Stereo Matching and Digital Surface Model Generation for Satellite Imagery: From Scanline Aggregation to Deep Learning with RAFTStereo

Yazgı Nur Sayın [1, 2], Ali Özgün Ok [1, 3]

[1] Hacettepe University, Dept. of Geomatics Engineering, Ankara, Türkiye –
yazgi.sayin@hacettepe.edu.tr, ozgunok@hacettepe.edu.tr
[2] TUBITAK Space Technologies Research Institute, Ankara, Türkiye – yazgi.sayin@tubitak.gov.tr
[3] National Intelligence Academy, Dept. of Computer Engineering, Ankara, Türkiye – a.ok@mia.edu.tr

Keywords: Stereo Matching, DSM Generation, Satellite Imagery, More Global Matching, Semi Global Matching, RAFTStereo.

Abstract

Digital Surface Model (DSM) generation from satellite stereo images is one of the key applications in both computer vision and photogrammetry. Recent progress in high-resolution satellite imaging and deep learning has favoured their applications for accuracy enhancement and automation in DSM generation. However, complex acquisition geometries from satellite imaging, regions of repetitive textures or patterns, and varying atmospheric conditions continue to complicate the process of dense stereo matching. This study presents a detailed framework for DSM generation: image preprocessing, epipolar rectification, disparity estimation, and 3D reconstruction. At the image matching stage of stereo images, it compares the traditional methods such as Semi-Global Matching (SGM) and More Global Matching (MGM) with a deep learning-based approach-RAFTStereo. Experimental results with WorldView-3 satellite stereo pairs of the Data Fusion Contest 2019 (DFC2019) dataset show that while SGM and MGM remain robust and computationally efficient, RAFTStereo performs better especially on radiometrically and geometrically complex scenes. MGM provides numerical errors at the lowest values ≈2-4 meters, while RAFTStereo offers more coherent disparity maps, with more smooth surfaces, and fewer artifacts. These results also point out the complementary nature of traditional approaches and learning-based methodologies.

1. Introduction

A Digital Surface Model (DSM) provides a three-dimensional representation of the Earth's surface, including both natural (topography) and man-made features like buildings and other infrastructure. DSMs are used in many tasks like terrain modelling, environmental modelling, urban planning, infrastructure monitoring, telecommunication planning, disaster risk management, and so on.

Stereo image analysis has been very significant in the geospatial domain, as it enables the derivation of high-resolution disparity from satellite imagery for the generation of accurate DSM. Despite the progress, accurate DSM generation from a satellite stereo image remains still challenging: agriculture or water surfaces void of texture, shadows, and repetitive patterns create ambiguity for disparity estimation, while atmospheric and radiometric variations between acquisitions make matching even more challenging. These challenges will arise because there is a pressing need for robust and general stereo matching algorithms/approaches that can work under various conditions of imagery.

The whole stereo processing usually consists of two parts: stereo rectification and dense disparity estimation. Rectification aligns the corresponding epipolar lines horizontally to simplify the matching process. Then, traditional or learning-based stereo matching methods estimate the disparity map, which encodes the pixel displacements between the image pair. For satellite images, traditional stereo matching methods such as Semi Global Matching (SGM) (Hirschmüller, 2007) are still popular due to their balance of efficiency and robustness (Rothermel et al., 2012). Through combining matching cost at pixel-wise level with smoothing constraints at multiple path aggregation, the SGM could provide reliable disparity maps featured by remote sensing (Xia et al., 2020). In addition, More Global Matching (MGM) improves upon SGM by incorporating additional information from previously visited neighbouring pixels, enhancing disparity consistency and robustness, particularly in angular directions (Facciolo et al., 2015). MGM also involves computing local matching costs, directional aggregation to reduce ambiguity, selecting minimum cost disparity, and post-processing refinement. Although very successful, traditional approaches suffer from complex scenes containing low/no texture and changing illumination characteristics, which are very typical seasonal scene properties also observed in spaceborne imagery datasets. To overcome these limitations, some fusion models have been proposed, obtaining large performance improvements for different satellite datasets (Gómez et al., 2023).

