Evaluating Learning-based Tie Point Matching for Geometric Processing of Off-Track Satellite Stereo

Shuang Song[1,2,4], Luca Morelli[5,6], Xinyi Wu[1,3], Rongjun Qin[1,2,3,4], *, Hessah Albanwan[7], Fabio Remondino[5]

[1] Geospatial Data Analytics Lab, The Ohio State University, Columbus, USA – Email: <song.1634><wu.4988><qin.324>@osu.edu
[2] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, USA
[3] Department of Electrical and Computer Engineering, The Ohio State University, Columbus, USA
[4] Translational Data Analytics Institute, The Ohio State University, Columbus, USA
[5] 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy – Email: <lmorelli><remondino>@fbk.eu
[6] Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy
[7] Civil Engineering Department, Kuwait University, Kuwait – Email: hessah.albanwan@ku.edu.kw

Commission II, WG II/1

Keywords: Multi-temporal analysis, Satellite relative orientation, SIFT, Superpoint, LightGlue, Image matching.

Abstract:
Tie-point matching of off-track stereo images is a very challenging task, which can impact bias compensation and digital surface model (DSM) generation. Compared to in-track stereo images, off-track stereo images are more complex primarily due to the radiometric differences caused by sun illumination, sensor responses, atmospheric conditions, and seasonal land cover variations, and secondly due to the longer baseline and larger intersection angle. These challenges significantly limit the use of the vast number of images in satellite archives for automated geometric processing and mapping. Recent advances in deep learning (DL) based matching show promising results against images with diverse illuminations, viewing angles and scales through learning examples. This paper evaluates the potentials of addressing the tie point matching problems in off-track satellite stereo images. Specifically, we focus on stereo pairs that failed or underperformed in classic matching algorithms (i.e., SIFT (scale invariant feature transform)), and evaluate the DL-based tie points matchers by its resulting geometric accuracy in relative orientation, and the generated DSM. The experiments are carried out using 40 off-track satellite stereo pairs from four different regions around the world. We conclude that DL-based methods provide a significant higher success rate in matching challenging multi-temporal stereo pairs, even if their matching accuracy is slightly lower than classic algorithms.

1. INTRODUCTION

Satellite stereo images offer unique advantages in 3D mapping, change detection, and building modelling due to their global coverage, low-cost per unit-area and periodic revisits (Bosch et al., 2019; Gui and Qin, 2021; Huang et al., 2022). Current commercial satellites offer images with ground sampling distance (GSD) up to 0.3 meters, potentially producing 1:10,000 topographic maps globally (Poli et al., 2015). Typical aerial sensors collect in-track stereo images, which are same-time images consistent in their lighting conditions, thus have less relative orientation problems. On the contrary, most satellite images are collected under multifaceted conditions and timings, under only partly controllable conditions. The satellite images often experience different sun illuminations, sensor responses, atmospheric conditions, anisotropic surfaces and seasonal landcover variations, as well as a larger baseline and intersection angle (Albanwan and Qin, 2022; Qin, 2019, 2016). Therefore, satellite stereo pairs from different times/tracks, namely off-track stereo images, face elevated challenges when using classic (hand-crafted) algorithms for tie points and dense image matching (Qin, 2019). As a result, the current practice still largely relies on collections that are designated for in-track stereo images, i.e., satellite images taken on the same track and minutes apart, leaving the vast number of satellite images significantly underutilized.

Deep learning (DL) based approaches are showing consistently progresses in image-matching problems and benchmarks (Jin et al., 2021; Remondino et al., 2022). Owning to its ability of learning complex features by samples, it was shown to be effective in addressing correspondence problem between images with drastic differences in scale, illumination, and colorimetry (Morelli et al., 2022; Morelli et al., 2024). However, their ability in addressing the compounded challenges in satellite off-track stereo pairs remain untested.

Therefore, in this paper, we aim to assess the advantages of these emerging methodologies in finding correspondences in multi-temporal satellite imagery. We collected full-frame satellite off-track stereo pairs from the IARPA CORE3D Public Dataset (Brown et al., 2018) where many traditional hand-crafted techniques, such as SIFT (Lowe, 2004), proved to be inadequate. The reported investigation employs the deep-image-matching library[1] (DIM) presented in Morelli et al. (2024), which is a Python-based tool that extends the capabilities of state-of-the-art local features and matchers to accommodate large-format images and datasets with rotations. We evaluate these matchers by the resulting geometric accuracy of the relative orientation, as well as the accuracy of the subsequently generated digital surface model (DSM) against LiDAR reference.

