Night and Day Aerial Photogrammetry

Luca Morelli [1], Giulio Perda [1], Salim Malek [1], Fabio Remondino [1], Klaus Legat [2]

[1] 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: (lmorelli, gperda, smalek, remondino)@fbk.eu

[2] AVT-Airborne Sensing Austria GmbH, Imst, Austria - Email: k.legat@avt.at

KEYWORDS: Day-night, Image matching, Deep Learning, Aerial triangulation, ALIKED, SuperPoint, LoFTR.

ABSTRACT:

Recent advancements in aerial imaging, including high-resolution sensors and integrated GNSS/IMU systems, have significantly enhanced photogrammetric methods for geospatial data acquisition. While most aerial data is captured during daylight, night-time imaging is increasingly being used in applications such as urban analysis and disaster assessment. However, automatic co-registration of day and night imagery remains challenging due to substantial radiometric differences. This study investigates the use of deep learning-based feature matching techniques for the alignment of multi-temporal, day-night aerial datasets. Experimental results show that feature extraction is highly sensitive to scale, with only a limited subset of deep learning (DL) methods—particularly ALIKED with LightGlue and SuperPoint with SuperGlue—proving robust under low-illumination conditions. Additionally, a U-Net-like model was trained to pre-process night-time images by approximating their radiometric characteristics to those of daytime images, enabling consistent feature matching across all tested methods. Among them, ALIKED with LightGlue offered the best balance between match quantity and computational efficiency. Object-space evaluations confirmed that the proposed pre-processing step significantly improves co-registration accuracy. The methodology offers a promising foundation for future multi-sensor and multi-modal image alignment tasks, including RGB-thermal and 2D-3D matching.

(a) (b)

Figure 1: Sample images of a day acquisition with a Vexcel UltraCam Dragon 4.1 camera (a) and a night acquisition with the IGI DigiCAM 150 camera (b).

1. INTRODUCTION

The improvements of aerial cameras and the integration of GNSS/IMU onboard sensors have been followed by numerous developments in automated photogrammetric methods for geospatial data generation and interpretation (Heipke and Rottensteiner, 2020; Kocaman et al., 2022).

Modern aerial platforms are equipped with cameras featuring large-format, high-spatial-resolution sensors, enabling GSD-level (ground sampling distance) reconstruction accuracy for both urban and rural landscapes. A common trend is the combined use of nadir and oblique imagery to enhance the reconstruction of vertical and sub-vertical surfaces (Remondino and Gerke, 2015; Toschi et al., 2017). These camera systems generally capture data in the visible spectrum, although there is an increasing request for multispectral or thermal bands acquisitions. This facilitates a broad range of applications, including the analysis of urban heat islands, land classification through spectral signature analysis, forestry assessments, etc. (Rodriguez et al., 2022; Beber et al., 2023). In some cases, these imaging systems are integrated with LiDAR sensors, which offer additional benefits such as the ability to penetrate vegetative cover and classify surfaces based on return time and other metrics (Toschi et al., 2019).

While most aerial data are usually acquired during daylight hours, the collection of night-time datasets (visible and thermal) is increasingly attracting interest and gaining traction in recent research and operational practices. For example, night-time imagery has been employed in the assessment of earthquake damage (Li et al., 2025) or, combined to daytime datasets, to support urban functional zone classification (Huang et al., 2021) or landscape monitoring (Santise et al., 2018).

Combining day and night imagery represents a specific multi-temporal acquisition scenario and the automatic co-registration of night-time imagery and daytime datasets poses challenges, primarily due to the substantial variations in illumination (Burdziakowski and Bobkowska, 2021). Conventional feature-matching algorithms, such as Scale-Invariant Feature Transform (SIFT - Lowe, 2004), often struggle under these conditions, necessitating the development or adaptation of more robust techniques.

1.1 Paper's Aim

This paper proposes a solution for the night and day co-registration problem in aerial photogrammetry. Indeed commercial (e.g. Metashape[1]) and open-source (e.g. COLMAP[2]) methods, based on SIFT-like feature matching, are unsuccessful in generating reliable correspondences between night and day image blocks.

The proposed approach is based on the extraction and matching of deep learning (DL) local features. These features are found with algorithms trained on large datasets that exhibit significant variation in viewing angles and - particularly relevant to this study - substantial radiometric differences caused by illumination changes (Jin et al., 2021). Deep learning-based local features

[1] https://www.agisoft.com/

[2] https://colmap.github.io/

https://doi.org/10.5194/isprs-archives-XLVIII-1-W4-2025-81-2025

have demonstrated considerable advantages in handling multi-temporal and multi-modal datasets (Song et al., 2024; Morelli et al., 2022). After initial experience with oblique blocks (Remondino et al., 2022), this work aims to evaluate the effectiveness of learning-based tie points for the co-registration of aerial night and day imagery using off-the-shelf pretrained models.

2. DATASETS

Two datasets (Table 1) over the city of Graz (Austria) are utilized:

- daytime RGB images (50) acquired with a Vexcel UltraCam Dragon 4.1[3] system (Farella et al., 2025);
- night-time RGB images (191) captured using the IGI DigiCAM camera.