The incorporation of Convolutional Neural Networks (CNN) into stereo matching, on the other hand, brought about a change in perspective. Early models, such as MC-CNN (Žbontar and LeCun, 2016), succeeded in learning matching costs from examples, which was followed by the end-to-end architectures like GANet (Zhang et al., 2019). The latter integrated SGM concepts into a differentiable framework, resulting in notable accuracy and efficiency gains (Xia et al., 2022; Gómez et al., 2022). Other more recent developments, most notably RAFTStereo (Lipson et al., 2021), rely on correlation volumes and the recurrence of refinement modules, inspired by the Recurrent All-Pairs Field Transforms (RAFT) model (Teed and Deng, 2020), to iteratively update disparities and enhance local-global consistency. Models of this type are promising for producing smoother and more complete DSMs, even in radiometrically challenging scenes. This study appraises two traditional scanline aggregation-based stereo matching approaches, SGM and MGM, together with the deep learning approach RAFTStereo, against benchmark data to identify its relative performance in disparity estimation and DSM production. Our aim is to highlight their respective strengths, limitations, and applicability to geospatial tasks and to evaluate

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-205-2025 | © Author(s) 2025. CC BY 4.0 License.

the efficacy of the methods for digital surface model utilising high-resolution stereo images. To do that, we investigated two benchmark datasets, i.e., the WHU Stereo satellite dataset (Li et al., 2022) and Data Fusion Contest 2019 (Bosch et al., 2019; Le Saux et al., 2019). The remainder of this paper is organised as follows: Section 2 introduces the three stereo image matching methods under investigation, namely SGM, MGM, and RAFTStereo, and the pipeline framework utilised. Section 3 introduces the GF-7 and WV-3 datasets and the processing steps, including training configuration. Section 4 presents the experimental setup and discusses the comparative results, emphasising the strengths and weaknesses of each approach. Finally, Section 5 concludes the paper with a summary of findings and directions for future research.

2. Methods Evaluated for Stereo Matching

Stereo matching, also known as correspondence matching, aims to derive a disparity map from rectified image pairs by measuring how objects shift between two viewpoints, providing indirect information about surface depth or elevation. In most workflows, the process follows several stages: computing the initial matching cost, aggregating or smoothing these costs, optimising disparities, and finally applying post-processing to refine the output.

Classical stereo approaches rely on low-level image features extracted from local neighbourhoods around each pixel (Hirschmüller and Scharstein, 2008). Their accuracy is often constrained because the matching cost functions depend on hand-crafted features. To overcome these limits, numerous deep-learning-based methods have been developed, offering substantial gains in depth estimation accuracy. Yet, deep architectures are memory-intensive and can be impractical for very-high-resolution satellite or aerial imagery, where the GPU must store large three-dimensional cost volumes. As a result, traditional algorithms remain attractive for processing relatively large datasets. Although deep learning is becoming mainstream across computer vision, most operational satellite stereo pipelines still rely on classical frameworks (Gómez et al., 2022).

2.1 SGM

SGM was developed to reduce the computational cost associated with fully global stereo methods (Hirschmüller, 2007). The algorithm is operating on epipolar-rectified image pairs and consists of four major steps: computation of matching costs, cost aggregation, disparity optimisation, and refinement. To derive a reliable depth or disparity map, SGM minimises a discontinuity-preserving energy function. Unlike traditional global methods, this energy is aggregated along multiple one-dimensional (1D) paths—such as left-to-right or top-to-bottom—rather than across a full two-dimensional (2D) grid.

Typically, information from 8 or 16 path directions is accumulated (Figure 1, left). For each pixel, costs from all directions are combined, and the disparity with the lowest total cost is chosen following the “winner-takes-all” strategy. While this approach is computationally efficient, it may introduce streak-like artefacts because cost aggregation along each path is performed independently of others.

2.2 MGM

MGM incorporates 2D contextual information into the 1D path-wise optimisation framework of SGM (Facciolo et al., 2015). This is effectively accomplished by utilising messages passed from previously visited pixels along the previous scanline (i.e., from pixels above). In the MGM algorithm, the matching decision at a given pixel is shaped not only by a few immediate neighbours but also by a broader region from the relevant quadrant (Figure 1 - right). This represents a major distinction from SGM, which only considers information propagated from specific directions.

MGM defines a dedicated scanning order for each direction (e.g., top-right, bottom-left) and accordingly computes directed cost accumulation values for each direction. The costs accumulated from all directions are then merged, correcting for redundant information. This merging is performed according to a formula, and the final matching is determined using the winner-takes-all method, selecting the disparity with the minimum total cost.