2. RELATED WORKS

Early works noted the unique challenges of off-track satellite stereo images, while most of them focus on evaluating different dense matching algorithms (Albanwan and Qin, 2022) or analyzing stereo configurations under varying acquisition conditions (d'Angelo et al., 2014; Facciolo et al., 2017; Qin, 2019). For example, Albanwan and Qin (2022) found that end-

---

* Corresponding author 2036 Neil Avenue, Columbus, Ohio, USA. qin.324@osu.edu
[1] https://github.com/3DOM-FBK/deep-image-matching

to-end DL based dense matchers can better process off-track stereo images, albeit it may suffer from generalization issues for unseen datasets (i.e., different sensors and resolutions). However, these studies neglected the fact that tie point matcher should be studied in the first place to ensure accurate geo-referencing / bundle adjustment, which were known to be extremely challenging for classic tie point extractor and matchers. In recent years, new approaches based on convolutional neural networks (CNNs) have been proposed to overcome the limitations of traditional hand-crafted local features, such as SIFT (Lowe, 2004) and ORB (Rublee et al., 2011). Conventional methods exhibit suboptimal performance when matching images characterized by substantial variations in illumination conditions and/or viewing angles. Typically, these CNNs are trained via self-supervised techniques, utilizing multi-temporal datasets derived from diverse sensors and including a broad spectrum of objects and environments (DeTone et al., 2018). Detection and description have been trained separately, e.g. Key.Net (Barroso-Laguna et al., 2019) and HardNet (Mishchuk et al., 2017), or jointly, as in SuperPoint (DeTone et al., 2018). Concurrently, there is a growing trend towards employing learned matchers, such as SuperGlue (Sarlin et al., 2020) and LightGlue (Lindenberger et al., 2023).

In the context of classical photogrammetric datasets, characterized by single-sensor acquisitions within a limited timeframe and substantial image overlap, the adoption of learned approaches offers minimal advantages and, in certain instances, may even result in reduced accuracy, as highlighted by Remondino et al. (2021). The advantage is instead evident in challenging multi-temporal datasets (Maiwald et al., 2021; Morelli et al., 2022) or under different viewing angles (Ioli et al., 2023). It is noteworthy that these approaches have inherent constraints, including the ability to execute predictions solely on images of limited dimensions determined by GPU capabilities, as well as limitations in rotation and scale invariance, as observed in Marelli et al. (2023).

3. METHODOLOGY

The processing workflow for finding image correspondences and evaluating the results is hereafter reported.

3.1 Processing and Evaluation Framework

The processing and evaluation framework, shown in Figure 1, aims to assess the performance of classic hand-crafted and DL-based tie point matching methods. Firstly, the satellite image stereo pairs are selected with proper convergence angle and a challenging appearance difference, from which tie points have been identified with both classic (i.e., SIFT) and DL-based local features and matches (Section 3.3). Considering that the localization accuracy of different methods varies, we refine these identified matches using Least Squares Matching (LSM) (Bellavia et al., 2024; Bethmann and Luhmann, 2010; Gruen, 1985). Using these tie points, RPC-based (Rational Polynomial Coefficients) relative orientation/bias compensation is performed using the RSP (RPC stereo processor) software (Qin, 2016), incorporating RANSAC, adjusting the RPC coefficients for the image pairs, using the matched points. The program's precision is then evaluated by the number of correctly matched points (inliers) and the epipolar error (y-parallax in the epipolar space) (see Section 3.4).

In addition, we also assess the accuracy of the resulting accuracy of the generated DSM. After completing the relative orientation step, dense stereo matching was performed to create DSM using the RSP software (Qin, 2016), which implements a typical Semi-Global Matching algorithm (Hirschmuller, 2008). We then compared this reconstructed DSM with a 3D ground truth DSM, created from USGS (United States Geological Survey) 3DEP airborne LiDAR products [2], from which we derive both the completeness and accuracy of the derived DSM (see Section 3.4).
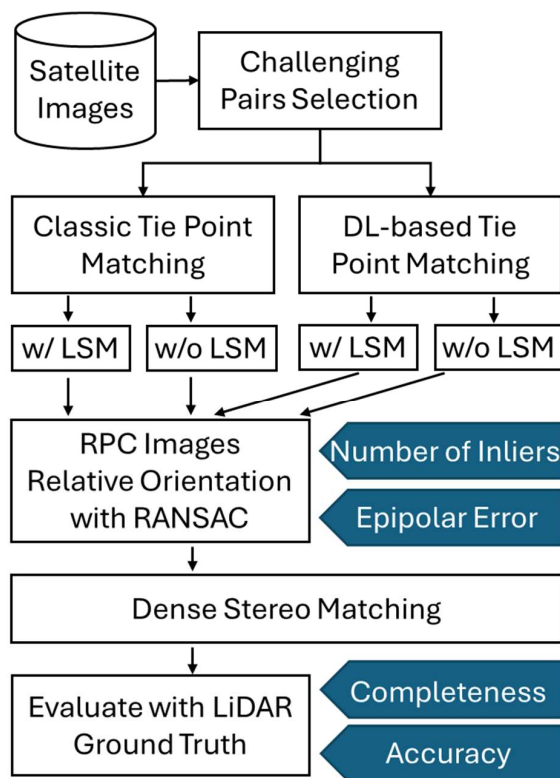

Figure 1. The evaluation workflow (LSM: Least-squares Matching, DL: Deep-Learning. White boxes denote data processing modules and dark blue denotes evaluation metrics).