A representative example of the imagery is shown in Figure 1. For both datasets, sensor position and orientation parameters, derived from the respective onboard navigation systems, are available and are used as initial approximations for the aerial triangulation process.

	Day RGB images	*Night RGB images*
Date	June 2024	November 2024
Camera	Vexcel UltraCam Dragon 4.1	IGI DigiCAM 150 (PhaseOne)
Sensor size	53.181 x 39.706 mm	53.407 x 40.051 mm
Sensor type	Sony IMX-411 (CMOS)	Phase One BSI (CMOS)
Original size	14144 x 10560 px	14204 x 10652 px
Reduced size	1024 x 764 px	1024 x 768 px
# images	50	191
Focal length	81 mm	40 mm
Above ground height	1060 m	860 m

Table 1: Night and day datasets information.

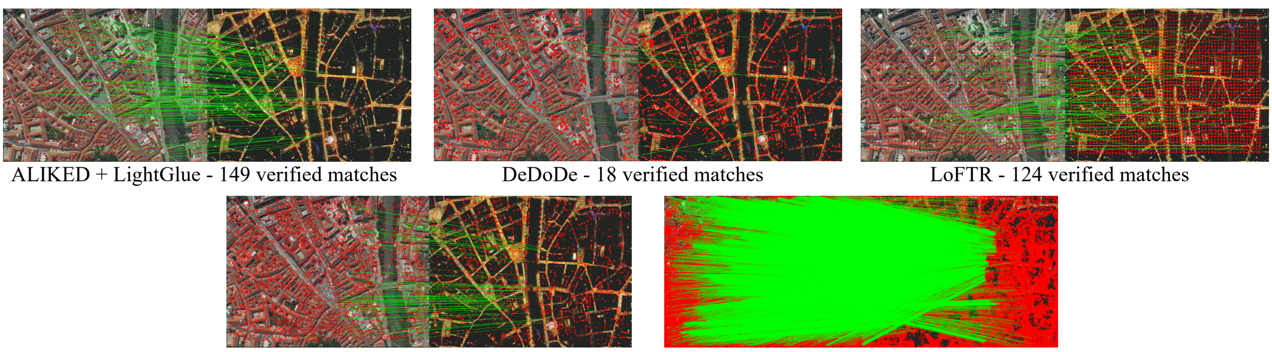
3. PRELIMINARY STUDY

Given the multiple learning-based methods available in the literature, preliminary experiments assessed the suitability of deep learning-based matching algorithms for handling significant radiometric and temporal differences between night and day images. A qualitative evaluation was conducted on a small dataset comprising nine images, with both night and day acquisitions covering the same urban area. The experiments employed the DIM[4] library (Morelli et al., 2024a), which provides a collection of both hand-crafted and deep learning-based local feature extractors and matchers, with the added capability of exporting correspondences directly into photogrammetric software. Given the high computational demands associated with processing high-resolution imagery, DIM supports high-resolution matching through a tiling strategy. The following DL methods were tested (Table 2): ALIKED (Zhao e t al., 2023) combined with LightGlue (Lindenberger et al., 2023), DISK (Tyszkiewicz et al., 2020) combined with LightGlue, DeDoDe (Edstedt et al., 2024a), SuperPoint (DeTone et al., 2018) combined with SuperGlue (Sarlin et al., 2020), SuperPoint combined with LightGlue, Key.Net (Barroso-Laguna et al., 2019) combined with HardNet (Mishchuk et al., 2017), LoFTR (Sun et al., 2021) and RoMa (Edstedt et al., 2024b). SIFT has been testes as representative of hand-crafted features.

However, no local feature method was able to generate a sufficient number of matches to triangulate the night and day multi-temporal images at full resolution. The only successful approach involved a significant downsampling of the images (Table 1). This suggests that discriminative correspondences, such as building corners or road intersections, can only be reliably detected at a lower resolution, where radiometric and structural differences between night and day images are less pronounced. Consequently, all subsequent experiments were conducted using downsampled image sets. Figure 2 presents a subset of methods that successfully produced matches passing the epipolar geometry verification. The figure shows that night and day matching is a complex task, which only a few DL local features can handle. Moreover, even among the methods that successfully detect tie points – such as DeDoDe, LoFTR and RoMa - a high number of outliers is observed.

Local features and matcher	*Computation time [s]*
SuperPoint + SuperGlue	33.96
SuperPoint + LightGlue	9.44
Disk + LightGlue	9.85
Aliked + LightGlue	8.18
DeDoDe + LightGlue	35.30
KeyNet + HardNet8	10.44
LoFTR	16.32
RoMa	488.92

Table 2: Computation time of different local features and matchers to extract tie points exhaustively in the subset of the full night and day block. Tests performed on 12th Gen Intel(R) Core(TM) i7-12700H 2.30 GHz, 32.0 GB RAM and NVIDIA GeForce RTX 3050.