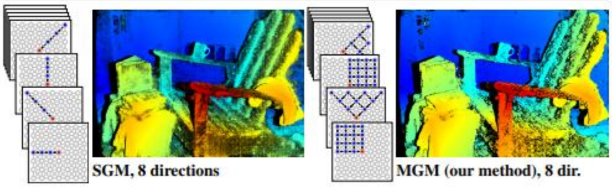

Figure 1. SGM and MGM approaches (Facciolo et al., 2015)

2.3 RAFTStereo

RAFTStereo is a deep learning framework tailored for stereo depth estimation (Lipson et al., 2021). It is adapted from the RAFT optical flow model (Teed and Deng, 2020), where optical flow defines the apparent motion between two consecutive images. In this stereo case, RAFTStereo estimates the horizontal shift-disparity-between rectified image pairs to reconstruct scene depth. Its key novelties are the introduction of multi-level GRUs, an efficient refinement strategy in the cost volume, and richer real-time performance.

The model uses a multi-scale GRU to efficiently propagate information between different spatial resolutions. Unlike traditional stereo networks, which originally adopted heavy 3D convolutional layers for cost-volume processing, RAFTStereo uses lighter 2D convolution with simpler updates to the cost volume. This architectural choice enables much faster inferences and makes the network practical for real-time or near-real-time applications.

RAFTStereo follows the dual encoder design, which consists of a feature encoder and a context encoder; each independently extracts features from Figure 2. The feature encoder takes both left and right input images and generates the dense feature maps, which form the correlation volume. This phase is implemented using residual blocks, with progressive downsampling by a factor of four or eight over the original spatial resolution, depending on the number of layers. Instance normalisation is used across these layers. The context encoder shares an overall similar structure, but with batch normalisation instead of instance normalisation, it is applied only to the left image. These features from the context encoder initialise the update module's hidden state and are repeatedly fed to the GRU at each iteration step.

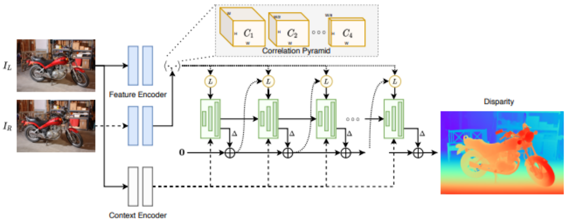

Figure 2. RAFTStereo Architecture (Lipson et al., 2021)

The correlation volume encodes the similarity information that allows the network to identify pixel correspondences between the left and right images. The visual similarity of every pixel in the reference image against pixels in the opposite view is computed, and a multidimensional volume is generated that will guide disparity prediction. To further ease this, RAFTStereo constructs a correlation pyramid by successively applying average pooling along the depth dimension; four correlation levels result at different spatial resolutions. This multiscale representation helps the network capture coarse structures at low resolution and finer details at higher resolution. Disparity estimates are iteratively refined across these scales to improve overall accuracy.

During the correlation lookup, one dimensionally spans a grid of integer offsets around the current disparity estimate. This grid serves as an index to sample correlation features from different levels of the pyramid. As these coordinates are continuous, not discrete, value sampling is done through bilinear interpolation to ensure smoothness. These sampled responses are aggregated into a single feature map. The network repeatedly updates the disparity map at multiple scales, normally 1/8, 1/16, and 1/32 of the original image resolution. With this multiresolution updating, RAFTStereo can handle broad texture regions as well as regions with subtle structural details.

While this design supports a GRU-based update process, RAFTStereo originally used different backbones to extract correlation and context features; however, experiments merged these into one backbone, accelerating the inference without sacrificing accuracy. The final RAFTStereo design adopts this single-backbone configuration, making a satisfactory balance between computational efficiency and predictive precision.

2.4 Stereo Processing Pipelines Utilised

Over the last years, quite a number of stereo processing pipelines have been proposed for satellite imagery, including the Satellite Stereo Pipeline (S2P). These systems emphasise end-to-end automation of pre-processing, image matching, and triangulation stages with the integration of domain-specific geometric corrections. S2P is a free, modular system designed at École Normale Supérieure Paris-Saclay in collaboration with CNES, allowing automatic generation of DSMs from high-resolution optical satellite stereo pairs in a repeatable manner (De Franchis et al., 2014).

Optimised for pushbroom sensor geometry, the pipeline consists of four major steps: geometric pre-processing, epipolar rectification, disparity estimation, and 3D triangulation (Figure 3). The S2P assumes the Rational Function Model (RFM) for approximations, which shows a nonlinear epipolar geometry inherent in pushbroom sensors, using first-order Taylor expansions to local affine transformations. Dense stereo matching is then performed with SGM or MGM using census-based matching costs to achieve robust disparity estimation.

