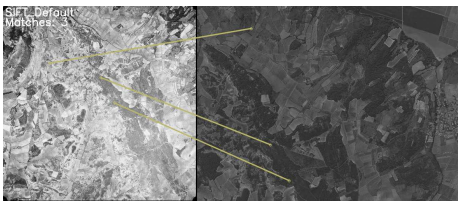
(a) Image pair



(b) Number of recovered matches

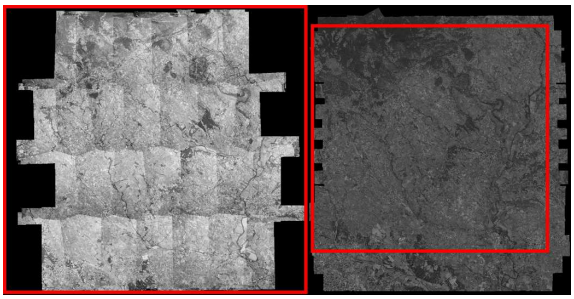


(c) $SIFT_{Adapted}^{RANSACInliers}$

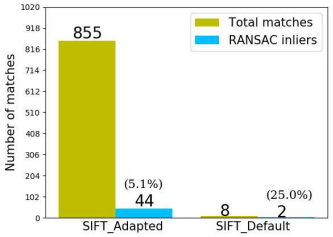


(d) $SIFT_{Default}^{TotalMatches}$

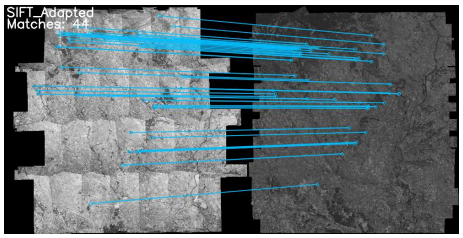
Figure 5.4: Comparison between $SIFT_{Adapted}$ and $SIFT_{Default}$ on a pair of images from Pezenas 1971 and Pezenas 2015 individually. (a) Image pair to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of total matches and RANSAC inliers of $SIFT_{Adapted}$ and $SIFT_{Default}$. (c) Visualization of RANSAC inliers based on $SIFT_{Adapted}$. (d)Visualization of total matches based on $SIFT_{Default}$.



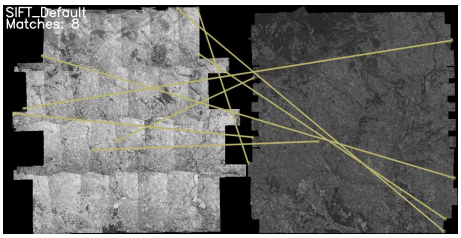
(a) Orthophotos



(b) Number of recovered matches



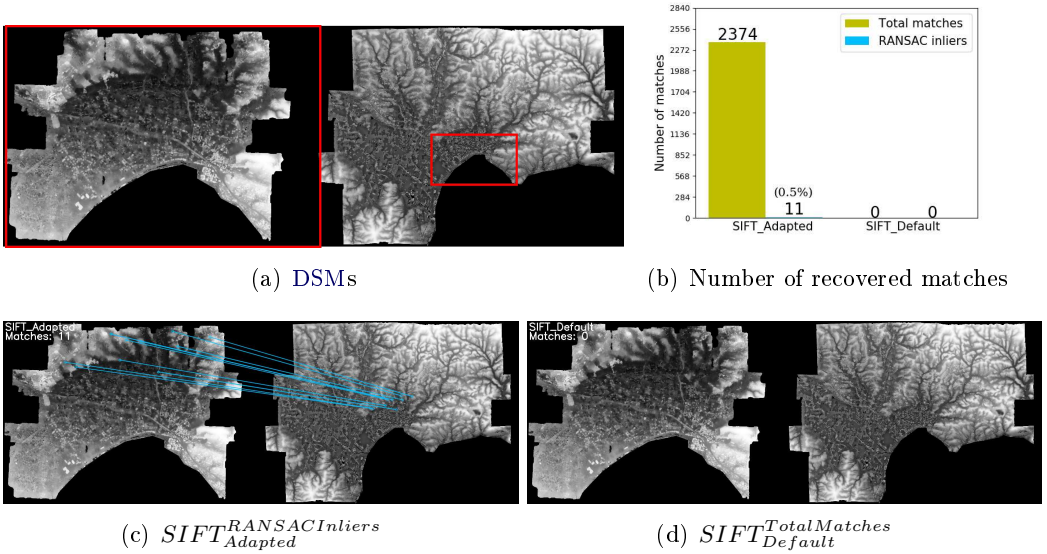
(c) $SIFT^{RANSACInliers}_{Adapted}$



(d) $SIFT^{TotalMatches}_{Default}$

Figure 5.5: Comparison between $SIFT_{Adapted}$ and $SIFT_{Default}$ on orthophotos from Pezenas 1981 and Pezenas 2015 individually. (a) Orthophotos to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of total matches and RANSAC inliers of $SIFT_{Adapted}$ and $SIFT_{Default}$. (c) Visualization of RANSAC inliers based on $SIFT_{Adapted}$. (d)Visualization of total matches based on $SIFT_{Default}$.

Results on matching DSMs. For pipeline DSM, we choose DSMs from Fréjus 1954 and 2014 individually. The results are displayed in Figure 5.6. As drastic scene changes are displayed in dataset Fréjus, $SIFT_{Default}$ fails to find any matches. $SIFT_{Adapted}$, however, recovers 11 good matches, even though the inlier ratio is dangerously low (i.e., 0.5%).



(a) DSMs (b) Number of recovered matches

(c) $SIFT_{Adapted}^{RANSACInliers}$ (d) $SIFT_{Default}^{TotalMatches}$

Figure 5.6: Comparison between $SIFT_{Adapted}$ and $SIFT_{Default}$ on DSMs from Fréjus 1954 and Fréjus 2014 individually. (a) DSMs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of total matches and RANSAC inliers of $SIFT_{Adapted}$ and $SIFT_{Default}$. (c) Visualization of RANSAC inliers based on $SIFT_{Adapted}$. (d)Visualization of total matches based on $SIFT_{Default}$.

In general, $SIFT_{Adapted}$ recovers enough good matches in all the 3 variants ($SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$), while $SIFT_{Default}$ fails. It is reasonable as inter-epoch images often look very different, $SIFT_{Default}$ generally recover very few matches. By downsampling the images, we are able to focus on the global outline of the scene to improve robustness. By relaxing the matching restriction of ratio test, right matches would be preserved while wrong matches would be removed in the subsequent RANSAC.

5.3.3 Comparison between $SuperGlue_{tiling}$ and $SuperGlue_{orig}$

In order to explore whether the *one-to-many tiling scheme* improves the performance of SuperGlue, we compare 2 sets of the results on matching multi-epoch orthophotos and DSMs with $SuperGlue_{tiling}$ and $SuperGlue_{orig}$. The former and latter stands

	orthophoto		DSM	
	Width [pix]	Height [pix]	Width [pix]	Height [pix]
E1970	899	618	3323	2394
E2014	1124	773	4154	2992

Table 5.1: Size of orthophotos and DSMs from Fréjus 1970 and 2014.

		orthophoto		DSM	
		Width	Height	Width	Height
E1970	$SuperGlue_{orig}$	1.4	1.3	1.8	1.6
	$SuperGlue_{tiling}$	2	2	2	2
E2014	$SuperGlue_{orig}$	5.2	5.0	6.5	6.2
	$SuperGlue_{tiling}$	2	2	2	2

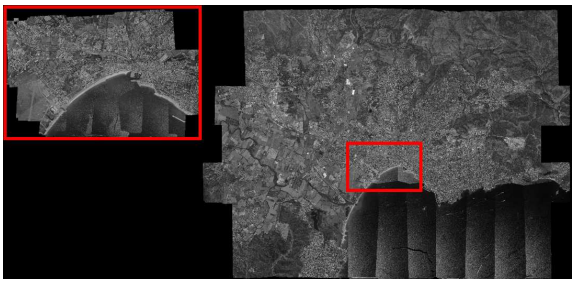
Table 5.2: Comparison of downsampling ratio between $SuperGlue_{tiling}$ and $SuperGlue_{orig}$ for both orthophotos and DSMs from Fréjus 1970 and 2014.

for SuperGlue combined with and without our *one-to-many tiling scheme*. We chose the orthophotos and DSMs from Fréjus 1970 and 2014 individually for testing.
The sizes of the orthophotos and DSMs are listed in Table 5.1. As mentioned in Section 5.3.1, the tile size $SZ_{one-to-many}$ is set to be 1280×960 pixels, and the image/tile pairs in both $SuperGlue_{tiling}$ and $SuperGlue_{orig}$ are downsampled to 640×480 pixels before entering SuperGlue. The comparison of downsampling ratio between $SuperGlue_{tiling}$ and $SuperGlue_{orig}$ for orthophotos and DSMs from Fréjus 1970 and 2014 is demonstrated in Table 5.2.

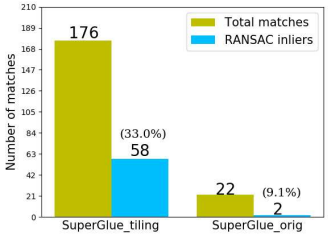
Results on matching orthophotos. Figure 5.7 displays the results of matching orthophotos with $SuperGlue_{tiling}$ and $SuperGlue_{orig}$. As can be seen, the former recovers 58 good matches with an inlier ratio reached 33%, while the latter fails to find any correct matches.

Results on matching DSMs. Figure 5.8 displays the results of matching DSMs with $SuperGlue_{tiling}$ and $SuperGlue_{orig}$. As can be seen, the former recovers 190 good matches with an inlier ratio reached 46.5%, while the latter fails to find any correct matches.

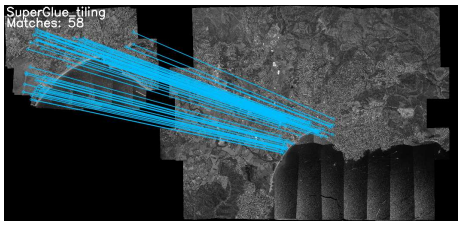
In general, $SuperGlue_{tiling}$ is able to recover enough good matches with high inlier ratio to guarantee stability in RANSAC, while $SuperGlue_{orig}$ fails to find any correct matches. In other words, our *one-to-many tiling scheme* improves the performance of SuperGlue significantly.



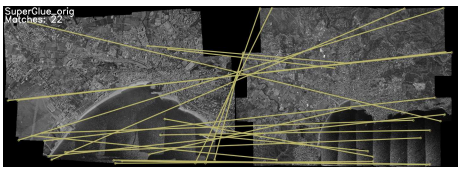
(a) Orthophotos



(b) Number of recovered matches

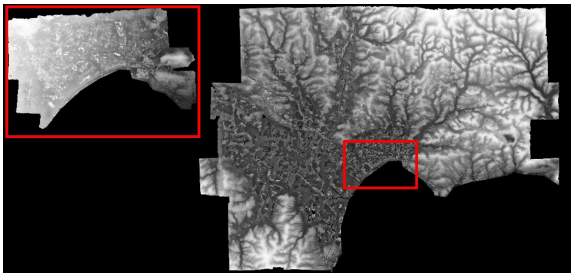


(c) $SuperGlue_{tiling}^{RANSACInliers}$

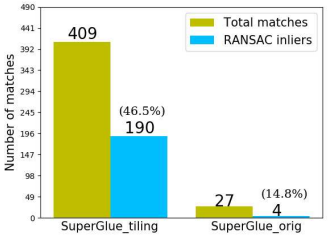


(d) $SuperGlue_{orig}^{TotalMatches}$

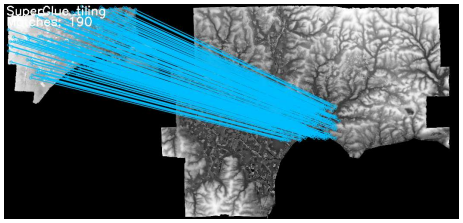
Figure 5.7: Comparison between $SuperGlue_{tiling}$ and $SuperGlue_{orig}$ on orthophotos from Fréjus 1970 and 2014 individually. (a) Orthophotos to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of total matches and RANSAC inliers of $SuperGlue_{tiling}$ and $SuperGlue_{orig}$. (c) Visualization of RANSAC inliers based on $SuperGlue_{tiling}$. (d)Visualization of total matches based on $SuperGlue_{orig}$.



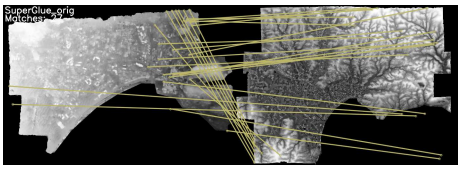
(a) DSMs



(b) Number of recovered matches



(c) $SuperGlue_{tiling}^{RANSACInliers}$



(d) $SuperGlue_{orig}^{TotalMatches}$

Figure 5.8: Comparison between $SuperGlue_{tiling}$ and $SuperGlue_{orig}$ on DSMs from Fréjus 1970 and 2014 individually. (a) DSMs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of total matches and RANSAC inliers of $SuperGlue_{tiling}$ and $SuperGlue_{orig}$. (c) Visualization of RANSAC inliers based on $SuperGlue_{tiling}$. (d)Visualization of total matches based on $SuperGlue_{orig}$.

5.3.4 Use case of matching guided by 2D similarity transformation

In this section we show an example where using 2D similarity model to guide matching is the only possible approach. The image pair to be matched is taken at the same time in Hofsjökull (c.f., Figure 5.9(a)). The overlapping zone is indicated with red rectangles. As can be seen, the area is fully covered with snow. It is extremely challenging to be matched as the whole image is weakly textured. However, the details revealed in the purple squares demonstrated that there are still some helpful information available. We compare the performance of *state-of-the-art* matching methods: (1) SIFT and (2) SuperGlue, as well as our matching strategy: (3) SIFT under the guidance of 2D similarity transformation model followed by RANSAC to remove outliers. No less than two matches are required to estimate the 2D similarity transformation parameters. In this particular case we need to measure 2 matching points manually. In less challenging scenarios one can use automated feature extractors for that purpose.

For each keypoint in the master image, our strategy uses 2D similarity transformation model to predict a location in the secondary image and search only its neighborhood of a circle with a radius S (in our experiment, S is set to be 45, 30 and 15 pixels respectively) to reduce ambiguity.
Figure 5.9(b-f) demonstrates the matching results of SIFT, SuperGlue and our matching strategy. As can be seen, SIFT and SuperGlue fail to find any correct matches, while our strategy obtains a large number of good matches with negligible manual labor. The number of matches increases as the search radius S changes from 45 to 15, which is reasonable due to decrease of ambiguity. However, false matches are introduced when the radius is too small (c.f., Figure 5.9(f)). The best balance is achieved with S set to be 30 pixels (c.f., Figure 5.9(e)).

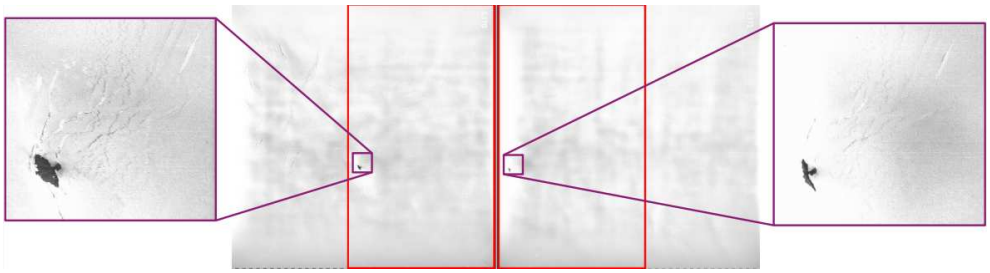
5.3.5 Comparison of 6 variants

In order to evaluate the results of the 6 variants qualitatively and quantitatively, the following criteria are applied:

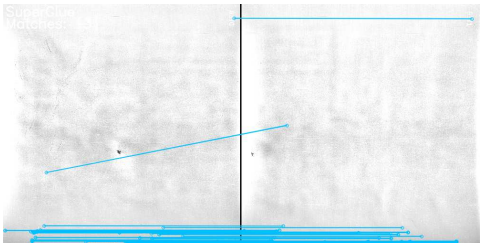
1. **Matches visualization**. The numbers of (1) total matches (i.e., matches before RANSAC) as well as (2) RANSAC inliers (matches that survived RANSAC) are displayed together in bar charts; in the meantime, the RANSAC inliers are visualized and demonstrated, from which we can tell whether the variants succeeded or failed.

2. **DoD**. For the variants that succeeded, we use the resulted orientations in the same frame to calculate DSMs in order to generate DoD. The visualization of DoD as well as the statistical information are displayed.

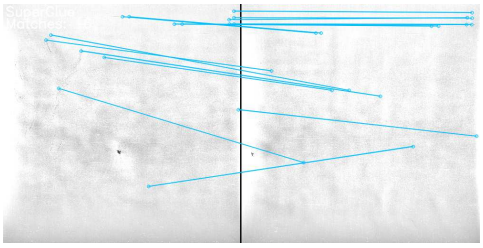
As the results reveal similar pattern on different datasets, for the sake of



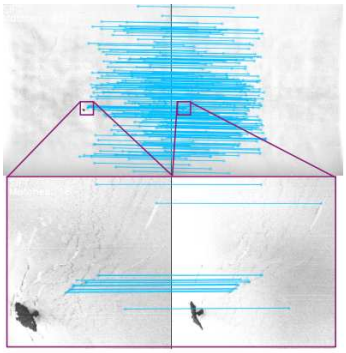
(a) Overlapping zone with details



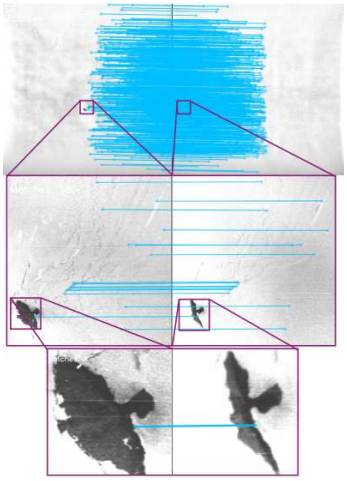
(b) SIFT



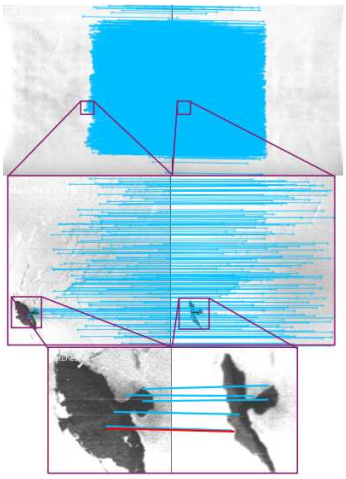
(c) SuperGlue



(d) Guided SIFT (S=45)



(e) Guided SIFT (S=30)



(f) Guided SIFT (S=15)

Figure 5.9: Matching results of an intra-epoch image pair from dataset Hofsjökull. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. Details are revealed in purple squares. (b) and (c) are matches recovered by SIFT and SuperGlue individually. (d-f) displays matches found by our matching strategy with search radius set to be 45, 30 and 15 pixels individually, with purple squares indicating the details.

simplicity, we only show the results of 2 sets of datasets (i.e., Fréjus and Alberona), as the former represents challenging case with drastic scene changes and the latter stands for general case, and we move the results of Kobe as well as Pezenas to Appendix A to keep the current section tidy.

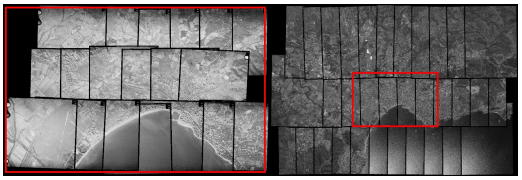
Matches visualization. For each dataset, each free epoch E_f is matched with reference epoch E_r with 6 variants (i.e., ① $SIFT_{ImgPairs}$, ② $SuperGlue_{ImgPairs}$, ③ $SIFT_{Ortho}$, ④ $SuperGlue_{Ortho}$, ⑤ $SIFT_{DSM}$ and ⑥ $SuperGlue_{DSM}$), the matches are visualized and displayed together for comparison. In Table 5.3 we display whether the 6 variants succeeded or failed on all the datasets.

	ImgPairs		Ortho		DSM	
	SIFT	SuperGlue	SIFT	SuperGlue	SIFT	SuperGlue
$Frejus^{1954}_{2014}$	Fail	**Succeed**	Fail	**Succeed**	**Succeed**	**Succeed**
$Frejus^{1966}_{2014}$	Fail	**Succeed**	Fail	**Succeed**	**Succeed**	**Succeed**
$Frejus^{1970}_{2014}$	Fail	**Succeed**	Fail	**Succeed**	**Succeed**	**Succeed**
$Pezenas^{1971}_{2015}$	**Succeed**	**Succeed**	**Succeed**	**Succeed**	**Succeed**	**Succeed**
$Pezenas^{1981}_{2015}$	**Succeed**	**Succeed**	**Succeed**	**Succeed**	**Succeed**	**Succeed**
$Pezenas^{1971}_{2014(Satellite)}$	/	/	Fail	Fail	**Succeed**	**Succeed**
$Pezenas^{1981}_{2014(Satellite)}$	/	/	Fail	**Succeed**	**Succeed**	**Succeed**
$Kobe^{1991}_{1995}$	Fail	**Succeed**	Fail	**Succeed**	**Succeed**	**Succeed**
$Alberona^{1954}_{2003}$	Fail	**Succeed**	Fail	**Succeed**	**Succeed**	**Succeed**

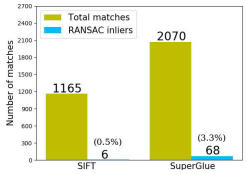
Table 5.3: Demonstration of 6 variants succeeding or failing on each dataset.

For Fréjus, the reference epoch E_r is 2014, the matches visualizations between free epochs E_f (i.e., epoch 1954, 1966 and 1970) and E_r are displayed in Figure 5.10, 5.11 and 5.12 individually. For Alberona, the reference epoch E_r is 2003, the matches visualizations between free epoch E_f (i.e., epoch 1954) and E_r are displayed in Figure 5.13.

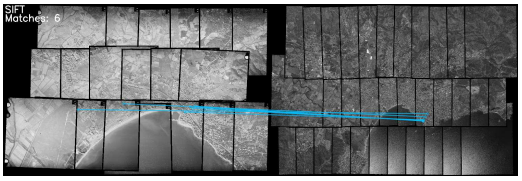
DoD. According to Table 5.3, by applying 6 variants (or 4 variants for satellite images) on each free epoch and the reference epoch, we got 50 testing cases. Among all the cases, there are 37 of them succeeded, which leads to 37 sets of co-registered orientations. For each set of resulted orientations, we use them to calculate DSMs in free epoch and reference epoch individually in order to generate DoD. 16 sets of DoDs belong to Fréjus and Alberona therefore are displayed in the current section (Figure 5.14 and 5.15). Their corresponding statistical information is displayed in Table 5.4. The rest 21 sets of DoDs are displayed in Appendix A.