3.2 Satellite Off-track Stereo Pairs – Data preparation

Classic tie point matching with hand-crafted approaches, such as SIFT, has been widely used in aerial datasets because of their robustness and efficiency (Ling et al., 2021). However, as mentioned earlier, it falls short in cases where drastic illumination, scale and/or view differences are observed. Our evaluation focuses on these challenging cases where images show significant appearance differences. To derive 3D geometry, we select stereo pairs with specific intersection angle in the range of 5° to 35° (Albanwan and Qin, 2022; Qin, 2019, 2016). In the meantime, these selected stereo pairs a ranked based on their seasonal and sun illumination differences, i.e., sun angle difference and month-of-year difference using attributes from metadata, respectively. An example where illumination change lead to a huge difference in appearance is shown in Figure 2 seasonal differences are shown in Figure 3. The month-of-year difference is computed with Equation 1, where $month_i$ refers to the month-of-year of two paired images.

$$\min(|month_1 - month_2|, \ 12 - |month_1 - month_2|) \quad (1)$$

Based on these criteria, we reduce the number of candidate pairs by selecting the top K. Finally, we perform matching on all the remaining candidates using SIFT as the representative for hand-crafted methods. Based on the number of matches, we finally selected the top 10 most challenging pairs for each test site.
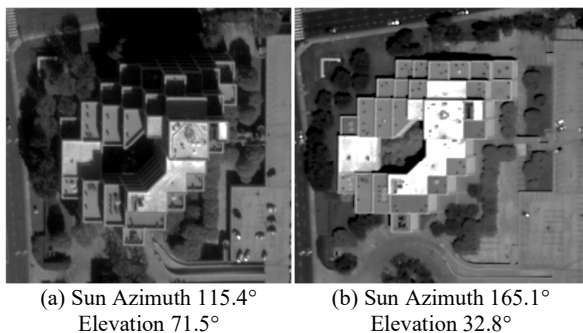

(a) Sun Azimuth 115.4°
Elevation 71.5°

(b) Sun Azimuth 165.1°
Elevation 32.8°

Figure 2. An example of illumination difference (UCSD, CA).

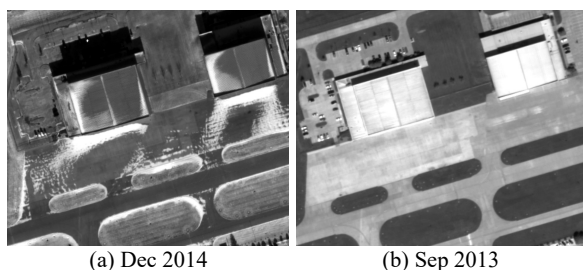(a) Dec 2014   (b) Sep 2013

Figure 3. An example of seasonal difference (OMA, NE).

3.3 Pair Matching with Hand-crafted and Deep Learning-based Local Features and Matchers

SIFT and SuperPoint are selected as the baseline methods for hand-crafted and learned local features, respectively. For matching SIFT features, the classic nearest neighbor approach is used with a ratio threshold equal to 0.95 instead of 0.80-0.85. Preliminary tests have shown that on these datasets affected by extreme seasonal and illumination changes, a too low ratio threshold is too restrictive in discarding ambiguous matches. With a higher threshold, more matches are retained, leaving the elimination of possible outliers to the test with epipolar geometry. SuperPoint is usually matched with SuperGlue or LightGlue. LightGlue has been chosen, since it is an optimized version of SuperGlue with a more permissive license. These algorithms are available in the DIM (Deep-Image-Matching) library (Morelli et al., 2024, 2022), which has been designed to process large-format images as the chosen 40 satellite image pairs detailed in Section 4.1.

The available 16-bit satellite images were normalized within the intensity range of 0-255, followed by a conversion from 16 to 8 bits to align with the pre-trained SuperPoint model which was trained on 8bit grayscale images (DeTone et al., 2018). This pre-processing was used for both learned and hand-crafted methods for a fair comparison. Other possible normalizations will be considered in future studies. Considering the large coverage of satellite images, it is necessary to process the data with both traditional and deep learning methodologies through tiling the images. Therefore, to match at satellite images at the full resolution, corresponding tiles are defined by the GRID approach implemented in DIM which matches only corresponding tiles.

This approach is possible since the two satellite images essentially cover the same ground area.

The number of local features and the precision of key points extracted influences the final accuracy (Jin et al., 2021), therefore the same number of features has been used for SIFT and SuperPoint. A maximum of 1000 feature points per tile were extracted, with each tile having a size of 2400 x 2000 pixels to allow our GPU to be able to run SuperPoint without sacrifices image resolution. These values were chosen to have good coverage of the images but not an excessive number of features on full-resolution images. From some preliminary experiments, it was seen that by varying the number of features the overall outcome of the study is not affected.
Some preliminary experiments show that the overall outcome of the study was not affected by varying the number of features.

3.4 Evaluation Metrics

As described in Section 3.1, the evaluation metrics are twofold: (1) statistics following RPC-based relative orientation and (2) a comparison of dense reconstruction to the ground truth DSM.