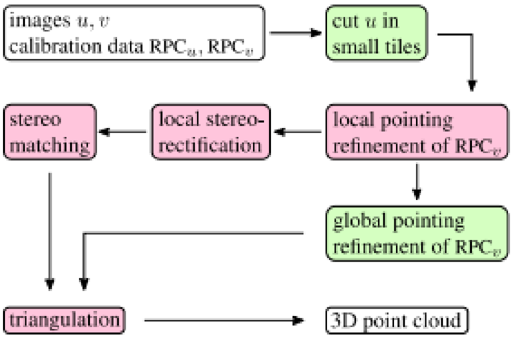
images u, v
calibration data RPC_u, RPC_v
cut u in small tiles
local pointing refinement of RPC_v
local stereo-rectification
stereo matching
global pointing refinement of RPC_v
triangulation
3D point cloud

Figure 3. S2P Stages (De Francis et al., 2014)

3. Benchmark Datasets

Li et al. (2022) generated the WHU-Stereo satellite benchmark dataset using the GF-7 satellite, which is equipped with a dual-line array stereoscopic camera system and provides high-resolution stereo imagery. High-quality in-track stereo pairs are acquired by the simultaneous panchromatic image capturing from forward and backward viewing angles of approximately −5° and +26° at a ground sampling distance (GSD) finer than 0.8 meters. This dataset is a challenging benchmark for stereo matching of high-resolution satellite images and performance evaluation of deep learning models.

The Data Fusion Contest 2019 (DFC2019) benchmark dataset (Bosch et al., 2019; Le Saux et al., 2019) includes multi-date, multi-view, and multi-spectral satellite imagery with high-resolution 3D ground truth. It provides over 20 GB of WorldView-3 imagery and LiDAR-derived DSMs and point clouds, covering ~320 km² of urban areas in Jacksonville, Florida, and Omaha, Nebraska, with 1.3 m VNIR and 35 cm panchromatic GSD.

Patil and Guo (2023) introduced the Stellar dataset which was a new and challenging large-scale benchmark for satellite stereo matching. This dataset integrates LiDAR-derived DSMs as ground truth for training and evaluating stereo matching algorithms.

SGM and its extensions, the MGM framework, have built-in support in the S2P pipeline. Besides these, it includes several algorithmic variants, such as tvl1, msmw, sgbm, and mgm_multi, that allow flexible adaptation of stereo processing to diverse terrain and radiometric conditions (De Franchis et al., 2014). Besides that, cost functions, regularisation strength, and consistency filtering are configurable with S2P, and all these options can be fine-tuned for disparity estimation accuracy and robustness without a need to change the core of the algorithmic structure.

RAFTStereo iteratively updates the disparity map by using a multi-level recurrent GRU-based architecture, leveraging correlation pyramids to maintain the global context. It works directly with high-resolution images.

Training sessions have been performed based on the WHU datasets for RAFTStereo. A total of 1,222 left and right stereo images along with disparity images at 0.8 meters in the pan band resolution and each measuring 1,024 x 1,024 pixels were shared along with ground truth information for generating disparity maps. No additional pre-processing was done on the WHU dataset, and the configuration was selected to keep spatial resolution while not exceeding memory consumption over the limits of available hardware. Mixed precision training was used to accelerate computation without loss in numerical stability or performance. In this study, the WHU-Stereo satellite benchmark dataset was used for training the RAFTStereo model, and thereafter, the DFC2019 benchmark dataset was used for testing the DSM generation. Since the ground truth DSMs provided in the DFC2019 dataset are not georeferenced, the LiDAR-derived DSMs from the Stellar dataset were employed as the reference ground truth.

All computational experiments and evaluations presented in this study were carried out on a high-performance computing system equipped with an Intel® Xeon® Platinum 8380 CPU running at 2.30 GHz, featuring 160 logical processors (40 cores per socket with 2 threads per core). The system also includes 503 GB of RAM and NVIDIA L4 24 GB GPUs, which provided significant acceleration for deep learning-based stereo matching and large-scale DSM generation tasks (Sayın and Ok, 2025).

4. Results and Discussion

The performance of DSM generation has been evaluated using both visual analysis and numerical measures. Table 1 presents the overall performance of the methods tested. Completeness (COMP) values are expressed in percentage (%), while Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values are given in meters (m).

Data	***SGM***			***MGM***		
	COMP	RMSE	MAE	COMP	RMSE	MAE
DFC JAX 68	97.30	4.10	1.79	95.25	**2.84**	1.63
Data	***RAFTStereo***					
	COMP	RMSE	MAE			
DFC JAX 68	100	4.12	2.78			
Data	***SGM***			***MGM***		
	COMP	RMSE	MAE	COMP	RMSE	MAE
DFC JAX 33	81.25	2.40	1.09	75.00	**2.12**	0.96
Data	***RAFTStereo***					
	COMP	RMSE	MAE			
DFC JAX 33	100	2.49	1.24			

Table 1. Overall performance of the different methods for the generated digital surface models. The best RMSE values are highlighted in bold.