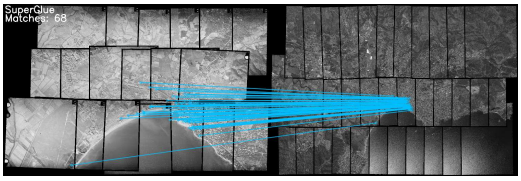
(a) Image pairs (19×33 pairs)



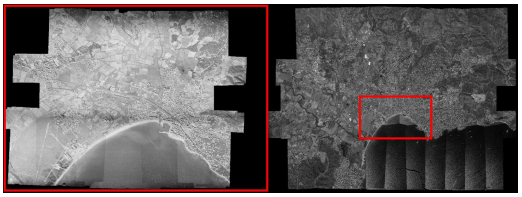
(b) Number of recovered matches(*ImgPairs*)



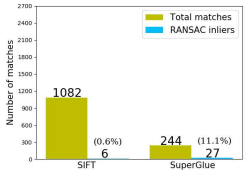
(c) $SIFT^{RANSACInliers}_{ImgPairs}$



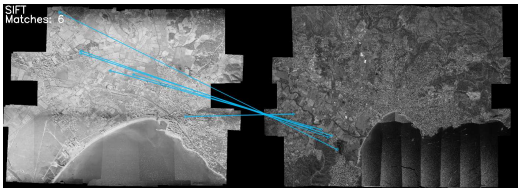
(d) $SuperGlue^{RANSACInliers}_{ImgPairs}$



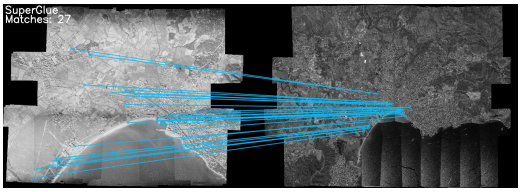
(e) Orthophotos



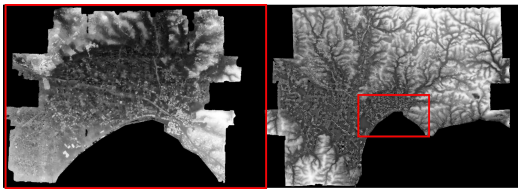
(f) Number of recovered matches(*Ortho*)



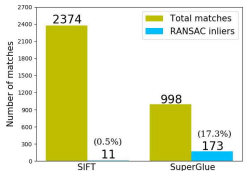
(g) $SIFT^{RANSACInliers}_{Ortho}$



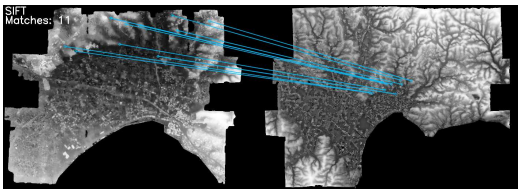
(h) $SuperGlue^{RANSACInliers}_{Ortho}$



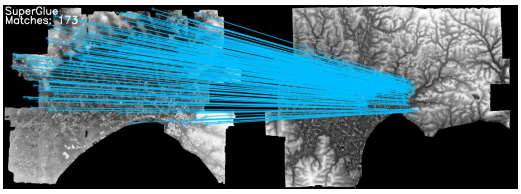
(i) DSMs



(j) Number of recovered matches(*DSM*)

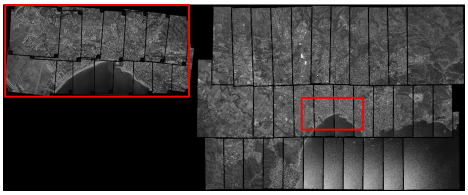


(k) $SIFT^{RANSACInliers}_{DSM}$

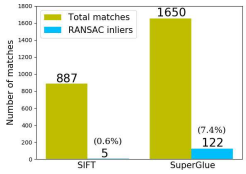


(l) $SuperGlue^{RANSACInliers}_{DSM}$

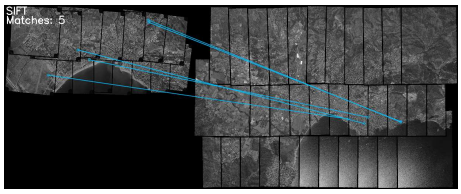
Figure 5.10: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Fréjus 1954 and 2014**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



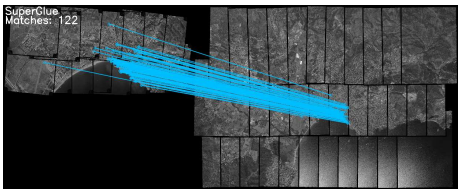
(a) Image pairs (15×33 pairs)



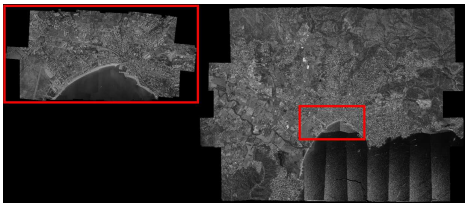
(b) Number of recovered matches(*ImgPairs*)



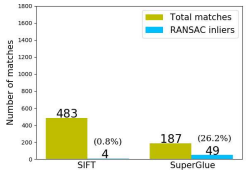
(c) $SIFT_{ImgPairs}^{RANSACInliers}$



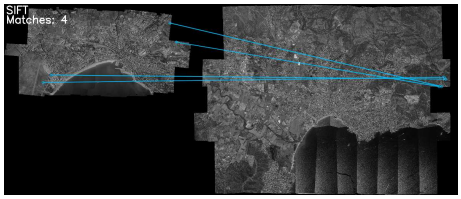
(d) $SuperGlue_{ImgPairs}^{RANSACInliers}$



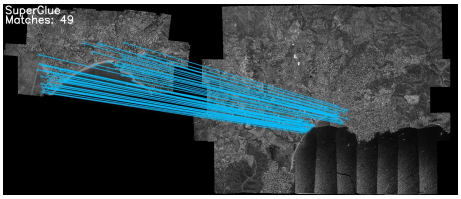
(e) Orthophotos



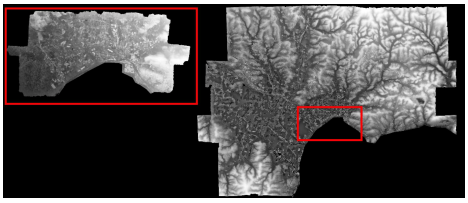
(f) Number of recovered matches(*Ortho*)



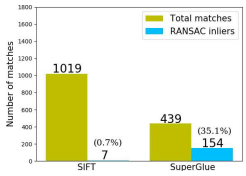
(g) $SIFT_{Ortho}^{RANSACInliers}$



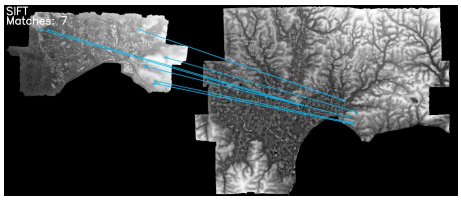
(h) $SuperGlue_{Ortho}^{RANSACInliers}$



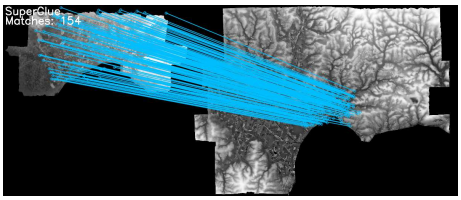
(i) DSMs



(j) Number of recovered matches(*DSM*)

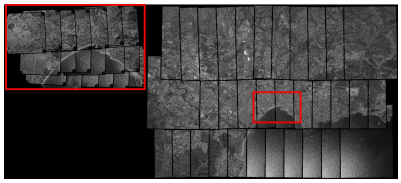


(k) $SIFT_{DSM}^{RANSACInliers}$

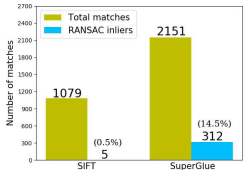


(l) $SuperGlue_{DSM}^{RANSACInliers}$

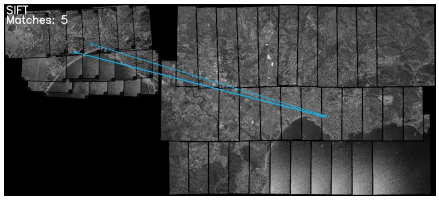
Figure 5.11: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Fréjus 1966 and 2014**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



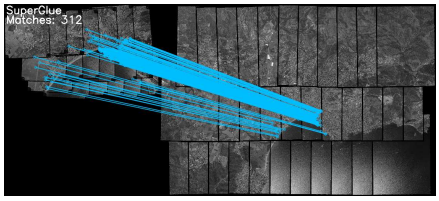
(a) Image pairs (19×33 pairs)



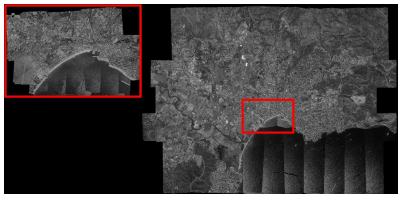
(b) Number of recovered matches(*ImgPairs*)



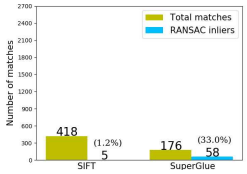
(c) $SIFT^{RANSACInliers}_{ImgPairs}$



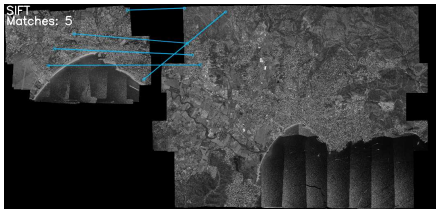
(d) $SuperGlue^{RANSACInliers}_{ImgPairs}$



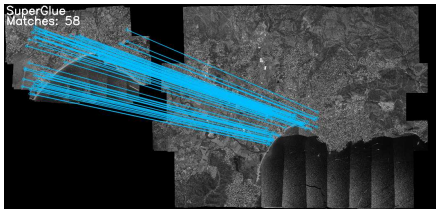
(e) Orthophotos



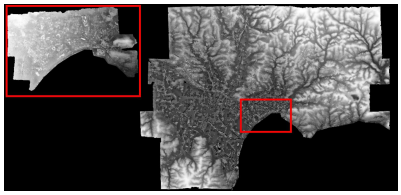
(f) Number of recovered matches(*Ortho*)



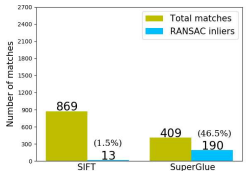
(g) $SIFT^{RANSACInliers}_{Ortho}$



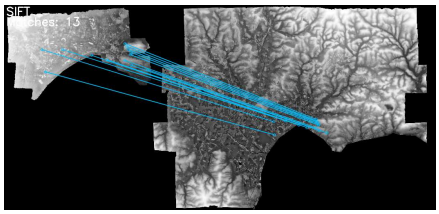
(h) $SuperGlue^{RANSACInliers}_{Ortho}$



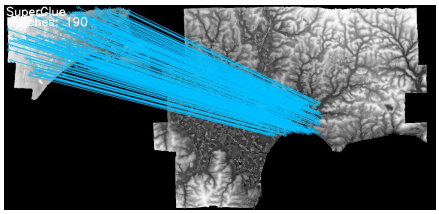
(i) DSMs



(j) Number of recovered matches(*DSM*)

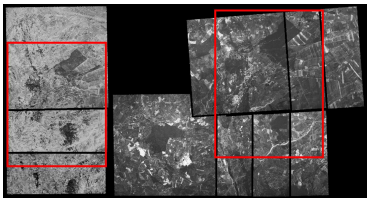


(k) $SIFT^{RANSACInliers}_{DSM}$

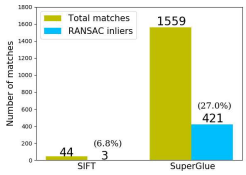


(l) $SuperGlue^{RANSACInliers}_{DSM}$

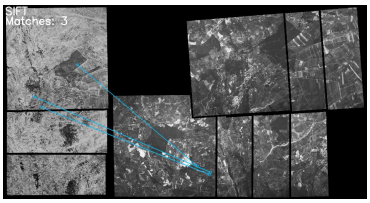
Figure 5.12: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Fréjus 1970 and 2014**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



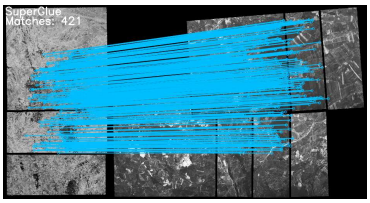
(a) Image pairs (3×7 pairs)



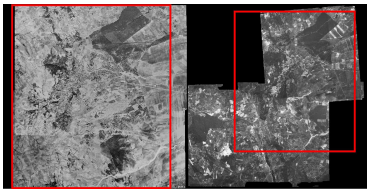
(b) Number of recovered matches(*ImgPairs*)



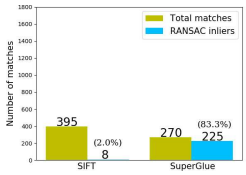
(c) $SIFT_{ImgPairs}^{RANSACInliers}$



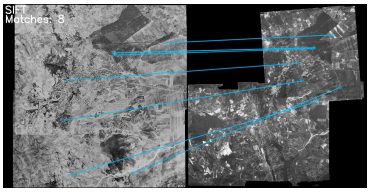
(d) $SuperGlue_{ImgPairs}^{RANSACInliers}$



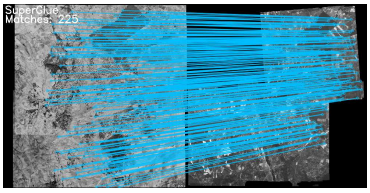
(e) Orthophotos



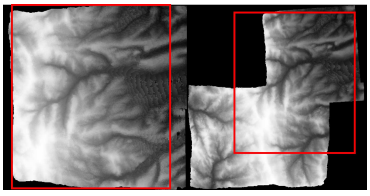
(f) Number of recovered matches(*Ortho*)



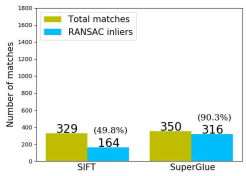
(g) $SIFT_{Ortho}^{RANSACInliers}$



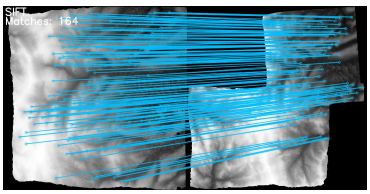
(h) $SuperGlue_{Ortho}^{RANSACInliers}$



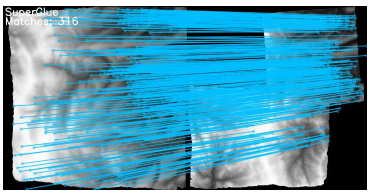
(i) DSMs



(j) Number of recovered matches(*DSM*)

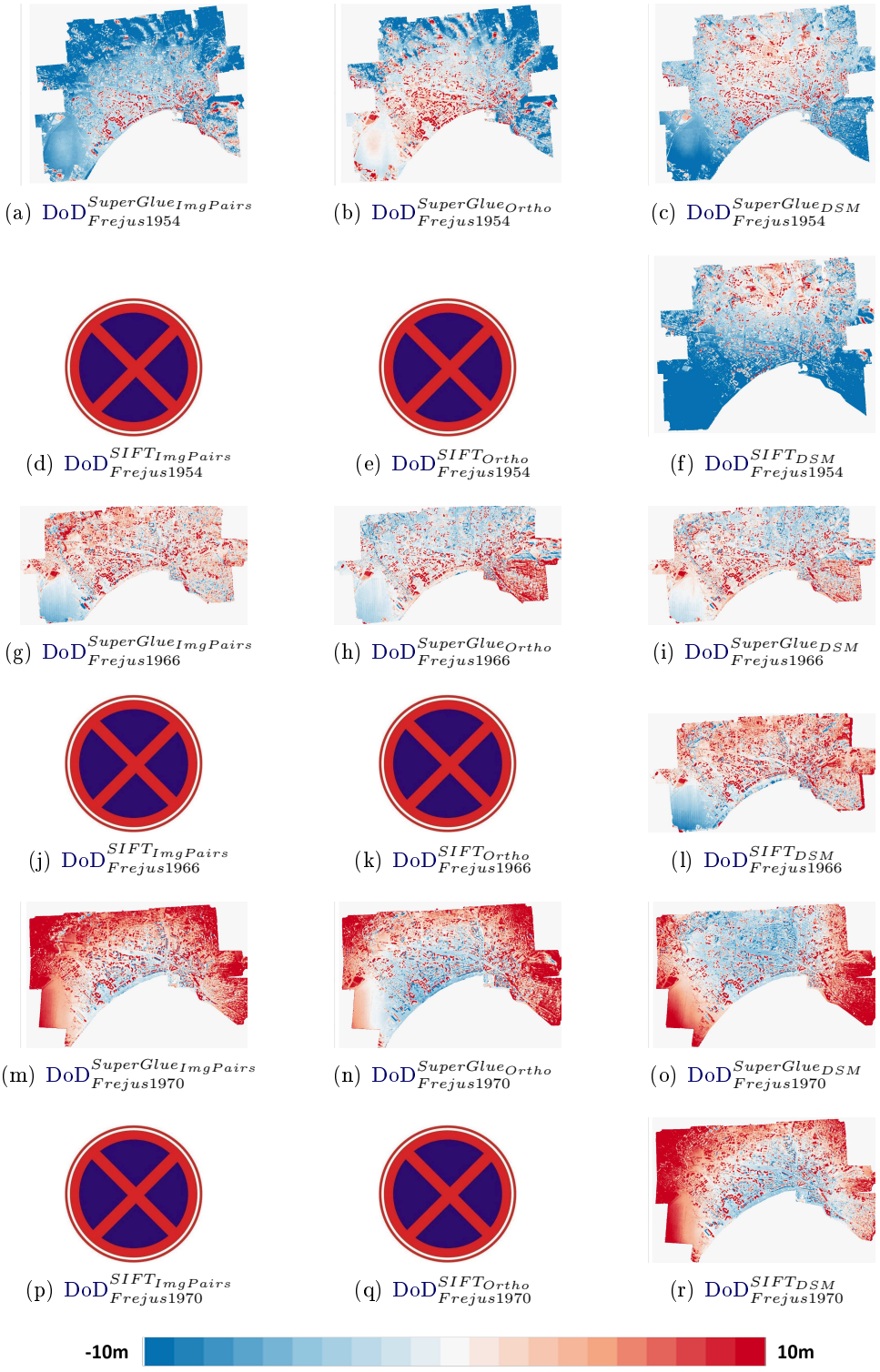


(k) $SIFT_{DSM}^{RANSACInliers}$



(l) $SuperGlue_{DSM}^{RANSACInliers}$

Figure 5.13: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Alberona 1954 and 2003**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



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

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

(g) $\text{DoD}^{SuperGlue_{ImgPairs}}_{Frejus1966}$ (h) $\text{DoD}^{SuperGlue_{Ortho}}_{Frejus1966}$ (i) $\text{DoD}^{SuperGlue_{DSM}}_{Frejus1966}$

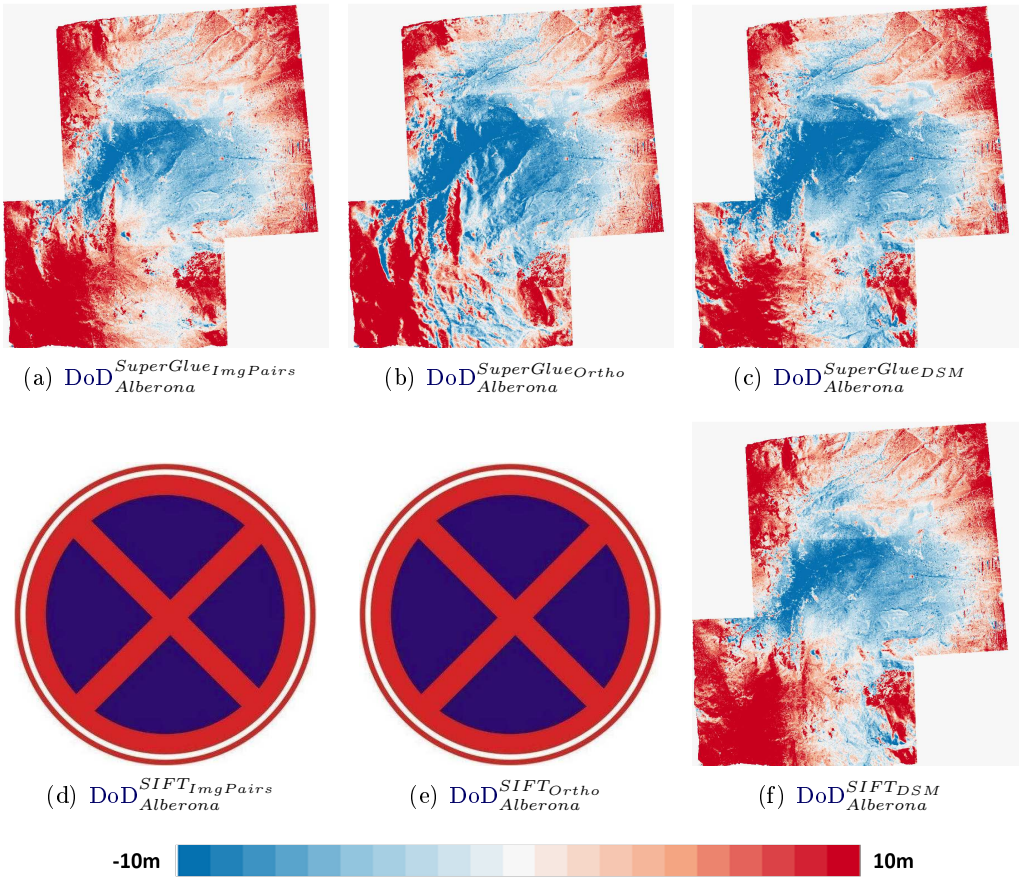
(j) $\text{DoD}^{SIFT_{ImgPairs}}_{Frejus1966}$ (k) $\text{DoD}^{SIFT_{Ortho}}_{Frejus1966}$ (l) $\text{DoD}^{SIFT_{DSM}}_{Frejus1966}$

(m) $\text{DoD}^{SuperGlue_{ImgPairs}}_{Frejus1970}$ (n) $\text{DoD}^{SuperGlue_{Ortho}}_{Frejus1970}$ (o) $\text{DoD}^{SuperGlue_{DSM}}_{Frejus1970}$

(p) $\text{DoD}^{SIFT_{ImgPairs}}_{Frejus1970}$ (q) $\text{DoD}^{SIFT_{Ortho}}_{Frejus1970}$ (r) $\text{DoD}^{SIFT_{DSM}}_{Frejus1970}$

(s) DoD legend

Figure 5.14: DoDs between free epoch **Fréjus 1954, 1966, 1970** and reference epoch **2014** with variants $SuperGlue_{ImgPairs}$ (a, g, m), $SuperGlue_{Ortho}$ (b, h, n), $SuperGlue_{DSM}$ (c, i, o), $SIFT_{ImgPairs}$ (d, j, p), $SIFT_{Ortho}$ (e, k, q) and $SIFT_{DSM}$ (f, l, r). The prohibition sign means the corresponding variant failed.



(a) DoD$^{SuperGlue_{ImgPairs}}_{Alberona}$ (b) DoD$^{SuperGlue_{Ortho}}_{Alberona}$ (c) DoD$^{SuperGlue_{DSM}}_{Alberona}$

(d) DoD$^{SIFT_{ImgPairs}}_{Alberona}$ (e) DoD$^{SIFT_{Ortho}}_{Alberona}$ (f) DoD$^{SIFT_{DSM}}_{Alberona}$

(g) DoD legend

Figure 5.15: DoDs between free epoch **Alberona 1954** and reference epoch **2003** with variants $SuperGlue_{ImgPairs}$ (a), $SuperGlue_{Ortho}$ (b), $SuperGlue_{DSM}$ (c), $SIFT_{ImgPairs}$ (d), $SIFT_{Ortho}$ (e) and $SIFT_{DSM}$ (f). The prohibition sign means the corresponding variant failed.

		μ [m]	σ [m]	$\lvert\mu\rvert$ [m]
$DoD^{Frejus}_{1954-2014}$	$SuperGlue_{ImgPairs}$	6.02	7.04	7.39
	$SuperGlue_{Ortho}$	2.55	7.31	5.29
	$SuperGlue_{DSM}$	2.24	5.34	**4.35**
	$SIFT_{ImgPairs}$	/	/	/
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	6.91	8.90	8.45
$DoD^{Frejus}_{1966-2014}$	$SuperGlue_{ImgPairs}$	-1.74	4.78	3.55
	$SuperGlue_{Ortho}$	-0.69	5.11	3.59
	$SuperGlue_{DSM}$	-0.80	4.71	**3.28**
	$SIFT_{ImgPairs}$	/	/	/
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	-2.12	5.66	4.42
$DoD^{Frejus}_{1970-2014}$	$SuperGlue_{ImgPairs}$	-5.60	6.16	6.54
	$SuperGlue_{Ortho}$	-3.34	6.50	5.48
	$SuperGlue_{DSM}$	-2.37	6.57	**5.35**
	$SIFT_{ImgPairs}$	/	/	/
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	-3.76	6.21	5.47
$DoD^{Alberona}_{1954-2003}$	$SuperGlue_{ImgPairs}$	-2.90	7.65	6.13
	$SuperGlue_{Ortho}$	-1.06	9.35	6.86
	$SuperGlue_{DSM}$	-0.46	7.70	6.08
	$SIFT_{ImgPairs}$	/	/	/
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	-2.11	7.39	**5.92**

Table 5.4: Average value μ, standard deviation σ, and absolute average value $|\mu|$ of all the DoDs in Figure 5.14 and 5.15.

Discussion. As can be seen, $SIFT_{ImgPairs}$ and $SIFT_{Ortho}$ fail to recover correct matches for most of the cases. It is reasonable as the appearance of inter-epoch RGB images often looks different and SIFT is not sufficiently invariant over time by its very nature.
$SIFT_{DSM}$ and $SuperGlue_{DSM}$ recover enough good matches for all the datasets, even the most challenging case of Fréjus with extreme scene changes, thanks to stable information on DSMs. However, the inlier ratio of $SIFT_{DSM}$ on Fréjus is dangerously low (around 1%), which makes the RANSAC procedure unstable and the rough co-registration result inferior.
$SuperGlue_{ImgPairs}$ and $SuperGlue_{Ortho}$ succeeded for almost all the testing cases, except for $SuperGlue_{Ortho}$ on matching Pezenas 1971 and 2014, as the overlapping zone is too limited to provide enough clues in context to ensure the matching performance.

For the DoDs, *dome* effect is present in all the datasets. This is caused by inaccurately estimated lens distortion parameters. It is acceptable for rough co-registration as our goal is only to provide guidance for precise matching.
It is worth noting that the DoDs between Fréjus 1954 and 2014 based on variant $SIFT_{DSM}$ (i.e., Figure 5.14 (f)) is accompanied with specifically obvious *dome* effect, with its absolute average value $|\mu|$ reached 8.04 meters as shown in Table 5.4. This is because the inlier ratio of matches in the RANSAC procedure is too low to guarantee reliable rough co-registration. According to the absolute average value $|\mu|$ displayed in Table 5.4, variant $SuperGlue_{DSM}$ performs the best for matching Fréjus 1954 and 1966 to 2014, where drastic scene changes are presented. In the meantime, variants $SuperGlue_{Ortho}$ and $SIFT_{DSM}$ perform the best for less challenging cases (i.e., matching Fréjus 1970 to 2014, and Alberona 1954 to 2003).