Our first metric is based on the statistics of relative orientation, specifically the inlier ratio and the epipolar error of the inliers. Instead of adjusting full RPC parameters (80 coefficients in total), we employed 1st order bias correction similar to previous work (Qin, 2019). The inlier ratio indicates the number of inliers after RANSAC (Fischler and Bolles, 1981) over the initial number of tie points, and assesses the effectiveness and precision of the feature-matching process. A higher number of inliers increases our confidence in the relative orientation results, as it suggests a lower number of erroneous matches. The epipolar errors (y-parallax) of inliers are calculated, which means for each tie point, the distance in pixels between a matched point and its corresponding epipolar line. We use the root mean squared epipolar error of all valid matches as metric, with a lower error indicating a better tie point quality. This metric has been particularly useful in evaluating matching quality when the number of inliers is too low to warrant a reliable relative orientation, potentially impacting the accuracy of the subsequent dense image matching, and DSM generation.

For image pairs where both classic and DL-based methods provide enough matching points for reliable orientation, we assess the RPCs' quality by creating a DSM through dense stereo matching and comparing it to the actual ground truth DSM. In this scenario the metric is composed by the completeness and the accuracy of the resulting DSM. The completeness of the DSM is defined as the percentage of the ground truth DSM's area that the derived DSM covers. Completeness values range from 0 to 1, with values closer to 1 indicating superior dense reconstruction. The accuracy of the DSM is the RMSE (Root Mean Square Error) between the derived DSM and the ground truth DSM. First, the two DSMs are aligned by applying a least squares surface matching for accurate co-registration (Qin, 2016). Then, the RMSE of pixel-wise distances is computed, excluding pixels classified as NaN from both the generated and ground truth DSM.

4. EXPERIMENTS AND EVALUATION

4.1 Datasets

Satellite pairs have been chosen from IARPA Multi-View Stereo 3D Mapping Challenge (Bosch et al., 2019) and the CORE3D (Brown et al., 2018) benchmark. These include stereo images captured by the WorldView-3 satellite sensor, which boasts a

spatial resolution of 0.3 meters. Additionally, the airborne LiDAR data for the same areas are available either directly through the Multi-View Stereo Challenge or publicly via the USGS 3DEP program. This data has been converted into a ground-truth Digital Surface Model (DSM) for performance assessment. The IARPA challenge provides 50 overlapping images covering 100 square kilometers near San Fernando, Argentina, collected between January 2015 and January 2016. Meanwhile, the CORE3D data set includes 26 images of Jacksonville (JAX), Florida, taken between October 2014 and February 2016; 43 images of Omaha (OMA), Nebraska, from September 2013 to November 2015; and 35 images of UCSD, California, captured between October 2014 and March 2016. Each CORE3D site spans approximately 200 square kilometers. The WorldView-3 images consist of a high-resolution panchromatic (PAN) image and a lower-resolution multispectral (MUL) image. In our research, we focus exclusively on the high-resolution PAN images, which have a ground sample distance (GSD) of roughly 30cm.

The four regional datasets are denoted as ARG, JAX, OMA, and UCSD (examples are shown in Figure 4). Employing the pair selection method outlined in Section 3.2, we chose 10 challenging image pairs from each region, totaling 40 pairs for our analysis.
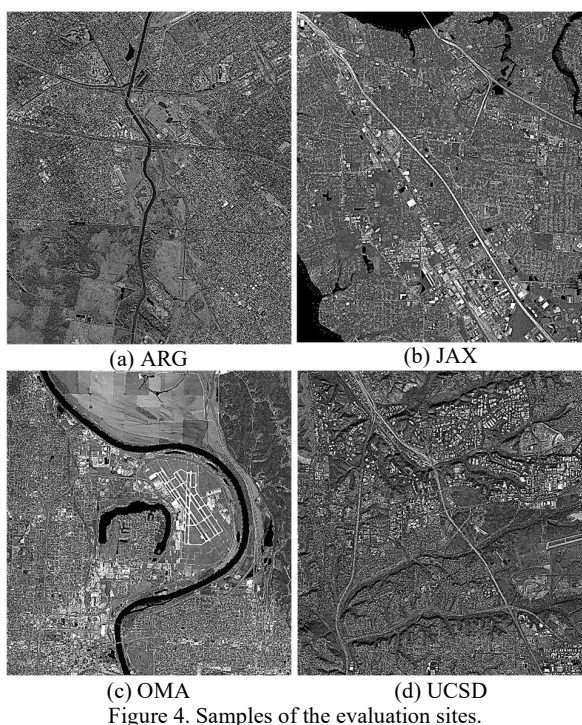

(a) ARG   (b) JAX

(c) OMA   (d) UCSD
Figure 4. Samples of the evaluation sites.