According to the numerical results presented in Table 1, the RMS distance is computed as ≈ 2-4 meters for both the traditional and deep learning methods. Figures 4 and 5 present the right and left stereo images (for JAX_068 and JAX_033 regions), the ground truths and the output DSMs produced by traditional and deep learning methods. The elevation values represent ellipsoidal heights.

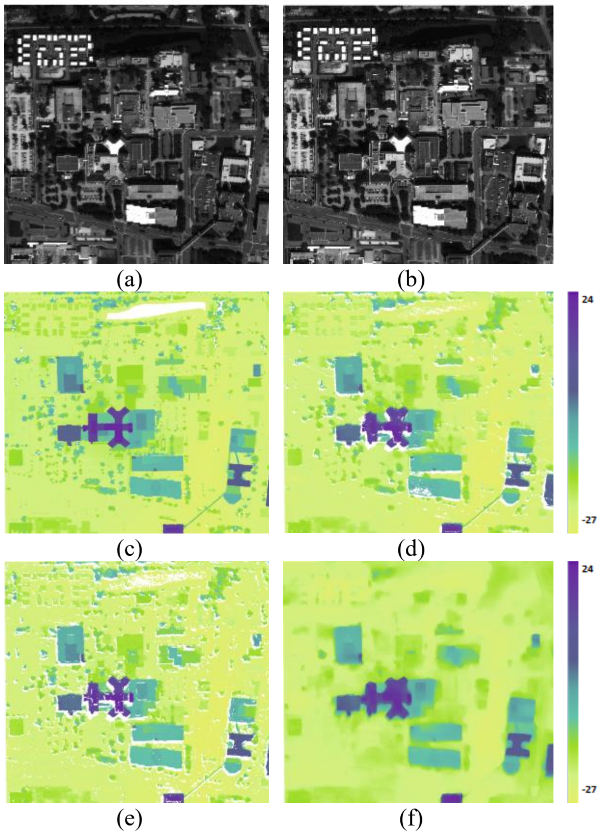
(a) (b)
(c) (d)
(e) (f)

Figure 4. JAX_068 region results. (a) Left stereo image, (b) right stereo image, (c) ground truth (GT) DSM, (d) output DSM using SGM, (e) output DSM using MGM, (f) output DSM using RAFTStereo model.

Figure 4 contrasts the ground truth DSM with SGM, MGM, and RAFTStereo algorithms for an urban area. The ground truth DSM has sharp structural boundaries with well-defined building outlines and thus reflects accurate elevation information. The SGM-based DSM, though retaining the general structural outline, contains significant amount of noise and discontinuities, particularly along building edges and vegetated areas, resulting in a blocky appearance. The DSM of MGM represents relatively smooth and continuous output information compared to SGM, with far more regular transitions in surface elevations and reduced noise, although there is still some minor distortion of fine structural details. The RAFTStereo-based DSM provided the most complete result, demonstrating smoothness of elevation transitions and generally well-preserved large-scale structures. However, due to the model's regularisation, some fine details are smoothed out. Overall, for this region, RAFTStereo seems to be visually and structurally consistent, while MGM represents the best balance between noise removal and sharpness of the structure and outperforms the traditional SGM approach.