5.4 Conclusion

We provide 2 strategies for multi-epoch rough co-registration: (1) match image pairs (i.e., *ImgPairs*) and (2) match orthophotos/DSMs (i.e., *Ortho* and *DSM*). For each pipeline, we test 2 feature matching methods (SIFT and SuperGlue), which leads to 6 variants (i.e., ① $SIFT_{ImgPairs}$, ② $SuperGlue_{ImgPairs}$, ③ $SIFT_{Ortho}$, ④ $SuperGlue_{Ortho}$, ⑤ $SIFT_{DSM}$ and ⑥ $SuperGlue_{DSM}$). We test the variants on 4 datasets (Fréjus, Pezenas, Kobe and Alberona), including the cases of (1) matching aerial epochs only and (2) matching aerial and satellite epochs mixed. Experiments show that:

1. $SIFT_{DSM}$ and $SuperGlue_{DSM}$ lead to more robust results than other variants, since landcover provides more reliable information as scene evolves.

2. SuperGlue is generally more reliable than SIFT, as the former is more invariant over time.

Chapter 6

Precise matching

Contents

6.1 Introduction

6.1.1 Motivation and objective

The rough co-registration stage elaborated in Section 5 laid a solid foundation for matching inter-epoch images, as it roughly aligned images from different epochs in a globally consistent way. However, the alignment is not accurate enough for high precision cartography. Therefore, we propose a precise matching stage to get matches with higher accuracy, which benefits from the guidance of rough co-registration to guarantee both robustness and precision. For each inter-epoch image pair I^{e_1} and I^{e_2} to be matched, our goal is to find precise matches $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$ ($\mathbf{K}^{e_i}$ represents keypoints extracted in image I^{e_i}). Based on the roughly co-registered orientations and DSMs resulted from Chapter 5, we can readily predict a potential matching point $\widetilde{\mathbf{K}}^{e_2}$ in I^{e_2} for keypoint $\mathbf{K}^{e_1}$. As rough co-registration provides robust yet imprecise alignment, the precise matching point for keypoint $\mathbf{K}^{e_1}$ should not be far away from the predicted point $\widetilde{\mathbf{K}}^{e_2}$. Therefore, we can narrow down the search space in precise matching stage by only considering the local neighborhood of the

predicted point $\widetilde{\mathbf{K}}^{e_2}$ to reduce ambiguity tremendously. For hand-crafted methods like SIFT, the strategy of predicting keypoints followed by narrowing down the search space can be readily applied. Besides, as SIFT provides explicitly the scale and rotation angle of the keypoints, we can take advantage of that and introduce an idea of *SclRotCheck*, which is to check if the scale and rotation of the keypoints coincide with the scale and rotation predicted by rough co-registration.
For learned methods like SuperGlue, it is not easy to modify the algorithm, as it inevitably involves retraining the model, which is not easy due to lack of training data. Therefore we propose an *one-to-one tiling scheme* (not to confuse with the *one-to-many tiling scheme* presented in Section 5) to feed roughly aligned patches into the model to reduce ambiguity. Its merits are twofold: (1) up-scaling the learning based feature matching algorithms to high resolution imagery, as directly feeding the original images often lead to inaccurate results; (2) narrowing down the searching space in an elegant way without modifying the model.

6.1.2 Contributions

Our contribution is to combine rough co-registration results, *one-to-one tiling scheme*, *SclRotCheck* and 3D-RANSAC into a reliable pipeline to recover both robust and precise matches, more specifically, we:

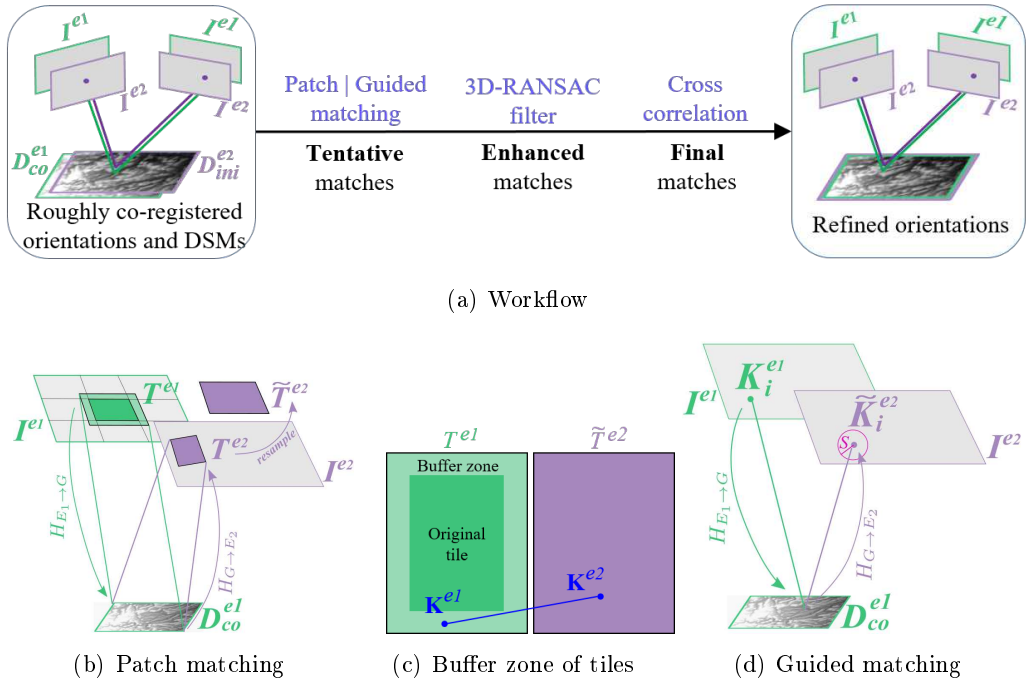
1. reduce the difficulty in precise matching under the guidance of co-registered orientations and DSMs by narrowing down the search space.
2. introduce *one-to-one tiling scheme* to (1) scale-up the deep learning methods and (2) reduce matching ambiguity without retraining the model.
3. introduce *SclRotCheck* to remove potential matches whose scale ratio and rotation difference are not consistent with the prediction of rough co-registration.
4. perform 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 (1D versus 2D constraint).

6.2 Methodology

To compute precise inter-epoch matches, we perform matching on original RGB images under the guidance of co-registered orientations and DSMs. It consists of extracting tentative inter-epoch matches, followed by a 3D-RANSAC filter and a cross correlation stage to remove outliers. The workflow is displayed in Figure 6.1(a).

We choose matching RGB images for precise matching instead of DSMs, as DSMs are (1) noisy due to errors inevitably introduced during calculating DSMs and (2) monotonous in flat terrain due to lack of textures. In Section 6.3.2 we

displayed the matching results on both RGB images and DSMs over the same area. It demonstrates that more matches are found in DSMs but the accuracy is inferior. As our goal is to recover accurate matches, the RGB images are more suitable than DSMs. Besides, it is more efficient as calculating high resolution DSMs is computationally demanding.



(a) Workflow

(b) Patch matching

(c) Buffer zone of tiles

(d) Guided matching

Figure 6.1: (a) Workflow of precise matching. It is carried out by performing patch or guided matching to obtain tentative matches, followed by 3D-RANSAC filter and cross correlation, giving rise to final matches. (b) and (d) illustrate toy-examples of the patch and guided matching, respectively. (c) displays the match where $\mathbf{K}^{e_1}$ exceeds the original tile size (dark green area) and is therefore abandoned.

6.2.1 Get tentative matches with patch/guided matching

We offer two alternatives to recover tentative matches: patch or guided matching. The former uses learned features, while the latter uses hand-crafted features. Patch matching often gives larger number of matches, while guided matching is in general more efficient.

Patch matching for learned features. For patch matching, we propose a *one-to-one tiling scheme* to improve matching performance of learned features and reduce ambiguity at the same time. It is illustrated in Figure 6.1(b), and elaborated

below:

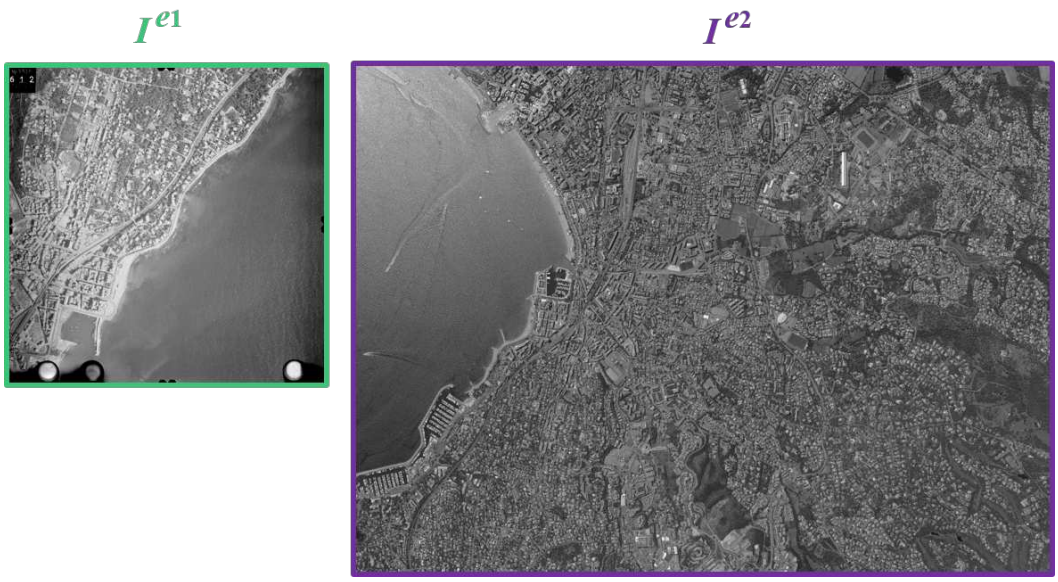
1. Crop the master RGB image I^{e_1} into M original tiles of size $SZ^{Orig}_{one-to-one}$, and expand them with a buffer zone of size SZ_{buffer} (as shown in Figure 6.1(c)), giving rise to M buffered tiles (T^{e_1}) of size $SZ_{one-to-one}$;
2. Project each buffered tile T^{e_1} onto the DSM $D^{e_1}_{co}$ and backproject to secondary RGB image I^{e_2} to find the corresponding tile T^{e_2};
3. Resample T^{e_2} to $\widetilde{T}^{e_2}$, so that the tile pair $P(T^{e_1}, \widetilde{T}^{e_2})$ is free from differences of rotation, scale and extent;

We apply SuperGlue on each tile pair $P(T^{e_1}, \widetilde{T}^{e_2})$ to find matches $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$ ($\mathbf{K}^{e_i}$ represents keypoints in image I^{e_i}), and merge the matches together by removing the ones with $\mathbf{K}^{e_1}$ located in the buffer zone. As the orientations and DSMs are only roughly co-registered, we take into account the margin of error when projecting tiles to overlapping images. This is why we add a buffer zone in the tile T^{e_1}.
For better understanding, in Figure 6.2 we display an example of an inter-epoch image pair, as well as the tile pairs resulted from the *one-to-one tiling scheme*.

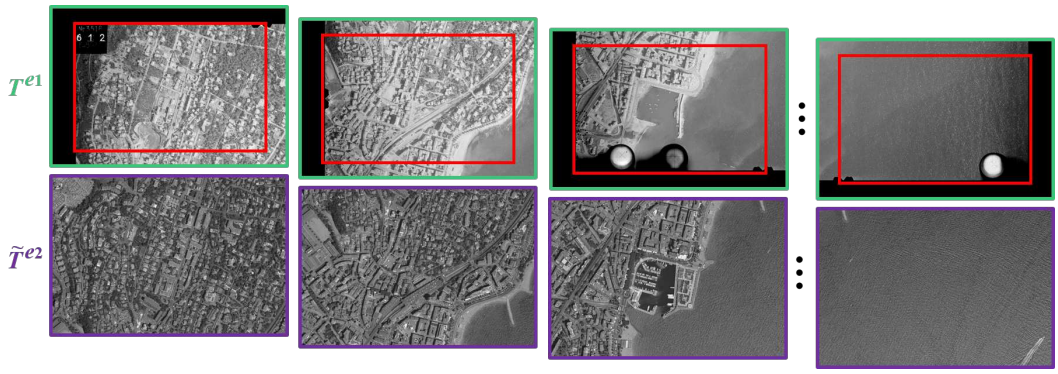
Our patch matching experiments are performed based on SuperGlue, however, other learned methods can be adopted readily.

Guided matching for hand-crafted features. The patch matching substitute orientated towards hand-crafted features is the guided matching, as shown in Figure 6.1(d). It leverages the positions of predicted keypoints, the known scale ratio and rotation differences to narrow down the list of the matching candidates. In our experiments, we use the SIFT points, but the pipeline is suitable to any hand-crafted extractor. It consists of the following steps:

1. Compute the scale ratio R_{scl} and the rotation D_{rot} between two images by sequentially projecting the I^{e_1} image corners to the co-registered DSM $D^{e_1}_{co}$ and to image I^{e_2};
2. Extract keypoints $\mathbf{K}^{e_1}$ in image I^{e_1} and $\mathbf{K}^{e_2}$ in image I^{e_2};
3. Intersect the keypoints $\mathbf{K}^{e_1}$ with the co-registered DSM $D^{e_1}_{co}$;
4. Back-project them to image I^{e_2}, giving rise to predicted keypoints $\widetilde{\mathbf{K}}^{e_2}$;
5. Search for a subset of points in $\mathbf{K}^{e_2}$ located within a radius S centered at the predicted positions $\widetilde{\mathbf{K}}^{e_2}$;
6. Remove candidate matches whose scales and rotations computed by SIFT are out of range $[(1\text{-}Th_{scl})\times R_{scl}, (1+Th_{scl})\times R_{scl}]$ and $[D_{rot}\text{-}Th_{rot}, D_{rot}+Th_{rot}]$;



(a) Example of an image pair



(b) Demonstration of tile pairs

Figure 6.2: Illustration of patch matching applied on an inter-epoch image pair. (a) The master image (I^{e_1}) and secondary image (I^{e_2}) are taken at Fréjus in 1954 and 2014 individually. (b) Tile pairs resulted from *one-to-one tiling scheme*, the tile zones before and after buffering are marked as red and green rectangles.

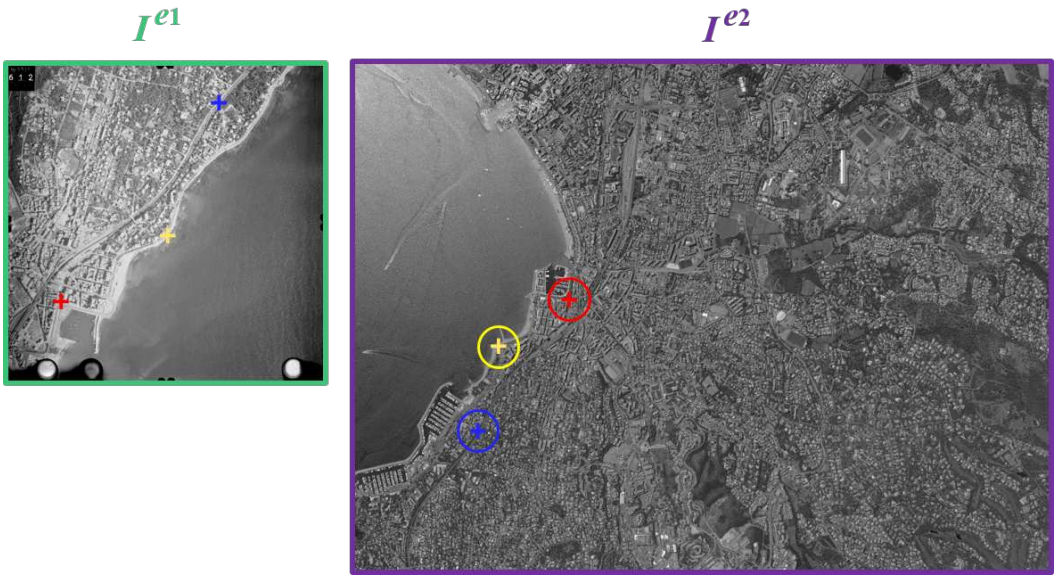


Figure 6.3: Illustration of keypoint prediction (cross symbols) accompanied with search space (circles), the master image (I^{e_1}) and secondary image (I^{e_2}) are taken at Fréjus in 1954 and 2014 individually.

7. Find the best matches with mutual nearest neighbor combined with the first to second nearest neighbor ratio test [Lowe 2004].

For better understanding, in Figure 6.3 we display an example of an inter-epoch image pair, with keypoint prediction (cross symbols) accompanied with search space (circles) superposed on them.

6.2.2 Get enhanced matches with 3D-RANSAC

To compute enhanced matches, we apply a 3D-RANSAC filter on the previously obtained tentative matches. More precisely, we do the following: (1) for each match $M(\mathbf{K}^{e_1}, \mathbf{K}^{e_2})$, the keypoints $\mathbf{K}^{e_1}$ and $\mathbf{K}^{e_2}$ are projected onto DSM $D^{e_1}_{co}$ and $D^{e_2}_{ini}$ individually to get 3D matching points $M(\mathbf{KG}^{e_1}, \mathbf{KG}^{e_2})$; and (2) the matches $M(\mathbf{KG}^{e_1}, \mathbf{KG}^{e_2})$ are iteratively sampled to compute the 3D Helmert transformation RANSAC model:

$$\begin{bmatrix} KG_x^{e_2} \\ KG_y^{e_2} \\ KG_z^{e_2} \end{bmatrix} = \lambda \cdot \mathbf{R} \cdot \begin{bmatrix} KG_x^{e_1} \\ KG_y^{e_1} \\ KG_z^{e_1} \end{bmatrix} + \begin{bmatrix} \Delta_x \\ \Delta_y \\ \Delta_z \end{bmatrix}. \tag{6.1}$$

where λ is the scale factor, $\mathbf{R}$ is the rotation matrix and $\begin{bmatrix} \Delta_x, \Delta_y, \Delta_z \end{bmatrix}^T$ is the translation vector. Matches within T_r of its predicted position (i.e., $|\mathbf{KG}^{e_2} - (\lambda \cdot \mathbf{R} \cdot \mathbf{KG}^{e_1} + \Delta)| < T_r$) are considered as inliers.

6.2.3 Get final matches with cross correlation

In the preceding step we got rid of a substantial number of outliers, however, we believe that not all outliers could be identified. Besides, our goal is to get a moderate number of reliable matches instead of many unreliable ones. Therefore we introduce a different filtering method (i.e., cross correlation) to further remove false matches. Even though cross correlation itself is not discriminative and efficient enough when used alone, it fits well in our pipeline as we already recovered many discriminative and well-distributed matches before applying it. Matches with their correlation scores below a predefined threshold T_c are discarded. The correlation window size was set to be large enough to take into account the context around a point (32×32 pixels in our experiment). Figure 6.4 shows an example of a false match (red) eliminated by cross correlation, while the true match (blue) is kept.

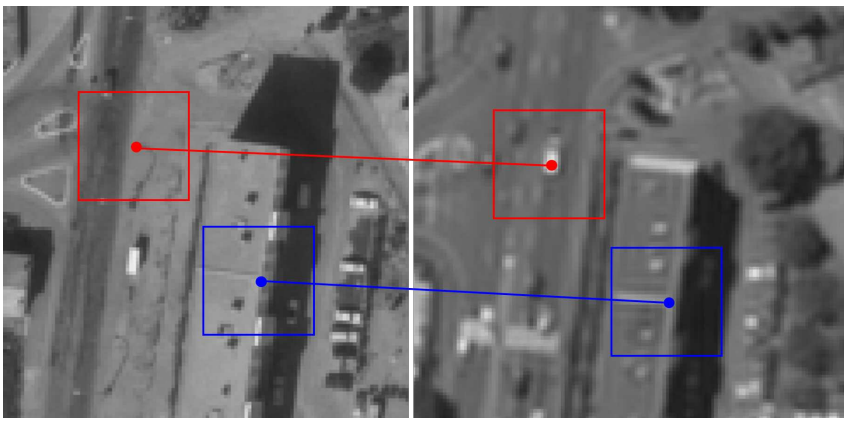


Figure 6.4: Demonstration of the validation with cross-correlation. Considering poor quality of historical images, the window size (blue and red rectangles) was set to 32×32 pixels. False match (red) is eliminated by cross correlation, while true match (blue) is kept.

6.2.4 Refine orientations

Based on the intra-epoch and inter-epoch matches, a free network BBA is performed to refine all the image orientations and camera calibrations. If the results need to be analyzed in a metric scale, a 3D Helmert transformation will be performed to move the refined acquisitions in an arbitrary reference frame to a metric one. If the precise orientations for one of the epochs were known, their parameters will be fixed during the BBA and the subsequent 3D Helmert transformation will be skipped. 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. Fraser is a radial model, with decentric and affine parameters, there

are 12 degrees of freedom: 1 for focal length , 2 for principal point, 2 for distortion center , 3 for coefficients of radial distorsion, 2 for decentric parameters and 2 for affine parameters. We choose this model because we want to test the versatility of our pipeline, instead of trying different model to achieve the best performance.

6.3 Experiments

As described in the previous section, our precise matching stage relies on RGB images and consists of 3 main steps to get the tentative, enhanced and final matches. In Section 6.3.2 we compare the matching results on RGB images and DSMs to explain why we choose the former over the latter to perform precise matching.
For obtaining tentative matches, there are 2 alternatives (i.e., patch or guided matching), leading to 2 precise matching variants:

1. *Patch*: recover tentative matches with patch matching, followed by 3D-RANSAC and cross correlation to remove outliers;
2. *Guided*: same as *Patch*, except replacing patch matching with guided matching.

For each dataset, we choose the rough co-registration results calculated with $SIFT_{DSM}$ and $SuperGlue_{DSM}$ individually (as they are the most robust variants for rough co-registration) to guide the precise matching *Patch* or *Guided*, leading to 4 sets of variants, which are referred to as:

1. $Patch_{SpGDSM}$
2. $Guided_{SpGDSM}$
3. $Patch_{SIFTDSM}$
4. $Guided_{SIFTDSM}$

We test our precise matching variants on all the multi-epoch datasets which are elaborated in Chapter 4: Fréjus, Pezenas, Kobe and Alberona. The results are demonstrated in Section 6.3.3.

For Fréjus, Kobe and Alberona, we keep all the epochs for experiments, as Fréjus displayed drastic scene changes, while Kobe and Alberona witnessed earthquake and landslide individually. In Pezenas, less changes are observed. Therefore we maximize the matching difficulty by choosing both aerial and satellite epoch accompanied with the largest time gap (i.e., aerial epoch 1971 and satellite epoch 2014).

The orientations of GT epochs (i.e., Pezenas 2014 and Fréjus 2014) were treated as fixed during the combined BBA since they were accurately known *a-priori*, while

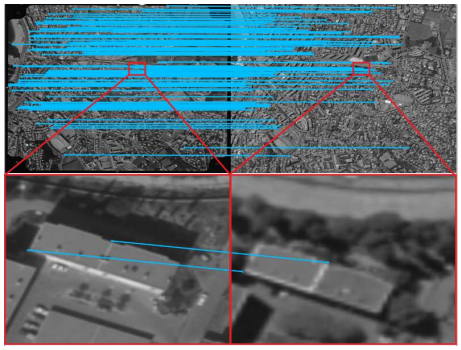
all the remaining orientations were considered as free parameters. At first, interior orientation parameters were shared among all images. Once stable initial values were known, interior parameters were further refined with image-dependent affine parameters. The affine component of the camera calibration is expected to model the shear of the analog films, at least partially.