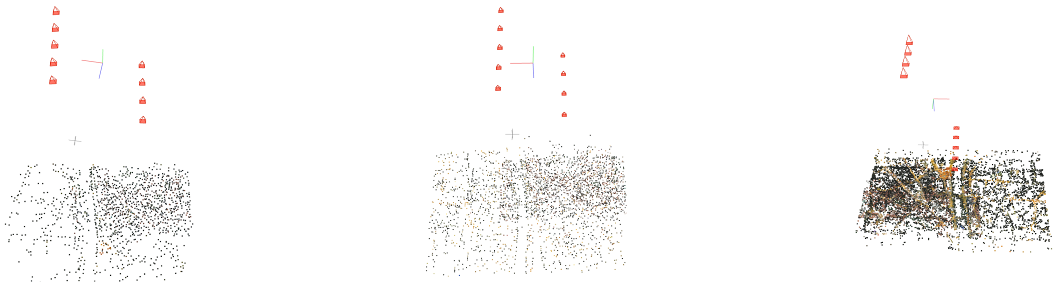


ALIKED + LightGlue - 149 verified matches
DeDoDe - 18 verified matches
LoFTR - 124 verified matches
SuperPoint + SuperGlue - 80 verified matches
RoMa - 8625 verified matches

Figure 2: Verified matches (RANSAC on epipolar geometry) visualized with green lines between a daytime and night-time image.

[3] https://www.vexcel-imaging.com/ultracam-dragon-4-1/

[4] https://github.com/3DOM-FBK/deep-image-matching



ALIKED + LightGlue   SuperPoint + SuperGlue   RoMa

Figure 3: The only three DL-based methods able to orient a subset of the night and day image block composed of 4 daytime (lower strip) and 5 night-time (upper strip).

Ultimately, the ALIKED and LightGlue combination was selected as the preferred matching method, because of its speed (Table 2) and capability of extracting valid correspondences and successfully orienting the entire subset (Figure 3). RoMa, a dense matcher, demonstrated the ability to orient the multi-modal block, but it was excluded due to its prohibitively high computational cost. SuperPoint combined with SuperGlue could be instead a valid alternative to ALIKED combined with LightGlue.

4. METHODOLOGY

Of the eight matching approaches evaluated in the preliminary study, only three produced reliable results for the co-registration of night and day imagery blocks. While this outcome is noteworthy due to its novel application in aerial photogrammetry, it is somewhat unexpected. Deep learning methods, which were included in the evaluation, are typically trained on challenging day-night image pairs, and a better ability to manage such datasets was expected. Therefore, an optional image enhancement step is introduced, which involves adjusting the radiometry of daytime images to resemble that of night-time images, with the goal of increasing the number of feature matches, hence the quality of the bundle adjustment. As anticipated in Section 3, the proposed solution for the co-registration of day-night multi-temporal image blocks is based on ALIKED combined with LightGlue, utilizing the pre-trained model provided by the original authors.

4.1 DL feature-based day-night image co-registration

Based on the initial investigation, all images are downsampled by a factor of approximately 14 - an empirically determined resolution at which correct tie points can be extracted using ALIKED+LightGlue, SuperPoint+SuperGlue and RoMa. The matching is conducted using the DIM library, employing a brute-force strategy that exhaustively attempts to match daytime and night-time images. No image tiling is required when working at this resolution.
Image orientation is performed within DIM using *pycolmap* – the Python bindings of COLMAP – keeping the interior parameters constant for the Vexcel UltraCam Dragon camera as provided in the calibration certificate. The IGI DigiCAM camera was instead processed in self-calibration allowing for radial distortion parameters to be estimated. DIM outputs the 3D tie points and camera orientation in a COLMAP reconstruction format which is loaded with *pycolmap* to finalize the orientation. The camera poses of the night images obtained from the onboard inertial navigation system of the aerial platform are utilized as constraints in a final bundle adjustment. To emphasize the co-registration errors of the night-time block on the daytime block, only daytime poses are used. In *pycolmap*, the bundle adjustment cost function minimizes at the same time the reprojection error of the tie points and the residuals on the night images positions with a prior accuracy of 5 cm.
The final output is hence a COLMAP reconstruction scaled and georeferenced in the reference system of the night images. The reconstruction is exported in Bundler format and imported into Metashape to independently generate dense point clouds for both night and day images. The proposed method does not rely on ground control points (GCPs), as it is designed to be a fully automatic co-registration approach. However, incorporating GCPs into the pre-oriented day and night image blocks would likely further enhance the accuracy of the reconstruction.

Figure 4: Example of 512x512 px patch pairs used for the training process and extracted from daytime and night-time orthophotos.

4.2 Image enhancement

Prior to extracting keypoints with ALIKED, an optional pre-processing step is proposed in which the radiometry of daytime images is transformed to more closely resemble that of night-time images. This is achieved using an in-house U-Net-like architecture (Ronneberger et al., 2015), i.e. a convolutional neural network trained in a supervised manner with paired RGB daytime and night-time images. The objective is to enable the prediction of a night-time-like version of any given daytime image at inference time. Given the difficulty of acquiring drone imagery captured from precisely the same viewpoint during both day and night, the network was trained using image tiles from georeferenced daytime and night-time orthophotos.