4.2 Analysis with Relative Orientation

The RPC-based relative orientation is evaluated in terms of inliers number and epipolar error. After RANSAC, if the number of inliers is less than 10, the relative orientation result is considered unreliable and therefore discarded. Based on this standard, the success rate of relative orientation is presented in Figure 5. To ensure a fair comparison, we only included successfully analyzed pairs using both SIFT and LightGlue methods.
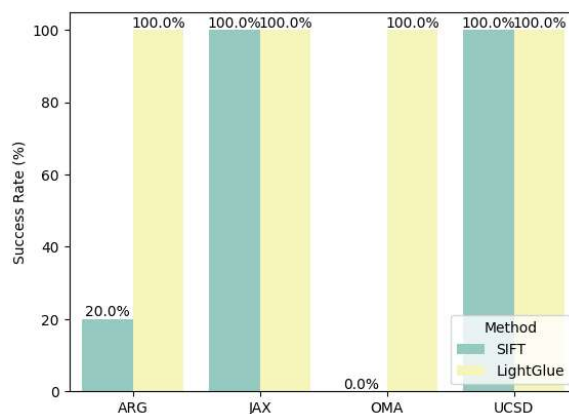

Figure 5. Success rates of relative orientation using SIFT and LightGlue.

An important finding is that SIFT matching was unsuccessful for most ARG pairs and all OMA pairs. This finding is illustrated in Figure 6, where a pair of images and their matches are reported. The failure is attributed to significant texture changes caused by seasonal differences.

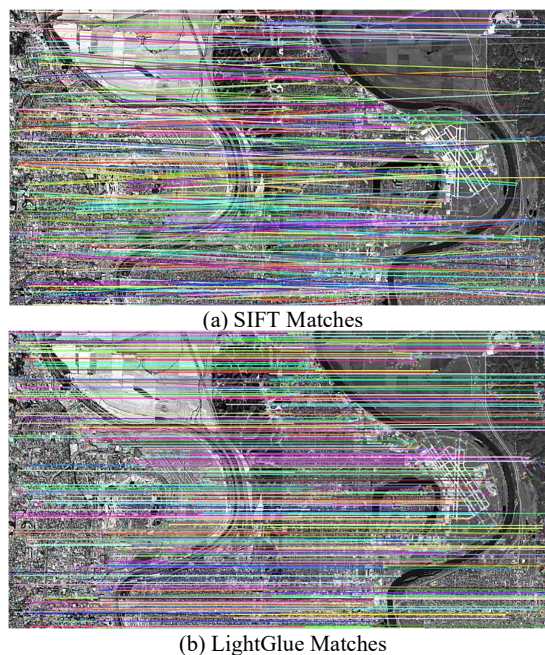(a) SIFT Matches

(b) LightGlue Matches
Figure 6. An example in OMA that SIFT failed in relative orientation due to too less inliers but LightGlue successes, highlighting the substantial difference between seasonal appearances.

Further examination of the inlier ratio statistics of tie-point matching methods, as shown in Figure 7, reveals that LightGlue consistently performs well across all regions. It shows a similar and focused distribution of inlier ratios, even achieving impressive ratios in the most challenging region (OMA), where some pairs exhibited over a 70% inlier ratio. In contrast, the classic descriptor, SIFT, exhibits variable performance; it outperforms LightGlue in regions like JAX and UCSD but is less robust against extreme appearance changes observed in ARG and OMA.
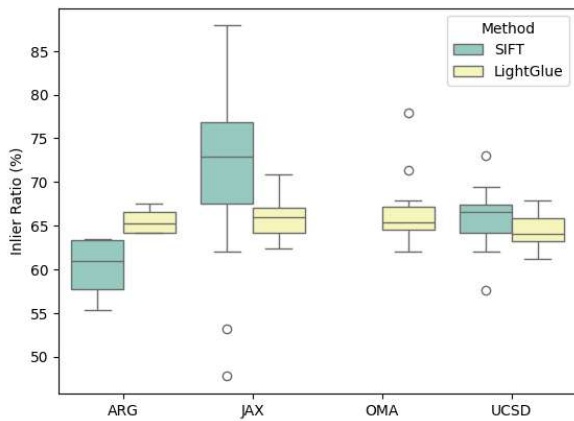
Figure 7. The box plot of inlier ratio after relative orientation with RANSAC. Higher ratios are preferred.

However, when evaluating the epipolar error, as depicted in Figure 8, SIFT inliers demonstrate a lower epipolar error than LightGlue across all regions being assessed. Epipolar errors range between 0.25 and 2.00 pixel a part for OMA, while RANSAC error threshold was set to 4 pixels for both SIFT and LightGlue. The higher epipolar error of LightGlue could be explained with the matches from SuperPoint that are extracted at pixel level, while SIFT extracts keypoints with sub-pixel accuracy. In future work a possible solution to decrease LightGlue epipolar error could be to choose a lower RANSAC error threshold (range 1-2 pixel).
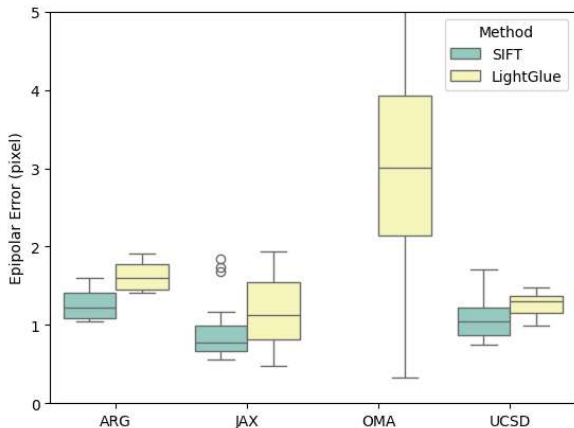

Figure 8. The box plot epipolar error after relative orientation with RANSAC. Lower errors are preferred.