6.3.1 Implementation details

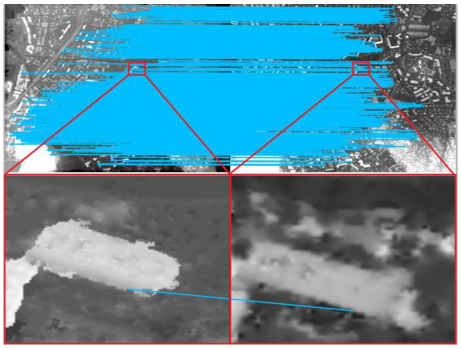
Same as Section 5.3.1, all input images are downsampled by a factor of 3 beforehand to improve efficiency except for dataset Alberona. To calculate the DSMs, we further downsample the images by a factor of 4 (different from 8 in Section 5.3.1), which amounts to a total downsampling factor of 12 with respect to the input images (total factor of 4 for Alberona). For example, the images in Fréjus 1970 are downsampled from [8766, 8763] to [730, 730] for calculating DSMs. Note that the DSMs serve 2 purposes in precise matching: (1) narrowing down the search space in finding tentative matches, and (2) providing 3D coordinates for 3D-RANSAC filter. A low resolution surface is good enough for these tasks and improves the efficiency.
For patch matching, the buffered tile size $SZ_{one-to-one}$ is set to be 640×480 pixels, the buffer size SZ_{buffer} is 10%×$SZ_{one-to-one}$ (i.e., widening 64 pixels on both left and right sides, 48 pixels on both upper and lower sides). Therefore, the original tile size $SZ^{Orig}_{one-to-one}$ is left to be 512×384 pixels. The tile pairs entering SuperGlue are not downsampled. For guided matching, the search radius S is set to be 100 pixels; the thresholds for checking scale and rotation (i.e. Th_{scl} and Th_{rot}) are set to be 0.2 and 30° individually. For the 3D-RANSAC procedure, we set the number of iteration to 1000, and T_r to 10×GSD where GSD is the mean ground sampling distance in the coordinate frame of reference epoch E_r. This distance is computed as the ground distance between two adjacent image pixels. The threshold T_c in cross correlation is set to be 0.6. To balance the number of the intra- and inter-epoch matches, we perform intra-epoch matches reduction available as command *Ratafia* in MicMac [Pierrot-Deseilligny *et al.* 2015]. If the intra-epoch matches after reduction are still obviously more than the inter-epoch ones, we further set the relative observation weight in the BBA. The matches reduction algorithm maximizes good spatial distribution, points' multiplicity and low reprojection error, it also helps to speed up the BBA.
Inter-epoch matches are extracted for every possible combination of 2 epochs and finally merged.

6.3.2 Comparison of precise matching on DSMs and original RGB images

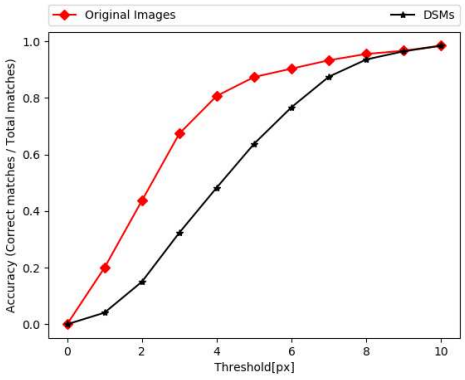
In order to decide which type of image (DSM or original RGB image) is more suitable for executing precise matching, we apply our variant *Patch* on both DSMs and RGB images of Fréjus 1970 and 2014 for comparison. The final matches are displayed in Figure 6.5 (a) and (b). To asses quantitatively the results, we created a GT depth map and calculated the accuracy (correct matches / total matches). In Figure 6.5 (c) we plot the accuracy curves while varying the reprojection error threshold from 0 to 10 pixels. Obviously the result using the RGB images is more accurate, even though the DSMs recovered more matches. This is because historical RGB images are inevitably accompanied with noise, and it gets worse in DSMs at full resolution (see the DSM shaded image in Figure 6.5 (d)) due to inevitably information loss and errors introduction during calculating the DSM. Therefore RGB images are more suitable for precise matching.



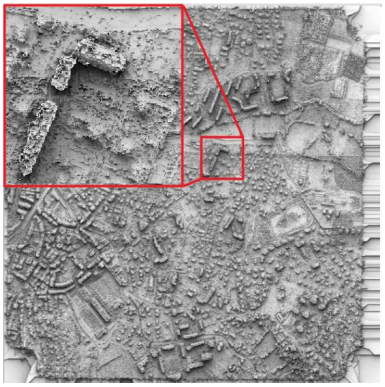
(a) Matches on RGB images



(b) Matches on DSMs



(c) Accuracy of (a) and (b)



(d) Shaded image of historical DSM

Figure 6.5: Comparison of precise matching on original RGB images and DSMs.

6.3.3 Comparison of 4 variants

In order to evaluate the results qualitatively and quantitatively, the following criteria would be applied:

1. **Matches visualization**. The number of tentative, enhanced and final matches are displayed together in bar charts; in the meantime, the final matches are visualized and demonstrated.

2. **DoD**. For each variant, the refined orientations would be used to calculate DSMs in order to generate DoD. The visualization of DoD as well as the statistical information are displayed. Since the orientations are refined with precise matches, DoDs with *dome* effect mitigated or even eliminated are expected.
   For Pezenas and Fréjus datasets, DoDs are calculated between historical epochs and the available GT epochs. For Kobe and Alberona datasets, there is no GT. Therefore we calculate the DoDs between every epochs instead.

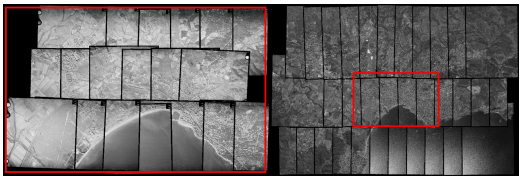
3. **Ground displacement**. For the dataset that witnessed an earthquake (i.e., Kobe), we: (1) calculate the DSMs; (2) orthorectify the images; and (3) perform 2D correlation of the respective orthophotos [Rosu *et al.* 2015] to see whether we can observe the slip of the tectonic plate.

For matches visualization and DoD, as the results show similar pattern on different datasets, we only display the results of Fréjus and Alberona in the current section for the sake of simplicity, and move the results of Kobe as well as Pezenas to Appendix B.

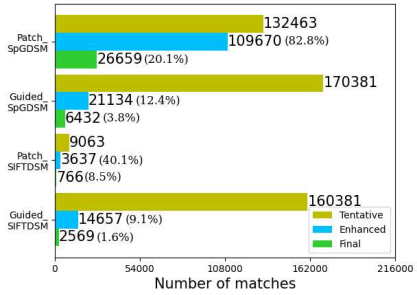
Matches visualization. For each dataset, we match every possible combination of 2 epochs with 4 variants (i.e., ① $Patch_{SpGDSM}$, ② $Guided_{SpGDSM}$, ③ $Patch_{SIFTDSM}$ and ④ $Guided_{SIFTDSM}$).

For Fréjus, there exist 4 epochs, leading to 6 sets of epoch combination. The visualizations of resulted matches are displayed in Figure 6.6, 6.7, 6.8, 6.9, 6.10 and 6.11.
For Alberona, there exist 2 epochs, leading to 1 set of epoch combination, the matches visualization is displayed in Figure 6.12.

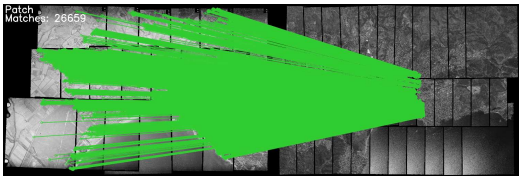
DoD. The DoDs for Fréjus and Alberona are demonstrated in Figure 6.13 and 6.14. In each figure, the roughly co-registered DoDs resulted from rough co-registration variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (elaborated in Chapter 5, hereinafter referred to as DoD^{SpGDSM} and $\text{DoD}^{SIFTDSM}$) are displayed as references, and the refined DoDs resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$,



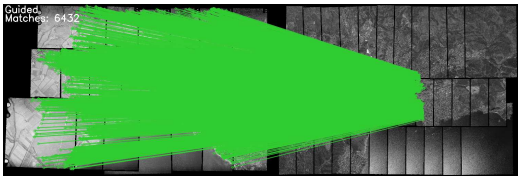
(a) Overlapping zone



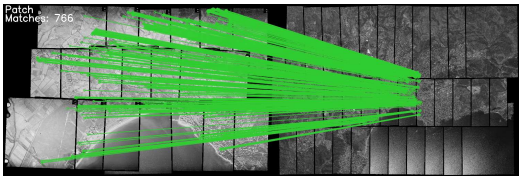
(b) Number of recovered matches



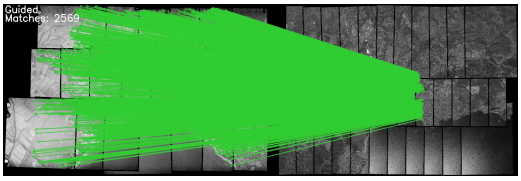
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$

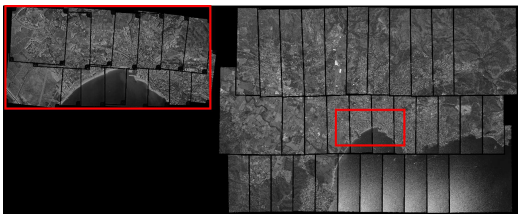


(e) $Patch_{SIFTDSM}$

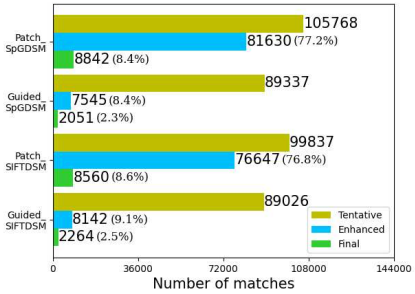


(f) $Guided_{SIFTDSM}$

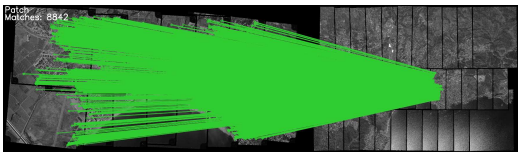
Figure 6.6: Precise matching visualization of **Fréjus 1954 and 2014**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



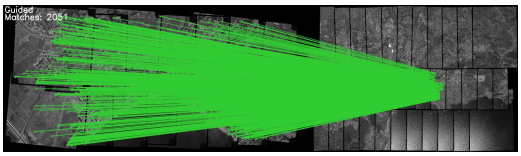
(a) Overlapping zone



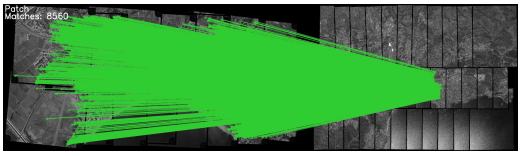
(b) Number of recovered matches



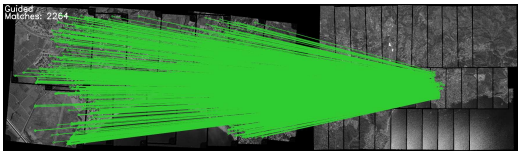
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$

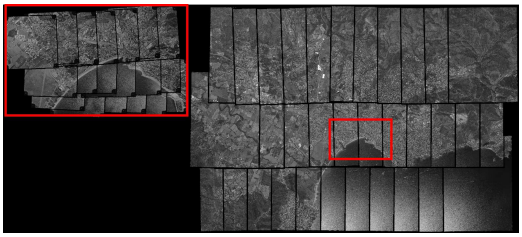


(e) $Patch_{SIFTDSM}$

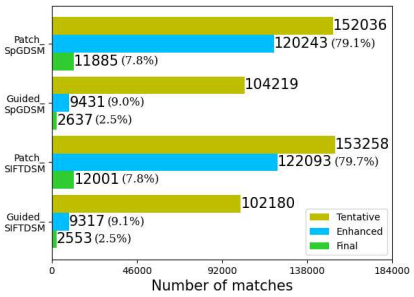


(f) $Guided_{SIFTDSM}$

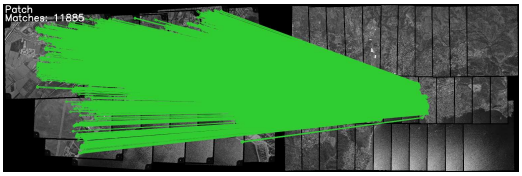
Figure 6.7: Precise matching visualization of **Fréjus 1966 and 2014**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



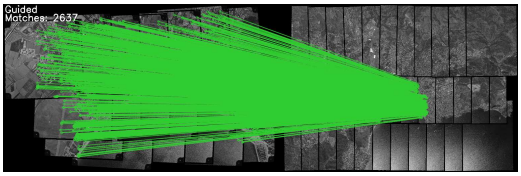
(a) Overlapping zone



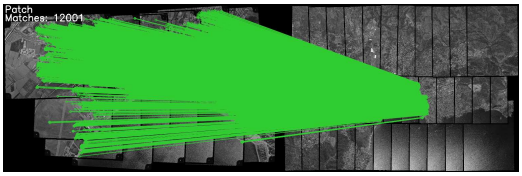
(b) Number of recovered matches



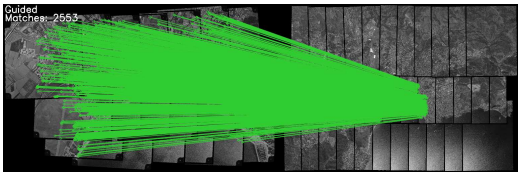
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$

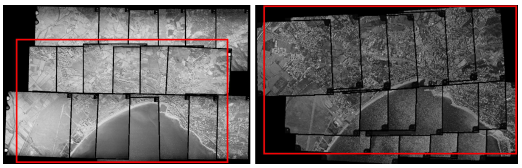


(e) $Patch_{SIFTDSM}$

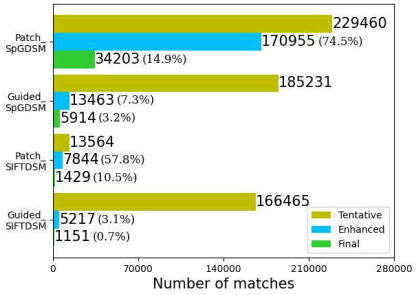


(f) $Guided_{SIFTDSM}$

Figure 6.8: Precise matching visualization of **Fréjus 1970 and 2014**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.

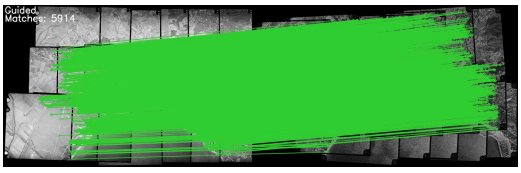


(a) Overlapping zone

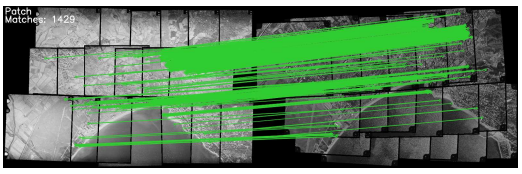


(b) Number of recovered matches

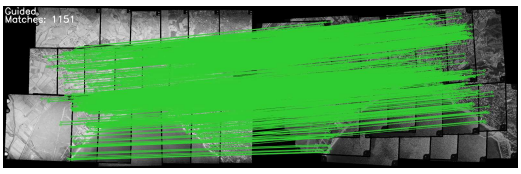
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$

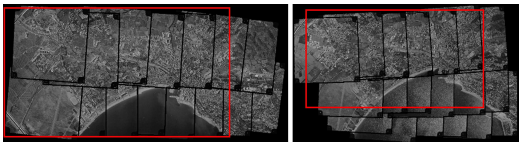


(e) $Patch_{SIFTDSM}$

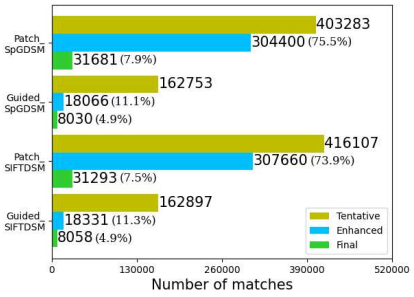


(f) $Guided_{SIFTDSM}$

Figure 6.9: Precise matching visualization of **Fréjus 1954 and 1970**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



(a) Overlapping zone



(b) Number of recovered matches

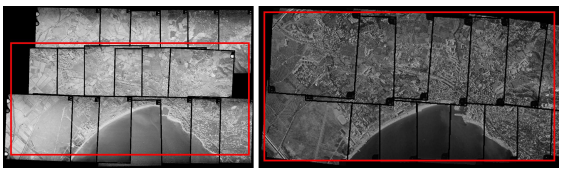
(c) $Patch_{SpGDSM}$

(d) $Guided_{SpGDSM}$

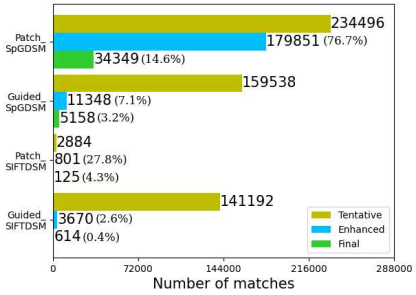
(e) $Patch_{SIFTDSM}$

(f) $Guided_{SIFTDSM}$

Figure 6.10: Precise matching visualization of **Fréjus 1966 and 1970**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



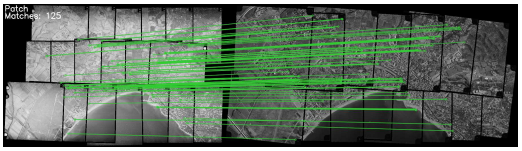
(a) Overlapping zone



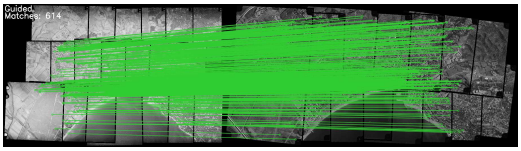
(b) Number of recovered matches

(c) $Patch_{SpGDSM}$

(d) $Guided_{SpGDSM}$

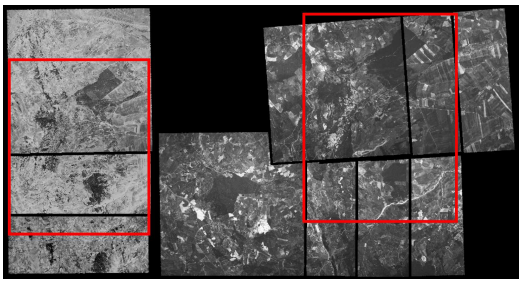


(e) $Patch_{SIFTDSM}$

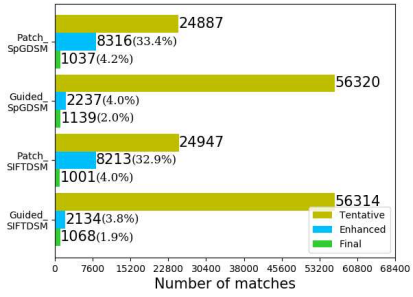


(f) $Guided_{SIFTDSM}$

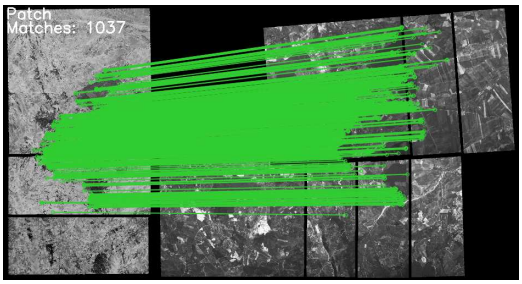
Figure 6.11: Precise matching visualization of **Fréjus 1954 and 1966**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



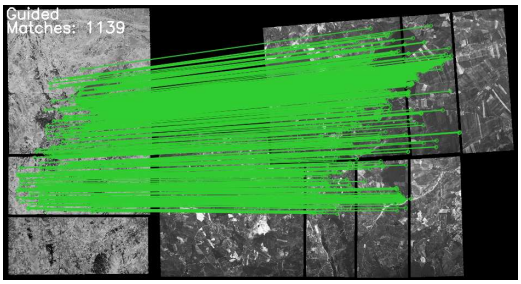
(a) Overlapping zone



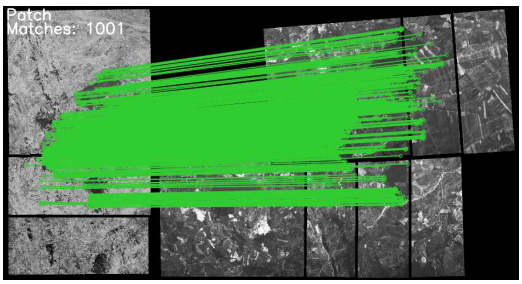
(b) Number of recovered matches



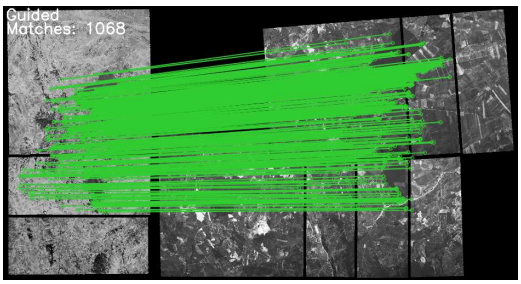
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$



(e) $Patch_{SIFTDSM}$



(f) $Guided_{SIFTDSM}$

Figure 6.12: Precise matching visualization of **Alberona 1954 and 2003**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.

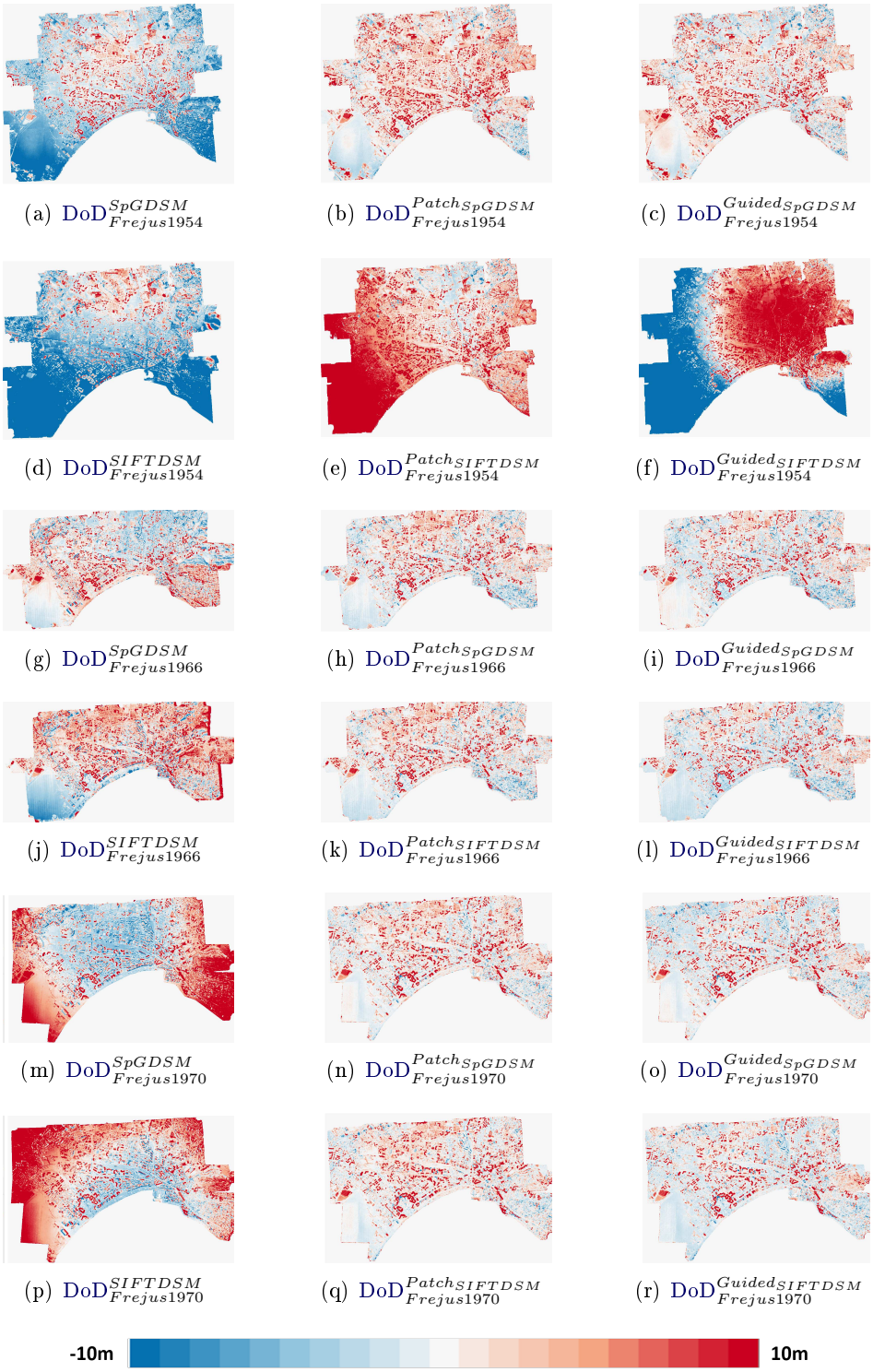
		μ [m]	σ [m]	$\lvert\mu\rvert$ [m]
$DoD^{Frejus}_{1954-2014}$	$SpGDSM$	2.24	5.34	4.35
	$Patch_{SpGDSM}$	-1.89	4.16	2.93
	$Guided_{SpGDSM}$	-1.29	4.03	**2.65**
	$SIFTDSM$	6.91	8.90	8.45
	$Patch_{SIFTDSM}$	-6.17	6.54	6.69
	$Guided_{SIFTDSM}$	3.72	17.12	12.90
$DoD^{Frejus}_{1966-2014}$	$SpGDSM$	-0.80	4.71	3.28
	$Patch_{SpGDSM}$	-0.93	4.00	2.57
	$Guided_{SpGDSM}$	-0.51	3.93	**2.45**
	$SIFTDSM$	-2.12	5.66	4.42
	$Patch_{SIFTDSM}$	-0.99	4.01	2.58
	$Guided_{SIFTDSM}$	-0.24	3.92	2.56
$DoD^{Frejus}_{1970-2014}$	$SpGDSM$	-2.37	6.57	5.35
	$Patch_{SpGDSM}$	-1.18	3.83	2.38
	$Guided_{SpGDSM}$	-0.46	3.83	**2.37**
	$SIFTDSM$	-3.76	6.21	5.47
	$Patch_{SIFTDSM}$	-1.28	3.92	2.45
	$Guided_{SIFTDSM}$	-0.43	3.79	2.40
$DoD^{Alberona}_{1954-2003}$	$SpGDSM$	-0.28	7.84	6.17
	$Patch_{SpGDSM}$	-0.80	4.44	3.22
	$Guided_{SpGDSM}$	-0.20	6.20	4.86
	$SIFTDSM$	-1.88	7.44	5.89
	$Patch_{SIFTDSM}$	-1.00	3.73	**2.75**
	$Guided_{SIFTDSM}$	2.11	5.86	4.63

Table 6.1: Average value μ, standard deviation σ, and absolute average value $|\mu|$ of all the DoDs in Figure 6.13 and 6.14.

$Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ (hereinafter termed as $\text{DoD}^{Patch_{SpGDSM}}$, $\text{DoD}^{Guided_{SpGDSM}}$, $\text{DoD}^{Patch_{SIFTDSM}}$ and $\text{DoD}^{Guided_{SIFTDSM}}$) are given for comparison. For the DoDs of Alberona, the extent of the landslide area is indicated with black lines based on the landslide inventory map, which is plotted by expert geomorphologists with visual interpretation of aerial photographs. The corresponding statistical information is displayed in Table 6.1.

Ground displacement (i.e., Gd). The northeastward Gd maps of Kobe dataset as well as the ground truth Gd provided by the Geospatial Information Authority of Japan are presented in Figure 6.15. The roughly co-registered Gds resulted from variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (i.e., Figure 6.15 (b) and (e), elaborated in Chapter 5) are displayed as references, and the refined Gds resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ (i.e., Figure 6.15 (c, d) and (f, g)) are given for comparison.

Discussion. As can be seen, both *Patch* and *Guided* recover a lot of matches, except for the ones involving epoch 1954 based on rough co-registration of $SIFT_{DSM}$ (Figure 6.6 (e, f), 6.9 (e, f) and 6.11 (e, f)). It is because the rough co-registration



(a) $DoD^{SpGDSM}_{Frejus1954}$ (b) $DoD^{Patch_{SpGDSM}}_{Frejus1954}$ (c) $DoD^{Guided_{SpGDSM}}_{Frejus1954}$

(d) $DoD^{SIFTDSM}_{Frejus1954}$ (e) $DoD^{Patch_{SIFTDSM}}_{Frejus1954}$ (f) $DoD^{Guided_{SIFTDSM}}_{Frejus1954}$

(g) $DoD^{SpGDSM}_{Frejus1966}$ (h) $DoD^{Patch_{SpGDSM}}_{Frejus1966}$ (i) $DoD^{Guided_{SpGDSM}}_{Frejus1966}$

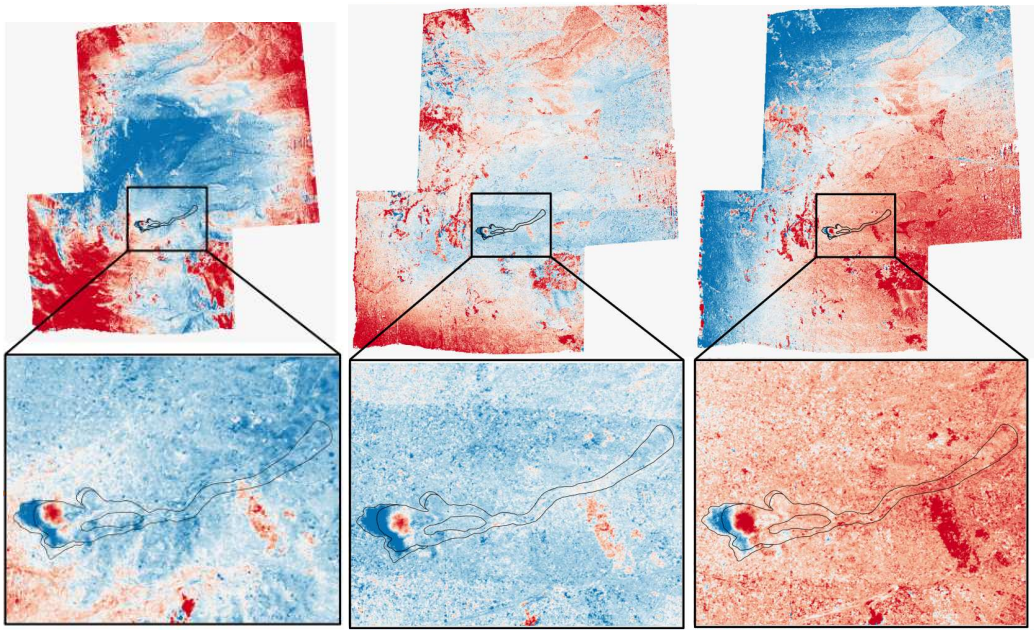
(j) $DoD^{SIFTDSM}_{Frejus1966}$ (k) $DoD^{Patch_{SIFTDSM}}_{Frejus1966}$ (l) $DoD^{Guided_{SIFTDSM}}_{Frejus1966}$

(m) $DoD^{SpGDSM}_{Frejus1970}$ (n) $DoD^{Patch_{SpGDSM}}_{Frejus1970}$ (o) $DoD^{Guided_{SpGDSM}}_{Frejus1970}$

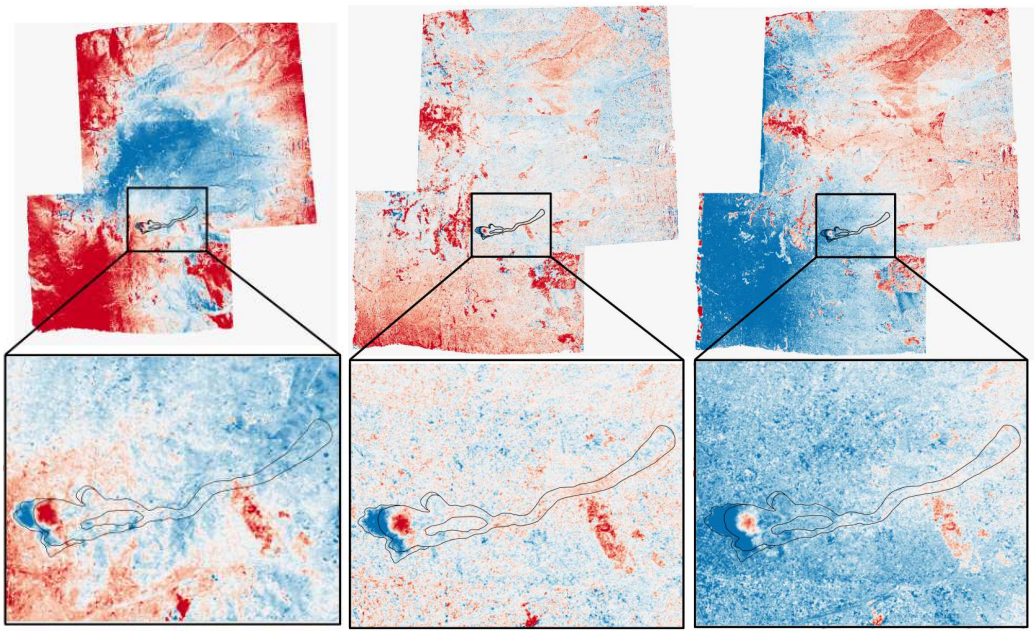
(p) $DoD^{SIFTDSM}_{Frejus1970}$ (q) $DoD^{Patch_{SIFTDSM}}_{Frejus1970}$ (r) $DoD^{Guided_{SIFTDSM}}_{Frejus1970}$

(s) DoD legend

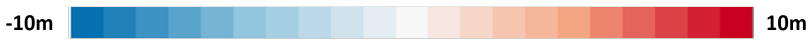
Figure 6.13: (a-f) DoDs between free epoch **Fréjus 1954** and reference epoch **2014**. (g-l) DoDs between free epoch **Fréjus 1966** and reference epoch **2014**. (m-r) DoDs between free epoch **Fréjus 1970** and reference epoch **2014**. (a, d, g, j, m, p) are roughly co-registered DoDs resulted from variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (elaborated in Chapter 5). (b, c, e, f, h, i, k, l, n, o, q, r) are refined DoDs resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



(a) $\text{DoD}^{SpGDSM}_{Alberona}$ (b) $\text{DoD}^{Patch_{SpGDSM}}_{Alberona}$ (c) $\text{DoD}^{Guided_{SpGDSM}}_{Alberona}$

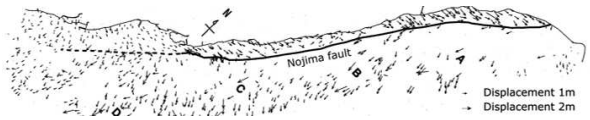


(d) $\text{DoD}^{SIFTDSM}_{Alberona}$ (e) $\text{DoD}^{Patch_{SIFTDSM}}_{Alberona}$ (f) $\text{DoD}^{Guided_{SIFTDSM}}_{Alberona}$



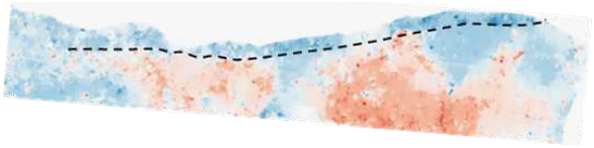
(g) DoD legend

Figure 6.14: DoDs between free epoch **Alberona 1954** and reference epoch **2003**. (a) and (d) are roughly co-registered DoDs resulted from variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (elaborated in Chapter 5). (b, c, e, f) are refined DoDs resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.

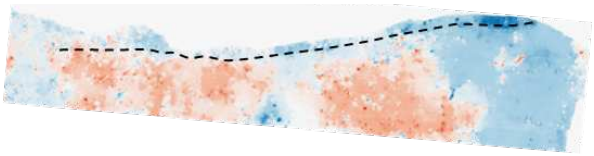


(a) Gd_{Kobe}^{GT}

(b) $\mathrm{Gd}_{Kobe}^{SpGDSM}$

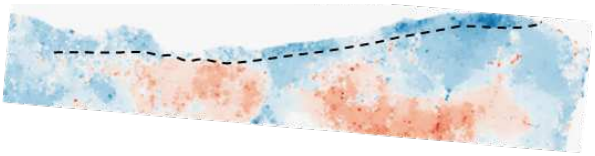


(c) $\mathrm{Gd}_{Kobe}^{Patch_{SpGDSM}}$

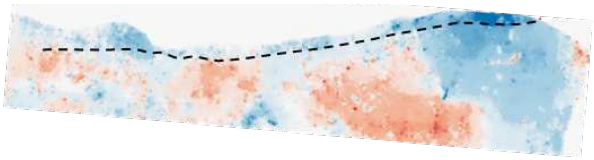


(d) $\mathrm{Gd}_{Kobe}^{Guided_{SpGDSM}}$

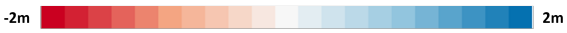
(e) $\mathrm{Gd}_{Kobe}^{SIFTDSM}$



(f) $\mathrm{Gd}_{Kobe}^{Patch_{SIFTDSM}}$



(g) $\mathrm{Gd}_{Kobe}^{Guided_{SIFTDSM}}$



(h) Gd legend

Figure 6.15: Ground displacement(Gd) between free epoch **Kobe 1991** and reference epoch **1995**. (a) is the ground truth Gd provided by the Geospatial Information Authority of Japan. (b) and (e) are roughly co-registered Gds resulted from variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (elaborated in Chapter 5). (c, d, f, g) are refined Gds resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.

result of $SIFT_{DSM}$ for epoch 1954 is unsatisfactory, as was mentioned in Section 5.3.5. Besides, 3D-RANSAC filter and cross correlation removed a considerable number of matches, at the same time enough matches survived, which guaranteed robustness of our method.

For the DoDs, the *dome* effect appears in all the DoD^{SpGDSM} and $\text{DoD}^{SIFTDSM}$ (i.e., the first column of the subgraphs in Figure 6.13 and 6.14), as the camera parameters are poorly estimated without the precise matches.
For most refined DoDs in Fréjus (i.e., the second and third columns of subgraphs in Figure 6.13 except for (e) and (f)), the *dome* effect is effectively mitigated, thanks to our numerous and precise matches. In the meantime, the real scene changes are preserved, such as the new buildings and seaports.
For Figure 6.13 (e) $\text{DoD}^{Patch_{SIFTDSM}}_{Frejus1954}$ and (f) $\text{DoD}^{Guided_{SIFTDSM}}_{Frejus1954}$, the *dome* effect is even worse than the roughly co-registered one (i.e., Figure 6.13 (d) $\text{DoD}^{SIFTDSM}_{Frejus1954}$) due to low quality of matches shown in Figure 6.6 (e, f), 6.9 (e, f) and 6.11 (e, f).
For the refined DoDs in Alberona, the *dome* effect is mitigated for the variant *Patch*, but not for *Guided*, as the images from different epochs showed various tone, which is challenging for *Guided*. Besides, the images are poorly preserved and scanned with non photogrammetric scanner, and limited number of images leads to a lack of redundant observation. Therefore, only the $\text{DoD}^{Patch_{SIFTDSM}}_{Alberona}$ (i.e., Figure 6.14 (e)) showed both useful signs in the landslide zone and limited systematic errors in the whole block, as it is based on good matches recovered with *Patch* variant under the good rough co-registration resulted from $SIFT_{DSM}$.

For the Gds, we displayed the GT in Figure 6.15 (a), and ours in Figure 6.15 (b-g). The GT is produced with a lot of manual work, which is very laborious. However, ours are completely automatic. According to the GT, there is a nojima fault caused by the earthquake (i.e. indicated as black line), and the displacement is indicated as small arrows. On the upper side of the fault, the arrows are generally rightward, while on the lower side they are leftward. In ours, the displacement is indicated as colors: blue represents rightward and red leftward. As can be seen, ours after refinement with precise matching (i.e. Figure 6.15 (d-g)) recovered the same signal as ground truth, which is rightward on the upper side of the fault (i.e. indicated as dashed line) and leftward on the lower side.

6.4 Conclusion

In this section we elaborate two variants for precise matching: *Patch* and *Guided*. We test each variant based on two sets of rough co-registration results: $SIFT_{DSM}$ and $SuperGlue_{DSM}$, which leads to 4 variants (i.e., ① $Patch_{SpGDSM}$, ② $Guided_{SpGDSM}$, ③ $Patch_{SIFTDSM}$ and ④ $Guided_{SIFTDSM}$. Experiments are performed on 4 sets of datasets (Fréjus, Pezenas, Kobe and Alberona), including the cases of (1) matching aerial epochs only and (2) matching aerial and satellite

epochs mixed. Experiments show that:

1. Both precise matching variants (i.e., *Patch* and *Guided*) are capable of recovering numerous and accurate matches, as long as the rough co-registration result is reliable.

2. By adopting the precise matches in a BBA routine, the systematic errors in the surfaces can be effectively mitigated while the real scene changes stay.

Chapter 7

Conclusion and Perspective

Contents

7.1 Conclusion

During the past decades, a large number of historical images were digitized, which signified huge potential for long-term environmental monitoring studies. Unfortunately, their value is overlooked as they are accompanied with special characteristics: analog films were probably inappropriately conserved, leading to poor radiometric quality; deformation caused by scanning; different resolutions and acquisition conditions, etc. The principal difficulty in processing multi-epoch historical images is feature matching. Often, no *a priori* about the camera geometry is available and a dense distribution of matches is required to model it a *posteriori*. Even though we have seen an emergence of software solutions capable of processing modern digital images in a 100% automated manner, the performance of these solutions degenerates when applied to multi-epoch historical images.

The thesis aims at matching historical images as well as modern digital images taken at different times. The goal is accomplished with the divide and conquer strategy, which is to decompose the task of recovering robust and precise matches on inter-epoch image pairs into 2 sub-tasks: (1) rough co-registration focusing on robustness, and (2) precise matching on accuracy.

Five representative sets of datasets for different applications are introduced in order to validate the suitability of our pipelines for various domains. They consist of mixed images (i.e., historical and modern, aerial and satellite images) with heterogeneous acquisition conditions.

Different strategies for rough co-registration are studied. The first attempt we made is matching each inter-epoch image pair separately followed by building a globally consistent transformation model over the whole block. It is not efficient and robust enough, leading us to another strategy: combining images from the

same epoch into entirety (i.e., orthophoto or DSM) and applying matching directly on the whole block. All the strategies are tested on four sets of multi-epoch datasets, based on which we come to a conclusion that the strategy of matching DSM provides the most robust results. Besides, different configurations of matching methods (i.e., SIFT and SuperGlue) are compared, and a use case of matching guided by 2D similarity transformation is presented.

Then, we propose and evaluate precise matching under the guidance of rough co-registration. Two variants are explored for obtaining tentative matches: (1) patch matching orientated towards learned features and (2) guided matching focused on hand-crafted features, followed by 3D-RANSAC and cross correlation to remove false matches. The most robust variants for rough co-registration (i.e., matching DSMs with SIFT and SuperGlue respectively) are chosen to guide the precise matching in the experiments, based on which we conclude that both patch and guided matching are capable of recovering a large number of accurate and robust matches as long as the rough co-registration result is reliable. Besides, comparison of precise matching on DSMs and original RGB images is performed to explain why we choose RGB images over DSMs for precise matching.

7.2 Perspective

Historical dataset benchmark There are a lot of benchmark datasets for feature matching, but none of them are multi-epoch historical images. In order to push forward the *state-of-the-art* in multi-epoch historical image processing, in the future we are interested in publishing the datasets used in this thesis, as well as collaborating with other scholars who are interested in processing historical images to build an open-access historical dataset benchmark (i.e., MultiHist). It should contain different scenes accompanied with ground truth orientations and DSMs, or even GCPs if possible. Different scenes consist of several epochs, probably organized as Figure 7.1.

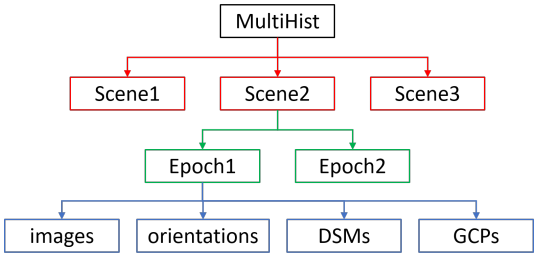
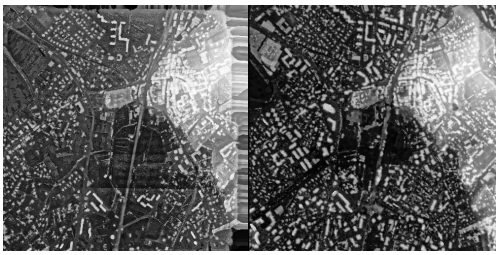


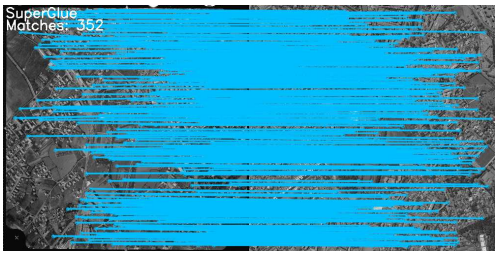
Figure 7.1: Organization of our benchmark.

Train a network with RGB images combined with DSMs Another direction of our future work is to use both RGB images and DSMs to train a neural network architecture in extracting robust features over time. As training data made from historical images is limited, it might be better to fine-tune existing models (e.g., SuperGlue). In order to validate if it improves matching performance to use RGB images and DSMs at the same time, we did a comparison of using off-the-shelf SuperGlue model to match (1) RGB images only, (2) corresponding full resolution DSMs only and (3) RGB images combined with DSMs by concatenating keypoints. We choose a pair of roughly aligned images and feed them directly into SuperGlue without applying any *tiling scheme* to keep the performance independent from irrelevant factors. The results are displayed in Figure 7.2(c), (d) and (e) respectively, with their accuracy compared in Figure 7.2(f). As can be seen, it provides more matches with better accuracy when simply feeding concatenated keypoints to the ready-made model, it is reasonable to expect better performance after we fine-tune the model.

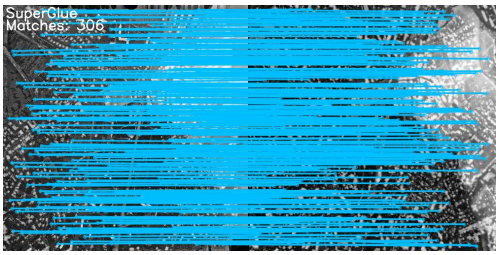
(a) RGB image pair



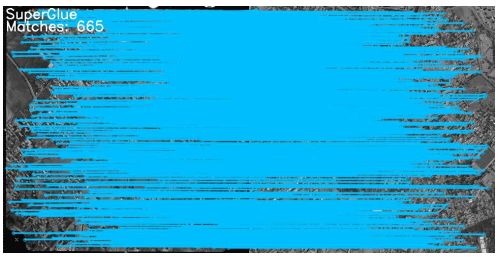
(b) DSM pair



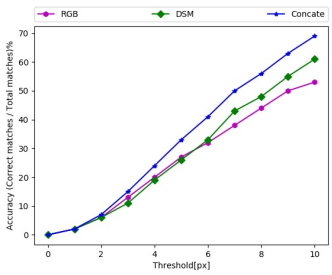
(c) Matches on RGB images



(d) Matches on DSMs



(e) Matches on concatenation



(f) Accuracy of (c-e)

Figure 7.2: Comparison of SuperGlue applied on RGB images (c), DSMs (d) and combined input by concatenating the keypoints (e).

Appendices

Appendix A

Result of rough co-registration

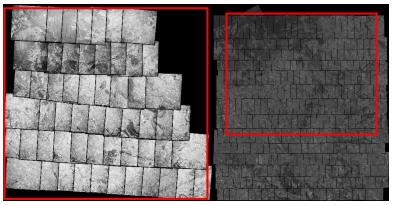
In Section 5.3.5, we demonstrate the matches visualization and DoDs of rough co-registration results on representative datasets Fréjus and Alberona. In this section the remaining results on other datasets (i.e., Pezenas and Kobe) are displayed.

A.1 Matches visualization

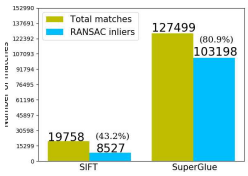
For Pezenas, there are 2 reference epochs E_r: aerial E_r (i.e., epoch 2015) and satellite E_r (i.e., epoch 2014). The matches visualizations between free epochs E_f (i.e., epoch 1971 and 1981) and both E_r are displayed in Figure A.1, A.2, A.3 and A.4. For Kobe, the reference epoch E_r is 1995, the matches visualizations between free epoch E_f (i.e., epoch 1991) and E_r are displayed in Figure A.5.
As can be seen, for both free epochs E_f in Pezenas, even the weakest variants $SIFT_{ImgPairs}$ and $SIFT_{Ortho}$ succeeded on matching them to the aerial reference epoch E_r, since neither drastic scene changes nor different image tones is shown in Pezenas, which lowers the matching difficulty. However, $SIFT_{Ortho}$ fails on satellite reference epoch E_r, as satellite epoch not only has more limited overlapping zone with the free epochs, especially for epoch 1971, but also is covered with clouds. For the rest 4 variants, we can see the same pattern as the results of Fréjus and Alberona (c.f., Section 5.3.5).

A.2 DoD

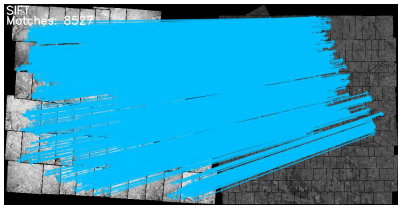
The visualizations of DoDs for datasets Pezenas and Kobe are displayed in Figure A.6, A.7 and A.8. The corresponding statistical information is given in Table A.1.
As can be seen, different epochs are roughly aligned with *dome* effect present in all the DoDs due to poorly estimated camera parameters, same pattern as the results of Fréjus and Alberona (c.f., Section 5.3.5).