The developed U-Net model is trained on a custom training dataset comprising aligned daytime (input) and night-time (output) image pairs. The initial dataset is composed of an aligned pair of images which is partitioned into three subsets: 70% for training, 10% for validation, and 20% for testing. The training dataset is then created by applying a random cropping of size 512×512 to the training area, as shown in Figure 4. In total 6400 patches of size 512×512 are generated for the training, half of them by using augmentation techniques based on Geometric transformations (flipping and rotation) and Color space transformations (brightness adjustment).
The U-Net is trained for a maximum of 100 epochs with a batch size of 8 and an initial learning rate of 1×10^{-4}, optimized using the ADAM method (Kingma & Ba, 2015). The loss function is defined as the mean absolute error (*MAE*) between the predicted image (y) and the ground truth night-time image (x).

$$MAE(x,y) = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|$$

where N is the total number of pixels.

5. RESULTS

5.1 Evaluation of image enhancement

RGB daytime images are transformed into night-time-like images to improve matching performance on day-night image pairs. The U-Net model effectively adjusts the radiometry of daytime images to resemble that of night-time images, as shown in Figure 5. Inaccuracies in the radiometric transformation spuriously appear as the exact negative of the expected outcome. Regardless, details over smaller well-lit objects (road marks, cars) and the geometries of buildings are preserved. To evaluate the effects of this radiometric transformation on the co-registration of the day-night image block, four daytime images included in the nine-image day-night block of the preliminary study are transformed with U-Net. Exhaustive image matching is then carried out with the matching algorithms listed in Section 3.

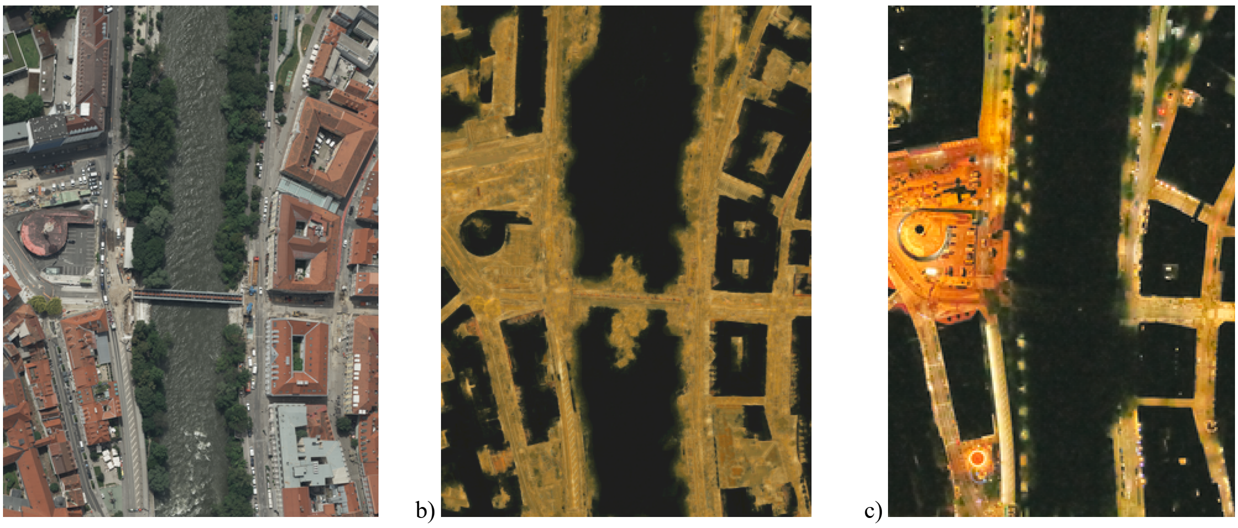

a) b) c)

Figure 5: Original daytime image (a) transformed to night-time with U-Net (b) to match the radiometry of an original night-time image over the same urban area (c).