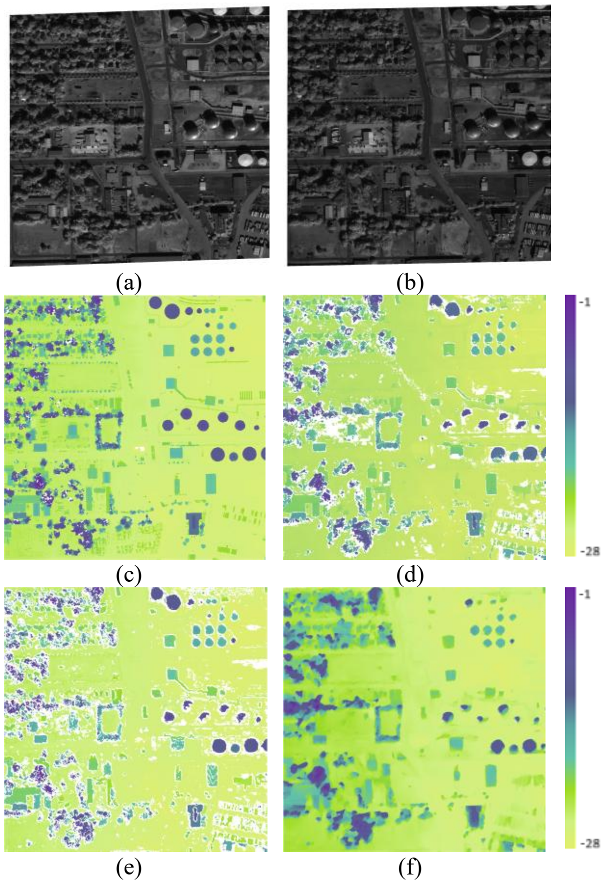

Figure 5. JAX_033 region results. (a) Left stereo image, (b) right stereo image, (c) ground truth (GT) DSM, (d) output DSM using SGM, (e) output DSM using MGM, (f) output DSM using RAFTStereo model.

Figure 5 shows the comparison between the ground truth DSM and the generated DSMs by SGM, MGM, and RAFTStereo for another urban-industrial area with large structures and open spaces. In the case of the ground truth DSM, building outlines are clear, and various circular and rectangular structures are distinct, showing an appropriate representation of the complex surface geometry. The DSM produced by SGM shows the general arrangement but possesses a lot of noise and discontinuities, mainly at edges and shadowed areas, contributing to irregular elevation. The MGM-based DSM did an improvement in avoiding random noises, and surfaces were smoother and more coherent. However, some fine structures remained fragmented. Contrarily, the RAFTStereo-based DSM makes it a more consistent and visually realistic surface reconstruction, making major geometric forms like circular tanks and building roofs sometimes better preserved while minimising noise in open areas. Once again, RAFTStereo yielded the smoothest results; MGM came second, while SGM had lesser stability and preservation of details, especially in heterogeneous surface regions.

In Figures 6 and 7, pixel-based difference maps between the ground truth DSM and the DSMs generated using the SGM, MGM, and RAFTStereo approaches reveal the spatial distribution and magnitude of elevation errors. We intentionally omit the elevation differences larger than 10 meters for better visualisation.

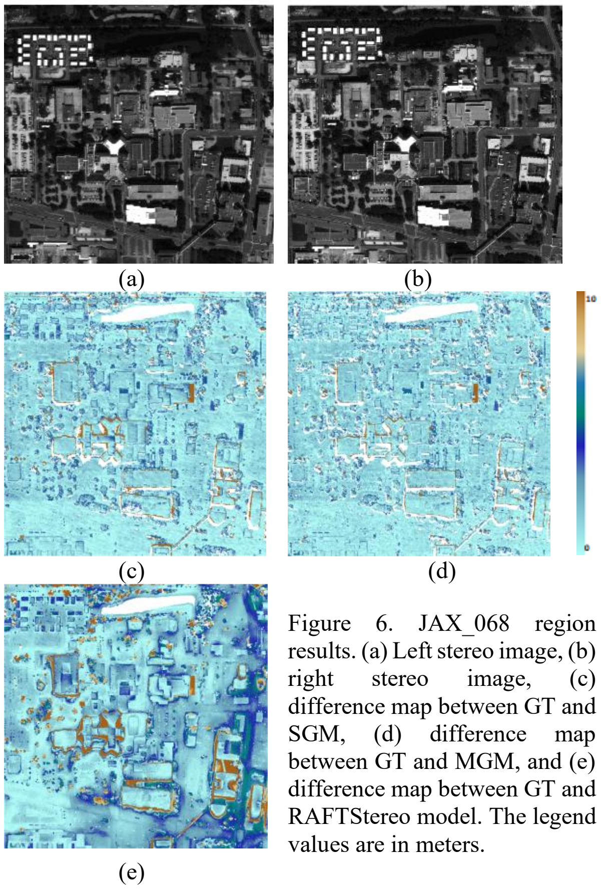

Figure 6. JAX_068 region results. (a) Left stereo image, (b) right stereo image, (c) difference map between GT and SGM, (d) difference map between GT and MGM, and (e) difference map between GT and RAFTStereo model. The legend values are in meters.