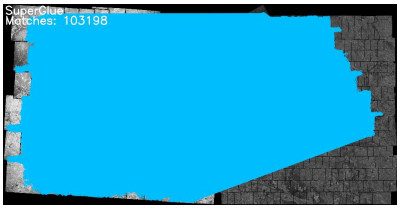
(a) Image pairs (57×382 pairs)



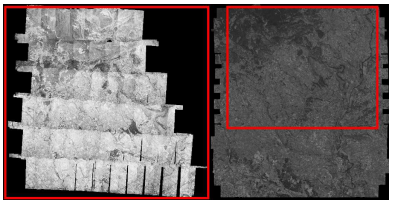
(b) Number of recovered matches(*ImgPairs*)



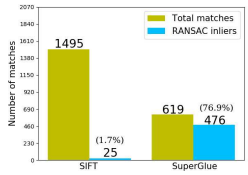
(c) $SIFT_{ImgPairs}^{RANSACInliers}$



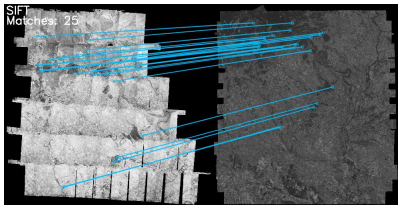
(d) $SuperGlue_{ImgPairs}^{RANSACInliers}$



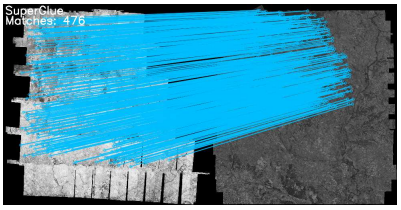
(e) Orthophotos



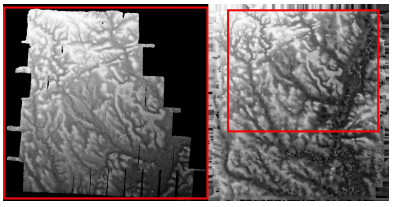
(f) Number of recovered matches(*Ortho*)



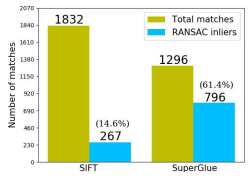
(g) $SIFT_{Ortho}^{RANSACInliers}$



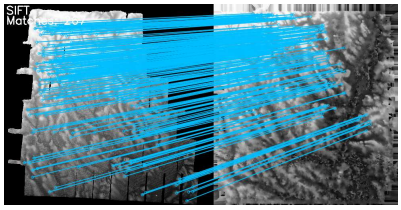
(h) $SuperGlue_{Ortho}^{RANSACInliers}$



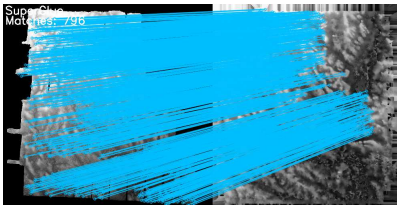
(i) DSMs



(j) Number of recovered matches(*DSM*)

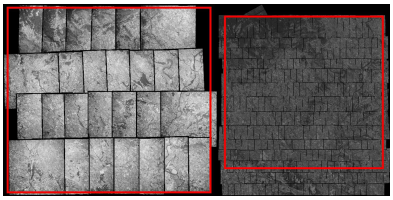


(k) $SIFT_{DSM}^{RANSACInliers}$

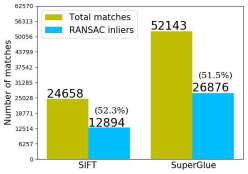


(l) $SuperGlue_{DSM}^{RANSACInliers}$

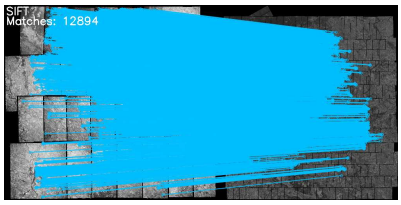
Figure A.1: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Pezenas 1971 and 2015**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



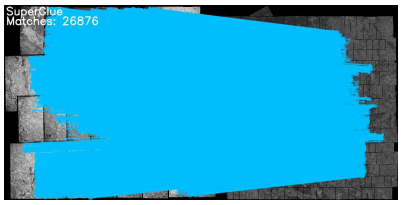
(a) Image pairs (27×382 pairs)



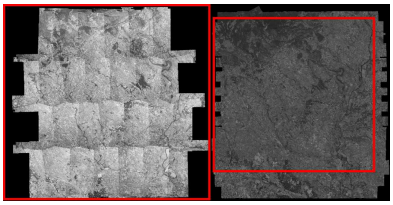
(b) Number of recovered matches(*ImgPairs*)



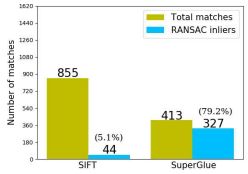
(c) $SIFT^{RANSACInliers}_{ImgPairs}$



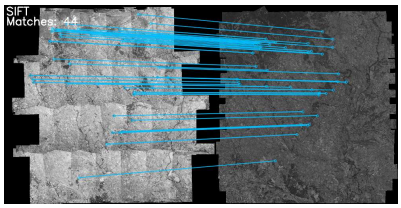
(d) $SuperGlue^{RANSACInliers}_{ImgPairs}$



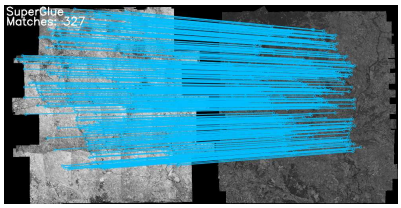
(e) Orthophotos



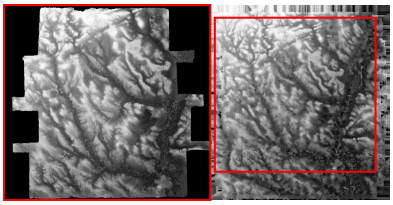
(f) Number of recovered matches(*Ortho*)



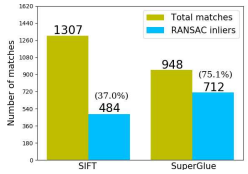
(g) $SIFT^{RANSACInliers}_{Ortho}$



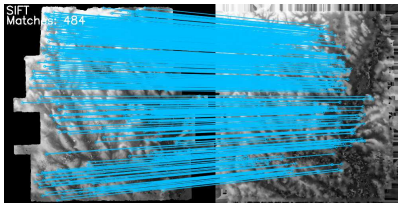
(h) $SuperGlue^{RANSACInliers}_{Ortho}$



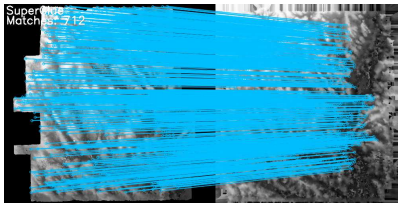
(i) DSMs



(j) Number of recovered matches(*DSM*)

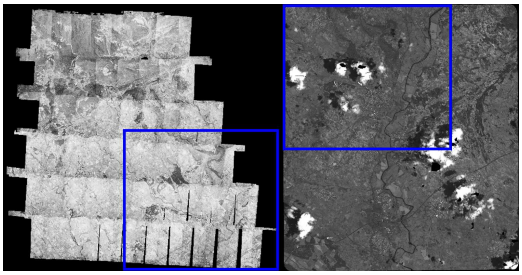


(k) $SIFT^{RANSACInliers}_{DSM}$

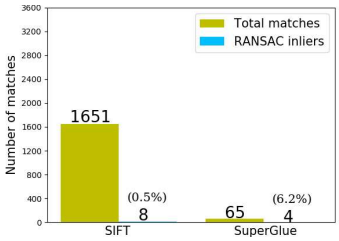


(l) $SuperGlue^{RANSACInliers}_{DSM}$

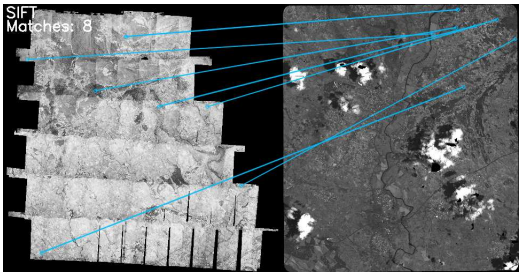
Figure A.2: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Pezenas 1981 and 2015**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



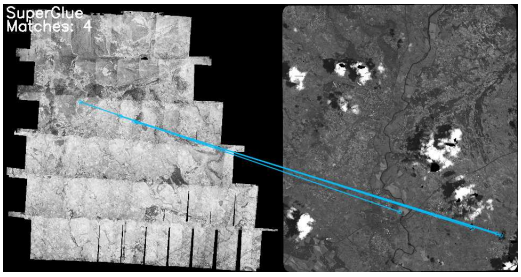
(a) Orthophotos



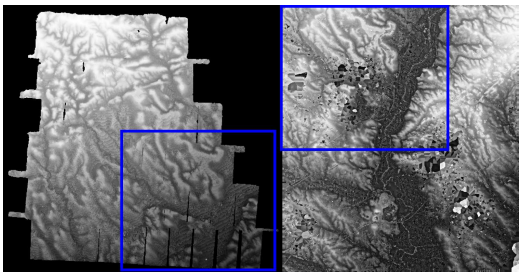
(b) Number of recovered matches(*Ortho*)



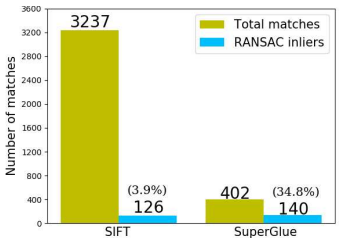
(c) $SIFT_{Ortho}^{RANSACInliers}$



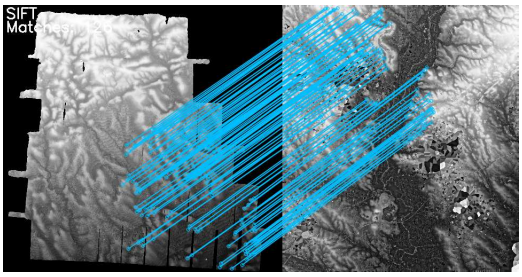
(d) $SuperGlue_{Ortho}^{RANSACInliers}$



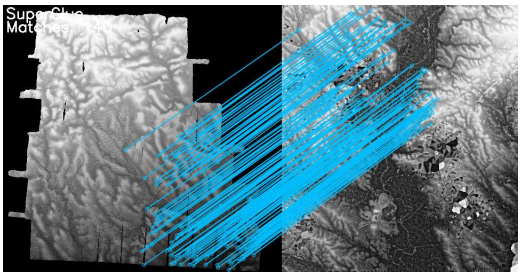
(e) DSMs



(f) Number of recovered matches(*DSM*)

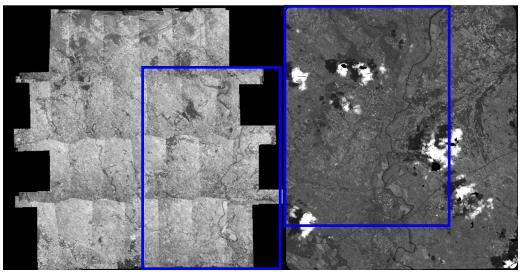


(g) $SIFT_{DSM}^{RANSACInliers}$

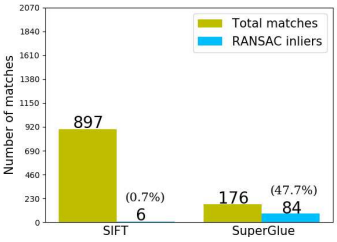


(h) $SuperGlue_{DSM}^{RANSACInliers}$

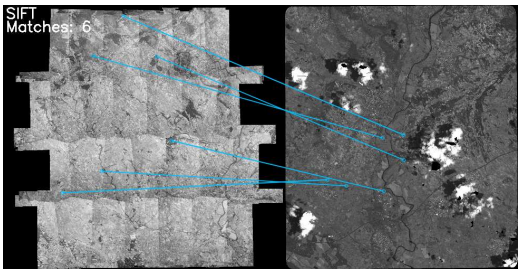
Figure A.3: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Pezenas 1971 and 2014 (Satellite)**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



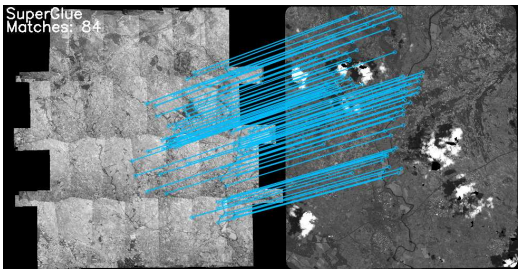
(a) Orthophotos



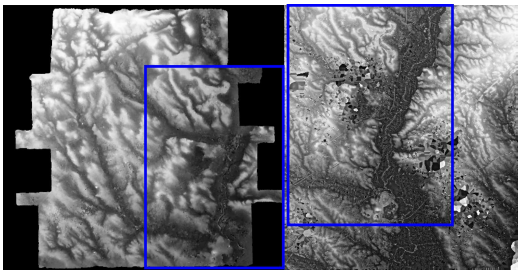
(b) Number of recovered matches(*Ortho*)



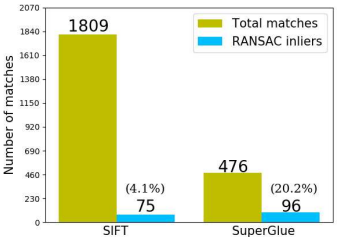
(c) $SIFT_{Ortho}^{RANSACInliers}$



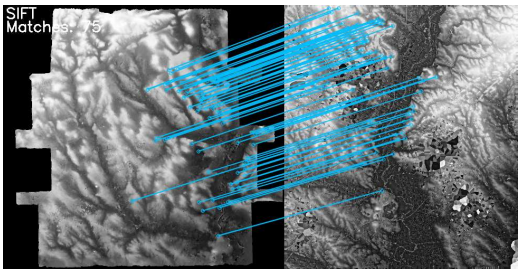
(d) $SuperGlue_{Ortho}^{RANSACInliers}$



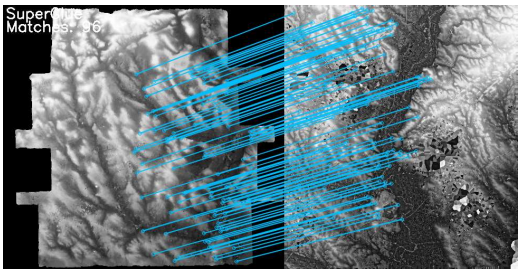
(e) DSMs



(f) Number of recovered matches(*DSM*)

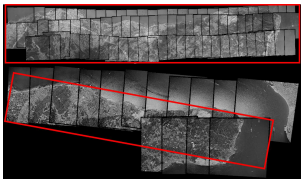


(g) $SIFT_{DSM}^{RANSACInliers}$

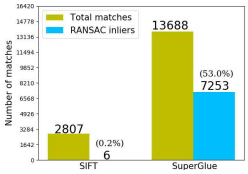


(h) $SuperGlue_{DSM}^{RANSACInliers}$

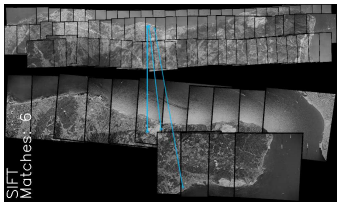
Figure A.4: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Pezenas 1981 and 2014 (Satellite)**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



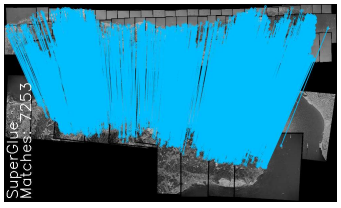
(a) Image pairs (15×83 pairs)



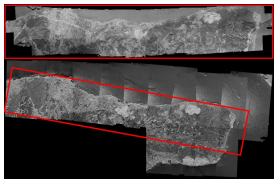
(b) Number of recovered matches(*ImgPairs*)



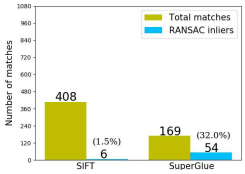
(c) $SIFT^{RANSACInliers}_{ImgPairs}$



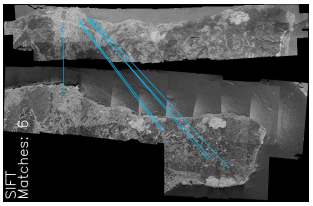
(d) $SuperGlue^{RANSACInliers}_{ImgPairs}$



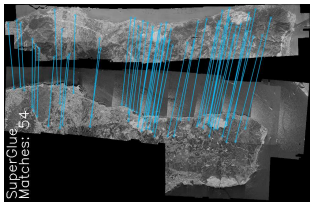
(e) Orthophotos



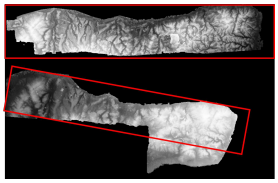
(f) Number of recovered matches(*Ortho*)



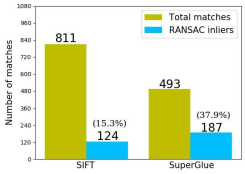
(g) $SIFT^{RANSACInliers}_{Ortho}$



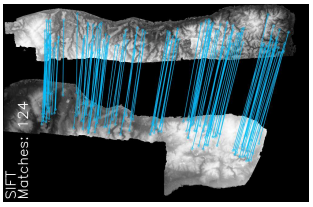
(h) $SuperGlue^{RANSACInliers}_{Ortho}$



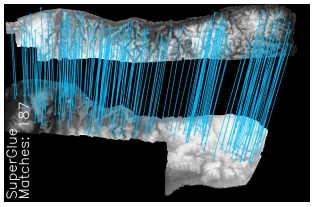
(i) DSMs



(j) Number of recovered matches(*DSM*)

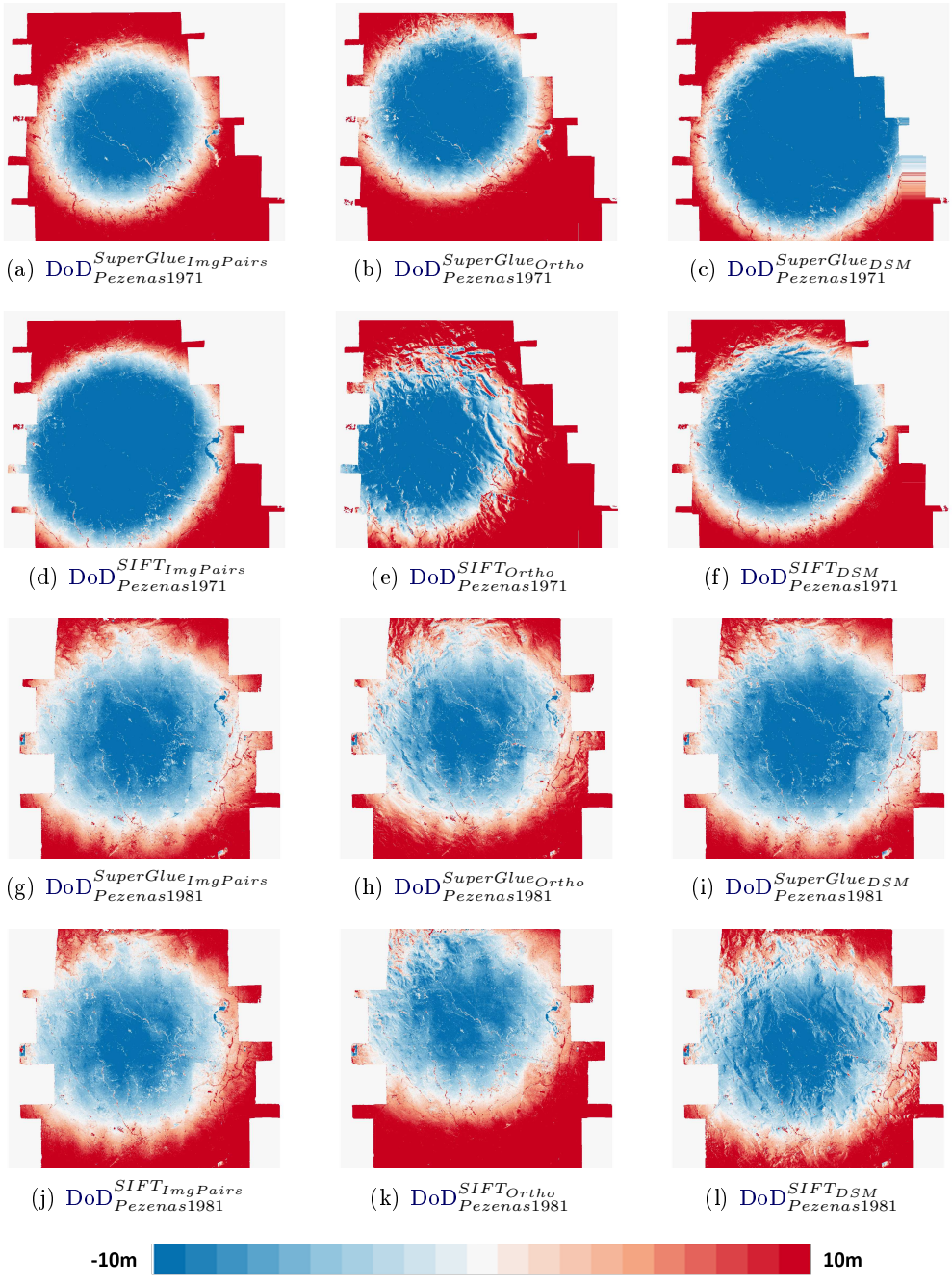


(k) $SIFT^{RANSACInliers}_{DSM}$



(l) $SuperGlue^{RANSACInliers}_{DSM}$

Figure A.5: Result of *ImgPairs* (a-d), *Ortho* (e-h) and *DSM* (i-l) on matching **Kobe 1991 and 1995**. (a, e, i) Image pairs/orthophotos/DSMs to be matched, with red rectangles indicating the overlapping zone. (b, f, j) Numbers of total matches and RANSAC inliers of both SIFT and SuperGlue on variants *ImgPairs*, *Ortho* and *DSM* individually. (c, g, k) Visualization of RANSAC inliers based on $SIFT_{ImgPairs}$, $SIFT_{Ortho}$ and $SIFT_{DSM}$. (d, h, l) Visualization of RANSAC inliers based on $SuperGlue_{ImgPairs}$, $SuperGlue_{Ortho}$ and $SuperGlue_{DSM}$.



(a) DoD$^{SuperGlue_{ImgPairs}}_{Pezenas1971}$ (b) DoD$^{SuperGlue_{Ortho}}_{Pezenas1971}$ (c) DoD$^{SuperGlue_{DSM}}_{Pezenas1971}$

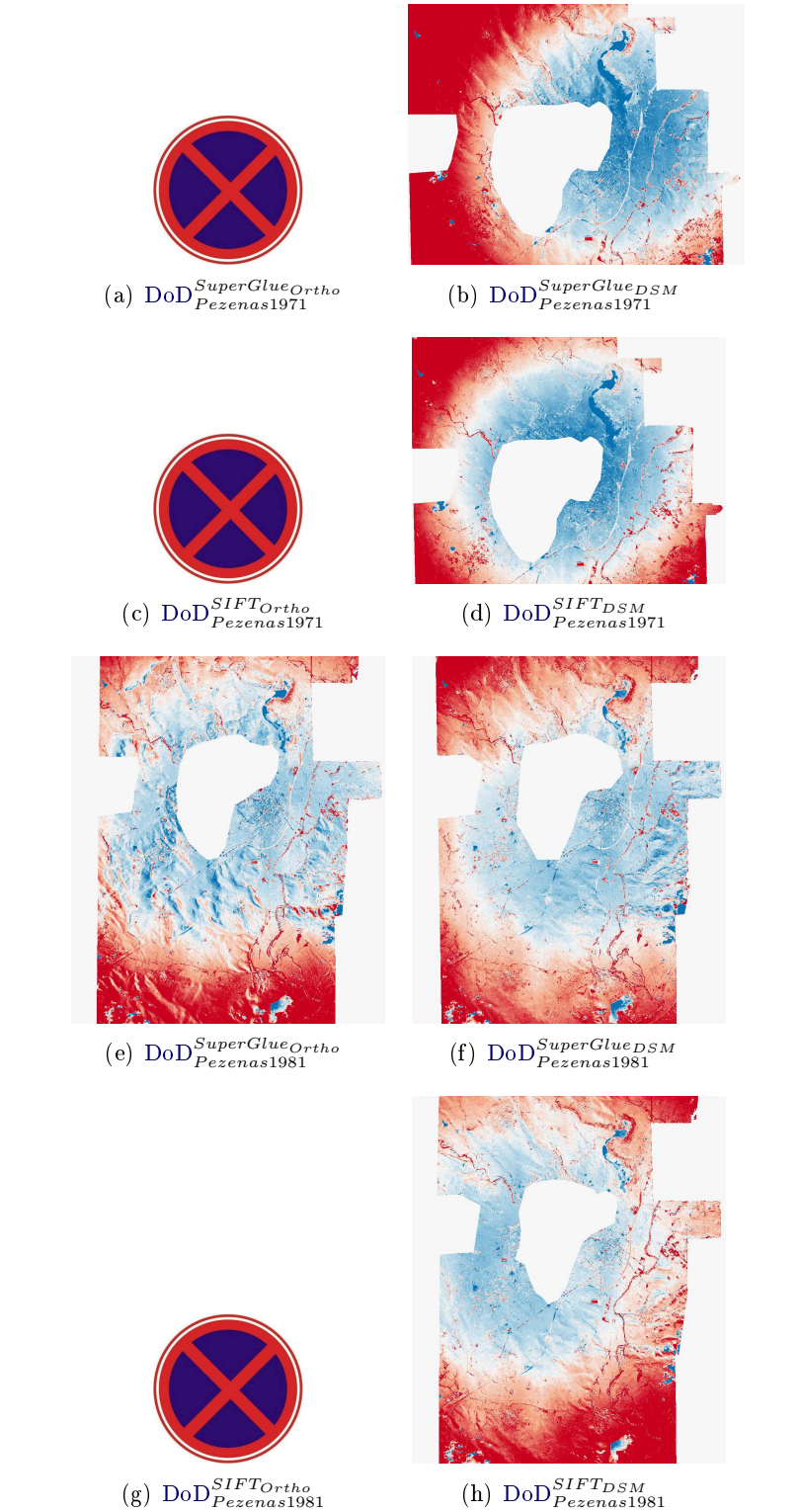
(d) DoD$^{SIFT_{ImgPairs}}_{Pezenas1971}$ (e) DoD$^{SIFT_{Ortho}}_{Pezenas1971}$ (f) DoD$^{SIFT_{DSM}}_{Pezenas1971}$

(g) DoD$^{SuperGlue_{ImgPairs}}_{Pezenas1981}$ (h) DoD$^{SuperGlue_{Ortho}}_{Pezenas1981}$ (i) DoD$^{SuperGlue_{DSM}}_{Pezenas1981}$

(j) DoD$^{SIFT_{ImgPairs}}_{Pezenas1981}$ (k) DoD$^{SIFT_{Ortho}}_{Pezenas1981}$ (l) DoD$^{SIFT_{DSM}}_{Pezenas1981}$

(m) DoD legend

Figure A.6: DoDs between free epoch **Pezenas 1971, 1981** and reference aerial epoch **2015** with variants $SuperGlue_{ImgPairs}$ (a, g), $SuperGlue_{Ortho}$ (b, h), $SuperGlue_{DSM}$ (c, i), $SIFT_{ImgPairs}$ (d, j), $SIFT_{Ortho}$ (e, k) and $SIFT_{DSM}$ (f, l). The prohibition sign means the corresponding variant failed.