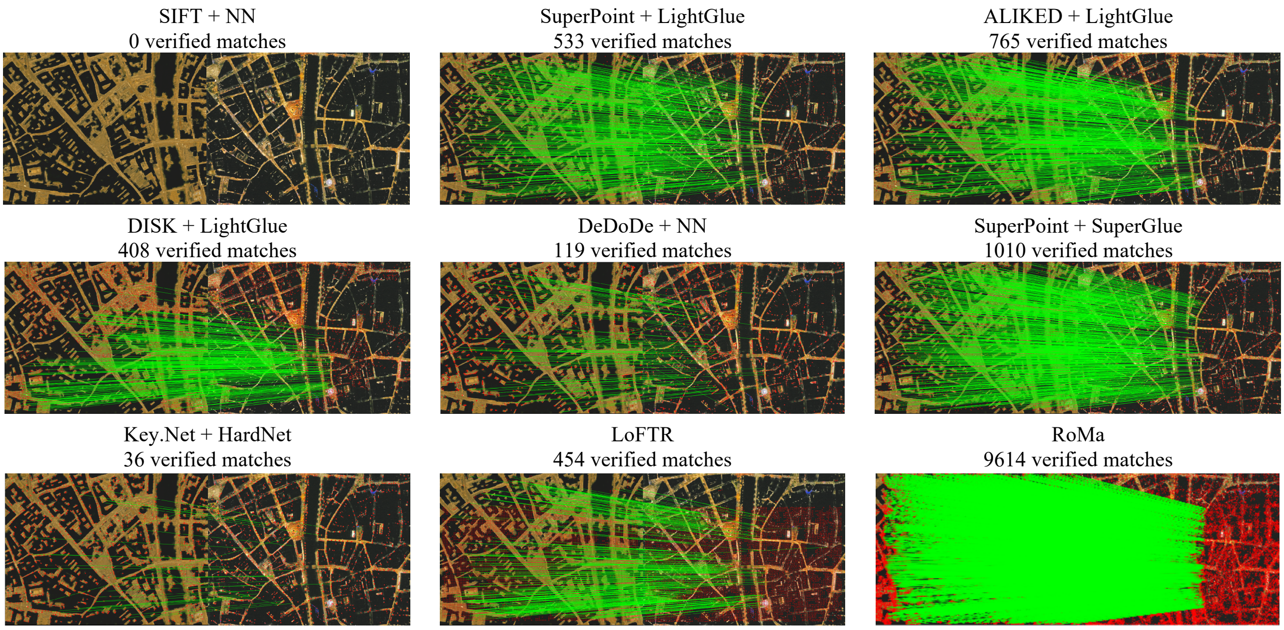


Figure 6: Example of feature matching between a transformed night-time image (from a daytime image, left) and an original night-time image (right), across different methods. NN= nearest neighbour matching.

Figure 6 presents image matching examples on an image pair, reporting the number of tie points which pass epipolar geometry verification. Compared to the preliminary analysis conducted on the same image pair, although there are clear differences in the number of tie points - ranging from just 36 for Key.Net+HardNet to 9614 for RoMa - all methods appear to benefit from the current enhancement, as they are now able to find correct matches.

For methods that had previously succeeded in producing tie points even without radiometric adjustment, a significant increase in the number of verified tie points is now observed: ALIKED+LightGlue increased from 149 to 765, DeDoDE from 18 to 119, LoFTR from 124 to 454, SuperPoint+SuperGlue from 80 to 1010, and RoMa from 8625 to 9614. Additionally, a qualitative reduction in outliers is noticeable in the results produced by LoFTR and RoMa.

These findings suggest that, although the tested matching algorithms are originally trained on images with substantial radiometric variation - including day-to-night pairs—they still benefit from the radiometric adjustment to achieve optimal matching performance.

Finally, Figure 7 illustrates the orientation of the 9-image block after the four daytime images were transformed to night-time using the U-Net model. All matching approaches, except for LoFTR, result in a qualitatively correct sparse reconstruction of the block. In the case of LoFTR, the daytime images fail to orient properly and are incorrectly aligned along a straight line, corresponding to the original flight path.