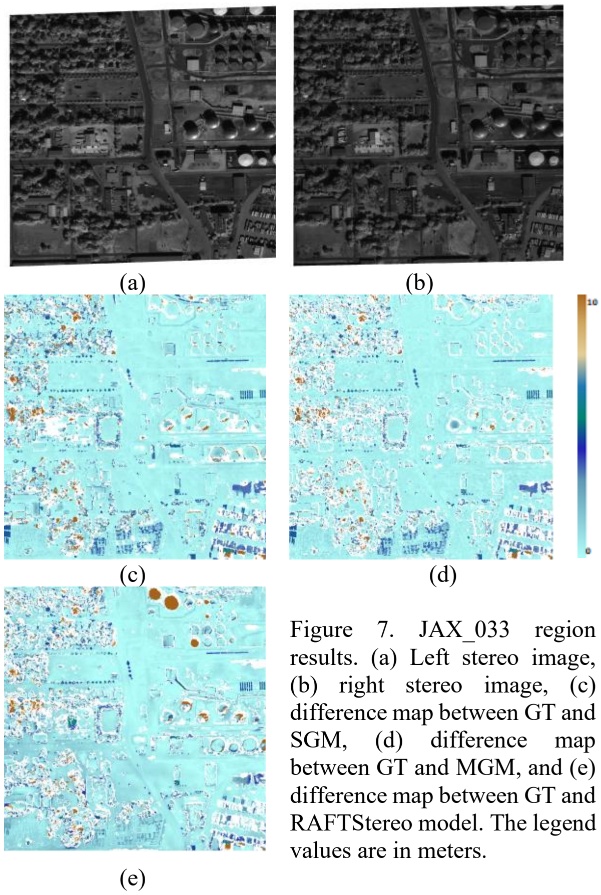

Figure 7. JAX_033 region results. (a) Left stereo image, (b) right stereo image, (c) difference map between GT and SGM, (d) difference map between GT and MGM, and (e) difference map between GT and RAFTStereo model. The legend values are in meters.

The difference maps of the SGM-based DSMs exhibit relatively high residuals, particularly around building edges and circular structures, where orange tones indicate notable height deviations. In addition, random noise and small-scale inconsistencies are visible across flat areas, demonstrating SGM's sensitivity to radiometric and geometric variations. The MGM-based difference maps show an overall improvement with reduced error intensity and a more uniform distribution of elevation differences. The edges of buildings and circular tanks are better defined, reflecting MGM's enhanced multi-directional regularisation and improved surface smoothness. The RAFTStereo-based difference maps demonstrate average magnitudes among the three methods. Most of the area in Figure 7 appears light blue, representing deviations from the ground truth, while only a few localised residuals remain around and in complex structural details and occluded regions. We must emphasise once again that in this study, the WHU-Stereo satellite benchmark dataset was used for training the RAFTStereo model, and thereafter, the DFC2019 benchmark dataset was used for testing the results of DSM generation. Therefore, the RAFTStereo model results are based on the combination of these two high-quality datasets; however, the performance of a model trained on one dataset may not necessarily translate well to another dataset due to differences in data distribution and quality. In this study, the consistency and compatibility of the WHU-Stereo and DFC2019 benchmark datasets provide a solid foundation for validating the effectiveness of the RAFTStereo model in such a stereo matching task. These observations confirm that RAFTStereo achieves interesting results, substantially achieving similar elevation errors compared to traditional SGM and its improved variant, MGM.

5. Conclusion

In this study, the process of generating DSM from high-resolution satellite stereo images using traditional stereo matching methods and a modern deep learning approach is investigated using two benchmark datasets, i.e., the WHU Stereo satellite dataset and Data Fusion Contest 2019.

Traditional stereo matching methods, particularly SGM and MGM, continue to provide reliable performance in structured urban environments and textured terrains, especially under limited computational resources. However, their ability to generalise to challenging scenarios such as homogeneous surfaces, seasonal variations, and significant radiometric inconsistencies remains limited. Recent advancements in deep learning-based stereo matching methods, such as RAFTStereo, have demonstrated notable performance in disparity estimation and DSM accuracy, generalisation and robustness across the tested satellite imagery conditions. Although deep learning methods outperform classical techniques in most conditions, their dependency on extensive training data and their generalisation across different satellite platforms without fine-tuning remains major concerns.

The RMS evaluations and difference map analyses presented in this study revealed complementary strengths and weaknesses among the tested methods. MGM consistently produced the lowest numerical errors, whereas RAFTStereo generated more visually coherent disparity maps with reduced noise and improved surface continuity, especially across homogeneous areas and natural surfaces such as water bodies and vegetation. SGM maintained stable performance in structured regions but showed systematic deviations along object boundaries. These results highlight the importance of jointly considering both numerical accuracy and spatial error distribution in the evaluation of stereo matching algorithms.