(a) $\mathrm{DoD}^{SuperGlue_{Ortho}}_{Pezenas1971}$ (b) $\mathrm{DoD}^{SuperGlue_{DSM}}_{Pezenas1971}$

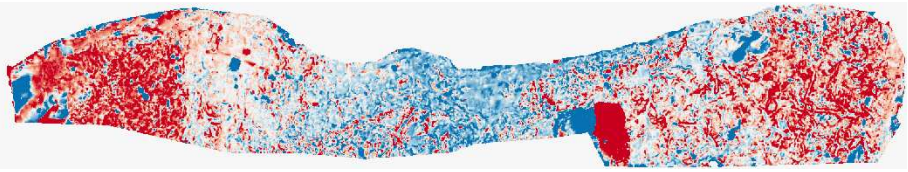
(c) $\mathrm{DoD}^{SIFT_{Ortho}}_{Pezenas1971}$ (d) $\mathrm{DoD}^{SIFT_{DSM}}_{Pezenas1971}$

(e) $\mathrm{DoD}^{SuperGlue_{Ortho}}_{Pezenas1981}$ (f) $\mathrm{DoD}^{SuperGlue_{DSM}}_{Pezenas1981}$

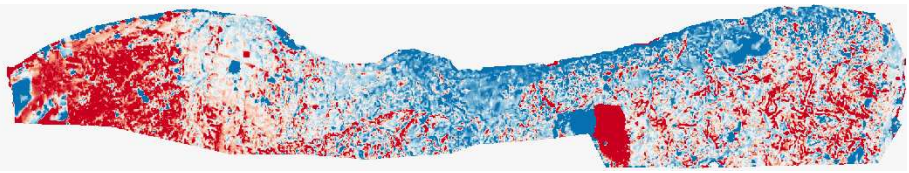
(g) $\mathrm{DoD}^{SIFT_{Ortho}}_{Pezenas1981}$ (h) $\mathrm{DoD}^{SIFT_{DSM}}_{Pezenas1981}$

(i) DoD legend

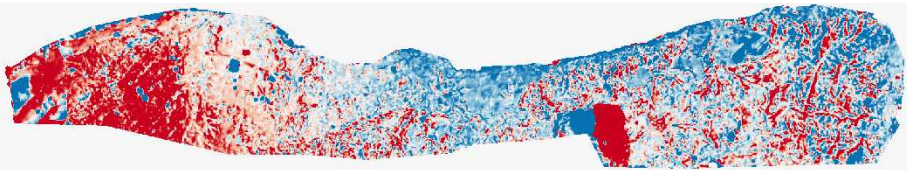
Figure A.7: DoDs between free epoch **Pezenas 1971, 1981** and reference satellite epoch **2014** with variants $SuperGlue_{Ortho}$ (a, e), $SuperGlue_{DSM}$ (b, f), $SIFT_{Ortho}$ (c, g) and $SIFT_{DSM}$ (d, h). The holes among them are areas covered with clouds which are masked out. The prohibition sign means the corresponding variant failed.



(a) $\text{DoD}_{Kobe}^{SuperGlue_{ImgPairs}}$



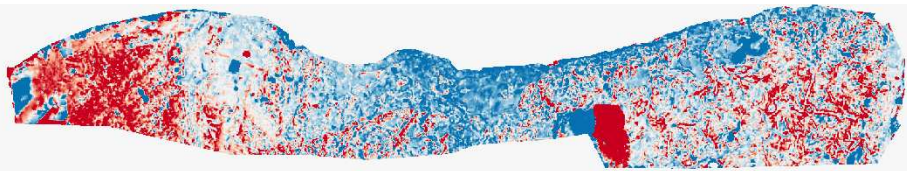
(b) $\text{DoD}_{Kobe}^{SuperGlue_{Ortho}}$



(c) $\text{DoD}_{Kobe}^{SuperGlue_{DSM}}$

(d) $\text{DoD}_{Kobe}^{SIFT_{ImgPairs}}$

(e) $\text{DoD}_{Kobe}^{SIFT_{Ortho}}$



(f) $\text{DoD}_{Kobe}^{SIFT_{DSM}}$

(g) DoD legend

Figure A.8: DoDs between free epoch **Kobe 1991** and reference epoch **1995** with variants $SuperGlue_{ImgPairs}$ (a), $SuperGlue_{Ortho}$ (b), $SuperGlue_{DSM}$ (c), $SIFT_{ImgPairs}$ (d), $SIFT_{Ortho}$ (e) and $SIFT_{DSM}$ (f). The prohibition sign means the corresponding variant failed.

		μ [m]	σ [m]	$\lvert\mu\rvert$ [m]
$DoD^{Pezenas}_{1971-2015}$	$SuperGlue_{ImgPairs}$	-7.46	16.20	**13.56**
	$SuperGlue_{Ortho}$	-8.25	21.89	16.47
	$SuperGlue_{DSM}$	4.73	16.06	14.03
	$SIFT_{ImgPairs}$	2.54	17.86	13.98
	$SIFT_{Ortho}$	-8.39	22.78	17.29
	$SIFT_{DSM}$	-0.85	17.81	**13.56**
$DoD^{Pezenas}_{1981-2015}$	$SuperGlue_{ImgPairs}$	-0.98	19.74	7.40
	$SuperGlue_{Ortho}$	-1.70	9.17	7.30
	$SuperGlue_{DSM}$	0.96	8.42	**6.97**
	$SIFT_{ImgPairs}$	-2.02	9.44	7.52
	$SIFT_{Ortho}$	-4.82	12.76	10.05
	$SIFT_{DSM}$	-0.72	8.96	7.21
$DoD^{Pezenas}_{1971-2014(Satellite)}$	$SuperGlue_{Ortho}$	/	/	/
	$SuperGlue_{DSM}$	-4.35	12.48	8.45
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	-1.45	11.24	**5.78**
$DoD^{Pezenas}_{1981-2014(Satellite)}$	$SuperGlue_{Ortho}$	-1.92	6.77	5.06
	$SuperGlue_{DSM}$	-1.81	5.85	**4.58**
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	-2.76	6.59	5.03
$DoD^{Kobe}_{1991-1995}$	$SuperGlue_{ImgPairs}$	-1.63	13.85	**7.24**
	$SuperGlue_{Ortho}$	-0.54	14.83	7.78
	$SuperGlue_{DSM}$	-0.75	14.62	7.95
	$SIFT_{ImgPairs}$	/	/	/
	$SIFT_{Ortho}$	/	/	/
	$SIFT_{DSM}$	0.27	14.40	7.57

Table A.1: Average value μ, standard deviation σ, and absolute average value $|\mu|$ of all the DoDs in Figure A.6, A.7 and A.8.

Appendix B

Result of precise matching

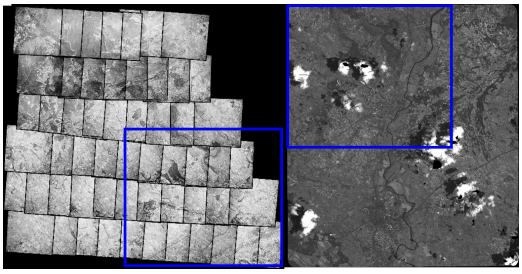
In Section 6.3.3, we demonstrate the matches visualization and DoDs of precise matching results on representative datasets Fréjus and Alberona. In this section the remaining results on other datasets (i.e., Pezenas and Kobe) are displayed.

B.1 Matches visualization

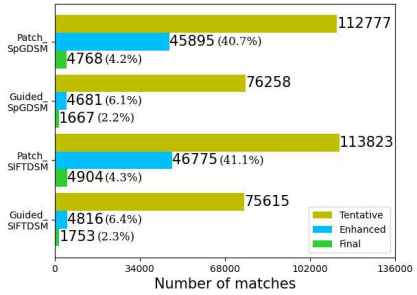
For both datasets, there exist 2 epochs, leading to 1 set of epoch combination for precise matching. Four variants (i.e., ① $Patch_{SpGDSM}$, ② $Guided_{SpGDSM}$, ③ $Patch_{SIFTDSM}$ and ④ $Guided_{SIFTDSM}$) are tested on both datasets, the resulted matches are visualized in Figure B.1 and B.2. As can be seen, patterns similar to Fréjus and Alberona (c.f., Section 6.3.3) are present.

B.2 DoD

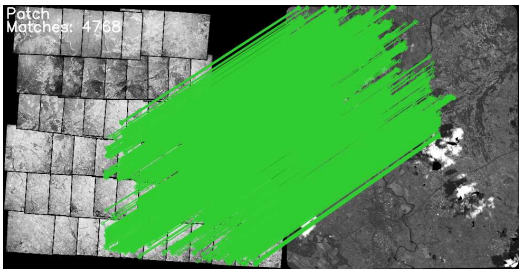
The DoDs for Pezenas and Kobe are demonstrated in Figure B.3 and B.4. The corresponding statistical information is displayed in Table B.1. As can be seen, patterns similar to Fréjus and Alberona (c.f., Section 6.3.3) are present.



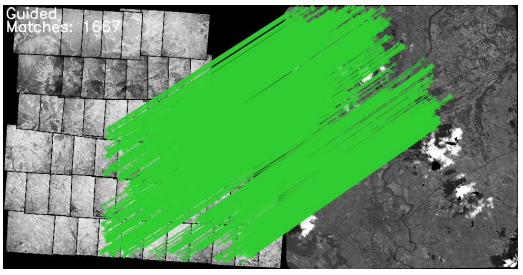
(a) Overlapping zone



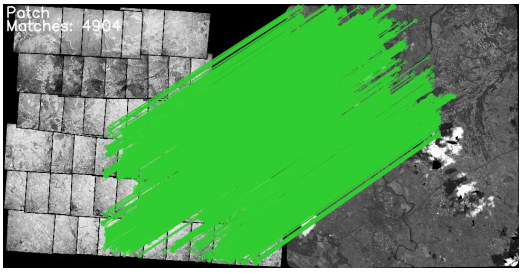
(b) Number of recovered matches



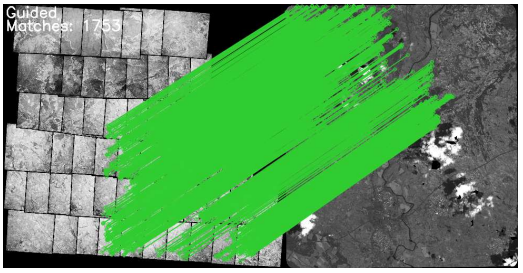
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$

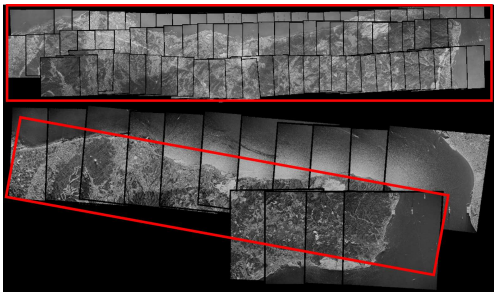


(e) $Patch_{SIFTDSM}$

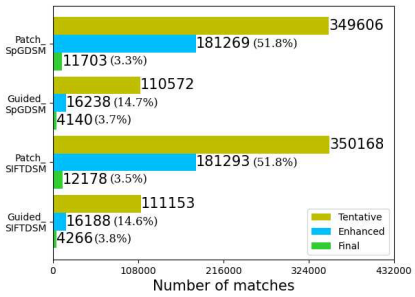


(f) $Guided_{SIFTDSM}$

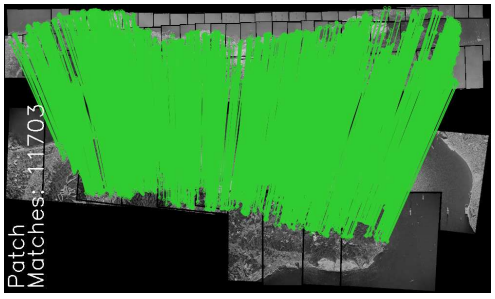
Figure B.1: Precise matching visualization of **Pezenas 1971 and 2014 (Satellite)**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



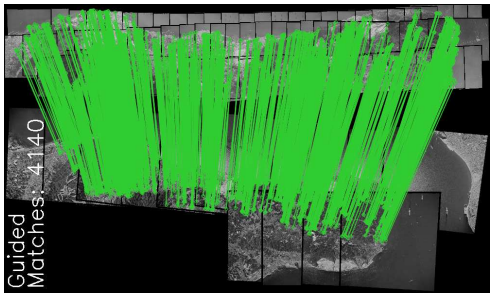
(a) Overlapping zone



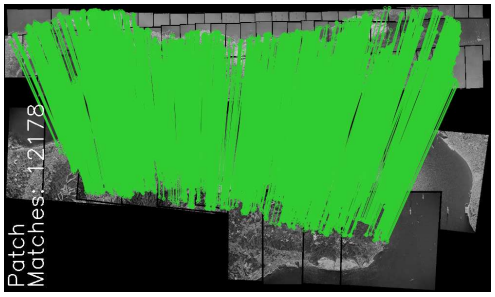
(b) Number of recovered matches



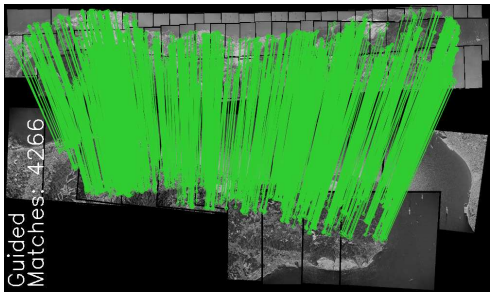
(c) $Patch_{SpGDSM}$



(d) $Guided_{SpGDSM}$

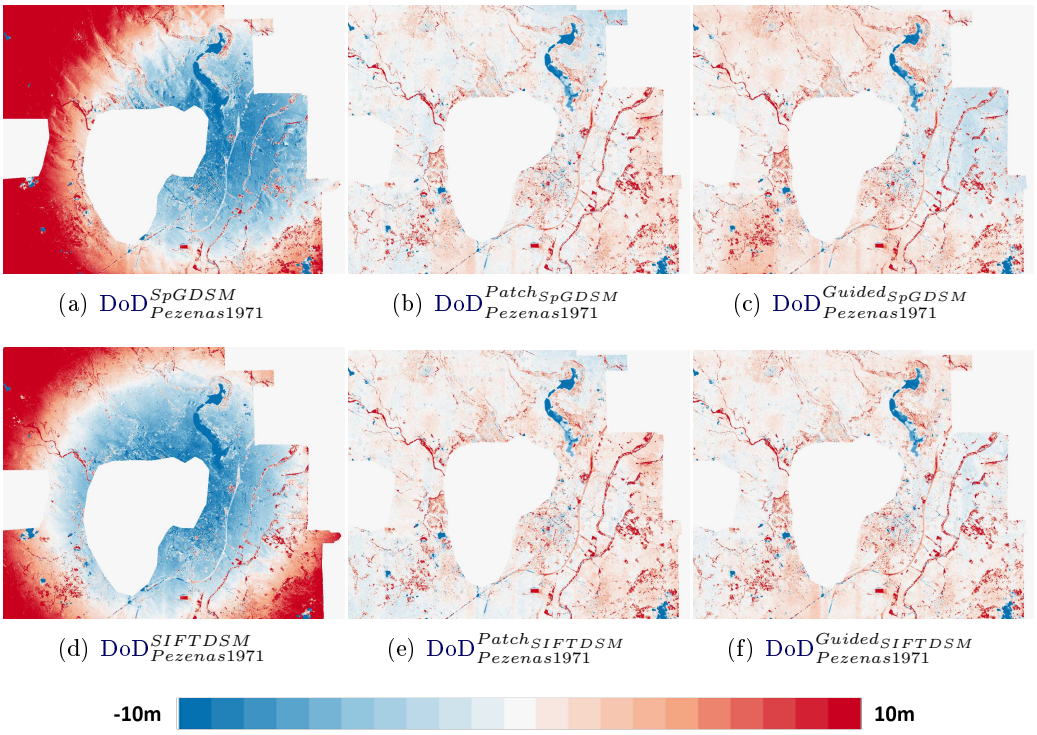


(e) $Patch_{SIFTDSM}$



(f) $Guided_{SIFTDSM}$

Figure B.2: Precise matching visualization of **Kobe 1991 and 1995**. (a) Image pairs to be matched, with red rectangles indicating the overlapping zone. (b) Numbers of tentative, enhanced and final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. (c-f) Visualization of final matches recovered with $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.



(a) $\text{DoD}^{SpGDSM}_{Pezenas1971}$ (b) $\text{DoD}^{Patch_{SpGDSM}}_{Pezenas1971}$ (c) $\text{DoD}^{Guided_{SpGDSM}}_{Pezenas1971}$

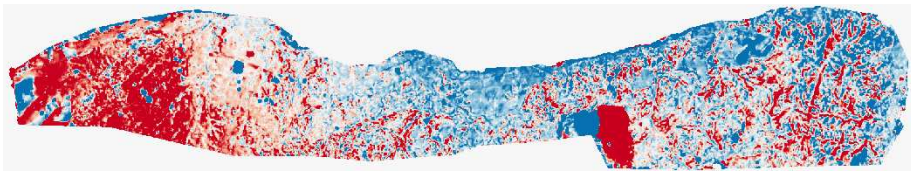
(d) $\text{DoD}^{SIFTDSM}_{Pezenas1971}$ (e) $\text{DoD}^{Patch_{SIFTDSM}}_{Pezenas1971}$ (f) $\text{DoD}^{Guided_{SIFTDSM}}_{Pezenas1971}$

(g) DoD legend

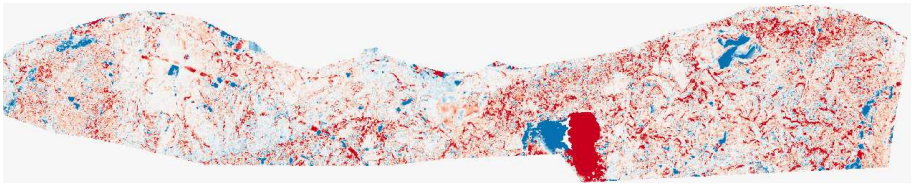
Figure B.3: DoDs between free epoch **Pezenas 1971** and reference satellite epoch **2014**. (a) and (d) are roughly co-registered DoDs resulted from variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (elaborated in Chapter 5). (b, c, e, f) are refined DoDs resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually. The holes among them are areas covered with clouds which are masked out.

		μ [m]	σ [m]	$\lvert\mu\rvert$ [m]
$DoD^{Pezenas}_{1971-2014(Satellite)}$	$SpGDSM$	-4.35	12.48	8.45
	$Patch_{SpGDSM}$	-0.46	3.73	1.72
	$Guided_{SpGDSM}$	-0.78	3.73	1.94
	$SIFTDSM$	-1.45	11.24	5.78
	$Patch_{SIFTDSM}$	-0.60	3.71	1.71
	$Guided_{SIFTDSM}$	-0.69	3.66	**1.68**
$DoD^{Kobe}_{1991-1995}$	$SpGDSM$	-0.75	14.62	7.95
	$Patch_{SpGDSM}$	1.93	10.26	3.99
	$Guided_{SpGDSM}$	2.03	11.74	4.30
	$SIFTDSM$	0.27	14.40	7.57
	$Patch_{SIFTDSM}$	1.80	10.36	4.00
	$Guided_{SIFTDSM}$	1.84	9.48	**3.87**

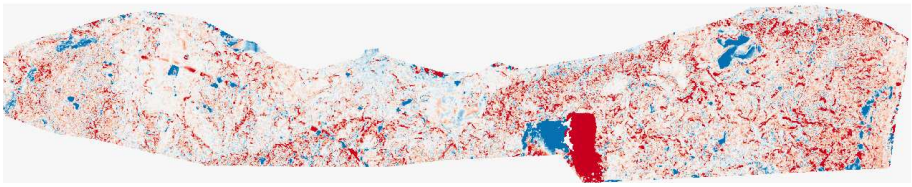
Table B.1: Average value μ, standard deviation σ, and absolute average value $|\mu|$ of all the DoDs in Figure B.3 and B.4.



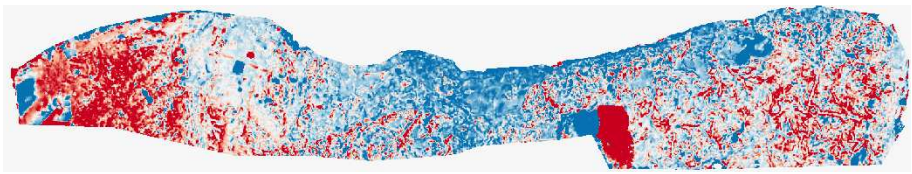
(a) $\text{DoD}_{Kobe}^{SpGDSM}$



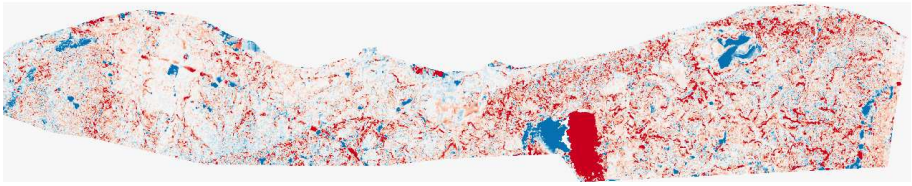
(b) $\text{DoD}_{Kobe}^{Patch_{SpGDSM}}$



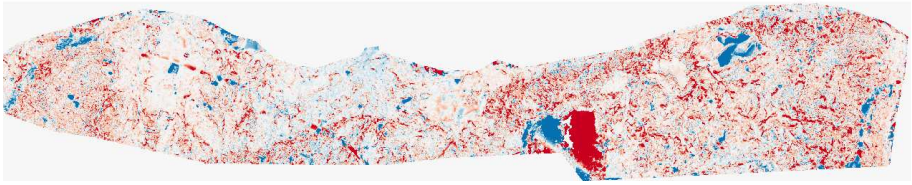
(c) $\text{DoD}_{Kobe}^{Guided_{SpGDSM}}$



(d) $\text{DoD}_{Kobe}^{SIFTDSM}$



(e) $\text{DoD}_{Kobe}^{Patch_{SIFTDSM}}$



(f) $\text{DoD}_{Kobe}^{Guided_{SIFTDSM}}$

(g) DoD legend

Figure B.4: DoDs between free epoch **Kobe 1991** and reference epoch **1995**. (a) and (d) are roughly co-registered DoDs resulted from variants $SuperGlue_{DSM}$ and $SIFT_{DSM}$ (elaborated in Chapter 5). (b, c, e, f) are refined DoDs resulted from variants $Patch_{SpGDSM}$, $Guided_{SpGDSM}$, $Patch_{SIFTDSM}$ and $Guided_{SIFTDSM}$ individually.

Appendix C

Tutorial of our pipeline

We provide two thorough tutorials [Zhang *et al.* 2021e], [Zhang *et al.* 2021d] with test datasets to familiarize users with our pipelines. The goal of the tutorials is to recover matches for multi-epoch images. The tutorial performs an intra-epoch processing, followed by an inter-epoch processing. The latter consists of 2 main steps: rough co-registration and precise matching. At the end, an evaluation part is presented to generate and display the resulted DoDs. The structure of the tutorial is as follows:

- Intra-epoch processing:

    1. **Feature matching**. Apply feature matching based on SIFT on images within the same epoch.
    2. **Relative orientation**. Compute relative orientations for each epoch.
    3. **DSM generation**. Compute DSM of each epoch based on relative orientations.