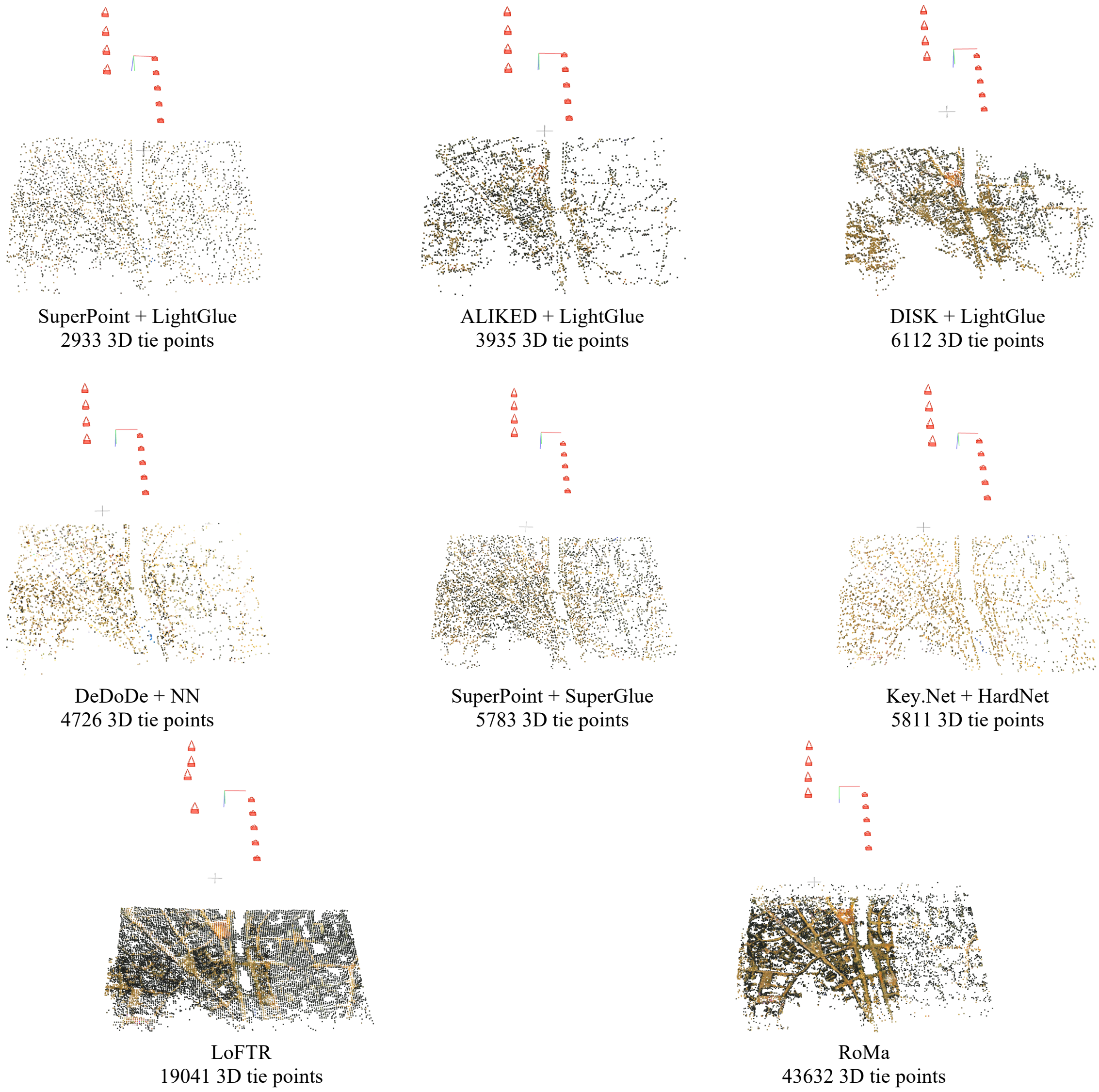


Figure 7: Orientation of a subset of the night and day image block, 4 daytime (lower strip) and 5 night-time (upper strip), after radiometric adjustment with U-net. NN= nearest neighbour matching.

5.2 Evaluation in object space

Figure 8 presents the results of night and day concurrent image triangulation using image correspondences extracted by coupling ALIKED and LightGlue. Initial experiments revealed that ALIKED outperformed other methods in both speed and the number of tie points detected. SuperPoint + SuperGlue also proved to be a viable alternative. While it extracted fewer tie points than ALIKED during the preliminary phase, it outperformed ALIKED in terms of tie point extraction following image enhancement with U-Net. However, SuperPoint + SuperGlue is significantly slower than ALIKED in tie point extraction. Consequently, given ALIKED's strong performance both with and without image enhancement, along with its computational efficiency, it was selected as the deep learning-based feature extractor for the object space analysis.

The accuracy of the co-registration is evaluated by comparing the day and night 3D dense reconstructions, since the objective of the proposed method is to automatically co-register night and day image blocks. A cloud-to-cloud (C2C) distance is used as metric. Figures 8a and 8b display the sparse point cloud and camera orientations in COLMAP with GNSS-INS positions used as constraint in the adjustment. The C2C distance computed between the day and night dense point clouds generated with the orientation results based on the original (Figure 8c) and with U-Net enhanced (Figure 8d) images. On average, the U-net camera network achieves a 8 cm lower distance between day and night points (Table 3), highlighting that the camera poses estimated with the transformed images are more accurate. Improved orientation is associated with a higher number of tie points (Jin et al., 2021) extracted on images enhanced with U-Net, which likely also enhances the robustness to outliers during filtering using RANSAC and epipolar geometry.