4.3 Analysis with Dense Stereo Matching

Figure 9 compares DSMs produced with adjusted RPCs using SIFT and LightGlue for four image pairs. LightGlue is able to find matches in all scenarios, even in extremely challenging situations (see Figure 9 (g)). SIFT fails to find enough tie points for relative orientation in the OMA dataset but obtained a significantly better dense reconstruction for the UCSD scenario, while ARG and JAX results are comparable (Figure 9 (a-d)). Completeness and accuracy of DSM are plotted in Figure 10. In terms of completeness, LightGlue and SIFT performances are comparable for the ARG and JAX dataset, while SIFT completeness is higher and less dispersed in the UCSD dataset. In terms of accuracy, SIFT is slightly more accurate for ARG and JAX, and significantly more accurate in the UCSD dataset. Figure 11 compares SIFT and LightGlue's overall performance

across different regions using all pairs, excluding any stereo pairs where either method failed in relative orientation. The results indicate that SIFT can achieve slightly better completeness and accuracy compared to LightGlue.

(a) SIFT RPCs (ARG) (b) LightGlue RPCs (ARG)
(c) SIFT RPCs (JAX) (d) LightGlue RPCs (JAX)
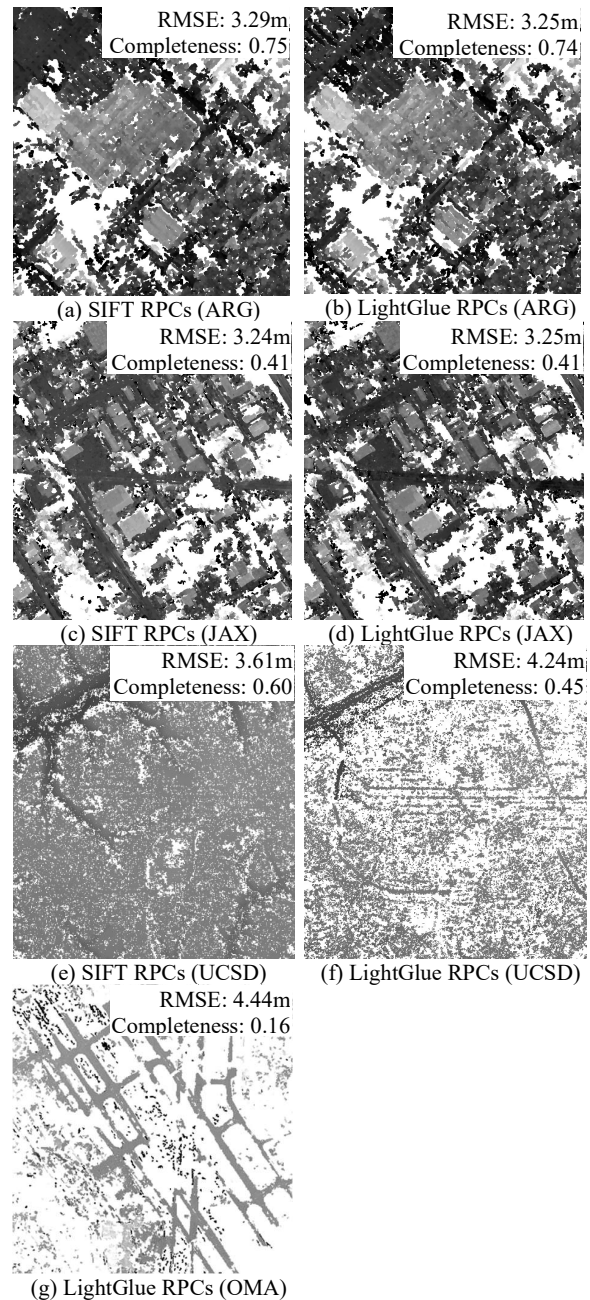(e) SIFT RPCs (UCSD) (f) LightGlue RPCs (UCSD)
(g) LightGlue RPCs (OMA)