- Inter-epoch processing:

    1. **Automated pipeline**. The automated pipeline will launch the whole inter-epoch processing pipeline by calling several subcommands.
    2. **Deep-dive in submodules**. We also provide deep-dive to explain all the submodules used in the automated pipeline. It consists of: (1) rough co-registration, which roughly co-register the DSMs and image orientations from different epochs; (2) precise matching, which obtains precise matches under the guidance of rough co-registration.

- Evaluation:

    1. **Roughly co-registered DoD**.
    2. **Refined DoD based on SuperGlue**.
    3. **Refined DoD based on SIFT**.

Take one tutorial (i.e., [Zhang *et al.* 2021e]) as example, in the following we display the commands used in the tutorial. The dataset used in the tutorial consists of 2 epochs (i.e., 1971 and 1981).

C.1 Intra-epoch processing

In this section, both epochs 1971 and 1981 go through the same commands individually. For the sake of simplicity, we take only epoch 1981 as an example to demonstrate the commands.

C.1.1 Feature matching

1. Recover tie-points with command *Tapioca*:

```
mm3d Tapioca MulScale OIS-Reech_IGNF_PVA_1-0__1981.*tif 500 -1 PostFix=_1981
```

2. Remove tie-points on the fiducial marks with command *HomolFilterMasq*:

```
mm3d HomolFilterMasq OIS-Reech_IGNF_PVA_1-0__1981.*tif
 GlobalMasq=Fiducial_marks_masq-1981-3.tif PostIn=_1981 PostOut=_1981-Masq
```

3. Tie-points reduction with command *Ratafia*:

```
mm3d TestLib NO_AllOri2Im OIS-Reech_IGNF_PVA_1-0__1981.*tif SH=_1981-Masq

mm3d Ratafia OIS-Reech_IGNF_PVA_1-0__1981.*tif SH=_1981-Masq Out=_1981-Ratafia
```

C.1.2 Relative orientation

Recover relative orientation with command *Tapas*:

```
mm3d Tapas FraserBasic OIS-Reech_IGNF_PVA_1-0__1981.*tif Out=1981 SH=_1981-Masq
```

C.1.3 DSM generation

Calculate DSM with command *Malt*:

```
mm3d Malt Ortho OIS-Reech_IGNF_PVA_1-0__1981.*tif 1981 NbVI=2
MasqImGlob=Fiducial_marks_masq-1981-3.tif DirMEC=MEC-Malt_1981 EZA=1 ZoomF=2
DoOrtho=0
```

C.2 Inter-epoch processing

C.2.1 Automated pipeline with command *TiePHistoP*

1. Option 1: SuperGlue:

```
mm3d TiePHistoP Ori-1971 Ori-1981 ImgList1971all.txt ImgList1981all.txt
 MEC-Malt_1971 MEC-Malt_1981 CoRegPatchLSz=[1280,960]
 CoRegPatchRSz=[1280,960] PrecisePatchSz=[1280,960] Feature=SuperGlue
```

2. Option 2: SIFT:

```
mm3d TiePHistoP Ori-1971 Ori-1981 ImgList1971all.txt ImgList1981all.txt
 MEC-Malt_1971 MEC-Malt_1981 PrecisePatchSz=[1280,960] Feature=SIFT
 SkipCoReg=1 CoRegOri1=1971_CoReg_SuperGlue
```

C.2.2 Deep-dive in the pipeline's submodules

1. Rough co-registration

   (1) DSM Equalization for each epoch with command *TestLib DSM_ Equalization*:

```
mm3d TestLib DSM_Equalization MEC-Malt_1981 DSMFile=MMLastNuage.xml
 OutImg=DSM1981-gray.tif

mm3d TestLib DSM_Equalization MEC-Malt_1971 DSMFile=MMLastNuage.xml
 OutImg=DSM1971-gray.tif
```

   (2) DSM Wallis filter for each epoch with command *TestLib Wallis*:

```
mm3d TestLib Wallis DSM1981-gray.tif Dir=MEC-Malt_1981
 OutImg=DSM1981-gray.tif_sfs.tif

mm3d TestLib Wallis DSM1971-gray.tif Dir=MEC-Malt_1971
 OutImg=DSM1971-gray.tif_sfs.tif
```

   (3) Matching DSM based on SuperGlue with 4 rotation hypotheses.

   (3.1) Rotate the secondary DSM four times and split DSM pairs into patch pairs with command *TestLib GetPatchPair*:

```
mm3d TestLib GetPatchPair BruteForce MEC-Malt_1971/DSM1971-gray.tif_sfs.tif
 MEC-Malt_1981/DSM1981-gray.tif_sfs.tif  OutDir=./Tmp_Patches-CoReg
 Rotate=1 PatchLSz=[1280,960] PatchRSz=[1280,960]
```

   (3.2) Hypothesis 0 °:

```
mm3d TestLib SuperGlue SuperGlueInput.txt  InDir=./Tmp_Patches-CoReg/
 OutDir=./Tmp_Patches-CoReg/ SpGOutSH=-SuperGlue

mm3d TestLib MergeTiePt ./Tmp_Patches-CoReg/  HomoXml=SubPatch.xml
 MergeInSH=-SuperGlue MergeOutSH=-SubPatch PatchSz=[1280,960]

mm3d TestLib RANSAC R2D MEC-Malt_1971.tif MEC-Malt_1981.tif
 Dir=./Tmp_Patches-CoReg/ 2DRANInSH=-SubPatch
 2DRANOutSH=-SubPatch-2DRANSAC
```

(3.3) Hypothesis 90 °:

```
mm3d TestLib SuperGlue SuperGlueInput_R90.txt  InDir=./Tmp_Patches-CoReg/
 OutDir=./Tmp_Patches-CoReg/ SpGOutSH=-SuperGlue

mm3d TestLib MergeTiePt ./Tmp_Patches-CoReg/  HomoXml=SubPatch_R90.xml
MergeInSH=-SuperGlue MergeOutSH=-SubPatch_R90
 PatchSz=[1280,960]

mm3d TestLib RANSAC R2D MEC-Malt_1971.tif MEC-Malt_1981.tif
 Dir=./Tmp_Patches-CoReg/ 2DRANInSH=-SubPatch_R90
 2DRANOutSH=-SubPatch_R90-2DRANSAC
```

(3.4) Hypothesis 180 °:

```
mm3d TestLib SuperGlue SuperGlueInput_R180.txt  InDir=./Tmp_Patches-CoReg/
 OutDir=./Tmp_Patches-CoReg/ SpGOutSH=-SuperGlue

mm3d TestLib MergeTiePt ./Tmp_Patches-CoReg/  HomoXml=SubPatch_R180.xml
 MergeInSH=-SuperGlue MergeOutSH=-SubPatch_R180 PatchSz=[1280,960]

mm3d TestLib RANSAC R2D MEC-Malt_1971.tif MEC-Malt_1981.tif
 Dir=./Tmp_Patches-CoReg/ 2DRANInSH=-SubPatch_R180
 2DRANOutSH=-SubPatch_R180-2DRANSAC
```

(3.5) Hypothesis 270 °:

```
mm3d TestLib SuperGlue SuperGlueInput_R270.txt  InDir=./Tmp_Patches-CoReg/
 OutDir=./Tmp_Patches-CoReg/ SpGOutSH=-SuperGlue

mm3d TestLib MergeTiePt ./Tmp_Patches-CoReg/  HomoXml=SubPatch_R270.xml
```

```
 MergeInSH=-SuperGlue MergeOutSH=-SubPatch_R270 PatchSz=[1280,960]

mm3d TestLib RANSAC R2D MEC-Malt_1971.tif MEC-Malt_1981.tif
 Dir=./Tmp_Patches-CoReg/ 2DRANInSH=-SubPatch_R270
 2DRANOutSH=-SubPatch_R270-2DRANSAC
```

(4) Create GCPs with command *TestLib CreateGCPs*:

```
mm3d TestLib CreateGCPs ./Tmp_Patches-CoReg MEC-Malt_1971.tif
 MEC-Malt_1981.tif ./ ImgList1971all.txt ImgList1981all.txt
 Ori-1971 Ori-1981 MEC-Malt_1971 MEC-Malt_1981
 CreateGCPsInSH=-SubPatch_R180-2DRANSAC Out2DXml1=OutGCP2D_epoch1971.xml
 Out3DXml1=OutGCP3D_epoch1971.xml Out2DXml2=OutGCP2D_epoch1981.xml
 Out3DXml2=OutGCP3D_epoch1981.xml
```

(5) 3D Helmert transformation with command *GCPBascule*:

```
mm3d GCPBascule "OIS-Reech_IGNF_PVA_1-0__1971.*tif" 1971 1981
 OutGCP3D_epoch1981.xml OutGCP2D_epoch1971.xml
```

2. Precise matching

(1) Get overlapped images with command *TestLib GetOverlappedImages*:

```
mm3d TestLib GetOverlappedImages 1971 1981 ImgList1971all.txt
 ImgList1981all.txt Para3DH=Basc-1971-2-1981.xml
```

(2) Get Patch Pair with command *TestLib GetPatchPair Guided*:

```
mm3d TestLib GetPatchPair Guided
 OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974.tif
 OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064.tif
 Ori-1971 Ori-1981 OutDir=./Tmp_Patches-Precise
 SubPXml=OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974_
OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064_SubPatch.xml
 ImgPair=OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974_
OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064
_SuperGlueInput.txt
 PatchSz=[1280,960] Para3DH=Basc-1971-2-1981.xml DSMDirL=MEC-Malt_1971
```

(3) Get tentative tie-points (option1: SuperGlue) with command *TestLib SuperGlue* and *TestLib MergeTiePt*:

```
mm3d TestLib SuperGlue
 OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974_
OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064
_SuperGlueInput.txt
 InDir=./Tmp_Patches-Precise/ OutDir=./Tmp_Patches-Precise/
 SpGOutSH=-SuperGlue CheckNb=100

mm3d TestLib MergeTiePt ./Tmp_Patches-Precise/
 HomoXml=OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974_
OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064_SubPatch.xml
 MergeInSH=-SuperGlue MergeOutSH=-SuperGlue  OutDir=./ PatchSz=[1280,960]
 BufferSz=[128,96]
```

(4) Get tentative tie-points (option1: SIFT) with command *TestLib Guided-SIFTMatch*:

```
mm3d TestLib GuidedSIFTMatch
 OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974.tif
 OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064.tif
 Ori-1971 Ori-1981  SkipSIFT=false DSMDirL=MEC-Malt_1971 DSMDirR=MEC-Malt_1981
 Para3DH=Basc-1971-2-1981.xml
```

(5) 3D-RANSAC with command *TestLib RANSAC R3D*:

```
mm3d TestLib RANSAC R3D
 OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974.tif
 OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064.tif
 Ori-1971 Ori-1981 Dir=./  DSMDirL=MEC-Malt_1971 DSMDirR=MEC-Malt_1981
 DSMFileL=MMLastNuage.xml DSMFileR=MMLastNuage.xml 3DRANInSH=-SuperGlue
 3DRANOutSH=-SuperGlue-3DRANSAC
```

(6) Cross correlation with command *TestLib CrossCorrelation*:

```
mm3d TestLib CrossCorrelation
 OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974.tif
 OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064.tif
 CCInSH=-SuperGlue-3DRANSAC CCOutSH=-SuperGlue-3DRANSAC-CrossCorrelation
 SzW=32 CCTh=0.6 PatchSz=[1280,960] BufferSz=[30,60]
 PatchDir=./Tmp_Patches-Precise
 SubPXml=OIS-Reech_IGNF_PVA_1-0__1971-06-21__C2844-0141_1971_FR2117_0974_
OIS-Reech_IGNF_PVA_1-0__1981-06-16__C2544-0021_1981_F2544-2644_0064_SubPatch.xml
```

C.3 Evaluation

C.3.1 Roughly co-registered DoD

1. Get DSM of epoch 1971:

```
mm3d Malt Ortho OIS-Reech_IGNF_PVA_1-0__1971.*tif 1981 NbVI=2
 DirMEC=MEC-Malt_1971_CoReg EZA=1 MasqImGlob=Fiducial_marks_masq-1971-3.tif
 ZoomF=4 DoOrtho=0
```

2. Calculate DoD with command *CmpIm*:

```
mm3d CmpIm MEC-Malt_1971_CoReg/Z_Num7_DeZoom4_STD-MALT.tif
 MEC-Malt_1981/Z_Num8_DeZoom2_STD-MALT.tif UseFOM=1 FileDiff=DoD-CoReg.tif
 16Bit=1
```

C.3.2 Refined DoD based on SuperGlue

1. Set weight of inter-epoch tie-points with command *TestLib TiePtAddWeight* :

```
mm3d TestLib TiePtAddWeight 10 InSH=-SuperGlue-3DRANSAC-CrossCorrelation
```

2. Txt to binary conversion with command *HomolFilterMasq*:

```
mm3d HomolFilterMasq "0.*tif" PostIn=-SuperGlue-3DRANSAC-CrossCorrelation-W10
 PostOut=-SuperGlue-3DRANSAC-CrossCorrelation-W10-dat ANM=1 ExpTxt=1
 ExpTxtOut=0
```

3. Merge intra- and inter-epoch tie-points with command *MergeHomol*:

```
mm3d MergeHomol "Homol_1971-Ratafia|Homol_1981-Ratafia
|Homol-SuperGlue-3DRANSAC-CrossCorrelation-W10-dat"
 Homol_Merged-SuperGlue
```

4. Run bundle adjustment with command *Campari*:

```
mm3d Campari "0.*tif" 1981 Campari_Refined-SuperGlue SH=_Merged-SuperGlue
AllFree=1 NbIterEnd=20 SigmaTieP=0.25
```

5. Get DSM of epoch 1981:

```
mm3d Malt Ortho OIS-Reech_IGNF_PVA_1-0__1981.*tif Campari_Refined-SuperGlue
 NbVI=2 DirMEC=MEC-Malt_1981_Refined-SuperGlue EZA=1
 MasqImGlob=Fiducial_marks_masq-1981-3.tif ZoomF=2 DoOrtho=0
```

6. Get DSM of epoch 1971:

```
mm3d Malt Ortho OIS-Reech_IGNF_PVA_1-0__1971.*tif Campari_Refined-SuperGlue
 NbVI=2 DirMEC=MEC-Malt_1971_Refined-SuperGlue EZA=1
 MasqImGlob=Fiducial_marks_masq-1971-3.tif ZoomF=4 DoOrtho=0
```

7. Calculate DoD:

```
mm3d CmpIm MEC-Malt_1971_Refined-SuperGlue/Z_Num7_DeZoom4_STD-MALT.tif
 MEC-Malt_1981_Refined-SuperGlue/Z_Num8_DeZoom2_STD-MALT.tif UseFOM=1
 FileDiff=DoD-Refined-SuperGlue.tif 16Bit=1
```

C.3.3 Refined DoD based on SIFT

1. Set weight of inter-epoch tie-points:

```
mm3d TestLib TiePtAddWeight 10 InSH=-GuidedSIFT-3DRANSAC-CrossCorrelation
```

2. Txt to binary conversion:

```
mm3d HomolFilterMasq "O.*tif" PostIn=-GuidedSIFT-3DRANSAC-CrossCorrelation-W10
 PostOut=-GuidedSIFT-3DRANSAC-CrossCorrelation-W10-dat ANM=1 ExpTxt=1
 ExpTxtOut=0
```

3. Merge intra- and inter-epoch tie-points:

```
mm3d MergeHomol "Homol_1971-Ratafia|Homol_1981-Ratafia
|Homol-GuidedSIFT-3DRANSAC-CrossCorrelation-W10-dat"
 Homol_Merged-GuidedSIFT
```

4. Run bundle adjustment:

```
mm3d Campari "O.*tif" 1981 Campari_Refined-GuidedSIFT SH=_Merged-GuidedSIFT
 AllFree=1 NbIterEnd=20 SigmaTieP=0.25
```

5. Get DSM of epoch 1981:

```
mm3d Malt Ortho OIS-Reech_IGNF_PVA_1-0__1981.*tif Campari_Refined-GuidedSIFT
 NbVI=2 DirMEC=MEC-Malt_1981_Refined-GuidedSIFT EZA=1
 MasqImGlob=Fiducial_marks_masq-1981-3.tif ZoomF=2 DoOrtho=0
```

6. Get DSM of epoch 1971:

```
mm3d Malt Ortho OIS-Reech_IGNF_PVA_1-0__1971.*tif Campari_Refined-GuidedSIFT
 NbVI=2 DirMEC=MEC-Malt_1971_Refined-GuidedSIFT
 MasqImGlob=Fiducial_marks_masq-1971-3.tif EZA=1 ZoomF=4 DoOrtho=0
```

7. Calculate DoD:

```
mm3d CmpIm MEC-Malt_1971_Refined-GuidedSIFT/Z_Num7_DeZoom4_STD-MALT.tif
 MEC-Malt_1981_Refined-GuidedSIFT/Z_Num8_DeZoom2_STD-MALT.tif UseFOM=1
 FileDiff=DoD-Refined-GuidedSIFT.tif 16Bit=1
```

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 20.)

[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 20.)

[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 20.)

[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 22.)

[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 22.)

[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 20.)

[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 23.)

[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 23.)

[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 23.)

[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 23.)

[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 22.)

[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 pages 2 and 12.)

[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 21.)

[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 21.)

[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 23.)

[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 23.)

[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 20 and 21.)

[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 20.)

[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 2 and 12.)

[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 23.)

[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 23.)

[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 21.)

[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 pages 2 and 12.)

[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 pages 2 and 12.)

[Fraser 1997] Clive S Fraser. *Digital camera self-calibration.* ISPRS Journal of Photogrammetry and Remote sensing, vol. 52, no. 4, pages 149–159, 1997. (Cited on page 69.)

[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 pages 1 and 11.)

[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 pages 22 and 23.)

[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 19.)

[IGN 2019] IGN. *remonterletemps*. https://remonterletemps.ign.fr/, 2019. (Cited on pages 1 and 11.)

[James & Robson 2014] Mike R James and Stuart Robson. *Mitigating systematic error in topographic models derived from UAV and ground-based image networks*. Earth Surface Processes and Landforms, vol. 39, no. 10, pages 1413–1420, 2014. (Cited on page 22.)

[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 21 and 22.)

[Leroy & Rousseeuw 1987] Annick M Leroy and Peter J Rousseeuw. *Robust regression and outlier detection*. Wiley, 1987. (Cited on page 21.)

[Lindenberger *et al.* 2021] Philipp Lindenberger, Paul-Edouard Sarlin, Viktor Larsson and Marc Pollefeys. *Pixel-Perfect Structure-from-Motion with Featuremetric Refinement*. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 5987–5997, 2021. (Cited on page 22.)

[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 3, 12, 20, 21, 37 and 68.)

[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 20.)

[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 20.)

[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 23.)

[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 23.)

[McGlone 2013] Chris McGlone. Manual of photogrammetry. American Society for Photogrammetry and Remote Sensing (ASPRS), 2013. (Cited on page 22.)

[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 2, 12 and 23.)

[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 20.)

[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 20 and 21.)

[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 23.)

[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 20.)

[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 22.)

[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 19.)

[Morel & Yu 2009] Jean-Michel Morel and Guoshen Yu. *ASIFT: A new framework for fully affine invariant image comparison.* SIAM journal on imaging sciences, vol. 2, no. 2, pages 438–469, 2009. (Cited on page 38.)

[Moulon *et al.* 2016] Pierre Moulon, Pascal Monasse, Renaud Marlet and Others. *OpenMVG.* `https://github.com/openMVG/openMVG`, 2016. (Cited on page 37.)

[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 20 and 21.)

[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 pages 2 and 12.)

[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 20 and 21.)

[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 23.)

[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 23.)

[Pierrot-Deseilligny & Cléry 2012] Marc Pierrot-Deseilligny and Isabelle Cléry. *Apero, an open source bundle adjusment software for automatic calibration and orientation of set of images.* ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVIII-5/W16, pages 269–276, 2012. (Cited on pages 22 and 37.)

[Pierrot-Deseilligny & Paparoditis 2006] Marc Pierrot-Deseilligny and Nicolas Paparoditis. *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, vol. 36, no. 1/W41, pages 1–5, 2006. (Cited on page 37.)

[Pierrot-Deseilligny *et al.* 2015] Marc Pierrot-Deseilligny, Ewelina Rupnik, Luc Girod, Jérémie Belvaux, Gregoire Maillet, Matthieu Deveau and Gérald Choqueux. *MicMac, Apero, Pastis and Other Beverages in a Nutshell.* MicMac documentation, vol. 4, 2015. (Cited on page 71.)

[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 23.)

[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 21.)

[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 20.)

[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 20.)

[Rosu *et al.* 2015] AM Rosu, M. Pierrot-Deseilligny, A. Delorme, R. Binet and Y. Klinger. *Measurement of ground displacement from optical satellite image correlation using the free open-source software MicMac.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 100, pages 48–59, 2015. (Cited on page 73.)

[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 pages 3, 12 and 21.)

[Sarlin *et al.* 2021] Paul-Edouard Sarlin, Ajaykumar Unagar, Mans Larsson, Hugo Germain, Carl Toft, Viktor Larsson, Marc Pollefeys, Vincent Lepetit, Lars Hammarstrand, Fredrik Kahl*et al.* *Back to the feature: Learning robust camera localization from pixels to pose.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3247–3257, 2021. (Cited on page 22.)

[Schonberger & Frahm 2016] Johannes L Schonberger and Jan-Michael Frahm. *Structure-from-motion revisited.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4104–4113, 2016. (Cited on pages 22 and 37.)

[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 21.)

[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 20.)

[Snavely *et al.* 2006] Noah Snavely, Steven M Seitz and Richard Szeliski. *Photo tourism: exploring photo collections in 3D.* In ACM Siggraph 2006 Papers, pages 835–846. 2006. (Cited on page 22.)

[Sonka *et al.* 2014] Milan Sonka, Vaclav Hlavac and Roger Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014. (Cited on page 21.)

[Souchon *et al.* 2010] Jean-Philippe Souchon, Christian Thom, Christophe Meynard, Olivier Martin and Marc Pierrot-Deseilligny. *The IGN CAMv2 System.* The Photogrammetric Record, vol. 25, no. 132, pages 402–421, 2010. (Cited on page 25.)

[Sweeney 2015] Chris Sweeney. *Theia Multiview Geometry Library: Tutorial & Reference.* `http://theia-sfm.org`, 2015. (Cited on page 37.)

[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 20 and 21.)

[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 20.)

[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 21.)

[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 page 21.)

[Truong Giang *et al.* 2018] Nguyen Truong Giang, Jean-Michaël Muller, Ewelina Rupnik, Christian Thom and Marc Pierrot-Deseilligny. *Second iteration of photogrammetric processing to refine image orientation with improved tie-points.* Sensors, vol. 18, no. 7, page 2150, 2018. (Cited on page 22.)

[USGS 2019] USGS. *earthexplorer.* https://earthexplorer.usgs.gov/, 2019. (Cited on pages 1 and 11.)

[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 20.)

[Wackrow *et al.* 2008] R Wackrow, JH Chandler and T Gardner. *Minimising systematic errors in DEMs caused by an inaccurate lens model.* International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, no. B5, pages 1–6, 2008. (Cited on page 22.)

[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 pages 2 and 12.)

[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 21.)

[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 20 and 21.)

[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 39.)

[Zhang *et al.* 2021a] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Code in micmac for matching multi-epoch images.* https://github.com/micmacIGN/micmac/tree/master/src/uti_phgrm/TiePHistorical, 2021. (Cited on pages 8 and 18.)

[Zhang *et al.* 2021b] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Project website for matching multi-epoch images.* https://www.umr-lastig.fr/ewelina-rupnik/historicalfeatures.html, 2021. (Cited on pages 9 and 18.)

[Zhang *et al.* 2021c] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Slides for matching multi-epoch images.* https://drive.google.com/uc?id=16sJZiY3sTZSRZKBzpy2IHVloCLj6pEwP, 2021. (Cited on pages 9 and 18.)

[Zhang *et al.* 2021d] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Tutorial of matching multi-epoch aerial and satellite images.* https://colab.research.google.com/drive/14okQ8bBhEZmy6EGRIQvazTqrN39oc_K5, 2021. (Cited on pages 8, 18 and 109.)

[Zhang *et al.* 2021e] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Tutorial of matching multi-epoch aerial images.* https://colab.research.google.com/drive/1poEXIeKbPcJT_2hyQOBhzcj1EEhO8OgD, 2021. (Cited on pages 8, 18 and 109.)

[Zhang *et al.* 2021f] Lulin Zhang, Ewelina Rupnik and Marc Pierrot-Deseilligny. *Video for matching multi-epoch images.* https://youtu.be/YnF-FOUJaSM, 2021. (Cited on pages 9 and 18.)