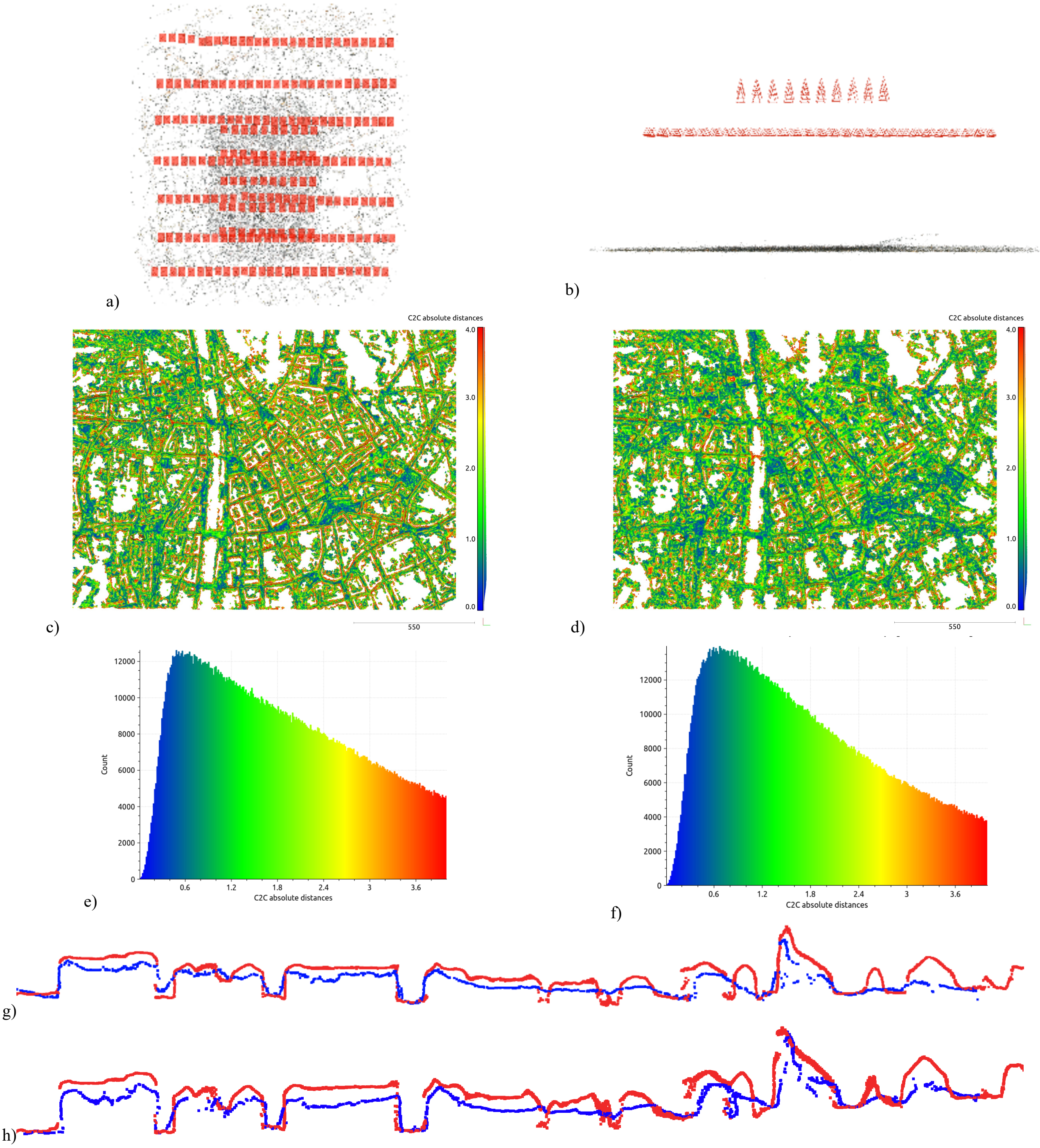


Figure 8: Results of the joint constrained adjustment of the night and day images using ALIKED + LightGlue in COLMAP (a – top view, b – side view). C2C distances between night and day point clouds generated with poses from the orientation of the original images (c) and from poses after U-Net processing (d). Histograms of C2C distances for the original images (e) and U-Net pre-processed images (f). Cross-section comparison in meters of day and night point clouds (**blue**: night image block; **red**: day image block) processing original images (g) and pre-processed with U-Net (h).

Applying an Iterative Closest Point (ICP) algorithm to the day-night point clouds results in negligible changes in the Cloud-to-Cloud (C2C) distance (Table 3). This indicates that the proposed automatic co-registration method achieves accuracy comparable to that obtained through manual co-registration followed by ICP refinement - a procedure commonly adopted in practice.
Additionally, a qualitative comparison was provided through a representative cross-sectional analysis. Figure 8g presents the cross-section for the original images, while Figure 8h corresponds to the images pre-processed using U-Net. Overall, the night-time point cloud (in blue) is frequently shifted in height, it shows much less details and is much noisier, primarily due to reduced illumination, particularly noticeable on building rooftops that are often completely dark in the night images.

C2C error [m]	*Before ICP*	*After ICP*
Original images	1.80	1.78
U-net enhancement	1.72	1.73

Table 3: C2C error between night and day point clouds using original or enhanced images (i.e. daylight images converted to night radiometry with U-Net). The orientation is performed with ALIKED + LightGlue.

6. CONCLUSIONS

Co-registering day and night aerial images presents a significant challenge. Preliminary investigations have demonstrated that successful feature extraction in this context depends heavily on the scale at which local features are detected. In particular, the high level of noise typically present in night-time imagery makes fine-scale feature extraction unreliable, forcing substantial downsampling to achieve consistent results. Even at coarser resolutions, only a limited subset of deep learning-based feature extractors - specifically ALIKED combined with LightGlue and SuperPoint combined with SuperGlue - proved effective for establishing night and day correspondences. Though RoMa also demonstrated the ability to extract suitable features, its dense matching strategy results in prohibitively long processing times.
To address the limited efficiency of deep learning-based methods, it has been demonstrated that a neural network—specifically, U-Net—can be trained to approximate the radiometric characteristics of daytime images to those of night-time images prior to matching. This additional pre-processing step enabled the orientation of the day-night image block using all eight deep learning-based methods evaluated.
With or without pre-processing, ALIKED combined with LightGlue demonstrated the best trade-off between the number of extracted matches and processing speed. Therefore, it was selected for metric evaluation in object space, where pre-processing with U-Net led to a significant improvement in the co-registration accuracy of the night-on-day image block. Future work will explore the possibility of extending this technique to the matching of images acquired from different sensors and modalities, such as RGB-thermal or 2D-3D (Morelli et al., 2024b).

REFERENCES

Barroso-Laguna, A., Riba, E., Ponsa, D. and Mikolajczyk, K., 2019. Key. net: Keypoint detection by handcrafted and learned CNN filters. Proc. *ICCV*, pp. 5836-5844.

Beber, R., Perda, G., Takhtkeshha, N., Remondino, F., Maffei, T., Poli, D., Moe, K., Cipriano, P., and Ciliberti, M., 2023. Multi-modal geospatial and thematic data to foster Green Deal applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-1/W3-2023, 9–16.

Burdziakowski, P., Bobkowska, K., 2021. UAV Photogrammetry under poor lighting conditions – Accuracy considerations. *Sensors*, 21, 3531.