Figure 9. Comparison between DSM using relative orientation from SIFT and LightGlue matches. SIFT RPCs of OMA is not included because they all failed in relative orientation.
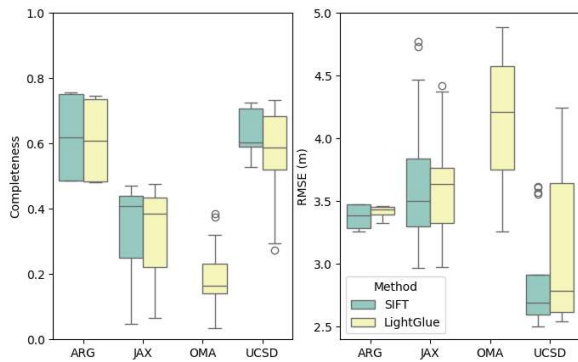
Figure 10. The box plots of DSM completeness and accuracy of SIFT and LightGlue on four regions. For ARG, only pairs where both methods successfully generated DSM are considered. As for OMA, all LightGlue pairs are considered.
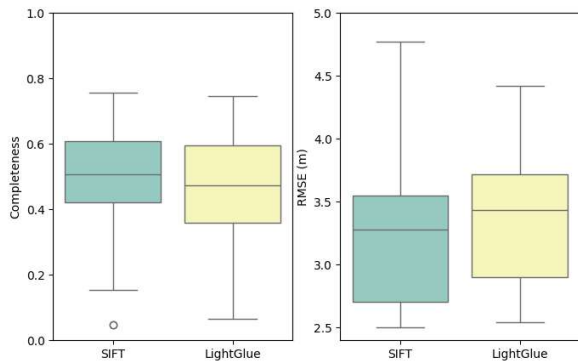

Figure 11. Comparison DSM quality of SIFT and LightGlue on all pairs where both methods successfully generated DSM in box plot.

4.4 Analysis of the LSM Effectiveness

Least Squares Matching (LSM) (Bellavia et al., 2024; Bethmann and Luhmann, 2010; Gruen, 1985) is a technique for patch-based point matching, and in practice, it is often used to refine the positions of tie points to achieve sub-pixel accuracy for geometric processing, i.e., relative orientation or bundle adjustment. Considering that tie-point extraction may be performed on a low-resolution layer of the pyramid (such as SIFT), in our experiment, we explore the effectiveness of using LSM to enhance the accuracy of the matches by adjusting the tie point locations. We assess the relative change in evaluation metrics (refer to Section 3.4) with and without LSM using Equation 2.

$$\text{Relative change} = \frac{m_{LSM} - m_{plain}}{m_{plain}} \times 100 \qquad (2)$$

where m is one of the previously defined metrics.

The relative changes (with and without applying the LSM) considers geometric processing statistics including inlier ratio, epipolar error, DSM completeness, and DSM accuracy across all pairs. The relative differences (by applying the LSM) are shown in Figure 12. It can be seen that SIFT and LightGlue statistics can be improved notably when being refined by LSM only in terms

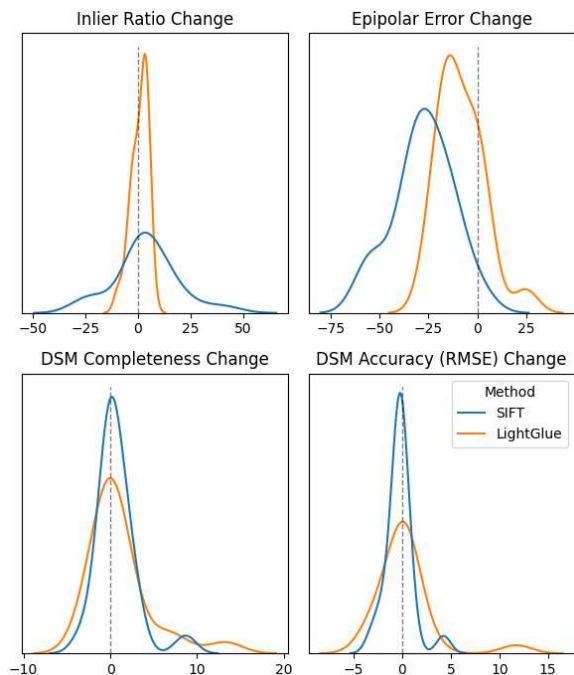of epipolar error change, with SIFT obtaining a gain almost twice than LightGlue.

Figure 12. The percentage changes in metrics due to applying LSM. For Inlier Ratio and DSM Completeness, higher values indicate better performance; whereas for epipolar error and DSM RMSE, lower values are better.

5. CONCLUSION

This work investigated the effectiveness of deep learning-based tie points methods in addressing geometric processing problems with off-track satellite stereo. Using a set of multi-date satellite images, we construct challenging stereo pairs, and assess the quality of tie points by assessing the resulting accuracy of relative orientation and, subsequentially the generated DSM.

Our findings revealed a noticeable improvement in the rate of DL-based successful matches compared to classic methods (i.e., SIFT). This was especially true in cases where the differences in sunlight and seasonal changes posed a challenge. Although DL-based methods provide more matches and are less sensitive to appearance changes, their overall matching quality in terms of epipolar error, completeness and RMSE on DSM, is slightly worse than classic algorithms. As results are promising, our future works aim to investigate the performance of other DL-based local features and matchers to support the extraction of geometric information from satellite offtrack stereo pairs.

REFERENCE

Albanwan, H., Qin, R., 2022. A comparative study on deep-learning methods for dense image matching of multi-angle and multi-date remote sensing stereo-images. The Photogrammetric Record 37, 385–409.

Barroso-Laguna, A., Riba, E., Ponsa, D., Mikolajczyk, K., 2019. Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5836–5844.