In summary, our analysis suggests that while traditional methods remain indispensable for specific conditions, the future of satellite stereo matching strongly leans towards deep learning-based and hybrid approaches, particularly for large-scale, high-precision disparity generation. Drawing upon the findings of this research and recent literature, future investigations should focus on designing memory-efficient deep architectures to efficiently process large satellite images and using hybrid methods. In addition, fine-tuning pre-trained models on satellite-specific datasets can significantly improve their ability to handle the unique characteristics of satellite imagery, including large disparity ranges, pushbroom sensor geometries, and multi-temporal variations. The availability of publicly available datasets remains essential for unbiased evaluation, and careful benchmark design is fundamental to driving progress in satellite-based disparity generation.

Acknowledgements

The presented work was supported by the Secretariat of Defence Industries (SSB) within the scope of the Defence Industry Artificial Intelligence Talent Cluster (SAYZEK) project.

References

Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G.D. and Brown, M., 2019. Semantic stereo for incidental satellite images. *arXiv Preprint*, arXiv:1811.08739. https://doi.org/10.48550/arXiv.1811.08739

De Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-3, 49–56. doi.org/10.5194/isprsannals-II-3-49-2014

Facciolo, G., De Franchis, C., Meinhardt, E., 2015. MGM: A significantly more global matching for stereovision. Proc. Brit. Mach. Vis. Conf. (BMVC), 1–12. doi.org/10.5244/C.29.90

Gómez, A., Randall, G., Facciolo, G., von Gioi, R.G., 2022. An experimental comparison of multi-view stereo approaches on satellite images. Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2937–2946. doi.org/10.1109/WACV51458.2022.00078

Gómez, A., Randall, G., Facciolo, G. and von Gioi, R.G., 2023. Improving the pair selection and the model fusion steps of satellite multi-view stereo pipelines. *HAL Preprint*, hal-04157016. https://doi.org/10.48550/arXiv.2307.06633

Hirschmüller, H., 2007. Stereo processing by semi-global matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2), 328–341. doi.org/10.1109/TPAMI.2007.1166

Hirschmüller, H., Scharstein, D., 2008. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9), 1582–1599. doi.org/10.1109/TPAMI.2008.221

Le Saux, B., Yokoya, N., Hänsch, R. and Brown, M., 2019. 2019 IEEE GRSS Data Fusion Contest: Large-scale semantic 3D reconstruction. *IEEE Geoscience and Remote Sensing Magazine*, 7(4), 33–36. https://doi.org/10.1109/MGRS.2019.2941454

Li, S., He, S.,Jiang, S., Jiang, W., Zhang, L., 2022. WHU-Stereo: A challenging benchmark for stereo matching of high-resolution satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-14. doi.org/10.1109/TGRS.2023.3245205

Lipson, L., Teed, Z., Deng, J., 2021. RAFT-Stereo: Multilevel recurrent field transforms for stereo matching. International Conference on 3D Vision (3DV), London, United Kingdom, 2021, pp. 218-227, doi.org/10.1109/3DV53792.2021.00032

Patil, S. and Guo, Q., 2023. STELLAR: A large satellite stereo dataset for digital surface model generation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-1-2023, pp. 433–442. https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-433-2023

Rothermel, M., Wenzel, K., Fritsch, D. and Haala, N., 2012, December. SURE: Photogrammetric surface reconstruction from imagery. Proceedings LC3D workshop, Berlin 8(2).

Sayın Y.N., Ok A.O., 2025. Advances in Stereo Matching for Disparity Estimation from Satellite Imagery: Traditional Scanline Aggregation Methods versus Deep Learning-Based RAFTStereo. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* (in press)

Teed, Z. and Deng, J., 2020. RAFT: Recurrent all-pairs field transforms for optical flow. *European Conference on Computer Vision (ECCV 2020)*, Cham: Springer International Publishing, pp. 402–419. https://doi.org/10.1007/978-3-030-58536-5_24

Xia, Y., d'Angelo, P., Tian, J. and Reinartz, P., 2020. Dense matching comparison between classical and deep learning based algorithms for remote sensing data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII–B2–2020, 521–525. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-521-2020

Xia, Y., d'Angelo, P., Fraundorfer, F., Tian, J., Reyes, M.F. and Reinartz, P., 2022. GA-Net-Pyramid: An efficient end-to-end network for dense matching. Remote Sensing, 14(8), 1942. https://doi.org/10.3390/rs14081942

Žbontar, J., LeCun, Y., 2016. Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research*, 17(65), 1–32.

Zhang, F., Prisacariu, V., Yang, R. and Torr, P.H.S., 2019. GA-Net: Guided aggregation net for end-to-end stereo matching. *arXiv Preprint*, arXiv:1904.06587. https://doi.org/10.48550/arXiv.1904.06587