Diederik P. Kingma, Jimmy Ba., 2015. Adam: A Method for Stochastic Optimization, *ICLR*.

DeTone, D., Malisiewicz, T. and Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. Proc. *CVPR*, pp. 224-236.

Edstedt, J., Bökman, G., Wadenbäck, M. and Felsberg, M., 2024a. DeDoDe: Detect, don't describe - Describe, don't detect for local feature matching. Proc. *3DV,* pp. 148-157.

Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M. and Felsberg, M., 2024b. RoMa: Robust dense feature matching. Proc. *CVPR,* pp. 19790-19800.

Farella, E.M., Morelli, L., Remondino, F., Qin, R., Schachinger, B. and Legat, K., 2025. Investigating the new Ultracam Dragon hybrid aerial mapping system. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLVIII-M-6, pp. 117-125.

Heipke, C., Rottensteiner, F., 2020. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geospatial Information Science*, Vol. 23.

Huang, X., Yang, J., Li, J. and Wen, D., 2021. Urban functional zone mapping by integrating high spatial resolution night-time light and daytime multi-view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *175*, pp.403-415.

Kocaman, S., Akca, D., Poli, D., Remondino, F., 2022. *3D/4D City Modelling - From Sensors to Applications.* Whittles Publishing, 224 pages.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M. and Trulls, E., 2021. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, *129*(2), pp.517-547.

Li, F., Liao, S., Fu, X. and Liu, T., 2025. NPP-VIIRS night-time lights illustrate the post-earthquake damage and subsequent economic recovery in Hatay Province, Turkey. *ISPRS International Journal of Geo-Information*, *14*(4), p.149.

Lindenberger, P., Sarlin, P.E. and Pollefeys, M., 2023. Lightglue: Local feature matching at light speed. Proc. *ICCV,* pp. 17627-17638.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, Vol., 60, pp. 91-110.

Mishchuk, A., Mishkin, D., Radenovic, F. and Matas, J., 2017. Working hard to know your neighbor's margins: Local descriptor learning loss. *Advances in neural information processing systems*, *30*.

Morelli, L., Bellavia, F., Menna, F. and Remondino, F., 2022. Photogrammetry now and then – from hand-crafted to deep-learning tie points. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLVIII-2/W1-2022, 163–170.

Morelli, L., Ioli, F., Maiwald, F., Mazzacca, G., Menna, F. and Remondino, F., 2024a. Deep-image-matching: a toolbox for multiview image matching of complex scenarios. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 48(2-W4-2024), pp.309-316.

Morelli, L., Perda, G., Ioli, F., Trybała, P., Sterpin, A., Rigon, S., Sutherland, N., Medici, M., Remondino, F. and Vitti, A., 2024b. Co-registering laser scanning point clouds and photogrammetric images with deep learning multi-modal matching. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 48*, pp.335-342.

Remondino, F., Gerke, M., 2015. Oblique aerial imagery: a review. *Proc. Photogrammetric Week*, 75-83.

Remondino, F., Morelli, L., Stathopoulou, E., Elhashash, M. and Qin, R., 2022. Aerial triangulation with learning-based tie points. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLIII-B2-2022, 77–84.

Rodríguez, M.V., Melgar, S.G. and Márquez, J.M.A., 2022. Assessment of aerial thermography as a method of in situ measurement of radiant heat transfer in urban public spaces. *Sustainable Cities and Society*, Vol. 87, 104228.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, Springer, Cham, 234-241.

Santise, M., Thoeni, K., Roncella, R., Diotri, F., Giacomini, A., 2018. Analysis of low-light and night-time stereo-pair images for photogrammetric reconstruction. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 1015–1022.

Sarlin, P.E., DeTone, D., Malisiewicz, T. and Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. Proc. *CVPR*, pp. 4938-4947.

Schonberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. Proc. *CVPR*, pp. 4104-4113.

Song, S., Morelli, L., Wu, X., Qin, R., Albanwan, H. and Remondino, F., 2024. Evaluating learning-based tie point matching for geometric processing of off-track satellite stereo. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, *48*, pp.393-400.

Sun, J., Shen, Z., Wang, Y., Bao, H. and Zhou, X., 2021. LoFTR: Detector-free local feature matching with transformers. In Proc. *CVPR*, pp. 8922-8931.

Toschi, I., Ramos, M.M., Nocerino, E., Menna, F., Remondino, F., Moe, K., Poli, D., Legat, K. and Fassi, F., 2017. Oblique photogrammetry supporting 3D urban reconstruction of complex scenarios. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* Vol. XLII-1-W1, pp. 519-526.

Toschi, I., Remondino, F., Hauck, T., Wenzel, K., 2019. When photogrammetry meets LiDAR: Towards the airborne hybrid era. *GIM International*, 17-21.

Tyszkiewicz, M., Fua, P. and Trulls, E., 2020. DISK: Learning local features with policy gradient. *Advances in Neural Information Processing Systems*, *33*, pp.14254-14265.

Zhao, X., Wu, X., Chen, W., Chen, P.C., Xu, Q. and Li, Z., 2023. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Transactions on Instrumentation and Measurement*, *72*, pp.1-16.