Bellavia, F., Morelli, L., Colombo, C., Remondino, F., 2024. Progressive Keypoint Localization and Refinement in Image Matching, in: Foresti, G.L., Fusiello, A., Hancock, E. (Eds.), Image Analysis and Processing - ICIAP 2023 Workshops, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 322–334.

Bethmann, F., Luhmann, T., 2010. Least-squares matching with advanced geometric transformation models. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 86–91.

Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G.D., Brown, M., 2019. Semantic Stereo for Incidental Satellite Images. Proc. IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1524–1532.

Brown, M., Goldberg, H., Foster, K., Leichtman, A., Wang, S., Hagstrom, S., Bosch, M., Almes, S., 2018. Large-scale public lidar and satellite image data set for urban semantic labeling, in: Laser Radar Technology and Applications XXIII. Presented at the Laser Radar Technology and Applications XXIII, SPIE, pp. 154–167.

d'Angelo, P., Rossi, C., Minet, C., Eineder, M., Flory, M., Niemeyer, I., 2014. High Resolution 3D Earth Observation Data Analysis for Safeguards Activities, in: Symposium on International Safeguards. Presented at the Symposium on International Safeguards, Vienna, Austria, pp. 1–8.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 224–236.

Facciolo, G., de Franchis, C., Meinhardt-Llopis, E., 2017. Automatic 3D Reconstruction From Multi-Date Satellite Images. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 57–66.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395.

Gruen, A., 1985. Adaptive Least Squares Correlation: A powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography Vol. 14, 175–187.

Gui, S., Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. ISPRS Journal of Photogrammetry and Remote Sensing 181, 1–19.

Hirschmuller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 328–341. https://doi.org/10.1109/TPAMI.2007.1166

Huang, D., Tang, Y., Qin, R., 2022. An evaluation of PlanetScope images for 3D reconstruction and change detection – experimental validations with case studies. GIScience & Remote Sensing 59, 744–761.

Ioli, F., Bruno, E., Calzolari, D., Galbiati, M., Mannocchi, A., Manzoni, P., Martini, M., Bianchi, A., Cina, A., De Michele, C., Pinto, L., 2023. A Replicable Open-Source Multi-Camera System for Low-Cost 4D Glacier Monitoring. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-1–2023, 137–144.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M., Trulls, E., 2021. Image Matching Across Wide Baselines: From Paper to Practice. Int J Comput Vis 129, 517–547.

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed, in: ICCV.

Ling, X., Huang, X., Qin, R., 2021. A Unified Framework of Bundle Adjustment and Feature Matching for High-Resolution Satellite Images. Photogrammetric Engineering & Remote Sensing 87, 485–490.

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110.

Maiwald, F., Lehmann, C., Lazariv, T., 2021. Fully Automated Pose Estimation of Historical Images in the Context of 4D Geographic Information Systems Utilizing Machine Learning Methods. ISPRS International Journal of Geo-Information 10, 748.

Marelli, D., Morelli, L., Farella, E.M., Bianco, S., Ciocca, G., Remondino, F., 2023. ENRICH: Multi-purposE dataset for beNchmaRking In Computer vision and pHotogrammetry. ISPRS Journal of Photogrammetry and Remote Sensing 198, 84–98.

Mishchuk, A., Mishkin, D., Radenovic, F., Matas, J., 2017. Working hard to know your neighbor's margins: Local descriptor learning loss, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Morelli, L., Bellavia, F., Menna, F., Remondino, F., 2022. Photogrammetry Now and Then–From Hand-Crafted to Deep-Learning Tie Points–. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 48, 163–170.

Morelli, L., Ioli, F., Maiwald, F., Mazzacca, G., Menna, F., Remondino, F., 2024. Deep-Image-Matching: A Toolbox for Multiview Image Matching of Complex Scenarios. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLVIII-2/W4-2024.

Poli, D., Remondino, F., Angiuli, E., Agugiaro, G., 2015. Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction. ISPRS Journal of Photogrammetry and Remote Sensing, High-Resolution Earth Imaging for Geospatial Information 100, 35–47.

Qin, R., 2019. A critical analysis of satellite stereo pairs for digital surface model generation and a matching quality prediction model. ISPRS Journal of Photogrammetry and Remote Sensing 154, 139–150.

Qin, R., 2016. Rpc Stereo Processor (rsp) – a Software Package for Digital Surface Model and Orthophoto Generation from Satellite Stereo Imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences III–1, 77–82.

Remondino, F., Menna, F., Morelli, L., 2021. Evaluating Hand-Crafted and Learning-Based Features for Photogrammetric Applications. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2021, 549–556.

Remondino, F., Morelli, L., Stathopoulou, E., Elhashash, M., Qin, R., 2022. Aerial Triangulation with Learning-Based Tie Points. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2022, 77–84.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF, in: 2011 International Conference on Computer Vision. Presented at the 2011 International Conference on Computer Vision, pp. 2564–2571.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. SuperGlue: Learning Feature Matching With Graph Neural Networks. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4938–4947.